eurac research

AUTONOME PROVINZ BOZEN - SÜDTIROL
PROVINCIA AUTONOMA DI BOLZANO - ALTO ADIGE
PROVINZIA AUTONOMA DE BULSAN - SÜDTIROL

unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

The necessity to consider the landslide data origin in statistically-based spatial predictive modelling

-

A landslide intervention index for South Tyrol (Italy)

Stefan Steger[1], Volkmar Mair[2], Christian Kofler[1,3], Stefan Schneiderbauer[1,4], and Marc Zebisch[1]

[1]Eurac Research, Institute for Earth Observation, Bolzano-Bozen, Italy (stefan.steger@eurac.edu)

[2]Office for Geology and Building Materials Testing, Autonomous Province of Bolzano-South Tyrol, Cardano, Italy

[3]Faculty of Science and Technology, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy

[4]United Nations University, Institute for Environment and Human Security, GLOMOS program

Introduction

Many statistically-based landslide susceptibility maps are supposed to depict the relative likelihood of an area to be affected by future landslides.

Current practice (according a literature review):

- Vital modelling decisions, such as the selection of explanatory variables, are frequently based on quantitative criteria (e.g. predictive performance)
- Models with an apparent high predictive performance are used to produce maps for spatial planning or to infer the causes of slope instability

Our hypothesis: Wrong conclusions are likely to follow whenever the origin/characteristics of the underlying landslide data is ignored and modelling decisions are based solely on predictive performance estimates

The **aims** were to (i) analyse available landslide inventory data in the context of its origin in order to (ii) highlight potential pitfalls of performance driven procedures and to (iii) develop a predictive model that takes landslide background information into account.

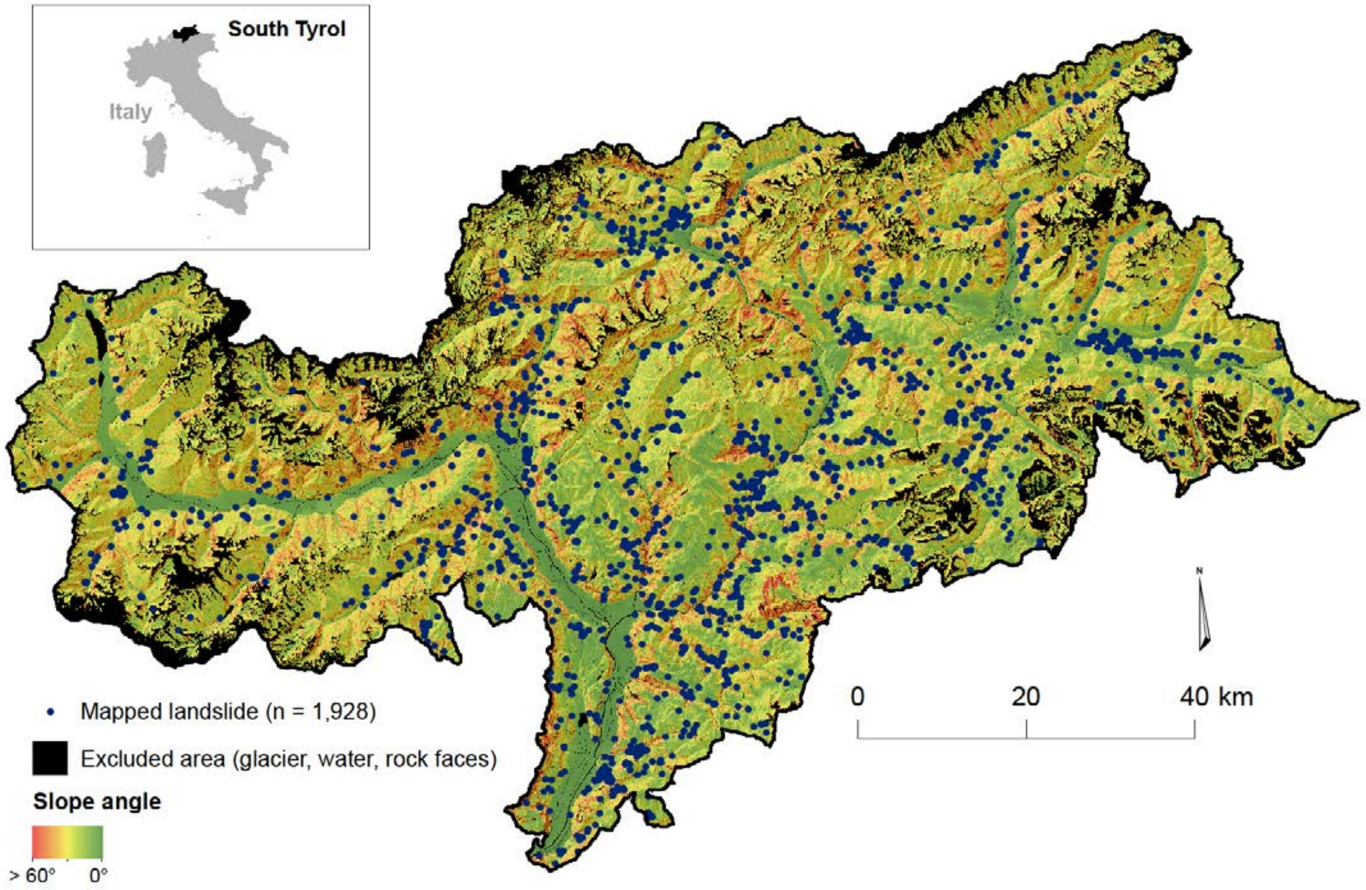
Study area and data

Study area

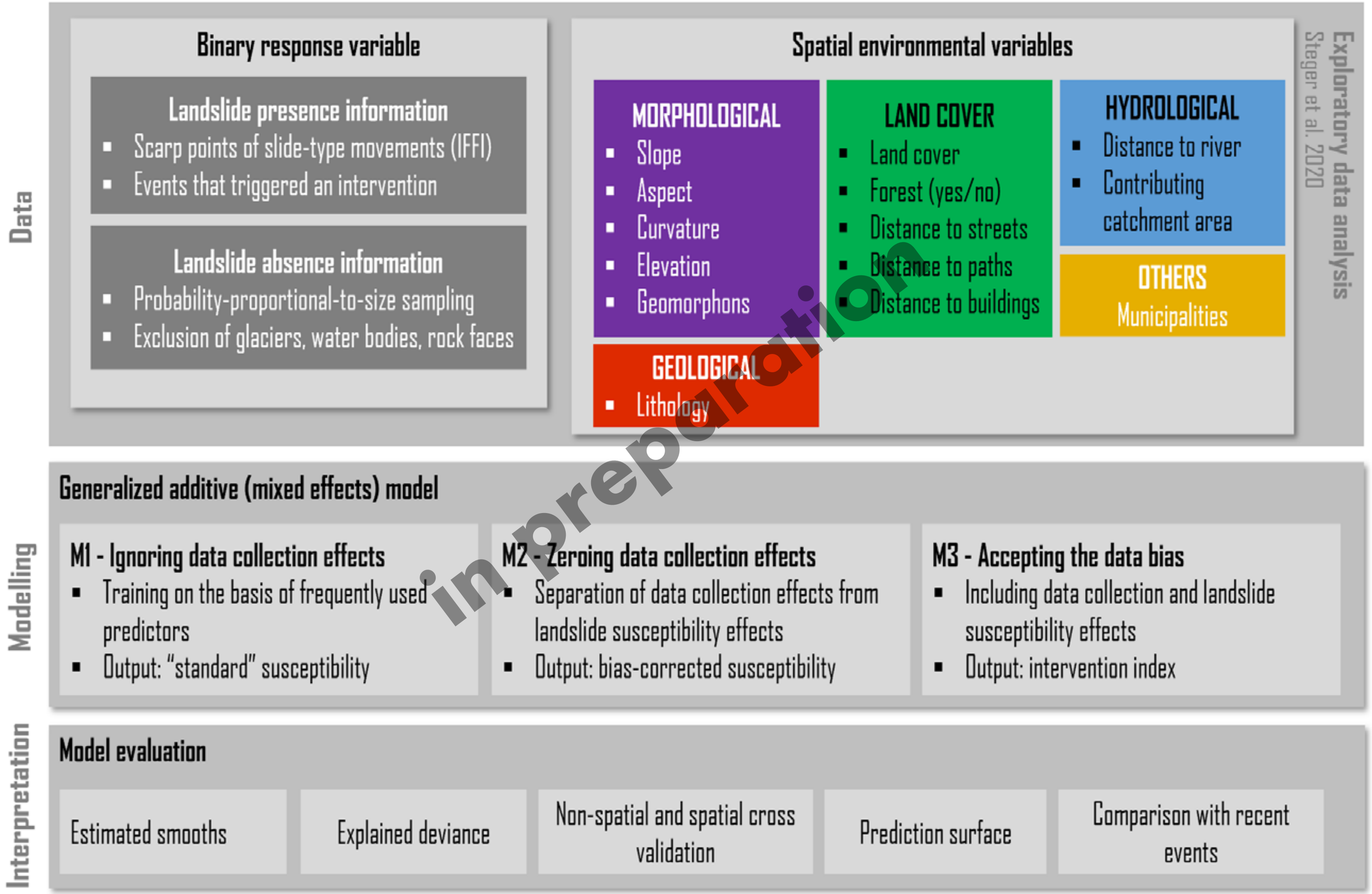
- Autonomous Province of South Tyrol (Northern Italy)
- Areal extent over 7,400 km²
- Characterized by a considerable heterogeneity in terms of geomorphology, geology, land cover, land use and climate

Landslide data

- Based on the Italian landslide inventory (IFFI project)
- For this study: shallow slide-type movements
- 1928 positionally accurate landslide scarp locations
- **Data origin and collection context**: The data relates to landslides that induced interventions by e.g. the road service or the geological office (→ **damage causing events**)



The approach



Steger et al. (in prep.)

3 Models

M1 – Ignoring data collection effects (frequent practice)

M2 – Zeroing data collection effects

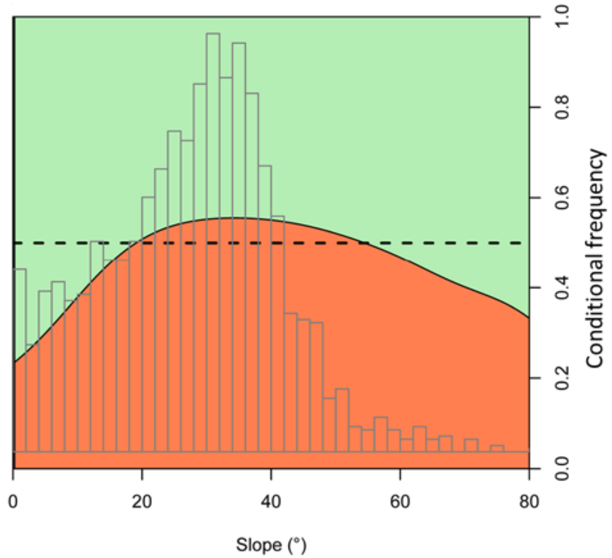
M3 – Including data collection and landslide susceptibility effects to produce a landslide intervention index

Geomorphic plausibility?, Bias? etc.

Some results: exploratory data analysis

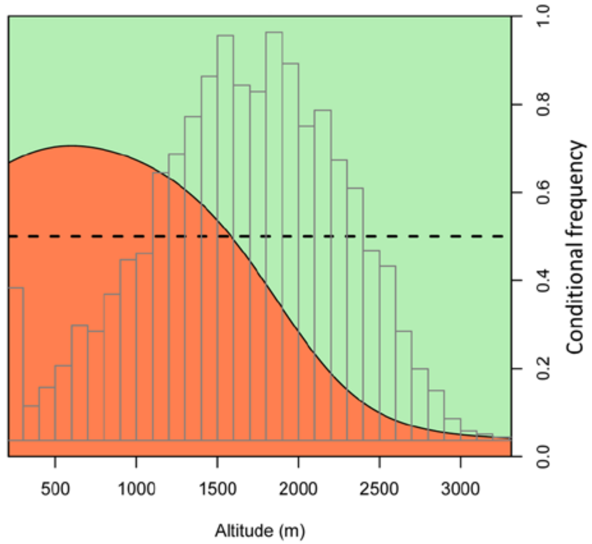
3 commonly applied predictors

Slope



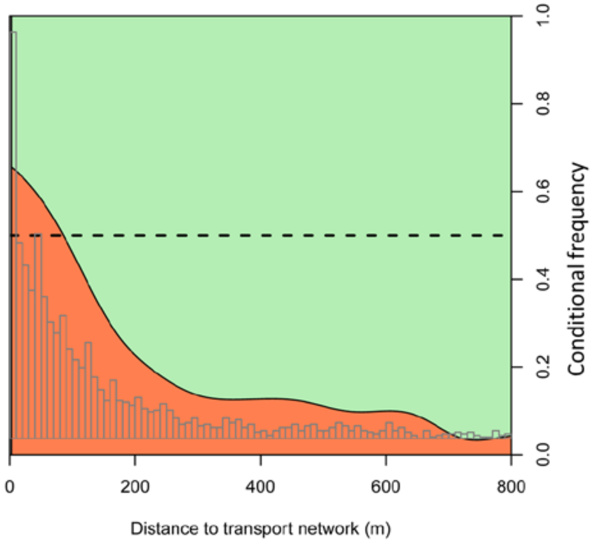
Landslides were frequently mapped for medium inclined slopes and seldomly for flat and very steep terrain → likely a **landslide susceptibility effect**

Altitude



Landslides were frequently mapped at lower altitudes and seldomly for high alpine areas → **most likely a data collection effect (or a mixture)**

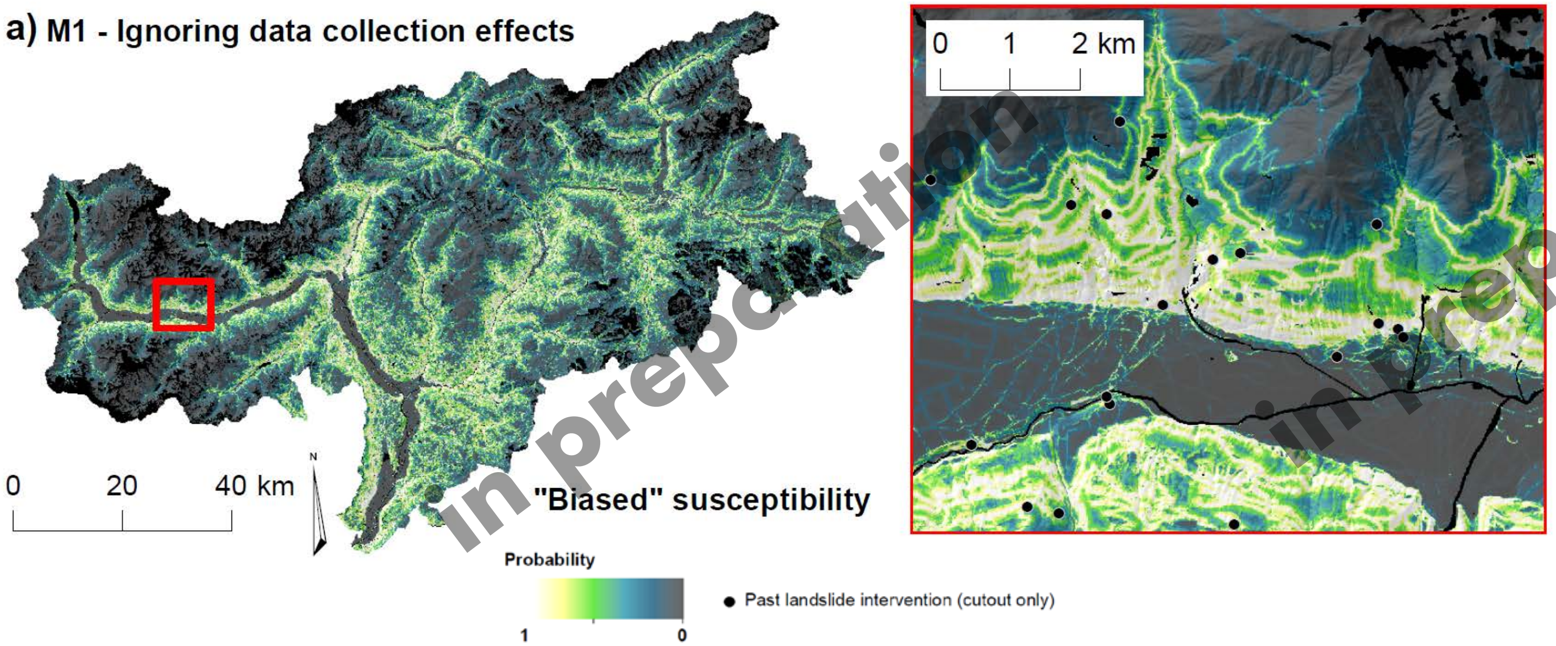
Distance to transport network

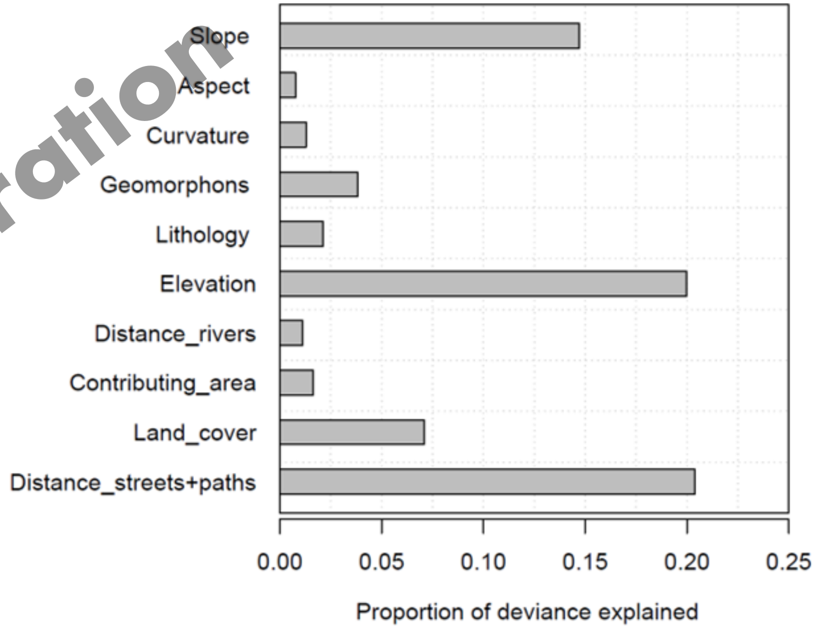


Landslides were much more frequently mapped in close distance to infrastructure and seldomly far away from it → **very likely a data collection effect**

modified from Steger et al. 2020 (accepted)

Some results: M1 “frequent practice”



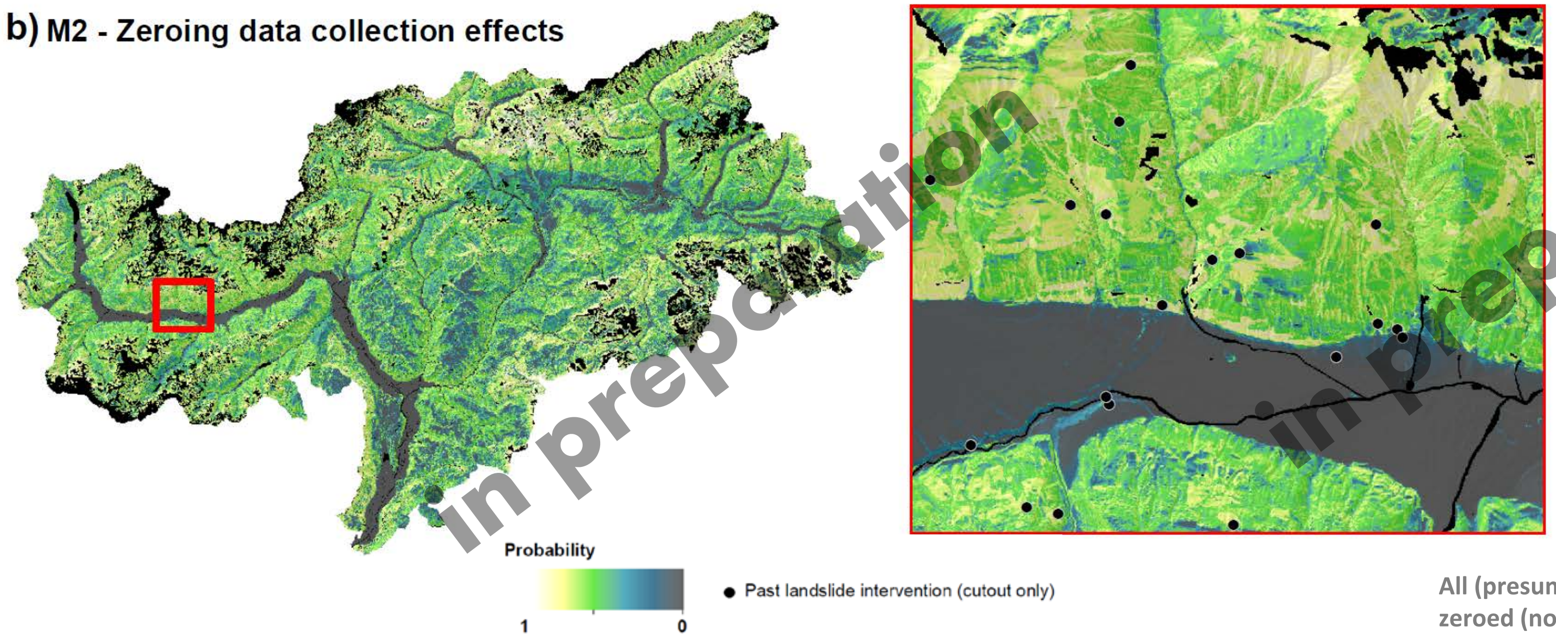


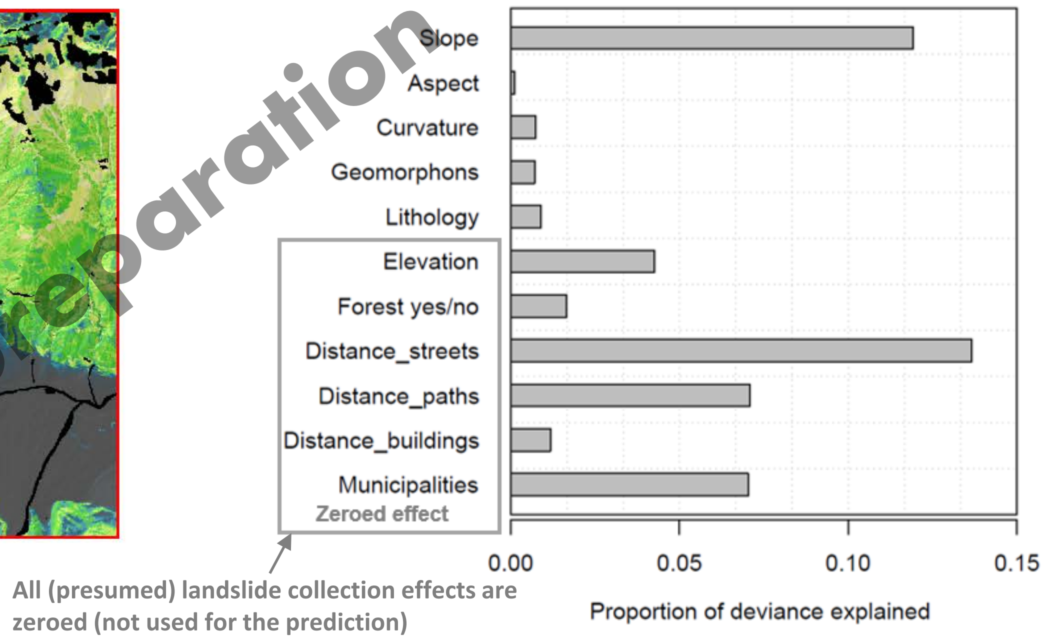
Steger et al. (in prep.)

- **Excellent predictive performance** (CV-AUROC: 0.87, SCV-AUROC: 0.86)
- Most “influential” variables: **distance to streets/paths, elevation, slope**
- Highest landslide likelihood: medium inclined slopes in close distance to infrastructure and at lower slope positions

Interpretation: Despite its excellent performance, the produced map **does not reflect landslide susceptibility** nor can the model be used to infer important landslide predisposing factors. The results are a **mixture of** both, “**landslide susceptibility effects**” (e.g. low likelihood of flat and very steep terrain) **and effects associated with** the provincial **landslide data collection** strategy (e.g. few interventions at high altitudes, increasing number of interventions with decreasing distance to infrastructure). From a geomorphic point of view, the **well-performing model is highly biased**.

Some results: M2 “bias-corrected susceptibility”



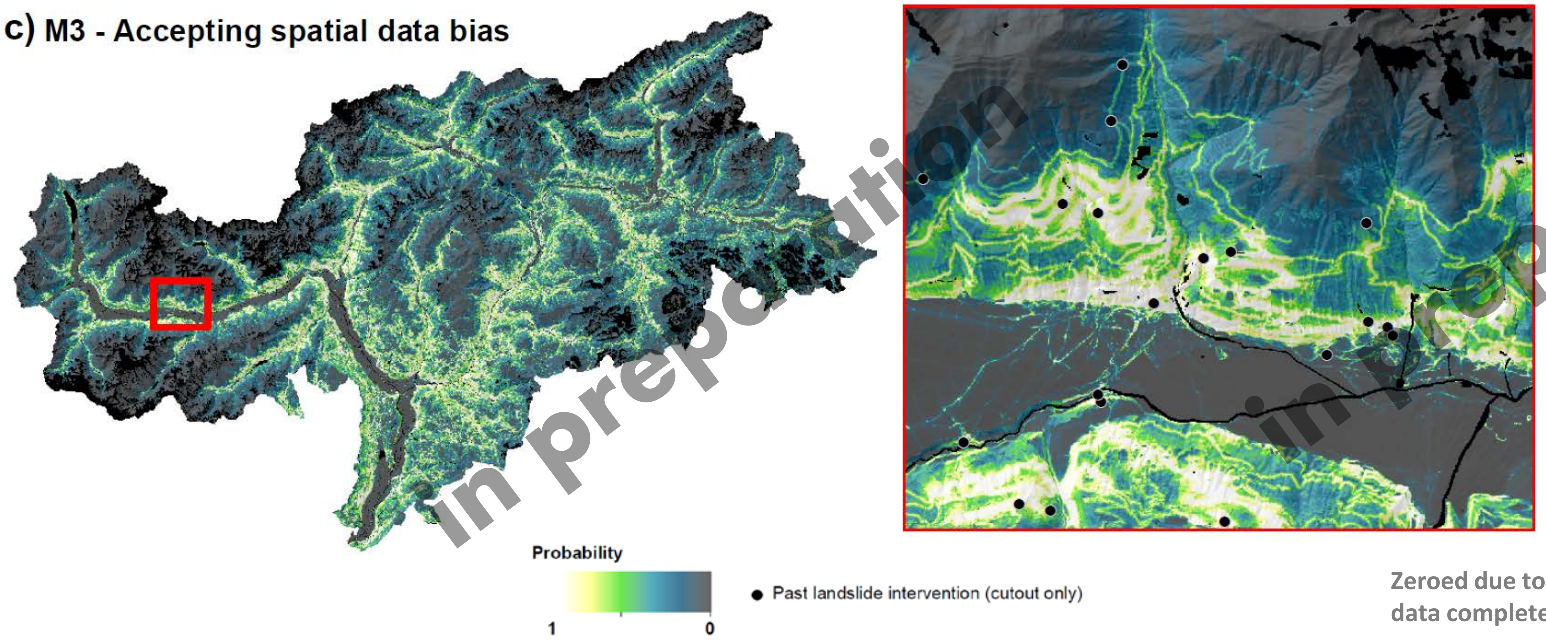


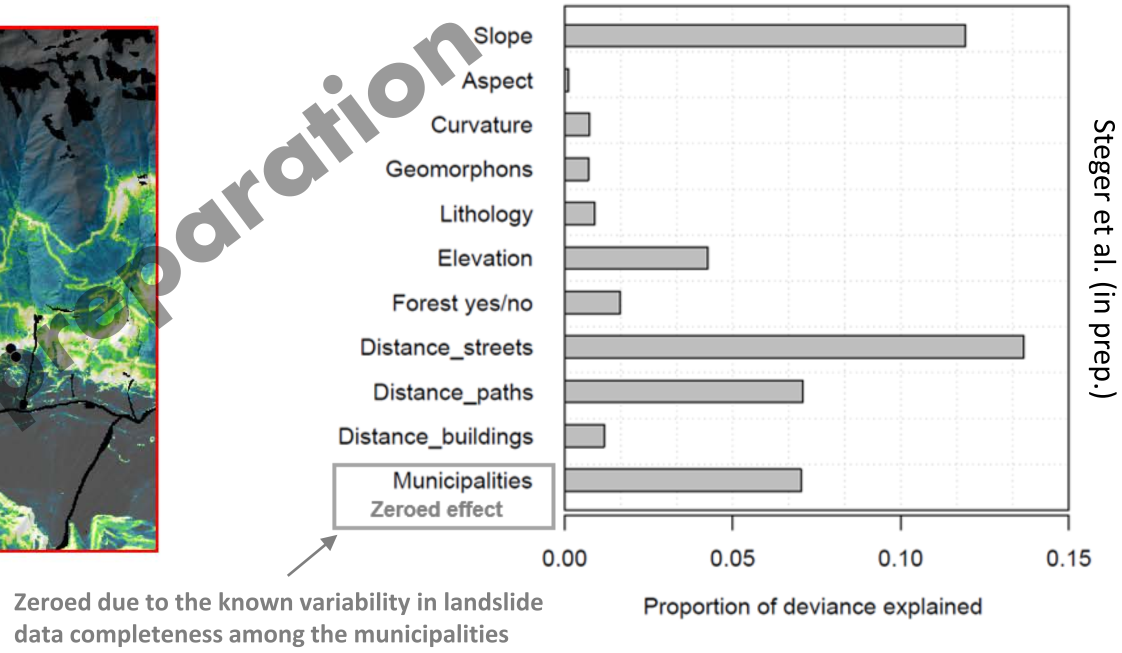
Steger et al. (in prep.)

- **Poor predictive performance** (CV-AUROC: 0.59, SCV-AUROC: 0.61)
- Rather **uniform spatial pattern at the hillsides:** slope dominates because other influential variables are averaged out (zeroed)

Interpretation: A sole focus on landslide susceptibility effects is **challenging** in case the underlying **spatial landside distribution reflects – to a very large extent – data collection effects** (here: landslides distant from infrastructure are ignored). To avoid an error propagation, many “influential” predictors that describe this bias have to be zeroed. Furthermore, some variables concurrently represent landslide susceptibility effects and data collection effects (e.g. altitude, land cover). All this renders a bias-correction using mixed-effects modelling particularly challenging (in this case unsuccessful).

Some results: M3 “landslide intervention index”





Steger et al. (in prep.)

- **Excellent predictive performance** (CV-AUROC: 0.88, SCV-AUROC: 0.87)
- A very high portion of spatially independent landslide interventions was predicted accurately

Interpretation: The model does not depict landslide susceptibility. However, it is **in line with the data collection procedure** and it describes simultaneously landslide susceptibility effects and data collection effects. The model allows – with high accuracy – to identify **areas where future interventions are likely to take place**. The results can help provincial authorities to allocate resources and to gain knowledge on where damage causing events can be expected in the future. Validation with landslides (n =64) that caused damage during a recent storm event (Nov. 2019) indicates its high predictive power (prediction rate 0.95).

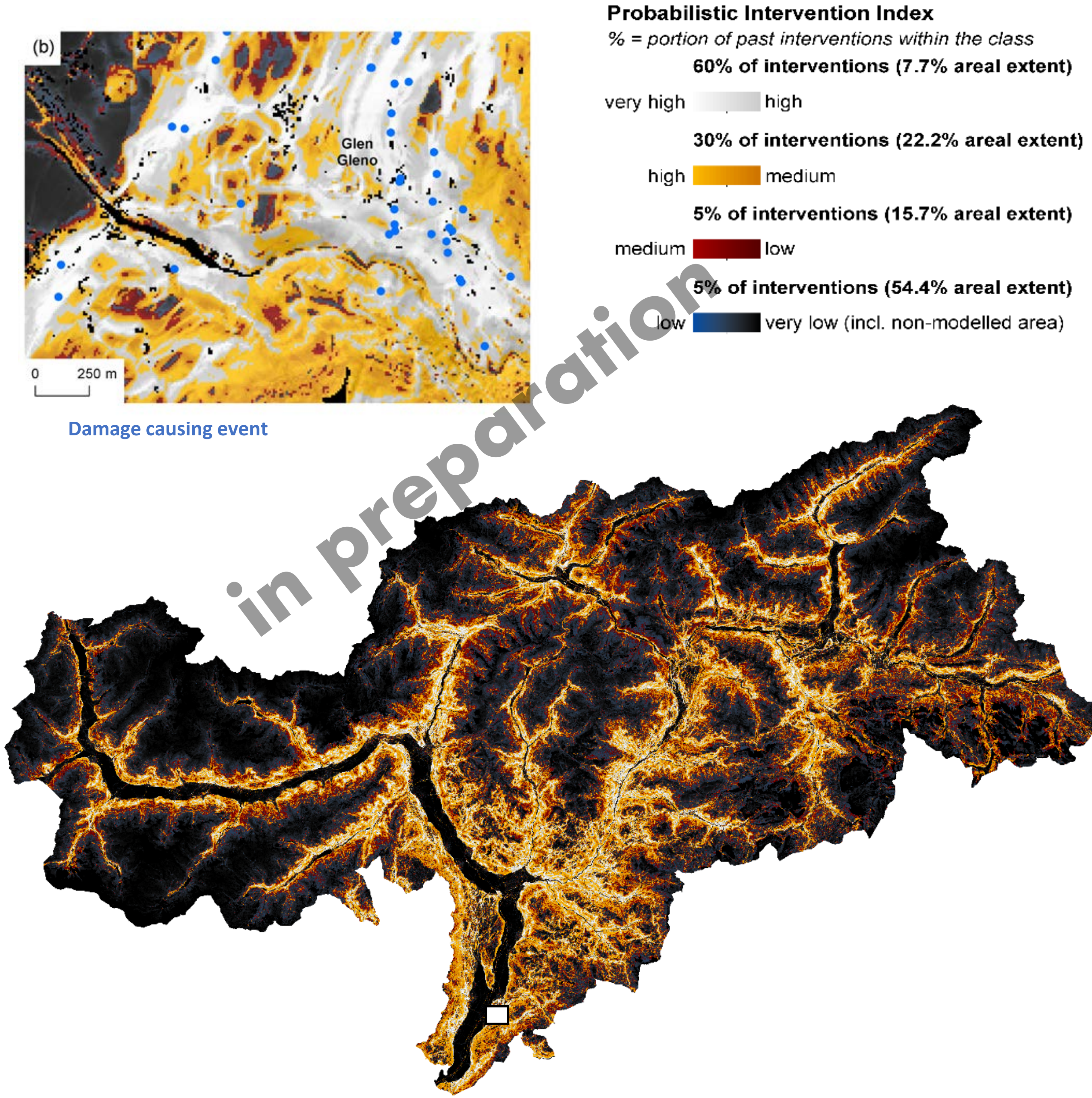
Conclusions

- **Wrong conclusions** can be drawn from excellently performing statistical models **whenever qualitative background information is disregarded**

- Variables that increase the predictive performance do not necessarily describe geomorphically plausible effects (e.g. distance to infrastructure, elevation, land cover)

- The **landslide intervention model** provides a spatial estimate on where future landslides that trigger interventions/damage are likely to be initiated (“Impact-focused assessment”)

- *Local authorities: “Excellent **tool** for the monitoring and **planning of ordinary and extraordinary maintenance** (of infrastructure and settlements)”*



Steger et al. (in prep.)

Thank you for your attention!

stefan.steger@eurac.edu