UNIVERSITÀ
DEGLI STUDI
FIRENZE

PhD in
Earth and Planetary Sciences

CYCLE - XXXVIII

Regional Risk Assessment of Shallow Landslides: Integrating Triggering and Runout Analysis

Academic Discipline (SSD) GEOS-03/B

Doctoral Candidate
Dr. Brilli Nicolò

Supervisor
Prof. Tofani Veronica

Co-Supervisor
Prof. Gigli Giovanni
Dr. Masi Elena Benedetta

Coordinator
Prof. Conticelli Sandro

2022-2025

Abstract

Landslides represent a major natural hazard, posing severe threats to human life and infrastructure. In this context, this thesis presents SADRI (Scenario-Aware Dynamic Landslide Risk & Impact), an innovative, automated, and event-based computational framework designed to quantify the risk (R) associated with rainfall-induced shallow landslides, potentially evolving into flow-like processes, through a dynamic assessment of triggering, propagation, and impact on exposed elements.

The first step of the procedure concerns the triggering assessment via the HIgh REsolution Slope Stability Simulator – HIRESSS (Rossi et al., 2013) model, which generates Failure Probability (P_f) maps in response to specific rainfall scenarios.

To identify unstable Slope Units, a dual threshold is applied based on P_f and the minimum number of unstable pixels within each unit, known as the Instability Diffusion Threshold (IDT).

Since a static and global P_f/IDT pair inevitably generates errors such as False Positives and False Negatives, the RE-ACT (Refine Alert Classification Thresholds) procedure was developed.

This innovative module implements a data-driven post-processing approach via Explainable AI (XAI), capable of quantifying the contribution of local environmental factors, such as slope and lithology, to model errors, thereby producing an optimized IDT threshold specific to each Slope Unit.

Subsequently, the propagation analysis is restricted to the identified unstable Slope Units and utilizes clusters of contiguous pixels above the P_f threshold as input for the DebrisDice model (Bregoli et al., 2018).

The rheological parameters of the model were defined by calibrating them on known events in back-analysis through a rigorous Bayesian optimization procedure implemented within the SEARCH module, identifying a calibrated parameter set for the study area.

These simulations provide the event intensity (I) as output, expressed as flow velocity (m/s). Vulnerability (V) is then calculated as the degree of expected loss, ranging from 0 to 1, via an S-shaped function evaluating the ratio between intensity and the structural resistance of the asset, while Exposure (E) quantifies the economic value of the building fraction actually intersected by the simulated event.

The framework provides two distinct risk indicators calculated for each individual building in euros ($R = H \times V \times E$).

The first is a Deterministic Indicator, which represents the economic damage value calculated through a single runout simulation using the calibrated rheological parameters defined during the calibration phase.

The second is a Probabilistic Indicator, obtained by assigning a variation range to the calibrated parameter set to quantify uncertainty; the system performs hundreds of Monte Carlo simulations, generating not a single result but a probability distribution of possible economic damages, from which the mean value and standard deviation are calculated.

The developed framework was applied to the Misa River Basin (Marche, Italy), simulating two rainfall scenarios with different characteristics: the event of 2-4 May 2014 and the event of 15-16 September 2022.

The results demonstrate the framework's capability to successfully differentiate between diverse rainfall scenarios, translating physical hazard into tangible economic metrics.

By providing quantitative financial estimates of potential damage rather than generic hazard classes, SADRI proves to be a robust operational tool for dynamic, regional-scale landslide risk assessment, effectively supporting impact-based decision-making for civil protection purposes.

Ringraziamenti

Arrivato alla fine di questo percorso, desidero ringraziare tutte le persone che, in modi diversi, mi sono state vicine e hanno contribuito al raggiungimento di questo importante traguardo.

Un primo e sincero ringraziamento va alla Professoressa Veronica Tofani, mia tutor, per avermi accompagnato durante questi anni di dottorato con grande disponibilità, fiducia e attenzione. Grazie per il supporto costante, per i consigli preziosi e per avermi dato la possibilità di crescere, confrontarmi e mettermi alla prova, sia dal punto di vista scientifico sia personale.

Ringrazio il Professor Giovanni Gigli, mio correlatore, per il contributo dato allo sviluppo di questo lavoro, per i suggerimenti e per il confronto sempre utile e costruttivo. Grazie per la disponibilità e per l'attenzione dimostrata durante questo percorso.

Un ringraziamento particolare va alla Dottoressa Elena Benedetta Masi, mia correlatrice, per l'enorme pazienza, la presenza costante e il grande aiuto dato in questi anni. Grazie per avermi seguito così da vicino, per il tempo, l'attenzione e le energie dedicate, soprattutto nei momenti più impegnativi. Il tuo supporto, sia scientifico sia umano, è stato fondamentale per portare avanti e concludere questa tesi.

Un sentito ringraziamento va al Professor Marcel Hürlimann e a tutto il suo team di ricerca, per l'accoglienza e l'enorme disponibilità ricevuta durante il periodo di ricerca svolto presso l'Universitat Politècnica de Catalunya a Barcellona. Grazie per l'ospitalità e l'aiuto ricevuto durante quei mesi. Questa esperienza è stata molto importante per il mio percorso di dottorato e ha rappresentato un'occasione preziosa di crescita scientifica, professionale e personale.

Il ringraziamento più grande va ai miei genitori, che mi hanno sempre sostenuto, incoraggiato e accompagnato in ogni momento di questo percorso. Senza la loro fiducia e il loro supporto costante, questo traguardo non sarebbe stato possibile.

Ringrazio mia sorella e le mie nipoti, Nicole ed Emily, per l'affetto, la vicinanza e per i momenti di leggerezza regalati anche nei periodi più intensi. Un grazie sincero va anche a tutta la mia famiglia, che mi è sempre stata accanto e mi ha fatto sentire il proprio sostegno.

Ringrazio infine tutti gli amici e i colleghi con cui ho condiviso questi anni: le giornate di lavoro, i confronti, le difficoltà, ma anche le risate e i momenti più belli. Ognuno di loro, a modo suo, ha reso questo percorso più ricco, più leggero e sicuramente più bello.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides the introduction and motivation for the study, presents the problem statement, and outlines the thesis's objectives and structure.

1.1 Context and Current Challenges in Landslide Risk Assessment

Landslides are a major hazard for communities across continents, repeatedly causing fatalities and substantial economic losses (Petley, 2012; Haque et al., 2016; Froude and Petley, 2018; Fidan et al., 2024).

The situation has become more severe due to three concurrent factors operating together: urban expansion and infrastructure encroachment into steep, unstable terrain (Grima et al., 2020; Dille et al., 2022); the growing frequency and intensity of heavy rainfall events (Gariano and Guzzetti, 2016); and land-cover conversions, often supported by inadequate planning, that degrade slope stability and alter catchment hydrology (Karsli et al., 2009; Pacheco Quevedo et al., 2023).

When severe storms strike landscapes already stressed by human modification, the results are often destructive and further amplified for people and assets through cascading effects (Marengo et al., 2020; Johnston et al., 2021; Ozturk et al., 2022).

In Italy, a combination of geological and geomorphological settings, specific rainfall regimes, and dense settlement patterns in exposed areas results in exceptionally high susceptibility and exposure.

The most recent national assessment by the Italian Institute for Environmental Protection and Research (ISPRA) confirms the growing extent of hydrogeological risk areas across the country.

According to ISPRA Report 415/2025 data (Trigila et al., 2025), there is a distinct deterioration compared to the previous reporting cycle.

As of the latest assessment, 94.5% of Italian municipalities include areas designated as being at risk (a 0.6% increase since 2021), whilst the territory classified under moderate-to-very-high hazard classes has expanded by 4.6%, now covering 23% of the nation.

The primary problem stems from the accelerating growth of landslide hazards, which show no indication of abating. As demonstrated in *Figure 1*, high-risk zones have increased by 9.2% relative to 2021 and now occupy 9.5% of the national territory.

The recent expansion has resulted in a substantial upsurge in the exposure of critical assets: approximately 742,000 buildings (+31%) and over 14,000 cultural heritage sites (+12%) are now located in high-risk zones, directly impacting 1.28 million residents. Regarding flood risks, current indicators still use the 2021 assessment, which estimated 6.8 million exposed residents in medium-probability scenarios.

These data are currently under revision to reflect recent changes, with the updated national mapping scheduled for release in 2026 (Trigila et al., 2021, 2025).

The mean direct damages from 2013 to 2020 are estimated at roughly €2.8 billion per year for combined landslide and flood events, a figure that underscores the urgent need for stronger mitigation, preparedness, and response capabilities (Greenpeace Italy, 2024). The key national indicators are summarised in *Table 1*.

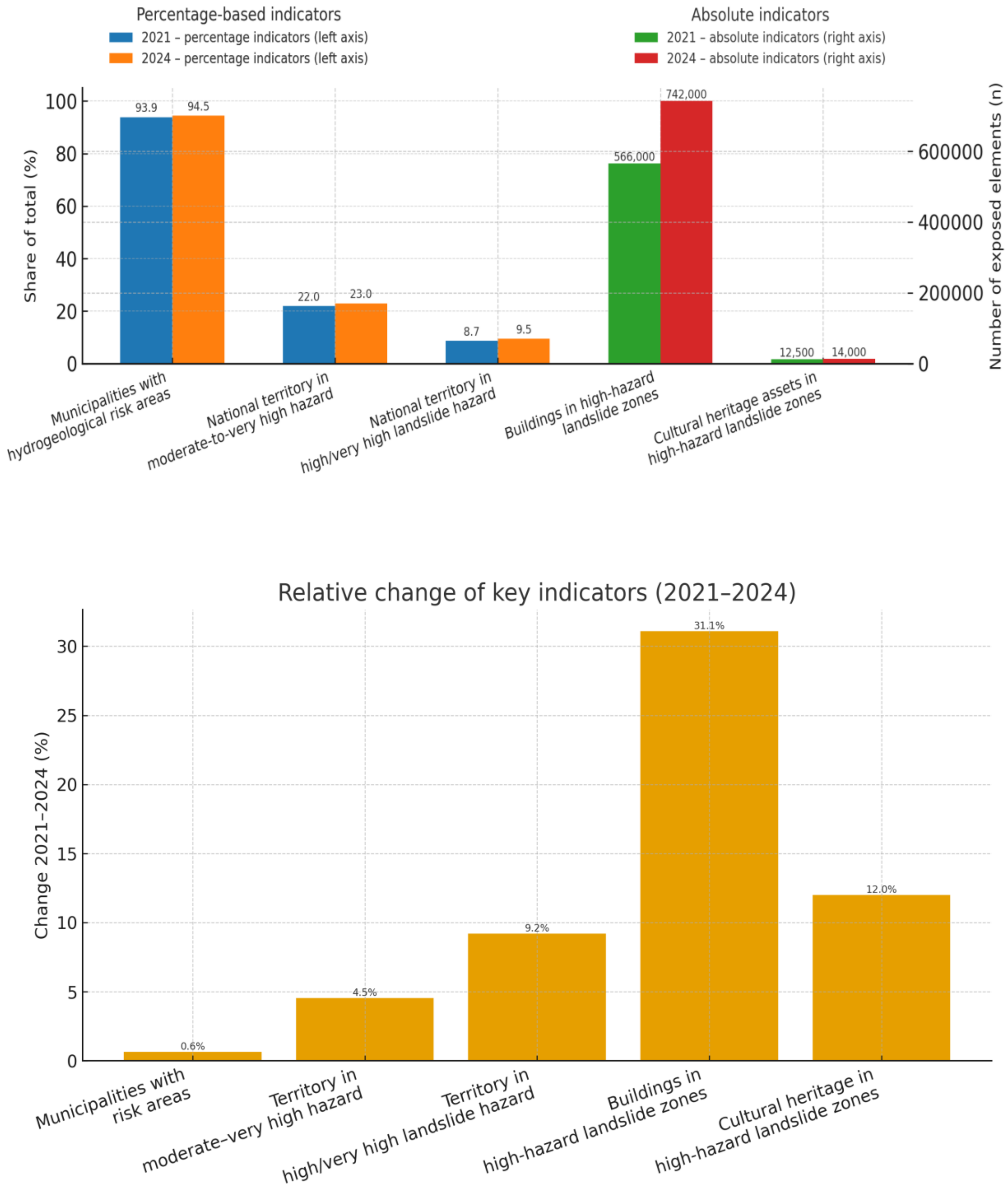


Figure 1. *Comparison of hydrogeological risk indicators in Italy between 2021 and 2024. The top graph shows both percentage and absolute values. The graph at the bottom illustrates the corresponding relative percentage change, thereby quantifying the increase in risk. Source: Elaboration on ISPRA data, Reports 356/2021 and 415/2025 (Trigila et al., 2021, 2025)*

Table 1. *Key hydrogeological risk indicators in Italy. The values shown correspond to the most recent available data*

Risk Indicator	**Landslide Risk**	**Flood Risk (Medium Scenario)**
Area classified in the high/very high landslide risk categories	9.5% of the national territory *	
Area in medium hydraulic hazard class		10% of the national territory
Municipalities with areas at hydrogeological risk (aggregate)	94.5% *	
Population at risk	1.28 million (2.2% of total) *	6.8 million
Households at risk	582,000 (2.2% of total) *	> 2.9 million
Buildings at risk	> 742,000 (5.1% of total) *	> 1.5 million (10.7% of total)
Businesses/services at risk	> 75,000 *	> 640,000
Cultural heritage at risk	> 14,000 *	> 34,000
Average annual direct damage (2013-2020, combined landslides + floods)	€2.8B (Greenpeace Italy, 2024 - data from Regional Authorities and the Italian Civil Protection Department)	

Notes: *Values marked with an asterisk (*) indicate data updated to 2024 (Trigila et al., 2025). All other values correspond to the 2021 assessment (Trigila et al., 2021)*

The challenge is further amplified by climate change. The IPCC Sixth Assessment Report (AR6) (IPCC, 2023) concludes that heavy precipitation events have already increased in intensity and frequency in many northern and central European and Mediterranean land regions. In these areas, wide-ranging instability is often observable because the weather produces short, powerful thunderstorms alternating with extended periods of stratiform rainfall. The non-stationarity of these factors under climate change strengthens the need for an event-specific assessment rather than a static mapping (Gariano and Guzzetti, 2016; IPCC, 2023).

This current pattern requires researchers to develop real-time risk assessment systems that integrate evolving environmental conditions (e.g. rainfall intensity and ground deformation) and predict potential impacts accordingly (Haque et al., 2019; Kendon et al., 2023).

The key operational limitation is that risk maps based mainly on historical data represent persistent hazard zones, but do not account for short-term, event-specific risk conditions.

Landslide risk (R), as originally defined by Varnes (1984), is commonly conceptualised as the combined product of hazard (H), vulnerability (V), and exposure (E). The original formula has received various modifications and additions since its introduction to better handle complex landslide phenomena (Fell et al., 2005, 2008; van Westen et al., 2006; Corominas et al., 2014). Over a domain with multiple elements at risk derived from multiple landslides, total risk (R_{tot}) can be written as the sum of asset-level contributions (van Westen et al., 2006) (*Figure 2*):

$$R_{tot} = \sum [\, H_e (V_e E_e) \,]$$

where H_e *denotes the landslide hazard associated with element e,* V_e *the expected damage ratio conditional on the event intensity at that element, and* E_e *the economic value of the exposed element.*

In the traditional approach, landslide hazard (H) is estimated to determine the location, timing and magnitude of an event (Guzzetti et al., 1999, 2005). This estimation is often expressed as:

$$H = S \times P_t \times (P_A \geq A_l) ,$$

where S indicates landslide susceptibility (the spatial likelihood that a mapping unit is prone to initiation), Pt the temporal probability (frequency/return period/exceedance probability within a given window), and $P_A \geq A_l$ *the magnitude probability (the chance that a landslide exceeds a specified size* A_l*) (Guzzetti et al., 2005).*

Landslide susceptibility can be established through qualitative and semi-quantitative evidence, including landslide inventories, expert geomorphological mapping, and heuristic weighting, or via quantitative models: (a) physically based models (PBMs) that couple a geotechnical stability model (e.g., Mohr-Coulomb, often in an infinite-slope form) with a hydrological module (e.g., Green-Ampt infiltration model (Green and Ampt, 1911) or 1-D Richards equation (Richards, 1931)), as implemented in SHALSTAB (Montgomery and Dietrich, 1994), SINMAP (Pack et al., 1998), TRIGRS (Baum et al., 2008), HIRESSS (Rossi et al., 2013), FSLAM (Medina et al., 2021), and (b) data-driven/ML approaches (GLM/GAM, RF, SVM, ANN, recent DL) that capture non-linear patterns but raise issues of transferability, uncertainty quantification, and interpretability (Breiman, 2001; Vapnik, 1995, 1999; Goetz et al., 2015; Tien Bui et al., 2016; Zêzere et al., 2017; Huang et al., 2017; Huang & Zhao, 2018).

The vulnerability (V) is linked to intensity (I) via a damage function V(I) so that expected loss (L) for an element is $L = V_{(I)} \times E$; in many operational products, however, I is only a coarse intensity class (or even unavailable), and (V) is assigned as fixed, asset-class values or generic tabular curves, i.e., not event-conditioned (Uzielli et al., 2008; Li et al., 2010; Corominas et al., 2014).

Exposure (E) is usually taken as the total value of assets within a mapping unit (e.g., €/m^2 for buildings, €/m for linear infrastructure, or population counts), rather than the value of the actually affected share (Fell et al., 2005; Pellicani et al., 2014).

This approach is useful for planning, but it often relies on relatively static input variables and limited temporal updating, resulting in fixed multi-year maps. These tools are invaluable for long-term land-use planning, as they answer the fundamental question:

"Which areas are most susceptible to slope failure?"

However, these systems have built-in restrictions in actual operational environments. For emergency management and real-time decision-making, event-specific answers are required to address more urgent questions:

"Given this specific rainfall event, where and when landslides may occur, and with what intensity and potential impact?"

Current operational products rarely provide full integration between these two domains. Consequently, they rarely provide the event-conditioned expected loss estimates needed by civil protection agencies for effective prioritisation.

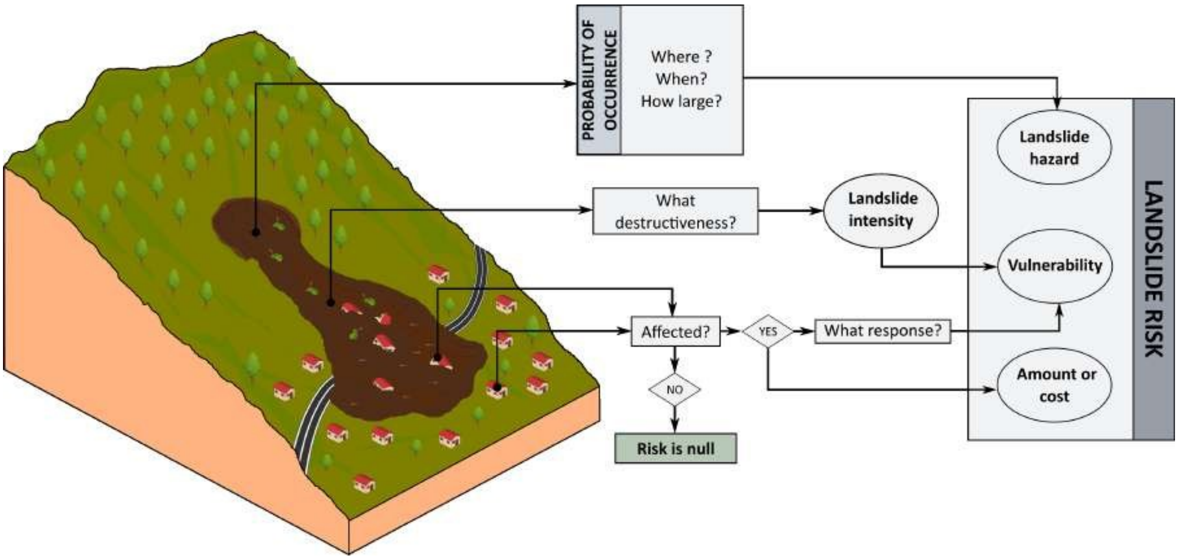


Figure 2. *Conceptual model of landslide risk (from: Caleca, 2024)*

1.2 Research Objectives

This thesis is primarily concerned with bridging the gap between traditional, static landslide risk zonation and the operational necessity of dynamic, event-specific assessments. This research, therefore, aims to design, develop, and validate a novel, fully automated procedure, hereafter referred to as the Scenario-Aware Dynamic Landslide Risk & Impact (SADRI) framework, capable of translating dynamic rainfall scenarios into quantitative, asset-level economic impact maps.

To reach this goal, the study addresses three distinct methodological gaps through the following specific objectives:

1) **Addressing the constraints of global alert thresholds:** Traditional landslide early warning systems rely on global instability thresholds to translate continuous probability maps into discrete alert levels (e.g., Yellow, Orange, Red). These are often defined using an approach that classifies a spatial unit as unstable based on a "failure probability-instability diffusion" pair of thresholds.

   The Failure Probability Threshold (FPT) quantifies the intensity of instability at a pixel, while the Instability Diffusion Threshold (IDT) determines its spatial extent within the spatial unit.

   However, applying a single, static IDT across geomorphologically heterogeneous territories leads to persistent classification errors. It fails to account for local variability in predisposing factors, resulting in False Positives (FP) and False Negatives (FN). This research aims to overcome this limitation by developing the Refine Alert Classification Thresholds (RE-ACT) procedure.

   The module begins with an intermediate operational baseline (e.g. the Orange alert level, which offers the optimal trade-off between sensitivity and specificity), using Explainable AI (XAI), specifically SHAP values, to diagnose the environmental root causes of classification errors (e.g., specific lithology, specific slope or land use patterns).

   The aim is to formulate a data-driven, spatially adaptive correction factor tailored to the geomorphological traits of each Slope Unit (SU). By dynamically correcting the IDT based on local drivers, the system minimises residual errors and enhances alert reliability;

2) **Integrating Source-Area Delineation with Rheological Calibration:** A critical challenge in modelling landslide flow is the subjective nature of delineating source areas when relying on generic historical inventories, which often do not distinguish between initiation and deposition zones.

   This research addresses this fundamental ambiguity by implementing the Source Evaluation and Rheological Calibration Hub (SEARCH), an innovative procedure driven by Hierarchical Bayesian optimisation. The innovation here lies in moving beyond simple parameter tuning; the system conducts a simultaneous, dual-layer search to identify the optimal source area for each landslide polygon. By optimising the source geometry within these shapes, the model rigorously identifies the optimal set of rheological parameters (e.g., friction factor, mass to drag ratio) for the study area.

   By wrapping this process in a k-fold cross-validation scheme, the framework reduces manual bias and supports model robustness and transferability;

3) **Quantifying the Economic Toll of Landslide Hazards:** The final objective is to convert physical model outputs into tangible economic metrics for civil protection, moving beyond the constraints of standard hazard or intensity maps. While these outputs depict physical variables such as flow velocity or debris thickness, they do not explicitly quantify potential economic damage.

   The objective is to operationalise an event-based form of the risk equation ($R = H \times V_{(I)} \times E$) to compute tangible economic consequences. By coupling dynamic hazard scenarios (H) with intensity-dependent vulnerability curves ($V_{(I)}$) and the economic value of exposed assets (E), the system generates event-specific loss scenarios expressed in euros. This shift from static, qualitative assessments to quantitative financial estimates provides decision-makers with precise metrics on prospective losses. This facilitates advanced cost-benefit analyses to evaluate the economic efficiency of different emergency responses and long-term mitigation strategies.

1.3 The proposed SADRI Framework

While the term "landslide" covers a broad spectrum of processes (Hungr et al., 2014), the current study focuses on rainfall-induced shallow landslides, which are triggered by transient increases in pore-water pressure and a corresponding reduction of effective stress during infiltration.
They typically involve failure depths ranging from decimetres to a few metres, commonly between 0.3 and 2 m (Campbell, 1975; Ellen et al., 1988; Crosta, 1998), and may rapidly evolve into flow-like processes, such as debris flows or mudflows, when the failed material mobilises downslope.
These events are of particular concern due to their sudden onset and lack of precursor signals, making their spatial and temporal prediction challenging.
Their capacity to propagate along gullies, thereby achieving elevated velocities and destructive impact pressures, underscores the necessity to develop operational methodologies. In this context, the thesis develops and validates SADRI, a reproducible workflow designed to translate dynamic rainfall scenarios into asset-level impact maps.
Within this framework, rainfall-induced shallow failures are treated as potential source areas for subsequent runout modelling. SADRI is therefore intended for rainfall-induced shallow landslides potentially evolving into debris-flow-like processes.
It is not designed to simulate deep-seated landslides, rockfalls, rock avalanches, structurally controlled bedrock instabilities, or complex landslide mechanisms unless additional process-specific models and calibration strategies are introduced.

For this class of processes, SADRI explicitly links the main stages of the hazard-to-impact modelling chain:

triggering → propagation (intensity evaluation) → impact assessment (exposure and vulnerability).

The framework reinterprets the standard risk equation ($R = H \times V_{(I)} \times E$) at the asset level. In SADRI, the three components are not treated as static map layers, but are evaluated dynamically according to the simulated rainfall scenario, the resulting runout intensity, and the actual portion of exposed assets intersected by the modelled footprint. The components are defined as follows:

- **Dynamic Hazard (H):** This component is shifted from a conventional *probability of occurrence* to an event-conditioned representation of the landslide process. Hazard is therefore described by the runout footprint generated by the coupled triggering–propagation chain, together with its associated intensity field.
  The rainfall-driven triggering model identifies where instability may initiate, while the calibrated runout model simulates the downslope propagation of the failed material.
  The resulting output provides the simulated peak flow velocity at each asset location, expressed in m/s, which is then used as the intensity parameter controlling the vulnerability function V(I);

- **Vulnerability (V):** Vulnerability provides the link between the physical intensity of the simulated event and the expected degree of damage to the exposed asset.
  In SADRI, vulnerability is not assigned as a fixed value for an asset class, but is calculated through an intensity-dependent damage function. Specifically, vulnerability is modelled using a S-shaped function modified from Li et al. (2010), which returns a damage ratio ranging from 0 to 1.
  This ratio is derived from the relationship between the simulated intensity, expressed as flow velocity (I), and the structural resistance of the asset (R_struct).
  In this way, assets exposed to the same runout footprint may experience different levels of expected damage depending on both the local impact velocity and their structural resistance;

- **Exposure (E):** Exposure represents the monetary value of the asset portion effectively located within the dynamic hazard footprint. For buildings, SADRI does not automatically assign the full asset value to the risk calculation. Instead, the simulated runout field is spatially intersected with the building geometry to identify the fraction of each asset actually affected by the event.

The affected surface is then multiplied by the corresponding unit economic value (€/m²), allowing the framework to quantify the exposed value directly involved in the simulated impact.

The resulting asset-level risk therefore represents an event-conditioned economic loss, expressed in euros, obtained by combining the simulated hazard intensity, the corresponding vulnerability value, and the monetary exposure of the affected building fraction.

The framework operates through a sequential workflow. The initial step involves the assessment of potential triggering using the High-Resolution Slope Stability Simulator model (HIRESSS; Rossi et al., 2013), which yields a spatial field of per-pixel Failure Probability (P_f) in response to specific rainfall scenarios.

To enable territorial alert assessment, the study area is partitioned into Slope Units (SUs), which are the basic decision-making spatial units.

These SUs represent geomorphologically significant divisions and were automatically delineated using the segmentation methodology developed by Alvioli et al. (2016).

The identification of unstable SUs is achieved through a dual-criterion approach that integrates the global pixel-level Failure Probability Threshold (FPT) with the spatial extent of instability (defined as Instability Diffusion Threshold, IDT).

The optimal threshold pair was defined by evaluating a comprehensive set of performance metrics, including the True Skill Statistic (TSS) and the Area Under the Receiver Operating Characteristic Curve (AUROC).

To ensure local accuracy, the framework incorporates the Refine Alert Classification Thresholds (RE-ACT) procedure, which provides, as previously indicated, an optimised IDT threshold specific to each SU, quantifying the contribution of local environmental variables to model errors.

Once the unstable Slope Units (SU) have been defined, the procedure moves on to identifying the source areas. The analysis is conducted exclusively on these SUs, identifying contiguous clusters of pixels that exceed the Failure Probability Threshold (FPT).

These clusters, defined by selectable minimum and maximum sizes, constitute the source areas and serve as the geometric input for the subsequent execution of the runout model DebrisDice (Bregoli et al., 2018).

To best simulate the flow propagation, the optimal set of rheological parameters for the study area was identified via back-analysis using the Source Evaluation and Rheological Calibration Hub (SEARCH) module, employing Bayesian optimisation with k-fold cross-validation.

The runout simulations provide the event intensity field, expressed as peak flow velocity (m/s). This output is then combined with the vulnerability and exposure components defined above to derive event-specific economic risk at the single-asset level.

Specifically, the framework provides two distinct economic risk indicators to quantify the damage in euros: firstly, a deterministic indicator, which is the value of the economic damage calculated through a single runout simulation using the optimal rheological parameters defined during the calibration phase; and secondly, a probabilistic indicator, representing the value obtained by assigning a plausible range of variation to the set of optimal parameters to quantify uncertainty.

For the latter, the system performs hundreds of Monte Carlo simulations (MC) (Metropolis and Ulam, 1949), generating a probability distribution of possible economic damages from which the mean value and standard deviation are calculated.

Finally, the SADRI framework concludes with the outcomes classification and report generation. The framework categorises the risk status of each SU (e.g., Direct Impact, Uncertainty) and automatically compiles the final deliverables, producing both geospatial maps and detailed technical PDF reports.

To demonstrate the operational applicability and robustness of the SADRI framework, the procedure was subjected to a rigorous validation strategy within a complex environmental setting.

The evaluation focused on assessing the system's sensitivity, computational efficiency, and reliability by simulating two contrasting rainfall scenarios:

- ✓ *Scenario A (Procedure Calibration Event):* The extreme convective event of 15-16 September 2022. Characterised by high-intensity, short-duration precipitation, this event served as the benchmark for the calibration phase. Supported by a detailed landslide inventory, it was utilised to optimise the rheological parameters and fine-tune the alert thresholds;

- ✓ *Scenario B (Procedure Test Event):* The prolonged stratiform event of 2-4 May 2014, marked by moderate but widespread precipitation on saturated soils. This scenario served as a test to verify the system's transferability and predictive capability on an independent dataset.

Through this comparative analysis, the thesis seeks to validate a methodology capable of supporting decision-making processes during civil protection emergencies, effectively bridging the gap between scientific modelling and practical risk management.

1.4 Thesis Structure

The thesis is organised into seven distinct chapters, each addressing a key component of the research:

- ❖ **Chapter 1:** Introduces the research motivation, defines the problem statement, and outlines the main contributions of the thesis;
- ❖ **Chapter 2:** Provides a critical review of the current state of the art in landslide modelling and risk assessment, covering PBMs, runout models, ML/PIML, and uncertainty;
- ❖ **Chapter 3:** Presents the proposed methodological framework (SADRI), including its architecture, data inputs, modules, and validation process;
- ❖ **Chapter 4:** Describes the study area and the rainfall scenarios simulated for the analysis;
- ❖ **Chapter 5:** Presents the applications and results, focusing on triggering, propagation, calibration outcomes, and dynamic risk mapping;
- ❖ **Chapter 6:** Discusses the implications of the results, addressing sensitivity, uncertainty, limitations, and the framework's transferability;
- ❖ **Chapter 7:** Concludes the thesis by summarising the key findings and proposing directions for future work.

Chapter 2

State of the art

This chapter presents a critical review of the state of the art in landslide modelling and risk assessment, analysing physically based triggering and runout models (PBMs), integration with machine learning approaches (ML), quantitative risk assessment (QRA), and uncertainty quantification (UQ) and management.

2.1 Introduction to Landslide Risk Assessment

Natural-hazard risk arises from the interaction between physical processes and exposed human, infrastructural, and economic systems.
The concept was progressively formalised during the first half of the 1980s, leading to the development of quantitative frameworks for risk assessment.
Varnes (1984) formalised landslide risk as the expected losses affecting people, property, infrastructure, and activities within a given area and time period.
Subsequent studies consolidated the three-component formulation of risk based on Hazard (H), Vulnerability (V) and Exposure (E) (Fell, 1994; Dai et al., 2002; Fell et al., 2005, 2008; van Westen et al., 2006; Corominas et al., 2014).
The international framework for disaster risk management identifies H-V-E factors and societal readiness as essential elements which determine risk development so non-structural measures serve as vital tools for reducing disaster impacts (UNISDR, 2009; UNDRR, 2019).
An early quantitative formulation of landslide risk assessment focused on a single element at risk, such as an individual structure, and expressed the expected damage through the following risk equation:

$$R = P(M_i) \cdot P(X_j|M_i) \cdot P(T|X_j) \cdot V_{ij} \cdot C \qquad [2.1]$$

This expression disaggregates total risk (R) into its fundamental components, thus providing the analytical basis for the simplified conceptual model $R = H \times V \times E$:

- **Evaluation of the Hazard (H)** factor requires an assessment of the probabilities inherent to the event itself, encompassing both its size and reach: $H \approx P(M_i) \cdot P(X_j|M_i)$. Here, $P(M_i)$ (Magnitude Probability) quantifies the likelihood of a landslide occurring with a specific size (M_i), while $P(X_j|M_i)$ (Spatial Reach and Intensity Probability) represents the conditional probability that the event reaches a specific location (X) and impacts it with a defined intensity (j);

- **Vulnerability (V)** is perhaps the most direct measure, represented by the term Vij: $V \equiv Vij$. This Specific Vulnerability is the expected degree of loss (a proportion between 0 and 1) that the element will suffer from an impact of magnitude i and intensity j.

- **Conversely, Exposure (E)** serves to quantify potential asset value and presence, defined as ≈ $P(T|X_j)\cdot C$. This incorporates $P(T|X_j)$ (Temporal Presence Probability), which is the conditional probability that the element is present at point X at the time T of the event's occurrence, multiplied by C (Value of the Element at Risk), the monetary or replacement cost of the asset;

While the detailed formula [2.1] is essential for single-site quantitative analysis, the focus shifts to aggregated total risk when assessing entire communities or large areas encompassing multiple exposed elements (van Westen et al. 2006):

$$\mathrm{R} = \sum_{j=1}^{n} \mathrm{Hj} \times \mathrm{Vj} \times \mathrm{Ej} \tag{2.2}$$

The summation process which uses the total number of exposed elements or units (n) connects the detailed risk assessment at one location to the complete risk assessment of the entire region.

Risk assessments for land-use planning have used static (multi-annual) hazard classes together with average exposure and generic vulnerability since the past which serve as useful tools for regulatory purposes and long-term mitigation efforts. The process of making operational decisions during storm evolution and crisis situations needs real-time estimates which depend on specific events and conditions to determine both time and location of initiation and mass movement paths and asset-intensity levels and expected loss amounts.

Current operational standards increasingly rely on the coupling of process-based models with probabilistic treatment of uncertainty (van Westen et al., 2006; Corominas et al., 2014).

A key theoretical consideration is the interdependence of these components. Although analytically factored as separate terms, intensity (I, part of H) and structural response (central to V) co-determine the effective forcing, creating an intrinsic interdependence that must be rigorously maintained in detailed, asset-level analyses (Corominas et al., 2014). Quantitative Risk Assessment (QRA) typically utilises two primary metrics for communication:

- ❖ Scenario Risk (Scenario Loss):
  Estimates the consequences derived from a single event of well-defined magnitude, providing the basis for designing defence works and emergency planning (AGS, 2007);

- ❖ Expected Annual Loss (EAL):
  Represents the long-term average financial risk, expressed in annualised monetary terms, and calculated by integrating estimated losses over the full frequency-magnitude spectrum (Hungr et al., 2008). The EAL is the standard metric for cost-benefit analysis, financial risk management, and insurance purposes (FEMA, 2023; UNDRR, 2015).

2.2 From Hazard to Risk Impact Assessment

2.2.1 Hazard (H)

Characterising landslide hazard (H) fully requires an examination of how its core pillars, spatial and temporal probability, magnitude, and intensity, interact. The subsequent analysis explores these individual factors, starting with the spatial probability of occurrence.

2.2.1.1 Landslide Susceptibility

Landslide susceptibility is defined as the spatial probability of terrain conditions conducive to landslide initiation within mapping units. This measure indicates where landslides are more likely to occur but represents only one component of hazard assessment (Brabb, 1984; Guzzetti et al., 2005). Most research methods rely on the assumption that present and past controlling factors remain informative for anticipating future conditions (Varnes, 1984; Carrara et al., 1991; Corominas et al., 2014).

The terms “susceptibility” and “hazard” have frequently been confused in small-scale planning contexts because susceptibility is often adopted directly as a planning tool (Carrara, 1983; Reichenbach et al., 2018; Corominas et al., 2023).

Landslide susceptibility methods can be grouped into qualitative, semi-quantitative and quantitative approaches (Corominas et al., 2014; Reichenbach et al., 2018).

Qualitative and semi-quantitative methods:

- Inventory-based mapping: This empirical qualitative method employs inventory-based maps, including landslide density and, when available, information on landslide activity or state, to extrapolate future spatial patterns from observed landslide distributions (Parise and Wasowski, 1999; Coe et al., 2000; Reid and Page, 2003);

- Expert geomorphological mapping: This knowledge-driven qualitative approach relies on the investigator's interpretation of slope instability, geomorphological evidence and geological settings, rendering the final outputs inherently dependent on expert judgement (van Westen et al., 2000; Reichenbach et al., 2005);

- Heuristic weighting and ranking: This knowledge-driven semi-quantitative approach translates expert knowledge and descriptive observations into numerical weights or hierarchical ranks assigned to causal factors (Ruff and Czurda, 2008; Stanley and Kirschbaum, 2017). The assessment process therefore depends on the selected weighting criteria, expert judgement, and the completeness and quality of the available inventory data.

Quantitative methods:

- Physically based models (PBMs): The first quantitative category includes physically based models (PBMs) which function as “white-box” systems that derive stability through the direct implementation of explicit physical principles. The geotechnical core performs shallow landslide analysis through an infinite-slope Mohr-Coulomb limit-equilibrium formulation.

  This formulation works in conjunction with a hydrological module to transform precipitation data into time-varying pore-pressure and saturation fields.

  These coupled components produce a temporally variable Factor of Safety FS(x,y,t) or Failure Probability (*Pf*) (Montgomery and Dietrich, 1994; Pack et al., 1998; Rigon et al., 2006; Baum et al., 2008; Rossi et al., 2013; Raia et al., 2014; Medina et al., 2021). Hydrological differentiation remains a key distinction between models: whereas steady-state formulations (e.g., SHALSTAB (Montgomery and Dietrich, 1994), SINMAP (Pack et al., 1998)) provide fast screening results but they do not resolve storm-scale temporal dynamics, transient models instead compute pore-pressure evolution through Richards-equation solutions (e.g., TRIGRS (Baum et al., 2008), HIRESSS (Rossi et al., 2013), GEOtop-FS (Rigon et al., 2006)).

  These systems can support near-real-time prediction but they require substantially higher computational resources to function.

  The 3-D limit-equilibrium solvers use DEM-based sampling to evaluate rotational slip surfaces which extend beyond planar failure mechanisms, as implemented in Scoops3D (Reid et al., 2015) and R.ROTSTAB (Mergili et al., 2014).

  The STEP-TRAMM model from Lehmann and Or (2012) represents an alternative approach which adopts a fibre-bundle progressive failure paradigm. The basic method of Monte Carlo sampling enables users to propagate parameter ranges into failure probability estimates when dealing with uncertain conditions (Metropolis and Ulam, 1949; Baum et al., 2008; Rossi et al., 2013);

- Data-driven models: The data-driven approach identifies statistical relationships between landslide occurrence and conditioning factors, thereby supporting probabilistic estimates of where landslides may occur. In practice, the model learns from past landslide inventories by analysing how geomorphological and environmental factors are distributed in areas affected by landslides, and then uses these relationships to predict susceptibility in other mapping units.
  The analysis includes model families ranging from linear GLMs (Atkinson and Massari, 1998) to non-linear GAMs (Brenning, 2008; Petschko et al., 2014), Random Forest (Catani et al., 2013), SVMs (Vapnik, 1995, 1999), and deep learning architectures (Bui et al., 2020; Azarafza et al., 2021).
  These models offer high predictive performance, but their operational reliability is frequently constrained by the accuracy, completeness, and representativeness of the landslide inventory data (Zêzere et al., 2017). In addition, estimates may be biased by spatial autocorrelation effects associated with specific sampling and data-collection strategies (Steger et al., 2016, 2017; Gaidzik and Ramírez-Herrera, 2021).

2.2.1.2 Temporal Probability: Frequency and Magnitude

The temporal occurrence of landslides can be described through three elements which include frequency (i.e., the average number of events within a specified interval), return period ($T = 1/\lambda$), and exceedance probability over a specified time window Δt. The exceedance probability method is often preferred in quantitative probabilistic hazard assessments. The four main categories to determine temporal occurrence include:

- Heuristic approaches: probabilities are assigned by experts based on causal factors and geomorphological setting;

- Indirect approaches: the return period of the trigger (e.g., rainfall, earthquakes) is assumed to represent that of the induced landslides; these approaches estimate "how often" events occur rather than "which slopes" will fail;

- Probabilistic models: event rates are inferred from historical records and translated into exceedance probabilities (e.g., binomial/Poisson or power-law formulations); a key limitation is the often-unrealistic assumption of independence and stationarity;

- Physically based models (PBMs): these models simulate the hydro-geotechnical response through time (e.g., FS(t), pore-water pressure, saturation), enabling both forecasting and hindcasting within a fully spatio-temporal modelling framework. They require large amounts of data and substantial computational resources and are therefore typically applied to individual slopes or small catchments over short time windows.

Recent trends emphasise spatio-temporal data analysis techniques to enhance susceptibility modelling through the integration of dynamic covariates, including antecedent rainfall and seasonal patterns, in order to capture time-dependent and location-based shifts in event distributions.

Their main limitation remains the sparse availability of landslide inventories with accurate event dating, which constrains the ability to fully understand how the timing and conditions of triggers influence landslide behaviour.

The timing of events introduces uncertainty which affects magnitude estimation because landslide dimensions expressed by area or volume depend on the conditions under which the event occurs.

The environmental state that prevails at the time of triggering, including soil saturation or antecedent rainfall intensity, influences both slope stability and the mobilisation of material.

Thus, accurate temporal occurrence characterisation is essential because this supports the prediction of magnitude distributions and improves the understanding of initiation and propagation processes (Guzzetti et al., 2005; Corominas et al., 2014; Catani et al., 2016).

The relationship between landslide magnitude and frequency patterns often shows similarities to the Gutenberg–Richter law used to describe earthquake scaling behaviour (Gutenberg and Richter, 1944; Guzzetti et al., 2002).

The power-law behaviour of non-cumulative representations typically shows three distinct features which include a power-law tail for medium-to-large events, a scaling transition point (cutoff), and a rollover at small sizes, frequently associated with inventory incompleteness or physiographic constraints (Stark and Hovius, 2001; Guthrie and Evans, 2004; Corominas and Moya, 2008; Tanyaş et al., 2018).

Exceedance probabilities P_{Al} can be estimated by fitting appropriate probability density functions (Malamud et al., 2004; Guzzetti et al., 2005; Catani et al., 2016), although these calculations are mostly performed at study-area scale rather than for individual mapping units.

The two modelling approaches enable landslide size prediction through PBMs which require information on initial volumes, material rheology, and forcing factors (Burton and Bathurst, 1998; Alvioli et al., 2014; Milledge et al., 2014; Domènech et al., 2019; van den Bout et al., 2022), while data-driven models statistically regress observed landslide areas against conditioning and trigger variables (Lombardo et al., 2021; Aguilera et al., 2022; Moreno et al., 2023).

The relationship between when events occur and how large landslides become forms an essential connection in hazard modelling because temporal dynamics control triggering conditions that strongly influence landslide size and evolution and integrated spatio-temporal approaches are therefore necessary for reliable hazard and risk assessments.

2.2.1.3 Landslide Intensity

Landslide intensity, which is defined as the destructive force of a landslide (Hungr, 1997), is a critical yet complex component in quantitative risk assessment.

It functions as the pivotal conduit between the tangible event and the elements that are susceptible to risk, as it serves as the principal input for determining vulnerability (Uzielli et al., 2008).

Intensity, in contrast to magnitude (size) or susceptibility (location), is not a static property; rather, it is a dynamic measure that can evolve as the event progresses along its trajectory (Corominas et al., 2014).

This dynamic nature of landslides necessitates the use of parameters that are contingent on the specific type of landslide under analysis (Hungr, 1997; Corominas and Moya, 2008; Corominas et al., 2014).

Furthermore, factors such as soil properties and volume play a pivotal role in the analysis (Tofani et al., 2017).

For events of a rapid nature, the preferred metric varies. In the case of rockfalls, the preferred metric is kinetic energy (Crosta and Agliardi, 2003; Jaboyedoff et al., 2005; Moos et al., 2017).

In the case of debris flows and shallow landslides, the intensity of such events is often modelled using various parameters, including peak discharge, flow depth, and velocity (Raetzo et al., 2002; Zanchetta et al., 2004; Hürlimann et al., 2008).

Conversely, the predominant indicator for slow-moving landslides is displacement, whether total or differential. The measurement of this phenomenon can be conducted through the utilisation of in-situ instruments (Gili et al., 2000; Stark and Choi, 2008), or, with increasing frequency, via remote sensing techniques (Tofani et al., 2014; Casagli et al., 2023).

These techniques encompass ground-based (Carlà et al., 2019) and satellite InSAR (Raspini et al., 2018; Solari et al., 2020). A strong correlation exists between the magnitude (size/volume) of a landslide and its resulting intensity, with larger sizes generally producing greater destructive potential (Lari et al., 2014).

It is evident that, in consideration of the aforementioned relationship, local-scale numerical models have the capacity to accurately simulate intensity parameters (Agliardi and Crosta, 2003; Hungr and McDougall, 2009; Pastor et al., 2015).

However, on complex slopes, intensity is represented by a distribution of values (Lari et al., 2014), and the use of simple statistics to summarise it can introduce approximations (Agliardi et al., 2009; Calvo and Savi, 2009; Lambert et al., 2012; Corominas et al., 2014).

In broader analyses, physical modelling is not feasible, and data-driven approaches are used instead. These approaches define intensity as the "number of events per mapping unit" (Di Napoli et al., 2023; Yadav et al., 2023). This method, however, approximates destructive power by substituting occurrence frequency for the actual physical dynamics of the event.

The assessment of intensity remains a significant challenge, primarily due to its highly site-specific nature and the numerous parameters involved.

2.2.1.4 Physically Based Modelling: Triggering and Runout

Physically based models (PBMs) for rainfall-triggered landslides combine a geotechnical description of slope stability with hydrological components that convert precipitation into time-varying pore-water.

The "white-box" nature of these models allows for the generation of physically interpretable outputs such as the Factor of Safety FS(t) or Failure Probability (P_f) by resolving the Limit Equilibrium Method (LEM) acting on potential failure surfaces, as they utilised complex stated equations and parameter sets.

The geotechnical formulations adopted by most triggering PBMs rely on an infinite-slope, Mohr–Coulomb limit-equilibrium approach to analyse planar and shallow failure mechanisms.

This formulation represents the core of TRIGRS (Baum et al., 2008), HIRESSS (Rossi et al., 2013), SHALSTAB (Montgomery and Dietrich, 1994), SINMAP (Pack et al., 1998), FSLAM (Medina et al., 2021), SHIA_Landslide (Aristizábal et al., 2016), GEOtop-FS (Rigon et al., 2006), tRIBS-VEGGIE (Ivanov et al., 2008), and SLIP (Montrasio and Valentino, 2008). Scoops3D (Reid et al., 2015) and R.ROTSTAB/r.slope.stability (Mergili et al., 2014) employ 3-D method-of-columns analyses to investigate failure geometry when instability is curved or deep-seated. These approaches evaluate FS by sampling rotational or ellipsoidal slip surfaces through the DEM and by using externally supplied groundwater conditions or pressure-head fields as hydraulic inputs. Hydrologically, models are differentiated by how they represent subsurface flow and infiltration processes. Steady-state formulations are computationally efficient and well suited to static susceptibility screening, although they lack the ability to resolve event timing and temporal dynamics. SHALSTAB (Montgomery and Dietrich, 1994) links stability to a steady shallow subsurface flow balance and identifies instability where the rainfall-to-transmissivity ratio q/T exceeds a threshold, whereas SINMAP (Pack et al., 1998) adopts a comparable steady-state framework but propagates parameter uncertainty to a Stability Index through the sampling of admissible parameter ranges.

The analysis of storm-related pore-pressure evolution requires transient formulations, which allow the investigation of scenario-based conditions and, where appropriate, real-time applications.

TRIGRS (Baum et al., 2008) couples the infinite-slope LEM with an analytical, linearised 1-D Richards solution for vertical infiltration under variably saturated conditions; its probabilistic extension, TRIGRS-P, applies Monte Carlo sampling to derive spatial patterns of failure probability.

HIRESSS (Rossi et al., 2013) likewise employs an infinite-slope LEM driven by an analytical approximation to transient Richards' equation under wet conditions and includes a native Monte Carlo engine for uncertainty quantification at high spatial resolution and large spatial extents.

The fully distributed implementation of GEOtop-FS derives pore-pressure and moisture states from GEOtop's numerical Richards' solver (Rigon et al., 2006) and can optionally include a coupled water–energy balance. This configuration enables the representation of snow and freeze–thaw feedbacks, thereby increasing physical realism at the expense of higher computational demand.

The ecohydrological tRIBS-VEGGIE framework (Ivanov et al., 2008) couples a distributed Richards-type hydrologic core with dynamic vegetation processes, including evapotranspiration, leaf area index, and phenology, in order to investigate the effects of land use and canopy dynamics on slope stability.

The SLIP model (Montrasio and Valentino, 2008) adopts a compact event-scale infiltration representation to simulate rapid "soil-slip" events that are typically dominated by short and intense storms, and it accounts for cohesion decay with saturation to compute time-evolving FS efficiently from laboratory to territorial scales.

Beyond classical LEMs, the STEP-TRAMM model developed by Lehmann and Or (2012) adopts a fibre-bundle/hexagonal-column paradigm to simulate progressive failure cascades through load redistribution mechanisms controlled by pore-pressure-dependent bond strength.

This framework enables the investigation of pre-failure damage accumulation and progressive weakening processes that are not explicitly represented by standard LEM-based formulations.

Model selection depends on three main factors, including the dominant failure mechanism, the availability of input data, and the operational constraints of the application. Steady-state tools provide appropriate methods for reconnaissance-level susceptibility assessment, transient Richards-based solvers enable event-conditioned forecasting through explicit handling of antecedent moisture conditions, and 3-D column methods address deep-seated rotational failures when suitable hydraulic inputs are available (Brilli et al., 2025). A review of prominent models in the literature is presented in *Table 2*.

Table 2. *Overview of Physically Based Triggering Models (PBMs)*

Model	**Geotechnical module**	**Hydrologic module**	**Outputs**	**References**
SHALSTAB	Infinite slope (Mohr-Coulomb)	Steady-state lateral flow (no Richards' eq.)	FS	Montgomery and Dietrich, 1994
SINMAP	Infinite slope (Mohr-Coulomb)	Steady-state lateral flow (no Richards' eq.)	FS / P_f	Pack et al., 1998
TRIGRS / P	Infinite slope (Mohr-Coulomb)	Transient infiltration (linearised Richards' eq.)	FS / P_f	Baum et al., 2008
HIRESSS	Infinite slope (Mohr-Coulomb)	Transient infiltration + lateral flow (approx. Richards' eq.)	FS / P_f	Rossi et al., 2013
GEOtop-FS	Infinite slope (Mohr-Coulomb)	Transient infiltration + lateral flow (numerical Richards' eq.)	FS / P_f	Rigon et al., 2006
tRIBS-VEGGIE	Infinite slope (Mohr-Coulomb)	Transient infiltration + lateral flow (numerical Richards' eq.) + dynamic vegetation	FS	Ivanov et al., 2008
SHIA-Landslide	Infinite slope (Mohr-Coulomb)	Transient flow (vertical + lateral) (conceptual, no Rich. eq.)	FS	Aristizábal et al., 2016
SLIP	Infinite slope (Mohr-Coulomb)	Transient infiltration (no lateral flow, Rich. eq.)	FS	Montrasio and Valentino, 2008
FSLAM	Infinite slope (Mohr-Coulomb)	Transient vertical flow + Steady-state lateral flow (no Richards' eq.)	FS / P_f	Medina et al., 2021
STEP-TRAMM	Soil-columns (fibre-bundle)	Transient infiltration (no Richards' eq.) + Lateral load redistribution	FS	Lehmann and Or, 2012
SCOOPS3D	3D LEM (grouped columns)	External hydraulic inputs (not a hydrologic model)	FS / 3D surfaces	Reid et al., 2015
R.ROTSTAB / r.slope.stability	3D LEM (ellipsoids sampling)	External hydraulic inputs (not a hydrologic model)	FS / P_f	Mergili et al., 2014

These models demand detailed and high-resolution input data to effectively simulate the complex hydro-mechanical processes leading to slope failure. Key among these inputs is the time-dependent rainfall intensity, denoted as R(t), which drives infiltration and pore-water pressure dynamics within the slope.

Precise representation of R(t) is essential because the temporal variation of rainfall influences the transient response of soil moisture and pore pressure, thereby impacting slope stability over time.

Beyond climatic inputs, these models require a comprehensive set of spatially distributed geotechnical parameters to accurately characterise material behaviour and subsurface conditions.

These parameters include the effective cohesion (c'), which represents the soil's inherent shear strength excluding pore pressure effects; the internal friction angle (ϕ'), controlling frictional resistance along potential failure surfaces; the failure depth (Z), indicating the depth at which shear failure initiates; and hydraulic diffusivity, which describes the rate at which pore pressures equilibrate through the soil matrix.

Each of these parameters must be mapped across the modelled area to capture heterogeneous conditions that strongly influence localised failure susceptibility (Tofani et al., 2017).

In recent developments, numerous frameworks have been incorporating uncertainty treatment, encompassing a range of methodologies.

These include parameter-range envelopes (e.g. SINMAP), stochastic approaches (e.g. FSLAM) and explicit Monte Carlo sampling (e.g. TRIGRS-P, HIRESSS). The integration of these techniques has enabled the delivery of probabilistic maps that are commensurate with the decision-making requirements.

This emphasis on triggering, nevertheless, constitutes merely the initial stage. Landslides are a diverse class of mass movements, characterised by the material (e.g. rock, debris, earth) and the style of movement (e.g. fall, topple, slide, lateral spread, flow).

The modernised Varnes taxonomy (Hungr, Leroueil, and Picarelli, 2014) harmonises terms and is widely adopted in engineering geology. The selection of a specific PBM is contingent upon the objectives of the application. Triggering is governed by the storm's intensity-duration history and antecedent soil moisture, which control the timing and extent of instability. It is imperative to recognise that these processes are time-dependent and spatially variable. Consequently, the implementation of transient, distributed modelling techniques is essential. Furthermore, to accurately assess the impact, these triggering models must be coupled with dynamic runout models, which provide intensity fields. The intensity that governs damage, such as flow depth, velocity, and derived dynamic pressure, is realised during propagation (Hungr, 1997; Raetzo et al., 2002; Corominas et al., 2014). For this purpose, runout models translate the released mass into inundation extents and intensity fields along the path. Several families of runout models exist in the literature, classified based on their underlying physical and computational complexity (*Table 3*):

- Empirical & GIS-based Models: These are the simplest approaches, often used for regional-scale screening. They do not solve complex physical equations but rely on geometric rules, allometric scaling, or stochastic paths. LAHARZ (Schilling, 1998) is a classic tool for volcanic hazards, specifically designed to delineate lahar inundation zones based on empirical allometric volume-area and volume-width scaling relationships; Flow-R (Horton et al., 2013) is a widely used GIS tool that determines propagation extents using user-defined reach-angles (energy lines) and multiple-flow (multi-flow) routing rules to simulate spreading; GPP (Wichmann, 2017) is a similar GIS-based framework that simulates process paths using a mass-point concept, combining stochastic (random walk) routing with various user-selected empirical friction models;

- Reduced-Complexity Models: Designed to balance computational speed with physics, these models are often used for regional hazard or scenario analysis. ProDF (Gorr et al., 2022) is a specialised model that estimates runout using simplified empirical mobility and depth relationships, specifically calibrated for post-wildfire debris flow scenarios; DebrisDice (Bregoli et al., 2018) adopts a fully stochastic (Monte Carlo) approach, routing mass packets governed by simplified Voellmy-Salm physics to generate maps of invasion probability rather than deterministic depths;

- Depth-Averaged, Single-Phase (1φ) Models: This is the most common class of physically based models. They treat the flow as a single-phase mixture and solve the depth-averaged conservation equations (shallow-water type). They are often grouped by their rheological approach:

    a) Geotechnical/Rheological Models: This group focuses on mass movement dynamics. RAMMS::DebrisFlow (Christen et al., 2010; WSL Institute for Snow and Avalanche Research SLF [WSL], 2024) is a widely-used model that solves the 1φ equations using the specific Voellmy-Salm rheology (Coulomb μ + velocity-squared ξ) (Voellmy, 1955) and is valued for its direct output of dynamic pressure.

    TITAN2D (Patra et al., 2005) is tailored for dry granular avalanches (like rock avalanches or pyroclastic flows), implementing the Savage-Hutter rheology (Savage and Hutter, 1989) based on granular kinetic theory.

    DAN3D (McDougall and Hungr, 2004) uses a meshless, Lagrangian SPH-based solver, while MassMov2D (Beguería et al., 2009) uses a more traditional Eulerian (grid-based) solver. Both are highly flexible, allowing the user to select from a library of rheologies (e.g., Bingham, Coulomb, Voellmy) to simulate different materials;

b) Hydraulic (Non-Newtonian) Models: These are derived from riverine hydraulic modelling and adapted for viscous flows. FLO-2D (O'Brien et al., 1993) is a finite-difference model, and the HEC-RAS module (USACE-HEC, 2021-2024) is a finite-volume solver.
Both are highly effective at simulating channelised and unconfined flows by implementing non-Newtonian rheologies (Bingham, Herschel-Bulkley) that require a yield stress;

- Depth-Averaged, Two-Phase (2φ) Mixture Model: These physically advanced models simulate the solid and fluid components of the mixture separately, capturing more complex physics.
  D-Claw (George & Iverson, 2014; Iverson & George, 2014) is a benchmark two-phase model that solves the coupled equations for the solid and fluid fractions.
  It explicitly models granular dilatancy and pore-pressure evolution, which governs flow mobility; r.avaflow (Mergili et al., 2017) also implements a two-phase (Voellmy/Coulomb) model and is uniquely designed to simulate entire process chains, including entrainment and the cascading of different processes (e.g., an avalanche transforming into a debris flow);

- Advanced Continuum & Particle-Based Models: These are the most computationally intensive models, often used in research to simulate the full 3D physics of the flow and its interaction with structures.
  This class includes Lagrangian (particle-based) methods like Smoothed Particle Hydrodynamics (SPH), which excels at modelling complex free-surface flows and structural impact (e.g. Pastor et al., 2015), and Eulerian/Hybrid methods like the Arbitrary Lagrangian-Eulerian (ALE) Finite Element Method (FEM) (e.g., Crosta et al., 2003) and the Material Point Method (MPM), which is highly effective at simulating the full transition from solid failure to fluid-like flow (e.g., Li et al., 2016; Troncone et al., 2020).

Table 3. *Overview of Landslide Propagation (Runout) Models*

Model	**Class**	**Physics / rheology**	**Outputs**	**References**
LAHARZ	Empirical scaling	No rheology - volume; area/width scaling	Extent map	Schilling, 1998
Flow-R	Empirical/GIS	No rheology - empirical DEM routing	Extent map, Pathway map	Horton et al., 2013
GPP	Empirical/GIS (stochastic)	No rheology - empirical/stochastic DEM routing (Multi-friction models)	Probability/Invasion map, Pathway map, (Velocity/Depth optional)	Wichmann, 2017
DebrisDice	Reduced-complexity / stochastic	Voellmy-Salm (stochastic approach)	Probability/Invasion map, Extent map, Velocity map	Bregoli, Medina and Bateman, 2018
ProDF	Reduced-complexity (post-wildfire)	No rheology - empirical mobility/depth relations	Extent map, Depth/Thickness map, Probability/Uncertainty map	Gorr et al., 2022
DAN3D	Depth-averaged 1φ	Multiple rheologies (Coulomb; Voellmy; Bingham)	Depth/Thickness map, Velocity map, Extent map	McDougall and Hungr, 2004
MassMov2D	Depth-averaged 1φ	Multiple rheologies (Coulomb; Voellmy; Bingham)	Depth/Thickness map, Velocity map, Extent map	Beguería, Van Asch & Malet, 2009

Model	Class	Physics / rheology	Outputs	References
RAMMS::DebrisFlow	Depth-averaged 1φ	Voellmy-Salm	Depth/Thickness map, Velocity map, Extent map, Dynamic-pressure/Impact map	Christen, Kowalski & Bartelt, 2010; WSL, 2024
TITAN2D	Depth-averaged (granular)	Savage-Hutter (granular Coulomb)	Depth/Thickness map, Velocity map, Extent map	Patra et al., 2005
FLO-2D	Hydraulic/2D FV (non-Newtonian)	Non-Newtonian rheology (Bingham; Herschel-Bulkley)	Depth/Thickness map, Velocity map, Extent map	O'Brien, Julien and Fullerton, 1993
HEC-RAS	Hydraulic 1D/2D + non-Newtonian	Non-Newtonian rheology (Bingham; Herschel-Bulkley)	Depth/Thickness map, Velocity map, Extent map (channelised)	USACE-HEC (2021-2024)
D-Claw	Two-phase 2φ (mixture)	Granular-fluid mixture with dilatancy	Depth/Thickness map (mixture), Velocity map, Extent map	George & Iverson, 2014; Iverson & George, 2014
r.avaflow	Two-phase 2φ / process chains	Voellmy/Coulomb (two-phase)	Depth/Thickness map, Velocity map, Extent map (cascading)	Mergili et al., 2017
SPH Models	Continuum Mechanics (Particle-based)	Non-Newtonian Fluid Dynamics (Lagrangian)	Particle velocity, pressure, impact forces	e.g. Pastor et al., 2015
ALE-FEM Models	Continuum Mechanics (Hybrid)	Solid/Fluid continuum mechanics (ALE)	Velocity field, Stress/Strain, Extent map	e.g., Crosta et al., 2003
MPM Models	Continuum Mechanics (Hybrid)	Solid-to-fluid transition mechanics (Particle-in-Cell)	Velocity field, Stress/Strain, Extent map	e.g., Li et al., 2016; Troncone et al., 2020

For debris flows, dynamic pressure is a key intensity metric. Empirical thresholds suggest that reinforced concrete buildings often avoid major damage below ~35 kPa but may suffer major damage above ~100-110 kPa, while masonry buildings can suffer major damage at ~15-30 kPa (Hu et al., 2012; Kang & Kim, 2016). Hence, runout solvers that return depth/velocity/pressure fields are essential precursors to quantitative damage estimates.

2.2.2 Vulnerability (V)

The analysis of landslide risk depends on vulnerability, which measures the expected level of damage or loss for elements located in landslide-prone areas (Leone et al., 1996; Corominas et al., 2014).
Vulnerability therefore expresses the degree of expected impact on exposed elements in the event of a landslide. The evaluation of vulnerability requires identifying the exposed elements (e.g., buildings, infrastructure, people) and then estimating the consequences associated with the impact.
For physical components, vulnerability is commonly expressed as a damage-to-value ratio, meaning the proportion of the total asset value that is lost due to damage (Li et al., 2010).
For human populations, vulnerability instead corresponds to the likelihood of fatality when exposed to a landslide event (Fell et al., 2008).

The parameter is dimensionless and generally ranges from 0 to 1: 0 indicates no damage or loss, whereas 1 represents complete loss. In some cases, values slightly above 1 may occur when repair costs exceed the replacement cost of the element (Remondo et al., 2008).

The concept of vulnerability consists of two separate elements, which include physical aspects and social aspects. Physical vulnerability refers to the direct damage experienced by buildings and infrastructure systems, and it depends on characteristics that can be described and measured.

These include construction materials, maintenance condition, number of storeys, foundation type, landslide characteristics (such as type and intensity), and the position of the asset relative to the landslide pathway (Corominas et al., 2014; Uzielli et al., 2008; Li et al., 2010; Winter et al., 2014).

The assessment of infrastructure vulnerability also considers functionality loss, meaning that an infrastructure element may become partially or completely unusable because it is blocked, interrupted, or unsafe, even when structural damage is limited.

For this reason, the analysis requires information on infrastructure categories, size, and usage intensity, so that the consequences of obstruction or damage can be evaluated in operational terms.

Social vulnerability concerns impacts on human life and well-being. These impacts depend on factors such as landslide intensity, individual physical capacity and mobility, risk awareness, demographic variables, and their spatial distribution.

Social vulnerability is dynamic, because it can change over time as population characteristics, preparedness levels, and exposure patterns evolve. This complexity makes its evaluation more uncertain and challenging than physical vulnerability (Glade, 2003; Guillard-Gonçalves and Zêzere, 2018).

Methods:

- Heuristic: Vulnerability classes or values are assigned based on qualitative observations and professional judgement, which may lead to subjective results (Cardinali et al., 2002; Bell and Glade, 2004; Michael-Leiba et al., 2005);

- Data-driven: These methods link landslide intensity to observed losses through regression-based and probabilistic models, allowing vulnerability to be estimated from empirical intensity–damage relationships (Fuchs et al., 2007; Uzielli et al., 2008; Li et al., 2010; Papathoma-Koehle et al., 2012; Totschnig and Fuchs, 2013; Zhang et al., 2018; Singh et al., 2019);

- Analytical/mechanics methods: these methods are less common because they are complex, but they provide a physically grounded way to connect landslide intensity to the probability of exceeding specific damage states through mechanistic models and reliability-based fragility curves (Galli and Guzzetti, 2007; Ozturk et al., 2016; Peduto et al., 2017; Del Soldato et al., 2019; Chen et al., 2021).

The research focuses mainly on buildings because they are static, clearly identifiable elements and their damage can be described in a relatively direct and consistent way. In contrast, social vulnerability is harder to quantify and remains less developed, often relying on composite indicators that summarise socio-economic conditions (Xiao et al., 2022; Nor Diana et al., 2021).

2.2.3 Exposure (E)

In landslide risk frameworks, exposure analysis identifies and quantifies the elements located in areas that could be affected by landslides. Exposure therefore identifies what is present inside potential impact zones, including people, buildings and infrastructure, and who may face harm if the event affects that location (Corominas et al., 2014; Pellicani et al., 2014).

In this context, exposure describes which and how many elements are located in areas susceptible to landslide impact, prior to evaluating expected damage levels (vulnerability). Researchers perform exposure assessment for fixed assets such as buildings and transport routes by intersecting hazard footprints with asset inventories through spatial analysis, and then adjusting the evaluation according to landslide type. Rockfall hazards typically affect narrow corridors along the block trajectory, and the exposed area depends on block size and on the direction and extent of movement. Debris flows instead affect wider zones, and exposure can be quantified through the ratio between impacted area and total area, which provides a practical measure of the spatial extent of affected assets (Budetta et al., 2016; Corominas et al., 2019).

The movement of slow-moving landslides exposes elements across different zones, which include the slide mass, scarps, and runout fields. In these cases, exposure is distributed across multiple parts of the landslide system rather than being concentrated along a single path.

The estimation of exposure becomes more complex for mobile elements such as inhabitants or vehicles, because their presence is not constant. For this reason, exposure for mobile elements requires probabilistic representations that describe how people and vehicles are distributed in both space and time, for example depending on daily activities, mobility patterns, or evacuation phases.

This can be achieved through agent-based models and other dynamic simulation tools, which reproduce movement patterns and evacuation behaviour (Bunce et al., 1997; Fell et al., 2005; Jaiswal et al., 2010; Ferlisi et al., 2012). These approaches allow researchers to identify exposure hotspots under evolving scenarios, meaning locations where exposure may concentrate as conditions change (Jumadi et al., 2017, 2020).

The process of measuring exposure values must complement the spatial identification of exposed elements. Human exposure is typically expressed as the number of people exposed, sometimes separated into sociodemographic categories, because assigning monetary values to human lives is ethically and methodologically problematic (Maes et al., 2017). Physical assets can instead be valued financially through the replacement cost method, which estimates the current expense required to rebuild or replace an asset (Schuster and Fleming, 1986; NOAA, 2023).

The accuracy of exposure assessment depends strongly on the availability, quality, and spatial resolution of the data, and recent progress has provided extensive high-resolution datasets at both regional and global scales (De Bono and Mora, 2014; Emberson et al., 2020; Scaini et al., 2024). Despite these advances, exposure remains relatively under-addressed in landslide risk compared to other natural hazards, mainly because of difficulties related to scale and data availability (Corominas et al., 2014; Kreibich et al., 2022).

2.3 Innovation in Landslide Modelling

Recent advances in landslide risk modelling can be characterised by three primary and increasingly convergent branches of innovation: The enhancement of physically based models (PBMs) to simulate complex process chains is the first of these. The second is the application of advanced data-driven Machine Learning (ML) models. The third is the development of Hybrid Physics-ML frameworks, an emerging field known as Physics-Informed Machine Learning (PIML).

2.3.1 Process Chains and Physically Based Models (PBMs)

A significant line of research focuses on enhancing the physical realism of process simulations. Modern quantitative risk assessment frameworks increasingly rely on tightly coupled PBM chains, which explicitly link the initiation phase to the runout phase.

Predicting both initiation and the subsequent runout are the two key steps in regional landslide risk assessment. An example of such an integrated approach is the probabilistic framework proposed by Wu et al. (2025), which combines an initiation model and a runout model to estimate the exceedance probability of a given intensity measure, accounting for uncertainties in both models via Monte Carlo simulations (MC).

The accuracy of downstream runout predictions is extremely sensitive to the precision with which the initial volume of mobilised material is estimated. If runout is ignored, landslide risk can be significantly underestimated, as the "impact zone" is often far more extensive than the source area alone. Advanced propagation models, such as those based on the Shallow Water Equations (SWE), incorporate entrainment processes - the ability of a moving landslide to erode and assimilate material along its path - which can dramatically increase the volume and destructive potential of the mass movement. Beyond single-event modelling, the research horizon has expanded to include the simulation of multi-hazard cascading events. A prime example is the catastrophic flooding from Landslide-Dam Outburst Floods (LDOFs). Unified modelling platforms like r.avaflow allow for the simulation of the entire process chain within a single, integrated environment (Mergili et al., 2017).

2.3.2 Data-Driven Models (ML) and Interpretability (XAI)

Parallel to physics-based advancements, the evolution of ML algorithms for landslide susceptibility has seen a transition from classical statistical models to robust ensemble methods, including Random Forests, Gradient Boosting Machines and LightGBM, as well as complex deep learning architectures, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) (Breiman, 2001; Friedman, 2001; Ke et al., 2017; LeCun et al., 1998; Hochreiter and Schmidhuber, 1997) (*Table 4*).

This evolution represents a crucial innovation, as machine learning offers the capacity to incorporate a vast range of geoenvironmental and climatic factors, thereby complementing and extending traditional rainfall-threshold-based alert systems, which have been widely used in operational early warning frameworks (Guzzetti et al., 2008; Peruccacci et al., 2017; Segoni et al., 2018a; Peres and Cancelliere, 2021).

Rather than replacing these systems, ML approaches can enhance their predictive capability by integrating additional spatial, geoenvironmental and climatic variables into the modelling process.

Such progress has enabled the creation of accurate pixel-based Landslide Hazard Maps (LHMs), effectively including both the spatial and temporal dimensions for alerting purposes.

The processing capability of these algorithms, despite their complexity, makes them suitable for operational territorial early warning systems (Nocentini et al., 2025).

This growing complexity has, however, intensified the "black-box" problem, making the adoption of eXplainable AI (XAI) techniques crucial for ensuring the transparency and trust required for operational use. In order to address this challenge, a number of model-agnostic interpretation methods have been developed. The following section will present the most prominent of these, as reported in *Table 4*:

- Partial Dependence Plots (PDPs) (Friedman, 2001) and Individual Conditional Expectation (ICE) Plots (Goldstein et al., 2015), show the marginal effect that one or two features have on the predicted output. PDPs are widely used in landslide susceptibility modelling to verify the plausibility of the results;

- Permutation Feature Importance (PFI), measures the importance of a feature by quantifying the degradation of model accuracy when the values of that feature are randomly shuffled (permuted) (Breiman, 2001). This method is often used for variable importance estimation in ensemble algorithms;

- Layer-wise Relevance Propagation (LRP): This is specific to Deep Learning models (e.g., CNNs) and distributes the model's prediction backward through the network to the input features, helping to visualise which regions of the data were most relevant for the final decision (Bach et al., 2015);

- Local Interpretable Model-Agnostic Explanations (LIME), proposed by Ribeiro, Singh, and Guestrin (2016), provides local explanations for individual predictions. It works by approximating the behaviour of a complex model in the immediate vicinity of a specific prediction with a simpler, inherently interpretable model;

- SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee (2017), offers a unified framework based on Shapley values from cooperative game theory. Its strength lies in its ability to provide both local explanations (quantifying the marginal contribution of each feature to a single prediction) and global explanations (visualising overall feature importance). SHAP values are founded on sound theoretical principles (efficiency, symmetry, consistency), making them a powerful tool for scientific validation in geosciences, allowing researchers to verify that the model is learning physically plausible and scientifically sound relationships (Lundberg and Lee, 2017).

2.3.3 Hybrid Physics-ML Frameworks (PIML)

The emerging field combines PBMs with ML to create a strategic convergence which merges the physical rigour of traditional models with the predictive power of data-driven algorithms.
The research aims to address two main limitations which have historically affected these approaches, namely limited generalisability and high computational demand.
In this context, generalisability refers to the difficulty of applying a model reliably to new areas or new conditions, while computational cost refers to the time and resources required to run complex simulations.

PBMs and ML therefore interact through multiple integration strategies, which are summarised in *Table 4*:

- Physics-as-Features: The hybrid method in this approach uses physically meaningful information as input for ML models, so that the data-driven algorithm is trained using variables that already describe the physical state of the slope.
  This includes PBM outputs such as the Factor of Safety (FS), which indicates whether a slope is stable or close to failure, as well as direct physical observations such as InSAR-derived velocities, which measure ground deformation and surface movement.
  These physically based features can then be used to train ML algorithms such as LightGBM (Ullah et al., 2025).
  This approach helps maintain a link between machine-learning predictions and physical processes, improving interpretability and potentially increasing data efficiency because the training relies on variables with explicit physical meaning;

- Physics-Constrained Models (PINNs): The Physics-Constrained Models (PINNs) represent an approach which integrates neural networks with physical constraints through the Physics-Informed Neural Network (PINN) framework (Raissi et al., 2019).
  In this approach, the governing physical equations are embedded directly into the ML architecture so that the optimisation process is constrained by known physical laws rather than being driven only by statistical fitting.
  The study by Dahal and Lombardo (2025) provides a key example through the development of a PINN architecture for coseismic landslide susceptibility prediction.
  The model incorporates geotechnical parameters as latent variables, meaning parameters that are not directly observed but are inferred during training, and the neural network output is then passed into a model layer that implements the Newmark displacement method.
  This structure forces the final susceptibility estimates to remain physically consistent because the predictions are linked to a physically based displacement calculation rather than being produced only by correlation patterns;

- ML for Computational Efficiency (Surrogates): Another crucial innovation is the use of ML-based surrogate models, which are designed to approximate the behaviour of complex PBMs while requiring much lower computational effort.
  A representative example is the use of Gaussian Process Emulators (GPEs) (Rasmussen and Williams, 2006), which are trained to reproduce the input–output response of high-fidelity physically based simulations (Zhao and Kowalski, 2022).
  The models therefore provide a computational shortcut because they can return results that are similar to PBM outputs without running the full expensive simulation each time.
  This acceleration makes large-scale Uncertainty Quantification (UQ) feasible through Monte Carlo (MC) simulations, which would otherwise be computationally prohibitive.
  In this way, UQ can shift from a post-hoc validation step, performed only after simulations are completed, into an active predictive design methodology that supports scenario exploration and reliability assessment.

Table 4. *Overview of Machine Learning (ML) Methods*

Method	Models	Key Function & Use in Landslide Analysis
Ensemble Methods	Random Forests, Gradient Boosting Machines, LightGBM	Boosts predictive accuracy and stability by combining multiple models. Used for susceptibility mapping.
Deep Learning	CNNs, LSTMs, PINNs	Learns complex patterns from spatial (CNNs) and temporal (LSTMs) data. PINNs add physics constraints.
Surrogate Models	Gaussian Process Emulators (GPEs)	Creates fast approximations of complex physical models to enable large-scale uncertainty analysis.
Interpretability (XAI)	PDP, PFI, LRP, LIME, SHAP	Explains "black-box" models to ensure predictions are transparent and scientifically plausible.

2.4 Reliability of the Landslide Risk Modelling Chain

The reliability of landslide risk assessment depends on the robustness of the entire modelling chain, from the identification of potential triggering zones to the estimation of runout intensity, vulnerability, exposure, and expected losses. Landslide Susceptibility Mapping (LSM) remains a primary tool for identifying areas prone to slope failure and for supporting risk analysis and mitigation planning (Reichenbach et al., 2018).
However, when susceptibility or hazard models are used within a quantitative risk framework, their reliability cannot be evaluated solely in terms of spatial predictive accuracy.
It also depends on the methodological rigour adopted for calibration, validation and uncertainty quantification across all components of the risk equation, including vulnerability, exposure, and economic loss estimation.

Calibration is the systematic adjustment of model parameters to minimise the discrepancy between simulated outputs and observed data. This process differs substantially according to the type of model considered. In physically based models (PBMs), calibration mainly concerns uncertain geotechnical, topographic, rheological, and hydrological input parameters.
A common tool for this purpose is back-analysis, which uses well-documented historical events to infer the set of material properties or model parameters that best reproduce observed outcomes, such as failure surfaces, runout extent, or deposit footprints (Zhang et al., 2010a).
Historically based on manual trial-and-error procedures, back-analysis has progressively evolved towards automated Decision Support Systems (DSS) capable of exploring large numbers of parameter combinations.
At regional scale, probabilistic back-analysis, particularly within a Bayesian framework, can be used to treat soil or rheological parameters as random variables and update their prior distributions using historical landslide evidence (Zhang et al., 2010b). In data-driven susceptibility modelling, calibration has a different meaning. Model training usually optimises classification performance, but the raw outputs of algorithms such as Random Forests do not necessarily represent true event probabilities (Niculescu-Mizil and Caruana, 2005).
Therefore, probabilistic calibration is required when these outputs are used in quantitative hazard or risk analysis. This post-processing step transforms model scores into calibrated probabilities that better reflect the observed likelihood of slope failure, using methods such as Platt Scaling (Platt, 1999) or Isotonic Regression (Zadrozny and Elkan, 2002). Calibration and parameterisation issues also affect the impact and loss components of the risk chain. Vulnerability functions translate landslide intensity into an expected degree of damage, but their parameters are often derived from limited empirical damage datasets, expert judgement, numerical simulations, or literature-based curves.
For debris flows and rapid landslides, different intensity descriptors can be adopted, including flow depth, velocity, impact pressure, or combined intensity indices, and the selected descriptor strongly influences the resulting damage estimate (Quan Luna et al., 2011, 2014; Ciurean et al., 2017; Luo et al., 2023). At regional scale, vulnerability parameterisation is further constrained by the limited availability of building-specific information, such as structural typology, construction quality, maintenance state, and local protection or

shielding effects. Exposure is not calibrated in the same way as physical or statistical models. Its reliability depends primarily on the quality and consistency of the data used to describe the elements at risk.

Each building or asset should be associated with clear information, including asset type, economic value, cadastral value, occupancy, and use. These data must also be compatible with the hazard and runout outputs, particularly in terms of coordinate system, positional accuracy, geometric validity, spatial resolution and economic units. Inaccuracies in asset location, footprint geometry, or unit economic value may introduce systematic bias into the estimation of exposed value, which is then propagated through the risk calculation and affects the reliability of the final loss estimate.

Following calibration, **validation** provides the main test of a model's reliability and its ability to generalise to new, unseen data. A model can be well calibrated yet still fail to generalise; validation must therefore be conducted on independent data. A critical challenge in spatial modelling and validation is represented by Spatial Autocorrelation (SAC), according to which observations located close to each other tend to be more similar than distant observations (Tobler, 1970).

Standard Random Cross-Validation (R-CV) ignores this spatial dependence and may therefore produce data leakage and systematically inflated performance estimates. Spatial Cross-Validation (S-CV) addresses this problem by partitioning data according to geographical location, thereby providing a more reliable assessment of predictive capability under spatial transfer conditions (Brenning, 2012).

Model validation also requires performance metrics that are appropriate for the modelling objective and data structure. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is widely used (Fawcett, 2006), but it can be overly optimistic when applied to the highly imbalanced datasets typical of landslide studies (Saito and Rehmsmeier, 2015).

In such cases, the Precision-Recall Curve (PRC) provides a more informative alternative because it focuses on the performance of the positive class (Davis and Goadrich, 2006).

For runout models, validation should instead assess the spatial correspondence between simulated and observed footprints. Spatial-overlap metrics derived from contingency-table comparisons, such as the Critical Success Index or Threat Score (CSI/TS), which are equivalent to the Jaccard Index or Intersection-over-Union (IoU) for binary footprint comparisons, are particularly suitable because they account for true positives, false positives, and false negatives without being dominated by true negatives (Mead et al., 2021).

Beyond statistical validation, a further practical challenge concerns the transformation of continuous probability maps into discrete operational warning maps, such as Yellow, Orange, and Red alert levels.

The literature proposes distinct methodological strategies for selecting these decision thresholds. These range from early empirical rainfall Intensity-Duration thresholds (Caine, 1980), which evolved to include antecedent rainfall conditions (Guzzetti et al., 2007; Segoni et al., 2018a), to dynamic hydro-meteorological thresholds (Marino et al., 2020). Other approaches rely on statistical optimisation, using ROC curve analysis to identify a single optimal threshold or testing different percentiles to define multi-level warning systems (Piciullo et al., 2017). More recent approaches have shifted the objective from maximum statistical accuracy to maximum operational reliability, where missed alarms, or false negatives, are considered unacceptable errors in civil protection applications (Nocentini et al., 2025).

The Double-Threshold Validation Tool (DTVT), originally proposed by Bulzinetti et al. (2021) and subsequently revised by Nocentini et al. (2025), is a framework designed for this purpose. It addresses the scale discrepancy between pixel-level model outputs and territorial warning units by aggregating results into larger Pixel Aggregation Units (PAUs). The classification of a PAU is based on a pair of thresholds:

1. The Failure Probability Threshold (FPT): The probability value above which a single pixel is considered unstable;
2. The Instability Diffusion Threshold (IDT): The percentage or absolute number of unstable pixels required to classify the entire PAU as unstable. The DTVT iterates combinations of FPT and IDT and evaluates them according to a priority hierarchy, such as requiring zero false negatives for low-level alerts or minimising false positives for high-level alerts.

When the modelling chain is extended from hazard and runout modelling to impact and risk assessment, validation becomes more complex because the target variable is no longer only the location of instability, the operational warning level, or the spatial match of a runout footprint, but the expected damage or economic loss affecting exposed assets. Ideally, risk estimates should be validated against post-event damage data, including the location of impacted buildings, the observed degree of damage, the affected building fraction and, where available, repair or reconstruction costs.

However, such datasets are rarely available with sufficient spatial completeness and reliability, especially at regional scale. Consequently, validation of landslide risk models often relies on indirect evidence, such as the spatial correspondence between simulated impact areas and reported damage, the consistency of municipal-scale impact patterns, or comparison with post-event inventories and civil protection reports.

The final pillar of model reliability is **Uncertainty Quantification (UQ)**. Uncertainty is an unavoidable component of modelling complex natural systems and should be explicitly quantified by distinguishing between aleatoric uncertainty, which reflects inherent randomness, and epistemic uncertainty, which reflects lack of knowledge (Der Kiureghian and Ditlevsen, 2009).

A widely used method for propagating input parameter uncertainty is Monte Carlo simulation (MC) (Metropolis and Ulam, 1949). Given that Monte Carlo simulation can be computationally expensive, emulators or surrogate models, such as Gaussian Process Emulators (GPEs), are often used to approximate the response of complex numerical models at a reduced computational cost (O'Hagan, 2006; Rasmussen and Williams, 2006). In landslide runout modelling, Zhao and Kowalski (2022) demonstrated how Bayesian inference, Gaussian process emulation and active learning can be integrated to support efficient probabilistic parameter calibration. These approaches are often coupled with Sensitivity Analysis (SA), such as Morris or Sobol methods, to identify which input parameters contribute most to the output variance (Morris, 1991; Sobol', 2001; Saltelli et al., 2008).

In quantitative landslide risk assessment, uncertainty should not be propagated only through triggering and runout models, but through the entire chain leading to economic loss. This includes uncertainties related to input data, model parameters, model structure, vulnerability functions, exposure geometry, and the valuation of exposed assets. In runout modelling, uncertainty is not limited to parameter variability. It may also arise from model-structure assumptions and process simplifications, particularly when reduced-complexity models are used for regional-scale applications. One relevant example concerns sediment entrainment and volumetric growth during propagation. In reduced-complexity propagation models, the mobilised volume is often controlled by the predefined source area and by calibrated propagation parameters, while erosion and sediment incorporation along the flow path are not dynamically simulated (Bregoli et al., 2018). This simplification may affect runout extent, velocity distribution, impact pressure, and damage estimates where channelised debris flows entrain substantial bed material during propagation. Therefore, when dynamic entrainment is not explicitly represented, model outputs should be interpreted as regional-scale impact scenarios rather than fully process-resolving simulations of local debris-flow dynamics.

Additional uncertainty enters the chain through vulnerability and exposure. Vulnerability functions may produce substantially different loss estimates depending on whether damage is related to flow depth, velocity, impact pressure, structural resistance, or combined intensity descriptors.

Previous debris-flow risk applications have shown that the choice of vulnerability function can significantly affect estimated direct monetary losses for buildings, demonstrating that vulnerability should be treated as a major source of uncertainty rather than as a deterministic post-processing step (Quan Luna et al., 2014).

Similarly, exposure uncertainty may arise from the spatial accuracy of building footprints, the completeness of asset databases, the reliability of unit economic values, and the estimated fraction of each asset intersected by the simulated runout footprint. These sources of uncertainty should be considered together with geotechnical, hydrological, and rheological parameter uncertainty when quantitative loss estimates are produced.

An integrated probabilistic framework therefore provides the most coherent basis for assessing the reliability of landslide risk estimates. By propagating uncertainty from triggering and runout modelling through vulnerability, exposure, and economic valuation, it becomes possible to express risk results not as single deterministic values, but as uncertainty-bounded estimates, for example through percentile-based uncertainty intervals (Rougier et al., 2013; Knevels et al., 2023).

This probabilistic approach improves the transparency of quantitative landslide risk assessment by making uncertainty explicit, and supports more robust asset-level loss estimation (Uzielli et al., 2009; Eidsvig et al., 2014) as well as risk-informed decision-making across local to regional scales (Corominas et al., 2014).

Chapter 3

The Methodological Framework

The following chapter explains how the SADRI integrated computational framework operates and outlines its architectural design, which represents the methodological core of this doctoral research.

SADRI is designed to provide quantitative estimates of the risk associated with rainfall-induced shallow landslides through a complete and automated modelling chain.

The system orchestrates a sequence of models and geospatial analyses that start from rainfall-driven shallow-landslide triggering assessment and proceed through the identification of unstable spatial units, source-area extraction, runout simulation, physical impact assessment, and the estimation of economic risk for exposed assets. The framework is implemented to generate a technical reporting bundle for each spatial unit under investigation.

3.1 General Architecture

The SADRI system functions as an orchestrated and modular framework that links rainfall-driven shallow-landslide initiation to runout propagation and asset-level economic risk estimation within a reproducible workflow. At the process-modelling level, the system architecture starts from the rainfall-driven triggering assessment performed by HIRESSS (Rossi et al., 2013).

In the proposed workflow, HIRESSS represents the physically based triggering component of the framework: dynamic rainfall scenarios are combined with spatially distributed topographic, geotechnical and hydrological parameters to simulate the transient slope response and to evaluate shallow landslide instability conditions.

The model output consists of spatially distributed Failure Probability (P_f) maps, which provide the event-conditioned link between rainfall forcing, infiltration-driven reduction in slope stability and potential shallow landslide initiation. These Failure Probability maps are not used directly as final hazard or risk maps. Instead, they are converted into operational instability information through the FPT/IDT classification procedure.

The Failure Probability Threshold (FPT) identifies pixels whose probability of failure exceeds the selected instability condition, whereas the Instability Diffusion Threshold (IDT) evaluates whether the spatial concentration of unstable pixels within each Slope Unit is sufficient to classify that unit as unstable.

This classification is subsequently refined by the RE-ACT module, which adjusts the instability criterion according to local geo-environmental conditions and model-error patterns. Only the Slope Units classified as unstable after this filtering and refinement stage are transferred to the propagation component of the framework. Within these units, contiguous clusters of pixels exceeding the FPT are extracted and treated as candidate source areas for runout simulation. These source clusters provide the initiation geometry for DebrisDice (Bregoli et al., 2018), which simulates debris-flow-like propagation through a DEM-based stochastic approach and produces runout intensity information, particularly peak flow velocity (m/s).

The resulting intensity field is then coupled with vulnerability and exposure data to estimate asset-level economic consequences. In this way, the general architecture explicitly connects rainfall forcing, physically based shallow-landslide triggering, spatial alert classification, runout propagation and dynamic economic risk assessment within a single reproducible modelling chain.

The individual components of this chain are detailed in the following sections: the HIRESSS triggering model in *Section 3.3.1*, the DebrisDice propagation model in *Section 3.3.2*, alert-level definition and RE-ACT refinement in *Section 3.4*, SEARCH calibration in *Section 3.5*, and the dynamic risk analysis in *Section 3.6*.

From an implementation perspective, SADRI is driven by a central configuration file (`config.json`), which defines the analysis context, input and output paths, model parameters, thresholds, metadata, and workflow options. The computational workflow is implemented through six functional Python modules, which cover configuration loading, data preparation, schema definition, simulation management, risk analysis, and reporting. The workflow is launched and coordinated by the separate entry-point script `main.py`, which acts as the orchestration layer rather than as one of the six functional modules. It controls the sequential execution of the workflow and preserves a clear separation between simulation outputs and downstream risk computation. This architectural organisation allows each component to be maintained and tested independently, while the overall workflow remains reproducible through a single, explicit configuration file.

The following sections describe the role of each component:

1. **Initial Configuration (`config.json and config_loader.py`):** The workflow is driven by the external configuration file `config.json`, which defines the operational settings of the analysis. These include input and output paths, model parameters, classification thresholds, metadata, workflow switches, and strictness controls for validation and raster/vector alignment.
   The Python module `config_loader.py` reads and validates this configuration file, ensuring that the required settings are consistently available to the other modules of the framework;
2. **Data Management (`data_manager.py`):** The `data_manager.py` module performs geospatial dataset preparation by loading and pre-processing vector and raster inputs, and by executing reprojection, geometry repair, and raster alignment checks.
   Alignment checks can be enforced as fatal errors, although users can modify strictness settings to downgrade these checks to warnings;
3. **Schema Definition (`constants.py`):** This component provides schema definitions by specifying canonical column names that are used across the workflow, ensuring consistency between modules (e.g., `BASIN_ID, econ_value, vuln_index`);
4. **Simulation Management (`simulation.py`):** The `simulation.py` module functions as a wrapper for the external solver DebrisDice, which prepares model inputs, executes deterministic and Monte Carlo simulations, and produces georeferenced outputs. The module also writes Monte Carlo summaries (JSON/CSV) under stable per-unit folders;
5. **Risk Analysis (analysis.py):** This module contains the scientific logic for source handling, either FPT-driven or manually defined, impact metrics, intensity filtering, vulnerability calculation, exposure quantification, and per-asset damage and risk estimation, consistent with the adopted fragility approach;
6. **Reporting (reporter.py):** The reporter.py module generates a reporting bundle for each spatial unit, including a PDF report, an HTML preview, and a machine-readable JSON context. The module exports standardised maps, including a deterministic overview, a probabilistic map when available, and a study-area overview.

A final terminological note concerns the use of the term basin within the computational scripts. In the code, this label is used as a generic placeholder for the spatial unit of analysis. Although the present application mainly operates on Slope Units, the same computational structure can be adapted to other aggregation units, such as municipalities or sub-basins, depending on the spatial resolution and operational objective of the analysis.

3.2 The Computational Workflow

The end-to-end procedure follows a six-phase operational sequence (*Figure 3*). The flowchart provides a compact view of the modelling chain and the main intermediate artefacts exchanged between phases, while the next sections describe each phase in detail, focusing on inputs, key parameters, and outputs.

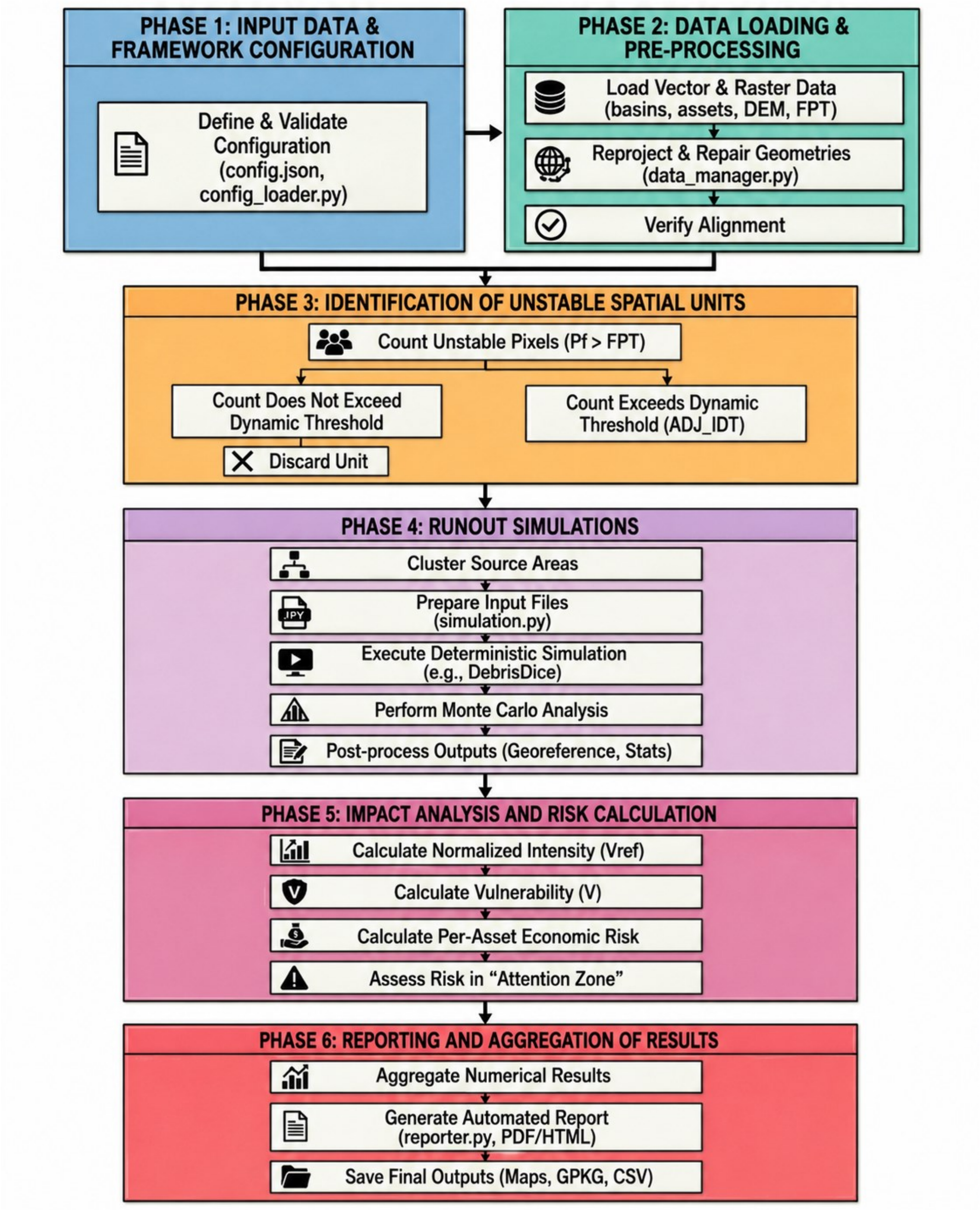


Figure 3. *Six-phase Flowchart of the SADRI (Scenario-Aware Dynamic Landslide Risk & Impact) modelling chain from configuration to simulation, risk computation, and reporting*

3.2.1 Phase 1: Input Data and Framework Configuration

The computational workflow depends on a specific set of input data which the `config.json` file manages through its centralised path and parameter structure.

This centralised approach ensures modularity and reproducibility, because analysts can adapt the framework to different study areas by modifying only the configuration file while keeping the same modelling chain.

The SADRI framework therefore requires each analysis to begin with the compilation of the `config.json` file, which functions as the main operational hub and control centre for all subsequent steps.

Table 5. *Input data required by the workflow*

Input Data	**Format**	**Description**
basins_shapefile	Vector File (.shp)	Polygons of the watersheds to be analysed
assets_shapefiles	Vector Files (.shp)	Dictionary of vector files representing exposed elements (e.g., buildings, infrastructure)
dem_raster	Raster File (.asc, .tif)	The Digital Elevation Model (DEM) of the study area
fpt_raster	Raster File (.tif, .asc)	The failure probability map used in P_f/FPT-based screening (optional when manual source definition is adopted)
source_areas_shapefile	Vector File (.shp)	(Optional/Alternative) Manual source areas used when manual source definition is adopted (`project.use_manual_sources`)
debrisdice_executable	Executable (.exe)	The executable file of the external simulation model (DebrisDice)
dynamic_thresholds_csv	CSV File (Optional)	A file containing the dynamic instability thresholds (ADJ_IDT) specific to each spatial unit
report_template	HTML File (Optional)	A custom HTML template for generating PDF reports
logo_dir	Folder (Optional)	A folder containing logos (e.g., .png) to be included in the report header

The file is organised into separate logical sections, and each section manages a specific part of the process through dedicated parameters and user-defined settings. These sections are fully user-configurable, and they are summarised below and reported in *Table 6*, while *Figure 4* provides an example of the configuration file:

- **Paths:** The Paths section contains all paths that point to input datasets and output directories. It includes the path to the Digital Elevation Model (`dem_raster`), the path to the failure probability map (`fpt_raster`), the vector file describing the spatial units (`basins_shapefile`), and the vector files describing exposed assets (`assets_shapefiles`).
  The same section also stores the path to the simulation model executable (debrisdice_executable), together with the output directory used to save final results (`output_dir`) and the directory reserved for temporary files (`temp_dir`). In addition, the configuration can include a `source_areas_shapefile` when the workflow adopts manual source definition instead of automated trigger-area identification;
- **Project:** The Project section contains project metadata together with global settings that apply to the entire workflow. This section defines the expected Coordinate Reference System (CRS) for all analyses (`expected_crs`), which ensures that every raster and vector layer is processed in a consistent spatial reference.
  The same section specifies the number of parallel processes that can be used to speed up computation (`max_workers`). It also collects the textual information that will appear in the final reporting outputs, such as the organisation name and the authors list;

- **Risk_parameters:** The Risk_parameters section gathers the thresholds and physical parameters required by the risk analysis stage. This section defines the Failure Probability Threshold (FPT) that is applied to the P_f map in order to identify relevant trigger zones (`fpt_threshold`).
  The same section also includes the velocity threshold (`velocity_threshold_ms`), the percentile adopted for intensity normalisation (`normalization_percentile`), and the buffer width expressed in metres that is used to define the "Attention Zone" around the runout areas (`safety_buffer_m`).
  Together, these parameters ensure that the hazard outputs are transformed into risk-relevant indicators using consistent and reproducible criteria;
- **Simulation_parameters:** The Simulation_parameters section governs how the simulation model behaves during execution. This section includes a DebrisDice subsection that contains the physical parameters of the runout solver, such as `friction_factor`, which controls resistance and influences propagation patterns.
  The same section also includes a `monte_carlo` (MC) subsection devoted to uncertainty analysis, where the number of iterations (`iterations`) and the random number generator seed (`random_seed`) are defined.
  These settings ensure that stochastic simulations can be executed in a controlled manner, and they guarantee that results remain reproducible when the same configuration is used;
- **Vulnerability:** The Vulnerability section defines the parameters required for the economic damage assessment.
  This section includes the maximum economic value that can be assigned to a single asset through a capping parameter (`exposure_max_value`), which prevents unrealistic loss estimates for individual elements.
  The same section also includes a dictionary of default values (`default_values`) that assigns a structural vulnerability index to different asset categories, such as "residential" and "commercial". This structure ensures that the damage estimation can remain consistent even when asset-specific vulnerability information is incomplete or not available;
- **Report:** The Report section controls the graphical and layout settings of the final PDF report. This section ensures that reporting outputs follow a standard structure and remain readable across different study areas.
  An inset subsection supports the customisation of the detail maps, and it defines their position, size, and the margin applied around the asset.
  These settings ensure that maps and figures provide an effective visual summary while remaining consistent with the reporting standards of the framework.

Table 6. *A comprehensive dictionary of the SADRI framework's configuration parameters as defined in the config.json file, with a concise description of their respective role*

Section	Parameter	Description
Paths		
	input_maps_folder	(Informational) Generic path for input data
	basins_shapefile	Vector file of watersheds to be analysed
	assets_shapefiles	Dictionary of asset vector file paths
	dem_raster	Path to the Digital Elevation Model (DEM)
	fpt_raster	Path to the Failure Probability Tensor (FPT) map
	source_areas_shapefile	(Optional/Alternative) Manual source polygons used when P_f/FPT-based source detection is not used
	debrisdice_executable	Path to the DebrisDice simulation executable
	output_dir	Main directory for all final outputs

	dynamic_thresholds_csv	(Optional) CSV file with spatial unit-specific initiation thresholds
	temp_dir	Directory for temporary simulation files
	report_template	(Optional) Path to a custom HTML report template
Project		
	expected_crs	Target Coordinate Reference System (CRS) for all spatial data
	max_workers	Number of parallel processes for the analysis
	organisation, authors, etc.	Metadata (e.g., institution, authors) for the PDF report
	logo_dir	(Optional) Directory containing logos for the report header
	debug_mode	If true, retains intermediate simulation files for inspection
	keep_mc_json_in_production	If true, saves the Monte Carlo summary JSON in production mode
	strict_validation	If true, enforces fail-fast validation of mandatory inputs
	strict_alignment	If true, treats DEM/FPT alignment mismatches as fatal errors (otherwise logged as warnings)
Risk_parameters		
	fpt_threshold	Probability threshold [0-1] to identify unstable pixels from the FPT map
	min_cluster_size, max_cluster_size	Minimum and maximum size for unstable pixel clusters to be considered valid source areas
	max_velocity_for_normalization	Fallback value for the reference velocity (Vref) used in normalisation
	velocity_threshold_ms	Velocity threshold (m/s) below which impacts are considered non-damaging
	safety_buffer_m	Buffer width (metres) around the impact area to define the "Attention Zone"
	save_runout_polygons, save_source_areas	If true, exports the geometries of runout and source areas
	fpt_data_format	Defines the FPT raster data format: "decimal" (0-1) or "percentage" (0-100)
	normalization_percentile	Percentile used to calculate the reference velocity (`Vref`) from all simulated maximum velocities
	clip_max_velocity	(Optional) Maximum velocity value (m/s) to cap outliers before percentile calculation
	default_min_unstable_pixels	Fallback for the minimum number of unstable pixels required to run a simulation
	attention_mc_stat	Statistic ("max" or "p95") to use for the MC backfill intensity calculation
	min_attention_intensity	Minimum normalised intensity assigned to assets in the "Attention Zone"
	use_attention_mc_backfill	If true, enables Monte Carlo re-evaluation of attention-zone assets
	tolerance_for_contact	Tolerance (m) to consider an asset impacted even if it only touches the runout area
Simulation_parameters		
	Debrisdice Subsection	
	debrisdice. [parameter: frictionm factor, viscosity, etc.]	Model parameters (e.g., friction_factor, density) passed to the DebrisDice solver
	Monte_Carlo Subsection	
	monte_carlo.iterations	Number of simulations to execute in the Monte Carlo analysis

	monte_carlo. uncertainty_factor	Perturbation factor (e.g., 0.1 for ±10%) applied to the model's physical parameters
	monte_carlo. parameters_to_perturb	Dictionary specifying which model parameters to perturb randomly
	random_seed	Seed for the random number generator to ensure the reproducibility of the analysis
Vulnerability		
	exposure_max_value	Maximum economic value (capping) per asset
	default_values	Default structural vulnerability index [0-1] assigned by asset type
	exposure_default_value	Default economic value assigned to assets without a specified value
Report		
	Inset Subsection	
	inset.[parameter: buffer_m, loc, height_frac, etc.]	Defines the inset's position, size, margin (buffer_m), and zoom limits (min_side_m)
	max_insets	Maximum number of detail maps (insets) to include per report. If 0, there is no limit

```
{
  "paths": {
    "input_maps_folder": "C:\\Users\\user\\Desktop\\SADRI_Project\\data",
    "basins_shapefile": "C:\\Users\\user\\Desktop\\SADRI_Project\\data\\basins.shp",
    "assets_shapefiles": {
      "edifici_test": "C:\\Users\\user\\Desktop\\SADRI_Project\\data\\buildings.shp",
      "infrastrutture_test": "C:\\Users\\user\\Desktop\\SADRI_Project\\data\\infrastructures.shp"
    },
    "dem_raster": "C:\\Users\\user\\Desktop\\SADRI_Project\\data\\dem.asc",
    "fpt_raster": "C:\\Users\\user\\Desktop\\SADRI_Project\\data\\basin_maps\\fpt.tif",
    "debrisdice_executable": "C:\\Users\\user\\Desktop\\SADRI_Project\\DebrisDice.exe",
    "output_dir": "C:\\Users\\user\\Desktop\\SADRI_Project\\output_runout",
    "dynamic_thresholds_csv": "C:\\Users\\user\\Desktop\\SADRI_Project\\adjusted_thresholds.csv",
    "temp_dir": "C:\\Users\\user\\Desktop\\SADRI_Project\\temp_runs",
    "report_template": "C:\\Users\\user\\Desktop\\SADRI_Project\\templates\\report_template.html"
  },

  "project": {
    "expected_crs": "EPSG:32632",
    "max_workers": 2,
    "organization": "University of Florence",
    "authors": "Dr. Nicolò Brilli",
    "report_version": "SADRI - Preliminary Analysis v1.1",
    "contact_email": "nicolo.brilli@unifi.it",
    "legal_notes": "This report was generated automatically. It is important to note that estimates are subject to uncertainty
    "logo_dir": "C:\\Users\\user\\Desktop\\SADRI_Project\\logos",
    "debug_mode": false,
    "keep_mc_json_in_production": true
  },
```

Figure 4*. Example of config.json file structure for two types of sections (paths and project) for SADRI framework*

3.2.2 Phase 2: Data Loading and Pre-processing

Once the configuration is validated, the orchestrator (`main.py`) invokes the `data_manager.py` module, the purpose of which is to prepare a consistent and reliable geospatial dataset for subsequent analyses.
This step is fundamental to ensuring the robustness of the entire process by minimising errors due to geometric or reference system inconsistencies:

1. **Vector Data Loading and Validation:** The process begins by loading the vector files defined in the paths section of the configuration file, specifically the spatial aggregation units (`basins_shapefile`) and the various exposed element layers (`assets_shapefiles`).

During loading, the system verifies the presence of mandatory columns (e.g., `BASIN_ID`) and halts execution if any are missing, thereby ensuring data integrity;

2. **Coordinate Reference System (CRS) Alignment:** An essential prerequisite for any geospatial analysis is the homogeneity of the coordinate system. The `data_manager.py` module ensures that all loaded vector layers are reprojected into the metric reference coordinate system defined by the `project.expected_crs` parameter (e.g., "EPSG:32632"). This operation is crucial for guaranteeing the validity of geometric calculations, such as areas, lengths, and distances (buffers);
3. **Geometry Repair and Integrity Assurance:** Geospatial data may contain invalid geometries (e.g., self-intersecting polygons) that can cause errors in subsequent operations. To prevent such issues, the framework implements an automatic repair procedure. Each geometry is validated and, if necessary, corrected using a zero-distance buffering technique (`buffer(0)`), a standard GIS operation for resolving topological problems without significantly altering the object's shape. Multi-part geometries are also handled by exploding them into single-part geometries for more granular management;
4. **Raster Alignment Verification:** In this phase, a preliminary consistency check is performed between the primary raster datasets: the Digital Elevation Model (DEM) and the Failure Probability (P_f) map. The `verify_raster_alignment` function checks that the two rasters share the same CRS, the same resolution (pixel size), and the same grid orientation. Any detected inconsistency is handled according to the project strictness settings: in strict mode mismatches are treated as fatal errors, otherwise they may be downgraded to warnings while preserving explicit logging for auditability.

3.2.3 Phase 3: Identification of Unstable Spatial Aggregation Units

In this phase, the framework selects the spatial aggregation units that will be subjected to runout analysis.

A crucial methodological aspect is the selective nature of this process: runout simulations are not performed on all spatial units present in the study area but are contingent upon a preliminary assessment of internal instability.

Only units that exceed a certain predefined critical threshold proceed to full runout modelling. This thresholding step acts as an operational gate between the triggering and propagation stages.

It avoids performing computationally demanding runout simulations in Slope Units that are stable according to the triggering model, or in units where only a limited number of pixels exceed the Failure Probability Threshold and therefore do not provide sufficient spatial evidence for a physically meaningful source area.

The operational criterion is defined through the FPT/IDT thresholding procedure, described in detail in *Section 3.4* and calibrated in the application phase in *Section 5.2.2.*

Its local refinement through RE-ACT is presented in *Section 5.3.* The FPT identifies unstable pixels within the Failure Probability map, whereas the IDT defines the minimum number of unstable pixels required to classify the entire spatial unit as unstable.

Therefore, the transition to runout modelling depends not only on the presence of individual unstable pixels, but on whether their spatial concentration within the unit is sufficient to support a plausible initiation condition.

In addition to P_f/FPT-driven screening, the framework also supports an alternative mode in which source areas are provided directly as a vector layer (manual source definition).

In that case, units proceed to simulation when at least one valid source polygon intersects the unit domain.

The first step, therefore, concerns the selection of spatial units that merit in-depth analysis. For each unit, the framework implements a two-level instability criterion:

1. **Unstable Pixel Count:** Firstly, the `analysis.py` module analyses the P_f raster (already clipped to the spatial unit's extent) and counts the total number of pixels whose value exceeds the Failure Probability Threshold (FPT). This threshold is initially defined in the `config.json` (`risk_parameters.fpt_threshold`) and is static for each spatial unit. The framework supports P_f maps stored either as decimals (0-1) or percentages (0-100) through the `fpt_data_format setting`;
2. **Dynamic IDT Threshold Verification:** The count obtained in step 1 is then compared against a second, operational threshold that is specific to each unit.

The system searches for the current `BASIN_ID` in the CSV file specified in `paths.dynamic_thresholds_csv` within the `config.json`, to find a customised threshold value named the *Adjusted Instability Detection Threshold* (ADJ_IDT).
This threshold is not a fixed value; rather, it is the result of an upstream optimisation procedure designed to correct the initial classification errors, namely False Positives (FP) and False Negatives (FN), associated with the initial FPT/IDT alert classification.
This external procedure analyses the environmental characteristics of each spatial unit, such as slope, root cohesion, lithology, elevation, and land use, and calculates their relative importance in generating classification errors using machine learning techniques, specifically Random Forest diagnostics interpreted through SHAP values.
The importance of these features is combined to calculate a score (WGT) that recalibrates the initially defined instability threshold (IDT), specifically for each erroneously classified spatial unit identified by the first threshold configuration, thus providing a new, optimised one (ADJ_IDT) to classify it as unstable. The system is designed to use a global fallback value defined in `risk_parameters.default_min_unstable_pixels` (`config.json`), equal to the initial non-optimised IDT threshold, should a specific value not be present in the ADJ_IDT column.

Only the spatial aggregation units in which the number of unstable pixels exceeds this dynamic threshold, or the fallback value corresponding to the initial IDT, are classified as "unstable" and proceed to the next phase. The remaining units are not forwarded to the runout simulation stage, thereby optimising computational resources.
This exclusion does not remove the local shallow-slope instability conditions from the HIRESSS outputs, where they remain represented in the Failure Probability maps.
It only prevents isolated or sparse unstable pixels from being propagated to the runout simulation stage when they do not exceed the IDT at the spatial-unit scale.
The framework therefore prioritises the most critical instability configurations, where the concentration of unstable pixels is sufficient to exceed the operational threshold and justify the activation of full propagation modelling.
This potential limitation is mitigated by calibrating the FPT/IDT configuration against the event inventory and by applying the RE-ACT correction procedure, which adapts the IDT according to local geo-environmental conditions and observed model-error patterns.
The threshold is therefore not an arbitrary exclusion rule, but a calibrated decision criterion that balances physical consistency, computational feasibility, and alert-classification performance.

3.2.4 Phase 4: Runout Simulations

Once an unstable spatial unit is selected, the `analysis.py` module proceeds to define the source areas from which to execute the runout simulations via an external solver (.exe). The main steps are reported below:

1. **Source Area Clustering:** The unstable pixels are grouped into spatially contiguous clusters, with variable dimensions defined by the user according to the case study and the type of landslides being investigated. It is indeed possible to define the minimum and maximum cluster size (`risk_parameters.min_cluster_size` - `.max_cluster_size`) in the `config.json` file to be used as source areas for the runout simulations.
   Clusters that exceed the maximum defined size are not eliminated but are subdivided into smaller, homogeneous sub-clusters that fall within the selected range.
   This subdivision is performed using the `AgglomerativeClustering` algorithm from the Scikit-learn library, which is constrained by a spatial connectivity graph (`kneighbors_graph`) to ensure the resulting sub-clusters maintain physical contiguity. The final set of valid clusters is then converted into polygonal geometries, representing the source areas for the subsequent simulations;
2. **Simulation Execution:** Managed by the simulation.py module, this phase begins with the automatic preparation of input files for the external executable specified in `config.json`.

Such inputs typically include ASCII grids of source areas, Digital Elevation Models (DEMs), and files defining physical or rheological parameters.
Each file is formatted to match the solver's requirements, ensuring consistent spatial extents across all datasets. Once the setup is complete, the system triggers the initial deterministic simulation, which models landslide propagation and generates a suite of output maps specific to the selected model;

3. **Monte Carlo Analysis:** The previously described phase is repeated N times, as indicated in `monte_carlo.iterations` (`config.json`).
   At each iteration, the model's physical parameters (friction coefficient, viscosity, etc.) specified in `monte_carlo.parameters_to_perturb` (`config.json`) are randomly perturbed within a user-definable range in `config.json` (`monte_carlo.uncertainty_factor`), generating a set of N plausible propagation scenarios.
   The entire stochastic process is made reproducible through the use of a seed for the random number generator (`random_seed`);
4. **Output Post-processing:** The simulation's primary output is georeferenced using the input DEM's metadata, producing a GeoTIFF raster for each simulation.
   These rasters are saved in a dedicated temporary subfolder (`mc_iter_*`), which is deleted at the end of the analysis unless `debug_mode` is active in `config.json` to allow for detailed inspection. Conversely, the final aggregated statistics, such as the mean and standard deviation of the risk calculated across all N simulated scenarios, are saved permanently in a single summary file (`montecarlo_summary.json`).
   Moreover, Monte Carlo outputs can be consolidated into per-basin hazard rasters (e.g., a probability-of-impact map expressed in 0-100% and a worst-case maximum-velocity map) to support probabilistic reporting and cartographic synthesis.

3.2.5 Phase 5: Impact Analysis and Risk Calculation

Once the georeferenced runout maps are obtained, the workflow enters the quantitative risk assessment phase, managed entirely by the `analysis.py` module. This crucial step translates the simulation results into an economic risk indicator (in Euros). The sequential steps are described in detail below:

1. **Intensity (I) Assessment:** The process begins with the quantification and subsequent normalisation of the phenomenon's intensity.
   In this study, intensity is defined as the maximum simulated flow velocity affecting a given asset. Operationally, for each exposed element intersecting the runout area, the system samples the velocity raster map (`velocity_georef.tif`) and identifies the maximum velocity pixel value within the asset geometry.
   This value is then filtered and normalised to obtain a dimensionless intensity indicator (`I_norm`), bounded between 0 and 1. The detailed implementation of velocity filtering, local reference velocity calculation, and normalisation is described in *Section 3.6.1*;

2. **Per-Asset Risk Calculation:** Once the intensity is normalised, the `analysis.py` module proceeds to calculate the economic risk for each individual asset that intersects the runout polygon.
   In the SADRI event-based workflow, the hazard component is already represented by the simulated runout footprint and by the local intensity used to derive vulnerability. Therefore, at the asset level, the deterministic economic loss is calculated as:

$$\text{Risk}\,[€] = \text{V}[\text{-}] \times \text{E}[€] \tag{3.1}$$

where V is the intensity-dependent vulnerability value and E is the exposed economic value. The detailed formulation of vulnerability and risk calculation is provided in Sections 3.6.2 and 3.6.4.

The components are calculated as follows:

- ✓ **Exposure (E):** Represents the economic value of the asset exposed to the event. Its calculation is based on the value present in the `econ_value` column, the presence of which in the asset vector files is verified during the data loading phase by the `data_manager.py` module.
  If this value is missing for a specific asset, the framework does not discard the element but assigns it a default value: the average economic value of all other assets of the same type (residential, industrial, etc.) present within the spatial unit under analysis. The final exposure calculation also depends on the asset's geometry:

  - For point assets, it corresponds to their total economic value;
  - For polygonal assets (e.g., buildings), it is calculated as a fraction of the total value, proportional to the area of the asset that intersects the runout polygon.

  For linear assets (e.g., road segments), exposure can be scaled proportionally to the intersected length fraction. Operationally, the exposed economic value is first scaled to account for the portion of the asset actually affected by the runout.
  Therefore, the effective exposed value in Euros (`Exposure_EUR`) is obtained by multiplying the asset's declared economic value (`econ_value`) by the fraction of its geometry impacted by the runout (`Impact_Fraction`);

- ✓ **Vulnerability (V):** Unlike a static value, vulnerability is calculated dynamically as a function that depends on both the asset's structural resistance and the intensity of the event impacting it:

  - **Structural Resistance (R_{struct}):** This is a dimensionless index ranging from 0 to 1, representing the asset's capacity to resist damage.
    The system reads this value from the `vuln_index` column of the asset file. If unavailable, it uses a default value based on the asset's type (e.g., "residential"), defined in the `vulnerability.default_values` dictionary in `config.json`;
  - **Normalised Intensity (`I_norm`):** This is the previously calculated dimensionless [0, 1] value representing the event's intensity;
  - **Vulnerability Function:** The final vulnerability is calculated via a piecewise "S-shaped" function that relates the ratio between the Normalised Intensity (`I_norm`) and the Structural Resistance (R_{struct}). The ratio I_{norm}/R_{struct} is therefore used as the input variable of the vulnerability function, rather than as the vulnerability value itself. This approach allows for modelling how an asset, even a robust one, can become highly vulnerable if struck by a high-intensity event;

- ✓ **Identification of Assets in the "Attention Zone":** In addition to assets directly impacted by the deterministic runout, the system also identifies those in a condition of potential danger due to their proximity. These are defined as "Assets in the Attention Zone."
  This zone is a buffer area around the runout perimeter, whose width in metres is controlled by the `risk_parameters.safety_buffer_m` parameter in `config.json`.
  For these assets, the framework does not merely identify them but performs a risk estimation procedure based on the Monte Carlo simulation results, if this is enabled via the `risk_parameters.use_attention_mc_backfill` flag.
  For each asset in the attention zone, the system analyses its performance across all *N* stochastic scenarios to calculate its impact intensity (`I_norm`) in a differentiated manner:

  - **Case 1: Asset impacted in at least one Monte Carlo simulation:** If the runout reaches the asset in one or more simulations, its normalised intensity (`I_norm`) is computed the same way as for directly impacted assets. Specifically, the system aggregates the impact velocities observed for that asset across the simulations in which it was hit and applies the selected statistic (e.g., maximum, 95th percentile) configured via `risk_parameters.attention_mc_stat` in `config.json`. The result is then normalised as previously described;

- **Case 2: Asset never impacted in Monte Carlo simulations.** If an asset lies within the attention zone but is never reached by the runout in any of the N simulations, its risk is set to zero by policy, and its exposure is recorded as the asset's full economic value solely as a diagnostic "nearby value," excluded from risk calculations.

This approach allows for the quantification of a more realistic and comprehensive risk for the attention zone, based on the impact probability and expected intensity that emerged from the uncertainty analysis.

3.2.6 Phase 6: Reporting and Aggregation of Results

The final phase of the workflow is dedicated to the synthesis, visualisation, and documentation of the results. This process begins with the aggregation of numerical results and culminates in the generation of a complete set of outputs, including a detailed technical report for each spatial aggregation unit.

Initially, the risk values calculated for each individual impacted asset are summed to obtain the total deterministic risk for the spatial unit. Concurrently, the statistics from the Monte Carlo (MC) analysis results are processed: the system calculates the mean (`Impact_Risk_MC_Mean`) and standard deviation (`Impact_Risk_MC_Std`) of the total risk, providing a quantitative measure of the expected risk and its uncertainty.

All these indicators - deterministic risk, the list of assets in the attention zone, and the Monte Carlo statistics - are consolidated into a results dictionary (`results_dict`), which serves as the input for the reporting module. Subsequently, the reporter.py module takes charge of this dictionary to generate the automated report, a process that orchestrates the creation of all summary elements.

First, the cartography is created in PNG format, which includes a main overview map (with the runout area and attention zone) and detailed maps with specific zooms for each directly impacted asset.

Simultaneously, the analytical content is prepared, such as detailed asset tables and the Executive Summary with key risk indicators. All these elements are then inserted into an HTML template via the Jinja2 templating engine and converted into a final PDF file using the WeasyPrint library. In addition to the PDF, the workflow can also persist an HTML preview and a machine-readable JSON context for auditability and downstream processing. At the conclusion of the entire process, the framework produces an organised set of outputs in the `output_dir folder` *(Table 7)*, which includes:

- **Risk Report (.pdf):** The main summary document for each spatial unit, integrating cartography, tables, and analysis;
- **Cartographic Maps (.png):** The high-resolution images of the main map and detail zooms, copied into a dedicated subfolder (pngs/) for easy consultation;
- **Geospatial Results (.gpkg, .shp):**

    - risk_assessment_results.gpkg: The primary vector output file, which joins the spatial unit polygons with all calculated risk indicators (e.g., `Impact_Risk, Risk_Class`);
    - Optional Vector Files: If enabled in the `config.json`, the system also exports the geometries of the runout polygons, source areas, and the attention zone;

- **Tabular Data (.json, .csv):**
    - montecarlo_summary.json: A file containing the aggregated statistics (mean, std. dev.) of the risk calculated during the Monte Carlo simulations;
    - MC_{BASIN_ID}.csv: A CSV file generated for each spatial unit, listing the individual risk values for each Monte Carlo iteration, allowing for a detailed audit.

Table 7. *A comprehensive view of SADRI generated outputs*

Output	**Format**	**Description**
Technical Report	.pdf	A document for each spatial unit, summarising all results including maps, risk indicators, analytical tables, and methodological parameters
Report Preview	.html	HTML version of the final report (same content as the PDF), for quick viewing/archiving (optional)
Report Context	.json	Structured data used to generate the report (indicators, tables, parameters/metadata), for auditability and reproducibility (optional)
High-Resolution Maps	.png	A series of high-resolution images, including a main map showing the runout and attention zones, and detailed zoom maps for each impacted asset
Geospatial Risk Results	.gpkg	The primary geospatial output file (risk_assessment_results.gpkg) containing spatial unit polygons enriched with all calculated risk indicators (e.g., Impact_Risk, Risk_Class)
Runout & Source Geometries	.gpkg	Vector files containing the geometries of the simulated runout polygons and the identified source areas
Monte Carlo Summary	.json / .csv	A file (montecarlo_summary.json/.csv) with detailed statistics from the simulations, including the mean, standard deviation, and risk values for each iteration
Probabilistic Hazard Rasters (MC)	.tif	Per-basin aggregated rasters such as probability-of-impact (0-100%) and worst-case maximum velocity (when Monte Carlo is enabled)

3.2.7 Operational Requirements

The `config.json` file provides users with multiple options to define data paths, yet the correct functioning of the SADRI framework still depends on compliance with a set of mandatory structural requirements.

For this reason, users must clearly distinguish between two types of components: those whose location and name are fixed and those whose location can be configured freely.

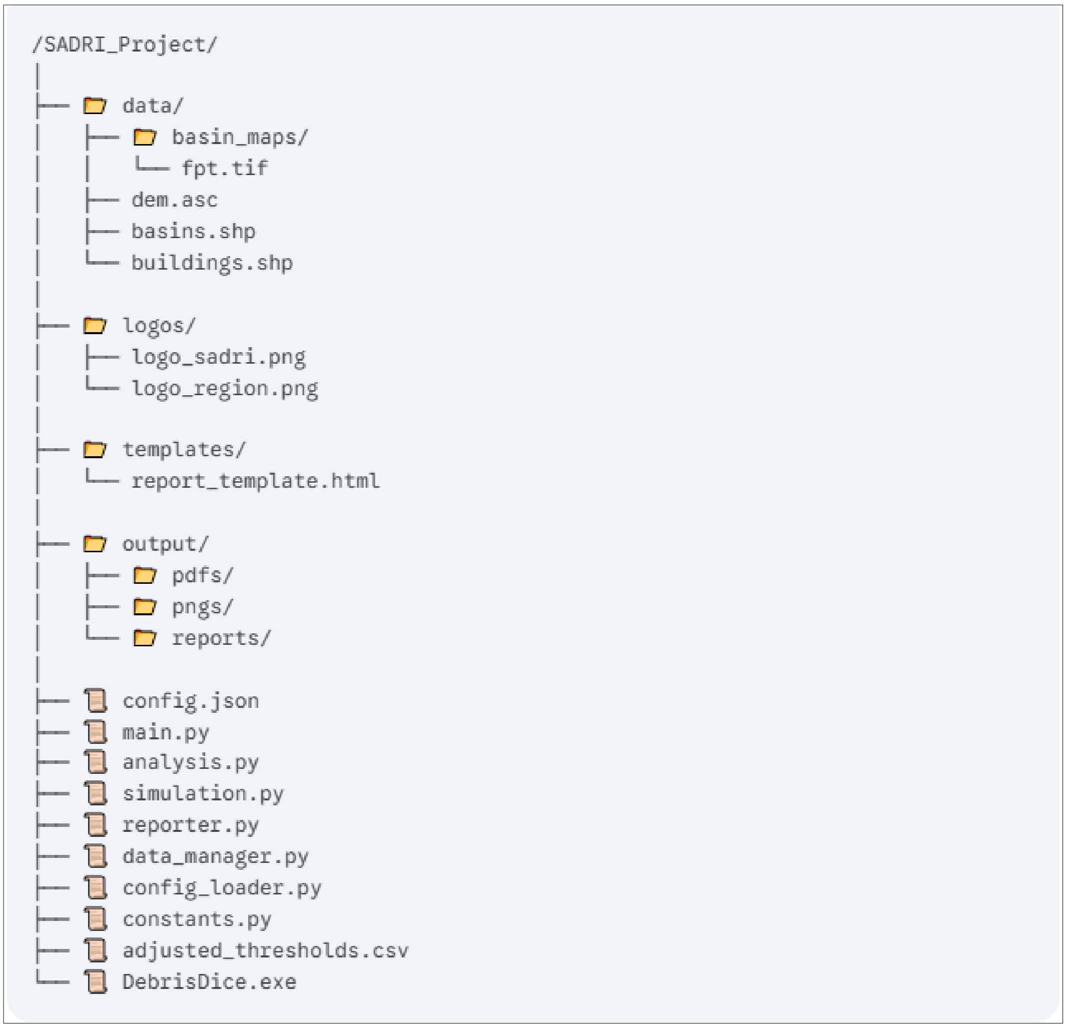
The program core, which includes the configuration file and all software modules, must remain inside the project root folder and must preserve its original names.

The configuration file must be named `config.json`, because this filename is referenced directly inside the startup script (`main.py`).

Similarly, the names of the Python modules (`main.py`, `analysis.py`, etc.) cannot be modified, because the scripts import each other using these exact filenames. Renaming them would therefore break the import mechanism and cause the program to fail at runtime.

The system also requires external resources such as geospatial datasets, the DebrisDice.exe executable, CSV files, templates, and logos, but these resources are location-independent as long as their absolute or relative paths are correctly specified in `config.json`.

The only folder whose existence is a prerequisite is the output directory (`output_dir`), although its location can also be chosen by the user. The following is a standardised project layout. The recommended structure is illustrated in *Figure 5*.



```
/SADRI_Project/
│
├── 📂 data/
│   ├── 📂 basin_maps/
│   │   └── fpt.tif
│   ├── dem.asc
│   ├── basins.shp
│   └── buildings.shp
│
├── 📂 logos/
│   ├── logo_sadri.png
│   └── logo_region.png
│
├── 📂 templates/
│   └── report_template.html
│
├── 📂 output/
│   ├── 📂 pdfs/
│   ├── 📂 pngs/
│   └── 📂 reports/
│
├── 📜 config.json
├── 📜 main.py
├── 📜 analysis.py
├── 📜 simulation.py
├── 📜 reporter.py
├── 📜 data_manager.py
├── 📜 config_loader.py
├── 📜 constants.py
├── 📜 adjusted_thresholds.csv
└── 📜 DebrisDice.exe
```

Figure 5. *Recommended hierarchical structure for the SADRI project's folders and files*

3.3 Physically Based Modelling

The SADRI framework integrates two distinct physically based models: one designed to describe landslide initiation (triggering) and one employed to simulate landslide propagation (runout).

3.3.1 Triggering Model: HIRESSS (HIgh REsolution Slope Stability Simulator)

The HIRESSS model (Rossi et al., 2013) was adopted as a physically based distributed model to evaluate the triggering phases of shallow landslides.
The model was developed for large-scale slope stability analysis with high spatial and temporal resolution, and it is designed to exploit parallel computing techniques (High-Performance Computing, HPC).
Because it can simulate time-evolving scenarios during rainfall events, HIRESSS is particularly useful for shallow landslide forecasting and for supporting early warning systems.
HIRESSS is divided into two interconnected modules: a hydrological and a geotechnical module (*Figure 6*):

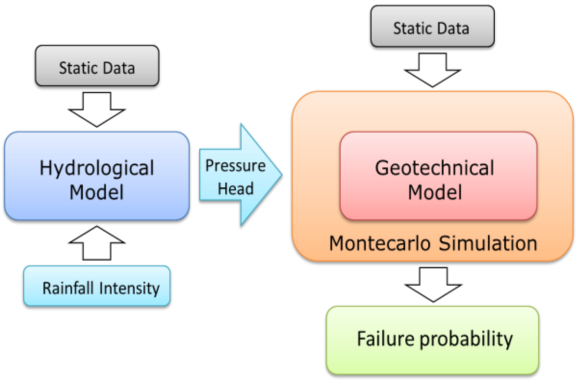


Figure 6. *HIRESSS Model flowchart*

- ✓ **The hydrological module:** This module is responsible for simulating the pore-water pressures that develop within the subsoil during rainfall events. A salient feature of this module is its capacity to solve, in parallel, a numerical approximation of the Richards equation, which governs water flow in unsaturated porous media. This formulation has been adapted specifically for use in conditions of high humidity, and includes hydraulic diffusivity in order to represent the propagation of pressure changes through the soil profile:

$$\frac{\partial h}{\partial t} = \left(\frac{D_0}{cos^2 \alpha}\right)\frac{\partial^2 h}{\partial Z^2}$$

[3.2]

where h is the pressure head, t is time, D_0 *is the maximum diffusivity under saturated condition, α is the slope angle, Z is depth measured from the soil surface.*

The hydrological module also computes the pressure load (h), using:

$$h = \frac{p}{\rho g}$$

[3.3]

where h is the pressure head (m), p is the pore water pressure (Pa), ρ is the water density (kg/m^3)*, and g is the gravitational acceleration* (m/s^2)*.*

In practice, the model solves the Richards equation in the spatio-temporal domain (z, t) by applying boundary conditions that represent rainfall infiltration and soil drainage, as described by Rossi et al. (2013). The resulting pressure load distribution is then transferred to the geotechnical module, which uses it to evaluate changes in stability over time.

- ✓ **The geotechnical module:** This module, on the other hand, evaluates slope stability by computing the safety factor (FS) using the infinite slope model (Skempton and Delory, 1957). The module includes key mechanisms that affect shallow slope stability, such as matric suction effects, soil density variations linked to changes in saturation, and the stabilising contribution of vegetation roots. In particular, root cohesion is included as an additional contribution to the total cohesion (c_{tot}), defined as (Masi et al., 2023):

$$c_{tot} = c' + c_r \quad \Rightarrow \quad c_r = kT_r\left(\frac{A_r}{A}\right)$$

[3.4]

where c' is effective cohesion (Pa), c_r *is root cohesion,* T_r *is root tensile, frictional, or compressive strength per unit area of soil,* A_r/A *is the root area ratio, and k is a coefficient that depends on the soil friction angle and on root orientation. This contribution can vary substantially depending on vegetation species and environmental conditions.*

The Factor of Safety (FS) is defined as the ratio between resisting forces and driving forces acting on the slope. Under unsaturated conditions it is expressed as:

$$\mathrm{FS} = \frac{\tan\varphi}{\tan\alpha} + \frac{c_{tot}}{\gamma_d\, y\, \sin\alpha} + \frac{\gamma_w\, h\, \tan\varphi}{\gamma_d\, y\, \sin\alpha} \left(\left[1 + (h_b^{-1}\,[h])^{\lambda+1} \right]^{\frac{\lambda}{\lambda+1}} \right)^{-1}$$

[3.5]

where φ is the friction angle, α is the slope angle, γ_d is the dry soil unit weight, y is the depth, γ_w is the water unit weight, h is the pressure head, h_b is the bubbling pressure, and λ is the pore size index.

In saturated conditions, the equation is simplified to:

$$FS = \frac{\tan\varphi}{\tan\alpha} + \frac{c_{tot}}{(\gamma_d(y-h) + \gamma_{sat}h)\sin\alpha} - \frac{\gamma_w h \tan\varphi}{(\gamma_d(y-h) + \gamma_{sat}h)\tan\alpha}$$

[3.6]

where γ_{sat} is the saturated soil unit weight.

Input Data

The HIRESSS model requires two classes of input data, both provided as raster maps with identical spatial resolutions. Static data describe the hydrogeological soil properties and are loaded once at the beginning of the simulations. Dynamic data represent the time-varying rainfall forcing and are updated continuously during the simulation *(Figure 7).*

Static Data

These parameters define the morphological, geotechnical, and hydrological properties of the study area:

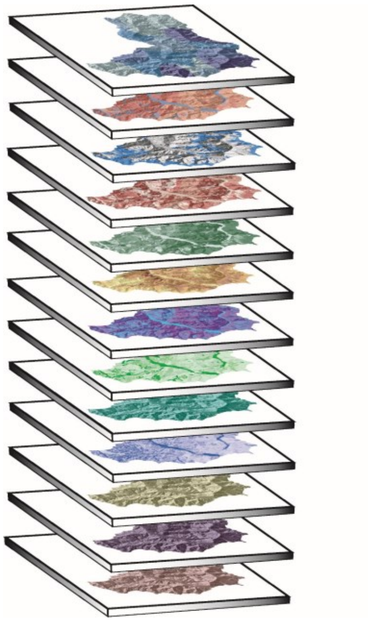
- *Soil Thickness (m):* The depth of the soil cover above the bedrock, which is crucial as the model calculates the slope stability as a function of depth;
- *Root Cohesion (Pa):* The additional shear strength provided to the soil by plant root systems. This parameter quantifies the stabilising effect of vegetation on the slope;
- *Slope Angle (°):* The slope steepness, derived directly from the DEM. It is a primary factor controlling the gravitational driving forces acting on the slope;
- *Outcrop Mask:* A binary map used to exclude areas of exposed rock or non-soil surfaces from the stability analysis, thereby optimising computational resources by focusing only on relevant areas;
- *Friction Angle (°):* The effective friction angle of the soil (φ), a key parameter from the Mohr-Coulomb failure criterion that defines the soil's shear strength;
- *Cohesion (Pa):* The effective cohesion of the soil (c), representing the soil's intrinsic shear strength when the effective normal stress is zero;
- *Dry Unit Weight (N/m³):* The unit weight of the soil in dry conditions (γ_d), used to calculate the gravitational forces acting on the soil mass;
- *Saturated Hydraulic Conductivity (m/s):* The saturated hydraulic conductivity (k_{sat}), which governs the rate of water infiltration and pore pressure propagation through the soil matrix;
- *Porosity (%):* The porosity of the soil (n), representing the volume of voids relative to the total volume, which determines the soil's water storage capacity;
- *Bubbling Pressure (m):* The bubbling pressure (h_b), a soil-water retention curve parameter from the Brooks and Corey model that influences hydraulic diffusivity in unsaturated conditions;
- *Pore Size Distribution Index (-):* pore-size index (λ), controlling retention and hydraulic response;

- *Residual Water Content (-):* The residual water content (θ_r), representing the minimum amount of water that remains in the soil pores under very dry conditions.

Dynamic Data

These parameters represent the time-varying trigger forcing slope stability:

- Rainfall Data (mm/h): hourly raster maps of rainfall intensity, which provide the key dynamic forcing for the hydrological module. These rasters can be obtained from meteorological radar products or by spatial interpolation of rain-gauge observations.



Static data

- Friction angle (°);
- Dry unit weight (N/m^3);
- Slope Angle (°);
- Cohesion (Pa);
- Root cohesion (Pa);
- Soil Thickness (m);
- Residual water content (θ_r);
- Porosity (%);
- Hydraulic conductivity (m/s);
- Bubbling pressure (m);
- Pore Size Index (λ);

Dynamic data

- Hourly rainfall intensity (mm/h).

***Figure** 7. HIRESSS static and dynamic input data*

Model Outputs

The primary outputs of HIRESSS consist of raster maps in ASCII GRID format, describing the spatial distribution and temporal evolution of slope stability conditions.
A key feature of the software is that it can produce outputs under both a deterministic and a probabilistic perspective, because it implements a Monte Carlo simulation procedure to manage parametric uncertainty.
The calculation is performed through a predefined number of iterations for each pixel.
In each iteration, geotechnical input parameters are sampled from predefined probability density functions (Gaussian), and a deterministic Factor of Safety (FS) is computed. After completing all runs, results are statistically aggregated to produce the final outputs.

The following maps are produced by this dual approach:

- **Mean Factor of Safety (FS) maps:** FS represents the central tendency of stability conditions.
  Even though each run produces a deterministic FS value (based on one specific parameter combination), the final map does not correspond to any single iteration.
  Instead, it reports the arithmetic mean of all FS values computed across the Monte Carlo runs. This mean reduces the effect of parameter variability and provides an estimate of the most probable

stability state for each pixel. A mean FS < 1.2 indicates that, on average, the pixel is prone to instability;

- **Failure Probability Maps (P_f, %):** This is the most significant output of the stochastic approach. HIRESSS produces temporal maps (typically 24-hourly maps for each simulated day), where each pixel expresses the probability that FS will fall below the critical threshold of 1.2.
  The probability is computed as the ratio between the number of runs in which FS < 1.2 and the total number of runs performed. In this way, P_f explicitly quantifies how parameter uncertainty influences the prediction of instability;

- **Hydrostatic Pressure Maps (Pressure Head):** The hydrological module can also export pressure-head maps (positive or negative/matric) at each time step.
  These maps are essential to interpret how rainfall infiltration modifies subsurface hydrological conditions and to link pore-pressure evolution directly to the corresponding stability changes.

Overall, using both mean stability outputs (average FS) and probabilistic outputs (P_f) allows the model to overcome the limitations of a purely deterministic analysis.
The framework provides, at the same time, an estimate of the expected stability condition and a quantitative measure of how robust that estimate is under the uncertainty that characterises natural systems.
The range of parameter variability around central values is user-selectable, and the adjustable parameters are reported in *Table 8*.

Table 8. *Static model parameters for which the range of variation can be entered*

Relative error	c' (Pa)	φ' (°)	Slope (°)	n (%)	γ_d (N/m^3)	Dtb (m)	k_{sat} (m/s)	λ (-)	h_b (m)	I. Pore (-)
(%)	-	-	-	-	-	-	-	-	-	-

Dtb = Depth to bedrock (Soil Thickness); I. Pore = Pore Size Index

3.3.2 Runout Model: DebrisDice

The DebrisDice model (Bregoli et al., 2018) simulates debris flow propagation by integrating physical laws with a stochastic approach applied to digital terrain data.
It is designed to estimate both the inundation probability and the primary flow dynamics, such as velocity, providing a robust basis for regional-scale hazard assessment.
More specifically, DebrisDice propagates predefined initiation cells or source areas over a DEM and estimates runout probability and flow velocity.
It does not include a dynamic entrainment module that updates the mobilised volume during propagation. Consequently, sediment entrainment and volumetric growth are not explicitly simulated during the propagation stage. The material propagated by the model is therefore controlled by the initial source areas and by the calibrated propagation parameters, rather than by a dynamic erosion or entrainment law acting along the flow path.

Rheology

The model's physical core is governed by the Voellmy rheological equation (Voellmy, 1955), a widely utilised equation that describes the motion of granular-fluid flows. The equation represents a dynamic balance between the forces driving the flow and those resisting it:

1. Driving Force: This is the force of gravity, which propels the flow downslope. Its magnitude is determined by the local slope angle (θ), derived directly from the Digital Elevation Model (DEM);

2. <u>Resisting Forces:</u> The Voellmy rheology (1955) models resistance as a combination of two distinct mechanisms:

   - Coulomb Friction (Dry Friction): A velocity-independent term representing the sliding friction at the base of the flow, controlled by the Friction Factor (μ);
   - Turbulent Resistance (Viscous Friction): A velocity-dependent term that models the internal turbulence and viscous resistance of the flow, controlled by the Mass-to-Drag ratio (k), also known as the turbulence coefficient.

The Voellmy's rheological equation for granular-fluid debris flows (Voellmy, 1955), is represented below, where flow velocity (v) along a path length (l) evolves as:

$$\frac{1}{2}\left(\frac{dv^2}{dl}\right) = g\,sin\theta - \,\mu\,g\,cos\theta - \frac{v^2}{k}$$

[3.7]

where g is gravity acceleration, θ is the local slope angle, μ represents sliding friction (Coulomb Friction coefficient), k is a turbulence-related drag coefficient controlling velocity-dependent resistance.

Stopping Criteria

Flow propagation is constrained by a series of stopping criteria, which are fundamental for defining the final runout extent. DebrisDice implements three main criteria:

1. **Minimum Slope (β):** The flow halts if it reaches a cell where the slope in all possible directions is less than a user-defined minimum threshold;

2. **Reach Angle (α):** The flow stops if the angle connecting the source to the flow front drops below the threshold value defined by the Reach Angle parameter (α). This geometric constraint limits the maximum runout distance based on the initial mobilised volume (V) via:

$$\mathrm{Tan}\,\alpha = \left(\frac{H}{L_{max}}\right) = 0.97V^{-0.105}$$

[3.8]

   where H is vertical drop and L_{max} is max runout distance. The mobilised volume (V) controls the maximum runout distance used by the reach-angle stopping criterion, but it is not updated along the flow path by an explicit entrainment process.

3. **Minimum Velocity (v):** In accordance with the Voellmy equation, the flow halts if its calculated velocity drops to zero or below.

Input Data

The model requires a series of inputs to execute a simulation, provided as ASCII text files. These files must be located in the same directory as the executable and must have the exact names with which they are provided to users:

- <u>Topography (topo.dat):</u> A Digital Elevation Model (DEM) in ESRI ASCII raster format representing the terrain elevation over which the flow will propagate, providing slope angles θ and elevation for flow routing;

- Initiation Points or Grid (ini.dat or coordinates): Defines the simulation's starting location. It can be provided in two ways, as specified in the input.dat file:

    1. Single point**:** By providing the coordinates (e.g., UTM) of a single initiation point;
    2. Initiation grid: By providing a binary raster file (ini.dat) where cells with a value of 1 represent the areas from which the flow can start;

- Physical and Simulation Parameters (input.dat): This text file controls the simulation's behaviour and the physical parameters of the flow:

    - *Number of Iterations (N_{iter}):* The number of stochastic simulations (random paths) the model must execute;
    - *Source Area:* A parameter in the input.dat file that defines whether the source is a single point or a grid;
    - *Coulomb Friction Coefficient (μ):* Resistance to sliding at the base of the flow (dry friction), related to material properties and surface roughness. It is one of the two key parameters of the Voellmy equation;
    - *Mass-to-Drag Ratio (k):* This parameter models the turbulent or viscous resistance, which varies with the flow's velocity. It is the second fundamental parameter of the Voellmy equation;
    - *Maximum Reach Angle (α):* A stopping criterion based on the overall flow trajectory. The flow stops if the angle between the starting point and its current position drops below this value, which is empirically related to the mobilised material volume. It is provided as the tangent of the angle;
    - *Minimum Slope Angle (β):* Another stopping criterion, based on local topography. The flow stops if it reaches a cell where the slope in all directions is below this minimum threshold. This value is also provided as a tangent;
    - *Initial Yield Stress (τ_y):* The minimum stress threshold (in Pascals) that must be overcome for the material to begin flowing. It is a typical parameter for Bingham fluids;
    - *Viscosity (η):* Represents the dynamic viscosity (in Pa·s) of the material once it has started to flow, describing its internal resistance to movement;
    - *Exponent (n):* A dimensionless parameter that modulates the relationship between stress and strain rate in non-Newtonian models (e.g., Herschel-Bulkley);
    - *Density (ρ):* The bulk density of the solid-liquid mixture (in kg/m³), which influences the gravitational driving force.

Model Outputs

The simulation outcomes are generated as ASCII raster grids provided to the users in .txt format, which detail the computed flow probabilities, velocities, and diagnostic layers:

- Inundation Probability Maps: Maps derived from Monte Carlo simulations of multiple stochastic flow paths, defined as:

$$P_{xy} = \left(\frac{n_{pass}}{n_{iter}}\right) \tag{3.9}$$

where n_{pass} is paths crossing cell (x,y) and n_{iter} total simulations. This quantifies uncertainty and variability of debris flow paths.

These outputs quantify uncertainty and variability of debris flow paths. For this purpose, the model generates two distinct probability maps, each based on a different flow-stopping criterion.

Both maps contain normalised values in the [0, 1] range, representing the probability that a cell will be traversed by a flow:

  - *Result_angle.txt:* This map shows the inundation probability calculated using the stopping criterion based on the Maximum Reach Angle (α). It is useful for assessing the maximum potential extent of the phenomenon in relation to the mobilised volume;
  - *Result_vel.txt:* This map shows the inundation probability calculated using the velocity-based stopping criterion (the flow stops when its velocity becomes non-positive).

- Velocity Map (Velocity.txt): A raster file that indicates, for each cell, the maximum simulated velocity (v, m/s) reached by the flow during the iterations;

- Diagnostic Maps: These grids do not directly represent the runout result but are useful tools for quality control of the input topography and for understanding the model's behaviour:

  - *Control.txt:* This map serves a dual purpose. First, it identifies topographical sinks present in the DEM, which could cause an anomalous stop of the flow.
    Second, it highlights the cells where the flow stopped because the local slope was below the set minimum threshold (Minimum Slope Angle).
  - *Maxpend.txt:* This is an optional output that traces, with a value of 1, the single path of maximum slope that the flow would follow starting from a single source.
    It is useful for visualising the most probable trajectory in the absence of lateral dispersion.

Probabilistic Approach

DebrisDice runs a large number of iterations simulating flow paths using random walk adjusted by local slope and velocity physics.
This multiplicity of simulated paths captures natural variability due to complex terrain and uncertain flow parameters.
The resulting probabilistic maps express spatial variability in debris flow likelihood, helping decision-makers assess hazard zones with quantified confidence.

3.4 Definition of Alert Levels

Landslide susceptibility models represent a powerful tool for the spatial prediction of slope instability, producing continuous maps of Failure Probability (P_f). Specifically, each pixel on the map is associated with a probability value, which indicates the likelihood of failure under the analysed conditions.
However, for civil protection operations, these continuous probability maps are not directly actionable for operational decision-making, because decision-making usually requires a clear and discrete classification. Operational systems must be able to state, in a direct way, whether a spatial unit (such as a river basin, slope unit, or municipality) should be considered “at risk” or “not at risk” during an ongoing triggering rainfall event. For this reason, P_f maps are commonly translated into a discrete alert system by applying two complementary thresholds:

- **FPT Threshold (Failure Probability Threshold):** a threshold applied to P_f values, which defines the minimum probability above which a single pixel is classified as unstable;

- **IDT Threshold (Instability Diffusion Threshold):** a threshold applied to the spatial unit scale, which defines the minimum number of unstable pixels required to classify an entire area as unstable.

The choice of each FPT/IDT threshold pair is a crucial step, because it directly determines the reliability and operational usefulness of the warning system. In practice, multiple threshold pairs are often associated with different colour-coded alert levels (e.g., Yellow, Orange, Red).

Within this logic, more conservative threshold pairs (lower thresholds) typically correspond to a lower alert level (e.g., Yellow), because they label instability more easily and therefore tend to generate more warnings.

Conversely, more restrictive pairs (higher thresholds) correspond to a higher alert level (e.g., Red), because they trigger warnings only when the evidence of instability is stronger and more widespread.

The definition and validation of these threshold pairs is not straightforward and is usually performed through back-analysis. This procedure evaluates how well each threshold pair performs by comparing predicted unstable areas with inventories of past landslide events, in order to measure the frequency of correct detections and classification errors. A fundamental limitation arises when a single global IDT threshold is applied uniformly over territories that are geomorphologically heterogeneous.

In this situation, the same number of unstable pixels may not represent the same level of risk everywhere, because predisposing factors such as lithology, morphology, or land use can vary substantially from one area to another.

As a consequence, a global threshold often becomes an approximation that cannot fully capture local variability, and this produces unavoidable classification errors.

Two key types of errors are commonly described:

- **False Positives (FP):** Instability is predicted for a spatial unit, but no landslide actually occurs;
- **False Negatives (FN):** A spatial unit is predicted to be stable, but landslide events occur in reality.

To address this limitation, the RE-ACT framework was developed as a data-driven post-processing strategy. The underlying hypothesis is that the errors of a physically based model are not randomly distributed, but instead are correlated with specific environmental conditions (for example, particular lithologies or land-use patterns) that the original physical model may not represent perfectly.

Importantly, the proposed solution is not to modify the physical model itself. Instead, the framework aims to improve the operational classification by optimising the IDT threshold locally, in a way that is spatially explicit. This means that, for each spatial unit, the IDT value is corrected using environmental factors that are statistically shown to be associated with the observed classification errors.

A key strength of the framework is that the correction is not treated as a "black box", because RE-ACT integrates Explainable AI (XAI) tools, specifically the SHAP (SHapley Additive exPlanations) algorithm. This approach improves predictive accuracy while keeping the correction process transparent, since SHAP quantifies which environmental variables contribute to the error, and how strongly they influence it.

In this way, the framework not only reduces misclassifications, but it also produces new interpretative insight into which factors control instability at the local scale.

The first methodological step is to define the initial alert levels by applying different threshold pairs derived from model results. In this thesis, three alert thresholds were defined following the methodology described by Nocentini et al. (2025).

Three FPT/IDT pairs were therefore selected to build a three-level warning system, each calibrated with a specific operational objective:

- ❖ <u>Low alert level (Yellow):</u> calibrated to be extremely cautious, with the main objective of achieving zero False Negatives (FN). This ensures that no real event is missed, even if this leads to a higher number of False Positives (FP);
- ❖ <u>High alert level (Red):</u> calibrated to be highly restrictive, with the objective of minimising False Positives (FP). This level is activated only under the most critical conditions, in order to avoid frequent false alarms;
- ❖ <u>Intermediate alert level (Orange):</u> designed as the best compromise between the two extremes, balancing the need to identify real events with the need to limit false alarms.

The subsequent step is to define a representative threshold level to which the correction pipeline can be applied. In practice, this means starting from a static IDT threshold and then assigning a new, locally corrected IDT value to each Slope Unit through the RE-ACT procedure, thereby improving the initial classification.

In this study, the correction pipeline was applied to the Orange threshold, because its balance between sensitivity and specificity represents the most realistic operational starting point.

The objective of RE-ACT is therefore to further refine this compromise threshold by correcting residual errors through spatially variable adjustments.

3.4.1 The RE-ACT Procedure: a pipeline for the optimisation of initial classification threshold

The RE-ACT framework is implemented in a fully reproducible Python pipeline, orchestrated by a main script (`RE-ACT.py`) and controlled via a single configuration file (`config.yml`).

All methodological choices, file paths and parameter values are specified in this file, which allows the user to reproduce and adapt the analysis with minimal manual intervention.

The main script is organised around three core classes:

- **ModelConfig:** centralises all configuration parameters, checks their validity, and verifies the consistency between the configuration and the input data;
- **EnvironmentalProcessor:** performs data assimilation and feature engineering, integrating environmental information into the spatial units;
- **ModelEngine and ResultVisualizer:** implement, respectively, the XAI-based diagnostic model and the generation of numerical and graphical outputs.

Execution is launched from the command line: `python RE-ACT.py --config path/to/config.yml`.

3.4.1.1 Input Data

The workflow is driven by a set of inputs, all specified in `config.yml`:

- **A spatial unit vector file** (e.g. slope units, basins), containing:
    - Geometry of each unit;
    - A column with the initial four-class classification (`TP`, `TN`, `FP`, `FN`) of the model (configured via `error_column`);
    - A column with the raw model score for each unit (configured via `count_column`), defined as the number of pixels whose failure probability exceeds the FPT threshold;
    - Optionally, a column with the count of exposed elements at risk (e.g. buildings), used when risk-aware corrections are enabled;
- **A set of environmental data layers, comprising both raster and vector datasets** (e.g. slope, elevation, lithology, land cover, root cohesion). Each dataset is associated, in `config.yml`, with:
    - A logical name (e.g. `slope`, `dem`, `litho`, `LULC`, `rootC`);
    - A file path (constructed from `base_input_path` and `env_files_relative`);
    - A type (`continuous` or `categorical`) defined in `env_types`.

All required input paths and the names of critical columns are exposed as configuration parameters and are not hard-coded in the script.

3.4.1.2 Configuration Structure

The behaviour of the RE-ACT pipeline is fully controlled by a single YAML configuration file (`config.yml`). This design choice keeps methodological decisions and data paths separate from the Python code, which helps preserve workflow reproducibility and makes the framework easier to apply to different case studies. In practical terms, users can adapt RE-ACT to a new study area by editing the `config.yml` file, without modifying the scripts that implement the pipeline.

All key elements of the analysis are explicitly declared in this configuration file. The document includes numerous parameters grouped into functional blocks that correspond to the main components of the RE-ACT workflow. The model settings block defines the analysis name, the initial four-class labelling (FP, FN, TP, TN), and the raw IDT-like score used as the starting point for the correction process.

The paths block specifies the locations of spatial units, environmental datasets, and output folders, ensuring that all required inputs and results are correctly referenced. CRS and alignment options control the harmonisation of spatial layers, ensuring that raster and vector datasets share a consistent coordinate reference system and remain properly aligned before the analysis starts.

The environmental-variables block regulates how predictors are managed depending on their type, distinguishing continuous variables from categorical variables. The IDT correction and weights section controls the environmental weight (WGT) applied during the correction step and defines the admissible range of IDT adjustments, so that the corrections remain bounded and do not introduce unrealistic threshold shifts.

Additional blocks define the settings for statistical screening, Random Forest and XAI configuration, validation procedures, and visualisation options.

When applying the framework to a new case study, only a limited subset of parameters typically needs to be updated. In most applications, users mainly revise the model settings, file paths, and environmental variables, while the more advanced options related to IDT correction, screening, XAI, validation, and plotting can often remain at their default values.

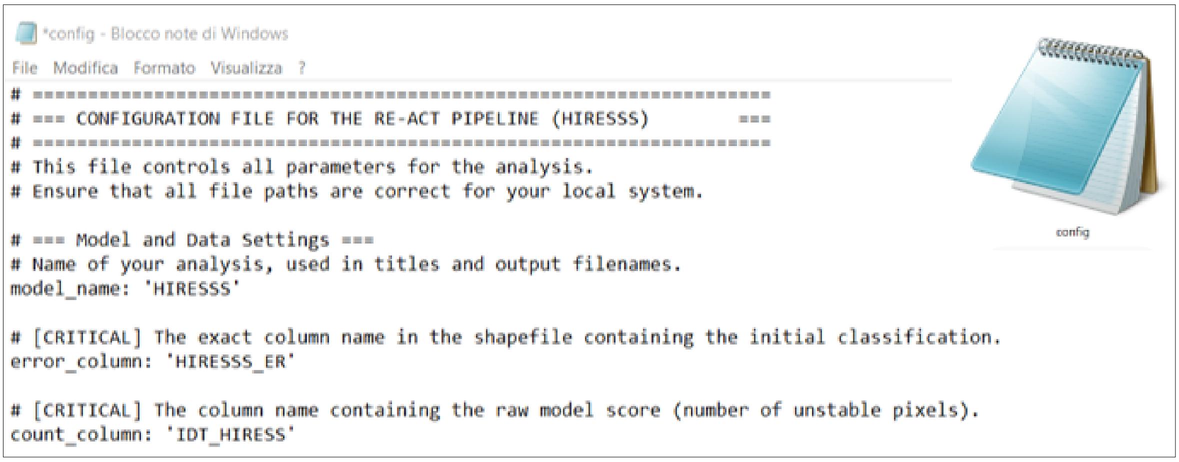
Table 9 provides a detailed parameter-by-parameter description of the configuration file following this structure, and *Figure 9* illustrates an example of the resulting YAML file.

Table 9. *A comprehensive dictionary of the RE-ACT framework's configuration parameters as defined in the config.yml file, with a concise description of their respective role*

Section	Parameter	Description
Model settings		
	`model_name`	Name of the analysis (e.g. HIRESSS); used in figure titles and as layer name in final_results.gpkg
	`error_column`	Column in the territorial-unit shapefile containing the initial four-class labels (TP, TN, FP, FN)
	`count_column`	Column with the number of unstable pixels per unit (raw IDT-like score)
	`error_fp_value`	String used in error_column to identify False Positives (e.g. 'FP')
	`error_fn_value`	String used in error_column to identify False Negatives (e.g. 'FN')
	`error_categories`	List of all allowed labels in error_column, usually ['TP','TN','FP','FN']
	`risk_elements_column`	Column with the count of exposed elements at risk (e.g. buildings) in each unit
	`use_risk_score`	If True, IDT correction is modulated by a normalised risk factor; if False, only environmental conditions are used
File paths		
	`base_input_path`	Folder containing all environmental datasets (rasters and vectors).

Section	Parameter	Description
	`input_shapefile`	Full path to the territorial-unit vector file (e.g. basins or slope units) with geometry, error_column and count_column
	`output_path`	Output directory where logs, tables, plots and geospatial layers are written
	`env_files_relative`	Mapping between logical variable names (e.g. slope, dem, litho) and filenames relative to base_input_path
CRS and alignment		
	`default_env_crs`	Default CRS assigned to environmental layers without a declared CRS (e.g. "EPSG:32632")
	`env_crs`	Optional CRS specified per environmental variable; overrides default_env_crs when present
	`basins_crs`	CRS assigned to the territorial-unit shapefile if no CRS is stored in the file
	`target_crs (optional)`	CRS to which all data are reprojected for analysis; if omitted, the spatial units CRS is used as master CRS
	`align_to`	Name of the reference raster (e.g. dem) used as template grid for aligning all continuous rasters
	`Target_resolution`	Target cell size for the template grid; if null, the resolution of the align_to raster after reprojection is used
Environmental variables		
	`env_types`	Declares each environmental variable as 'continuous' or 'categorical' to route it to the correct processing pipeline
	`env_columns_orig`	List of continuous variables (e.g. ['slope','dem','rootC']) used as numerical predictors in RF/SHAP
	`env_columns_cat`	List of prefixes for fractional-cover variables derived from categorical layers (e.g. ['lt_','clc_'])
	`cat_ref_columns`	For each categorical dataset (e.g. litho, LULC), name of the attribute containing the raw category codes
	`category_maps`	Maps raw codes or labels from categorical datasets to internal class names (e.g. 1 → lt_1, "Residential" → clc_5)
	`cat_expected (optional)`	For each categorical variable, list of internal class names to keep; only these classes are turned into fractional-cover columns
	`cat_group_names`	Human-readable names for each categorical prefix (e.g. lt_ → "Lithology", clc_ → "Land Use / Land Cover")
	`cat_prefix_map (optional)`	Explicit mapping from logical variable names (e.g. litho, LULC) to their prefixes (e.g. lt_, clc_) when defaults are not used
Labels		
	`cont_display_names`	Publication-ready labels for continuous variables (e.g. "Slope (°)", "DEM (m)"), used in plots and maps
	`cat_display_names`	Readable labels for categorical classes (e.g. "Alluvial", "Residential"), used in legends and SHAP plots
IDT correction		
	`original_idt`	Global baseline IDT threshold of the original model (in this work: Orange alert level)
	`max_idt_change_fp`	Maximum allowed increase of the IDT threshold in FP units

Section	Parameter	Description
	`max_idt_change_fn`	Maximum allowed decrease of the IDT threshold in FN units
	`idt_step`	Increment used to round the adjusted thresholds (e.g. 1 pixel)
	`max_allowed_idt`	Absolute upper limit for the IDT threshold (number of unstable pixels)
Weights		
	`wgt_base_nonerror`	Baseline environmental weight assigned to correctly classified units (usually 0.0)
	`wgt_normalization`	Scheme used to rescale raw weights (e.g. 'per_class_minmax', 'robust_minmax', 'zscore_soft', 'minmax')
	`wgt_q_low - wgt_q_high`	Lower and upper quantiles used to clip extreme weights in robust normalisation schemes
Statistical screening		
	`disable_statistical_filters`	If True, all environmental variables are used; if False, only those passing the statistical screening are retained
	`statistical_test_p_value`	Target significance level (e.g. 0.05) for FDR-adjusted q-values
	`min_prevalence`	Minimum fraction of units in which a categorical class must be present to be considered
	`min_abs_count`	Minimum number of units in which a categorical class must occur to be considered
	`min_cramers_v`	Minimum effect size (Cramér's V) required for categorical variables
	`min_cliffs_delta`	Minimum effect size (Cliff's δ) required for continuous variables
	`min_fraction_area`	Fractional area threshold defining presence/absence of a categorical class within a unit
Random Forest / XAI		
	`rf_max_depth`	Maximum tree depth in the Random Forest diagnostic models (null = no explicit limit)
	`rf_min_samples_split`	Minimum number of samples required to split an internal node
	`rf_n_estimators (internal)`	Number of trees; set internally by the script (default 300)
Validation and robustness		
	`holdout_test_size`	Fraction of units used for the test set in each hold-out split (e.g. 0.3)
	`holdout_random_state`	Random seed for the reference train–test split
	`holdout_random_states`	List of seeds for repeated hold-out runs used in the robustness analysis
Visualisation		
	`max_features_plot`	Maximum number of features shown in global-importance and SHAP summary plots
	`Visualization_dpi`	Resolution for saved plots (e.g. 300 dpi)
	`plot_palette`	Colours assigned to TP, TN, FP, FN in maps and scatter plots
	`boxplot_palette`	Colours used for original vs. adjusted IDT in the threshold boxplot



```
# ===================================================================
# === CONFIGURATION FILE FOR THE RE-ACT PIPELINE (HIRESSS)        ===
# ===================================================================
# This file controls all parameters for the analysis.
# Ensure that all file paths are correct for your local system.

# === Model and Data Settings ===
# Name of your analysis, used in titles and output filenames.
model_name: 'HIRESSS'

# [CRITICAL] The exact column name in the shapefile containing the initial classification.
error_column: 'HIRESSS_ER'

# [CRITICAL] The column name containing the raw model score (number of unstable pixels).
count_column: 'IDT_HIRESS'
```

***Figure 8**. Example of `config.yml` file structure for RE-ACT framework*

3.4.1.3 Generated Outputs

The workflow produces a structured set of outputs, managed by the `ResultVisualizer class`, ready for scientific reporting and operational use; *Table 10* provides a comprehensive overview of the outputs generated by the RE-ACT module:

- Performance_report.txt: A text report summarising the key performance metrics before and after correction, along with the results of the spatial analysis;
- Final_results.gpkg: The main geospatial output, a GeoPackage file containing the spatial units enriched with all original and newly calculated columns;
- Adjusted_thresholds.csv: A table listing the original and the new adjusted threshold (`ADJ_IDT`) for each spatial unit;
- Visualisations Folder: A suite of high-resolution scientific plots and maps, including SHAP summary plots, performance curves (ROC), and thematic maps, ready for publication.

***Table 10.** A comprehensive view of RE-ACT generated outputs*

Output name	**File name / pattern**	**Format**	**Description**
Performance report	`performance_report`	TXT	Summary of train/test sizes, initial vs. final metrics, percentage improvements and Moran's I results for the reference run
Final geospatial results	`final_results`	GPKG	Spatial units with geometry, original and final classification, environmental features, WGT, optional NRISK and ADJ_IDT
Adjusted thresholds table	`Adjusted _thresholds`	CSV	For each unit: BASIN_ID, raw score (count_column), adjusted threshold (ADJ_IDT), final label and main diagnostic fields

Output name	File name / pattern	Format	Description
Classification comparison layer	`Classification_comparison`	SHP	Layer with initial and final labels per unit and a status field (e.g. "Changed (FP → TN)", "Unchanged")
Robustness metrics	`Robustness_metrics`	CSV	Performance metrics for all hold-out runs, including differences relative to the reference seed
Statistical tests (run-wise)	`statistical_tests`	CSV	Statistical screening results for the reference run (test type, q-values, effect sizes, pass/fail)
Statistical tests (all runs)	`statistical_tests_all_runs`	CSV	Same as above, aggregated over all seeds with an additional seed column
Execution log	`pipeline`	LOG	Chronological log of the pipeline execution (info, warnings, non-fatal errors)
Crash report (if any)	`crash_log`	TXT	Stack trace saved automatically in case of fatal errors
Emergency data dump (if any)	`CRASH_DUMP`	GPKG	Snapshot of the working GeoDataFrame at the time of a critical failure
Global importance plots	`importance_FP` `importance_FN`	PNG/PDF	SHAP-based global feature importance for FP and FN diagnostic models
SHAP summary (beeswarm)	`shap_summary_FP,` `shap_summary_FN`	PNG/PDF	Beeswarm plots showing the distribution of SHAP values per feature for FP and FN
SHAP dependence plots	`shap_dependence_<variable>_FP,FN`	PNG/PDF	Dependence of SHAP values on each continuous predictor (FP and FN models)
SHAP categorical bar plots	`shap_bar_group_FP,` `shap_bar_group_FN`	PNG/PDF	Importance ranking of categorical classes (e.g. lithology, land use) based on mean
Threshold boxplot	`boxplot_thresholds`	PNG/PDF	Distribution of original vs. adjusted IDT thresholds for corrected units
Operating-point comparison	`operating_point_comparison`	PNG/PDF	Initial vs. final operating points (TPR, FPR) in ROC space, with their balanced accuracy
Precision-Recall curve	`precision_recall_final`	PNG/PDF	Precision–Recall curve and Average Precision for an auxiliary RF classifier (landslide vs. non-landslide units)
Map of changed units	`changed_Slope_Units_map`	PNG/PDF	Map highlighting units that changed class after correction (FP→TN, FN→TP) and those unchanged
Initial classification map	`Initial_classification_map`	PNG/PDF	Spatial distribution of the initial four-class classification
Final classification map	`Final_classification_map`	PNG/PDF	Spatial distribution of the final four-class classification after adaptive IDT correction

Output name	File name / pattern	Format	Description
Threshold map	`Threshold_Map`	PNG/PDF	Spatial distribution of the adjusted thresholds (ADJ_IDT)
Environmental weight map	`Weights_map`	PNG/PDF	Spatial distribution of environmental weights (WGT) per unit
Risk vs. weight scatter	`scatter_density_errors`	PNG/PDF	Scatter (and optional density contours) of NRISK vs. WGT for FP and FN units (when risk is used)
Moran's I scatterplot	`moran_scatterplot_errors`	PNG/PDF	Moran scatterplot of residual errors, annotated with Moran's I and its p-value

3.4.1.5 The Computational Workflow

The RE-ACT pipeline operationalises the concepts introduced in the previous sections, fixed FPT/IDT thresholds, local environmental conditioning and risk modulation, into a fully reproducible computational workflow.

The entire procedure is implemented in the main script (`RE-ACT.py`), controlled through the configuration file (`config.yml`), which specifies all methodological choices, data paths and parameter values. Conceptually, the pipeline is organised into six phases (*Figure 8*), which are executed sequentially.

In Phase I (*Data assimilation and feature engineering*), the framework constructs a harmonised geospatial dataset in which each spatial unit is enriched with continuous and categorical environmental descriptors, ready for statistical analysis and machine learning.

Phase II (*Statistically guided feature selection*) identifies those environmental variables that exhibit a statistically robust and practically relevant association with FP and FN errors, thereby improving interpretability and reducing noise in the diagnostic models.

Phase III (*Explainable AI diagnostics*) trains separate Random Forest models for FP and FN and uses SHAP values to quantify the contribution of each predictor to the propensity for misclassification.

In Phase IV (*Spatially adaptive threshold correction*), this diagnostic information is translated into unit-specific IDT thresholds, optionally modulated by risk, and the four-class classification is recomputed.

Phase V (*Validation and performance assessment*) evaluates the robustness and generalizability of the corrections through repeated hold-out validation, summary metrics and a Precision-Recall analysis.

Finally, Phase VI (*Spatial analysis of residual errors*) investigates whether remaining misclassifications are spatially random or still exhibit systematic clustering, by means of Moran's I and related spatial diagnostics.

Together, these phases transform an initial, globally calibrated IDT threshold into a set of spatially adaptive thresholds, while explicitly documenting which environmental conditions are associated with FP/FN errors and how much predictive performance improves under repeated *hold-out validation*.

In the following, each phase is described in detail in terms of specific objectives, methods and outputs.

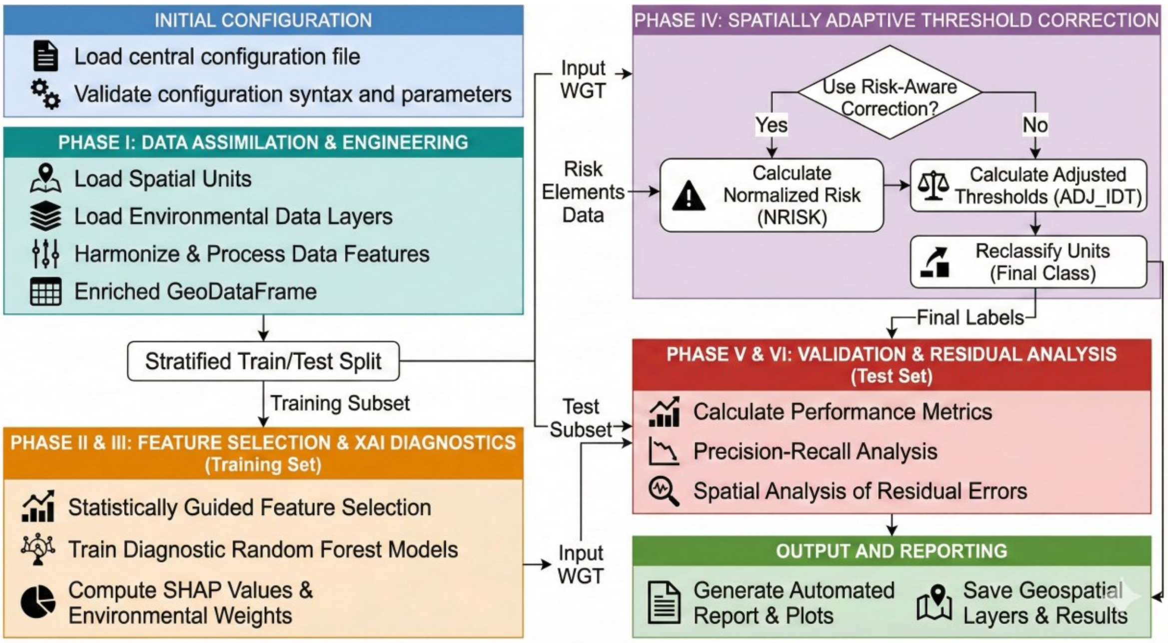


Figure 9. *Flowchart of RE-ACT procedure. The diagram illustrates the computational workflow for optimising the initial classification threshold. It is organised into six phases: data assimilation (I), statistically guided feature selection (II), XAI diagnostics (III), spatially adaptive threshold correction (IV), validation (V), and residual error analysis (VI). This procedure transforms a globally calibrated threshold into spatially adaptive thresholds to correct FP and FN errors*

Phase I: Data Assimilation and Feature Engineering

Objective: To construct a geospatial dataset in which each spatial unit is enriched with a coherent set of validated environmental attributes, ready for statistical analysis and machine learning.

This phase is implemented by the `EnvironmentalProcessor` class and proceeds as follows:

1. Geometry loading and validation: The spatial units (e.g. slope units or basins) are loaded using `geopandas`. The `ModelConfig` object checks that the input file exists and that critical columns (`BASIN_ID`, `error_column`, `count_column`, and optionally `risk_elements_column`) are present. The method `_validate_geometries` then performs a robust topological cleaning:
    - Invalid or empty geometries are detected and repaired using a combination of `geom.buffer(0)` and, where available, `shapely.validation.make_valid`;
    - Units whose geometry cannot be repaired are discarded; the number of removed units is logged;
    - An `area_tot` attribute is computed for each retained unit, used later to normalise categorical fractions;

   The coordinate reference system (CRS) of the units is checked. If missing, it is assigned using `basins_crs` (or `default_env_crs`). If a `target_crs` is specified, geometries are reprojected accordingly. This CRS is stored as the **master CRS** (`master_crs`) for all subsequent operations.

2. Harmonisation of raster data: The system harmonises continuous rasters on a common grid through the following steps:

  - A "template" raster is selected via the `align_to` parameter (e.g. the DEM). If necessary, its CRS is defined using `env_crs` or `default_env_crs`;
  - Using `rasterio` and `calculate_default_transform`, the template is reprojected to the master CRS, with a resolution equal to `target_resolution` (if provided) or its native resolution;
  - For each continuous variable, `_prepare_raster_for_alignment` reprojects the source raster to the template grid, ensuring common CRS, extent and pixel alignment. The harmonised rasters are stored in a temporary folder within the output directory.

3. <u>Processing of continuous variables:</u> Continuous environmental variables (e.g. slope, elevation, root cohesion) are processed by `_process_continuous`:

   - For each harmonised raster, `rasterstats.zonal_stats` computes the mean value within each spatial unit;
   - Missing values are imputed using the median of observed values;
   - For specific variables (e.g. `rootC`), physically implausible values are clipped (e.g. negatives are set to zero);
   - The resulting values are stored as additional numeric columns (e.g. `slope`, `dem`, `rootC`) in the GeoDataFrame.

4. <u>Processing of categorical variables:</u> Categorical variables, such as lithology and land use, are handled by `_process_categorical`:

   - The corresponding vector maps are loaded and, if necessary, assigned a CRS using `env_crs` or `default_env_crs`, then reprojected to the master CRS;
   - Raw category codes (e.g. lithology classes `1-7`, land-use descriptions) are mapped to standardised internal categories (e.g. `lt_1,…,lt_7`, `clc_1,…,clc_10`) via `category_maps`;
   - A spatial intersection between the categorical map and the units is computed using `geopandas.overlay`, and *the area of each category within each unit* is calculated;
   - Areas are normalised by `area_tot` to yield fractions (between 0 and 1) representing the proportion of each category in each unit;
   - the method `_finalize_data` ensures that:

     - All expected categorical columns (e.g. all `lt_*` and `clc_*` classes) are present, adding columns filled with zero where a class is absent;
     - Fractions are clipped to [0, 1], missing values are set to 0, and fractions for each prefix group (e.g. lithology, land use) are renormalised so that they sum to 1 within numerical tolerance.

Outcome: A single, clean `GeoDataFrame` in the master CRS, where each row corresponds to a spatial unit and contains:

a) Its unique identifier (`BASIN_ID`), geometry and area;
b) The initial classification (`TP`, `TN`, `FP`, `FN`);
c) The raw model score (`count_column`);
d) The environmental predictors: continuous attributes (e.g. `slope`, `dem`, `rootC`) and fractional covers (e.g. `lt_1,…,lt_7`, `clc_1,…,clc_10`).

This enriched dataset forms the basis for the subsequent statistical and XAI analyses.

Phase II: Statistically Guided Feature Selection

Objective: To identify the environmental variables that exhibit a statistically robust relationship with classification errors, thereby improving interpretability and reducing noise in the diagnostic models.
This phase is handled by the `perform_statistical_tests` method of `EnvironmentalProcessor`. It operates exclusively on the training subset (see Phase V) to avoid information leakage into the test set.
If the flag `disable_statistical_filters` is set to `True`, this phase is bypassed and all environmental features are passed to the subsequent steps. Otherwise, the following procedure is applied:

1. Separation by error block: The data are partitioned into two "error blocks":
   - **FP/TN block**: units labelled as `FP` or `TN`;
   - **FN/TP block**: units labelled as `FN` or `TP`.

   This allows the framework to detect features that are specifically associated with false positives or false negatives, rather than errors in general.

2. Tests for categorical variables: For each categorical fractional variable (e.g. `lt_1`, `clc_5`), a binary notion of "presence" is defined by a minimum fractional area (`min_fraction_area`). The following tests are then applied:
   - **2×2 contingency tests**: For each error block, a contingency table is constructed comparing "presence" vs. "absence" of the category and error vs. correct classification (e.g. FP vs. TN).

     Two alternatives are considered:

     - If expected or observed counts are small, a *Fisher test* is used;
     - Otherwise, a *Chi-square test of independence* is applied.

     The effect size is measured via Cramér's V (equivalent to $|\phi|$ for 2×2 tables).

   - **Mann-Whitney tests on fractions**: For robustness, the exact fractional values are also compared between error and correct units using the *Mann-Whitney U test*, with *Cliff's* δ as effect size.

   For a given variable and error block, the most informative test (lowest p-value) is retained. To ensure practical relevance, a "gate" is imposed: a variable is considered only if the category attains either a minimum prevalence (`min_prevalence`) or a minimum absolute count (`min_abs_count`) within the block.

3. Tests for continuous variables: For continuous environmental variables (e.g. `slope`, `dem`, `rootC`), the distributions in error vs. correct units are compared using the *Mann-Whitney U test* in each block (FP vs. TN, FN vs. TP). The effect size is again quantified using *Cliff's* δ.

4. Multiple testing and selection: For each error block (FP and FN) separately, the p-values are adjusted using the *Benjamini-Hochberg false discovery rate (FDR)* procedure. A variable is retained for that block if:
   - Its FDR-corrected *q-value* is below `statistical_test_p_value`;
   - The effect size exceeds `min_cramers_v` (for categorical variables) or `min_cliffs_delta` (for continuous variables);
   - For categorical variables, the prevalence/count gate is satisfied.

The full table of tests, effect sizes and decisions is written to `statistical_tests.csv`, and an extended file `statistical_tests_all_runs.csv` accumulates results across all validation runs. If no feature passes the criteria, the framework reverts to using all available features, and a warning is issued.

Outcome: A statistically subset of predictors, potentially different for FP and FN, that are demonstrably associated with misclassification patterns and are ready for use in the XAI diagnostic models.

Phase III: Explainable AI (XAI)

Objective: To quantify, in an interpretable way, the contribution of each environmental variable to the propensity of a spatial unit to be misclassified as FP or FN.

This phase is implemented in the `ModelEngine` class.

1. Diagnostic Random Forest models: For each error type, a separate **binary Random Forest classifier** is trained on the training subset:

    - **FP-diagnostic model**: discriminates between `FP` and `TN` units;
    - **FN-diagnostic model**: discriminates between `FN` and `TP` units.

    The predictors are: the statistically selected features for the corresponding error block (if available), or the global set of significant features, or all environmental variables, in a robust fallback mode. The Random Forest hyperparameters are tuned via *an internal grid search with cross-validation*, constrained by the minimum number of examples per class to prevent over-fitting.

2. SHAP-based feature importance: To open the “black box” of the Random Forest, RE-ACT computes SHAP values using the `shap` library:

    - SHAP values are obtained for each prediction and each feature, for both FP and FN diagnostic models;
    - The mean absolute SHAP value for each feature is used as a measure of its global importance;
    - In case SHAP computation is not feasible (e.g. degenerate datasets), the pipeline falls back to normalised *Gini importance from the Random Forest*.

    Crucially, the implementation ensures that the final importance dictionary includes all environmental variables present in the dataset: variables not selected or not used by the diagnostic model are explicitly assigned an importance of zero, so that preliminary but non-informative predictors still appear in the plots for transparency.

3. Definition of the Environmental Weight (WGT): For each spatial unit, the SHAP-based importances are combined into a *scalar environmental weight*, WGT, that quantifies the propensity of the unit to be misclassified:

    - A separate weight is computed for FP and FN models, based on the environmental profile of each unit (continuous variables are normalised; categorical fractions are used directly);
    - Only misclassified units (FP or FN) receive non-zero raw weights; correctly classified units (TP, TN) are assigned a baseline `wgt_base_nonerror` (typically zero).

    The vector of raw weights is then *normalised* using a user-selectable method (`wgt_normalization`), with `per_class_minmax` as default:

- In `per_class_minmax`, FP and FN weights are individually rescaled between 0 and 1 within each error class, while TP/TN retain the baseline value;
- In `robust_minmax` and `zscore_soft`, extreme values are clipped to user-defined quantiles (`wgt_q_low`, `wgt_q_high`) to limit the influence of outliers.

Outcome: *A per-unit environmental weight*, WGT∈[0, 1], for each error type (FP, FN), which measures how strongly the local environmental configuration is associated with the occurrence of that error, according to the diagnostic model.

Phase IV: Spatially Adaptive Threshold Correction

Objective: To exploit the diagnostic information produced in Phase III in order to derive unit-specific IDT thresholds, so as to reduce both False Positives (FP) and False Negatives (FN) while preserving the original logic of the susceptibility-based classification model. In Phase IV, the global threshold is replaced by a spatially adaptive threshold, computed independently for each unit, in order to account for local environmental controls that systematically influence misclassification patterns.This phase is implemented through the functions `correct_errors` and `_calculate_adjusted_idt` within ModelEngine, which calculate the adjusted threshold ADJ_IDT(b) for each spatial unit *b* and then apply the updated thresholds in the final classification stage:

1. Risk-neutral correction: When the risk-aware option is disabled (`use_risk_score = False`), the framework applies a correction based exclusively on environmental information. The fundamental assumption is that classification errors generated by the baseline model are not randomly distributed, but are instead associated with specific combinations of environmental conditions (e.g., lithology, land use, morphology) that the physical model may not fully represent. In this setting, the correction strategy modifies the IDT threshold by using the environmental error weights estimated in Phase III through the XAI model, using the following equations:

   - **For False Positives (FP)**:

$$ADJ_{IDT(b)} = IDT_{orig} + (WGT_{FP}(b) \times \Delta_{maxFP})$$

[3.10]

Where: ADJ_IDT(b) is the new, adjusted threshold computed for spatial unit b; IDT_orig is the initial, static threshold applied uniformly to all spatial unit; $WGT_{FP}(b)$ *is the Environmental Weight represents the environmental propensity of unit b to behave as FP;* Δ_{maxFP} *is the user-defined maximum allowable increase of the threshold, acting as a control parameter;*

This correction increases the minimum number of unstable pixels required to classify the unit as unstable. As a consequence, spatial units with a higher FP propensity require stronger evidence of spatially distributed instability before an alert is triggered, thereby reducing false alarms;

   - **For False Negatives (FN):**

$$ADJ_{IDT(b)} = IDT_{orig} - (WGT_{FN}(b) \times \Delta_{maxFN})$$

[3.11]

Where: $WGT_{FP}(b)$ *is the Environmental Weight represents the environmental propensity of unit b to behave as FN;* Δ_{maxFN} *is the user-defined maximum allowable decrease of the threshold;*

This correction lowers the unstable-pixel requirement for classification. Operationally, spatial units with a higher FN propensity are allowed to reach an unstable classification even when instability is expressed through a smaller number of unstable pixels, thus increasing sensitivity and reducing missed events;

2. Risk-aware correction: When the risk-aware option is enabled (`use_risk_score = True`), the framework modulates the environmental correction by incorporating a unit-specific risk factor NRISK(b).
   The rationale is that the same correction magnitude may be acceptable in units with limited exposure, while it may be inappropriate or unsafe in units where the potential consequences of misclassification are higher. The normalised risk score is defined as NRISK(b)∈[0, 1], and is derived from the count of exposed elements (`risk_elements_column`) through a robust transformation (logarithmic scaling followed by min–max normalisation), with explicit treatment of zero values and constant distributions. This formulation ensures that risk variability across units is represented consistently, without being dominated by extreme values:

   - **For False Positives (FP):**

$$ADJ_{IDT(b)} = IDT_{orig} + [(WGT_{FP}(b) \times \Delta_{maxFP} \times (1 - \mathrm{NRISK}_b)] \qquad [3.12]$$

where: NRISK$_b$ is the normalised risk score for spatial unit b, which moderates the correction;

Here, (1 − NRISK_b) attenuates the correction in highly exposed units. This produces a conservative behaviour in high-risk areas: as NRISK_b → 1, the correction magnitude tends to zero, limiting the increase in IDT and therefore preventing the system from becoming overly permissive in locations where false reassurance would be operationally critical.

   - **For False Negatives (FN)**:

$$ADJ_{IDT(b)} = IDT_{orig} - [(WGT_{FN}(b) \times \Delta_{maxFN} \times (1 + \mathrm{NRISK}_b)] \qquad [3.13]$$

In FN units, the factor (1 + NRISK_b) amplifies the correction. This means that in highly exposed areas the framework further lowers the IDT threshold, increasing the probability that a potentially critical instability condition will be flagged. Consequently, as NRISK_b increases, the correction becomes more stringent in terms of safety, prioritising the recovery of missed events.

Once the adjusted thresholds have been computed, the method `final_reclassification` applies the updated thresholds ADJ_IDT(b) to recompute the final class of each unit. The classification rule is:

- A unit is predicted as “Landslide” if the unstable-pixel count within the unit is ≥ ADJ_IDT(b);
- Otherwise, the unit is predicted as “No Landslide”;

The four-class label is then updated and stored as `FINAL_ERR`, allowing the framework to track how the correction affects the original error categories. In particular:

- **FP → TN** when the adjusted threshold eliminates a false alarm;
- **FN → TP** when the adjusted threshold recovers a previously missed event.

Finally, the variable `STATO_FINALE` summarises whether the unit remains unchanged or transitions from FP to TN or from FN to TP.

Outcome: The procedure produces a spatially adaptive and unit-specific threshold ADJ_IDT(b) (optionally modulated by exposure-based risk). This enables systematic reduction of residual misclassifications through a controlled threshold adjustment mechanism that remains consistent with the structure of the original susceptibility-based decision rule.

Phase V: Validation and Performance Metrics

Objective: To rigorously evaluate the robustness and generalizability of the correction pipeline through repeated hold-out validation and an independent test set.

1. Stratified train-test splits and repeated hold-out: The complete dataset of spatial units is divided into a training subset and a test subset using `sklearn.model_selection.train_test_split`. The split is configured so that:

    - The training fraction is equal to 1 − holdout_test_size;
    - The split is stratified by the initial four-class label (TP, TN, FP, FN), meaning that each subset preserves approximately the same proportion of true positives, true negatives, false positives, and false negatives observed in the full dataset. This stratification is essential to avoid biased performance estimates caused by unbalanced class distributions.

    This train/test splitting procedure is repeated for every random seed listed in `holdout_random_states`, so that the evaluation is not dependent on a single partition of the data. For each repeated run:

    - Phase II (statistical screening) and Phase III (XAI diagnostics) are executed only on the training set, so that all modelling choices are learned without access to the test data;
    - The threshold correction (Phase IV) is then applied to all spatial units, using the correction rules derived from the training phase;
    - Performance is evaluated only on the test set, both before correction (baseline configuration) and after correction (RE-ACT adjusted configuration), ensuring that the baseline and corrected configurations are evaluated under the same independent test conditions;

2. Performance metrics: For each run, classification performance is computed through a dedicated function (`calculate_metrics`) based on the confusion matrix derived from the four-class labels. The reported metrics include: Overall accuracy, which measures the proportion of correctly classified units; Recall (true positive rate) with respect to landslide occurrence, which quantifies how well the system identifies areas where landslides actually occurred; Precision, which measures the fraction of predicted landslide units that are truly landslide units; F1-score, which balances precision and recall in a single indicator; Balanced Accuracy (BA), computed as the average between the true positive rate and the true negative rate, and therefore useful when class proportions are uneven.

    For each run, these metrics are stored twice: a) For the initial configuration (before RE-ACT correction) evaluated on the test set; b) for the final configuration (after correction) evaluated on the same test set.

    All repeated-run results are collected in a single summary table (`robustness_metrics.csv`). In addition to reporting the metric values, the table includes, for each metric, the absolute difference and percentage difference with respect to the reference run (defined as the run associated with the first seed);

3. Precision-Recall analysis: To complement the metrics that depend on a fixed decision rule, an additional Precision–Recall (PR) analysis is performed to describe performance in probabilistic terms. The procedure is structured as follows:

   - An auxiliary Random Forest classifier is trained on the training set to discriminate between: “landslide units”, where a landslide actually occurred (TP and FN), and “non-landslide units”, where no landslide occurred (TN and FP);
   - The model outputs predicted probabilities on the test set, and these probabilities are used to compute: the Precision–Recall curve, and the Average Precision (AP) score, which summarises the overall PR behaviour as a single value.

     The resulting PR curve is saved as `precision_recall_final`. This provides a benchmark for the separability achievable with a probabilistic classifier, against which the fixed IDT-based decision rule can be compared in the operational pipeline.

4. Integrated performance report: For the reference run (first seed), the framework produces a consolidated text summary (`performance_report.txt`) that collects the main validation outputs in a single readable format, including:

   - Train/test sizes and hold-out parameters;
   - Initial vs. final metrics on the test set (accuracy, recall, precision, F1, BA);
   - Percentage improvements after correction;
   - Number and percentage of units corrected from FP→TN and FN→TP;
   - Global Moran’s I result (see Phase VI);
   - The list of random seeds used for robustness analysis.

Outcome: A *multi-run, hold-out based validation* that quantifies the gain in predictive performance and its stability across different random partitions of the data.

Phase VI: Residual Errors

After applying the correction procedure, a fraction of misclassifications can still remain. For this reason, Phase VI carries out a spatial diagnostic analysis to understand whether these residual errors occur as isolated cases across the study area, or whether they still follow a structured spatial pattern:

1. Residual error indicator and spatial weights: The `ResultVisualizer` class constructs a binary residual-error indicator as follows:
   - $e_i = 1$ for units that are still misclassified after correction (`FINAL_ERR` ∈ {FP, FN});
   - $e_i = 0$ for units that are correctly classified after correction (`FINAL_ERR` ∈ {TP, TN}).

   To investigate whether residual errors show spatial organisation, the pipeline then constructs a Queen-contiguity spatial weights matrix using `libpysal.weights.Queen.from_dataframe`. Under Queen contiguity, two spatial units are considered neighbours if they share either a boundary segment (edge) or a corner (vertex). This adjacency definition is appropriate for polygon-based spatial units, because it captures both direct contacts along edges and corner connections between neighbouring polygons.

2. Moran’s I analysis: Using `esda.moran.Moran`, the framework computes a global spatial autocorrelation test on the residual error indicator. The reported quantities consist of:

   - Moran’s I statistic, which indicates whether residual errors tend to be spatially clustered or spatially dispersed;

- Permutation-based pseudo p-value (p), which evaluates whether the observed Moran's I could plausibly arise by chance under spatial randomness;
- Standardised z-score, expressing the deviation from spatial randomness in standardised units.

The results are also examined using a Moran scatterplot (`splot.esda.moran_scatterplot`), which represents standardised residual errors against their spatial lag (i.e., the average value of neighbouring units). This plot supports interpretation by showing whether misclassifications tend to occur in spatial proximity.

Outcome: A spatial diagnostic framework that identifies whether residual errors are randomly distributed or still show systematic clustering, providing guidance for further model refinement and for targeted improvements in data and predictors.

3.5 Runout Model Calibration: The Source Evaluation and Rheological Calibration Hub (SEARCH)

This section describes the methodological pipeline developed to calibrate complex runout models using automatic optimisation algorithms.

The traditional back-analysis approach to calibration is often slow, subjective, and inefficient when it comes to exploring the large, strongly coupled, and high-dimensional parameter space of physically based models.

This limitation is further amplified by a very common and critical issue in landslide inventories (especially those derived from field surveys and public datasets): historical landslide records frequently consist of a single areal polygon that merges the source area and the runout/deposition zone, without providing any separation between the two. This creates both a practical ambiguity and a methodological bottleneck.

On this basis, a standard propagation model cannot be calibrated using this 'areal' information alone, because it requires a defined source area to initialise the simulation.

The manual delineation of the source area, for instance by applying a fixed slope threshold, is not an appropriate solution. It is inherently subjective and arbitrary, because the chosen delineation rule introduces an unknown and unquantified user bias before the optimisation even begins.

In addition, this approach is not scalable, because manually defining sources for hundreds or thousands of events is prohibitively time-consuming and often inconsistent across the dataset.

Importantly, the final calibrated rheological parameters are highly sensitive to this initial source geometry. If the source is defined in an arbitrary way, the calibration is inevitably biased, and this can force the optimisation towards non-physical parameter values regardless of the sophistication of the optimisation algorithm.

The calibration workflow is therefore built around the Source Evaluation and Rheological Calibration Hub (SEARCH), which was developed specifically to resolve this source-area ambiguity.

Within SEARCH, runout simulation is used as an evaluation tool to test the plausibility of each candidate source geometry. This shifts source delineation from a subjective pre-processing choice to a central, optimisable component of the calibration workflow, through a hierarchical optimisation strategy that is designed to determine, at the same time, two distinct sets of parameters:

- The model's Global Rheological Parameters (e.g., `friction_factor`, `range_angle`, `mass_to_drag`), which are assumed to be constant across the entire study area;

- The Local Geometric Parameters (e.g., slope criteria, consolidation methods, offset cells), which are optimised separately for each landslide and control how the source area is extracted from the observed landslide polygon.

To evaluate one set of global physical parameters (e.g. P_1 = [range_angle = 0.6; friction_factor = 0.1; mass_to_drag = 5.5]), the framework must first obtain the best local source geometry that can be paired with it. This is achieved by launching an internal source-shape optimisation loop in which P_1 is kept fixed, while a set of candidate sources (Source A, Source B, …) is generated and tested in a systematic way.

For each candidate source, the external runout model is fully executed using the same global parameters, Source A + P_1, Source B + P_1, etc. The continuous model output, such as velocity, flow probability or flow depth depending on the selected runout model, is then converted into a simulated runout polygon, and a composite performance score is computed using multiple evaluation metrics.

Among all tested candidates, only the best-performing source geometry (e.g. Source B + P_1) is retained, and the score of that winning simulation becomes the official score used to assess P_1 as a global parameter vector. In this way, global parameters are always evaluated together with their own optimised source geometry, rather than under an arbitrary source defined by the user, which ensures that different parameter sets are compared in an objective and reproducible manner.

Two key methodological principles underpin the SEARCH module: Bayesian Optimisation to explore the global parameter space efficiently, and K-Fold Cross-Validation to obtain robust out-of-sample performance estimates:

- ✓ Bayesian Optimisation: The main optimisation loop (Phase III) adopts an "intelligent search" strategy to identify the best global set of physical parameters, P_{best}, that maximises the mean performance across all landslides in the calibration subset. It progressively builds and updates a statistical surrogate model, a Gaussian Process, that maps global parameters to their average performance score. At each iteration, the surrogate proposes new promising parameter combinations, which are then evaluated through the hierarchical simulation described above, including the internal source optimisation loop;

- ✓ K-Fold Cross-Validation: To ensure that the calibrated global parameters, P_{best}, are robust and generalisable, the historical landslide inventory is divided into K distinct, spatially-aware folds. For each fold, one subset of landslides is reserved as validation data, while the remaining events form the calibration subset. The entire hierarchical optimisation cycle, inner source optimisation + outer Bayesian search, is repeated K times, producing a fold-specific P_{best} together with the corresponding set of optimised source geometries. Performance metrics are always computed on the fold that was held out and never used during optimisation, providing a valid out-of-sample assessment.

With the combination of these techniques, the SEARCH framework provides a scalable, objective, reproducible, and computationally efficient solution for the hierarchical calibration and validation of physically based runout models.

It should also be noted that an ideal back-analysis for propagation modelling would be better constrained by direct observations of runout dynamics, including flow velocity, flow depth, travel time, or impact pressure. These data would provide stronger constraints on the rheological parameters that control flow mobility, energy dissipation, and velocity evolution. However, such dynamic observations are rarely available in regional landslide inventories, which typically consist of static landslide polygons rather than time-dependent information on landslide motion.

To address these data limitations, the SEARCH framework adopts a spatial back-analysis strategy tailored to the information realistically available at the regional scale. Calibration is performed by evaluating the agreement between simulated and observed runout geometries through spatial-overlap, distance, and geometric performance metrics. Consequently, the resulting parameter set is optimized to replicate the observed spatial extent and geometry of the landslide, while the simulated velocity field is subsequently utilized as the intensity input for impact and risk assessments. An important architectural aspect of SEARCH is its ability to run different external propagation models within the same calibration workflow. In this thesis, DebrisDice (Bregoli et al., 2018) is selected as the external runout model because its stochastic formulation and computational efficiency allow thousands of simulations to be executed within practical computational times.

However, the workflow is not specific to DebrisDice. Other propagation models, such as Flow-R, RAMMS, GRASS-based workflows or r.randomwalk, can also be used, provided that they can be launched automatically and that their input and output files can be described in the configuration file (`config.yml`).

The configuration file defines which parameters are calibrated, how the input files are created, how the external model is called, and which output files are read to compute the calibration metrics.

The optimisation search space (`optimizer_space`) defines parameter names and admissible ranges (for example, friction angle or viscosity), and input generation (`file_templates`) is specified through dynamic text templates filled at runtime.

The same configuration also defines the execution logic (`command_template`), describing how the target executable (e.g., flowr.exe, DebrisDice.exe) is called with the correct flags, and it defines output interpretation (`output_config`), describing how simulation outputs are parsed and converted into spatial data for calibration. SEARCH is built on an abstract RunoutModel interface supported by concrete implementations: GenericModel (for command-line executables, such as RAMMS or Flow-R) and GrassModel (for GRASS-based workflows, such as r.randomwalk). At runtime, the appropriate implementation is instantiated based on `model_name`. To avoid relying on a single overlap measure, SEARCH adopts a multi-dimensional evaluation strategy designed to capture geometric differences and shape similarity. Vector-based metrics such as IoU, precision, recall, F1-score, relative area error, scaled Hausdorff distance, and Cohen's κ are computed from the simulated and observed polygons, while threshold-independent raster metrics (ROC AUC and Precision–Recall AUC) are computed on a regular grid within a configurable Region of Interest (ROI).

In this workflow, the ROI is defined as a limited spatial window, typically a buffer around the observed landslide polygon, and it can optionally be expanded to include the simulated footprint, so that pixel-based quantities are evaluated only where meaningful.

By restricting raster-based analysis to this ROI, the framework reduces background bias, i.e., the artificial inflation or degradation of performance scores that would occur if large stable areas far from the landslide were included. To enable automatic optimisation, all metrics are standardised onto a common direction where higher values always correspond to better performance, and they are aggregated into a single objective function using user-defined weights. Finally, SEARCH supports the validation of calibration results through bootstrap confidence intervals and summary HTML reports, providing a quantitative assessment of parameter uncertainty.

3.5.1 Input Data and Initial Configuration

The entire SEARCH workflow is orchestrated by the main script (`SEARCH.py`) and controlled through a single YAML configuration file (`config.yml`).

The user interacts almost exclusively with this file, which encapsulates all scientific choices (model parameters, metrics, validation strategy) and all technical details (paths, logging, plotting) without requiring any modification of the Python code. Geospatial inputs: the framework requires a small but essential set of datasets, with their paths specified in the paths section of `config.yml`:

- Historical landslide inventory (`paths.sources`): a vector layer (e.g. ESRI Shapefile or GeoPackage) containing polygons that represent the observed landslide runout footprints. Each feature corresponds to a single landslide and acts as a record with at least a unique identifier (e.g. *id*);
- Digital Elevation Model (`paths.dem`): a gridded DEM (preferably a GeoTIFF dataset), providing the topographic basis used for all simulations and slope estimates.

This same set of paths also defines the filesystem locations that the workflow relies on, as follows:

- A model working directory (`paths.model_working_dir`), where temporary, job-specific files generated during the simulation runs;
- A primary output directory (`paths.output_folder`), where final results, logs, tables, plots and GeoPackages are stored.

The `model_params` section is designated to save model-specific settings. For example, in this work and for the DebrisDice case study, the block set `DebrisDice_config` defines: (i) the global physical properties to be optimised (e.g., `friction_factor, range_angle, minimum_slope_angle, mass_to_drag, thr_threshold`), (ii) the text templates used to generate the DebrisDice input file at runtime (`input.dat`) via parameter placeholders, (iii) the command line required to invoke the target executable (DebrisDice.exe), and (iv) the mapping between logical output names (e.g., `profile_velocity`) and the corresponding physical output files (i.e., `Velocity.txt`), together with standard parsing instructions.

The setup is then parsed and validated against a defined schema at startup (a base `BASE_SCHEMA` and the model-specific schema for GenericModel or GrassModel), and saved in a typed Config object. This means that, once loaded and checked, the configuration becomes a structured object that provides direct access to all sections (`model_name`, `model_config`, `paths`, `processing`, `weights`, `bayes`, `validation`, `logging`, `plotting`).

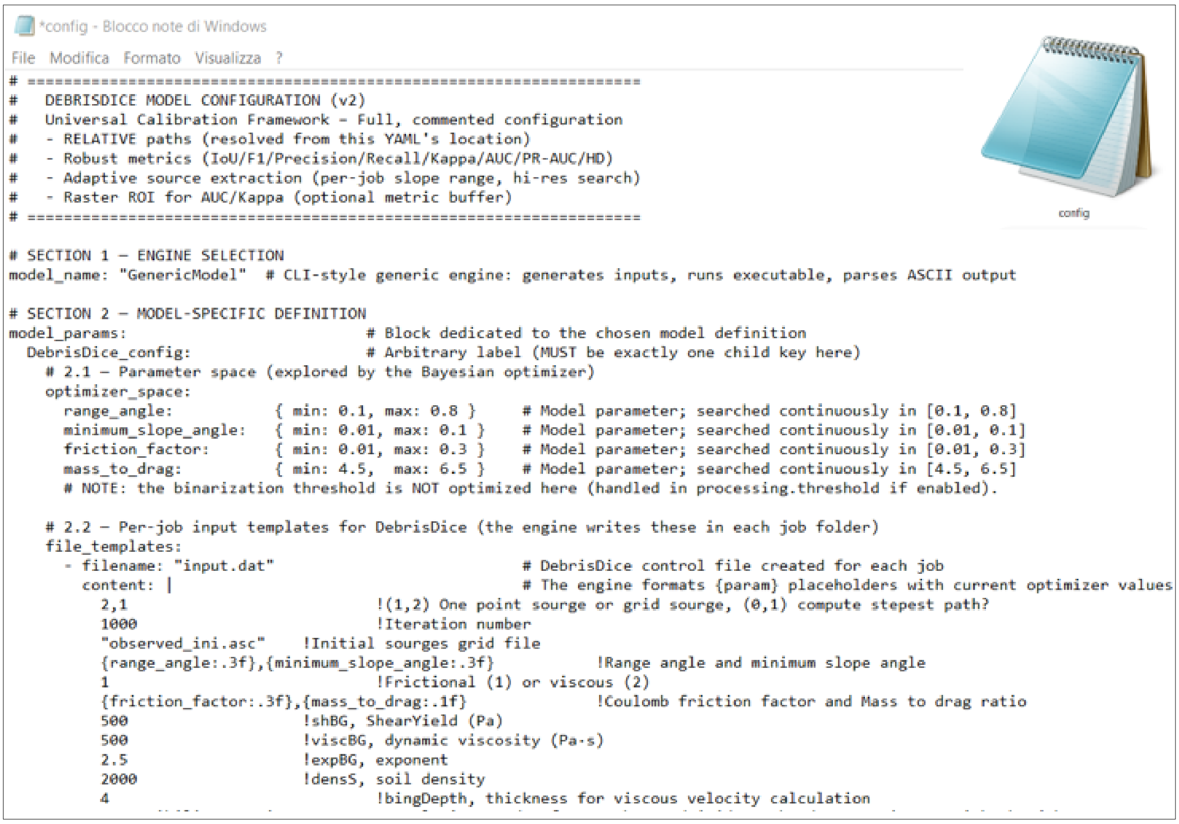
The structure and meaning of the primary configuration sections for this study are summarised in *Table 11*, while a version of the `config.yml` file is shown in *Figure 10*.

Operationally, when interfacing a new runout model via an appropriate `model_params` block, the SEARCH procedure for a new case study typically translates into: (i) updating the paths to the DEM and the inventory of landslides, (ii) defining the optimal `optimizer_space` for the model's physical parameters, and (iii) tuning the metric weights and validation strategy. Any other advanced options (such as resampling, source-shape search, ROI-based metrics, bootstrap settings) can generally be kept at their defaults, and then refined progressively as more experience is gained with the framework.

Table 11. *A comprehensive dictionary of the SEARCH framework's configuration parameters as defined in the config.yml file, with a concise description of their respective role*

Section	Parameter	Detailed Description
Model engine selection		
	`model_name`	Selects which model engine to use (e.g. `GenericModel`, `GrassModel`)
Model-specific definition		
	`model_params`	Top-level block containing the chosen model's specific configuration
	`optimizer_space`	Defines which global parameters are optimised and their min–max bounds
	`file_templates`	Templates for input files generated at runtime with parameter placeholders
	`command_template`	Command line to run the model executable and its arguments
	`output_config`	Specifies target output file(s) and format used to compute metrics
Paths		
	`sources`	Path to the landslide inventory vector layer (with polygon IDs)
	`dem`	Path to the main DEM used for slope analysis and simulations
	`output_folder`	Directory where all outputs (logs, tables, plots, GeoPackages) are saved
	`model_working_dir`	Root directory for temporary, per-job working subfolders
Processing - global options		
	`thr_threshold`	Global threshold to binarise model output into runout / non-runout
	`threshold`	Controls if and how the threshold is optimised or set per fold
	`buffer_distance` / `use_buffer_for_metrics`	Optional buffer applied to observed polygons when computing metrics

Section	Parameter	Detailed Description
	`auto_clip_buffer_m`/ `job_clip_buffer_m`	Distances used to clip DEM around the inventory and each event
Processing - source delineation		
	`source_interior_offset_cells`	Number of DEM cells eroded inside the footprint to position the source
	`source_area_consolidation`	Rule to merge fragmented source patches into a single polygon
	`min_source_area_cells`	Minimum number of DEM cells required for a valid source
	`source_area_cap`	Maximum source area as a fraction of the observed landslide
	`source_delineation_fallback_radius_m`	Radius for a fallback circular source if slope-based delineation fails
Processing - slope range and DEM handling		
	`slope_range_mode`	Method to define the slope range used for source carving
	`slope_percentiles`	Percentiles used to define slope range when using percentile mode
	`adaptive_slope_range_per_job`/ `adaptive_slope`	Enables per-event refinement of the slope range with safeguards
	`hires_source_search`	Controls use of higher-resolution DEMs for source delineation
	`resampling`	Settings for automatic DEM resampling and target resolutions
Processing - source shape search and volume		
	`local_1m_buffer_m`	Additional local buffer used when aligning DEM and polygons
	`source_shape_search`	Parameters controlling the inner source-shape optimisation loop
	`source_volume`	Settings for optional volume-based constraints on source size
Metric weights		
	`weights`	Weights assigned to each metric in the composite objective score
Optimisation and validation		
	`bayes`	Controls Bayesian optimisation (n. of calls, initial points, checkpointing)
	`validation`	Defines validation method, K-Fold/hold-out settings and ROI/CI options
	`metrics_roi`	Optional ROI definition for pixel-based metrics and AUC computation
	`spatial_block_kfold`	Enables spatially grouped K-Fold to limit spatial leakage
Logging and plotting		
	`logging`	Controls log verbosity, console output and log file location.
	`plotting`	Enables plots and sets basic style options (DPI, fonts, etc.)



```
# ====================================================================
#   DEBRISDICE MODEL CONFIGURATION (v2)
#   Universal Calibration Framework - Full, commented configuration
#   - RELATIVE paths (resolved from this YAML's location)
#   - Robust metrics (IoU/F1/Precision/Recall/Kappa/AUC/PR-AUC/HD)
#   - Adaptive source extraction (per-job slope range, hi-res search)
#   - Raster ROI for AUC/Kappa (optional metric buffer)
# ====================================================================

# SECTION 1 – ENGINE SELECTION
model_name: "GenericModel"  # CLI-style generic engine: generates inputs, runs executable, parses ASCII output

# SECTION 2 – MODEL-SPECIFIC DEFINITION
model_params:                         # Block dedicated to the chosen model definition
  DebrisDice_config:                  # Arbitrary label (MUST be exactly one child key here)
    # 2.1 – Parameter space (explored by the Bayesian optimizer)
    optimizer_space:
      range_angle:          { min: 0.1, max: 0.8 }     # Model parameter; searched continuously in [0.1, 0.8]
      minimum_slope_angle:  { min: 0.01, max: 0.1 }    # Model parameter; searched continuously in [0.01, 0.1]
      friction_factor:      { min: 0.01, max: 0.3 }    # Model parameter; searched continuously in [0.01, 0.3]
      mass_to_drag:         { min: 4.5,  max: 6.5 }    # Model parameter; searched continuously in [4.5, 6.5]
      # NOTE: the binarization threshold is NOT optimized here (handled in processing.threshold if enabled).

    # 2.2 – Per-job input templates for DebrisDice (the engine writes these in each job folder)
    file_templates:
      - filename: "input.dat"                          # DebrisDice control file created for each job
        content: |                                     # The engine formats {param} placeholders with current optimizer values
          2,1                       !(1,2) One point sourge or grid sourge, (0,1) compute stepest path?
          1000                      !Iteration number
          "observed_ini.asc"    !Initial sourges grid file
          {range_angle:.3f},{minimum_slope_angle:.3f}          !Range angle and minimum slope angle
          1                         !Frictional (1) or viscous (2)
          {friction_factor:.3f},{mass_to_drag:.1f}             !Coulomb friction factor and Mass to drag ratio
          500               !shBG, ShearYield (Pa)
          500               !viscBG, dynamic viscosity (Pa·s)
          2.5               !expBG, exponent
          2000              !densS, soil density
          4                         !bingDepth, thickness for viscous velocity calculation
```

Figure 10. *The following example illustrates the configuration file structure for the SEARCH framework in relation to the DebrisDice model*

3.5.2 Generated Outputs

As a result of the workflow described above, the framework produces a coherent set of outputs that cover all levels of analysis: tabular files (CSV/XLSX) that summarise calibration history, validation metrics and recommended parameters, geospatial layers (GPKG) containing the optimised source areas and the simulated runouts, and graphical products (PNG and HTML) that document convergence, parameter sensitivity and overall performance in a publication-ready form.

Table 12 provides a concise overview of these outputs, specifying for each file its name, format and main content.

Table 12. *A comprehensive view of SEARCH framework generated outputs*

Output	**File name / pattern**	**Format**	**Description**
Run log	`run.log`	TXT	Full text log of the execution (info, warnings, errors)
Skipped landslides	`skipped_landslides`	CSV	List of events not processed and the corresponding reasons
Optimisation history (hold-out)	`Optimization_history`	CSV	All tested parameter sets and scores for simple hold-out
Optimisation history (K-Fold)	`crossval_optimization_history`	CSV	Same as above, aggregated for all folds with fold labels

Output	File name / pattern	Format	Description
Best parameters per fold	`crossval_best_params_per_fold`	CSV	Optimal global parameter set identified in each fold
Recommended best parameters	`selected_best_parameters`	JSON	Final recommended global parameters and primary selection metric
Per-event metrics (hold-out)	`Validation_metrics`	CSV	Validation metrics for each landslide in simple hold-out
Per-event metrics (K-Fold)	`crossval_validation_metrics_all_folds`	CSV	Validation metrics for all landslides across all folds
Metric summary	`validation_metric_summary`	CSV	Summary statistics (mean, median, etc.) for all validation metrics
Bootstrap confidence intervals	`validation_metric_bootstrap_ci`	CSV	Bootstrap-based confidence intervals for selected metrics
Parameter statistics across folds	`parameter_stats_across_folds`	CSV	Distribution of each global parameter over all folds
Calibration summary workbook	`Calibration_summary`	XLSX	Multi-sheet summary of parameters, metrics, history and settings
Simulated runouts (hold-out)	`validation_simulated_runouts`	GPKG	Final simulated runout polygons for hold-out validation events
Simulated runouts (K-Fold)	`crossval_simulated_runouts`	GPKG	Final simulated runout polygons for all K-Fold validation events
Optimised source areas (hold-out)	`validation_source_areas`	GPKG	Optimised source polygons for hold-out validation events
Optimised source areas (K-Fold)	`crossval_source_areas`	GPKG	Optimised source polygons for all K-Fold validation events
Convergence plot (hold-out)	`plot_convergence`	PNG	Objective score vs. iteration for simple hold-out
Convergence plot (K-Fold)	`plot_crossval_convergence`	PNG	Objective score vs. iteration for all folds combined
Parameter pair plots	`plot_parameter_pairs*`	PNG	Pairwise parameter plots coloured by objective score
Metric boxplots	`plot_validation_metrics_boxplot, plot_crossval`	PNG	Boxplots of validation metrics (hold-out and K-Fold)
Metric bar plots	`plot_validation_metric_bars`	PNG	Bar chart of mean ± std for main validation metrics
Parameter histograms	`plot_param_hist_<name>`	PNG	Histograms of explored values for each global parameter
Executive HTML report	`executive_summary` / `final_report`	HTML	Compact graphical summary of calibration results and diagnostics

3.5.3 The Computational Workflow

The framework, which operates through the automated pipeline depicted in the flowchart in *Figure 11,* executes a sophisticated computational workflow organised into six distinct phases.
The entire process is encapsulated within a K-Fold Cross-Validation loop to ensure that the final results are statistically robust and generalisable.

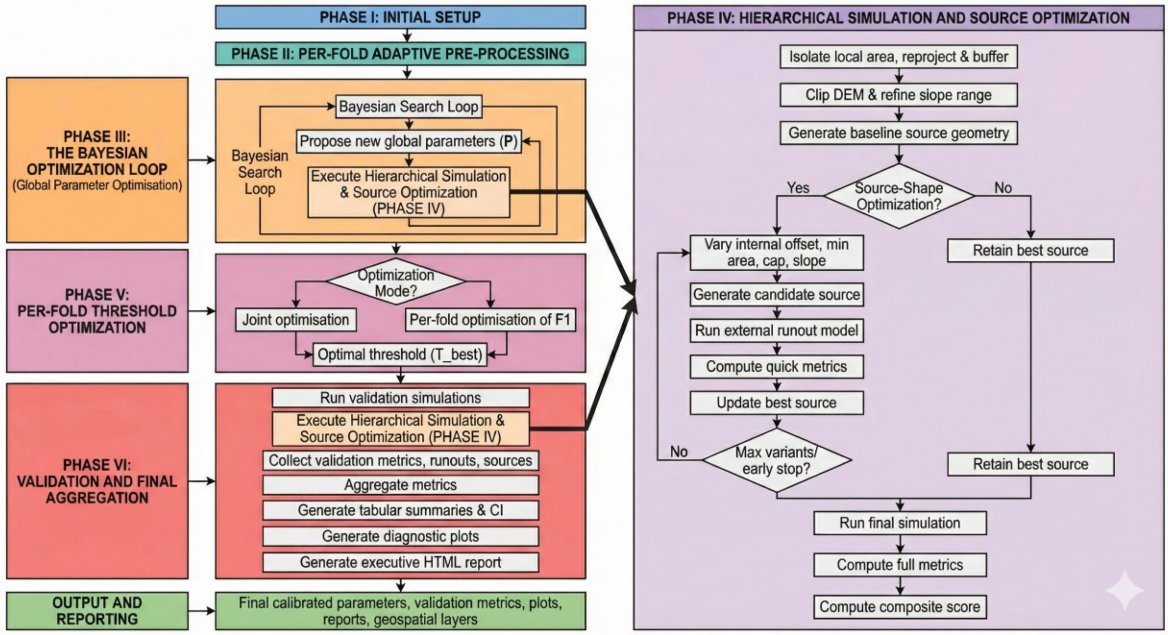


Figure 11. *Flowchart of the SEARCH framework. The diagram illustrates the computational workflow for automated hierarchical calibration. The left column displays the main sequence of six phases, including global Bayesian parameter optimisation and K-Fold validation. The right column details Phase IV, which performs the simulation and local source geometry optimisation, iteratively called from Phases III and VI*

Phase I: Initial Setup

This first step establishes the entire process, and enables it to proceed in a controlled, reproducible manner. The steps are:

- Configuration loading and schema validation: The `main()` function reads `config.yml` and validates it against `BASE_SCHEMA` and the appropriate model-specific schema (for `GenericModel` or `GrassModel`), before instantiating a `Config` object.
  Any missing or inconsistent parameters (e.g., incompatible bounds, missing paths) are detected at this stage and informative error messages are displayed;
- Environment setup: The script initialises the logging system (console output and file handler writing to run.log) and enforces conservative limits on BLAS/OMP threads (to avoid oversubscription and maintain robust behaviour when parallel simulation is used);
- Input data loading: The process involves reading the landslide inventory into a GeoDataFrame.
  When duplicate IDs are found, they are logged and written to a diagnostic GeoPackage for inspection. Subsequently, the system checks the DEM and converts it to GeoTIFF when necessary, ensuring an explicit CRS and aligned resolution. A DEM that has been clipped to the spatial extent of the inventory is prepared and its path is updated within the configuration (`paths.dem`);
- Model handler selection: The RunoutModel subclass (e.g. GenericModel for DebrisDice, based on `model_name`) is selected and its optimiser space (`optimizer_space`) is extracted. Legacy

aliases (e.g., `threshold → thr_threshold`) are normalised, and non-optimised parameters (e.g. the obsolete `slope_threshold`) are removed from the search space;

- Validation partitioning: The landslide inventory is partitioned into calibration and validation subsets according to `validation.method`. For K-Fold, a GroupKFold or KFold iterator is constructed, whereas a single random split produces a simple hold-out.

Outcomes: At the end of Phase I, SEARCH has a validated configuration, a harmonised DEM, a loaded inventory and a selected model engine, ready for fold-wise calibration.

Phase II: Per-Fold Adaptive Pre-processing

To avoid information leakage and to preserve the spatial structure of the data, all pre-processing operations that affect source-area parameterisation, such as the definition of `slope_range`, are applied independently within each calibration subset, whereas the alignment and resampling of the DEM are applied once on the entire inventory:

- Fold-specific slope range estimation: For each fold (or for the calibration subset in simple hold-out), `analyze_slope_distribution` computes a representative slope range using only the landslides included in the corresponding calibration subset.
  It first computes a robust slope raster from the DEM, and extracts slope pixels inside each observed polygon.
  Depending on `slope_range_mode`, the procedure either uses all pixels within the polygons (percentiles mode) or only the upper-elevation portion of each landslide (`highparts_minmax mode`), identified by masking pixels above the mean elevation of the polygon.
  The file `skipped_landslides.csv` logs events that are skipped (e.g. outside DEM extent, no valid slope pixels, or no elevation pixels).
  The parameter `processing.slope_range_deg` is defined based on the global min and max (or selected percentiles) across all calibration events for that fold;

- DEM resampling and alignment: The function `align_crs_and_resolution` ensures that the DEM and the inventory share the same metric CRS, assesses the characteristic spatial scale of landslides and, when needed, produces finer DEM versions at user-defined target resolutions (usually finer than the original one, e.g. 5 m, 2 m, 1 m).
  The resampled DEMs are cached on disk to avoid redundant processing. The aligned DEM is stored in the configuration and used as a reference grid for subsequent simulations.

Outcomes: This step ensures that slope criteria used for source delineation are derived exclusively from the corresponding calibration subset and that the DEM resolution is always appropriate for the characteristic landslide size.

Phase III: The Bayesian Optimisation Loop (Global Parameter Optimisation)

At the heart of the calibration is a Bayesian optimisation loop across the global parameter space:

1. Optimiser initialisation and objective function: An Optimiser object from `scikit-optimize` is initialised based on the bounds defined in `optimizer_space`.
   The objective function is defined as the negative of the mean composite score over all calibration landslides so that the optimiser can minimise it. For each suggested parameter vector P, the objective function:
   - Updates the working configuration (such as `thr_threshold` if threshold optimisation is enabled);

- Launches runout simulations for all calibration landslides in parallel using `joblib`;
- Calculates for each event the composite score returned by Phase IV, i.e. the weighted combination of performance metrics as defined in the weights section of `config.yml`;
- Returns the negative of the average score over events, ignoring simulations that failed.

The objective function minimised by the Bayesian optimiser is, thus, simply the negative of the mean composite score on all calibration landslides.

2. Bayesian search loop and checkpointing: The optimiser starts with a small set of random evaluations (`bayes.n_initial`) and iteratively generates new parameter vectors for each batch (updating the surrogate model). Intermediate values can be saved as checkpoints where calibration can be resumed later if necessary.

Outcomes: By the end of Phase III, SEARCH has identified, for each fold or for the full calibration set in the case of a simple hold-out strategy, a candidate set of global parameters P_{best}, which maximises the average composite score when combined with locally optimised source geometries.

Phase IV: Hierarchical Simulation and Source Optimisation

This is the most computationally intensive phase. During calibration it is invoked inside Phase III for every proposed set of global parameters, and the same logic is reused during validation.
Its purpose is to compute, for each individual landslide, a **single objective score** given a fixed set of global physical parameters **P** (e.g. P_1).
To achieve this, the framework does not run a single simulation per landslide. Instead, it performs a local optimisation of the source geometry: for each landslide, several alternative source configurations are generated and tested, and only the one that best reproduces the observed runout is retained.
The final performance score for that landslide and that parameter set **P** is then computed using a full set of metrics and the metric weights defined in `config.yml`.
Operationally, for each event, the outer procedure calls the function `_run_single_simulation(P, landslide)`, which implements the following steps:

- Local isolation and metric CRS: The observed landslide polygon is extracted, converted into a GeoDataFrame and reprojected to a metric coordinate reference system consistent with the DEM. A dedicated job-specific working directory is created under `paths.model_working_dir` so that all temporary files related to that event and that parameter set are kept separate, ensuring safe and fully parallel execution;
- Optional metric buffer for evaluation: Depending on the configuration flags `use_buffer_for_metrics` and `buffer_distance`, the observed polygon can be expanded by a fixed distance before computing the performance metrics.
  This "evaluation buffer" accounts for mapping uncertainties and small georeferencing errors. Importantly, this buffered polygon is used **only** for metric computation; the source delineation always uses the original mapped footprint;
- DEM clipping and optional high-resolution delineation: A subset of the DEM is extracted around the landslide, using the distance specified by `job_clip_buffer_m`.
  If high-resolution options are enabled (`hires_source_search`), this subset can be replaced by a finer DEM previously generated in Phase II.
  On this local DEM, the slope field is computed and, when required, refined by `compute_local_slope_range` so that the slope thresholds used to define the source better reflect the local topography of that specific event;
- Source generation and shape search: A first, "baseline" candidate source is delineated inside the observed polygon using the current slope range, the interior offsets

(`source_interior_offset_cells`) and the consolidation strategy (`source_area_consolidation`).

If `processing.source_shape_search.enabled` is set to True, SEARCH then enters an inner source-shape optimisation loop, in which multiple alternative source geometries are generated by varying:

a) The internal offset (e.g. 0 or 1 cell);
b) The minimum allowed source size;
c) The maximum source area ratio (area cap);
d) The widening or narrowing of the slope range.

For each source variant:

- The source polygon can be adjusted to satisfy volumetric constraints (if `source_volume.enabled`), using event-specific thickness or volume attributes from the inventory;
- The selected runout model (the `RunoutModel` engine, e.g. DebrisDice via `GenericModel`) is executed: the framework builds the required input files from `file_templates`, performs basic consistency checks (e.g. DebrisDice header checks) and runs the external executable;
- The model output is read and converted into a raster field (e.g. a velocity grid), which is then binarised using the current `thr_threshold` and polygonised to obtain the simulated runout footprint for that variant.

- Quick scoring and early stopping (metrics for source selection): For the purpose of source-shape search, a reduced set of "quick" metrics (typically Intersection-over-Union, F1-score and precision, as specified in `processing.source_shape_search.quick_metrics`) is computed by comparing each candidate simulated polygon with the observed (or buffered) polygon.
  These quick metrics are used only to decide which source geometry is better for a given landslide under the current global parameters P; they do not directly define the global optimum.
  The quick metrics are normalised and combined into a single "quick score" using the same normalisation rules and the same weight map as the full metrics, but restricted to the selected quick metrics.
  If a new variant improves the best quick score found so far, it becomes the current best source for that landslide; otherwise, a counter of consecutive non-improvements is incremented.
  The inner loop stops either when the maximum number of tested variants (`max_variants`) is reached or when the number of successive non-improving trials exceeds the early-stopping threshold (`early_stop_patience`). At this point, only the best source geometry for that landslide is retained; all others are discarded;
- Full metric computation (metrics for global parameter optimisation): Once the best source geometry has been identified, the framework runs a *single, definitive evaluation* for that event and computes the ***full set of metrics*** for the corresponding simulation. These include:
  - Overlap and classification metrics: IoU, precision, recall, F1-score;
  - Size-related metrics: relative area error between simulated and observed runouts;
  - Shape and agreement metrics: Hausdorff distance (scaled by `hd_scale_m`) and Cohen's κ;
  - Threshold-independent metrics: ROC AUC and Precision-Recall AUC, computed on a raster grid within the Region of Interest (ROI) defined by `validation.metrics_roi`.

In practice, the ROI is constructed as a binary mask on the simulation grid: by default, it contains the rasterised observed polygon, and it can optionally be expanded to include the simulated runout and to be dilated by a metric buffer (`buffer_m`, as specified in `validation.metrics_roi.buffer_m`). Pixel-based quantities (Cohen's κ, ROC AUC, PR AUC) are then computed only inside this mask, thus excluding distant background areas.

The metrics listed above are those for which the user can assign explicit weights in the `weights` section of `config.yml` (e.g. iou, f1_score, precision, recall, relative_error, hausdorff_distance, cohen_kappa, roc_auc, pr_auc, and optionally source_area_ratio). In a first step, each metric is normalised to the [0, 1] range and, if necessary, transformed so that higher values always correspond to better performance (for instance, errors and distances are converted into "1 - normalised error").

In a second step, these normalised metrics are aggregated into a single composite score for the landslide and the parameter set P by a weighted average based on the user-defined weights. Metrics with zero weight do not contribute to the composite score; metrics not listed in the weights block are ignored for optimisation purposes.

The composite score returned by `_run_single_simulation` therefore represents, for that landslide and that particular set of global parameters **P**, the **best performance that can be achieved** once the source geometry has been optimised locally using the quick metrics.

This value is passed back to the outer optimisation or validation loop and is used, in Phase III, to compute the average score over all calibration landslides for the current global parameter set.

In this way, global rheological parameters are always assessed under their own, internally optimised source geometry, while the **global optimum of the model** is determined by the weighted combination of the full set of metrics specified by the user.

Outcomes: This phase operationalises the key idea of SEARCH: global parameters are always evaluated together with the best source geometry that can be constructed for each event under those parameters and processing rules, and the final optimal parameter set is defined by a user-controlled, weighted multi-metric objective function.

Phase V: Per-Fold Threshold Optimisation

The binarisation threshold, which transforms continuous model output into a binary runout footprint, is very important for the trade-off between precision and recall, because it defines exactly which predicted values are counted as "runout" (values above the threshold) and which ones are counted as "no runout" (values below the threshold).

The SEARCH module has two complementary strategies, controlled by `processing.threshold.mode`:

- Joint optimisation (`mode: 'optimize'`): Here `thr_threshold` is present directly in the Bayesian optimiser search space, so it is tested and updated together with the other global model parameters during calibration. P_{best} includes the optimised value of `thr_threshold`, meaning that the final optimal parameter set includes the threshold value that gives the highest calibration performance.
  After calibration this value is placed back onto `processing.thr_threshold` and is used for all subsequent validation runs;
- Per-fold optimisation of F1 (`mode: 'per_fold_best_f1'`): Here `thr_threshold` is not a part of `optimizer_space`. Once P_{best} has been chosen, the function `choose_best_threshold_per_fold` performs a one-dimensional search of a candidate threshold grid, i.e. it tests a sequence of threshold values while keeping the physical parameters fixed to P_{best}.
  Simulations are re-run for each candidate threshold on the calibration subset; the mean F1-score is calculated (where the F1-score is the harmonic mean of precision and recall).
  For that fold, the threshold that maximises the mean F1-score is chosen as the trade-off between precision and recall (T_{best}) which is used during validation.

Outcomes: In particular, the second strategy decouples the optimisation of physical parameters from the post-processing threshold, because the physical parameters are optimised first, and the threshold is selected afterwards by choosing the value that maximises the mean F1-score (i.e., the harmonic balance between precision and recall) on the calibration subset.

Phase VI: Validation and Final Aggregation

The last stage assesses the fully calibrated model using independent validation data, i.e., landslides not used during calibration, and aggregates results across folds:

- Validation simulations: The function `run_validation` applies the calibrated parameters (and, if relevant, T_{best}) to all landslides in the validation subset for each fold.
  It runs `_run_single_simulation` for each event, computes the validation metrics, and saves both the simulated runout footprint and the optimised source geometries.
  The unioned simulated polygons are then stored jointly with the event ID for each event to avoid multi-part geometry complexity;
- Metrics and geospatial outputs per fold: All validation landslides and folds are combined (`crossval_validation_metrics_all_folds.csv`) into a single table in K-Fold mode, while simulated runouts and sources are collated in `crossval_simulated_runouts.gpkg` and `crossval_source_areas.gpkg`. In simple hold-out mode, analogous files are written as `validation_metrics.csv`, `validation_simulated_runouts.gpkg` and `validation_source_areas.gpkg`;
- Summarised statistics and uncertainty: The `export_professional_summaries_kfold` and `export_professional_summaries_holdout` functions provide a collection of tabular and graphical summaries:

  - Table of the best parameters per fold (`crossval_best_params_per_fold.csv`);
  - Descriptive statistics for each metric (means, medians, standard deviations, `validation_metric_summary.csv`);
  - A table of parameter statistics across folds (`parameter_stats_across_folds.csv`);
  - Bootstrap confidence intervals for the computed metrics (`validation_metric_bootstrap_ci.csv`), and a JSON file (`selected_best_parameters.json`) containing the recommended global parameter set, along with the selection metric used.

  - Excel summary and plots: A multi-sheet Excel workbook (`calibration_summary.xlsx`) provides a summary of the recommended parameters, per-fold parameters, summary metrics, optimisation history, metric weights, bootstrap confidence intervals and a flattened snapshot of the most important configuration options.
    Convergence curves, parameter pair plots, metric boxplots, bar plots of mean ± standard deviation, and parameter histograms are typical diagnostics that are saved as PNG files;

  - Executive HTML report: Finally, SEARCH produces an executive-style HTML report (`executive_summary.html`, also saved as `final_report.html` for backward compatibility), embedding the main plots as base64-encoded images and summarising the recommended parameters, the primary performance metric and the final validation results.
    This report is designed for quick inspection and communication with non-technical stakeholders.

3.6 Dynamic Risk Analysis: From Hazard to Impact

Once the propagation models have generated the runout intensity rasters, primarily represented by velocity fields, the SADRI framework executes the risk analysis chain.

This process, managed primarily by the `analysis.py` module, translates the physical simulation outputs into impact scenarios and quantitative estimates of economic risk for each spatial unit. In this workflow, exposed assets are considered only after the propagation simulation has been completed.

They are not included in DebrisDice as physical obstacles and therefore do not modify the flow routing, the velocity field or the stopping criteria.

The simulated runout and velocity rasters are first generated independently; asset geometries are then overlaid on these rasters to evaluate local impact intensity and economic consequences.

The analysis is divided into four fundamental steps: intensity calculation, exposure quantification, vulnerability modelling, and risk estimation, including uncertainty quantification.

3.6.1 Intensity Calculation (I)

The first parameter to be quantified is the event intensity (I), defined at the asset level as the maximum simulated flow velocity (`V_asset`, m/s) affecting each exposed element.

This value is extracted from the georeferenced runout velocity raster (`Velocity_georef.tif`) by sampling the highest pixel value within the asset footprint.

This conservative choice is implemented through the `_sample_intensity_on_asset` function in the `analysis.py` module.

It is important to note that `V_asset` represents the flow velocity at the asset location without considering physical flow–structure interaction, as buildings and other obstacles are not explicitly included in the propagation simulation.

To ensure data quality, the sampled velocity is first compared against a minimum threshold (`risk_parameters.velocity_threshold_ms`, m/s) to filter out numerical noise.

If `V_asset` falls below this limit, the flow is considered practically stopped at the asset location and the velocity is set to 0 m/s. The resulting velocity value is then normalised to convert the absolute velocity in m/s into a dimensionless relative intensity (`I_norm`).

This step accounts for geomorphological variability, since the same velocity may have a different relative meaning depending on whether it occurs in a steep or a gentle basin.

The normalisation uses a local reference velocity (`V_ref`, m/s), which serves as a benchmark for each spatial unit.

`V_ref` is derived from the distribution of maximum velocities simulated across both the deterministic run and all Monte Carlo iterations for the same spatial unit.

It is calculated as a user-defined high-end percentile of this distribution, specified in the configuration file through `risk_parameters.normalisation_percentile`; for example, a value of 98 corresponds to the 98th percentile.

Finally, the Normalised Intensity (`I_norm`) per asset is calculated using the filtered `V_asset` value as:

$$I\_norm = \min\left(1, \frac{V\_asset}{V\_ref}\right)$$

[3.14]

where `V_asset` is the filtered impact velocity, expressed in m/s, and `V_ref` is the characteristic high-end velocity of the corresponding spatial unit, also expressed in m/s. The resulting `I_norm` is dimensionless and bounded between 0 and 1, expressing the asset-level impact intensity relative to the local dynamic behaviour of the flow.

3.6.2 Vulnerability Calculation (V)

Vulnerability (V) is modelled as the expected degree of loss for an exposed asset, given a certain level of event intensity (I) and a certain structural resistance (R_struct) of the asset. The SADRI framework does not use a static vulnerability value; instead, it calculates it dynamically in two steps:

1. Structural Resistance (`R_struct`) Estimation: First, the `_get_structural_vulnerability` function determines the asset's intrinsic robustness.
   It attempts to read a structural resistance index ranging from 0 to 1 from the `vuln_index` column, as defined in the `constants.py` module, of the asset file.
   If this data is missing, the system assigns a default value based on the asset's type, for example "residential" or "commercial", retrieving it from the `vulnerability.default_values` dictionary in `config.json`. These default values represent the reference structural resistance assigned to each asset category when asset-specific information is unavailable;

2. Vulnerability Function Application: The final vulnerability is calculated by the `_calculate_vulnerability_li_et_al` function (which implements a modified version of the equation system introduced by Li et al. (2010)), modelling an S-shaped curve *(Figure 12)*.
   This function, defined as follows, relates the ratio between the Normalised Intensity (`I_norm`) and the Structural Resistance (`R_struct`):

$$x = \left(\frac{I\_norm}{R\_struct}\right)$$

$$V = f(x) = \begin{cases} 2x^2 \\ 1 - 2(1-x)^2 \\ 1.0 \end{cases} \longrightarrow \begin{pmatrix} if\ x \leq 0.5 \\ if\ 0.5 < x \leq 0.9 \\ if\ x > 0.9 \end{pmatrix}$$

[3.15]

where V ranges from 0 to 1. Low values indicate limited expected damage, whereas values close to 1 indicate severe or complete expected loss.

This approach realistically models how an asset, even a robust one, can become highly vulnerable if struck by an event of extreme intensity.

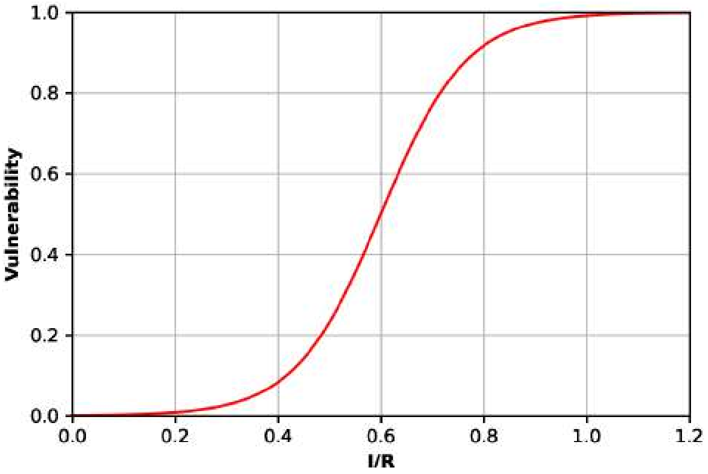


Figure 12. *Graphical representation of the S-shaped vulnerability curve modified from Li et al. (2010)*

3.6.3 Exposure Quantification (E)

Exposure quantifies the economic value of the elements located within the impact area. This calculation is managed by the `_calculate_exposure` function in `analysis.py` as follows:

1. Economic Value Identification: The asset's value is read from the `econ_value` column (defined in module `constants.py`) of the vector file. The presence of this column is verified at startup by the `data_manager.py` module;
2. Missing Data Management: If the `econ_value` for a specific asset is missing, the framework does not discard the element but assigns it a fallback value. This fallback is the average economic value of all other assets of the same type (e.g. residential, commercial) present within the spatial unit being analysed;
3. Proportional Quantification: The final exposure is not the asset's total value but is calculated proportionally to the physical impact, depending on the asset's geometry:
    - Polygons (e.g., buildings): Exposure is the total economic value multiplied by the fraction of the asset's area that intersects the runout polygon;
    - Points: Exposure is the total economic value.

3.6.4 Risk Estimation (R) and Uncertainty Quantification

In the general risk formulation, economic risk is commonly expressed as the combination of hazard (H), vulnerability (V), and exposure (E).

In the event-based SADRI application, the hazard term is not introduced as an additional multiplicative factor at this stage, because the occurrence and intensity of the hazardous process have already been explicitly simulated. The runout footprint identifies whether an asset is impacted, while the local velocity-derived intensity (`I_norm`) is used to calculate vulnerability.

Therefore, for each impacted asset, the deterministic economic loss is computed as the product of vulnerability and exposure, as implemented in the `calculate_impact_risk_detailed` function:

$$\text{Risk } [€] = V\ [-] \times E\ [€]$$

[3.16]

where V is the intensity-dependent vulnerability value, ranging from 0 to 1, and E is the exposed economic value in euros. In this formulation, the hazard information is already embedded in the impact/no-impact condition defined by the runout footprint and in the intensity term used to derive V.

The total deterministic risk for the spatial aggregation unit is the sum of the risks for all impacted assets (`Impact_Risk`).

To account for the inherent uncertainty in the runout model parameters, even if already calibrated via the `runout_calibrate` workflow, a Monte Carlo procedure is implemented in the `simulation.py` module. As specified in `config.json`, the entire simulation pipeline is repeated *N* times (`monte_carlo.iterations`).

In each iteration, key physical parameters (e.g., friction_factor, viscosity) are perturbed by sampling them from a defined distribution around their optimal value.

This process, managed by main.py, generates a distribution of N possible total risk values (`mc_risks`), rather than a single deterministic value. From this distribution, the system finally calculates the mean (`Impact_Risk_MC_Mean`) and standard deviation (`Impact_Risk_MC_Std`).

These indicators, saved in the `montecarlo_summary.json` file and the final GeoPackage, provide a quantitative measure of the uncertainty associated with the risk estimation.

Chapter 4

Study Area

This chapter provides a detailed characterisation of the study area, the Misa River basin, and describes the two distinct rainfall scenarios simulated: the extreme rainfall event of September 2022 and the prolonged rainfall event of May 2014, selected to analyse the hydrological response within the SADRI framework.

4.1 Misa River Basin

The study area is the Misa River basin, which is located in the Marche region, in central Italy. The basin, which covers an area of 384 km², extends from the south-western slopes of the mountain ranges of the Umbrian-Marchean Apennines, where the river originates (in the municipality of Genga in the province of Ancona), to the Adriatic coast, where it flows into the sea near Senigallia (AN) (*Figure 13*).

The altitude varies from sea level up to 824 m a.s.l. The mountainous sectors exhibit the steepest slopes, with maximum gradients reaching 67.7°.

From a geological perspective, the Misa basin is considered to be part of the Umbrian-Marchean Apennines sector of the Apennine chain, which was formed during the Cenozoic era as a consequence of the convergence between the African and Eurasian plates (Carmignani et al., 2004; Carminati and Doglioni, 2012).

The tectonic evolution of the region involves the interaction of compressive and extensional processes, which has produced folds and fault systems, together with the accumulation of sedimentary deposits of different origins. The basin includes seven main lithological units, which comprise alluvial, eluvial and colluvial deposits, poorly lithified clastic rocks, siliciclastic rocks, well-cemented clastic rocks, carbonate rocks, marlstone, and evaporite (Figure 14).

The majority of the basin is composed of clastic rocks with poor lithification, classified as the Argille Azzurre Formation (Vai, 2001; Conti et al., 2020). The hydrological response of the basin is strongly influenced by the widespread occurrence of low-permeability materials, as these substantially restrict infiltration and promote rapid surface runoff during intense rainfall events.

From a geomorphological perspective, the study area can be divided into two separate geographical zones. The first area is a mountainous region to the south-west, which is part of the Umbrian-Marchean Apennines, and is characterised by the presence of rock outcrops, primarily composed of marly limestones.

The second region extends as a flatter piedmont area towards the sea, where thick sequences of recent alluvial sediments and eluvio-colluvial deposits are observed (Centamore et al., 1979; Coltori, 1997).

The territory is characterised by a dense hydrographic network that conveys sediments downstream. The area is defined by the presence of the Misa River, which is joined by its primary tributary, the Nevola River.

The two main rivers are fed by numerous secondary and minor streams which follow a dendritic pattern throughout the basin, contributing to a rapid hydrological response during periods of intense precipitation.

The main land use consists of agricultural areas, with intensively cultivated fields, while urban development is concentrated primarily along the coastal sector (CORINE Land Cover 2018, Copernicus Land Monitoring Service & European Environment Agency, 2020).

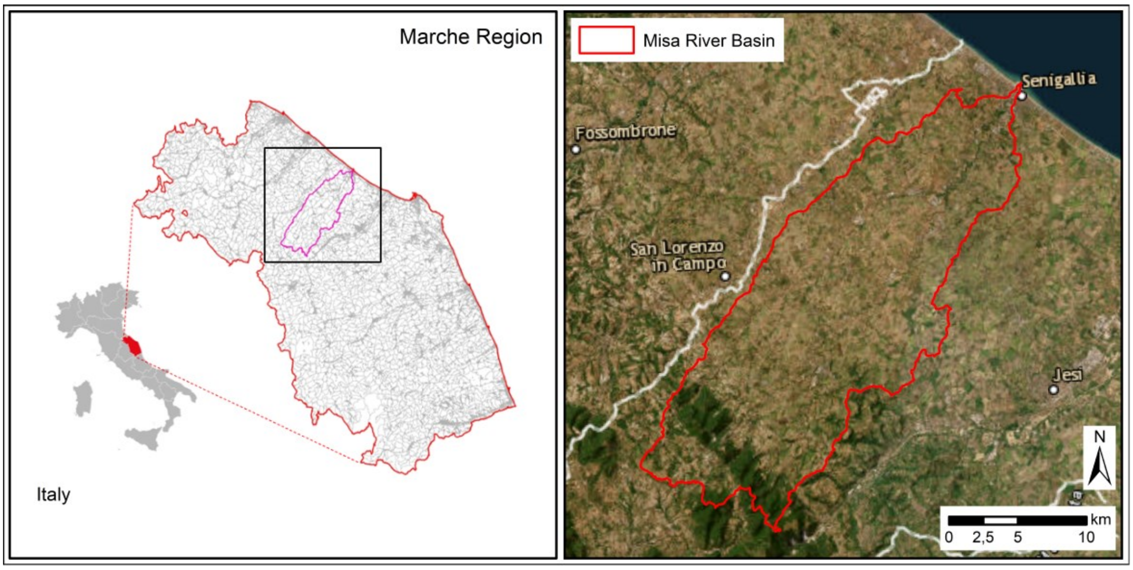


Figure 13. *Study area – Misa River Basin, Marche Region (Italy)*

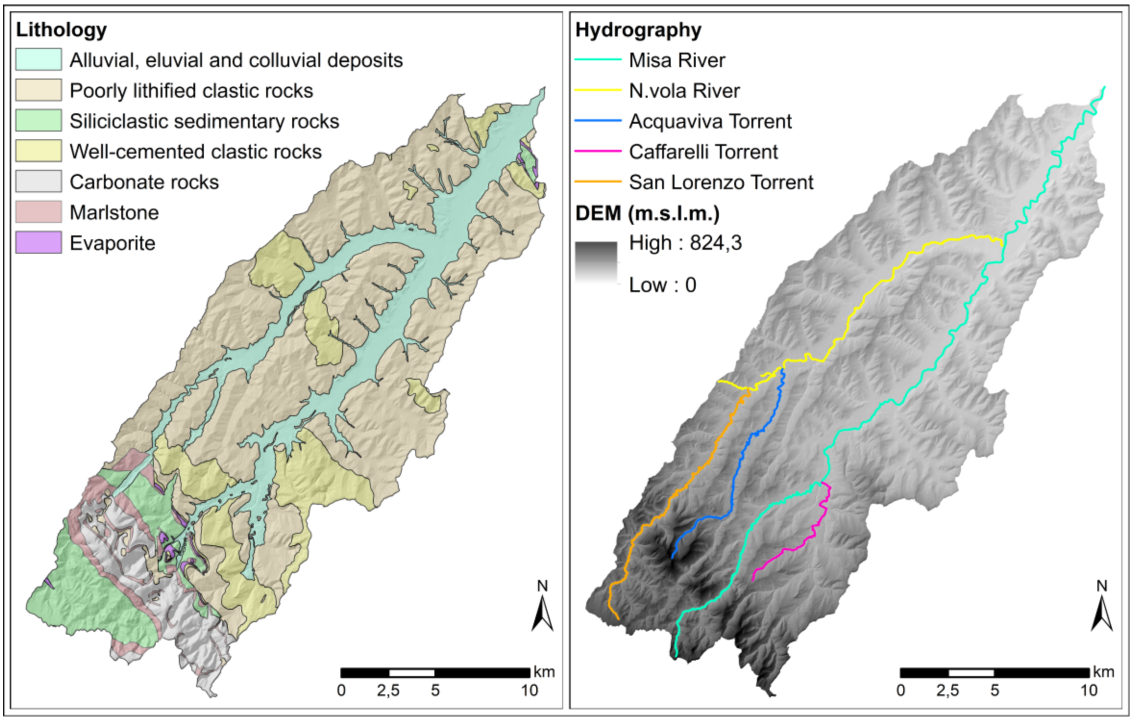


Figure 14. *Lithology and Digital Elevation Model (DEM) with Hydrography of the study area*

4.2 Rainfall Scenarios

In order to validate the SADRI framework under a range of hydrological conditions, two distinct rainfall events affecting the Misa River basin were selected, thus representing different scenarios.

4.2.1 Scenario A: Rainfall event of 15-16 September 2022

The rainfall event that affected the Marche region between the 15th and the 16th of September 2022 was classified as an "outlier event" due to its unprecedented intensity and exceptional magnitude (Donnini et al., 2023). The event is particularly significant because it occurred after several months of rainfall deficit, which had led to reduced soil moisture and low antecedent saturation levels across the region (Marche Region Functional Centre, 2022a). The episode was characterised by intense and persistent precipitation, with accumulated totals exceeding 400 mm within a few hours in the mountainous sectors (*Figure 15*).

This led to a multitude of shallow landslides, debris flows, and widespread erosion phenomena (Donnini et al., 2023). The severe meteorological conditions developed as a result of a combination of atmospheric drivers, including a pronounced inflow of water vapour (IWV – Integrated Water Vapor) from the western Mediterranean and intense atmospheric forcing that promoted rapid cloud development (Torcasio et al., 2023; Donnini et al., 2023). In more detail, a deep depression over Scandinavia steered a slow-moving, humid, and unstable south-westerly flow across central Italy.

The first convective storms progressively weakened as they approached the coastline; however, during the late afternoon of 15 September, a stationary and self-regenerating convective system developed.

Satellite imagery has been utilised to illustrate the evolution of a self-regenerating "V"-shaped convective thunderstorm (McCann, 1983; Duffourg and Ducrocq, 2011; Ducrocq et al., 2014) over the eastern sector of the central Apennines.

The system remained stationary and highly intense between approximately 17:00 and 20:00 local time, focusing rainfall over a narrow area that included the Candigliano, Cesano, Misa, and Sentino River basins.

Daily precipitation on September 15th, computed as the mean value across all pixels within the Misa basin, was approximately 60 mm (*Figure 16*). The most intense rainfall was recorded at the vertex of the "V" supercell, near the Cantiano station, where peak accumulations reached 419 mm in only 12 hours (Marche Region Functional Centre, 2022b). Return-time analyses based on several key rain gauges confirmed the exceptional nature of the event, exceeding the rainfall amounts associated with 1000-year return periods in multiple locations. Over a six-hour interval, the stations of Barbara, Colle, and Arcevia recorded 121.2 mm, 186.4 mm, and 117.8 mm, respectively, all surpassing the estimated 1000-year threshold (*Table 13*). The hydrological response was immediate, producing a severe flood that impacted several municipalities, including Ostra, Pianello di Ostra, Barbara, Serra de' Conti, Ostra Vetere, Castelleone di Suasa, Sassoferrato, and Senigallia, resulting in 12 fatalities and one missing person (Donnini et al., 2023). The destructive force of the floodwaters is also evidenced by the destruction of critical monitoring infrastructure: the hydrometric stations at Corinaldo (Nevola River) and San Vittore (Sentino River) were completely washed away, while the Pianello di Ostra station on the Misa River was severely damaged (Marche Region Functional Centre, 2022b).

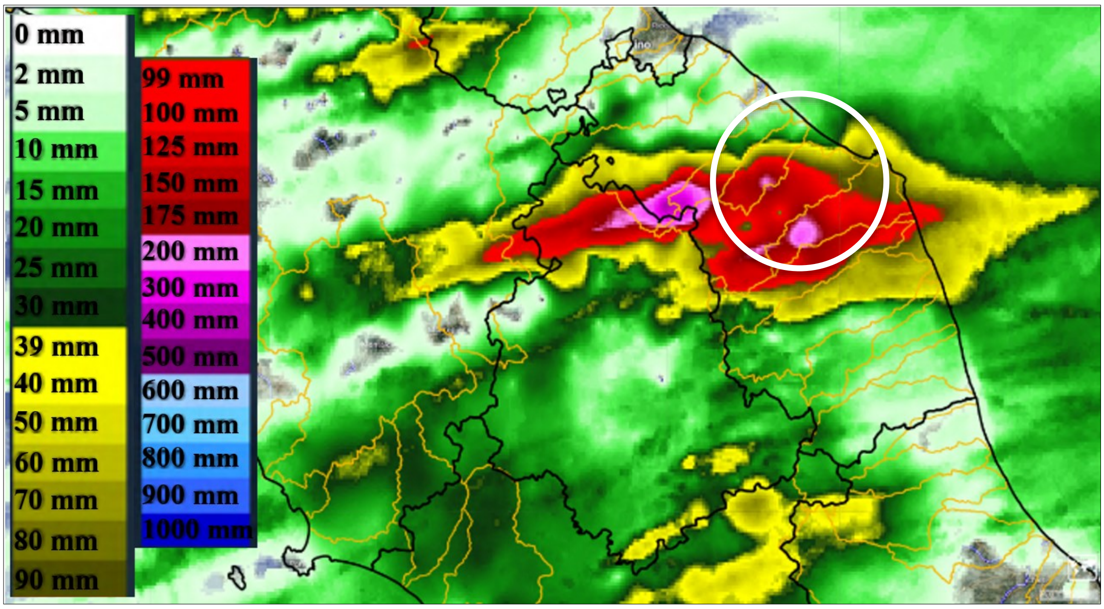


Figure 15. *Cumulative precipitation map (mm) for September 15, 2022, obtained using the "merging of radar data with rainfall data" provided by the National Department of Civil Protection - Dewetra platform. Black line: regional borders. yellow: river basins. white circle: Misa River Basin (modified from Marche Region Functional Centre, 2022b)*

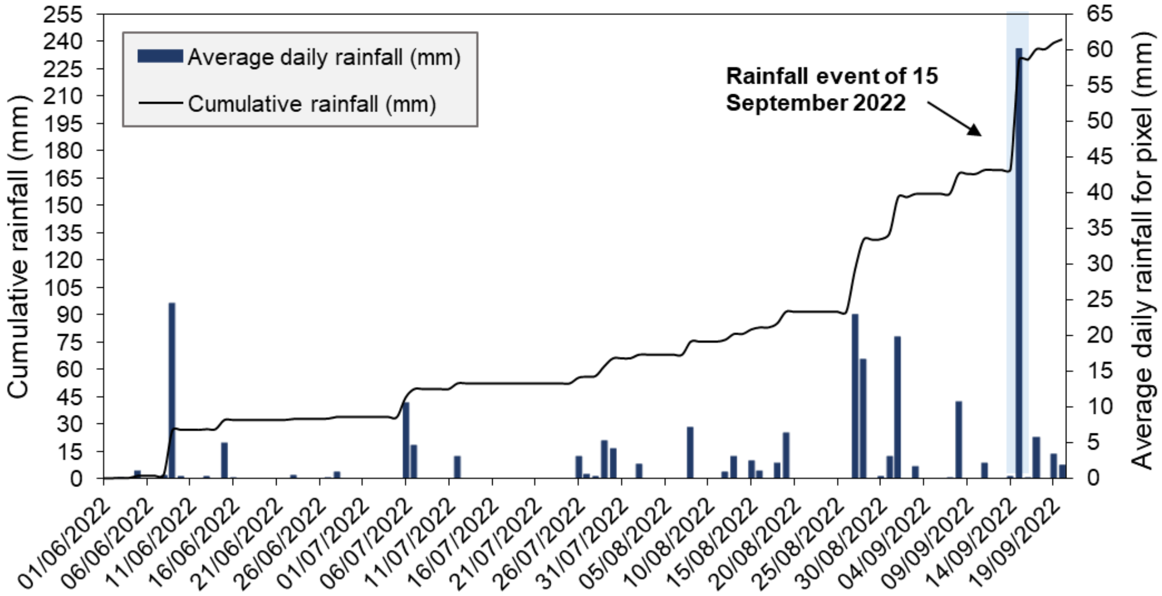


Figure 16. *Cumulative and average daily rainfall per pixel from 01/06/2002 to 20/09/2022 in the Misa River basin, the rainfall event is highlighted in light blue (rainfall radar data from the Italian National Radar Network available at: https://rischi.protezionecivile.gov.it/it/meteo-idro/monitoraggio/rete-radar/)*

Table 13. *Estimated return period of precipitation recorded by significant rain gauges for the 15-16 September rainfall event in the Misa River basin*

Station	**Basin**	**Cumulative rainfall (mm)**				**Return Period (year)**			
		Max 3h	*Max 6h*	*Max 12h*	*Max 24h*	*3h*	*6h*	*12h*	*24h*
Arcevia	Misa	94.8	117.8	128.8	129.2	>1,000	>1,000	910	200
Barbara	Misa	111.4	121.2	127.0	127.2	>1,000	>1,000	>1,000	140
Colle	Misa	162.4	184.4	204.0	204.0	>1,000	>1,000	>1,000	>1,000

A dedicated field and remote sensing campaign was conducted to map the storm's geomorphological effects. The initiative, as reported in Confuorto et al., 2025, resulted in the mapping of a total of 805 rainfall-induced landslides, 701 within the Misa River basin (*Figure 17-18*).

The analysis revealed that shallow landslides were the predominant instability mechanism, accounting for 64.5% of the mapped events. Sliding landslides represented 16.4% of the total, while mudflows accounted for 4.1%. The remaining events consisted of bank landslides (9.4%) and landslides with an undetected failure mechanism (5.6%).

This pervasive landslide activity was a direct consequence of the extreme rainfall intensity, which rapidly saturated the shallow soil layers, caused extensive slope failures, and contributed sediment to the river channels, thereby increasing the destructive power of the flood.

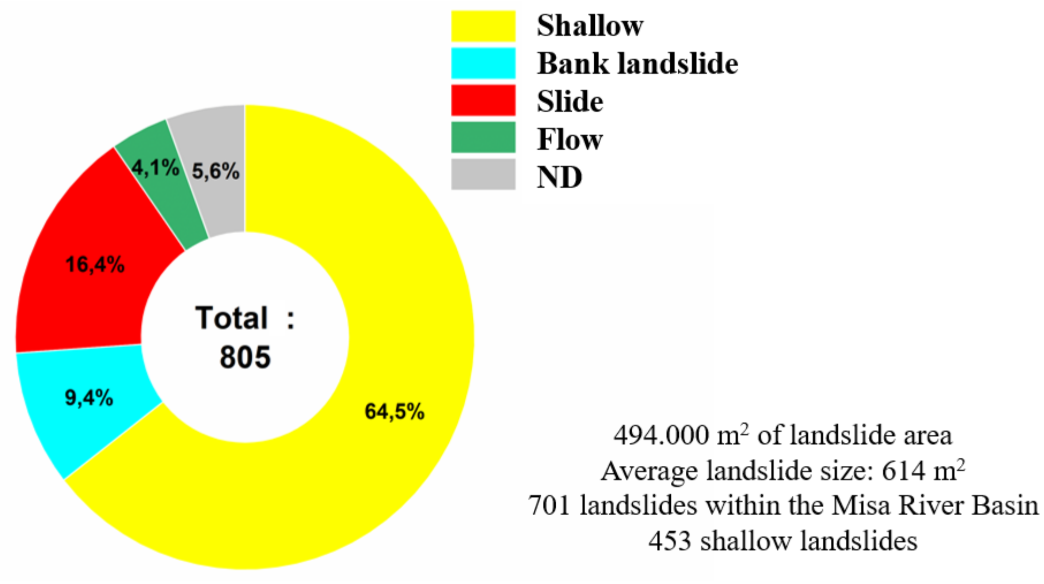


Figure 17. *Graph showing the total number and typology of landslides triggered during the rainfall event (modified from Confuorto et al., 2025)*

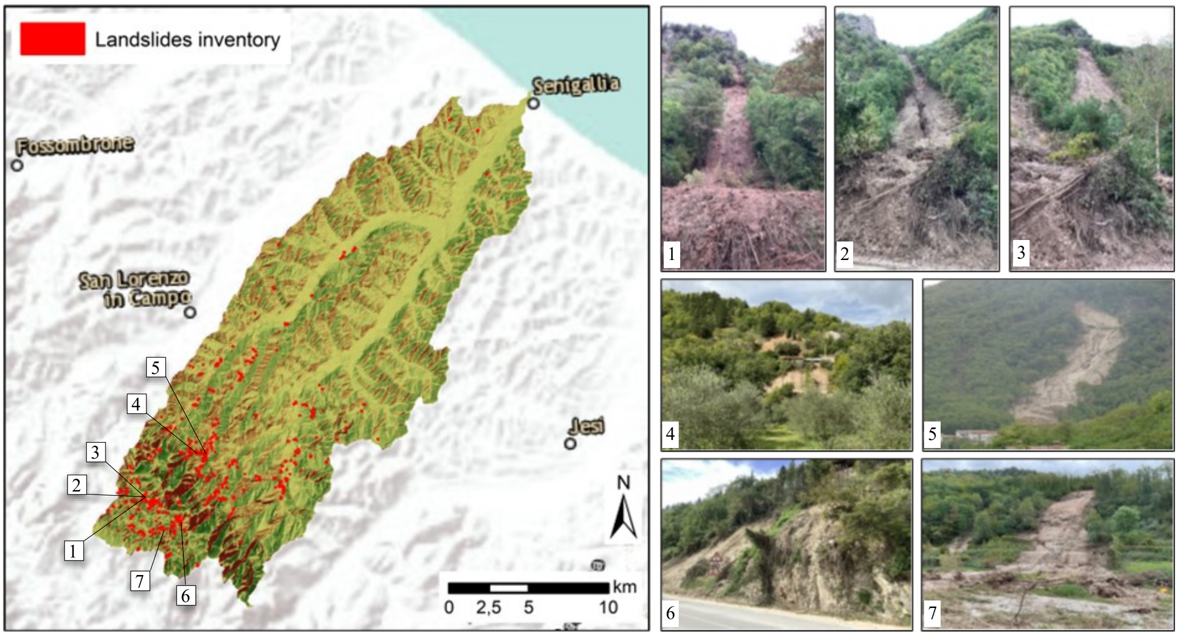


Figure 18. *Inventory of landslides triggered during the September 2022 rainfall event in the Misa River Basin. The representative examples shown on the right are identified on the inventory map by the corresponding number labels*

4.2.2 Scenario B: Rainfall event of 2-4 May 2014

The Marche region was affected by an extreme rainfall event between 2 and 4 May 2014, whose damaging consequences were amplified by antecedent conditions and an exceptionally high concentration of precipitation.

The event occurred after a prolonged rainy period that resulted in complete soil saturation, with an average of approximately 99 mm of rainfall recorded over the preceding 30 days (Marche Region Functional Centre, 2014). This antecedent condition substantially reduced the soil infiltration capacity, thereby increasing the susceptibility of the area to surface runoff. A deep atmospheric trough over the Tyrrhenian Sea advected warm and moist air from the Adriatic, favouring the development of intense and widespread thunderstorms (*Figure 19*). Within the 48-hour period under consideration, the most affected basins were the Arzilla, Misa, and Cesano basins. Rainfall amounts in these basins were 36.7 mm, 33.8 mm, and 31.6 mm, respectively, while cumulative totals over the two-day period reached 86.2 mm, 71.3 mm, and 75.5 mm (Marche Region Functional Centre, 2014).The most critical phase occurred over a very short time window, spanning the six hours between midnight and 6:00 a.m. on 3 May.

A detailed analysis of rainfall-station records was performed to quantify the severity of the event, with particular focus on the mid-hilly sector of the Misa basin. Data from the Agency for Agri-Food Services of the Marche Region (ASSAM) indicate that, on 3 May, the Serra de' Conti station recorded 104.6 mm of precipitation, followed by Barbara (76.8 mm), Senigallia (59.8 mm), and Corinaldo (59.4 mm).

Rainfall intensity is further highlighted by the peaks reported by the Civil Protection MIR Network: during the six hours of maximum accumulation, the highest values were measured at Corinaldo (68.4 mm) and Barbara (63.0 mm). These totals increased to 82.4 mm and 71.4 mm, respectively, over the 12-hour interval (*Table 14*). Over the entire 48-hour period, several stations exceeded 100 mm, with notable accumulations recorded at Foglia2-Montellabate (161.2 mm), Montecchio (149.6 mm), San Lorenzo in Campo (102.6 mm), Fermo (102.8 mm), Tesino (99.8 mm), and Spinetoli (102.2 mm). Furthermore, daily precipitation on 3 May, computed as the mean value across all pixels within the Misa basin, was approximately 55 mm (*Figure 20*).

A preliminary hydrological analysis indicated that rainfall intensity corresponded to a return period of approximately 10 years.

Nevertheless, the resulting hydrological response was disproportionately severe, as the flood wave exhibited a markedly higher return period than suggested by rainfall statistics alone (Marche Region Functional Centre, 2014). It is evident that, due to the already saturated soil conditions, precipitation was almost entirely converted into surface runoff, thereby generating an enhanced flood wave.

The consequences for the area were catastrophic, with widespread inundation and the most critical impacts along the Misa River, which flooded large portions of the city of Senigallia.

In addition, the event triggered numerous landslides, particularly on low-permeability hillslopes, leading to the obstruction of local road infrastructure within the affected area (*Figures 21-22*). The event produced widespread urban flooding, disruption of essential services, and resulted in three fatalities, with an estimated restoration cost of €366.181 million (Trigila et al., 2015).

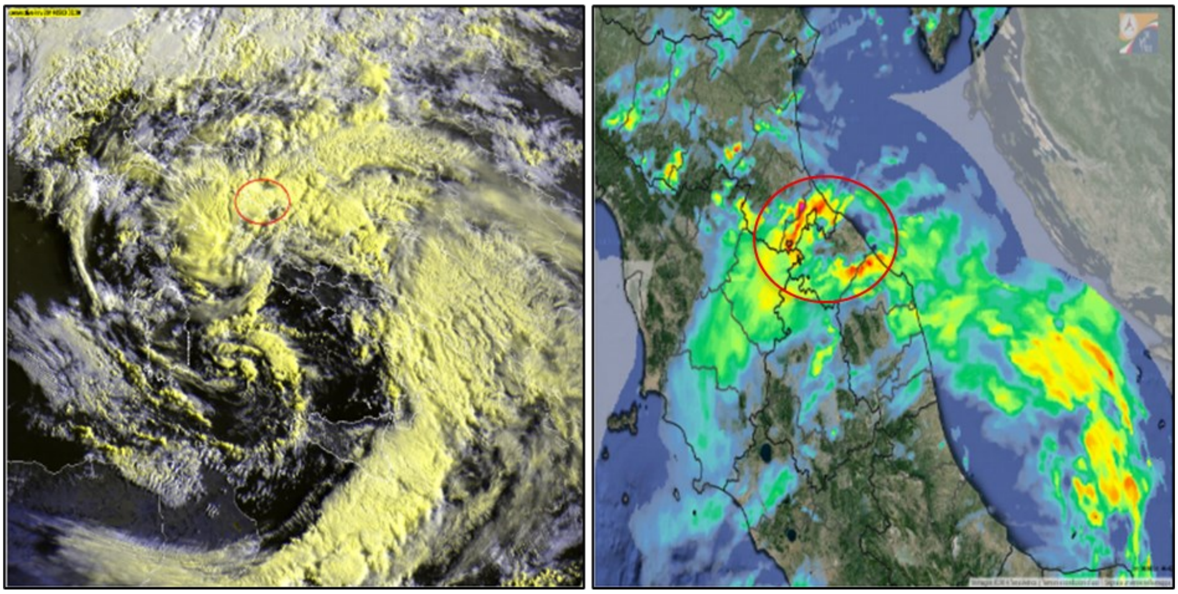


Figure 19. *Convective-hrv satellite image with the convective clouds that affected the northern coast of the Marche region (red circle (left)) and radar image - Vertical Maximum Intensity (right) referred to h 05:30 UTC of 3 May 2014, red circle represents the study area (modified from Marche Region Functional Centre, 2014)*

Table 14. *Estimated return period of precipitation recorded by significant rain gauges for the 2-4 May rainfall event in the Misa River basin (data from: Marche Region Functional Centre, 2014)*

Station	**Basin**	**Cumulative rainfall (mm)**		**Return Period (year)**	
		Max 6h	*Max 12h*	*6h*	*12h*
Arcevia	Misa	17.6	20.8	10	10
Barbara	Misa	63.0	71.4	>10	>10
Corinaldo	Misa	68.4	82.4	>10	>10
Senigallia	Misa	21.4	35.4	10	10

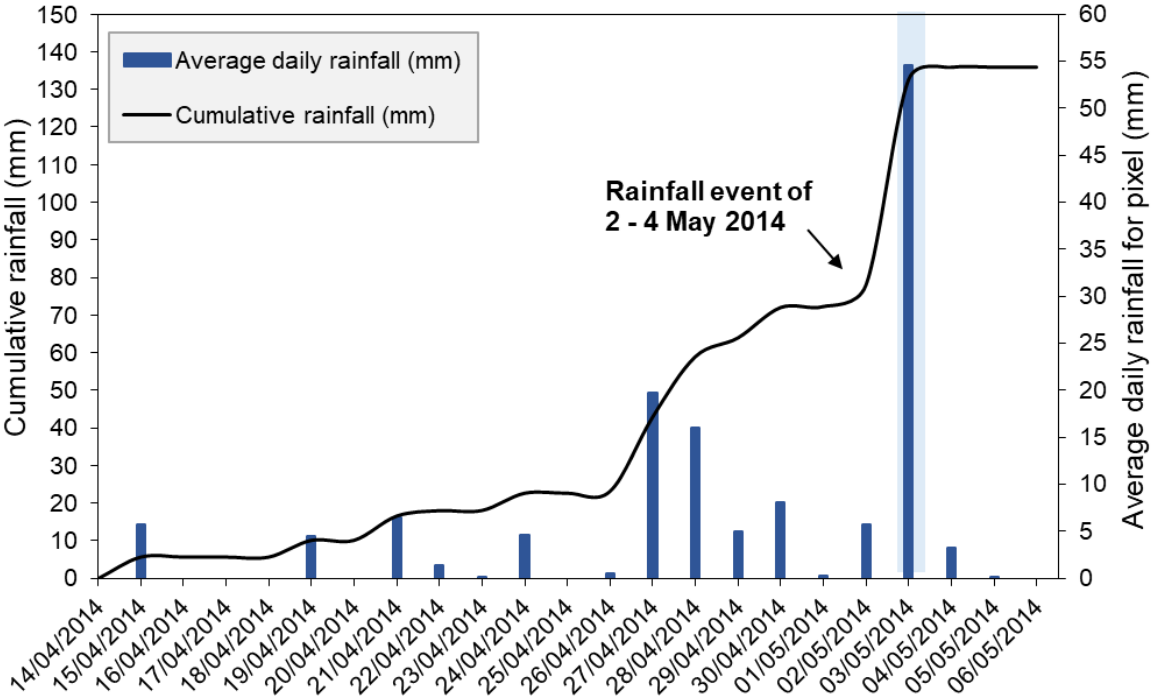


Figure 20. *Cumulative and average daily rainfall per pixel from 14/04/2014 to 06/05/2014 in the Misa River basin, the rainfall event is highlighted in light blue (data from rain gauges)*

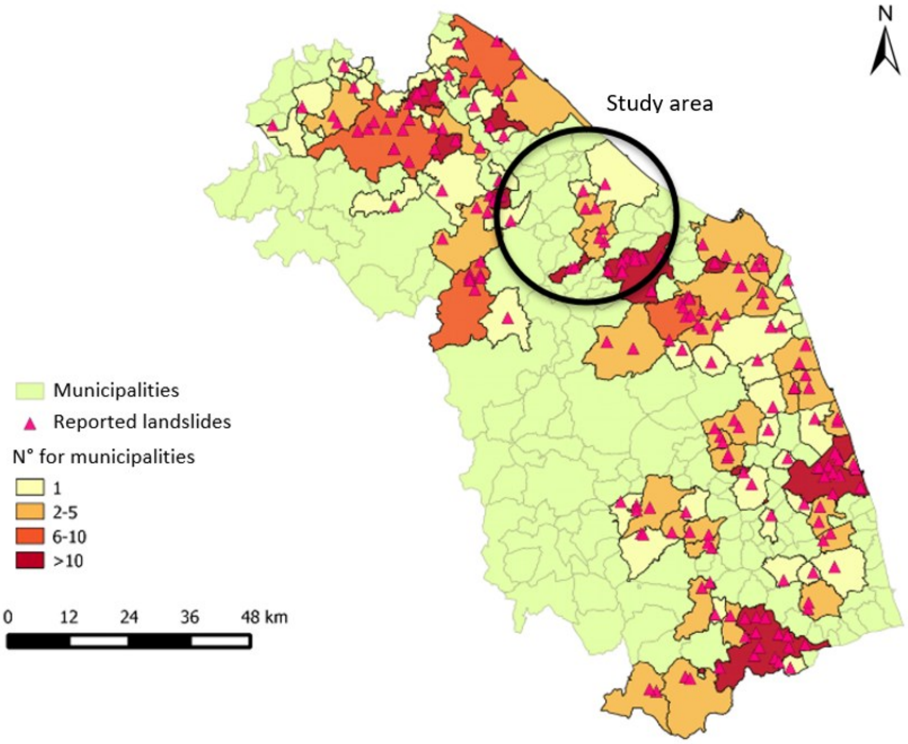


Figure 21. *Reported landslides during the event (modified from Marche Region Functional Centre, 2014)*

Figure 22. *Examples of reported landslides during the event (data from: Marche Region Functional Centre, 2014)*

Chapter 5

Application and Results

This chapter describes the application of the SADRI framework within the Misa River basin. The analysis follows a sequential workflow, starting with the preparation of input data, proceeding to the calibration of physical models and alert thresholds, and culminating in the dynamic simulation of risk scenarios.

The framework was applied by leveraging the September 2022 event as ground truth for defining model parameters. Once the activation thresholds and rheological properties were calibrated, they were used for the dynamic simulation of two historical scenarios (2022 and 2014), allowing for an assessment of the system's sensitivity to different rainfall events.

5.1 HIRESSS Model

5.1.1 Static Data

All data, both static and dynamic, were georeferenced and provided to the model in raster format with a uniform spatial resolution of 10 m. The parametrisation strategy adopted a lithology-based approach (Bicocchi et al. 2016; Tofani et al. 2017; Salvatici et al., 2018; Masi et al., 2023; Brilli et al., 2025), wherein each soil unit is assigned a specific set of geotechnical and hydrological values derived from the parent rock material, while accounting for relative uncertainties (see Section 3.3.1).

The geological characterisation relies on the national lithological map of Italy (Bucci et al., 2022), which was reclassified into seven distinct lithological classes to suit the modelling scale (*Figure 14*).

The contribution of vegetation to slope stability was modelled by incorporating root cohesion (c_r) into the Mohr-Coulomb failure criterion (Waldron 1977; Wu et al. 1979; Gray and Leiser 1982), as root systems primarily enhance soil cohesion without significantly altering the internal friction angle (Waldron and Dakessian 1981; Gray and Ohashi 1983; Operstein and Frydaman 2000; Giadrossich et al. 2010).

Root cohesion values were spatially distributed based on the CORINE Land Cover (CLC 2018) dataset, assigning specific values to dominant species derived from literature (Bischetti 2009; Burylo 2010; Vergani 2013).

A detailed Land Use Land Cover (LULC) map was then generated by harmonising the original CLC 2018 classification with the regional land-use map “Carta Uso del Suolo (1:10,000) - ADS40 2007” (available at: https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica/Cartografia/Repertorio/Cartausosuolo10000_2007), resulting into a ten macro-class map, consistent with the groupings adopted for the root-cohesion parametrisation (*Figure 23*).

Hydraulic properties, including bubbling pressure (h_b), pore size index (λ), and residual water content (θ_r), were derived using the empirical relationships proposed by Rawls et al. (1982).

Specifically, the hydraulic parameters associated with the Sandy Clay Loam textural class were assigned to lithologies characterised as sandy/silty clays (i.e., Poorly Lithified and Well-Cemented Clastic Rocks, Evaporites, and Siliciclastic Sedimentary Rocks). Similarly, the values corresponding to the Sandy Loam class were applied to Alluvial Deposits and Marlstones (clayey/silty sands), while the parameters characteristic of the Sand class were adopted for the sandy/silty gravels of Carbonate Rocks (*Figure 24*).

These parameters are critical for accurately modelling soil moisture dynamics and hydraulic behaviour during rainfall events. The slope angle map was derived from the 10 m resolution TINITALY Digital Elevation Model

(Tarquini et al., 2007). Geotechnical and hydrogeological properties, including cohesion (c), friction angle (φ°), hydraulic conductivity (k_s), porosity (n), and dry unit weight (γ_d), were determined through a review of literature, site-specific data, and expert criteria, as no ad hoc investigation campaigns for the present study were conducted to characterise the soils.

Specifically, parameter assignment relied on the integration of two main sources: a) data from seismic microzonation studies conducted in the local municipalities (available at https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica), which were analysed to derive mean representative values for the corresponding soil types; b) the calibrated input parameters reported in *Table 3* by Gioia et al. (2015) , which were adopted as reference values for compatible lithologies, being the result of a statistical analysis of literature data applied to the nearby Esino River basin.

Soil thickness was estimated using Geomorphologically Indexed Soil Thickness (GIST - Catani et al. 2010), a watershed-scale empirical model that integrates multiple morphometric attributes (such as curvature and relative slope position), geomorphological factors (topography, hydrology), and biological interactions (such as vegetation and animal activity) to calculate the soil accumulation potential for each grid cell, expressed as a dimensionless index between 0 and 1.

This index was subsequently linearly scaled to actual soil thickness (in metres) using minimum and maximum depth values observed for each specific lithology in the seismic microzonation logs (available at https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica).

Finally, to optimise the modelling process, a mask was created to exclude exposed rock outcrops, and flat areas, deemed insignificant for the purposes of the study due to low soil thickness or insufficient angles for landslides. This approach was developed using land use and soil thickness maps, slope data, and satellite imagery, merged via specific GIS procedures.

The result is an analysis that focuses on areas of greater relevance, thereby enhancing the accuracy of the results. *Table 15* provides a comprehensive summary of the input parameters, data sources, and methodologies adopted in the study for their assessment. The specific values assigned to the static input raster maps are detailed in *Table 16*, while *Figure 25* illustrates their spatial distribution.

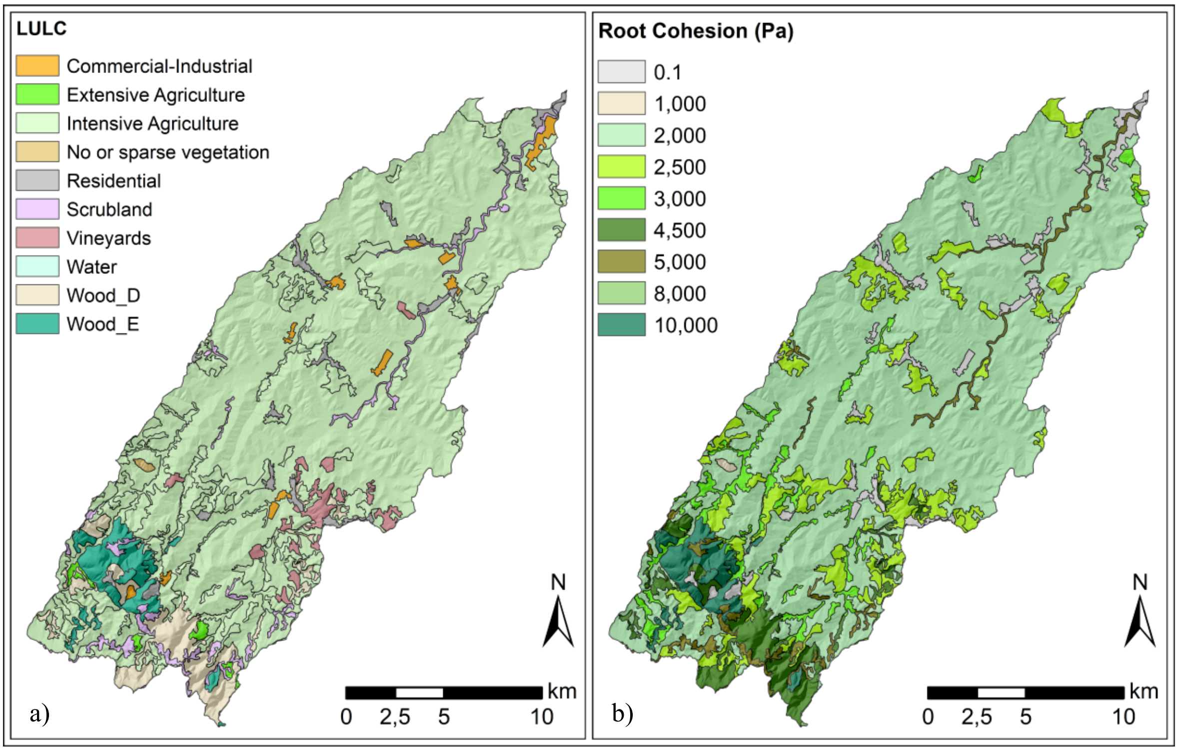
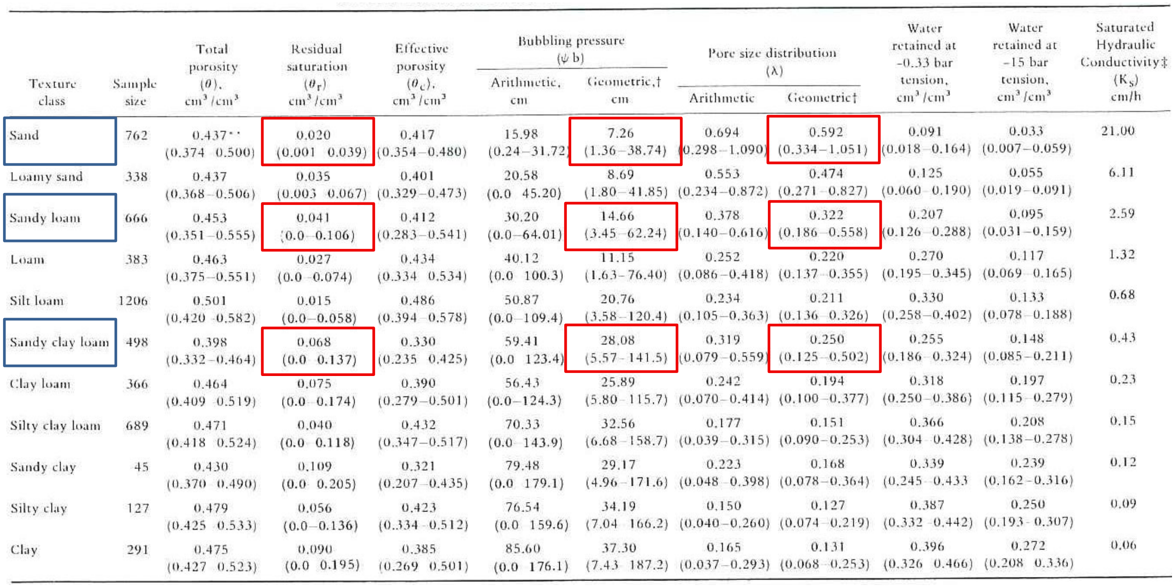


Figure 23. *Figure a) Land Use Land Cover (LULC) map derived by aggregating the original CLC 2018 data with the regional land-use map "Carta Uso del Suolo - ADS40 2007" into ten classes. Figure b) Root Cohesion (c_r) values for the different forest types. Wood_D=Wood (Deciduous & Mixed); Wood_E=Wood (Evergreen & Mixed)*



Texture class	Sample size	Total porosity (θ), cm³/cm³	Residual saturation (θ_r) cm³/cm³	Effective porosity (θ_e), cm³/cm³	Bubbling pressure (ψ b)		Pore size distribution (λ)		Water retained at −0.33 bar tension, cm³/cm³	Water retained at −15 bar tension, cm³/cm³	Saturated Hydraulic Conductivity‡ (K_s) cm/h
					Arithmetic, cm	Geometric,† cm	Arithmetic	Geometric†			
Sand	762	0.437** (0.374–0.500)	0.020 (0.001–0.039)	0.417 (0.354–0.480)	15.98 (0.24–31.72)	7.26 (1.36–38.74)	0.694 (0.298–1.090)	0.592 (0.334–1.051)	0.091 (0.018–0.164)	0.033 (0.007–0.059)	21.00
Loamy sand	338	0.437 (0.368–0.506)	0.035 (0.003–0.067)	0.401 (0.329–0.473)	20.58 (0.0–45.20)	8.69 (1.80–41.85)	0.553 (0.234–0.872)	0.474 (0.271–0.827)	0.125 (0.060–0.190)	0.055 (0.019–0.091)	6.11
Sandy loam	666	0.453 (0.351–0.555)	0.041 (0.0–0.106)	0.412 (0.283–0.541)	30.20 (0.0–64.01)	14.66 (3.45–62.24)	0.378 (0.140–0.616)	0.322 (0.186–0.558)	0.207 (0.126–0.288)	0.095 (0.031–0.159)	2.59
Loam	383	0.463 (0.375–0.551)	0.027 (0.0–0.074)	0.434 (0.334–0.534)	40.12 (0.0–100.3)	11.15 (1.63–76.40)	0.252 (0.086–0.418)	0.220 (0.137–0.355)	0.270 (0.195–0.345)	0.117 (0.069–0.165)	1.32
Silt loam	1206	0.501 (0.420–0.582)	0.015 (0.0–0.058)	0.486 (0.394–0.578)	50.87 (0.0–109.4)	20.76 (3.58–120.4)	0.234 (0.105–0.363)	0.211 (0.136–0.326)	0.330 (0.258–0.402)	0.133 (0.078–0.188)	0.68
Sandy clay loam	498	0.398 (0.332–0.464)	0.068 (0.0–0.137)	0.330 (0.235–0.425)	59.41 (0.0–123.4)	28.08 (5.57–141.5)	0.319 (0.079–0.559)	0.250 (0.125–0.502)	0.255 (0.186–0.324)	0.148 (0.085–0.211)	0.43
Clay loam	366	0.464 (0.409–0.519)	0.075 (0.0–0.174)	0.390 (0.279–0.501)	56.43 (0.0–124.3)	25.89 (5.80–115.7)	0.242 (0.070–0.414)	0.194 (0.100–0.377)	0.318 (0.250–0.386)	0.197 (0.115–0.279)	0.23
Silty clay loam	689	0.471 (0.418–0.524)	0.040 (0.0–0.118)	0.432 (0.347–0.517)	70.33 (0.0–143.9)	32.56 (6.68–158.7)	0.177 (0.039–0.315)	0.151 (0.090–0.253)	0.366 (0.304–0.428)	0.208 (0.138–0.278)	0.15
Sandy clay	45	0.430 (0.370–0.490)	0.109 (0.0–0.205)	0.321 (0.207–0.435)	79.48 (0.0–179.1)	29.17 (4.96–171.6)	0.223 (0.048–0.398)	0.168 (0.078–0.364)	0.339 (0.245–0.433	0.239 (0.162–0.316)	0.12
Silty clay	127	0.479 (0.425–0.533)	0.056 (0.0–0.136)	0.423 (0.334–0.512)	76.54 (0.0–159.6)	34.19 (7.04–166.2)	0.150 (0.040–0.260)	0.127 (0.074–0.219)	0.387 (0.332–0.442)	0.250 (0.193–0.307)	0.09
Clay	291	0.475 (0.427–0.523)	0.090 (0.0–0.195)	0.385 (0.269–0.501)	85.60 (0.0–176.1)	37.30 (7.43–187.2)	0.165 (0.037–0.293)	0.131 (0.068–0.253)	0.396 (0.326–0.466)	0.272 (0.208–0.336)	0.06

***Figure 24**. Hydraulic parameters (h_b, λ, θ_r) obtained from Rawls et al. (1982) for the texture classes identified with respect to each lithological unit. The blue rectangle identifies the reference soil texture class, whereas the red rectangle indicates the corresponding parameter values adopted*

***Table 15.** Summary of the input data and methodology employed*

Input parameters	Methodology
Slope Angle	Derived from TINITALY DEM 10 m (Tarquini et al., 2007)
Soil Thickness	GIST model (Catani et al., 2010) calibrated based on local seismic logs available at: https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica
Geotechnical and Hydrogeological	Data Integration from Gioia et al. (2015) and local microzonation studies available at: https://www.regione.marche.it/Regione-Utile/Paesaggio-Territorio-Urbanistica
Hydraulic	Empirical relationships by Rawls et al. (1982) based on soil texture
Root Cohesion	Plant species distribution from the CLC 2018 land cover map, cohesion values from literature (Bischetti 2009; Burylo 2010; Vergani 2013)
Mask	GIS-based exclusion of rock outcrops and flat areas

***Table 16.** Input parameters values assumed for the lithological classes in the HIRESSS simulations*

Lithological class	Soil Texture	γ_d (kN/m^3)	θ_r (-)	h_b (mH$_2$O)	λ (-)	n (%)	k_s (m/s)	φ (°)	c (kPa)
Alluvial deposits	Clayey/silty sands	19.6	0.041	0.1466	0.322	25	1.0E-06	31	0
Poorly lithified clastic rocks	Sandy/silty clay	19.4	0.068	0.2808	0.250	30	1.0E-06	23	1
Well-cemented clastic rocks	Sandy/silty clays	19.4	0.068	0.2808	0.250	30	1.0E-06	23	2
Evaporite	Sandy/silty clays	19.4	0.068	0.2808	0.250	30	1.0E-06	23	1
Siliciclastic sedimentary rocks (Flysch)	Sandy/silty clays	19.4	0.068	0.2808	0.250	25	1.0E-07	23	1
Marlstone	Clayey/silty sand	19.6	0.041	0.1466	0.322	25	1.0E-06	31	0
Carbonate rocks	Sandy/silty gravel	19.8	0.020	0.0726	0.592	25	1.0E-04	35	0

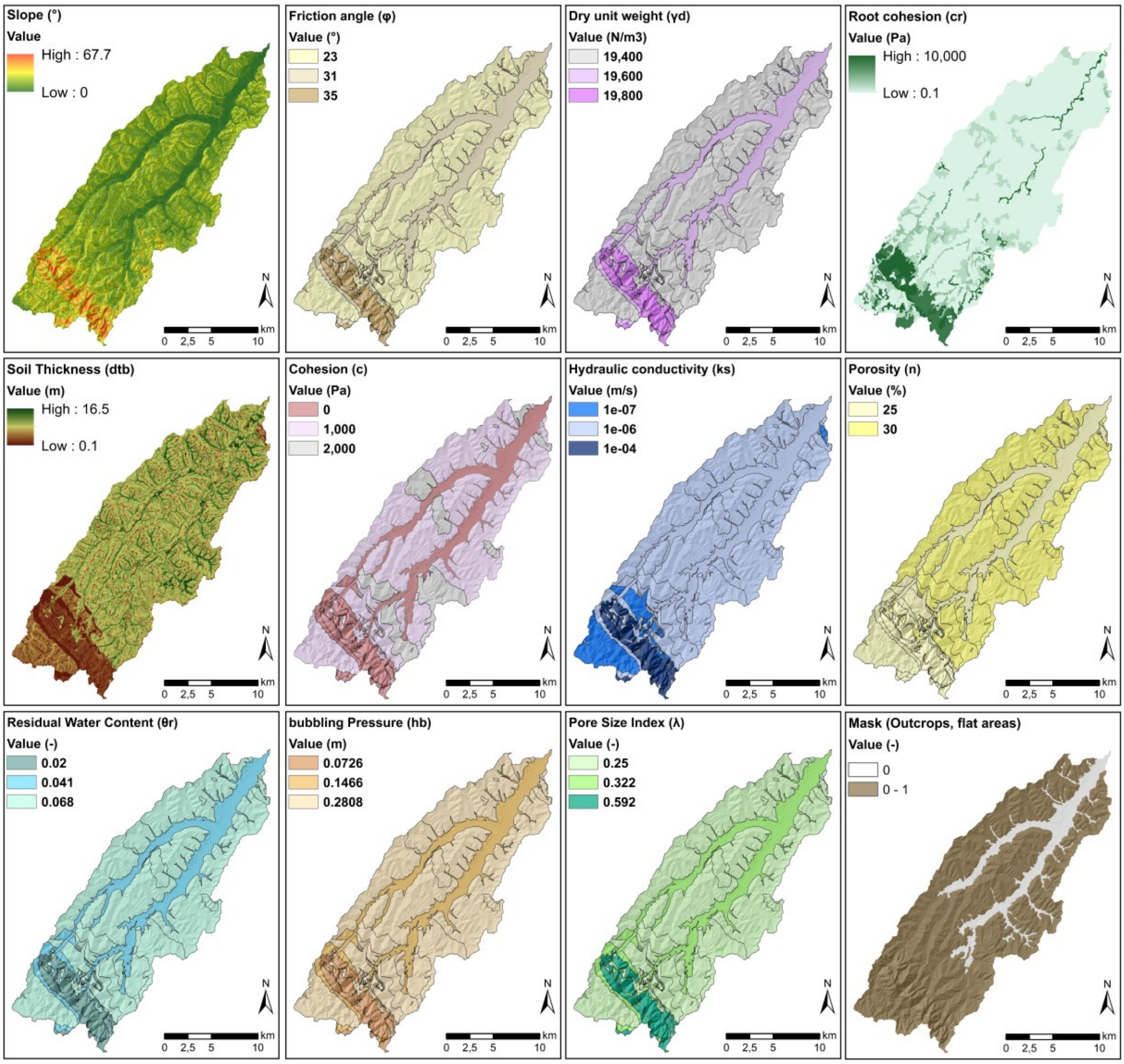


***Figure 25**. Raster maps of the static input data used for the HIRESSS simulations*

5.1.2 Dynamic Data

The study utilised two distinct hourly rainfall datasets, representing two opposing event types, to assess the sensitivity of both HIRESSS and the SADRI system to different hydrometeorological dynamics:

- **Scenario A** (Calibration Event): The event of September 15-16, 2022. This case featured a stationary V-shaped convective system that discharged exceptional precipitation amounts (>400 mm in 12 hours) over a localised area. The widespread landslides triggered across the basin led to the selection of this event for calibrating the entire SADRI system. The resulting landslide inventory, mapped according to Confuorto et al. (2025), provides a robust dataset that enables a rigorous and accurate validation of the model results;

- **Scenario B** (Test Event): The event of May 2-4, 2014. In contrast to the calibration event, this scenario was characterised by widespread, moderate-intensity, yet prolonged precipitation (lasting up to 48 hours) affecting soils already saturated by antecedent rainfall (Marche Region Functional Centre, 2014). Consequently, this scenario serves as a validation test to assess model performance under distinct hydrological conditions.

Due to historical data availability, the source of rainfall data differs between the two scenarios. For Scenario A (September 2022), rainfall data were retrieved from the Italian National Radar Network (available at: https://rischi.protezionecivile.gov.it/it/meteo-idro/monitoraggio/rete-radar/). The analysis processed the raw radar maps (originally provided with a 10-minute temporal resolution and values in decimillimetres) by aggregating them into hourly cumulative values and converting the data to millimetres to ensure consistency with the physical model requirements.In contrast, the analysis for Scenario B (2014) relied on data recorded by a network of seven specific rain gauge stations (*Table 17*; *Figure 26*). The use of rain gauge records was necessitated by significant gaps and inconsistencies in the available radar data for the May 2014 period, which made the retrieval of a continuous radar dataset impossible. Therefore, rainfall time series were retrieved from the Marche Region Civil Protection website (available at: http://app.protezionecivile.marche.it/sol/indexjs.sol?lang=it) with a 15-minute temporal resolution and subsequently aggregated into hourly maps. To transform point-source rainfall data into raster maps spatially consistent with the static input parameters, the Thiessen polygon technique (Thiessen, 1911), was applied (*Figure 26*). This method divides the basin into discrete influence zones based on the nearest neighbour principle: the rainfall intensity recorded by a gauge is extended to the surrounding area for which that station represents the closest reference point. *Figure 27* displays two examples of hourly rainfall maps, corresponding to specific time steps during the peak intensity of the events.

Table 17. *Geographical localisation of the rain gauges used in the simulations*

Sensor code	Longitude	Latitude	Elevation (m)	Station code	Station name
2736	13°12'	43°42'	6.00	606	Senigallia
2637	13°09'	43°39'	26.47	26	Bettolelle
2964	13°02'	43°39'	218.00	705	Corinaldo
2144	13°02'	43°38'	203.00	1790	Corinaldo
1270	13°03'	43°32'	350.00	119	Colle
2858	13°01'	43°34'	186.00	618	Barbara
1295	12°56'	43°29'	535.00	123	Arcevia

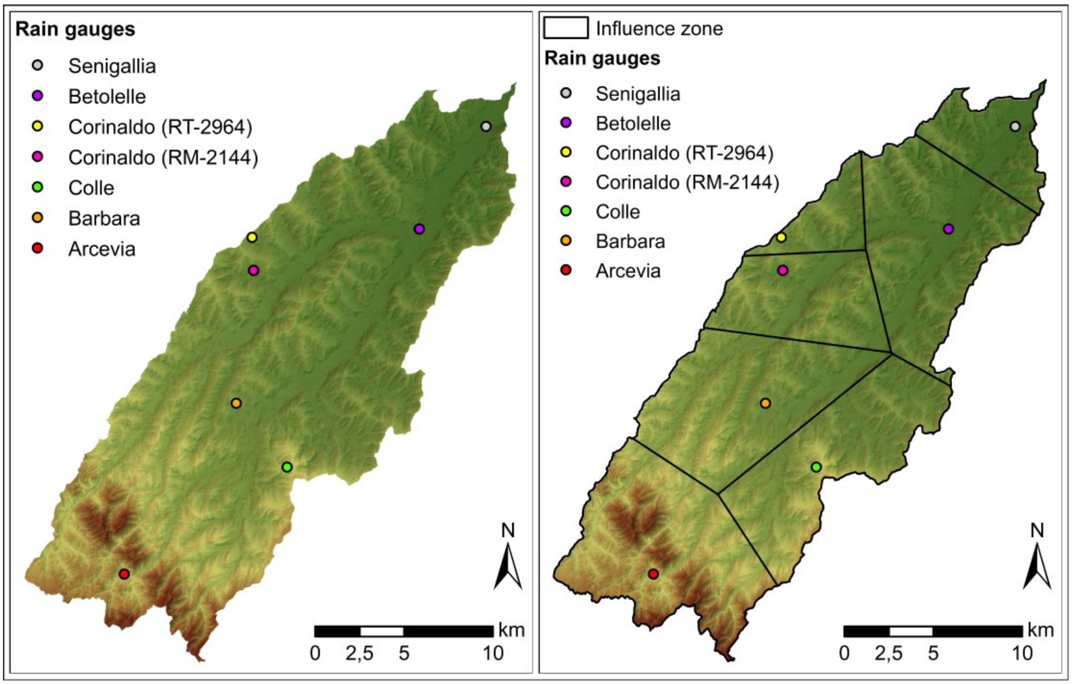


***Figure 26**. Thiessen polygon methodology applied to seven rain gauges in study area*

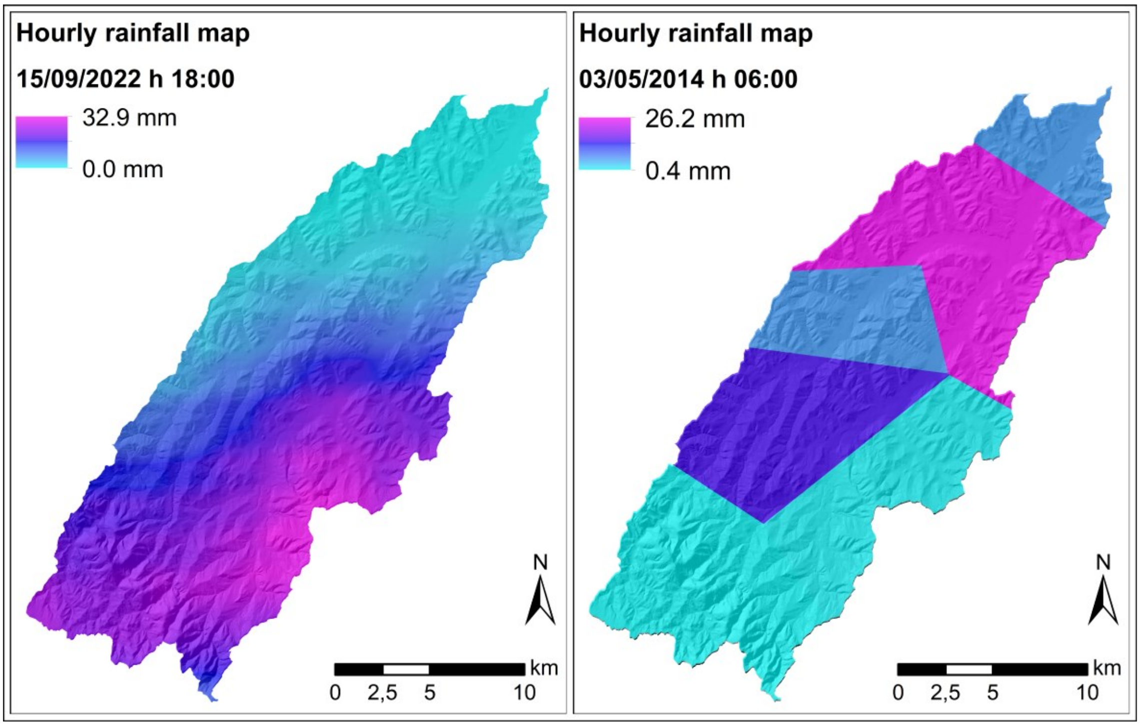


Figure 27. *Dynamic input data. The map on the left represents an example of a raster map of hourly precipitation for the time step "September 15, 2022, 6:00 PM" (Calibration Event, Scenario A), while the one on the right represents the time step "May 3, 2014, 6:00 AM" (Test Event, Scenario B)*

5.1.3 HIRESSS simulation results

As aforementioned, utilising the static and dynamic input raster data that has been previously described, the HIRESSS model generates the following distributed maps for each day of the simulated scenario:

- ✓ Twenty-four-hourly maps of the Failure Probabilities (P_f) for the pixels (probability of the FS being lower than 1) and 24-hourly maps of the average FS computed for each pixel during the Monte Carlo iterations;
- ✓ A daily map showing the maximum failure probability of each pixel calculated within 24 hours and a daily map showing the lowest average FS calculated for each pixel during the day (i.e., for each pixel, the FS of the time step resulting in the lowest value in the day is reported).

The technical specifications and primary characteristics of the two simulations are outlined in *Table 18*.

Table 18. *Main characteristics and results of the HIRESSS simulations concerning the days of the rainfall scenarios*

	Scenario A **15-16 September 2022**	**Scenario B** **02-04 May 2014**
HIRESSS Model		
Spatial Resolution	10 m	10 m
Time Step	1 h	1 h
Rainfall Data (h)	mm	mm
N° of Soil Layers	5	5
Monte Carlo Iterations	1,000	1,000
FS- P_f Hourly Maps	72	48
FS- P_f Daily Maps	3	2

The HIRESSS model performs stability analysis by discretising the vertical soil profile into user-defined depth intervals. Specifically for this study, the subsurface was segmented into 5 homogeneous layers.

The algorithm operates iteratively, computing the Factor of Safety (FS) of each pixel at the base of each layer.

To ensure a conservative stability assessment, the model identifies the minimum FS value across the entire vertical profile and assigns it to the pixel as the representative critical indicator.

Due to the lack of distributed field measurements, the simulation was initialised with a 'dry' state, setting the water table at the bedrock interface.

To reconstruct a realistic moisture profile, a two-week period preceding each rainfall event was simulated using antecedent rainfall.

The adequacy of this timeframe is substantiated by both the landslide characteristics and the model's observed dynamics.

The failures were predominantly shallow slides evolving into mud-debris flows (2-5 m thick in the triggering area) involving granular deposits.

Crucially, minimal flow accumulation values ($10\text{-}10^3$ m^2) were recorded in the source areas (Section 5.2.1), indicating a hydrological system driven by direct vertical infiltration rather than lateral flow.

This dominance of vertical infiltration, coupled with moderate permeability (mostly 10^{-6}) and limited soil thickness, ensures a rapid hydraulic response, consistent with the HIRESSS simulation results, which demonstrated the rapid dissipation of excess pore pressures during dry intervals.

Consequently, a two-week forcing period is considered sufficient for the system to reach dynamic equilibrium.

Regarding the parameterisation of geotechnical and hydrogeological properties, the official mean values adopted in the model are listed in *Table 16*.

In the absence of site-specific field measurements to quantify the local standard deviation, the variability ranges required for the Monte Carlo simulation were established based on the findings of Rossi (2010).

This study observed that the relative variability of soil parameters remains consistent across spatial scales. Consequently, relying on this empirical evidence, the probability distributions for Scenarios A and B were generated by perturbing the reference mean values according to the standard relative errors listed in *Table 19*.

Furthermore, the number of Monte Carlo iterations was set to 1,000.

This value is adopted as the optimal compromise between the statistical convergence of the calculated failure probability and the computational efficiency required for regional-scale simulations (Rossi, 2010; Masi et al., 2023)

Table 19. *Relative Variation of Static Parameters*

Rel. Error	c' (Pa)	φ' (°)	Slope (°)	n (%)	γ_d (N/m^3)	Dtb (m)	k_{sat} (m/s)	λ (-)	h_b (m)	I. Pore (-)
(%)	40	20	20	20	20	20	60	30	20	30

Dtb = Depth to bedrock (Soil Thickness); I. Pore = Pore Size Index

Scenario A: Calibration Event (15-16 September 2022)

The results obtained from the simulation of the September 15-16, 2022 event enabled the validation of the physically based model's behaviour under extreme rainfall conditions.

The spatial distribution of the predicted instability (*Figure 28*) highlights a clear concentration of high failure probability values (P_f > 90%) in the central-southern sector of the Misa River basin, generally reflecting the area most severely affected by the storm.

Detailed observations reveal a significant spatial agreement in high-susceptibility zones, where probability values frequently overlap or lie in proximity to the observed detachment zones, while northern regions correctly exhibit low instability.

In particular, a highly susceptible band coinciding with the marlstone lithology is observed, clearly highlighted by the model, suggesting a strong influence of local geotechnical characteristics on the response to the event.

Regarding the temporal dynamics, shown in *Figure 29a,* the number of unstable pixels remains negligible during the warm-up phase but exhibits a sharp, vertical increment coinciding exactly with the critical rainfall on September 15.

This confirms the model's rapid hydrological response and its alignment with the actual timing of the disaster. The magnitude of this forcing is further quantified in *Figure 29b*, which contrasts the near-zero distribution of pixels with P_f values greater than 80% in the pre-event phase with the surge observed at the event peak.

It is important to note that, in the aftermath of the event, the situation reverts to pre-event levels, thereby effectively simulating the dissipation of excess pore pressures once the rainfall forcing has been terminated.

Finally, the spatial accuracy of the model was quantitatively assessed using a buffer-based validation, as presented in *Figure 30*.

The graph plots the percentage of correctly identified landslides (y-axis) against increasing Failure Probability Thresholds (FPT) (x-axis), comparing the strict pixel-by-pixel intersection (0 m buffer) with spatial tolerances of increasing radii (buffers).

The analysis highlights two key trends:

- <u>Effect of Spatial Tolerance:</u>
  The strict pixel-by-pixel validation (lowest curve, *Figure 30*) captures a moderate portion of the events, but the detection rate improves significantly when a 50 m buffer is applied.
  This marked improvement in performance indicates that, while precise pixel location is constrained by the input data resolution (10 m DEM), the model effectively identifies instability within the immediate vicinity of the slope unit;

- <u>Sensitivity to Probability Thresholds:</u>
  A thorough examination of the curves along the x-axis reveals that the model exhibits a high recognition rate for probability thresholds up to 50%.
  However, a decline in performance is observed when restricting the validation to pixels with $P_f > 50\%$. This finding suggests that, while the model accurately identifies the majority of landslides as "unstable," it does not consistently attain the highest certainty levels ($P_f > 80\text{-}90\%$) for all events.

This behaviour suggests that the physical simplification (infinite slope) and the DEM resolution tend to smooth out the steepest real slopes, resulting in a systematic underestimation of the absolute risk values.

Consequently, while the model accurately identifies critical areas, it is conservative. Therefore, a robust warning system must rely on detecting the onset of instability (medium P_f values) rather than waiting for extreme probability peaks.

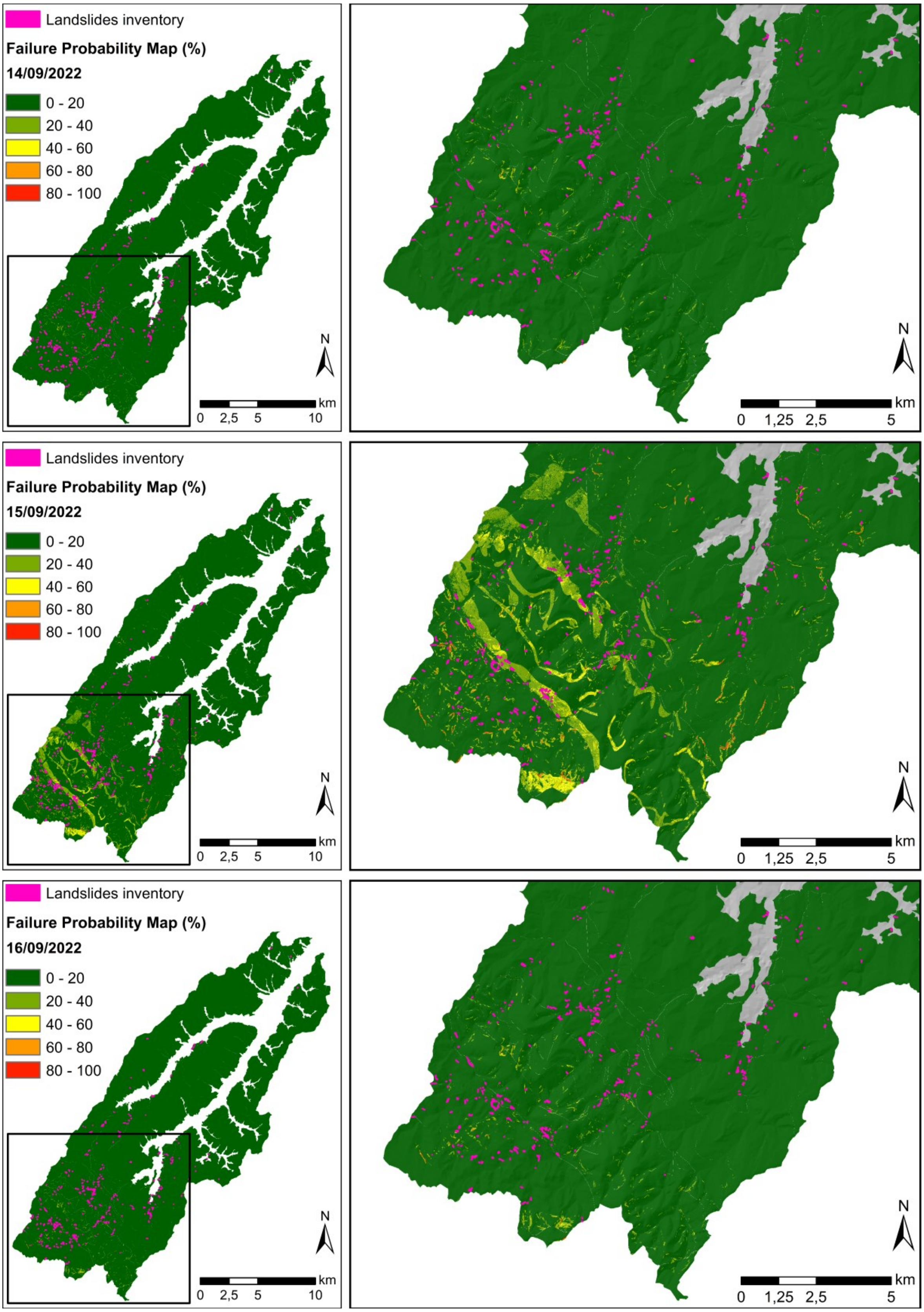


Figure 28. *Failure probability maps for the September 2022 event. Purple polygons indicate reported landslides, while yellow, orange, and red areas represent increasing failure probabilities*

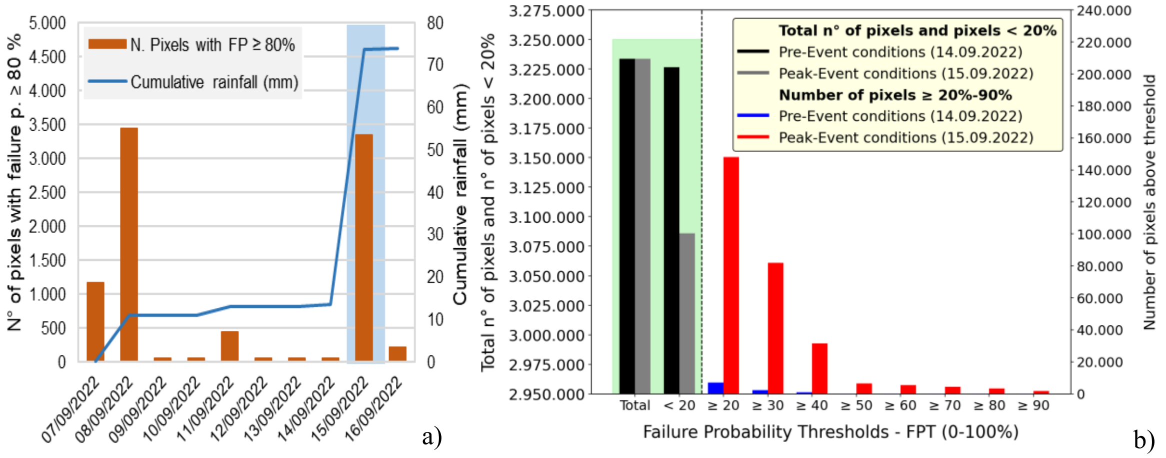


***Figure 29**. Temporal dynamics of instability for Scenario A. a) Daily evolution of unstable pixels (P_f > 80%) in relation to precipitation intensity; b) Comparison of pixel counts above increasing failure probability thresholds, highlighting the shift in instability levels from pre-event to post-event conditions*

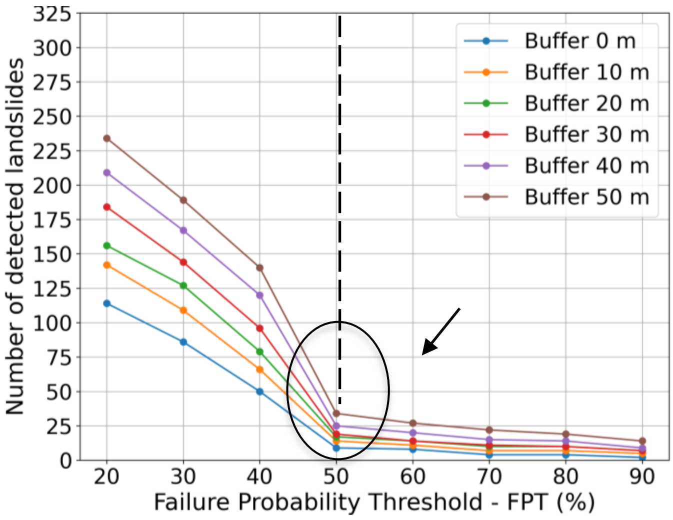


***Figure 30**. Relationship between the landslide detection rate and Failure Probability Thresholds (FPT) for increasing spatial buffers around the reported original landslide polygon (0-50 m). The black circle highlights the sharp spatial drop in landslide detection rate*

Scenario B: Test Event (02-04 May 2014)

The rainfall event of May 2-4, 2014 was simulated as a functional test. Unlike the calibration scenario, the system operated under prolonged precipitation and antecedent saturated soil conditions, which enabled an analysis of the system's behaviour under different hydrometeorological forcing.

As illustrated in *Figure 31*, the spatial distribution of the maximum Failure Probability (P_f) shows a hazard pattern spatially coherent with the susceptibility characteristics of the Misa River basin.

The areas most prone to landslide triggering are similar between the two simulated rainfall scenarios, however, unlike the September 2022 event, this episode shows hotspots with P_f values exceeding 90% in the central-eastern sectors of the basin, which accurately reflect the distinct rainfall footprint of this event.

In contrast to the calibration event, the analysis of temporal dynamics for this scenario (*Figure 32*) demonstrates a more complex antecedent behaviour.

The number of unstable pixels (P_f > 80%) showed a substantial increase several days before the main event, indicating the influence of minor antecedent rainfall leading to soil pre-saturation conditions. However, a drastic and temporary decrease in the number of unstable pixels is recorded on the day immediately preceding the arrival of the principal storm front (*Figure 32a*). Subsequently, in accordance with the onset of the rainfall event on May 2-3, there was a drastic increase in the number of critical pixels, following the peak of rainfall intensity (*Figure 32b*). The model shows notable sensitivity to variations in rainfall input which enables it to effectively simulate the transient hydraulic response of the slope. The system shows rapid dissipation of excess pore pressures which results in P_f values shifting toward stability levels when the system transitions from phases of intense precipitation to dry periods.

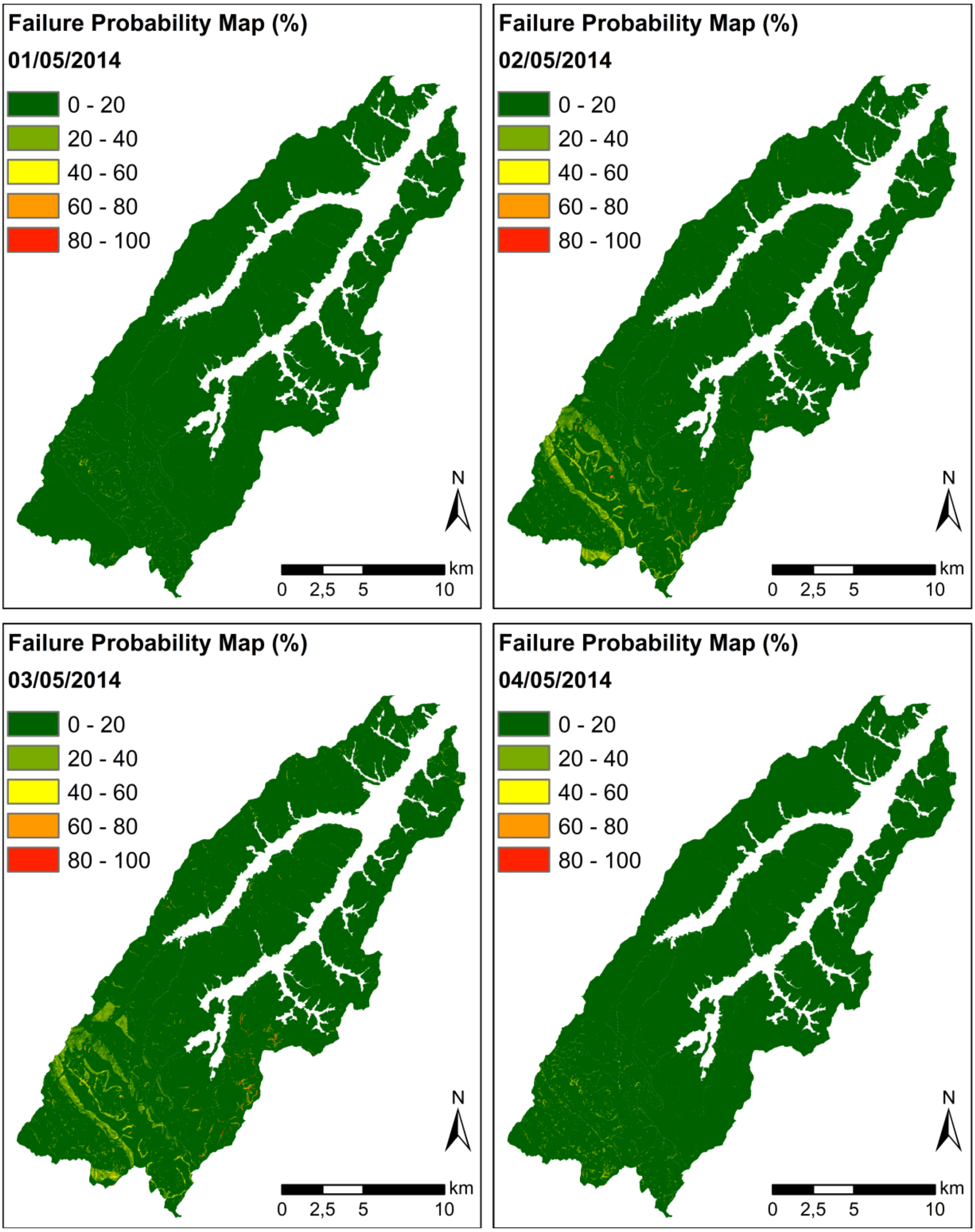
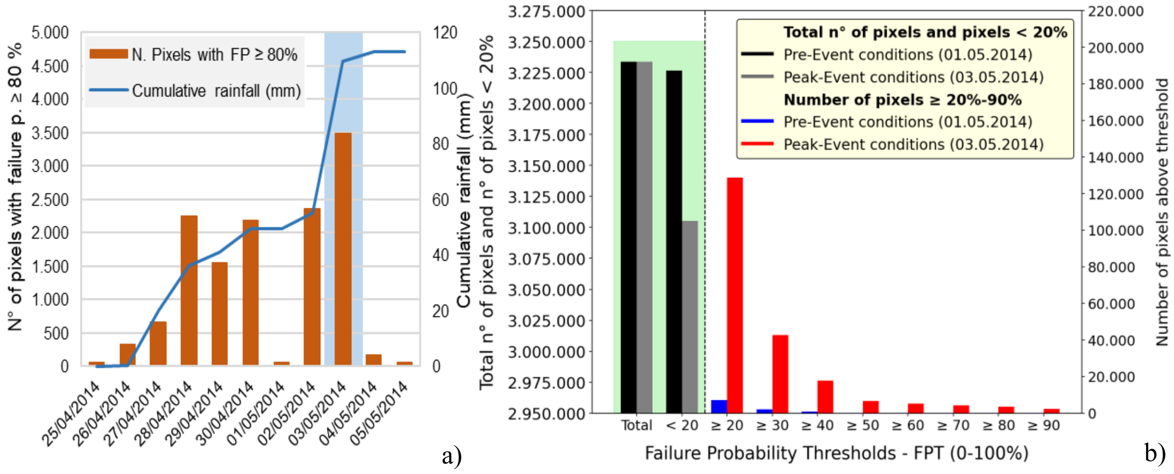


***Figure 31**. Failure probability maps for the May 2014 event. Purple polygons indicate reported landslides, while yellow, orange, and red areas represent increasing failure probabilities*



a) b)

Figure 32. *Temporal dynamics of instability for Scenario B. a) Daily evolution of unstable pixels ($P_f > 80\%$) in relation to precipitation intensity; b) Comparison of pixel counts above increasing failure probability thresholds, highlighting the shift in instability levels from pre-event to post-event conditions*

5.2 Definition of Alert Levels

Developing a binary warning system (stable/unstable) from continuous probability maps requires precise calibration. For the purposes of this study, the 15–16 September 2022 rainfall scenario in the Misa River basin was used as the ground truth for training the system, due to the availability of a high-quality landslide inventory.

5.2.1 Landslides inventory

Establishing reliable warning thresholds requires comparing model predictions against actual observed landslides. The validation dataset used in this thesis is derived from the event-based landslide inventory documented by Confuorto et al. (2025), as described in Section 4.2.1 (*Figure 17-18*).

The inventory was constructed through a mixed-methodological approach, combining the semi-automatic analysis of pre- and post-event Very High Resolution (VHR) satellite imagery with in-situ field surveys for validation.

The dataset includes a total of 805 landslides, of which 701 occurred within the Misa River basin, covering an area of approximately 494,000 square metres and with an average landslide size of 614 square metres.

In this thesis, only landslides with triggering mechanisms suitable for HIRESSS simulation were selected from the inventory, resulting in a subset of 568 shallow landslides and mudflows.

Landslides were initially mapped as generic polygons without distinguishing between source, detachment, and accumulation zones.

Therefore, the procedure outlined in Section 3.5 was applied to automatically extract their source areas. The resulting areas were then analysed to identify the factors controlling their occurrence.

This step evaluated the relationship between the mapped landslides and key environmental variables, including altitude, slope, and flow accumulation (derived from the 10 m resolution DEM), land use (CLC 2018), and lithology.

As illustrated in *Figure 33*, the majority of triggered events share a distinct set of environmental characteristics. From a morphometric perspective, instability is concentrated at altitudes ranging from 150 to 500 metres and on slopes between 5° and 40°, typically in areas characterised by minimal flow accumulation (10-10^3 m^2). Regarding land cover and geological setting, the analysis highlights a prevalence of failures in areas of intensive agriculture, typically associated with root cohesion values between 2 and 3 kPa, and primarily involving Quaternary deposits overlying Flysch (Siliciclastic Sedimentary Rock).

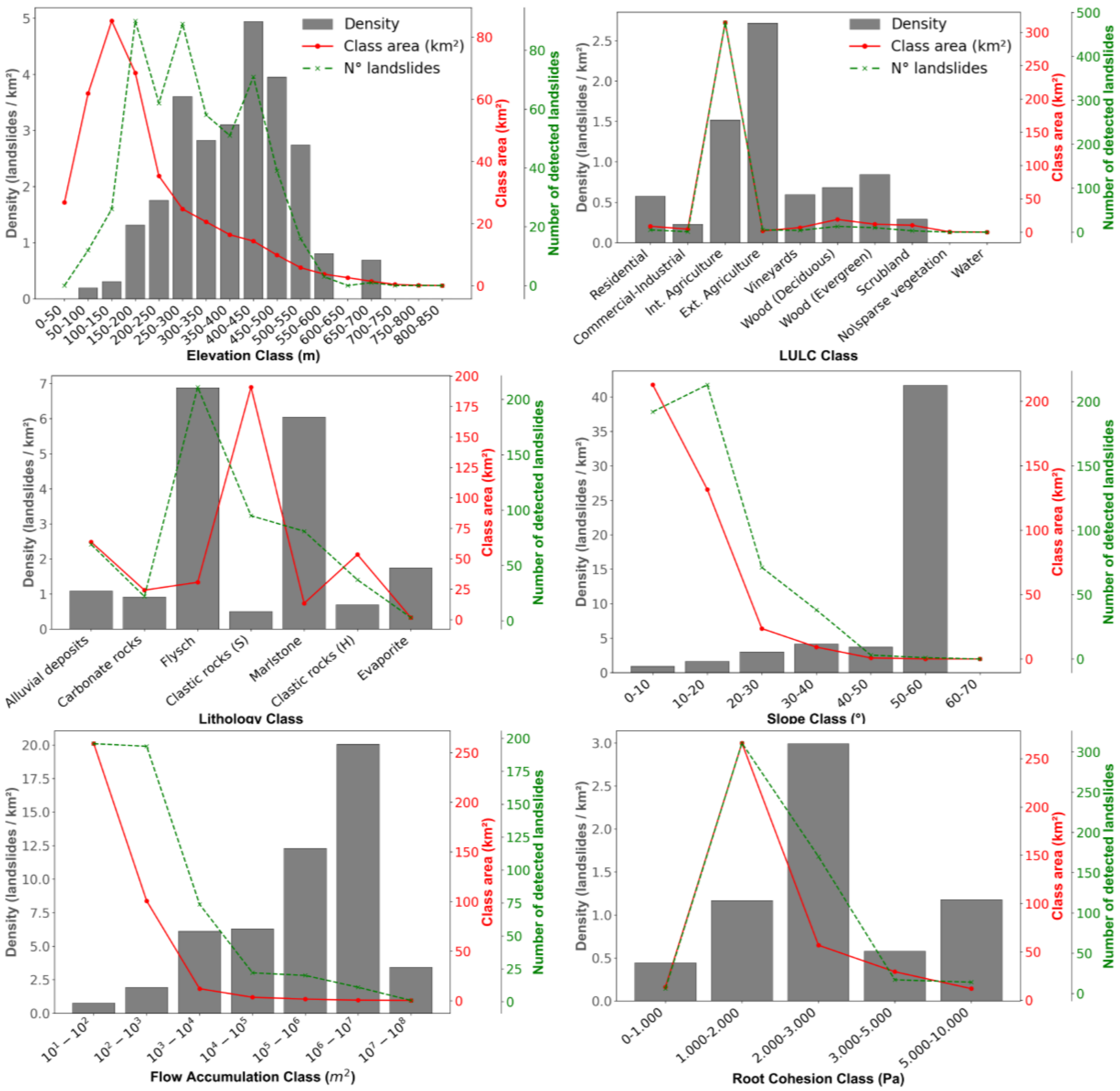


Figure 33. *Statistical analysis of the environmental characteristics of landslide source areas defined with the SEARCH framework (as described in Section 3.5), based on the inventory of the September 15-16, 2022 event in the Misa River basin. The data shows the distribution of landslides as a function of key environmental variables (altitude, slope, flow accumulation, land use, and lithology). The red line represents the total area (m^2) of the landslide source areas for each variable class. The green dashed line indicates the number of detected landslide events for each class. The grey column represents the landslide density, expressed as landslides/km^2*

5.2.2 Definition of Operating Thresholds

The process begins by identifying the spatial units to be used for the aggregation, which serve as the basis for calibrating thresholds and validating the HIRESSS model results.

Validation focuses on assessing the model's capability to correctly discretise stable and unstable conditions based on defined pairs of probability (FPT) and spatial (IDT) thresholds.

By comparing the outcomes of these threshold combinations against the actual landslide distribution recorded in September 2022 rainfall event, the system's performance is quantified using a contingency table (TP, TN, FP, FN).

The calibration workflow consists of three distinct steps:

a) Definition of spatial aggregation units: Establishment of the fundamental spatial domains that constitute the operational resolution of the warning system. These units define the spatial scale at which alert thresholds are applied and where model validation is performed against the landslide inventory;

b) Global Failure Probability Threshold (FPT) selection: Selection of the pixel-level instability threshold. This process converts the continuous probability maps into a binary classification by fixing an optimal cut-off. Consequently, individual pixels with probability values exceeding the FPT are classified as "unstable," whereas those falling below the threshold remain “stable”;

c) Operational warning levels definition: Defining the unit-level spatial threshold. With the FPT threshold previously selected, this step determines the spatial extent of instability within a spatial unit required to trigger a warning.

Spatial Aggregation Units

The preliminary step in establishing the validation framework involves defining the spatial aggregation units. To evaluate the model's performance on a geomorphologically meaningful scale, Slope Units (SUs) were adopted as the fundamental spatial domains within the SADRI framework.

The delineation of these units was performed using the *r.slopeunits* software, following the adaptive partitioning methodology proposed by Alvioli et al. (2016).

This approach allows for the automatic segmentation of the territory into hydrological and geomorphological units starting from a Digital Elevation Model (DEM). For the Misa River basin, the topographic input consisted of a DEM with a resolution of 10 × 10 m.

To ensure that the resulting units were morphometrically consistent and suitable for the specific slope-scale analysis of this study, the configuration parameters were selected in accordance with the optimisation strategy applied by Caleca (2024). The specific parameters employed for the shapefile generation are as follows:

- Minimum Area: 15,000 m^2;
- Circular Variance: 0.4;
- Clean Size: 10,000 m^2;
- Threshold: 20,000 m^2;
- Reduction Factor: 10.

This procedure resulted in the subdivision of the basin into Slope Units representing homogeneous hydrographic sub-basins or distinct slope portions. These entities serve as the functional link between the triggering analysis (HIRESSS) and the subsequent runout propagation.

The automated segmentation initially yielded 2,049 SUs, characterised by an average area of 0.2 km^2 and a size variability of 0.19 km^2 (expressed as a single standard deviation).

Subsequently, a post-processing filtering phase was applied. In accordance with the computational mask defined during the input data generation (see Section 5.1), units falling within excluded areas (e.g., flatlands, dense urban areas, or water bodies) were removed.

This refinement reduced the final dataset to 1,768 Slope Units (*Figure 34*), which constitute the valid domain for the instability calculations.

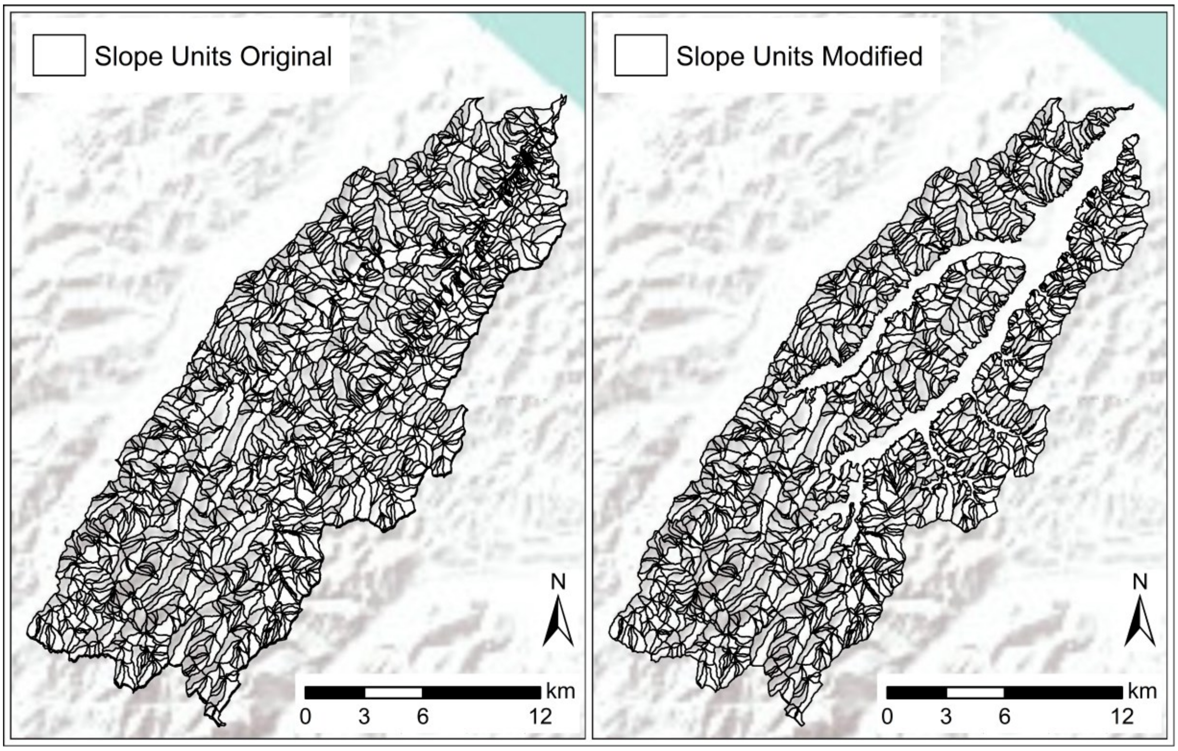


__Figure 34.__ Definition and refinement of the Slope Units (SUs) utilised as the spatial aggregation and validation units for the warning system. Left figure: The initial Slope Units (SUs) obtained by applying the adaptive partitioning methodology by Alvioli et al. (2016) to the Misa River basin. Right figure: The final Slope Units after the post-processing filtering phase.

Global Alert Thresholds Definition

In this second phase, the Double-Threshold Validation Tool (DTVT) (Bulzinetti et al., 2021) methodology was employed to evaluate the model's performance in discerning True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) outcomes.

This advanced validation framework employs a rigorous, data-driven calibration process. As previously mentioned, the core principle is the systematic analysis of the interaction between two independent control thresholds that define the warning condition: Failure Probability Threshold (FPT) and Instability Diffusion Threshold (IDT).

The optimal threshold pairs were determined through an iterative grid-search procedure. A matrix of combinations (i, j) was tested, where i represents the tested FPT values (range 50-95%, step 1%), while j represents the IDT values (range 1-100 pixels, step 1). For each pair (FPT_i, IDT_j), the system performed a complete validation routine against the 2022 landslide inventory.

This exhaustive calibration process yielded a matrix of performance values covering the entire parameter space. The quantitative evaluation of the numerous configurations was conducted by monitoring three key performance indicators:

- True Positive Rate (Sensitivity, TPR). This metric quantifies the system's ability to correctly detect instability. It is critical for the definition of the precautionary level (critical for the yellow level):

$$TPR = \frac{TP}{TP + FN}$$

[5.1]

- False Positive Rate (1 – Specificity, FPR). This metric indicates the rate of false alarms. Minimising this value is fundamental for high-criticality alerts (critical for the red level):

$$FPR = \frac{FP}{FP + TN} \tag{5.2}$$

- True Skill Statistic (TSS). Calculated as the difference between the hit rate and the false alarm rate, this index provides an unbiased measure of the model's overall discriminative capability (used to identify the best overall compromise, Orange level):

$$TSS = TPR - FPR \tag{5.3}$$

The analysis led to the definition of an *optimal Failure Probability Threshold (FPT) of 55%.*
This strategic choice was driven by the need to establish a robust baseline for the subsequent RE-ACT module, ensuring that the local correction algorithm can effectively fine-tune the spatial threshold (IDT) based on site-specific environmental features (see Section 5.3).
By providing a balanced starting point, this configuration preserves the necessary spatial instability patterns, allowing the system to refine warning levels without the risk of data loss in the preliminary phase. Consequently, this strategy directly addresses the following critical aspects:

- **Minimising Irreversible False Negatives (FN):** A higher threshold (e.g., 60%) was found to cause a significant drop in Sensitivity (TPR) for critical events.
  Once an event is filtered out by a high FPT at the pixel level, it is excluded from the spatial aggregation process and cannot be recovered by the correction module. This filtering results in the irreversible loss of critical instability patterns;

- **Maximising the Correction Potential:** By selecting the more conservative FPT of 55%, the system deliberately accepts a higher initial False Positive Rate (FPR).
  Nevertheless, this precautionary overestimation is manageable because the RE-ACT module is designed to selectively increase the spatial threshold (IDT) in environments recognised as stable.
  By raising the minimum IDT required to trigger an alarm, the module effectively suppresses false positives: Slope Units initially classified as unstable are reclassified as stable because the identified instability fails to satisfy the more rigorous spatial requirements adaptively applied by the correction module.
  Conversely, the module is structurally less effective at reconstructing missed alarms (False Negatives) caused by excessive probability filtering.
  Therefore, FPT 55% represents the necessary "inclusive baseline" that ensures the local correction algorithm has sufficient data to operate effectively, prioritising instability detection (sensitivity) over raw precision.

The performance of the model across the entire range of parameters is illustrated in *Figures 35-36* (FPT values range from 50 to 95%, with increments of 1%, while IDT values range from 1 to 100 pixels, with increments of 1 pixel).
Furthermore, the comparative trend analysis demonstrates that increasing the spatial constraint (IDT) leads to a divergence between TSS and AUC metrics, guiding the selection of specific operational points along the performance curve.

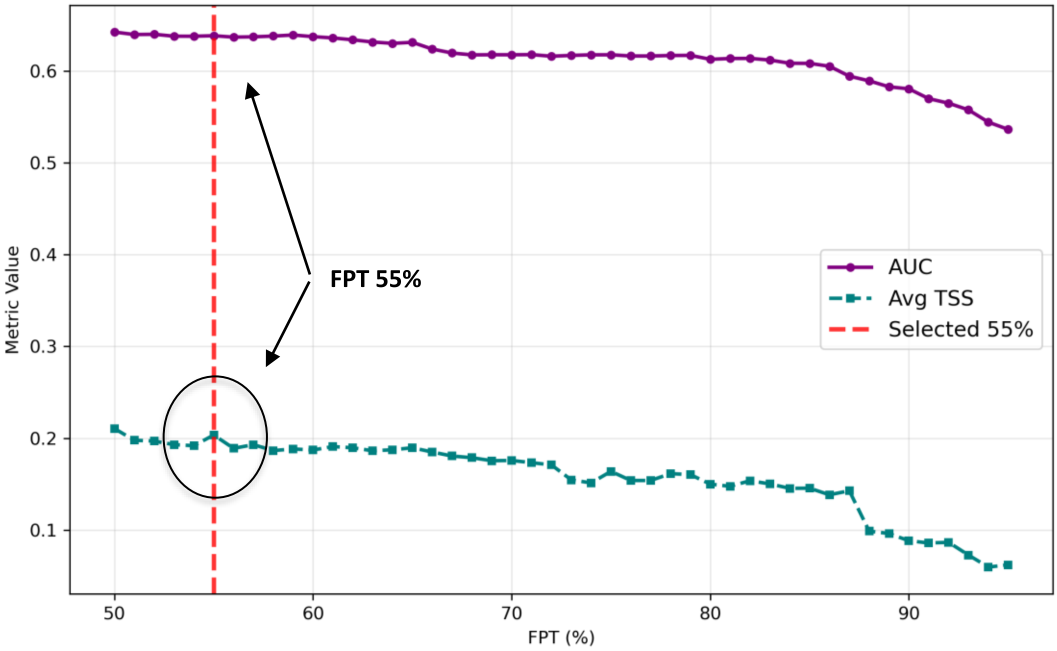


Figure 35. *Simultaneous analysis of the trend of TSS (True Skill Statistic, blue line) and AUC (Area Under the Curve, purple line) as a function of the Instability Diffusion Threshold (IDT) for a fixed Failure Probability Threshold (FPT) of 55%*

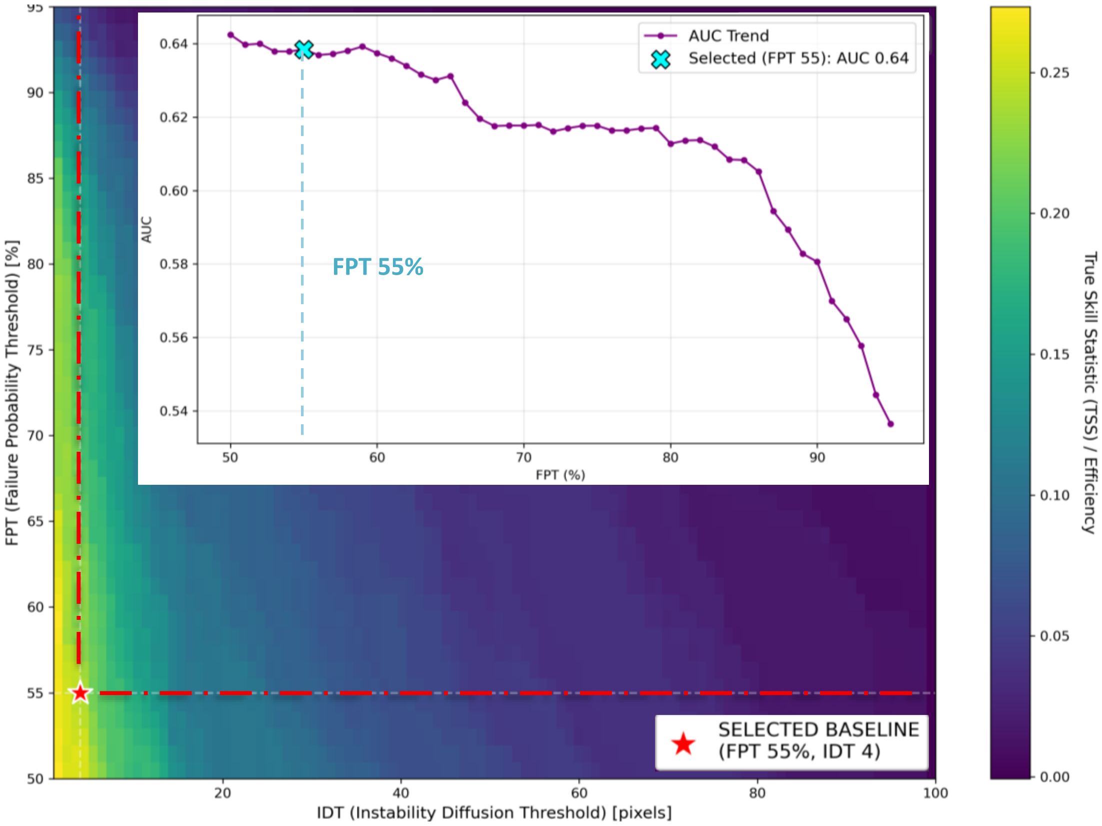


Figure 36. *3D visualisation of the True Skill Statistic (TSS) Response Surface as a function of the Failure Probability Threshold (FPT, x-axis) and the Instability Diffusion Threshold (IDT, y-axis). The colour gradient of the surface highlights the optimal configuration zones where the model maximises the balance between Sensitivity and Specificity. The top box shows the AUC (Area Under the Curve) performance trend for the full range of FPT (50-95%) and IDT (1-100 pixels) tested*

Landslide Early Warning Systems (LEWS) are typically organised into hierarchical criticality levels (Yellow, Orange, Red), which scale with the expected spatial magnitude of instability.

With the probability threshold fixed at 55%, three operational alert levels were defined by selecting three points along the ROC curve obtained by varying the Instability Diffusion Threshold (IDT).

Each level is associated with a distinct spatial requirement for instability within each Slope Unit, thereby establishing a specific trade-off between sensitivity (TPR) and specificity (1 − FPR). *Figure 37* summarises the selected configurations and their performance statistics.

It is important to note that, in this dataset, no combination of IDT with FPT ≥ 50% yields zero false negatives (FN = 0).

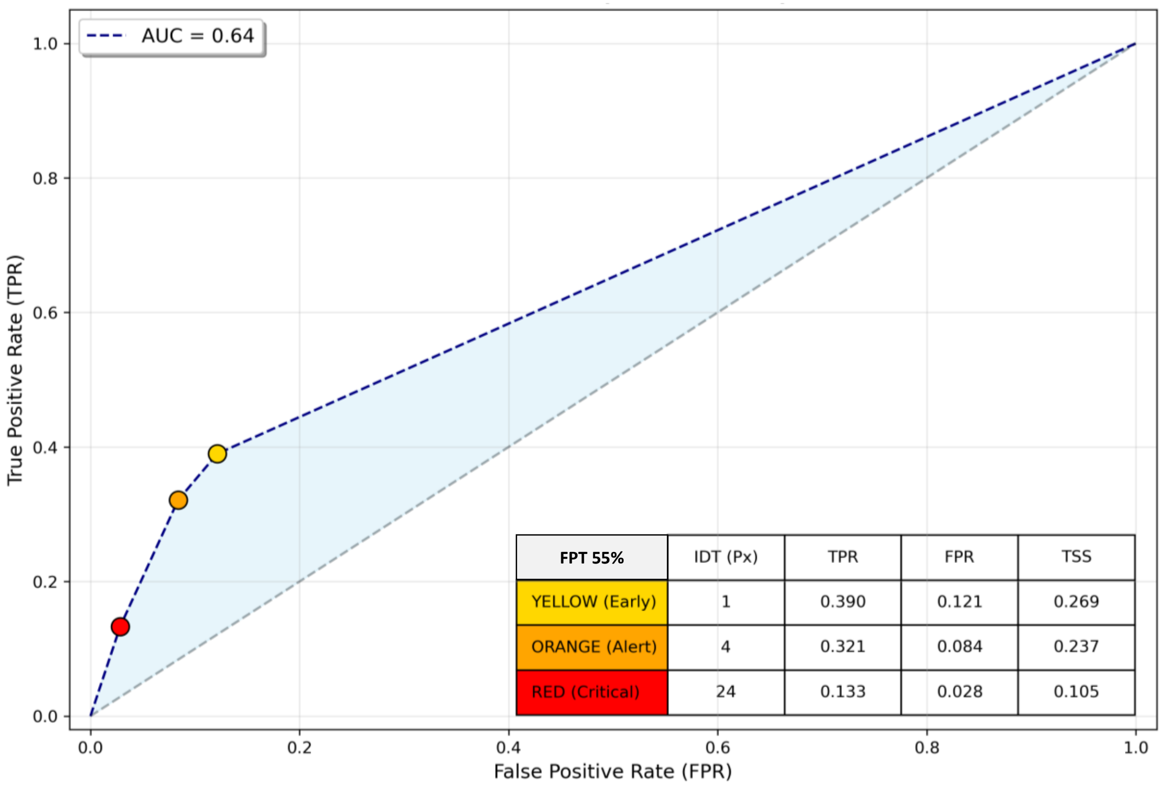
Achieving FN = 0 would require lowering FPT below 50%, i.e., classifying pixels as "unstable" even when their predicted probability of failure is lower than their probability of remaining stable; this option was deliberately excluded from the calibration.

Therefore, the three alert levels were defined by investigating the trade-off between TPR and FPR at a constant FPT of 55%, while solely varying the spatial threshold IDT:

- **Activation Level (Yellow Code):**
    - Configuration: IDT = 1 pixel;
    - Performance: TPR = 0.390, FPR = 0.121, TSS = 0.269 (TP = 85, FP = 188, FN = 133, TN = 1362);
    - Interpretation: This configuration represents the maximum sensitivity achievable under the constraint FPT ≥ 50%. Even a minimal signal of instability (a single unstable pixel) is sufficient to trigger the activation state.
      This setup minimises missed events across all practically usable thresholds, approximating a "zero-miss" objective at the expense of a relatively high intrinsic False Positive rate;

- **Pre-Alarm Level (Orange Code):**
    - Configuration: IDT = 4 pixels;
    - Performance: TPR = 0.321, FPR = 0.084, TSS = 0.237 (TP = 70, FP = 130, FN = 148, TN = 1420);
    - Interpretation: This setting corresponds to the local maximum of the True Skill Statistic (TSS) along the ROC curve (at FPT = 55%). It represents the best global compromise between sensitivity and specificity that the raw model can achieve without local calibration;
    - Strategic role: The Orange level serves as the baseline reference for the RE-ACT procedure for two reasons: (i) it provides a balanced dataset of true and false alarms, which is suitable for training the AI to recognise error drivers; and (ii) since it represents the maximum accuracy of the static model (highest TSS), it provides a robust benchmark to demonstrate the added value of the dynamic correction approach.

- **Alarm Level (Red Code):**
    - Configuration: IDT = 24 pixels.
    - Performance: TPR = 0.133, FPR = 0.028, TSS = 0.105 (TP = 29, FP = 44, FN = 189, TN = 1506).
    - Interpretation: This threshold is reserved for widespread and severe instability scenarios. By requiring an extended cluster of unstable pixels within each Slope Unit, it strongly suppresses false alarms (FPR ≈ 2.8%), aligning with the high-specificity requirements typically associated with red-alert decisions and large-scale civil-protection actions.



FPT 55%	IDT (Px)	TPR	FPR	TSS
YELLOW (Early)	1	0.390	0.121	0.269
ORANGE (Alert)	4	0.321	0.084	0.237
RED (Critical)	24	0.133	0.028	0.105

Figure 37. *Placement of the three operational alert levels (Yellow, Orange, Red) on the ROC (Receiver Operating Characteristic) curve for a fixed FPT of 55%. The three levels are defined by the selection of distinct Instability Diffusion Threshold (IDT) values, which determine the necessary trade-off between True Positive Rate (TPR/Sensitivity) and False Positive Rate (FPR/1-Specificity)*

5.3 RE-ACT Module

5.3.1 Procedure Setup

The RE-ACT framework, detailed in Section 3.4, was applied to refine the initial four-class classification results (True Positives, TP; True Negatives, TN; False Positives, FP; False Negatives, FN) obtained using the globally calibrated thresholds.
The Orange alert level configuration (FPT = 55%, IDT = 4 pixels) was utilised as the operational baseline for this procedure.
The refinement was conducted on the Slope Unit (SU) dataset (Section 5.2), which was enriched with a common set of environmental descriptors derived from five base layers: Elevation (DEM), Slope, Root Cohesion, Lithology, and Land Use Land Cover (LULC).
To ensure that these descriptors were suitable for the subsequent statistical analysis, aimed at identifying relevant error drivers, and for the Random Forest (RF) machine learning model, spatial aggregation to the Slope Unit scale was necessary.

This required differentiating the data treatment based on the intrinsic nature of the environmental variables:

- Continuous Variables (Elevation, Slope, Root Cohesion): These variables represent measurable quantities that vary gradually across the landscape.

To capture the typical condition of the SU while maintaining their quantitative meaning, they were summarised by calculating their mean value within each unit.
This mean value allows the model to correlate the average topographic and mechanical conditions of the unit with the probability of misclassification (FP or FN);

- Categorical Variables (Lithology and LULC): These variables represent discrete, distinct classes (e.g., rock types or specific land uses). Since averaging class labels is not mathematically meaningful, they were converted into fractional areal covers (ranging from 0 to 1).
  This procedure transforms the qualitative presence of a category into a quantifiable measure that reflects its spatial diffusion or proportion within the Slope Unit.
  For example, a fractional cover of 0.70 for "Flysch" serves as a numerical input feature, allowing the model to assess the influence of the relative abundance of that category on error generation.

The first operational step consists of a statistically guided selection of these environmental predictors. The goal is to identify which variables are systematically associated with classification errors in the baseline HIRESSS configuration (Orange alert level), retaining only those predictors that exhibit a robust and significant relationship with False Positives (FP) and False Negatives (FN). The analysis is performed by isolating two error blocks:

- **FP/TN block:** Compares False Positives (FP) to True Negatives (TN). This block focuses on stable units where the baseline model overestimates instability;

- **FN/TP block:** Compares False Negatives (FN) to True Positives (TP). This block focuses on unstable units where the baseline model underestimates instability.

This design isolates, in each block, an “error” class (FP or FN) and a “no-error” class (TN or TP) under the same ground truth (absence or presence of landslides), which is consistent with the subsequent use of RE-ACT to apply separate corrections for FP-dominated and FN-dominated settings.
The spatial distribution of errors is highly unbalanced. To avoid instability caused by the sparsity of error units, a repeated hold-out strategy was adopted.
In each run, the units were randomly split (with stratification by the four-class label TP/TN/FP/FN) into a 70% training set and a 30% test set. The procedure was repeated for five independent runs, using different random seeds (42, 101, 202, 303, 404).
This design guarantees, in expectation, that each run contains a sufficient number of FP and FN units, and allows the stability of the results to be evaluated across multiple random partitions.
Within each training set, a rigorous statistical screening was performed to quantify the association between environmental predictors and classification errors (*Figure 38*).
Continuous variables were analysed using the Mann-Whitney U test, and the corresponding effect size was expressed as Cliff’s δ.
Categorical variables (lithology and land cover) were first converted to slope-scale fractional covers (0-1) and then tested in two ways: via Pearson’s χ^2 or Fisher’s exact test on presence/absence counts, with Cramér’s V as effect size, and via Mann-Whitney U tests on the fractional covers, with $|\delta|$ as effect size.
Raw p-values were corrected using the Benjamini-Hochberg False Discovery Rate (FDR) procedure to control for multiple testing.
A predictor was retained only if it satisfied rigorous criteria across four aspects calibrated on the basis of an initial exploratory run without filters (*Table 20*): Presence, Validity (minimum prevalence), Reliability (FDR-adjusted significance $q < 0.05$), and Relevance (minimum effect size thresholds for Cramér’s V and $|\delta|$). *Figure 38* shows a complete flowchart of the statistical screening procedure.

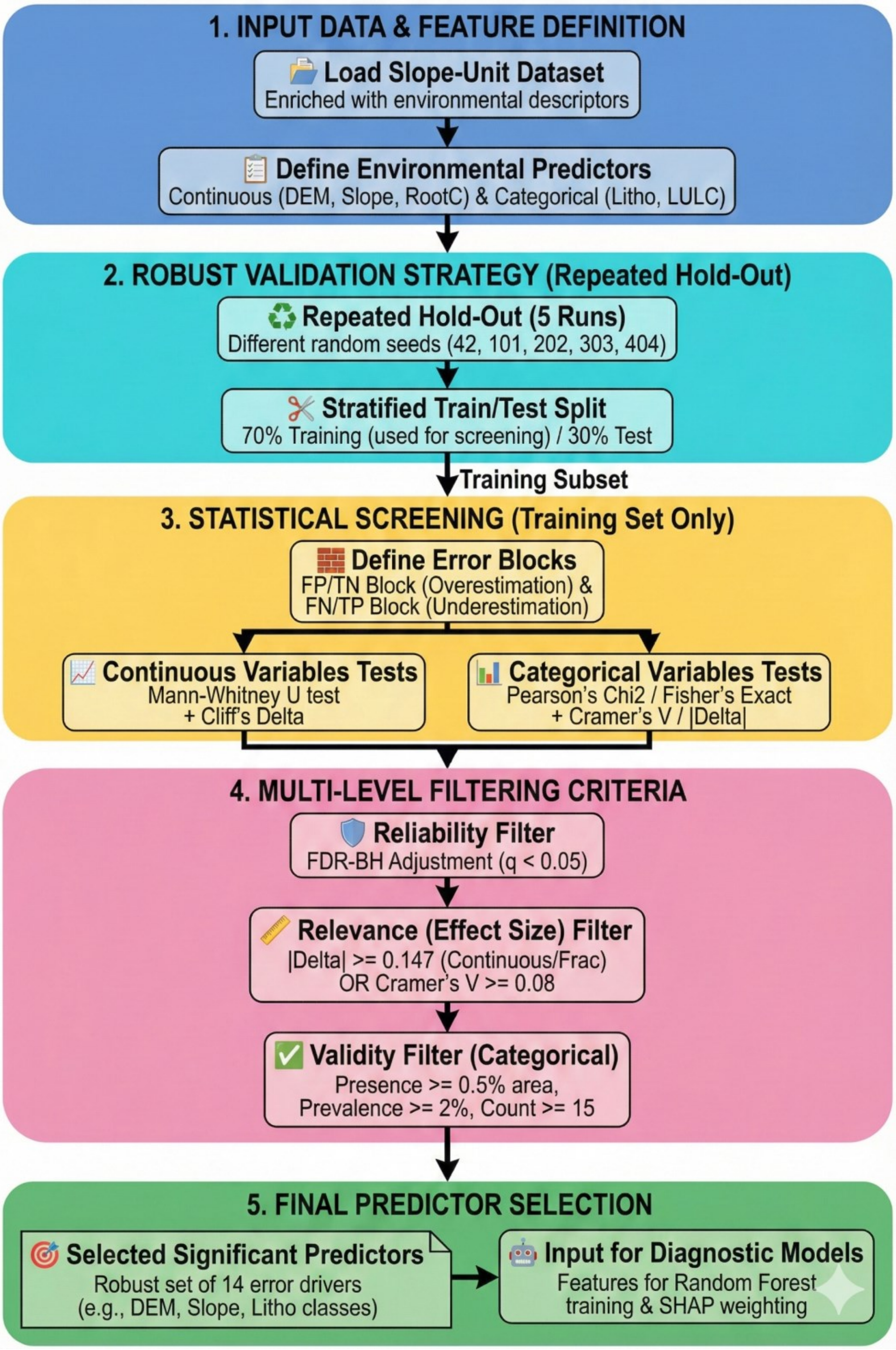


Figure 38. *Flowchart of the Statistical Screening Procedure (RE-ACT). The diagram illustrates the process of selecting significant environmental variables, starting from data loading and predictor definition (1), through a robust validation strategy (Repeated Hold-Out) (2), and statistical screening on error blocks (3). Variables are filtered via multiple criteria of reliability, relevance, and validity (4) to obtain the final set of predictors (5)*

Table 20. *Statistical screening criteria applied to environmental predictors in the RE-ACT framework. The table summarises the filters used to assess presence (definition of "class present" for categorical variables), validity (minimum prevalence and absolute count), reliability (FDR-adjusted significance, q < 0.05) and relevance (minimum effect size) of each predictor within the FP/TN and FN/TP error blocks*

Filtering step	**Scope**	**Metric**	**Threshold**	**Rationale**
Presence				
	Categorical	Fractional area of class (per unit)	Class considered "present" if fraction ≥ 0.005	Avoids spurious presence due to negligible areal fractions (<0.5% of unit area).
	Continuous	-	Not applied	Continuous variables are analysed on the full sample.
Validity				
	Categorical	Prevalence (relative) - Count (absolute)	Valid if prevalence ≥ 2% of units in block - Absolute count ≥ 15 SU	Excludes very rare classes that provide insufficient support and may yield unstable or spurious associations.
	Continuous	-	Not applied	No prevalence filter: all continuous values contribute to the tests.
Reliability				
	All variables	Adjusted p-value (q, FDR-BH)	q < 0.05 (per error block: FP/TN, FN/TP)	Controls the False Discovery Rate under multiple testing (Benjamini–Hochberg), retaining only statistically reliable associations.
Relevance (effect size)				
	Categorical	Cramér's V or |δ| (fractions)	Cramér's V ≥ 0.08 - |δ| ≥ 0.147	Filters out associations that are statistically significant but practically negligible (threshold ≈ "small" effect).
	Continuous	Cliff's δ	|δ| ≥ 0.147	Ensures that differences between error and non-error units have at least a small practical magnitude, not only statistical significance.

The screening process revealed 14 essential predictors which included seven lithological classes and four land-cover types and three continuous variables that met the filtering requirements in at least one error block across the five runs (Table 21).

Continuous variables (DEM, Slope, Root Cohesion) show the strongest and most stable statistical associations, with large |δ| values (typically > 0.40) and extremely small q-values in both data blocks.

Several categorical classes exhibit marked asymmetry between FP and FN: some are almost exclusively associated with false positives (e.g. Carbonate and Clastic_H), whereas others are mainly linked to false negatives (e.g. Evaporite and Extensive Agriculture).

These 14 predictors form the statistically robust set of error drivers used as the input feature space for the subsequent diagnostic RF models and for calculating the slope-specific environmental weights (WGT) used in the RE-ACT threshold adjustment.

Table 21. *Summary of the statistical screening results for environmental predictors. For each variable, the table reports its type (categorical or continuous), the number of hold-out runs (out of five) in which it passes all screening criteria in the FP/TN and FN/TP blocks ("Significant runs FP/FN"), and the corresponding median effect size and median FDR-adjusted q-value across runs. Effect sizes are expressed as Cramér's V for categorical predictors and Cliff's δ for continuous predictors (positive δ indicating higher values in error units and negative δ the opposite)*

Variable	Type	Significant runs	Effect size	Median q	Significant runs	Effect size	Median q
		FP	FP	FP	FN	FN	FN
Alluvial deposits	categorical	5/5	0.280	8.0e-07	5/5	0.361	3.7e-04
Carbonate rocks	categorical	5/5	0.197	3.0e-12	0/5	0.015	8.9e-01
Flysch	categorical	5/5	0.252	1.5e-15	5/5	0.433	1.6e-05
Clastic_S	categorical	5/5	0.383	1.4e-09	5/5	0.272	8.0e-03
Marlstone	categorical	5/5	0.216	2.1e-12	1/5	0.147	1.2e-01
Clastic_H	categorical	4/5	0.119	1.3e-03	0/5	0.038	7.6e-01
Evaporite	categorical	0/5	0.048	1.8e-02	4/5	0.176	1.3e-03
Ext_Agric	categorical	0/5	0.077	2.0e-09	2/5	0.127	3.7e-03
Int_Agric	categorical	5/5	0.579	5.3e-21	5/5	0.300	7.4e-03
Scrubland	categorical	5/5	0.256	1.1e-15	2/5	0.163	8.1e-02
Wood_D	categorical	5/5	0.322	5.5e-25	5/5	0.300	4.0e-04
DEM	continuous	5/5	0.666	1.7e-25	5/5	-0.472	3.0e-05
Root cohesion	continuous	5/5	0.669	8.4e-28	5/5	-0.326	3.2e-03
Slope	continuous	5/5	0.661	4.0e-25	5/5	-0.435	8.3e-05

Notes*: Clastic_S = Poorly Lithified Clastic Rocks; Clastic_H = Well-Cemented Clastic Rocks; Ext/Int_Agric = Extensive and Intensive Agriculture respectively; Wood_D = Wood (Deciduous & Mixed)*

5.3.2 RE-ACT Configuration

Once the relevant environmental variables were defined in the error model classification the procedure proceeded with training the Random Forest model to evaluate the importance of this statistically-based relationship. The complete configuration of the procedure is reported in *Table 22*.

The bounds on IDT variation were not chosen arbitrarily, but explicitly anchored to the global warning levels defined in Section 5.2.2.

In particular, the maximum allowed increase for False Positives (Max ΔFP = 20) and the maximum allowed decrease for False Negatives (Max ΔFN = 3), ensure that all locally adjusted thresholds remain confined between the Yellow (FPT = 55%, IDT = 1) and Red (FPT = 55%, IDT = 24) operating threshold points.

The risk-aware variant of RE-ACT was disabled to isolate the effect of local environmental controls on predictive skill.

Table 22. *Main configuration parameters of the RE-ACT implementation for Scenario A (15-16 September 2022)*

Parameter	**Value**	**Description**
FPT Threshold	55%	Fixed Probability Threshold for pixel instability
Original IDT	4	Baseline spatial threshold (Orange level) to be optimised
Train/Test Split	70/30%	Partitioning ratio for model training and independent testing
Random Seeds	5	Iterations performed (Seeds: 42, 101, 202, 303, 404)
P-value limit	0.05	Significance threshold for Mann-Whitney/Chi-squared tests
RF Estimators	100	Number of trees in the Random Forest (RF) classifier
RF Max Depth	10	Maximum depth of trees to prevent overfitting
Max ΔFP	+20	Max threshold increase allowed to correct False Positives
Max ΔFN	-3	Max threshold decrease allowed to correct False Negatives
Risk Awareness	False	Correction modulated by the normalised risk score.

5.3.3 Outputs Generated

The analysis first examined the environmental controls on model misclassification using SHAP (SHapley Additive exPlanations) applied to the FP and FN diagnostic Random Forest models. *Figure 39* presents the SHAP summary (beeswarm) plots for False Positives (FP/TN block) and False Negatives (FN/TP block), respectively.

For false positives (*Figure 39a*), the three continuous variables (elevation, slope, and root cohesion), dominate the SHAP ranking. Positive SHAP values are observed for high elevations, steep slopes, and strong root cohesion (red-purple dots on the right side of the axis), thereby pushing the model towards a landslide prediction in Slope Units that are in reality stable.

Conversely, low values of these same variables (blue dots on the left) produce negative contributions and help attenuate the False Positive tendency.

The Intensive Agriculture (Int_Agric) and Wood (Deciduous & Mixed - Wood_D) categorical predictors display a complex, non-monotonic behaviour.

Their impact varies significantly because high fractions of these classes can be associated with both positive and negative SHAP values, depending strongly on the surrounding topographic and lithological context.

In contrast, high concentrations of Alluvial deposits, Carbonate Rocks, and marlstone consistently produce negative SHAP values, indicating that the model treats these elements as comparatively stabilising and that they help suppress false alarms.

The Flysch and Scrubland classes behave differently, as high fractions significantly increase the likelihood of False Positives. Finally, the Poorly Lithified (Clastic_S) and Well-Cemented Clastic Rocks (Clastic_H) exert only a minor influence, remaining concentrated around SHAP ≈ 0.

The same continuous variables remain the main contributors for false negatives (*Figure 39b*), but DEM and slope show a reversed pattern with respect to the FP block.

Low elevations and gentle slopes (blue dots on the right) are now associated with positive SHAP values, whereas high elevations and steep slopes (red-purple on the left) contribute negatively and pull the prediction towards the stable class, favouring missed events.

Root cohesion shows a less clear pattern because both high and low values appear on both sides of the axis, making its effect weaker and less directional than DEM or slope.

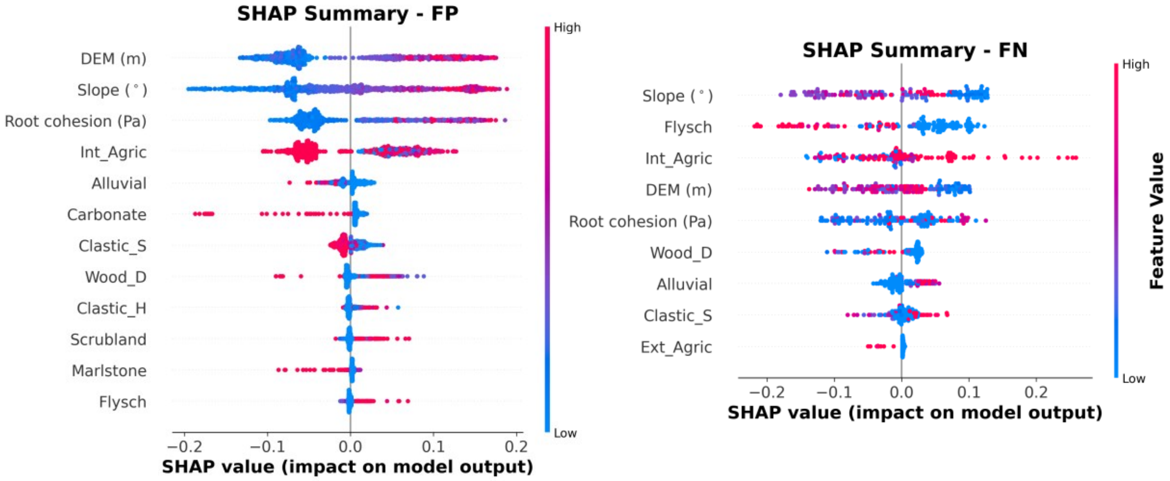
Two specific classes dominate the SHAP structure among the categorical predictors: Flysch and Intensive Agriculture (Int_Agric).

High areal fractions of Flysch are mostly linked to negative SHAP values, while low fractions are predominantly positive.

Int_Agric shows the opposite pattern, with higher fractional cover generally giving positive contributions and lower values tending to be neutral or slightly negative.

The remaining categorical variables play a secondary role. Alluvial and Clastic_S units are associated with mildly positive SHAP values when present in larger fractions.

On the other hand, high fractions of Ext_Agric or dense woodland more often yield small negative contributions, further biasing the model towards the stable class in some FN Slope Units.



a) **b)**

Figure 39. *SHAP summary (beeswarm) plots for the diagnostic Random Forest models. Figure a) refers to the FP/TN block (false positives vs. true negatives), Figure b) to the FN/TP block (false negatives vs. true positives). Each point represents a slope unit, with the x-axis showing the SHAP value (impact on the model output) and the colour scale indicating the relative feature value (blue = low, red = high). Features are ordered by decreasing mean |SHAP| value. Positive SHAP values increase the probability of belonging to the error class (FP or FN), while negative values favour the corresponding correct class (TN or TP)*

The statistical and SHAP diagnostics are translated into local corrections of the spatial threshold (IDT).

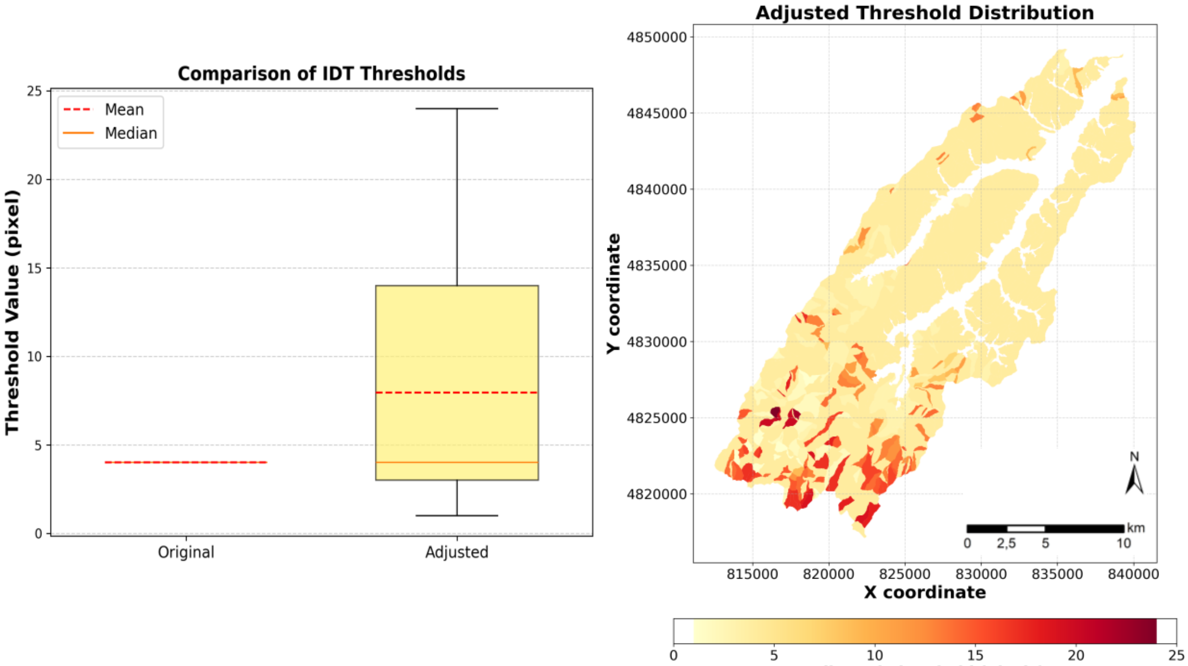
As shown by the boxplot in *Figure 40a*, all slope units start from the same baseline value (IDT = 4, Orange level), whereas the adjusted thresholds span a much wider range, from about 1 to 24 pixels.

The median of the adjusted distribution remains close to 4, but the mean shifts to approximately 8 pixels, indicating that RE-ACT mainly increases IDT in FP-prone units, while only a limited number of FN-prone units receive small downward corrections, in line with the asymmetric bounds imposed on ΔFP and ΔFN in the initial configuration.

The corresponding map of adjusted thresholds (*Figure 40b*) highlights the spatial units where IDT has been most strongly modified.

Elevated IDTs (up to 24 pixels) are concentrated in the southern and south-western sectors, where the diagnostic models indicate a high propensity for false alarms, and here the higher threshold acts as a spatial filter on spurious warnings.

Slight reductions towards values close to the Yellow level (IDT ≈ 1) are confined to a smaller set of units, mainly in lower-relief zones with gentle slopes, where conditions are more prone to missed events.



a)

b)

Figure 40. *Effect of the RE-ACT correction on the spatial threshold (IDT). Figure a) shows the boxplot of original (uniform IDT = 4, Orange level) and adjusted thresholds for misclassified units; Figure b) maps the adjusted IDT values across the study area, highlighting clusters where the threshold is substantially increased (up to 24 pixels) or slightly reduced towards the 'Yellow' level (IDT ≈ 1)*

The effectiveness of the RE-ACT correction is visually confirmed by the Changed Slope Units Classification Map (*Figure 41a*), which highlights the slope units that switched class after the adjustment: 56 Slope Units (approximately 3.2% of the 1,768 units) change from False Positive (FP) to True Negative (TN, shown in blue), while 9 (approximately 0.5%) change from False Negative (FN) to True Positive (TP, shown in light green).

The impact of the adaptive thresholds on the four-class classification is also evident when comparing the initial and final maps (*Figures 41b-c*).

This asymmetric behaviour depends on the starting configuration. Since the baseline spatial threshold (IDT) is set to 4 pixels, the system can only lower it by a maximum of 3 pixels before reaching the minimum possible value of 1.

In contrast, there is ample room to raise the threshold (up to +20 pixels) to suppress false positives. Therefore, the corrections mathematically forced to be larger for false alarms than for missed events.

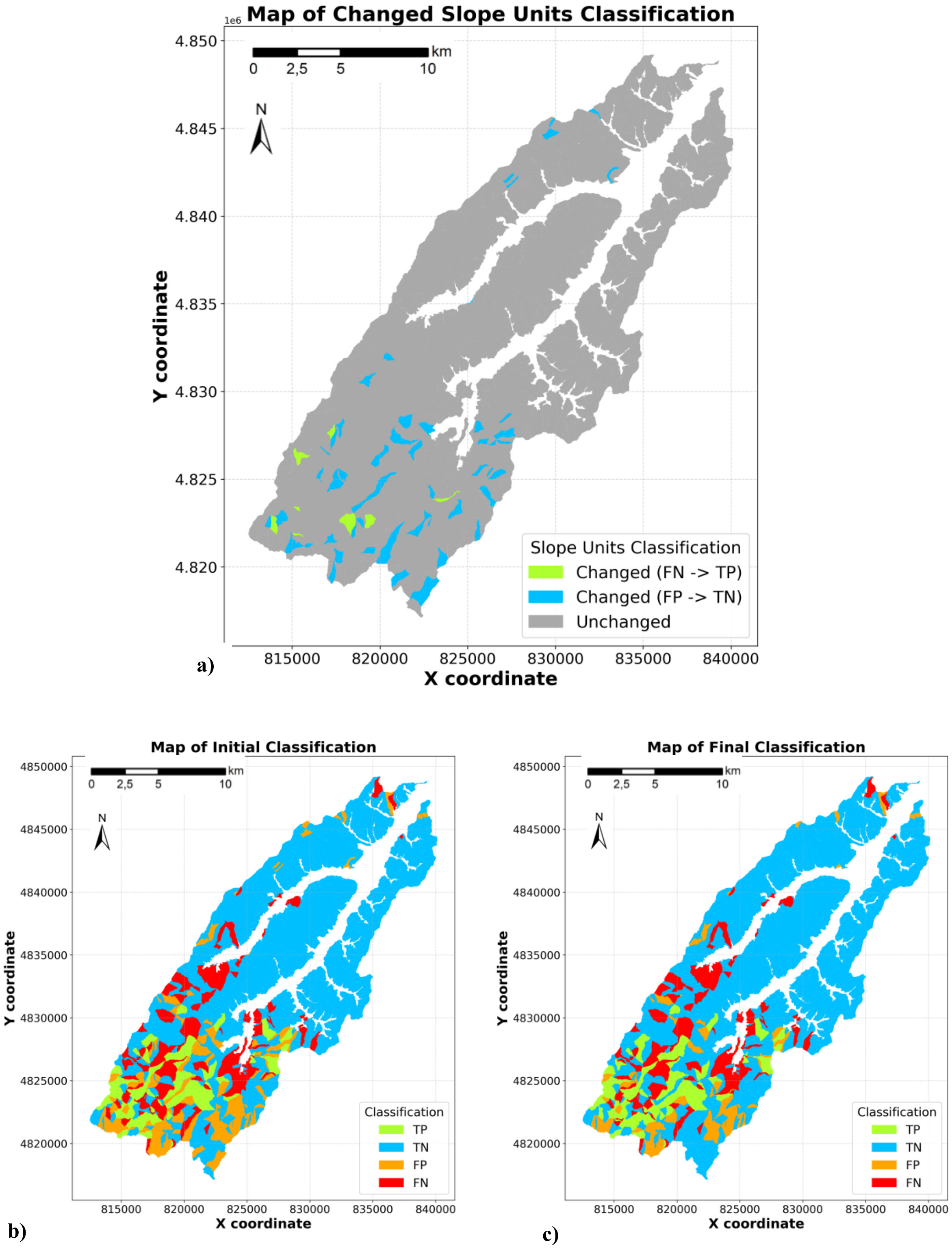


Figure 41. *Effect of the RE-ACT correction on the four-class classification. Figure a) shows the Changed Slope Units map, highlighting only those units whose label is modified after the adaptive IDT adjustment. Figure b) and c) display, respectively, the initial and final four-class maps (TP, TN, FP, FN)*

5.3.4 Performance Validation

The quantitative impact of the RE-ACT correction on the independent test set is summarised in *Table 23*. Overall, RE-ACT improves global classification performance, with increases in Accuracy, Precision, Recall, F1-score and Balanced Accuracy.

Table 23. *Summary of performance improvements on the independent test set (Scenario A)*

Metric	**Baseline (Global)**	**Corrected (RE-ACT)**	**Δ Change**
Accuracy	84.40%	87.60%	+3.20%
Precision	35.00%	48.95%	+13.95%
Recall (TPR)	32.31%	35.40%	+3.09%
F1-Score	33.60%	41.10%	+7.50%
Balanced Accuracy (BA)	61.98%	65.12%	+3.14%
False Positives (Count)	-	-	-56 (3.2% of total)
False Negatives (Count)	-	-	-9 (0.5% of total)

To further evaluate the quality of the corrected configuration, two complementary diagnostic analyses were carried out:

(i) a Precision-Recall analysis, to quantify the ability to discriminate landslide vs. non-landslide units under class imbalance;
(ii) a spatial autocorrelation analysis of the residual errors based on Moran’s I, to assess the degree of spatial clustering in the remaining misclassifications.

Their graphical outputs are shown in *Figure 42*, and the main results are described below.

The Precision-Recall curve in *Figure 42a* refers to an auxiliary Random Forest classifier trained to discriminate landslide vs. non-landslide mapping units using the subset of environmental predictors selected as described in the previous section.

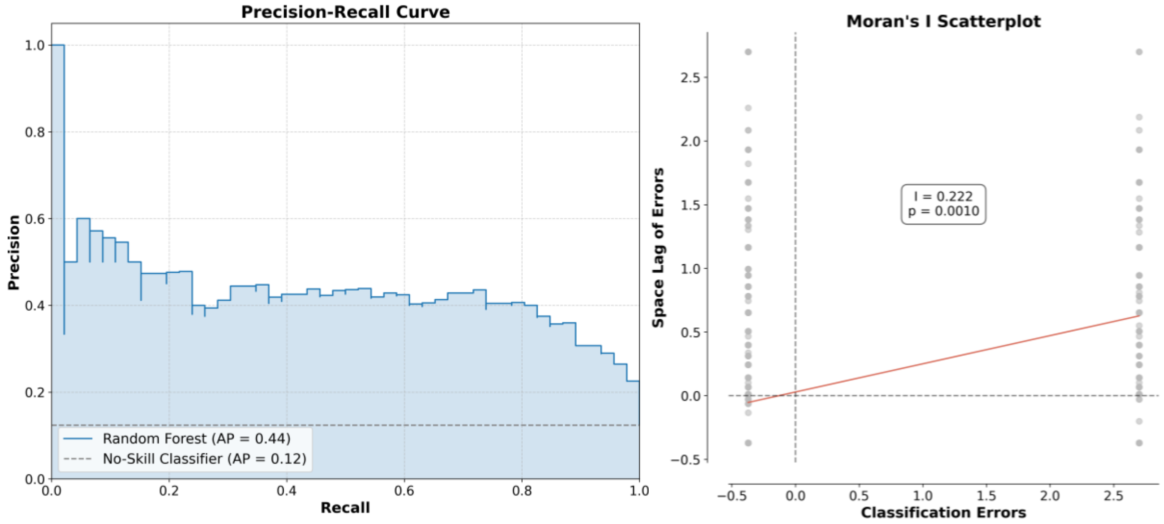
The curve yields an Average Precision (AP) of 0.44, clearly higher than the no-skill baseline (AP = 0.12) indicated by the horizontal dashed line.

This confirms that the environmental covariates used in RE-ACT carry substantial predictive information and that the corrected configuration operates in a region of the precision-recall space consistent with a non-trivial discrimination between unstable and stable units.

The spatial organisation of the remaining errors after correction was examined using Moran’s I (*Figure 42b*). The scatterplot relates, for each unit, its standardised binary error indicator (FP/FN vs. TP/TN) to the spatial lag of this indicator computed with a Queen-contiguity weights matrix.

The resulting Moran’s I is 0.222, with a permutation-based p-value of 0.0010, indicating a weak but statistically significant positive spatial autocorrelation.

This weak spatial clustering (Moran’s I = 0.222) indicates that while RE-ACT effectively removes the systematic, environment-driven component of the error, a residual local component persists, likely associated with site-specific factors not resolved at the current working resolution.



a) b)

Figure 42. *Figure a) Precision-Recall curve of the auxiliary Random Forest classifier for landslide vs. non-landslide mapping units. Figure b) Moran's I scatterplot of the binary classification errors (FP/FN vs. TP/TN) after RE-ACT correction*

5.4 Runout Model Calibration Results (SEARCH module)

5.4.1 Statistical analysis of landslide runout

For the calibration of the runout model DebrisDice (Bregoli et al., 2018), a first essential step was to perform an exploratory statistical analysis of the available landslide inventory.

The aim of this phase was not to fit a predictive model, but to understand the behaviour of the full dataset and to identify a physically robust sub-dataset suitable for calibrating the rheological runout parameters.

The analysis is based on a 10 m resolution Digital Elevation Model (DEM), resampled to 5 m using bilinear interpolation to ensure sufficient spatial sampling for smaller landslide polygons, and the landslide inventory described in Section 5.2.1, containing 568 landslides (shallow landslides and mudflows).

A dedicated automated workflow was developed to process this geospatial data. The procedure first harmonises the coordinate systems of the inputs (reprojecting vector data to match the DEM in UTM 32N) and extracts the relevant morphometric parameters for each event via geometric intersection. For each landslide polygon, the workflow identifies the source point (highest elevation cell) and the toe point (lowest elevation cell). Based on these anchor points, the following kinematic and geometric variables are computed:

- **Landslide Area:** The total polygon area (m^2);
- **Runout Length (L):** The planimetric distance between the source and the toe (m);
- **Height Drop (ΔH):** The elevation difference between source and toe (m);
- **Reach Angle (α):** The angle connecting the source to the toe, representing average mobility (°).

A physical consistency check was applied to filter out invalid geometries, such as polygons with no valid DEM cells or non-descending profiles (where source elevation < toe elevation).

This filtering process resulted in a final validated dataset of 567 landslides. On this consistent dataset, a comprehensive statistical evaluation was conducted, including descriptive statistics (boxplots and histograms in log scale of Landslide Area, Runout Length, Height Drop and Reach Angle, correlation matrices, and regression analysis (linear and quadratic) of Runout Length versus Reach Angle.

Additionally, to aid in the interpretation of the dataset structure, unsupervised machine learning (ML) techniques were employed.

Isolation Forest was used to flag potential statistical outliers (identifying approximately 5% of extreme cases), while DBSCAN was applied to the 3D variable space (Area, L, α) to visualise potential clustering of similar landslide types. *Table 24* summarises the main descriptive statistics of the four key runout descriptors. All statistics and graphics reported below therefore refer to the full set of 567 physically consistent landslides.

Table 24. *Descriptive statistics of runout parameters (567 landslides)*

Variable	**min**	**5th perc.**	**25th perc.**	**median**	**75th perc.**	**95th perc.**	**max**
Landslide area (m²)	10.0	36.7	112.2	267.4	569.2	1,764.8	19,456
Runout length (m)	5.0	11.2	18.0	30.4	54.1	120.8	389.1
Height drop ΔH (m)	0.02	0.83	2.65	5.44	12.24	29.03	92.9
Reach angle (°)	0.01	2.2	6.06	10.38	15.19	27.69	43.2

The boxplot of the four variables (*Figure 43a*) clearly demonstrates the pronounced right skewness of the distributions: the majority of landslides are small and short, with a restricted Height Drop, whereas a mere handful of events exhibit exceedingly vast areas, protracted runouts and elevated Reach Angles.

The log-scaled histograms (*Figure 43b*) confirm this pattern and highlight the following observations: most landslides have areas of a few hundred square metres (median ≈ 270 m², 75th percentile ≈ 570 m²), with a long tail of larger events up to almost 20,000 m².

This subset represents the longest and steepest events; runout lengths cluster mainly between 10 and 100 m, with a tail extending beyond 150 m; height drops are typically within 2-20 m; reach angles mostly range between 5° and 20°, with a smaller number of very mobile (low-angle) and very steep events.

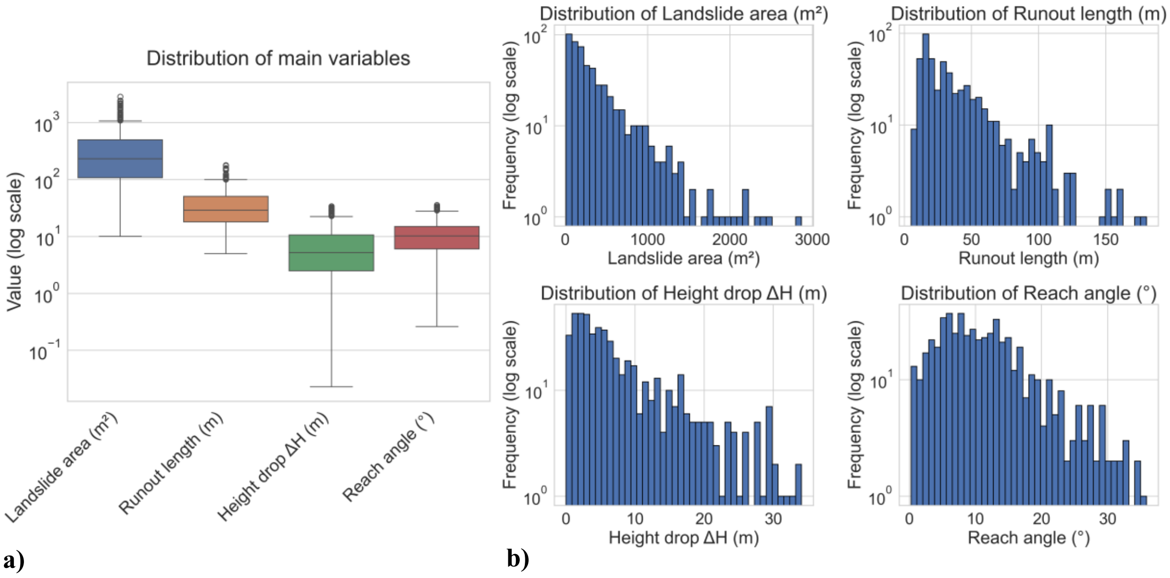


Figure 43. *Exploratory distributions of the main runout variables. Figure a) Boxplots of Landslide Area, Runout Length (L), Height Drop (ΔH) and Reach Angle (α) (log-scaled y-axis). Figure b) Histograms of the same variables with log-scaled frequency*

The correlation matrix (*Figure 44a*) demonstrates the correlation between runout length and both area and height drop. The data indicates a positive association; larger and steeper landslides travel farther.

The Reach Angle demonstrates a moderate correlation with Height Drop, yet a negligible correlation with Runout Length, thus reflecting its inherent nature as a ratio between Height Drop and Length.

The 3D scatter plot with DBSCAN clusters (*Figure 44b*) in the space of Landslide Area, Runout Length and Reach Angle highlights a dense main cloud of "typical" events and a small number of scattered points corresponding to very large or very long runouts.

These cases demonstrate a considerable degree of overlap with the anomalies identified by the Isolation Forest.

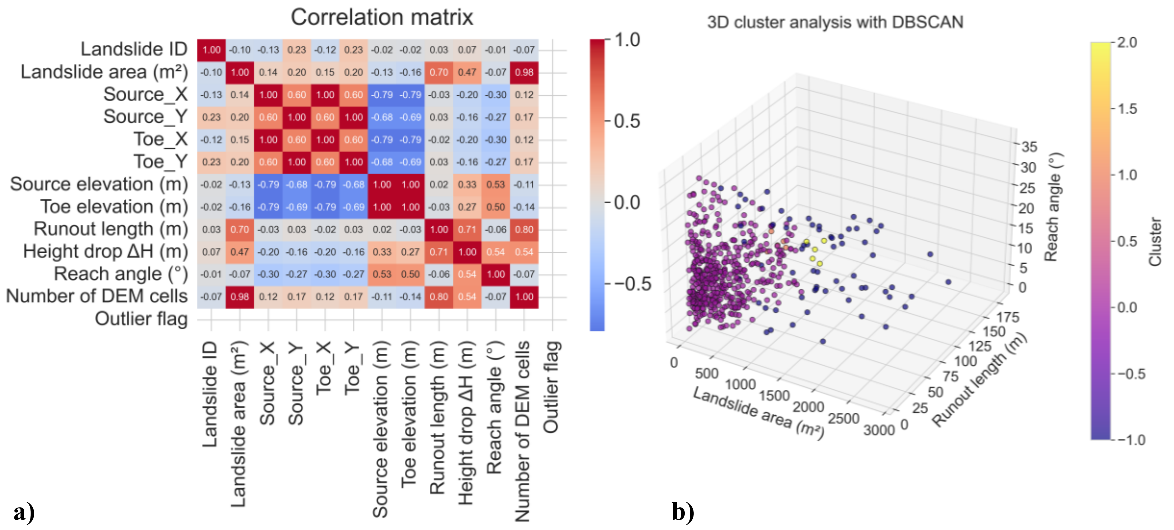


Figure 44. *Multivariate representation of the landslide runout dataset. Figure a) Pearson correlation matrix of all analysed variables, including the number of DEM cells and the Isolation Forest outlier flag. Figure b) 3-D scatter plot of landslide area, Runout Length (L) and Reach Angle (α), coloured by DBSCAN cluster*

Finally, the regression plots *(Figure 45)* illustrate the relationship between Reach Angle and Runout Length. Both the linear (*Figure 45a*) and quadratic (*Figure 45b*) fits show a weak negative trend, indicating that, within this dataset, longer runouts tend to be associated with slightly lower Reach Angles, but with considerable scatter due to varying local morphologies and volumes.

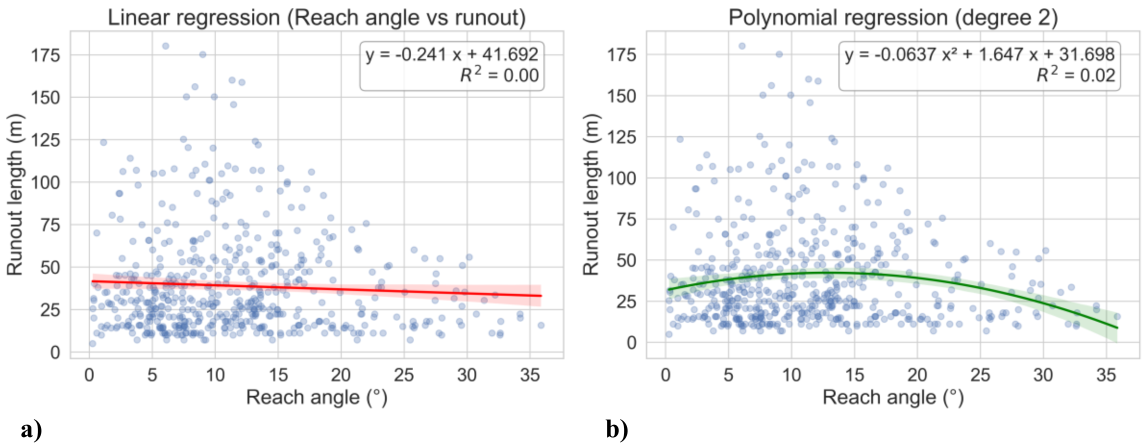


Figure 45. *Regression analysis between Reach Angle (α) and Runout Length (L). Linear (a) and quadratic (b) regressions, with best-fit curves, corresponding equations and R^2 values*

5.4.2 Runout Model Calibration

Data Subset Definition for Model Calibration

The statistical analysis reveals that the landslide inventory is predominantly composed of small-scale events: the median area is only 267.4 m², and the median runout length is just 30.4 m.
These values indicate that more than half of the recorded events involve extremely short propagation paths. To achieve a physically valid calibration of DebrisDice (Bregoli et al., 2018), the dataset was refined into a subset satisfying four practical conditions:

1. Computational Feasibility: The SEARCH algorithm couples DebrisDice (Bregoli et al., 2018) with an iterative global optimisation scheme, evaluating up to O(10^3) parameter combinations on the entire calibration set.
   Since the computational cost scales linearly with the number of events, running the full optimisation on the initial landslides dataset would result in prohibitive runtimes. Filtering the dataset allows for a more efficient exploration of the parameter space without compromising statistical significance;

2. DEM Resolution and Path Stability: Given the 5 m grid resolution, a runout of 50 m corresponds to 10 DEM cells. While the runout-length distribution shows that 50 m is close to the 75th percentile (54.1 m), this threshold was further refined.
   To reliably simulate the transition from acceleration to deposition, the model requires a minimum number of cells to calculate local gradients (slope and curvature) without being dominated by the boundary conditions of the source and toe.
   Consequently, the minimum runout requirement was increased to 70 m (≈14 cells). This ensures that the "transport zone" is numerically independent of the initiation and deposition geometries, allowing the rheological parameters to be constrained on a fully developed flow, as supported by investigations on flow development lengths (Hürlimann et al., 2003) and grid-size dependency in debris flow modelling (Stolz and Huggel, 2008; Boreggio et al., 2022);

3. Elevation Drop Significance: Since the 25th percentile of the elevation drop is 2.65 m, a threshold of $\Delta \geq 3$m was adopted. This limit, set slightly above the statistical reference, ensures that the elevation difference between source and toe reflects an actual slope geometry, distinguishing physical topography from potential vertical artefacts or precision limits inherent to the digital model;

4. Geometric Consistency: The 5th percentile of the reach angle is approximately 2.2°. Therefore, a minimum reach angle of 3° was selected. This threshold excludes the flattest ≈8% of events, thereby removing geometries which are more likely to be affected by mapping or geometric uncertainty.

Applying these thresholds to the full dataset yields 54 landslides, which constitute the calibration sub-dataset. This subset preserves a substantial number of events while focusing on landslides with sufficiently long and steep runouts, where DEM resolution and elevation differences are adequate to support a meaningful calibration of DebrisDice (Bregoli et al., 2018) rheological and runout parameters.

Configuration of the Optimisation Search Space

Based on the filtered landslide inventory, the SEARCH-DebrisDice calibration was configured using a search space grounded in both local statistical evidence and established literature benchmarks.
The geometric stop criteria (Reach Angle α, and Minimum Slope Angle β), expressed in radians, were established to capture the topographic control on runout.
Aligned with the exploratory statistical analysis (local Reach Angle 95th percentile - 27.69°), the search range for α was defined between 0.05 and 0.60 to encompass the observed kinematic variability, in accordance with the mobility values reported in literature (Corominas, 1996; Hürlimann et al., 2008; Bregoli et al., 2018).

Conforming to these geometric constraints, *β* was bounded between 0.04 and 0.20 to include typical slope thresholds for deposition (Benda and Cundy, 1990; Hürlimann et al., 2008).

The ranges for *μ* and *k* reflect effective basal resistance values reported in Voellmy-Rheology simulations of debris flows and avalanches, avoiding extreme combinations that would systematically produce either negligible propagation or unrealistic overshooting.

The range for the Friction Factor (*μ*) was defined between 0.04 and 0.35 in agreement with the best-fit values reported in literature (Revellino et al. 2004; Hürlimann et al., 2008; Hussin et al., 2012; Bregoli et al., 2018). The Mass-to-Drag Ratio (*k*) was explored between 1.5 and 12.0 to provide the necessary flexibility for the high variability of the turbulence term; since this parameter lacks a direct correlation with mobilised volume, it was calibrated site-specifically to accurately reflect local flow resistances.

Finally, the velocity threshold (v_{thr}) was constrained between 0.01 and 0.10 m/s. As suggested by Rickenmann et al. (2006), velocities on the order of centimetres per second (cm/s) can be negligible to approximate the flow as stationary, thereby effectively preventing numerical drift and ensuring computational stability.

This threshold is optimised by maximising the F1-score for each fold to ensure spatial consistency between the simulated results and the mapped footprint, establishing a dual calibration process: while the rheological parameters are tuned to resolve the flow propagation dynamics, the velocity threshold is concurrently optimised to define the precise stopping position.

The rationale for selecting the F1-score as the guiding metric for this specific parameter lies in its ability to penalise the two opposing numerical artefacts that govern the definition of the depositional area. Since v_{thr} acts as a hydrodynamic cutoff, a value set too low would fail to arrest slow-moving pixels, leading to unrealistic numerical drift (or "creeping") where the simulated mass expands indefinitely beyond the true limits (resulting in low Precision).

Conversely, a value set too high would cause premature deposition, freezing the flow on steeper slopes before it reaches the fan toe (resulting in low Recall). By maximising the F1-score the optimisation algorithm identifies the specific velocity cutoff that perfectly balances momentum dissipation and static deposition, yielding a runout geometry that matches the observed landslide boundaries.

Overall, the configuration in *Table 25* represents a compromise between physical realism, coverage of the observed behaviour and computational efficiency.

The weighting hierarchy is primarily anchored on Intersection over Union (IoU: 1.0) and F1-score (0.8) to drive the optimisation towards high geometric fidelity and balanced spatial accuracy.

A marginal preference is assigned to Recall (0.5) over Precision (0.4) to introduce a conservative safety factor, aiming to minimise false negatives without compromising the model's reliability through excessive overestimation.

Additionally, the Relative Error weight (0.5) strictly constrains the runout distance to plausible physical limits. *Table 25* summarises the parameters explored by SEARCH, the validation strategy and the metric weights used in the objective function.

Table 25. *Calibration setup for the Misa River basin: parameter search space and metric weights*

Category	Parameter / Metric	Symbol	Range / Value
Physical parameters (optimised)			
	Reach Angle	α	0.05 - 0.60 (rad.)
	Min. Slope Angle	β	0.04 - 0.20 (rad.)
	Friction Factor	μ	0.04 - 0.35 (-)
	Mass-to-Drag Ratio	k	1.50 - 12.0 (m)
Mapping parameter (optimised-F1score)			
	Velocity Threshold	v_{thr}	0.01 - 0.10 (m/s)
Validation strategy			
	Spatial K-Fold	k	5
	Bayesian calls per fold	N	80

Category	Parameter / Metric	Symbol	Range / Value
Objective function (metric weights)			
	Intersection over Union	*IoU*	Weight: 1.0
	F1-Score	*F1*	Weight: 0.8
	Relative Error	*RE*	Weight: 0.5
	Recall (Sensitivity)	*TPR*	Weight: 0.5
	Precision (PPV)	*PPV*	Weight: 0.4
	Hausdorff Distance	*HD*	Weight: 0.4
	Cohen's Kappa	*kappa*	Weight: 0.2
	PR AUC	*PR-AUC*	Weight: 0.2
	Source Area Ratio	*Sratio*	Weight: 0.3

Calibration Parameters

The robustness of the calibration was evaluated using a 5-fold spatial block cross-validation, where landslides are grouped into 500 m blocks and entire blocks are assigned to either calibration or validation folds.

This strategy reduces spatial dependence between folds and provides a realistic estimate of predictive skill for unsurveyed parts of the basin.

For each fold, SEARCH identifies a set of "best" parameters that maximises the weighted objective function on the calibration events. *Table 26* reports the values obtained in each spatial fold and their mean and standard deviation (σ) across folds.

Table 26. *Optimal parameters identified for each spatial fold and aggregated statistics on calibration subset*

Fold ID	Reach Angle (α)	Min. Slope Angle (β)	Friction Factor (μ)	Mass-to-Drag (k)	Threshold (v_{thr})
Fold 1	0.050	0.040	0.0481	9.47	0.0526
Fold 2	0.050	0.049	0.040	9.75	0.0621
Fold 3	0.600	0.040	0.040	9.13	0.0953
Fold 4	0.600	0.048	0.040	9.16	0.0100
Fold 5	0.600	0.040	0.0476	9.56	0.0384
Mean ± σ	0.380 ± 0.301	0.0436 ± 0.0049	0.0431 ± 0.0043	9.415 ± 0.266	0.0517 ± 0.0526

The analysis reveals a bimodal behaviour for the Reach Angle (α) parameter, which assumes extreme values (0.05 vs. 0.60) in different folds.

This divergence reflects the composite nature of the inventory: events with reduced mobility (low Reach Angles) are effectively modelled ($\alpha \approx 0.05$), while more mobile landslides (high Reach Angles) require values close to 0.60. Despite this bimodality, the key rheological parameters (Minimum Slope Angle, Friction Factor, Mass-to-Drag) show minimal variation ($\sigma \leq 0.005$) compared to the initial wide ranges (*Table 25*).

This stability indicates that the calibration does not excessively depend on local topographic peculiarities, but identifies a consistent basin-scale rheological signature for the September 2022 event.

The temporal evolution of the objective score is illustrated in *Figure 46*. The convergence plot shows that the weighted score stabilises after approximately 40-50 iterations in all folds, with only marginal improvements thereafter. This confirms that the Gaussian process surrogate used by the Bayesian algorithm efficiently explores the high-dimensional parameter space compared to naïve grid or random search.

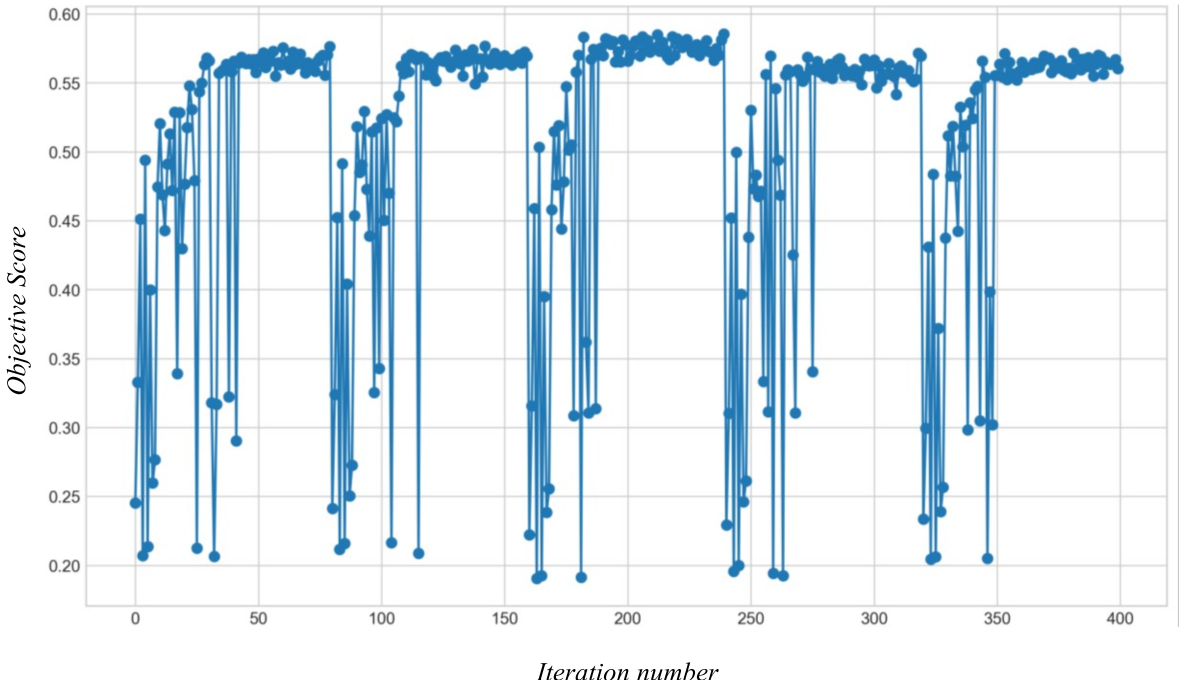


Figure 46. *Convergence of the optimiser during Bayesian calibration. Evolution of the weighted objective score versus iteration number for each of the five spatial folds*

For subsequent forward simulations, a single global parameter set was derived by taking, for each parameter, the median of its values across the five folds (*Table 27*).

The median was chosen for its robustness; in the case of the bimodal Reach Angle, it selects the value associated with the more mobile landslides, which is conservative for hazard-mapping purposes.

This consolidated set represents the “official” configuration adopted for scenario modelling in the Misa River basin.

Table 27. *Official rheological parameter set adopted for scenario modelling in the Misa River basin*

Parameter	Value
α	0.60;
β	0.04;
μ	0.04;
k	9.47;
v_{thr}	0.053.

Validation Performance Analysis

The predictive performance of the calibrated model was evaluated on the validation subsets of each fold (out-of-sample landslides).

For each event, the simulated runout polygon was compared to the mapped polygon within a 10 m buffer region of interest, and the metrics listed in *Table 25* were computed and aggregated across all validation events. *Table 28* summarises the mean and standard deviation (σ) of these metrics.

Table 28. *Aggregated validation metrics across all cross-validation folds*

Metric	Mean Value	Std. Dev. (σ)	95 % Bootstrap CI	Interpretation
IoU (Jaccard)	0.449	± 0.12	0.415 - 0.479	Overall spatial agreement
F1-Score	0.609	± 0.132	0.569 - 0.642	Balance between Precision and Recall
Recall (TPR)	0.601	± 0.191	0.544 - 0.649	Fraction of observed runout area correctly captured
Precision (PPV)	0.703	± 0.156	0.661 - 0.742	Fraction of simulated runout area that corresponds to reality
Relative Error	0.352	± 0.280	0.281 - 0.430	Over-/under-estimation of total area (in%)
Cohen's Kappa	0.394	± 0.123	0.362 - 0.426	Reliability of classification corrected for chance
Hausdorff Distance (m)	56.4	± 25.0	50.6 - 63.9	Maximum spatial discrepancy
ROC AUC	0.70	± 0.06	0.683 - 0.717	Pixel-wise discrimination between runout and non-runout
PR AUC	0.673	± 0.079	0.653 - 0.693	Discrimination performance in terms of Precision–Recall.

Bootstrap confidence intervals (500 resamples) confirm the robustness of these estimates (e.g. IoU = 0.45, 95% CI ≈ 0.415-0.479; F1-score = 0.609, 95% CI ≈ 0.569-0.642). In the Misa application, the metrics point to a consistent and moderately conservative performance of the calibrated model.

The Recall metrics is moderately high (0.601), which is coherent with the safety-oriented weighting in the objective function and ensures that underestimation of impacted areas is limited. IoU and F1-score around 0.449-0.609 indicate that both the extent and the shape of observed runouts are reproduced with good reliability.

Moderate Precision (0.703) and Relative Error (0.352) suggest some over-spreading. This is coherent with the calibration strategy, which was designed to minimise false negatives, deliberately favouring conservative overestimation rather than under-prediction. These aggregated performance statistics are summarised in *Figure 47*.

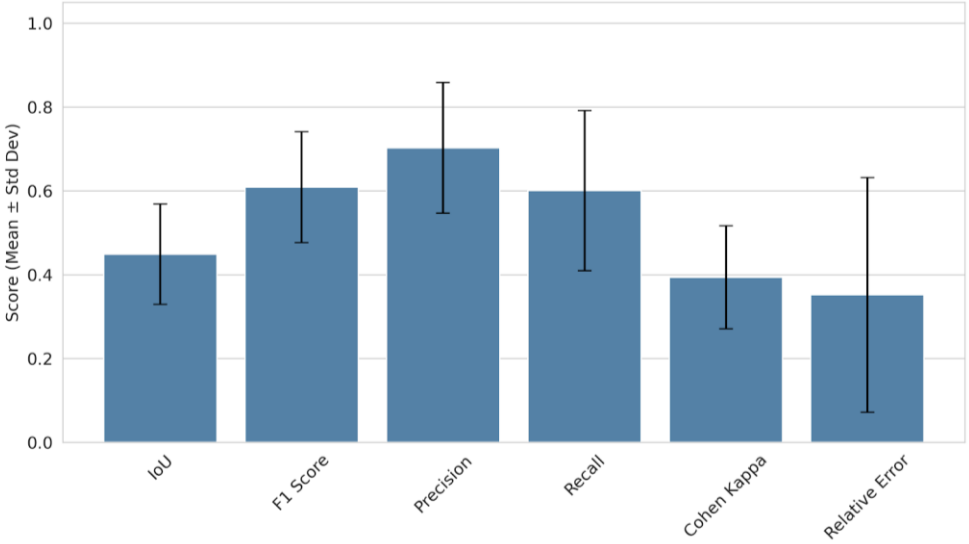


Figure 47. *Summary of validation metrics (mean ± standard deviation) across all cross-validation folds*

To assess global performance stability, *Figure 48* displays the distribution of validation metrics all landslides. The results indicate a robust behaviour for most events, as shown by the compact interquartile ranges. Outliers are limited to specific cases characterised by highly irregular flow paths or local DEM artefacts.

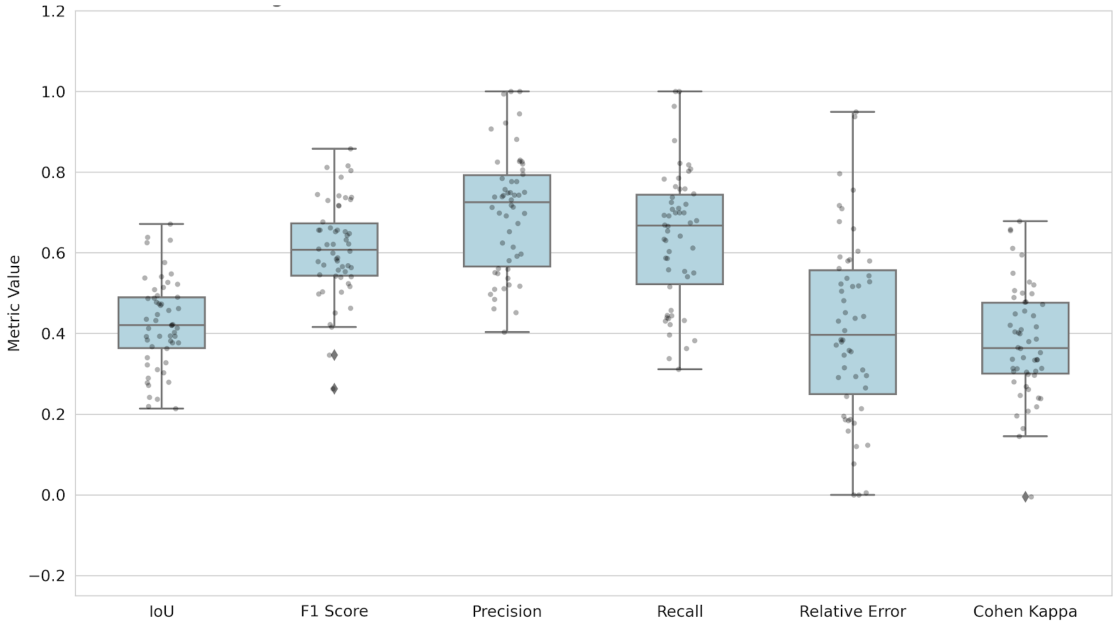


Figure 48. *Distribution of validation metrics across all landslides. The boxplots illustrate the median (central line), interquartile range (box), and full variability (whiskers and points) for each performance indicator*

Optimised Source Geometry Analysis

A key innovation of the SEARCH framework is its ability to automatically identify source areas. Usually, historical landslide maps show a single polygon that merges both the source and the deposition area.

SEARCH eliminates the need to manually separate these zones, a task that often depends too much on the operator's personal judgment.

Instead, the algorithm independently finds the most likely source area by following three practical steps:

a. *Isolate high-elevation areas:* it selects only the topographically highest parts of each polygon;
b. *Specific slope calculation:* the algorithm analyses the terrain inclination based exclusively on the landslides included in the active calibration fold. This prevents "information leakage" and ensures that the model learns only from the specific training data;
c. *Test different shapes:* it tries various sizes for the source area (up to 50% of the total polygon) until it finds the one that generates the most realistic movement.

Figure 49 illustrates this process. The left panel shows the entire dataset of 54 landslides used for calibration, represented by the original inventory polygons. The right panel provides a detailed zoom on three specific landslides (purple square) to show the results: the optimised source area is in red, the simulated runout path is in yellow, and the original landslide perimeter is in green.

It is important to note that the simulated flow appearing lateral to the source area is due to the resampling of the DEM from 10 m to 5 m.

This process introduces topographic artefacts, small digital variations in the terrain surface, which act as local irregularities, allowing the debris to spread laterally and involve areas situated to the sides of the source area.

This result proves that the algorithm resolves the uncertainty of the initial data, creating credible initiation zones without human intervention.

From a modelling standpoint, this means that the source shape and the landslide movement are optimised together.

This makes the final parameters much more reliable for predicting future hazard scenarios in the Misa River basin.

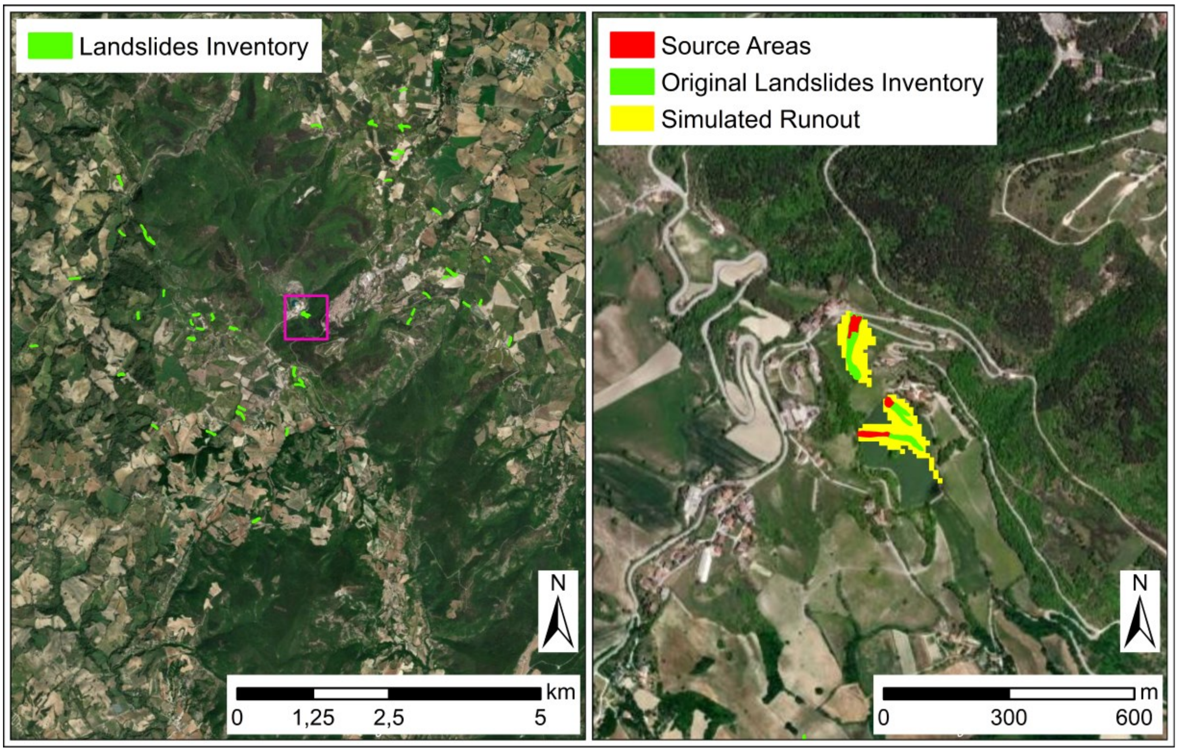


Figure 49. *Automated source delineation and runout simulation. Example output from the SEARCH framework. The panel on the left displays the original inventory polygons merging source and deposition area for the 54 landslides used for calibration. The panel on the right provides a detailed zoom on three specific landslides (purple square) to showcase the results: the optimised source area identified by the algorithm is shown in red, the corresponding simulated runout footprint in yellow, and the original reported landslide polygon in green*

5.5 Economic Risk Assessment

This section details the risk estimation and calculation phase for the buildings located within the study area. Implemented through the SADRI framework, the procedure translates physical impact scenarios into a quantification of expected economic damage by integrating the simulation results with a detailed database of exposed elements.

The analysis is articulated in three consecutive phases:

1. Construction of the asset dataset: This involves the parameterisation of exposure and vulnerability, identifying at-risk elements and assigning specific damage functions to each asset category;
2. System configuration: This consists of setting up the calculation parameters;
3. Application of the risk calculation: The execution of the analysis on two distinct modelling scenarios (Calibration and Test) to evaluate the distributed economic impact at the catchment scale, in both deterministic and probabilistic terms.

5.5.1 Asset Database Construction

In order to characterise the elements at risk, it was necessary to integrate various spatial and census data sources to define the Exposed Economic Value (E) and Structural Resistance (`R_struct`) for each building within the Misa basin.

Exposure (E): Assets Economic Value

The cartographic basis for building identification consists of the vector dataset "EDIFICATO.shp" provided by the Marche Region in Open Data format (available at the following address: https://www.regione.marche.it/Regione-Utile/Ambiente/Cartografia-e-informazioni-territoriali/Repertorio).
This dataset was derived from the GeoTopographic Database (DBGT) Update CTR2019, based on aerial photogrammetric restitution from the AGEA 2019 flight.
The polygonal geometries, essential for calculating the impact area, feature an acquisition scale of 1:10,000 and are georeferenced in the ETRF 2000 system at epoch 2008.0 (EPSG: 7792).
To ensure a reliable economic quantification of the exposed assets, the unitary market values (€/m^2) were derived from the Real Estate Market Observatory (OMI) database, which is managed by the Italian Revenue Agency (https://www.agenziaentrate.gov.it/portale/schede/fabbricatiterreni/omi/forniture-dati-omi).
The OMI database serves as the official reference for the Italian property market, ensuring a standardised and spatially consistent valuation framework.
The OMI system subdivides the national territory into homogeneous territorial zones (OMI Zones). Each zone represents a continuous portion of a municipality, often coinciding with or aggregating multiple census tracts, characterised by uniformity in economic appreciation, environmental conditions, and socio-economic status.
For every zone, the database provides a range of market quotations (minimum and maximum trade values), which are updated semi-annually and differentiated by building typology (e.g., Residential, Commercial, Productive/Industrial) and state of conservation. In this study, the assignment of economic value followed a spatial intersection procedure:

1. Spatial Join: The centroid of each building footprint was spatially overlaid with the OMI zonal map to identify the corresponding market zone;

2. Value Calculation: For each building, a specific unitary value was computed by averaging the minimum and maximum quotations provided by the OMI database for the relevant building typology (e.g., "Civil Dwelling" for residential assets) within that specific zone;

3. Exposure Estimation: This mean unitary value was then multiplied by the building's footprint area to derive the total economic exposure (E).

Finally, to correlate the economic assessment with the physical simulation, a spatial intersection was performed between the simulated runout polygon and the building footprints.
For each intersected building, the representative intensity (I_{asset}) was conservatively defined as the maximum pixel value of the flow velocity extracted from the overlapping area.
Consequently, the economic value considered in the final risk calculations is not the total value of the asset, but it represents only the portion of the building intersected by the simulated runout.

Structural Resistance (R_struct) and Vulnerability (V)

To define the buildings' propensity for damage, it was necessary to estimate their intrinsic resistance to impact stresses. Since point-data for individual buildings was unavailable, the methodology proposed by Caleca (2024) was adopted, which derives a structural resistance index (R_struct) by aggregating ISTAT census data at the census tract scale.

The Structural Resistance (R_struct) was modelled by combining three key factors, adapting the equation by Li et al. (2010): structural typology (ε_{sty}), state of maintenance (ε_{smn}), and building height (ε_{sht}) and is defined as follow:

$$R_{\_struct}\,[-] = \sqrt[3]{\varepsilon_{sty} \times \varepsilon_{smn} \times \varepsilon_{sht}}$$

[5.4]

Where ε_{sty} is a factor related to the structural type (e.g., masonry, reinforced concrete), ε_{smn} is a factor related to the state of maintenance (e.g., excellent, average, very poor), and ε_{sht} is a factor related to the height of the building (number of floors).

The numerical values assigned to these coefficients vary depending on the type of landslide considered (slow-motion, rapid-motion, or collapse landslides).
Considering that the simulations performed with DebrisDice (Bregoli et al., 2018) reproduce flow phenomena comparable to rapid debris flows, the specific coefficients for "fast-moving landslides" were adopted, as defined in Caleca, 2024. *Table 29* reports the values used to parameterise building resistance in the Misa River basin.

Table 29. *Coefficients adopted for the calculation of Structural Resistance (R_struct), specific for fast-moving landslide scenarios (adapted from Caleca, 2024)*

Factor (ϵ)	**ISTAT Category**	**Value**
Typology (ε_{sty})		
Residential	Reinforced Concrete	1.95
Residential	Load-bearing Masonry	0.90
Residential	Light Structures / Other	0.25
Productive / Commercial		0.10
Maintenance (ε_{smn})		
Residential	Very good	1.50
Residential	Good	0.60
Residential	Medium	0.10
Residential	Poor	0.10
Productive / Commercial		0.10
Height (ε_{sht})		
Residential	1 floor	0.40
Residential	2 floors	0.85
Residential	3 floors	1.00
Residential	≥ 4 floors	1.20
Productive / Commercial		0.10

The resulting index was normalised to be dynamically used by the SADRI framework: for each simulation cell, the final vulnerability (V) is calculated by comparing the local flow intensity (velocity, m/s) with the resistance of the impacted building, according to the damage curves for rapid flows implemented in the code (Section 3.6.2).

5.5.2 SADRI Framework Results

This section presents the results obtained by applying the SADRI (Scenario-Aware Dynamic Landslide Risk & Impact) framework to the two rainfall events simulated in the Misa River basin.
The objective of the analysis is to demonstrate the system's capability in producing high-resolution economic risk maps from rainfall scenarios. The data sources, system configuration, and simulation parameters adopted to perform the risk analysis for both scenarios are detailed below.

Input Data and System Configuration

The implementation of the SADRI procedure is based on the integration of five primary datasets, summarised in *Table 30.*

Table 30. *Summary of the Input datasets adopted for the risk analysis*

Input parameters	Methodology
DEM	Derived from TINITALY DEM 10 m (Tarquini et al., 2007)
Slope Units (SUs)	Generated via r.slopeunits (Alvioli et al., 2016) based on the 10 m DEM, serving as the fundamental spatial domains for the analysis (see Section 5.2.2)
Triggering Maps (FPT Maps)	HIRESSS Simulation Outputs: Raster maps defining unstable pixels (P_f > FPT) for Scenario A (2022) and Scenario B (2014)
Buildings	Dataset "EDIFICATO.shp" (Regione Marche, CTR 2019 Update), enriched with structural resistance values (R_struct) and economic value (€) derived from OMI
IDT Threshold	A dataset defining the specific Instability Diffusion Threshold (IDT) for each Slope Unit. This file dictates the minimum spatial extent of instability required to classify a basin as in an "Alert State"

The execution of the framework is orchestrated by a single centralised configuration file (config.json). This file regulates the entire simulation procedure, defining both the rheological parameters for the runout model and the logical criteria for the impact assessment. *Table 31* details the specific SADRI configuration setup adopted for both simulated scenarios.
The procedure employs a hierarchical filtering strategy to isolate unstable Slope Units.
These are specifically identified by comparing the aggregate area of pixels exceeding the fixed Failure Probability Threshold (FPT > 55%, Section 5.2) against the basin-specific Instability Diffusion Threshold (IDT). As detailed in Section 5.3, the IDT is defined via the RE-ACT methodology and calibrated to account for the specific local environmental context. Within these active units, the system identifies potential source areas by clustering contiguous unstable pixels.
The cluster dimensions were set to a range of three to six pixels. This configuration was chosen to align the native DEM resolution (10 metres) with the mean landslide source areas identified using the SEARCH model (300 m^2) and the mean landslide size (614 m^2) recorded during the event (Scenario A).
For the runout simulation, the rheological parameters were adopted directly from the SEARCH optimisation procedure described in Section 5.4.2. The number of DebrisDice iterations and Monte Carlo simulations (N) was set to 100, in accordance with the specifications for regional-scale analysis provided by Bregoli et al. (2018). The framework assesses the potential impact on buildings by providing two distinct risk indicators, calculated at the single-building level in economic terms (euros):

- **Deterministic Indicator (Reference Economic Indicator):**
  This metric quantifies the economic damage derived from a single runout simulation, executed using the optimal rheological parameters calibrated via the SEARCH Model (Section 5.4).

Concurrently, the system defines a spatial "Attention Zone" around the simulated runout. This area is established using a 20 m buffer, a threshold consistent with the proximity distance employed by Confuorto et al. (2025) to identify impacted assets.

However, the framework applies a functional distinction for assets located within the 20 m buffer, classifying this area as the "Attention Zone".

Since these structures are not directly hit by the flow, the Risk (R) is not computed. Instead, their total economic value is reported as "proximal exposure," quantifying the buildings located in the immediate vicinity of the hazard without assuming direct impact;

- **Probabilistic Indicator (Uncertainty Analysis):**

  To account for the intrinsic uncertainty in both the input data and the modelling process, the analysis is expanded using a stochastic approach.

  Specifically, the optimal rheological parameters used to derive the Reference Economic Indicator are perturbed by a variation factor, explicitly set to 20% (0.2) in this study.

  The procedure executes a set of Monte Carlo simulations, generating a probability distribution of potential economic losses rather than a single deterministic value.

  From this distribution, the Mean Expected Risk and the Standard Deviation are calculated, providing a statistical measure of the economic impact that accounts for potential variability in flow propagation.

Using probabilistic simulation allows for the evaluation of different possible scenarios, since varying the rheological parameters consequently changes the simulated runout and the expected consequences.

Therefore, the number of buildings affected by the landslide in the probabilistic analysis may increase or differ significantly from the number initially obtained through the deterministic simulation.

Consequently, structures may be impacted in the initial simulation (with damage evaluated using the Deterministic Indicator) or manifest subsequently through the uncertainty analysis. The affected buildings are further divided into two categories:

- **Assets impacted in MC iterations from initially Attention Zone:** identifies buildings located within the 20 m buffer (Attention Zone) which, although not intersected by the reference deterministic simulation, are affected by the simulated flow in one or more MC iterations;

- **Asset impacted only in MC iterations:** defines structures not initially identified as being at risk, whose exposure emerged exclusively through the probabilistic analysis.

  These represent the Residual Risk: assets impacted by the simulated flow in scenarios characterised by a lower probability of occurrence, yet remaining physically plausible.

Table 31. *SADRI Framework Configuration Parameters*

Parameter	Symbol	Value
Cluster Size (min/max)	-	3/6
FPT Threshold	FPT	0.55
Friction Factor	μ	0.04
Mass-to-Drag Ratio	k	9.47
Reach Angle	α	0.6
Min. Slope Angle	β	0.04
Iterations	N	100
Uncertainty Factor	-	0.20
Velocity Threshold	v_{thr}	0.053
Attention Zone (Buffer, m)	-	20.0

Simulations Results

This section presents the results obtained by applying the SADRI framework to the two distinct rainfall events simulated in the Misa River basin.

Beyond a simple descriptive analysis, the focus here is on quantifying economic impacts and investigating how parameter uncertainty serves to expand the hazard footprint, often reaching significantly further than the boundaries established by a single deterministic analysis.

The rainfall context, which drives the analyses, is summarised in *Figure 50*, highlighting the recorded 6-hour and 12-hour rainfall peaks for both scenarios.

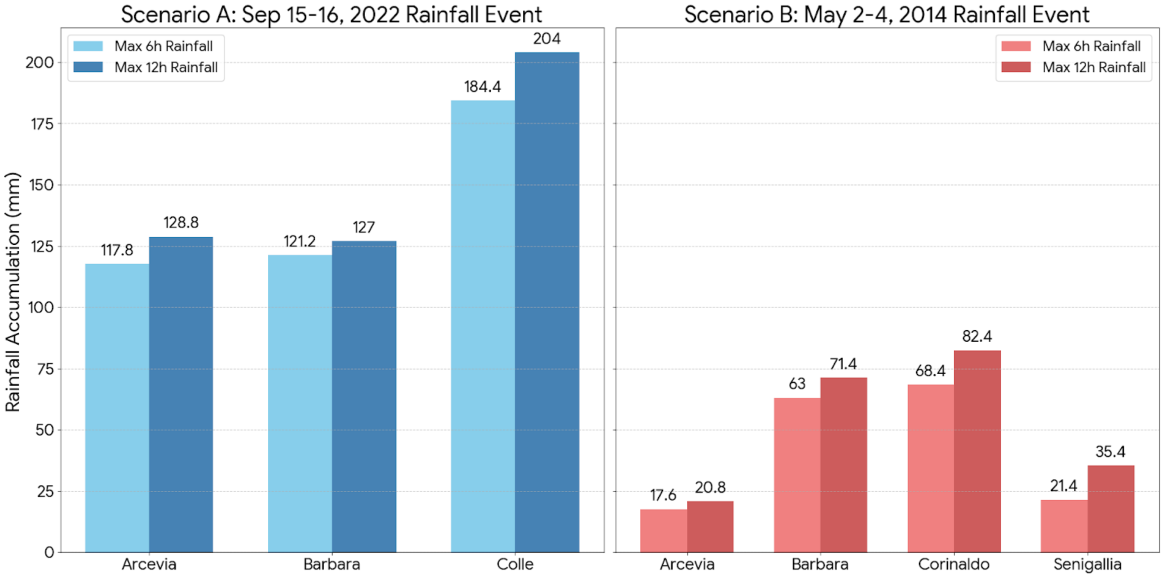


Figure 50. *Comparison of cumulative rainfall for the reference stations in the Misa River basin, highlighting the 6-hour and 12-hour peaks. (Left) Scenario A (September 2022); (Right) Scenario B (May 2014). The x-axis represents the names of the municipalities where the stations are located*

Scenario A Risk Assessment

As previously detailed, the September 15-16, 2022 rainfall event served as the calibration benchmark. The first step involved a deterministic simulation based on the fixed rheological parameters obtained through the SEARCH Model.

Subsequently, a probabilistic scenario (Monte Carlo) was executed by introducing a 20% perturbation to the physical rheological parameters over N=100 iterations.

A clear distinction between the two approaches emerges from the results, as highlighted in *Figure 51.* This figure presents the data aggregated at the Slope Unit level, allowing for a direct comparison of the modelling outcomes: the upper panels contrast the count of impacted assets between the deterministic (left) and Monte Carlo simulations (right), while the corresponding economic calculated damage in euros (€) is detailed in the lower sections.

Looking at the deterministic results (*Figure 51a*), a total of 32 impacted buildings were detected during the event. The spatial analysis shows a diffuse distribution where significant Slope Unit clustering is largely absent, with many units involving only 1 or 2 buildings.

Primarily situated in the southern study area, the units with the highest impact counts act as isolated peaks of vulnerability rather than forming a continuous hazard zone in this single-run simulation.

In contrast, previously undetected risk scenarios appear through the probabilistic approach (*Figure 51b*), which raises the total number of involved assets to 58.

A distinct clustering effect is visible in the south-western and central-southern sectors under the probabilistic lens. In these specific areas, several Slope Units shift from moderate to high impact classes (>5 buildings, in dark red).

This demonstrates that accounting for rheological uncertainty allows for the identification of building clusters situated at the margins of deterministic paths that are impacted by the simulated flow due to potential runout deviations.

The sensitivity of risk estimation to runout variability is illustrated by converting physical impacts into economic damages (€) (*Figure 51c-d*). While deterministic risk yields a total loss of €0.89M, the probabilistic mean risk increases to €0.98M (+10.1%).

The spatial transition between *Figures 51c* and *Figure 51d* demonstrates the movement of several southern basins into higher loss classes (indicated in dark blue), thereby quantifying additional plausible losses that a deterministic single-run simulation would underestimate.

A detailed analysis of the results of the two approaches reveals that 26 buildings were impacted exclusively in the probabilistic simulations (*Table 32*).

These assets are representative of the False Negative (FN) of the deterministic analysis and are classified as follows:

- Assets impacted in MC iterations from initially Attention Zone:

  This category represents 46.2% of the additional impacts (12 assets). These buildings were initially located within the 20 m buffer (Attention Zone) after the deterministic simulation.

  The fact that they were hit during Monte Carlo iterations confirms that the buffer is a useful precautionary tool, as it correctly anticipated areas where the flow could expand. The incremental mean damage for these assets is €1,320;

- Asset impacted only in MC iterations (Not detected in deterministic simulations):

  This category represents 53.8% of the increase (14 assets).

  These buildings are located outside even the 20 m buffer, and their involvement emerged exclusively through probabilistic analysis.

  They represent the Residual Risk: assets hit by the flow only in rare scenarios (with low probability) that are nonetheless physically possible.

  These vulnerabilities appear only by studying the propagation of uncertainty, which reveals runout extensions that deterministic analysis is unable to visualise, with an incremental risk of €55,620.

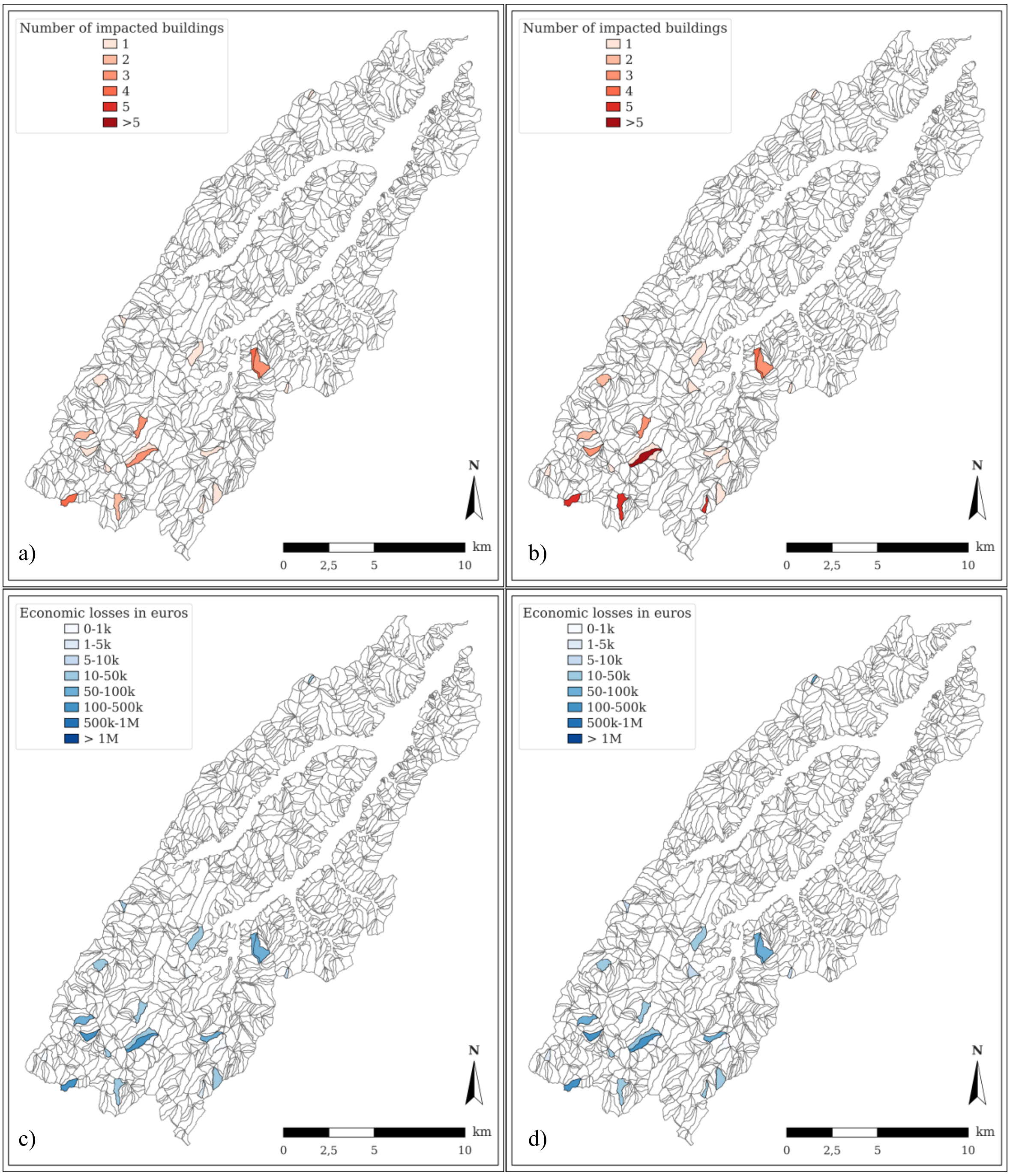


Figure 51. *Comparative impact and risk assessment for Scenario A. Panels (a, b) show the number of impacted buildings, while (c, d) represent economic losses (€), comparing deterministic (left) and probabilistic (right) results*

The subsequent section presents a comprehensive risk assessment for Slope Unit 1300 in the Arcevia municipality, thereby illustrating the model's internal calculation logic.

This Slope Unit was selected due to the presence of a substantial concentration of vulnerable elements along the potential runout path.

The complete set of results (comprising identifiers, categories, economic values and risk figures) can be found in the final PDF report in Appendix A.

Table 32. Comparative assessment of key risk metrics between the Deterministic Scenario and the Probabilistic Scenario (Monte Carlo). The table summarises the total count of impacted assets and the associated economic Risk, highlighting the variation (Δ) and the percentage increase (%)

Key Indicator	Deterministic Scenario	Probabilistic Scenario	Variation (Δ)	Variation (%)
Impacted Assets (Count)	32	58	+26	+81.2%
Economic Risk (Total)	€0.89M	€0.98M	+€0.09M	+10.1%
Assets impacted in MC iterations from initially Attention Zone				
Count	n.a.	12	-	46.2%
Risk	n.a.	€1,320	-	-
Asset impacted only in MC iterations (Not detected in deterministic simulations)				
Count	n.a.	14	-	53.8%
Risk	n.a.	€55,620	-	-

The comparison between the deterministic and the probabilistic results illustrated in Figure 52 demonstrate how the transition from a static to a probabilistic approach affects hazard zoning.

In this figure the buildings are represented as polygons categorised by their interaction with the hazard: green signifies assets never impacted in any simulation (Not impacted assets), red indicates those hit in the deterministic run (Impacted assets (det.)), and orange denotes structures within the initial Attention Zone (Assets in attention zone).

Additionally, black identifies buildings from the Attention Zone that are subsequently impacted during the Monte Carlo iterations (Impacted from initially Attention Zone), while fuchsia represents assets hit exclusively during probabilistic simulations, thereby identifying the specific structures that constitute the Residual Risk (New impacted (MC-only)).

The top-right panel presents the deterministic results, identifying one directly hit asset in red and three others within the 20 m Attention Zone in orange, leading to an estimated economic risk of approximately €46,377 for the impacted asset, while the assets in the attention zone represent a total economic value of €98,307 (total economic value in attention, not calculated risk).

In contrast, the lower panel provides a comprehensive perspective through the Monte Carlo simulations, where each pixel represents the impact probability, quantifying the statistical frequency of flow traversal across all iterations using a colour scale from blue to yellow representing a 1% to 100% likelihood. In this instance, the probabilistic map is derived from 100 Monte Carlo iterations.

The impact probability assigned to each pixel is therefore a direct frequency: for example, a 1% value means that the flow crossed that pixel in 1 out of 100 runs, while higher percentages indicate locations reached more consistently across the stochastic simulations.

The economic impact analysis reveals a crucial distinction between exposure value and expected economic loss. It is important to note that the system correctly differentiates between the total economic value of the asset and the calculated risk value; in fact, the economic value of the building is €71,310, while the calculated risk value is €46,377.

This is indicative of the logic behind the system, whereby the risk value is estimated based on the actual portion of the building involved in the simulated flow, rather than considering the entire value of the building a priori.

The report provides a detailed description of the variation in risk across the stochastic outcomes.

It also offers Monte Carlo summary statistics, which are exclusively computed for the simulations where an impact occurs (hits only).

The average risk is determined to be €61,531, with a standard deviation of €17,836. These values quantify, respectively, the typical expected loss during impact cases and how strongly that expected loss fluctuates from one impact case to another.

Finally, the probabilistic analysis confirms a stable hazard scenario. The Monte Carlo simulations did not identify any newly impacted assets or extend the flow into the Attention Zone. This highlights that for Slope Unit 1,300, the deterministic simulation provided a robust estimation, as the flow variability was insufficient to generate additional Residual Risk. This stability is also confirmed by the post-simulation exposure accounting: after Monte Carlo iterations the impacted set remains 1 asset, with 0 buildings moving from the Attention Zone to the impacted class and 0 new impacted buildings only in MC simulations (Monte-Carlo-only, Residual Risk). Consistently, the aggregate exposure totals remain unchanged, with an impacted-asset value of €71,310 and an Attention-Zone total value of €98,307.

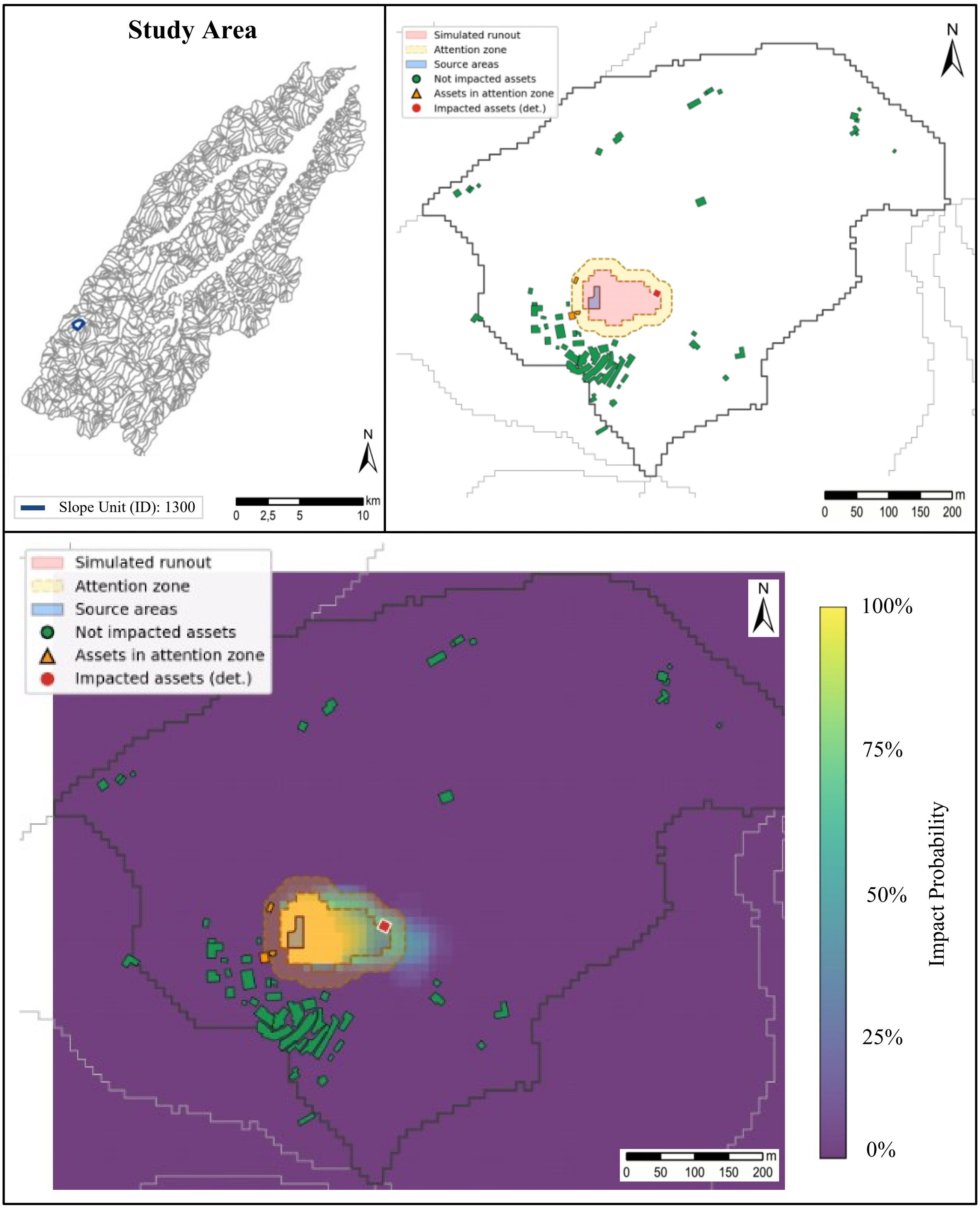


Figure 52. *Comparative analysis of runout simulations for a critical Slope Unit. The top-right panel displays the extent of the reference deterministic simulation. The bottom panel illustrates the probabilistic impact map, where the colour scale represents the pixel-based hit probability (frequency of flow occurrence across all MC iterations)*

Scenario B Risk Analysis

The second scenario pertains to the evaluation of the sensitivity of the SADRI framework to distinct rainfall events. The analysis focuses on the occurrence of 2-4 May 2014, distinguished by the presence of long-duration stratiform rainfall over a basin characterised by elevated levels of antecedent soil saturation.

The event is of fundamental importance in demonstrating the event-based nature of the procedure, which estimates risk based on the specific spatial localisation of rainfall and its interaction with the distribution of vulnerable elements.

It is important to highlight that, despite exhibiting lower hourly rainfall intensity peaks compared to Scenario A, the May 2014 event generated an economic impact risk value approximately five times higher.

This finding can be attributed to the fact that the rainfall in Scenario B triggered instabilities in sectors of the basin characterised by a significantly higher density of buildings and infrastructure.

This confirms the system's ability to conduct basin-specific analyses, where the risk is influenced not only by the severity of the hazard but also by the precise alignment between the triggering areas and the most populous parts of the area.

Specifically, the deterministic simulation identifies 67 impacted buildings, carrying a direct economic risk of €4.93M, a figure more than double that observed in Scenario A.

This damage scenario is further expanded by the probabilistic Monte Carlo analysis, which identifies a total of 106 assets and estimates an expected economic risk of €5.14M. *Figure 53* illustrates the spatial distribution of these impacts, revealing a diffuse damage pattern distinct from Scenario A, with a notable hotspot concentrated in the south-eastern part of the study area.

The Slope Units highlighted in red and orange in the upper panels (a, b) denote areas where the high concentration of buildings directly translates into a higher impact count.

The lower panels (c, d) confirm this trend, showing a marked increase in estimated economic losses in the south-eastern sector, where the probabilistic approach highlights previously undetected risk scenarios that shift several basins into the highest loss classes.

The comparative analysis reveals that 39 assets initially classified as not impacted by the deterministic model were identified as impacted following the Monte Carlo simulations (*Table 33*).

Among these additional assets, 18 were already located within the predefined 20 m Attention Zone. In these cases, the safety buffer functioned correctly by anticipating model uncertainty and proactively flagging vulnerability, with an incremental mean risk of €1,920.

Conversely, the remaining 21 assets were impacted while being located entirely outside the Attention Zone buffer. The latter cases delineate the domain of Residual Risk, wherein the uncertainty in the flow path resulted in the extension of runout in areas not encompassed by the deterministic simulation, thereby introducing a new risk factor, which leads the total calculated risk of about €89,880.

Table 33. *Comparative assessment of key risk metrics between the Reference Scenario (Deterministic) and the Probabilistic Scenario (Monte Carlo). The table summarises the total count of impacted assets and the associated economic risk, highlighting the variation and the percentage increase*

Key Indicator	Deterministic Scenario	Probabilistic Scenario	Variation (Δ)	Variation (%)
Impacted Assets (Count)	67	106	+39	+58.2%
Economic Risk (Total)	€4.93M	€5.14M	+€0.21M	+4.3%
Attention Zone → impacted in MC iterations				
Count	n.a.	18	-	46.2%
Risk	n.a.	€1.92	-	-
MC-only impacted (outside Deterministic Impact and Attention Zone)				
Count	n.a.	21	-	53.8%
Risk	n.a.	€89,880	-	-

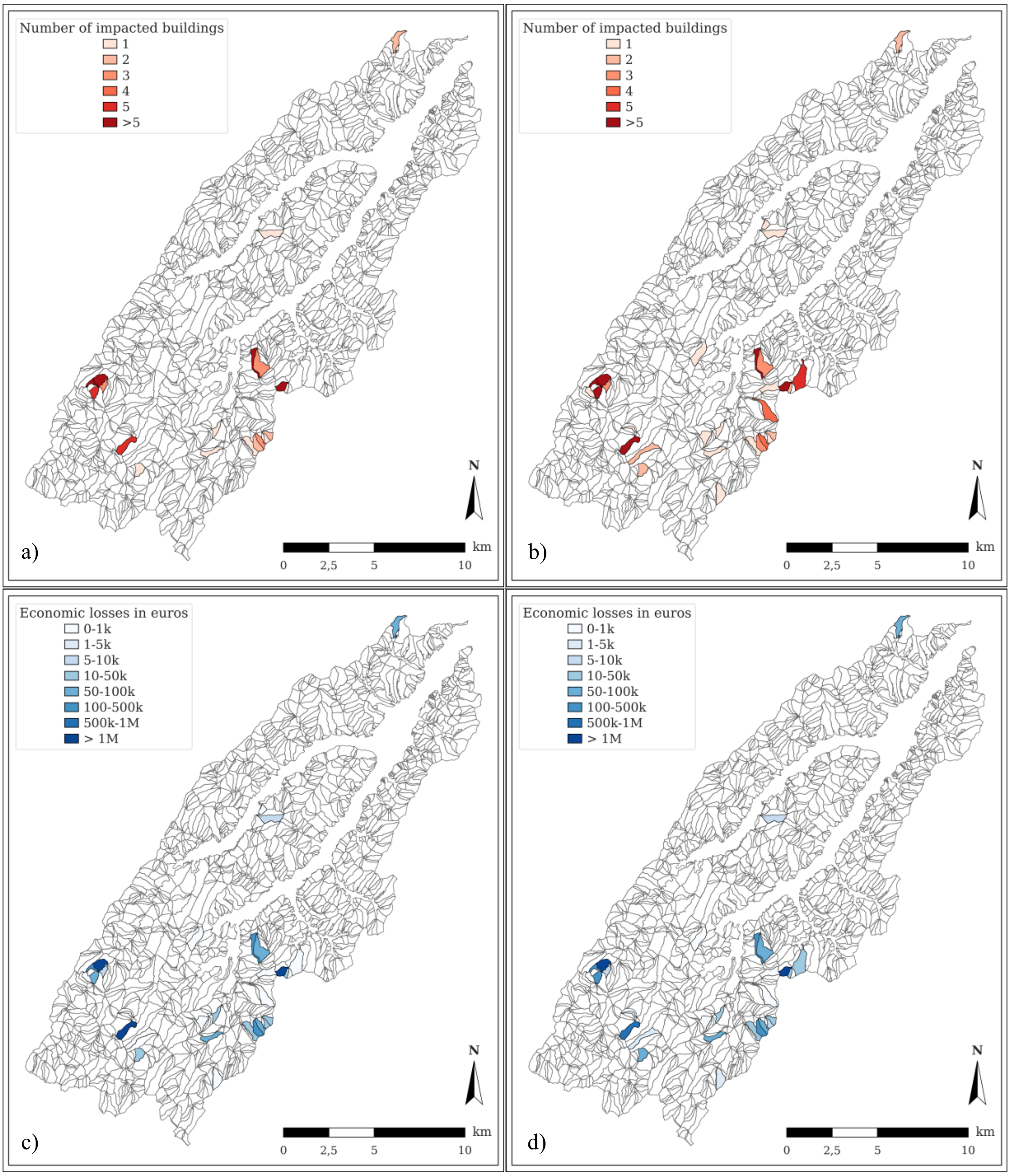


Figure 53. *Comparative impact and risk assessment for Scenario B. Panels (a, b) show the number of impacted buildings, while (c, d) represent economic losses (€), comparing deterministic (left) and probabilistic (right) results*

A detailed risk assessment is presented for Slope Unit 1309 in the Montecarotto municipality (*Figure 54*).
This specific unit was selected because it represents a critical area with a particularly high density of building assets along the potential runout paths, providing a clear comparison between the different modelling methodologies.
As in the previous example (*Figure 52*), buildings in *Figure 54* are shown as polygons and classified using the same colour-based categories.

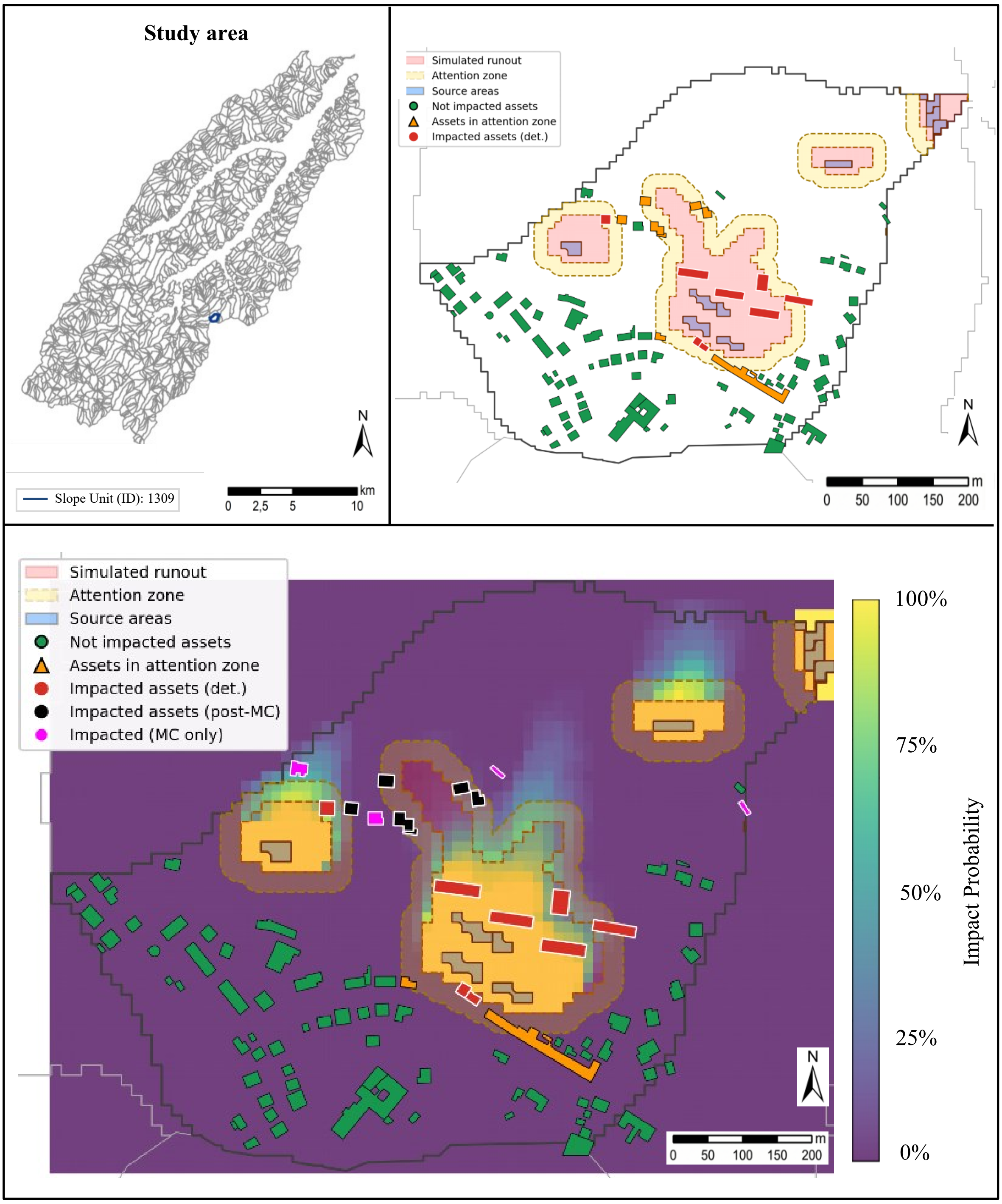


Figure 54. *Comparative analysis of runout simulations for a critical Slope Unit. The top-right panel displays the extent of the reference deterministic simulation. The bottom panel illustrates the probabilistic impact map, where the colour scale represents the pixel-based hit probability (frequency of flow occurrence across all MC iterations)*

The deterministic map (top-right) provides the baseline footprint: it shows 8 buildings reached by the flow and 8 additional buildings within the 20 m Attention Zone (close to the modelled runout but not intersected by it). The economic summary reflects this separation: the 8 impacted buildings have a total economic value of €1.97M and an estimated economic risk (expected loss) of €1.30M, while the 8 Attention-Zone buildings are reported as economic value only, with a total value of €1.69M (no deterministic risk is calculated for these assets).

The lower panel expands the analysis by running 100 Monte Carlo iterations and translating the results into an impact probability map (0-100%), where the pixel-based impact probability is computed over the 100 Monte Carlo iterations.

When this probabilistic evidence is used to reclassify exposure, the number of impacted assets increases to 18 impacted buildings, and the total economic value of impacted assets rises to €2.96M.

At the same time, the value remaining in the Attention Zone decreases to €1.01M, because part of the original buffer population is no longer “only near” the runout but is actually reached in some iterations.

The report also summarises risk variability across the impact cases (“hits only”): the average risk is €1.39M and the standard deviation is €221,725, meaning that when impacts occur, the expected loss is not constant but changes noticeably between different stochastic outcomes.

A key outcome of this unit is the emergence of 10 buildings that are not impacted in the deterministic run but appear as impacted once variability is introduced.

Six of them were initially within the deterministic Attention Zone and become impacted in the stochastic runs (outlined in black): together they account for €683,023 of economic value and €327.08 of additional risk.

Four more buildings (fuchsia) are impacted only in the Monte Carlo results, outside both the deterministic runout and the 20 m buffer, and contribute €48,621 of additional risk. Overall, these 10 “new” buildings add €48,948 of risk in the post-simulation impacted-asset accounting.

For a detailed, single-asset breakdown (IDs, categories, values, and risk figures), refer to Appendix A.

5.5.3 Qualitative validation of risk results

In order to assess the reliability of the SADRI framework, the generated output was validated using the rainfall event of 15-16 September 2022 within the Misa River basin. The strategy is founded on the event-specific inventory provided by Confuorto et al. (2025), and is characterised by a two-fold approach.

1. Spatial Consistency Validation: This validation was made at the municipality scale to verify if the areas identified as "High Risk" by the procedure correspond to the zones that actually experienced the highest instability.
   Since the observed data (landslide counts) and the model outputs (economic loss in Euros) are expressed in different units of measure, the validation evaluates the correspondence of the priority order (ranking) rather than the absolute values;

2. Quantitative Impact Validation: This involves a direct comparison between the damaged building counts reported in Confuorto et al., (2025) and those estimated by the SADRI procedure.
   The goal is to assess the framework's ability to quantify physical impacts using both the deterministic and probabilistic (Monte Carlo) approach.

To ensure a consistent spatial comparison between the model outputs and the observed data, a specific aggregation logic was applied.

Landslides, originally mapped as polygons in the event inventory, were represented as centroid points to facilitate spatial counting within administrative boundaries. Similarly, the modelled economic risk was aggregated from the SADRI slope units to the municipal scale by summing the deterministic losses and the Monte Carlo mean losses for all slope units whose centroids fall within each municipal boundary.

Observed landslide activity was quantified as the number of mapped events per municipality, allowing for a direct cross-referencing of the SADRI-estimated economic losses against the distribution of reported landslides.

The first step of the validation focuses on the prioritisation of the administrative units, as visualised in *Figure 55*, which presents a side-by-side comparison between the observed landslide frequency and the estimated economic risk. According to the observed data, Arcevia recorded the highest number of instability events (474 landslides), followed by Serra de’ Conti (46 landslides).

This specific hierarchy is perfectly mirrored by the SADRI results, where Arcevia exhibits the highest expected economic loss (€0.66M) and Serra de' Conti the second highest (€0.14M). Consequently, the framework correctly prioritises the most critical targets where the majority of the event occurred.

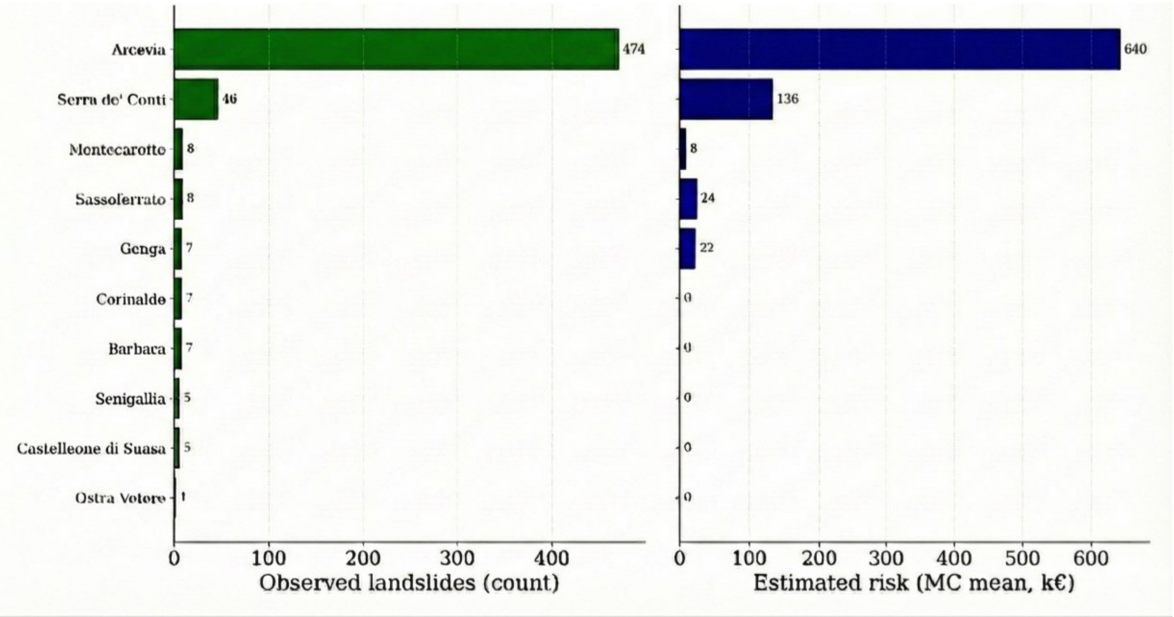


Figure 55. *Comparative analysis of the observed landslides and estimated risk at the municipal scale. Left: Frequency distribution of observed landslides per municipality, derived from the event-specific inventory. Right: Estimated economic risk (Monte Carlo Mean) calculated by the SADRI framework*

Spatially, this correspondence finds further confirmation in the multi-panel cartographic analysis presented in *Figure 56*. The top-left panel shows the municipal territories of the basin with the observed landslides overlaid in black, visualising the geographical distribution of the landslides event across the different administrative units.

This distribution is quantified in the top-right panel of the *Figure 56*, where the observed landslide frequency is mapped using a red colour scale, confirming the concentration of phenomena in the southern municipalities of the basin.

The estimated economic loss maps for municipality are reported in the bottom-left panel (Deterministic) and bottom-right panel (Monte Carlo), depicted using a blue colour scale.

While the spatial pattern is consistent, a difference can be observed between the two simulations: the Monte Carlo approach (bottom-right) exhibits a more consolidated risk distribution compared to the deterministic one, effectively filling gaps in the high-risk zones.

Turning to the quantitative impact on the built environment, a detailed comparison focused on the number of damaged buildings (*Table 34*) reported in Confuorto et al. (2025).

These were identified using a 20 m buffer around mapped landslides to account for mapping uncertainties against the impacted buildings highlighted by the SADRI procedure.

In the SADRI framework, the 20 m proximity buffer corresponds to the "Attention Zone" (buildings close to but not touched by the deterministic runout).

The use of the Monte Carlo (MC) approach proves essential in this context: by varying the rheological parameters, the MC simulations explore a wider range of runout scenarios.

Consequently, buildings that are merely in the "Attention Zone" in a deterministic run are frequently identified as "Impacted" (Loss > 0) in the probabilistic simulation.

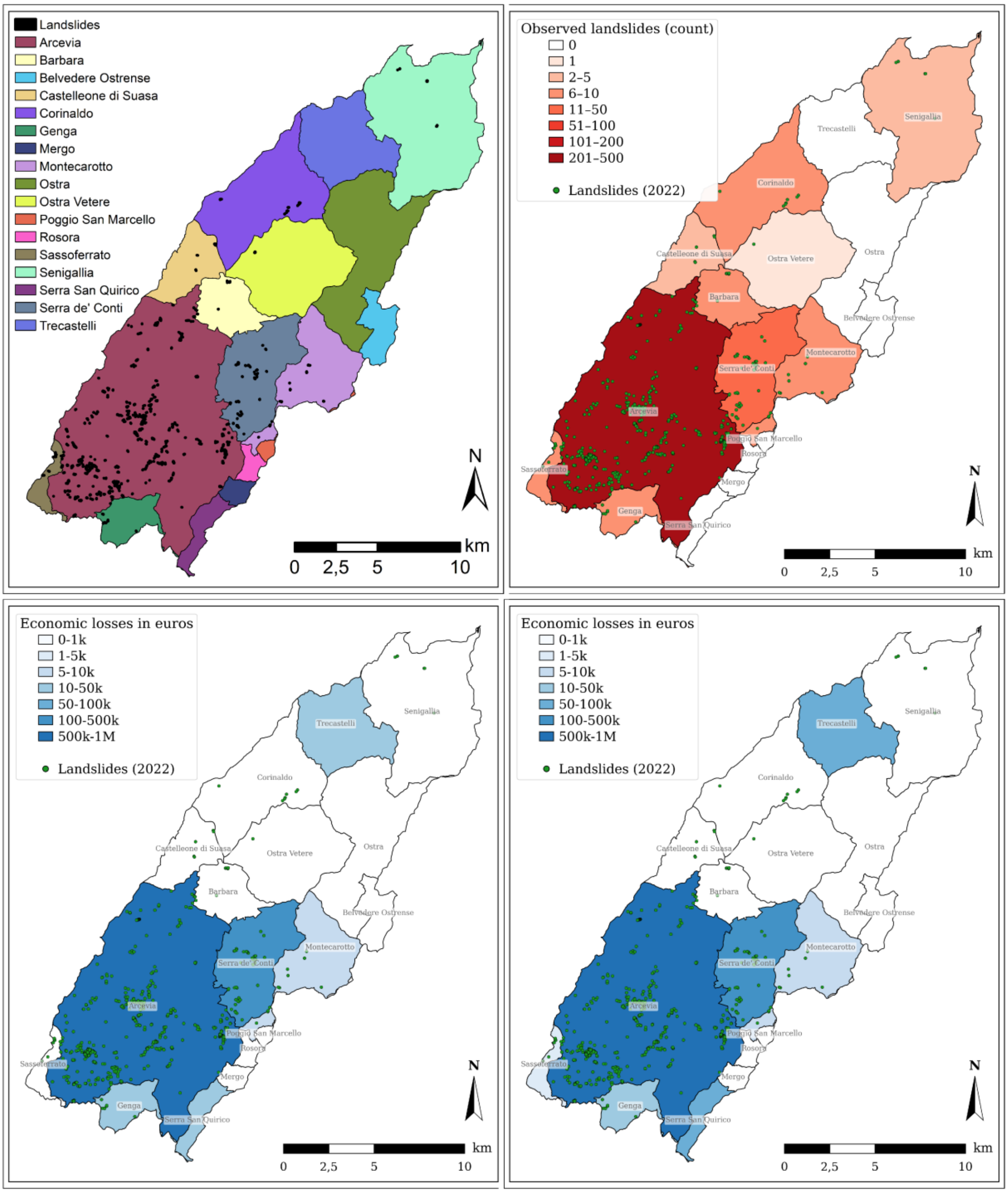


Figure 56. Spatial validation at the municipal scale. Top-Left: Map of the Misa basin showing the municipal subdivisions and the observed 2022 landslides; Top-Right: Number of observed landslides per municipality; Bottom-Left: Economic damages estimated using the Deterministic approach; Bottom-Right: Economic damages estimated using the Probabilistic (Monte Carlo Mean) approach

Table 34. *Comparison between the observed impacts of the September 2022 event and the SADRI model estimates. The table highlights the variations between the Deterministic (Det.) and Probabilistic (MC Mean) simulations. The "Var. (%)" column indicates the percentage increase in detected impacts when utilising the Monte Carlo approach*

Municipality	**Observed Reality**		**SADRI Economic Risk**	**SADRI Impact Analysis (Buildings)**		
	Landslides	Impacted Buildings	MC Mean (k€)	Deterministic	MC Mean	Variation (%)
Arcevia	474	74	640	19	37	+94.7%
Serra de' Conti	46	18	136	7	7	0.0%
Montecarotto	8	4	8	1	1	0.0%
Sassoferrato	8	2	24	0	1	New
Barbara	7	6	0	0	0	0.0%
Genga	7	6	22	2	5	+150.0%
Trecastelli	0	0	76	1	1	0.0%
Serra San Quirico	0	0	74	2	6	+200.0%

As illustrated in *Figure 57*, a comparison is made at a municipal scale between the estimated economic risk calculated by the procedure and the observed impacted building.

The deterministic simulation, constrained by fixed parameters, produces a conservative estimate, identifying a mere 19 impacted buildings in Arcevia.

The framework's capacity to accommodate spatial uncertainty inherent to slope stability is facilitated by the introduction of rheological variability via Monte Carlo.

The resultant data show a marked increase: the mean Monte Carlo count reaches 37 affected buildings (+94.7% over the deterministic run), bringing the figure much closer to the 74 cases documented by Confuorto et al. (2025). Similarly, in municipalities with fewer events like Genga, the probabilistic approach successfully identifies impacts that were missed by the deterministic simulation.

Furthermore, in Sassoferrato, the probabilistic analysis proved capable of detecting impacts that the deterministic run completely failed to identify (new impacted), recovering potential False Negatives.

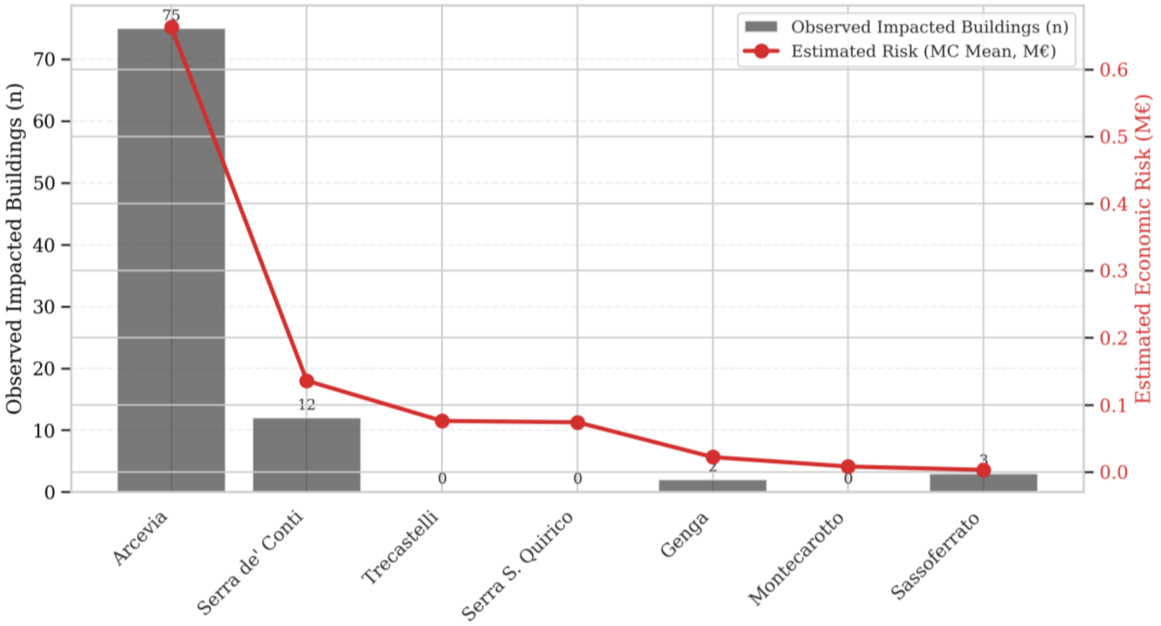


Figure 57. *Graph illustrating the relationship between the physical damage observed during the 2022 event and the risk modelled by SADRI*

Chapter 6

Discussion

This chapter presents a critical synthesis of the research findings, offering a comprehensive evaluation of the SADRI framework and its optimisation modules, RE-ACT and SEARCH.
Beyond reporting the results, the analysis interprets their operational significance for event-based landslide risk management. The discussion follows the logical order of the SADRI framework.
It begins with an overview of the shift from static hazard mapping to dynamic, event-based impact assessment. Then, it examines the specific components of the chain: the reliability of the warning classification (RE-ACT), the objective calibration of the runout (SEARCH), and the economic valuation of risk. Finally, the chapter outlines the future developments of this research.
Each thematic section follows a systematic logic, linking key findings to operational relevance, identifying limitations, and outlining necessary future developments.

6.1 SADRI overview

The SADRI framework was developed to address a persistent challenge in regional landslide risk management. The existing set of risk products is largely limited to two types: static hazard maps, which offer little utility during an active emergency, and hazard assessments that stop at the probability level without delivering impact-based metrics.
This gap is particularly consequential, given that climate change, by increasing the frequency of extreme weather events, is rendering conventional static mapping progressively less effective for operational needs.
To address this, SADRI utilises an end-to-end simulation chain that directly links specific rainfall scenarios with their physical and economic consequences. The system moves beyond standard hazard mapping to provide concrete economic damage indicators in euros.
Crucially, the system provides statistical indicators, such as mean expected risk and standard deviation, to quantify economic uncertainty and propagation variability relative to deterministic results.
In this context, 'risk' is strictly characterised through scenario-based assessment; it quantifies the losses arising from a specific rainfall event and its associated modelling chain, rather than relying on probabilistic expected losses integrated over time-dependent return periods.
This distinction remains fundamental because the severity of a storm is not merely a function of rainfall peaks. Instead, the final impact is dictated by the spatial coincidence of the storm footprint, antecedent soil saturation, and the distribution of exposed assets.
This conceptual framework is substantiated through the comparative evaluation of two distinct scenarios within the Misa River basin: Scenario A (2022), a high-intensity, short-duration event utilised for calibration, and Scenario B (2014), a prolonged event on saturated soil used for functional testing.
The analysis reveals a significant paradox: the lower-intensity rainfall event (Scenario B) resulted in substantially higher economic losses.
Such a finding confirms that impact is not strictly linear to rainfall intensity; rather, it is primarily driven by the intersection of the event with areas characterised by higher economic value and population density.

6.2 Triggering and Alerts in Modelling Chain

RE-ACT addresses a core limitation of landslide warning classification: the reliance on global thresholds (FPT/IDT) to convert continuous triggering probabilities map into discrete warning levels.

In standard practice, this conversion follows a sequential two-step logic defined at the basin scale.

First, the physically based model estimates the failure probability (P_f) for each pixel. A Failure Probability Threshold (FPT) is applied to classify individual pixels as unstable or stable.

Only after this pixel-level classification is established, the system evaluates the spatial extent of the instability. Within each slope unit, the number of unstable pixels is counted, and a warning is issued only if this count exceeds the Instability Diffusion Threshold (IDT).

This structure creates a rigid dependency: the final warning relies entirely on the accumulation of pixels filtered by the first threshold.

If the physical model systematically overestimates or underestimates stability in certain areas, this error propagates directly into the pixel count and, consequently, into the warning classification.

In a heterogeneous catchment, a single fixed pair of thresholds (FPT and IDT) cannot work equally well for steep bedrock slopes and gentle soil-covered hills, failing to capture local differences driven by lithology, land cover, and terrain morphology. The consequence is a recurrent misclassification pattern, identified by cross-referencing model predictions with the observed landslide inventory.

Specific environmental settings are repeatedly flagged as unstable, yet they exhibit no actual failure during the event (false alarms).

Conversely, other sectors fail to trigger a warning despite the documented presence of landslides in the basin (missed events). This implies that the reliability of the forecast is not uniform; rather, it is systematically conditioned by local environmental factors that the rigid threshold logic fails to consider.

The innovation of RE-ACT is not simply using machine learning (ML), but using it as an error-correction layer within the warning chain. RE-ACT does not replace the physical model; instead, it learns where the physical model tends to make mistakes. It acts as a diagnostic tool that refines the warning classification without altering the physical meaning of the probability.

Crucially, RE-ACT maintains consistency with the existing warning logic. Where the baseline procedure produces systematic false alarms, RE-ACT increases the required unstable area (IDT), reducing the number of slope units triggered. Conversely, where instabilities are often missed, it slightly decreases the required area to recover sensitivity. Because these adjustments are bounded, the warning classes remain distinct and consistent with the original calibration.

6.2.1 Strengths

The primary strength of RE-ACT is its ability to adjust warning criteria based on the local environment. Unlike rigid global calibration methods that apply the same rule everywhere, RE-ACT dynamically relaxes or tightens the warning thresholds depending on site-specific conditions. This allows the system to resolve the conflict between detecting all events (sensitivity) and avoiding false alarms (specificity).

Crucially, this is achieved while remaining anchored to the established warning framework. Although the full protocol defines three alert levels (Yellow, Orange, Red), the SADRI procedure utilises the Orange level (IDT = 4 pixels) as the single operational cut-off to classify a slope unit as unstable.

This starting point was chosen because, during the global calibration, the Orange threshold provided the best overall balance (maximising the TSS score). Furthermore, the system relies on a low probability threshold (FPT = 55%) to retain as many potential instabilities as possible from the start, minimising the risk of missing dangerous events (False Negatives).

This specific calibration strategy explains the asymmetric nature of the correction observed in the results. Because the system starts with a very inclusive setting (low threshold), it generates very few missed events but many false alarms. Consequently, RE-ACT acts mainly as a filter:

- To fix False Alarms: It has ample room to raise the threshold (up to +20 pixels) to suppress the noise, effectively "cleaning" the map;
- To fix Missed Events: Since the starting threshold is already very low (4 pixels), the system can only lower it by a maximum of 3 pixels before hitting the minimum definition of a warning (1 pixel). However, since the baseline was already designed to catch almost everything, large downward corrections are rarely needed.

By forcing the adjusted thresholds to stay inside these operational limits (between 1 and 24 pixels), RE-ACT ensures that the warning logic remains consistent with the civil protection protocols, simply optimising the specific trigger point used by the modelling chain.

To ensure robustness, the method was evaluated using a repeated hold-out strategy. In each run, slope units were split into a training set and an independent test set.

This separation is crucial: it proves that RE-ACT is not simply memorising the specific errors of the training data. Instead, the results on the independent test set demonstrate that the system has learned general correction rules that work effectively even on new, previously unseen portions of the basin. The results demonstrate a substantial improvement. Precision increased from 35.00% to 48.95%.

Tangibly, this means that before the correction, nearly two out of three warnings were false alarms; after RE-ACT, the reliability significantly increases, meaning a much higher fraction of warnings corresponds to actual danger. Recall remained stable (rising slightly from 32.31% to 35.40%), ensuring that this filtering did not accidentally suppress real alerts. Consequently, the F1-score improved from 33.60% to 41.10%.

Operationally, the correction removed 56 false positives and recovered 9 false negatives. The reduction of false positives is crucial because it mitigates "alarm fatigue", the tendency of decision-makers to distrust warnings when too many turn out to be false. Furthermore, it prevents the waste of computational resources by avoiding runout simulations for implausible scenarios. At the same time, recovering the 9 false negatives is equally vital. Although the number is smaller, these represent real hazardous events that the baseline model had missed. By recovering them, RE-ACT ensures that these valid source areas enter the hazard-to-impact chain, preventing the system from underestimating the actual risk and the associated potential losses.

A second strength is interpretability. RE-ACT does not apply a blind "black box" correction. Using SHAP-based explanations, it identifies which environmental factors drive the adjustment. This means the method does not only improve the scores, but acts as a diagnostic tool: it reveals exactly where the physical chain tends to fail (over-predicting or under-predicting) and which environmental settings are responsible.

Finally, the correction is conservative by design. This means the system is programmed to be cautious. By enforcing the asymmetric bounds discussed earlier, RE-ACT prevents the warning scheme from becoming too permissive. This ensures that every alert issued remains solid and trustworthy, allowing targeted recovery only where the data strongly supports it.

6.2.2 Methodological limitations

While the results confirm the robustness of RE-ACT, a rigorous assessment must also address its boundaries. The following analysis focuses on the structural limits imposed by the threshold design and the statistical constraints driven by the rarity of errors.

The limited capacity to recover missed events is a direct consequence of the initial strategic choice. The system was intentionally configured with a low probability threshold (FPT) to minimise False Negatives from the start. Therefore, the baseline already captures the vast majority of instabilities. Consequently, the geometric constraint (the fact that the threshold can only be lowered slightly, from 4 down to 1 pixel) is consistent with the operational design: since the system is "inclusive" by nature, it does not require large downward corrections. Instead, the effort is focused on the much wider margin available to filter out False Positives (where the threshold can rise up to 24 pixels), effectively cleaning the signal without risking missed alarms.

A second limitation concerns the validation strategy, which was strictly dictated by the data imbalance. As initially detailed in the methodology, the error classes (False Positives and False Negatives) are rare and spatially clustered. Under these conditions, standard validation methods, such as k-fold cross-validation, proved unsuitable, as they would likely generate test folds containing zero error cases, rendering the diagnostic model unstable. This necessitated the adoption of the repeated stratified hold-out strategy. This approach guarantees that rare errors are statistically represented in every test run. However, it represents a pragmatic compromise: while it stabilises the error estimates, it cannot guarantee full spatial independence, meaning that the performance metrics may still benefit from local similarities between training and testing data.

Finally, the residual error analysis reveals limits in the spatial and temporal resolution of the input data. Even after correction, misclassifications retain a weak spatial clustering (Moran's I = 0.222). This indicates that RE-ACT successfully removes systematic errors linked to geology, average slope, or other predictors available at Slope Unit scale, but cannot correct errors driven by information that is not represented in the input layers. These include local micro-topography, small terrain breaks, anthropogenic controls such as roads or drainage works, and localised rainfall patterns that may not be captured by the available rainfall data: future improvements must focus on acquiring higher-resolution terrain and rainfall information, rather than only refining the mathematical model.

6.2.3 Implications for the SADRI chain

Crucially, RE-ACT represents more than a marginal refinement; it secures the reliability of the initial "gatekeeper" step in the entire hazard-to-impact chain. Within the warning system, the slope-unit map functions as the primary filter: only units classified as "unstable" are passed on to the physical simulator to calculate runout and impact.

Consequently, any error at this stage multiplies downstream. False positives and false negatives do not remain isolated dots on a map; they fundamentally change the shape of the simulated hazard and the list of assets counted as impacted. Specifically, a more reliable warning map enhances the chain in two ways:

- Computational Efficiency: By removing false alarms, RE-ACT stops the system from wasting computer power on simulating runout for spurious sources. This keeps the modelling focused only on physically plausible triggering areas;
- Integrity of Impact Scenarios: Filtering out false positive sources reduces the risk of generating "phantom hazards". These are runout footprints triggered by spurious source areas that would otherwise intersect assets and inflate loss estimates.

Conversely, recovering false negatives ensures that real hazardous events are included in the chain. This prevents the system from missing valid runout footprints and underestimating the actual damage. Ultimately, RE-ACT strengthens the credibility of the entire process: it ensures that the reported impacts given to decision-makers are driven by plausible source areas, not by model noise.

6.2.4 Future developments

The most direct next step is to test RE-ACT on multiple rainfall events and adjacent catchments. This is necessary to distinguish between event-specific noise (temporary errors linked to one specific storm) and stable bias patterns (permanent limits of the physical model).

A second, fundamental development is the activation of the risk-aware component, which was technically implemented in the RE-ACT architecture but disabled for this specific study (Section 5.3). While the current analysis focused exclusively on environmental factors to correct physical model errors, the full framework allows for modulating thresholds based on the elements at risk present in the basin.

In this configuration, the system calculates a normalised risk score for each slope unit, which acts as a second weighting factor alongside the environmental drivers. Practically, this ensures that the warning sensitivity is dynamically adjusted according to the potential impact: the model can automatically lower the triggering threshold in areas with high exposure (prioritising safety for critical assets) while retaining higher thresholds in remote or low-value areas to suppress irrelevant false alarms.

This technical implementation provides the necessary foundation to apply a Cost-Loss analysis. By integrating the risk score, the system evolves from a purely diagnostic tool (fixing statistical inaccuracies) into a decision-support tool. It optimises warnings by weighing the operational cost of a false alarm (mobilisation expenses) against the much higher potential loss of a missed event, effectively minimising the total expected cost to society.

6.3 Runout simulation and calibration (SEARCH)

A major strength of SEARCH is that it solves a well-known problem in simulating landslides: parameter ambiguity. In traditional simulations, it is often possible to get the correct result (a landslide shape that looks like reality) using different inputs. For example, a user might draw the starting source area slightly differently or choose different friction values, and still get a similar final map.

This makes the process subjective: the result depends too heavily on the operator performing the simulation. If two different analysts study the same landslide, they might select different source areas and rheological parameters, producing results that are hard to compare or replicate. SEARCH addresses this issue by reducing operator subjectivity. It transforms the definition of the source area from a subjective manual choice into an objective mathematical calculation.

The effectiveness of this objective approach is confirmed by the calibration results in the Misa River basin, which provide two key insights regarding stability and safety.

First, the calibration proved to be extremely stable across the 5 spatial folds. When the model was tested on different portions of the basin, the key physical parameters (specifically the Friction Factor and the Mass-to-Drag ratio) showed very little variation (very low standard deviation).

This is a critical finding: it proves that the model identified a consistent physical behaviour for the debris flows valid for the entire basin, rather than producing random values that change depending on the location.

The only parameter that showed a variation was the Reach Angle, which split into two distinct groups of values (bimodal behaviour). This is not a model error, but a correct reflection of the reality: the inventory contains two types of landslides, some that stop quickly and others that travel long distances.

Faced with this dual behaviour, the choice to use the median value for the final simulation is a deliberate conservative strategy. It ensures that the model is calibrated to capture the more mobile (and dangerous) landslides, preventing the underestimation of the runout distance.

Secondly, the validation metrics confirm the success of the conservative calibration strategy applied in the objective function. By assigning a higher weight to Recall during the optimisation, the results show a high Recall (the model captures most of the real hazard areas) balanced by a moderate Precision (the model generates some over-estimation). In the context of civil protection, this trade-off is intentional and positive.

It indicates that the system is designed to avoid missing any potential hazard (False Negatives), even if this implies simulating a slightly larger impact area (False Positives). This confirms that SEARCH successfully prioritises the safety of the population over purely geometric perfection.

6.3.1 Strengths

As previously highlighted, the first SEARCH model strength is the automated optimisation of the source area. The model treats the shape of the source as a variable to be solved, just like the physical parameters (friction and turbulence). The workflow uses a hierarchical strategy: for every combination of rheological parameters it tests, it also automatically explores different variations of the source shape area.

It then selects the combination that works best. This means the rheological parameters are not compensating for a subjective manual delineation. Instead, the rheological parameters and the geometry of source areas are optimised together. This ensures identical results when performing the calibration again or if done by another operator (reproducibility).

The second strength lies in the evaluation strategy. Standard methods often rely on a single score (like simple overlap) to decide if a simulation is accurate. SEARCH, instead, uses a multi-metric objective function.

It aggregates multiple distinct indicators, including IoU, F1-score, Hausdorff distance, and PR-AUC, into a single final result. Crucially, this system is flexible: it is possible to assign different weights to each metric depending on the specific goals of the simulation.

This allows the operator to decide whether to prioritise the geometric accuracy of the edges (by increasing the weight of the Hausdorff distance) or the overall coverage of the area (by focusing on overlap metrics), making the calibration process adaptable to the specific needs of each case study.

Additional strengths concern the practical usability and the robustness of the workflow. From a practical point of view, the system is designed to be portable.

The entire simulation and calibration process is controlled by a single text file (config.yml). This design choice simplifies the workflow significantly: it makes it easy to copy the setup to another computer or share it with other researchers, guaranteeing that the experiment can be replicated exactly without ambiguity.

Finally, SEARCH solves a common technical problem related to digital elevation maps (DEM). When simulating small landslides on a coarse map (large grid cells), there are often not enough pixels to define the source area correctly. While standard software often requires the user to manually fix this issue, SEARCH automates this step. It uses a controlled resampling technique to create a smooth, coherent source area even on coarse maps. This contribution is critical because it prevents errors caused by the map resolution from propagating into the simulation of the landslide flow.

6.3.2 Methodological limitations

Despite these strengths, there are three clear limitations to consider regarding the data, the map resolution, and the computational cost.

The first limitation concerns the representativeness and completeness of the calibration data. The model learns strictly from the examples provided during the calibration. In an ideal back-analysis, rheological calibration would be constrained not only by mapped footprints, but also by observed velocity, flow-depth, or travel-time information. In the present regional-scale application, however, such dynamic observations were not available. Calibration was therefore constrained by footprint and runout-geometry agreement, and the calibrated parameters should be interpreted as effective regional values rather than as uniquely constrained physical rheological constants. A related limitation concerns the thematic representativeness of the calibration inventory. If the dataset contains only one specific type of event, such as rainfall-induced shallow landslides and mudflows, the calculated parameters will be optimised only for that specific setting.

Consequently, if the same parameters are applied to a different type of terrain or landslide process, the model will likely perform poorly. This is a limit of the available data, not a flaw in the software itself.

A second limitation arises from the resolution of the Digital Elevation Model (DEM). Even if SEARCH optimises the source area, the flow simulation depends entirely on the quality of the terrain representation. If the DEM is too coarse to represent small details such as narrow gullies, channels, road cuts, or local terrain breaks, the model acts as if these features do not exist.

To compensate for this simplified terrain representation, the optimisation algorithm might force the friction parameters to be lower than they should be, simply to make the simulated landslide reach the observed distance. This implies that the calculated parameters may be mathematically effective for reproducing the observed footprint, but physically misleading because they partly compensate for missing terrain features.

Finally, the method has a high computational cost. Because SEARCH tests thousands of combinations, exploring different source shapes, physical parameters, and cross-validation loops, it requires significant processing time. In the Misa River case study, running the full calibration required thousands of simulations, making it much more time-consuming and demanding than a standard manual analysis.

6.3.3 Future developments

The next steps for improving SEARCH focus on two key aspects: verifying if the model works in similar environments (regional transferability) and understanding how much map errors affect the result (uncertainty).

The most direct step is to test the spatial transferability to adjacent or geologically similar catchments. Future work should verify if the calibrated parameters can validly simulate landslides in neighbouring basins that share the same lithology and climate as the Misa River. This test is crucial to distinguish between regional validity and local overfitting. If the model works well in a similar nearby area, it means the calibrated parameters correctly represent the physical properties of the materials in this region.

Conversely, if it fails, it implies that the parameters were forced to compensate for the specific local anomalies of the training site, limiting their use to strictly local scenarios.

A second development involves quantifying the uncertainty of the Digital Elevation Model (DEM). To ensure that small errors in the DEM are not distorting the results, SEARCH should be tested using an ensemble of DEMs. In simple terms, this means running the simulation on a collection of different DEMs (for example, DEMs with slightly different resolutions or artificially added noise) rather than relying on a single one.

This approach allows us to measure the stability of the result: if the simulated landslide path changes drastically with a small DEM variation, the result is uncertain. This moves the analysis from providing just one "optimal" result to providing a transparent range of possible outcomes.

6.4 Risk Assessment

6.4.1 Exposure and Vulnerability Modelling

The definition of the economic module involved an evaluation of various methodologies for risk estimation. Following this assessment, a deliberate choice was made to align the SADRI framework with the exposure and vulnerability models previously established for the Misa River basin by Caleca (2024). This decision is scientifically strategic and maintains continuity with existing workflows for two primary reasons: reproducibility and comparability.

First, aligning with the protocol defined by Caleca (2024) ensures reproducibility. The exposure assessment relies on the official OMI database (Real Estate Market Observatory), which provides verifiable minimum and maximum property quotations categorised by territorial zone and building typology.

Following this standardised procedure, the economic value is calculated by matching each building's location with its corresponding OMI zone and multiplying the mean unit value by the building's footprint area. This rigorous use of public data allows for a transparent and replicable valuation process.

Second, this alignment guarantees comparability between results. By keeping the economic calculation fixed to the baseline set by Caleca (2024), the analysis isolates the variables related to the physical process.

This methodological consistency ensures that any variation in the final loss estimates is attributable strictly to improvements in the landslide simulation, specifically in the triggering and runout phases, rather than being conflated with methodological changes in the economic valuation.

However, adopting this basin-scale workflow implies specific limitations regarding data resolution. Building footprints are occasionally incomplete in rural areas, and OMI quotations represent average market values aggregated over large zones.

Consequently, the calculated exposure serves as a consistent basin-scale approximation rather than a precise commercial valuation for individual properties.

Similarly, the vulnerability model relies on census data to derive general structural information. While this approach allows for a spatial distinction between different construction types, such as masonry and reinforced concrete, it cannot capture building-specific details like recent retrofitting or maintenance status.

Therefore, the vulnerability values represent a scalable estimation suitable for large-scale planning, but they do not constitute a direct prediction of damage for a single specific house.

A necessary advancement for future research involves introducing economic uncertainty into this framework. Instead of relying on fixed mean values, the model should sample a range of market prices and vulnerability parameters. This development would ensure that the final risk calculation reflects not only the physical uncertainty of the landslide hazard but also the inherent economic uncertainty of the exposed assets.

6.4.2 Risk Analysis

A central outcome of the analysis is that deterministic outputs, while simple to use, do not represent the full range of plausible impacts. The deterministic simulation provides a single footprint and therefore a single set of impacted assets and losses. In contrast, the probabilistic approach (Monte Carlo), implemented through Monte Carlo perturbation of rheological parameters (20% over $N = 100$ iterations), produces an ensemble of runout outcomes that translates into a distribution of affected assets and economic losses.

To quantify the added value of the proposed framework, the results of the two scenarios were analysed to observe the variations generated by the probabilistic approach relative to the deterministic one.

The aggregate data show that the deterministic simulation, by relying on fixed parameters, yields a more conservative output that functions as a baseline representing a restricted extent of potential impact.

Conversely, the probabilistic approach provides a different risk perspective by accounting for the aleatory uncertainty inherent in flow propagation, which in these specific cases led to an increase in the calculated risk. It is essential to clarify that these results represent two distinct modelling philosophies rather than a definitive "correct" or "incorrect" solution. While the probabilistic method explores a wider range of possibilities and identifies higher exposure, neither approach has been externally validated against field data to determine which is objectively more accurate. Ultimately, these complementary modelling approaches serve as a critical tool for evaluating and visualising a diverse range of possible risk scenarios.

This difference is evident in both simulated rainfall scenarios. In Scenario A (September 2022) the deterministic simulation identifies 32 impacted buildings with an associated economic loss of €0.89M, while the probabilistic approach increases the number of impacted assets to 58, pushing the mean economic loss to €0.98 million. The aggregate data (summarised in *Table 35* and *Figure 58*) reveals a specific behaviour: the economic loss grows moderately (+10.1%), but the number of affected assets nearly doubles (+81.2%). This suggests that uncertainty primarily affects the flow boundaries.

Different combinations of rheological parameters produce scenarios that are more or less expansive, impacting a variable number of buildings compared to the single-run simulation.

It is important to note that all these outcomes represent plausible scenarios, as they are generated by varying the parameters within a physically valid range. Consequently, many buildings located just outside the path of the fixed simulation are impacted in the more critical scenarios. The deterministic simulation, by showing only the central trajectory, effectively hides the risk for these assets located at the margins. In Scenario B, the deterministic simulation already identifies a critical scenario, with losses reaching €4.93M involving 67 assets. However, the probabilistic analysis significantly expands the impact assessment: the mean loss rises to €5.14M, and the number of involved structures increases considerably, reaching 106 impacted assets.

Comparing the two events highlights a crucial finding: despite having lower rainfall peaks compared to Scenario A, the May 2014 event generated an economic impact five times higher.

This indicates that risk is not controlled by rainfall intensity alone. What dominates is the spatial overlap (the geographic intersection) between the landslide and the exposure.

In Scenario A, the triggered flows often initiated in areas not in immediate proximity to urban settlements, or simply did not extend far enough to reach the buildings, even where present.

Conversely, in Scenario B, the rainfall triggered instabilities in sectors characterised by a high density of buildings located directly along the potential runout paths.

This comparison confirms that the SADRI framework correctly captures the *event-based* nature of risk: the final damage depends on exactly where the hazard intersects with the populated areas, not just on the magnitude of the storm.

Table 35. *Comparative analysis of risk metrics and asset impact between the Deterministic and Probabilistic approaches*

Metric	Scenario A (Calibration)	Scenario B (Test)
Impacted Assets (Deterministic simulation)	32	67
Impacted Assets (Probabilistic simulation)	58	106
Probabilistic Increment (N° Additional impacted assets)	+26	+39
Attention Zone (20 m buffer)	12 (46%)	22 (56%)
Residual Risk (Assets initially not listed as at risk)	14	17
Economic Risk (Deterministic simulation)	€0.89M	€4.93M
Economic Risk (Probabilistic simulation)	€0.98M	€5.14M

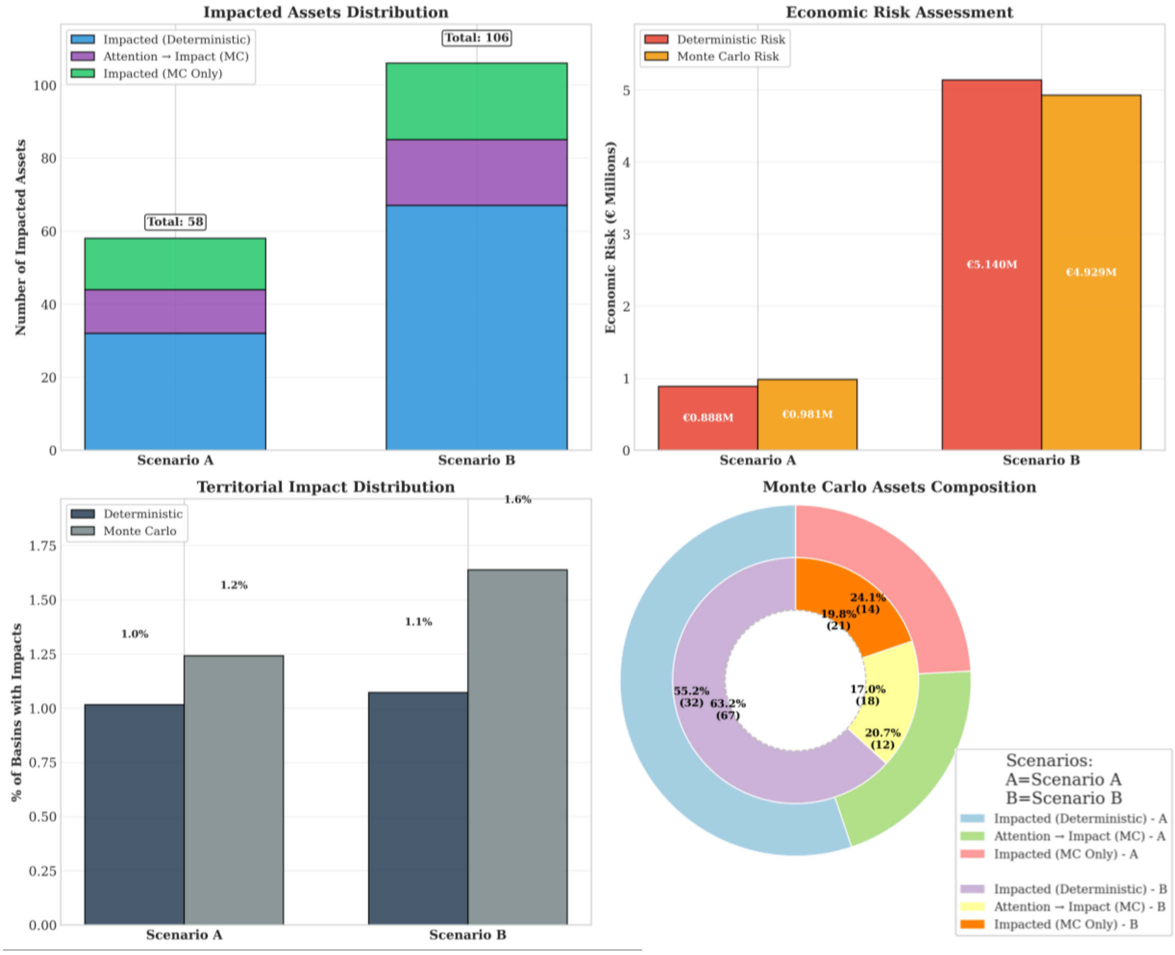
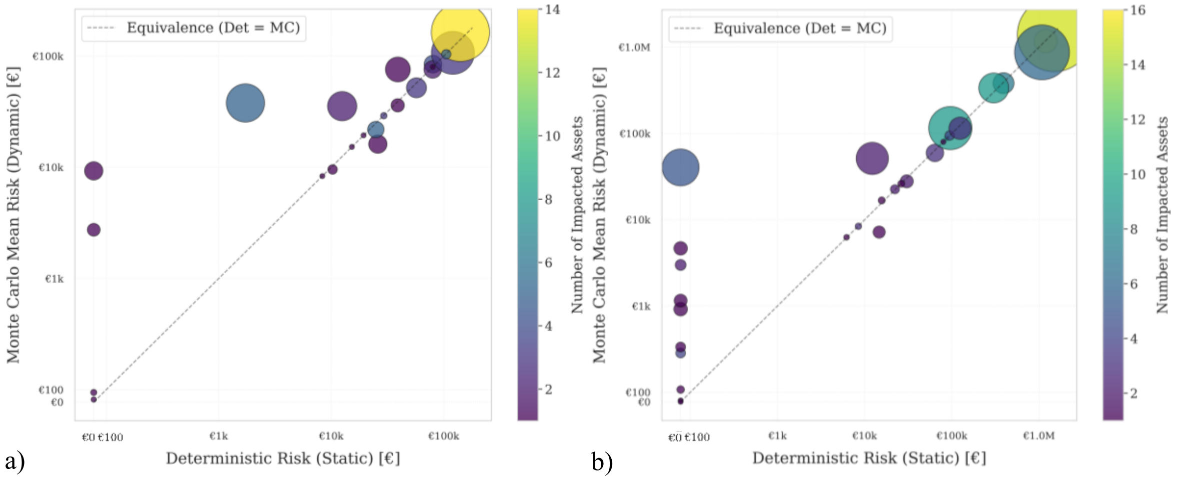


Figure 58. *Summary of Comparative Risk Metrics. The multi-panel figure illustrates the aggregate results for Scenario A and Scenario B*

This divergence is visualised in the scatter plots in *Figure 59*, which correlate the risk calculated by the deterministic simulation (x-axis) against the probabilistic mean risk (y-axis).
The dashed diagonal line represents perfect agreement. Since most data points are positioned above this line, the graph confirms visually that the probabilistic risk is systematically higher:

- Scenario A (*Figure 59a*): Most points cluster near the diagonal, indicating that for many basins the simulations converge.
  However, significant outliers in the upper section represent specific high-value areas where the single-run simulation failed to detect the potential impact, which emerges only when parameter uncertainty is considered;

- Scenario B (*Figure 59b*): The divergence is much more pronounced, with data points shifting markedly towards the top-right corner (high risk).
  This visually confirms the severe underestimation in the densely populated sectors discussed previously. The probabilistic approach captures the risk caused by the lateral spreading of the flow in these complex areas.
  The large yellow clusters highlight specific basins where the single-run simulation overlooks significant economic losses.



a) b)

Figure 59. *Risk Assessment Comparison Scatter Plots. (Left) Scenario A; (Right) Scenario B. The size of the bubbles represents the number of impacted assets. Points above the diagonal line indicate Slope Units where the Probabilistic Risk exceeds the Deterministic Risk*

Finally, the results clarify the role of the Attention Zone. The 20-meter buffer used in the deterministic simulation is designed to manage proximal exposure.

This term refers to assets located in the immediate vicinity of the landslide path which, although not directly hit by the principal flow, are close enough to be considered potentially at risk due to small deviations.

The analysis demonstrates that this buffer is effective but not sufficient.

Specifically, the buffer successfully intercepted 12 of the additional assets identified by the probabilistic model in Scenario A (46% capture rate) and 22 assets in Scenario B (56% capture rate).

However, a significant number of buildings were impacted entirely outside this safety distance: 14 in Scenario A and 17 in Scenario B.

These assets represent residual risk, defined as the potential impact that remains unaccounted for even after the application of standard safety margins.

It refers to those buildings located far from the expected path (outside the 20-meter buffer) which are hit only when the landslide extends unexpectedly due to parameter variations.

This demonstrates that a fixed geometric margin cannot fully represent the physical uncertainty of a landslide. Operational monitoring should therefore evolve from static buffers to dynamic attention perimeters derived directly from the probabilistic dispersion of the model.

6.4.3 Scope and limitations

A critical aspect of the discussion concerns the validation strategy adopted for the risk metrics. The validation was necessarily constrained by the nature of the available post-event data.

While the landslide inventory provided by Confuorto et al. (2025) offers a precise spatial distribution of the phenomena and a count of the involved structures, a coherent database of the actual monetary damages suffered by private and public assets is currently unavailable.

Consequently, a direct economic validation (i.e., comparing estimated Euro losses vs. paid insurance claims or reconstruction costs) was not feasible.

To overcome this limitation, the validation relied on proxy indicators: the hierarchical ranking of municipalities (i.e., ordering the administrative units from the most affected to the least affected) and the physical count of impacted buildings. This approach assumes that while the absolute economic values estimated by SADRI are model-based projections, the number of impacted buildings should mirror the observed reality.

The correct alignment between the observed landslide frequency and the modelled economic risk ranking confirms the validity of this assumption, as shown in *Figure 57*.

This indicates that, even in the absence of quantitative economic ground truth, the framework is robust enough to correctly identify the priority areas for intervention, which is the primary operational requirement for civil protection purposes.

The interpretation of results depends heavily on the quality of the rainfall data used to drive the simulations. Scenario A (2022) utilises radar data, which provides a continuous, detailed grid of rainfall values across the entire basin. Conversely, Scenario B (2014) relies on data from scattered rain gauges.

To assign a rainfall value (mm/h) to every point of the study area in Scenario B, the Thiessen polygons method was used. This technique divides the basin into specific "areas of influence": every point falling within a polygon is assigned the exact same rainfall value recorded by the corresponding gauge.

The limitation of this method is geometric: if there are few gauges, the resulting polygons become very large. Consequently, a slope located many kilometres away from the recording station is treated as if it received the same rainfall intensity as the station itself.

In reality, rainfall can vary significantly over such distances. This artificial "flattening" of the data means that the model might predict high risk in a distant area simply because it falls within the large polygon of a station that recorded a peak. Therefore, the divergence in the results is partly driven by this coarse spatialisation of the input data rather than solely by the physical model itself.

A second limitation lies in the scope of the exposure module. Currently, the system assesses risk exclusively for buildings. This exclusion was primarily dictated by the scarcity of homogeneous and up-to-date datasets suitable for a basin-wide economic valuation of linear assets.

Furthermore, estimating damages for infrastructure presents a higher degree of complexity compared to residential stocks: it requires not only calculating the physical reconstruction costs, which depend on detailed and often unavailable engineering specifications, but also accounting for the significant indirect economic losses caused by service disruption and network severance.

However, landslides at the catchment scale frequently disrupt critical infrastructure. Therefore, a clear priority for future development is extending the inventory to include road networks, bridges, railways, and utility lines, which are major components of the overall economic damage.

A further limitation concerns the spatial resolution of the input topography. Both the triggering and runout components of SADRI are affected by the resolution of the terrain data.

In this application, the modelling chain is based on a 10 m Digital Elevation Model (DEM), which is suitable for basin-scale analysis but cannot fully represent small topographic and anthropogenic features, such as minor roads, road cuts, retaining walls, culverts, drainage works, embankments, small channels and building-level geometric details.

This limitation mainly affects the first physical output of the chain: the Failure Probability (Pf) maps produced by HIRESSS. If small scarps, artificial cuts, narrow drainage lines or local slope variations are smoothed within the 10 m grid, some potential initiation conditions may be simplified or displaced in the triggering map.

The FPT/IDT classification, and consequently the set of Slope Units later analysed by RE-ACT, inherit this uncertainty from the initial physical modelling step.

RE-ACT can reduce systematic classification errors when they are associated with the environmental information available at Slope Unit scale, but it cannot recover local topographic controls that are not represented in the input data.

The effect is more direct for SEARCH, because runout calibration depends on the DEM used to define source areas, route the simulated flow, and compare simulated and observed footprints.

If narrow channels, road embankments or local terrain breaks are not represented, the simulated path, runout distance, and overlap metrics may be affected. This reinforces the regional nature of the calibration: the parameter set is appropriate for basin-scale scenario modelling, but should not be interpreted as able to reproduce all site-specific topographic controls.

These limits also influence the reliability of asset-level economic estimates. Buildings are intersected with the simulated runout footprint and velocity raster after the propagation simulation.

At local scale, small spatial shifts in the footprint or velocity field may change whether an asset is classified as impacted, the exposed fraction of the building and the maximum velocity assigned to it.

Therefore, the framework is best used as a basin-scale decision-support tool to identify plausible impact zones and guide further local investigations, rather than as a high-resolution building-scale model.

6.5 Model Transferability

The transferability of the SADRI framework must be evaluated by distinguishing between its methodological workflow, the calibration phase, and its operational application.

While the modelling workflow is inherently transferable, as its modular structure can be adapted to different study areas, the calibrated thresholds and rheological parameters are site-specific.

They cannot be automatically extrapolated from one basin to another without local recalibration. This process requires a suitable local landslide inventory, which serves two primary functions:

- In RE-ACT, the inventory allows the comparison between initial model classifications and observed landslide occurrences, enabling the identification of systematic errors and the fine-tuning of instability thresholds;

- In SEARCH, the inventory provides the observed geometries required for back-analysis, source-area testing, and calibration of the runout parameters.

Once this calibration has been completed for a specific basin and for the landslide processes represented in the dataset, the system becomes an operational tool.

For instance, after calibrating the Misa River basin for rainfall-induced shallow landslides and mudflows, SADRI can process subsequent rainfall events within the same context without repeating the calibration phase. New rainfall data are processed through the triggering model, and the outputs are analysed using the established thresholds, runout parameters, and exposure data.

However, these calibrated values should not be assumed valid in different geological or geomorphological settings.

In a new basin, the relationship between triggering-model errors and environmental variables may differ, and runout behaviour may depend on distinct material properties, slope morphology, channelisation, or landslide mechanisms.

Consequently, applying SADRI to a new region requires a local inventory to recalibrate both alert thresholds and rheological parameters.

A further constraint concerns process consistency. Calibration must be performed on landslide types that are compatible with the underlying modelling chain.

In this thesis, the focus is on rainfall-induced shallow landslides and mudflows. If a study area is dominated by different phenomena, such as deep-seated landslides, rockfalls, or rock avalanches, the existing parameters would not be physically appropriate.

In such cases, the inventory must be disaggregated by process type, and calibration should be restricted to those mechanisms that the triggering and runout models can realistically represent.

In conclusion, while SADRI is transferable as a modelling architecture, its parameters remain site- and process-specific.

Once locally calibrated, the system can be used repeatedly for dynamic landslide impact and risk assessment within the same geographical and physical context.

6.6 Synthesis and Future Development

The results obtained in the Misa River basin show that an integrated modelling chain that propagates uncertainty from triggering through runout and, subsequently, to economic consequences provides clear methodological value.

This means that the proposed framework does not treat the landslide as an isolated physical event, but rather as a continuous flow of information, from the initial slope failure to the final impact on the local economy.

SADRI provides its main advantage because it converts physically based hazard data into decision-relevant information: it identifies not only the areas of greatest hazard, but also the sectors subject to the most severe ground effects, and it quantifies expected damage through tangible economic indicators.

Instead of producing susceptibility maps alone, the framework provides quantified risk metrics expressed in monetary terms, which are more directly usable for decision-makers.

Future developments should focus on three main directions. First, the uncertainty introduced by spatial discretisation should be quantified more explicitly.

In this context, spatial discretisation refers to the fact that terrain, source areas and runout paths are represented on a raster grid, so the result may change when the DEM resolution changes.

This could be tested by using higher-resolution DEMs and, where available, DEM ensembles, i.e., sets of alternative terrain representations.

Such tests would help evaluate how stable source extraction, runout propagation and local impact estimates are with respect to the terrain representation.

Second, the risk-aware component of RE-ACT should be activated. In the present application, threshold correction is based on environmental error patterns; in future applications, the correction could also account for the potential consequences on exposed assets, lowering the threshold where exposure is high and keeping it more selective where exposure is low.

Third, exposure and vulnerability should be treated as uncertain inputs rather than fixed values. This means that asset values, structural resistance parameters and vulnerability functions should not be represented by a single deterministic value, but by plausible ranges or probability distributions.

For example, building values could vary within the minimum and maximum OMI quotations, structural resistance could account for uncertainty in construction type or maintenance state, and vulnerability curves could include alternative damage–intensity relationships.

In this way, SADRI would propagate uncertainty not only from triggering and runout modelling, but also through the exposure and vulnerability components that control the final economic loss estimate.

Chapter 7

Conclusion

This final chapter summarises the main outcomes of the research and relates them to the objectives and research questions defined at the beginning of the thesis. It also discusses the methodological contribution of the proposed framework and outlines the main directions for future development.

This doctoral research addressed one of the most complex challenges in landslide risk management: bridging the gap between advanced Physically Based Models (PBMs) and the operational requirements of Civil Protection early warning systems.

The primary objective of this work was to overcome current limitations by developing SADRI (Scenario-Aware Dynamic Landslide Risk & Impact): a modular, fully automated, and event-based computational framework.

This dynamic approach is of particular importance in the context of climate change, because the increasing frequency of extreme rainfall events is making traditional static risk maps progressively less effective for operational decision-making.

Such maps provide a fixed picture of hazard based on past statistics, but they cannot anticipate the specific impact of a real-time storm.

The thesis demonstrates that the success of the simulation chain relies on the strategic integration of heterogeneous techniques, combining the rigour of classical physics (which provides the fundamental, immutable laws governing stability) with the versatility of data science (which allows the system to "learn" from data, correcting errors and optimising parameters).

In parallel, the increasing availability of high-resolution structured databases, ranging from updated environmental variables to technical-structural characteristics of the built environment and updated cartography, does not merely enrich input data but constitutes the fundamental prerequisite for applying advanced algorithms.

It is precisely the convergence of physical modelling rigour, the diagnostic capability of Machine Learning (ML), and Bayesian optimisation techniques that enables a methodological leap: the transition from a static susceptibility assessment to a dynamic and economic risk quantification.

The final result is no longer limited to a qualitative Probability of Failure map (which indicates only where a landslide is likely to trigger), but extends to dynamic impact maps that display quantitative estimates of expected damages on the ground, resulting from the full physical evolution of the specific rainfall event.

The scientific value of this research lies in the engineering of a complete simulation chain, where each critical phase was resolved through the development of specific modules working in synergy to reduce overall uncertainty:

1. Dynamic Threshold Optimisation (RE-ACT Module): To translate physical maps into reliable Civil Protection alerts, the framework moved beyond the logic of static global warning thresholds (fixed probability values applied blindly to the whole basin to trigger an alarm).
   Through the use of Explainable AI (XAI) techniques, the system identifies local geomorphological conditions that induce uncertainty in the physical model, generating dynamic, site-specific alert thresholds, effectively relaxing or tightening the warning criteria based on the local environment.

This module acts as a necessary spatial filter to clear the physical signal of false positives before proceeding to economic assessment;

2. Objective Rheological Parameter Calibration (SEARCH Module): A reverse calibration procedure based on Bayesian optimisation was formalised to identify optimal parameter sets.
   This approach markedly reduces the subjectivity associated with the manual definition of geotechnical and rheological parameters. By deriving the parameter set through back-analysis of historical reference events, the procedure improves reproducibility and provides a transparent basis for regional runout modelling;

3. Event-Based Economic Evaluation: The fundamental contribution of the thesis is the completion of the risk chain (R = H × V × E). The framework translates physical hazard (flow velocity and extent) into monetary impact, clearly differentiating the effects of diverse rainfall scenarios (e.g., convective vs. stratiform events). The output is provided via two indicators:
   - *A deterministic indicator,* for a direct estimate of expected damage;
   - *A probabilistic indicator,* obtained via Monte Carlo simulations, which quantifies estimation uncertainty, offering decision-makers a confidence range regarding the potential economic impact.

The operational validity of this architecture was demonstrated through its application to the Misa River Basin, which highlighted how the proposed approach changes decision support in Civil Protection contexts:

- Risk Communication: SADRI allows moving beyond generic hazard classes (e.g., High, Medium, Low, often difficult to interpret operationally) in favour of communication based on the economic value of expected damage.
  This quantitative language is significantly more effective for justifying preventive measures (such as preventive evacuations) and for dimensioning the economic and logistical resources required for emergency management;

- Reliability and Prioritisation: Thanks to the spatial filtering performed by dynamic thresholds and targeted calibration, the system substantially reduces the phenomenon of event overestimation (False Positives).
  This mitigates the risk of "operational fatigue", the tendency of operators to ignore alarms after repeated false warnings, ensuring that attention is prioritised on areas where the predicted economic impact is genuinely significant.

Looking forward, transforming the framework into a full-scale operational system requires addressing certain limitations that emerged during the study, outlining the path for future developments:

- Topographic and Structural Data Resolution: The analysis highlighted that the current DEM resolution (10 m) represents a constraint for detailed modelling.
  It fails to capture anthropogenic micro-elements (such as retaining walls, roads, or culverts) that influence real flow propagation, nor does it allow for a precise geometric characterisation of individual structures.
  The future integration of high-resolution LiDAR data (1 m) and a building database with more detailed structural attributes is the necessary step to further refine damage estimation at the single-building level;

- Geographic Transferability: Currently, the system is validated on a pilot basin. To extend its application to other geomorphological contexts, future work should test the framework on a broader set of historical events and recalibrate both the alert thresholds and the runout parameters where local inventories are available;

- Software Integration and Reusability: Future developments should further consolidate SADRI as a reproducible and reusable computational workflow.
  This means standardising input data formats, defining clear interfaces between the modules, automating simulation execution and quality-control checks, and exporting results in formats that can be used by other decision-support systems.
  In this perspective, SADRI is not only a scientific modelling chain, but also a software environment that can support operational integration, programmatic reuse and future application to other territories or platforms;

- Multi-Risk Approach: The analysed events confirm the frequent concomitance of landslide and flood phenomena. The natural evolution of SADRI involves coupling with hydraulic models to simulate cascading scenarios (e.g., river damming and subsequent outburst floods), evolving toward an integrated multi-hazard decision-support system.

In conclusion, this thesis proposes a reproducible methodological framework for regional-scale quantitative landslide risk assessment. SADRI demonstrates that the intelligent integration of physically based modelling with multi-source data allows moving beyond traditional hazard maps, providing concrete, dynamic operational responses based on the real economic impact of extreme events.

References

AGS (2007). Practice note guidelines for landslide risk management. *Journal and News of the Australian Geomechanics Society*, 42(1), pp. 63–114.

Agliardi, F. and Crosta, G. B. (2003). High resolution three-dimensional numerical modelling of rockfalls. *International Journal of Rock Mechanics and Mining Sciences*, 40(4), pp. 455–471.

Agliardi, F., Crosta, G. B. and Frattini, P. (2009). Integrating rockfall risk assessment and counter measure design by 3D modelling techniques. *Natural Hazards and Earth System Sciences*, 9(4), pp. 1059–1073.

Aguilera, Q., Lombardo, L., Tanyas, H. and Lipani, A. (2022). On the prediction of landslide occurrences and sizes via Hierarchical Neural Networks. *Stochastic Environmental Research and Risk Assessment*, 36(8), pp. 2031–2048.

Alvioli, M., Guzzetti, F. and Rossi, M. (2014). Scaling properties of rainfall induced landslides predicted by a physically based model. *Geomorphology*, 213, pp. 38–47.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, *9*(11), 3975-3991.

Aristizábal, E., Vélez, J. I., Martínez, H. E., & Jaboyedoff, M. (2016). SHIA_Landslide: a distributed conceptual and physically based model to forecast the temporal and spatial occurrence of shallow landslides triggered by rainfall in tropical and mountainous basins. *Landslides, 13(3), 497-517.*

Atkinson, P. M. and Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. *Computers & Geosciences*, 24(4), pp. 373–385.

Azarafza, M., Azarafza, M., Akgün, H., Atkinson, P. M. and Derakhshani, R. (2021). Deep learning based landslide susceptibility mapping. *Scientific Reports*, 11(1), 24112.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R. and Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, 10(7), e0130140.

Baum, R. L., Savage, W. Z. and Godt, J. W. (2008). TRIGRS: a Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0. *US Geological Survey Open-File Report*, 2008-1159.

Beguería, S., Van Asch, T. W. and Malet, J. P. (2009). A GIS-based numerical model for simulating the kinematics of mud and debris flows over complex terrain. *Natural Hazards and Earth System Sciences*, 9(6), pp. 1897-1909.

Bell, R. and Glade, T. (2004). Quantitative risk analysis for landslides–Examples from Bíldudalur, NW-Iceland. *Natural Hazards and Earth System Sciences*, 4(1), pp. 117–131.

Benda, L. and Cundy, T. W. (1990). Predicting deposition of debris flows in mountain channels. *Canadian Geotechnical Journal*, 27(4), pp. 409–417.

Bicocchi, G., D'Ambrosio, M., Rossi, G., Rosi, A., Tacconi-Stefanelli, C., Segoni, S., Nocentini, M., Vannocci, P., Tofani, V., Casagli, N., & Catani, F. (2016). Shear strength and permeability in situ measures to improve landslide forecasting models: a case study in the Eastern Tuscany (Central Italy). *Landslides and engineered slopes. Experience, theory and practice*, 419-424.

Bischetti, G. B., Chiaradia, E. A., Epis, T. and Morlotti, E. (2009). Root cohesion of forest species in the Italian Alps. *Plant and Soil*, 324(1-2), pp. 71-89.

Boreggio, M., Bernard, M. and Gregoretti, C. (2022). Does the topographic data source truly influence the routing modelling of debris flows in a torrent catchment?. *Earth Surface Processes and Landforms*, 47(11), pp. 2683–2703.

Brabb, E. E. (1984). Innovative approaches to landslide hazard and risk mapping. In *International landslide symposium proceedings*, Toronto, Canada, volume 1, pp. 17–22.

Bregoli, F., Medina, V., & Bateman, A. (2018). TXT-tool 3.034-2.1: A Debris Flow Regional Fast Hazard Assessment Toolbox. In *Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools: Volume 2: Testing, Risk Management and Country Practices* (pp. 225-234). Cham: Springer International Publishing.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, pp. 5–32.

Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, *19*(23-32), 410.

Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest. *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5372–5375.

Brilli, N., Masi, E. B., Rossi, G. and Tofani, V. (2025). Slope stability modelling of shallow landslides at a regional scale. *Geoenvironmental Disasters*, 12, p. 18.

Bucci, F., Santangelo, M., Fongo, L., Alvioli, M., Cardinali, M., Melelli, L. and Marchesini, I. (2022). A new digital lithological map of Italy at the 1:100 000 scale for geomechanical modelling. *Earth System Science Data*, 14(9), pp. 4129–4151.

Budetta, P. D. L. C., De Luca, C., & Nappi, M. (2016). Quantitative rockfall risk assessment for an important road by means of the rockfall risk management (RO. MA.) method. *Bulletin of Engineering Geology and the Environment*, *75*(4), 1377-1397.

Bui, D. T., Tsangaratos, P., Nguyen, V. T., Van Liem, N. and Trinh, P. T. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *Catena*, 188, 104426.

Bulzinetti, M. A., Segoni, S., Pappafico, G. and Tofani, V. (2021). A tool for the automatic aggregation and validation of the results of physically based distributed slope stability models. *Water*, 13(17), 2313.

Bunce, C., Cruden, D. and Morgenstern, N. (1997). Assessment of the hazard from rock fall on a highway. *Canadian Geotechnical Journal*, 34(3), pp. 344–356.

Burylo, M., Hudek, C. and Rey, F. (2011). Soil reinforcement by the roots of six dominant species on eroded mountainous marly slopes (Southern Alps, France). *Catena*, 84(1-2), pp. 70-78.

Burton, A. and Bathurst, J. (1998). Physically based modelling of shallow landslide sediment yield at a catchment scale. *Environmental Geology*, 35, pp. 89–99.

Caine, N. (1980). The rainfall intensity-duration control of shallow landslides and debris flows. *Geografiska Annaler: Series A, Physical Geography*, 62(1/2), pp. 23–27.

Caleca, F. (2024). *Development of new methodologies for a multi-scale landslide quantitative risk assessment*. PhD Thesis, University of Florence.

Calvo, B. and Savi, F. (2009). A real-world application of Monte Carlo procedure for debris flow risk assessment. *Computers & Geosciences*, 35(5), pp. 967–977.

Campbell, R. H. (1975). Soil slips, debris flows, and rainstorms in the Santa Monica Mountains and vicinity, southern California. *US Geological Survey Professional Paper*, 851.

Cardinali, M., Reichenbach, P., Guzzetti, F., Ardizzone, F., Antonini, G., Galli, M. and Salvati, P. (2002). A geomorphological approach to the estimation of landslide hazards and risks in Umbria, Central Italy. *Natural Hazards and Earth System Sciences*, 2(1/2), pp. 57-72.

Carlà, T., Tofani, V., Lombardi, L., Raspini, F., Bianchini, S., Bertolo, D., Thuegaz, P. and Casagli, N. (2019). Combination of GNSS, satellite InSAR, and GBInSAR remote sensing monitoring to improve the understanding of a large landslide in high alpine environment. *Geomorphology*, 335, pp. 62–75.

Carmignani, L., Conti, P., Cornamusini, G., Meccheri, M., & Crescenti, U. (2004). The internal Northern Apennines, the northern Tyrrhenian sea and the Sardinia-Corsica block. *Geology of Italy. Special Volume, Italian Geological Society, IGC*, *32*, 59-77.

Carminati, E. and Doglioni, C. (2012). Alps vs. Apennines: The paradigm of a tectonically asymmetric Earth. *Earth-Science Reviews*, 112(1–2), pp. 67–96.

Carrara, A. (1983). Multivariate models for landslide hazard evaluation. *Mathematical Geology*, 15, pp. 403–426.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V. and Reichenbach, P. (1991). GIS techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms*, 16(5), pp. 427-445.

Casagli, N., Intrieri, E., Tofani, V., Gigli, G. and Raspini, F. (2023). Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment*, 4(1), pp. 51–64.

Catani, F., Segoni, S. and Falorni, G. (2010). An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. *Water Resources Research*, 46(5).

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13(11), pp. 2815–2831.

Catani, F., Tofani, V. and Lagomarsino, D. (2016). Spatial patterns of landslide dimension: a tool for magnitude mapping. *Geomorphology*, 273, pp. 361–373.

Centamore, E., Chiocchini, U., Cipriani, N., Deiana, G. and Micarelli, A. (1979). The minor basins in the context of the Umbro-Marchean region tectonic-sedimentary evolution during Middle-Upper Miocene. In *Ann. Geol. Pays Hellen, Tome hors serie*, pp. 247–251.

Chen, M., Tang, C., Zhang, X., Xiong, J., Chang, M., Shi, Q., Wang, F. and Li, M. (2021). Quantitative assessment of physical fragility of buildings to the debris flow on 20 August 2019 in the Cutou gully, Wenchuan, southwestern China. *Engineering Geology*, 293, 106319.

Christen, M., Kowalski, J. and Bartelt, P. (2010). RAMMS: Numerical simulation of dense snow avalanches in three-dimensional terrain. *Cold Regions Science and Technology*, 63(1-2), pp. 1-14.

Ciurean, R. L., Hussin, H., van Westen, C. J., Jaboyedoff, M., Nicolet, P., Chen, L., Frigerio, S. and Glade, T. (2017). Multi-scale debris flow vulnerability assessment and direct loss estimation of buildings in the Eastern Italian Alps. Natural Hazards, 85(2), pp. 929–957.

Coe, J., Michael, J., Crovelli, R. and Savage, W. (2000). Preliminary map showing landslide densities, mean recurrence intervals, and exceedance probabilities as determined from historic records, Seattle, Washington. *Open-File Report*.

Coltori, M. (1997). Human impact in the Holocene fluvial and coastal evolution of the Marche region, Central Italy. *Catena*, 30(4), pp. 311–335.

Confuorto, P., Franceschini, R., Scarpitta, L., Casagli, N., Morelli, S., Raspini, F., Tofani, V. and Moretti, S. (2025). Event-based landslide inventory through very high-resolution optical images and field surveys. *Geoenvironmental Disasters*, 12(1), 23.

Conti, P., Cornamusini, G. and Carmignani, L. (2020). An outline of the geology of the Northern Apennines (Italy), with geological map at 1:250,000 scale. *Italian Journal of Geosciences*, 139(2), pp. 149–194.

Copernicus Land Monitoring Service and European Environment Agency (2020). *CORINE Land Cover 2018 (CLC2018)*. Version 2020_20u1.

Corominas, J. (1996). The angle of reach as a mobility index for small and large landslides. *Canadian Geotechnical Journal*, 33(2), pp. 260–271.

Corominas, J. and Moya, J. (2008). A review of assessing landslide frequency for hazard zoning purposes. *Engineering Geology*, 102(3), pp. 193–213.

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.

Corominas, J., Matas, G., & Ruiz-Carulla, R. (2019). Quantitative analysis of risk from fragmental rockfalls. *Landslides*, *16*(1), 5-21.

Corominas, J., Guzzetti, F., Lan, H., Macciotta, R., Marunteranu, C., McDougall, S. and Strom, A. (2023). Revisiting landslide risk terms: IAEG commission C-37 working group on landslide risk nomenclature. *Bulletin of Engineering Geology and the Environment*, 82(12), 450.

Crosta, G. (1998). Regionalization of rainfall thresholds: an aid to landslide hazard evaluation. *Environmental Geology*, *35*(2), 131-145.

Crosta, G. and Agliardi, F. (2003). A methodology for physically based rockfall hazard assessment. *Natural Hazards and Earth System Sciences*, 3(5), pp. 407–422.

Crosta, G. B., Imposimato, S., & Roddeman, D. G. (2003). Numerical modelling of large landslides stability and runout. *Natural Hazards and Earth System Sciences*, *3*(6), 523-538.

Dahal, A., & Lombardo, L. (2025). Towards physics-informed neural networks for landslide prediction. *Engineering Geology*, *344*, 107852.

Dai, F., Lee, C. F. and Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, 64(1), pp. 65–87.

Davis, J. and Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 233–240.

De Bono, A. and Mora, M. G. (2014). A global exposure model for disaster risk assessment. *International Journal of Disaster Risk Reduction*, 10, pp. 442–451.

Del Soldato, M., Solari, L., Poggi, F., Raspini, F., Tomás, R., Fanti, R. and Casagli, N. (2019). Landslide-induced damage probability estimation coupling InSAR and field survey data by fragility curves. *Remote Sensing*, 11(12), 1486.

Der Kiureghian, A. and Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?. *Structural Safety*, 31(2), pp. 105-112.

Di Napoli, M., Tanyas, H., Castro-Camilo, D., Calcaterra, D., Cevasco, A., Di Martire, D., Pepe, G., Brandolini, P. and Lombardo, L. (2023). On the estimation of landslide intensity, hazard and density via data-driven models. *Natural Hazards*, 119(3), pp. 1513–1530.

Dille, A., Dewitte, O., Handwerger, A. L., d'Oreye, N., Derauw, D., Ganza Bamulezi, G., Ilombe Mawe, G., Michellier, C. et al. (2022). Acceleration of a large deep-seated tropical landslide due to urbanization feedbacks. *Nature Geoscience*, 15(12), pp. 1048–1055.

Domènech, G., Fan, X., Scaringi, G., van Asch, T. W., Xu, Q., Huang, R. and Hales, T. C. (2019). Modelling the role of material depletion, grain coarsening and revegetation in debris flow occurrences after the 2008 Wenchuan earthquake. *Engineering Geology*, 250, pp. 34–44.

Donnini, M., Santangelo, M., Gariano, S. L., Bucci, F., Peruccacci, S., Alvioli, M. et al. (2023). Landslides triggered by an extraordinary rainfall event in Central Italy on September 15, 2022. *Landslides*, 20, pp. 2199–2211.

Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A. et al. (2014). HyMeX-SOP1: The Field Campaign Dedicated to Heavy Precipitation and Flash Flooding in the Northwestern Mediterranean. *Bulletin of the American Meteorological Society*, 95(7), pp. 1083–1100.

Duffourg, F. and Ducrocq, V. (2011). Origin of the moisture feeding the Heavy Precipitating Systems over Southeastern France. *Natural Hazards and Earth System Sciences*, 11(4), pp. 1163–1178.

Eidsvig, U. M. K., Papathoma-Köhle, M., Du, J., Glade, T. and Vangelsten, B. V. (2014). Quantification of model uncertainty in debris flow vulnerability assessment. Engineering Geology, 181, pp. 15–26.

Ellen, S. D., Algus, M. A., Cannon, S. H., Fleming, R. W., Lahr, P. C., Peterson, D. M., & Reneau, S. L. (1988). Description and mechanics of soil slip/debris flows in the storm. *US Geological Survey, Professional Paper*, *1434*, 63-111.

Emberson, R., Kirschbaum, D. and Stanley, T. (2020). New global characterisation of landslide exposure. *Natural Hazards and Earth System Sciences*, 20(12), pp. 3413–3424.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), pp. 861–874.

Fell, R. (1994). Landslide risk assessment and acceptable risk. *Canadian Geotechnical Journal*, 31(2), pp. 261–272.

Fell, R., Ho, K. K., Lacasse, S. and Leroi, E. (2005). A framework for landslide risk assessment and management. *Landslide Risk Management*, pp. 13–36.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W. Z. et al. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102(3-4), pp. 85–98.

FEMA (2023). *Benefit-Cost Analysis Guide*. Washington, DC: Federal Emergency Management Agency.

Ferlisi, S., Cascini, L., Corominas, J. and Matano, F. (2012). Rockfall risk assessment to persons travelling in vehicles along a road: the case study of the Amalfi coastal road (southern Italy). *Natural Hazards*, 62, pp. 691–721.

Fidan, S., Tanyaş, H., Akbaş, A., Lombardo, L., Petley, D. N. and Görüm, T. (2024). Understanding fatal landslides at global scales: a summary of topographic, climatic, and anthropogenic perspectives. *Natural Hazards*, 120(7), pp. 6437–6455.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), pp. 1189–1232.

Froude, M. J. and Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18(8), pp. 2161–2181.

Fuchs, S., Heiss, K. and Hübl, J. (2007). Towards an empirical vulnerability function for use in debris flow risk assessment. *Natural Hazards and Earth System Sciences*, 7(5), pp. 495–506.

Gaidzik, K. and Ramírez-Herrera, M. T. (2021). The importance of input data on landslide susceptibility mapping. *Scientific Reports*, 11(1), 19334.

Galli, M. and Guzzetti, F. (2007). Landslide vulnerability criteria: a case study from Umbria, Central Italy. *Environmental Management*, 40(4), pp. 649–665.

Gariano, S. L. and Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, 162, pp. 227–252.

George, D. L., & Iverson, R. M. (2014). A depth-averaged debris-flow model that includes the effects of evolving dilatancy. II. Numerical predictions and experimental tests. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 470*(2170), 20130820.

Giadrossich, F., Preti, F., Guastini, E., & Vannocci, P. (2010). Metodologie sperimentali per l'esecuzione di provedi taglio diretto su terre rinforzate con radici Experimental methodologies for the direct shear testson soils reinforced by roots. *GEOLOGIA TECNICA & AMBIENTALE*, *4*, 5-12.

Gili, J. A., Corominas, J. and Rius, J. (2000). Using Global Positioning System techniques in landslide monitoring. *Engineering Geology*, 55(3), pp. 167–192.

Gioia, E., Speranza, G., Ferretti, M., Godt, J. W., Baum, R. L., & Marincioni, F. (2016). Application of a process-based shallow landslide hazard model over a broad area in Central Italy. *Landslides*, *13*(5), 1197-1214.

Glade, T. (2003). Vulnerability assessment in landslide risk analysis. *Erde* 134(2), 123–146.

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

Goldstein, A., Kapelner, A., Bleich, J. and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), pp. 44-65.

Gorr, A. N., McGuire, L. A., Youberg, A. M. and Rengers, F. K. (2022). A progressive flow-routing model for rapid assessment of debris-flow inundation. *Landslides*, 19(9), pp. 2055-2073.

Gray, D. H. and Leiser, A. T. (1982). *Biotechnical slope protection and erosion control*. Van Nostrand Reinhold Company Inc.

Gray, D. H. and Ohashi, H. (1983). Mechanics of fiber reinforcement in sand. *Journal of Geotechnical Engineering*, 109(3), pp. 335-353.

Green, W. H. and Ampt, G. A. (1911). Studies on soil physics. Part I. The flow of air and water through soils. The Journal of Agricultural Science, 4(1), pp. 1–24.

Greenpeace Italy (2024). *Quanto costa all'Italia la crisi climatica?*. Rome: Greenpeace Italy.

Grima, N., Edwards, D., Edwards, F., Petley, D. and Fisher, B. (2020). Landslides in the Andes: Forests can provide cost-effective landslide regulation services. *Science of the Total Environment*, 745, 141128.

Guillard-Gonçalves, C. and Zêzere, J. L. (2018). Combining social vulnerability and physical vulnerability to analyse landslide risk at the municipal scale. *Geosciences*, 8(8), 294.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4), pp. 185–188.

Guthrie, R. H. and Evans, S. G. (2004). Magnitude and frequency of landslides triggered by a storm event, Loughborough Inlet, British Columbia. *Natural Hazards and Earth System Sciences*, 4(3), pp. 475–483.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31(1-4), pp. 181-216.

Guzzetti, F., Malamud, B. D., Turcotte, D. L. and Reichenbach, P. (2002). Power-law correlations of landslide areas in central Italy. *Earth and Planetary Science Letters*, 195(3-4), pp. 169–183.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1-4), pp. 272–299.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics*, 98, pp. 239–267.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: an update. *Landslides*, *5*(1), 3-17.

Haque, U., Blum, P., Da Silva, P. F., Andersen, P., Pilz, J., Chalov, S. R., Malet, J.-P. et al. (2016). Fatal landslides in Europe. *Landslides*, 13, pp. 1545–1554.

Haque, U., Da Silva, P. F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A. et al. (2019). The human cost of global warming: Deadly landslides and their triggers (1995–2014). *Science of the Total Environment*, 682, pp. 673-684.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), pp. 1735–1780.

Horton, P., Jaboyedoff, M., Rudaz, B. E. A. and Zimmermann, M. (2013). Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale. *Natural Hazards and Earth System Sciences*, 13(4), pp. 869-885.

Hu, K. H., Cui, P., & Zhang, J. Q. (2012). Characteristics of damage to buildings by debris flows on 7 August 2010 in Zhouqu, Western China. *Natural Hazards and Earth System Sciences*, *12*(7), 2209-2217.

Huang, F., Yin, K., Huang, J., Gui, L. and Wang, P. (2017). Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. *Engineering Geology*, 223, pp. 11–22.

Huang, Y. and Zhao, L. (2018). Review on landslide susceptibility mapping using support vector machines. *Catena,* 165, pp. 520–529.

Hungr, O. (1997) Some methods of landslide hazard intensity mapping. In *Landslide risk assessment*, eds D. Cruden and R. Fell, pp. 215–226. Routledge

Hungr, O., McDougall, S., Wise, M. and Cullen, M. (2008). Magnitude–frequency relationships of debris flows and debris avalanches in relation to slope relief. *Geomorphology*, 101(1-2), pp. 63–70.

Hungr, O. and McDougall, S. (2009). Two numerical models for landslide dynamic analysis. *Computers & Geosciences*, 35(5), pp. 978–992.

Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, *11*(2), 167-194.

Hürlimann, M., Rickenmann, D. and Graf, C. (2003). Field and monitoring data of debris-flow events in the Swiss Alps. *Canadian Geotechnical Journal*, 40(1), pp. 161-175.

Hürlimann, M., Rickenmann, D., Medina, V. and Bateman, A. (2008). Evaluation of approaches to calculate debris-flow parameters for hazard assessment. *Engineering Geology*, 102(3-4), pp. 152-163.

Hussin, H. Y., Quan Luna, B., van Westen, C. J., Christen, M., Malet, J. P. and Van Asch, T. W. (2012). Parameterization of a numerical 2-D debris flow model with entrainment: a case study of the Faucon catchment, Southern French Alps. *Natural Hazards and Earth System Sciences*, 12(10), pp. 3075-3090.

IPCC (2023). *Climate Change 2023: Synthesis Report*. Geneva: IPCC.

Iverson, R. M., & George, D. L. (2014). A depth-averaged debris-flow model that includes the effects of evolving dilatancy. I. Physical basis. *Proceedings of the royal society a: mathematical, Physical and Engineering Sciences*, *470*(2170), 20130819.

Ivanov, V. Y., Bras, R. L. and Vivoni, E. R. (2008). Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks. *Water Resources Research*, 44(3).

Jaboyedoff, M., Dudt, J. and Labiouse, V. (2005). An attempt to refine rockfall hazard zoning based on the kinetic energy, frequency and fragmentation degree. *Natural Hazards and Earth System Sciences*, 5(5), pp. 621–632.

Jaiswal, P., van Westen, C. J. and Jetten, V. (2010). Quantitative assessment of direct and indirect landslide risk along transportation lines in southern India. *Natural Hazards and Earth System Sciences*, 10(6), pp. 1253–1267.

Johnston, E. C., Davenport, F. V., Wang, L., Caers, J. K., Muthukrishnan, S., Burke, M. and Diffenbaugh, N. S. (2021). Quantifying the effect of precipitation on landslide hazard in urbanized and non-urbanized areas. *Geophysical Research Letters*, 48(16), e2021GL094038.

Jumadi, Carver, S. and Quincey, D. (2017). A Conceptual Design of Spatio-Temporal Agent-Based Model for Volcanic Evacuation. *Systems*, 5(4), 53.

Jumadi, J., Malleson, N., Carver, S. and Quincey, D. (2020). Estimating Spatio-Temporal Risks from Volcanic Eruptions using an Agent-Based Model. *Journal of Artificial Societies and Social Simulation*, 23(2).

Kang, H. S., & Kim, Y. T. (2016). The physical vulnerability of different types of building structure to debris flow events. *Natural Hazards*, *80*(3), 1475-1493.

Karsli, F. E. Z. İ., Atasoy, M., Yalcin, A., Reis, S., Demir, O. Ğ. U. Z. H. A. N., & Gokceoglu, C. (2009). Effects of land-use changes on landslides in a landslide-prone area (Ardesen, Rize, NE Turkey). *Environmental monitoring and assessment*, *156*(1), 241-255.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Kendon, E. J., Fischer, E. M. and Short, C. J. (2023). Variability conceals emerging trend in 100yr projections of UK local hourly rainfall extremes. *Nature Communications*, 14, 1133.

Knevels, R., Petschko, H., Proske, H., Leopold, P., Mishra, A. N., Maraun, D. and Brenning, A. (2023). Assessing uncertainties in landslide susceptibility predictions in a changing environment (Styrian Basin, Austria). *Natural Hazards and Earth System Sciences*, 23, pp. 205–229.

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N. et al. (2022). The challenge of unprecedented floods and droughts in risk management. *Nature*, 608(7921), pp. 80-86.

Lambert, C., Thoeni, K., Giacomini, A., Casagrande, D. and Sloan, S. (2012). Rockfall hazard analysis from discrete fracture network modelling with finite persistence discontinuities. *Rock Mechanics and Rock Engineering*, 45(5), pp. 871-884.

Lari, S., Frattini, P. and Crosta, G. (2014). A probabilistic approach for landslide hazard analysis. *Engineering Geology*, 182, pp. 3–14.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp. 2278–2324.

Lehmann, P. and Or, D. (2012). Hydromechanical triggering of landslides: From progressive local failures to mass release. *Water Resources Research*, 48(3).

Leone, F., Asté, J. P. and Leroi, E. (1996). Vulnerability assessment of elements exposed to mass-movement: working toward a better risk perception. In: Senneset, K. (ed.) Landslides / Glissements de terrain: Proceedings of the 7th International Symposium on Landslides, Trondheim, Norway. Rotterdam: A. A. Balkema, Vol. 1, pp. 263–269.

Li, Z., Nadim, F., Huang, H., Uzielli, M. and Lacasse, S. (2010). Quantitative vulnerability estimation for scenario-based landslide hazards. *Landslides*, 7, pp. 125–134.

Li, X., Wu, Y., He, S., & Su, L. (2016). Application of the material point method to simulate the post-failure runout processes of the Wangjiayan landslide. *Engineering Geology*, *212*, 1-9.

Lombardo, L., Tanyas, H., Huser, R., Guzzetti, F. and Castro-Camilo, D. (2021). Landslide size matters: A new data-driven, spatial prototype. *Engineering Geology*, 293, 106288.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Luo, H. Y., Zhang, L. M., Zhang, L. L., He, J. and Yin, K. S. (2023). Vulnerability of buildings to landslides: The state of the art and future needs. Earth-Science Reviews, 238, 104329.

Maes, J., Kervyn, M., de Hontheim, A., Dewitte, O., Jacobs, L., Mertens, K. et al. (2017). Landslide risk reduction measures: A review of practices and challenges for the tropics. *Progress in Physical Geography*, 41(2), pp. 191–221.

Malamud, B. D., Turcotte, D. L., Guzzetti, F. and Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, 29(6), pp. 687–711.

Marche Region Functional Centre (2014). *Event Report: 2–4 May 2014*. Ancona: Regione Marche – Centro Funzionale.

Marche Region Functional Centre (2022a). *Climatic and hydrological analysis, September 2021 – July 2022*. Ancona: Regione Marche – Centro Funzionale.

Marche Region Functional Centre (2022b). *Preliminary Event Report: 15–17 September 2022*. Ancona: Regione Marche – Centro Funzionale.

Marengo, J. A., Alves, L. M., Ambrizzi, T., Young, A., Barreto, N. J. and Ramos, A. M. (2020). Trends in extreme rainfall and hydrogeometeorological disasters in the Metropolitan Area of São Paulo: a review. *Annals of the New York Academy of Sciences*, 1472(1), pp. 5–20.

Marino, P., Peres, D. J., Cancelliere, A., Greco, R. and Bogaard, T. A. (2020). Soil moisture information can improve shallow landslide forecasting using the hydrometeorological threshold approach. *Landslides*, 17(9), pp. 2041–2054.

Masi, E. B., Tofani, V., Rossi, G., Cuomo, S., Wu, W., Salciarini, D., Caporali, E. and Catani, F. (2023). Effects of roots cohesion on regional distributed slope stability modelling. *Catena*, 222, 106853.

McCann, D. W. (1983). The enhanced-V: A satellite observable severe storm signature. *Monthly Weather Review*, *111*(4), 887-894.

McDougall, S. and Hungr, O. (2004). A model for the analysis of rapid landslide motion across three-dimensional terrain. *Canadian Geotechnical Journal*, 41(6), pp. 1084-1097.

Mead, S. R., Procter, J. and Kereszturi, G. (2021). Quantifying location error to define uncertainty in volcanic mass flow hazard simulations. Natural Hazards and Earth System Sciences, 21, pp. 2447–2460.

Medina, V., Hürlimann, M., Guo, Z., Lloret, A. and Vaunat, J. (2021). Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *Catena, 201, 105213.*

Mergili, M., Marchesini, I., Rossi, M., Guzzetti, F. and Fellin, W. (2014). Spatially distributed three-dimensional slope stability modelling in a raster GIS. *Geomorphology*, 206, pp. 178–195.

Mergili, M., Fischer, J. T., Krenn, J. and Pudasaini, S. P. (2017). r.avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows. *Geoscientific Model Development*, 10(2), pp. 553-569.

Metropolis, N. and Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), pp. 335–341.

Michael-Leiba, M., Baynes, F., Scott, G. and Granger, K. (2005). Quantitative landslide risk assessment of Cairns, Australia. *Landslide Hazard and Risk*, pp. 621–642.

Milledge, D. G., Bellugi, D., McKean, J. A., Densmore, A. L. and Dietrich, W. E. (2014). A multidimensional stability model for predicting shallow landslide size and shape across landscapes. *Journal of Geophysical Research: Earth Surface*, 119(11), pp. 2481–2504.

Montgomery, D. R. and Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. *Water Resources Research*, 30(4), pp. 1153–1171.

Montrasio, L. and Valentino, R. (2008). A model for triggering mechanisms of shallow landslides. *Natural Hazards and Earth System Sciences*, 8, pp. 1149–1159.

Moos, C., Dorren, L. and Stoffel, M. (2017). Quantifying the effect of forests on frequency and intensity of rockfalls. *Natural Hazards and Earth System Sciences*, 17(2), pp. 291–304.

Moreno, M., Steger, S., Tanyas, H. and Lombardo, L. (2023). Modeling the area of co-seismic landslides via data-driven models: The Kaikōura example. *Engineering Geology*, 320, 107121.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), pp. 161–174.

Niculescu-Mizil, A. and Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 625–632.

NOAA (2023). *Billion-Dollar Weather and Climate Disasters*. National Oceanic and Atmospheric Administration.

Nocentini, N., Segoni, S., Rosi, A. and Fanti, R. (2025). Double-threshold validation tool (DTVT): From landslide hazard maps to operational early warning systems. *International Journal of Disaster Risk Reduction*, 129, 105786.

Nor Diana, M. I., Muhamad, N., Taha, M. R., Osman, A. and Alam, M. M. (2021). Social vulnerability assessment for landslide hazards in Malaysia: A systematic review study. *Land*, 10(3), 315.

O'Brien, J. S., Julien, P. Y. and Fullerton, W. T. (1993). Two-dimensional water flood and mudflow simulation. *Journal of Hydraulic Engineering*, 119(2), pp. 244-261.

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91(10–11), pp. 1290–1300.

Operstein, V. and Frydman, S. (2000). The influence of vegetation on soil strength. *Proceedings of the Institution of Civil Engineers-Ground Improvement*, 4(2), pp. 81-89.

Ozturk, U., Tarakegn, Y. A., Longoni, L., Brambilla, D., Papini, M., & Jensen, J. (2016). A simplified early-warning system for imminent landslide prediction based on failure index fragility curves developed through numerical analysis. *Geomatics, Natural Hazards and Risk*, *7*(4), 1406-1425.

Ozturk, U., Bozzolan, E., Holcombe, E. A., Shukla, R., Pianosi, F. and Wagener, T. (2022). How climate change and unplanned urban sprawl bring more landslides. *Nature*, 608(7922), pp. 262–265.

Pacheco Quevedo, R., Velastegui-Montoya, A., Montalván-Burbano, N., Morante-Carballo, F., Korup, O. and Daleles Rennó, C. (2023). Land use and land cover as a conditioning factor in landslide susceptibility: a literature review. *Landslides*, 20(5), pp. 967–982.

Pack, R. T., Tarboton, D. G. and Goodwin, C. N. (1998). The SINMAP approach to terrain stability mapping. In: Moore, D. and Hungr, O. (eds.) Proceedings of the 8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada, 21–25 September 1998. Rotterdam: A. A. Balkema, Vol. 2, pp. 1157–1166.

Papathoma-Koehle, M., Keiler, M., Totschnig, R. and Glade, T. (2012). Improvement of vulnerability curves using data from extreme events: debris flow event in South Tyrol. *Natural Hazards*, 64, pp. 2083–2105.

Parise, M. and Wasowski, J. (1999). Landslide activity maps for landslide hazard evaluation: three case studies from Southern Italy. *Natural Hazards*, 20(2), pp. 159–183.

Pastor, M., Blanc, T., Haddad, B., Drempetic, V., Morles, M. S., Dutto, P., ... & Merodo, J. F. (2015). Depth averaged models for fast landslide propagation: mathematical, rheological and numerical aspects. *Archives of Computational Methods in Engineering*, *22*(1), 67-104.

Patra, A. K., Bauer, A. C., Nichita, C. C., Pitman, E. B., Sheridan, M. F. et al. (2005). Parallel adaptive numerical simulation of dry avalanches over natural terrain. *Journal of Volcanology and Geothermal Research*, 139(1-2), pp. 1-21.

Peduto, D., Ferlisi, S., Nicodemo, G., Reale, D., Pisciotta, G. and Gullà, G. (2017). Empirical fragility and vulnerability curves for buildings exposed to slow-moving landslides at medium and large scales. *Landslides*, 14, pp. 1993–2007.

Pellicani, R., van Westen, C. J., & Spilotro, G. (2014). Assessing landslide exposure in areas with limited landslide information. *Landslides*, *11*(3), 463-480.

Peres, D. J. and Cancelliere, A. (2021). Comparing methods for determining landslide early warning thresholds: potential use of non-triggering rainfall for locations with scarce landslide data availability. *Landslides*, 18, pp. 3135–3147.

Perkins, N. J. and Schisterman, E. F. (2006). The Inconsistency of "Optimal" Cutpoints Obtained using Two Criteria based on the Receiver Operating Characteristic Curve. *American Journal of Epidemiology*, 163(7), pp. 670–675.

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M. and Guzzetti, F. (2017). Rainfall thresholds for possible landslide occurrence in Italy. *Geomorphology*, 290, pp. 39-57.

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, 40(10), pp. 927–930.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T. (2014). Assessing the quality of landslide susceptibility maps–case study Lower Austria. *Natural Hazards and Earth System Sciences*, 14(1), pp. 95–118.

Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., & Calvello, M. (2017). Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides*, *14*(3), 995-1008.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, *10*(3), 61-74.

Quan Luna, B., Blahut, J., van Westen, C. J., Sterlacchini, S., van Asch, T. W. J. and Akbas, S. O. (2011). The application of numerical debris flow modelling for the generation of physical vulnerability curves. Natural Hazards and Earth System Sciences, 11, pp. 2047–2060.

Quan Luna, B., Blahut, J., Camera, C., van Westen, C. J., Apuani, T., Jetten, V. and Sterlacchini, S. (2014). Physically based dynamic run-out modelling for quantitative debris flow risk assessment: a case study in Tresenda, northern Italy. Environmental Earth Sciences, 72(3), pp. 645–661.

Raetzo, H., Lateltin, O., Bollinger, D. and Tripet, J. (2002). Hazard assessment in Switzerland–Codes of Practice for mass movements. *Bulletin of Engineering Geology and the Environment*, 61, pp. 263–268.

Raia, S., Alvioli, M., Rossi, M., Baum, R. L., Godt, J. W. and Guzzetti, F. (2014). Improving predictive power of physically based rainfall-induced shallow landslide models: a probabilistic approach. *Geoscientific Model Development*, 7(2), pp. 495-514.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686-707.

Rasmussen, C. E. and Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press.

Raspini, F., Bianchini, S., Ciampalini, A., Del Soldato, M., Solari, L., Novali, F. et al. (2018). Continuous, semi-automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific Reports*, 8(1), 7253.

Rawls, W. J., Brakensiek, D. L. and Saxtonn, K. E. (1982). Estimation of soil water properties. *Transactions of the ASAE*, 25(5), pp. 1316-1320.

Regione Marche (2014). *Rapporto di Evento 2-4 maggio 2014*. Ancona: Dipartimento per le Politiche Integrate di Sicurezza e per la Protezione Civile.

Reichenbach, P., Galli, M., Cardinali, M., Guzzetti, F. and Ardizzone, F. (2005). Geomorphological mapping to assess landslide risk: Concepts, methods and applications in the Umbria region of central Italy. *Landslide Hazard and Risk*, pp. 429-468.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M. and Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, pp. 60–91.

Reid, L. and Page, M. (2003). Magnitude and frequency of landsliding in a large New Zealand catchment. *Geomorphology*, 49(1-2), pp. 71–88.

Reid, M. E., Christian, S. B., Brien, D. L. and Henderson, S. T. (2015). Scoops3D: software to analyze 3D slope stability throughout a digital landscape (No. 14-A1). *US Geological Survey*.

Remondo, J., Bonachea, J. and Cendrero, A. (2008). Quantitative landslide risk assessment and mapping on the basis of recent occurrences. *Geomorphology*, 94(3-4), pp. 496–507.

Revellino, P., Hungr, O., Guadagno, F. M. and Evans, S. G. (2004). Velocity and runout simulation of destructive debris flows and debris avalanches in pyroclastic deposits, Campania region, Italy. *Environmental Geology*, 45(3), pp. 295-311.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1, pp. 318-333.

Rickenmann, D., Laigle, D., McArdell, B. W. and Hübl, J. (2006). Comparison of 2D debris-flow simulation models with field events. *Computational Geosciences*, 10(2), pp. 241-264.

Rigon, R., Bertoldi, G. and Over, T. M. (2006). GEOtop: A distributed hydrological model with coupled water and energy budgets. *Journal of Hydrometeorology*, 7(3), pp. 371-388.

Rossi, G. (2010). *A physically based distributed slope stability simulator to analyze shallow landslides triggering in real time and on a large scale*. PhD thesis, University of Florence.

Rossi, G., Catani, F., Leoni, L., Segoni, S. and Tofani, V. (2013). HIRESSS: a physically based slope stability simulator for HPC applications. *Natural Hazards and Earth System Sciences*, 13(1), pp. 151-166.

Rougier, J., Sparks, S. and Hill, L. (eds.) (2013). *Risk and Uncertainty Assessment for Natural Hazards*. Cambridge: Cambridge University Press.

Ruff, M. and Czurda, K. (2008). Landslide susceptibility analysis with a heuristic approach in the Eastern Alps (Vorarlberg, Austria). *Geomorphology*, 94(3-4), pp. 314–324.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Saito, T. and Rehmsmeier, M. (2015). The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432.

Salvatici, T., Tofani, V., Rossi, G., D'Ambrosio, M., Stefanelli, C. T., Masi, E. B. et al. (2018). Application of a physically based model to forecast shallow landslides at a regional scale. *Natural Hazards and Earth System Sciences*, 18(7), pp. 1919-1935.

Savage, S. B. and Hutter, K. (1989). The motion of a finite mass of granular material down a rough incline. Journal of Fluid Mechanics, 199, pp. 177–215.

Scaini, C., Tamaro, A., Adilkhan, B., Sarzhanov, S., Ismailov, V., Umaraliev, R., Safarov, M., Belikov, V., Karayev, J. and Faga, E. (2024). A new regionally consistent exposure database for Central Asia: population and residential buildings. Natural Hazards and Earth System Sciences, 24, pp. 929–945.

Schilling, S. P. (1998). *LAHARZ;* GIS programs for automated mapping of lahar-inundation hazard zones (No. 98-638). US Geological Survey; Information Services [distributor],.

Schuster, R. L. and Fleming, R. W. (1986). Economic losses and fatalities due to landslides. *Bulletin of the Association of Engineering Geologists*, 23(1), pp. 11–28.

Segoni, S., Piciullo, L. and Gariano, S. L. (2018a). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, 15(8), pp. 1483-1501.

Segoni, S., Rosi, A., Lagomarsino, D., Fanti, R., & Casagli, N. (2018b). Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Natural Hazards and Earth System Sciences*, *18*(3), 807-812.

Singh, A., Kanungo, D. P., & Pal, S. (2019). A modified approach for semi-quantitative estimation of physical vulnerability of buildings exposed to different landslide intensity scenarios. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *13*(1), 66-81.

Skempton, A. and DeLory, F. (1957). Stability of natural slopes in London clay. *Proceedings of the 4th International Conference on Soil Mechanics and Foundation Engineering*, 2, pp. 378-381.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1–3), pp. 271–280.

Solari, L., Bianchini, S., Franceschini, R., Barra, A., Monserrat, O., Thuegaz, P., Bertolo, D., Crosetto, M. and Catani, F. (2020). Satellite interferometric data for landslide intensity evaluation in mountainous regions. *International Journal of Applied Earth Observation and Geoinformation*, 87, 102028.

Stanley, T. and Kirschbaum, D. B. (2017). A heuristic approach to global landslide susceptibility mapping. *Natural Hazards*, 87, pp. 145–164.

Stark, C. P. and Hovius, N. (2001). The characterization of landslide size distributions. *Geophysical Research Letters*, 28(6), pp. 1091–1094.

Stark, T. D. and Choi, H. (2008). Slope inclinometers for landslides. *Landslides*, 5, pp. 339–350.

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

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

Stolz, A. and Huggel, C. (2008). Debris flows in the Swiss National Park: the influence of different flow models and varying DEM grid size on modeling results. *Landslides*, 5(3), pp. 311–319.

Tanyaş, H., Allstadt, K. E. and van Westen, C. J. (2018). An updated method for estimating landslide-event magnitude. *Earth Surface Processes and Landforms*, 43(9), pp. 1836–1847.

Tarquini, S., Isola, I., Favalli, M., Mazzarini, F., Bisson, M., Pareschi, M. T. and Boschi, E. (2007). TINITALY/01: a new triangular irregular network of Italy. *Annals of Geophysics*.

Thiessen, A. H. (1911). Precipitation averages for large areas. *Monthly weather review*, *39*(7), 1082-1089.

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*(2), 361-378.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), pp. 234–240.

Tofani, V., Ventisette, C. D., Moretti, S. and Casagli, N. (2014). Integration of remote sensing techniques for intensity zonation within a landslide area: A case study in the northern Apennines, Italy. *Remote Sensing*, 6(2), pp. 907-924.

Tofani, V., Bicocchi, G., Rossi, G., Segoni, S., D'Ambrosio, M., Casagli, N. and Catani, F. (2017). Soil characterization for shallow landslides modeling: a case study in the Northern Apennines (Central Italy). *Landslides*, 14(2), pp. 755-770.

Torcasio, R. C., Papa, M., Del Frate, F., Dietrich, S., Toffah, F. E. and Federico, S. (2023). Study of the Intense Meteorological Event Occurred in September 2022 over the Marche Region with WRF Model: Impact of Lightning Data Assimilation on Rainfall and Lightning Prediction. *Atmosphere*, 14, 1152.

Totschnig, R. and Fuchs, S. (2013). Mountain torrents: quantifying vulnerability and assessing uncertainties. *Engineering Geology*, 155, pp. 31–44.

Trigila, A., Iadanza, C., Berti, D. and Lucarini, M. (2015). *Principali eventi franosi e alluvionali in Italia nel periodo 2003-2014*. Rome: ISPRA, Technical Report RT/SUO-IST 002/2015.

Trigila, A., Iadanza, C., Lastoria, B., Bussettini, M. and Barbano, A. (2021). *Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio - Edizione 2021*. Rome: ISPRA, Reports 356/2021.

Trigila, A., Lastoria, B., Iadanza, C., Bussettini, M., Mariani, S., D'Ascola, F. et al. (2025). *Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio - Edizione 2024*. ISPRA, Rapporti 415/2025.

Troncone, A., Pugliese, L., & Conte, E. (2020). Run-out simulation of a landslide triggered by an increase in the groundwater level using the material point method. *Water*, *12*(10), 2817.

Ullah, I., Chen, Z., Hussain, M. A., Shah, S. U. and Ali, N. (2025). Hybrid Machine Learning and SBAS-InSAR Integration for Landslide Susceptibility Mapping Along the Balakot–Naran Route, Pakistan. *Remote Sensing*, 17(20), 3464.

UNDRR (2015). *Sendai Framework for Disaster Risk Reduction 2015-2030*. United Nations Office for Disaster Risk Reduction.

UNDRR (2019). *Global Assessment Report on Disaster Risk Reduction 2019*. United Nations Office for Disaster Risk Reduction.

UNISDR (2009). *Terminologia sobre reducción del riesgo de desastres*. Estrategia Internacional para la Reducción de Desastres de las Naciones Unidas.

USACE-HEC (2021–2024). *HEC-RAS River Analysis System, 2D Modeling User's Manual*. Versions 6.0-6.4. Hydrologic Engineering Center, US Army Corps of Engineers.

Uzielli, M., Nadim, F., Lacasse, S. and Kaynia, A. M. (2008). A conceptual framework for quantitative estimation of physical vulnerability to landslides. *Engineering Geology*, 102(3-4), pp. 251–256.

Uzielli, M., Lacasse, S. and Nadim, F. (2009). Probabilistic risk estimation for geohazards: a simulation approach. In: Honjo, Y., Suzuki, M., Hara, T. and Zhang, F. (eds.) Geotechnical Risk and Safety: Proceedings of the 2nd International Symposium on Geotechnical Safety and Risk, Gifu, Japan, 11–12 June 2009. London: CRC Press, pp. 355–362.

Vai, G. B. and Martini, I. P. (eds.) (2001). *Anatomy of an Orogen: The Apennines and Adjacent Mediterranean Basins*. Dordrecht: Kluwer Academic Publishers.

van den Bout, B., Tang, C., van Westen, C. and Jetten, V. (2022). Physically-based modelling of co-seismic landslide, debris flow and flood cascade. *Natural Hazards and Earth System Sciences Discussions*, pp. 1-56.

van Westen, C., Soeters, R. and Sijmons, K. (2000). Digital geomorphological landslide hazard mapping of the Alpago area, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 2(1), pp. 51–60.

van Westen, C. J., Van Asch, T. W., & Soeters, R. (2006). Landslide hazard and risk zonation—why is it still so difficult?. *Bulletin of Engineering geology and the Environment*, *65*(2), 167-184.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.

Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory* (2nd Edition). New York: Springer-Verlag.

Varnes, D. J. (1984). Landslide hazard zonation. A review of principles and practice*. UNESCO Natural Hazard Series*.

Vergani, C., Bassanelli, C., Rossi, L., Chiaradia, E. A., & Bischetti, G. B. (2013). The effect of chestnut coppice forests abandon on slope stability: a case study.

Voellmy, A. (1955). Über die Zerstörungskraft von Lawinen. *Schweizerische Bauzeitung*, 73, pp. 212-285.

Waldron, L. J. (1977). The shear resistance of root-permeated homogeneous and stratified soil 1. *Soil Science Society of America Journal*, 41(5), pp. 843-849.

Waldron, L. J. and Dakessian, S. (1981). Soil reinforcement by roots: calculation of increased soil shear resistance from root properties. *Soil Science*, 132(6), pp. 427-435.

Wichmann, V. (2017). The Gravitational Process Path (GPP) model (v1.0) - a GIS-based simulation framework for gravitational processes. *Geoscientific Model Development*, 10(9), pp. 3309-3327.

Winter, M., Smith, J., Fotopoulou, S., Pitilakis, K., Mavrouli, O., Corominas, J. and Argyroudis, S. (2014). An expert judgement approach to determining the physical vulnerability of roads to debris flow. *Bulletin of Engineering Geology and the Environment*, 73, pp. 291–305.

WSL Institute for Snow and Avalanche Research SLF (2024). RAMMS::Debrisflow User Manual. Davos: WSL Institute for Snow and Avalanche Research SLF.

Wu, T. H., McKinnell III, W. P. and Swanston, D. N. (1979). Strength of tree roots and landslides on Prince of Wales Island, Alaska. *Canadian Geotechnical Journal*, 16(1), pp. 19-33.

Wu, C. G., Zhang, J., Chen, H. X., Wang, H. A., & Lu, D. G. (2025). Physically based probabilistic prediction of regional rainfall-induced shallow landslides: from initiation to runout analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *19*(2), 233-246.

Xiao, Y., Tang, X., Li, Y., Huang, H. and An, B.-W. (2022). Social vulnerability assessment of landslide disaster based on improved TOPSIS method: Case study of eleven small towns in China. *Ecological Indicators*, 143, 109316.

Yadav, R., Huser, R., Opitz, T. and Lombardo, L. (2023). Joint modelling of landslide counts and sizes using spatial marked point processes with sub-asymptotic mark distributions. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(5), pp. 1139–1161.

Zadrozny, B. and Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02)*, pp. 694–699.

Zanchetta, G., Sulpizio, R., Pareschi, M., Leoni, F. and Santacroce, R. (2004). Characteristics of May 5–6, 1998 volcaniclastic debris flows in the Sarno area (Campania, southern Italy): relationships to structural damage and hazard zonation. *Journal of Volcanology and Geothermal Research*, 133(1-4), pp. 377–393.

Zhang, J., Tang, W. H. and Zhang, L. M. (2010a). Efficient probabilistic back-analysis of slope stability model parameters. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(1), pp. 99–109.

Zhang, L. L., Zhang, J., Zhang, L. M. and Tang, W. H. (2010b). Back analysis of slope failure with Markov chain Monte Carlo simulation. *Computers and Geotechnics*, 37(7), pp. 905–912.

Zhang, S., Zhang, L., Li, X. and Xu, Q. (2018). Physical vulnerability models for assessing building damage by debris flows. *Engineering Geology*, 247, pp. 145–158.

Zhao, H. and Kowalski, J. (2022). Bayesian active learning for parameter calibration of landslide run-out models. *Landslides*, 19(8), pp. 2033–2045.

Zêzere, J., Pereira, S., Melo, R., Oliveira, S. and Garcia, R. A. (2017). Mapping landslide susceptibility using data-driven methods. *Science of the Total Environment*, 589, pp. 250–267.

Appendix A - SADRI Technical Reports

This appendix contains the technical reports automatically generated by the SADRI framework in PDF format for the two Slope Units selected in *Section 5.5.2*, to clarify the analysis and estimation process underlying the reported economic risk values. Given the high number of Slope Units in the study area, only these two reports are included:

- **A.1 Technical Report - Slope Unit 1300**: Arcevia municipality (September 2022 event case study);
- **A.2 Technical Report - Slope Unit 1309**: Montecarotto municipality (May 2014 high-criticality case study).

For consistency with National Civil Protection standards, the reports use the Italian decimal comma, while the rest of this thesis follows English numerical notation.

A.1 Technical Report: Slope Unit 1300 (Arcevia municipality - Scenario A)

TECHNICAL REPORT — SLOPE UNIT 1300

Organization: DST - UniFi
Authors: Dr. Nicolò Brilli
Version: SADRI - Preliminary Analysis V1.1

Run date/time: 2026-01-12 13:09:02
Contact: nicolo.brilli@unifi.it

Executive Summary

Deterministic baseline (no Monte Carlo)

> **Impacted assets:** 1
> **Attention-Zone assets:** 3
> **Total economic value (impacted assets):** € 71.309,67
> **Total risk (impacted):** € 46.376,87
> **Total economic value (attention zone):** € 98.307,25

> **Policy note:** attention-zone assets do *not* generate an economic risk estimate (value only).

Monte Carlo summary (probabilistic)

> **Impacted assets (after simulations):** 1
> **Impacted from initially Attention Zone:** 0
> **New impacted (MC-only):** 0
>
> ---
>
> **Total economic value of impacted assets (post-MC classification):** € 71.309,67
> **Average risk when impact occurs (hits only):** € 61.531,49
> **Risk variability when impact occurs (standard deviation, hits only):** € 17.836,34
> **Total economic value in attention zone (post-MC classification):** € 98.307,25

Legal / operational notes

> This report was generated automatically. Estimates are subject to uncertainty and must be validated by an expert technician.

Simulation parameters

Methodology / risk parameters

FPT threshold	0.55
Velocity threshold (noise) [m/s]	0.053
Attention buffer [m]	20.0
V_ref [m/s]	10.297504425048828

Simulation configuration

Expected CRS	EPSG:32632
Max workers	11
Monte Carlo	Active (N=100)
Random seed	42
Selected runout analysis model	DebrisDice
DebrisDice model_type	1
DebrisDice source_mode	2,1
DebrisDice iterations	100
DebrisDice range_angle	0.6
DebrisDice minimum_slope_angle	0.04
DebrisDice friction_factor	0.04
DebrisDice mass_to_drag	9.47
DebrisDice viscosity	500.0
DebrisDice shear_yield	500
DebrisDice exponent	2.5
DebrisDice density	2000
DebrisDice bing_depth	4
Monte Carlo iterations	100
Monte Carlo uncertainty_factor	0.2
Monte Carlo parameters_to_perturb	friction_factor, mass_to_drag, minimum_slope_angle, range_angle, shear_yield

Study area overview

Figure 1: Study area overview showing the Slope Unit under analysis (Slope Unit 1300).

Deterministic simulation map (baseline)

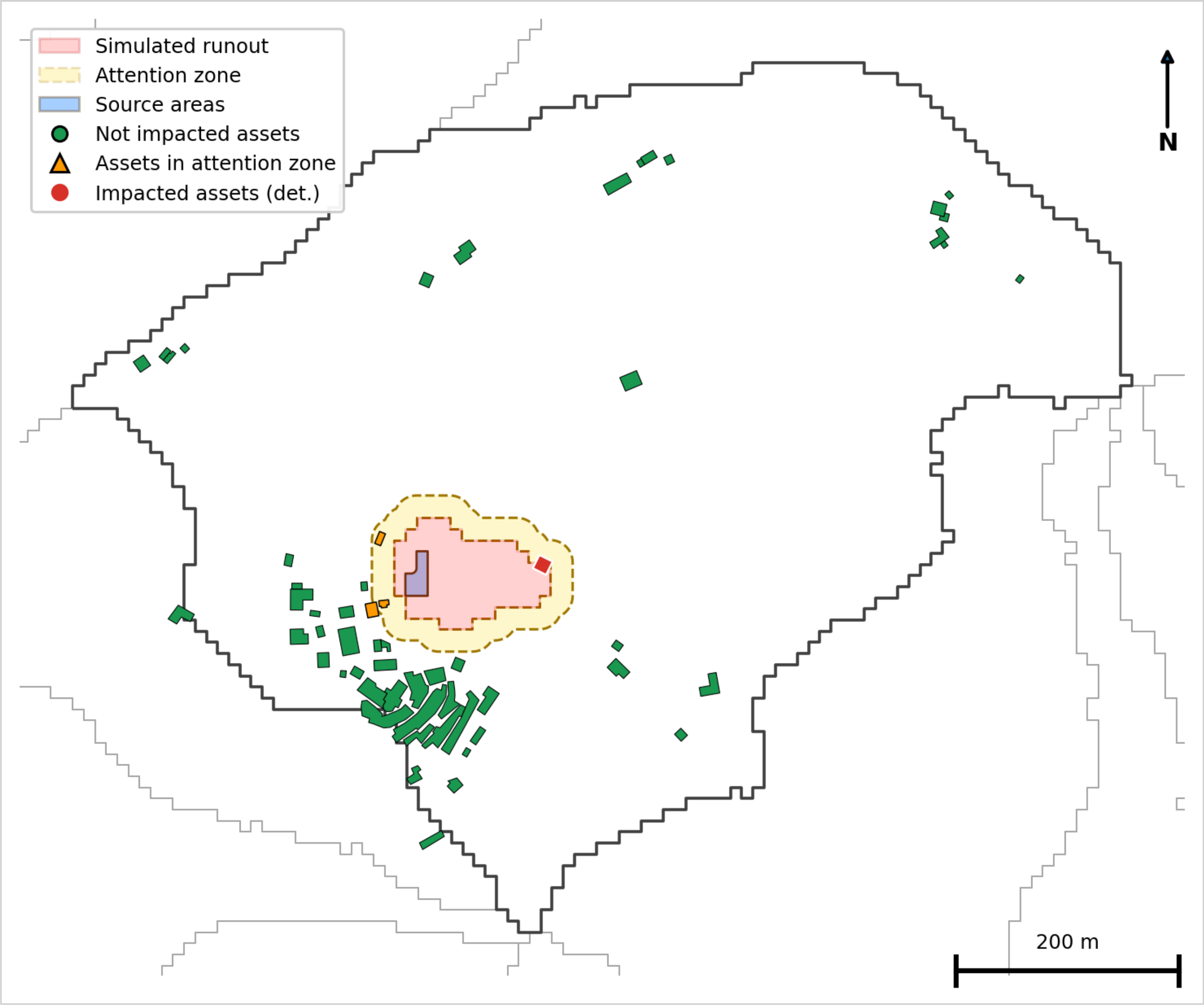


Figure 2: Deterministic runout map with impacted and attention-zone assets (baseline, no MC perturbation).

Monte Carlo probabilistic simulation map

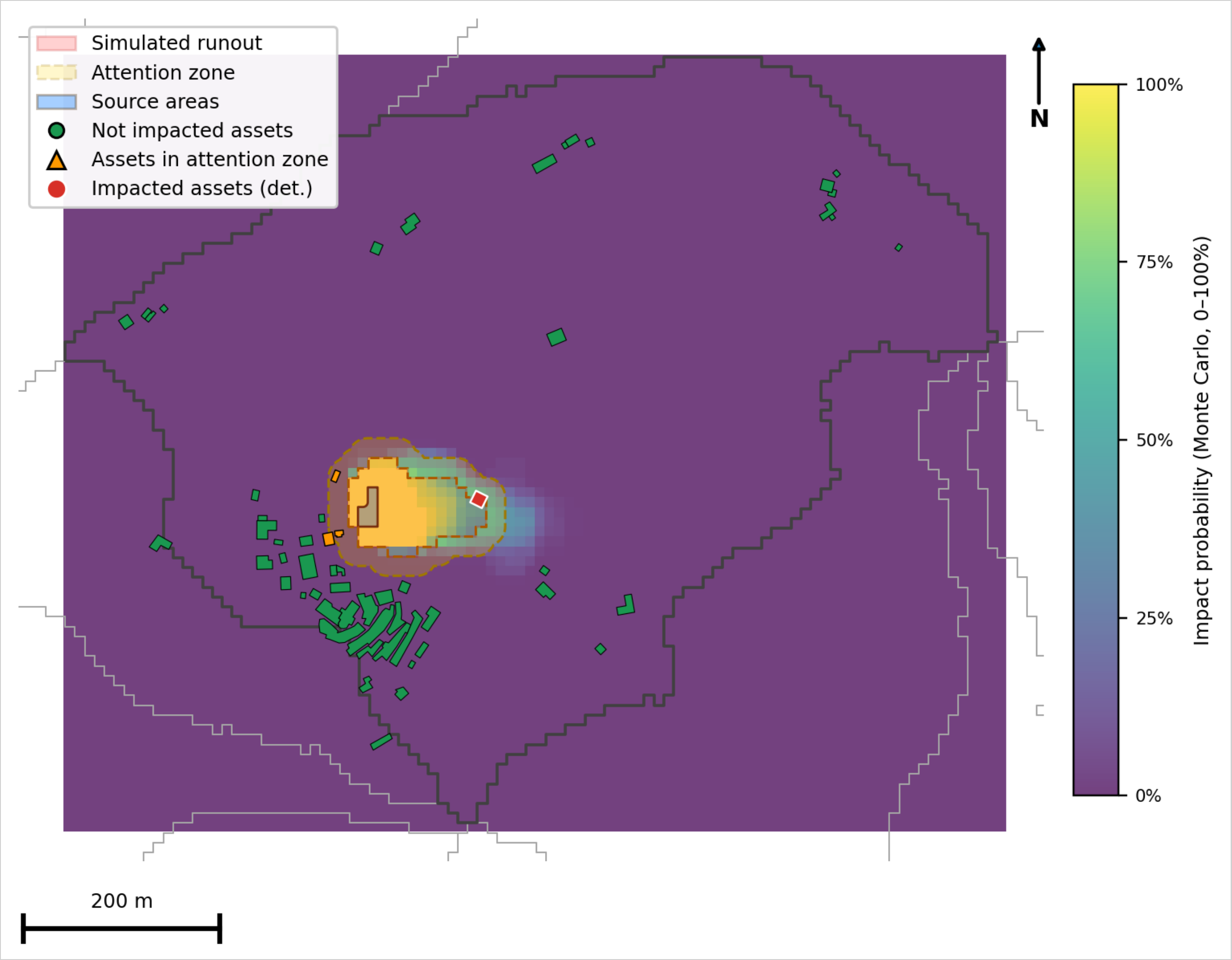


Figure 3: Monte Carlo probabilistic runout and impacted assets classified by impact probability (hit rate).

Municipality risk summary (post-MC classification)

Municipality	Risk (EUR)
Arcevia	€ 46.376,87

Asset tables — deterministic baseline (no MC)

Baseline deterministic impacted assets. No Monte Carlo frequencies are shown here.

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
215465	Residential Building	0.378	€ 46.376,87	Arcevia	residential

Baseline deterministic attention-zone assets (value only by policy; no risk).

Asset ID	Description	Total economic value (EUR)	Structural vulnerability	Municipality	type
215439	Residential Building	€ 51.300,09	0.395	Arcevia	residential
20716	Shed / Canopy	€ 20.520,73	0.395	Arcevia	other
215441	Residential Building	€ 26.486,43	0.395	Arcevia	residential

Asset tables — Monte Carlo (post-simulation classification)

> Impacted assets after MC: includes promoted from deterministic attention zone and MC-only / new impacted assets. “Impact probability” is the hit rate across MC iterations.

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
215465	Residential Building	0.378	€ 46.376,87	Arcevia	residential

> Attention-zone assets after MC: assets that were in deterministic attention zone and were never impacted in Monte Carlo.

Asset ID	Description	Total economic value (EUR)	Structural vulnerability	Municipality	type
215439	Residential Building	€ 51.300,09	0.395	Arcevia	residential
20716	Shed / Canopy	€ 20.520,73	0.395	Arcevia	other
215441	Residential Building	€ 26.486,43	0.395	Arcevia	residential

---

SADRI System — Debris-flow risk assessment report
Generated on 2026-01-12 13:09:02 —

A.2 Technical Report: Slope Unit 1309 (Montecarotto municipality - Scenario B)

TECHNICAL REPORT — SLOPE UNIT 1309

Organization: DST - UniFi
Authors: Dr. Nicolò Brilli
Version: SADRI - Preliminary Analysis V1.1

Run date/time: 2026-01-12 14:06:15
Contact: nicolo.brilli@unifi.it

Executive Summary

Deterministic baseline (no Monte Carlo)

Impacted assets: 8
Attention-Zone assets: 8
Total economic value (impacted assets): € 1.974.927,72
Total risk (impacted): € 1.303.323,59
Total economic value (attention zone): € 1.694.683,21

Policy note: attention-zone assets do *not* generate an economic risk estimate (value only).

Monte Carlo summary (probabilistic)

Impacted assets (after simulations): 18
Impacted from initially Attention Zone: 6
New impacted (MC-only): 4

Total economic value of impacted assets (post-MC classification): € 2.964.793,43
Average risk when impact occurs (hits only): € 1.385.453,78
Risk variability when impact occurs (standard deviation, hits only): € 221.724,91
Total economic value in attention zone (post-MC classification): € 1.011.659,96

Legal / operational notes

This report was generated automatically. Estimates are subject to uncertainty and must be validated by an expert technician.

Simulation parameters

Methodology / risk parameters

FPT threshold	0.55
Velocity threshold (noise) [m/s]	0.053
Attention buffer [m]	20.0
V_ref [m/s]	10.970475196838379

Simulation configuration

Expected CRS	EPSG:32632
Max workers	11
Monte Carlo	Active (N=100)
Random seed	42
Selected runout analysis model	DebrisDice
DebrisDice model_type	1
DebrisDice source_mode	2,1
DebrisDice iterations	100
DebrisDice range_angle	0.6
DebrisDice minimum_slope_angle	0.04
DebrisDice friction_factor	0.04
DebrisDice mass_to_drag	9.47
DebrisDice viscosity	500.0
DebrisDice shear_yield	500
DebrisDice exponent	2.5
DebrisDice density	2000
DebrisDice bing_depth	4
Monte Carlo iterations	100
Monte Carlo uncertainty_factor	0.2
Monte Carlo parameters_to_perturb	friction_factor, mass_to_drag, minimum_slope_angle, range_angle, shear_yield

Study area overview

Figure 1: Study area overview showing the Slope Unit under analysis (Slope Unit 1309).

Deterministic simulation map (baseline)

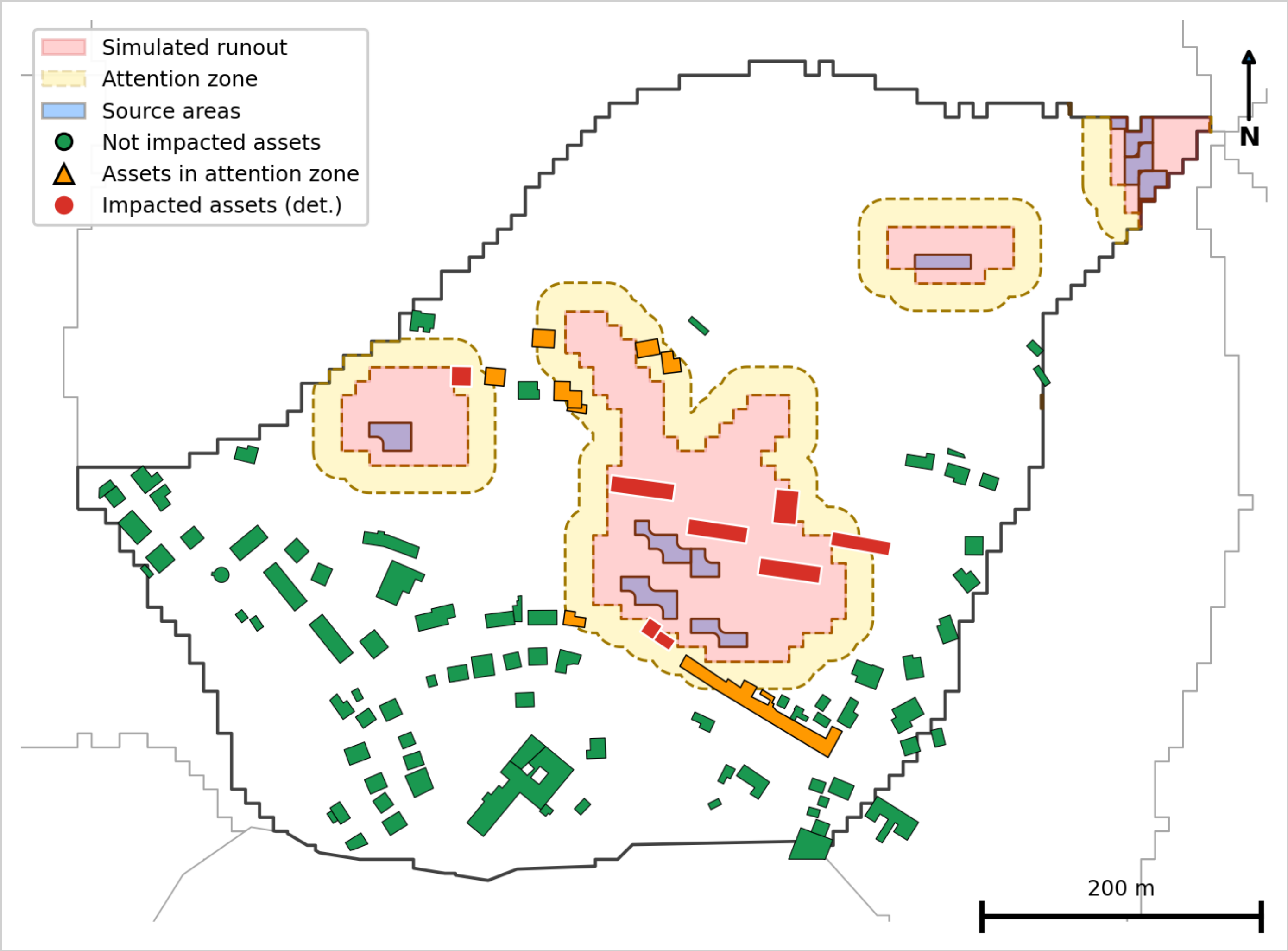


Figure 2: Deterministic runout map with impacted and attention-zone assets (baseline, no MC perturbation).

Monte Carlo probabilistic simulation map

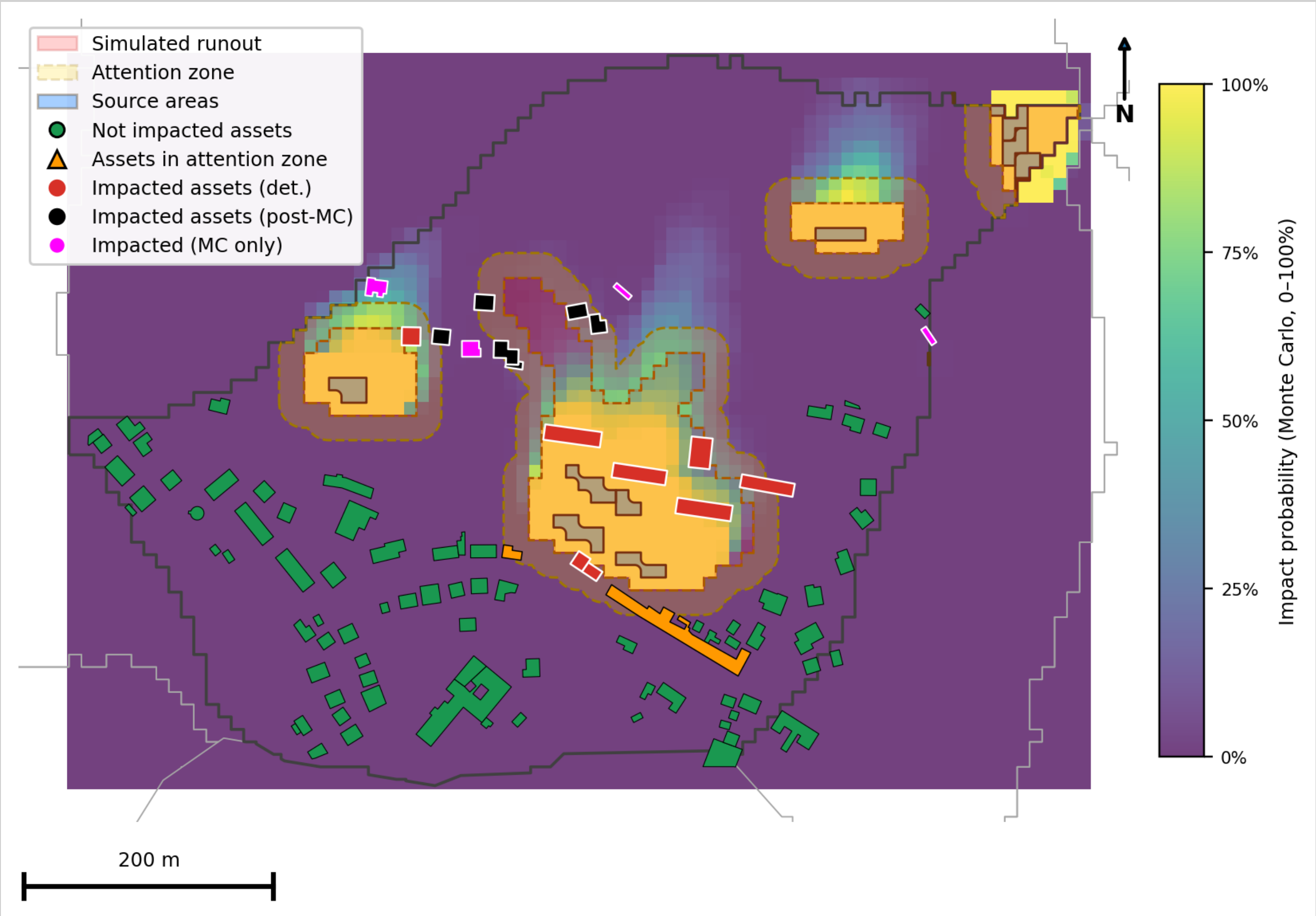


Figure 3: Monte Carlo probabilistic runout and impacted assets classified by impact probability (hit rate).

Municipality risk summary (post-MC classification)

Municipality	Risk (EUR)
Montecarotto	€ 1.352.271,33

Asset tables — deterministic baseline (no MC)

> Baseline deterministic impacted assets. No Monte Carlo frequencies are shown here.

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
79293	Industrial Plant	0.250	€ 3.124,74	Montecarotto	commercial
229401	Residential Building	0.250	€ 46.344,81	Montecarotto	residential
228274	Residential Building	0.358	€ 1.175,90	Montecarotto	residential
228273	Residential Building	0.358	€ 372.597,14	Montecarotto	residential

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
228275	Residential Building	0.358	€ 335.890,14	Montecarotto	residential
228277	Residential Building	0.358	€ 154.224,01	Montecarotto	residential
228276	Residential Building	0.358	€ 374.626,32	Montecarotto	residential
228281	Residential Building	0.358	€ 15.340,53	Montecarotto	residential

> Baseline deterministic attention-zone assets (value only by policy; no risk).

Asset ID	Description	Total economic value (EUR)	Structural vulnerability	Municipality	type
229448	Residential Building	€ 925.263,39	—	Montecarotto	residential
228498	Residential Building	€ 86.396,57	—	Montecarotto	residential
522178	Other	€ 33.228,44	0.358	Montecarotto	other
228278	Residential Building	€ 175.875,35	0.358	Montecarotto	residential
228280	Residential Building	€ 112.838,53	0.358	Montecarotto	residential
228284	Residential Building	€ 112.636,39	0.358	Montecarotto	residential
228283	Residential Building	€ 117.202,61	0.358	Montecarotto	residential
228282	Residential Building	€ 131.241,93	0.358	Montecarotto	residential

Asset tables — Monte Carlo (post-simulation classification)

> Impacted assets after MC: includes promoted from deterministic attention zone and MC-only / new impacted assets. “Impact probability” is the hit rate across MC iterations.

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
79293	Industrial Plant	0.250	€ 3.124,74	Montecarotto	commercial
229401	Residential Building	0.250	€ 46.344,81	Montecarotto	residential
228274	Residential Building	0.358	€ 1.175,90	Montecarotto	residential
228273	Residential Building	0.358	€ 372.597,14	Montecarotto	residential
228275	Residential Building	0.358	€ 335.890,14	Montecarotto	residential

Asset ID	Description	Structural vulnerability	Risk Eur	Municipality	type
228277	Residential Building	0.358	€ 154.224,01	Montecarotto	residential
228276	Residential Building	0.358	€ 374.626,32	Montecarotto	residential
228281	Residential Building	0.358	€ 15.340,53	Montecarotto	residential
522178	Other	0.358	€ 31,50	Montecarotto	other
228278	Residential Building	0.358	€ 32,84	Montecarotto	residential
228280	Residential Building	0.358	€ 78,82	Montecarotto	residential
228284	Residential Building	0.358	€ 118,23	Montecarotto	residential
228283	Residential Building	0.358	€ 59,12	Montecarotto	residential
228282	Residential Building	0.358	€ 6,57	Montecarotto	residential
228279	Residential Building	0.358	€ 288,90	Montecarotto	residential
228380	Residential Building	0.360	€ 7,35	Montecarotto	residential
522321	Other	0.358	€ 358,90	Montecarotto	other
229569	Residential Building	0.360	€ 47.965,51	Montecarotto	residential

> Attention-zone assets after MC: assets that were in deterministic attention zone and were never impacted in Monte Carlo.

Asset ID	Description	Total economic value (EUR)	Structural vulnerability	Municipality	type
229448	Residential Building	€ 925.263,39	0.250	Montecarotto	residential
228498	Residential Building	€ 86.396,57	0.250	Montecarotto	residential

---

SADRI System — Debris-flow risk assessment report
Generated on 2026-01-12 14:06:15 — © SADRI

La borsa di dottorato è stata finanziata con risorse dell'Unione europea – NextGeneration EU
Piano Nazionale di Ripresa e Resilienza (PNRR)
Missione 4 – "Istruzione e Ricerca", Componente 2 – "Dalla ricerca all'impresa"
Linea di investimento 3.1 – "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione"
Decreto Direttoriale 28 dicembre 2021, n. 3264
Avviso pubblico per la presentazione di proposte progettuali per "Rafforzamento e creazione di Infrastrutture di Ricerca"
Progetto IR0000032 – Itineris
CUP B53C22002150006