



2020 Kyoto Japan

A Statistical Exploratory Analysis of Inventoried Slide-Type Movements for South Tyrol (Italy)

Stefan Steger, Volkmar Mair, Christian Kofler, Massimiliano Pittore, Marc Zebisch, and Stefan Schneiderbauer

Abstract

Landslides of the slide-type movement represent common damaging phenomena in the Italian province of South Tyrol. Up to January 2019, the landslide inventory of the province lists 1928 accurately mapped landslides that required intervention by e.g. the local road service or the provincial geological survey. Thus, this landslide data set mainly includes events that caused damage. The aim of this contribution was to investigate and critically interpret statistical associations between the inventoried slide-type movements and a variety of spatial environmental variables. The assessment of conditional frequencies and the discriminatory power of single variables revealed conditions that are typically present at landslide mapping

locations, e.g. topography, land cover, rock types, and proximity to infrastructure. A critical interpretation of the statistical results highlighted the need to consider the landslide data origin (i.e. background information) in order to avoid misleading statements and wrong inferences. The findings of the here presented work show that the availability of detailed landslide information does not always ensure that valid process-related conclusions can be drawn from subsequent statistical analyses (e.g. identification of important landslide controls). Despite considerable methodical advancements in the field of statistical data analysis and machine learning, we conclude that the principle ‘*correlation does not necessarily imply (geomorphic) causation*’ remains of particular relevance when exploiting available landslide information.

S. Steger (✉) · C. Kofler · M. Pittore · M. Zebisch · S. Schneiderbauer
Institute for Earth Observation, Eurac Research, Viale Druso 1,
39100 Bolzano, Italy
e-mail: stefan.steger@eurac.edu

C. Kofler
e-mail: christian.kofler@eurac.edu; christian.kofler1@natec.unibz.it

M. Pittore
e-mail: massimiliano.pittore@eurac.edu

M. Zebisch
e-mail: marc.zebisch@eurac.edu

S. Schneiderbauer
e-mail: stefan.schneiderbauer@eurac.edu; schneiderbauer@ehs.unu.edu

V. Mair
Office for Geology and Building Materials Testing, Autonomous
Province of Bolzano-South Tyrol, 39053 Cardano, Italy
e-mail: volkmar.mair@provinz.bz.it

C. Kofler
Faculty of Science and Technology, Free University of
Bozen-Bolzano, 39100 Bozen-Bolzano, Italy

S. Schneiderbauer
GLOMOS Program, United Nations University, Institute for
Environment and Human Security, Bonn, Germany

Keywords

Landslide inventory · South Tyrol · Exploratory data analysis · IFFI · Sampling bias · Susceptibility

Introduction

The Italian province of South Tyrol is an Alpine area frequently affected by natural hazard phenomena, such as floods, snow avalanches and different landslide types. Landslides of the slide-type movement (cf. Cruden and Varnes 1996; Hungr et al. 2014) cause damage on infrastructure, agricultural fields and private properties every year. The most extensive landslide inventory for South Tyrol is managed by the provincial geological survey and contains 1928 positionally accurate points that relate to the scarp location of slide-type movements.

This contribution aims to investigate statistical associations between inventoried shallow slide-type movements and frequently applied spatial environmental variables. Particular

emphasis is placed on a critical interpretation of the results by taking information on the underlying landslide data collection explicitly into account. Potential implications for the statistical analysis of already available inventory data are discussed.

Study Area and Data

South Tyrol

South Tyrol extends over 7400 km² and is situated at the northernmost part of Italy (Fig. 1). The area is characterized by a substantial morphological, geological and climatic heterogeneity (e.g. annual precipitation from <500 mm up to >1000, altitude from 200 to 3900 m a.s.l.) (Stingl and Mair 2005; Piacentini et al. 2012).

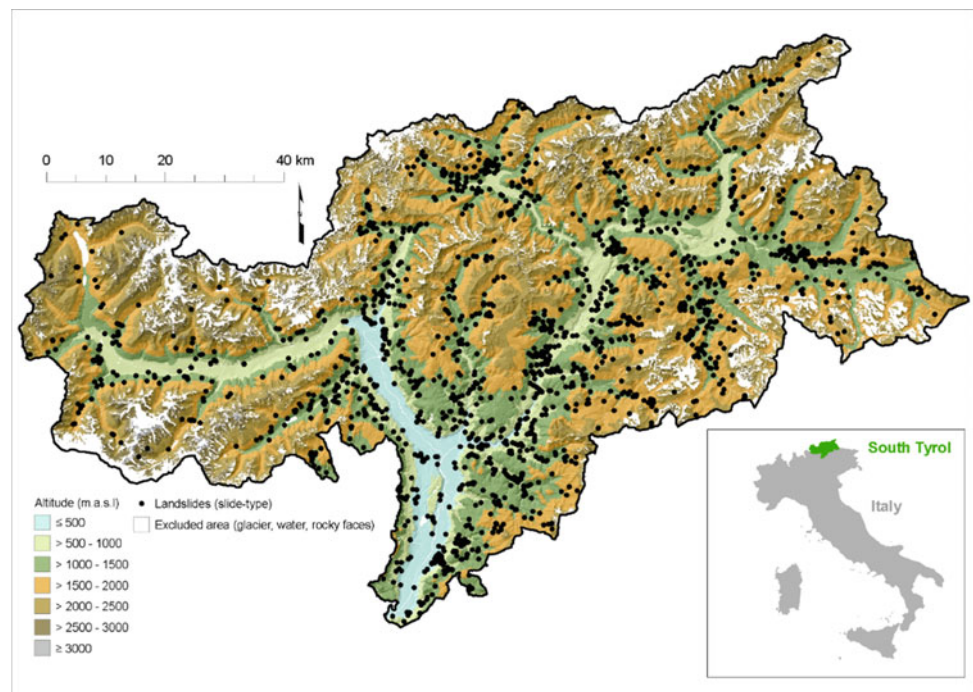
In the study area, landsliding is controlled by an interplay of manifold (de)stabilizing variables. Besides predisposing factors related to lithology, the properties of weathered material and topography, also slowly changing preparatory factors (e.g. vegetation, prolonged rain and snow melt) and potential triggers (e.g. heavy rainfall, intensive snow melt, storms) are known to affect the location and frequency of slope instability. Human activities are as well an important landslide influencing factor, particularly due to land use practices and construction works (Tasser et al. 2003; Corsini et al. 2005; Stingl and Mair 2005; Borgatti and Soldati 2010; Piacentini et al. 2012).

Landslide Data and Environmental Factors

The available landslide inventory is based on the Italian Landslide Inventory project (IFFI: Inventario dei Fenomeni Franosi in Italia) which was launched in 1999 (Trigila et al. 2010). Up to January 2019, the South Tyrolean IFFI data consists of 7573 registered events, among them 1928 shallow landslides of the slide-type movement. Positionally accurate points (i.e. GPS-based mapping in the field) at the landslide scarp positions represent these slide-type movements (Fig. 1). Deep-seated landslides were not considered, because of the necessity to analyse deep-seated phenomena separately from their shallow counterparts and a low sample size of deep-seated movements. Due to the underlying provincial landslide mapping strategy only events that required an intervention by the geological survey for the purpose of documentation, mitigation and/or prevention are registered within this data set. Therefore, the inventory mainly relates to events that caused damage.

Information on spatial environmental variables were accessed via the provincial Geodata platform (Geokatalog 2019). Topographic information was derived from a 2.5 m × 2.5 m Light Detection and Ranging (LiDAR) digital terrain model (DTM) which was resampled to 10 m spatial resolution in order to achieve an adequate generalization (e.g. smoothing of noise) and to avoid a close description of post-failure topography (Steger et al. 2020). Spatial information on infrastructure (i.e. roads, pathways, buildings footprints), perennial water courses, geology (i.e.

Fig. 1 Overview of the study area and spatial distribution of inventoried slide-type movements (one point per landslide scarp). Areas shown in white (i.e. glaciers, water bodies, rocky faces) were a-priori excluded from the analysis



‘*Geologische Übersichtskarte Südtirol*’) and land use/cover (‘*Realnutzungskarte*’ version 2015; 1:10,000) were also extracted from the open Geodata platform (Geokatalog 2019).

Methods

Data Preparation

A binary variable that consist of (i) landslide presences (i.e. inventory data) and (ii) randomly sampled landslide absence locations builds a common basis to investigate landslide controls or landslide susceptibility (Brenning et al. 2015; Schmaltz et al. 2017; Reichenbach et al. 2018). For this study, 1928 absence locations (1:1 ratio presences to absences) were randomly distributed across the study site while glaciers, water bodies, rock faces and locations in close proximity to inventoried landslides (i.e. 50 m buffer) were a-priori disregarded (Fig. 1). Trigonometric calculations based on a slope angle map were used to control sampling intensity of absences in order to account for discrepancies between the dimension of an area in planar view and the actual surface area (i.e. the steeper the terrain, the higher the sampling probability of a cell). This procedure was assumed to be of utility to counterbalance bias related to the presence of “trivial” terrain (cf. Steger and Glade 2017) and an overrepresentation of landslides in steep terrain.

SAGA GIS (Conrad et al. 2015) and the R package RSAGA (Brenning 2008) were used to calculate the commonly utilized DTM derivatives slope angle, slope aspect, convergence index and the upslope contributing catchment area. The latter variable was transformed logarithmically to decrease skewness (Brenning et al. 2015). GRASS GIS (Lennert 2017) was utilized to derive 10 initial geomorphological phenotypes (cf. Jasiewicz and Stepinski 2013) which were heuristically merged to derive six distinct landforms. Euclidean distances were calculated to obtain 10 m raster layers that represent the distance to rivers (i.e. perennial rivers and streams), roads (i.e. asphalt roads, highways), paths (i.e. unpaved roads and trails), transport networks (i.e. streets and paths merged) and to buildings. Nearest neighbour resampling was applied to rasterize land use/cover information and the geological map (Geokatalog 2019).

Exploratory Data Analysis

Relationships between landslide presences/absences and spatial variables were elaborated by means of conditional frequency plots (c.f. Brenning 2008; Steger et al.

2016). The plots depict how the observed landslide presence to absence ratio varies across variable values. Due to the balanced number of observation in the initial sample (1:1 presence to absence ratio), a conditional frequency of 0.5 can be interpreted as equal-ratio-threshold that separates ‘*relative high landslide densities*’ (>0.5) from ‘*low densities*’ (<0.5).

Insights into the efficiency of single variables to ‘*explain*’ the distribution of landslide presences and absences were gained by evaluating the individual discriminatory power using the Area Under the Receiver Operating Characteristic curve (AUROC) in analogy to Steger et al. (2020). In summary, predictions based on a Generalized Additive Model (Wood 2006) were used to evaluate the ability of single explanatory variables to separate landslides presences from absences (AUROC of 0.5: random separation, 1: perfect separation).

Results and Discussion

Conditional frequencies showed that medium inclined terrain (Fig. 2a), south and east exposed sides (Fig. 2b), foot-slopes (Fig. 2c), areas below 1600 m a.s.l (Fig. 2d), and areas close to rivers (Fig. 2f) were associated with comparably high conditional landslide frequencies. In contrast, landslide scarps were more sparsely registered for flat areas (Fig. 2a, c) and very steep terrain (Fig. 2a), for north facing slopes (Fig. 2b), for higher slope positions (i.e. ridges, spurs) and high altitudes (Fig. 2c, d). Low landslide frequencies were also observed for extraordinarily convex and concave shaped terrain (Fig. 2e), for locations distant from rivers (Fig. 2f) and for areas that exhibit a very high contributing area (Fig. 3a). The plots that depict the distance to infrastructure (Fig. 3b–e) provided clear evidence that landslides were frequently registered for areas close to asphalt roads (Fig. 3b), pathways (Fig. 3c) and buildings (Fig. 3e).

Considering that the underlying landslide data represents mainly events documented by the geological survey for intervention, mitigation and/or prevention, we inferred that the observed associations have to be interpreted with caution. For instance, the observed high landslide frequencies close to infrastructure and at lower slope positions might mainly be attributed to the provincial landslide mapping strategy, which concentrates on events that led to interventions and neglects areas far from potentially exposed assets (e.g. roads, major infrastructures, buildings).

In this context, we deduced that the observed varying conditional landslide frequencies among rock types (Fig. 3f) and land cover units (Fig. 4a) do not only reflect slope stability effects, but can partly be ascribed to the underlying landslide mapping design. We expect an overrepresentation of inventoried landslide occurrence within the units (e.g.

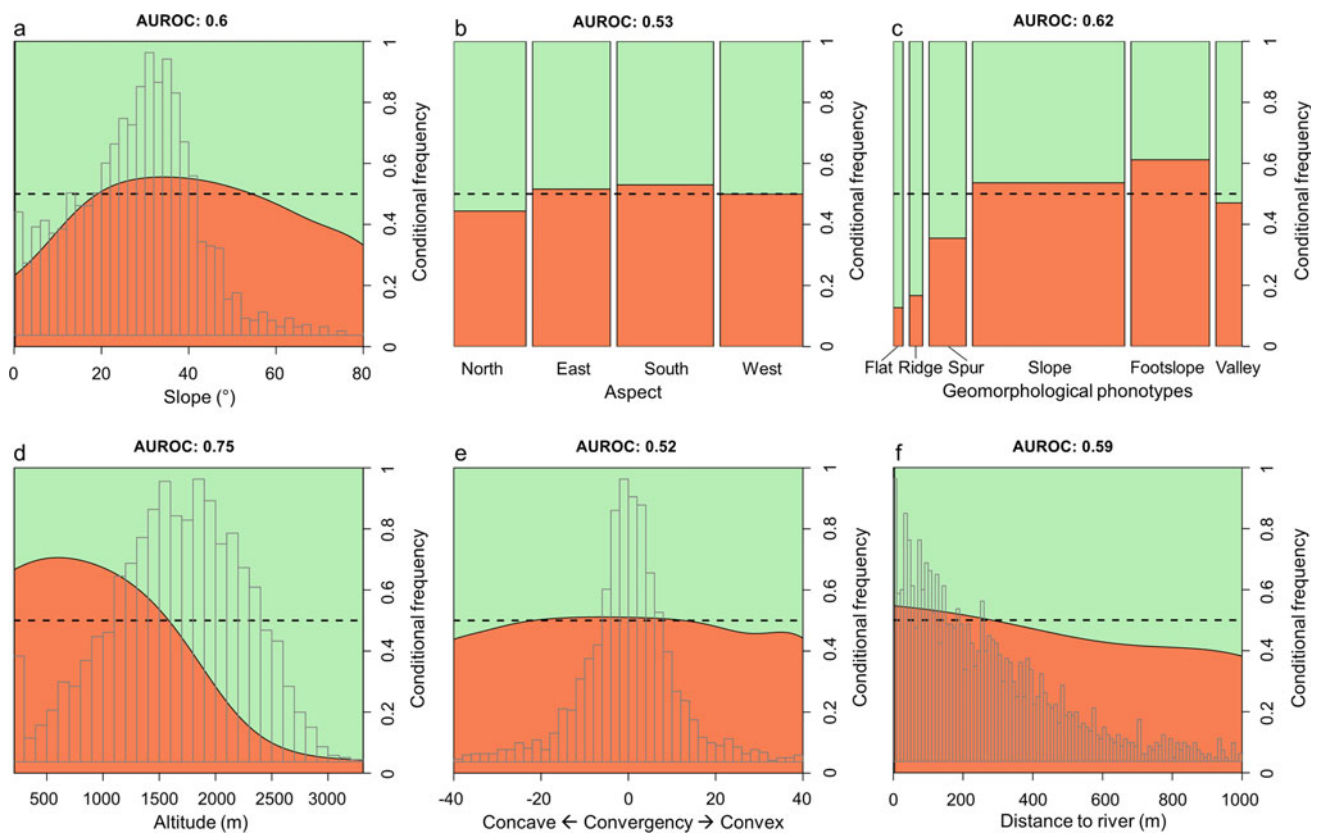


Fig. 2 Exploratory analysis for the variables slope angle (a), aspect (b), geomorphic phenotypes (c), altitude (d), convergence index (e) and distance to river (f). The conditional frequency illustrates the ratio of landslide presences (orange area) to absences (green area) for the respective variable. The histograms (grey bars) relate the distribution of

continuous variables (a, d, e, f) and the width of the bars depicts data density for categorical variables (b, c). The AUROC shows the discriminatory power for single variables (AUROC 0.5: random separation, 1: perfect separation)

geological classes, land cover types) that spatially coincide with densely built hillsides (i.e. denser road network, many buildings).

Comparisons of the statistical associations among environmental variables (e.g. slope angle vs. land cover in Fig. 4c, d) underlined that medium inclined terrain was simultaneously associated with both a higher portion of forested area (Fig. 4d) and a high landslide density (Fig. 2a). Thus, the observed slightly above average landslide density in forests (Fig. 4b) does not necessarily highlight that forested areas are slightly less stable than their non-forested equals. We would rather conclude that in South Tyrol forests are likely located on slope inclinations that favour slope instability, while also the dense network of forest roads might play a role on why forest were associated with comparably high landslide mapping densities. It can be assumed

that the observed low landslide density for steep forested areas above ~ 1600 m a.s.l. and at a greater distance from agricultural land and infrastructure/settlements strongly reflects the underlying data collection process. The fact that farmers consciously avoid using unstable zones for agricultural purposes also influences the results.

Ultimately, such interpretations highlight the need for a joint consideration of multiple variables when assessing and interpreting statistical relationships (e.g. multiple-variable models) and the necessity to take the underlying data background explicitly into account.

General guidelines for the AUROC interpretation (cf. Hosmer and Lemeshow 2000) revealed that altitude was the only topographic factor with an acceptable ability to separate landslide presences from absence observations (AUROC > 0.7 , Fig. 2d). Acceptable to excellent AUROCs were

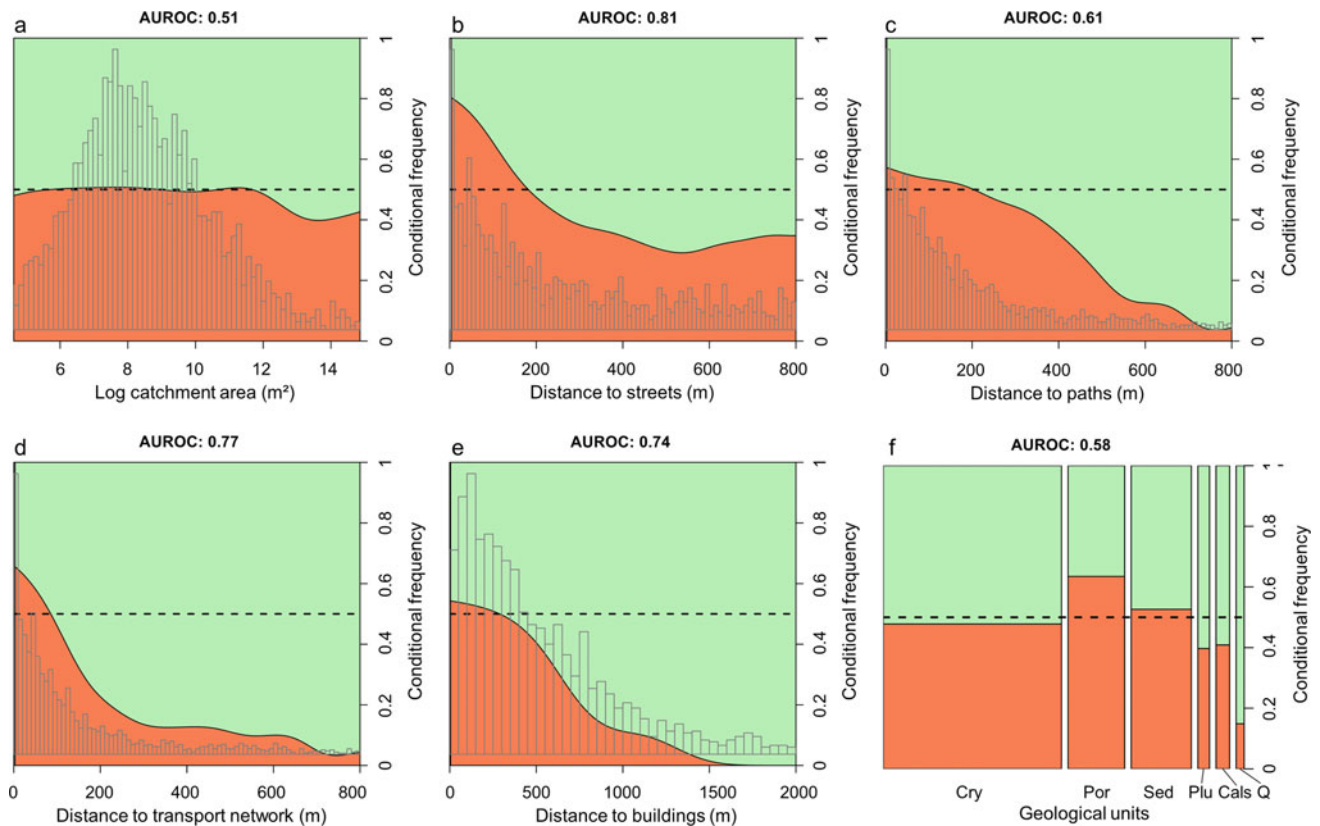


Fig. 3 Exploratory analysis for the variables upslope contributing catchment area (a), distance to streets (b), distance to pathways (c), distance to transport network (streets and pathways; d), distance to buildings (e) and rock types (f). The conditional frequency illustrates the ratio of landslide presences (orange area) to absences (green area) for the respective variable. The histograms (grey bars) relate the

distribution of continuous variables (a–e) and the width of the bars in f depicts data density for geological units (Cry: Crystalline basement, Por: Porphyry, Sed: Sedimentary rocks, Plu: Plutonite, Cals: Calc-schists with ophiolites, Q: Quaternary). The AUROC shows the discriminatory power for single variables (AUROC 0.5: random separation, 1: perfect separation)

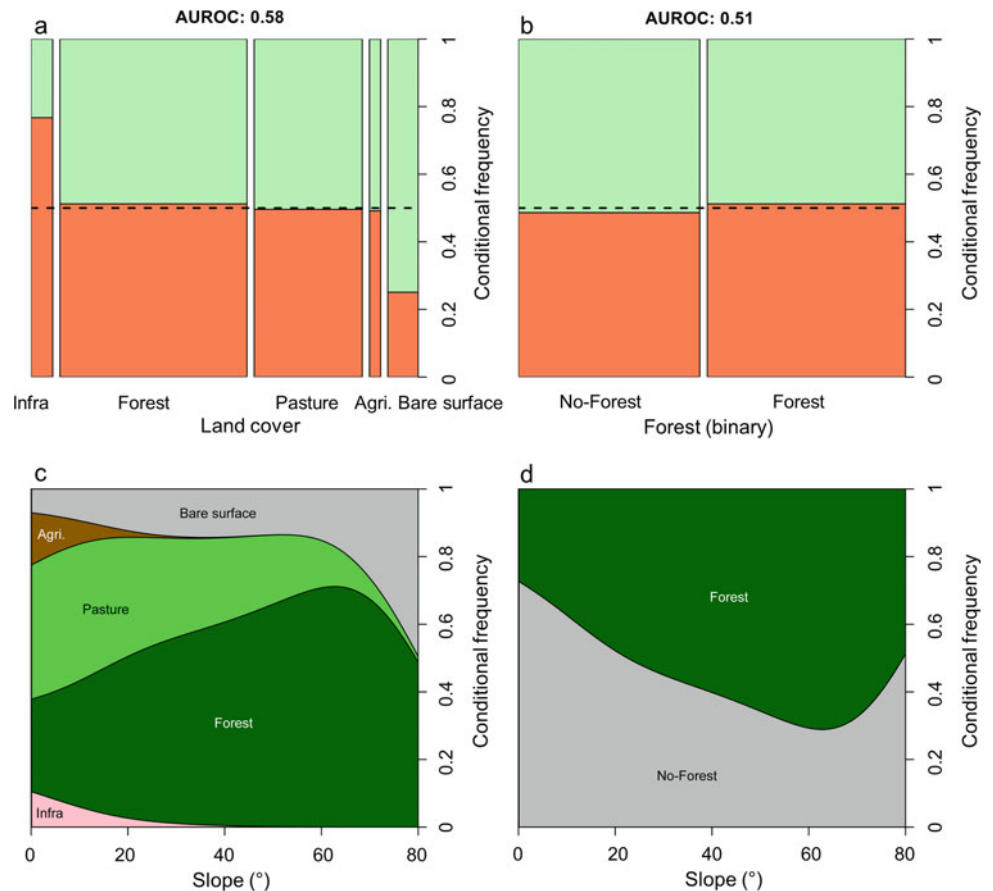
observed for the variables that depict the proximity to important infrastructure: distance to buildings (0.74, Fig. 3e) and distance to roads (0.81, Fig. 3b).

These numbers do not provide evidence that undercutting or overloading of hillslopes due to infrastructure construction are important landslide predisposing factor for South Tyrol. The obtained AUROCs reflect the provincial mapping strategy that explicitly focuses on important infrastructure and settlements. We conclude that statistically ‘important’ or ‘significant’ associations between landsliding and environmental factors have to be interpreted in the light of the underlying data characteristics and by questioning their geomorphological plausibility (Steger et al. 2016).

Conclusion and Outlook

This contribution highlights how erroneous conclusions are likely to follow from statistical analyses when the background information associated with available landslide data is ignored. A look at published literature indicates that recalling the well-known principle ‘*correlation does not imply (geomorphic) causation*’ seems to be of particular relevance in the field of statistically-based landslide analysis (e.g. susceptibility modelling), also because (i) the vast majority of available landslide inventories exhibit a spatially heterogeneous completeness and because (ii) analysts still

Fig. 4 Exploratory analysis of the variables land cover (a) and forest/no-forest (b) and interrelations among land cover and slope angle (c, d). The conditional frequency in a and b illustrates the ratio of landslide presences (orange area) to absences (green area) for the respective class while the width of the bars depicts data density. The AUROC relates to the discriminatory power of the variables (AUROC 0.5: random separation, 1: perfect separation). Conditional frequencies in c and d show the fraction of land cover units conditional on slope angle



need to assign meaning and geomorphological plausibility to the obtained numbers (Steger et al. 2016, 2017).

Based on the results, we will focus on the co-development (i.e. modellers and landslide data provider) of non-linear multiple-variable spatial landslide models that explicitly account for the underlying data origin. The modelling results will also be presented at the World Landslide Forum 5.

Acknowledgements The authors thank the Autonomous Province of South Tyrol for providing spatial input data.

References

- Borgatti L, Soldati M (2010) Landslides as a geomorphological proxy for climate change: a record from the Dolomites (northern Italy). *Geomorphology* 120:56–64
- Brenning A (2008) Statistical geocomputing combining R and SAGA: the example of landslide susceptibility analysis with generalized additive models. *Hamburger Beiträge Phys Geogr Landschaftsökol* 19:410
- Brenning A, Schwinn M, Ruiz-Páez AP, Muenchow J (2015) Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Nat Hazards Earth Syst Sci* 15:45–57
- Conrad O, Bechtel B, Bock M et al (2015) System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci Model Dev* 8:1991–2007
- Corsini A, Pasuto A, Soldati M, Zannoni A (2005) Field monitoring of the Corvara landslide (Dolomites, Italy) and its relevance for hazard assessment. *Geomorphology* 66:149–165
- Cruden DM, Varnes DJ (1996) Landslide types and processes: investigation and mitigation. TRB special report, 247. National Academy Press, Washington, pp 36–75
- Geokatalog (2019) Open geodatabase of the Autonomous Province of South Tyrol. Accessible via <https://geokatalog.buergemetz.bz.it/geokatalog/>. Accessed 20 Feb 2019
- Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York
- Hungr O, Leroueil S, Picarelli L (2014) The Varnes classification of landslide types, an update. *Landslides* 11:167–194
- Jasiewicz J, Stepinski TF (2013) Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology* 182:147–156
- Lennert M (2017) GRASS development team. Geographic resources analysis support system (GRASS) software, version 7
- Piacentini D, Troiani F, Soldati M et al (2012) Statistical analysis for assessing shallow-landslide susceptibility in South Tyrol (south-eastern Alps, Italy). *Geomorphology* 151–152:196–206
- Reichenbach P, Rossi M, Malamud BD et al (2018) A review of statistically-based landslide susceptibility models. *Earth Sci Rev* 180:60–91
- Schmaltz EM, Steger S, Glade T (2017) The influence of forest cover on landslide occurrence explored with spatio-temporal information. *Geomorphology* 290:250–264

- Steger S, Glade T (2017) The challenge of “trivial areas” in statistical landslide susceptibility modelling. In: Proceedings of the 4th world landslide forum, 29 May–2 June, Ljubljana. Springer, Ljubljana
- Steger S, Brenning A, Bell R et al (2016) Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. *Geomorphology* 262:8–23
- 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
- Steger S, Schmaltz E, Glade T (2020) The (f)utility to account for pre-failure topography in data-driven landslide susceptibility modelling. *Geomorphology* 354:107041
- Stingl V, Mair V (2005) Einführung in die Geologie Südtirols: [aus Anlass des 32. Internationalen Geologischen Kongresses im Sommer 2004 in Florenz]. Autonome Provinz Bozen-Südtirol. Amt Geol Baustoffprüfung
- Tasser E, Mader M, Tappeiner U (2003) Effects of land use in alpine grasslands on the probability of landslides. *Basic Appl Ecol* 4:271–280
- Trigila A, Iadanza C, Spizzichino D (2010) Quality assessment of the Italian Landslide Inventory using GIS processing. *Landslides* 7:455–470
- Wood SN (2006) Generalized additive models: an introduction with R. Chapman and Hall/CRC, Taylor and Francis Group, Boca Raton