

DIPARTIMENTO DI SCIENZE BIOLOGICHE, GEOLOGICHE
E AMBIENTALI

Corso di Laurea Magistrale in Geologia e Territorio

Tesi di Laurea Magistrale

Landslide susceptibility mapping using a
physically-based model: The May 2023 Emilia-
Romagna meteorological event

Candidato:
Enea Fiori

Relatore:
Prof. Matteo Berti

Correlatore:
Prof. Marcel Hürlimann

Abstract

Rainfall-induced shallow landslides are an increasingly frequent hazard, with a direct impact on the land and on the people who live on it. Physically-based models assess their susceptibility by coupling a mechanical stability analysis with a hydrological description of the slope. The difficulty lies in the choice of the input parameters. The geotechnical and hydrological properties these models require cannot be measured at catchment scale: they have to be assumed within plausible intervals and calibrated on the observed landslides.

Two major rainfall events struck the Emilia-Romagna region of Italy in May 2023. They triggered thousands of shallow landslides in the Apennine sector, mostly debris slides and debris flows, on slopes where no instability had been recorded before. The inventory compiled immediately afterwards records the conditions under which those failures occurred. This thesis applies the Fast Shallow Landslide Assessment Model (FSLAM) to the municipal territory of Casola Valsenio (84 km², 100 to 960 m a.s.l.), calibrating and evaluating it on that inventory. The purpose is not the best performance the model can give, but a measure of what each modelling choice is worth.

Three elements were fixed at the outset: a balanced validation dataset built from the inventory, with 4948 points per class, the search space of the parameters, and the rule for selecting a solution. Six configurations were then compared on that common basis. They differ in the calibration strategy and in the structure of the model.

Across the six, accuracy ranges from 0.653 to 0.665 and the area under the ROC curve from 0.703 to 0.740, while the parameter sets that produce them differ by orders of magnitude. What the choices do change is the operating point of the model, the plausibility of the calibrated parameters and the consistency of the state of the slopes before the event. That last quantity shows the widest difference of the whole comparison, and it is the one the calibration objective does not measure.

Accuracy alone therefore cannot select the parameters. Physical plausibility has to be set as a criterion before the calibration, not read off the results afterwards. What limits the model here is not its formulation, but the spatial information it can draw on.

Contents

Abstract	2
1 Introduction	5
1.1 Rationale	5
1.2 Objectives	5
1.3 Structure and Workflow	5
2 Shallow Landslides and Susceptibility Assessment	7
2.1 Shallow Landslides: Definition, Classification and Triggering	7
2.1.1 Definition and Classification of Landslides (Cruden & Varnes) .	7
2.1.2 Characterisation by Failure Depth	8
2.1.3 Failure Mechanism and Triggering	9
2.2 The Infinite Slope Stability Model	11
2.3 Landslide Susceptibility: Concept and Assessment Approaches	13
2.4 Statistical (Data-Driven) Methods	15
2.5 Physically-Based Methods	15
3 The Study Area and the May 2023 Rainfall Event	17
3.1 Geographical Overview of the Study Area	17
3.2 Geological Setting	18
3.2.1 Lithostratigraphy	18
3.2.2 Structural Setting	22
3.2.3 Geological Control on Slope Instability	22
3.3 Pre-Event Landslide Inventory	23
3.4 The May 2023 Extreme Event	24
3.4.1 Antecedent Rainfall Conditions	24
3.4.2 The First Meteorological Event (May 1–3, 2023)	25
3.4.3 The Second Meteorological Event (May 16–17, 2023)	26
3.4.4 Cumulative Rainfall and Return Periods	27
3.5 The May 2023 Post-Event Landslide Inventory	28
3.5.1 Regional-Scale Inventory: Mapping Methods, Magnitude, and Landslide Typologies	28
3.5.2 Focus on Casola Valsenio	29
4 Methodology	31
4.1 The FSLAM Model	31
4.1.1 Overview and Rationale	31
4.1.2 Geotechnical Component	31
4.1.3 Hydrological Model	31
4.1.4 Structure and Required Inputs	33
4.1.5 Deterministic / Stochastic Approach	34
4.1.6 Output	34
4.1.7 Modifications	35
4.2 Input Data and Preparation	37
4.2.1 Digital Elevation Model (DEM)	37

4.2.2	Soil Units	39
4.2.3	Land Use and Land Cover Classes	41
4.2.4	Rainfall	43
4.2.5	Soil Thickness	45
4.3	Validation Framework and Criteria	49
4.3.1	Validation Dataset	49
4.3.2	Performance Metrics	51
4.4	Calibration Strategies	53
4.4.1	Calibrated Parameters and Search Space	54
4.4.2	Manual Calibration	56
4.4.3	Automatic Calibration	56
4.4.4	Calibration of the Modified Configurations	57
5	Model Application and Results	59
5.1	Base Configuration – Manual Calibration	59
5.2	Base Configuration – Automatic Calibration	66
5.2.1	Single-Objective Calibration	66
5.2.2	Multi-Objective Calibration	69
5.3	Accumulation-Length Threshold Configuration	74
5.4	Distributed Soil-Thickness Configuration	80
5.5	Comparative Analysis	86
6	Conclusions	91
6.1	Main Results	91
6.2	The Role of Antecedent Rainfall	92
6.3	Limitations	93
6.4	Perspectives	93
	References	95

1 – Introduction

1.1 Rationale

Landslide susceptibility is the propensity of an area to generate landslides. In formal terms, it is the probability that a failure occurs at a given location, given the geo-environmental conditions that characterise it. Its assessment rests on two assumptions (Section 2.3). The first is that the past is a guide to the future, so that areas already affected by landsliding are those where further landsliding is most likely. The second is that the same holds for areas with topography, geology and geomorphology similar to those where landslides have been triggered.

A susceptibility assessment indicates where the terrain conditions make failure more likely, not when a landslide will occur or how often. The time frame remains outside the analysis.

In May 2023 two exceptional rainfall events struck the central and eastern part of the Emilia-Romagna region (Chapter 3). They caused widespread flooding in the plain, thousands of landslides in the Apennine sector, casualties and severe damage across the whole area affected. After the event a detailed mapping of the instability became possible, and with it the compilation of a landslide inventory.

A physically-based model requires geotechnical and hydrological parameters that cannot be measured at the catchment scale: they have to be estimated within plausible intervals and calibrated on the observed failures. The inventory records the conditions under which they occurred, and makes that calibration possible.

1.2 Objectives

The present work carries out an assessment of shallow landslide susceptibility with a physically-based model, calibrated and evaluated on that inventory. The municipal territory of Casola Valsenio was chosen as the study site, an area that stands out within the inventory for the density of the failures and for the accuracy of the mapping (Section 3.5).

The aim of the work is not the most accurate result obtainable, but a measure of how much each variation weighs. Different calibration methods and structural modifications of the model are therefore put to the test, and the configurations are compared under identical conditions, so that the differences observed can be attributed to the choices that distinguish them.

1.3 Structure and Workflow

The work was carried out in the following steps.

1. Framing of shallow landslides and of their triggering mechanism, review of the approaches to susceptibility assessment, and choice of a physically-based model.
2. Characterisation of the study area and of the May 2023 rainfall event, and description of the landslide inventory compiled after it.

3. Detailed description of the model adopted and of its structure.
4. Preparation of the model input data: digital elevation model, litho-technical classes, land cover, antecedent and event rainfall, and normalised map of the soil thickness.
5. Construction of the validation dataset from the inventory, and definition of the search space of the parameters and of the selection rule, both established before the calibration.
6. Calibration of the model along three branches: manual, automatic single-objective and automatic multi-objective.
7. Implementation and recalibration of the two modifications to the formulation: the accumulation-length threshold and the distributed soil thickness.
8. Running and evaluation of the resulting configurations across the different hydrological conditions, and their comparison.

The chapters follow the same sequence: the first two steps correspond to Chapters 2 and 3, the five that follow to Chapter 4, and the last to Chapter 5.

2 – Shallow Landslides and Susceptibility Assessment

2.1 Shallow Landslides: Definition, Classification and Triggering

The assessment of landslide susceptibility requires a preliminary definition of the phenomenon under study, and its placement within a consistent classification framework. This section introduces the terminology adopted throughout the thesis. It moves from the general definition and classification of landslides to the specific characterisation of shallow landslides. These were the predominant failure type triggered by the May 2023 event in Emilia-Romagna (Berti et al., 2025), and are therefore the focus of the present work.

2.1.1 Definition and Classification of Landslides (Cruden & Varnes)

A landslide is commonly defined as “the movement of a mass of rock, debris or earth down a slope” (Cruden & Varnes, 1996). This definition is deliberately broad. It encompasses a wide range of gravity-driven slope processes, which differ markedly in mechanics, velocity and material. The scheme proposed by Cruden & Varnes (1996) organises this diversity. It has become the most widely adopted international reference for landslide description, and remains the standard framework in both scientific and applied contexts.

The classification is built on two independent criteria: the *type of movement* and the *type of material*. Five principal types of movement are distinguished. *Falls* involve the detachment of soil or rock from a steep slope, and the subsequent descent of the detached mass mainly through the air. *Topples* consist of the forward rotation of a mass about a pivot located below its centre of gravity. *Slides* are downslope displacements along one or more well-defined rupture surfaces. They are further divided into *rotational* slides, which develop along a concave surface, and *translational* slides, which occur along an approximately planar surface. *Spreads* entail the lateral extension of a cohesive mass over a weaker underlying layer. Finally, *flows* are spatially continuous movements in which the displaced material behaves like a viscous fluid.

The second criterion subdivides the displaced material into three classes: *rock*, the intact bedrock prior to failure; *debris*, predominantly coarse soils; and *earth*, predominantly fine soils. Combining movement and material yields compound terms that identify each phenomenon unambiguously, such as *rock fall*, *debris flow* or *earth slide* (Table 2.1). The framework is further complemented by descriptors of the state of activity and by a velocity scale, ranging from extremely slow to extremely rapid (Cruden & Varnes, 1996). It was subsequently revised and expanded by Hungr et al. (2014), to reflect advances in landslide characterisation.

Table 2.1: Abbreviated classification of slope movements, adapted from Cruden & Varnes (1996).

Type of movement	Type of material		
	Bedrock	Soils	
		Predominantly coarse	Predominantly fine
Fall	Rock fall	Debris fall	Earth fall
Topple	Rock topple	Debris topple	Earth topple
Slide	Rock slide	Debris slide	Earth slide
Spread	Rock spread	Debris spread	Earth spread
Flow	Rock flow	Debris flow	Earth flow

2.1.2 Characterisation by Failure Depth

The classification framework introduced in the previous section organises landslides by type of movement and type of material, but it does not consider the depth at which failure occurs. This dimension is central to the present work. The depth of the rupture surface is a criterion essentially orthogonal to the Cruden and Varnes scheme (Cruden & Varnes, 1996): a landslide of a given movement–material type, such as a debris slide, may in principle develop along a slip surface located anywhere from a few decimetres to several tens of metres below the ground surface. Sidle & Ochiai (2006) accordingly organise soil mass movements into broad process categories, distinguished among other criteria by the depth and the rate of the displacement. These range from shallow, rapid landslides to slower, deep-seated failures. Shallow landslides are those in which failure takes place close to the ground surface. No single threshold is universally adopted, but they are commonly characterised by a soil mantle less than about 2 m thick, developed in soils with low cohesion (Sidle & Ochiai, 2006, p. 26).

The shallow nature of these failures is governed by the structure of the slope cover. The soil mantle generally rests on a sharp discontinuity, either the underlying bedrock or a less permeable horizon such as a compact till (Sidle & Ochiai, 2006, p. 26). This contact acts as the failure plane, and marks a strong contrast in hydraulic conductivity. It localises the rupture surface: it is a plane of mechanical weakness and, being comparatively impermeable, a boundary at which infiltrating water tends to accumulate. For these reasons the slip surface of shallow landslides typically develops at or near the base of the soil mantle, most commonly at the soil–bedrock interface, rather than within the bedrock itself.

Applied to soil-mantled slopes, the depth criterion identifies a small set of recurrent types: in coarse-grained materials, debris slides, debris avalanches and debris flows; in fine-grained materials, earth slides and earth flows (Sidle & Ochiai, 2006, p. 26). The update of the Varnes classification proposed by Hungr et al. (2014) provides precise definitions for the coarse-grained types most relevant to the present study. A *debris slide* is the sliding of a mass of granular material on a shallow, planar surface parallel with the ground; the sliding mass is usually a veneer of colluvium or weathered soil resting on a stronger substrate. A *debris avalanche* is a very rapid to extremely rapid shallow flow of saturated debris on a steep, unconfined slope. A *debris flow* is a comparably rapid, surging flow of saturated debris confined within an established channel (Figure 2.1).

These types are genetically linked rather than mutually exclusive (Figure 2.1). Shallow failures frequently originate as comparatively slow debris slides. As they incorporate water and lose internal cohesion, they mobilise into much faster debris avalanches and, where liquefaction and channelisation occur, into debris flows (Sidle & Ochiai, 2006, p. 27). This slide-to-flow continuum is the behaviour documented for the May 2023 Emilia-Romagna event (Berti et al., 2025).

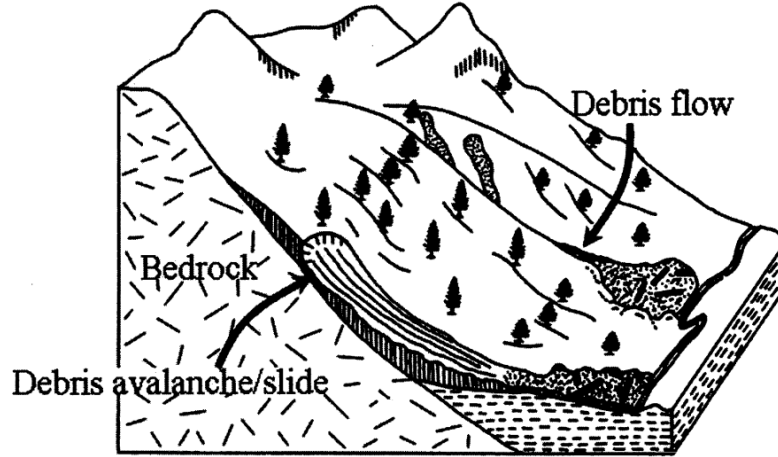


Figure 2.1: Schematic of a shallow, rapid landslide on a steep hillslope, where a thin soil cover overlies relatively impermeable bedrock. The debris slide/avalanche triggers a debris flow that extends into a headwater channel. From Sidle & Ochiai (2006, p. 26).

2.1.3 Failure Mechanism and Triggering

The occurrence of a shallow landslide reflects the interplay of two distinct groups of factors. *Predisposing* (or conditioning) factors define the intrinsic susceptibility of a slope and evolve over long time spans. They include the gradient and shape of the hillslope, the thickness and shear strength of the soil mantle, the underlying lithology and the land cover. *Triggering* factors, by contrast, are the short-term perturbations that bring an already predisposed slope to failure. For shallow landslides the dominant trigger is intense or prolonged rainfall, occasionally combined with, or replaced by, rapid snowmelt or seismic shaking. Rainstorm-induced shallow failures are believed to account for the greatest number of slope movements, and the largest volume of mass wasting, worldwide (Sidle & Ochiai, 2006, p. 29). The following discussion therefore focuses on the rainfall-driven failure mechanism, which is the one relevant to the May 2023 Emilia-Romagna event.

The mechanical basis of rainfall-induced failure lies in the principle of effective stress. The shear strength that a soil can mobilise along a potential slip surface is governed by the Mohr–Coulomb criterion. It depends on the effective normal stress, that is, on the difference between the total stress and the pore-water pressure. During prolonged or intense rainfall, water infiltrating the soil cover raises the pore-water pressure (Sidle & Ochiai, 2006, p. 29), and dissipates the matric suction that characterises the partially saturated cover. Both processes reduce the effective stress acting on the incipient failure plane, and with it the available shear strength. Failure is initiated once the shear strength falls below the gravitational shear stress that drives the soil mass downslope.

Two additional contributions to strength are particularly relevant for shallow failures in soil covers, and both are sensitive to the same processes. The apparent cohesion provided by matric suction is lost as the soil approaches saturation. The root reinforcement supplied by vegetation can be reduced by deforestation, wildfire or root decay (Hungre et al., 2014). Thin soil covers persist on steep slopes largely by virtue of these two contributions, so their reduction during a major rainstorm is often decisive. This explains why shallow landslides may be triggered even on densely forested slopes, as widely observed during the May 2023 event (Berti et al., 2025).

The pore-water pressures responsible for failure are generated by the downward infiltration of rainfall and the advance of a wetting front through the soil cover. Because the base of the cover coincides with a marked permeability contrast, infiltrating water tends to accumulate above it, promoting the rapid build-up of a perched water table and of slope-parallel seepage (Hungre et al., 2014). Iverson (2000) showed that the pore-pressure field controlling shallow landsliding can be separated into two components acting on different timescales. A *steady*, background component is established over long periods, from days to decades, and is governed by the long-term drainage and the topographic setting of the slope. It sets the predisposition of the hillslope, but does not by itself trigger failure. Superimposed on it is a *transient* component, generated during and immediately after a storm over much shorter timescales, from minutes to months. It is this transient pressure pulse, once transmitted to the potential slip surface, that triggers the landslide (Iverson, 2000).

This framework also clarifies the role of antecedent conditions. A slope that is already wet at the onset of a storm starts from a higher background pressure, so that a smaller transient input is sufficient to initiate failure. It further accounts for the rapidity of shallow failures: the characteristic pore-pressure response time increases with the square of the failure depth, so thin soil covers respond almost immediately to intense rainfall (Iverson, 2000).

At the regional scale, the link between rainfall and shallow landsliding is most commonly expressed through empirical rainfall thresholds. These identify the minimum rainfall conditions, typically combinations of mean intensity and duration, above which failures are expected. Building on the pioneering work of Caine (1980), Guzzetti et al. (2008) compiled a global dataset of rainfall events associated with shallow landslides and debris flows. They confirmed that the minimum intensity required to trigger failure decreases with increasing rainfall duration, following a power-law (log-log linear) relationship in intensity-duration space. Such thresholds provide an empirical, operational counterpart to the physical mechanism described above, and are widely used for forecasting and early warning.

For the study region, Berti et al. (2012) derived probabilistic rainfall thresholds for the Emilia-Romagna Apennines through a Bayesian approach, returning a landslide probability for given combinations of event duration, intensity and total rainfall. The May 2023 rainfall sequence, which triggered the shallow landslides analysed in this thesis, is examined in detail in Chapter 3.

Reproducing this triggering mechanism in a spatially distributed manner is the aim of the physically-based susceptibility models adopted in this thesis, whose theoretical basis is developed in the following sections.

2.2 The Infinite Slope Stability Model

The shallow landslides characterised in the previous section develop along a slip surface that is thin, approximately planar, and oriented parallel to the ground surface (Sidle & Ochiai, 2006, p. 26). The failed mass is markedly elongated, with depth-to-length ratios typically smaller than 0.1 (Cruden & Varnes, 1996). Because the mass is much longer and wider than it is thick, the lateral and upslope–downslope boundary effects on its equilibrium become negligible. The slope can therefore be idealised as an infinitely long, uniform incline, and its stability assessed on a single representative soil column, bounded below by a slip surface parallel to the topography. This is the *infinite slope model* (Figure 2.2).

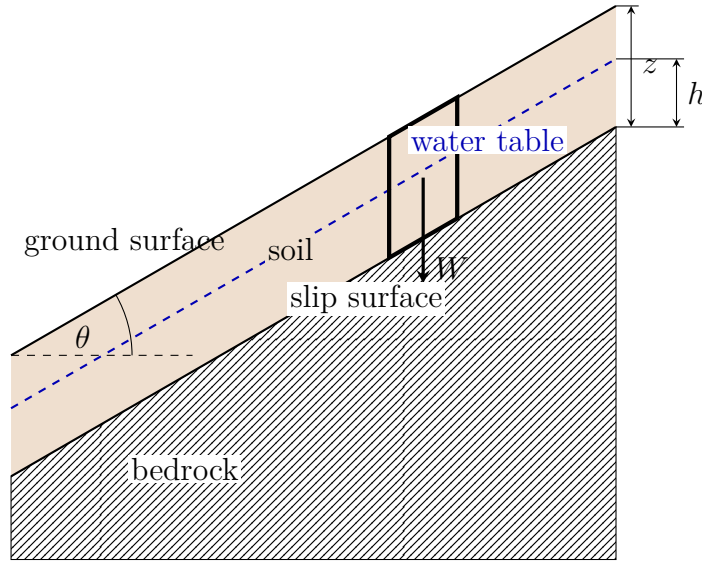


Figure 2.2: Schematic of the infinite slope model. A soil layer of thickness z , inclined at the slope angle θ , rests on an impermeable bedrock that defines the slip surface; the water table lies at a height h above it, with seepage parallel to the slope. W is the weight of a representative soil column.

This idealisation is well suited to shallow, translational failures over a lithic or otherwise low-permeability contact, for which the assumption of seepage parallel to the slope is likewise appropriate. On this basis, the infinite slope model underpins widely used physically-based models of shallow landsliding (Borga et al., 1998; Montgomery & Dietrich, 1994). The model rests on a set of simplifying assumptions: the soil cover is homogeneous and of uniform thickness; the slip surface is planar and parallel to an infinitely long slope; seepage, where present, is slope-parallel; and conditions are invariant along and across the slope, so that the problem reduces to a one-dimensional analysis of a single column. Although idealised, these assumptions capture the essential mechanics of shallow slope failure, and support the factor-of-safety formulation developed below.

The stability of the soil column is quantified by the *factor of safety* FS , defined as the ratio between the shear strength available along the slip surface and the shear stress mobilised on it. The slope is stable for $FS > 1$, at limit equilibrium (incipient failure) for $FS = 1$, and unstable for $FS < 1$. Consider a soil column of vertical thickness z on a slope of inclination θ . Resolving its weight into the components normal and

tangential to the slip surface yields a normal stress $\sigma = \rho_s g z \cos^2 \theta$ and a driving shear stress $\tau = \rho_s g z \sin \theta \cos \theta$, where ρ_s is the bulk soil density and g the gravitational acceleration. The shear strength is expressed in terms of effective stress through the Mohr–Coulomb criterion, $\tau_f = C + (\sigma - u) \tan \varphi$, in which C is the cohesion, φ the effective friction angle and u the pore-water pressure on the slip surface. The factor of safety then becomes (Lambe & Whitman, 1969; Knappett & Craig, 2012):

$$FS = \frac{C + (\rho_s g z \cos^2 \theta - u) \tan \varphi}{\rho_s g z \sin \theta \cos \theta}. \quad (2.1)$$

For a dry, cohesionless soil ($u = 0$, $C = 0$), Equation 2.1 reduces to its simplest form:

$$FS = \frac{\tan \varphi}{\tan \theta}. \quad (2.2)$$

This shows that such a slope is stable as long as its inclination remains below the friction angle ($\theta < \varphi$). The use of this effective-stress, infinite-slope framework to analyse the stability of natural slopes was pioneered by Skempton & DeLory (1957). In real shallow covers two terms modify this limiting condition: the cohesion C , which raises the available strength, and the pore-water pressure u , which lowers it by reducing the effective stress. Because u provides the explicit link to the slope hydrology, its evaluation is addressed below.

In a shallow cover with a water table standing at a vertical height h above the slip surface and seepage parallel to the slope, the pore-water pressure on the failure plane is $u = \rho_w g h \cos^2 \theta$, where ρ_w is the density of water. Substituting this expression into Equation 2.1 and rearranging yields the working form of the infinite slope factor of safety,

$$FS = \frac{C}{\rho_s g z \sin \theta \cos \theta} + \left(1 - \frac{h}{z} \frac{\rho_w}{\rho_s}\right) \frac{\tan \varphi}{\tan \theta}, \quad (2.3)$$

in which the stabilising contribution of cohesion (first term) is separated from the frictional one (second term), the latter modulated by the relative water-table height h/z . This is the formulation on which most physically-based shallow-landslide models are built (Montgomery & Dietrich, 1994; Borga et al., 1998).

The ratio h/z measures the degree of saturation of the soil cover, and varies between 0 and 1. As the water table rises, the term $(h/z)(\rho_w/\rho_s)$ grows, the effective stress and the frictional resistance decrease, and the factor of safety falls; the fully saturated condition, $h = z$, is the most critical. For a given soil and slope, the factor of safety is therefore governed by the position of the water table, which is in turn set by the slope hydrology. In physically-based models, Equation 2.3 is coupled with a hydrological model that predicts h from the balance between recharge and topographically-controlled drainage (Montgomery & Dietrich, 1994; Borga et al., 1998). Consistently with the triggering mechanism discussed earlier, this water table integrates a steady, antecedent component and a transient, event-driven one (Iverson, 2000). The same slope may therefore remain stable under ordinary conditions, and reach failure when intense rainfall drives h towards the surface.

The cohesion term C in Equation 2.3 gathers the strength contributions already introduced when discussing the failure mechanism: the effective cohesion of the soil, C_s , and the apparent cohesion provided by root reinforcement, C_r , so that $C = C_s + C_r$. Both are modest in the granular, weathered covers typical of shallow failures, and the root contribution is transient, being lost where vegetation is removed. Some regional

models therefore neglect cohesion and adopt the cohesionless limit ($C = 0$), regarded as representative of coarse debris covers (Borga et al., 1998; Montgomery & Dietrich, 1994). Others retain it explicitly, particularly the root term, which is decisive on the forested slopes where many of the May 2023 failures occurred.

The infinite slope model owes its usefulness to its simplicity, but the same simplicity defines its limitations. It assumes a homogeneous cover of uniform thickness and a planar slip surface of infinite lateral extent. It therefore neglects the three-dimensional effects that may be significant at the scale of an individual landslide: lateral root cohesion, buttressing, and flow convergence in topographic hollows. The hydrological description is likewise idealised, the pore pressure being referred to a water table in steady, slope-parallel seepage rather than to the fully transient infiltration process. These assumptions are acceptable for a first-order, regional assessment of shallow-landslide susceptibility, but they should be kept in mind when its results are interpreted.

Applied to a single column, Equation 2.3 describes the stability of one point on a hillslope. Its value for susceptibility assessment emerges when it is evaluated over an entire territory. The factor of safety, or equivalently the critical rainfall or the probability of failure, is computed for each grid cell. The infinite slope model thus becomes the mechanical core of the physically-based, spatially distributed susceptibility models (Borga et al., 1998; Montgomery & Dietrich, 1994) on which the approach adopted in this thesis is based, as detailed later in this chapter.

2.3 Landslide Susceptibility: Concept and Assessment Approaches

Landslide susceptibility is the propensity of an area to generate landslides: in formal terms, the probability of spatial occurrence of known slope failures, given a set of geo-environmental conditions (Guzzetti et al., 2006). It does not forecast the time at which a landslide will occur, how often it will recur, or how large and destructive the failure will be. It therefore predicts where landslides are expected, not when (Guzzetti et al., 2006). On this basis, the predisposing factors introduced earlier become the spatial predictors of instability.

That step rests on two assumptions (Fell et al., 2008). The first is that the past is a guide to the future, so that areas which have experienced landsliding are likely to experience it again. The second is that areas with similar topography, geology and geomorphology are also likely to experience landsliding, and it is this that allows the information of an inventory to be extended beyond the slopes it records. Both are usually reasonable, but they admit exceptions, for instance where the material available to fail has been exhausted by earlier failures.

Susceptibility is the first level of a nested hierarchy adopted in landslide zoning for land-use planning (Fell et al., 2008). Susceptibility zoning classifies existing and potential landslides by area, volume and spatial distribution, but does not estimate how often they occur: the time frame is explicitly left out of the analysis. Hazard zoning takes the outcome of the susceptibility assessment and assigns to it an estimated frequency, that is, an annual probability. Risk zoning takes the outcome of the hazard assessment in turn, and evaluates the damage expected on the elements at risk, accounting for their vulnerability and for the spatial and temporal probability. The present work is confined to the first level.

The output of a quantitative susceptibility assessment may be relative or absolute (Fell et al., 2008). Data-treatment techniques first weigh the individual parameters, then search for the combination that best explains the spatial distribution of the mapped landslides; the resulting scores are reclassified into classes of high, medium and low susceptibility. Absolute susceptibility is obtained instead through deterministic approaches such as slope stability models, and its descriptor is the factor of safety. The distinction has a practical consequence: relative susceptibility holds only within the area assessed, and the same class may correspond to very different absolute susceptibilities in different areas.

Whatever the method, a susceptibility assessment rests on three ingredients. The first is a set of conditioning factors: the thematic layers that describe the terrain attributes controlling slope stability (Table 2.2). The second is a landslide inventory, which records, for the landslides of an area, their location and classification, their volume, their state of activity and the date on which they occurred, together with their other characteristics (Fell et al., 2008); for the study area this role is played by the regional inventories described in Chapter 3. The third is the mapping unit into which the territory is partitioned and over which susceptibility is evaluated. The units proposed in the literature fall into seven groups, among them grid cells and slope units (Reichenbach et al., 2018), and the grid cell is generally the one preferred for heuristic, statistical and physically-based modelling (Guzzetti et al., 1999).

Table 2.2: Main categories of conditioning factors used in landslide susceptibility assessment, with representative examples. Adapted from Reichenbach et al. (2018).

Category	Representative factors
Morphometric (DEM-derived)	Slope gradient, aspect, plan and profile curvature, elevation, contributing area
Geological	Lithology, structural setting, distance to faults
Hydrological	Distance to drainage, drainage density, topographic wetness index
Land cover / soil	Land use, vegetation cover (NDVI), soil type

The outcome is a map that divides the land into homogeneous domains and ranks them by degree of susceptibility (Fell et al., 2008). The reliability of the map is established through validation, by measuring how far the predicted susceptibility levels match the distribution and abundance of the landslides known in the same region (Guzzetti et al., 2006). Since the year 2000, success- and prediction-rate curves and ROC curves have become widely used for this purpose (Reichenbach et al., 2018). They are also the ones adopted here, and are described in Chapter 4.

The methods proposed in the literature fall into five main categories: geomorphological hazard mapping, the analysis of landslide inventories, heuristic or index-based methods, functional statistically based models, and geotechnical or physically-based models (Guzzetti et al., 1999). Geomorphological mapping is a direct and qualitative method, which relies on the ability of the investigator to recognise actual and potential slope failures. The analysis of inventories predicts future patterns of instability from the past and present distribution of landslide deposits, typically through density maps. The indirect methods proceed in steps instead: the landslides are first recognised and

mapped over a training area, the physical factors correlated with instability are identified, their relative contribution is estimated, and the land surface is classified into domains of different degree. Among them, the heuristic approach remains qualitative: it rests on the a priori knowledge of the instability factors of the area, which are ranked and weighted according to their assumed or expected importance. The statistical and the physically-based methods produce numerical estimates, and are examined in the two following sections; the latter is the approach adopted in the present work.

2.4 Statistical (Data-Driven) Methods

Statistical methods work as black boxes: they analyse the functional relationships between the instability factors and the distribution of past and present landslides (Guzzetti et al., 1999). The term indicates that the link is established without passing through the physical mechanism of failure. Such a model is built on the assumption that the factors responsible for past failures in a region will also generate the landslides of the future (Guzzetti et al., 1999). The relationship is expressed as a general linear model, in which the presence of landslides in a sampling unit, or the proportion of its area affected, is related to a set of predictor variables through coefficients estimated from the data (Guzzetti et al., 1999).

Of the multivariate techniques applied to the various mapping units, the most widely used are discriminant analysis, linear and logistic regression, and neural networks (Guzzetti et al., 1999). In recent years the preference has shifted increasingly towards machine-learning methods (Reichenbach et al., 2018), among them fuzzy logic systems, support vector machines and forest trees.

The quality of a statistical model depends on the inventory used to calibrate it: the model reproduces the conditions represented in that inventory, and its predictive power declines where failures occur in settings that the inventory samples poorly (Reichenbach et al., 2018).

This limitation is particularly relevant to the present study. The May 2023 event produced mostly first-failure landslides (Berti et al., 2025), whose distribution and classification are described in Section 3.5. A model calibrated on the historical inventory would therefore be poorly equipped to anticipate this population of failures.

2.5 Physically-Based Methods

Physically-based methods evaluate the infinite slope factor of safety introduced in Section 2.2 for every cell of a digital elevation model, coupling the stability analysis with a hydrological model that supplies the pore-water pressure (Guzzetti et al., 1999; Medina et al., 2021). The outcome depends on how the input parameters are treated: a single value of the factor of safety where they are assigned deterministically, a probability of failure where they are described by probability distributions (Pack et al., 1998). The models developed on this basis differ mainly in the treatment of these two components, the slope hydrology and the parameters.

Steady-state formulations with single-valued parameters, such as SHALSTAB (Montgomery & Dietrich, 1994), return a deterministic outcome. Those that assign uncertainty ranges to the parameters, such as SINMAP (Pack et al., 1998), return the probability that a location is stable, and reduce to the former when the uncertainty

bounds coincide and the cohesion is set to zero. Others describe infiltration in transient conditions, letting the factor of safety evolve during the storm, as in TRIGRS (Baum et al., 2008).

The strength of these methods lies in the explicit description of the failure process. Data-driven approaches disregard the complex physical processes involved in initiation, whereas physically-based models compute stability from the physical properties that govern it (Medina et al., 2021). They can therefore assess slopes on which no failure has been recorded, which is the situation discussed in the previous section. An inventory is not required by the computation, but where one exists it serves to calibrate the parameters, and this is the use made of it in the present work. The rainfall, moreover, enters as an input, so that stability can be assessed for a specific event (Medina et al., 2021).

These advantages come with demanding data requirements. Physically-based models need spatially distributed estimates of the geotechnical and hydrological parameters, which are hard to obtain at the regional scale and carry a considerable uncertainty (Medina et al., 2021); the high computational cost is in turn a recognised drawback of the family (Medina et al., 2021). Their validity is moreover confined to the shallow, translational failures for which the infinite slope assumption holds (Section 2.2), and their hydrological description is necessarily simplified.

For the present study the balance of these considerations favours the physically-based approach, for the reasons set out in the previous section. The model adopted is the Fast Shallow Landslide Assessment Model (FSLAM), which combines a probabilistic formulation with a deliberately simplified structure (Medina et al., 2021). The probabilistic treatment of the parameters mitigates the principal weakness of the family, propagating their uncertainty into the predicted probability of failure. Its formulation and its application to the study area are presented in Chapter 4.

3 – The Study Area and the May 2023 Rainfall Event

3.1 Geographical Overview of the Study Area

The study area encompasses the municipal territory of Casola Valsenio, located in the southeastern part of the Emilia-Romagna region, within the Ravenna province (Figure 3.1). The municipality shares its administrative borders with Castel Bolognese, Riolo Terme, and Brisighella to the east and the north-east, Fontanelice and Borgo Tossignano (Metropolitan City of Bologna) to the west, and Palazzuolo sul Senio (Metropolitan City of Florence, Tuscany region) to the south.

According to the latest demographic data, the municipality has a population of approximately 2500 inhabitants (ISTAT, resident population at February 2026), mainly concentrated in the primary urban centre and scattered in smaller rural settlements throughout the valley.

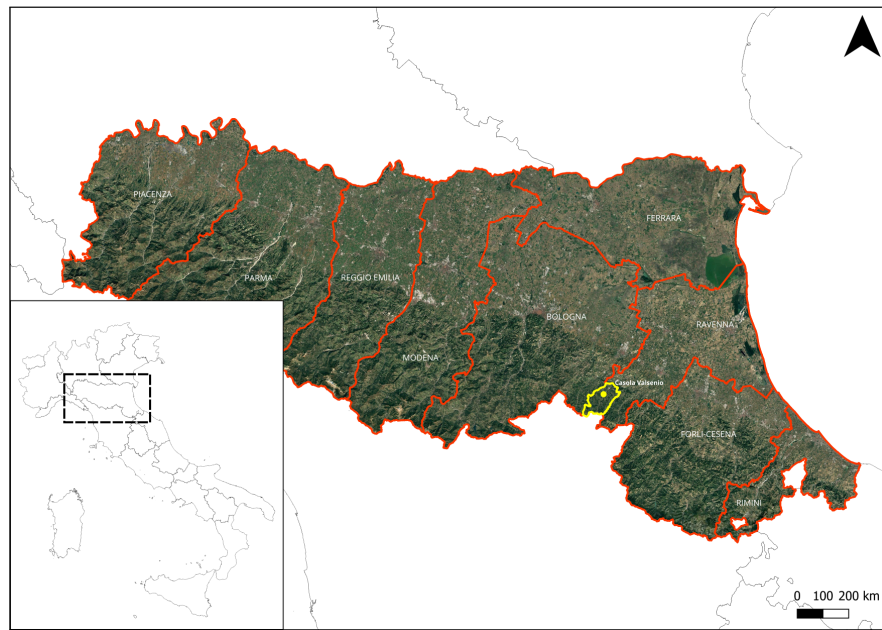


Figure 3.1: Administrative framework and geographical location of the Emilia-Romagna region, with a specific focus on the Casola Valsenio municipal area.

The territory is located in the upper valley of the Senio River and covers a total area of approximately 84 km². The landscape is characterised by a typical Apennine setting, with elevations ranging from approximately 100 m a.s.l. along the main valley floor to approximately 960 m a.s.l. at the highest ridge. The drainage network is hierarchically organised around the main course of the Senio River, which flows from south-west to north-east (Figure 3.2).

Forest is the dominant land cover, occupying about two thirds of the municipal territory, while farmland covers roughly a quarter. The two occupy distinct parts of the landscape. The median slope under forest is 28°, against 13° under farmland, which lies on the valley floor and on the gentler ground.

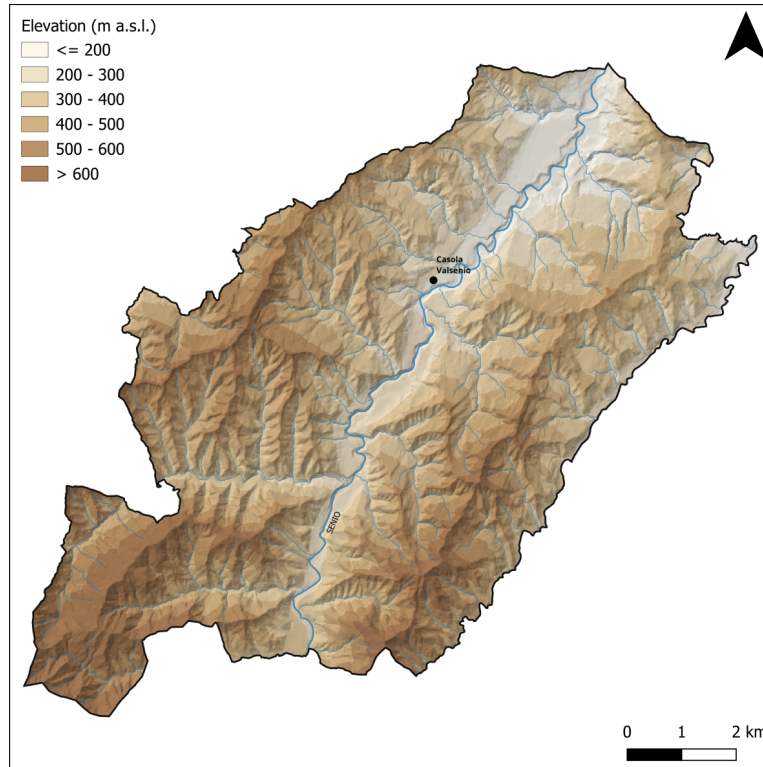


Figure 3.2: Elevation of the Casola Valsenio municipal area with the drainage network of the Senio River.

3.2 Geological Setting

The study area lies in the Romagna sector of the Northern Apennines, the north-east-verging fold-and-thrust belt of the Italian peninsula. The bedrock of the area belongs to the *Umbria-Marche-Romagna succession* and consists almost entirely of the *Marnoso-Arenacea Formation* (FMA) (Benini et al., 2009b; Antolini et al., 2001). The FMA is one of the thickest and most widespread turbidite units of the Northern Apennines. It represents the fill of a foredeep basin that was active from the late Burdigalian to the late Tortonian (Miocene), and that migrated progressively north-eastward as the chain advanced (Benini et al., 2009b; Antolini et al., 2001). Sedimentation was dominated by turbidity currents, fed largely by Alpine-derived clastic input. These entered the narrow, elongated basin from the north-north-west and were channelled south-eastward along its axis (Benini et al., 2009b). The resulting deposits are the rhythmic alternations of sandstone and pelite, with subordinate hemipelagic marls, that give the formation its name (Benini et al., 2009b). On the basis of the sandstone-to-pelite ratio, the mean bed thickness and the stratigraphic position, the formation has been subdivided at the regional scale into fourteen members (Benini et al., 2009b), whose local expression in the Casola Valsenio area is described in the following section

3.2.1 Lithostratigraphy

Within the study area the Marnoso-Arenacea Formation is represented by twelve of its fourteen regional members. These crop out from the Corniolo Member (FMA2, Langhian) at the base to the terminal Borgo Tossignano Member (FMA14, late Tor-

tonian) at the top; FMA1 and FMA6 are not exposed in the area. The succession is overlain by the Ghioli di Letto and the Gessoso-Solfifera, which mark the evaporitic closure of the basin (Figure 3.3). The members are presented below in stratigraphic order, taking the sandstone-to-pelite ratio (A/P) as the leading lithological criterion. The age and A/P character of each unit are summarised in Table 3.1. The regional geological mapping distinguishes lithofacies within the members, as well as *guide levels* (Benini et al., 2009b; Antolini et al., 2001). Lithofacies are not distinguished here, and the guide levels are shown collectively in Figure 3.3.

The lower part of the local succession consists mainly of pelite-dominated members, within which a first sandstone-dominated interval is intercalated. At the base, the Corniolo Member (FMA2) has an A/P ratio of about 1/3 to 1/2. Its reduced thickness and lateral extent reflect its progressive replacement by the overlying Premilcuore Member. The Premilcuore Member (FMA3) interrupts the pelitic trend as a subordinate sandstone-dominated interval, with an A/P ratio ranging from slightly below 1 to about 2, locally up to 6 in amalgamated horizons. The pelite-dominated regime is re-established upward by the Galeata (FMA4) and Collina (FMA5) members. The former is pelitic-arenaceous (A/P generally below 1, mostly 1/3 to 1/2) and rich in carbonate levels. The latter is the most pelite-rich member of this lower interval (A/P commonly between 1/5 and 1/3), with abundant hemipelagic marls and thin-bedded turbidites. The interval is closed by the Monte Bassana Member (FMA7), of pelitic-arenaceous character (A/P at most 1), which encloses the Casaglia chaotic unit: a body of resedimented Ligurian and sub-Ligurian material, whose significance is addressed below together with the other intraformational chaotic bodies of the succession (Benini et al., 2014).

Above, the Nespoli Member (FMA8) is the first predominantly sandstone member of the sequence: its A/P ratio is almost always greater than 1, mostly between 1/2 and 2, and it increases westward to about 20 in the Firenzuola area. The upper boundary of the member corresponds to the base of the Visignano chaotic unit, whose emplacement is discussed below. The overlying members record a return towards more balanced and then pelite-dominated conditions. The Civitella Member (FMA9) is pelitic-arenaceous (A/P generally below 1, about 1/3 to 1/2), with metre- to decametre-thick sandstone horizons of A/P greater than 6, locally amalgamated, some of which are mapped as guide levels. The Dovadola Member (FMA10) is arenaceous-pelitic (A/P between 1/2 and 2), with comparable coarser sandstone packages. The Modigliana Member (FMA11) restores a predominantly pelitic regime (A/P always below 1, between 1/5 and 1/2), within which subordinate sandstone packages are intercalated (Benini et al., 2014).

Above the Modigliana Member the succession culminates in its two coarsest and thickest members, which represent the arenaceous acme of the formation. The Castel del Rio Member (FMA12), the most areally extensive member of the whole formation, is prevalently arenaceous-pelitic, with an A/P ratio greater than 3 that locally exceeds 10 in its upper part; the sandstone beds show a general increase in thickness towards the top. The overlying Fontanelice Member (FMA13) is more arenaceous still and higher in energy (A/P from 3 to 20 and above). It consists of coarse to medium sandstones in thick to very thick beds, locally up to 10 m. This arenaceous phase is abruptly closed by the Borgo Tossignano Member (FMA14), the terminal unit of the formation: a fining-upward sequence of marls and marly clays, with thin subordinate sandstone interbeds that decrease in frequency and thickness upward (A/P from 1/3 to

1/10). It records the end of turbiditic sedimentation in the basin and passes, upward and laterally, into the restricted, anoxic environment of the overlying Ghioli di Letto (Benini et al., 2014; Benini et al., 2009b; Benini et al., 2009a).

The closure of the marine basin is recorded by the two formations that overlie the Marnoso-Arenacea Formation. The Ghioli di Letto (GHL) are an alternation of bituminous pelites and marly biosiltites, with an A/P ratio generally below 1/5 that decreases upward to about 1/10, and a maximum thickness of about 100 m. Deposited between the late Tortonian and the Messinian, they are capped by the Gessoso-Solfifera Formation (GES), the evaporitic top of the local succession: selenitic gypsum and gypsarenites with subordinate bituminous pelites, up to about 150 m thick, deposited in a lagoonal *sabkha*-type environment during the Messinian (Benini et al., 2009b; Benini et al., 2009a).

A recurrent feature of the succession is the presence of chaotic bodies that record synsedimentary instability of the basin floor. Two of these carry extrabasinal material: the Casaglia chaotic unit, of Ligurian and sub-Ligurian affinity, and the Visignano chaotic unit, of Ligurian, epi-Ligurian and sub-Ligurian affinity. The erosive base of the former cuts the Monte Bassana Member (FMA7) and locally reaches the top of the Collina Member. The latter, up to more than 100 m thick, is intercalated between the Nespoli (FMA8) and Civitella (FMA9) members and passes eastward into a predominantly marly slump. A third, more local body is the Bedetta slump, up to 40 m thick, which is intercalated in the lower part of the Nespoli Member and consists of dominant marls with subordinate sandstones (Benini et al., 2014; Benini et al., 2009b).

Taken as a whole, the local Marnoso-Arenacea succession is a rhythmic stack of members alternating between pelite-dominated and sandstone-dominated intervals, from the pelitic lower members, through the arenaceous acme of the Castel del Rio and Fontanelice members, to the terminal Borgo Tossignano marls. It is this alternation, captured by the sandstone-to-pelite ratio rather than by the absolute thickness of each unit, that constitutes the primary lithological control on the predisposition of the slopes to failure, an aspect taken up in Section 3.2.3 (Antolini et al., 2001). Its influence, however, becomes fully apparent only in combination with the structural attitude of the succession, which is described in the following section.

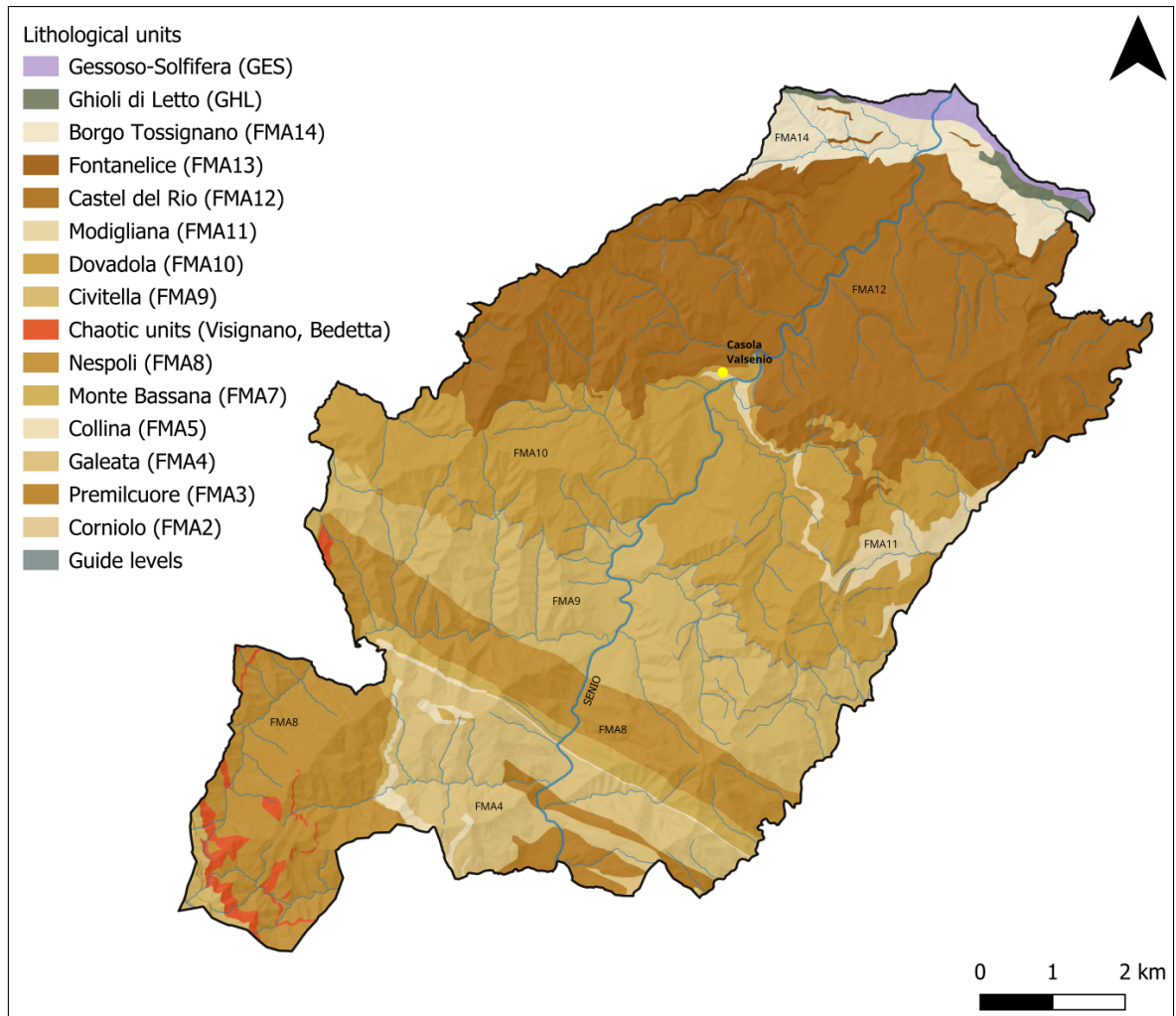


Figure 3.3: Lithological map of the Casola Valsenio area. The Marnoso-Arenacea Formation is represented by twelve of its fourteen members, shaded by their sandstone-to-pelite ratio (A/P) from pelite-dominated (light) to sandstone-dominated (dark), and overlain by the Ghioli di Letto and the Gessoso-Solfifera that mark the evaporitic closure of the basin (Table 3.1). Drainage network and shaded relief derived from the 10 m DEM (Section 4.2.1).

Table 3.1: Members of the Marnoso-Arenacea Formation and the overlying units exposed in the Casola Valsenio area, listed in stratigraphic order (youngest at the top). Compiled from the CARG explanatory notes of sheets 253 Marradi, 239 Faenza, 238 Castel San Pietro Terme and 254 Modigliana (Benini et al., 2014; Benini et al., 2009b; Benini et al., 2009a; Antolini et al., 2001).

Unit (symbol)	Age	A/P
Gessoso-Solfifera (GES)	Messinian	—
Ghioli di Letto (GHL)	late Tortonian – Messinian	$< 1/5$ ($\rightarrow 1/10$)
Borgo Tossignano (FMA14)	late Tortonian	$1/10 - 1/3$
Fontanelice (FMA13)	late Tortonian	$3 - 20+$
Castel del Rio (FMA12)	middle–late Tortonian	> 3 ($\rightarrow >10$)
Modigliana (FMA11)	early–middle Tortonian	$1/5 - 1/2$
Dovadola (FMA10)	early Tortonian	$1/2 - 2$
Civitella (FMA9)	early Tortonian	$1/3 - 1/2$
<i>Visignano chaotic unit (vs)</i>	latest Serravallian–earliest Tortonian	—
Nespoli (FMA8)	late Serravallian	$1/2 - 2$
Monte Bassana (FMA7)	Serravallian	≤ 1
Collina (FMA5)	early Serravallian	$1/5 - 1/3$
Galeata (FMA4)	early Serravallian	< 1 ($1/3 - 1/2$)
Premilcuore (FMA3)	Langhian–early Serravallian	$\sim 1 - 2$ ($\rightarrow 6$)
Corniolo (FMA2)	Langhian	$1/3 - 1/2$

3.2.2 Structural Setting

In the study area the Marnoso-Arenacea Formation is autochthonous, since the succession has not undergone any major tectonic translation. It belongs to the Romagna domain, which corresponds to the upper part of the Umbria-Marche-Romagna succession and the Messinian evaporites, and in outcrop it shows a regular monoclinial attitude dipping towards the north-north-east. In this sector much of the Apennine chain is buried beneath recent alluvial deposits, so that the tectonic architecture is best resolved by combining surface and subsurface data. The main structures are north- to north-east-verging thrusts, active mainly during the Pliocene and the early Pleistocene, and later south- to south-west-verging back-thrusts, along which the Quaternary deformation developed (Benini et al., 2009b).

The attitude of the strata is among the factors that influence the local geomorphology, and dip-slope bedding in particular plays an essential role in the development of slide-type landslides. The more arenaceous members, such as FMA3, FMA8 and FMA13, form the steepest relief, with slope gradients frequently exceeding 50%. Where the bedding is concordant with the hillslope, the valley flanks tend to be gentler and more extensively developed than those on the opposite side, and are also more affected by landsliding (Antolini et al., 2001).

3.2.3 Geological Control on Slope Instability

According to the CARG survey, the lithology of the members is only one of the components that control the distribution and type of slope failures: the others are the attitude of the bedding with respect to the slope, the degree of fracturing and weathering of

the material, and its degree of saturation (Antolini et al., 2001). On this basis each slope requires its own assessment, and the picture given below is a general frame of reference rather than a classification of landslide propensity. It covers the whole range of movement types, from rockfalls to planar and rotational slides and earth flows, and refers only to the units that make up the succession of the study area.

In the pelite-dominated members (FMA5, FMA9, FMA11 and FMA14) the high pelite content and the associated surface weathering favour earth flows, which are particularly frequent in the Borgo Tossignano Member (FMA14). Rockfalls occur where thicker arenaceous beds are undermined by the yielding of the pelites that support them, as in the Modigliana Member (FMA11). The Collina Member (FMA5) locally develops subvertical walls more than 20 m high, from which blocks and pebbles detach (Antolini et al., 2001).

In the sandstone-dominated members (FMA12 and FMA13) the decisive factor is the attitude of the bedding with respect to the slope. Planar slides prevail on dip-slopes, where even the subordinate marly interbeds are sufficient to set thick sandstone slabs in motion (Antolini et al., 2001). In these translational slides the movement takes place along the bedding surfaces, and locally mobilises rock blocks that preserve their internal coherence and recognisable stratigraphic features (Benini et al., 2009b). The steepest faces generate rockfalls instead, most conspicuous in the Fontanelice Member (FMA13), where blocks detach from subvertical walls exceeding 50 m in height (Antolini et al., 2001). Where these intervals and the Messinian gypsum banks give rise to rock scarps, the rockfalls are generally of limited extent (Benini et al., 2009b).

The members with balanced alternations (FMA2, FMA3, FMA4, FMA7, FMA8 and FMA10) constitute the most heterogeneous terrain, in which all failure types are represented. The largest slides again develop on dip-slopes, favoured by thick hemipelagites in contact with arenaceous beds. Planar slides prevail, with the Galeata Member (FMA4) behaving much like the Corniolo, Premilcuore and Monte Bassana members, whereas rockfalls, locally accompanied by rotational movements, are frequent in the Dovadola Member (FMA10). Despite an arenaceous fraction of about 50%, the Nespoli Member (FMA8) offers no particular guarantee of stability: its dense network of fractures in the sandstone beds promotes water circulation and the disaggregation of apparently massive rock bodies, and gives rise to step-like morphologies and block movements (Antolini et al., 2001).

The picture outlined above concerns slope instability in general, and describes mainly movements controlled by the attitude of the bedding and by the fracturing of the rock mass. The present study focuses instead on the shallow, rainfall-triggered failures that characterised the May 2023 event, whose distribution is addressed in Section 3.5.

3.3 Pre-Event Landslide Inventory

To assess the susceptibility of the study area, the distribution of landslides recorded prior to the May 2023 event was analysed. At regional level, the Italian Landslide Inventory (IFFI), compiled by ISPRA together with the regions, records 80,335 pre-event landslides in Emilia-Romagna. In the national mosaic of the River Basin Plans (PAI), 14.6% of the regional territory is classified at high and very high landslide hazard (ISPRA, 2023). Figure 3.4 shows the distribution of landslide bodies recorded prior to May 2023 within the municipality of Casola Valsenio. This pre-event record,

however, concerns a different population of movements from that of May 2023: in the coarse-grained units, to which the Marnoso-Arenacea Formation belongs, nearly all the landslides of the event were first-time failures on slopes previously free of documented landslides (Berti et al., 2025).

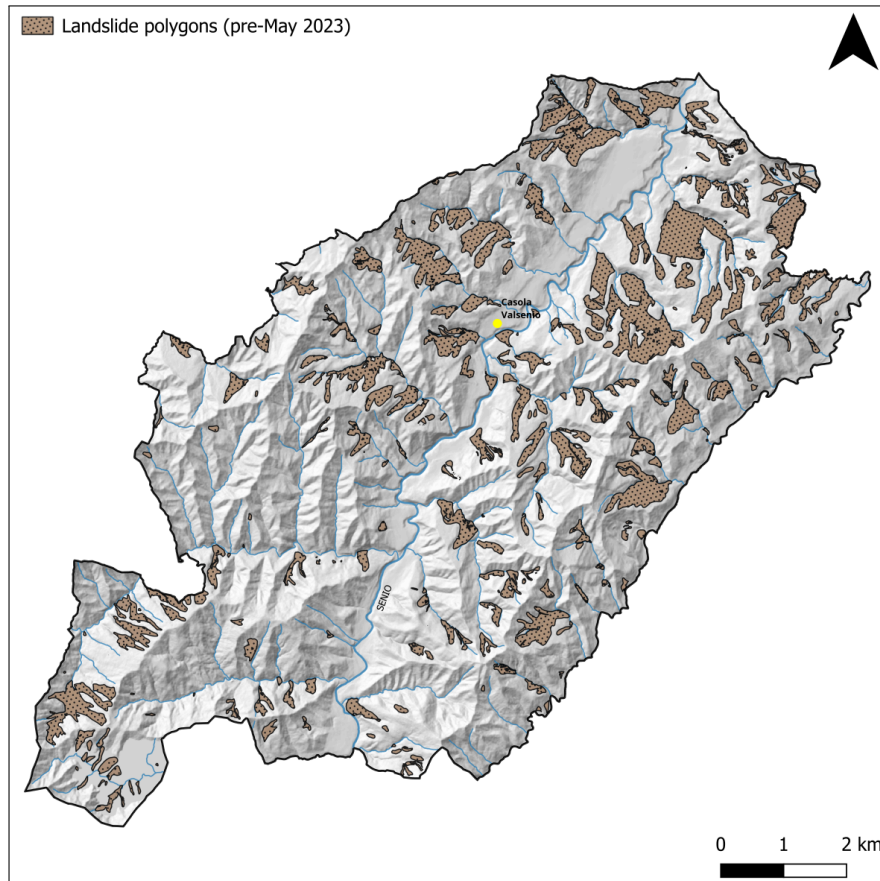


Figure 3.4: Pre-event landslide distribution within the Casola Valsenio municipal area, based on the Italian Landslide Inventory (IFFI).

3.4 The May 2023 Extreme Event

The rainfall event of May 2023 was among the most severe in the modern climatological history of the Emilia-Romagna region. To assess the conditions that triggered the landslides on the slopes, the event must be examined in its entirety, from the antecedent conditions to the two critical peaks, at both regional and catchment scales within the study area.

3.4.1 Antecedent Rainfall Conditions

According to the Arpae monthly bulletins (Arpae Emilia-Romagna, 2023a; Arpae Emilia-Romagna, 2023b; Arpae Emilia-Romagna, 2023c; Arpae Emilia-Romagna, 2023d), in the early months of 2023 the Emilia-Romagna region experienced a phase of progressive rainfall deficit. In January the average total regional rainfall reached 87.3 mm, 30.8 mm above the 1991–2020 climatological mean. Rainfall was concentrated in the middle two weeks of the month, with a fairly uniform geographical

distribution and positive anomalies almost everywhere, most pronounced in Romagna.

The picture gradually reversed in the following months. In February the regional average total fell to 27.7 mm, less than half the climatological mean. The rainfall, which fell almost entirely in the last few days of the month, mainly affected Romagna, where the monthly totals were nonetheless within the normal range. In March the total reached 50.3 mm, 17.2 mm below the climatological mean, a shortfall of 25.5%. Geographically, the anomalies were either zero or slightly positive in Romagna and negative in the rest of the region, with values progressively worsening from east to west.

The deficit peaked in April, with a regional average of just 26.7 mm, 52.2 mm below the climatological mean, a shortfall of 66.2%. This was the third-lowest figure since 1991, after 2007 and 2011. The anomalies were severely negative across the entire region, between 50% and 75% below the mean, with slightly less severe values in the Ferrara area.

Overall, on the eve of the May event, spring 2023 was characterised by a marked and progressive rainfall deficit, culminating in an exceptionally dry April (Figure 3.5).

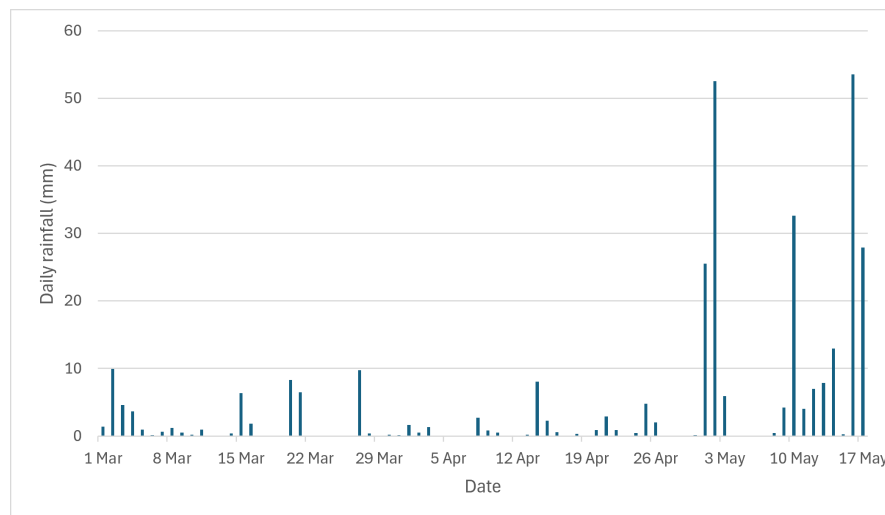


Figure 3.5: Daily rainfall averaged over the rain-gauge network of the Emilia-Romagna region, from 1 March to 17 May 2023. Rain-gauge data from Arpae (Arpae-SIMC, 2023).

3.4.2 The First Meteorological Event (May 1–3, 2023)

The first event began on 1 May and was characterised by persistent rainfall, mainly of light to moderate intensity, which produced low hourly totals throughout the event while generating significant overall accumulations (Arpae Emilia-Romagna, 2023f). On 1 May daily totals remained modest, with a maximum of 66.8 mm/24h at Guiglia (MO).

The bulk of the rainfall was concentrated on 2 May, when numerous stations in the central-eastern sector exceeded 100 mm/24h. The highest accumulations were measured in the Bologna area and in Romagna, with daily maxima of 206.4 mm at Le Taverne (Fontanelice, Santerno catchment), 187.6 mm at Trebbio (Modigliana, Lamone catchment) and 183.2 mm at Monte Albano (Casola Valsenio, Senio catchment). Also in the Senio catchment, the Casola Valsenio station recorded 172.4 mm. The western part of the region, by contrast, was only marginally affected. On 3 May the rainfall gradually subsided, without producing any significant figures.

Over the event as a whole, the greatest accumulations, locally above 200 mm, affected the whole central-eastern part of the region (Figure 3.6). Although the event did not reach exceptional local intensities, falling short of the maximum levels recorded elsewhere in the region in recent years, it stood out for the exceptional duration and the extent of the area affected.

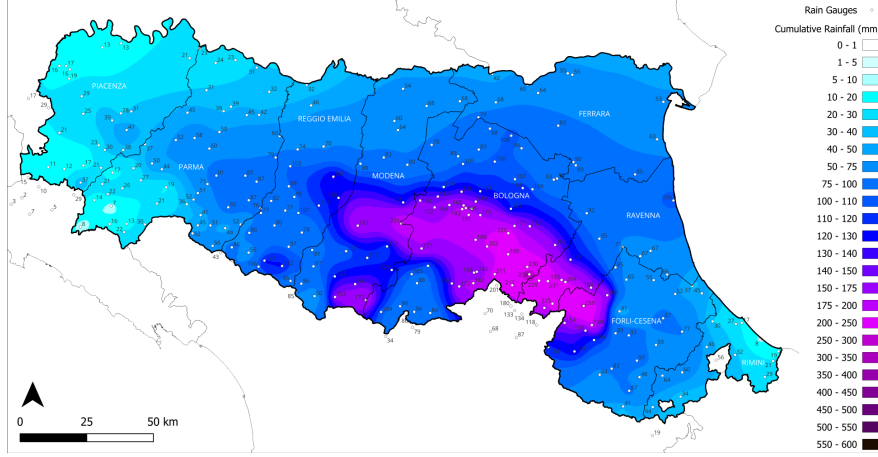


Figure 3.6: Regional distribution of cumulative rainfall in Emilia-Romagna during the first event (May 1–3, 2023). Rain-gauge data (white dots) from Arpae (Arpae-SIMC, 2023), interpolated as described in Section 4.2.4.

3.4.3 The Second Meteorological Event (May 16–17, 2023)

Between the two main events, light residual rainfall and thunderstorms occurred on 9–10 and 12–13 May. Over the whole period leading up to the second event (1–15 May), which also included the first, cumulative rainfall exceeded 200 mm in the hilly catchments from the Samoggia to the Montone, including the Senio. The soils therefore reached total saturation before the second event began, unlike the first (Arpae Emilia-Romagna, 2023e).

The second event was concentrated on 16 May, with rainfall widely exceeding 100 mm. On 17 May the totals were lower, as the event died out in the afternoon. The provinces most affected were Forlì-Cesena, Ravenna, Bologna and Rimini. On 16 May the daily maximum rainfall reached 205.6 mm at Trebbio (Modigliana) and 196.8 mm at San Cassiano (Brisighella), both in the Lamone catchment.

Over the two days, the greatest accumulations, locally above 200 mm, affected much of the central-eastern Apennine sector of the region, from the Bologna to the Rimini area (Figure 3.7). The maxima were 260.8 mm at Monte Albano (Senio catchment), 254.8 mm at Trebbio and 254.6 mm at San Cassiano (Lamone catchment). In the Modena area rainfall remained below 100 mm, and in the Ferrara area it was around 50–70 mm. The event affected the same catchments already hit between 1 and 3 May, but in the Senio, Lamone and Montone catchments the 24-hour rainfall of 16–17 May exceeded the all-time highs just set by the first event. The event as a whole was exceptional both in intensity and in the extent of the territory affected (Arpae Emilia-Romagna, 2023e).

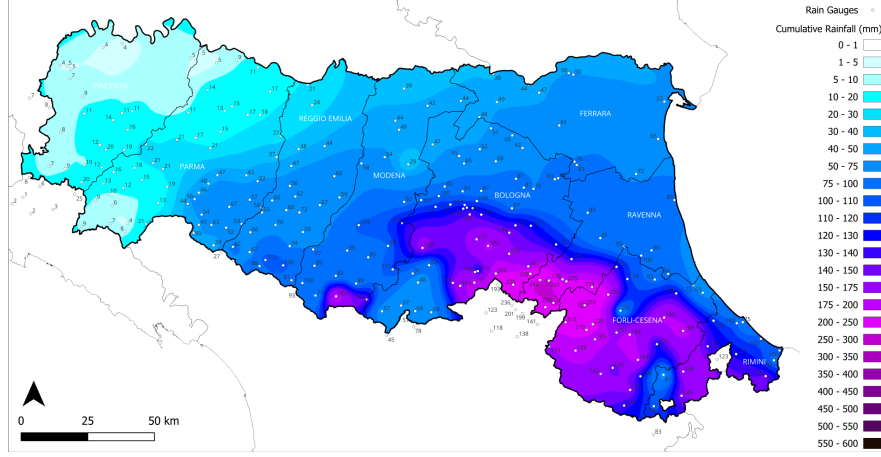


Figure 3.7: Regional distribution of cumulative rainfall in Emilia-Romagna during the second event (May 16–17, 2023). Rain-gauge data (white dots) from Arpae (Arpae-SIMC, 2023), interpolated as described in Section 4.2.4.

3.4.4 Cumulative Rainfall and Return Periods

Taking the event as a whole (1–17 May), the average regional cumulative rainfall was 221.40 mm, the highest such value ever recorded over a 17-day period since 1961 (Figure 3.8). The highest figures were recorded at the Trebbio station (Modigliana, Lamone catchment) with 609.8 mm and at Le Taverne (Fontanelice, Santerno catchment) with 563.4 mm. Across the Apennine areas of Bologna and Romagna, the two events together accounted for around a quarter of the annual climatological total (Arpae Emilia-Romagna, 2023e).

Extreme-value analysis estimates return periods exceeding 500 years for daily and two-day durations in the Senio, Lamone, Montone and Ronco catchments, while for sub-daily durations, estimated from a much shorter series, the figures are more modest and mostly below 100 years (Brath et al., 2023). The Casola Valsenio area, falling within the Senio catchment, is therefore among the areas most severely affected by the event.

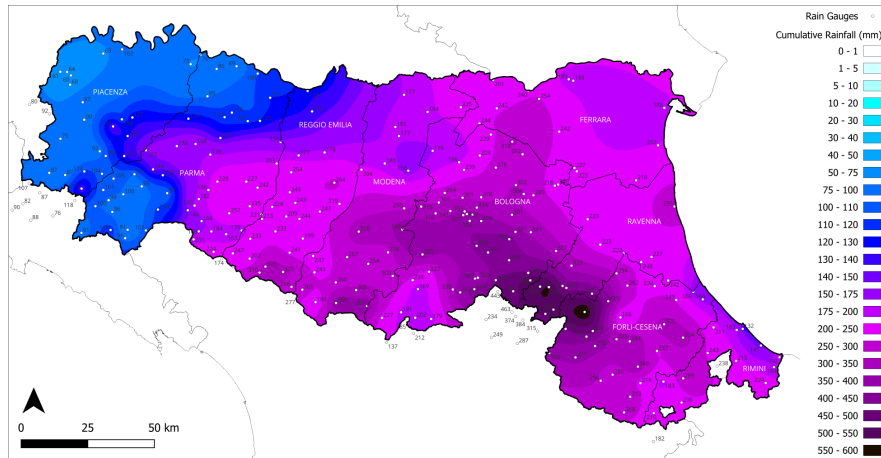


Figure 3.8: Regional distribution of total cumulative rainfall in Emilia-Romagna during the entire critical period (May 1–17, 2023). Rain-gauge data (white dots) from Arpae (Arpae-SIMC, 2023), interpolated as described in Section 4.2.4.

3.5 The May 2023 Post-Event Landslide Inventory

The exceptional rainfall sequence described above triggered widespread slope instability across the entire southern sector of Emilia-Romagna, from the province of Reggio Emilia to that of Rimini. The systematic census of the resulting failures is documented in the RER2023 inventory (Berti et al., 2025). This inventory was officially designated as the reference landslide map for the May 2023 event by the Emilia-Romagna Region and the Po River Authority, and it underpins the *Piano Speciale Dissesto Idrogeologico*, the special plan governing the post-event reconstruction.

3.5.1 Regional-Scale Inventory: Mapping Methods, Magnitude, and Landslide Typologies

The inventory was compiled through manual photo-interpretation, carried out by expert technicians, of high-resolution aerial images (0.2 m) acquired on 23 May 2023, shortly after the second rainfall event, by comparing the pre- and post-event conditions of the slopes within a GIS environment. The entire project was then subjected to a systematic quality control process, which ensured a high level of accuracy in the final dataset (Berti et al., 2025).

A total of 80,997 landslide polygons were mapped over an area exceeding 6000 km², with local densities reaching about 200 landslides per square kilometre (Berti et al., 2025) (Figure 3.9). Following the Cruden & Varnes (1996) scheme, the phenomena were classified as debris slides, debris flows and rock-block slides in the coarse-grained units, and as earth slides and earth flows in the fine-grained units. Debris slides are by far the dominant type, about 66% by number (Figure 3.10). Typically small ($< 1000 \text{ m}^2$) and developed on slopes steeper than 25–30°, they were mostly fast and highly mobile. Together with debris flows, a further 15% and consistently initiated by debris slides, they account for 81% of the total, confirming that shallow sliding of the weathered soil cover was the prevailing failure mechanism of the event. By area the picture shifts: debris slides fall to 49% and debris flows rise to 29%, so the flows are fewer in number but far more extensive. Rock-block slides, though less than 2%, produced the most severe local impacts, developing as translational planar slides along the bedding planes of the Marnoso-Arenacea Formation on comparatively gentle slopes ($< 15^\circ$). Earth slides and earth flows, confined to the fine-grained units, largely reactivated pre-existing dormant landslides.

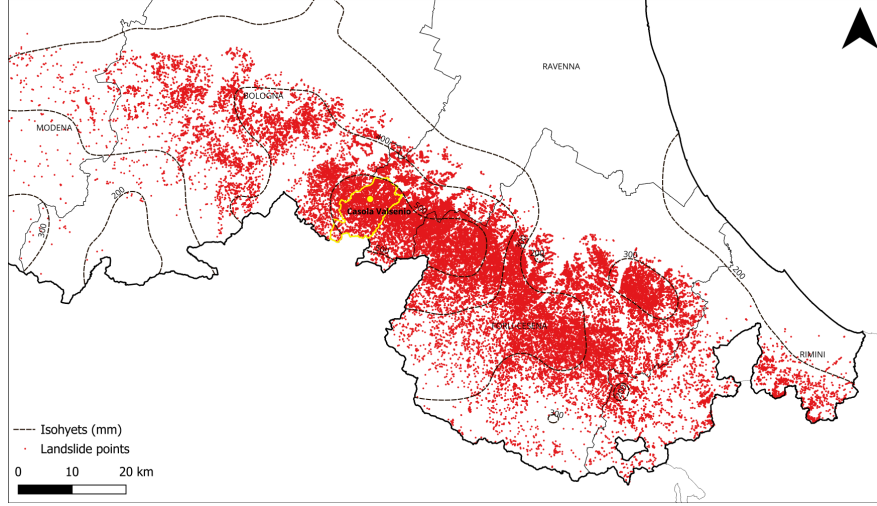


Figure 3.9: Distribution of the May 2023 landslides (Berti et al., 2025), with isohyets of cumulative rainfall (mm) for 1–17 May 2023 (Arpae Emilia-Romagna, 2023e). Casola Valsenio is outlined in yellow.

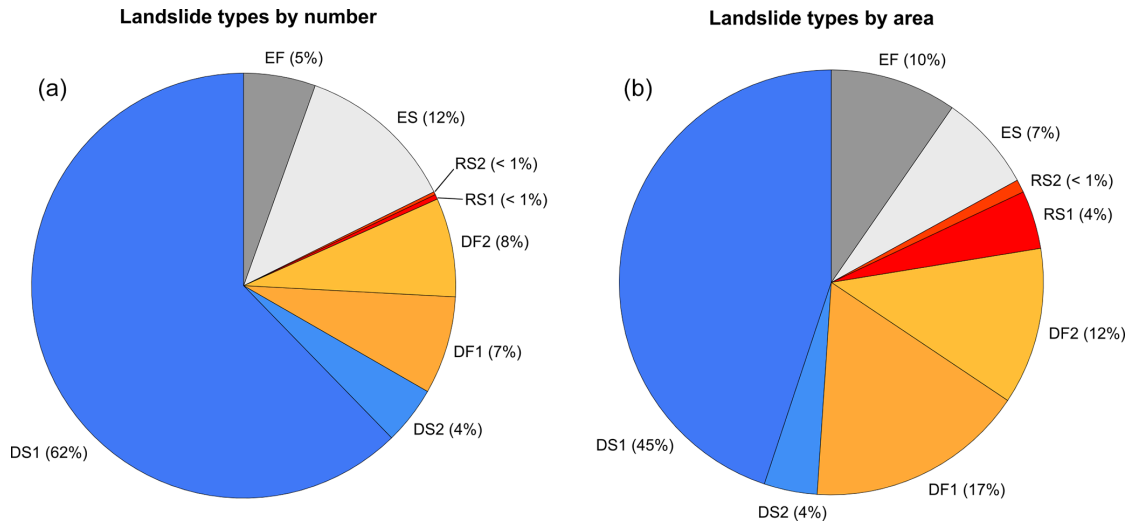


Figure 3.10: Percentage of landslide types in the RER2023 inventory, by number (a) and by area (b). DS: debris slides; DF: debris flows; RS: rock-block slides; ES: earth slides; EF: earth flows. Classes 1 and 2 denote high and low mobility respectively. From Berti et al. (2025).

3.5.2 Focus on Casola Valsenio

The municipality of Casola Valsenio lies within the Romagna sector, where the Marnoso-Arenacea Formation predominates (Section 3.2), and it was one of the areas with the highest landslide density during the event. Within the RER2023 inventory the area also stands out for the accuracy of its mapping, to the extent that it was used as a reference in the construction of the dataset (Berti et al., 2025), which makes the local inventory particularly reliable.

A total of 4,156 debris slides and 1,115 debris flows were mapped in the area, almost exclusively first-time failures on slopes carved in the Marnoso-Arenacea Formation

(Berti et al., 2025) (Figure 3.11). Casola Valsenio is therefore the site on which the susceptibility analysis of this thesis is based.

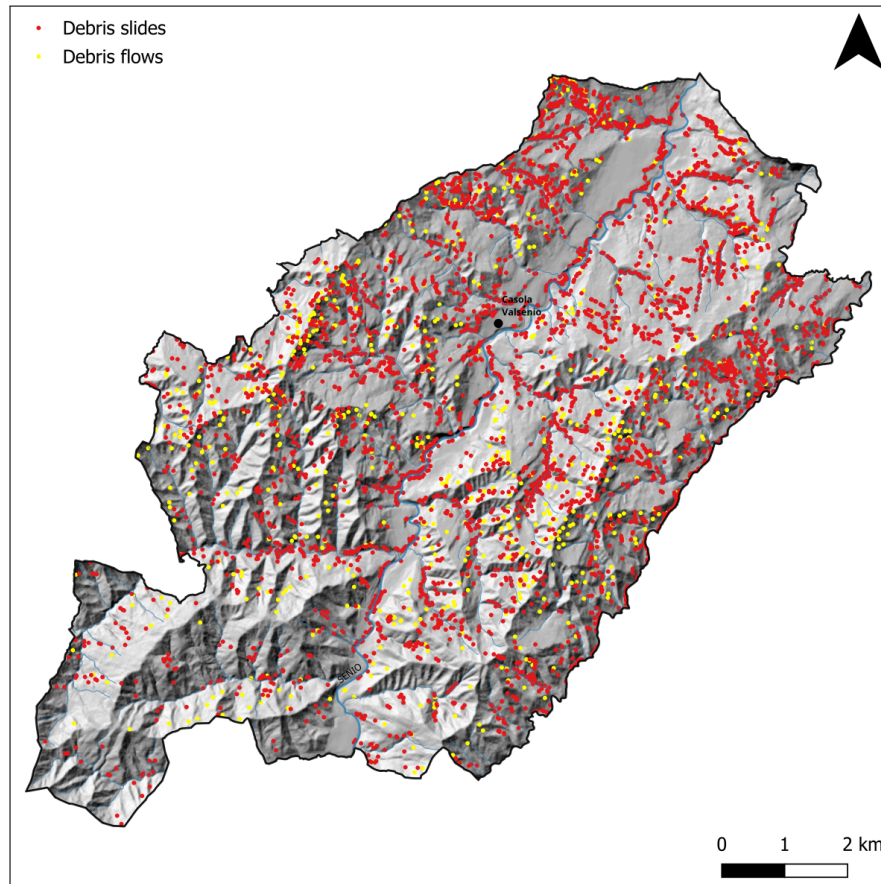


Figure 3.11: Initiation points of the debris slides and debris flows mapped within the Casola Valsenio municipal area after the May 2023 event (Berti et al., 2025).

4 – Methodology

4.1 The FSLAM Model

4.1.1 Overview and Rationale

The model adopted in this study is the Fast Shallow Landslide Assessment Model (FSLAM), a physically-based model (Section 2.5) for assessing the susceptibility to rainfall-induced shallow landslides at the regional scale. It was designed to perform calculations over large areas from high-resolution topographic data, while minimising the computation time. Structurally, FSLAM couples the infinite slope theory with a simplified hydrological model that describes the pore-water pressure conditions within the soil cover under rainfall. The results are provided both in deterministic form, as a factor of safety (FS) map, and in stochastic form, as a probability of failure (PoF) map, the model's main output. The geotechnical, hydrological and probabilistic components of the model are described, in that order, in the following paragraphs.

4.1.2 Geotechnical Component

The geotechnical component of FSLAM is the infinite slope theory, which the model adopts in the form derived in Section 2.2 (Equation 2.3), evaluating the factor of safety cell by cell (Medina et al., 2021). The geotechnical formulation is therefore the standard one, as already defined in Chapter 2 (including $C = C_s + C_r$ and the degree of saturation h/z); the distinctive feature of FSLAM lies in the hydrological determination of the water table height h (Section 4.1.3) and in the stochastic treatment of the parameters (Section 4.1.5).

4.1.3 Hydrological Model

The hydrological model governs the degree of saturation (h/z) by determining the water table height h that appears in the factor-of-safety equation. In FSLAM, this height is obtained as the sum of two contributions corresponding to distinct time scales:

$$h = h_a + h_e, \quad (4.1)$$

where h_a is associated with the antecedent rainfall and is estimated using a medium-to long-term lateral flow approach, whilst h_e is associated with the event rainfall and is estimated using a short-term vertical flow approach (Medina et al., 2021). The two contributions are described separately in the following paragraphs.

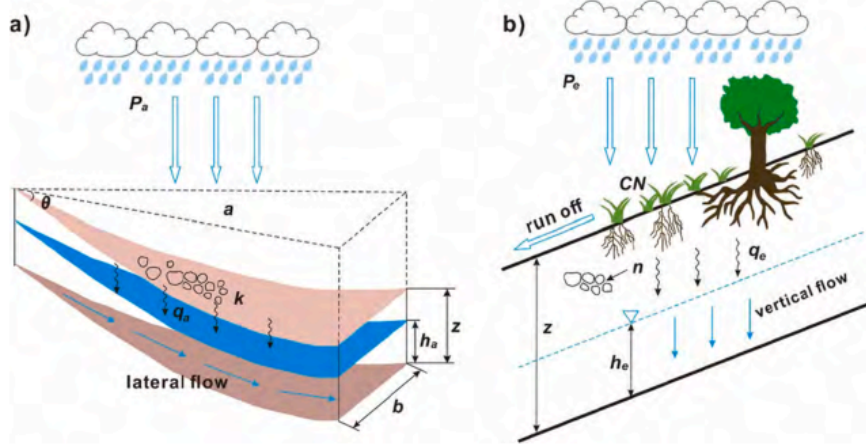


Figure 4.1: The two hydrological models incorporated in FSLAM. (a) The antecedent rainfall P_a feeds the lateral-flow component (Equation 4.2). (b) The event rainfall P_e feeds the vertical-flow component (Equation 4.3). From Medina et al. (2021).

Antecedent component (lateral flow). The term h_a represents the water table level determined by the precipitation preceding the event. The antecedent rainfall P_a is introduced as an input parameter and directly represents a long-term effective infiltration (mm/d), i.e. the fraction of precipitation that actually infiltrates into the soil, net of runoff and evapotranspiration. As no water balance model is available, in FSLAM the recharge q_a coincides with P_a and must therefore be provided in effective form from the available data (Medina et al., 2021). The value of h_a is computed according to the formulation of Montgomery & Dietrich (1994) (Equation 4.2), as a function of the drainage area a , the cell size b , the horizontal hydraulic conductivity K , the slope angle θ and the antecedent recharge q_a :

$$h_a = \left(\frac{a}{b}\right) \frac{q_a}{K \sin \theta \cos \theta}. \quad (4.2)$$

Event component (vertical flow). The term h_e represents the rise in the water table caused by the event rainfall P_e , calculated using a vertical flow approach and neglecting recharge from upstream areas (Figure 4.1): at the short-term timescale of the event, vertical flow is in fact predominant over lateral flow (Iverson, 2000).

Since a dynamic and transient description of infiltration in the unsaturated zone, based on Richards' equation, requires numerous parameters and is computationally intensive at the regional scale, FSLAM adopts a simplified static formulation: the model does not track the temporal evolution of the triggering process, but evaluates the final value of the factor of safety. Assuming the soil is close to saturation at the start of the event, infiltration is instantaneously translated into a rise in the water table through the soil porosity (Medina et al., 2021):

$$h_e = \frac{q_e}{n}. \quad (4.3)$$

Unlike the preceding component, in which P_a is already given as a long-term effective infiltration (mm/d, i.e. a rate), the event rainfall P_e is a raw precipitation depth expressed in millimetres (mm) that must be converted into infiltration q_e , also in mm. The conversion is carried out using the SCS-CN model (USDA Soil Conservation Ser-

vice, 1986), which estimates the surface runoff, and implicitly the infiltration, from a single parameter, the curve number (CN):

$$q_e = P_e - \frac{(P_e - (5080/CN - 51))^2}{P_e + 4(5080/CN - 51)}. \quad (4.4)$$

By substituting the contributions of the antecedent and event rainfall into the factor-of-safety equation (Equation 2.3), the final equation applied by the model to each cell of the study area is obtained:

$$FS = \frac{C}{g \rho_s z \cos \theta \sin \theta} + \left(1 - \left(\left(\frac{a}{b} \right) \frac{q_a}{K z \sin \theta \cos \theta} + \frac{q_e}{n z} \right) \left(\frac{\rho_w}{\rho_s} \right) \right) \frac{\tan \varphi}{\tan \theta}. \quad (4.5)$$

The result is therefore a map of the factor of safety in deterministic form or, when the stochastic variability of the parameters is taken into account, a probability of failure (PoF) map.

4.1.4 Structure and Required Inputs

FSLAM operates on a raster grid and computes its outputs cell by cell. The input data required by the model consist of five raster files and two parameter files, whose organisation is summarised in the flowchart of Figure 4.2.

The five rasters provide the spatially distributed information required by the model. From the DEM, the model derives, for each cell, the slope angle θ and the drainage area (flow accumulation) a . The SOIL raster identifies the soil classes and the LULC raster (land use and land cover) the land-use classes. Finally, the two rainfall rasters RAIN_ ANT and RAIN_ EVENT provide the antecedent rainfall P_a and the event rainfall P_e , respectively (Section 4.1.3).

The rasters are accompanied by two parameter files. The `soil.csv` file associates each soil class with its mechanical and hydrological parameters (soil cohesion C_s , friction angle φ , density ρ_s , hydraulic conductivity K , thickness z and porosity n), together with the hydrologic soil group (HSG). The two parameters that most influence stability, the cohesion C_s and the friction angle φ , are provided as ranges (minimum and maximum values) rather than as single values: this is the basis for the stochastic treatment described in Section 4.1.5. The `hmtu.csv` file, in turn, is linked to the LULC raster: it assigns the root cohesion C_r (also as a range) to each land-use class and merges the SOIL and LULC rasters into the Hydrological-Mechanical Terrain Units (HMTU), for which a combination matrix between the HSG and the land use defines the curve number CN used by the SCS-CN model (Section 4.1.3).

Taken together, these inputs provide all the parameters that appear in the model equations (Equations 4.2 – 4.5). The values actually adopted for the study area are presented in Section 4.2, whilst the deterministic and stochastic use of the parameters is described in the following section.

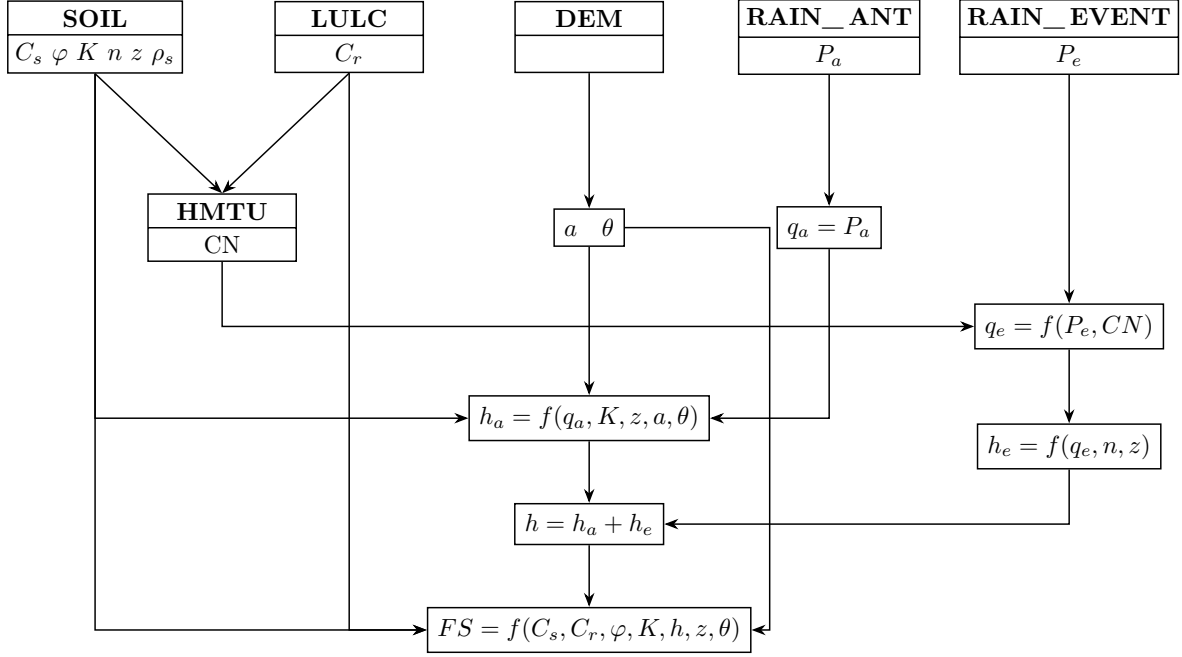


Figure 4.2: Flowchart of the FSLAM factor-of-safety computation: from the five input rasters and two parameter files to the hydrological and geotechnical equations. Adapted from Medina et al. (2021).

4.1.5 Deterministic / Stochastic Approach

The model employs two different approaches, deterministic and stochastic. In deterministic mode, the factor of safety is computed for each cell according to Equation 4.5, using the mean values of the ranges provided as input (Section 4.1.4); the result is a spatially distributed map of the factor of safety (Medina et al., 2021).

Since, at the regional scale, the determination of these parameters is subject to considerable uncertainty, they can be provided as ranges and treated as probability distributions. The stochastic treatment applies to the most sensitive parameters, the cohesions (soil and root) and the friction angle, whilst the others are treated deterministically (Section 4.1.4). Consequently, the factor of safety itself becomes a random variable and the model returns a map of the probability of failure, defined as the probability that the factor of safety is less than one, $PoF = P(FS < 1)$. This PoF map constitutes the main output of the model (Medina et al., 2021).

4.1.6 Output

In addition to the distinction between the deterministic and stochastic approaches described in the previous section, the model considers three different scenarios; Table 4.1 summarises its main outputs. The distinction between the scenarios reflects three different hydrological conditions in which the slopes may be found (Medina et al., 2021):

- dry condition: in the absence of a groundwater table, with zero pore pressures, stability is governed solely by the geotechnical model (§4.1.2);
- initial condition: the water table height is determined solely by the antecedent rainfall (P_a) and represents the state of the slopes before the event;

- final condition: both hydrological components (P_a and P_e) determine the water table height and, together with the geotechnical model, the stability of the slopes; this is the most critical condition.

Table 4.1: Outputs of the FSLAM model under the three progressive conditions.

Condition	Hydrological tribution	Con-	Factor of safety	Probability of failure
Dry	absent		✓	—
Initial (pre-event)	P_a		✓	✓
Final (post-event)	$P_a + P_e$		✓	✓

In addition to the scenarios involving rainfall, the model identifies cells whose stability is independent of the hydrological conditions, classifying them as unconditionally stable and unconditionally unstable (Medina et al., 2021). A cell is unconditionally unstable when it is already unstable under dry conditions, that is, when the shear strength alone is not sufficient to ensure its equilibrium; in this case, no rainfall is required to trigger failure. Conversely, a cell is unconditionally stable when it does not reach failure even under conditions of maximum saturation ($h/z = 1$), remaining stable even under the most intense rainfall event.

Since the model operates in stochastic form, both classes are returned as probability maps. In particular, the probability that a cell is unconditionally unstable coincides with the probability of failure under dry conditions, $P(FS_{dry} < 1)$, that is, the stochastic counterpart of the deterministic solution under dry conditions.

During the calculation, the model also determines the slope angle, the contributing area, runoff, infiltration and the water table height within the soil cover, expressed by the degree of saturation h/z (Section 2.2), the latter being calculated for both the initial and final conditions. These quantities are returned as spatially distributed rasters, allowing each factor to be examined individually across the study area.

4.1.7 Modifications

In addition to the base version of the code, described in the previous sections, two modifications to the model, developed by the authors of FSLAM and not yet described in any publication, were employed in the present work. The first introduces a length threshold on the flow accumulation, which affects the calculation of the antecedent component of the hydrological model; the second, implemented by the authors upon specific request for this work, allows the soil thickness to be provided as a spatially distributed raster, rather than as a constant value per soil class. Each of the two modifications was evaluated as a distinct model configuration, subject to its own calibration (Section 4.4). The two are not independent, however: the distributed-thickness configuration also includes the accumulation-length threshold (Section 4.4.4).

Accumulation-length threshold. In the original formulation of the model, the term a/b is derived from the flow accumulation calculated on the DEM using a D8 algorithm (Guo et al., 2022). For each cell, a represents the total drainage area, that

is, the set of all cells that drain into it, up to the divide. As noted in Equation 4.2 (Section 4.1.3), it is assumed that all the antecedent recharge q_a falling on the drainage area reaches the cell, without any distance limit. The drainage area increases downstream, reaching its maximum values at hollows and valley floors (Guo et al., 2022).

However, this assumption is unrealistic at the timescale of the antecedent condition, particularly in the presence of low-permeability lithologies. The distance that water infiltrating the subsoil can actually travel is limited. In the original formulation, unlimited accumulation therefore tends to overestimate h_a in cells characterised by a large drainage area, which are systematically close to saturation and, consequently, unstable.

The modification introduces a length threshold on the accumulation: a value, expressed in metres, which limits the distance, measured along the drainage path, within which upstream cells contribute to the drainage area. Consequently, when calculating a for a given cell, only those cells whose flow path towards it does not exceed the threshold are taken into account, and the antecedent recharge is accumulated over the same area. The accumulation computed with this limit is referred to below as truncated. The threshold is assigned per soil class, via an additional column (`acc_length`) in the `soil.csv` file: this allows the accumulation length to be differentiated according to soil properties. The modification affects only the antecedent component h_a ; the event component h_e , based on a purely vertical flow, remains unchanged. Finally, it should be noted that the calculation of the truncated accumulation, which is iterative in nature, entails a computational cost that increases with the threshold value; already in the original formulation, the calculation of the flow accumulation constitutes the most demanding geoprocess of the entire model (Guo et al., 2022).

The threshold values adopted are given in Section 4.4.4, together with the reasons for them; the resulting configurations are presented in Chapter 5.

Distributed soil thickness. In the base version of the model, the soil thickness z is a constant parameter for each soil class, assigned via the relevant column in the `soil.csv` file (Section 4.1.4). In the case under consideration, where the domain is almost entirely occupied by a single litho-technical unit (Section 4.2.2), this results in a virtually uniform thickness across the entire study area.

However, thickness exerts a strong influence on stability: it appears in both terms of Equation 4.5, in the denominator of the cohesive contribution and in the degree of saturation h/z . Catani et al. (2010) identify it as one of the most significant parameters in determining the stability of the soil cover, particularly for thicknesses of less than 1.5 m, within which small variations produce very rapid changes in FS, and observe that it varies continuously as a function of hillslope morphology. A constant value per soil class is unable to represent that variability.

The modification allows the thickness to be supplied to the model as a spatially distributed raster: a sixth input raster (`z.asc`), which is added to the five described in Section 4.1.4. When present, it replaces the constant class value on a cell-by-cell basis; all other parameters remain assigned via `soil.csv`. The construction of the raster used in this work, obtained with the GIST model (Catani et al., 2010), is described in Section 4.2.5.

4.2 Input Data and Preparation

The study area, defined geographically in Section 3.1 and geologically in Section 3.2, is the municipality of Casola Valsenio. This section describes, for each of the five raster inputs required by the base version of the model (Section 4.1.4), the data sources and the processing carried out. The two parameter files (`soil.csv` and `hmtu.csv`), which assign mechanical and hydrological parameters to soil and land-use classes, are instead presented later in relation to their calibration (Chapter 5).

Soil thickness, which is normally assigned as a constant value per soil class via `soil.csv`, can also be provided as a sixth raster input in the distributed-thickness variant of the model (Section 4.1.7); the construction of this raster is described in Section 4.2.5.

All layers were standardised to a common resolution of 10 m, matching that of the digital elevation model (Section 4.2.1) used as the reference grid, and projected onto a single reference system (ETRS89 / UTM Zone 32N, EPSG:25832). The analyses were carried out in a GIS environment (QGIS) and in an R environment (RStudio).

The calculation domain was not limited to the municipal boundary but was extended to a wider area comprising the complete catchments draining into the study area (Figure 4.3). This choice is dictated by the hydrological calculations: since the lateral flow contribution depends on the catchment area (flow accumulation) of each cell (Section 4.1.3), cropping the rasters to the municipal boundary alone would truncate the catchment area of the boundary cells, thereby distorting the result. By extending the domain to the complete catchments, the flow accumulation is calculated correctly for every cell in the study area, to which the model results are subsequently cropped.

However, the extension of the domain concerns only the DEM. As flow accumulation is a purely topographical quantity, it is the DEM that must cover the entire extended domain; the other four inputs, which are used in the cell-by-cell calculation, are required only within the study area, since the results are cropped to this area and their values outside it do not influence the outcome.

The domain thus extended also straddles the administrative boundary between Emilia-Romagna and Tuscany. As no single regional DEM covers the entire domain on its own, the DEM was created by mosaicking the models provided by the two regions into a single continuous surface (Section 4.2.1). It should be noted that the study area lies entirely on the Emilia-Romagna side, whilst the Tuscan portion constitutes only the upstream extension necessary to enclose the catchments: the data from the Emilia-Romagna Region therefore represent the primary source, whilst the Tuscan data serve solely to complete the DEM.

4.2.1 Digital Elevation Model (DEM)

The Digital Elevation Model (DEM) forms the topographical basis of the analysis and was derived by combining two sources (Section 4.2): the Emilia-Romagna Region's DEM, with a resolution of 10 m, derived from the 1:5,000-scale CTR (2015 edition), and the Tuscany Region's DEM, also with a resolution of 10 m, derived from the 1:10,000-scale CTR (surveys from 1993–1998). Both models are derived from the digitisation of the regional technical maps and predate the May 2023 event: the topographic surface used therefore represents the morphology prior to the failures.

Since the two sources share a native resolution of 10 m, the mosaicking process

preserves the resolution of the reference grid, without the need for resampling. The mosaicked surface was then processed using the geomorphometric algorithms of WhiteboxTools (v2.4.3) (Wu & Brown, 2025), run in the R environment. Firstly, depression filling removed local minima, ensuring a continuous drainage network and thus a correct calculation of flow accumulation. Starting from the surface thus processed, the calculation domain was defined as the set of catchments draining through the study area (Figure 4.3), and the DEM was cropped to this domain.

The DEM obtained in this way serves as the 10 m reference grid for the entire procedure, to which all other input layers are aligned (Section 4.2).

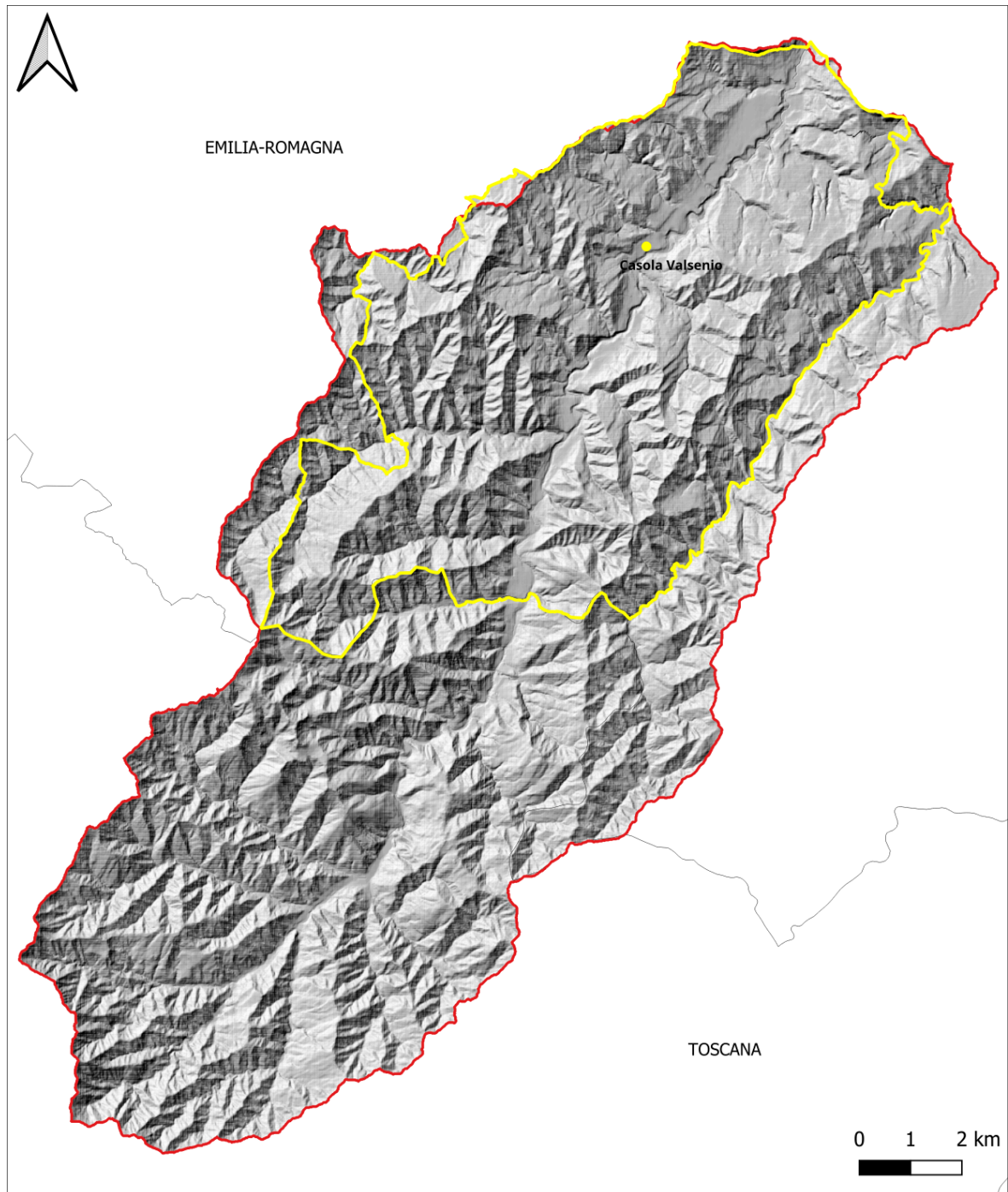


Figure 4.3: Extent of the calculation domain. The red line delimits the extended domain, comprising the complete catchments that drain through the study area, whilst the yellow line marks the municipal boundary of Casola Valsenio (the study area). The shaded relief is the mosaicked 10 m DEM (Section 4.2.1).

4.2.2 Soil Units

The SOIL spatial layer is a categorical raster that assigns each cell a litho-technical unit associated with the bedrock. The classification is the one adopted by the RER2023 inventory (Berti et al., 2025), which groups the more than 600 geological formations distinguished at regional level into eight litho-technical units, on the basis of their

lithological composition and structural domain (Table 4.2); the underlying cartography is the Geological Map of the Emilia-Romagna Region (1:10 000 scale). The layer was adopted directly from RER2023, which also ensures consistency with the landslide inventory on which Section 3.5 is based.

In the absence of detailed mapping of the overburden, the soil classes are derived from the bedrock lithology, on the assumption that the geotechnical behaviour of the soil cover is inherited from the parent material. This is the same rationale underlying the RER2023 classification, whose coarse-grained units (classes 4–8) produce granular, debris-type soils and whose fine-grained units (classes 1–3) produce fine, earth-type soils (Berti et al., 2025).

Of the eight units of the regional classification, only three outcrop within the Casola Valsenio study area (Figure 4.4): the Marnoso-Arenacea Formation (class 7), largely dominant and to which the calibration is restricted (Section 4.4); the clay shales, tectonized clays and olistostromes (class 3), corresponding to the intraformational chaotic bodies described in Section 3.2; and the massive-rock class (class 4), represented in the area exclusively by the Gessoso-Solfifera evaporites (Section 3.2). The remaining five units are not present.

The layer was therefore rasterised onto the 10 m reference grid and cropped to the calculation domain (Section 4.2); outside the study area, consistent with what was stated in the introduction, the value of the layer does not affect the result.

The SOIL raster identifies the units to which the `soil.csv` file assigns mechanical and hydrological parameters and, together with the LULC layer, helps to define the Hydrological-Mechanical Terrain Units (HMTU; Section 4.1.4). The parameter values for each unit are presented alongside the calibration results (Chapter 5).

Table 4.2: Litho-technical units (bedrock-based) of the RER2023 inventory (Berti et al., 2025), adopted for the SOIL layer; underlying cartography: Geological Map of the Emilia-Romagna Region (1:10 000 scale).

Class	Description
1	Clays, silty clays, and marly clays
2	Marls and marly clays
3	Clay shales, clay breccias, tectonized clays, olistostromes
4	Massive rocks: basalts, serpentines, limestones, arenites
5	Flysch rocks made of rhythmic alternations of sandstones, limestones, pelites, and shales
6	Weakly cemented sandstones and conglomerates
7	Flysch rocks made of rhythmic alternations of sandstones and pelites
8	Weakly cemented sandstones with interbedded pelitic layers

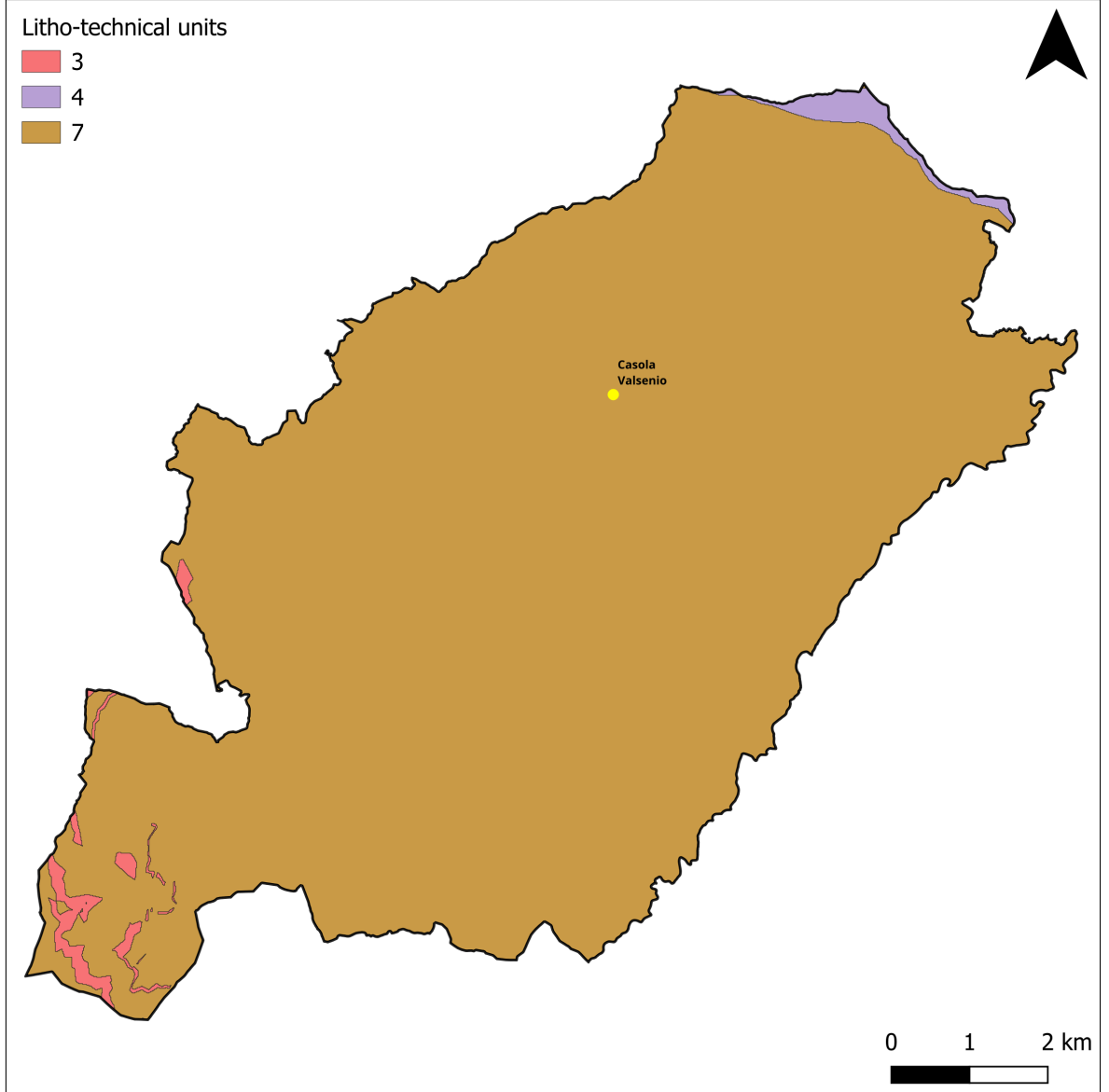


Figure 4.4: Litho-technical units of the study area, after the RER2023 classification (Table 4.2).

4.2.3 Land Use and Land Cover Classes

The LULC spatial layer is a categorical raster that assigns a land use and land cover class to each cell. The base map is the SU2014 dataset (Land Use in the Emilia-Romagna Region) (Regione Emilia-Romagna, 2018), classified according to the Corine Land Cover directive into twelve classes, as adopted in the RER2023 inventory (Berti et al., 2025). Land use and land cover influence the model in two ways (Section 4.1.4): vegetation provides root cohesion (C_r), which contributes to the resistance of the soil cover, whilst land use controls the partitioning between infiltration and runoff, and thus the curve number (CN).

However, the model parameter file (`hmtu.csv`) follows the formulation of Medina et al. (2021), which relates root cohesion and the curve number to a different set of land cover classes (*forest, shrubs, grassland, bare rock, scree, bare soil, farmland, water, urban area*). The twelve Corine classes have therefore been reclassified into these

classes, according to the correspondence shown in Table 4.3. In the study area (Figure 4.5), land cover is clearly dominated by the *forest* and *farmland* classes, with a minor presence of *shrubs* and, along the Senio valley floor, of *water*; the *scree* class is not present.

The reclassified layer was then rasterised onto the 10 m reference grid and cropped to the calculation domain (Section 4.2).

The LULC raster identifies the land cover classes to which the `hmtu.csv` file assigns parameters; the values for each class are presented alongside the calibration (Chapter 5).

Table 4.3: Correspondence between the Corine land-cover classes of the SU2014 dataset (Regione Emilia-Romagna, 2018) and the model land-cover classes (Medina et al., 2021), used to reclassify the LULC layer.

Class	Model land cover classes	Corine classes (SU2014)
1	Forest	Forested areas (311–313)
2	Shrubs	Shrub and/or herbaceous areas (321–323)
3	Grassland	Permanent grasslands (231)
4	Bare rock	Extraction areas, quarries, artificial soils (131–133)
5	Scree	—
6	Bare soil	Areas with sparse or no vegetation (331–334)
7	Farmland	Arable land (211–213); vineyards, orchards, tree crops (221–224); heterogeneous agricultural areas (241–243)
8	Water	Water bodies (511–521); wetlands (411–422)
9	Urban area	Urban areas (111–124); artificial green areas, parks (141–143)

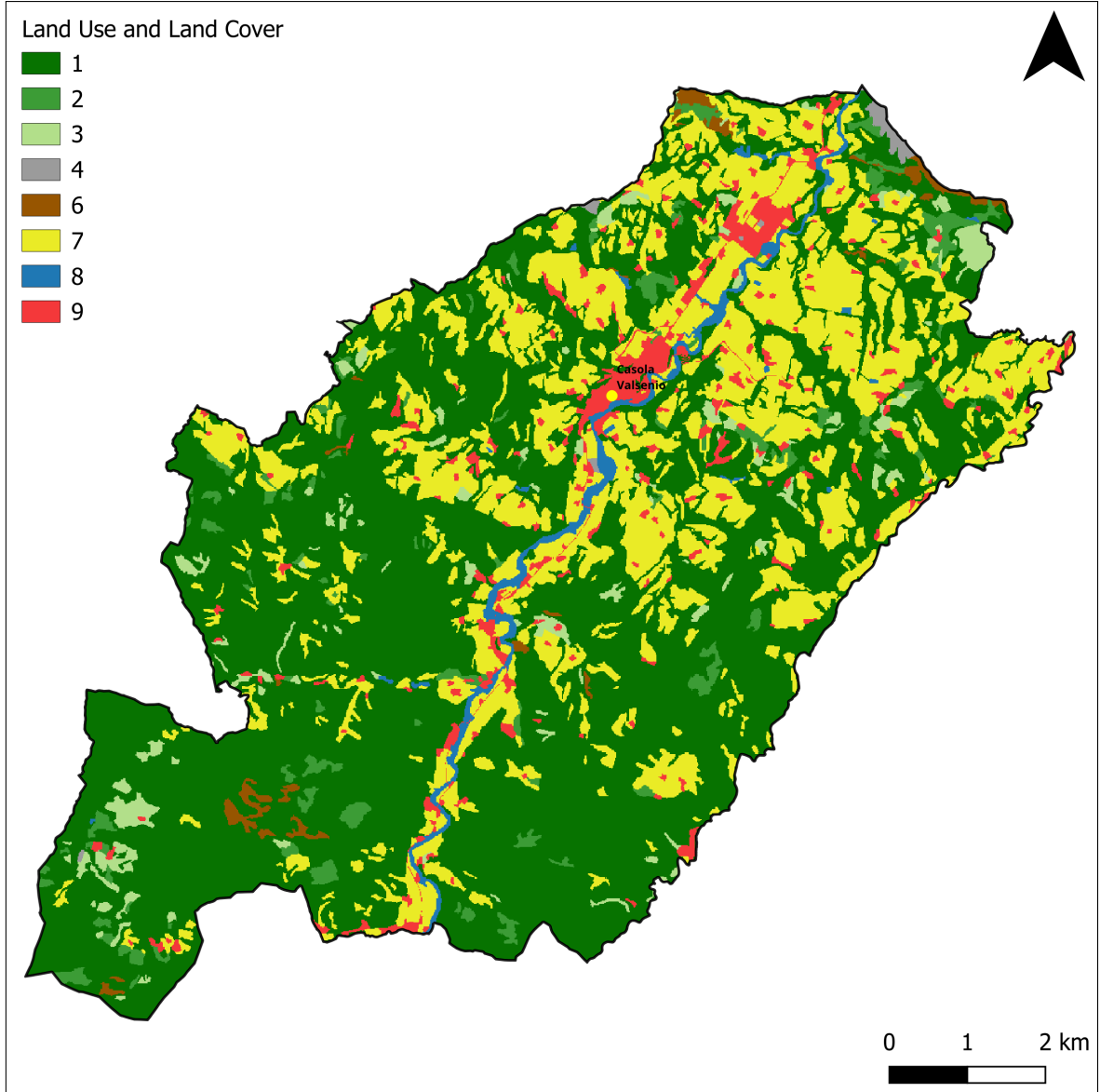


Figure 4.5: Land use and land cover of the study area, reclassified into the model land-cover classes (Table 4.3).

4.2.4 Rainfall

The model's two rainfall inputs, the antecedent rainfall (RAIN_ANT , P_a) and the event rainfall (RAIN_EVENT , P_e), were derived by spatially interpolating the data recorded by the rain-gauge stations of the RIRER regional network, managed by Arpae-SIMC and extracted via the Dext3r portal (Arpae-SIMC, 2023). Of the 264 stations available at the regional level, 241 were used, after excluding those with missing data.

The interpolation was carried out in the R environment using the `gstat` package (v2.1-5) (Pebesma, 2004), via universal kriging with elevation as external drift, derived from the DEM (Section 4.2.1). The quality of the interpolation was assessed by cross-validation: for the event cumulative rainfall, universal kriging reduces the RMSE from 39.3 to 36.4 mm compared with ordinary kriging, which justifies its adoption. The interpolation was performed at the regional scale, and the resulting rasters were cropped to the calculation domain and resampled to the 10 m reference grid (Section 4.2). The

precipitation distribution maps presented in Section 3.4 for descriptive purposes were obtained using ordinary kriging; the model inputs, however, use universal kriging, with virtually equivalent results.

The antecedent rainfall RAIN_ANT represents the effective recharge resulting from the rainfall preceding the event and, as discussed in Section 4.1.3, must be provided as effective infiltration (mm/d) (Medina et al., 2021). It was derived from the rainfall recorded between 24 and 30 April 2023, immediately prior to the event (Section 3.4), by applying an effective infiltration coefficient of 30%. This value was adopted following the indication of the model authors, based on their previous applications (M. Hürli-mann, personal communication). It is also consistent with the infiltration coefficients reported in the literature for the flysch of the northern Apennines, ranging between 0.4 and 0.7 of the effective precipitation (Civita 2005, as reported by Cervi et al. (2012)), once the additional fraction lost to surface runoff has been taken into account. The adoption of a fixed coefficient is nonetheless a simplification: FSLAM incorporates no water balance model (Section 4.1.3), so the antecedent recharge must be supplied already in effective form. The resulting raster provides the value of P_a which feeds the model's antecedent lateral-flow component.

The event rainfall RAIN_EVENT represents the raw cumulative rainfall (mm) recorded over the period 1–17 May 2023. This window coincides with the one adopted in Section 3.4.4 for the cumulative-rainfall and return-period analysis, and it encompasses both meteorological pulses of the event. The choice to treat the whole period as a single triggering rainfall, rather than the two pulses separately, is dictated by the nature of the RER2023 inventory on which the model is calibrated and evaluated: the landslides were mapped from imagery acquired after the end of the event, with no survey of the intervening period (Section 3.5), so that individual landslides cannot be attributed to the first or the second pulse. The raster therefore provides the value of P_e , which the model converts into event infiltration q_e through the SCS-CN model (Section 4.1.3).

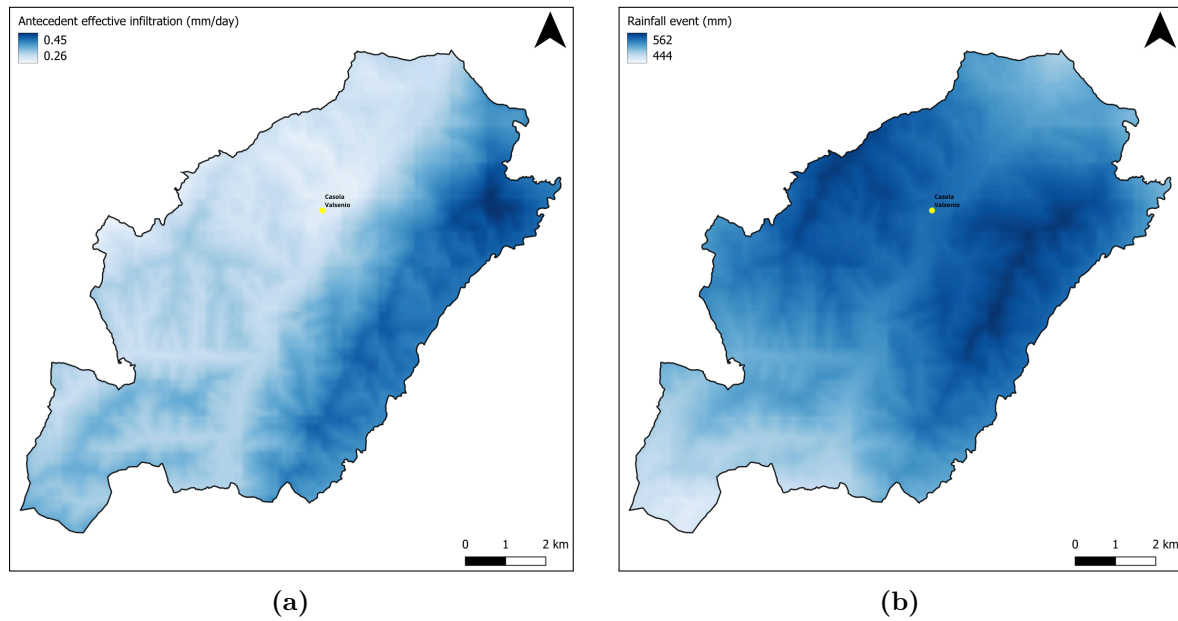


Figure 4.6: Rainfall inputs over the study area. (a) Antecedent effective infiltration (P_a , mm/d) for the period 24–30 April 2023, with the 30% effective-infiltration coefficient applied. (b) Event rainfall (P_e , mm) cumulated over 1–17 May 2023.

4.2.5 Soil Thickness

In the model's base configuration, the soil thickness z is a constant parameter, assigned by soil class (Section 4.1.4). In the variant with distributed thickness (Section 4.1.7), it is instead provided as a raster. The spatial distribution of soil thickness was estimated using the GIST (*Geomorphologically Indexed Soil Thickness*) model of Catani et al. (2010), an empirical approach that derives soil thickness at catchment scale from geomorphometric attributes of the terrain. Its structure was retained, while the definition of the individual factors was adapted to the information available for the study area.

In the original formulation, thickness is the product of four normalised factors (Catani et al., 2010):

$$z = C \cdot P \cdot S \cdot K_c, \quad (4.6)$$

where C depends on profile curvature, P on hillslope position, S on slope gradient, and K_c is the maximum thickness. Following Gupta et al. (2024), the position factor P was omitted, as no clear correlation between thickness and hillslope position emerges in complex terrains; the adopted formula therefore reduces to:

$$z = C \cdot S \cdot K_c. \quad (4.7)$$

The factor S expresses the control of slope on thickness. As the gradient increases, erosion and mass transport progressively remove the soil cover, until the substrate is exposed (Catani et al., 2010; Gupta et al., 2024). It is defined by two thresholds (Gupta et al., 2024):

$$S = \begin{cases} 1 & \beta \leq \varphi \\ \frac{1}{1 + \tan \beta} & \varphi < \beta \leq \theta \\ 0 & \beta > \theta \end{cases} \quad (4.8)$$

where β is the local slope and φ and θ are the lower and upper thresholds: below φ the soil cover retains its full thickness ($S = 1$); between φ and θ the thickness decreases as the slope increases; above θ the substrate is exposed and the thickness becomes zero ($S = 0$).

The factor C , in turn, expresses the control of curvature on the thickness of the soil cover. On mature hillslopes, local soil thickness is inversely correlated with profile curvature. Convex forms, subject to erosion, develop thinner covers, whereas concave forms, where loose material accumulates at the base, retain thicker ones (Catani et al., 2010; Gupta et al., 2024). In the original formulation, the factor is obtained by linearly rescaling the profile curvature (c_p) between 0 and 1 over the whole range of values (Catani et al., 2010). Gupta et al. (2024) restrict the linear correlation to a symmetric central interval. This interval is defined from the curvature distribution of their own study area, and the extreme values are excluded as anomalous. Pixels outside the interval are assigned the limit values: 0 for convex pixels and 1 for concave ones.

Both GIST and the version of Gupta et al. (2024) calibrate these factors using direct field data (borehole thickness measurements for K_c , shear tests for φ , and in-situ surveys for θ), which are not available in this case. The only information that could be used is the DEM (Section 4.2.1), from which curvature and slope are derived, and the inventory of observed initiation points (Section 3.5).

In their absence, the construction was constrained by a physical criterion. Where a shallow landslide was triggered, a soil cover was present; the map must therefore

not assign zero thickness to that point. Since $K_c > 0$, Equation 4.7 shows that the thickness vanishes only if $C = 0$ or $S = 0$. At every observed initiation point both factors must therefore be positive. Denoting by β_{max} the maximum observed initiation slope:

$$\theta \geq \beta_{max}, \quad \min C > 0. \quad (4.9)$$

The first condition requires the upper threshold of the factor S (Equation 4.8) to be no lower than the maximum observed initiation slope. The second satisfies the requirement on C by imposing a positive lower bound over the whole domain, rather than by moving its threshold beyond the most convex curvature values. These are isolated extremes, and a threshold placed beyond them would compress the normalisation over the rest of the domain. The criterion therefore acts on these two limits alone. No thickness value is interpolated from the observed landslides, and the geometry of the map follows the curvature and the slope of the DEM. The inventory nonetheless sets the limits within which those two variables act, so the map is not independent of the dataset later used as the target of the validation.

The two thresholds of the factor S (Equation 4.8) were derived from the slope distribution at the initiation points of the inventory (Figure 4.7). The range in which the observed initiations are concentrated was assumed to approximate the range within which slope controls the thickness of the soil cover.

The upper threshold is set to $\theta = 60^\circ$, the first integer value above the maximum observed initiation slope (58.9°). The criterion would also be satisfied by $\theta = \beta_{max}$, but the inventory constrains this threshold from below without determining it. Above the threshold the thickness vanishes; by construction, no initiation point of the inventory falls in that range.

The lower threshold, by contrast, is not constrained by the criterion. Below it $S = 1$, the soil cover is fully retained, and no initiation point can therefore be cancelled. A percentile of the distribution is sufficient here, rather than its extreme value. The 5th percentile was adopted, $\varphi = 15.9^\circ$, which delimits the range in which the observed initiations are concentrated. The threshold retains the symbol of the original formulation but not its meaning: it replaces the internal friction angle, which cannot be measured here, with a topographic criterion.

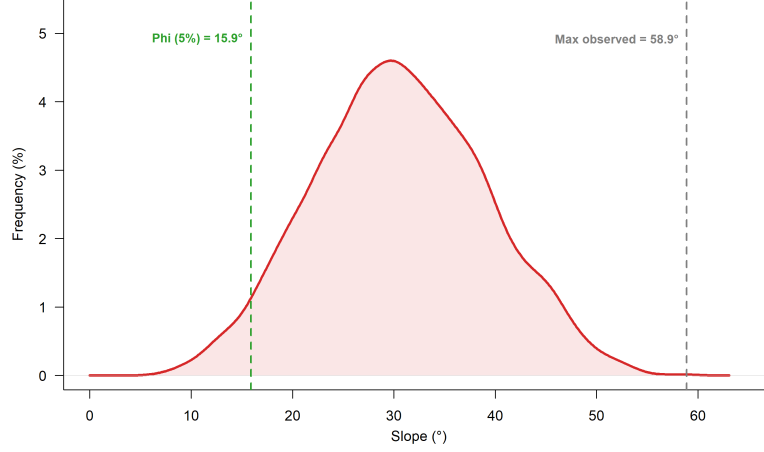


Figure 4.7: Slope distribution at the RER2023 landslide initiation points. The dashed lines mark the 5th percentile and the maximum observed initiation slope.

The profile curvature was computed on the DEM (Section 4.2.1). The numerical value with which Gupta et al. (2024) delimit the interval is not transferable, since profile curvature depends on the units adopted and on the local morphology. The threshold was therefore defined by setting t equal to the standard deviation of the local distribution ($t = 0.007$), within which 79.8% of the pixels of the domain fall.

$$C_{norm} = \begin{cases} 1 & c_p \leq -t \\ \frac{t - c_p}{2t} & -t < c_p < t \\ 0 & c_p \geq t \end{cases} \quad C = c_{min} + (1 - c_{min}) C_{norm}. \quad (4.10)$$

Without a positive lower bound, the factor C would vanish on the most convex sectors. The map would then assign zero thickness to 15% of the observed initiation points (751 out of 4948), which fall on cells with $c_p \geq t$. A value $c_{min} = 0.2$ was adopted: convex sectors and ridges retain one fifth of the thickness assigned to concave sectors.

In the original formulation (Equation 4.6), K_c is the maximum thickness. Being a scalar, however, it determines only the magnitude of the thickness and not its spatial distribution, which depends on C and S alone. The two components were therefore separated: the spatial distribution is left to GIST, while the scale is treated as a calibration parameter (Section 4.4). To this end, the map is normalised with respect to its own mean:

$$z_{norm}(\mathbf{x}) = \frac{C(\mathbf{x}) S(\mathbf{x})}{\overline{C \cdot S}}, \quad \overline{C \cdot S} = \frac{1}{N} \sum_{j=1}^N C(\mathbf{x}_j) S(\mathbf{x}_j), \quad (4.11)$$

where the mean is computed over the whole calculation domain (Section 4.2). The quantity z_{norm} is dimensionless and has unit mean by construction. The thickness in metres is obtained by multiplying it by a scale factor k , which replaces K_c and is determined during calibration (Section 4.4).

The resulting normalised map is shown in Figure 4.8 and its statistics are reported in Table 4.4. The thickest covers are found in concave sectors and on gently sloping

hillsides, the thinnest on ridges and on steep slopes. The cells with zero thickness, which can therefore be regarded as exposed substrate, are 279 (0.01% of the domain) and all lie above 60° : they thus derive from the factor S alone. At the observed initiation points the minimum value of z_{norm} is 0.218: no initiation point falls on cells without a soil cover, and the criterion is satisfied with a clear margin.

The thickness map in metres, obtained with the calibrated scale factor, is presented in Section 5.4.

Table 4.4: Summary statistics of the normalised map z_{norm} over the calculation domain. Values are dimensionless.

Statistic	Value (–)
Minimum	0.000
1st quartile	0.652
Median	0.960
Mean	1.000
3rd quartile	1.313
Maximum	2.266

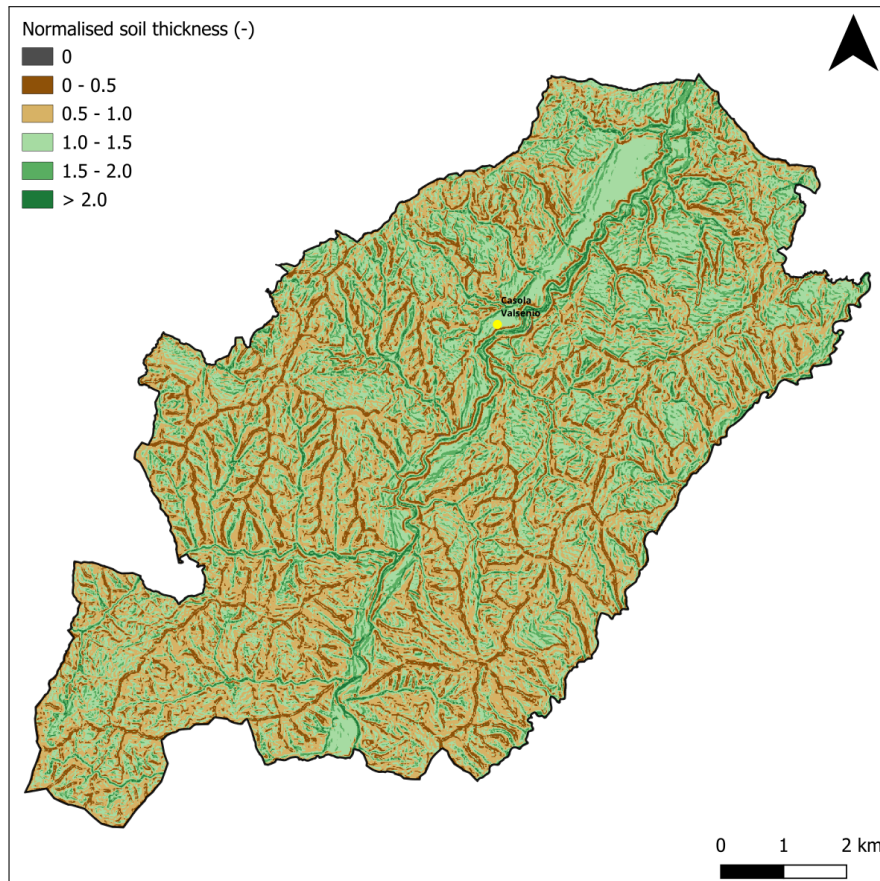


Figure 4.8: Normalised soil-thickness map z_{norm} obtained with the GIST model (Catani et al., 2010).

4.3 Validation Framework and Criteria

The FSLAM model produces a probability of failure (PoF) map as its main output (Section 4.1.6). To assess its predictive capability, this output was compared with the RER2023 inventory (Section 3.5) using a binary classification scheme: each point in the validation dataset is associated with an observed condition (landslide / non-landslide), which is compared with the class predicted by the model based on the PoF value. Performance is quantified using the confusion matrix and the ROC curve, standard tools for validating landslide susceptibility (Reichenbach et al., 2018) and previously employed for FSLAM (Medina et al., 2021; Guo et al., 2022).

4.3.1 Validation Dataset

The dataset used for validation is the same as that used for calibration (Section 4.4): it was not split into independent calibration and validation sets. Guzzetti et al. (2006) note that the quality of a susceptibility model can be ascertained either with the same landslide data used to obtain the estimate, or with independent landslide information not available when the model was built. The first strategy serves to evaluate the role of the input information and the ability of the model to cope with variations in it; the second serves to determine the prediction skill on new landslides. The design adopted here corresponds to the first. This choice is consistent with the aim of the analysis, which is the relative comparison of different model configurations under identical conditions, rather than the estimation of an absolute value of predictive accuracy. Since all configurations are evaluated on the same dataset, their comparison remains valid; in the absence of an independent test set, the absolute values of the metrics should therefore be interpreted primarily as comparisons between the configurations.

The validation is based on a dataset of points specifically constructed from the RER2023 inventory, balanced between landslide and non-landslide points (4,948 for each class). The entire procedure (Figure 4.9) was implemented in R to ensure its reproducibility.

Both classes of points are confined to the same suitability domain, defined by the intersection of three criteria, so that landslide and non-landslide points are directly comparable. The first criterion is lithological: only cells belonging to the Marnoso-Arenacea Formation (unit 7; Section 4.2.2) are considered. This is the largely dominant litho-technical unit and the only one for which the model has been calibrated. The second criterion is topographical: cells with a slope of less than 10° , where shallow landslides cannot be triggered, are excluded; the threshold lies well below the range of slopes at which the observed initiations are concentrated (Section 4.2.5, Figure 4.7) and therefore excludes no actual landslides. The third criterion concerns land use: only the classes in which the triggering of shallow landslides is plausible (forest, shrubs, grassland, bare soil and farmland) are retained, whereas the classes water (where the recorded events are bank collapses, not attributable to the modelled process), urban (human-modified areas where the model does not apply) and bare rock (lacking a soil mantle; the only area so classified is a forested quarry slope, with no mapped initiations) are excluded.

Landslide points. The landslide points are derived from the initiation points mapped in the RER2023 inventory (Section 3.5), selecting only debris slides and de-

bris flows: these are the shallow landslide types that constitute the target of FSLAM and which represent the predominant part of the phenomena triggered by the event (Berti et al., 2025). A 2.5 m buffer was applied to each initiation point falling within the suitability domain, generating a polygon representative of the detachment area. This choice takes into account the average size of the detachment areas, estimated through visual analysis of the landslide polygons, and the cell size (10 m). For each polygon, the PoF value assumed is the maximum among the intersected cells, so as to represent the most critical condition within the initiation area.

Non-landslide points. The same number of non-landslide points as landslide points was generated, in order to obtain a balanced dataset. They were randomly sampled within the same suitability domain, with the additional exclusion of all landslide polygons from the RER2023 inventory: in this way, they fall within stable terrain, unaffected by any mapped instability. Unlike the landslide points, no buffer was applied, as these single points do not represent a physical failure area, and the PoF value is that of the individual cell in which each point lies.

The final dataset therefore consists of 4,948 landslide points and 4,948 non-landslide points.

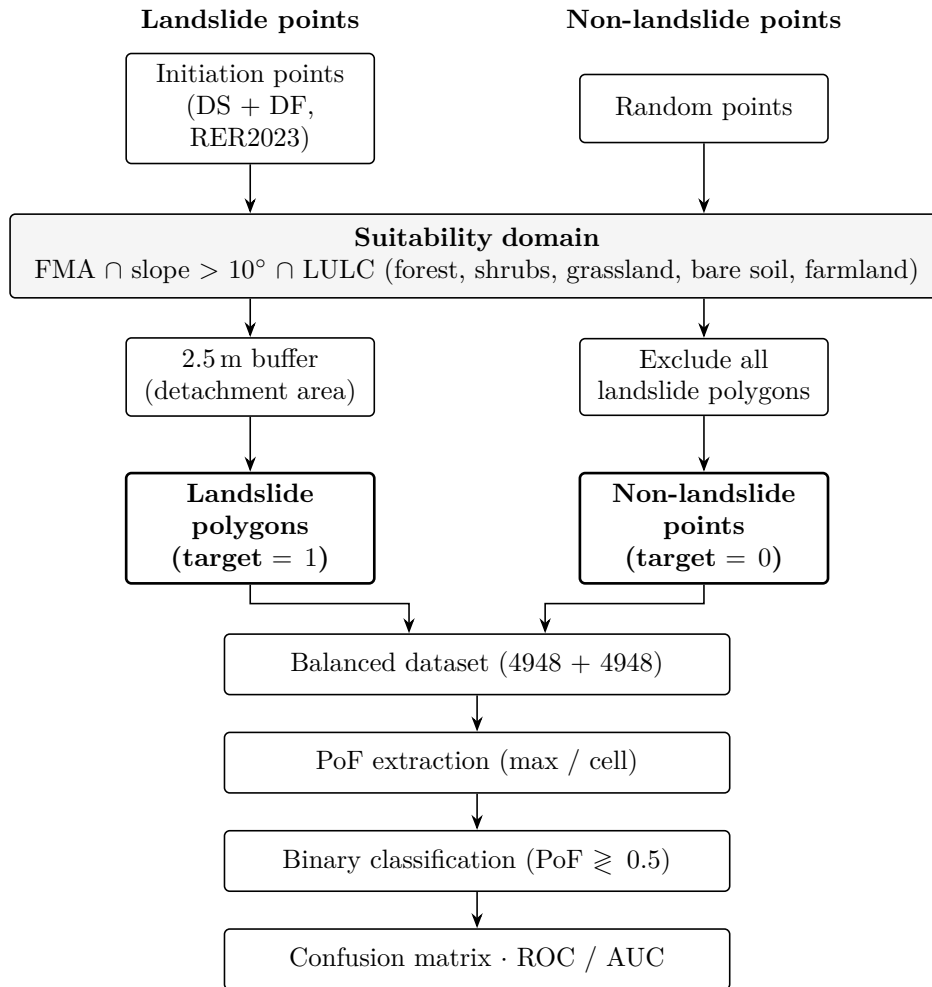


Figure 4.9: Workflow of the validation dataset construction and evaluation. The individual selection criteria are detailed in the text above.

4.3.2 Performance Metrics

Each point is assigned a predicted class by comparing the PoF value with a threshold of 0.5: the point is classified as predicted unstable if $\text{PoF} > 0.5$, and as predicted stable otherwise. Since the PoF is defined as the probability that the factor of safety is less than one (Section 4.1.5), a threshold of 0.5 identifies the cells in which more than half of the combinations of the parameters treated stochastically give a factor of safety less than one. Equivalently, it is the value at which the median of the distribution of the factor of safety equals one. The authors of the model adopt the same threshold, on the same grounds (Medina et al., 2021). The threshold was fixed in advance and was not optimised on the metrics. Its effect on the comparison between the configurations is examined in Section 5.5.

The comparison between the observed and predicted classes, extended to the entire dataset, is summarised in the confusion matrix (Figure 4.10), which divides the points into four categories. True positives (TP) are landslide points correctly predicted as unstable, and true negatives (TN) are non-landslide points correctly predicted as stable; false positives (FP) are non-landslide points incorrectly predicted as unstable, and false negatives (FN) are landslide points incorrectly predicted as stable. The first two categories represent correct classifications, whilst the other two represent the model's errors.

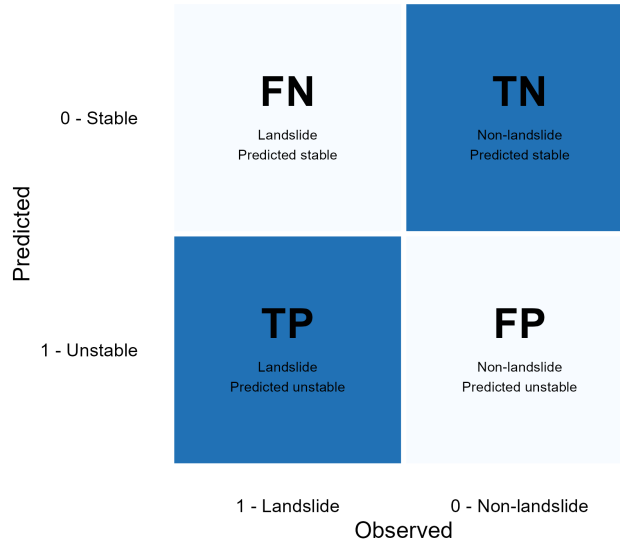


Figure 4.10: Structure of the confusion matrix adopted for the validation.

This breakdown gives rise to the two rates that define the ROC space. The true positive rate (TPR) is the proportion of landslides correctly identified, i.e. the model's ability to recognise unstable slopes:

$$\text{TPR} = \frac{TP}{TP + FN}. \quad (4.12)$$

The false positive rate (FPR) is the proportion of stable points erroneously classified as unstable, i.e. the model's false alarms:

$$\text{FPR} = \frac{FP}{FP + TN}. \quad (4.13)$$

Since TPR and FPR depend on the threshold chosen, the ROC (*Receiver Operating Characteristic*) curve is used for an evaluation independent of the threshold; this curve plots TPR against FPR as the classification threshold varies (Figure 4.11). The confusion matrix, evaluated at the fixed threshold of 0.5, corresponds to a single point on this curve, whereas the full curve is obtained by varying the threshold across its entire range. A curve close to the diagonal indicates a model no better than random classification, whilst a curve approaching the top-left corner indicates almost perfect discrimination. The area under the curve (AUC) summarises this ability in a single value: it corresponds to the probability that, given a landslide point and a non-landslide point chosen at random, the model will assign the former a PoF higher than that of the latter, and ranges from 0.5 (random classification) to 1 (perfect discrimination). The AUC is the primary metric for comparing model configurations.

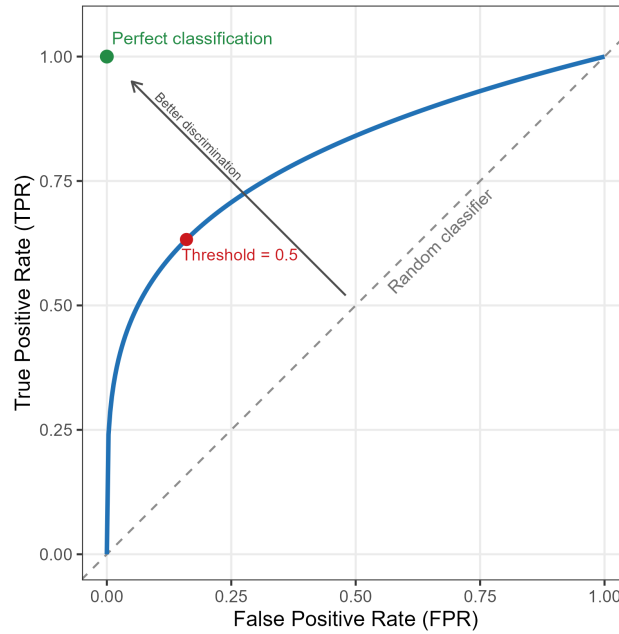


Figure 4.11: Schematic representation of the ROC space.

The validation is based on the confusion matrix, from which the true positive rate (TPR), the true negative rate (TNR) and the overall accuracy (ACC) are derived. These are complemented by the ROC curve and its area (AUC), for a threshold-independent evaluation, and by the PoF distribution as a diagnostic plot. The true negative rate, the complement of the false positive rate used for the ROC ($TNR = 1 - FPR$), is the proportion of stable points correctly identified:

$$TNR = \frac{TN}{TN + FP}, \quad (4.14)$$

while the accuracy is the overall proportion of correctly classified points:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (4.15)$$

The distribution of the PoF values (Figure 4.12) shows the frequency of PoF values separately for landslide and non-landslide points. A model with good discriminatory power concentrates landslide points towards high PoF values and non-landslide points

towards low values, with limited overlap between the two distributions; the degree of separation provides an immediate indication, complementary to the AUC, of the model's ability to distinguish between the two conditions.

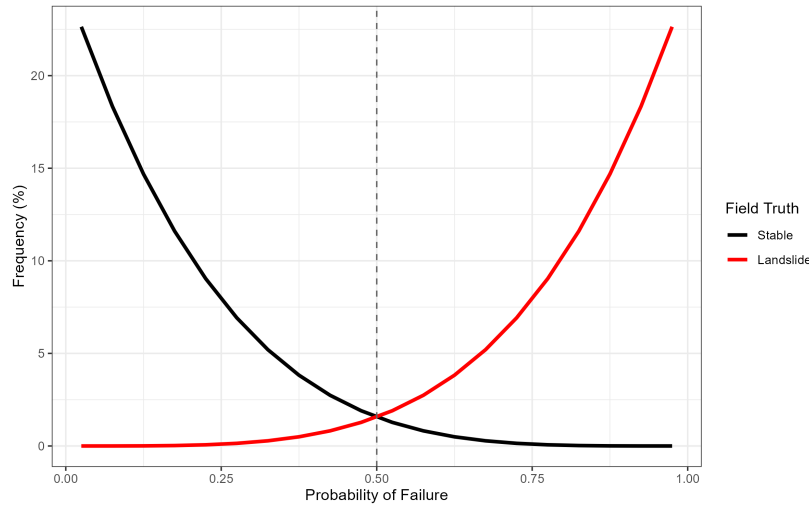


Figure 4.12: Schematic representation of the PoF distribution for the two classes of points, illustrating a good separation between the two distributions.

The confusion matrix and derived rates were calculated in the R environment using the `caret` package (Kuhn, 2008), whilst the ROC curve and AUC were obtained using the `pROC` package (Robin et al., 2011). The entire procedure was applied to the three hydrological scenarios (dry, initial, final; Section 4.1.6) and to the six model configurations. The results and their comparison are presented in Chapter 5.

4.4 Calibration Strategies

FSLAM requires the characterisation of the soil classes (`soil.csv`) and the land-cover classes (`hmtu.csv`) through the selection of their respective mechanical and hydrological parameters (Section 4.1.4). At the regional scale, these parameters cannot be measured directly over the entire area, and their estimation is therefore subject to considerable uncertainty. They must consequently be calibrated, within physically plausible intervals, so that the model reproduces the observed distribution of landslides, and the conditions that preceded them, as closely as possible. In the case study, the calibration is restricted to the Marnoso-Arenacea Formation (class 7, Section 4.2.2) as regards the soil classes.

Three calibration branches were compared. In the first, the parameters were adjusted manually, by trial and error, following the approach traditionally used with FSLAM (Asurza et al., 2026; Guo et al., 2022; Medina et al., 2021). In the second, the search was carried out automatically, with the R-FSLAM optimisation tool (Asurza et al., 2026) and a single objective. In the third, the automatic search was constrained by a second objective, defined on the pre-event state of the slopes.

The comparison between single- and multi-objective calibration follows the approach of Asurza et al. (2026). From the manual to the single-objective branch, the objective is unchanged and only the search procedure differs. The comparison therefore measures the contribution of a systematic exploration of the parameter space. From

the single- to the multi-objective branch, the objective function changes and, with it, the search algorithm and the rule by which the solution is selected. The three are not independent: optimising two objectives at once requires a different algorithm, and the Pareto front it returns requires a rule to choose among its solutions (Section 4.4.3). The comparison measures the overall effect of adding the pre-event constraint, not that of the objective function alone; the attribution of the individual effects is discussed in Section 5.5.

The three branches were compared on the base configuration of the model. The two modifications described in Section 4.1.7 were subsequently recalibrated with the multi-objective branch (Section 4.4.4). The resulting parameter values and the performance of each configuration are presented in Chapter 5.

4.4.1 Calibrated Parameters and Search Space

Not all the parameters populating `soil.csv` and `hmtu.csv` (Section 4.1.4) are subject to calibration. The calibrated parameters are the physical properties of the soil and of the cover that cannot be measured at the scale of the study area: the soil cohesion C_s , the root cohesion C_r , the friction angle φ , the thickness z , the porosity n and the hydraulic conductivity K .

The remaining parameters are held fixed. The density ρ_s is kept at 2000 kg/m³ for all classes, as in Medina et al. (2021), since its influence on the factor of safety is negligible. The curve number and the hydrologic soil group are not free physical parameters but conventional values of the SCS-CN method (USDA Soil Conservation Service, 1986), assigned according to Medina et al. (2021). The antecedent rainfall P_a and the event rainfall P_e are not parameters but input data, measured and interpolated (Section 4.2.4).

The calibrated parameters are represented in two ways (Section 4.1.5). The cohesions and the friction angle are given as ranges, with a minimum and a maximum value, and treated as stochastic variables; the remaining ones are given as a single deterministic value. In the automatic calibration, only the upper limit of the two cohesions is calibrated: $C_{s,min}$ and $C_{r,min}$ are kept at zero, while $C_{s,max}$ and $C_{r,max}$ are scaled. Both limits of the friction angle are instead calibrated, independently of each other.

The initial parameter set was derived from the literature. The Marnoso-Arenacea Formation is an alternation of arenaceous and pelitic terms (Section 3.2) and, as in the RER2023 classification (Section 4.2.2), the behaviour of the soil cover is inherited from the bedrock material. Among the classes available in the literature, the closest is the *conglomerate, sandstone and shale* class of Guo et al. (2022): it was therefore adopted as the starting point for all the calibration branches (Table 4.5).

The soil parameters were taken from that class, with the sole exception of the hydraulic conductivity. Hürlimann et al. (2022) report the two end members of the arenaceous–pelitic alternation separately, with $K = 10^{-4}$ m/s for the soil cover on sandstone and 10^{-6} m/s for that on mudstone: their geometric mean, 10^{-5} m/s, was therefore adopted. The root cohesion values adopted for the land-cover classes come instead from Medina et al. (2021), the source of the `hmtu.csv` file.

The values reported by the three sources are parameters of models calibrated on their respective case studies, not independent measurements, and are to be understood as a plausibility reference.

The set of physically plausible intervals constitutes the search space of the calibra-

tion, that is, the set of parameter combinations that may be explored. It is common to the three branches: the automatic calibration uses it as a set of explicit limits, whereas in the manual calibration the values are kept within it. The intervals adopted are reported in Table 4.5.

Table 4.5: Initial values from the literature, search intervals and parameters held fixed. (a) Soil parameters, referring to class 7. (b) Land-cover class parameters (`hmtu.csv`).

(a) Soil parameters							
Parameter	Symbol (unit)	Initial value	Admissible interval				
Soil cohesion, minimum	$C_{s,min}$ (kPa)	0	fixed				
Soil cohesion, maximum	$C_{s,max}$ (kPa)	2	1.0 – 5.0				
Friction angle, minimum	φ_{min} (°)	30	20 – 30				
Friction angle, maximum	φ_{max} (°)	40	32 – 40				
Thickness	z (m)	1.0	0.8 – 2.0				
Scale factor [†]	k (m)	1.0	0.8 – 2.0				
Hydraulic conductivity	K (m/s)	1.0×10^{-5}	$6.25 \times 10^{-7} - 6.25 \times 10^{-5}$				
Porosity	n (–)	0.35	0.25 – 0.40				
Density	ρ_s (kg/m ³)	2000	fixed				
Hydrologic soil group	HSG	A	fixed				
Antecedent rainfall	P_a (mm/d)	Section 4.2.4	fixed				
Event rainfall	P_e (mm)	Section 4.2.4	fixed				

(b) Land-cover class parameters							
Class	Cover	C_r (kPa)	$C_{r,max}$ interval (kPa)	CN_A	CN_B	CN_C	CN_D
1	Forest	0 – 5	3.0 – 10.0	36	60	73	79
2	Shrubs	0 – 3	1.8 – 6.0	43	65	76	82
3	Grassland	0 – 2	1.2 – 4.0	49	69	79	84
4	Bare rock	0	0	77	86	91	94
6	Bare soil	0	0	77	86	91	94
7	Farmland	0 – 1	0.6 – 2.0	56	68	80	84
8	Water	*	*	100	100	100	100
9	Urban	0 – 1	0.6 – 2.0	90	92	96	98

[†]The scale factor k applies only to the distributed-thickness configuration (Section 4.2.5); its calibration is described in Section 4.4.4. *The *water* class is assigned a conventional value

$C_r = 999$ kPa, which makes it unconditionally stable.

The interval of the hydraulic conductivity spans two orders of magnitude around the initial value. It includes the pelitic end member of the alternation (10^{-6} m/s) and remains below the arenaceous one (10^{-4} m/s), which would assign to the unit the permeability of only one of the two terms.

The intervals of the two limits of the friction angle are defined so that they cannot cross: φ_{min} does not exceed 30° and φ_{max} does not fall below 32° . The constraint is necessary because in the first trials the optimiser produced solutions with $\varphi_{min} > \varphi_{max}$.

Two limitations should finally be stated. The calibration of the soil parameters is restricted to class 7 (Section 4.2.2), which covers almost the entire study area and on which all the validation points fall. The root cohesion is instead scaled by a single multiplier, applied simultaneously to all the land-cover classes: it is therefore not

possible to calibrate the forest independently of the others, and the ratios between the classes remain those of the source.

4.4.2 Manual Calibration

The manual calibration is a trial-and-error procedure, carried out without a search algorithm: the parameter values are chosen by the operator and evaluated one at a time. The objective is the accuracy in the final condition (Section 4.3.2), the same one adopted by the single-objective automatic branch.

Maximising the true positives leads almost inevitably to overprediction: by widening the area predicted as unstable, their number increases by construction, at the expense of the stable points. The accuracy constrains both classes instead, and cannot be improved simply by extending the predicted unstable area.

The values were varied within the admissible intervals (Table 4.5) and evaluated with the same metrics used for the other branches (Section 4.3). The procedure was stopped when further variations no longer produced appreciable improvements in the objective.

4.4.3 Automatic Calibration

The automatic calibration was performed with R-FSLAM (Asurza et al., 2026), a tool developed in R for optimising FSLAM parameters. Within the search space defined in Section 4.4.1, the tool searches for the values that maximise the objective functions, evaluated on the validation dataset (Section 4.3). The initial values are those taken from the literature, the same from which the manual branch starts.

To make a systematic search feasible, the tool does not run the model at each iteration. It evaluates the probability of failure at the validation points only, applying the model equations (Equation 4.5) and the same stochastic treatment of the parameters (Section 4.1.5). Each point, landslide and non-landslide, therefore requires its slope, drainage area, antecedent and event rainfall, soil class and land-cover class.

These values are taken from the same input and output rasters used by the model, and the probability of failure so computed was compared with the one produced by the complete model at the same points, confirming that the analytical evaluation is a faithful surrogate of the model.

The tool also operates on the single cell containing each point, whereas the validation applies a 2.5 m buffer to the landslide points and takes the maximum PoF among the intersected cells (Section 4.3.1). The accuracy values maximised during calibration are therefore slightly lower than those reported in Chapter 5, which are the validated ones.

R-FSLAM implements two calibration modes, single- and multi-objective, which differ in the objective function and in the search algorithm, and correspond to the two automatic branches applied to the base configuration.

Single-objective calibration. The search is carried out with the Dynamically Dimensioned Search (DDS) algorithm. The objective function is the accuracy in the final condition, the same one adopted by the manual branch (Section 4.4.2).

Since the algorithm is stochastic, a single run may settle on a local optimum. The search was therefore repeated over eight independent runs, and the parameter set that

achieved the highest accuracy was adopted.

Multi-objective calibration. The search is carried out with the NSGA-II genetic algorithm (*Elitist Non-dominated Sorting Genetic Algorithm II*), which optimises two objective functions simultaneously. In the final condition the objective function is the accuracy, with the points treated according to their actual status. In the initial condition it is the true negative rate, computed by treating all points as stable, including the initiation points: before the event no slope had yet failed, and the model must therefore not predict their instability.

The second objective responds to a requirement of physical plausibility. A calibration based on the final condition alone can reproduce the observed landslides well but remain implausible in the pre-event state, predicting as already unstable slopes that were in fact stable.

The two objective functions conflict: a parameterisation that improves the final condition tends to worsen the initial one. There is therefore no single optimal solution, but a set of non-dominated solutions (the Pareto front), each corresponding to a different trade-off between the two objectives (Asurza et al., 2026).

The solution was selected according to a rule defined in advance. The first criterion is a constraint: the true negative rate in the initial condition must exceed 0.9, the value recommended by Asurza et al. (2026), who below that threshold observe probability-of-failure maps in the initial condition showing widespread instability, similar to those of the final condition. Among the admissible solutions, the one with the highest final accuracy was adopted and, where the accuracy was comparable, the one with the more plausible partition between the two hydrological components.

The true negative rate appearing in the constraint is the one defined above, computed over all points treated as stable. It is a distinct quantity from the true negative rate reported among the validation metrics in Chapter 5, which is computed on the non-landslide points alone.

4.4.4 Calibration of the Modified Configurations

The two modifications described in Section 4.1.7 were not evaluated with the parameters of the base configuration, but recalibrated, using the multi-objective branch. The comparison thus takes place between the respective optimal calibrations.

In the configuration with the accumulation-length threshold, the threshold value is not itself a calibrated parameter. It was set in advance at 20 and 50 m, that is, at two and five cells of the reference grid (Section 4.2.1), so as to compare a strong truncation with a moderate one. For each of the two values the remaining parameters were recalibrated, giving two distinct configurations.

The drainage area required by the tool is, in this case, the one truncated at the threshold under examination. For each of the two values a preliminary FSLAM run was therefore performed, from which the accumulation raster was extracted and, from it, the drainage area of each point. The accumulation depends only on the topography and on the threshold value, not on the parameterisation: the preliminary run can therefore be carried out with any parameter set.

The distributed-thickness configuration runs on the modified version of the code, the one with the accumulation-length threshold, and was evaluated at a threshold

of 50 m. This allows a direct comparison with the constant-thickness configuration calibrated at the same threshold.

In this configuration the constant thickness is replaced by the normalised map z_{norm} (Section 4.2.5), and the calibrated parameter is the scale factor k :

$$z(\mathbf{x}) = \max(k z_{norm}(\mathbf{x}), z_{min}). \quad (4.16)$$

Since z_{norm} has unit mean by construction, k coincides with the mean thickness of the soil cover over the domain. It is therefore directly comparable with the constant thickness of the other configurations, and was assigned the same search interval (Table 4.5). The minimum value $z_{min} = 0.1$ m is applied downstream of the scaling. It is necessary because in the cells where the map vanishes the factor of safety (Equation 4.5) would not be defined.

This calibration required an extension of the tool. R-FSLAM assigns the thickness as a constant value per soil class; in the version used here the tool receives instead, for each point, the normalised value of the map, to which it applies the scale factor being calibrated.

5 – Model Application and Results

This chapter presents the results of the six model configurations defined in Chapter 4. The first three share the original formulation of FSLAM and differ only in the calibration strategy (Section 4.4): the manual calibration (Section 5.1) and the two automatic ones, single- and multi-objective (Section 5.2). The remaining three adopt the modifications introduced in Section 4.1.7, each recalibrated with the multi-objective branch (Section 4.4.4): the accumulation-length threshold, set at 20 and 50 m (Section 5.3), and the distributed soil thickness (Section 5.4).

Sections 5.1–5.4 follow a common scheme. Each section first reports the parameters obtained from the calibration, compared with the literature values from which the calibration starts (Section 4.4.1). It then presents the maps returned by the model and the corresponding domain statistics, quantifies the performance of the configuration on the validation dataset (Section 4.3) and examines its progression across the three hydrological conditions (Section 4.1.6). It closes with the interpretation of the results, highlighting strengths and limitations. Section 5.3 departs from this scheme in one respect. It opens with the direct effect of the threshold, measured with the parameters fixed, because that is the only way to isolate it.

The comparison between the six configurations is reserved for Section 5.5.

All the configurations start from the same set of literature parameters (Table 4.5), the reference against which the effect of the calibration is assessed. Applied to the model without any calibration, this set correctly identifies 52.3% of the observed landslides and 73.9% of the stable points, with an accuracy of 0.631 and an area under the ROC curve of 0.712. In the initial condition the true negative rate is 0.994, and only 6.7% of the initiation points are already predicted as unstable. The literature set therefore describes the pre-event state in a physically consistent way, but captures only slightly more than half of the observed failures: the calibration acts on this imbalance, addressing a different aspect in each configuration.

5.1 Base Configuration – Manual Calibration

This section presents the results of the base configuration of the model, with the parameters obtained from the manual calibration (Section 4.4.2), whose objective is the accuracy in the final condition.

Calibrated parameters. The parameters of class 7 are reported in Table 5.1, compared with the initial literature values from which the calibration started.

Table 5.1: Parameters of class 7 obtained from the manual calibration, compared with the initial values from the literature.

Parameter	Symbol (unit)	Literature	Calibrated
Soil cohesion, maximum	$C_{s,max}$ (kPa)	2	2
Friction angle, minimum	φ_{min} ($^{\circ}$)	30	22
Friction angle, maximum	φ_{max} ($^{\circ}$)	40	34
Thickness	z (m)	1.00	1.25
Hydraulic conductivity	K (m/s)	1.0×10^{-5}	4.0×10^{-6}
Porosity	n (-)	0.35	0.35
Density	ρ_s (kg/m ³)	2000	2000
Hydrologic soil group	HSG	A	A

The calibration modified only three parameters: the friction angle, reduced by six to eight degrees in both limits; the thickness, increased by 25%; and the hydraulic conductivity, reduced by a factor of 2.5. The porosity and the soil cohesion remained at the literature values. The root cohesion also remained unchanged: the values adopted are those of Table 4.5(b), and this is the only configuration in which the vegetation cover is not modified by the calibration.

None of the calibrated parameters lies at the boundary of the search space (Table 4.5). The optimum is therefore internal, and not an effect of the limits imposed on the search.

Model output. Figures 5.1 and 5.2 show, respectively, the probability of failure (PoF) map and the factor of safety (FS) map in the final condition. The two maps represent the same result in its two forms, stochastic and deterministic (Section 4.1.5). Instability is concentrated on the steeper slopes, whereas the valley floors and the gently sloping sectors remain stable.

Over the municipal territory, 41.96% of the cells are unstable in the final condition ($PoF > 0.5$) and 29.76% are essentially stable ($PoF \leq 0.01$); the remaining 28.28% falls within the intermediate band of uncertainty. The cells at near-certain failure ($PoF > 0.99$) amount to 8.01%.

Through two dedicated outputs (Section 4.1.6), the model also identifies the cells whose stability does not depend on rainfall: 23.81% are unconditionally stable (probability > 0.99 , stable even at full saturation) and 0.43% unconditionally unstable (probability > 0.99 , unstable already under dry conditions). The comparison between the 8.01% of cells at near-certain failure in the final condition and the 0.43% unconditionally unstable directly measures the effect of the event rainfall on the initiation of failure.

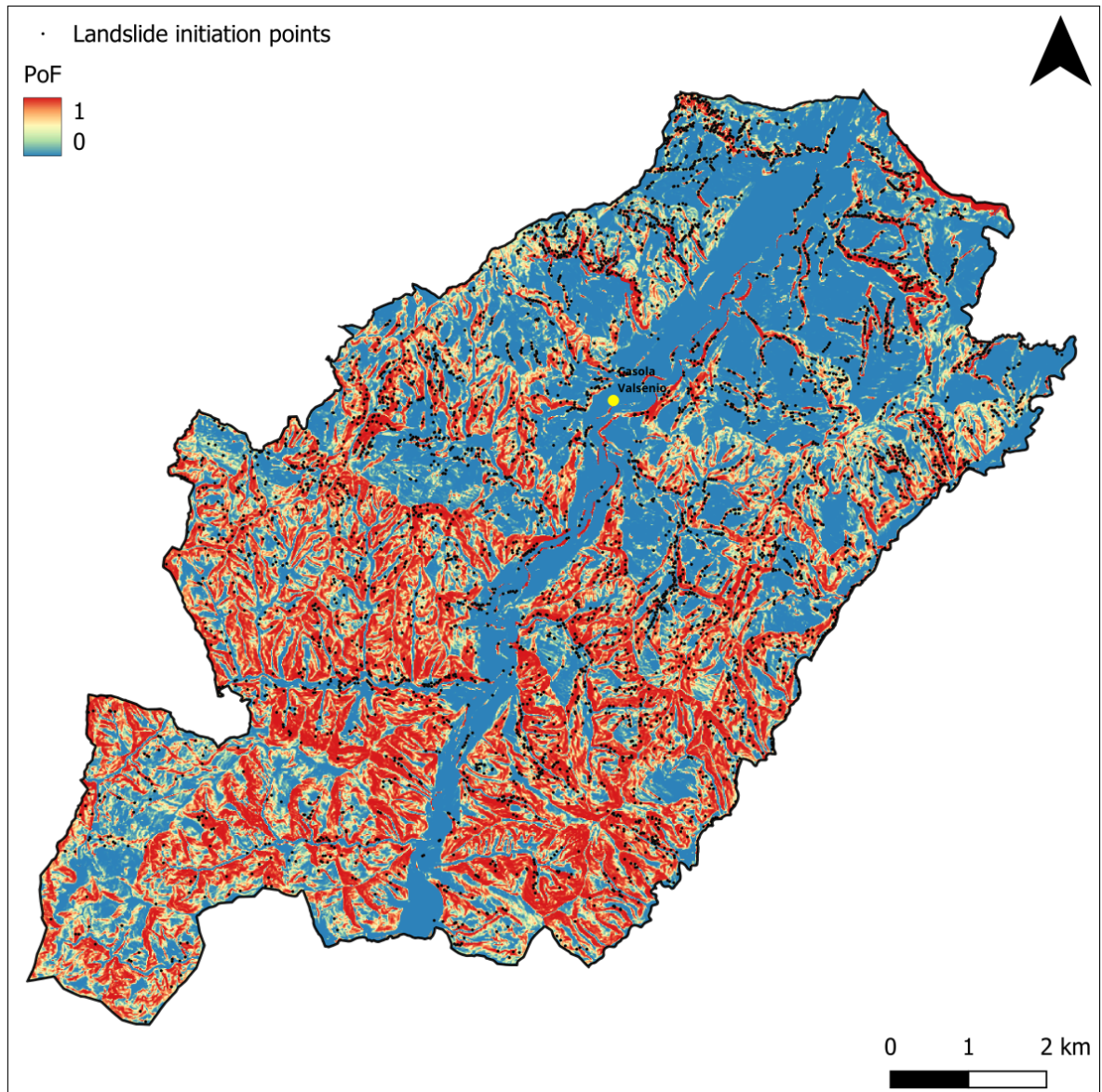


Figure 5.1: Probability of failure (PoF) map in the final condition for the base configuration with manual calibration. In black, the initiation points of the May 2023 inventory (DS + DF).

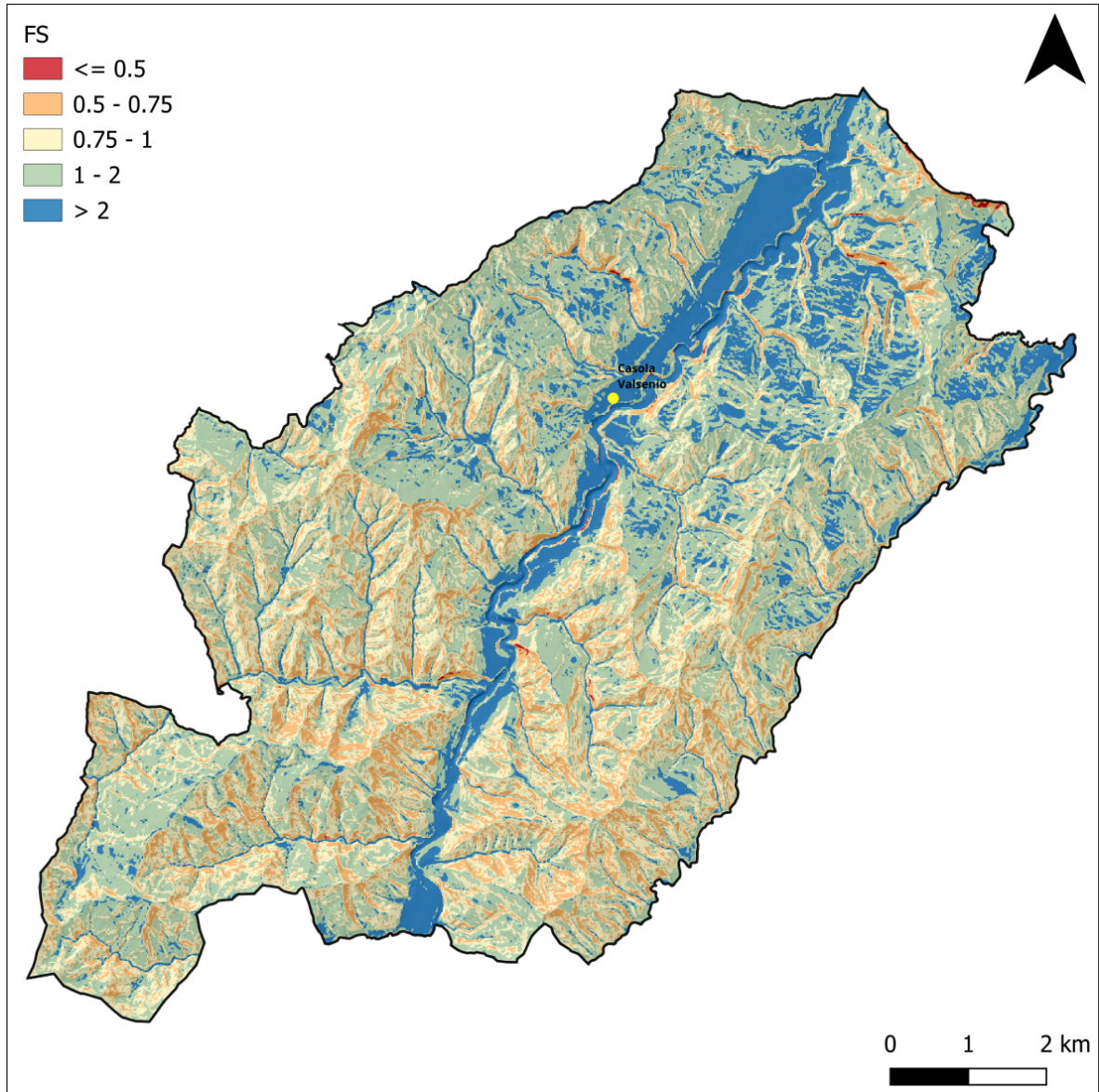


Figure 5.2: Factor of safety (FS) map in the final condition for the base configuration with manual calibration, reported as an example of the deterministic form of the result; in the following, the configurations are presented through the PoF alone.

Validation in the final condition. The performance is quantified on the validation dataset described in Section 4.3.1, comparing the PoF in the final condition with the observed inventory. The confusion matrix is reported in Figure 5.3. The model correctly identifies 81.3% of the observed landslides ($\text{TPR} = 0.813$), but only 50.9% of the stable points ($\text{TNR} = 0.509$). The accuracy is 0.661, markedly higher than the 50% expected from random classification on a balanced dataset, and the area under the ROC curve (AUC) is 0.731, indicative of a moderate discriminating capacity. These values are reported, together with those of the two preceding hydrological conditions, in Table 5.2.

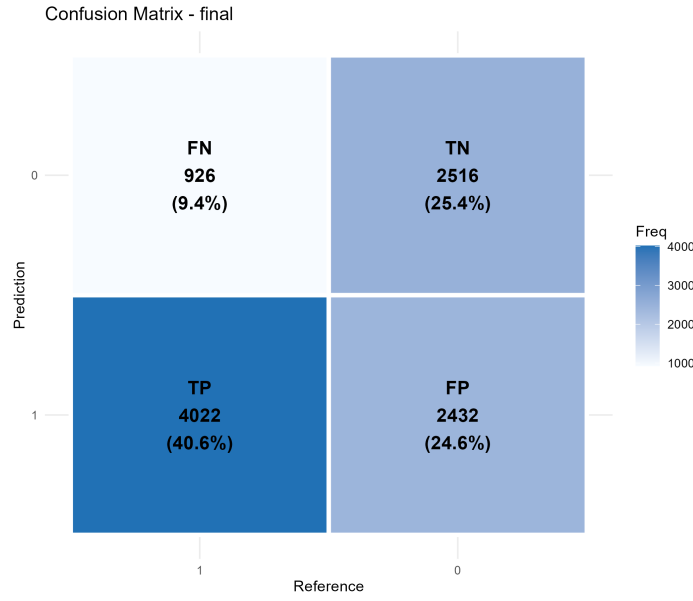


Figure 5.3: Confusion matrix of the base configuration with manual calibration, in the final condition.

The distribution of slope by category (Figure 5.4) clarifies the nature of the error. The points that the model predicts as unstable, both those that failed and those that did not, have almost coincident median slopes (31.7° and 31.6°) and largely overlapping interquartile ranges, clearly distinct from those of the points predicted as stable (18.0° and 20.6°): the separation drawn by the model effectively coincides with a slope threshold. The mean probabilities of failure of the true and false positives are likewise almost coincident (0.91 and 0.86), against 0.12 and 0.19 for the two negative categories. The overprediction is therefore not an artefact of the threshold adopted on the PoF: the model captures the predisposition associated with slope, but does not distinguish, among the slopes that possess it, those that failed from those that did not.

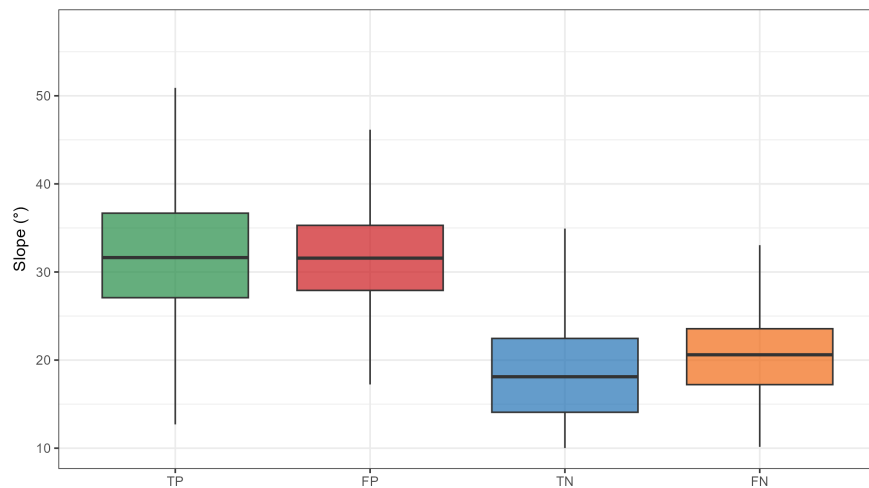


Figure 5.4: Distribution of slope by classification category (TP, FP, TN, FN), in the final condition for the base configuration with manual calibration.

The PoF distribution over the validation points (Figure 5.5) makes the overprediction evident. The landslide points cluster towards high values, consistently with the

high TPR, whereas the stable points do not concentrate at low values as they would in a well-discriminating model, but display a marked secondary mode at high probabilities. The overlap between the two distributions in that region is the graphical counterpart of the low TNR and the moderate AUC. The PoF values are also distributed mainly towards the extremes, with few points near the threshold: the classification is therefore weakly sensitive to the adopted threshold value.

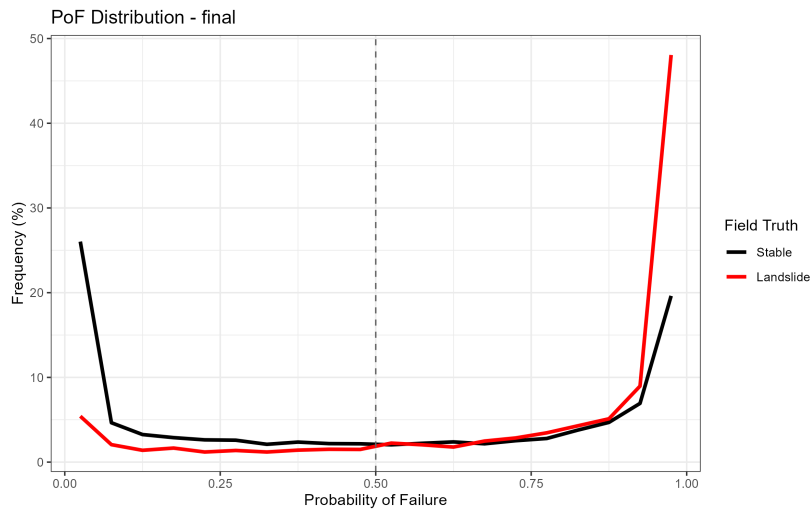


Figure 5.5: PoF distribution in the final condition for the landslide and non-landslide points.

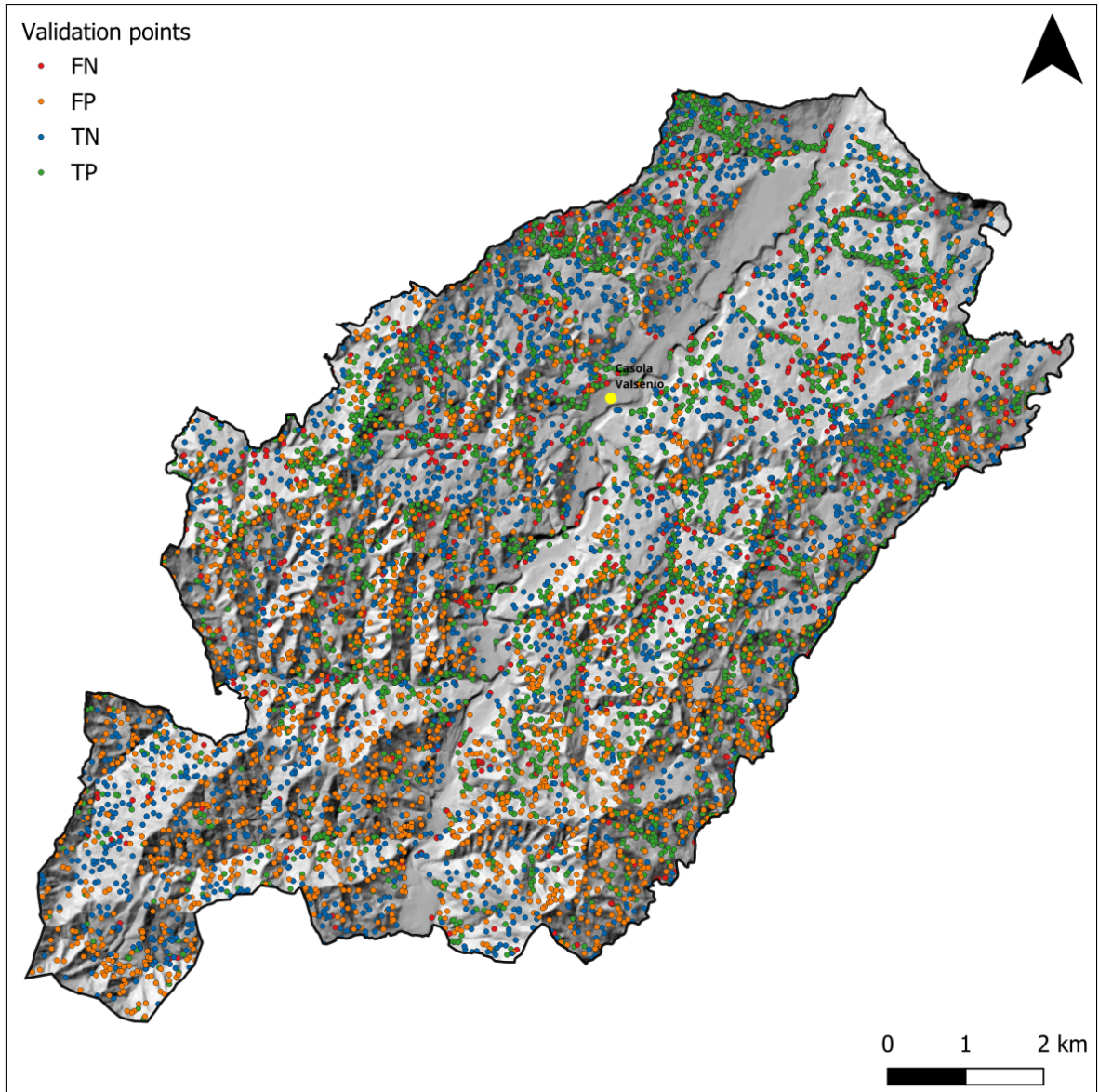


Figure 5.6: Classified validation points (TP, TN, FP, FN) overlaid on the hillshade, in the final condition for the base configuration with manual calibration.

Progression of the hydrological conditions. Applying the validation also to the dry and initial conditions (Section 4.1.6) yields the progression reported in Table 5.2. The TPR rises from 0.297 to 0.418 and then to 0.813, whilst the TNR decreases from 0.912 to 0.856 and then to 0.509. The AUC varies between 0.716 and 0.732. As saturation increases, a growing number of cells exceeds the failure threshold: what shifts is the operating point of the model, not its capacity to rank the cells according to their proneness to failure. The full identification of the landslides is achieved only in the final condition, which confirms the event rainfall as the triggering factor.

Table 5.2: Validation metrics of the base configuration with manual calibration, in the three hydrological conditions (Section 4.1.6).

Condition	TPR	TNR	ACC	AUC
Dry	0.297	0.912	0.604	0.716
Initial (pre-event)	0.418	0.856	0.637	0.732
Final (post-event)	0.813	0.509	0.661	0.731

Interpretation: strengths and limitations. The base configuration with manual calibration achieves an accuracy of 0.661 and a moderate discriminating capacity ($\text{AUC} = 0.731$). It detects a high proportion of the observed landslides ($\text{TPR} = 0.813$), but its specificity is low ($\text{TNR} = 0.509$). Three limitations emerge, differing in nature.

The first concerns the state preceding the event. Already in the initial condition, 41.8% of the initiation points are predicted as unstable, and 29.7% even under dry conditions. Since these are states prior to the initiation, in which the slopes that later failed were still stable, such values ought to be close to zero. The calibration therefore leads to an overestimation of the instability in the pre-event state, because by optimising the final state alone it places no constraint on the initial one.

The second limitation concerns the determination of the parameters. In the course of the procedure, sets of values differing considerably from one another returned accuracies within a range of a few thousandths: the objective adopted therefore discriminates only weakly between alternative parameterisations. In addition, the procedure explores the parameter space through successive trials and is stopped according to an empirical criterion, without any guarantee of having reached the global optimum. The values reported in Table 5.1 therefore represent one of the many sets that achieve equivalent performance.

The third limitation is of a different nature, as it does not depend on the parameter values. The separation drawn by the model effectively coincides with a slope threshold, and among the steeper slopes the capacity to distinguish those that failed from the others is almost nil. Slope is the predisposing factor best represented by the model, but it is not sufficient to determine which slopes failed.

5.2 Base Configuration – Automatic Calibration

This section presents the results of the base configuration of the model, with the parameters obtained from the automatic calibration. The R-FSLAM tool was used in its two calibration modes, which adopt a different objective function and a different search algorithm (Section 4.4.3).

The first branch adopts the accuracy in the final condition as its objective. It is the same objective as that of the manual calibration (Section 4.4.2), but the successive trials are replaced by a systematic search of the parameter space. The second branch adds a constraint on the state preceding the event.

5.2.1 Single-Objective Calibration

The search was repeated over eight independent runs (Section 4.4.3), and the solution adopted is the one that achieved the highest accuracy.

Calibrated parameters. The solutions of the eight runs are reported in Table 5.3, sorted by decreasing accuracy.

Table 5.3: Solutions of the eight independent runs of the DDS algorithm, sorted by decreasing accuracy. The first row is the adopted solution.

Seed	ACC ^a	$C_{s,max}$ (kPa)	C_r forest (kPa)	φ_{min} (°)	φ_{max} (°)	z (m)	K (10 ⁻⁶ m/s)	n (-)
42	0.6372	2.65	7.53	20.3 ^b	36.6	1.97 ^b	0.91 ^b	0.370
1	0.6367	1.40	5.17	29.5 ^b	33.2	1.47	1.25	0.386 ^b
123	0.6361	1.63	5.86	28.9	35.0	1.85	0.73 ^b	0.358
7	0.6352	1.50	5.16	22.7	34.4	1.25	3.53	0.393 ^b
2024	0.6346	2.19	4.74	28.7	34.4	1.65	1.68	0.380
5	0.6338	2.17	4.07	26.5	33.6	1.36	1.68	0.396 ^b
11	0.6305	1.23 ^b	3.46 ^b	27.1	39.7 ^b	0.95	0.83 ^b	0.258 ^b
99	0.6297	1.20 ^b	3.64 ^b	24.7	38.8	0.87 ^b	1.00	0.290

^aAccuracy computed by the analytical evaluation of the tool on the single cell containing each point (Section 4.4.3); it is not directly comparable with the validated accuracies reported in the rest of the chapter. ^bValue falling within the outer 10% of the corresponding admissible interval (Table 4.5).

The eight runs converge on accuracy values between 0.629 and 0.638. The parameters that produce them are, by contrast, very different from one another: the thickness varies from 0.87 to 1.97 m, the hydraulic conductivity by a factor of almost five, and the root cohesion of the forest class from 3.46 to 7.53 kPa. The objective adopted therefore does not identify the parameters. Very different parameterisations, obtained with the same procedure on the same search space and the same dataset, return equivalent performance.

The eight solutions differ, however, in physical plausibility. The adopted solution, which is the one with the highest accuracy, has three parameters close to the limits of the search space. Runs 7 and 2024 are lower in accuracy by two and three thousandths respectively, but their parameters lie almost entirely within the intervals and are closer to the literature values. The selection rule was applied as defined in advance (Section 4.4.3), but accuracy alone provides no criterion for preferring a physically more credible solution to a statistically equivalent one.

Model output. Over the municipal territory, 40.11% of the cells are unstable in the final condition ($PoF > 0.5$) and 24.51% are essentially stable ($PoF \leq 0.01$); the remaining 35.39% falls within the intermediate band of uncertainty. The cells at near-certain failure ($PoF > 0.99$) amount to 5.56%. The cells whose stability does not depend on rainfall are 0.36% unconditionally unstable and 20.76% unconditionally stable (probability > 0.99 in both cases).

The probability of failure map is not reported: at the scale of representation adopted it is not distinguishable from that of the manual configuration (Figure 5.1). The difference between the two configurations does not lie in the spatial distribution of the instability, but in the gradation of the probabilities. Compared with the manual configuration, the cells at near-certain failure decrease from 8.01% to 5.56% and the intermediate band rises from 28.28% to 35.39%.

Validation in the final condition. The model correctly identifies 80.4% of the observed landslides ($\text{TPR} = 0.804$) and 52.6% of the stable points ($\text{TNR} = 0.526$), with an accuracy of 0.665 and an area under the ROC curve of 0.740. The confusion matrix is reported in Figure 5.7.

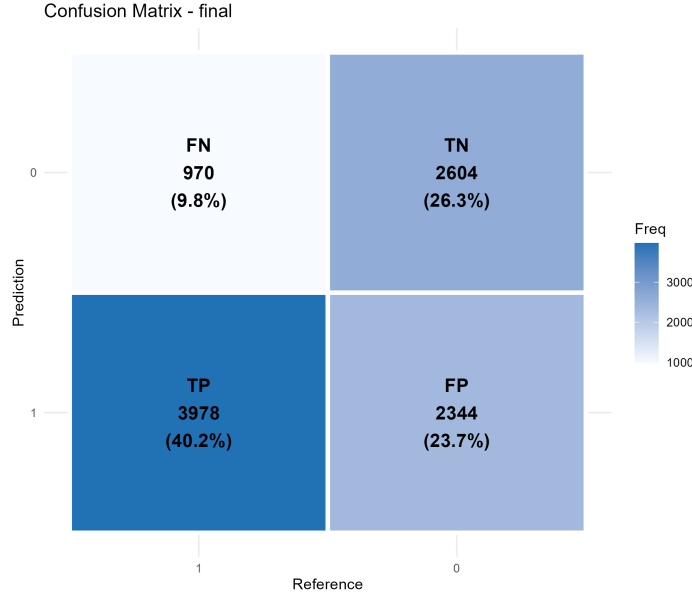


Figure 5.7: Confusion matrix of the base configuration with single-objective calibration, in the final condition.

The behaviour described in Section 5.1 is confirmed: the true and false positives have mean probabilities of failure of 0.89 and 0.83, and mean slopes of 32.0° and 31.5° .

Compared with the manual calibration, the accuracy rises from 0.661 to 0.665 and the area under the ROC curve from 0.731 to 0.740. Both differences are of the order of a few thousandths.

Progression of the hydrological conditions. The progression across the three hydrological conditions is reported in Table 5.4.

Table 5.4: Validation metrics of the base configuration with single-objective calibration, in the three hydrological conditions (Section 4.1.6).

Condition	TPR	TNR	ACC	AUC
Dry	0.296	0.912	0.604	0.719
Initial (pre-event)	0.603	0.734	0.669	0.741
Final (post-event)	0.804	0.526	0.665	0.740

The state preceding the event is markedly worse than in the manual calibration. In the initial condition, 60.3% of the initiation points are predicted as unstable, against 41.8% for the manual configuration, although none of those slopes had yet failed. The algorithm reaches this configuration because the objective evaluates the final condition alone, in which a high antecedent water table helps to detect the observed landslides.

Interpretation: strengths and limitations. The systematic search optimises the declared objective more effectively than the manual procedure. The gain, however, amounts to a few thousandths of accuracy: with the same objective, replacing the successive trials with a search algorithm does not appreciably change the performance of the model.

The limitations concern physical plausibility. The parameters adopted are pushed closer to the edges of the search space than those of other solutions of equal accuracy. The pre-event state, already poorly represented by the manual calibration, deteriorates further. Both aspects depend on the objective adopted, which does not take them into account, and motivate the constraint introduced in the following subsection.

5.2.2 Multi-Objective Calibration

The multi-objective branch optimises two objective functions simultaneously: the accuracy in the final condition and the true negative rate in the initial condition, which must exceed 0.9. The solution adopted is the one identified by the selection rule defined in advance (Section 4.4.3). The true negative rate of the constraint is measured over the whole sample treated as stable. It therefore does not coincide with the value reported among the validation metrics, which counts only the non-landslide points.

Calibrated parameters. The constraint is satisfied by 86 solutions, reported in Figure 5.8 in the space of the two objective functions.

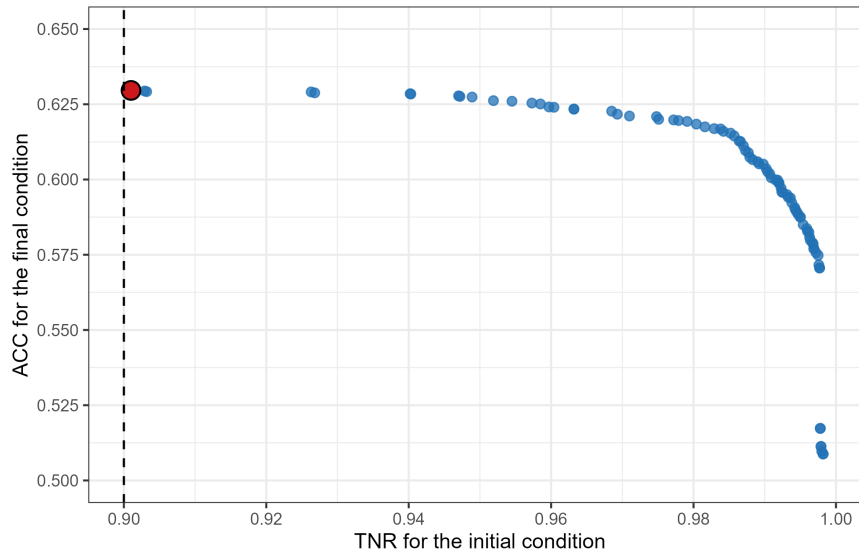


Figure 5.8: The Pareto front composed of the admissible solutions of the multi-objective calibration; the dashed line indicates the constraint and the red point the adopted solution.

The trade-off between the two objectives is markedly asymmetric. Up to a true negative rate of about 0.97 in the initial condition, the final accuracy decreases only from 0.630 to 0.622. Beyond that value it falls rapidly, reaching 0.509 at 0.998. The consistency of the pre-event state is therefore obtainable at almost no cost, provided the constraint is not pushed too far.

The adopted solution lies at the lower end of the admissible region, with a true negative rate of 0.901 in the initial condition and a final accuracy of 0.630. The

corresponding parameters are reported in Table 5.5.

Table 5.5: Parameters of class 7 obtained from the multi-objective calibration, compared with the initial values from the literature.

Parameter	Symbol (unit)	Literature	Calibrated
Soil cohesion, maximum	$C_{s,max}$ (kPa)	2	1.34
Root cohesion, forest	$C_{r,max}$ (kPa)	5	3.95
Friction angle, minimum	φ_{min} ($^{\circ}$)	30	22.4
Friction angle, maximum	φ_{max} ($^{\circ}$)	40	36.4
Thickness	z (m)	1.00	0.82
Hydraulic conductivity	K (m/s)	1.0×10^{-5}	3.34×10^{-5}
Porosity	n (-)	0.35	0.25

Unlike the manual calibration, which modified three of them, the multi-objective calibration acts on all the parameters. The soil cover becomes thinner, with lower cohesion and a lower friction angle, but more permeable and less porous. The porosity coincides with the admissible minimum of 0.25, and the thickness, equal to 0.82 m, is close to the minimum of 0.8 m.

The comparison with the single-objective calibration is, however, the most significant result. Although the two parameterisations give comparable accuracies, they are opposite: the thickness passes from 1.97 to 0.82 m and the hydraulic conductivity from 9.1×10^{-7} to 3.34×10^{-5} m/s, a factor of 37.

Model output. Figure 5.9 shows the probability of failure map in the final condition.

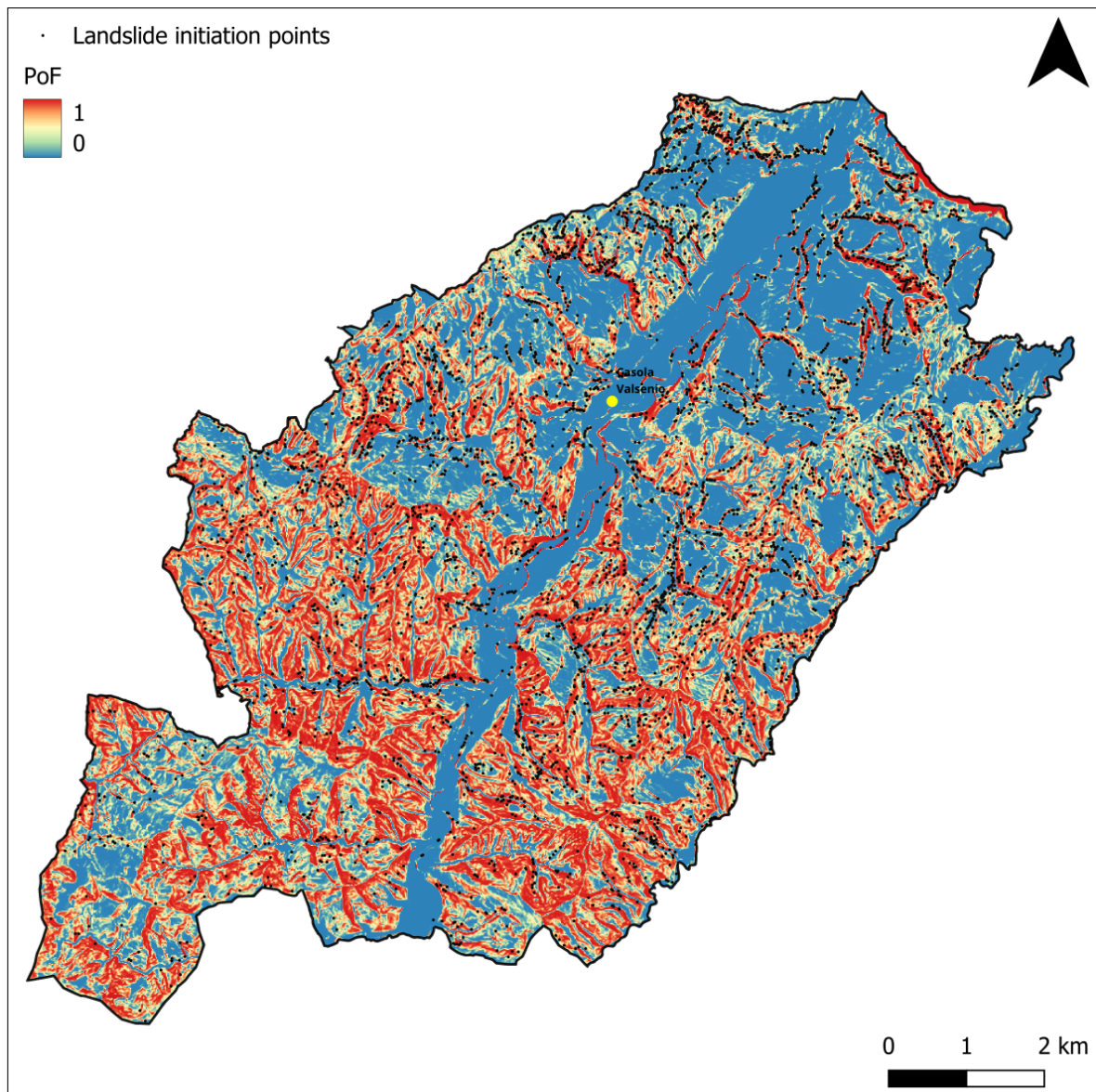


Figure 5.9: Probability of failure (PoF) map in the final condition for the base configuration with multi-objective calibration. In black, the initiation points of the May 2023 inventory (DS + DF).

Over the municipal territory, 42.11% of the cells are unstable in the final condition ($PoF > 0.5$) and 27.23% are essentially stable ($PoF \leq 0.01$); the remaining 30.67% falls within the intermediate band of uncertainty. The cells at near-certain failure ($PoF > 0.99$) amount to 4.59%. The cells whose stability does not depend on rainfall are 0.28% unconditionally unstable and 25.35% unconditionally stable (probability > 0.99 in both cases). Compared with the manual configuration, the extent of the predicted instability is almost identical (42.11% against 41.96%), but the cells at near-certain failure are halved, from 8.01% to 4.59%. For the same predicted unstable area, the probability surface is therefore less polarised.

Validation in the final condition. The model correctly identifies 81.4% of the observed landslides ($\text{TPR} = 0.814$) and 50.5% of the stable points ($\text{TNR} = 0.505$), with an accuracy of 0.659 and an area under the ROC curve of 0.732. The confusion matrix is reported in Figure 5.10.

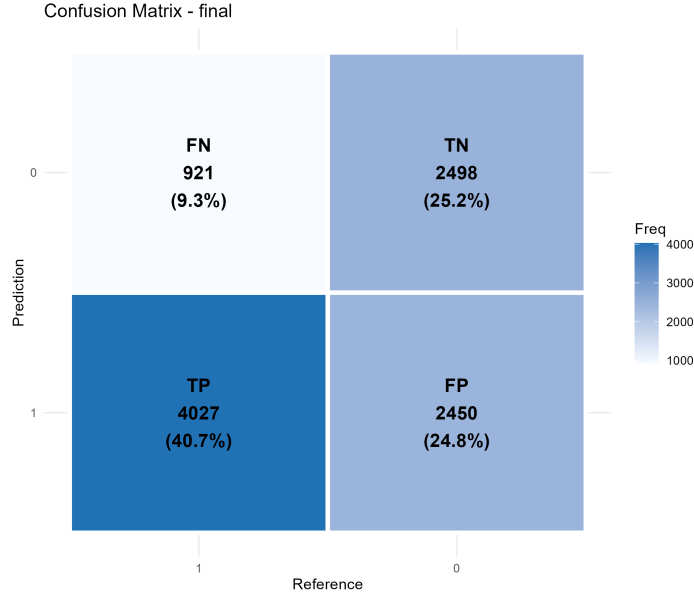


Figure 5.10: Confusion matrix of the base configuration with multi-objective calibration, in the final condition.

Here too the true and false positives remain indistinguishable, with mean probabilities of failure of 0.89 and 0.84 and mean slopes of 31.9° and 31.4° . Compared with the single-objective calibration, the accuracy decreases from 0.665 to 0.659 and the area under the ROC curve from 0.740 to 0.732.

Progression of the hydrological conditions. The progression across the three hydrological conditions is reported in Table 5.6.

Table 5.6: Validation metrics of the base configuration with multi-objective calibration, in the three hydrological conditions (Section 4.1.6).

Condition	TPR	TNR	ACC	AUC
Dry	0.191	0.970	0.580	0.718
Initial (pre-event)	0.203	0.966	0.585	0.724
Final (post-event)	0.814	0.505	0.659	0.732

The constraint has the intended effect. In the initial condition, 20.3% of the initiation points are predicted as unstable, against 60.3% for the single-objective calibration and 41.8% for the manual one, and the validated true negative rate is 0.966. Before the event the model classifies almost all the slopes as stable, including those that would later fail. The comparison between the two automatic calibrations in the initial condition is shown in Figure 5.11.

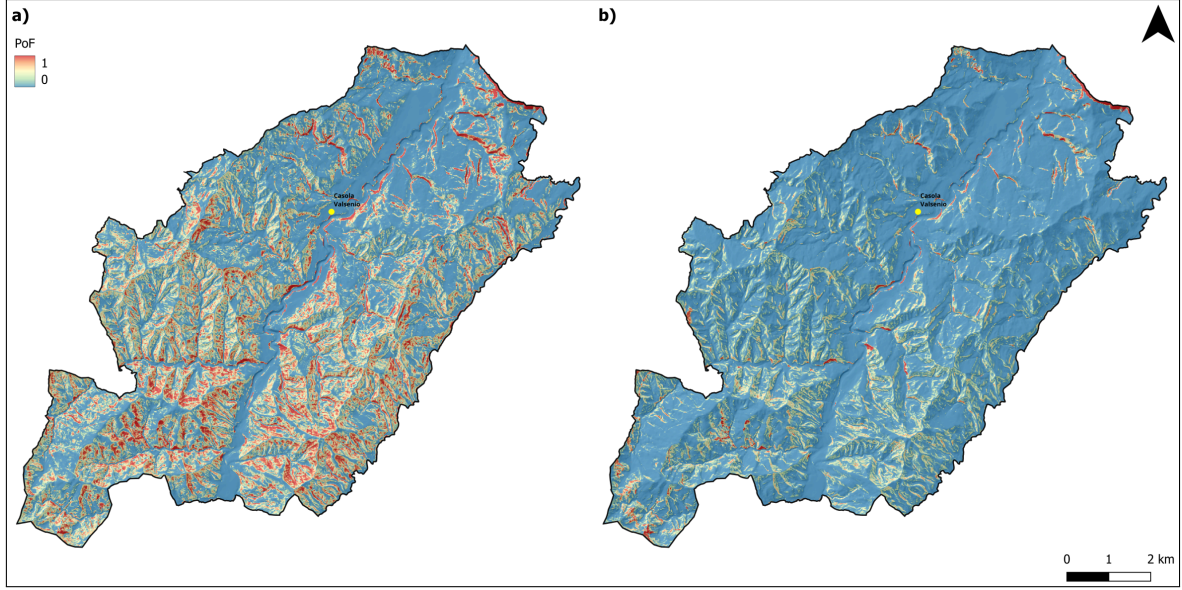


Figure 5.11: Probability of failure (PoF) map in the initial condition: (a) single-objective calibration; (b) multi-objective calibration. The two maps share the same colour scale.

The partition of the water table rise between the two hydrological components (Section 4.1.3) describes the state of the slope before the event. At the validation points the antecedent rainfall produces a median rise equal to 2% of the thickness of the soil cover ($h_a/z = 0.02$), whereas the event infiltration determines 97% of it ($h_e/z = 0.97$). The water table is therefore low before the event, and it is the May rainfall that brings it close to saturation. The table confirms this: between the dry and the initial condition the TPR rises only from 0.191 to 0.203. In this configuration, as in the initial parameter set, the antecedent rainfall does not appreciably modify the state of the slopes.

Interpretation: strengths and limitations. The second objective achieves what it was introduced for. The state preceding the event is now represented in a physically consistent manner, and the instability is determined almost entirely by the May rainfall, as expected for a triggering event.

The cost in performance is almost negligible, but there is no gain either: the second objective yields physical consistency, not predictive capacity.

Finally, two parameters lie at the lower edge of the search space: the porosity coincides with the limit and the thickness lies two centimetres above it. In both cases the value was not chosen by the calibration, but imposed by the limit of the search interval. The two modifications presented in the following sections start from this observation. The first acts on the antecedent hydrological component, to verify whether a more accurate representation of it allows comparable performance with parameters closer to the literature values. The second replaces the constant thickness with a distributed map, to assess to what extent the spatial variability of the soil cover affects the prediction of slope instability.

5.3 Accumulation-Length Threshold Configuration

The modification truncates the drainage area at a fixed distance upslope (Section 4.1.7), and therefore acts on the antecedent component of the hydrological model alone. The two values, 20 and 50 m, are not calibrated parameters: they were set in advance (Section 4.4.4), and for each of the two the remaining parameters were recalibrated with the multi-objective branch, following the same selection rule as the base configuration.

The comparisons that follow are made against the base configuration with multi-objective calibration (Section 5.2.2), on which both modifications are built. The section first examines the effect of the threshold with the parameters fixed, and then the two recalibrated configurations.

Direct effect of the threshold. Equation 4.2 shows directly what the threshold acts on. The water table height produced by the antecedent rainfall is proportional to the contributing area per unit width and inversely proportional to the hydraulic conductivity. The truncation acts on the first term and can only reduce it, and can therefore only reduce h_a . The purpose of this configuration is thus to reduce the contribution of the antecedent component to slope stability, where that component weighs too heavily on the final result.

The effect of the threshold can be measured directly only with the parameters fixed. In the two recalibrated configurations, presented below, the truncation and the parameterisation change together and the two effects cannot be separated. The threshold was therefore varied over a range of values while keeping the parameters at the literature values. The results are reported in Table 5.7.

Table 5.7: Direct effect of the accumulation-length threshold, with the parameters fixed at the literature values: drainage area at the validation points, hydrological partition and classification metrics.

Threshold	Median area (m ²)	Mean area (m ²)	h_a/z	h_e/z	TPR ^a	TNR ^a	ACC ^a
10 m	100	100	0.01	0.90	0.412	0.765	0.589
20 m	196	197	0.02	0.90	0.418	0.760	0.589
50 m	414	450	0.04	0.89	0.429	0.749	0.589
100 m	577	793	0.06	0.89	0.443	0.742	0.592
200 m	639	1291	0.06	0.88	0.449	0.739	0.594
unlimited ^b	654	2215	0.07	0.88	0.453	0.739	0.596

^aMetrics computed by the analytical evaluation of the tool on the single cell containing each point (Section 4.4.3). ^bValue computed with the original version of the code, which does not implement the threshold.

The truncation reduces the drainage area at the validation points, but not uniformly. Without a threshold the mean is more than three times the median: most of the points collect from a few upstream cells, whereas a minority, located where the flow converges, collect from very large areas. It is on this minority that the truncation acts. At 50 m the mean falls by 80% while the median loses 37%, which shows that the cut affects mainly the values in the tail; for most of the points, which already drain from a few cells, there is little left to remove. At the shorter thresholds a second effect comes into

play: the maximum value that can be reached is imposed by the threshold itself, and at 20 m every point is reduced to two cells. Mean and median then almost coincide.

The relation between contributing area and water table height follows what the formulation requires. Between the 10 m threshold and unlimited accumulation the median area goes from 100 to 654 m², that is, by a factor of 6.5, and the median h_a/z from 0.01 to 0.07, by a factor of seven. The water table height produced by the antecedent rainfall nevertheless remains between 1% and 7% of the thickness of the soil cover over the whole range of thresholds, against an h_e/z stable between 0.88 and 0.90.

With the parameters unchanged, the effect on the metrics is negative in sign. Passing from unlimited accumulation to the 10 m threshold, the accuracy decreases from 0.596 to 0.589. The true positive rate falls by 0.041 while the true negative rate rises by 0.026: shortening the threshold loses more observed landslides than it recovers stable points.

A final observation concerns the information carried by the variable on which the threshold acts. At the validation points the contributing area has a median of 663 m² where the landslides were triggered and 644 m² where they were not, and the two distributions cannot be distinguished. The model, for its part, predicts as unstable the cells with a small area: in the base configuration the false positives have a median of 499 m² and the true positives 607, against 934 for the true negatives and 1116 for the false negatives.

Calibrated parameters. The constraint on the true negative rate in the initial condition is satisfied by 94 solutions for the 20 m threshold and by 95 for the 50 m one. In both cases the solution adopted is the one with the highest final accuracy among the admissible ones, as for the base configuration. The parameters are reported in Table 5.8, compared with the literature values and with those of the base configuration.

Table 5.8: Parameters of class 7 obtained from the recalibration of the two threshold scenarios, compared with the initial values from the literature and with those of the base configuration.

Parameter	Symbol (unit)	Literature	Base	20 m	50 m
Soil cohesion, maximum	$C_{s,max}$ (kPa)	2	1.34 ^a	1.02 ^a	1.05 ^a
Root cohesion, forest	$C_{r,max}$ (kPa)	5	3.95	6.67	4.43
Friction angle, minimum	φ_{min} (°)	30	22.4	22.7	24.0
Friction angle, maximum	φ_{max} (°)	40	36.4	32.3 ^a	34.7
Thickness	z (m)	1.00	0.82 ^a	0.89 ^a	0.84 ^a
Hydraulic conductivity	K (10 ⁻⁶ m/s)	10.0	33.4	7.47	31.0
Porosity	n (-)	0.35	0.25 ^a	0.27	0.25 ^a

^aValue falling within the outer 10% of the corresponding admissible interval (Table 4.5).

In the 50 m scenario the values differ from those of the base configuration, but the overall picture remains much the same. The porosity coincides with the lower limit of the interval, the thickness stays close to it, and the hydraulic conductivity is still about three times the literature value. The largest difference is in the soil cohesion, lower by a fifth. A structural modification followed by a complete recalibration therefore returns a parameterisation qualitatively equivalent to that of the base configuration.

The 20 m scenario does shift the parameterisation, and not in a single direction. Two parameters move closer to the literature values: the hydraulic conductivity goes from 3.3 times higher to 1.3 times lower, and the porosity moves away from the lower limit of the interval. Two others move in the opposite direction, until they press against their own limits. The maximum friction angle stops three tenths of a degree above the admissible minimum, and the soil cohesion two hundredths of a kPa above its own.

The solutions adopted are not, however, the only ones available. The admissible set of the 20 m scenario includes a solution with the same conductivity and the same thickness as the base configuration, within one hundredth of accuracy of the adopted one. All three sets, that of the base configuration included, also contain one with a conductivity essentially equal to the literature value, likewise within one hundredth. The accuracies quoted here are the cell accuracies computed by the analytical evaluation of the tool, as in Table 5.7, and are not comparable with the validated ones reported below. Even with the strongest truncation, finally, the admissible set contains solutions in which the antecedent component rises to 19% of the thickness of the soil cover. The threshold therefore places no ceiling on h_a .

Model output. The 50 m scenario returns domain statistics that coincide with those of the base configuration: 42.10% of the cells are unstable in the final condition ($\text{PoF} > 0.5$) against 42.11%, and 4.31% are at near-certain failure ($\text{PoF} > 0.99$) against 4.59%. The probability of failure map is not reported, as it is not distinguishable from that of the base configuration (Figure 5.9).

The 20 m scenario, by contrast, produces a markedly different picture. The unstable cells fall to 34.36% and those at near-certain failure to 1.56%, while the intermediate band of uncertainty rises to 36.72%. The model is therefore more stable overall and less polarised. The unconditionally unstable cells rise from 0.28% to 0.46% and the unconditionally stable ones fall from 25.35% to 24.17% (probability > 0.99 in both cases).

Figure 5.12 shows the difference between the probabilities of failure of the base configuration with multi-objective automatic calibration (Section 5.2.2) and of the configuration with the 20 m threshold. The difference is absent along the valley floors and in the gently sloping sectors, and is widespread and positive over the hillslopes, where the probability of failure is higher in the base configuration.

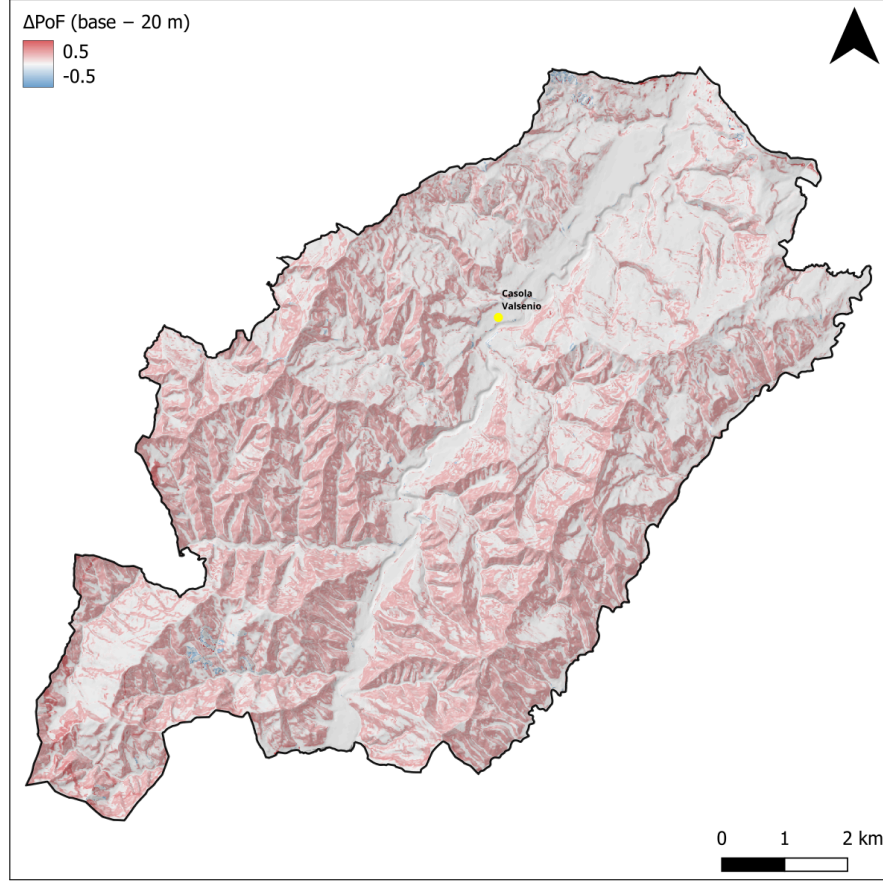


Figure 5.12: Difference between the probability of failure of the base configuration and that of the 20 m threshold configuration, in the final condition. Positive values indicate a higher probability in the base configuration.

Table 5.9: Median difference between the probabilities of failure of the two configurations (base – 20 m threshold), by classes of equal size. (a) Slope classes, over the whole domain. (b) Drainage-area classes, restricted to the cells with slope between 20 and 35°; the first row reports the median fraction of the drainage area removed by the 20 m threshold in each class.

(a) Slope classes					
Slope (°)	0–12.8	12.8–19.5	19.5–26.1	26.1–32.7	> 32.7
Median ΔPoF	0.00	0.00	0.05	0.20	0.15

(b) Drainage-area classes, slope 20–35°					
Drainage area (m ²)	100–300	300–470	470–720	720–1290	> 1290
Fraction removed at 20 m	23%	51%	66%	78%	91%
Median ΔPoF	0.17	0.18	0.18	0.15	0.12

Table 5.9 reports the median difference by slope and drainage-area classes. The magnitude of the difference is governed by slope: zero below 20°, highest in the band between 26 and 33°, slightly lower on the steepest slopes. The drainage area does not produce a comparable pattern. With the slope restricted to between 20 and 35°, so as to remove the effect just described, the difference remains between 0.12 and 0.18 in

all classes, whereas the truncation acts on them very unevenly, removing from 23% to 91% of the drainage area. If the difference between the two configurations came from the truncation, it should follow that progression. It runs the other way, and is largest where the truncation removes least.

The map and the table say the same thing in two different ways. The drainage area is largest in the hollows and valley floors, where the truncation removes nine tenths of the upstream contribution, and it is precisely there that the map shows no difference. Over the hillslopes, where the truncation removes less than a quarter, the difference is widespread and systematic.

Validation in the final condition. The 50 m scenario correctly identifies 81.5% of the observed landslides ($\text{TPR} = 0.815$) and 50.5% of the stable points ($\text{TNR} = 0.505$), with an accuracy of 0.660 and an area under the ROC curve of 0.734. The values coincide with those of the base configuration to within a few thousandths, and the counts of the confusion matrix differ from those of the base by no more than three points in each of the four categories. The matrix is therefore not reported.

The 20 m scenario correctly identifies 72.7% of the observed landslides ($\text{TPR} = 0.727$) and 60.0% of the stable points ($\text{TNR} = 0.600$), with an accuracy of 0.664 and an area under the ROC curve of 0.737. The confusion matrix is reported in Figure 5.13. These values are reported, together with those of the two preceding hydrological conditions, in Table 5.10.

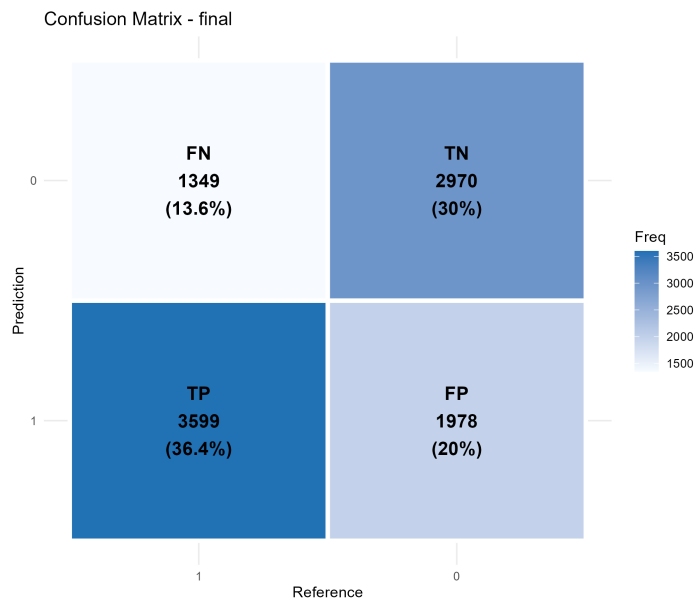


Figure 5.13: Confusion matrix of the 20 m threshold configuration, in the final condition.

In this configuration too the true and false positives remain indistinguishable, with mean probabilities of failure of 0.82 and 0.74 and mean slopes of 32.6° and 32.5° . Both mean probabilities are lower than those of the base configuration (0.89 and 0.84), consistently with the general shift towards stability.

Compared with the base configuration, the true positive rate falls by 0.087 and the true negative rate rises by 0.095, while the accuracy increases by 0.004 and the area under the ROC curve by 0.005. A shift of about nine percentage points in the two rates, in opposite directions, therefore translates into four thousandths of accuracy.

Progression of the hydrological conditions. The progression across the three hydrological conditions is reported in Table 5.10.

Table 5.10: Validation metrics of the two threshold scenarios, in the three hydrological conditions (Section 4.1.6).

Configuration	Condition	TPR	TNR	ACC	AUC
50 m threshold	Dry	0.198	0.968	0.583	0.713
	Initial (pre-event)	0.202	0.965	0.584	0.714
	Final (post-event)	0.815	0.505	0.660	0.734
20 m threshold	Dry	0.154	0.982	0.568	0.717
	Initial (pre-event)	0.165	0.980	0.572	0.718
	Final (post-event)	0.727	0.600	0.664	0.737

Both scenarios retain the consistency of the pre-event state obtained with the constraint. In the initial condition the initiation points predicted as unstable are 20.2% in the 50 m scenario, a value that coincides with the 20.3% of the base configuration, and 16.5% in the 20 m scenario: the shorter threshold is therefore slightly more conservative.

The partition of the water table rise between the two hydrological components (Section 4.1.3) is essentially the same in the three configurations: at the validation points h_a/z lies between 0.02 and 0.03 and h_e/z between 0.97 and 0.98. This is not a coincidence. The 20 m scenario adopts a hydraulic conductivity 4.5 times lower than that of the base configuration, which on its own would raise h_a by the same factor, while the truncation reduces its contributing area by 3.3 times. The two effects nearly cancel, as Equation 4.2 requires. The partition is therefore unchanged, but not because the threshold left it untouched.

The transition from the dry to the initial condition, which directly measures the contribution of the antecedent rainfall, confirms this: the TPR increases by 0.013 in the base configuration, by 0.005 in the 50 m scenario and by 0.011 in the 20 m scenario. In none of the three configurations does the antecedent rainfall appreciably modify the state of the slopes.

The area under the ROC curve remains between 0.713 and 0.737 across all six combinations of configuration and hydrological condition.

Interpretation: strengths and limitations. The 50 m scenario is inert. A structural modification followed by a complete recalibration returns a configuration virtually indistinguishable from the base one on every measure: domain statistics, validation metrics, counts of the confusion matrix. The parameters, however, do not coincide, and the soil cohesion differs by a fifth. It is not the same configuration: it is a different parameter set returning the same outcome.

The 20 m scenario, by contrast, appreciably shifts the operating point. The threshold is the only input that differs, and it is therefore what led the calibration to a different parameter set. The difference between the two maps, however, is not produced by the truncation of the drainage area, but by the parameters that the recalibration adopted. Three observations support this. The difference from the base configuration follows slope and not drainage area, and across the area classes it runs in the opposite direction to the fraction truncated. The unconditionally unstable cells increase, and they

are unstable under dry conditions, in which the antecedent component plays no part. Finally, with the parameters fixed, the effect of the threshold on the accuracy is almost zero and negative in sign. What is observed therefore comes from the parameter set adopted, and in particular from the reduction of the maximum friction angle, which acts as a function of slope.

The limited effectiveness of the modification has two reasons. The first is contained in the formulation: the truncation acts in the right direction, but it can only reduce a term that was already small. In the configurations compared here h_a/z does not exceed 0.03 against an h_e/z close to 0.97, and with the literature parameters it stops at 0.07 against 0.89. Removing a fraction of a contribution of this size cannot modify the stability of a cell.

The second reason concerns the nature of the error. The modification corrects the overestimation of the water table in the cells with a large drainage area, but that is not where the model errs: the false positives have a median contributing area of 499 m² against 654 m² for the whole sample, whereas the true negatives lie at 934 m². The threshold therefore acts on a minority mechanism, recovering the false positives located in convergent positions, which are worth a few points of true negative rate, and leaves everything else unchanged.

The modification therefore does what it was designed to do, reducing the contributing area and with it the antecedent component, but in this study area that reduction has no direct consequences. And its recalibration shows, as the previous ones did, that the objective adopted does not identify the parameters. The intervention on the antecedent component has given what it could give; the modification presented in the following section acts on another term of the stability equation, the thickness of the soil cover.

5.4 Distributed Soil-Thickness Configuration

The modification allows the soil thickness to be supplied as a distributed raster, in place of the constant value per soil class (Section 4.1.7). The map used is the normalised one obtained with the GIST model (Section 4.2.5); the calibrated parameter is no longer the thickness, but the scale factor k that multiplies it (Equation 4.16). The remaining parameters, together with k , were recalibrated with the multi-objective branch (Section 4.4.4), following the same selection rule adopted for the preceding configurations.

The configuration runs on the version of the code that implements the accumulation-length threshold, with the threshold set at 50 m. The comparison is therefore with the constant-thickness configuration calibrated at the same threshold (Section 5.3), which is also the one for which the hydrological modification proves inert. Both configurations were recalibrated in full, however, so the effect of the map cannot be separated from that of the recalibration.

Calibrated parameters. The constraint is satisfied by 80 solutions. The parameters adopted are reported in Table 5.11, compared with the literature values and with those of the constant-thickness configuration at the same threshold.

Table 5.11: Parameters of class 7 obtained from the recalibration of the distributed-thickness configuration, compared with the initial values from the literature and with those of the constant-thickness configuration with a 50 m threshold.

Parameter	Symbol (unit)	Literature	50 m	Distributed
Soil cohesion, maximum	$C_{s,max}$ (kPa)	2	1.05 ^a	1.00 ^a
Root cohesion, forest	$C_{r,max}$ (kPa)	5	4.43	3.83
Friction angle, minimum	φ_{min} (°)	30	24.0	26.7
Friction angle, maximum	φ_{max} (°)	40	34.7	38.8
Thickness ^b	z, k (m)	1.00	0.84 ^a	1.694
Hydraulic conductivity	K (10 ⁻⁶ m/s)	10.0	31.0	3.97
Porosity	n (-)	0.35	0.25 ^a	0.25 ^a

^aValue falling within the outer 10% of the corresponding admissible interval (Table 4.5). ^bIn the first two columns the value is the constant thickness z ; in the third it is the scale factor k , which coincides with the mean thickness of the soil cover over the domain, since z_{norm} has unit mean (Section 4.4.4).

The friction angle moves appreciably closer to the literature values, from 24.0–34.7° to 26.7–38.8°, against the 30–40° of the literature. The hydraulic conductivity, by contrast, remains of the same order of magnitude as the literature value.

The scale factor of the thickness calls for a separate consideration. At 1.694 it lies in a central position within the admissible interval, whereas in the constant-thickness configurations the thickness was pressed against the minimum of 0.8 m: the calibration is therefore no longer constrained by the limit of the search space. The value nevertheless corresponds to a mean thickness of 1.69 m over the domain, further from the literature value of one metre than the 0.84 m of the constant-thickness configuration.

The soil cohesion and the porosity remain at their limits, both exactly at the admissible minimum.

The resulting thickness map in metres is shown in Figure 5.14. The spatial distribution is the one already described in Section 4.2.5: thicker covers in the concave sectors and on the gently sloping hillsides, thinner ones on the ridges and on the steep slopes.

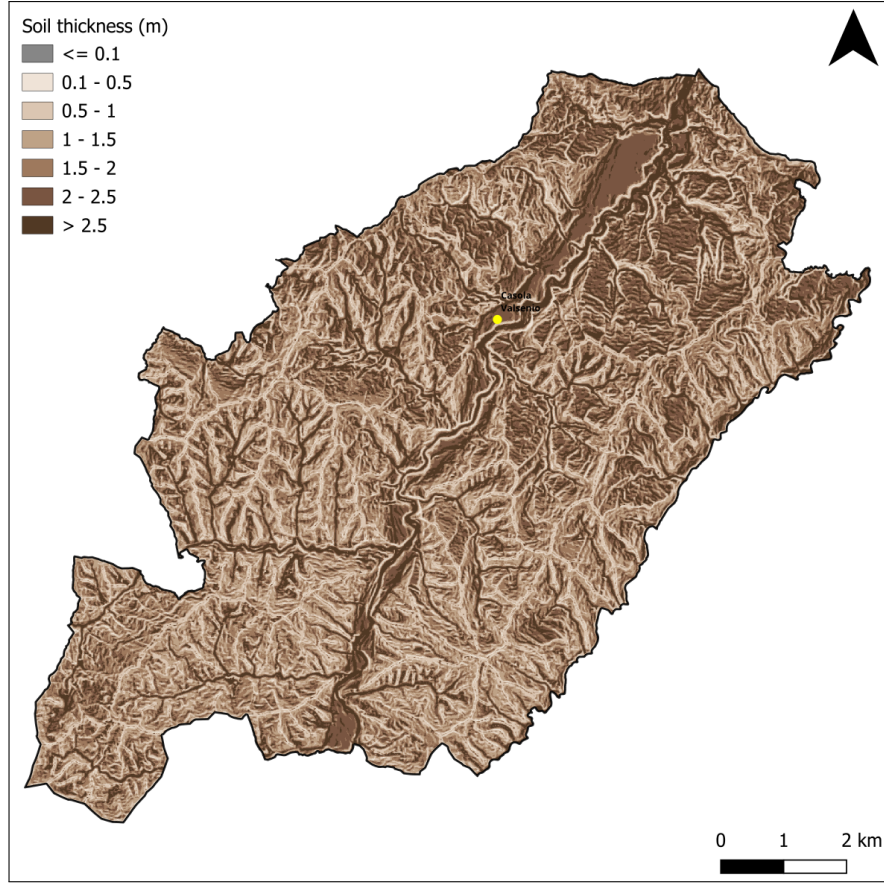


Figure 5.14: Soil thickness in metres, obtained by scaling the normalised map of Figure 4.8 with the factor $k = 1.694$ determined by the calibration. The minimum value of 0.10 m is the lower limit imposed by the model (Equation 4.16).

Over the calculation domain the mean thickness is 1.69 m and the maximum 3.84 m. At the validation points the map distinguishes the two observed classes only slightly (Table 5.12): the landslide points have a median of 1.28 m against 1.57 m for the stable points, a difference of 0.29 m; with the slope restricted to between 20 and 35°, so as to remove the effect of slope, the difference falls to 0.08 m, against an overall interquartile range of 0.83 m.

Table 5.12: Thickness of the soil cover at the validation points, distributed-thickness configuration: distribution by observed class, with and without a restriction on slope (20 – 35°).

Sample	Class	n	Q1 (m)	Median (m)	Q3 (m)
All	Landslide	4948	0.83	1.28	1.77
	Non-landslide	4948	1.14	1.57	2.11
Slope 20–35°	Landslide	2902	0.94	1.38	1.81
	Non-landslide	2605	1.06	1.46	1.83

Model output. Figure 5.15 shows the probability of failure map in the final condition.

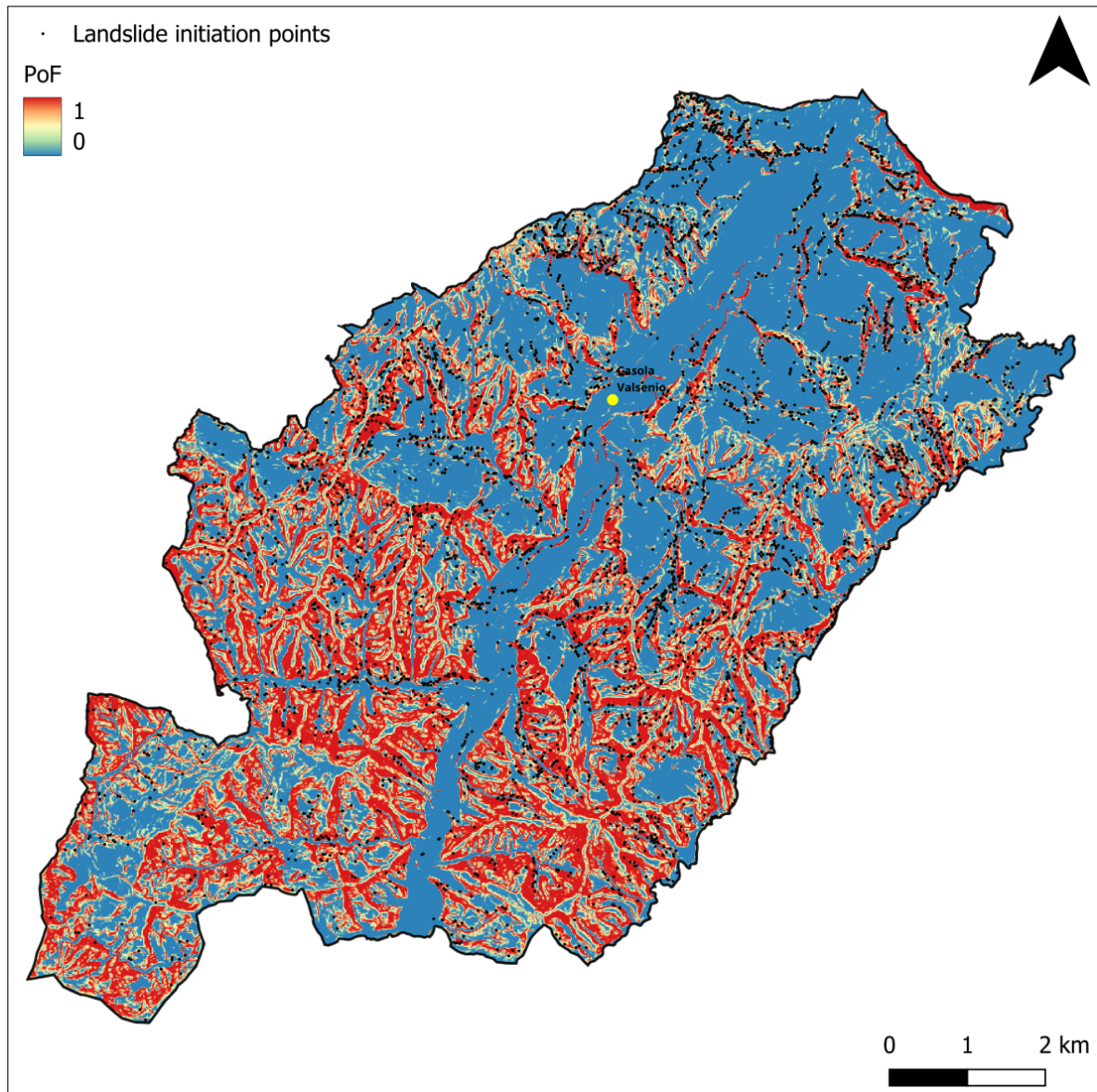


Figure 5.15: Probability of failure (PoF) map in the final condition for the distributed-thickness configuration. In black, the initiation points of the May 2023 inventory (DS + DF).

Over the municipal territory, 37.68% of the cells are unstable in the final condition ($PoF > 0.5$), against 42.10% for the constant-thickness configuration, while the essentially stable cells ($PoF \leq 0.01$) rise from 29.82% to 41.25%. The cells at near-certain failure ($PoF > 0.99$) increase from 4.31% to 14.60%. The cells whose stability does not depend on rainfall are 0.26% unconditionally unstable and 27.25% unconditionally stable (probability > 0.99 in both cases).

The distinctive feature of this configuration is the polarisation of the probability surface: the cells falling into one of the two extreme classes ($PoF \leq 0.01$ or $PoF > 0.99$) rise from 34.1% to 55.9% of the municipal territory. The map consequently shows a sharply contrasted picture, in which the instability is arranged in continuous bands

along the steep slopes, whereas the broad interfluvies and the valley floors are clearly stable, with a narrow transition zone.

Validation in the final condition. The model correctly identifies 75.2% of the observed landslides ($\text{TPR} = 0.752$) and 55.4% of the stable points ($\text{TNR} = 0.554$), with an accuracy of 0.653 and an area under the ROC curve of 0.703. The confusion matrix is reported in Figure 5.16. These values are reported, together with those of the two preceding hydrological conditions, in Table 5.13.

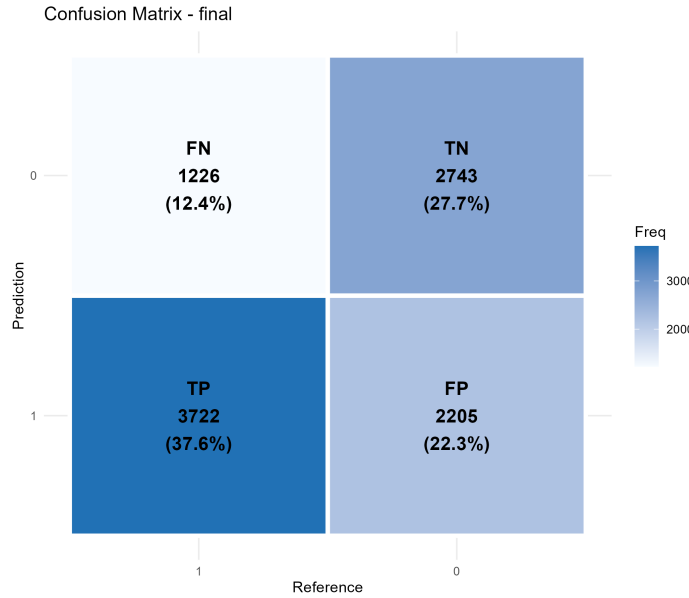


Figure 5.16: Confusion matrix of the distributed-thickness configuration, in the final condition.

The true and false positives remain indistinguishable, with mean probabilities of failure of 0.92 and 0.89 and mean slopes of 32.6° and 32.1° . The gap between the two mean probabilities, however, narrows to three hundredths, against five in the constant-thickness configuration. The two negative categories also move downwards, to 0.09 and 0.14: all four categories therefore move closer to the extremes, consistently with the polarisation observed over the domain.

The model also responds markedly to the thickness: in the $20\text{--}35^\circ$ slope band, the points predicted as unstable have mean thicknesses of 1.30 and 1.39 m, against 1.63 and 1.77 m for those predicted as stable.

Compared with the constant-thickness configuration, the true positive rate falls by 0.062 and the true negative rate rises by 0.049; the accuracy decreases by seven thousandths, whereas the area under the ROC curve decreases by thirty-two.

Progression of the hydrological conditions. The progression across the three hydrological conditions is reported in Table 5.13.

Table 5.13: Validation metrics of the distributed-thickness configuration, in the three hydrological conditions (Section 4.1.6).

Condition	TPR	TNR	ACC	AUC
Dry	0.159	0.969	0.564	0.704
Initial (pre-event)	0.194	0.953	0.574	0.708
Final (post-event)	0.752	0.554	0.653	0.703

The state preceding the event is preserved: in the initial condition 19.4% of the initiation points are predicted as unstable, against 20.2% for the constant-thickness configuration.

The partition of the water table rise between the two hydrological components (Section 4.1.3), by contrast, changes markedly. At the validation points h_a/z passes from 0.02 to 0.08 and h_e/z from 0.98 to 0.79. The change derives from the two parameters that govern the ratio: the hydraulic conductivity has fallen by almost an order of magnitude, and h_a is inversely proportional to it (Equation 4.2), while the greater thickness of the soil cover only partly offsets its effect.

The transition from the dry to the initial condition confirms this on an independent measure: the true positive rate passes from 0.159 to 0.194, an increase of 0.035, against 0.005 for the constant-thickness configuration.

The antecedent rainfall therefore takes on a non-negligible weight without the pre-event state being compromised. In the single-objective calibration a high antecedent component was accompanied by a physically implausible pre-event instability (Section 5.2.1); here the true positive rate in the initial condition remains at 0.194.

The area under the ROC curve remains between 0.703 and 0.708 across the three conditions, as in all the configurations examined.

Interpretation: strengths and limitations. The distributed representation of the thickness lowers the discriminating capacity of the model. The area under the ROC curve falls from 0.734 to 0.703, whereas the accuracy decreases by seven thousandths. The two metrics therefore do not describe the same change: what worsens is the ranking of the points. This is the opposite of the behaviour of the accumulation-length threshold (Section 5.3), which left the area under the curve unchanged.

The mechanism can be read in the validation data. The model responds markedly to the thickness, but the thickness does not separate the observed classes. The probabilities are therefore shifted considerably, and on the basis of a variable that does not distinguish the points that failed from those that did not: the ranking changes without moving closer to the observed one. The polarisation described above is the same effect seen at the scale of the domain: the model becomes sharper without becoming more correct.

The two effects, however, cannot be separated. The distributed configuration has both a different representation of the thickness and different parameters, since both configurations were recalibrated in full; in Section 5.3 the same ambiguity was resolved by keeping the parameters at the literature values, a control that is not available here. The

comparison between the observed classes, by contrast, involves no calibrated parameter: it can therefore be stated that the map carries little information on the observed class, and that this is consistent with the loss observed, but it cannot be established how much of that loss comes from the map and how much from the recalibration.

The balance on the plausibility of the parameters is mixed: the friction angle moves closer to the literature values, the mean thickness and the root cohesion move away from them, whereas the soil cohesion and the porosity remain at the admissible minimum. At the initiation points, moreover, 16% of the modelled covers exceed the 2 m indicated in Section 2.1.2 as the thickness characteristic of shallow landslides.

The hydrological partition does not provide a counterweight. The antecedent component gains weight, but the selection rule adopted (Section 4.4.3) treats as more plausible a partition with a small antecedent component, which is the one all the other configurations return; the change follows from the lower hydraulic conductivity, and nothing in the results indicates that the greater weight is the more correct. The trade-off offered by this configuration is therefore a measurable loss of discriminating capacity against a partial gain in the plausibility of the parameters: the opposite of what is observed for the accumulation-length threshold, where the modification cost nothing and yielded a small gain in plausibility. This is the first time in the chapter that performance and plausibility move in opposite directions.

The result does not show that the thickness of the soil cover matters little. It shows that this map, which derives the thickness from curvature and slope (Section 4.2.5), does not carry the information needed to identify the slopes that failed: the limitation concerns the map adopted, and the result does not allow any conclusion on the variable itself.

5.5 Comparative Analysis

The preceding sections have examined the six configurations one at a time. A comparison between them is possible because they all share the validation dataset, the search space and the selection rule.

Across six configurations that differ in calibration strategy, hydrological formulation and representation of the soil thickness, the accuracy varies by 0.012 and the AUC by 0.037 (Table 5.14). The latter range is almost entirely produced by a single configuration. Without it the spread narrows to 0.009. What does change appreciably is of another kind: the operating point of the model, the physical plausibility of the calibrated parameters and the consistency of the pre-event state.

The variation in accuracy is limited in turn. Since the validation dataset is balanced, with 4948 points per class, the accuracy coincides with the mean of TPR and TNR, and its 95% confidence interval is ± 0.009 , virtually identical for all six configurations. The calculation treats the points as independent, whereas they are spatially correlated, so the interval is a lower bound on the true uncertainty. The six accuracies therefore fall within the width of a single interval.

A final check concerns the threshold at which the points are classified (Section 4.3.2). When the accuracy is computed as a function of that threshold, its maximum falls between 0.47 and 0.73 for all six configurations, and the value at 0.5 departs from that maximum by at most 0.006, smaller than the confidence interval just given. The ranking of the six is also the same whether the common threshold of 0.5 or the optimum

of each configuration is adopted: fixing the threshold in advance, rather than optimising it configuration by configuration, therefore does not alter the comparisons that follow.

The overall picture. The validation metrics of the six configurations are reported in Table 5.14, the calibrated parameters in Table 5.15. Both tables also include the uncalibrated literature set, as a reference.

In the final condition the true positive rate ranges from 0.727 to 0.815 and the true negative rate from 0.505 to 0.600, spanning 0.087 and 0.095 respectively. In the initial condition the true positive rate varies from 0.165 to 0.603.

The calibrated parameters cover much wider intervals. The hydraulic conductivity ranges from 9.1×10^{-7} to 3.34×10^{-5} m/s, a factor of 37, and the soil thickness from 0.82 to 1.97 m, a factor of 2.4.

The contrast between the two sets of numbers is the first outcome of the comparison: the parameters vary by factors, the performance by thousandths.

Table 5.14: Validation metrics of the six configurations in the final condition, with the true positive rate in the initial condition. The first row is the uncalibrated literature set, reported as a reference.

Configuration	TPR	TNR	ACC	AUC	TPR _{initial}
Literature ^a	0.523	0.739	0.631	0.712	0.067
Manual calibration	0.813	0.509	0.661	0.731	0.418
Single-objective	0.804	0.526	0.665	0.740	0.603
Multi-objective	0.814	0.505	0.659	0.732	0.203
20 m threshold	0.727	0.600	0.664	0.737	0.165
50 m threshold	0.815	0.505	0.660	0.734	0.202
Distributed thickness	0.752	0.554	0.653	0.703	0.194

^aNot calibrated; parameter set taken from the literature (Section 4.4.1).

Table 5.15: Calibrated parameters of the six configurations. The first row is the uncalibrated literature set, reported as a reference.

Configuration	$C_{s,max}$ (kPa)	C_r forest (kPa)	φ_{min} (°)	φ_{max} (°)	z (m)	K (10^{-6} m/s)	n (–)
Literature ^a	2.00	5.00	30.0	40.0	1.00	10.0	0.350
Manual calibration	2.00	5.00	22.0	34.0	1.25	4.00	0.350
Single-objective	2.65	7.53	20.3 ^b	36.6	1.97 ^b	0.91 ^b	0.370
Multi-objective	1.34 ^b	3.95	22.4	36.4	0.82 ^b	33.4	0.250 ^b
20 m threshold	1.02 ^b	6.67	22.7	32.3 ^b	0.89 ^b	7.47	0.270
50 m threshold	1.05 ^b	4.43	24.0	34.7	0.84 ^b	31.0	0.250 ^b
Distributed thickness	1.00 ^b	3.83	26.7	38.8	1.694 ^c	3.97	0.250 ^b

^aNot calibrated. ^bValue falling within the outer 10% of the corresponding admissible interval (Table 4.5). ^cScale factor k applied to the normalised thickness map, not a constant thickness (Section 4.2.5).

Variation of the operating point. The differences between configurations do not concern how accurate the model is, but where its operating point falls. The true positive

and true negative rates move in opposite directions, exchanging about nine percentage points along an accuracy that remains flat: the configuration with the 20 m threshold has the lowest TPR, 0.727, and at the same time the highest TNR, 0.600. Unlike those in accuracy, these variations are well above the uncertainty of the respective estimates.

The operating point is determined by the parameter set rather than by the direct effect of the structural modifications. For the accumulation-length threshold the comparison at fixed parameters shows this directly: moving the threshold from 20 to 50 m shifts the true positive rate by 0.011, whereas between the two recalibrated configurations the difference is 0.087 (Section 5.3). For the distributed thickness an equivalent control is not available, because the scale factor is itself a calibrated parameter, and the two effects cannot be separated (Section 5.4). The comparison of thickness between the observed classes, which involves no calibrated parameter, nevertheless shows that this variable carries little discriminating information. Since each configuration was recalibrated, every modification was in any case evaluated with the parameters most favourable to it.

The shift is also visible over the whole domain: the proportion of the municipality classified as unstable ($\text{PoF} > 0.5$) ranges from 34.4% to 42.1% depending on the configuration.

The largest range, however, is not in the final condition. In the initial condition the true positive rate varies from 0.165 to 0.603, a range of 0.438. It is the largest difference of the entire comparison, and it concerns the state of the slopes before the event.

It is also the quantity that the objective adopted in the first two branches does not measure. Both the manual and the single-objective calibrations maximise the accuracy in the final condition, a criterion that leaves the pre-event state free to fall anywhere. This is why in the multi-objective branch it was constrained by a second objective (Section 5.2.2). What the metric does not measure, the calibration does not control.

The three calibration branches. The first three branches differ in the calibration procedure and in the objective on which it is carried out. The two transitions are not of the same kind. The transition from manual to automatic single-objective calibration changes only the procedure, and produces a small but solid gain in discriminating capacity: the AUC rises from 0.731 to 0.740 (+0.009). In the same transition, however, the proportion of initiation points predicted as unstable before the event rises from 41.8% to 60.3%. The transition from the single- to the multi-objective branch changes the objective, and with it the algorithm and the selection rule (Section 4.4.3). It gives back almost exactly what the first had gained: the AUC returns to 0.732 (−0.008), while the proportion of initiation points already unstable falls to 20.3%. Of the two effects, only the fall in the proportion of points already unstable is attributable to the objective. The true negative rate in the initial condition does not enter the objective function of the single-objective branch, and a search cannot optimise what it does not measure. The difference in AUC cannot be attributed in the same way, since in it the three differences act together. At the two ends of the sequence, between the manual and the multi-objective calibration, the AUC is the same and the proportion of initiation points already unstable has halved.

That the calibration strategy is not what limits the model emerges from three independent observations, all of them already presented. The eight runs of the single-objective algorithm, started from different seeds, converge within 0.0075 of cell accu-

racy while adopting very different parameters (Section 5.2.1). The single- and multi-objective solutions have opposite parameterisations, with thicknesses of 1.97 and 0.82 m and conductivities a factor of 37 apart, and differ in AUC by eight thousandths. The Pareto front of the multi-objective branch contains 86 non-dominated solutions, that is, as many equivalent compromises between the two objectives (Section 5.2.2). Very different parameterisations produce very similar performance: choosing between them is not choosing accuracy.

The ceiling of the AUC. The AUC of the six configurations lies between 0.703 and 0.740. Figure 5.17 superimposes the six ROC curves in the final condition: five remain within a band of nine thousandths, while the curve of the distributed-thickness configuration runs below the others. Unlike the operating point, the capacity to rank the points is not appreciably increased by any of the modelling choices examined.

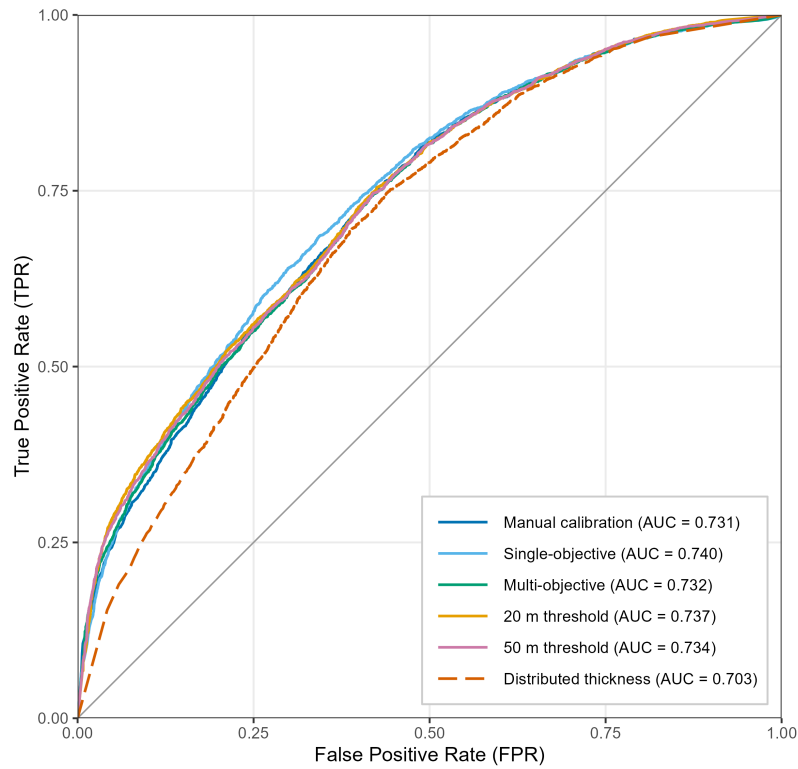


Figure 5.17: ROC curves of the six configurations in the final condition, computed on the validation dataset (Section 4.3.1). The area under each curve is reported in the legend.

Where that capacity reaches its limit can be read in the mean slopes of the four classification categories. In the multi-objective configuration the true positives fall on slopes of 31.9° on average and the false positives on slopes of 31.4° , while the true negatives lie at 18.6° and the false negatives at 21.1° . The model separates steep from gentle slopes clearly, but among the steep slopes it does not distinguish those that failed from those that did not. This is the third limitation observed in Section 5.1, here measured across all the categories.

The contribution of the hydrological components can be quantified, because the three conditions are evaluated on the same points and with the same rule. In the dry condition, in which no water is involved, the AUC already lies between 0.704 and 0.719. Adding the antecedent and the event rainfall brings it to between 0.703 and

0.741. Over the same progression the true positive rate goes from 0.15 to 0.82. The hydrology moves the operating point, not the ranking.

The distributed-thickness configuration shows that adding a spatial variable does not necessarily raise the ceiling. The thickness map, derived from the same digital elevation model from which the slope comes, brings the AUC to 0.703, below even that of the uncalibrated literature set. It is also the only configuration in which the transition from the dry to the final condition does not improve the ranking. The resulting map is more polarised: 14.6% of the cells are at near-certain failure ($\text{PoF} > 0.99$), against values between 1.6% and 8.0% in the other five, and 41.3% are essentially stable ($\text{PoF} \leq 0.01$) against 24.5 to 29.8%. Sharper, not more correct. The same ordering is found in the calibration, where the maxima of the four Pareto fronts are 0.6296, 0.6294 and 0.6291 for the constant-thickness configurations and 0.6176 for the distributed one.

The three arguments converge. The discriminating capacity is not limited by the calibration strategy, which leaves it where it is. It is not limited by the hydrological formulation, which is worth two hundredths of AUC. And it is not raised by adding a variable derived from the same topographic source. The ceiling is set by the spatial information available: the variables at the model's disposal describe well which slopes are predisposed to failure, and much less well which of the predisposed ones actually failed.

Accuracy and physical plausibility. The comparison has shown that performance does not separate the six configurations. The accuracy varies within the precision with which it is known, and the AUC within 0.037. What separates them lies elsewhere: the physical plausibility of the calibrated parameters and the consistency of the state of the slopes before the event.

Reichenbach et al. (2018) observe that a model with a higher AUC has better statistical performance than one with a lower AUC, but that the second may prove more meaningful from a geomorphological perspective, because its errors may be less severe or distributed over the territory in a more sensible way. The present case is an example. The six configurations differ little in statistical terms, but markedly in the way they represent the state of the slopes before the event, where the proportion of initiation points already predicted as unstable goes from 16.5% to 60.3%.

On this basis the base configuration with multi-objective calibration is the one to be preferred. The choice does not fall on the highest AUC, which belongs to the single-objective branch: selecting the highest value after the fact would mean adopting a criterion different from the one declared in advance for all six configurations. It falls instead on the only branch whose selection rule included a criterion of physical plausibility, fixed before the results were observed, and which yields a consistent pre-event state at the cost of a difference in accuracy smaller than the precision of the measurement.

No configuration, however, dominates the others in every respect. The 20 m threshold obtains a hydraulic conductivity closer to the literature values, but that value was not selected: it is the by-product of a recalibration (Section 5.3) whose distinctive feature, the threshold, proved ineffective in this study area. This is the distinction that the comparison suggests adopting: physical plausibility counts as a criterion when it is declared in advance, not when it is incidental.

6 – Conclusions

This work has applied a physically-based model to the assessment of shallow landslide susceptibility over the municipal territory of Casola Valsenio. The model was calibrated on the inventory of the landslides triggered by the May 2023 event (Berti et al., 2025), and its fit was measured against the same data. The aim was not to obtain the highest possible performance from FSLAM, but to test different calibration strategies and structural modifications, and to assess their effects on the present application.

To this end six configurations were built, in two groups. The first three use the model as published and differ in the calibration strategy (Section 4.4). From the manual calibration to the automatic single-objective one, only the search procedure changes; from the latter to the multi-objective one, the pre-event constraint is added, together with the algorithm and the selection rule that it entails. The other three introduce two modifications to the formulation (Section 4.1.7), the length threshold on the flow accumulation and the soil thickness supplied as a distributed map: both were implemented in the code and recalibrated with the multi-objective branch. Three elements were defined before the results were observed and were not adjusted afterwards: the validation dataset (Section 4.3.1) and the search space of the parameters (Section 4.4.1), common to all six configurations, and the selection rule (Section 4.4.3), applied to the four configurations calibrated on the Pareto front.

It is this last aspect that makes the six outcomes a comparison, and not six separate applications. The differences observed between the configurations can be attributed to the choices that distinguish them; the absolute values, by contrast, remain tied to the validation and calibration design adopted.

6.1 Main Results

The calibration strategy is not what limits the model. Parameter sets that differ widely produce equivalent performance, and nothing in the performance indicates which of them to adopt. Three independent results point the same way. The eight runs of the single-objective algorithm converge on almost coincident accuracies, although they start from parameters far apart (Section 5.2.1). The single- and multi-objective solutions have opposite parameters and close metrics (Section 5.5). The Pareto front of the multi-objective calibration contains dozens of equivalent compromises (Section 5.2.2).

The multi-objective calibration was introduced in order to constrain the pre-event state (Section 4.4.3). The proportion of initiation points already predicted as unstable before the event falls markedly. The cost in accuracy is smaller than the precision with which the accuracy is known. The predictive capacity, however, does not improve: it remains that of the other two branches. The gain is one of physical consistency, and it was the declared aim of the method adopted.

The discriminating capacity was not raised by any of the choices examined. The hydrological components move the operating point, not the ranking: from the dry to the final condition the true positive rate is multiplied by more than five, while the AUC stays where it was. A distributed thickness map, derived from the same source as the slope, adds no information and worsens the ranking (Section 5.4). The error that remains is always the same: among the steep slopes the model does not recognise

which of them failed, and the false positives fall at a mean slope virtually identical to that of the true positives. The limit therefore does not lie in the modifications to the formulation examined here, but in the spatial information available to the model.

The last result is methodological, and it is the only one that holds beyond this case study. Accuracy is not enough to choose the parameters. The calibration returns one set among many possible ones, all of them equivalent in the metrics. The metric does not say which of those sets is the physically most credible. Physical plausibility must therefore be declared as a criterion before the calibration, and not recognised as an outcome afterwards.

6.2 The Role of Antecedent Rainfall

One result concerns the study area rather than the model, and it holds within the range of parameters examined. At the initiation points the antecedent component of the water table is negligible compared with the event component (Section 4.1.3), and it already was with the literature parameters, before any calibration (Section 5.3). The two branches calibrated on the final accuracy alone are the exception (Section 4.4): there the calibration inflated it, to the point of saturating a substantial fraction of the soil cover. This is an outcome of the objective adopted: where the constraint on the pre-event state is present, the antecedent component remains small.

There are two reasons, and the first is geomorphological. The antecedent component grows with the contributing area upslope (Equation 4.2), and at the validation points that area is of a few hundred square metres: these are hillslope positions, not hollows (Section 5.3). Cells that collect from large areas do exist, but they are a minority, and on all the others the long-term recharge cannot raise the water table appreciably.

The second concerns the hydrological conditions that preceded the event. The spring of 2023 closed with an exceptionally dry April (Section 3.4.1), and the antecedent rainfall supplied to the model is derived from the last week of that month (Section 4.2.4). The rainfall of the first May pulse does not enter that term: the whole period 1–17 May is supplied to the model as event rainfall. What the antecedent component represents is therefore the state of the slopes at the end of April. The event rainfall, by contrast, was the largest ever recorded in the region over a period of that length (Section 3.4.4). The difference in scale between the two components is such that the antecedent one cannot be decisive.

What has been said so far concerns the partition of the model inputs adopted in this work. A possible alternative is to take the first May pulse as antecedent rainfall and the second as event rainfall. This would first require the observed landslides to be attributed to a specific pulse. The present inventory does not allow this: the mapping is based on images taken after both pulses, and the distinction is not included in the dataset (Berti et al., 2025). The same work notes that it would be feasible only in limited areas, where high-resolution imagery was acquired after the first event.

It would also require the evolution of the water table between the two pulses to be represented. The model does not provide for this: it is formulated for a single short rainfall event, with an antecedent component intended for a steady long-term recharge, and it includes no water balance (Section 4.1.3). These are the two conditions under which the hypothesis could be tested.

6.3 Limitations

The limitations of the work have been stated in the preceding chapters, each where it arose. It is worth gathering them here, because together they set the boundaries of what the results can support.

- Calibration and validation rest on the same inventory, without an independent test set (Section 4.3.1). The absolute values of the metrics are therefore optimistic: they measure how closely the model fits the data on which it was calibrated, not how well it would predict future events.
- The validation design adopted, with a buffer applied to the initiation points and the stable points sampled at random within the same suitability domain, makes the values not directly comparable with those of other applications of the model, built on different inventories and criteria.
- The calibration is restricted to a single litho-technical unit, on which all the validation points fall (Section 4.4.1). The soil cohesion and the friction angle are therefore constant over the whole sample, and cannot contribute to distinguishing the slopes that failed from those that did not.
- Each modified configuration was recalibrated in full (Sections 5.3 and 5.4). The effect of the modification and that of the new parameter set cannot therefore be separated, except where a comparison with the parameters fixed was possible.
- The thickness of the soil cover is a geomorphological proxy, obtained from the digital elevation model that also yields the slope (Section 4.2.5). The two variables are not independent. Nor is the map independent of the inventory that sets the limits of both its factors, and against which the model's performance is then measured. It is used in one configuration only, and that is the one with the lowest AUC, so the dependence cannot have inflated the results.
- Two parameters were not calibrated freely: the root cohesion is scaled by a single multiplier applied to all the land-cover classes (Section 4.4.1), and the effective infiltration coefficient is fixed at 30% without a water balance (Section 4.2.4).
- The event rainfall spans seventeen days and two pulses two weeks apart, and enters the model as a single input, converted into one instantaneous rise of the water table (Sections 4.1.3 and 4.2.4). The static formulation does not represent the drainage that took place between them, so the partition between the two hydrological components does not follow the actual sequence of wetting.

6.4 Perspectives

In addition to the partition between antecedent and event rainfall, discussed in Section 6.2, the work leaves two further directions open.

The first concerns the thickness of the soil cover. The distributed-thickness configuration is its experimental motivation (Section 5.4): the map adopted derives from the topography, with thresholds set from the initiation points, and carries no information able to distinguish the observed classes. A campaign of direct measurements, even

on a limited number of points, would replace a geomorphological proxy with measured values, and would give the model, among the predisposed slopes, information that none of its present inputs provides.

The second is the geotechnical characterisation of local samples in the laboratory. The geotechnical parameters adopted do not come from measurements: they derive from a regional litho-technical classification, and their values are those of literature classes, stated in Chapter 4 as a plausibility reference and not as independent measurements. This is the condition that Medina et al. (2021) describe as the general one, since detailed geotechnical maps of the soil properties are not normally available at regional scale and the information has to be adapted from geological or other maps. Laboratory or field tests would replace those hypotheses with measurements, and would also allow them to be differentiated within the formation, which a single litho-technical unit does not permit. The same authors note that values obtained in this way can be used directly as input parameters, without passing through the calibration.

Future work can rest on the same validation criteria, the same search space and the same selection rule. Any new configuration can then be compared with the six presented in this thesis.

A susceptibility map expresses the propensity of a territory to produce landslides from its geo-environmental conditions (Guzzetti et al., 2006), and it is the first level of zoning: it indicates where landslides are expected, not when, and does not estimate their frequency (Fell et al., 2008). The greatest limitation that emerges from these results lies inside that “where”. The model separates the predisposed slopes clearly from those that are not, but among the predisposed ones it distinguishes little, because that is where the spatial information available runs out. The map therefore remains reliable in delimiting the set of slopes on which failure is possible, and it is in this form that it enters land-use planning and the setting of priorities for intervention.

References

- Antolini P. et al., 2001. Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, Foglio 254 – Modigliana. Roma: ISPRA, Servizio Geologico d'Italia.
- Arpae Emilia-Romagna, 2023a. Bollettino mensile. Anno IV, n. 1, gennaio 2023. Bollettino mensile. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- 2023b. Bollettino mensile. Anno IV, n. 2, febbraio 2023. Bollettino mensile. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- 2023c. Bollettino mensile. Anno IV, n. 3, marzo 2023. Bollettino mensile. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- 2023d. Bollettino mensile. Anno IV, n. 4, aprile 2023. Bollettino mensile. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- 2023e. Rapporto degli eventi meteorologici di piena e di frana del 16–18 maggio 2023. Rapporto d'evento. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- 2023f. Rapporto degli eventi meteorologici di piena e di frana dell'1–4 maggio 2023. Rapporto d'evento. Bologna: Struttura Idro-Meteo-Clima, Arpae Emilia-Romagna.
- Arpae-SIMC, 2023. Dext3r — estrazione dati meteo della rete di rilevamento regionale RIRER. Agenzia regionale per la prevenzione, l'ambiente e l'energia dell'Emilia-Romagna, Servizio IdroMeteoClima (Arpae-SIMC). URL: <https://simc.arpae.it/dext3r/>.
- Asurza F.A., Abancó C., Hürlimann M., Medina V., 2026. Calibration scenarios for physically based rainfall-induced landslide modelling at regional scale. Application to Vall d'Aran (Central Pyrenees, Spain). *Bulletin of Engineering Geology and the Environment* 85, 27.
- Baum R.L., Savage W.Z., Godt J.W., 2008. TRIGRS — A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, Version 2.0. Open-File Report 2008-1159. U.S. Geological Survey.
- Benini A. et al., 2009a. Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, Foglio 238 – Castel San Pietro Terme. Roma: ISPRA, Servizio Geologico d'Italia.
- 2009b. Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, Foglio 239 – Faenza. Roma: ISPRA, Servizio Geologico d'Italia.
- 2014. Note Illustrative della Carta Geologica d'Italia alla scala 1:50.000, Foglio 253 – Marradi. Roma: ISPRA, Servizio Geologico d'Italia.
- Berti M., Martina M.L.V., Franceschini S., Pignone S., Simoni A., Pizziolo M., 2012. Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface* 117, F04006.
- Berti M., Pizziolo M., Scaroni M., Generali M., Critelli V., Mulas M., Tondo M., Lelli F., Fabbiani C., Ronchetti F., Ciccacese G., Dal Seno N., Ioriatti E., Rani R., Zuccherini A., Simonelli T., Corsini A., 2025. RER2023: the landslide inventory dataset of the May 2023 Emilia-Romagna meteorological event. *Earth System Science Data* 17, 1055–1074.
- Borga M., Dalla Fontana G., Da Ros D., Marchi L., 1998. Shallow landslide hazard assessment using a physically based model and digital elevation data. *Environmental Geology* 35 (2-3), 81–88.

- Brath A., Casagli N., Marani M., Mercogliano P., Motta R., 2023. Rapporto della Commissione tecnico-scientifica per l'analisi degli eventi meteorologici estremi del mese di maggio 2023. Technical report. Regione Emilia-Romagna. 147 pp.
- Caine N., 1980. The rainfall intensity–duration control of shallow landslides and debris flows. *Geografiska Annaler: Series A, Physical Geography* 62 (1-2), 23–27.
- Catani F., Segoni S., Falorni G., 2010. An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. *Water Resources Research* 46 (5), W05508.
- Cervi F., Ronchetti F., Martinelli G., Bogaard T.A., Corsini A., 2012. Origin and assessment of deep groundwater inflow in the Ca' Lita landslide using hydrochemistry and in situ monitoring. *Hydrology and Earth System Sciences* 16, 4205–4221.
- Cruden D.M., Varnes D.J., 1996. Landslide Types and Processes. In: *Landslides: Investigation and Mitigation*. Ed. by Turner A.K., Schuster R.L. Transportation Research Board Special Report 247. U.S. National Academy of Sciences, Washington D.C., 36–75.
- Fell R., Corominas J., Bonnard C., Cascini L., Leroi E., Savage W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology* 102 (3-4), 85–98.
- Guo Z., Torra O., Hürlimann M., Abancó C., Medina V., 2022. FSLAM: a QGIS plugin for fast regional susceptibility assessment of rainfall-induced landslides. *Environmental Modelling and Software* 150, 105354.
- Gupta K., Satyam N., Segoni S., 2024. A comparative study of empirical and machine learning approaches for soil thickness mapping in the Joshimath region (India). *Catena* 241, 108024.
- Guzzetti F., Carrara A., Cardinali M., 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), 181–216.
- 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.
- Guzzetti F., Reichenbach P., Ardizzone F., Cardinali M., Galli M., 2006. Estimating the quality of landslide susceptibility models. *Geomorphology* 81 (1-2), 166–184.
- Hungr O., Leroueil S., Picarelli L., 2014. The Varnes classification of landslide types, an update. *Landslides* 11 (2), 167–194.
- Hürlimann M., Guo Z., Puig-Polo C., Medina V., 2022. Impacts of future climate and land cover changes on landslide susceptibility: regional scale modelling in the Val d'Aran region (Pyrenees, Spain). *Landslides* 19, 99–118.
- ISPRA, 2023. Report ISPRA 17 maggio 2023 su evento alluvionale Emilia-Romagna 16-17 maggio 2023. Istituto Superiore per la Protezione e la Ricerca Ambientale. URL: https://www.isprambiente.gov.it/files2023/notizie/pdf24_merged.pdf.
- Iverson R.M., 2000. Landslide triggering by rain infiltration. *Water Resources Research* 36 (7), 1897–1910.
- Knappett J.A., Craig R.F., 2012. *Craig's Soil Mechanics*. 8th ed. London: Spon Press.
- Kuhn M., 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28 (5), 1–26.
- Lambe T.W., Whitman R.V., 1969. *Soil Mechanics*. Series in Soil Engineering. New York: John Wiley & Sons.

- Medina V., Hürlimann M., Guo Z., Lloret A., Vaunat J., 2021. Fast physically-based model for rainfall-induced landslide susceptibility assessment at regional scale. *Catena* 201, 105213.
- Montgomery D.R., Dietrich W.E., 1994. A physically based model for the topographic control on shallow landsliding. *Water Resources Research* 30 (4), 1153–1171.
- Pack R.T., Tarboton D.G., Goodwin C.N., 1998. The SINMAP Approach to Terrain Stability Mapping. In: *Proceedings of the 8th Congress of the International Association of Engineering Geology*. Vancouver, Canada.
- Pebesma E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30 (7), 683–691.
- Regione Emilia-Romagna, 2018. Uso del suolo 2014 — Coperture vettoriali di dettaglio. Scala 1:10.000, edizione 2018; licenza CC BY 4.0. Regione Emilia-Romagna, Servizio Statistica e Sistemi Informativi Geografici. URL: https://metasfera.regione.emilia-romagna.it/ricerca_metadato?uuid=r_emiro:2018-11-09T144433.
- Reichenbach P., Rossi M., Malamud B.D., Mihir M., Guzzetti F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews* 180, 60–91.
- Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J.-C., Müller M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77.
- Sidle R.C., Ochiai H., 2006. Landslides: Processes, Prediction, and Land Use. 18. *Water Resources Monograph*. Washington, D.C.: American Geophysical Union.
- Skempton A.W., DeLory F.A., 1957. Stability of natural slopes in London Clay. In: *Proceedings of the 4th International Conference on Soil Mechanics and Foundation Engineering*. 2. London, 378–381.
- USDA Soil Conservation Service, 1986. Urban Hydrology for Small Watersheds. Technical Release 55. Washington, D.C.: U.S. Department of Agriculture, Soil Conservation Service, Engineering Division.
- Wu Q., Brown A., 2025. whitebox: 'WhiteboxTools' R Frontend. R package version 2.4.3.