Dear editor,

please find below a marked-up manuscript version that highlights all changes made. Point-by-point responses to the referees can be found within the interactive discussion section.

Best regards,

Stefan Steger

THE PROPAGATION OF INVENTORY-BASED POSITIONAL ERRORS INTO STATISTICAL LANDSLIDE SUSCEPTIBILITY MODELS

Stefan Steger[1], Alexander Brenning[2], Rainer Bell[1], and Thomas Glade[1]

5 1 Department of Geography and Regional Research, University of Vienna, Universitätsstraße 7, A-1010 Vienna, Austria
2 Department of Geography, Friedrich Schiller University Jena, Loebdergraben 32, D-07743 Jena, Germany

Correspondence to: Stefan Steger (stefan.steger@univie.ac.at)

Abstract. There is unanimous agreement that a precise spatial representation of past landslide occurrences is a prerequisite to produce high quality statistical landslide susceptibility models. Even though perfectly accurate landslide inventories rarely

10 exist, investigations of how landslide inventory-based errors propagate into subsequent statistical landslide susceptibility models are scarce.

The main objective of this research was to systematically examine whether and how inventory-based positional inaccuracies of different magnitudes influence modelled relationships, validation results, variable importance and the visual appearance of landslide susceptibility maps. The study was conducted for a landslide-prone site located in the districts of Amstetten and

15 Waidhofen/Ybbs, Eastern Austria, where an earth-slide point inventory (n = 591) was available.

The methodological approach comprised an artificial introduction of inventory-based positional errors into the present landslide data set and an in-depth evaluation of subsequent modelling results. Positional errors were introduced by artificially changing the original landslide position by a mean distance of 5, 10, 20, 50 and 120 m. The resulting differently precise response variables were separately used to train logistic regression models. Odds ratios of predictor variables

20 provided insights into modelled relationships. Cross-validation and spatial cross-validation enabled an assessment of predictive performances and permutation-based variable importance. All analyses were additionally carried out with synthetically generated data sets to further verify the findings under rather controlled conditions.

The results revealed that an increasing positional inventory-based error was generally related to increasing distortions of modelling and validation results. However, the findings also highlighted that interdependencies between inventory-based

25 spatial inaccuracies and statistical landslide susceptibility models are complex. The systematic comparisons of 12 models provided valuable evidence that the respective error-propagation was not only determined by the degree of positional inaccuracy inherent in the landslide data, but also by the spatial representation of landslides and the environment, landslide magnitude, the characteristics of the study area, the selected classification method and an interplay of predictors within multiple variable models. Based on the results, we deduced that a direct propagation of minor to moderate inventory-based

30 positional errors into modelling results can be partly counteracted by adapting the modelling design (e.g. generalization of input data, opting for strongly generalizing classifiers). Since positional errors within landslide inventories are common and

subsequent modelling and validation results are likely to be distorted, the potential existence of inventory-based positional inaccuracies should always be considered when assessing landslide susceptibility by means of empirical models.

Keywords: Landslide susceptibility, Logistic regression, Landslide inventory, Validation, Variable importance, Error propagation~~;~~.

5   1. Introduction

The analysis of landslide susceptibility has been a highly active research topic during the past decades (Wu et al., 2015). ~~The term landslide susceptibility refers to the likelihood of a certain location to be affected by upcoming landslide events (Brabb, 1984; Corominas et al., 2013; Fell et al., 2008; Guzzetti et al., 1999, 2005).~~ The term landslide susceptibility refers to the likelihood of a certain location to be affected by upcoming landslides without taking into account the potential temporal
10   occurrence or magnitude of landslide events (Brabb, 1984; Corominas et al., 2013; Fell et al., 2008; Guzzetti et al., 1999, 2005).

For large regions or areas where detailed geotechnical information is missing, statistically-based approaches are most commonly applied to generate landslide susceptibility maps (Cascini, 2008; Corominas et al., 2013; ~~van~~Van Westen et al., 2008). The underlying concept is based on a modified uniformitarian principle (Hutton, 1788) by assuming that future
15   landslides evolve more likely under those environmental conditions that led to past slope movements (Brabb, 1984; Ermini et al., 2005; Fabbri et al., 2003; Fell et al., 2008; Glade et al., 2005).

In practice, this idea is operationalized by fitting a binary classification model to a data set containing spatial information on the presence and absence of past landslides (response) and a number of associated static preparatory environmental factors (predictors; e.g. slope, lithology). The resulting classification rule allows to identify conditions that were likely to promote
20   past slope movements, and conditions that favoured slope stability. When the constructed model is considered to appropriately describe the underlying relationship, a likelihood of landslide occurrence can be assigned to all spatial units containing information on these environmental conditions. Thus, a specific location shows a high landslide susceptibility whenever the respective environmental setting is similar to the conditions observed for past landslide occurrences (Ermini et al., 2005; Goetz et al., 2015b; Petschko et al., 2014a; Van Den Eeckhaut et al., 2006).

25   The quality and applicability of a landslide susceptibility map is regularly inferred from predictive performance estimates which compare the predicted susceptibility pattern with an independent data set that is supposed to mimic future landslide locations (Beguería, 2006; Brenning, 2005; Chung and Fabbri, 2003; Frattini et al., 2010; Remondo et al., 2003). An issue found by many authors is that the quality of the presented landslide susceptibility map is directly related to the quality of the input data used to construct the underlying model (Cascini, 2008; Corominas et al., 2013; Fressard et al., 2014; Guzzetti et
30   al., 2006; ~~van~~Van Westen et al., 2008). In this context, an accurate representation of past landslide activities (landslide inventory) is regularly considered as a crucial component to produce reliable results (Ardizzone et al., 2002; Fressard et al., 2014; Galli et al., 2008; Guzzetti et al., 2012; Harp et al., 2011; Hussin et al., 2016; Steger et al., 2015, 2016). However,

2

perfect landslide inventories are known to be rarely available (Malamud et al., 2004; Petschko et al., ~~2015~~2016). Several studies emphasize that the positional accuracy of a landslide inventory can vary greatly and is influenced by the type (e.g. aerial photograph, shaded-relief image) and quality (e.g. spatial resolution) of available base maps, the type and sizes of mapped landslides, human factors (e.g. accuracy of public reports, covered time period) and the processing of analogue

5   information sources (e.g. digitization) (Ardizzone et al., 2002; Harp et al., 2011; Petschko et al., ~~2015~~2016; Santangelo et al., 2015).

A number of authors compared landslide susceptibility models generated from different inventories (e.g. Ardizzone et al., 2002; Fressard et al., 2014; Galli et al., 2008; Steger et al., 2015, 2016; Zêzere et al., 2009). However, differences between the respective landslide inventories could only be approximated in relative terms due to a variety of underlying data sources

10 and an unavailability of quantitative information on the specific positional errors. Thus, specific statements on how a particular positional error may propagate into the final models and maps could not be made.

This study aims to shed more light on the impact of positionally erroneous landslide inventories on the results of statistical landslide susceptibility models by artificially controlling for the extent of positional inaccuracy inherent in the respective data set. The main objective was to systematically examine the influence of positionally inaccurate landslide inventories on

15 modelled relationships, validation results and the appearance of landslide susceptibility maps. All analyses were also conducted with synthetically generated data to further validate the results under controlled conditions.

2. Study area

The study site extends over 100 km² (20 km x 5 km) and is located in the eastern part of Austria (Fig. ~~1A~~1a). The area belongs to the Lower Austrian districts Amstetten and Waidhofen/Ybbs. Altitudes range from less than 350 m a.s.l. in the

20 northern valley bottoms to 790 m a.s.l. at the southern hilltops. The landslide-prone area (5.9 landslides / km²) can be characterized as a hilly undulating landscape with an average slope gradient of 11° (Fig. ~~1~~1b).

The morphological and lithological condition of the study site plays a major role for the propensity of the respective slopes to fail. The major part of the area (81 km²) relates to alternating sediment sequences of the Rheno-Danubian Flysch Zone (Fig. ~~1B~~1c). Several publications emphasize that the high density of shallow landsliding within the Flysch is related to the

25 widespread presence of deeply weathered soils consisting of silts and clayey material (Petschko et al., ~~2015~~2016; Schwenk, 1992; Wessely et al., 2006). A minor portion of the terrain (3 km²) can be assigned to clastic sediments of the Molasse Zone. The Molasse lies along the northern side of the Eastern Alps and is regularly associated with several meter thick clayey and marly soils which tend to be susceptible to landsliding, even at lower slope gradients (Wessely et al., 2006; Petschko et al., ~~2015~~2016). Quaternary sediments (16 km²) that relate to alluvial deposits and fluvial terraces dominantly cover the valley

30 floors of the area. Landslide density is low within this area, mainly due to its prevalent flat topography (compare Fig. ~~1B~~1b with Fig. ~~1C~~1c).

3

In the light of the general litho-morphological predisposition of the area, landslide triggering is primarily related to a direct water-influx into the subsurface. Most landslides of the area occurred after intensive precipitation and snow-melting events (Petschko et al., ~~2015~~2016; Schweigl and Hervás, 2009; Schwenk, 1992). The resulting reduction of internal friction and increase in water-related loading are considered to be the main driving forces along a regularly distinct sliding plane.

5 However, also human interventions were observed to be an increasingly important landslide influencing factor (Schwenk, 1992; Wessely et al., 2006). The southern lying hilly areas are for the most parts used for intensive cattle farming (e.g. pastures) or covered by forests while arable land is prevalent in the northern lying flat lowlands (Eder et al., 2011). The area lies in a transitional climate zone (oceanic influences from the West and continental influences from the East) with mean annual precipitation rates around 1,000 mm (Skoda and Lorenz, 2007).

10 The majority of landslides in the area are relatively small and shallow. Petschko et al. (~~2015~~2016) investigated landslide sizes for the districts Amstetten, Waidhofen/Ybbs and Baden and found a median size of landslides of the slide-type movement of 787 m² for the Rheno-Danubian Flysch Zone, 1,189 m² for the Molasse Zone and 1,260 m² for Quaternary sediments. Based on an analysis of landslide archive entries (i.e. Building Ground Registry), Bell et al. (2014a) estimated a median landslide depth of 1.7 m (mean 2.2 m) for 142 landslides recorded for the district Waidhofen/Ybbs.

15 **3. Data**

3.1 Landslide inventory

The historical landslide inventory used for this study represents a subsample of landslides mapped by Petschko et al. (~~2015~~2016). The original data set consists of 13,166 point features which represent the initiation zone of landslides of the slide-type movement (one point per landslide) and was compiled by analyzing mainly topographic derivatives of ~~an~~a 1 m x 1 m airborne laser scanning (ALS) digital terrain model (DTM; ~~1 m x 1m~~flight campaign: 2006-2009), supported by interpreting two orthophotos~~.~~ (flight campaigns: 2000-2004 and 2007-2008). The point-inventory was generated for susceptibility assessments while also several other studies endorse to use one point per landslide observation (Atkinson et al., 1998; Goetz et al., 2015a; Petschko et al., 2014a; Van Den Eeckhaut et al., 2006). Since the respective points stand representative for the main scarp location (cf. Petschko et al., 2016), subsequent landslide susceptibility maps provide an estimate on where landslides are more likely to be initiated. The sample used for this study consists of 591 landslides. Further information on the mapping of this inventory and its related advantages and disadvantages are described by Petschko et al. (~~2015~~2016).

3.2 Predictor variables

The number and type of predictors used within statistical landslide susceptibility analyses varies greatly depending on the scope of the study, the type of the investigated landslides, the characteristics of the area and data availability (Guzzetti et al., 1999; ~~van~~Van Westen et al., 2008). To enhance traceability of modelling results, we opted to include only few but widely

used predictors within the analyses. All models were generated with a predictor set consisting of a lithology layer (Fig. ~~1B~~1c) and three topographic variables, namely slope (Fig. ~~1C~~1b) and two variables representing slope orientation (Fig. ~~1D,E~~1d,e).

The inclination of a slope is directly related to shear stresses and therefore almost always used as a predictor within statistical landslide susceptibility modelling. Slope aspect is regularly introduced as a proxy for a varying degree of insolation and evapotranspiration influencing moisture and weathering conditions or, in some cases, as a representative for the dipping of geologic structures (Gorsevski et al., 2006; ~~van~~Van Westen et al., 2008). The two slope azimuth predictors were directly derived from a resampled (10 m x 10 m) ALS-based DTM. Classification of the continuously scaled aspect layer (originally ranging from 0° to 360°) was avoided by calculating its cosine (representing the degree of north-exposedness) and sine (east-exposedness) (Brenning, 2009, 2012a). This transformation ensured that the applied linear models treated similarly oriented slopes (e.g. 1° vs. 360°) similarly. At regional scales, the parent material of the soils is usually represented by lithological maps (Gorsevski et al., 2006; ~~van~~Van Westen et al., 2008). This information was obtained from a digital geological map of Lower Austria (GK200) available at a scale of 1:200,000 and resampled to the modelling resolution of 10 m x 10 m.

We decided to exclude land cover from the analysis for two reasons. Firstly, the available recent land cover data does not necessarily correspond to the land cover present at the time of landslide occurrence, due to constantly ongoing land use changes and the fact that landslide age is expected to vary substantially within the present inventory (Petschko et al., 2014b; ~~van~~Van Westen et al., 2008). Secondly, an omission of land cover as a predictor was expected to reduce the chances that the suspected land cover related-bias (e.g. overrepresentation of landslides in forested areas, cf. Bell et al., 2012; Petschko et al., ~~2015~~2016) is directly propagated into the final results (cf. Steger et al., 2016)~~.~~

. However, land cover was introduced as a predictor within the synthetic data set, because the respective landslide data set was not defined to be affected by a systematic error, while land cover was specified to be static in time (cf. Sect. 4.6).

4. Methods

In this study we aimed to analyze all data in a tangible and traceable way. Therefore, we avoided using less interpretable classifiers or a high number of predictors. For classification we opted for the well-established logistic generalized linear model, also known as logistic regression (cf. Sect. 4.2), while we included only a small number of frequently applied predictors (cf. Sect. 3.2). The general methodological framework consisted of an artificial introduction of positional inaccuracies into the present landslide inventory (cf. Sect. 4.1) and a subsequent in-depth evaluation of modelling results. The final response variable included within each model was based on an equal number of landslide presence-observations (landslide inventory) and randomly sampled absence-observations (1:1 sampling) (Goetz et al., 2015a; Heckmann et al., 2014). All analyses were additionally performed with synthetic data in order to further verify the observations under rather controlled conditions (cf. Sect. 4.6).

4.1 Artificial introduction of positional errors into the landslide inventory

Inventory-based positional errors were introduced by artificially changing the original landslide scarp position according to a normal distribution in a random direction (Fig. 2). The mean distance to the original position was specified as 5, 10, 20, 50 and 120 m (Fig. 2C2c). Smaller positional inaccuracies (e.g. 5 or 10 m) may be present when digitizing landslide inventories mapped through interpretation of aerial photographs (Santangelo et al., 2015) while the highest value (120 m) was defined according to an analysis made for the Flysch Zone. InWithin this caseanalysis, 65 of the 681 analyzed landslide database entries of the Building Ground Registry (more information refer to Schwenk, 1992 and Steger et al., 2016) contained a quantitative estimate on the positional accuracy of the respective reported landslide (point location. The derived mean positional error: of 120 m, (standard deviation: of 84 m).) was directly adopted to specify the inaccuracy of the most erroneous inventory.

4.2 Logistic regression and odds ratios

Logistic regression is probably the most widely used classifier to generate landslide susceptibility maps (Wu et al. 2015). Many studies confirmed its high suitability for the analysis of landslide susceptibility due to its ability to predict a binary response variable (absence and presence of landslides) using both, continuously and categorically scaled predictors while ensuring a high generalizability, interpretability and smooth prediction surfaces (Atkinson and Massari, 1998; Felicísimo et al., 2013; Goetz et al., 2015a; Steger et al., 2016; Van Den Eeckhaut et al., 2006). Logistic regression models of this study were fitted by using the R package "stats" (R Core Team, 2014). The probability of a landslide occurring $P(Y = 1)$ was predicted by:

$$logit(P(Y = 1)) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{1}$$

in which, β_0 represents the intercept and $\beta_1 \ldots \beta_p$ the regression coefficients of predictors $X_1 \ldots X_p$. The associations between predictors and the response within multiple variable regression models can be expressed as odds ratios (OR). In contrast to an interpretation of probabilities, OR allow to express this relation with a single number while accounting for the influence of other predictors (Brenning et al., 2015; Hosmer and Lemeshow, 2000). For instance, OR estimated for the predictor lithology display the estimated chances that a certain lithological unit is affected by landsliding while accounting for the possible confounding effect of slope angle (Steger et al., 2016).

The odds of an event occurring (in this case a landslide) are defined as the probability of this event occurring $P(Y = 1)$, divided by the probability that this event is not occurring $P(Y = 0)$. The ratio between the odds after a one-unit increase in the predictor X_p and the original odds is referred to as odds ratio:

$$OR(X_p) = \frac{Odds(X_p+1)}{Odds(X_p)} = \frac{Odds \; after \; a \; one \; unit \; increase \; in \; the \; predictor}{Original \; odds} \tag{2}$$

For logistic regression based models, OR can be directly derived from exponentiated regression coefficients ($\beta_1 \dots \beta_p$). OR of > 1 point to a positive relation between a continuously scaled predictor and landslide occurrence while ORs of < 1 express a negative association. Within this study, OR obtained for all continuously scaled predictors were estimated for "meaningful" increments (Brenning, 2012a; Brenning et al., 2015; Hosmer and Lemeshow, 2000). For example, ORs presented for the predictor slope angle refer to the changes in the odds for a slope angle increase of $10°$ ($X_p + 10$). OR estimated for categorically scaled predictors have to be interpreted in relation to a specified reference level.

4.3 Comparison with reference models

An ~~expert-based~~additional evaluation of the final results was conducted by comparing all modelled relationships and maps with the results ~~of those models that~~obtained by the reference models that were generated with the original (i.e. unmodified) inventory. These references were assumed to be less affected (i.e. reference model for the real data) or unaffected (i.e. reference for the synthetic data) by inventory-based positional errors. ~~Consequently, we determined that~~We considered a model ~~and the consequent~~or map ~~may~~to be strongly affected by an inventory-based positional inaccuracy~~, whenever~~ if the ~~underlying~~ modelled ~~associations~~relationships (i.e. OR of predictors) respectively the spatial pattern of the maps differed substantially from ~~the respective reference models~~their previously defined references. We therefore considered the respective positional error to have little effect in all cases where the estimated OR and the susceptibility maps were similar to their references.

4.4 Assessment of the predictive performance

~~The prediction skills of the present models were estimated by calculating the AUROC by applying two partitioning techniques, implemented in the R package "sperrorest" (Brenning, 2012b), namely k-fold cross validation (CV) and k-fold spatial cross validation (SCV). CV consists of a repeated non-spatial random partitioning of the original sample into k subsamples. Within each iteration, a performance measure (e.g. AUROC) is estimated for one of the k subsamples while the remaining sample (k-1) is used to fit the model. In contrast to CV, the partitioning within SCV is performed spatially and consists of a repeated spatial partitioning of the training sample and test sample (Ruß and Brenning, 2010). In the context of landslide susceptibility modelling, SCV can be applied to estimate the spatial transferability of modelling results (Goetz et al., 2015a; Petschko et al., 2014a). In this study, CV and SCV were repeated 50 times with 10 folds per repetition.~~

The prediction skills of the models were estimated by calculating the AUROC by applying two partitioning techniques, implemented in the R package "sperrorest" (Brenning, 2012b), namely k-fold cross validation (CV) and k-fold spatial cross validation (SCV). In contrast to single hold-out validation, CV and SCV are not based on one single split of the training and test sample (e.g. 80% for calibration and 20% for validation), but on a repeated partitioning of the original sample into k subsamples. In each iteration, a performance measure (e.g. AUROC) is estimated for one of the k subsamples, while the remaining (k-1) subsamples are combined into a training set that is used to calibrate the model. Thus, validation results that

are based on CV and SCV are not dependent on one specific sample split. In fact, CV as well as SCV allow that all available data can be used to validate and to calibrate the final models. CV is based on a repeated non-spatial random splitting, whereas SCV is performed spatially and consists of a repeated spatial partitioning of the training sample and test sample (Brenning, 2012b; Ruß and Brenning, 2010). In this study, the predictive performance of all models was estimated by repeating CV and SCV 50 times with 10 folds per repetition. More specifically, within each of the 50 repetitions, each observation of the response variable was applied nine times to calibrate the model and once to test the predictive performance. The presented AUROC values refer to the median of these 500 estimates. Previous studies emphasized the suitability of CV and SCV in the context of landslide susceptibility modelling (Goetz et al., 2015a; Petschko et al., 2014a; Steger et al. 2016).

Furthermore, ~~the AUROC~~an additional validation strategy was ~~also used~~applied in order to quantitatively evaluate and compare the ability of all models to ~~differentiate~~"predict" landslide presences and landslide absences of the unmodified response ~~variables (=~~variable. For this purpose, the AUROC was used to compare the predictions of each model with the unmodified landslide inventory (i.e. unaffected by an artificially introduced positional error). This metric relates to the goodness of model fit ~~for~~, whenever the respective models ~~generated~~were calibrated with ~~the~~an unmodified data set~~)~~.

4.5 Estimation of predictor importance

Permutation-based variable importance assessments assume that the importance of a specific predictor is directly related to the decrease in classification accuracy or predictive performance after randomly reordering (i.e. permuting) the values of this predictor (Strobl et al., 2007). In landslide susceptibility modelling, permutation-based variable importance was previously applied by Goetz et al. (2015a).

In this study, a predictor importance evaluation was conducted to assess whether and how the apparent importance of a specific predictor changes when the positional accuracy of the inventory changes. Thus, the AUROC decrease of each predictor was assessed on non-spatial (CV) and spatial cross-validation partitions (SCV) using the R package "sperrorest" (Brenning, 2012b; Ruß and Brenning, 2010). Herewith, each predictor was permuted 50 times within each fold leading to a total number of 25,000 permutations per predictor (50 repetitions times 10 folds times 50 permutations) within each partitioning technique (CV and SCV) for each of the 12 models.

4.6 Generation of synthetic data

None of the susceptibility models generated for the present study area can be considered to reflect a true and unbiased relation between landslides and environmental conditions, especially due to the unavailability of perfect and accurate spatial information (i.e. landslide data and predictors) and the inherent subjectivity involved during model construction (e.g. predictor and classifier selection). Thus, the conducted comparison of a reference model with models generated with erroneous inventories may only provide an indication of the effect of inventory-based inaccuracies on modelling results. Thus, all analyses were additionally performed with synthetically generated data. This procedure allowed us to define a

"true" association between landslides and their predisposing factors while further controlling for environmental conditions (e.g. spatial distribution and interrelations between predictors) and model specific parameters (e.g. sample size).

The preparation of synthetic data consisted of three major steps: (i) designing the environmental conditions of the study area (= predictors), (ii) specifying a "true" relation between landslides and predictors and (iii) simulating a landslide distribution
5   within the study area according to the predefined "true" relation.

The areal extent of the synthetic study site (20 km x 5 km) was adopted from the Lower Austrian study area while a topographically diverse terrain was generated by generalizing an already available DTM (cf. 3D image in Fig. 3A3a). The topographic predictors slope (Fig. 3A3a) and both aspect layers (Fig. 3B,C3b,c) were directly derived from this smoothed DTM, which was expected to ~~represents~~be unaffected by local data noise and to represent undisturbed pre-failure conditions
10   (cf. Van Den Eeckhaut et al., 2006). Lithology and land cover layers were generated by systematically intersecting and reclassifying randomly generated raster files produced at different spatial scales. The resulting ~~pixilated~~noisy appearing grid files were smoothed and resampled (Fig. 3D,E3d,e).

The three lithological classes (33 km² each) were positioned in the Western (unit A) and Eastern half (unit B) of the study site, while unit C was similarly spread across the area (Fig. 3D3d). The spatial distribution of land cover classes (33 km²
15   each) was conditioned on the spatial distribution of slope inclinations as regularly observed within European regions (Rickli et al., 2002; Steger et al., 2016; Van Den Eeckhaut et al., 2006). More specifically, we defined that the flattest areas (1st slope tercile) consist of arable lands (66%) and pastures (33%) while the steepest parts (3rd slope tercile) are dominated by forests (66%) and pastures (33%). Only medium inclined slopes (2nd slope tercile) were equally covered by all three land cover classes. Landslide susceptibility was defined to be solely dependent on five factors, namely slope, northness, eastness,
20   lithology and land cover (Tab. 1). For the definition of the effect of topographic variations, we oriented ourselves on observed associations for the Lower Austrian study area. Pastures and Arable lands were defined to be equally prone to landsliding (OR: 1) while the stabilizing effects of forests on shallow landslide activity was expected to decrease the chances of landsliding (OR: 0.5). For the lithological units we specified that the units A and B should be equally prone to landsliding (OR: 1) while unit C was defined to be less susceptible (OR: 0.5).

25   Relationships depicted in Table 1 (first row) were back-transformed to regression coefficients while a rare events correction of the intercept (King and Zeng, 2001) was conducted to consider the low portion of landslide-presences in the area (~ 0.002 % of the cells were envisaged to represent landslide initiation) in relation to the 50 % of presence-data in the final response variable (1:1 sampling strategy). The resulting logistic regression equation was then transferred to each raster cell of the area to obtain a probability raster expressing the predefined associations (Fig. ~~3F). A spatially balanced landslide data set was~~
30   ~~generated according to Theobald et al. (2007). This procedure allowed us to favor particular sampling locations according to an inclusion probability raster. The resulting sample (n = 2000; Fig. 3F~~3f). A sampling approach developed by Theobald et al. (2007) was adopted to spatially distribute landslide initiation zones (i.e. represented by points) according to the predefined relationships. More precisely, the mentioned probability raster was used to control sampling intensity during the generation of 2000 spatially balanced landslide initiation points (i.e. raster cells with high probabilities are more likely

9

selected as landslide location) (Theobald et al., 2007). This comparably high number of landslide points was chosen to assure a high explanatory power of the empirical results while simultaneously assuring computational feasibility. The resulting sample (Fig. 3f) was considered to represent an unbiased and accurate landslide inventory and used as presence data within all previously mentioned analyses (cf. Sect. 4.1 to 4.5). The similarity of prescribed relationships (OR in the first row of Tab. 1) and model estimates (OR in the second row of Tab. 1) indicated that the previously established associations were successfully describable by the logistic regression model fitted with artificially generated data sets (i.e. response variable and predictors). Thus, we further considered this model as a useful reference that reflects the "true" relation between the response and the respective environmental conditions.

5. Results

A first inspection of real and synthetic data sets revealed that landslides were more frequent on flatter slopes and less frequent on steeper slopes with a decreasing positional accuracy of the respective inventory (Fig. 2B, 3G2b, 3g). On average, slope angles of the original landslide locations were 4.1° (real data) and 4.6° (synthetic data) higher compared to the most inaccurate inventory (mean error: 120 m). Landslide densities within lithological units changed comparably little when introducing the highest positional error (mean 120 m) into the inventories. Landslide densities within the Flysch decreased from 7.2 to 7 landslides per km² while landslide densities observed for the Molasse remained unchanged (1 landslide per km²). An increasing landslide density was detected for the Quaternary unit by introducing a mean positional error of 120 m (from 0.4 to 0.7 landslides per km²). The following changes in landslide densities were observed for the lithological units of the synthetically generated data set by introducing a high positional error (mean: 120 m): The number of landslides per km² for unit A (respectively B) decreased from 33.3 (B: 14.1) to 31.4 (B: 12.7), while landslide density within the unit C increased from 13.3 to 14.9.

5.1 Modelled relationships

Modelled relationships obtained for the predictors slope angle, northness and eastness (for both data sets) and land cover (only synthetic data) indicated that the strength of association with landslide occurrence generally decreased with an increasing positional error of the inventory since the respective OR tended to approach the neutral value of 1 (Fig. 4A,C,D,F,G4a,c,d,f,g). Only the predictor lithology (Fig. 4B,E4b,e) did not show this trend.

Modelled associations remained nearly unchanged when simulating the smallest positional error (mean: 5 m) (Fig. 4). OR changes became noticeable for a simulated mean positional error of 10 m (real data) respectively 50 m (synthetic data; compare OR with the dashed line in Fig. 4). Especially, OR of the predictors slope (Fig. 4A,D4a,d) and lithology (Fig. 4B,E4b,e) provided quantitative evidence that the models generated with real data exhibit a higher sensitivity to medium positional errors of the inventory (e.g. mean: 10 m, 20 m) compared to the apparently more robust synthetic models.

OR further exposed that a decreasing accuracy of landslide inventories was related to decreasing sensitivity of modelling results on slope angle (Fig. ~~4A,D~~4a,d). For instance, the susceptibility model generated with real data and with the unmodified inventory showed that the chances of a 10° steeper slope to be affected by future landsliding are 7.3 times higher compared to its 10° flatter counterparts. The respective model generated with the most inaccurate inventory (mean error: 120 m) predicted that the chances of landslide occurrence solely rises by a factor of 1.9 for such a slope angle increase (Fig. ~~4A~~4a). Nevertheless, each produced model still predicted a positive relation (OR > 1) between landslide occurrence and the inclination of a slope. OR obtained for the lithology layers revealed major changes for the Molasse (real data) respectively the lithological unit B (synthetic data) (Fig. ~~4B,E~~4b,e). The chances that those units will be affected by future landslides were modelled to be considerably lower (i.e. most distant from the threshold of 1) whenever the respective model was generated with the most inaccurate data sets.

5.2 Visual appearance of maps

A visual inspection of landslide susceptibility maps generally confirmed that the respective modelled relationships were visually recognizable in the predicted susceptibility patterns. Thus, ~~an increasing~~a high positional error of the inventory (mean error: 120 m) was observed to result in susceptibility maps that were less influenced by local ~~topographic~~slope angle and slope aspect variations. Ultimately, this led to more uniformly appearing susceptibility patterns at slope scale (e.g. Fig. ~~5F~~5f). Flatter slopes were generally predicted as more susceptible and steepest slopes as less susceptible with an increasing positional inventory-based error (Fig. 5). The declining effect of local slope variations was accompanied by an increasing visibility of lithological transitions. Especially the model generated with real data and the most inaccurate inventory (Fig. ~~5F~~5f) depicted this tendency by strongly accentuating the silhouettes of the Flysch Zone (cf. Fig. ~~1B~~1c).

5.3 Predictive performance

Predictive performances decreased with a decreasing positional accuracy of the inventories (CV and SCV in Fig. 5). However, models generated with a mean inventory-based error of 50 m still produced AUROCs of 0.76 (CV; real data) respectively 0.83 (synthetic data). Non-spatially assessed median AUROCs were observed to decrease from 0.85 (real data) and 0.86 (synthetic data) to 0.69 (real data) and 0.75 (synthetic data) (CV in Fig. 5). The general trend of decreasing prediction skills with an increasing positional error was exposed by both partitioning techniques, namely CV and SCV. In analogy to observations of the modelled relationships, we observed a first distinct performance drop (AUROC decrease of > 0.1) for the models generated with an inventory-based positional mean error of 10 m (real data) respectively 50 m (synthetic data).

A comparison of predicted susceptibility patterns with the respective unmodified response variable revealed that all models performed well to forecast the original landslide position ("0" in Fig. 5). It was remarkable that models generated with highly inaccurate inventories (mean error: 120 m) achieved AUROCs of 0.84 (real data) and 0.86 (synthetic data).

11

5.4 Predictor importance

The findings provided quantitative indications that the results of variable importance assessments obtained by multiple variable statistical landslide susceptibility models are not only related to underlying geomorphic processes, but also to positional errors inherent in the underlying landslide inventories. The general ranking of predictors within the models generated with real data changed first for a mean positional accuracy of 10 m (CV) respectively 5 m (SCV) (Fig. ~~6A,C~~6a,c). In contrast, the ranking of predictors remained unchanged for the models generated with synthetic data sets up to a mean positional error of 120 m (Fig. ~~6B,D~~6b,d).

Slope had the greatest variable importance in all situations, revealing that this environmental factor played the central role to predict landslide susceptibility within each model. However, an interpretation of AUROC decreases (Fig. 6) also suggested that the dominance of this predictor diminished as positional error increased (e.g. from 0.25 to 0.08 in Fig. ~~6A~~6a). An opposite, but less pronounced tendency was observed for the predictor lithology (e.g. from 0.01 to 0.05 in Fig. ~~6A~~6a). The importance of the predictors slope (CV: 0.08; SCV: 0.07) and lithology (CV: 0.05; SCV: 0.06) was similar whenever the models were generated with the most inaccurate inventory and with real data.

6. Discussion

In general, our results were in agreement with studies stating that inventory-based errors modify the results of a landslide susceptibility analysis (Ardizzone et al., 2002; Fressard et al., 2014; Galli et al., 2008). We observed that an increasing positional error of landslide locations was generally related to an increasing distortion of subsequent modelling results. Thus, this study provides quantitative evidence for the statement that the explanatory power of a landslide susceptibility analysis increases with an increasing quality of the underlying landslide data set (Blahut et al., 2010; Fressard et al., 2014; Galli et al., 2008; Guzzetti et al., 2006; Harp et al., 2011; Petschko et al., ~~2015~~2016; Steger et al., 2016; ~~van~~Van Westen et al., 2008). However, the systematic comparisons also suggested that interdependencies between inventory-based errors and subsequent modelling and validation results are complex. Thus, an identification of a generally valid threshold (i.e. the positional error should not exceed x meters), which separates acceptable models from unacceptable ones, was considered to be inappropriate. The findings provided valuable indications that the propagation of inventory-based errors is not only determined by the degree of positional inaccuracy inherent in a landslide data set, but also by a combination of additional factors. Those aspects include (i) the spatial representation of landslides and the environment (cf. Sect. 6.2), (ii) landslide sizes (cf. Sect. 6.2), (iii) the characteristics of the study area (cf. Sect. 6.3), (iv) the selection ~~respectively parameterization~~ of a classification method (cf. Sect. 6.4) and (v) an interplay of predictors within multiple variable models (cf. Sect. 6.5).

6.1 Usefulness of an additional modelling with synthetic data

The spatial information available for this study was, as presumably every "real" data set used within regional landslide susceptibility studies, considered to be imperfect. Apart from the assumption that the present inventory was affected by a

land-cover related incompleteness (Petschko et al., ~~2015~~2016), also the available environmental data was not expected to ~~flawlessly~~ represent the full spectrum of landslide predisposing factors for the area. Thus, statements on how a specific data modification (in this case: using positionally-inaccurate inventories) reduced or increased the respective model's quality should be treated with caution, especially because increasing predictive performances do not necessarily reflect an increasing

5 quality of a landslide susceptibility model (Lobo et al., 2008; Steger et al., 2016). In this respect, synthetically generated data allowed us to define a reference model that depicted a "true" association between landslides and the environment. Subsequent modelling proved valuable to gain deeper insights into the effect of inventory-based errors on modelling results and cross-check our findings. Furthermore, observed differences between modelling results obtained from the real data sets and the synthetically generated data (e.g. differences in OR-drops; cf. Fig. 4) were helpful to move our attention to related

10 important model-influencing aspects (e.g. characteristics of the study area; cf. Sect. 6.3).

6.2 The spatial representation of input data and landslides size

The available landslide point inventory represents the scarp locations of mainly smaller landslide features, which tend to exhibit a distinct steep and concave morphology in comparison to its surroundings (Petschko et al., ~~2015~~2016).

15 Consequently, already small changes in the point position (e.g. 5 m) may lead to the tendency that the respective features are, in the case of high-resolution slope data (e.g. 1 m x 1 m), displaced into their usually flatter vicinities. However, those immediate neighbourhoods may be represented by identical lithologies, land cover classes or soil types due to a typically coarser spatial resolution of the underlying data sets (Cascini, 2008; ~~van~~Van Westen et al., 2008).

We argue that the applied modelling resolution of input data might not only affect the relative importance of predictors

20 within a landslide susceptibility model (Catani et al., 2013), but also how spatial inventory-based errors are propagated into the final modelling results. For instance, the resampling of the present DTM from 1 m x 1m to the modelling resolution of 10 m x 10 m was related to an enlargement of the area covered by one grid cell (i.e. 1 m² to 100 m²) and a generalization of the topography. This combination increased the chances that nearby observations (e.g. 5 m distance from original location) were represented by an identical grid cell or cells with similar values (e.g. of slope) to the respective original location. Therefore,

25 slope angles of landslide locations (real data) derived from a 10 m-raster were on average just 0.2° lower for a mean inventory-based inaccuracy of 5 m compared to the unmodified inventory. These similarities were expected to mainly cause the observed effect that an average inventory-based error of 5 m did almost not affect modelled relationships (Fig. 4). Thus, we argue that a resampling of predictors to a coarser resolution might not only help to reduce the impact of local topographic variations related to post-failure conditions (conditions after landslide occurrence) or data noisiness (Petschko et al., 2014a;

30 Van Den Eeckhaut et al., 2006; ~~van~~Van Westen et al., 2008), but also to reduce the impact of smaller inventory-based positional errors on the results of a grid-based statistical landslide susceptibility model.

A very high mean positional error of an inventory (e.g. 120 m) may require a correspondingly higher generalization of the ~~topography~~topographic variables, such as slope angle or slope aspect, to enhance the chances that the respective landslide

observation is still represented by an identical respectively similar grid cell. However, those high simplifications may as well lead to the effect that actual important landslide-influencing morphological features disappear within the derivatives of a strongly resampled terrain model.

The chances that an inaccurately mapped landslide-point may still be located within the boundaries of a corresponding landslide may as well increase with an increasing size of the respective landslides. However, statistically predicting landslide susceptibility for mainly larger events may nonetheless require a substantial generalization of a grid-based topography, because the respective pre-failure conditions may considerably differ from the conditions after landslide occurrence (Hussin et al., 2016; Süzen Lüfti and Doyuran, 2004; Van Den Eeckhaut et al., 2006). Again, corresponding simplifications (i.e. resampling) might result in topographical predictors of little practical use. Such discrepancies (i.e. between pre-failure and post-failure topographies) may be counteracted for accurately mapped inventories by approximating the terrain before landslide occurrence (e.g. Van Den Eeckhaut et al., 2006). However, such approaches are likely to be inapplicable whenever only positionally erroneous inventories are available.

We finally recommend that an expected small to moderate positional inaccuracy inherent in a landslide inventory (e.g. 5 to 20 m) might partly be counteracted by resampling the respective grid-based data sets to a spatial resolution higher than the expected mean positional inaccuracy. An alternative spatial representation of the environment (e.g. terrain units, first-order catchments) (Alvioli et al., 2016; Bell et al., ~~2014~~2014b; Guzzetti et al., 1999) might be most expedient whenever modelling with highly inaccurate landslide inventories or erroneous inventories that mainly consist of large events is envisaged. The associated high generalization of environmental conditions within terrain-unit based models might be one reason for earlier observations that susceptibility maps generated from inventories with considerable positional disagreements appeared and performed similarly (Ardizzone et al., 2002). In this respect, recent developments which allow an automatic, but controllable (e.g. sizes of units) subdivisions of the terrain into slope units (Alvioli et al., 2016) might prove highly useful. It is, however, important to ensure that the respective landslides are still located within the unit where they were initiated (Bell et al., ~~2014~~2014b). Which approach to apply also depends on the ultimate aim of the final susceptibility maps, e.g. either to be used in spatial planning strategies or as part of regional landslide early warning systems.

6.3 The characteristics of the study area

For the real data, the modelled relationship between landslide occurrence and topographic characteristics weakened for positional inventory errors > 10 m (Fig. ~~4A,C~~4a,c). This was attributed to the effect of positional error reducing observed topographic differences between landslide and non-landslide locations (e.g. slope angles in Fig. ~~2B~~2b). In the more strongly generalized synthetic area, in contrast, such changes in modelled relationships were hardly noticeable up to a mean positional error of 50 m (3D view in Fig. ~~3A~~3a; Fig. ~~4D~~4d). The differences between the synthetic data and the real data were interpreted as evidence that the geomorphic characteristics of a study area plays as well a major role on how inventory-based inaccuracies may be propagated into modelling results. We inferred that as long as the resolution of the data is generalized to mainly present pre-failure conditions, grid-based models generated for undulating and smooth landscapes (e.g.

14

Flysch Zone of Lower Austria) might tend to be less affected by smaller positional errors of the inventory compared to sites featuring sharp topographic transitions (e.g. terraced landscapes, cuesta landscapes, limestone areas). Based on the results we inferred that the influence of an identical inventory-based mean positional error (e.g. 20 m) on subsequent modelling results may be considerably higher for areas exhibiting a topographically more complex terrain. This complexity is on one hand
5   dependent on how a study site is represented (cf. Sect. 6.2), but also on the characteristics of the study area itself.

6.4 An additional argument on why model overfitting should be avoided

An aspect worth further consideration is whether different classification techniques are to a different extent affected by positional inaccuracies of landslide inventories. In particular, non-linear statistical models (e.g. Generalized additive models, multivariate adaptive regression splines) or flexible machine learning techniques (e.g. Support Vector Machines, Decision
10 Trees, Random Forest) are increasingly used because of possible gains in predictive performance (Ballabio and Sterlacchini, 2012; Catani et al., 2013; Felicísimo et al., 2013; Tien Bui et al., 2012), but they bear a higher risk of overfitting to training data (Brenning, 2005, 2012a; Goetz et al., 2015a; Steger et al., 2016). In analogy to a generalized representation of the study area (cf. Sect. 6.2), we expect that a higher degree of model generalization, as provided by the less flexible techniques, may provide a certain level of protection against learning, i.e. overfitting to, errors originating from a landslide inventory.
15 Thus, this study supports the statement that an application of strongly generalizing classifiers might still be valuable in the context of landslide susceptibility modelling (Brenning, 2012a; Goetz et al., 2015a; Steger et al., 2016) while we suspect that strongly overfitting classifiers are likely to reproduce inventory-based errors. This assumption is supported by earlier research, which showed that logistic regression-based landslide susceptibility models generated with rather different landslide inventories produced similar predictive performances and landslide susceptibility maps (Zêzere et al., 2009).
20 Furthermore, Ardizzone et al. (2002) concluded that statistical models may as well substantially minimize the effect of data errors on landslide susceptibility models.

6.5 Considerations of an interplay of predictors within multiple variable models

Logistic regression models, like other multiple variable classifiers, try to maximize the probability of obtaining the observations (landslide presence or landslide absence) according to the data provided (Hosmer and Lemeshow, 2000). The
25 present findings (i.e. OR and predictor importance) provided quantitative indications that the applied models counteracted the diminishing explanatory power of topographic variations (especially slope angle) by adjusting the weights assigned to the coarser scaled predictor lithology.

For instance, the logistic regression models produced with real data and with the unmodified inventory did not strongly accentuate differences in landslide susceptibility (cf. OR in Fig. 4B4b) between the Flysch (7.2 landslides per km²) and the
30 Molasse (1 landslide per km²) since the respective models accounted for the fact that the Flysch Zone is considerably steeper (mean slope 12.3°) compared to the Molasse (5.1°). Even though landslide densities observed for those units remained nearly unchanged when simulating an inventory-based-error (i.e. Flysch 7 and Molasse 1 landslide per km² for the 120 m

15

mean error), modelled relationships increasingly emphasized differences between those units (Fig. 4B4b). Thus, we inferred that the respective models were decreasingly capable to account for local topographic variations (OR of slope are approaching the threshold of 1) with an increasing positional error of the inventory and subsequently were increasingly dependent on differences describable by other predictors (e.g. landslide densities for lithological units). Those tendencies

5 were also discernible for the synthetic data set, ultimately due to deteriorating topographical effects (cf. Fig. 4) and dissimilar landslide densities between lithological units (cf. Sect. 5). Variable importance assessments (e.g. decreasing importance of slope and increasing importance of lithology in Fig. 6) and the appearance of the final maps (e.g. striking appearance of the Flysch in Fig. 5F5f) further exposed these distortions. This study underlines that besides the applied raster resolution (Catani et al., 2013), also the inventory-based positional accuracy influences how strongly a predictor appears to

10 discriminate landslide-presences from landslide-absences. Consequently, we argue that potential inventory-based positional errors should additionally be taken into account, when a process-based geomorphic interpretation of statistical landslide susceptibility models is envisaged (Brenning et al., 2015; Vorpahl et al., 2012).

6.6 Misleading performance estimates

In general we found evidence that a point-based landslide inventory that is increasingly affected by positional errors

15 approaches a distribution of complete spatial randomness (Cressie, 2015). Thus, statistically discriminating landslide locations from non-landslide locations becomes increasingly difficult since non-landslide observations are regularly represented by random spatial observations (e.g. Conoscenti et al., 2016; Goetz et al., 2015a). The observed decreasing predictive performances provided quantitative evidence for this assumption (CV and SCV in Fig. 5). The tendency towards diminishing differences between landslide presences and absences were also reflected by conditional frequencies of slope

20 angles (Fig. 2B2b; Fig. 3G3g) and OR of topographic parameters (Fig. 4).

The spatial pattern of landslide susceptibility maps (e.g. cutouts in Fig. 5) visually highlighted that all models established a positive relationship between landslide occurrence and slope angle (ORs > 1.9 in Fig. 4A,D4a,d) while the lithological units Flysch (real data) respectively the unit A (synthetic data) were consistently predicted as most susceptible (Fig. 4B,E4b,e). Thus, all models predicted that the steepest parts of the Flysch Zone (respectively unit A) are highly prone to landsliding,

25 even though the underlying landslide-presences (white dots in Fig. 5) were increasingly located on considerably flatter slopes (e.g. Fig. 2B2b; Fig. 3G3g). In analogy, the flattest parts covered by Quaternary sediments (respectively unit C) were consistently predicted as the most stable locations. From this perspective, it might become clear that models generated with erroneous inventories performed well (all AUROCs > 0.84) in predicting the original landslide position (cf. "0" in Fig. 5). This observation is consistent with our suggestion that strongly generalizing classifiers might also reduce the effect of

30 inventory-based errors on modelling results (cf. Sect. 6.4; Ardizzone et al., 2002). However, validation results that were based on a landslide subsample of the data set used to generate the model (as within most studies) revealed apparently lower predictive performances (CV, SCV in Fig. 5). Thus, this study provides another example that misleading performance

estimates may follow whenever a landslide susceptibility model is generated and validated with erroneous landslide inventories (Steger et al., 2016).

This study further underlined that the predictive performance of a statistical landslide susceptibility model is just one of many indicators of the quality and reliability of a spatial prediction model (Fressard et al., 2014; Guzzetti et al., 2006; Lobo

5   et al., 2008; Petschko et al., 2014a; Rossi et al., 2010; Steger et al., 2016). Thus, we consider a differentiated evaluation of input data qualities and modelling results (i.e. quantitative and expert-based) as an necessary supplement to get insights into limitations and the reliability of subsequent modelling results (Demoulin and Chung, 2007; Fressard et al., 2014; Guzzetti et al., 2006; Petschko et al., 2014a; Steger et al., 2016).

7. Conclusion and final recommendations

10   There is no doubt that the explanatory power of a landslide susceptibility analysis increases with an increasing quality of the underlying landslide inventory (Fressard et al., 2014; Galli et al., 2008; Guzzetti et al., 2006; Petschko et al., ~~2015~~2016; Steger et al., 2016; ~~van~~Van Westen et al., 2008). The present findings highlighted that positionally erroneous landslide inventories affected modelled relationships, variable importance assessments and the explanatory power of conventional predictive performance estimates of a statistical landslide susceptibility model. We found valuable evidence that the impact

15   of inventory-based positional errors on modelling results is not only dependent on the degree of inaccuracy inherent in the respective landslide-presence data, but also on an interplay of other model-influencing aspects: Besides the size (i.e. small vs. large landslides) and spatial representation of landslides (i.e. landslide scarp vs. landslide body), also the spatial representation of the environment (i.e. raster resolution, terrain unit) was expected to determine how a specific positional error was propagated into the final results. Furthermore, also the morphological characteristics of the study area (i.e.

20   undulating landscape vs. landscape with sharp transitions) and the applied classification method (i.e. strongly generalizing vs. highly flexible classifiers) are believed to control the extent of potentially misleading modelling results.

The present findings indicate that the propagation of inventory-based positional inaccuracies into modelling results can be reduced by selecting an appropriate study design. Thus, we advise to accept inventory-based errors as unavoidable and then seek to obtain in-depth information on the potential limitations of present data sets. A subsequent improvement of the

25   respective data set (e.g. updated mapping) should be the first step. For the likely case that modelling with positionally imperfect inventories cannot be avoided (e.g. Bell et al., ~~2014~~2014b; Eeckhaut et al., 2011), we recommend to generalize input data to a coarser scale (e.g. resampling of predictors or using an alternative representation of the environment; cf. Sect. 6.2) while opting for modelling techniques that simplify observed associations (e.g. logistic regression; cf. Sect. 6.4). The potential presence of inventory-based errors should always be taken into account since subsequent performance estimates

30   (cf. Sect. 6.6), modelled relationships and variable importance (cf. Sect. 6.5) are likely to be distorted whenever the underlying models were generated with positionally erroneous landslide inventories.

17

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F. and Guzzetti, F.: Automatic delineation of geomorphological slope-units and their optimization for landslide susceptibility modelling, Geoscientific Model Development Discussions, 1–33, doi:10.5194/gmd-2016-118, 2016.

10 Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F. and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, Natural Hazards and Earth System Science, 2(1/2), 3–14, 2002.

Atkinson, P., Jiskoot, H., Massari, R. and Murray, T.: Generalized linear modelling in geomorphology, Earth Surface Processes and Landforms, 23(13), 1185–1195, 1998.

Atkinson, P. M. and Massari, R.: Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy,

15 Computers & Geosciences, 24(4), 373–385, 1998.

Ballabio, C. and Sterlacchini, S.: Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy, Mathematical Geosciences, 44(1), 47–70, doi:10.1007/s11004-011-9379-9, 2012.

Beguería, S.: Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management, Natural Hazards, 37(3), 315–329, doi:10.1007/s11069-005-5182-6, 2006.

20 Bell, R., Petschko, H., Röhrs, M. and Dix, A.: Assessment of landslide age, landslide persistence and human impact using airborne laser scanning digital terrain models, Geografiska Annaler: Series A, Physical Geography, 94(1), 135–156, doi:10.1111/j.1468-0459.2012.00454.x, 2012.

Bell, R., Petschko, H., Proske, H., Leopold, P., Heiss, G., Bauer, C., Goetz, J., Granica, K. and Glade T.: Methodenentwicklung zur Gefährdungsmodellierung von Massenbewegungen in Niederösterreich – MoNOE, Final project

25 report, pp. 224, Vienna., 2014a.

Bell, R., Cepeda, J. and Devoli, G.: Landslide susceptibility modeling at catchment level for improvement of the landslide early warning system in Norway., in Conference Proceedings of the World Landslide Forum 3, vol. 4: Discussion Session, Beijing, China., 20142014b.

Blahut, J., vanVan Westen, C. J. and Sterlacchini, S.: Analysis of landslide inventories for accurate prediction of debris-flow

30 source areas, Geomorphology, 119(1-2), 36–51, doi:10.1016/j.geomorph.2010.02.017, 2010.

Brabb, E. E.: Innovative approaches to landslide hazard mapping. 4th International Symposium on Landslides, 16-21 September, pp. 307–324, Toronto., 1984.

Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, Natural Hazards and Earth System Science, 2005.

35 Brenning, A.: Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. Remote Sensing of Environment, 113(1):239-247, Remote Sensing of Environment, 113(1), 239–247, doi:10.1016/j.rse.2008.09.005, 2009.

Brenning, A.: Improved spatial analysis and prediction of landslide susceptibility: Practical recommendations., in Landslides and Engineered Slopes: Protecting Society through Improved Understanding., edited by: Eberhardt, E., Froese, C., Turner,

40 A. K., and Leroueil, S., Tylor and Francis, Banff, Alberta, Canada., pp. 789–795, 2012a.

Brenning, A.: 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. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6352393 (last access 22 April 2016), 20122012b.

Brenning, A., Schwinn, M., Ruiz-Páez, A. P. and Muenchow, J.: 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(1), 45–57, doi:10.5194/nhess-15-45-2015, 2015.

Cascini, L.: Applicability of landslide susceptibility and hazard zoning at different scales, Engineering Geology, 102(3-4), 164–177, doi:10.1016/j.enggeo.2008.03.016, 2008.

Catani, F., Lagomarsino, D., Segoni, S. and Tofani, V.: Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues, Natural Hazards and Earth System Science, 13(11), 2815–2831, doi:10.5194/nhess-13-2815-2013, 2013.

Chung, C.-J. F. and Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping, Natural Hazards, 30(3), 451–472, 2003.

Conoscenti, C., Rotigliano, E., Cama, M., Caraballo-Arias, N. A., Lombardo, L. and Agnesi, V.: Exploring the effect of absence selection on landslide susceptibility models: A case study in Sicily, Italy, Geomorphology, 261, 222–235, 2016.

Corominas, J., ~~van~~Van Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J. and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, Bulletin of Engineering Geology and the Environment, doi:10.1007/s10064-013-0538-8, 2013.

Cressie, N.: Statistics for spatial data, John Wiley & Sons., 2015.

Demoulin, A. and Chung, C.-J. F.: Mapping landslide susceptibility from small datasets: A case study in the Pays de Herve (E Belgium), Geomorphology, 89(3-4), 391–404, doi:10.1016/j.geomorph.2007.01.008, 2007.

Eder, A., Sotier, B., Klebinder, K., Sturmlechner, R., Dorner, J., Markart, G., Schmid, G. and Strauss, P.: Hydrologische Bodenkenndaten der Böden Niederösterreichs - Endbericht, HydroBodNÖ. BAW, BFW, 2011.

Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J.-P., Montanarella, L. and Nadim, F.: Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data, Landslides, 9(3), 357–369, doi:10.1007/s10346-011-0299-z, 2011.

Ermini, L., Catani, F. and Casagli, N.: Artificial Neural Networks applied to landslide susceptibility assessment, Geomorphology, 66(1-4), 327–343, doi:10.1016/j.geomorph.2004.09.025, 2005.

Fabbri, A. G., Chung, C.-J. F., Cendrero, A. and Remondo, J.: Is prediction of future landslides possible with a GIS?, Natural Hazards, 30(3), 487–503, 2003.

Felicísimo, Á. M., Cuartero, A., Remondo, J. and Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10(2), 175–189, doi:10.1007/s10346-012-0320-1, 2013.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. and Savage, W. Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land use planning, Engineering Geology, 102(3-4), 85–98, doi:10.1016/j.enggeo.2008.03.022, 2008.

Frattini, P., Crosta, G. and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, Engineering Geology, 111(1-4), 62–72, doi:10.1016/j.enggeo.2009.12.004, 2010.

Fressard, M., Thiery, Y. and Maquaire, O.: 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(3), 569–588, doi:10.5194/nhess-14-569-2014, 2014.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F. and Reichenbach, P.: Comparing landslide inventory maps, Geomorphology, 94(3-4), 268–289, doi:10.1016/j.geomorph.2006.09.023, 2008.

Glade, T., Anderson, M. and Crozier, M. J., Eds.: Landslide hazard and risk, J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ., 2005.

Goetz, J. N., Brenning, A., Petschko, H. and Leopold, P.: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling, Computers & Geosciences, 81, 1–11, doi:10.1016/j.cageo.2015.04.007, 2015a.

Goetz, J. N., Guthrie, R. H. and Brenning, A.: Forest harvesting is associated with increased landslide activity during an extreme rainstorm on Vancouver Island, Canada, Natural Hazards and Earth System Science, 15(6), 1311–1330, doi:10.5194/nhess-15-1311-2015, 2015b.

Gorsevski, P. V., Gessler, P. E., Foltz, R. B. and Elliot, W. J.: Spatial prediction of landslide hazard using logistic regression and ROC analysis, Transactions in GIS, 10(3), 395–415, 2006.

19

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, Geomorphology, 31(1-4), 181–216, 1999.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale, Geomorphology, 72(1-4), 272–299, doi:10.1016/j.geomorph.2005.06.002, 2005.

5 Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81(1-2), 166–184, doi:10.1016/j.geomorph.2006.04.007, 2006.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.-T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112(1-2), 42–66, doi:10.1016/j.earscirev.2012.02.001, 2012.

Harp, E. L., Keefer, D. K., Sato, H. P. and Yagi, H.: Landslide inventories: the essential part of seismic landslide hazard
10 analyses, Engineering Geology, 122(1), 9–21, 2011.

Heckmann, T., Gegg, K., Gegg, A. and Becht, M.: Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows, Natural Hazards and Earth System Science, 14(2), 259–278, doi:10.5194/nhess-14-259-2014, 2014.

Hosmer, D. W. and Lemeshow, S.: Applied Logistic Regression, John Wiley & Sons, New York., 2000.

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

Hutton, J.: X. Theory of the Earth; or an Investigation of the Laws observable in the Composition, Dissolution, and Restoration of Land upon the Globe., Earth and Environmental Science Transactions of the Royal Society of Edinburgh,
20 1(02), 209–304, doi:10.1017/S0080456800029227, 1788.

King, G. and Zeng, L.: Logistic regression in rare events data, Political analysis, 9(2), 137–163, 2001.

Lobo, J. M., Jiménez-Valverde, A. and Real, R.: AUC: a misleading measure of the performance of predictive distribution models, Global Ecology and Biogeography, 17(2), 145–151, doi:10.1111/j.1466-8238.2007.00358.x, 2008.

Malamud, B. D., Turcotte, D. L., Guzzetti, F. and Reichenbach, P.: Landslide inventories and their statistical properties,
25 Earth Surface Processes and Landforms, 29(6), 687–711, doi:10.1002/esp.1064, 2004.

Petschko, H., Brenning, A., Bell, R., Goetz, J. and Glade, T.: Assessing the quality of landslide susceptibility maps – case study Lower Austria, Natural Hazards and Earth System Science, 14(1), 95–118, doi:10.5194/nhess-14-95-2014, 2014a.

Petschko, H., Bell, R. and Glade, T.: Relative Age Estimation at Landslide Mapping on LiDAR Derivatives: Revealing the Applicability of Land Cover Data in Statistical Susceptibility Modelling, in Landslide Science for a Safer Geoenvironment,
30 edited by K. Sassa, P. Canuti, and Y. Yin, pp. 337–343, Springer International Publishing. Available from: http://link.springer.com/chapter/10.1007/978-3-319-05050-8_53 (last access 26 July 2016), ~~2014~~2014b.

Petschko, H., Bell, R. and Glade, T.: Effectiveness of visually analyzing LiDAR DTM derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling, Landslides, 13(5), 857–872, doi:10.1007/s10346-015-0622-1, ~~2015~~2016.

35 R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna. Available from: http://www.R-project.org/, 2014.

Remondo, J., González, A., De Terán, J. R. D., Cendrero, A., Fabbri, A. and Chung, C.-J. F.: Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain, Natural Hazards, 30(3), 437–449, 2003.

Rickli, C., Zürcher, K., Frey, W. and Lüscher, P.: Wirkungen des Waldes auf oberflächennahe Rutschprozesse,
40 Schweizerische Zeitschrift für Forstwesen, 153(11), 437–445, 2002.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C. and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114(3), 129–142, doi:10.1016/j.geomorph.2009.06.020, 2010.

Ruß, G. and Brenning, A.: Spatial Variable Importance Assessment for Yield Prediction in Precision Agriculture, in Advances in Intelligent Data Analysis IX, edited by Cohen, P. R., Adams, N. M. and Berthold, M. R., pp. 184–195, Springer
45 Berlin Heidelberg. Available from: http://link.springer.com/chapter/10.1007/978-3-642-13062-5_18 (last access: 12 August 2016), 2010.

Santangelo, M., Marchesini, I., Bucci, F., Cardinali, M., Fiorucci, F. and Guzzetti, F.: An approach to reduce mapping errors in the production of landslide inventory maps, Natural Hazards and Earth System Science, 15(9), 2111–2126, doi:10.5194/nhess-15-2111-2015, 2015.

Schweigl, J. and Hervás, J.: Landslide mapping in Austria, JRC Scientific and Technical Reports, European Commission Joint Research Centre, Institute for Environment and Sustainability, Italy. Available from: http://eusoils.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/other/EUR23785EN.pdf (last access: 20 August 2016), 2009.

Schwenk, H.: Massenbewegungen in Niederösterreich 1953-1990, Geologische Bundesanstalt, Wien., 1992.

5 Skoda, G. and Lorenz, P.: Mittlere Jahresniederschlagshöhe - Modellrechnung mit unkorrigierten Daten, Hydrologischer Atlas Österreich 2007, Erläuterungsblatt 2.2, 2007.

Steger, S., Bell, R., Petschko, H. and Glade, T.: Evaluating the Effect of Modelling Methods and Landslide Inventories Used for Statistical Susceptibility Modelling, in Engineering Geology for Society and Territory - Volume 2, edited by G. Lollino, D. Giordan, G. B. Crosta, J. Corominas, R. Azzam, J. Wasowski, and N. Sciarra, pp. 201–204, Springer International

10 Publishing. Available from: http://link.springer.com/chapter/10.1007/978-3-319-09057-3_27 (last access 16 August 2016), 2015.

Steger, S., Brenning, A., Bell, R., Petschko, H. and Glade, T.: Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps, Geomorphology, 262, 8–23, doi:10.1016/j.geomorph.2016.03.015, 2016.

15 Strobl, C., Boulesteix, A.-L., Zeileis, A. and Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution, BMC Bioinformatics, 8, 25, doi:10.1186/1471-2105-8-25, 2007.

Süzen Lüfti, M. L. and Doyuran, V.: A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate, Environmental Geology, 45(5), 665–679, doi:10.1007/s00254-003-0917-8, 2004.

Theobald, D. M., Stevens, D. L., White, D., Urquhart, N. S., Olsen, A. R. and Norman, J. B.: Using GIS to Generate

20 Spatially Balanced Random Survey Designs for Natural Resource Applications, Environmental Management, 40(1), 134–146, doi:10.1007/s00267-005-0199-x, 2007.

Tien Bui, D., Pradhan, B., Lofman, O. and Revhaug, I.: Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naive Bayes Models, Mathematical Problems in Engineering, 2012, e974638, doi:10.1155/2012/974638, 2012.

25 Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G. and Vandekerckhove, L.: Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium), Geomorphology, 76(3-4), 392–410, doi:10.1016/j.geomorph.2005.12.003, 2006.

Vorpahl, P., Elsenbeer, H., Märker, M. and Schröder, B.: How can statistical models help to determine driving factors of landslides?, Ecological Modelling, 239, 27–39, doi:10.1016/j.ecolmodel.2011.12.007, 2012.

30 Wessely, G., Draxler, I., Gangl, G., Gottschling, P., Heinrich, M., Hofmann, T., Lenhardt, W., Matura, A., Pavuza, R., Peresson, H. and Sauer, R.: Geologie der österreichischen Bundesländer - Niederösterreich, Geologische Bundesanstalt, Wien., 2006.

van Van Westen, C. J., Castellanos, E. and Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, Engineering Geology, 102(3-4), 112–131, doi:10.1016/j.enggeo.2008.03.010, 2008.

35 Wu, X., Chen, X., Zhan, F. B. and Hong, S.: Global research trends in landslides during 1991–2014: a bibliometric analysis, Landslides, 12(6), 1215–1226, doi:10.1007/s10346-015-0624-z, 2015.

Zêzere, J., Henriques, C. S., Garcia, R. A. C. and Piedade, A.: Effects of landslide inventories uncertainity on landslide susceptibility modelling, In book: Landslide Processes: From Geomorphologic Mapping to Dynamic Modelling., Edition: Strasbourg, Publisher: CERG Editions, Editors: Mallet, J.-P.; Remaitre, A.; Boggard, pp.81-86, 2009.
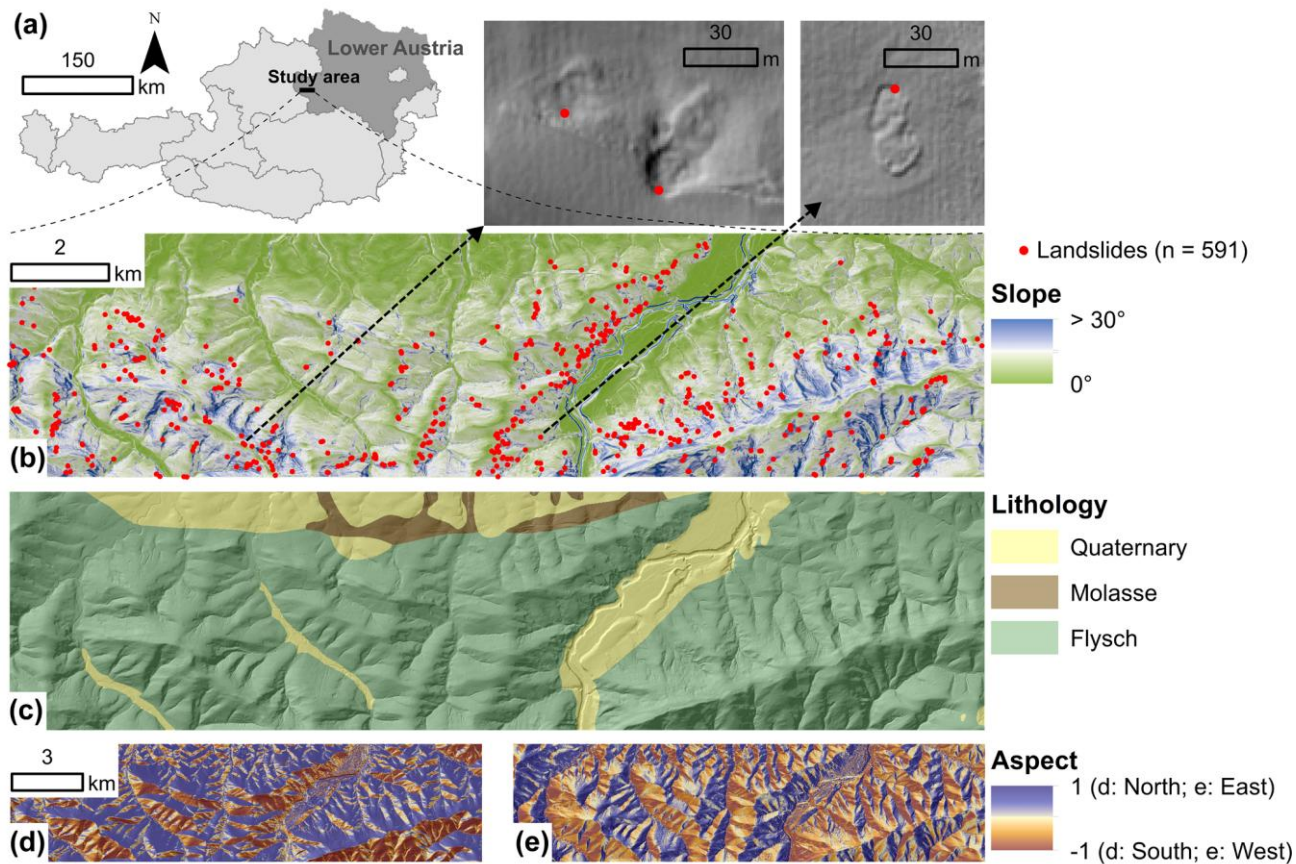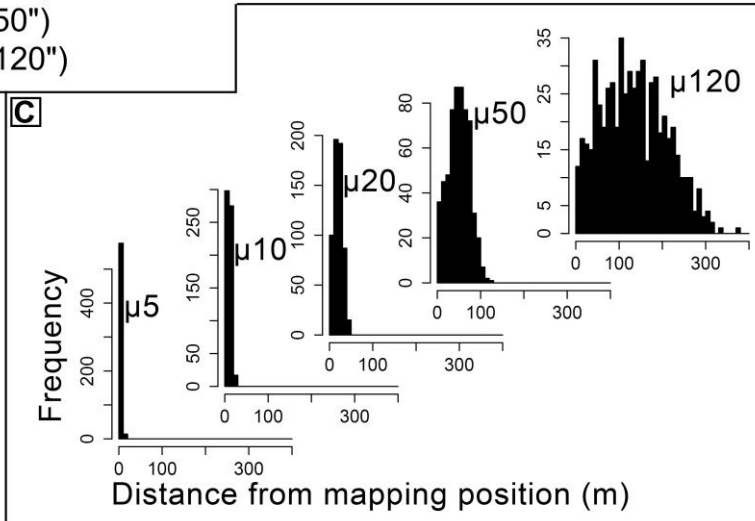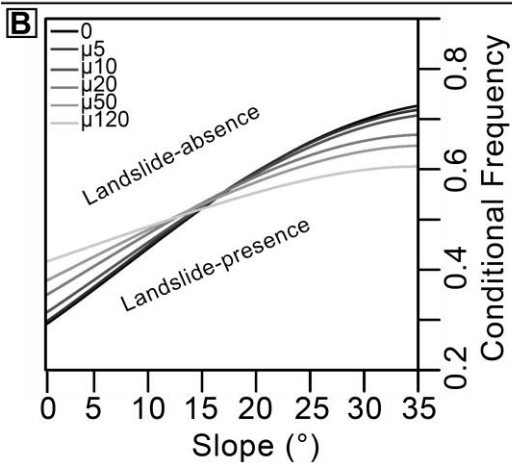
40
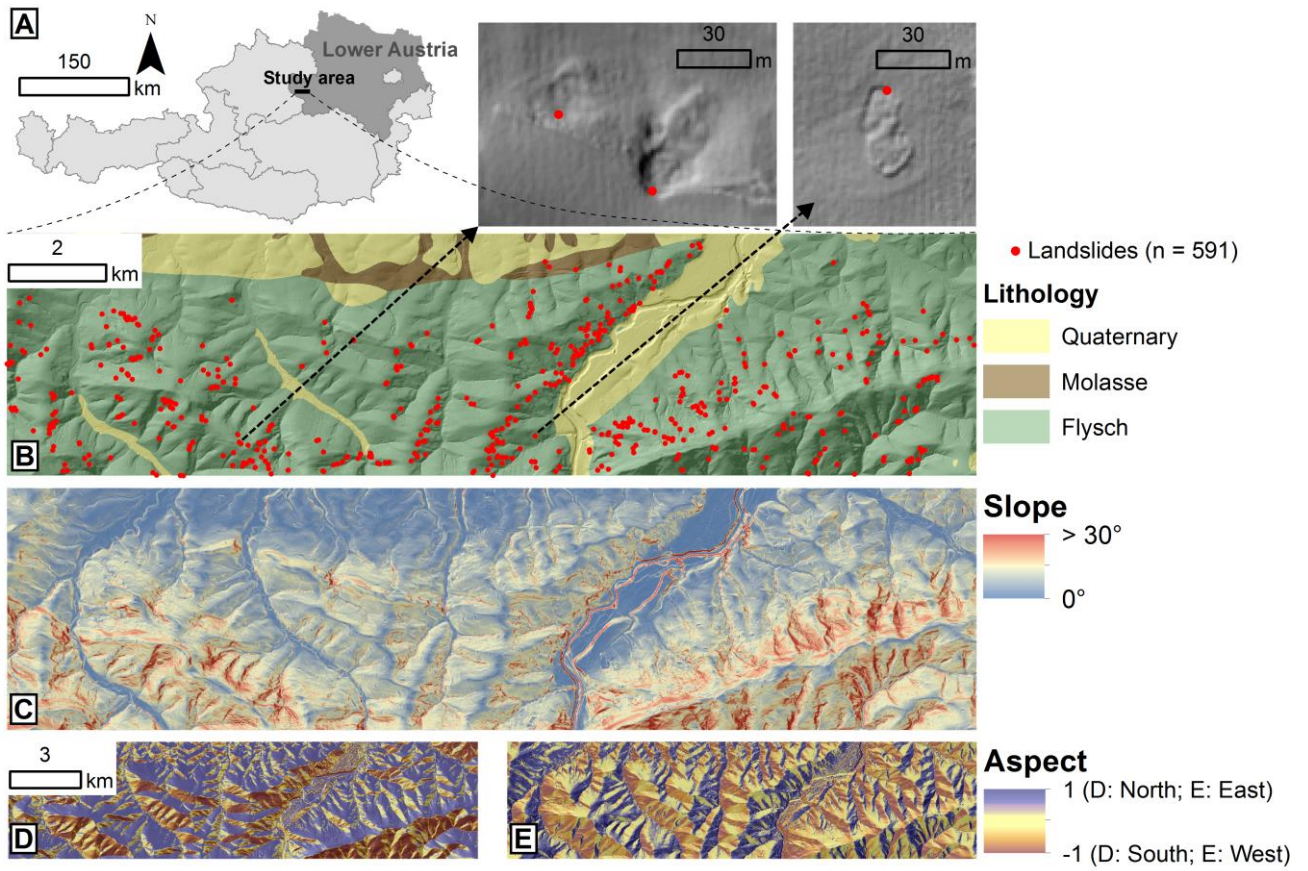
Figure 1. Location of the study area within Austria (~~A~~a), shaded relief image of the area with an overlaid ~~lithological units~~slope angle map and spatial distribution of landslide scarps (~~B~~b). The shaded relief images (top right) show typical landslides of the study area and the corresponding mapping location. ~~Slope angles~~Lithological units are given in ~~C~~"c" while slope orientation is represented as northness (~~D~~d) and eastness (~~E~~e).
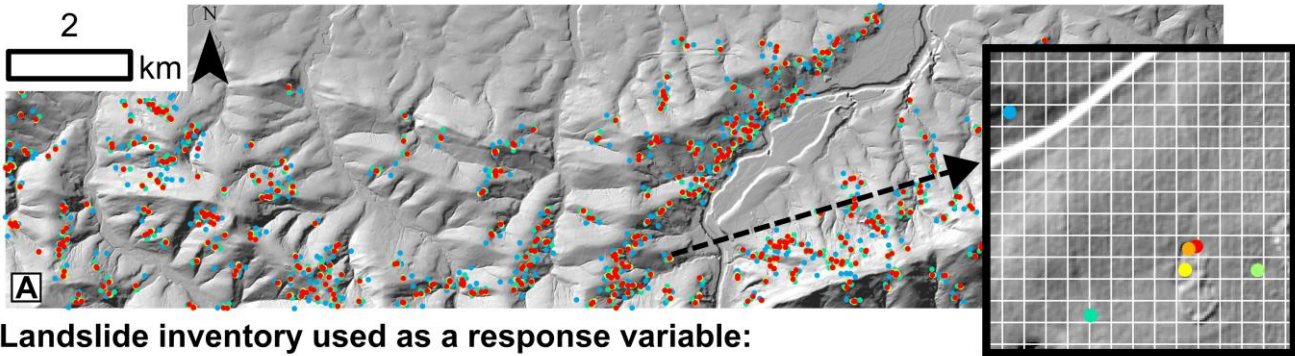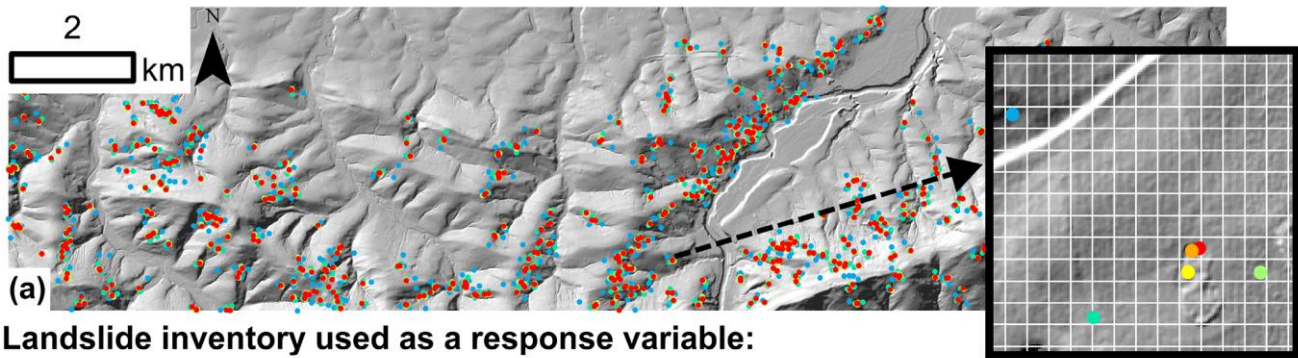
Landslide inventory used as a response variable:

- Unmodified ("0")
- Inaccurate ("μ5")
- Inaccurate ("μ10")
- Inaccurate ("μ20")
- Inaccurate ("μ50")
- Inaccurate ("μ120")
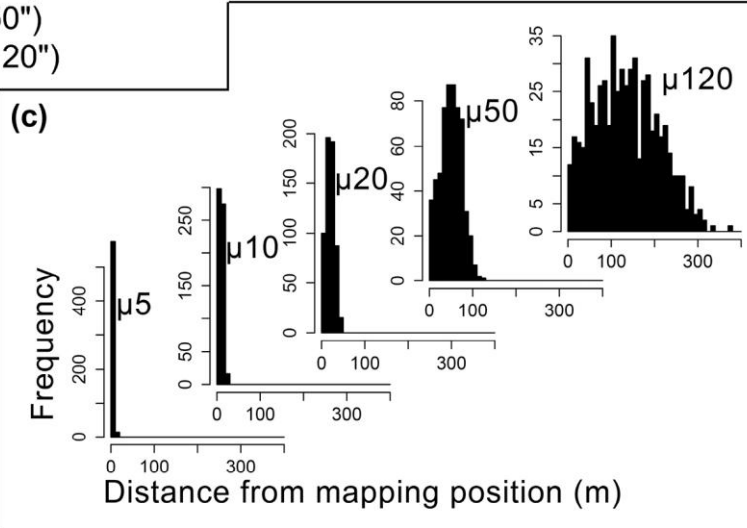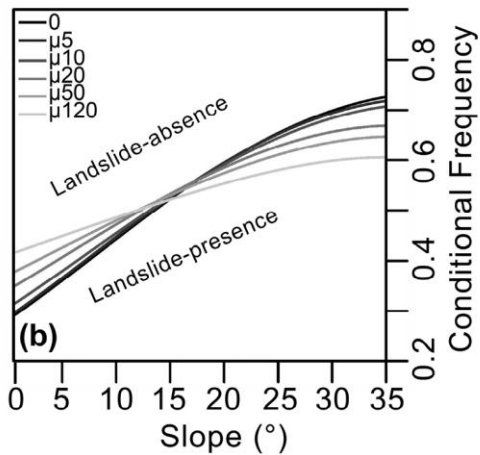
Figure 2. Overview of the simulated positional inaccuracy of the inventory for the study area (real data). The excerpt in ~~A~~"a" shows an example of an original mapping position of a landslide scarp (red dot) and corresponding simulated positional inaccuracies (the white grid corresponds to the modelling resolution of 10 m). The conditional frequency plot in ~~B~~"b" shows that a decreasing positional accuracy of the inventory led to the effect that landslides were more likely located on flatter slopes. The histograms in ~~C~~"c" (bin width 10 m) depict the frequency of landslides (n = 591) in relation to the distance from the original mapping position for the respective simulated positional inaccuracy.
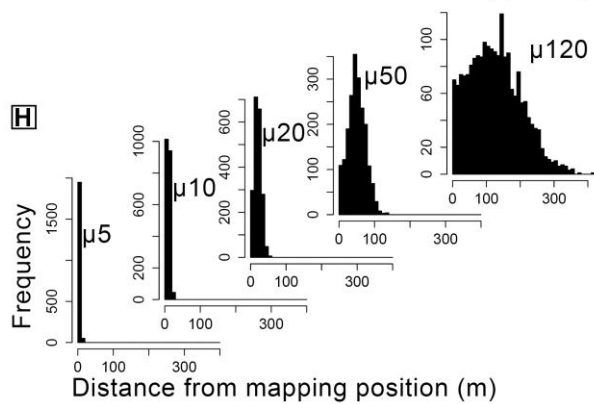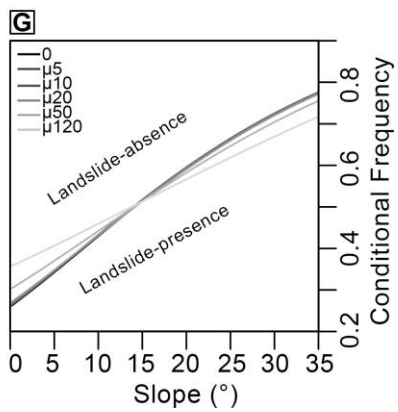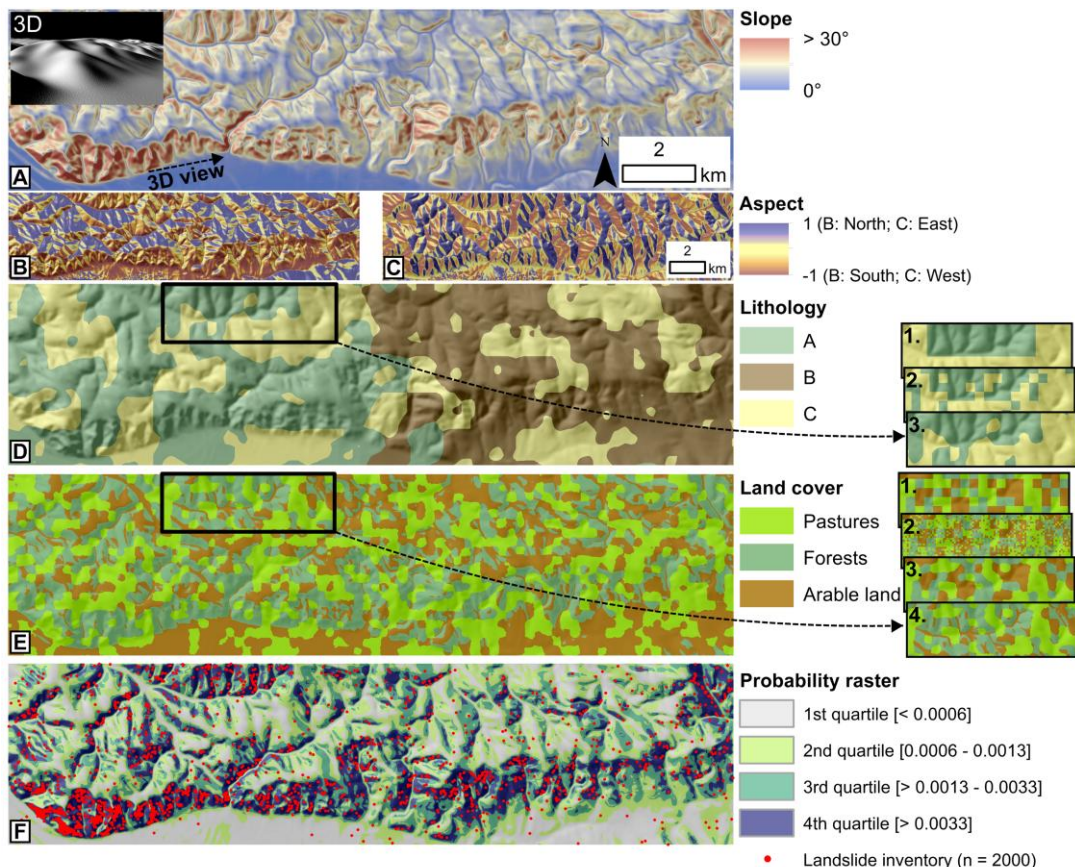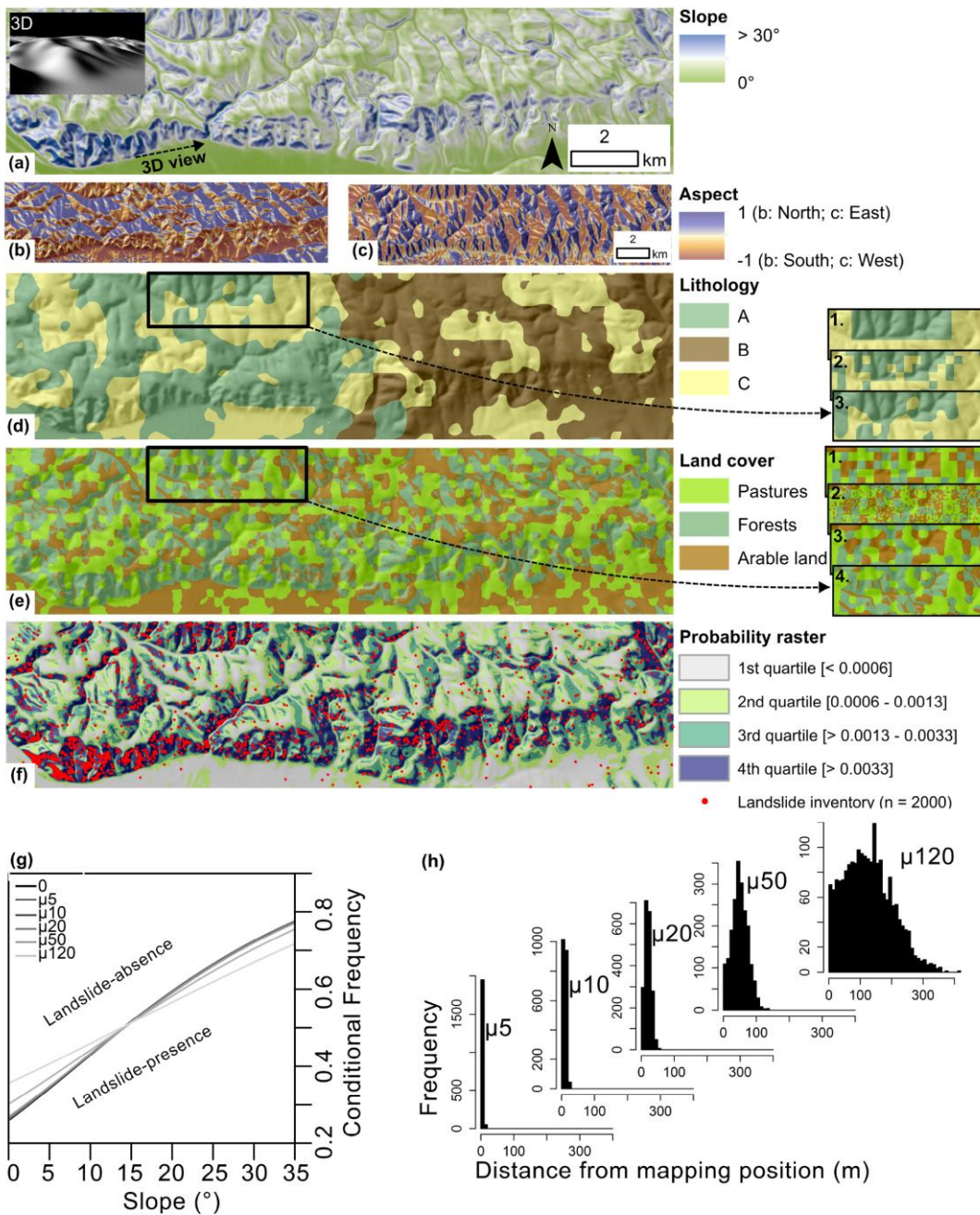
Figure 3. Spatial representation of the synthetically generated data set. Topographic parameters (~~A,B,C~~a,b,c); Lithological and land cover units (~~D,E~~d,e) with excerpts depicting the generation of this data sets (1: initial random raster; 2: intersection with a random raster produced at a higher spatial resolution; 3: moving window-based smoothing; 4: introduction of a slope-dependency for land cover). Landslide susceptibility according the predefined relationship and subsequent landslide sample further used as landslide presences for model realizations (~~F~~f). The plot in ~~G~~"g" depicts the conditional frequency of landslide occurrence on slope angle for the respective inventories. The histograms in ~~H~~"h" (bin width 10 m) show the frequency of landslides (n = 2000) in relation to the distance from the original mapping position.

Table 1. Prescribed relationship between the predictors and the response variable for the synthetic data set expressed as odds ratios (OR) (first row). The model estimate (bottom row) was based on the synthetically generated landslide sample (cf. Fig. 3F3f) and a logistic regression model trained with the predictors slope, northness, eastness, lithology, land cover. OR for the predictor slope relate to a slope angle increase of 10°, OR for both components of slope aspect (northness, eastness) to an orientation change of 45°.

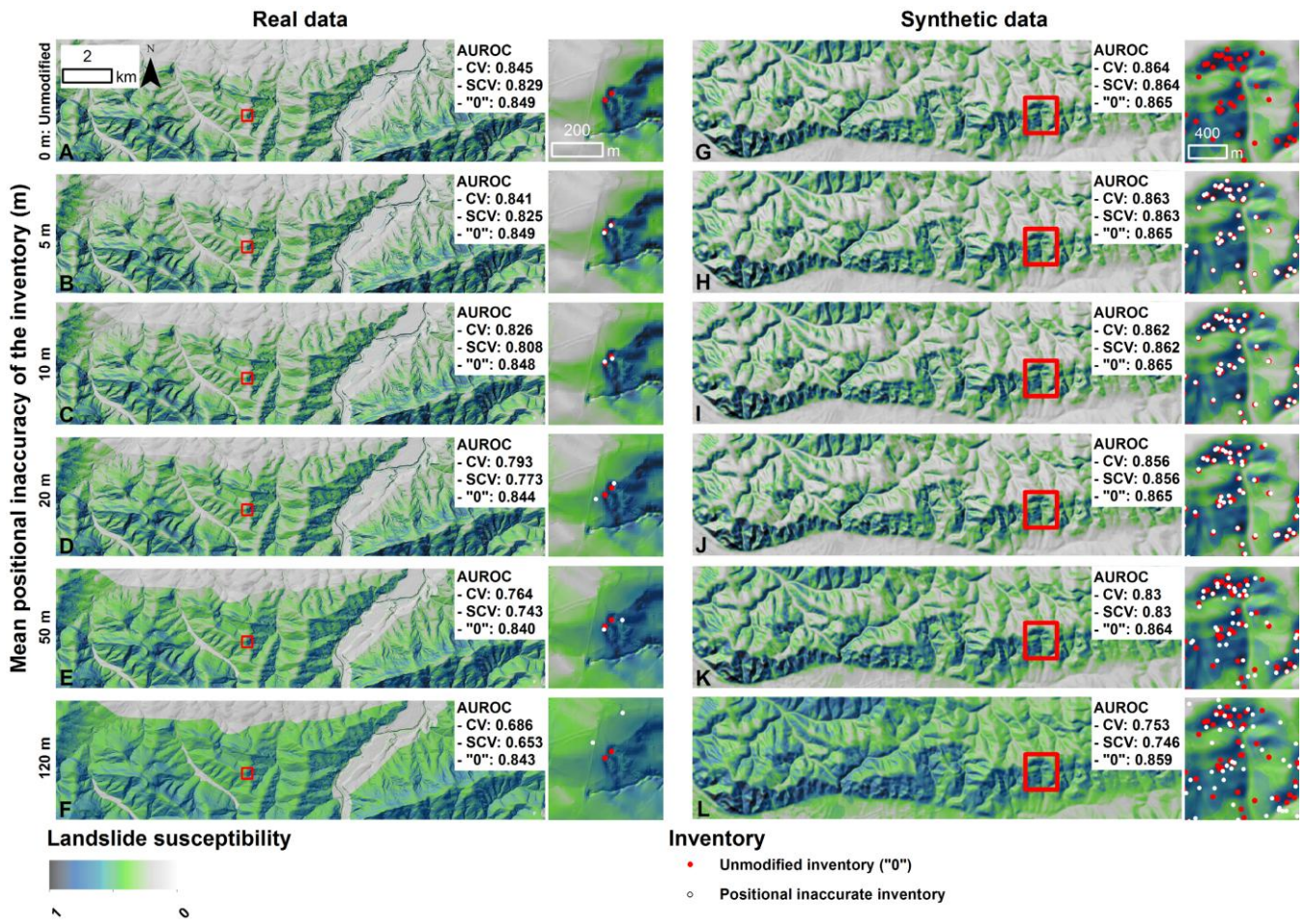|	Slope	Northness	Eastness	Litho. A	Litho. B	Litho. C	Pastures	Forests	Arable land
OR (prescribed)	6	0.8	1.4	1	1	0.5	1	0.5	1
OR (reference model)	6.22 *5.4 / 7.2	0.76 *0.7 / 0.8	1.37 *1.3 / 1.5	1 ref. class	0.86 *0.7 / 1	0.45 *0.4 / 0.6	1 ref. class	0.51 *0.4 / 0.6	0.86 *0.7 / 1.1

* 95% confidence interval
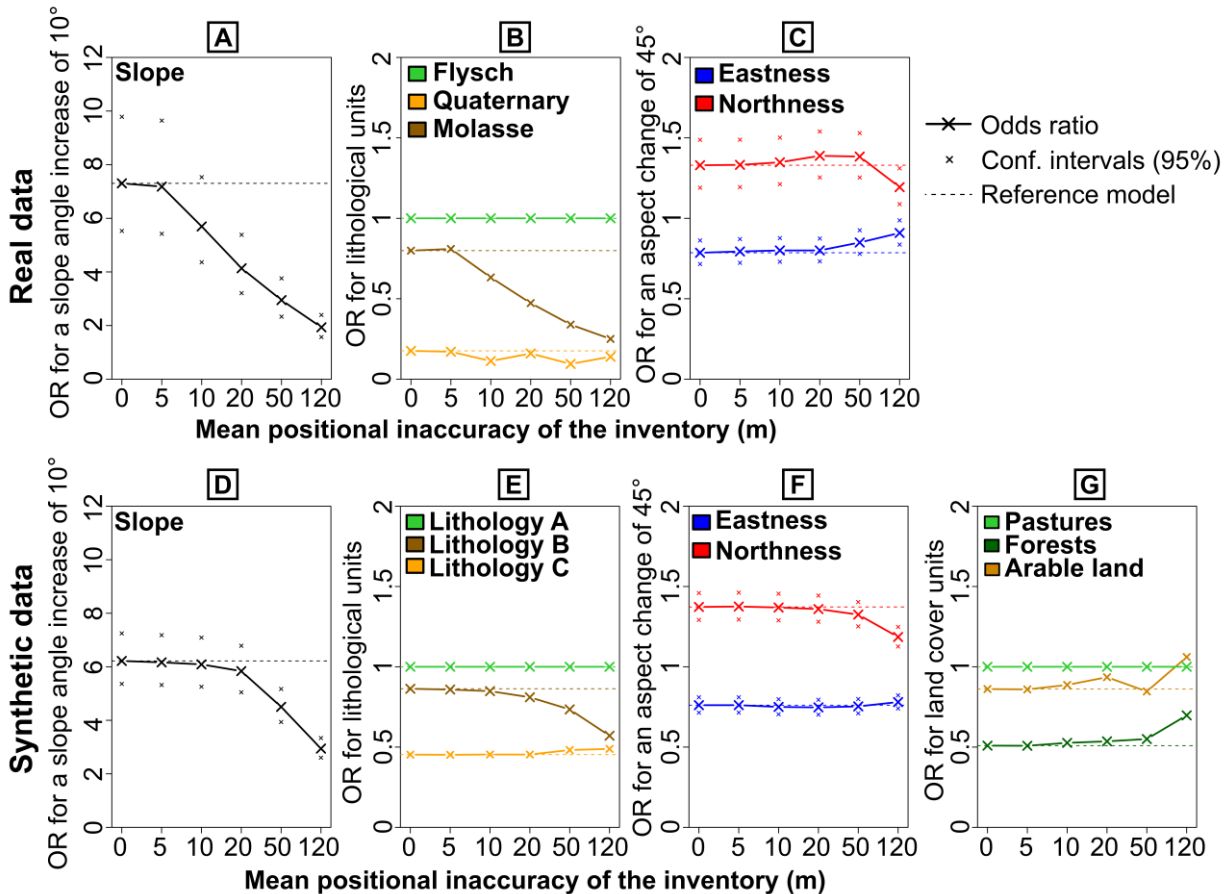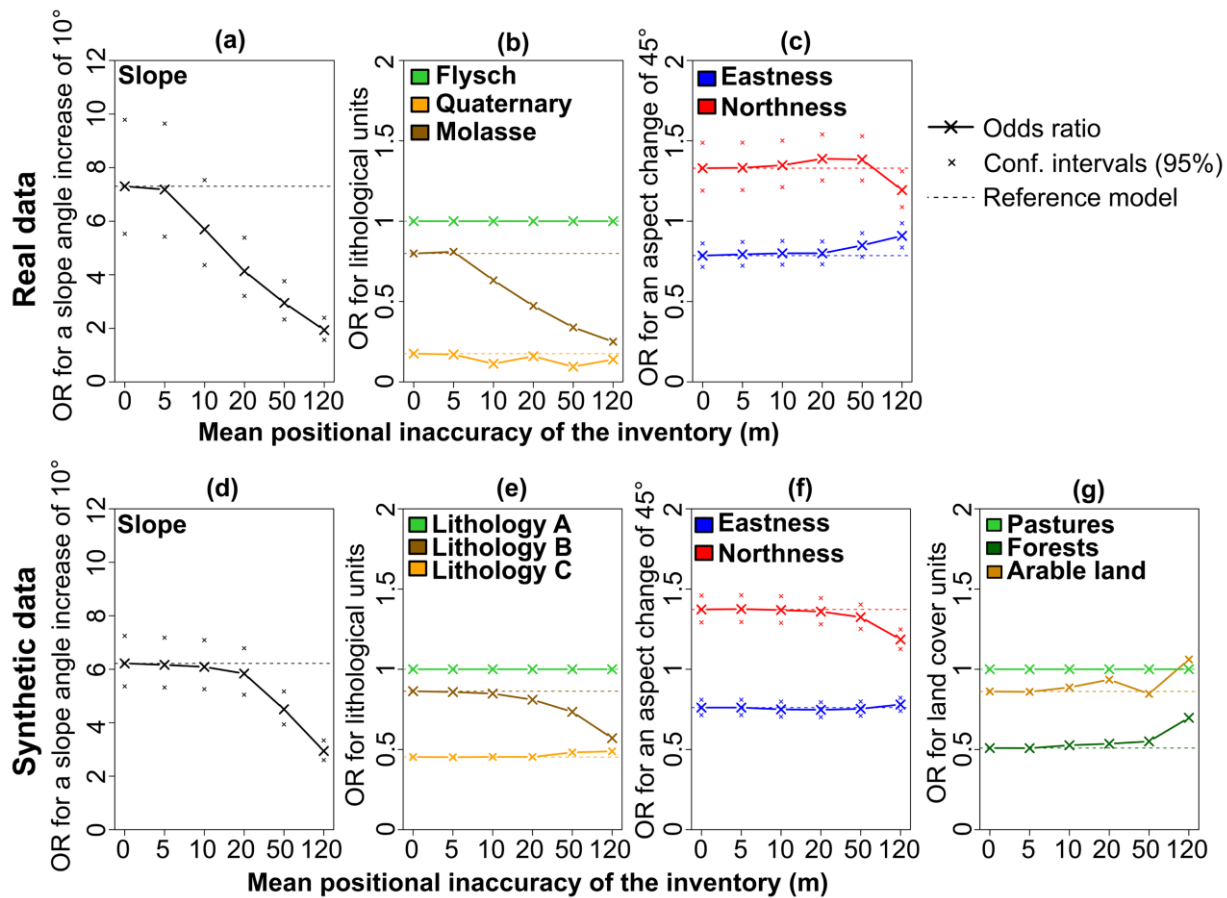
Figure 4. Estimated odds ratios (OR) for the models generated with differently accurate landslide inventories. ORs obtained for the predictors slope (~~A,D~~a,d), northness (~~C,F~~c,f), eastness (~~C,F~~c,f) and land cover (~~G~~g) suggested a decreasing estimated effect of topographic and land cover variations on modelling results with an increasing positional error of the inventory (i.e. ORs are approaching the no-effect threshold value 1). ORs of the lithology classes revealed decreasing modelled chances that the Molasse (~~B~~b) respectively the unit B (~~E~~e) may be affected by future landsliding.

29

Landslide susceptibility

Inventory

- •   Unmodified inventory ("0")
- ○   Positional inaccurate inventory
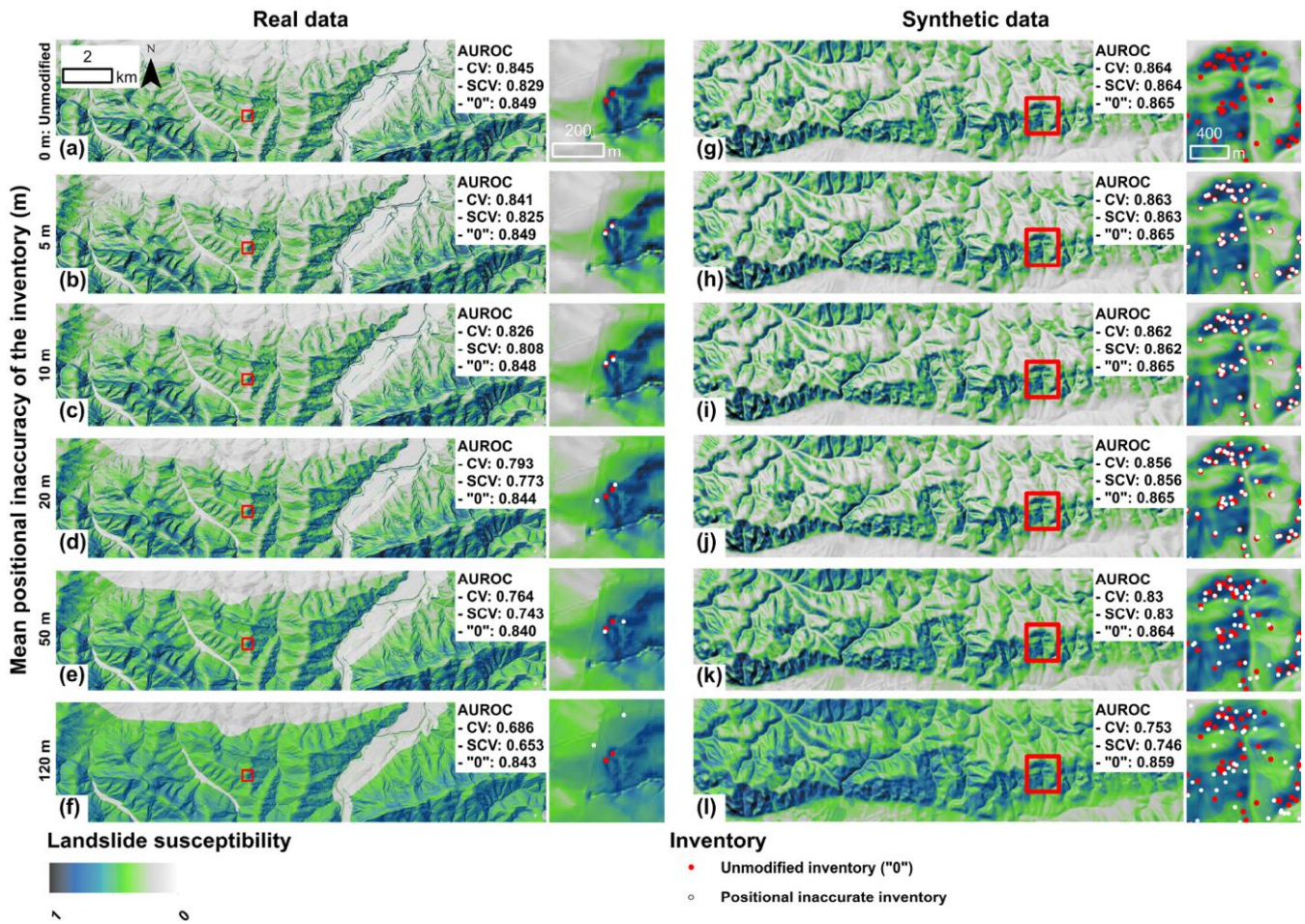
Figure 5. Landslide susceptibility maps (for the location of the cutouts refer to the red boxes) of models generated with differently accurate inventories (rows) for the real data (left column) and the synthetic data (right column). Corresponding predictive performances (median AUROC for CV and SCV) and the quantitative comparison with the original landslide positions ("0") are given in the respective white boxes. The cutouts additionally show the location of the positionally erroneous inventories (white dots) used as input for the respective models and the position of the original inventory (red dot). Note that maps generated with the most inaccurate inventory (~~F,L~~f,l) are characterized by a locally more uniform spatial pattern.
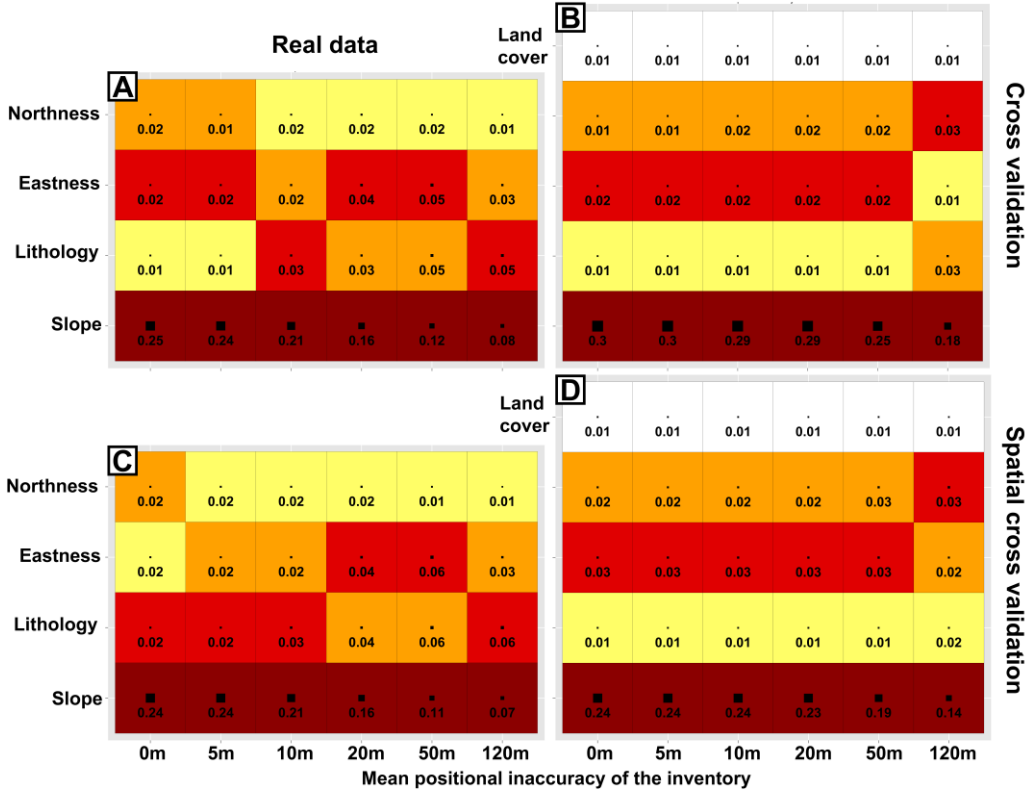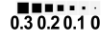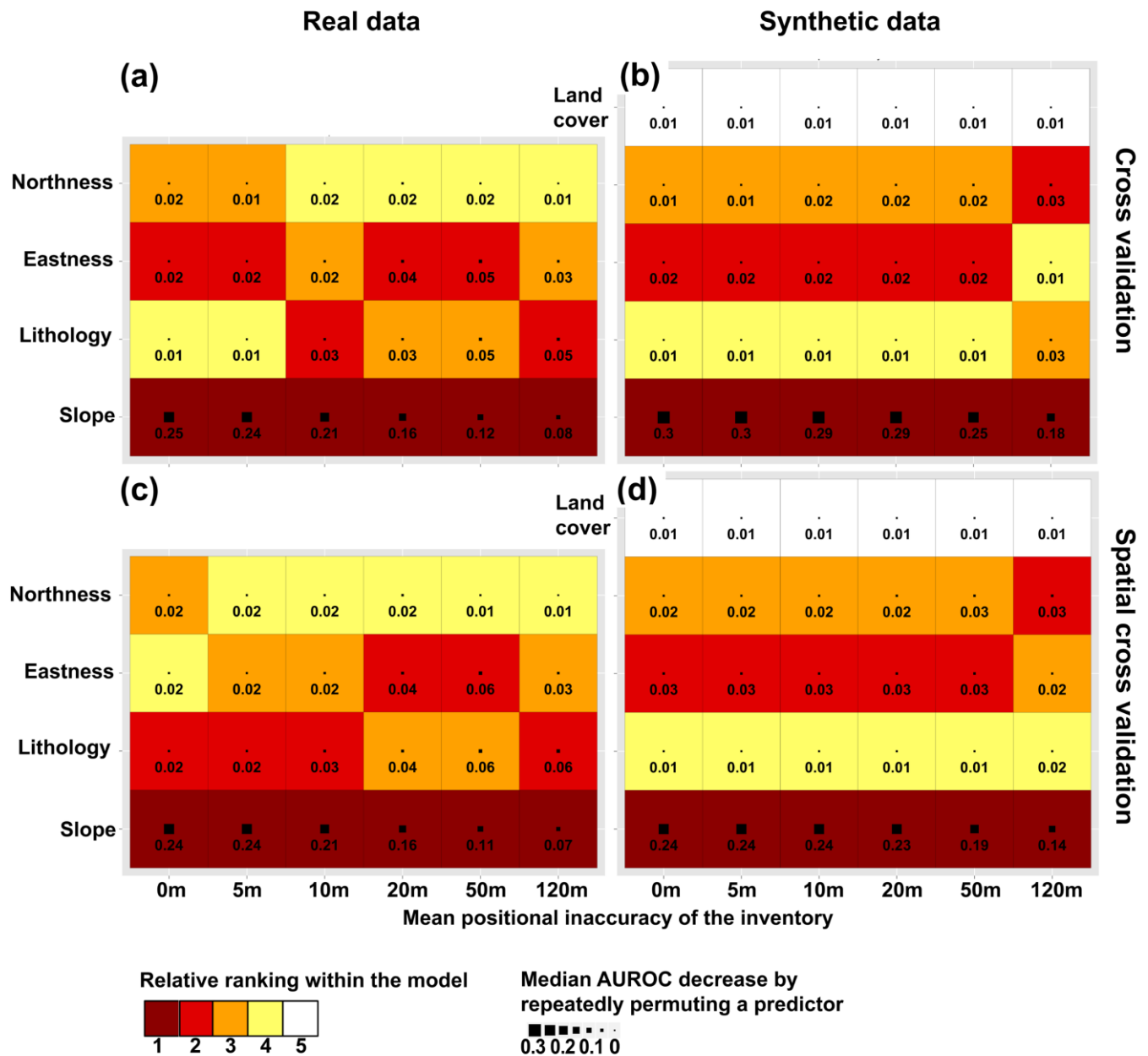
31

Synthetic data
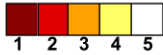
Real data

Figure 6. Permutation-based variable importance for the respective reference models (0 m) and models generated with positionally erroneous inventories (5 m – 120 m)~~.),~~ **for the real data sets (a,c) and the synthetic data (b,d) obtained by CV (a,b) and SCV (c,d).** The relative ranking indicates that slope was the most influential predictor for each model while its estimated importance constantly decreased with an increasing inventory-based positional error (cf. **Median** AUROC decrease). The estimated importance of lithology slightly increases with a growing inventory-based positional inaccuracy.