Science of the Total Environment 776 (2021) 145935

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

# Correlation does not imply geomorphic causation in data-driven landslide susceptibility modelling – Benefits of exploring landslide data collection effects

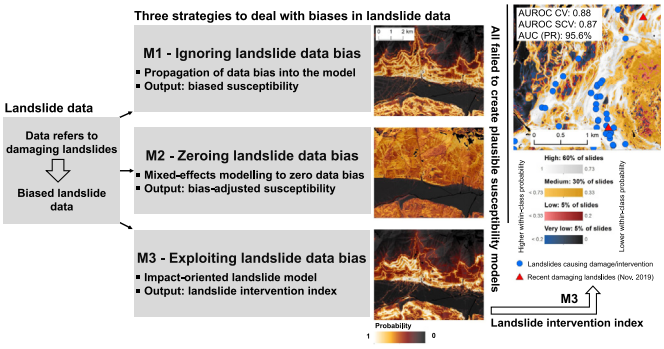
Stefan Steger [a,*], Volkmar Mair [b], Christian Kofler [a], Massimiliano Pittore [a], Marc Zebisch [a], Stefan Schneiderbauer [a,c]

[a] Eurac Research, Institute for Earth Observation, Bolzano-Bozen, Italy
[b] Office for Geology and Building Materials Testing, Autonomous Province of Bolzano-South Tyrol, Cardano-Kardaun, Italy
[c] UNU-EHS, GLOMOS Programme, Bolzano-Bozen, Italy

## HIGHLIGHTS

- Landslide susceptibility models are commonly based on biased landslide data.
- The available landslide data systematically refers to damage causing events.
- The created models represent three diverse strategies to handle such data biases.
- Despite partly high performances, all failed to reflect landslide predisposition.
- Exploiting data bias allowed modelling areas likely affected by damaging landslides.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 23 November 2020
Received in revised form 9 February 2021
Accepted 10 February 2021
Available online 18 February 2021

Editor: Paulo Pereira

*Keywords:*
Generalized additive model
Landslide inventory
Validation
South Tyrol
Landslide exposure

## ABSTRACT

Data-driven landslide susceptibility models formally integrate spatial landslide information with explanatory environmental variables that describe predisposing factors of slope instability. Well-performing models are commonly utilized to identify landslide-prone terrain or to understand the causes of slope instability. In most cases, however, the available landslide data is affected by spatial biases (e.g. underrepresentation of landslides far from infrastructure or in forests) and does therefore not perfectly represent the spatial distribution of past slope instabilities. Literature shows that implications of such data flaws are frequently ignored.

This study was built upon landslide information that systematically relates to damage-causing and infrastructure-threatening events in South Tyrol, Italy (7400 km$^2$). The created models represent three conceptually different strategies to deal with biased landslide information. The aims were to demonstrate why an inference of geomorphic causation from apparently well-performing models is invalid under common landslide data bias conditions (Model 1), to test a novel bias-adjustment approach (Model 2) and to exploit the underlying data bias to model areas likely affected by potentially damaging landslides (Model 3; intervention index), instead of landslide susceptibility. The study offers a novel perspective on how biases in landslide data can be considered within data-driven models by focusing not only on the process under investigation (landsliding), but also on the circumstances that led to the registration of landslide information (data collection effects). The results were evaluated in terms of statistical relationships, variable importance, predictive performance, and geomorphic plausibility.

* Corresponding author at: Eurac Research, Institute for Earth Observation, Viale Druso 1/Drususallee 1, 39100 Bolzano/Bozen, Italy.
*E-mail address:* stefan.steger@eurac.edu (S. Steger).

https://doi.org/10.1016/j.scitotenv.2021.145935

The results revealed that none of the models reflected landslide susceptibility. Despite partly high predictive performances, the models were unable to create geomorphically plausible spatial predictions. The impact-oriented intervention index, however, enabled to identify damage-causing landslides with high accuracy. We conclude that the frequent practice of inferring geomorphic causation from well-performing models without accounting for data limitations is invalid.



## 1. Introduction

Landslides of the slide-type movement (Cruden and Varnes, 1996; Hungr et al., 2013) represent damage-causing geomorphic phenomena in most populated mountain environments, such as the Italian province of South Tyrol (Piacentini et al., 2012). The efficiency of landslide risk reduction measures is well known to be dependent on the knowledge of where future slope instabilities can be expected (Glade et al., 2005). Thus, landslide susceptibility maps are considered valuable as they portray the propensity of an area to slope instability, without accounting for temporal components (e.g. timing, recurrence interval) or process magnitudes (Guzzetti et al., 2005; Lombardo et al., 2020). Statistical or machine learning supervised classification algorithms (henceforth termed 'data-driven') are widely applied to derive landslide susceptibility maps for large areas (Cascini, 2008; Goetz et al., 2015b; Zêzere et al., 2017; Reichenbach et al., 2018). They are used to support land management and spatial planning activities (Guillard and Zêzere, 2012; Piacentini et al., 2012; Fressard et al., 2014; Petschko et al., 2014) and to obtain insights into the factors that control slope stability (Vorpahl et al., 2012; Brenning et al., 2015; Goetz et al., 2015a; Persichillo et al., 2017; Pisano et al., 2017). Susceptibility assessments serve also as an input for complementary analyses, such as the elaboration of susceptibility-dependent critical rainfall thresholds, the development of regional landslide early warning systems or landslide hazard and risk zonation (Guzzetti et al., 2005; Remondo et al., 2005; Pereira et al., 2016; Krøgli et al., 2018; Monsieurs et al., 2019).

Most data-driven landslide susceptibility models are built upon multiple spatial environmental variables that are often considered stationary in time. These variables usually act as proxies for landslide predisposing factors (Reichenbach et al., 2018). For instance, slope angle maps describe a spatially varying downslope forcing while other frequently applied morphometric variables, such as slope aspect, elevation, curvature indices, surface roughness may rather be seen as surrogates for landslide predisposition (Reichenbach et al., 2018). Land cover/use maps are used to represent spatially varying hydrological and mechanical vegetation effects on slope stability or to describe human impact due to land use practice (Van Westen et al., 2008; Dou et al., 2015). Maps that depict geological or soil type units are used as proxies for the prevailing subsurface characteristics (Van Westen et al., 2008). Proximity maps that depict the distance of a location to specific linear or point elements (e.g. roads, rivers, springs, earthquake epicenters, fault lines) are included to reflect a variety of landslide controls, such as undercutting and oversteepening of hillslopes due to fluvial incision or road construction (Bui et al., 2011; Schicker and Moon, 2012; Brenning et al., 2015).

It is often stressed that the explanatory power of such data-driven landslide assessments relies on the quality of the landslide information which is used to create a link with environmental variables (Ardizzone et al., 2002; Guzzetti et al., 2012; Van Den Eeckhaut et al., 2012; Fressard et al., 2014; Steger et al., 2016a; Hearn and Hart, 2019; Depicker et al., 2020; Jacobs et al., 2020). In its simplest form, a landslide inventory map depicts the location of recorded slope failures. The presence of ancillary information on, e.g., the underlying mapping purpose/strategy, inventoried landslide types, potential uncertainties and known errors substantially amplifies the usability of landslide data (Malamud et al., 2004; Galli et al., 2008; Guzzetti et al., 2012; Steger et al., 2016a; Zieher et al., 2016; Samia et al., 2017). The positional precision and spatial representativeness of inventoried information is of special concern for the production of meaningful landslide susceptibility models (Steger et al., 2016a; Jacobs et al., 2020). A regional landslide inventory never contains all landslides that ever occurred in an area. For landslide susceptibility modelling, a systematic underrepresentation of landslides within areas describable by explanatory variables (e.g. specific land cover classes or altitudinal belts) is likely to result in biased models (Steger et al., 2017). For large areas, the quality of inventoried landslide data is influenced by the scope of the landslide mapping campaign, the applied mapping strategy and the underlying data source quality (Bell et al., 2012; Guzzetti et al., 2012; Marc et al., 2018; Reichenbach et al., 2018).

Aerial photo interpretation for instance is likely to result in an underrepresentation of landslide information in forest-covered terrain and specific slope orientations due to shadowing effects (Brardinoni et al., 2003; Jacobs et al., 2016) while landslide mapping based on high resolution light detection and ranging (LiDAR) digital terrain models (DTMs) may be prone to an overestimation of forest-preserved landslides (Bell et al., 2012). Although field-based mappings can favor an accurate delineation of terrain features, a spatially consistent identification of landslides can be hampered by the presence of badly visible or difficult to access terrain (Bornaetxea et al., 2018; Knevels et al., 2020). Inventories compiled from archive data or public reports can be positionally inaccurate and biased towards an overrepresentation of events that caused damage (Guzzetti et al., 1999; Steger et al., 2016b). These examples highlight that the spatial distribution of inventoried landslides is not only reliant on environmental factors that control slope stability, but also on the rationale and methodical approach behind the collection of information on past slope instabilities. The applied data collection strategy codetermines the type and degree of bias inherent in the landslide inventory.

Currently, the quality of most landslide susceptibility models and underlying statistical relationships is assessed based on performance metrics that entail a comparison of the model predictions (i.e. landslide susceptibility score) with independent landslide and non-landslide observations (i.e. test set) (Chung and Fabbri, 2003; Guzzetti et al., 2006; Frattini et al., 2010). The majority of published landslide susceptibility research avoids the critical questioning of calculated performance indicators and relies on the guiding principle 'higher predictive performance equals more meaningful results' (Steger et al., 2016b). Literature indicates that model performance estimates are used to prove the explanatory power of the results or to justify an ensuing interpretation of causal relationships between landslide occurrence and environmental variables. However, drawing direct conclusions from models with a high statistical performance may be misleading, because spatial inconsistencies in the landslide training data or the inclusion of a large portion of easy-to-classify terrain can lead to a vastly overoptimistic performance estimates, alterations in modelled associations and an underestimation of the role of epistemic uncertainties (Steger et al., 2017; Steger and Glade, 2017).

An increase in the availability of spatial environmental data, classification algorithms and computational resources has facilitated the implementation of increasingly complex models to be leveraged for assessing landslide susceptibility. Although it is known that such data-driven models can learn and reproduce biases from training data,

potential implications of widespread landslide data limitations are frequently ignored (Steger et al., 2017). Indeed, the plethora of currently published landslide susceptibility research seems to be merely guided by model performance metrics (Reichenbach et al., 2018). Since bias is almost always present in a study, researchers must consider how it affects subsequent interpretation possibilities (Gerhard, 2008; Pannucci and Wilkins, 2010).

This research builds upon landslide data from the Italian province of South Tyrol (7400 km$^2$) that systematically represents damage-causing and infrastructure-threatening landslides. The created spatial landslide models (*M1* to *M3*) correspond to three conceptually diverse strategies to deal with heterogeneously complete landslide information (i.e. underrepresentation of landslide data within specific areas). The aims were to (i) demonstrate why an inference of geomorphic causation from well-performing models can be invalid under common landslide data bias (*M1*), (ii) to test a novel bias-adjustment approach under severe landslide data bias conditions (*M2*) and to (iii) exploit landslide data peculiarities to map areas which are more and less likely affected by damaging landslides (*M3*). The *M3* approach led to the final model and was based on the assumption that the spatial distribution of registered landslides is not only determined by landslide predisposing factors (further termed susceptibility effects), but additionally by effects associated with the underlying landslide data collection procedure (further termed data collection effects). This research offers a novel viewpoint on how bias in available landslide information can be exploited by modelling not only the geomorphic phenomena under investigation (susceptibility effects), but also the spatial particularities that led to the registration of landslide information (data collection effects). The derived final landslide intervention index (*M3*) does not represent conventional landslide susceptibility as it depicts areas where damage-causing or infrastructure-threatening landslides are more and less likely to take place. All models were evaluated in terms of modelled relationships, variable importance, spatial pattern, predictive performance (non-spatial, spatial and temporal) and geomorphic plausibility.

## 2. Materials and methods

### 2.1. Study area

The Italian autonomous province of South Tyrol is located in the Eastern Alps (Fig. 1). The area extends over 7400 km$^2$ and is characterized by a strong heterogeneity in terms of geomorphology, geology, land cover and climate. The available DTM data reflects the mountainous character of the study site with a considerable difference in altitude (from 200 m to 3900 m a.s.l.) and a mean slope angle of 27°. More than 78% of the area is situated above 1200 m a.s.l. The general geological and structural setting of South Tyrol can be assigned to three main tectonic units, which are divided by prominent tectonic fault systems, such as the Periadriatic and the Brenner line. The Southalpine Units include the Dolomites and are located in the Southeast. This unit is mainly composed of volcanic rocks of the Permian Athesian Volcanic Complex and the covering sedimentary rocks (i.e. limestones and dolomites, marls). The metamorphic basement is represented by phyllites and granitic intrusions. The western and northern parts of South Tyrol can be assigned to the Austroalpine Units which are dominated by metamorphic rocks, such as schists, marbles, amphibolites, ortho-and paragneisses. The northeastern part of the area is formed by the units of the Tauern Window, which consist of calcareous mica schists and other metasediments and large metagranites, the so-called Central Gneisses (Stingl and Mair, 2005).

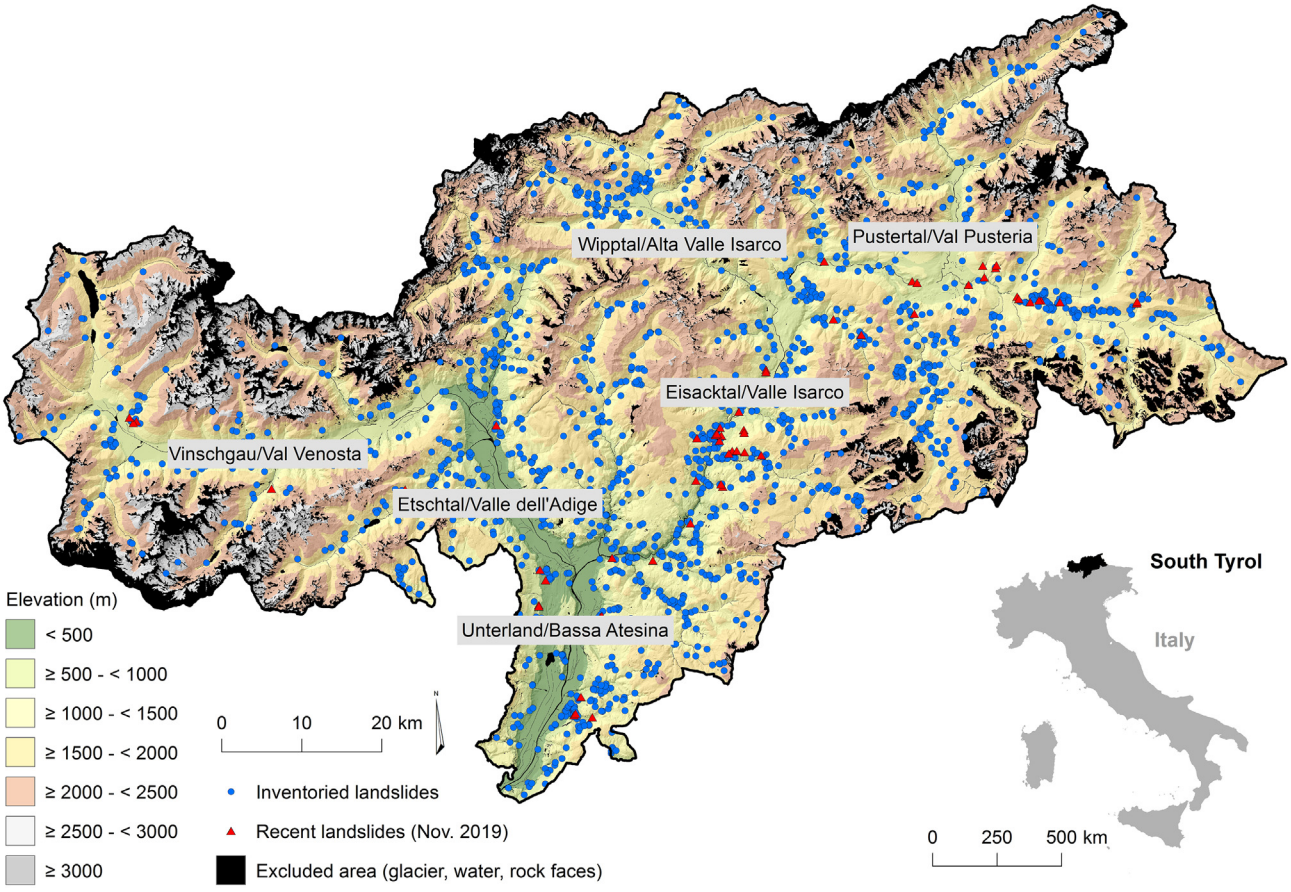



**Fig. 1.** The study area and the spatial distribution of registered landslide locations (blue points; $n = 1928$). The red triangles depict the initiation zone of damage causing shallow landslides ($n = 60$) that occurred during a severe weather event in November 2019. The area omitted from the analysis (excluded area) relates to water bodies, glaciers and rock faces.

The inner Alpine climate of South Tyrol can be described as moderately dry compared to other Alpine regions. A large portion of South Tyrol, including the majority of the central part (Etschtal/Valle dell' Adige, Eisacktal/Valle Isarco; Fig. 1), receives mean annual precipitation rates between 600 mm and 1000 mm. Lowest annual precipitation rates of around 500 mm are prevalent in the westward situated valleys of the Vinschgau/Val Venosta whilst higher rates of more than 1200 mm can be observed for the mountainous parts in the North and Northeast (Pustertal/Val Pusteria, Wipptal/Alta Valle Isarco). Elevation differences, slope orientation and associated shadowing effects play a major role in modifying temperature, insolation and precipitation regimes (Mergili and Kerschner, 2015; Lewińska et al., 2018; Zebisch et al., 2018).

Information from the year 2018 reveals that the land cover of the territory is dominated by woodlands (43% of the area: around 90% conifers, 7% mixed, 3% deciduous) and agricultural land (36% of the areal extent) (Autonomous Province of South Tyrol, 2018; Lewińska et al., 2018). Forests are predominantly located at the hillsides while agricultural land and settlements are widespread on flat terrain. The proportion of bare surface increases substantially for slope inclinations above 60° (Steger et al., 2021). A considerable share of the ~533,000 inhabitants (ASTAT, 2019) is living in municipalities and smaller towns located in the valley floors, with Bozen/Bolzano being the provincial capital (~108,000 inhabitants). A dense network of roads and paths links the smaller municipalities and dispersed farms at the hillsides.

Landsliding represents a common geomorphic phenomenon in the province of South Tyrol. Extensive research has already been conducted to investigate deep-seated movements or debris flows in South Tyrol (Corsini et al., 2005; Strozzi et al., 2005; Scheidl and Rickenmann, 2010; Casagli et al., 2016; Marra et al., 2016; Darvishi et al., 2018). Comparably few researchers tackled the topic of shallow slide-type movements in the area (Tasser et al., 2003; Piacentini et al., 2012; Steger et al., 2021). Shallow landsliding in South Tyrol is controlled by an interplay of manifold environmental factors. Besides hillslope morphology and the properties of weathered slope material also aspects related to vegetation cover and medium-term weather conditions (e.g. prolonged rain, wet seasons) were reported to affect landslide occurrence (Tasser et al., 2003; Schlögel et al., 2020). Natural landslide triggering in the area can mainly be attributed to heavy rainfall and intensive snow melt events. In recent decades, also human activities were highlighted to play an increasing role, especially due to construction activities and land use practices in hillslope terrain (Tasser et al., 2003; Stingl and Mair, 2005; Borgatti and Soldati, 2010; Piacentini et al., 2012; Steger et al., 2021).

## 2.2. Data

### 2.2.1. Spatial environmental variables

The data-driven identification of areas prone to slope instability is reliant on spatial environmental information that is observed at landslide locations and landslide-free zones. A wide range of spatial environmental variables have already been tested to model landslide susceptibility (Reichenbach et al., 2018). The number of variables and their spatial resolution varies greatly among published landslide susceptibility studies. For instance, Rotigliano et al. (2011) based their analysis on three explanatory variables whilst Rossi et al. (2010) opted for an initial set of 51 variables. The selected cell size does affect the importance of explanatory variables within a model as well as the spatial prediction pattern (Catani et al., 2013). Cell sizes from 10 × 10 m to 50 × 50 m are common whereas higher resolutions may not always be the best choice (Arnone et al., 2016; Steger et al., 2016a). The spatial variables used for this study were supposed to either represent landslide predisposing factors (susceptibility effects) or variability associated with the underlying landslide data collection procedure (collection effects; more details in Section 2.3.1). The variables can be assigned to five thematic clusters as presented in Reichenbach et al. (2018): (i) morphological variables, (ii) land cover, (iii) hydrological variables (iv) geology and (v) others (Fig. 2 and Table A.1 in supplementary material). The basic data was obtained from the South Tyrolean open Geodata platform (Geokatalog, 2019). A bilinearly resampled 10 m × 10 m LiDAR-DTM, SAGA GIS (Conrad et al., 2015) and GRASS GIS (GRASS Development Team, 2019) were used to derive the morphological variables slope, aspect, contributing catchment area, curvature and geomorphons (i.e. terrain forms as described by Jasiewicz and Stepinski, 2013). Spatial land cover information was utilized to obtain a land cover map (classes: infrastructure, forest, pasture, agricultural land, bare surface) and forest presence-absence information ('Realnutzungskarte Südtirol' version 2015). Euclidean distances were calculated for linear and areal features to derive proximity-to-feature

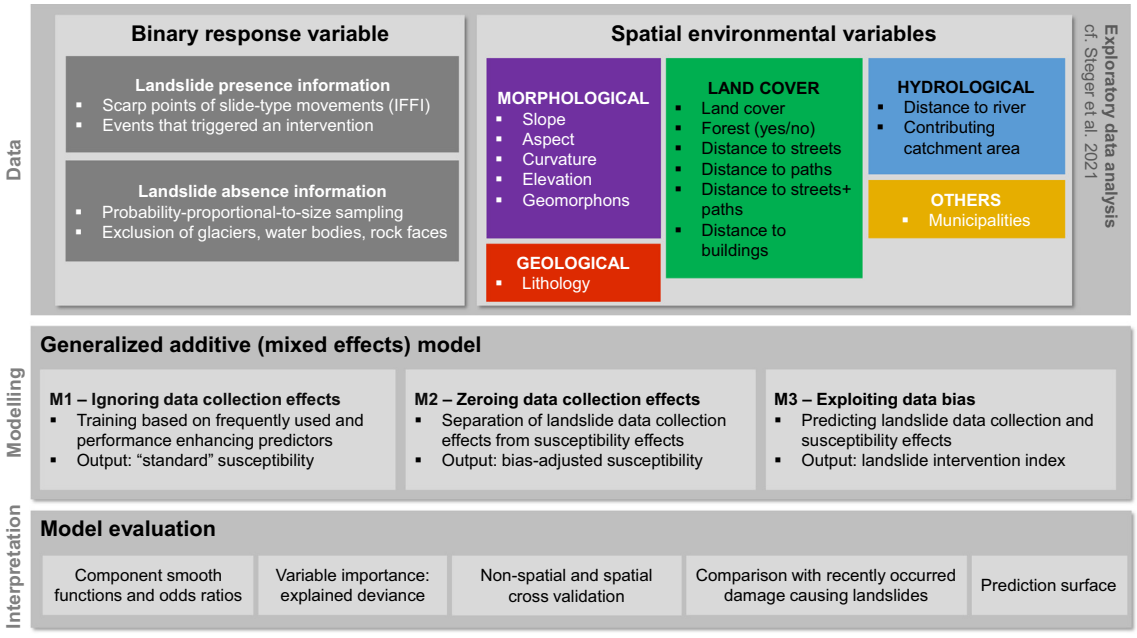



**Fig. 2.** Overview of the methodical framework. Different input data combinations (details in Table A.1 in supplementary material) were used to produce three spatial landslide models (*M1–M3*). The interpretation was based on a critical inspection of multiple metrics and model visualization techniques.

information: distance to streets (asphalt roads, highways), distance to paths (unpaved roads and trails), distance to linear infrastructure (streets and paths merged), distance to rivers (perennial water courses) and distance to buildings (building footprints). The analyses also included a geological overview map ('Geologische Übersichtskarte Südtirol') and a shapefile of administrative units (i.e. municipality ID). More details on the spatial data sets can be found in Steger et al. (2021).

#### 2.2.2. Landslide inventory

The spatial landslide information is based on the Italian landslide inventory (Inventario dei fenomeni franosi in Italia: IFFI). The national IFFI project was launched in 1999 with the aim to record landslides of different movement types for the entire Italian territory using a standardized approach. The IFFI activities are coordinated by the institute for Environmental Protection and Research (ISPRA) while currently each Italian region and self-governing province (e.g. South Tyrol) is responsible to assemble landslide data for its area. It is known that the quality and characteristics of IFFI information varies across the individual mapping domains, i.e. regions and self-governing provinces. Statistical analyses indicated a heterogenous completeness of landslide information among the mapping domains and a varying completeness of landslide information in areas far from urban settlements, since some regions focus more on urban areas than others (Trigila et al., 2010). Differences in the applied data collection strategy (e.g. aerial photo interpretations vs. archived damage reports vs. field surveys) among the mapping domains and the resulting heterogenous landslide information hampers a straightforward application of IFFI data to assess landslide susceptibility at national scale (Trigila et al., 2013).

Up to January 2019, the South Tyrolean IFFI data contains 7573 inventoried landslides of different movement types. For this research, 1928 shallow landslides of the slide-type movement were extracted from this data set (blue points in Fig. 1). The available point information relates to landslide scarp positions which were frequently mapped with high precision using differential Global Positioning Systems (dGPS). The South Tyrolean IFFI data covers so called space-relevant areas by design ('raumrelevante Flächen'). This implies that landslides that triggered an intervention by the provincial authorities (e.g. geological survey, road service, forest service) for documentation, mitigation and prevention are systematically registered while landslides that did not pose a threat or caused a damage are usually not documented (Steger et al., 2021). The majority of registered events damaged or threatened asphalt roads, and to a lower extent unpaved paths and trails. A comparable low portion of inventoried landslides caused direct damage on buildings. Most landslides were inventoried within the last three decades with an increasing number of registrations since the introduction of the IFFI project around the turn of the millennium. The comparable high number of registered landslides in the recent decades is codetermined by a more rigorous documentation and digitalization of damaging landslides in South Tyrol.

The data also contains landslides which were investigated to compile the official natural hazard plans at municipality level. In South Tyrol, the natural hazard planning formally focuses on built-up areas (incl. a 300 m buffer around the buildings), small settlements, single houses, infrastructure, and main lines of public interest (e.g. roads outside built-up areas) (Gefahrenzonenplan Südtirol, 2021). Variation in the completeness of landslide information among the municipalities can be expected, also due differences in the phases of the natural hazard plan development. At present, around half of the municipalities already have an approved hazard plan in place while others have only started the analysis phase (Gefahrenzonenplan Südtirol, 2021). Furthermore, the fact that different experts are in charge for compiling landslide information for the municipalities may also introduce spatial data heterogeneities.

In summary, the available landslide information mainly relates to slope instabilities that caused damage or threatened buildings and infrastructure. The results of an initial statistical exploratory analysis build the foundation for this research (cf. Steger et al., 2021). Associated univariate statistics highlighted that landslides were frequently registered for medium inclined slopes, low altitude areas and terrain forms classified as slopes and footslopes. The analysis also showed that landslides are much more frequently registered for areas close to streets, forest roads and buildings. In addition, the three models presented here were also validated with newly available temporally independent shallow landslide data ($n = 60$). The associated points depict the scarp location of landslides that caused damage during a heavy weather period in November 2019.

### 2.3. Methods

#### 2.3.1. Methodical framework: Susceptibility effects vs. data collection effects

Data on inventoried landslide occurrence and spatial environmental variables were used to create and evaluate three spatial landslide models (Fig. 2). The research design built upon the assumption that the spatial distribution of available landslide data is not only determined by landslide causing factors (susceptibility effects), but also by the underlying landslide data collection strategy (data collection effects).

The general approach relied on (i) a close exchange between the modelling team and the landslide data manager/provider (i.e. Office for Geology and Building Materials Testing) and on (ii) a recently conducted analysis of inventoried slide-type movements (Steger et al., 2021) to identify whether a spatial variable represented a plausible susceptibility effect or rather a data collection effect. For example, slope angle was considered a susceptibility effect, because of its observed plausible relation to landslide occurrence while we found no evidence that differently inclined slopes were treated differently during the landslide data collection procedure. In contrast, the elaborated statistics and the known overrepresentation of inventoried landslides in close proximity to infrastructure endorsed the decision to consider the variable 'distance to streets and paths' as a data collection effect. The three derived models (*M1*, *M2*, *M3*) represent three different options to deal with heterogeneously complete landslide data (Fig. 2):

- The ***M1 approach*** ignores geomorphically implausible relationships and potential implications of heterogeneously complete landslide information on purpose by treating each variable that enhances the performance of the model as a landslide susceptibility effect. The *M1* approach represents a 'frequent practice' in landslide susceptibility modelling that builds its prediction upon commonly applied and performance enhancing explanatory variables (cf. Reichenbach et al., 2018) while neglecting consequences of erroneous input data. The *M1* modelling procedure was expected to provide a striking example on how a strict maximization of predictive performances can pave the way to modelling results with little geomorphic explanatory power.
- The ***M2 approach*** aims to produce a bias-adjusted landslide susceptibility map by separating landslide susceptibility effects from the effects that describe the data collection procedure. Novel mixed-effects modelling was implemented to predict landslide susceptibility while simultaneously accounting for variability evolving from landslide data bias.
- The ***M3 approach*** tackles the topic of spatial landslide modelling from a new perspective by exploiting specificities associated with the landslide data collection strategy. The aim was not to model landslide susceptibility, but rather to identify areas where future landslides are more likely to be recorded in the context of the systematic provincial landslide mapping strategy. The *M3* map is based on a simultaneous prediction of both, landslide susceptibility effects and data collection effects. The derived landslide intervention index highlights areas where damage-causing or infrastructure-threatening landslides are more or less likely to occur because (i) the underlying mapping systematically focused on landslides that triggered an intervention and (ii) because the model setup allowed to also describe spatial particularities of this landslide mapping procedure.

A detailed description of the modelling methods can be found in Section 2.3.3 while the associated variables are presented in the supplementary material (Table A.1). The results were evaluated at the model level (predictive performance, relative variable importance, prediction surface) and at the level of single variables (component smooth functions, odds ratios). The final prediction patterns were also compared with temporally independent information on damaging landslides (red triangles in Fig. 1).

*2.3.2. Landslide presence data and probability-proportional-to-size sampling of landslide-absence*

Several strategies have already been applied to represent landslide presences and absences for grid-based landslide susceptibility models (Van Den Eeckhaut et al., 2012; Regmi et al., 2014; Conoscenti et al., 2016; Hussin et al., 2016; Bornaetxea et al., 2018). Sampling one point for each landslide is beneficial to lessen undesired effects related to spatial autocorrelation, to save computational resources and to treat differently sized landslides equally (Atkinson et al., 1998; Van Den Eeckhaut et al., 2006; Petschko et al., 2014; Goetz et al., 2015b; Zêzere et al., 2017). For this research, landslide presence locations were directly derived from the available point-based IFFI landslide inventory that contains 1928 landslide scarps (Section 2.2.2).

In South Tyrol, shallow landslides cover just a small portion of the entire area which is why a random sampling of grid-cells that do not relate to mapped landslide locations is reasonable to summarize the characteristics of landslide-free zones for the purpose of statistical analyses (Blahut et al., 2010; Goetz et al., 2015b; Steger and Glade, 2017). Strategies applied to sample landslide-free zones are usually based on simple random point sampling within a planarly projected study area. However, such sampling ignores the discrepancy between the planar dimension of an area, as shown in a GIS planar view, and the actual surface area. Seen from above (planar view), the portion of visible surface area becomes smaller with an increasing slope inclination (Fig. 3a). Thus, randomly sampled landslide-absence locations underrepresent steep terrain.

Probability-proportional-to-size (PPS) sampling (Singh and Mangat, 1996) allows to take advantage of auxiliary information to define selection probabilities. For this study, the selection probability of a raster cell was set proportional to a size measure that describes the 'actual' surface area of each cell ($A_{surface}$). A trigonometric function was applied to calculate $A_{surface}$ on the basis of the known planar surface area of a grid cell ($A_{planar}$; in our case 100 $m^2$) and the local slope angle S. $A_{surface}$ is the ratio between the nominal grid cell area and the cosine of slope inclination in radians ($S_r$; derived from the 10 m DTM).

$$A_{surface} = {}^{A_{planar}}/_{\cos S_r} \quad (1)$$

The sampling probabilities were represented by $A_{surface}$. The likelihood of a raster cell to be sampled was defined to be dependent on its inclination with increasing probabilities for steeper terrain as exemplified in Fig. 3. For instance, the chances of a cell to be selected as landslide free zones are two times higher for 60° steep terrain compared to its flat (0°) counterpart. An equal number of landslide absence and presence data (1928 observations for each class) was selected to create the binary response. The landslide presence information was considered unaffected by such a projection-related sampling bias, also because the landslides were directly mapped in the field.

Areas related to water bodies (i.e. lakes, rivers), glaciers and rock faces were not used as potential candidates for the PPS landslide absence sampling to lessen undesirable effects associated with the inclusion of trivial terrain (Steger and Glade, 2017). Areas within close distance (50 m radius) to inventoried landslide points were also excluded from the non-landslide sampling procedure to minimize the chance of drawing samples from known landslide terrain.

In combination with classifiers that allow accounting for nonlinear relationships, the applied PSS sampling is expected to be of utility for areas where – in comparison to the general slope angle distribution in the study area – medium inclined slopes are most susceptible to landsliding. Compared to a simple GIS-based random sampling procedure, this strategy results in a higher portion of non-landslide samples in steeper terrain which counterbalances (i) undesired effects related to the presence of trivial flat terrain and (ii) an underrepresentation of landslide-free zones in steep terrain leading to e.g. too high susceptibility scores in very steep terrain, such as rock outcrops) (Steger and Glade, 2017). For South Tyrol, differences in the model response are particularly relevant for very steep terrain (Fig. 3d).

*2.3.3. Classification algorithm and variable selection*

A generalized linear model (GLM) with a logistic link function (also known as logistic regression) is the most popular classifier for assessing landslide susceptibility (Budimir et al., 2015; Wu et al., 2015; Reichenbach et al., 2018). A semi-parametric generalized additive model (GAM) can be considered an extension of a GLM. In contrast to a GLM, a GAM allows to fit smoothing functions through continuously scaled data to reflect nonlinear associations. Since GAMs combine efficient statistical learning, a high flexibility and a high transparency they are also termed interpretable machine learners (Zuur et al., 2009; Wood, 2017; Molnar, 2020). Model transparency and interpretability is particularly useful to detect bias and debug models, because flexible classifiers are likely to learn errors inherent in the training data (Steger et al., 2017; Molnar, 2020). GAMs have successfully been applied to model landslide susceptibility or to get insights into (de)stabilizing factors in a variety of environmental contexts (Petschko et al.,

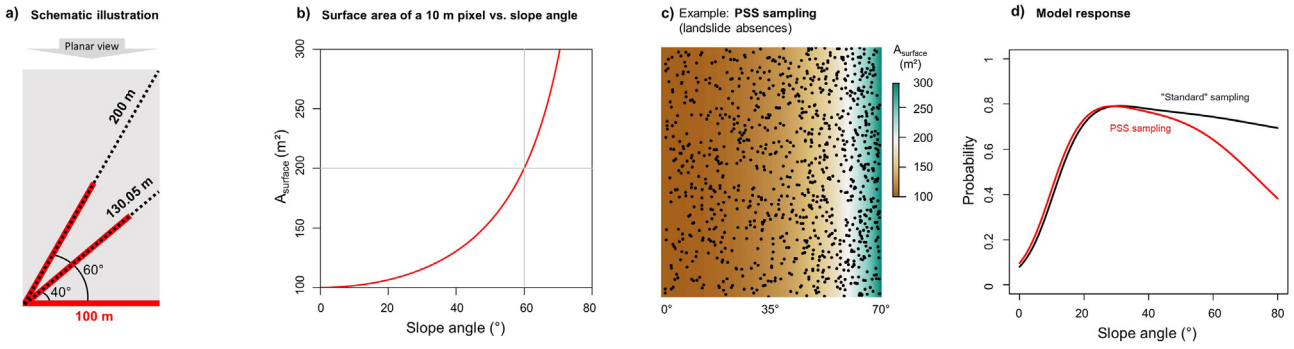



**Fig. 3.** Visualization of the implemented PPS sampling approach. The schematic illustration in (a) shows that a 60° (40°) steep slope with a length of 200 m (130.05 m) relates to a visible horizontal length of 100 m in a planar projection. The plot (b) shows the relationship between the surface area of a 10 m raster cell and slope inclination (e.g. a 10 m pixel in 60° steep terrain represents a surface area of 200 $m^2$). PPS sampling of 1000 points from flat to steep terrain is shown in (c). The plot (d) exemplifies predicted susceptibility scores (y-axis) as a function of slope angle (x-axis) and sampling strategy (line color): the data relates to the study area while the underlying generalized additive model was trained using the predictor slope angle and elevation (elevation was fixed at 1000 m for prediction).

2014; Brenning et al., 2015; Goetz et al., 2015a; Bordoni et al., 2020; Brock et al., 2020; Knevels et al., 2020).

Most algorithms applied to model landslide susceptibility can be assigned to the group of 'fixed effects models', where the considered variables are used to model the direct influence on the response. Instead, mixed effects models contain a fixed effect part and a random effect part. Random effects allow to account for hierarchical data structures and to capture variation related to nuisance variables (Bolker et al., 2009; Zuur et al., 2009). Mixed-effects GLMs have recently been applied to tackle the challenge of modelling landslide susceptibility using systematically incomplete landslide information (Steger et al., 2017).

Using conventional 'fixed effects models', a direct propagation of a systematic inventory-based bias can be reduced by simply ignoring bias-describing variables. For instance, too low predicted landslide probabilities in forested terrain due to an underrepresentation of landslide data within forests can partly be counteracted by neglecting variables such as 'forest yes/no' or 'land cover' during modelling. However, in environmental research, variables are usually interrelated in space, such as forests are primarily located at medium inclined slopes in South Tyrol. Thus, modelled relationship between landslide occurrence and 'non-bias describing variables' (e.g. slope angle in this case) can be confounded. In the context of this example, model distortions will follow if the bias-describing variable land cover does not account for this bias-related data variability during model parameter estimation, ultimately because the variability arising from the landslide data bias is explained by the spatially interrelated variable slope angle. In this case, the consequence would be a too low predicted landslide probability at medium inclined slopes. In other words, introducing a bias-describing variable as a random effect (e.g. land cover if the landslide data is heterogeneously complete among land cover units) enables to account for the associated variation during model parameter estimation. The prediction can then be based on the fixed effects part alone while the bias-describing random effects are zeroed (i.e. averaged-out). This procedure allows to avoid (i) a direct propagation of the bias into the final predictions and (ii) possibly confounding effects due to a bias-describing lurking variable (Steger et al., 2017). The potential of generalized additive mixed effects models (GAMMs) is unexplored in the field of spatial landslide prediction.

For this research, GAMs (fixed effects only; *M1*) and GAMMs (fixed and random effects; *M2* and *M3*) as implemented in the 'mgcv' R package were applied while the associated smoothing functions were automatically fitted using internal cross validation (Wood, 2017). The exclude argument in the predict function of the mgcv package was used to zero (i.e. average-out) specific terms in the models *M2* and *M3* for the spatial prediction (marked as 'Zero' in Table A.1 in supplementary material).

Table A.1 in supplementary material lists the variables and their use within the respective models. In this context, the expression 'predict' marks variables which were used for model fitting and spatial prediction (the conventional approach in susceptibility modelling). The term 'zero' relates to variables which were only used for the estimation of model parameters, but zeroed during prediction to avoid a bias-propagation and confounding effects. The expression 'not considered' marks variables that were ignored within the respective model. The variables slope angle, aspect, curvature, geomorphons and lithology were used for the prediction within all models (*M1–M3*), because these variables (i) increased the proportion of explained deviance and (ii) because no evidence was found that this data describes variability associated with the landslide data collection procedure.

The *M1* approach intentionally ignores potential effects of data bias on the results, which is why variables that are likely to relate to a landslide data incompleteness were included. *M1* additionally comprised as direct explanatory variables the frequently used and performance-enhancing variables land cover, distance to linear infrastructure, distance to rivers, upslope contributing area and elevation.

An explicit consideration of landslide data collection effects was envisaged for building the models *M2* (data collection effects are zeroed) and *M3* (data collection effects are predicted). We opted to include a simplified version of the land cover layer (forest yes/no) to account for the suspected effect that the occurrence and mapping frequency of damage causing landslides differs among forested and non-forested terrain. The anticipated effect of an overproportional reporting of landslides nearby buildings was represented by the 'distance to buildings' layer. For constructing the models *M2* and *M3*, we opted to split the former 'distance to infrastructure' layer into one variable that represents the distance to streets and another one that describes the distance to paths. This enabled considering that landslide reporting frequency in the area is codetermined by the relevance of affected infrastructure (i.e. landslide damage is known to be very systematically recorded for higher ranked roads). Since mapping of landslides was not systematically conducted in high-altitude areas, we considered the variable 'elevation' to be mainly related to a data collection effect (further discussion in Section 4.2).

The variable 'municipality ID' was included in the *M2* and *M3* model fitting to account for a nuisance (random) effect associated with a varying completeness of landslide data among these administrative units. However, a spatial prediction of this data collection effect with an explicit dependence on the municipality was deemed detrimental, because future landslides (*M2*) or future landslide interventions (*M3*) are not necessarily expected to be more frequent in municipalities with a more rigorous documentation of past events. Thus, the municipality ID effect was zeroed for the final prediction. The two variables 'upslope contributing area' and 'distance to rivers' were not considered for building *M2* and *M3*, because these variables did not depict a geomorphically plausible relationship to landslide occurrence (e.g. negative correlation to landslide occurrence) nor did they relate to the underlying data collection procedure.

#### 2.3.4. Model evaluation and visualization of the landslide intervention index (M3)

The result interpretation was supported by numerous model evaluations. The relative importance of variables was estimated by comparing the explained deviance of a full model (all variables) with the drop in explained deviance from a model excluding the variable of interest. This variable importance indicator was expressed as the proportion of deviance explained. The relative importance of a variable was considered higher if the model without the variable of interest exhibited a lower goodness-of-model-fit compared to the original model. Identical smoothing parameters were applied to the variables throughout the analysis to avoid inconsistencies due to changes in the smoothing parameter estimation (i.e. the smooths were fixed and identical for the full and reduced models) (Goetz et al., 2018; Knevels et al., 2020).

Transformation functions (i.e. component smooth functions) were visualized to uncover modelled predictor-response relationships for continuously scaled variables while accounting for effects of the remaining variables (Zuur et al., 2009; Wood, 2017; Knevels et al., 2020; Molnar, 2020). Odds ratios (ORs), derived from exponentiated model coefficients, depicted modelled relationships for categorical variables. ORs can be considered a relative measure of effect size that illustrate the estimated chance of a certain variable class (e.g. forest) to be affected by a landslide in comparison to a predefined reference class (OR = 1, e.g. pasture) while considering the effects of the other variables in the model (Steger et al., 2017).

The fitting and predictive performance was assessed by calculating the area under the receiver operating characteristic curve (AUROC) (Swets, 1988). Partitioning of the data into multiple training sets (i.e. fitting performance) and test sets (i.e. predictive performance) was achieved by means of k-fold cross validation (CV) and k-fold spatial cross validation (SCV). CV is based on a non-spatial repeated random splitting of training and test data whereas SCV builds upon on a k-means cluster algorithm to achieve spatially explicit partitions of

training and test areas (Brenning, 2012; Steger et al., 2016b; Schratz et al., 2018). In this context, 500 AUROCs (50 repetitions times 10 folds) were computed for each model and data partitioning design. SCV interquartile ranges were interpreted as an indicator for the spatial transferability and robustness of the spatial model predictions (Petschko et al., 2014; Steger et al., 2017).

Prediction rate curves (Chung and Fabbri, 2003) and the associated area under the curve (AUC) were used to confront the final prediction pattern with the scarp location of 60 shallow landslides that caused damage during a recent heavy weather event. The respective curves depict the cumulative proportion of true positive observations as a function of the proportion of the total area at decreasing predicted susceptibility scores (Jacobs et al., 2020).

The conducted geomorphic plausibility check (Steger et al., 2016b) was not based on a confrontation of visible landslide features with spatially predicted susceptibility scores because (i) landowners frequently remove visible landslides footprints in the area and because (ii) terrain with no geomorphic evidence of former landslide activity is not necessarily protected against future slope instabilities. The geomorphic plausibility check of the models was inspired by Oreskes et al. (1994) and the concept of biological plausibility (e.g. Holland, 1986). A model was therefore considered geomorphically implausible if the underlying modelled associations (i.e. component smooth functions and ORs) directly reflected known data flaws (i.e. data collection effects) or if the modelled associations were in obvious disagreement with ascribed susceptibility effects (cf. Table A.1 in supplementary material). High predictive performance scores should confirm the efficiency of the model to classify out-of-model landslide observations.

The intervention index map (*M3*) will further be utilized by local authorities (Section 4.3). The practical applicability and interpretability of this map was enhanced by grouping the continuously-scaled probability scores into four color-coded classes following the example of Petschko et al. (2014). The selected class-thresholds relate to the portion of registered landslides falling within each class: 90% of inventoried past landslides fall into the classes 'high' (60% of landslides; probabilities 1 to 0.73) and 'medium' (30% of landslides; probabilities <0.73 to 0.33) while 10% of registered slides fall into the classes 'low' (5% of landslides; probabilities <0.33 to 0.2) and 'very low' (remaining 5% of landslides; probabilities <0.2 to 0). A further visual within-class differentiation was achieved by modifying the brightness of the color-coded classes according the within-class probability score: the brighter the color the higher the within-class landslide probability score and vice versa.

## 3. Results

### 3.1. Important variables and modelled relationships

The proportion of explained deviance provided insights into the relative importance of each variable (Fig. 4). For *M1*, the four variables distance to streets/paths (0.2, rank 1), elevation (0.2, rank 2), slope angle (0.15, rank 3), land cover (0.07, rank 4) and geomorphons (0.03, rank 5) were associated with the highest computed variable importance (Fig. 4a). The component smooth functions of the three highest ranked variables (Fig. 5 top row) indicated that the modelled landslide likelihood was highest for medium inclined hillslopes (approx. > 20° and < 55°), at lower altitudes (approx. < 1500 m) and in close distance to linear infrastructure (approx. < 100 m from streets and paths). While accounting for the influence of the other variables in the model, forested areas were associated with the lowest chance of being mapped as a landslide location (OR 0.4) compared to infrastructure (OR 1.6), agricultural land (OR 1.4), bare surface (OR 2) and pasture (OR 1, reference class). Terrain forms (geomorphons) classified as slopes (OR 1.7) and footslopes (OR 1.6) were related to the highest modelled odds of landsliding followed by spurs (OR 1.3), ridges (OR 1), valleys (OR 1, reference class) and flat terrain (OR 0.4) (Table A.2 in supplementary material).

The model fit behind *M2* and *M3* was built to simultaneously account for landslide susceptibility effects (first five variables in Fig. 4b) and landslide data collection effects (remaining variables in Fig. 4b). The highest portion of explained deviance was obtained for the variables distance to streets (0.14, rank 1), slope angle (0.12, rank 2), distance to paths (0.07, rank 3), the categorical variable municipality ID (0.07, rank 4) and elevation (0.04, rank 5). Among these five most 'important' variables, only slope angle represented a landslide susceptibility effect. Thus, the major share of data variability was explained by factors that relate to the underlying data collection procedure and not to the actual geomorphic process under investigation. Component smooth functions (Fig. 5 bottom row) and the differences in explained deviance (distance to streets vs. distance to paths, Fig. 4b) indicated that landslides were very frequently mapped close to streets and regularly close to unpaved paths. The comparably low importance of the variable distance to buildings exemplified that the interpretation of the modelling results should be conducted with care, because the included variables may frequently compete for the same shared variability (e.g. an area close to a building is also close to linear infrastructure). In

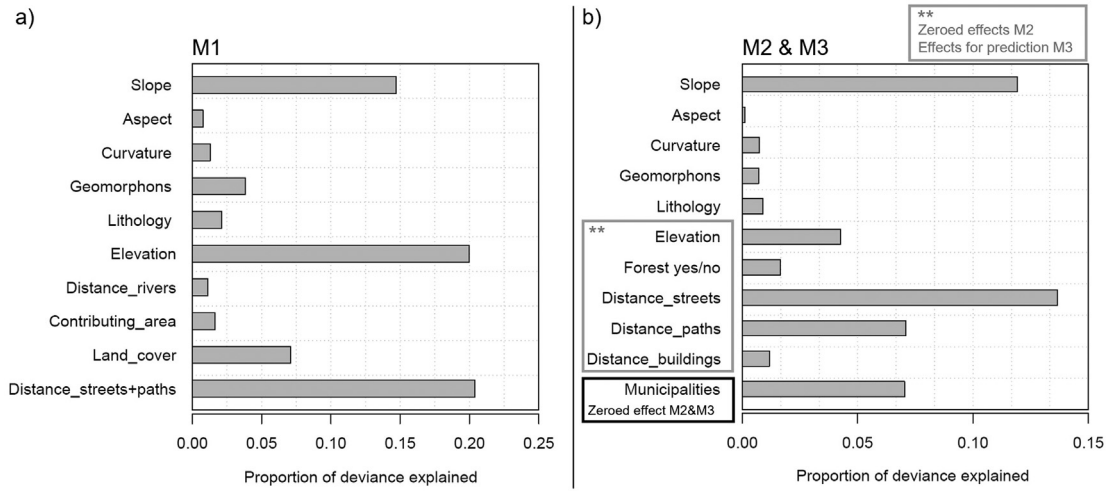



**Fig. 4.** Variable importance expressed as the proportion of deviance explained. The model M1 (a) relates to a GAM where all variables were used for the spatial prediction. The models M2 and M3 (b) are based on an identical GAMM fit while the variables in the boxes were treated differently (zeroed vs. prediction) for the spatial prediction (cf. Table A.1 in supplemenatry material).

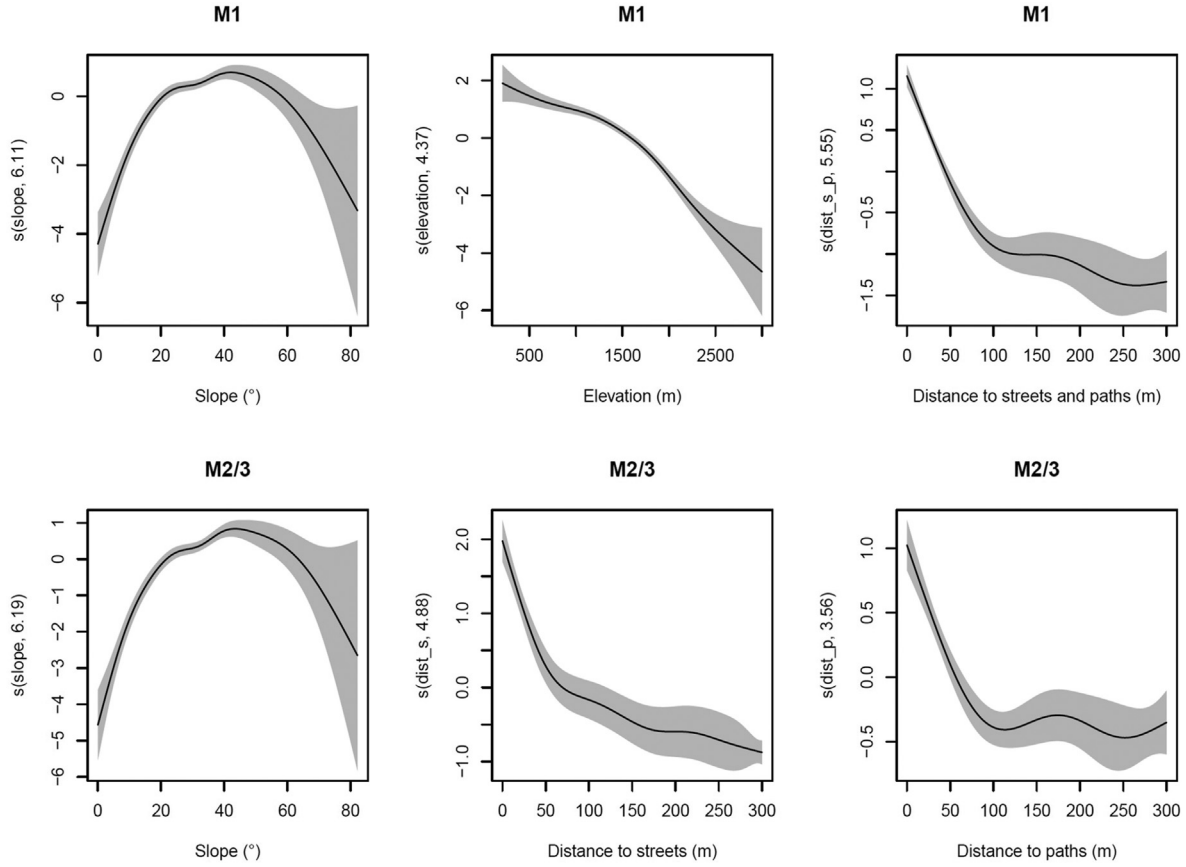



**Fig. 5.** Component smooth functions (centered transformations) and confidence bands (95%) for the three most important variables for the model M1 (top row) and M2/3 (bottom row). The plots visualize the modelled relationships between inventoried landslide occurrence and continuously scaled variables (y-axis values >0 depict an estimated above average modelled likelihood of landslide occurrence and vice versa).

analogy to *M1*, higher slope positions, convex terrain forms and higher elevations (particularly above ~1500 m a.s.l.) were associated with estimated low chances of landslide occurrence. The high importance rank of the variable municipality (rank 4) indicated that the spatial distribution of mapped landslides can partly be explained by these administrative units, even if the influence of the other variables is considered. This further supported the suspicion that the available landslide information is heterogeneously complete among the municipalities (cf. Section 2.2.2). Lower altitudes were estimated to be less likely affected by landsliding. ORs of the variable forest yes/no (rank 6) showed that the modelled chances of landslide occurrence were 1.9 times higher for non-forests compared to forests (OR = 1) (Table A.2 in supplementary material).

### 3.2. Spatial prediction pattern

The maps shown in Fig. 6 are a spatial representation of the modelled relationships as a function of the variable values observed for each raster cell ('Predict' in Table A.1 in supplementary material). The shown probability scores range between 0 and 1 and depict whether the predicted local conditions were similar (values closer to 1) or dissimilar (values closer to 0) to the conditions observed for registered landslides. Inspection of the spatial prediction patterns revealed similarities between the models *M1* and *M3* and a rather different pattern for the model *M2*. A visual comparison of the maps (*M1* vs. *M3*) indicated that ignoring the presence of data collection effects (*M1*) or explicitly predicting data collection and susceptibility effects (*M3*) did not cause substantial differences in the final prediction pattern. Both maps displayed highest predicted landslide probabilities at medium-inclined, concave-shaped and unforested hillslopes which are close to linear infrastructure and located at lower slope positions. In this context, the *M3* map accentuated positions close to streets while *M1* did not differentiate between streets and unpaved paths. Particularly low probabilities were predicted for flat and very steep terrain and for high altitude areas.

Averaging-out the effects associated with the landslide data collection procedure (*M2*) led to a low spatial variation of predicted landslide probabilities at the hillslope areas (Fig. 6b). This pattern reflected the strong dependency of the model prediction on slope angle variations. Flat and very steep terrain were associated with low probabilities while medium inclined slopes were related to medium and high predicted landslide probability scores. In contrast to the M1 and *M3* maps, the *M2* spatial pattern depicted high alpine areas as prone to landsliding. Visually discernible spatial heterogeneities at the hillsides were associated with changes in the terrain form variables curvature and geomorphons. Higher probabilities were predicted for concave-shaped terrain at slope and footslope positions. From a regional viewpoint, the effect of lithology was also discernible at the transitions between lithological classes.

In a strict sense, the *M3* map reflected areas that are more or less likely to become a landslide mapping location if the current provincial landslide mapping strategy is maintained. Since the mapping was systematically conducted for landsides that caused a damage or triggered an intervention, the final spatial pattern depicts the relative spatial likelihood of damage-causing or infrastructure-threating landslides. Fig. 6c shows the unclassified version of this landslide intervention index. The raw spatial pattern exhibits considerable similarities to the *M1* map. Highest probability scores were observed for medium inclined, concave-shaped and unforested terrain at lower slope positions that are close to infrastructure. The separate treatment of the variables distance to streets and distance to pathways and the inclusion of the variable distance to buildings led to a more pronounced accentuation of

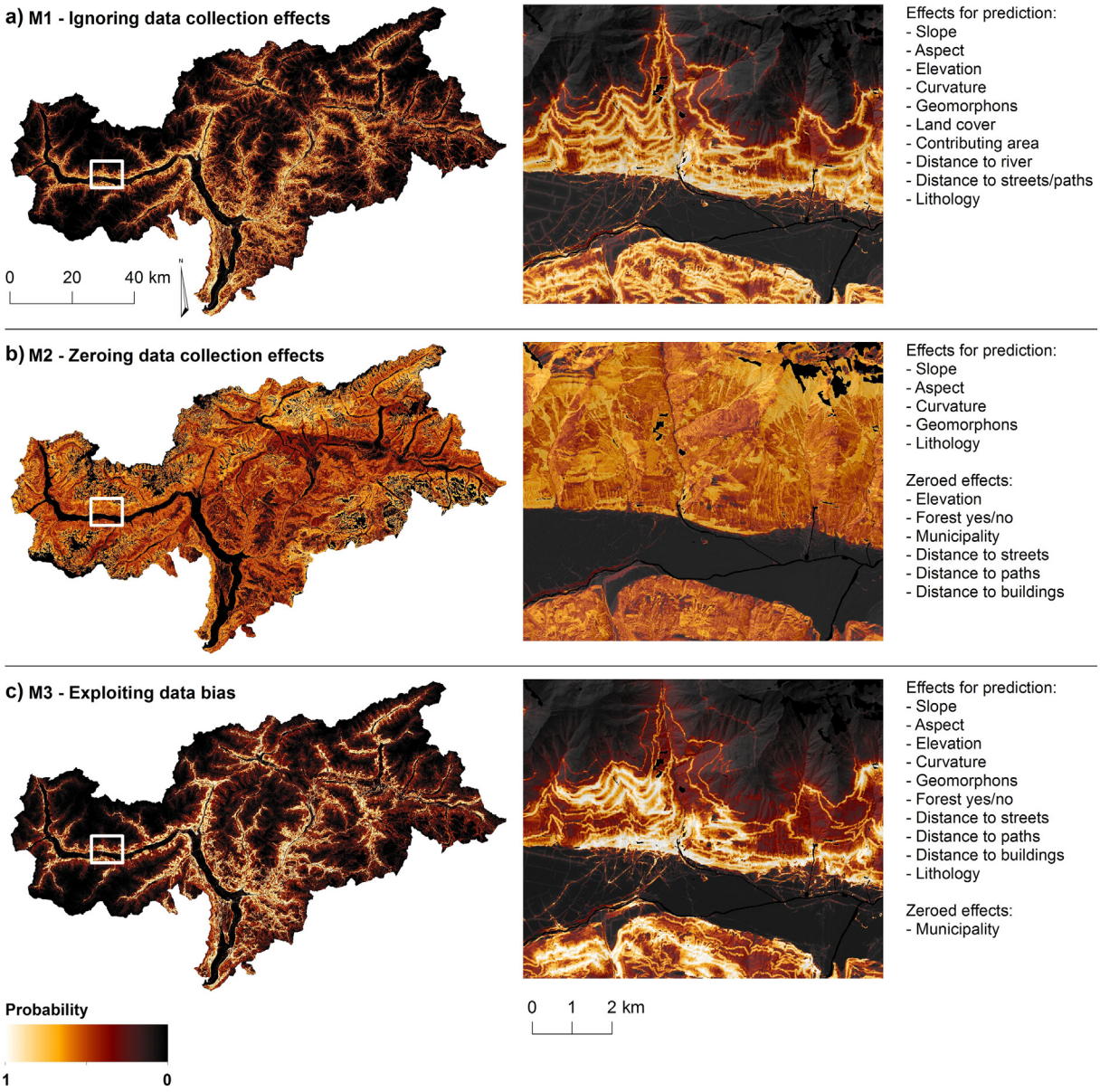



**Fig. 6.** Spatial pattern of the raw model predictions. Probability values close to one (zero) indicate that the local environmental conditions are similar (dissimilar) to the conditions observed at inventoried landslide locations. The M1 approach (a) ignored the presence of landslide data bias on purpose and was built upon commonly used variables. The M2 approach (b) aimed to account for landslide data limitations using mixed-effects modelling. The M3 map (c) represents the unclassified landslide intervention index whose classified version is presented in Fig. 8 and in the supplementary material.

locations close to streets and buildings (white areas in Fig. 6c). Lower probabilities were generally observed near less relevant infrastructure, such as forest roads or trails.

### 3.3. Model performance

K-fold CV and k-fold SCV described the ability of the models to correctly predict out-of-sample landslide presence and absence observations (Fig. 7a). The AUROCs indicated a similar and high predictive performance for the models *M1* and *M3* and a low performance for *M2*. In numbers, median AUROC scores of 0.87 (CV) and 0.86 (SCV) were observed for *M1*. *M2* was associated with a median AUROC score of 0.59 (CV) and 0.61 (SCV) while *M3* exhibited the highest performance with median AUROCs of 0.88 (CV) and 0.87 (SCV).

The comparably small SCV interquartile range for *M1* (0.026) and *M3* (0.028) provided quantitative evidence that the model setup (classifier-variable combination) produced spatially transferable and robust predictions (SCV box size in Fig. 7a). Instead, the high SCV interquartile range associated with *M2* (0.127) revealed the high variability of estimated AUROCs in response to iterative changes in the training and test regions. The discrimination ability of *M2* could be deemed acceptable ($0.7 \leq \text{AUROC} < 0.8$) for some of these data partitions, while for others it was estimated to be worse than random guess ($< 0.5$). This underscored that the *M2* model setup was incapable to establish spatially robust modelled relationships.

The prediction rate curves (Fig. 7b) confront the probability-ranked spatial predictions with the location of 60 temporally independent landslides that caused damage during a heavy weather event in November 2019. The curve shape for *M1* and *M3* showed that a large portion of these landslides occurred in areas with high predicted probability scores. The strong clustering of recently occurred damage causing landslides in areas of high probability was reflected by the high AUC values. *M3* was associated with the highest AUC (95.6%) closely followed by *M2* (95%). By dividing the prediction surface into two groups that relate to

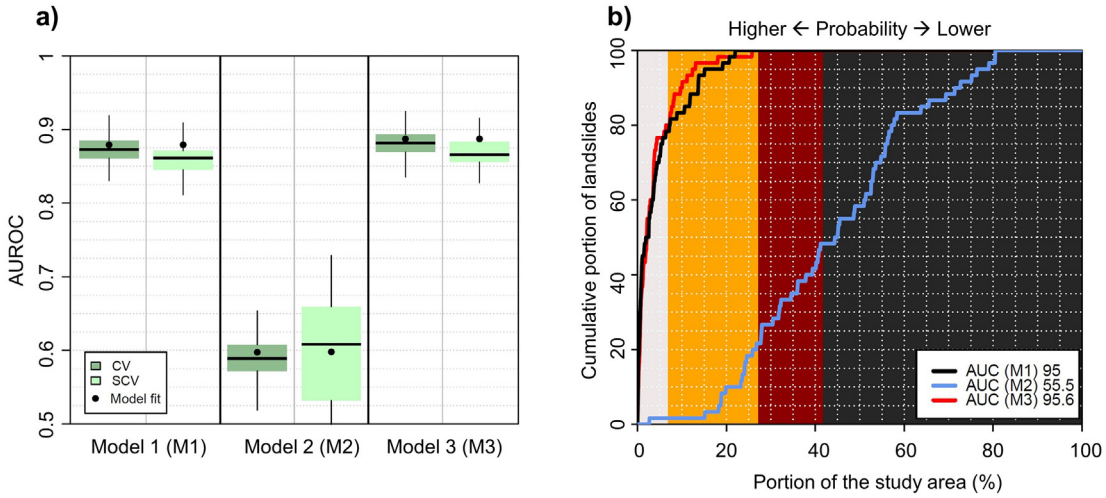



**Fig. 7.** Validation results for non-spatial (CV) and spatial (SCV) data partitions (a) and prediction rates based on 60 temporally independent damage-causing landslides from 2019 (b). Each boxplot in (a) depicts 500 AUROCs (0.5 random model, 1 perfect discrimination). The curves in (b) depict the cumulative portion of true positives (x-axis) against the portion of the total area at decreasing probability scores (y-axis) for each model. The background refers to the color-coded classes of the final landslide intervention index map (Fig. 8 and supplementary material).

higher (group 1, < 30% in Fig. 7b) and lower predicted probabilities (group 2, 30–100%), none of the 60 recently occurred landslides were initiated in the second group. Thus, all newly observed damage-causing landsides were predicted by the 30% high probability zone. 80% of landslide initiation points ($n = 48$) were located within the *M3* class 'high' (7.7% areal extent) while the remaining 20% ($n = 12$) were observed for the *M3* class 'medium' (22.2% areal extent).

In contrast, the *M2* AUC of 55.5% and the respective curve shape indicated no substantial clustering of recently initiated landslides within high probability zones. Only 34 of the 60 landslides (57%) were located within the high probability area when dividing the area into two equally sized halves that represent higher and lower predicted landslide probabilities (50% threshold in Fig. 7b). This indicated a near random prediction pattern for *M2*.

## 4. Discussion

### *4.1. M1: Why the well-performing model M1 did not totally reflect landslide susceptibility*

A landslide inventory affected by a systematic spatial incompleteness is likely to provoke erroneous landslide susceptibility models. However, literature suggests that such data biases are regularly disregarded when assessing landslide susceptibility using data-driven procedures (Steger et al., 2017). In the context of the *M1* modelling approach, the underlying landslide inventory bias was ignored on purpose to shed light on possible implications of disregarding common flaws in the data used to train and validate the models. AUROC interpretation guidelines (i.e. Hosmer et al., 2013) suggest that the model *M1* exhibited an excellent ability to discriminate between positive (i.e. landslide presence) and negative (i.e. absences) test set observation. However, a careful examination of the numerical results led to the conclusion that the model did not allow to create a plausible representation of landslide susceptibility. We argue that the subsequent spatial prediction pattern expressed, to a large extent, empirical relationships that can be attributed to the underlying landslide data collection procedure rather than to the process under investigation (i.e. landsliding). For example, the known underrepresentation of landslide data at high altitudes and far from infrastructure was directly reflected by a very low predicted probability in these areas. The results hence did not necessarily imply that shallow landsliding in South Tyrol occurs more frequently at lower altitudes or close to infrastructure. The findings rather confirmed that an inclusion environmental variables that are frequently used to model landslide susceptibility, such as altitude or the distance to roads, should not be justified by solely referring to obtained model performances or to a preconceived relation to landslide occurrence (Steger et al., 2017; Hearn and Hart, 2019). In fact, the consideration of the landslide data collection background provided valuable evidence that the model *M1* gave very limited insights into the interplay of environmental conditions that give rise to landsliding. Ultimately, this example pinpointed that relying solely on predictive performance can lead to misleading conclusions.

The propagation of landslide data bias into the *M1* results was caused by building the final prediction upon variables that systematically described the landslide data collection bias. The main culprits for the *M1* model distortions were the influential (Fig. 4a) and bias-describing variables 'elevation', 'distance to streets and paths' and, to a lower extent, also land cover. It can be expected that similar error-promoting mechanisms are common in the field of data-driven spatial landslide modelling, because frequently used environmental variables may often describe landslide data bias (see Table A.1 in supplementary material). For instance, landslide mapping completeness based on aerial or satellite imagery can be affected by shadow effects related to slope aspect (Schlögel et al., 2015), a variable frequently included in landslide susceptibility modelling. The commonly applied variable land cover may also relate to systematic landslide data heterogeneities (Brardinoni et al., 2003; Bell et al., 2012; Jacobs et al., 2016). The effectively surveyed area during field-based landslide mappings and the consistency of archive-based inventories (e.g. damage reports) may critically depend on the accessibility of the terrain describable by the common variables distance to roads and topographic indices (Bornaetxea et al., 2018; Steger et al., 2021).

The obtained high predictive performances for the *M1* model (median AUROCs 0.86 for CV and 0.87 for SCV, prediction rate 95%) casted further doubts on the frequent practice of justifying landslide susceptibility models as soon as the validation implies success (Steger et al., 2016b). From a geomorphic viewpoint, the *M1* model performance was considered overoptimistic, ultimately because it simply reflected a self-affirmation of landslide data bias. In particular, the underestimation of landslide occurrence far from infrastructure was of systematic nature and thus also inherent in each subsequent test data set, irrespective of the applied strategy to split training and test data. The high portion of true negatives was therefore traced back to the low landslide susceptibility scores far from infrastructure and an insufficient portion of landslide test observations in these zones. In analogy, the high fraction of true positives was codetermined by the very high predicted landslide susceptibility scores nearby linear infrastructure and a spatially

coinciding overrepresentation of landslide test data in these areas. The very high prediction rate of 95%, which was based on temporally independent landslide observations, was subject to the same kind of distortion, because only landslides that damaged infrastructure were registered during this heavy weather event. Based on these findings, we argue that a dogmatic optimization of model performance may even obstruct the view on the broader modelling context and push modelers to stumble into the McNamara fallacy during model construction. The fallacy consists in the supposition that useful decisions are exclusively based on data and associated metrics while non-quantifiable aspects are of little relevance (O'Mahony, 2017). From our viewpoint, the current trend of benchmarking algorithms from a pure quantitative perspective (Reichenbach et al., 2018) without considering the potential implications of the underlying data characteristics might foster misleading interpretations. We argue that in the light of a consistently increasing data availability and flexibility of modelling techniques, a critical evaluation of landslide data quality will remain crucial for the development of meaningful data-driven landslide susceptibility models.

In summary, systematic distortions of modelling and validation results may be no rarity in the field of data-driven spatial landslide modelling, because (i) heterogeneously complete landslides inventories are a rule rather than an exception (Ardizzone et al., 2002; Brardinoni et al., 2003; Trigila et al., 2010; Bell et al., 2012; Guzzetti et al., 2012; Van Den Eeckhaut et al., 2012; Petschko et al., 2014; Knevels et al., 2020), (ii) frequently used environmental variables can introduce such biases in a model (Table A.1 in supplementary material) and (iii) an interpretation of quantitative model performance metrics in the sense of 'high performance equals meaningful results' is common practice in the field (Steger et al., 2016b).

### 4.2. M2: Why bias-adjustment using mixed-effects modelling was of limited success

The *M2* approach aimed to assess landslide susceptibility by adjusting for variability that relates to the landslide data bias following Steger et al. (2017). The results indicate that bias-adjustment using mixed-effects modelling may reach its limits when the spatial distribution of inventoried landslides is primarily controlled by variables that do not relate to landslide susceptibility and have therefore to be averaged-out. In our experiments, most of the data variability was explained by variables related to the landslide data collection and not the supposed phenomena of interest (i.e. landslide occurrence).

The efficiency of bias-adjustment was further hampered by the circumstance that some variables concurrently represented the data collection procedure and factors that influence landslide susceptibility. This restricted the unambiguous assignment of the labels 'susceptibility effect' (to be used for prediction) or 'data collection effect' (to be zeroed) to each variable. For example, the variables 'elevation' and 'forest yes/no' were zeroed to avoid a direct propagation of the underlying landslide mapping bias (see the *M1* example). From a geomorphic perspective, however, elevation differences and variations in land cover were still expected to play a role for explaining the actual spatial distribution of landslide occurrence in South Tyrol. The zeroing of several influential variables, such as elevation, rendered the *M2* prediction particularly dependent on one remaining susceptibility effect, namely slope angle. The resulting *M2* map principally reflected that flat and very steep terrain is less susceptible to landsliding than the prevalent medium inclined hillslopes. It can be questioned whether decision makers can take advantage of such trivial conclusions.

The calculated low *M2* model performance (median AUROC of 0.59 for CV and 0.61 for SCV, prediction rate 55.5%) can principally be explained by the uniform spatial prediction pattern and by the actual unavailability of representative landslide test data for the area: the prevalent probability scores of around 0.5 in large parts of the hillslopes induced that inventoried test landslides were rarely observed for highly susceptible terrain while landslide-absence samples were seldomly observed for low susceptible slopes. The lack of landslide test data in high altitude terrain or far from infrastructure led to an increase in apparent false negatives and a decrease in the true positive rate. Thus, availability of a spatially representative landslide test data would likely improve the *M2* model performance metrics. In essence, this underlines that the explanatory power of model performance estimates critically depends on the degree of bias inherent in the landslide test data. An interpretation of predictive performance should therefore always consider the quality of the test data used to calculate the respective metric.

Ultimately, the *M2* findings also highlight that mixed-effects modelling is no panacea to tackle the problem of systematically incomplete landslide inventories in landslide susceptibility modelling. Although an adaptation of the sampling or modelling design is beneficial when dealing with common error-prone landslide information (Van Den Eeckhaut et al., 2012; Steger et al., 2017; Bornaetxea et al., 2018), we argue that a subsequent inference of cause-and-effect relationships has still to be conducted with great care.

### 4.3. M3: Changing the perspective to predict landslide interventions

The *M3* model design was inspired by the South Tyrolean natural hazard planning procedure, which focuses on natural hazards close to potentially exposed elements (e.g. buildings, roads, paths) while neglecting processes that are less significant in terms of risk (e.g. natural hazards at high altitudes). The key idea behind the *M3* approach was built upon the premise that the available landslide data can be considered systematic and representative for slope instabilities that caused damage or an intervention. This data characteristic was exploited to map the propensity of areas to be affected by damage-causing or infrastructure-threating landslides. This impact-focused perspective was facilitated by (i) the fact that the underlying data collection only focused on landslides that were relevant in terms of risk and by (ii) explicitly predicting both, the effects that describe spatial specificities of the data collection procedure (collection effects) and effects that relate to the process under investigation (susceptibility effects).

Changing the subject to be modelled, from landslide susceptibility to areas likely affected by damaging landslides, implied that most of the previously discussed landslide data bias issues (*M1*, *M2*) did not negatively affect the *M3* modelling approach. Instead, it allowed us to instrumentalize the otherwise detrimental ability of data-driven models to reproduce training data characteristics including its biases. The final *M3* index showed that the likelihood of a landslide intervention was highest for medium steep, concave-shaped landforms at lower slope positions nearby infrastructure (Fig. 8 and supplementary material).

Quantitative landslide risk assessments are challenging to implement over large areas, also due to a critical lack of input data on, e.g., the frequency and magnitude of past landslides (to characterize the hazard) or the physical vulnerability of exposed elements. Thus, the identification of priority areas for risk mitigation activities is frequently based on a simplified approach that overlays spatial data on landslide predisposition (i.e. susceptibility) with geo-spatial information on the exposed assets (e.g. buildings, streets, paths) (e.g., Pellicani et al., 2014). The approach used to derive the landslide intervention index (*M3*) has some similarities with a landslide exposure assessment since it integrates information on potential landslide areas with data on physical assets. The *M3* map allows identifying potential landslides occurrence locations that are likely to endanger infrastructure and necessitate mitigation measures (i.e. landslide interventions). Whereas in landslide exposure analysis the importance of asset type can be assigned using qualitative weights (Pellicani et al., 2014), in the presented landslide intervention index (*M3*) an empirical ranking is emerging from the statistics in the input data, for instance resulting in the stronger accentuation of areas close to roads, rather than to unpaved paths (e.g. wider brighter lines vs. narrower darker lines within Fig. 6c). This behavior ultimately reflects the higher relevance and prioritization

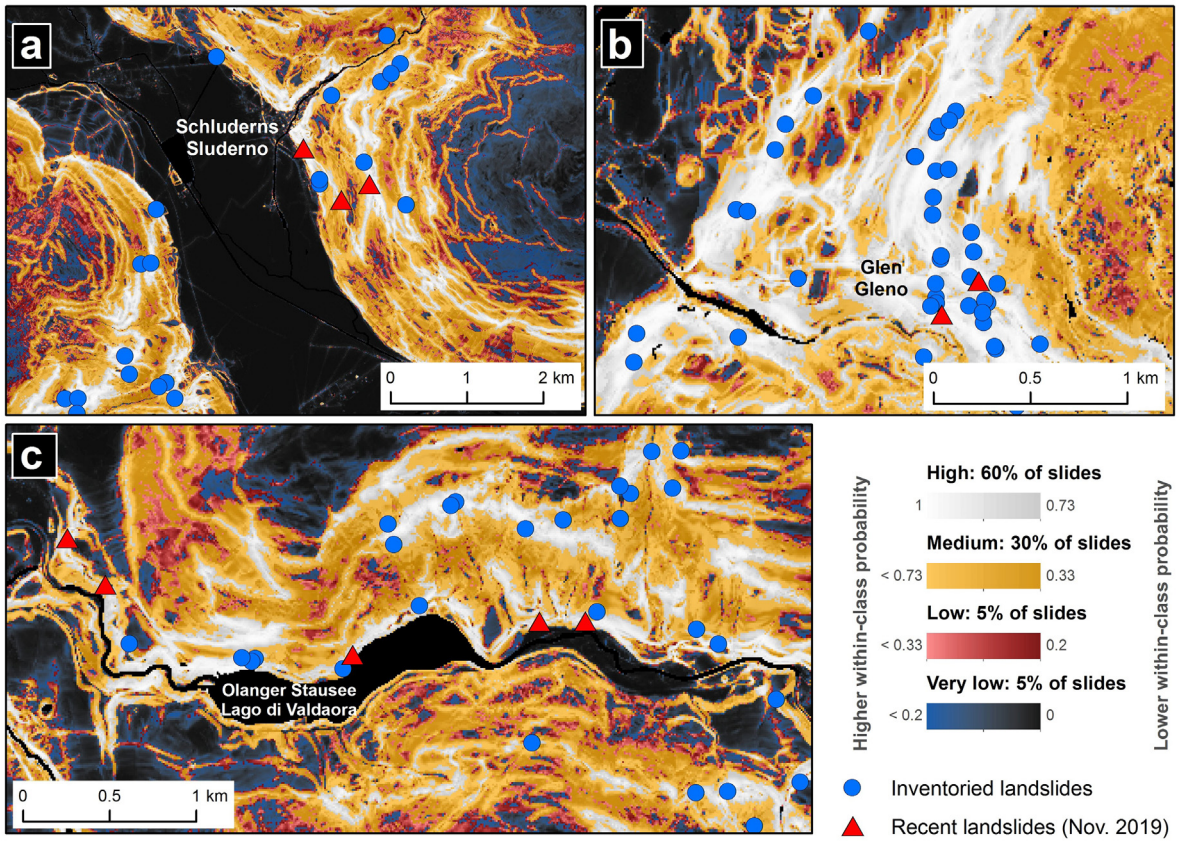



**Fig. 8.** The classified landslide intervention index (*M3*) for selected areas. Inventoried landslides (points) and temporally independent damage-causing landslides of the November 2019 event (triangles) are superimposed. The map for the entire area can be viewed in the supplementary material (incl. The locations of the shown examples: a, b and c).

of such infrastructure in terms of landslide-induced interventions. The final landslide intervention index will further be employed by the Office for Geology and Building Materials Testing as a tool for allocation of resources (e.g. informing first responders in case of forecasted storms) and for the monitoring and planning of ordinary and extraordinary maintenance.

In contrast to the *M1* and *M2* validation results, the informative value of the calculated *M3* AUROCs was considered high since the landslide test data spatially represented the population of interest in a systematic way (i.e. landslides that caused an intervention). Thus, the high performance scores (median AUROC of 0.88 for CV and 0.87 for SCV, prediction rate 95.6%) reflected that the spatial predictions repeatedly coincided with independent information on damaging landslides (Fig. 8). From this viewpoint, we argue that the explanatory power of model performance metrics is high in case the predictions are evaluated against a reasonable representation of the objects under investigation. Instead, modelling and validating with a landslide data that is spatially biased with respect to the object of interest is likely to foster wrong conclusions (see the discussion on *M1*).

The spatial prediction of almost all variables (except 'municipality ID', see Section 2.3.3) allowed to circumvent the previously described dilemma of having to assign an exclusive label to each variable (i.e. susceptibility effect to be predicted or data collection effect to be zeroed). For the envisaged *M3* spatial prediction task, it was irrelevant to know the degree to which a modelled effect referred to landslide predisposition or to the underlying landslide data collection since both of these effects were considered for the prediction. For instance, ascertaining whether the negative relationship between landslide occurrence and elevation was mainly caused by a more thorough registration of landslides at lower altitudes or rather by elevation-dependent natural influences was considered unimportant when deriving the landslide intervention index. The chance that a landslide related intervention takes place in non-forested terrain was estimated to be 1.9 times higher compared to forests (Table A.2 in supplementary material) while accounting for numerous confounders. Even though the stabilizing effect of forest on shallow landslides is described in literature, it remained unclear to which extent the model reflected the stabilizing effects of trees or simply a lower priority of forested terrain in terms of interventions associated with landsliding.

The substantial similarity between the biased *M1* susceptibility map and the intervention index map *M3* was related to the fact that for both models the predicted variables represented a mixture of landslide susceptibility and landslide data collection effects. This was detrimental to develop a landslide susceptibility model (*M1*), but beneficial to create the intervention index (*M3*). The comparable and high validation results (*M1* and *M3*) underlined that models must be evaluated not only from a statistical viewpoint, but also against common sense and domain knowledge. Even though this sounds trivial, it is remarkable how often the data and modelling context is ignored in the field (Hearn and Hart, 2019). Finally, the results further underscored challenges in understanding key landslide controls using statistical procedures (Vorpahl et al., 2012), particularly under data bias conditions.

## 5. Conclusion

This study showed, on the basis of three conceptually different models (*M1*–*M3*), that accounting for landslide data bias is paramount in data-driven spatial landslide modelling. It was demonstrated that the spatial distribution of landslide occurrence can not only be explained by landslide predisposing factors (i.e. landslide susceptibility effects), but also by commonly used variables that actually described and reproduced the underlying bias in landslide observations and recording practice (i.e. data collection effects). Ascertaining whether environmental variables and subsequent models reasonably reflected landslide

predisposition or rather a data collection bias is not always trivial, particularly because data bias can increase the apparent model performance.

By intentionally ignoring obvious flaws in the available landslide information, we could illustrate how failure to account for widespread spatial landslide data bias can lead to (i) distorted estimates on where landslides can be expected, (ii) wrong inference regarding the underlying predisposing environmental factors and (iii) overoptimistic model performance estimates (*M1*). In this study, bias-adjustment using novel mixed-effects modelling (*M2*) was ineffective, mainly because most of the influential environmental variables were related to the landslide data bias, and not to the phenomena of interest (i.e. landslide susceptibility). The third approach (*M3*) focused on the simultaneous prediction of landslide susceptibility effects and effects that described spatial particularities associated with the systematic registration of damage causing landslides. The derived landslide intervention index model allowed identifying areas where damage-causing and infrastructure-threatening landslides are likely to occur. The efficiency of this impact-oriented assessment was confirmed using temporally independent data on damage-causing landslides.

We argue that, despite the application of increasingly flexible models and purely number-oriented procedures in the field of data-driven landslide modelling, scrutinizing landslide data background and model plausibility remains a prerequisite towards meaningful results. Understanding the non-trivial balance between data quality, model flexibility and associated interpretation possibilities will remain a challenge in the era of artificial intelligence. We conclude that under serious data bias conditions, researchers may have to accept that *"the data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data"* (Tukey, 1986, p.74–75).

## CRediT authorship contribution statement

**Stefan Steger**: Conceptualization, Methodology, Coding, Validation, Data curation, Writing - Original Draft, Visualization. **Volkmar Mair**: Conceptualization, Validation, Data curation, Writing - Review & Editing. **Christian Kofler**: Data curation, Coding, Visualization, Writing - Review & Editing. **Massimiliano Pittore**: Conceptualization, Validation, Writing - Review & Editing. **Marc Zebisch**: Conceptualization, Validation, Writing - Review & Editing. **Stefan Schneiderbauer**: Conceptualization, Writing - Review & Editing, Funding acquisition.

## Declaration of 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.

## Acknowledgments

The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs. This research was conducted in the context of the GreenRisk4Alps project (ASP635). GreenRisk4Alps has been financed by Interreg Alpine Space, one of the 15 transnational cooperation programs covering the whole of the European Union (EU) in the framework of European Regional policy. The authors thank the Autonomous Province of Bolzano/Bozen for providing access to environmental input data and the three anonymous reviewers and the editor for their valuable comments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.145935.

## References

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., Reichenbach, P., 2002. Impact of mapping errors on the reliability of landslide hazard maps. Natural Hazards and Earth System Science 2, 3–14.

Arnone, E., Francipane, A., Scarbaci, A., Puglisi, C., Noto, L.V., 2016. Effect of raster resolution and polygon-conversion algorithm on landslide susceptibility mapping. Environ. Model. Softw. 84, 467–481. https://doi.org/10.1016/j.envsoft.2016.07.016.

ASTAT, 2019. Provincial statistics Institute of the Autonomous Province of South Tyrol. Link: https://astat.provinz.bz.it/de/bevoelkerung.asp (last access: 20 July 2020).

Atkinson, P., Jiskoot, H., Massari, R., Murray, T., 1998. Generalized linear modelling in geomorphology. Earth Surf. Process. Landf. 23, 1185–1195.

Autonomous Province of South Tyrol, 2018. South Tyrol in figures. Provincial Statistics Institute. Link: https://astat.provinz.bz.it/downloads/Siz_2019-eng.pdf (last access: 23 July 2020).

Bell, R., Petschko, H., Röhrs, M., Dix, A., 2012. Assessment of landslide age, landslide persistence and human impact using airborne laser scanning digital terrain models. Geografiska Annaler: Series A, Physical Geography 94, 135–156. https://doi.org/10.1111/j.1468-0459.2012.00454.x.

Blahut, J., van Westen, C.J., Sterlacchini, S., 2010. Analysis of landslide inventories for accurate prediction of debris-flow source areas. Geomorphology 119, 36–51. https://doi.org/10.1016/j.geomorph.2010.02.017.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24, 127–135. https://doi.org/10.1016/j.tree.2008.10.008.

Bordoni, M., Galanti, Y., Bartelletti, C., Persichillo, M.G., Barsanti, M., Giannecchini, R., Avanzi, G.D., Cevasco, A., Brandolini, P., Galve, J.P., Meisina, C., 2020. The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models. CATENA 193, 104630. https://doi.org/10.1016/j.catena.2020.104630.

Borgatti, L., Soldati, M., 2010. Landslides as a geomorphological proxy for climate change: a record from the Dolomites (northern Italy). Geomorphology, Landslide geomorphology in a changing environment 120, 56–64. https://doi.org/10.1016/j.geomorph.2009.09.015.

Bornaetxea, T., Rossi, M.C.C., Marchesini, I., Alvioli, M., 2018. Effective surveyed area and its role in statistical landslide susceptibility assessments. Nat. Hazards Earth Syst. Sci. 18, 2455–2469. https://doi.org/10.5194/nhess-18-2455-2018.

Brardinoni, F., Slaymaker, O., Hassan, M.A., 2003. Landslide inventory in a rugged forested watershed: a comparison between air-photo and field survey data. Geomorphology 54, 179–196. https://doi.org/10.1016/S0169-555X(02)00355-0.

Brenning, A., 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest, in: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International. pp. 5372–5375.

Brenning, A., Schwinn, M., Ruiz-Páez, A.P., Muenchow, J., 2015. Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. Natural Hazards and Earth System Science 15, 45–57. https://doi.org/10.5194/nhess-15-45-2015.

Brock, J., Schratz, P., Petschko, H., Muenchow, J., Micu, M., Brenning, A., 2020. The performance of landslide susceptibility models critically depends on the quality of digital elevations models. Geomatics, Natural Hazards and Risk 11, 1075–1092. https://doi.org/10.1080/19475705.2020.1776403.

Budimir, M.E.A., Atkinson, P.M., Lewis, H.G., 2015. A systematic review of landslide probability mapping using logistic regression. Landslides 12, 419–436. https://doi.org/10.1007/s10346-014-0550-5.

Bui, D.T., Lofman, O., Revhaug, I., Dick, O., 2011. Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat. Hazards 59, 1413–1444. https://doi.org/10.1007/s11069-011-9844-2.

Casagli, N., Cigna, F., Bianchini, S., Hölbling, D., Füreder, P., Righini, G., Del Conte, S., Friedl, B., Schneiderbauer, S., Iasio, C., Vlcko, J., Greif, V., Proske, H., Granica, K., Falco, S., Lozzi, S., Mora, O., Arnaud, A., Novali, F., Bianchi, M., 2016. Landslide mapping and monitoring by using radar and optical remote sensing: examples from the EC-FP7 project SAFER. Remote Sensing Applications: Society and Environment 4, 92–108. https://doi.org/10.1016/j.rsase.2016.07.001.

Cascini, L., 2008. Applicability of landslide susceptibility and hazard zoning at different scales. Eng. Geol. 102, 164–177. https://doi.org/10.1016/j.enggeo.2008.03.016.

Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. Natural Hazards and Earth System Science 13, 2815–2831. https://doi.org/10.5194/nhess-13-2815-2013.

Chung, C.J.F., Fabbri, A.G., 2003. Validation of spatial prediction models for landslide hazard mapping. Nat. Hazards 30, 451–472.

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N.A., Lombardo, L., Agnesi, V., 2016. Exploring the effect of absence selection on landslide susceptibility models: a case study in Sicily, Italy. Geomorphology 261, 222–235. https://doi.org/10.1016/j.geomorph.2016.03.006.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Corsini, A., Pasuto, A., Soldati, M., Zannoni, A., 2005. Field monitoring of the Corvara landslide (Dolomites, Italy) and its relevance for hazard assessment. Geomorphology, Geomorphological hazard and human impact in mountain environmentsRelations between Man and the Mountain Environment in terms of Geomorphological Hazards and Human Impact in Europe 66, 149–165. https://doi.org/10.1016/j.geomorph.2004.09.012.

Cruden, D.M., Varnes, D.J., 1996. Landslide Types and Processes, TRB Special Report. National Academy Press, Washington.

Darvishi, M., Schlögel, R., Bruzzone, L., Cuozzo, G., 2018. Integration of PSI, MAI, and intensity-based sub-pixel offset tracking results for landslide monitoring with X-band corner reflectors—Italian Alps (Corvara). Remote Sens. 10, 409. https://doi.org/10.3390/rs10030409.

Depicker, A., Jacobs, L., Delvaux, D., Havenith, H.-B., Maki Mateso, J.-C., Govers, G., Dewitte, O., 2020. The added value of a regional landslide susceptibility assessment: the western branch of the east African rift. Geomorphology 353, 106886. https://doi.org/10.1016/j.geomorph.2019.106886.

Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10, e0133262.

Frattini, P., Crosta, G., Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111, 62–72. https://doi.org/10.1016/j.enggeo.2009.12.004.

Fressard, M., Thiery, Y., Maquaire, O., 2014. Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the pays d'Auge plateau hillslopes (Normandy, France). Natural Hazards and Earth System Science 14, 569–588. https://doi.org/10.5194/nhess-14-569-2014.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008. Comparing landslide inventory maps. Geomorphology 94, 268–289. https://doi.org/10.1016/j.geomorph.2006.09.023.

Gefahrenzonenplan Südtirol, 2021. Official documents related to the provincial natural hazard zone planning (in German and Italian). Link: http://www.provinz.bz.it/natur-umwelt/natur-raum/planung/gefahrenzonenplan.asp (last access: 8 January 2021).

Geokatalog, 2019. Open Geodatabase of the Autonomous Province of South Tyrol. Link: https://geoportal.buergernetz.bz.it/geodaten.asp (last access: 5 October 2020).

Gerhard, T., 2008. Bias: considerations for research practice. Am. J. Health Syst. Pharm. 65, 2159–2168.

Glade, T., Anderson, M., Crozier, M.J. (Eds.), 2005. Landslide Hazard and Risk: Issues, Concepts and Approach. John Wiley, Chichester.

Goetz, J., Guthrie, R.H., Brenning, A., 2015a. Forest harvesting is associated with increased landslide activity during an extreme rainstorm on Vancouver Island, Canada. Natural Hazards and Earth System Science 15, 1311–1330. https://doi.org/10.5194/nhess-15-1311-2015.

Goetz, J., Brenning, A., Petschko, H., Leopold, P., 2015b. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11. https://doi.org/10.1016/j.cageo.2015.04.007.

Goetz, J., Brenning, A., Marcer, M., Bodin, X., 2018. Modeling the precision of structure-from-motion multi-view stereo digital elevation models from repeated close-range aerial surveys. Remote Sens. Environ. 210, 208–216. https://doi.org/10.1016/j.rse.2018.03.013.

GRASS Development Team, 2019. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. Link: http://grass.osgeo.org (last access: 12 July 2020).

Guillard, C., Zêzere, J., 2012. Landslide susceptibility assessment and validation in the framework of municipal planning in Portugal: the case of Loures municipality. Environ. Manag. 50, 721–735. https://doi.org/10.1007/s00267-012-9921-7.

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, 181–216.

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.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Galli, M., 2006. Estimating the quality of landslide susceptibility models. Geomorphology 81, 166–184.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., Chang, K.-T., 2012. Landslide inventory maps: new tools for an old problem. Earth Sci. Rev. 112, 42–66. https://doi.org/10.1016/j.earscirev.2012.02.001.

Hearn, G.J., Hart, A.B., 2019. Landslide susceptibility mapping: a practitioner's view. Bull. Eng. Geol. Environ. 78, 5811–5826. https://doi.org/10.1007/s10064-019-01506-1.

Holland, P.W., 1986. Statistics and causal inference. J. Am. Stat. Assoc. 81, 945. https://doi.org/10.2307/2289064.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression, Third edition. ed, Wiley series in probability and statistics. Wiley, Hoboken, New Jersey.

Hungr, O., Leroueil, S., Picarelli, L., 2013. The Varnes classification of landslide types, an update. Landslides 11, 167–194. https://doi.org/10.1007/s10346-013-0436-y.

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2016. Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology 253, 508–523. https://doi.org/10.1016/j.geomorph.2015.10.030.

Jacobs, L., Dewitte, O., Poesen, J., Maes, J., Mertens, K., Sekajugo, J., Kervyn, M., 2016. Landslide characteristics and spatial distribution in the Rwenzori Mountains, Uganda. J. Afr. Earth Sci. 134, 917–930. https://doi.org/10.1016/j.jafrearsci.2016.05.013.

Jacobs, L., Kervyn, M., Reichenbach, P., Rossi, M., Marchesini, I., Alvioli, M., Dewitte, O., 2020. Regional susceptibility assessments with heterogeneous landslide information: Slope unit- vs. pixel-based approach. Geomorphology 356, 107084. doi:10.1016/j.geomorph.2020.107084.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156. https://doi.org/10.1016/j.geomorph.2012.11.005.

Knevels, R., Petschko, H., Proske, H., Leopold, P., Maraun, D., Brenning, A., 2020. Event-based landslide modeling in the Styrian Basin, Austria: accounting for time-varying rainfall and land cover. Geosciences 10, 217. https://doi.org/10.3390/geosciences10060217.

Krøgli, I.K., Devoli, G., Colleuille, H., Boje, S., Sund, M., Engen, I.K., 2018. The Norwegian forecasting and warning service for rainfall- and snowmelt-induced landslides. Nat. Hazards Earth Syst. Sci. 18, 1427–1450. https://doi.org/10.5194/nhess-18-1427-2018.

Lewińska, K.E., Ivits, E., Schardt, M., Zebisch, M., 2018. Drought impact on phenology and green biomass production of Alpine Mountain Forest—case study of South Tyrol 2001–2012 inspected with MODIS time series. Forests 9, 91. https://doi.org/10.3390/f9020091.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., Huser, R., 2020. Space-time landslide predictive modelling. Earth Sci. Rev. 209, 103318. https://doi.org/10.1016/j.earscirev.2020.103318.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004. Landslide inventories and their statistical properties. Earth Surf. Process. Landf. 29, 687–711. https://doi.org/10.1002/esp.1064.

Marc, O., Stumpf, A., Malet, J.-P., Gosset, M., Uchida, T., Chiang, S.-H., 2018. Initial insights from a global database of rainfall-induced landslide inventories: the weak influence of slope and strong influence of total storm rainfall. Earth Surf. Dynam. 6, 903–922. https://doi.org/10.5194/esurf-6-903-2018.

Marra, F., Nikolopoulos, E.I., Creutin, J.D., Borga, M., 2016. Space–time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship. Journal of Hydrology, Flash floods, hydro-geomorphic response and risk management 541, 246–255. https://doi.org/10.1016/j.jhydrol.2015.10.010.

Mergili, M., Kerschner, H., 2015. Gridded precipitation mapping in mountainous terrain combining GRASS and R. Norsk Geografisk Tidsskrift - Norwegian Journal of Geography 69, 2–17. https://doi.org/10.1080/00291951.2014.992807.

Molnar, C., 2020. Interpretable Machine Learning. Lulu.com. ISBN: 9780244768522.

Monsieurs, E., Dewitte, O., Demoulin, A., 2019. A susceptibility-based rainfall threshold approach for landslide occurrence. Nat. Hazards Earth Syst. Sci. 19, 775–789. https://doi.org/10.5194/nhess-19-775-2019.

O'Mahony, S., 2017. Medicine and the McNamara fallacy. J R Coll Physicians Edinb 47, 281–287. https://doi.org/10.4997/jrcpe.2017.315.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646.

Pannucci, C.J., Wilkins, E.G., 2010. Identifying and avoiding bias in research. Plast. Reconstr. Surg. 126, 619–625. https://doi.org/10.1097/PRS.0b013e3181de24bc.

Pellicani, R., Van Westen, C.J., Spilotro, G., 2014. Assessing landslide exposure in areas with limited landslide information. Landslides 11, 463–480. https://doi.org/10.1007/s10346-013-0386-4.

Pereira, S., Garcia, R.A.C., Zêzere, J.L., Oliveira, S.C., Silva, M., 2016. Landslide quantitative risk analysis of buildings at the municipal scale based on a rainfall triggering scenario. Geomatics, Natural Hazards and Risk 0, 1–25. doi:https://doi.org/10.1080/19475705.2016.1250116.

Persichillo, M.G., Bordoni, M., Meisina, C., 2017. The role of land use changes in the distribution of shallow landslides. Sci. Total Environ. 574, 924–937. https://doi.org/10.1016/j.scitotenv.2016.09.125.

Petschko, H., Brenning, A., Bell, R., Goetz, J., Glade, T., 2014. Assessing the quality of landslide susceptibility maps – case study Lower Austria. Natural Hazards and Earth System Science 14, 95–118. https://doi.org/10.5194/nhess-14-95-2014.

Piacentini, D., Troiani, F., Soldati, M., Notarnicola, C., Savelli, D., Schneiderbauer, S., Strada, C., 2012. Statistical analysis for assessing shallow-landslide susceptibility in South Tyrol (south-eastern Alps, Italy). Geomorphology 151–152, 196–206. https://doi.org/10.1016/j.geomorph.2012.02.003.

Pisano, L., Zumpano, V., Malek, ž., Rosskopf, C.M., Parise, M., 2017. Variations in the susceptibility to landslides, as a consequence of land cover changes: a look to the past, and another towards the future. Sci. Total Environ. 601–602, 1147–1159. https://doi.org/10.1016/j.scitotenv.2017.05.231.

Regmi, N.R., Giardino, J.R., McDonald, E.V., Vitek, J.D., 2014. A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. Landslides 11, 247–262. https://doi.org/10.1007/s10346-012-0380-2.

Reichenbach, P., Rossi, M., Malamud, B.D., 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.

Remondo, J., Bonachea, J., Cendrero, A., 2005. A statistical approach to landslide risk modelling at basin scale: from landslide susceptibility to quantitative risk assessment. Landslides 2, 321–328. https://doi.org/10.1007/s10346-005-0016-x.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., Peruccacci, S., 2010. Optimal landslide susceptibility zonation based on multiple forecasts. Geomorphology 114, 129–142.

Rotigliano, E., Agnesi, V., Cappadonia, C., Conoscenti, C., 2011. The role of the diagnostic areas in the assessment of landslide susceptibility models: a test in the sicilian chain. Nat. Hazards 58, 981–999. https://doi.org/10.1007/s11069-010-9708-1.

Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., Rossi, M., 2017. Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. Landslides 14, 547–558. https://doi.org/10.1007/s10346-016-0739-x.

Scheidl, C., Rickenmann, D., 2010. Empirical prediction of debris-flow mobility and deposition on fans. Earth Surf. Process. Landf. 35, 157–173. https://doi.org/10.1002/esp.1897.

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.

Schlögel, R., Malet, J.-P., Reichenbach, P., Remaître, A., Doubre, C., 2015. Analysis of a landslide multi-date inventory in a complex mountain landscape: the Ubaye valley case study. Natural Hazards and Earth System Science 15, 2369–2389. https://doi.org/10.5194/nhess-15-2369-2015.

Schlögel, R., Kofler, C., Gariano, S.L., Van Campenhout, J., Plummer, S., 2020. Changes in climate patterns and their association to natural hazard distribution in South Tyrol (Eastern Italian Alps). Sci. Rep. 10, 5022. https://doi.org/10.1038/s41598-020-61615-w.
Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2018. Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. arXiv:1803.11266 [cs, stat].
Singh, R., Mangat, N.S., 1996. Elements of Survey Sampling. Springer Science & Business Media. ISBN 978-94-017-1404-4.
Steger, S., Glade, T., 2017. The challenge of "trivial areas" in statistical landslide susceptibility modelling. In: Mikos, M., Tiwari, B., Yin, Y., Sassa, K. (Eds.), Advancing Culture of Living with Landslides. Springer International Publishing, Cham, pp. 803–808.
Steger, S., Brenning, A., Bell, R., Glade, T., 2016a. The propagation of inventory-based positional errors into statistical landslide susceptibility models. Nat. Hazards Earth Syst. Sci. 16, 2729–2745. https://doi.org/10.5194/nhess-16-2729-2016.
Steger, S., Brenning, A., Bell, R., Petschko, H., Glade, T., 2016b. Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. Geomorphology 262, 8–23. https://doi.org/10.1016/j.geomorph.2016.03.015.
Steger, S., Brenning, A., Bell, R., Glade, T., 2017. The influence of systematically incomplete shallow landslide inventories on statistical susceptibility models and suggestions for improvements. Landslides 14, 1767–1781. https://doi.org/10.1007/s10346-017-0820-0.
Steger, S., Mair, V., Kofler, C., Pittore, M., Zebisch, M., Schneiderbauer, S., 2021. A statistical exploratory analysis of inventoried slide-type movements for South Tyrol (Italy), in: Guzzetti, F., Mihalić Arbanas, S., Reichenbach, P., Sassa, K., Bobrowsky, P.T., Takara, K. (Eds.), Understanding and Reducing Landslide Disaster Risk, Volume 2 From Mapping to Hazard and Risk Zonation. Springer Nature Switzerland AG 2021 (ISBN 978–3–030-60226-0) (in print).
Stingl, V., Mair, V., 2005. Einführung in die Geologie Südtirols. Autonome Provinz Bozen-Südtirol, Amt f. Geologie u. Baustoffprüfung. ISBN 9788870734089.
Strozzi, T., Farina, P., Corsini, A., Ambrosi, C., Thüring, M., Zilger, J., Wiesmann, A., Wegmüller, U., Werner, C., 2005. Survey and monitoring of landslide displacements by means of L-band satellite SAR interferometry. Landslides 2, 193–201. https://doi.org/10.1007/s10346-005-0003-2.
Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.
Tasser, E., Mader, M., Tappeiner, U., 2003. Effects of land use in alpine grasslands on the probability of landslides. Basic and Applied Ecology 4, 271–280. https://doi.org/10.1078/1439-1791-00153.
Trigila, A., Iadanza, C., Spizzichino, D., 2010. Quality assessment of the Italian landslide inventory using GIS processing. Landslides 7, 455–470. https://doi.org/10.1007/s10346-010-0213-0.
Trigila, A., Frattini, P., Casagli, N., Catani, F., Crosta, G., Esposito, C., Iadanza, C., Lagomarsino, D., Mugnozza, G.S., Segoni, S., Spizzichino, D., Tofani, V., Lari, S., 2013. Landslide susceptibility mapping at national scale: The Italian case study. In: Margottini, C., Canuti, P., Sassa, K. (Eds.), Landslide Science and Practice. Landslide Inventory and Susceptibility and Hazard Zoning vol. 1. Springer, Berlin, Heidelberg, pp. 287–295. https://doi.org/10.1007/978-3-642-31325-7_38.
Tukey, J.W., 1986. Sunset Salvo. Am. Stat. 40, 72–76. doi:10.2307/2683137.
Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006. Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76, 392–410. https://doi.org/10.1016/j.geomorph.2005.12.003.
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. https://doi.org/10.1007/s10346-011-0299-z.
Van Westen, C.J., Castellanos, E., Kuriakose, S.L., 2008. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. Eng. Geol. 102, 112–131. https://doi.org/10.1016/j.enggeo.2008.03.010.
Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B., 2012. How can statistical models help to determine driving factors of landslides? Ecol. Model. 239, 27–39. https://doi.org/10.1016/j.ecolmodel.2011.12.007.
Wood, S.N., 2017. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.
Wu, X., Chen, X., Zhan, F.B., Hong, S., 2015. Global research trends in landslides during 1991–2014: a bibliometric analysis. Landslides 12, 1215–1226. https://doi.org/10.1007/s10346-015-0624-z.
Zebisch, M., Vaccaro, R., Niedrist, G., Schneiderbauer, S., Streifeneder, T., Weiß, M., Troi, A., Renner, K., Pedoth, L., Baumgartner, B., 2018. Klimareport Südtirol 2018. Eurac Research, Bozen/Bolzano, Italy.
Zêzere, J.L., Pereira, S., Melo, R., Oliveira, S.C., Garcia, R.A.C., 2017. Mapping landslide susceptibility using data-driven methods. Sci. Total Environ. 589, 250–267. https://doi.org/10.1016/j.scitotenv.2017.02.188.
Zieher, T., Perzl, F., Rössel, M., Rutzinger, M., Meißl, G., Markart, G., Geitner, C., 2016. A multi-annual landslide inventory for the assessment of shallow landslide susceptibility–two test cases in Vorarlberg, Austria. Geomorphology 259, 40–54.
Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R, 2009th ed. Springer Science and Business Media, New York, NY. https://doi.org/10.1007/978-0-387-87458-6.