



Project “Mass Movements in Germany” and its implications for nationwide landslide susceptibility assessment

Jewgenij Torizin¹ · Michael Fuchs¹ · Dirk Balzer¹ · Dirk Kuhn¹ · Stefan Glaser² · Dominik Ehret³ · Johannes Wiedenmann³ · Peter Dommaschk⁴ · Stefan Henscheid⁵ · Roland Strauß⁵

Received: 19 July 2023 / Accepted: 11 April 2024
© The Author(s) 2024

Abstract

The paper presents the framework and discusses the outcomes of the “Mass Movements in Germany” project—a collaboration of five State Geological Surveys (SGSs) and the Federal Institute for Geosciences and Natural Resources (BGR) to conduct a feasibility study for a nationwide landslide susceptibility map in Germany using harmonized datasets. Therefore, distinct methods representing heuristic, physically-based, and data-driven approaches were employed. In 14 consecutive modeling cases, we investigated parameter selection and preparation, model building, evaluation, and transferability issues. The results show that nationally harmonized datasets, deemed high-quality, exhibited data generalization at regional scales, affecting the reliability of the landslide susceptibility patterns. While using regional datasets on larger scales mitigated some generalization effects, heterogeneities in inventory datasets among distinct federal states pose challenges in creating a nationwide applicable model. Heterogeneous data require locally adjusted model designs affecting the model's transferability and comparability of data-driven models across large regions. Heuristic methods operate without observational data but require in-depth knowledge of the regional-geological conditions by involving expert groups rather than single experts. The physically-based model shows promising results with parametrization based on the pedotransfer functions and soil database supplementing the national soil map. However, this approach is limited to shallow translational landslides. Thus, considering all mass movement types, a single comprehensive approach for creating a nationwide landslide susceptibility map for Germany is currently not feasible. Close collaboration and further harmonization of datasets and methods involving all SGSs are needed.

Keywords Nationwide · Landslide susceptibility · Data-driven · Heuristic · Physically-based · Harmonized datasets

✉ Jewgenij Torizin
jewgenij.torizin@bgr.de

Michael Fuchs
michael.fuchs@bgr.de

Dirk Balzer
dirk.balzer@bgr.de

Dirk Kuhn
dirk.kuhn@bgr.de

Stefan Glaser
stefan.glaser@lfu.bayern.de

Dominik Ehret
dominik.ehret@rpf.bwl.de

Johannes Wiedenmann
johannes.wiedenmann@rpf.bwl.de

Peter Dommaschk
peter.dommaschk@smul.sachsen.de

Stefan Henscheid
stefan.henscheid@gd.nrw.de

Roland Strauß
roland.strauss@gd.nrw.de

¹ Federal Institute for Geosciences and Natural Resources (BGR), 30655 Hannover, Germany

² Bayerisches Landesamt Für Umwelt, 86179 Augsburg, Germany

³ Landesamt Für Geologie, Rohstoffe Und Bergbau, 79104 Freiburg, Germany

⁴ Landesamt Für Umwelt, Landwirtschaft Und Geologie, 01311 Dresden, Germany

⁵ Geologischer Dienst Nordrhein-Westfalen, 47803 Krefeld, Germany

Introduction

Landslides are significant natural hazards frequently associated with earthquakes, torrential rainfall events, storms, and riverine floods, causing several billion USD in annual losses and claiming thousands of casualties worldwide (Froude and Petley 2018; Sim et al. 2022). Though the casualties are distributed unevenly, with about 95% in developing countries (Freeman et al. 2003; Lacasse and Nadim 2014), economic loss by landslide events is prevalent in all mountainous regions, affecting, in particular, transportation infrastructure (e.g., Jaiswal et al. 2010; Vranken et al. 2013, Martino et al. 2019). Klose et al. (2015) estimated Germany's direct and indirect losses at 300 million USD annually, with about 80 – 90 million USD solely on national highway infrastructure (Klose et al. 2016).

There is increasing scientific evidence that climate change influences the frequency and intensity of extreme weather events, posing significant challenges for communities worldwide (e.g., Coumou and Rahmstorf 2012; Clarke et al. 2021; IPCC 2023). The total losses by natural hazards in 2022 were about 270 billion USD. Most of those losses are attributed to meteorological and hydrological hazards (Munich Re 2023a), with a rising trend over the past ten years (Munich Re 2023b). Considering these worrying trends, policymakers would be well-advised to develop adaptation strategies to face the challenges ahead. This adaptation process implies, among others, employing reliable and effective hazard assessment tools at different scales. In the following, we will use the term scale in its cartographic sense, which denotes the ratio between a distance on a map and the corresponding actual distance on the ground. In the following, large-scale refers to maps with scales of 1:10,000 or greater, where the map depicts a smaller area with a high level of detail. Conversely, small-scale refers to maps with scales less than 1:500,000, which cover larger geographical areas but with less detail. Regional scale maps lie between the small and large scales, offering a balance of detail and area coverage suitable for representing states, provinces, or sizable metropolitan areas.

Regional scale hazard assessments are critical in providing policymakers with a broad overview, identifying areas of high hazard and risk potentials, and supporting action plans for more detailed investigations. These assessments aim to provide a comprehensive understanding of the hazards in a specific region, which is essential for developing informed and effective adaptation strategies.

Compared to other natural hazards, landslides are typically confined to specific local conditions (e.g., Fell et al. 2008). Therefore, accurate landslide hazard assessments at regional scales can be highly challenging,

as the conditions responsible for landslide occurrence are difficult to replicate or to transfer to other regions (e.g., Petschko et al. 2014). Additionally, the scarcity of observational multitemporal datasets further complicates assessing landslide hazards (e.g., Van Westen et al. 2006).

Because of these challenges, landslide hazard assessment is often substituted by landslide susceptibility assessment (LSA) (Brabb 1985). Unlike the landslide hazard, landslide susceptibility does not consider the temporal occurrence and magnitude of the events but depicts their spatial probability of occurrence given the presence of specific controlling factors (e.g., Guzzetti et al. 2005; Tian et al. 2017). Nevertheless, it is still valuable for identifying landslide-prone areas and land use planning and management decisions.

LSA includes heuristic (knowledge-based), physically-based, data-driven approaches (statistical and machine learning methods), and hybrid concepts combining those methods. While many options exist to assess landslide susceptibility, there is no global standard for approaching the issue at international, national, and even regional levels. Although extensive efforts were made in pan-European projects to create guidelines for harmonized mapping areas at landslide risk (e.g., Hervás et al. 2007; Malet et al. 2007; Reichenbach et al. 2007; Van Den Eeckhaut et al. 2012; Günther et al. 2014; Wilde et al. 2018), the employed concepts remain very different at the country level calling for a common legal framework for dealing with landslides (Herrera et al. 2018). Notable, among these past ventures at the European level, Germany was commonly underrepresented. The latter is due to Germany's decentralized federal structure, in which natural disaster prevention is the responsibility of every single federal state.

Consequently, every federal state pursues its own hazard and risk assessment procedures. A national legal framework or guideline defining standards for conducting regional LSA does not exist yet. The demand to assess landslide hazards and related risks varies with the prevalent geomorphological conditions. Thus, federal states with mountainous regions invest more in analyzing and mapping landslides than those with predominantly flat terrain. The different demands and the generally available coping capacities across federal states lead to principally heterogeneous methodical data collection and processing. Therefore, from the German perspective, generating a national landslide susceptibility map is as challenging as acting internationally, bringing states with specific interests and varying capacities to the common ground. Though German federal states act independently, they operate not isolated. State Geological Surveys (SGSs) have professional collaboration on various fundamental topics such as soil (Ad-hoc-AG Boden), geology (Ad-hoc-AG Geologie), and hydrogeology (Ad-hoc-AG Hydrogeologie) in so-called ad-hoc working groups which involve participants of all 16 SGDs and the Federal Institute for Geosciences and

Natural Resources (BGR). For decades, these joint working groups have compiled thematic national overview maps for soil, geology, and hydrogeology at a scale of 1: 200,000. However, a close project collaboration on comparably specific topics, such as nationwide landslide susceptibility assessment that could offer robust findings (e.g., a national map), has never been implemented due to the different importance of this hazard type in distinct federal states and thus, different prioritization and allocation of resources and capacities. The collaboration regarding landside assessment in Germany is currently limited to basic recommendations for landslide characterization issued in 2008 and updated in 2016 (Ad-hoc-AG Geologie 2016). However, it is essential to note that these recommendations are not obligatory. As a result, only a few SGDs have implemented the recommended standards yet.

The German academic community has addressed the landslide issue in the past. Dikau and Glade (2003) introduced a small-scale national overview map for landslide susceptibility as a part of the national atlas. Günther and Thiel (2009) proposed a preliminary heuristic landslide susceptibility map for Germany in the European Soil Thematic Strategy framework. Further research assessed the regional landslide susceptibility in several parts of Germany (e.g., Neuhäuser and Terhorst 2007; Terhorst and Kreja 2009; Damm et al. 2009; Kaynia et al. 2008). Klose et al. (2016) discussed the impacts of landslide hazards in Germany and their historical and socioeconomic perspectives.

Moreover, there are attempts to establish a national landslide database (e.g., Damm and Klose 2015; Kreuzer et al. 2017).

However, a common weak point in the mentioned academic studies is the limited involvement of the designated SGSs. This leads to a notable gap in legal recognition, particularly for maps produced at scales relevant to decision-makers. The debate often centers around the differing levels of responsibility assumed by academic researchers and the SGSs, sparking significant discussions. Academic researchers tend to prefer advanced, albeit complex and resource-demanding, methodologies. In contrast, SGSs lean towards more pragmatic, cost-efficient strategies that can be implemented more easily over extensive areas. Moreover, when academics publish landslide susceptibility maps, they typically do not bear legal responsibility for the implications of their use. SGDs, however, must thoroughly evaluate all disseminated information, acknowledging that decisions based on their data could have profound impacts.

Besides pure academic research, institutional collaborations between SGSs and the BGR led to the creation of regional landslide susceptibility maps currently used by some SGSs (e.g., Günther and Thiel 2009).

The project "Mass Movements in Germany" was initiated by four state geological surveys from Baden-Württemberg,

Bavaria, North Rhine-Westphalia, and Saxony, accompanied by the BGR. These state institutions agreed to conduct a comprehensive feasibility study for a nationwide LSA at regional scales based on five distinct study areas representing different data situations across the federal states. The feasibility study should consider the available thematic geoinformation nationwide, evaluate its applicability in different LSA methods by comparing it with regionally available data, and set up a frame for a possible national landslide susceptibility overview map supplemented by recommendations. This paper briefly presents the technical framework and discusses the project's outcomes and implications for a nationwide landslide susceptibility assessment in Germany.

Study areas

The feasibility study covers five study areas: the Swabian Alb and Foreland (SAF) in Baden-Württemberg (BW), Franconian Alb and Foreland (FAF), Simbach (SI) in Bavaria (BY), Elbe Valley Trench (EVT) in Saxony (SN), and Sieg Valley (SV) in North Rhine-Westphalia (NRW) (Fig. 1).

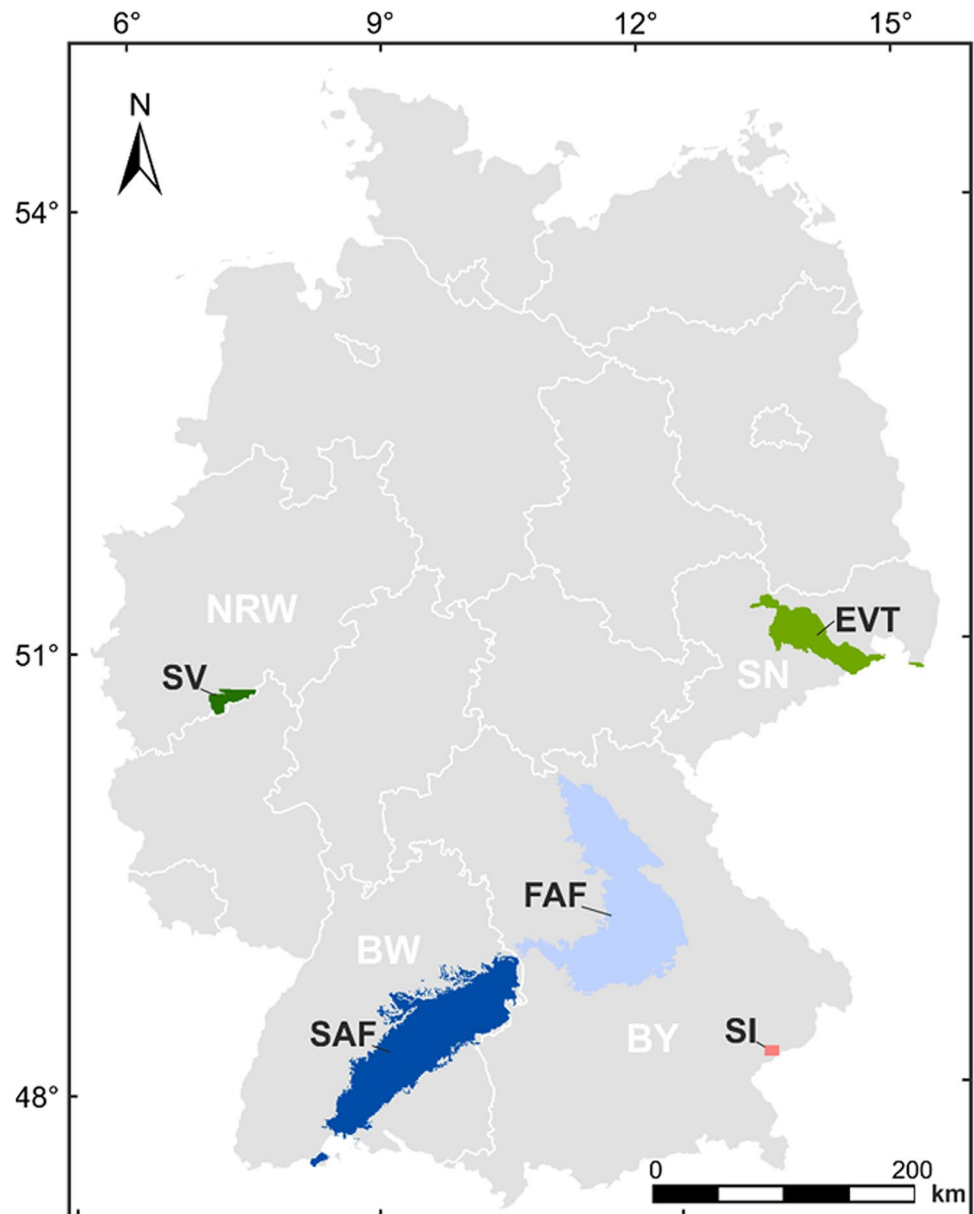
For the selection of the study areas, at least one of the following criteria was applicable:

- latent risk of mass movements;
- high priority for investigation in the respective federal state;
- sufficient information available for modeling;
- assumed to be suitable for comparing different modeling methods and testing their transferability between areas.

SAF, FAF, and EVT boundaries were defined based on the hydrogeological regions (Ad-hoc-AG Hydrogeologie 2016). The boundaries of SI and SV were delimited manually based on landslide events, e.g., triggered by extreme local rainfalls.

The SAF and FAF cover two mountain ranges in southern Germany, each with a similar size of about 8,000 km² and a cuesta landscape with elevations ranging from 240 to 689 m above sea level (a.s.l.) in FAF and 242 to 1015 m a.s.l. in SAF, respectively. The lithostratigraphic sequence in these ranges consists of clayey, calcareous, and, to a lesser extent, sandy sedimentary rocks from the Keuper to the youngest Jurassic period. The Jurassic deposits of the Swabian Alb dip to the SE towards the Danube under Tertiary molasse deposits. In contrast, on the eastern flank of the Franconian Alb, they are overlain by clayey-sandy Cretaceous deposits. The sedimentary rocks are frequently covered by Pleistocene loess and loess loam corresponding to Würm and, in some parts, Riss glaciation. The thickness of these deposits can reach several meters. Additionally, debris flow deposits and alluvial soils are

Fig. 1 Project study areas: SAF – Swabian Alb and Foreland in BW – Baden-Württemberg, FAF – Franconian Alb and Foreland, and SI – Simbach in BY – Bavaria, EVT—Elbe Valley Trench in SN – Saxony, SV – Sieg Valley in NRW – North Rhine-Westphalia



widespread in the area. Mass movements of various types occur at the step-like edge of the mountain ranges.

The EVT covers an area of about 2,000 km². The study area traces old NW–SE oriented tectonic lineaments and includes the Elbe River valley. The elevation in the study area varies between 53 and 790 m a.s.l. The valley's shape is narrow in the south, widening to a broad floodplain with thick Quaternary deposits in the north. The different sub-areas are delineated by structural geological features and characterized by a wide range of igneous, metamorphic, and sedimentary rocks. In the north of the Elbe Valley, magmatic rocks geomorphologically occur as domes. The southern part of the area reveals a rugged erosional landscape with canyon-like valleys within several hundred meters thick

Cretaceous sandstones. The steep flanks of the valleys are mainly affected by rockfalls.

The SV study area is located in the Rhenish Massif east of Bonn and covers the middle course of the Sieg River with an area of about 177 km². It is characterized by hilly to mountainous terrain with elevations ranging from about 70 to 400 m a.s.l. The bedrock in this area is composed of an interbedding of mudstone, siltstone, and sandstone layers of the Lower Devonian age. Periglacial talus layers, loess-influenced weathered soils, and recent younger weathering soil formations of varying thicknesses overlie the slopes. Clayey-silty and sandy-loamy soils are widespread in the study area. The most frequent type of mass movements that occur are translational slides.

The SI study area is located in the Bavarian Molasse Basin, covering approximately 90 km². The hilly terrain exhibits elevation heights between 300 and 550 m a.s.l. Miocene units comprising deposits of marine, brackish-fluvial, and limnic origin dominate the lithological sequence. Unconsolidated or weakly consolidated, except for a conglomerate rock called Nagelfluh, Quaternary deposits build the top of the lithological sequence. In June 2016, this area was affected by a torrential rainfall event (up to 180 mm/h), causing a riverine flood with enormous tangible damages and five fatalities (Rimböck et al. 2018). The rainfall event triggered more than 120 shallow landslides.

Methods

In the past, many methods have been introduced for regional LSA. These methods can be roughly categorized as physically-based, heuristic (knowledge-driven), and data-driven approaches (e.g., Aleotti and Chowdhury 1999; Guzzetti et al. 1999).

Physically-based methods directly analyze physical processes using force equilibrium (e.g., Pack et al. 2005; Jibson et al. 2000; Mergili et al. 2014) or more rigorous approaches. However, these methods require comprehensive knowledge of slope geometry, geotechnical material properties, and local hydrological conditions. The resulting model generates results in safety factors for a particular slope or a failure probability, indicating its vulnerability to failure.

In heuristic or knowledge-driven concepts, one expert or a group of experts assesses the importance of specific factors assumed to control the occurrence of mass movements (e.g., Stevenson 1977; Schleier et al. 2014; Kirschbaum et al. 2016; Stanley and Kirschbaum 2017). A weighted overlay of assessed factors generates a landslide susceptibility map. Therefore, heuristic analyses do not need observational data to establish the model and are suitable in areas with poor landslide inventories.

In data-driven approaches, the importance of the factors is evaluated based on observational data using statistical models either established by experts or directly learned from data by machine learning (ML) algorithms. The data-driven LSA is a binary classification, which allows for the application of many different statistical and ML methods and their derivatives (e.g., Reichenbach et al. 2018; Torizin et al. 2022). The data-driven methods can be subdivided into bivariate and multivariate methods, differing in the way they integrate independent variables (also contributing factors) within the model. Bivariate methods analyze the relationship between the dependent variable (landslide) and one contributing factor at a time, making them well-suited for understanding direct, pairwise correlations (e.g., Bonham-Carter 1994; Dahal et al. 2008). In contrast,

multivariate models simultaneously incorporate multiple independent variables to predict or explain a dependent variable, offering a more complex and comprehensive analysis that captures the interaction of various factors (e.g., Menard 1995; Schicker and Moon 2012; Lombardo and Mai 2018). The complexity of multivariate methods varies from linear models, such as logistic regression, to non-linear models, such as artificial neural networks.

In this study, we picked representative methods of each category that have gained significant attention in academic research in recent decades and have been applied in numerous case studies worldwide. The following briefly introduces the methods used.

Analytical hierarchy process

The Analytical Hierarchy Process (AHP) is a multi-criteria decision-making method introduced by Saaty (1977). It provides a systematic pairwise comparison of criteria to structure and evaluate hierarchical decision problems, resulting in numerical estimates representing each factor's weight in decision-making (e.g., Saaty 2008). The method determines the relative importance of these factors, and it performs a weighted overlay analysis to generate a landslide susceptibility map. Different studies demonstrated AHP's effectiveness in integrating multiple factors using expert knowledge to produce reliable LSA results (e.g., Hung et al. 2016; Persson et al. 2014).

Infinite slope model

The Infinite Slope Model (ISM) is a one-dimensional limit-equilibrium model that evaluates the force balance acting on a soil slab resting on an inclined plane (slope). The model makes several assumptions regarding the properties of the slab and acting forces. So, the slab has an infinite extent (no lateral forces at the end of the slab), and the water table is parallel to the slope. These assumptions make the model best applicable to shallow translational landslides (e.g., Mankelov and Murphy 1998). Due to its simplicity, it is suitable for implementation in grid-based GIS analyses, computing the force balance for each grid independently. Although it appears simplistic, researchers have successfully used the model in various approaches by coupling it with infiltration models such as steady-state recharge or transient infiltration for unsaturated soils (e.g., Pack et al. 2005; Baum et al. 2010). The model has also been applied to derive critical seismic loads (Jibson et al. 2000) or critical rainfall thresholds (Claessens et al. 2005). Stochastic Monte Carlo simulation typically extends ISM to account for soil properties' spatial variability at regional scales (e.g., Hammond 1992; Luzy et al. 2000; Fuchs et al. 2014).

Weight of evidence

Weights of Evidence (WoE) is a bivariate statistical method that analyzes the relationships between the independent variables (factors) and dependent observational data. It operates by analyzing each independent factor separately using the Bayesian formulation of conditional probabilities, which yields likelihood ratios for landslide occurrence based on the presence or absence of the factor (e.g., Bonham-Carter et al. 1989; Bonham-Carter 1994; Dahal et al. 2008). WoE assumes conditional independence of factors. This assumption simplifies the computation of the likelihoods, resulting in high transparency and generally well-interpretable results, but also makes the analysis sensitive to multicollinearities (e.g., Agterberg and Cheng 2002; Torizin 2016). Finally, the individually weighted factors are combined in a linear additive model to generate a landslide susceptibility map.

At the BGR, the WoE method has been successfully used for many years in various projects worldwide (e.g., Teerarungsigul et al. 2015; Torizin et al. 2017; Torizin et al. 2018) and is, therefore, a reference methodology for other data-driven methods involved into the study.

Logistic regression

Logistic Regression (LR) is a multivariate statistical method commonly used to model the relationship between a dependent dichotomous target variable and related independent factors (Menard 1995). The method involves fitting a logistic function to the data, which describes the relationship between the probability of landslide occurrence and the values of the independent factors (e.g., Ayalew and Yamagishi 2005; Can et al. 2005). A trained model is applied to predict the probability of landslide occurrence for a given set of conditions and to generate a susceptibility map based on the predicted probabilities. LR belongs to popular LSA methods implemented worldwide in the past three decades (e.g., Bernknopf et al. 1988; Dai and Lee 2003; Lombardo and Mai 2018).

Artificial neural network

Artificial Neural Network (ANN) is an ML algorithm inspired by the structure and function of the human brain. It uses layers of interconnected nodes (neurons) to perform input data computations and produce output. ANNs train to recognize patterns, make predictions, and perform other tasks by adjusting the weights of the connections between neurons through a process called backpropagation (e.g., Rumelhart et al. 1986). The most frequent type of ANN applied for LSA is the feed-forward ANN with one hidden layer (e.g., Ermini et al. 2005; Lee and Evangelista 2006; Pradhan and Lee 2010). Like LR, the trained ANN model

predicts the probability of landslide occurrence given a set of conditions.

Data

Parameters

In Germany, harmonized data exist as an overview map series for geology, hydrology, and soil at 1:200,000 and 1:250,000 scales, published in the joint institutional work of the SGS working groups and the BGR. Geology and soil overview maps are based on more detailed maps at scales of 1:25,000 and 1:50,000. Regional overview maps incorporate consistent symbol keys across Germany for geology, hydrogeology, and soil, serving as a foundation for aggregation and generalization to small-scale maps such as the Geological Overview Map (GÜK1000) at a scale of 1:1000,000. Comprehensive factual data accompany these map series. Beyond the essential attributes needed to characterize legend units, such as lithology and stratigraphy in the GÜK200, there is a factual database in the BÜK200, which includes standardized soil profiles and their proportional representation in each map legend unit for the upper two meters.

In addition to the geoscientific information layers, the Federal Agency for Cartography and Geodesy (BKG) provides digital elevation models (DEM) with ground resolutions of 25 m and 10 m, the Digital Landscape Model at scale 1:250,000 (DLM250) and Corine Land Cover (CLC). The DLM250 describes topographic objects of the landscape and the earth's surface relief in vector format. The dataset includes various object types and their key attributes, such as roads, paths, railways, water bodies, and settlements. The CLC program in Germany is part of a European initiative to provide consistent and comparable information on European land cover. The CLC data for Germany includes detailed mapping of various land cover types, such as agricultural areas, forests, urban regions, and water bodies.

The German Meteorological Service (DWD) offers regionalized precipitation data known as REGNIE, encompassing various temporal aggregates like daily, monthly, and yearly averages across three decades. Among these, REGNIE8110 specifically denotes the average annual precipitation calculated from 1981 to 2010, reflecting long-term precipitation trends over these 30 years. Table 1 gives an overview of the datasets used in the study.

Inventories

Landslide inventories can provide valuable observational data for understanding the frequency, magnitude, and distribution of landslide events in a specific region (e.g., Hervas and Bobrowsky 2009). Therefore, ideally, they

Table 1 National and regional datasets applied in the project

Information layer	Spatial extent	Scale/Resolution	Data Type	Link
Soil overview map of Germany (BÜK200) with soil database	National	1: 200,000	Polygon	https://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-laufend/BUEK200/BUEK200.html
Geological overview map of Germany (GÜK250)	National	1: 250,000	Polygon, Linie	https://www.bgr.bund.de/DE/Themen/Sammlungen-Grundlagen/GG_geol_Info/Karten/Deutschland/GUEK250/guek250_node.html
Geological overview map of Germany (GK1000)	National	1: 1000,000	Polygon	https://www.bgr.bund.de/DE/Themen/Sammlungen-Grundlagen/GG_geol_Info/Karten/Deutschland/GK1000/gk1000_inhalt.html
Soil overview map 1000 (BÜK1000)	National	1: 1000,000	Polygon	https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/buek1000_node.html
Corine Landcover (CLC10)	National	10 ha	Polygon	http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=22&gdz_user_id=0
Digital Elevation Model 25 (DEM25)	National	25 m × 25 m	Raster	https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/digitale-gelandedmodelle.html
Digital Elevation Model 10 (DEM10)	National	10 m × 10 m	Raster	https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/digitale-gelandedmodelle.html
Digital Landscape Model (DLM250)	National	1: 250,000	Polygon, Linie	https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/digitale-landschaftsmodelle.html
Precipitation (REGNIE)	National	1 km × 1 km	Raster	https://opendata.wd.de/climate_environment/CDC/grids_germany/daily/regnie/regnie_info.html
Soil map of Saxony (BK50)	Regional	1: 50,000	Polygon	https://luis.sachsen.de/boden/bk50.html
Geological map of Lausitz-Jizera-Karkonosze GK100 (L-J-K)	Regional	1: 100,000	Polygon	https://gdk.gdi-de.org/geonetwork/srv/api/records/02396a22-e6c2-477d-89af-77603a8f97ba
Geological map of Bavaria (dGK25)	Regional	1: 25,000	Polygon	https://www.lfu.bayern.de/umweltdaten/geodatendienste/index_wms.htm#Geologie

Table 1 (continued)

Information layer	Spatial extent	Scale/Resolution	Data Type	Link
Overview soil map of Bavaria (UeBK25)	Regional	1: 25,000	Polygon	https://www.lfu.bayern.de/umweltdaten/geodatendienst/pretty_download/oaddienst.htm?dld=uebk25
Soil map of North Rhine-Westfalia (BK50)	Regional	1: 50,000	Polygon	https://www.opengeodata.nrw.de/produkte/geologie/boden/BK/ISBK50/

Table 2 Used landslide inventories of the study areas (provided by the respective participating federal states)

Study area	Event count	Geometry	Event type
SAF	12,338	Polygon (undifferentiated)	Slide, Rockfall
FAF	2,202	Point (center of scarp) Polyline (center of scarp), Polygon (deposit area)	Rotational slide, Translational slide, Rockfall
EVT	463	Point	Slide, Rockfall, Flow
SI	121	Point	Translational slides
SV	8	Point	-

comprise detailed descriptions of the landslide events, such as the process type, time of occurrence, and geometrical information. Well-maintained landslide inventories are essential for developing and validating predictive data-driven models (e.g., Lima et al. 2021).

Respective SGSs provided landslide inventories for the project study areas. These inventories indicate significant differences in quality and acquisition strategies, which may pose a substantial source of uncertainty in the subsequent analyses (e.g., Lima et al. 2021; Loche et al. 2022). For SAF, the landslide inventory was mapped from the hillshade of DEM with a one-meter ground resolution providing polygon shapes. One single polygon consolidates the entire landslide from the depletion area to the deposit toe. The distinguished types are slides (undifferentiated) and rockfalls. In FAF, landslides were detected by field mapping and interpretation of airborne imagery. The FAF inventory distinguishes the mass movement types as translational slides, rotational slides, and rockfalls. Also, it considers different geometrical representations of scarp and deposition areas (see Table 3). For EVT, the mass movement inventory is based on field surveys. Points mark the depletion area's top (scarp's center) with a field mapping-related uncertainty. The few mass movements for SV are historical records of uncertain origin (no detailed description available), which generally reduce the reliability of the inventory for model generation and validation, making it potentially unusable. Field mapping and the interpretation of high-resolution airborne imagery also build the source for the landslide inventory in the SI study area. Table 2 provides an overview of the inventory features in the project's study areas.

Analysis

The comprehensive feasibility study started in the FAF area, which presents a moderately sized landslide inventory. This dataset was assessed as substantial

Table 3 Overview of the case studies with corresponding modeling cases, applied methodology, and formulated research issues

No. Case study	Study area	Modeling case	Methods	Research issues
1	FAF	1.1	WoE	Which of the available nationwide information layers is suitable for data-driven LSA?
		1.2	WoE	Can we reduce the generalization effects encountered in 1.1 by locally or regionally available information layers at scale 1: 25,000? Is it only a scale issue?
		1.3	WoE	Can we reduce the bias introduced through generalization in thematic information layers available nationwide by adjusting the geometric location of mass movements?
		1.4	WoE	Can we improve the model reliability by substituting GÜK250 with the adjusted parameter distance to the lithostratigraphic boundary to cope with flawed associations due to generalization effects?
		1.5	WoE	Can we reduce the significance of the parameter slope by masking the flat area, and what effect does it have on other parameters?
	SAF	1.6	WoE	Can we transfer a model to another region (e.g., a neighboring region) with comparable geomorphological and geological conditions?
2	EVT	2.1	WoE	Can we replicate the findings in case study 1 in another region?
		2.2	AHP	How comparable are the heuristic AHP results with a data-driven approach (WoE from 2.1)?
		2.3	ISM	Can we use a regional soil map (1: 50,000) to parametrize the ISM using pedotransfer functions?
		2.4	ANN	Can we utilize ANN in a region that has comparably poor data? What is the performance compared to the bivariate statistical method?
		2.5	ANN	A comparison study to 2.4 was performed by Schumann (2020)
3	SV	3.1	ISM	Repetition of the ISM based on the national soil map
4	SI	4.1	ISM	How comparable are the results when parametrizing ISM utilizing a regional soil map (UeBK25) and national soil map BÜK200?
		4.2	WoE LR ANN ISM	How comparable are different LSA models?

enough to develop and test data-driven models, offering insights applicable to regions with both larger and smaller landslide inventories. We employed the WoE method to create a foundational case study (case study 1 in Table 3), wherein we explored the applicability of various parameters within data-driven LSA. The low complexity and high transparency of the WoE method facilitated engaging discussions with local experts, helping to pinpoint additional research questions.

The general analysis workflow followed a cyclical pattern, beginning with data preparation, then model development, and culminating in an evaluation using statistical metrics. This cycle concluded with expert discussions, where model predictions were contrasted against expert expectations. We refined the model whenever there was a discrepancy, testing different model designs and alternative parameters. The iterative workflow led to various sequential modeling cases designed to probe specific issues through targeted research questions. Notably, the primary objective of these small modeling cases was not merely to produce the most precise landslide susceptibility maps. Instead, the focus was on understanding how changes in the model affected the outcomes, gaining insights into the underlying processes and their implications for a potential comprehensive nationwide LSA.

The significant findings of the first case study were replicated in the EVT study area (case study 2, Table 3) to manifest their general validity (modeling case 2.1). Additionally, in the EVT area, which has an insufficient inventory for specific mass movement types, the heuristic AHP method and ISM were tested. Both methods do not require observational data to build the model (modeling cases 2.2. and 2.3). Further, ANN was applied in EVT to investigate the applicability of non-linear data-driven approaches in regions with poor observational data. Because there are no established standards for applying ANN in LSA, we invited a research group from the Technical University Berlin to conduct a comparability study (Schumann 2020).

In the third case study conducted in the SV region (Table 3), we tested the ISM with parameterization based on the BÜK200. However, the SV area presented limited opportunities to comprehensively evaluate the model's performance. Consequently, we extended our discussion of the ISM to a fourth case study in the SI region. This area was chosen due to its significantly higher record of events, which could be attributed to a single triggering event, enabling a more thorough and convenient model evaluation. In the SI area, ISM, we compared parametrized ISM models based on a regional soil map at a scale of 1:25,000 and BÜK200.

Given the data consistency in the SI study area, we compared data-driven methodologies with the ISM. Acknowledging the potential limitation of a sparse inventory for implementing complex ANN, we also conducted a linear multivariate LR analysis. This additional step was taken to gauge the broader applicability of multivariate data-driven techniques in areas characterized by limited data availability.

Table 3 summarizes the modeling cases in the five study areas and highlights the formulated research issue.

The subsequent sections detail the specific analytical steps involved in data-driven, physically-based, and heuristic methodologies and model evaluation and zonation techniques.

Data-driven analysis

We used the LSAT PM Software (Torizin et al. 2022) and other routines written in R and Python to conduct the data-driven analyses.

At the beginning of the analyses, we randomly split the landslide inventory into training and test datasets for all data-driven methods in a ratio of 80:20. This division followed a common practice in machine learning (ML), usually adopting a ratio between 70:30 and 80:20 (e.g., Joseph and Vakayl 2022; Thien and Yeo 2022). We chose an 80:20 split, often associated with the Pareto principle, to balance the training dataset's representativeness with the test dataset's adequacy. Notably, differently sized samples may inherently introduce a bias that should be considered, e.g., via k-fold cross-validation or Monte-Carlo cross-validation when evaluating the model with smaller test datasets (e.g., Torizin et al. 2021; Torizin et al. 2022).

We implemented different data preparation procedures for data-driven analyses depending on the method. For the bivariate WoE, we reclassified continuous variables such as slope, distance to roads, or distance to tectonic features into discrete variables. LSAT PM supports the conversion procedures by providing different options for building classes.

For LR and ANN, the parameters were normalized following the standard procedures in ML. We scaled continuous data using the min–max scaler to the value range of 0 to 1. This scaling ensures that no feature dominates the model just because of the scale of its values. Categorical datasets were one-hot-encoded. The one-hot encoding transfers multiclass categorical variables in a set of binary variables (also known as dummy or indicator variables in statistical analyses), characterizing the presence or absence of a single categorical feature (e.g., Bishop 2006; Pedregosa et al. 2011). This encoding is necessary because linear models such as LR assume a linear relationship between independent variables and the dependent variable. When used directly, factor variables imply an ordinal relationship, suggesting a meaningful order or ranking. However, this assumption is

not applicable for categories on a nominal scale, such as land use or lithology.

While WoE involves all data in the analysis (unless masking is applied to eliminate some effects or data parts, e.g., flat areas in modeling case 1.5), we tested sampling procedures for the employed ML algorithms (LR and ANNs). For these methods, random sampling of non-landslide areas was considered. The sampling is usually done due to the inherent imbalance of the target variable class (few ones and many zeros) because landslide events are comparably rare compared to the overall areas under consideration. In ML, the class imbalance is frequently considered problematic (e.g., Japkowicz and Stephen 2002) and is the focus of current research (e.g., Krawczyk 2016). In the sampling process for our analysis, ratios of landslide to non-landslide pixels between 1:3 and 1:8 were considered by drawing non-landslide pixels randomly across the entire study area.

While WoE and LR outputs provide metrics that can be well-interpreted, such as log weights, coefficients, and p-values, ANNs are usually considered "black box" approaches. Different techniques have been introduced in the past few decades to make the outcomes of ML models explainable, such as feature importance based on permutation (Breiman 2001; Fisher et al. 2019), Shapley value (Lundberg and Lee 2017), or the Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al. 2016). In this study, we use the permutation techniques implemented in the *iml* R-Package (Molnar et al. 2018) to assess the significance of each feature by observing the increase in the model's prediction error after shuffling the feature's values, which effectively breaks the association between the feature and the actual outcome. A feature is considered significant if shuffling its values results in an increased error in the model, indicating that the model relied on that feature for making predictions.

Physically-based analysis

The ISM model was performed based on our own Python application (unpublished) in three study areas: EVT, SV, and SI. To parametrize the ISM, we utilized soil maps and their respective databases at various scales. We derived the geotechnical material properties from the soil units and associated information using pedotransfer functions (PTF). PTFs comprise a set of rules and empirical relations to translate soil texture and structure into soil-specific properties such as cohesion, angle of internal friction, bulk density, and hydraulic conductivity (e.g., McBratney et al. 2002; Wadoux et al. 2021). We used the PTF defined by the working group for soil (Ad-hoc AG Boden 2000, 2005) and the working group of the German Soil Society (Renger et al. 2008) tailored for the soil classification utilized in Germany's soil maps.

Specifically, we estimated the mean bulk density based on the soil type using empirical relation after Renger et al. (2008):

$$\rho = \rho_{max} - 0.005C - 0.001Si$$

where ρ is the mean bulk density, C represents the percentage of clay content, and Si denotes the percentage of silt content in a soil type, as determined by a soil texture diagram by Ad-hoc-AG Boden (2005). Parameter cohesion was derived using the tables from the so-called linkage rule 1.8 (Ad-hoc-AG Boden 2000) and internal friction angle utilizing the linkage rule 1.10 (Ad-hoc-AG Boden 2000). The linkage rules take the soil type and the soil structure as input. While soil type can be derived from the soil map, the values are averaged over all types of soil structure to account for given uncertainty.

Each soil map unit can be composed of several soil types. The portions of the soil types in the map unit are given in the corresponding database. To estimate the mean and standard deviation for the map unit based on the weighted distributions of the included soil types, we use the equation:

$$\mu = \frac{\sum_{i=1}^N m_i n_i}{N},$$

$$\sigma = \frac{\sum n_i (m_i - \mu)^2}{N}$$

where μ is the mean of the dataset, n_i is the frequency of the i^{th} group, m_i is the midpoint of the i^{th} group, σ the standard deviation, and N is the sample size.

$$P(FS \leq 1) = \frac{1}{N} \sum_{i=1}^N H_i \left\{ 1 - \left[\frac{f(c)}{f(\rho_s) g z \sin \vartheta \cos \vartheta} + \frac{\tan f(\varphi)}{\tan \vartheta} - w \frac{\rho_w \tan f(\varphi)}{f(\rho_s) \tan \vartheta} \right] \right\},$$

where $P(FS \leq 1)$ is the failure probability, N is the number of simulation iterations, $f(x)$ denotes the probability density function for parameters cohesion c [kPa], soil density ρ_s , g is the gravitational acceleration with 9.81 ms^{-2} , and internal friction φ [°], ϑ is the slope angle [°], and w is the wetness index taking values between 0 for dry and 1 for fully saturated conditions. H is the Heavyside step function defined as:

$$H : \mathbb{R} \rightarrow \{0, 1\}; H_i(x) = \begin{cases} 1 & : x \geq 0 \\ 0 & : x < 0 \end{cases},$$

Returning zero for values smaller than zero (this is the case when the FS is greater than 1) and values of one for all values greater or equal to zero. The summation of positive outcomes divided by the total number of iterations effectively estimates the failure probability.

MC simulations for extensive areas can be computationally demanding. To alleviate this, we integrate the concept of unique condition units with ergodic assumptions, enhancing computational efficiency. We treat continuous parameters such as slope as categorical by assigning fine-grained integer classes of one degree, reducing them to manageable integer rasters. These rasters are combined to identify unique condition classes defined by slope value and corresponding soil map unit. Within these classes, we assume identical behavior across the distribution of parameters, regardless of their spatial location. This approach enables the conducting of MC simulations for unique conditions and the application of the results across all corresponding spatial locations. By focusing on limited unique combinations, we can significantly compress large raster datasets for calculations, freeing up computational resources for additional tasks such as more simulation repetitions. The estimated mean and standard deviation of parameters, which provided a range of values assumed by Gaussian distribution, were used in ISM extended by MC simulation. We randomly drew the parameters—soil density, cohesion, and angle of internal friction—from the estimated distributions. We employed different uniformly distributed wetness scenarios to compute the ISM without integrating a realistic precipitation scenario and a subsequent hydrological infiltration model.

As a result, we obtained a probability of factor of safety (FS) less or equal to one given the conditions associated with the respective soil unit of the soil map and wetness scenario. The computation equation for the given task can be formulated as follows:

Heuristic analysis

The AHP was conducted using the application designed by Goepel (2013), employing a simple query to rank the importance of factors such as slope and lithology regarding their influence on rockfall events in the EVT test area by geoscientists involved.

We utilized the GK1000 to derive lithological classes. GK1000 exhibits nine distinct lithological classes in the study area. In contrast, the GÜK250 dataset contains approximately 85 lithostratigraphic units in the same area, which would require aggregation or additional hierarchical grouping to enable ranking using the AHP. Therefore, GK1000 was selected to enable a more practical choice for this assessment.

Seven geoscientists with different experience levels and knowledge of the regional conditions participated in this assignment. After joining the different expert rankings, we

estimated the consensus of the rankings. Finally, the averaged priorities were applied to the mapping units and overlaid to a susceptibility map.

Model evaluation and zoning

We evaluated all models using the receiver operating characteristics (ROC) curve. The ROC curve is a tool to estimate the goodness of the binary classifier (e.g., Fawcett 2006) depicting the relation of the True Positive Rate against the False Positive Rate of the model. A quantitative measure to compare ROCs is the area under the ROC curve (AUROC). To assess uncertainties in data-driven models, we introduced Monte Carlo cross-validation (MCCV) (e.g., Picard and Cook 1984, Torizin et al. 2021). The traditional cross-validation divides the data into a fixed number of equally sized folds, and the model is trained and tested on each fold in turn. MCCV randomly draws training sets for each iteration, allowing a more comprehensive evaluation of the model's performance, as it helps to avoid biases arising from using fixed data partitioning. MCCV is particularly useful when working with small datasets or when the data distribution is highly variable. Averaging the performance over 100 random splits provides a robust estimate of the model's performance.

Finally, we used validation ROC curves to divide the obtained landslide susceptibility into five zones as follows:

- Very high: about 50% of all known/predicted events
- High: about 30% of all known/predicted events
- Moderate: about 15% of all known/predicted events
- Low: about 3% of all known/predicted events
- Very low: less than 2% of known/predicted events.

Results and discussion

The initial case study in FAF served as preliminary work for the subsequent investigations. The analysis indicated the exceptional discriminatory power of the slope information layer obtained from DEM25. As a result, susceptibility models for various mass movement types have achieved good to excellent accuracy, as shown in Table 5. However, while slope gradient is sufficient to characterize rockfalls, rotational and translational landslides exhibit distributions extending to shallower slopes. Thus, additional parameters are needed to improve the spatial discrimination of susceptible areas for these mass movement types. Including geological or soil maps at a scale of 1: 250,000 and 1: 200,000 in the model enhances the identification of landslide-prone areas for rotational and translational landslides. These layers strongly correlate with the

Table 4 Parameters derived from the DEM25 and corresponding AUROC values for different mass movement types

Parameter	AUROC for Training and (Test)		
	Rotational Slide	Translational Slide	Rockfall
Slope	0.93 ± 0.01 (0.93)	0.89 ± 0.04 (0.87)	0.96 ± 0.02 (0.96)
Aspect	0.57 ± 0.02 (0.54)	0.60 ± 0.03 (0.61)	0.62 ± 0.03 (0.59)
Curvature	0.79 ± 0.01 (0.79)	0.74 ± 0.03 (0.77)	0.80 ± 0.02 (0.79)
GÜK250	0.87 ± 0.02 (0.87)	0.90 ± 0.04 (0.83)	0.80 ± 0.03 (0.80)
BÜK200	0.78 ± 0.01 (0.79)	0.85 ± 0.02 (0.79)	0.84 ± 0.02 (0.85)
BÜK1000	0.70 ± 0.01 (0.69)	0.73 ± 0.03 (0.67)	0.73 ± 0.02 (0.75)
CLC10	0.81 ± 0.01 (0.80)	0.80 ± 0.02 (0.82)	0.82 ± 0.02 (0.81)
REGNIE8110	0.70 ± 0.02 (0.71)	0.70 ± 0.03 (0.72)	0.60 ± 0.03 (0.53)

distribution of landslides and provide traceably interpretable classifications. Vector maps at minor scales (e.g., GK1000) or raster layers with coarser ground resolution (e.g., REGNIE8110) demonstrate generally lower performance and are more challenging to interpret accurately. Information layers at a scale of 1:250,000 also show effects that we attributed to the generalization of information.

Table 4 shows the parameters derived from the DEM25 and other thematic layers and corresponding AUROC values for different mass movement types. The AUROC values depict the training dataset with the corresponding variance based on MCCV. The value in brackets specifies the AUROC value for the test dataset.

The best combination of parameters is slope and lithology from GÜK250, providing AUROC on the test dataset of 0.96 for rotational slides, 0.94 for translational slides, and 0.97 for rockfalls. Figure 2 shows a small excerpt of the FAF study area with susceptibility patterns for different mass movement types based on the abovementioned parameters.

Findings from the first case study (see also Table 4) suggest that geoscientific information layers, initially not created for analyzing the spatial distribution of mass movement potentials, represent mapping units with varying degrees of abstraction depending on their original thematic objective. The geometric generalization, e.g., by simplifying or smoothing map unit borders, and conceptual generalization, e.g., aggregation of map units, in the transition between the large and small scales can lead to incorrect spatial attribution of conditions when superimposing generalized information layers with landslide locations. Also, modifiable areas of map units across the scales directly influence the estimates for

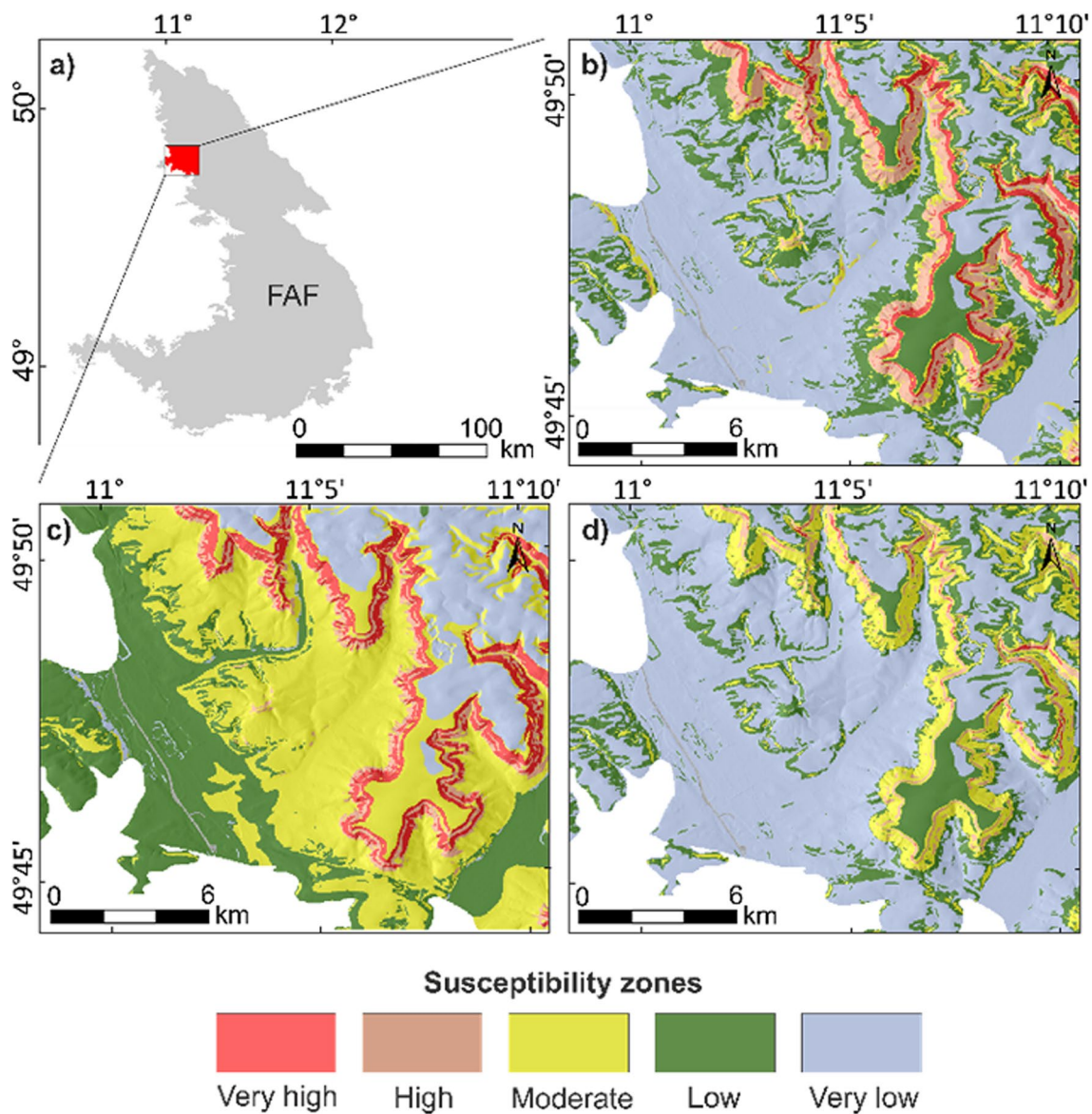


Fig. 2 Susceptibility pattern based on the parameters slope and lithology from GÜK250 for b) rotational slides, c) translational slides, and d) rockfalls

conditional probabilities. These errors become more evident as the degree of generalization increases, significantly limiting the reliability of data-driven analyses at small scales such as 1: 1000,000. Notably, statistical model metrics such as AUROC may not reveal those effects. Closer scrutiny of obtained weights and their interpretation is advisable to identify potentially inaccurate attributions.

Different strategies to cope with generalization effects at regional scales were tested in the following modeling cases (1.2 – 1.4). By integrating information on a larger scale in modeling case 1.2 (Table 4), we can suppose that observed generalization effects, such as errors in the spatial attribution between events and factor classes, are naturally reducible.

We deduce this from the weights ratios obtained for specific lithological units such as sandstones and limestones forming the sequence of the SAF area (see also Fig. 3). At larger scales, the weights for incompetent sand- and claystone increase, better reflecting the expert's expectations based on field experience. However, the significant positive weights for limestones remain at larger scales, pointing out that the generalization is also present on a scale of 1: 25,000, and another reason must exist, such as biases in the inventory.

Spatial attribution errors that may arise from mass movements' geometrical delineation (e.g., point or polygon) were scrutinized in the modeling case 1.3. As suggested by the results of the previous two modeling cases, landslides,

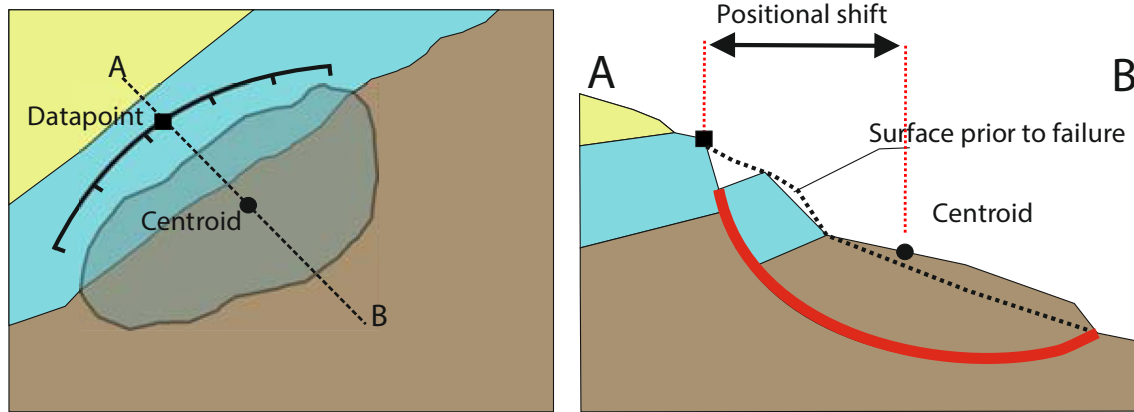


Fig. 3 Principal sketch of the regional conditions and the available landslide data

Table 5 Selected parameters with corresponding AUROC values after masking with a specific slope value

Type	Mask slope value [°]	Parameter	AUROC
Rotational slide	< 10	GÜK250	0.82 (0.81)
		Slope	0.69 (0.68)
		Curvature	0.70 (0.69)
Translational slide	< 5	GÜK250	0.88 (0.80)
		Slope	0.81 (0.80)
		Curvature	0.63 (0.64)
Rockfall	< 12	GÜK250	0.74 (0.73)
		Slope	0.86 (0.87)
		Curvature	0.62 (0.60)

depicted as a point at the scarp, do not necessarily capture the causal conditions of the failure supposed by local experts. The initial WoE model, considering these landslide positions, estimates that the limestones and dolomites that form the prominent step in the landscape are very susceptible since a significant portion of landslide points are located within it. However, local expertise shows that while the limestones are involved in the process, the failure is initiated by the incompetent rocks underlying it (Fig. 3).

The shifting of the landslide points to the centroid of the mass deposits (available as a polygon for rotational landslides) provided minor variations to the AUROC of the model. However, the procedure improved the interpretation of the factor weights, making them more acceptable to local experts since more significant weights better considered the underlying sandstones. Similar results were achieved in the modeling case 1.4, introducing alternative parameters, such as the distance to lithostratigraphic unit boundaries, without modifying the original landslide point position. However, in terms of statistical metrics, we also observed no significant changes here. Hence, locally adjusted designs improve model interpretability

but might affect the model's transferability to other regions since the specific design conditions might not be replicable.

Modeling case 1.5 addressed a specific issue frequently appearing in LSA at regional scales. The parameter slope has extremely discriminative characteristics, suppressing the possibility of evaluating the contribution of other parameters. Approaching the LSA from causal interpretations, we postulate that slope is a necessary cause. Therefore, if a critical slope gradient is not reached (depending on mass movement type), a failure will not occur, disregarding the presence of other contributing factors. Masking flat areas, i.e., excluding them from the ROC space, reduces the slope parameter's discrimination power, giving a chance to other contributing parameters. Table 5 shows selected parameters with corresponding AUROC values after masking with a specific slope value (see also Table 4). Within masked flat areas, the slope appears less critical in distinguishing rotational and translational landslides, and the GÜK250 parameter becomes a more effective discriminator. The mask could include steeper slopes to improve the performance of other parameters for rockfalls. However, it is essential to note that slope values may not always accurately represent the slope gradient due to the surface generalization effect of the DEM25. Increasing the threshold for the mask may lead to masking out steeper areas, which appears smoothed on DEM's ground resolution.

In the modeling case 1.6, we explored the feasibility of applying a susceptibility model developed for one region to another region with comparable geological and morphological characteristics. Specifically, we attempted to transfer models designed for rockfalls and rotational landslides from FAF to SAF. Both regions are part of the South German Scarplands and exhibit similar geomorphological features. Model transferability is contingent on the target region's characteristics that are recognizable to the model. Geomorphometric features such as slope gradients are generally transferable, provided the training region encompasses the

full range of slope values for model calibration. However, transferring distinct lithological conditions poses significant challenges. Despite FAF and SAF belonging to the same geomorphological structure, only 15% of SAF's area could be characterized using the petrological IDs from GÜK250 found in FAF. This limitation stems mainly from the varying levels of detail in describing specific lithostratigraphic units across the two regions despite identical scales of the base maps. So, after a closer look at base maps on a 1:25,000 scale in some regions, we observed that geological formations exhibit different aggregation levels receiving different unique IDs. To mitigate this issue, we aggregated the lithostratigraphic units using a more generalized symbolic key (LBEG 2015), enabling us to characterize approximately 90% of SAF's area with weights derived from FAF. However, this aggregation reduced the number of distinct lithostratigraphic map units from the original 140 used in FAF to 63. This additional aggregation may pose an issue as it considers stratigraphic properties rather than material properties, generating complex units composed of varying lithology, and the interpretation becomes challenging. The remaining 10% of the area in SAF represents a lithological sequence of the middle Keuper not exposed in FAF.

Evaluating the transferred models using the ROC curve provides an AUROC value of 0.98 for rockfalls and 0.90 for rotational slides. We attribute the slightly better results for rockfalls to the generally more rugged terrain in the SAF and the acquisition procedure of the landslide inventory in this area in which the landslides were mapped based on a DEM hillshade generally exhibiting better spatial alignment to the morphometric parameters derived from a DEM even if DEM used in the LSA is of lower resolution. Also, this shows that the single parameter slope is sufficient to represent this mass movement type at regional overview scales, specifically if the entire area (also flat regions) is considered. The lower AUROC for slides may have different reasons. First, this might stem from the undifferentiated inventory of slides in SAF, which comprises rotational and translational slides. Also, the geometry of the landslides involved in the analysis differs as it consists of polygons depicting all parts of the mass movement. Further, when closely analyzing the contribution of the slope parameter for translational and rotational landslides in FAF (Table 4), we see that AUROC of 0.90 is approximately the average when putting these types together. Based on this, we could suppose that through the additional generalization of lithostratigraphic units, the contribution of the lithology parameter was further diminished so that, finally, we majorly observe only the contribution of the slope.

We could replicate the key findings from the initial FAF case study for rockfalls in the EVT study area in the second case study. Other landslide types are underrepresented in

the inventory and do not provide sufficient data for a data-driven analysis.

As mentioned, the combination of the slope and GÜK250 parameters has proved to be the most effective for analyzing rockfall events, with the same trend that the parameter lithology becomes increasingly significant for EVT when flat areas are excluded from the analysis. Otherwise, the parameter slope dominates the susceptibility pattern. Application of other locally available parameters, such as the L-J-K geological map (Table 2), distance from roads (derived from DLM250), and tectonic features, were examined without improving the models.

While the generalization effects in FAF could be described qualitatively by interpreting the alteration in the weights of contributing parameter lithology, we could directly and explicitly quantify them in EVT based on the landslide inventory. The crucial point in EVT is that the landslide inventory collected through field surveys includes a detailed field description of the lithological layers in which the landslides occurred. This field description was compared with attributes obtained by superimposing the landslide locations onto geological maps. As a result, we could identify the wrong association of lithological attributes for approximately 31% of slides and about 10% of rockfalls. We corrected this bias by shifting the event counts to the correct lithological class. Comparably to the results of modeling cases 1.2–1.4 in FAF, we noted that while the statistical performance of the model did not significantly change, the correction of the associations altered the order of importance for distinct lithological units. The latter suggests that relying solely on statistical metrics is inadequate for evaluating the model's accuracy and reinforces findings and arguments of earlier studies, e.g., Steger et al. (2021), emphasizing that correlation does not imply causation. Furthermore, the results are congruent with the arguments presented by Lima et al. (2021), advocating that data biases should not be ignored. Instead, they suggest adapting the model design to address and rectify data inaccuracies effectively.

The AHP method in modeling case 2.2 revealed different challenging points. As described in the analysis, applying the method to features exhibiting many classes will require a complex hierarchy, and the analysis becomes generally cumbersome due to the need to make numerous pairwise comparisons. However, even breaking the complexity down to a few classes, we observed a high degree of subjectivity inherent in involving experts' judgment. While the priorities for the parameter slope indicated a consensus of about 98%, reflecting the general physical process understanding of all involved experts regarding the causal interaction between rockfall potential and the parameter slope, the assessment of the GK1000 parameter classes showed an apparent disagreement reflected in a consensus of only

55%. Therefore, we can conclude that labeling rock types without contextualizing them within a regional geological framework is inadequate for accurate and unambiguous weighting toward the stated objective. Different subjective associations (based on the expert's working experience in different regions) regarding discontinuity patterns or binding agents and the composition of specific rock categories influenced individual experts' decisions. A comparison with the data-driven assessment based on WoE performed with the same parameters (Fig. 4c) reveals that the expert-based assessment categorizes larger areas as moderate to highly susceptible (Fig. 4b). The proportions of the very high susceptibility class are comparable in both models. The applied overlay procedure in AHP partly explains the pattern difference. In AHP, slope and lithology factors were equally weighted for the overlay. In WoE, the weights for parameter slope are naturally significantly higher, gaining better contrast and suppressing lithology weights in flat areas. This effect can also be achieved in AHP if the parameters get additional weights (e.g., parameter slope has a much higher priority than lithology).

The Artificial Neural Network (ANN) was employed in the EVT (Fig. 4d) and SI. Schumann (2020) realized an alternative ANN approach for comparative purposes in EVT. The approach by Schumann (2020) shows significant differences in the data preparation process (feature engineering), sampling strategies, and the number of parameters used. However, the efficiency of the final susceptibility models was comparable. Finally, both studies concluded that in regions with limited data availability, the complexity involved in preparing data for ANN and the subsequent effort required to interpret these models, e.g., through additional analyses for understanding the model's outcomes (e.g., feature importance tests), did not justify the knowledge gained in specific case studies. The inherent flexibility of ANNs makes them susceptible to overfitting when the model is trained with many inputs relative to the size of the landslide inventory. Alternative, more straightforward methods, such as WoE, have yielded comparable results with less computational effort and higher levels of transparency and interpretability. Therefore, considering the qualitative and quantitative limitations of the available datasets, employing ANNs for comprehensive, nationwide modeling is not deemed the most effective strategy under current conditions. This stance is predicated on the notion that unless there is a substantial improvement in the data's volume and quality, the advantages of using ANN in landslide susceptibility studies remain marginal.

We applied ISM in three study areas: EVT, SV, and SI (Fig. 5), parametrizing the model based on soil maps of varying scales. The proposed workflow was well applicable in the three study areas, pointing to the good feasibility of

the approach given the available nationwide data. However, different challenging aspects also need to be highlighted.

In the parametrization process, we observed that clayish soils may exhibit various soil structures (Ad-hoc AG Boden 2000, 2005), generally not known for specific soil types derived from the soil map unit. This uncertainty significantly increases the variance of the cohesion and internal friction values. Because ISM is sensitive to variance in cohesion, estimates of values that are too high may lead to a general underestimation of the failure probability for map units dominated by clayish soil types located on gentle slopes.

Another issue that arises in areas with poor observational data is the limited possibility to evaluate the model's performance. While model generation is possible, we lack reliable datasets to estimate its accuracy in areas such as SV. Generally, we noted that fully saturated models agreed better with the recorded events, while models under dry conditions performed weakly. These results could imply that the events were likely triggered by rainfall. However, given all other uncertainties, such as the missing type of landslide and possibly uncertain location, they are too large to provide a confident model evaluation.

The situation is better in the SI area, where the landslide type is known, and all events are attributed to a specific trigger. In SI, the ISM model performs well with AUROC values of about 0.79 for the national soil map (BÜK200) and 0.81 for the regional soil map (UeBK25) (Fig. 5). The slightly better performance of the regional soil map is mainly due to the more detailed geometrical resolution of the larger map scale. Comparing the inputs, we count about 897 combinations of 1-degree slope classes and soil map units from BÜK200 and about 1364 unique combinations for UeBK25. As indicated above, significant uncertainties occur on clayish soils estimated as overconfidently stable due to their high cohesion in the ISM model. In these areas, characterized as stable, eight landslides were detected.

The results characterize failure probability and LSA zonation under uniformly distributed wetness conditions, excluding infiltration or precipitation models. While this approach might initially seem impractical and unrealistic, the primary objective was to separate the geotechnical model from the uncertainties of the infiltration model, focusing on the impact of the variability of geotechnical parameters derived from soil maps. The practical use case would involve creating a series of models for various uniform wetness conditions to facilitate the development of a stacked foundational model. The stacked model can subsequently be integrated with diverse wetness scenarios, thereby avoiding the need to recalculate failure probabilities. Realistic wetness scenarios can be applied as spatial queries

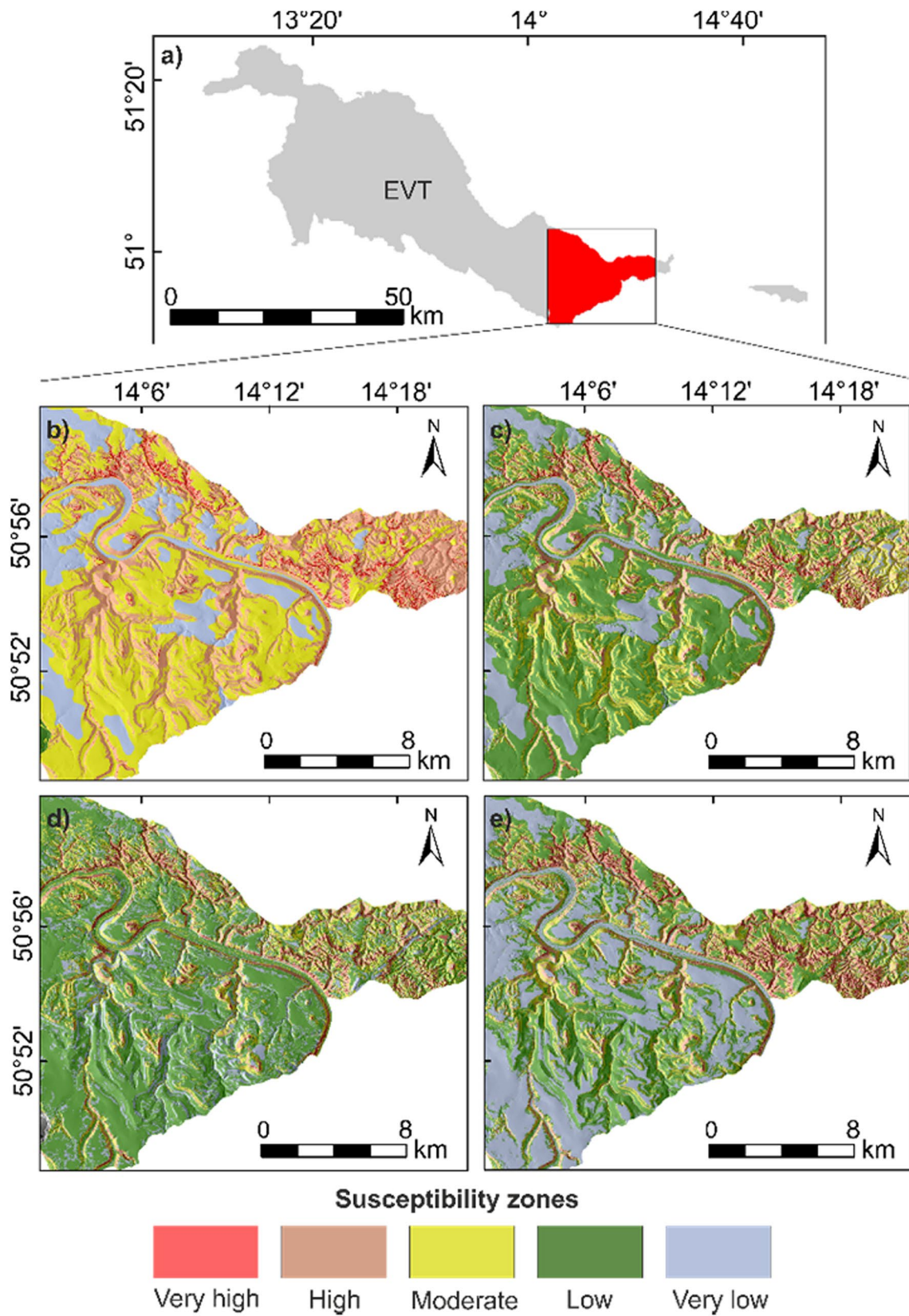


Fig. 4 a) Shape of the EVT study area; b) Zoning based on AHP utilizing parameters slope and GK1000; c) Zoning based on WoE for slope and GK1000; d) Zoning based on ANN with parameters slope and GÜK250; e) Zoning based on WoE with parameters slope and GÜK250

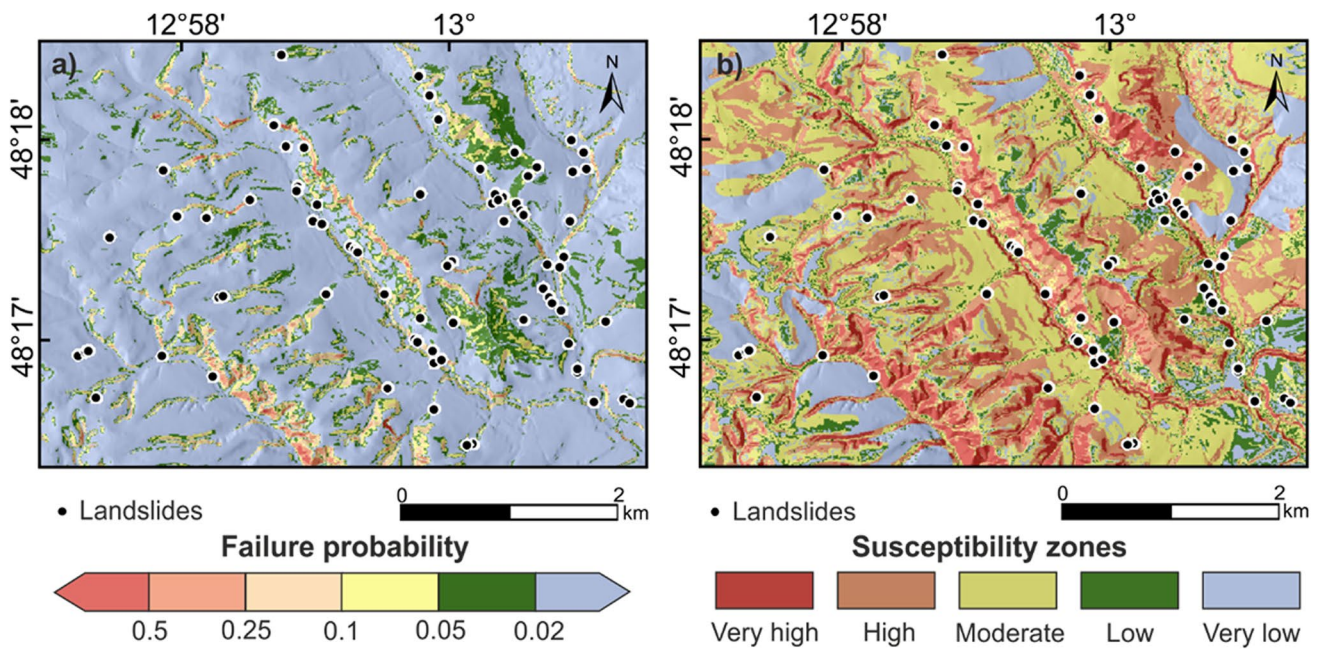


Fig. 5 Results of the ISM for Simbach area: a) Estimated failure probability for fully saturated conditions; b) Corresponding susceptibility zoning based on the ROC curve

to this model framework to ascertain the likelihood of failure in specific raster cells at particular wetness levels.

The comparison of ISM and data-driven models in the SI study area shows that the applied soil maps are suitable for characterizing shallow landslides. ISM achieves AUROC values of about 0.79 parametrized with BÜK200 and AUROC of 0.81 parametrized by UeBK25. Data-driven methods show a similar trend, emphasizing that the fine-grained UeBK25 performs better with AUROCs of 0.88 ± 0.02 for models including BÜK200 and 0.90 ± 0.02 for UeBK25 on training data. Generally, all data-driven models define slope as an essential feature with the most significant effect. The contribution order of the soil map units differs among the data-driven models. Figure 6 shows the importance of the features for ANN based on the permutation ranked according to their importance to the ANN model. Additionally, colored rectangles depict variables that show positive coefficients in LR (red) and WoE (green). Notably, the map unit MU40, representing the "clayey-loamy molasse and other Tertiary materials with a loamy top layer," has the highest coefficient in the linear LR among the soil map classes but is not significantly considered in bivariate WoE and the non-linear ANN.

While random sampling employed in multivariate methods (LR and ANN) did not markedly affect the training

AUROC of the models, it significantly influenced the models' generalization capabilities on the test dataset. For the test dataset, the performance significantly dropped to values around 0.80 for sampling low sampling ratios. The most robust results, with patterns comparable to WoE, were obtained using the imbalanced dataset, including all non-landslide areas (Fig. 7). Also, we observed a notable difference in the ranking of soil map classes considered significant through the sampling models, underscoring the heightened sensitivity of ML algorithms to both the sampling but also the overall volume of data. This latter becomes especially noticeable in LR models, where the p-values for the coefficients' estimates often exceed the standard significance threshold of 0.05, sometimes approaching one for specific soil map units. High p-values like these indicate a weak statistical significance for the corresponding variables, suggesting that the observed effects might be due to random chance rather than genuine predictive relationships. This implies that data availability limitations could influence the model's coefficients more than their relevance as predictors.>

For the zoning, the most significant differences occur for areas depicted as moderate, low, and very low (Fig. 7).

Overall, the conducted feasibility study investigated the opportunities of a harmonized approach for generating a national landslide susceptibility map in Germany. While we tried to select study areas that could reliably represent the data



Fig. 6 Feature importance from the ANN model, including parameter slope and map units (MU) from BÜK200 and no sampling, with highlighted features considered also significant in WoE and LR models; with MU33: loamy-sandy molasse material, partly with a loamy flow-soil cover, MU38: clayey-loamy molasse material with a loamy top layer, MU34: silty to loamy molasse material, mainly with a flow-soil cover, MU42: loamy and sandy-loamy valley deposits, MU35: silty material of the freshwater brackish molasse, locally with loess-

loam cover, MU31: gravelly molasse material, MU11: silty-loamy washout materials, MU30: Loess loam with molasse material, MU14: River marl or loamy valley deposits over carbonate-rich gravel, MU52: loess loam over loess loam flow soil, MU40: clayey-loamy molasse and other Tertiary materials with a loamy top layer, MU12: river marl over carbonate-rich gravel, MU5: sandy to loamy over gravelly floodplain deposits, MU3: gravelly, silty, and clayey floodplain deposits

situation among different federal states, the findings might still not be exhaustive. Nevertheless, our results reveal the most critical issues to address when approaching the task of nationally harmonized LSA and are in general agreement with the experiences of other recent studies in Europe targeting a national landslide susceptibility (e.g., Loche et al. 2022; Lima et al. 2021).

A notable finding is that the quantity of functional parameters diminishes as the scale becomes more regional and for smaller-scale overview maps, as illustrated in Fig. 8. This trend is predominantly influenced by the discriminatory capability of geomorphometric variables, most notably the slope parameter.

Conclusions

We can draw the following conclusions based on the feasibility study results. A data-driven national susceptibility map is feasible for rockfalls based on the single parameter slope. At regional overview scales, the discriminative power of slope parameters is overwhelming. The contribution of other parameters is statistically not reliably detectable when applying metrics such as AUROC. Because the slope is a generally well-transferable parameter, the model is trainable in selected pilot areas with good data coverage. Likewise, the definition of critical thresholds is feasible with heuristic methods.

The slope is also an essential but insufficient parameter for characterizing translational and rotational slide distribution. Other factors for which regional

model transferability is limited, e.g., due to local characteristics of factors (e.g., lithological conditions), are to be considered (Fig. 8). On the other hand, a holistic nationwide data-driven assessment of those factors is currently not feasible due to heterogeneous inventory datasets hosted by the respective SGSs that would ultimately lead to biased estimates of landslide susceptibility.

Employing heuristic methods for areas with insufficient observational data would require intensive and long-term cooperation among all SGSs, taking into account detailed expert knowledge of regional geology.

The physically-based modeling utilizing a nationwide soil map with PTF for model parametrization provides a promising approach for shallow translational landslides. In Germany, shallow landslides are predominantly related to torrential rainfall events. Therefore, the physically-based approach could also provide dynamic and scenario-based models considering the climatic change projections. Further tests in areas with good observational data and well-known spatiotemporal characteristics are needed to understand possible model limitations better. Further in-depth investigations on the model parametrization from the soil database, model extension by infiltration models, and elaboration of critical rainfall thresholds would be an asset.

The physically-based modeling of deep-seated rotational landslides demands rigorous models with detailed input data (e.g., 3-D underground models), which are generally unavailable in the required quality for larger areas.

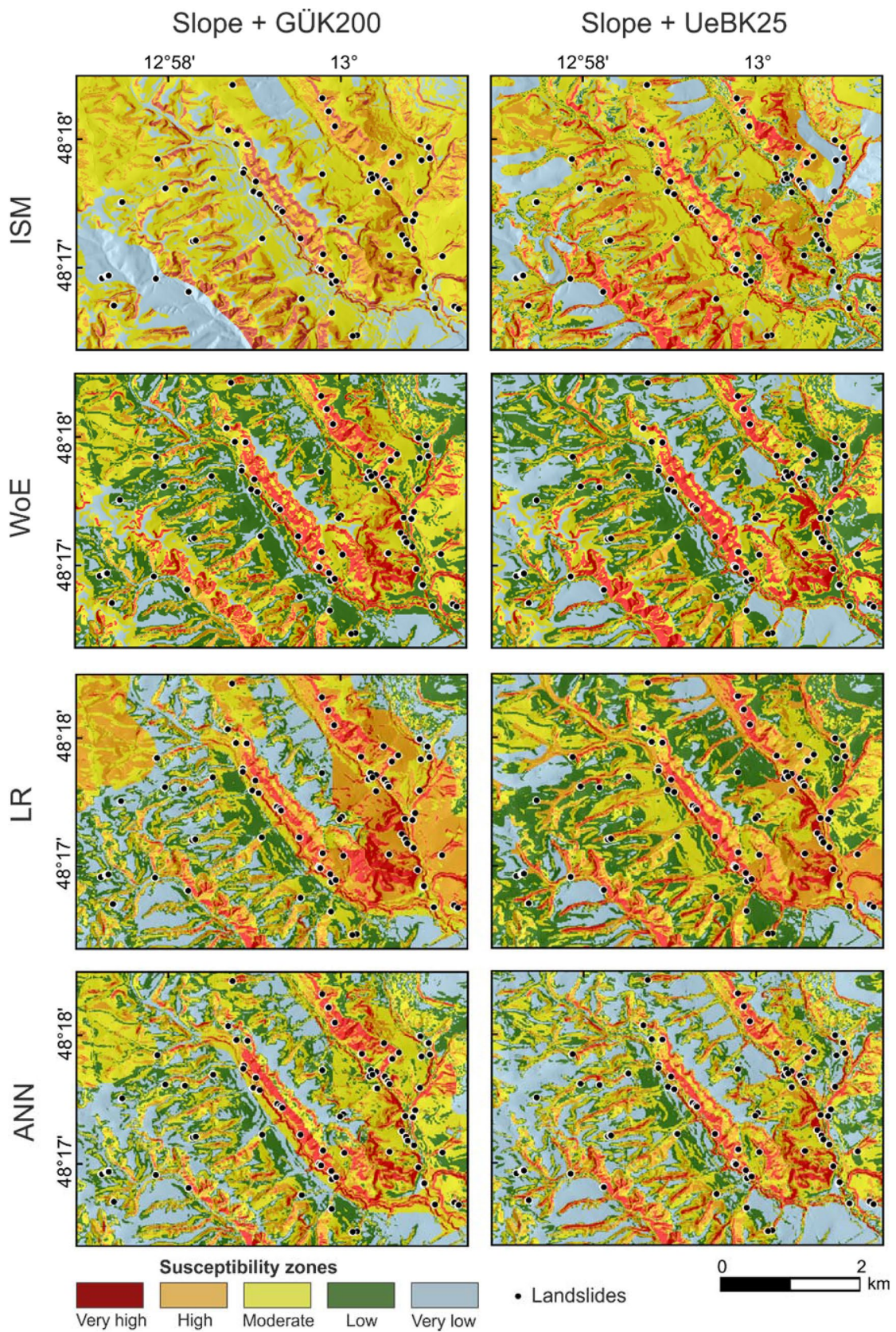


Fig. 7 Comparison of the LSA models parametrized with slope and soil maps BÜK200 and UeBK25

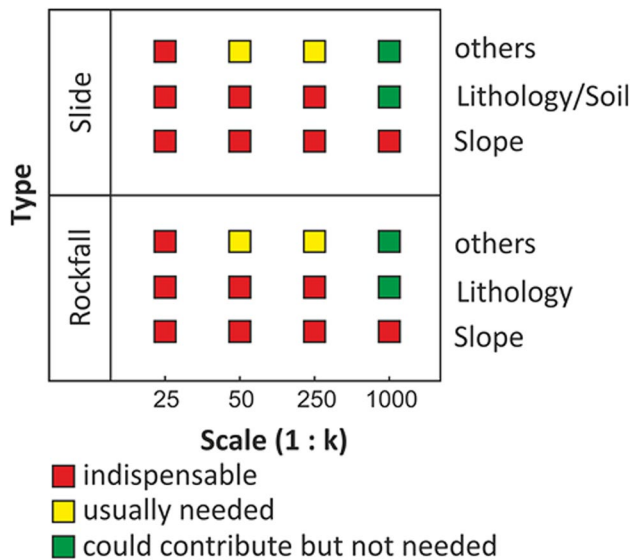


Fig. 8 Schematic overview of parameter layers needed for LSA depending on mass movement type and analysis scale

Based on the conclusions, we propose actions that could foster the development of national procedures. An initial step would be the generation of non-susceptibility maps to exclude areas not susceptible (unconditionally stable areas due to the absence of the necessary causes) to landslides following the examples given by Marchesini et al. (2014) and Jia et al. (2021). This step would significantly reduce the areas for data collection.

A further step could include introducing harmonized data acquisition, storage, and processing procedures among SGSs. From the hazard assessment point of view, establishing a harmonized national landslide database would be a quantum leap for Germany.

In particular, for the parameter lithology, additional work needs to be done regarding the reinterpretation of lithostratigraphic map units into meaningful material strength classes based on the lithological classes, their weathering conditions, and their deformation history. This is crucial to increase the model transferability among several regional-geological units.

Finally, we conclude that nationwide assessment and mapping of landslide susceptibility in Germany still needs considerable effort by the geoscientific community. According to their legal mandate, the SGDs play a key role in this context.

Acknowledgements We sincerely thank Felix Schumann for performing the comparative study for the EVT area. His contribution was essential in enhancing the quality of our research.

We also thank Prof. Dr. Tomas Manuel Fernandez-Steeger and Dr. Anika Braun for their fruitful discussions on employing artificial neural networks. Their insights and expertise greatly influenced our understanding and approach in this field.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ad-hoc-AG Boden (2000) Methodendokumentation Bodenkunde – Auswertungsmethoden zur Beurteilung der Empfindlichkeit und Belastbarkeit von Böden, vol Geologisches Jahrbuch, No 1. Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten in der Bundesrepublik Deutschland, Schweizerbart Science Publishers, Stuttgart. http://www.schweizerbart.de/publications/detail/isbn/9783510958702/Sonderheft_SG_1_Geol_Jahrb_Reihe_G_
- Ad-hoc AG Boden (2005) Bodenkundliche Kartieranleitung KA5, 5th edn. Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten, Hannover
- Ad-hoc-AG Geologie (2016) Gefahrenhinweiskarten geogener Naturgefahren in Deutschland - ein Leitfaden der Staatlichen Geologischen Dienste (SGD), vol Geologisches Jahrbuch, no 164. Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten in der Bundesrepublik Deutschland, Schweizerbart Science Publishers, Stuttgart
- Ad-hoc-AG Hydrogeologie (2016) Regionale Hydrogeologie von Deutschland, vol Geologisches Jahrbuch, no 163. Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten in der Bundesrepublik Deutschland, Schweizerbart Science Publishers, Stuttgart
- Agterberg FP, Cheng Q (2002) Conditional independence test for weight-of-evidence modeling. *Nat Resour Res* 11(4):249–255. <https://doi.org/10.1023/A:1021193827501>
- Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. *Bull Eng Geol Envir* 58:21–44. <https://doi.org/10.1007/s100640050066>
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic Regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains Central Japan. *Geomorphology* 65(1–2):15–31. <https://doi.org/10.1016/j.geomorph.2004.06.010>
- Baum RL, Godt JW, Savage WZ (2010) Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. *J Geophys Res: Earth Surf* 115(F3). <https://doi.org/10.1029/2009JF001321>

- Bernknopf RL, Cambell RH, Brookshire DS, Shapiro CD (1988) A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with applications for economic evaluation. *Bull Int Assoc Eng Geol* 25:39–56. <https://doi.org/10.2113/gseegeosci.xxv.1.39>
- Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
- Bonham-Carter GF, Agterberg FP, Wright DF (1989) Weights of evidence modelling: a new approach to mapping mineral potential. *Stat Appl Earth Sci Geol Survey Can Paper* 89(9):171–183
- Bonham-Carter GF (1994) Geographic information systems for geoscientists: Modelling with GIS, Pergamon Press, Ottawa. <https://doi.org/10.1016/C2013-0-03864-9>
- Brabb EE (1985) Innovative approaches to landslide hazard and risk mapping. In: Proceedings of ivth International conference and field workshop in landslides, vol 1. Japan Landslide Society, pp 17–22
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Can T, Nefeslioglu HA, Gokceoglu C, Sonmez H, Duman TY (2005) Susceptibility assessments of shallow earth flows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* 72:250–271. <https://doi.org/10.1016/j.geomorph.2005.05.011>
- Claessens L, Heuvelink GBM, Schoorl JM, Veldkamp A (2005) DEM resolution effects on shallow landslide hazard and soil redistribution modelling. *Earth Surf Process Landf* 30:461–477
- Clarke BJ, Otto FEL, Jones RG (2021) Inventories of extreme weather events and impacts: Implications for loss and damage from and adaptation to climate extremes. *Clim Risk Manag* 32:100285
- Coumou D, Rahmstorf S (2012) A decade of weather extremes. *Nat Clim Chang* 2:491–496. <https://doi.org/10.1038/nclimate1452>
- Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ Geol* 54:311–324. <https://doi.org/10.1007/s00254-007-0818-3>
- Dai FC, Lee CF (2003) A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic Regression. *Earth Surf Process Landf* 28:527–545. <https://doi.org/10.1002/esp.456>
- Damm B, Klose M (2015) The landslide database for Germany: Closing the gap at national level. *Geomorphology* 249:82–93. <https://doi.org/10.1016/j.geomorph.2015.03.021>
- Damm B, Varga K, Heckmann T, Becht M (2009) The impact of bedrock stratification on landslide susceptibility – An example of GIS-based landslide modelling in the Bunter Sandstone areas of northern Hesse and southern Lower Saxony (Germany). *Erde* 140:175–193
- Dikau R, Glade T (2003) Nationale Gefahrenhinweiskarte gravitativer Massenbewegungen. In: Liedtke H, Mäusbacher R, Schmidt K-H (eds) Relief, Boden und Wasser.- Nationalatlas Bundesrepublik Deutschland. Relief, Boden und Wasser.- Nationalatlas Bundesrepublik Deutschland. Institut für Länderkunde, Leipzig, pp 98–99
- Ermini L, Catani F, Casagli N (2005) Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* 66:327–343. <https://doi.org/10.1016/j.geomorph.2004.09.025>
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102:85–98. <https://doi.org/10.1016/j.enggeo.2008.03.022>
- Fisher A, Rudin C, Dominici F (2019) All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J Mach Learn Res* 20:1–81
- Freeman P, Martin L, Mechler R, Warner K (2003) A methodology for incorporating natural Catastrophes into macroeconomic projections. In: Van Der Veen AA, Arellano V, Nordvik JP (eds) Proceedings: point NEDIES and University of Twente Workshop - in search of a common methodology for damage estimation. Office for Official Publications of the European Communities, European Union, pp 51–165
- Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat Hazards Earth Syst Sci* 18:2161–2181. <https://doi.org/10.5194/nhess-18-2161-2018>
- Fuchs M, Torizin J, Kühn F (2014) The effect of DEM resolution on the computation of the factor of safety using an infinite slope model. *Geomorphology* 224:16–26
- Goepel K (2013) Implementing the Analytic Hierarchy Process as a Standard Method for Multi-Criteria Decision Making in Corporate Enterprises – A New AHP Template with Multiple Inputs. Proceedings of the International Symposium on the Analytic Hierarchy Process. Kuala Lumpur/Malaysia. <https://doi.org/10.13033/isahp.y2013.047>
- Günther A, Thiel C (2009) Combined rock slope stability and landslide susceptibility assessment of the Jasmund cliff area (Rügen Island, Germany). *Nat Hazard* 9:687–698. <https://doi.org/10.5194/nhess-9-687-2009>
- Günther A, Van den Eeckhaut M, Malet J, Reichenbach P, Hervás J (2014) Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information. *Geomorphology* 224:69–85. <https://doi.org/10.1016/j.geomorph.2014.07.011>
- 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. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Guzzetti F, Reichenbach P, Cardinali M, Galli M, Ardizzone F (2005) Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* 72:272–299. <https://doi.org/10.1016/j.geomorph.2005.06.002>
- Hammond C, Hall D, Miller S, Swetik P (1992) Level I stability analysis (LISA) documentation for version 2.0. Forest Service, Intermountain Research Station, US Department of Agriculture, Ogden
- Herrera G, Mateos R, Garcia-Devalillo J et al (2018) Landslide databases in the Geological Surveys of Europe. *Landslides* 15:359–379. <https://doi.org/10.1007/s10346-017-0902-z>
- Hervas J, Bobrowsky P (2009) Mapping: Inventories, Susceptibility, Hazard and Risk. In: Sassa K, Canuti P (eds) Landslides – Disaster Risk Reduction. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-540-69970-5>
- Hérvás J, Günther A, Reichenbach P, Chacón J, Passuto A, Malet J-P, Hobbs P, Maquaire O, Tagliavini F, Guerrieri L, Montanarella L (2007) Recommendations on a common approach for mapping areas at risk of landslides in Europe. In: Hérvás J (ed) Guidelines for mapping areas at risk of landslides in Europe. Experts meeting, JRC, Ispra, Italy. Office for Official Publications of the European Communities, Luxembourg, pp 41–45
- Hung LQ, Van NTH, Duc DM et al (2016) Landslide susceptibility mapping by combining the analytical hierarchy process and weighted linear combination methods: a case study in the upper Lo River catchment (Vietnam). *Landslides* 13:1285–1301. <https://doi.org/10.1007/s10346-015-0657-3>
- IPCC (2023) Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing

- Team, Lee H, Romero J (eds) IPCC, Geneva, Switzerland, pp 35–115. <https://doi.org/10.59327/IPCC/AR6-9789291691647>
- Jaiswal P, Van Westen CJ, Jetten V (2010) Quantitative assessment of direct and indirect landslide risk along transportation lines in southern India. *Nat Hazard* 10:1253–1267. <https://doi.org/10.5194/nhess-10-1253-2010>
- Japkowicz N, Stephen S (2002) The Class Imbalance Problem: A Systematic Study. *Intell Data Anal* 6(5):429–449. <https://doi.org/10.3233/IDA-2002-6504>
- Jia G, Alvioli M, Gariano SL, Marchesini I, Guzzetti F, Tang Q (2021) A Global Landslide Non-Susceptibility Map. *Geomorphology* 389:107804. <https://doi.org/10.1016/j.geomorph.2021.107804>
- Jibson RW, Harp EL, Michael JA (2000) A method for producing digital probabilistic seismic landslide hazard maps. *Eng Geol* 58:271–289. [https://doi.org/10.1016/S0013-7952\(00\)00039-9](https://doi.org/10.1016/S0013-7952(00)00039-9)
- Joseph VR, Vakayil A (2022) Split: An optimal method for data splitting. *Technometrics* 64(2):166–176. <https://doi.org/10.1080/00401706.2021.1921037>
- Kaynia AM, Papatoma-Köhle M, Neuhäuser B, Ratzinger K, Wenzel H, Medina-Cetina Z (2008) Probabilistic assessment of vulnerability to landslide: application to the village of Lichtenstein, Baden-Württemberg, Germany. *Eng Geol* 101:33–48. <https://doi.org/10.1016/j.enggeo.2008.03.008>
- Kirschbaum D, Stanley T, Yatheendradas S (2016) Modeling landslide susceptibility over large regions with fuzzy overlay. *Landslides* 13:485–496. <https://doi.org/10.1007/s10346-015-0577-2>
- Klose M, Damm B, Terhorst B (2015) Landslide cost modeling for transportation infrastructures: a methodological approach. *Landslides* 12:321–334. <https://doi.org/10.1007/s10346-014-0481-1>
- Klose M, Maurischat P, Damm B (2016) Landslide impacts in Germany: a historical and socioeconomic perspective. *Landslides* 13:183–199. <https://doi.org/10.1007/s10346-015-0643-9>
- Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 5:221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Kreuzer TM, Wilde M, Terhorst B, Damm B (2017) A landslide inventory system as a base for automated process and risk analyses. *Earth Sci Inform* 10:507–515. <https://doi.org/10.1007/s12145-017-0307-5>
- Lacasse S, Nadim F (2014) Landslide risk assessment and mitigation strategy. In: Duan L (ed) *Bridge engineering handbook-superstructure design*, 2nd edn. CRC Press, Boca Raton, pp 31–61. <https://doi.org/10.1201/b15621>
- LBEG (2015) *Symbolschlüssel Geologie*. Schweizerbart Science Publishers, Stuttgart http://www.schweizerbart.de/publications/detail/isbn/9783510968510/Symbolschlüssel_Geologie_4_vollig
- Lee S, Evangelista D (2006) Earthquake-induced landslide susceptibility mapping using an artificial neural network. *Nat Hazard* 6:687–695. <https://doi.org/10.5194/nhess-6-687-2006>
- Lima P, Steger S, Glade T (2021) Counteracting flawed landslide data in statistically based landslide susceptibility modelling for very large areas: a national-scale assessment for Austria. *Landslides* 18:3531–3546. <https://doi.org/10.1007/s10346-021-01693-7>
- Loche M, Alvioli M, Marchesini I, Bakka H, Lombardo L (2022) Landslide susceptibility maps of Italy: Lesson learnt from dealing with multiple landslide types and the uneven spatial distribution of the national inventory. *Earth Sci Rev* 232:104125. <https://doi.org/10.1016/j.earscirev.2022.104125>
- Lombardo L, Mai PM (2018) Presenting logistic regression-based landslide susceptibility results. *Eng Geol* 244:14–24. <https://doi.org/10.1016/j.enggeo.2018.07.019>
- Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30:4765–4774. <https://arxiv.org/abs/1705.07874>
- Luzy L, Pergalani F, Terlien MTJ (2000) Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information system. *Eng Geol* 58(3–4):313–336. [https://doi.org/10.1016/S0013-7952\(00\)00041-7](https://doi.org/10.1016/S0013-7952(00)00041-7)
- Malet J-P, Maquaire O, Thiery Y, Puissant A, Van Beek LPH, Van Asch TWJ, Remaître A (2007) Landslide risk zoning—what can be expected from model simulations? A tentative application in the South French Alps. In: Hervas J (ed) *Guidelines for mapping areas at risk of landslides in Europe*. Experts meeting, JRC, Ispra, Italy. Office for Official Publications of the European Communities, Luxembourg, pp 31–34
- Mankelov J, Murphy W (1998) Using GIS in the probabilistic assessment of earthquake triggering landslide hazards. *J Earthq Eng* 2(4):593–623. <https://doi.org/10.1080/13632469809350336>
- Marchesini I, Adrizzone F, Alvioli M, Rossi M, Guzzetti F (2014) Non-susceptible landslide areas in Italy and the Mediterranean region. *Nat Hazards Earth Syst Sci Discuss* 2:2813–2849. <https://doi.org/10.5194/nhessd-2-2813-2014>
- Martino S, Bozzano F, Caporossi P, D’Angiò D, Della Seta M, Esposito C, Fiorucci M, Giannini LM, Iannucci R, Marmoni GM, Mazzanti P, Missori C, Moretto S, Rivellino S, Schilirò L, Troiani F (2019) Impact of landslides on transportation routes during the 2016–2017 Central Italy seismic sequence. *Landslides* 16:1221–1241. <https://doi.org/10.1007/s10346-019-01162-2>
- McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer functions to soil inference systems. *Geoderma* 109(1):41–73. [https://doi.org/10.1016/S0016-7061\(02\)00139-8](https://doi.org/10.1016/S0016-7061(02)00139-8)
- Menard, Scott (1995) *Applied logistic regression analysis*. In: *Quantitative applications in social sciences*, 2nd edn, no 106. Sage Publications, Thousand Oaks
- Mergili M, Marchesini I, Rossi M, Guzzetti F, Fellin W (2014) Spatially distributed three-dimensional slope stability modelling in a raster GIS. *Geomorphology* 206(1):178–195. <https://doi.org/10.1016/j.geomorph.2013.10.008>
- Molnar C, Bischl B, Casalicchio G (2018) iml: An R Package for Interpretable Machine Learning. *J Open Source Softw* 3(26):786. <https://doi.org/10.21105/joss.00786>
- Munich Re (2023a) NatCatService. https://www.munichre.com/content/dam/munichre/mrwebsitespressreleases/nat-cat-2022-world-map.pdf/_jcr_content/renditions/original/nat-cat-2022-world-map.pdf. Accessed on 15.05.2023
- Munich Re (2023b) Natural disaster risk. Rising trend in losses. <https://www.munichre.com/en/risks/natural-disasters.html>. Accessed on 22.05.2023
- Neuhäuser B, Terhorst B (2007) Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the jurassic escarpment (SW-Germany). *Geomorphology* 86(1–2):12–24. <https://doi.org/10.1016/j.geomorph.2006.08.002>
- Pack RT, Tarboton DG, Goodwin CN, Prasad A (2005) SINMAP 2. A stability index approach to terrain stability hazard mapping, technical description and user's guide for version 2.0. Utah State University <https://hydrology.usu.edu/sinmap2/sinmap2.PDF>
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Persson M, Stevens R, Lemoine Å (2014) Spatial quick-clay predictions using multi-criteria evaluation in SW Sweden. *Landslides* 11:263–279. <https://doi.org/10.1007/s10346-013-0385-5>
- Petschko H, Brenning A, Bell R, Goetz J, Glade T (2014) Assessing the quality of landslide susceptibility maps – case study Lower Austria. *Nat Hazards Earth Syst Sci* 14:95–118. <https://doi.org/10.5194/nhess-14-95-2014>

- Picard RR, Cook RD (1984) Cross-validation of regression models. *J Am Stat Assoc* 79(387):575–583. <https://doi.org/10.2307/2288403>
- Pradhan B, Lee S (2010) Landslide susceptibility assessment, and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ Model Softw* 25:747–759. <https://doi.org/10.1016/j.envsoft.2009.10.016>
- Reichenbach P, Günther A, Guzzetti F (2007) Criteria for the identification of landslide risk areas in Europe: the Tier 2 approach. In: Hervas J (ed) *Guidelines for mapping areas at risk of landslides in Europe*. Experts Meeting, JRC, Ispra, Italy. Office for Official Publications of the European Communities, Luxembourg, pp 41–44
- Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Renger M, Bohne K, Facklam M, Harrach T, Riek W, Schäfer W, Wessolek G, Zacharias S (2008) Ergebnisse und Vorschläge der DBG-Arbeitsgruppe “Kennwerte des Bodengefüges” zur Schätzung bodenphysikalischer Kennwerte. Technische Universität Berlin, Berlin
- Ribeiro MT, Singh S, Guestrin C (2016) Why Should I Trust You?: Explaining the Predictions of Any Classifier. *Knowl Discov Databases* 16:1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Rimböck, Andreas, Hübl Johannes, Höhne Rainer (2018) Extreme torrential flooding at Simbach on June 1st, 2016 - key findings of a detailed event analysis. In: *Symposium Proceedings of the INTERPRAENENT 2018 in the Pacific Rim*. Toyama, pp 170–176
- Rumelhart D, Hinton G, Williams R (1986) Learning representations by back-propagating errors. *Nature* 323:533–536. <https://doi.org/10.1038/323533a0>
- Saaty TL (1977) A scaling method for priorities in hierarchical structures. *J Math Psychol* 15:234–281. [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5)
- Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Serv Sci* 1:83–98. <https://doi.org/10.1504/IJSSCI.2008.017590>
- Schicker R, Moon V (2012) Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology* 161–162:40–57. <https://doi.org/10.1016/j.geomorph.2012.03.036>
- Schleier M, Bi R, Rohn J, Ehret D, Xiang W (2014) Robust landslide susceptibility analysis by combination of frequency ratio, heuristic GIS methods and ground truth evaluation for a mountainous study area with poor data availability in the Three Gorges Reservoir area, PR China. *Environ Earth Sci* 71:3007–3023. <https://doi.org/10.1007/s12665-013-2677-4>
- Schumann F (2020) Investigation of approaches for landslide susceptibility mapping in the Elbtalgraben using methods of machine learning. Technical University of Berlin, Berlin
- Sim KB, Lee ML, Wong SY (2022) A review of landslide acceptable risk and tolerable risk. *Geoenviron Disasters* 9:3. <https://doi.org/10.1186/s40677-022-00205-6>
- Stanley T, Kirschbaum D (2017) A heuristic approach to global landslide susceptibility mapping. *Nat Hazards* 87:145–164. <https://doi.org/10.1007/s11069-017-2757-y>
- Steger S, Mair V, Kofler C, Pittore M, Zebisch SS (2021) Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling—Benefits of exploring landslide data collection effects. *Sci Total Environ* 776:145935. <https://doi.org/10.1016/j.scitotenv.2021.145935>
- Stevenson PC (1977) An empirical methods for the evaluation of relative risk. *Bull Int Assoc Eng Geol* 16:69–72. <https://doi.org/10.1007/BF02591451>
- Teerarungsikul S, Torizin J, Fuchs M et al (2015) An integrative approach for regional landslide susceptibility assessment using weight of evidence method: a case study of Yom River Basin, Phrae Province Northern Thailand. *Landslides* 13(5):1151–1165. <https://doi.org/10.1007/s10346-015-0659-1>
- Terhorst B, Kreja R (2009) Slope stability modelling with SINMAP in a settlement area of the Swabian Alb. *Landslides* 6(4):309–319. <https://doi.org/10.1007/s10346-009-0167-2>
- Thien TF, Yeo WS (2022) A comparative study between PCR, PLSR, and LW-PLS on the predictive performance at different data splitting ratios. *Chem Eng Commun* 209(11):1439–1456. <https://doi.org/10.1080/00986445.2021.1957853>
- Tian T, Balzer D, Wang L, Torizin J, Wan L, Li X, Chen L, Li A, Kuhn D, Fuchs M, Lege T, Tong B (2017) Landslide hazard and risk assessment Lanzhou, Province Gansu, China - Project introduction and outlook. In: Mikoš M, Tiwari B, Yin Y, Sassa K. (eds) *Advancing culture of living with landslides*, pp 1027–1033. WLF 2017. Springer, Cham. <https://doi.org/10.1007/s10346-015-0659-1>
- Torizin J (2016) Elimination of informational redundancy in the weight of evidence method: an application to landslide susceptibility assessment. *Stoch Environ Res Risk A* 30(2):635–651. <https://doi.org/10.1007/s00477-015-1077-6>
- Torizin J, Fuchs M, Awan AA, Ahmad I, Akhtar SS, Sadiq S, Razzak A, Weggenmann D, Fawad F, Khalid N, Sabir F, Khan AH (2017) Statistical landslide susceptibility assessment of the Mansehra and Thorgar districts, Khyber Pakhtunkhwa Province Pakistan. *Nat Hazards* 89(2):757–784. <https://doi.org/10.1007/s11069-017-2992-2>
- Torizin J, Wang L, Fuchs M, Tong B, Balzer D, Wan L, Kuhn D, Li A, Chen L (2018) Statistical landslide susceptibility assessment in a dynamic environment: a case study for Lanzhou City, Gansu Province NW China. *J Mt Sci* 15(6):1299–1318. <https://doi.org/10.1007/s11629-017-4717-0>
- Torizin J, Fuchs M, Kuhn D, Balzer D, Wang L (2021) Practical accounting for uncertainties in data-driven landslide susceptibility models. Examples from the Lanzhou case study. In: Guzzetti F, Mihalić Arbanas S, Reichenbach P, Sassa K, Bobrowsky PT, Takara K (eds) *Understanding and reducing landslide disaster risk*. WLF 2020. ICL contribution to landslide disaster risk reduction, Springer, Cham. https://doi.org/10.1007/978-3-030-60227-7_27
- Torizin J, Schüßler N, Fuchs M (2022) Landslide Susceptibility Assessment Tools v1.0.0b – Project Manager Suite: a new modular toolkit for landslide susceptibility assessment. *Geosci Model Dev* 15:2791–2812. <https://doi.org/10.5194/gmd-15-2791-2022>
- Van Den Eeckhaut M, Hervás J, Jaedicke C, Malet J-P, Montanarella L, Nadim F (2012) Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data. *Landslides* 9:357–369
- Van Westen C, van Asch T, Soeters R (2006) Landslide hazard and risk zonation—why is it still so difficult? *Bull Eng Geol Environ* 65:167–184. <https://doi.org/10.1007/s10064-005-0023-0>
- Vranken L, Van Turnhout P, Van Den Eeckhaut M, Vandekerckhove L, Poesen J (2013) Economic valuation of landslide damage in hilly regions: A case study from Flanders, Belgium. *Sci Total Environ* 447:323–336. <https://doi.org/10.1016/j.scitotenv.2013.01.025>
- Wadoux AMJ-C, Odeh IOA, McBratney AB (2021) Overview of Pedometrics. Reference Module in Earth Systems and Environmental Sciences. Elsevier. <https://doi.org/10.1016/B978-0-12-822974-3.00001-X>

Wilde M, Günther A, Reichenbach P, Malet JP, Hervás J (2018) Pan-European landslide susceptibility mapping: ELSUS version 2. J

Maps 14:97–104. <https://doi.org/10.1080/17445647.2018.1432511>