



# Assessing the quality of landslide susceptibility maps – case study Lower Austria

H. Petschko<sup>1</sup>, A. Brenning<sup>2</sup>, R. Bell<sup>1</sup>, J. Goetz<sup>1,2</sup>, and T. Glade<sup>1</sup>

<sup>1</sup>Department of Geography and Regional Research, University of Vienna, Austria

<sup>2</sup>Department of Geography and Environmental Management, University of Waterloo, Ontario N2L 3G1, Canada

Correspondence to: H. Petschko (helene.petschko@univie.ac.at)

Received: 27 February 2013 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 10 April 2013

Revised: 11 October 2013 – Accepted: 27 November 2013 – Published: 16 January 2014

**Abstract.** Landslide susceptibility maps are helpful tools to identify areas potentially prone to future landslide occurrence. As more and more national and provincial authorities demand for these maps to be computed and implemented in spatial planning strategies, several aspects of the quality of the landslide susceptibility model and the resulting classified map are of high interest. In this study of landslides in Lower Austria, we focus on the model form uncertainty to assess the quality of a flexible statistical modelling technique, the generalized additive model (GAM). The study area (15 850 km<sup>2</sup>) is divided into 16 modelling domains based on lithology classes. A model representing the entire study area is constructed by combining these models. The performances of the models are assessed using repeated *k*-fold cross-validation with spatial and random subsampling. This reflects the variability of performance estimates arising from sampling variation. Measures of spatial transferability and thematic consistency are applied to empirically assess model quality. We also analyse and visualize the implications of spatially varying prediction uncertainties regarding the susceptibility map classes by taking into account the confidence intervals of model predictions. The 95 % confidence limits fall within the same susceptibility class in 85 % of the study area. Overall, this study contributes to advancing open communication and assessment of model quality related to statistical landslide susceptibility models.

## 1 Introduction

Landslides occur in mountainous as well as in hilly or coastal regions worldwide and have often been an underestimated hazard. In general, people and governing authorities are not sufficiently aware of the potential locations and consequences of landslides (Hervás, 2003). However, in Austria authorities and residents have become more aware of landslide hazards because of recent major landslide events, which affected many residents, caused significant damage to infrastructure and private properties and were reported and discussed on the local media (Damm et al., 2010). These include events in August 2005 in Gasen and Haslau and incidents in 2009 in the district of Feldbach and in the province of Lower Austria, where about 4000 landslides occurred in total (Schwarz and Tilch, 2008; Hornich and Adelwöhrer, 2010; Abteilung Feuerwehr und Zivilschutz, Amt der NÖ Landesregierung, 2010; BMLFUW, 2010). While in the past in Austria practices of post-disaster recovery and reconstruction were applied, recently more and more national and provincial authorities have been demanding pre-disaster mitigation tools, which help to prevent future damage caused by natural hazards such as landslides. Aiming at avoidance of the hazard, zonation plans can be facilitated in prospective land use planning to prevent development in undesirable locations or undesirable types of development (Schwab et al., 2005). In this context landslide susceptibility maps have proven to be a powerful tool, as they give coherent information on the spatial probability on where landslides, or landslide scarps, might occur (Brabb, 1984; Glade et al., 2005; Guzzetti et al., 2000, 2006; Varnes, 1984). The term landslide susceptibility is hereby defined by Brabb (1984) as the likelihood of

a landslide occurring in an area with given local terrain attributes. The frequency or time of occurrence of the future landslides is not assessed within susceptibility zoning (Fell et al., 2008).

This study was embedded in a project preparing landslide susceptibility maps for the province of Lower Austria (15 850 km<sup>2</sup>) with an aimed output map scale of 1 : 25 000 (Glade et al., 2012). Generally, landslide hazard zonation for determining slope instability due to landslides is based on simple landslide inventories or heuristic, statistical (including machine learning) or deterministic approaches (Soeters and Van Westen, 1996). Although these approaches are dealing with landslide hazards (spatial and temporal occurrence probability of landslides) the same categorization can be applied for susceptibility maps. Statistical landslide susceptibility models are particularly useful for modelling large areas on a medium scale (1 : 25 000–1 : 500 000) to get an overview of which slopes or slope sections might be prone to landslides in future (Fell et al., 2008). According to the resulting susceptibility maps, hot spots can be identified where more detailed analysis of the slope stability should follow (Van Westen et al., 1997). For this study a statistical approach was regarded as suitable considering the aimed output map scale, the size of the study area, the available data and the limited availability of additional resources for preparing input data (such as a more detailed landslide inventory or soil parameters).

The central assumption of landslide susceptibility modelling is based on James Hutton's (1726–1797) concept of uniformitarianism, “the past and the present are keys to the future” (Orme, 2002). Thus, statistical estimation of the possible future location of landslides is usually based on the conditions (e.g. local terrain attributes) of past landslides (Varnes et al., 1984; Carrara et al., 1993). Presently, several statistical and machine learning techniques are available for application to landslide susceptibility modelling. The most common are logistic regression (e.g. Atkinson et al., 1998; Ayalew and Yamagishi, 2005; Van Den Eeckhaut et al., 2006), bivariate models such as weights of evidence (e.g. Neuhäuser and Terhorst, 2007; Schicker and Moon, 2012) and machine learning techniques such as artificial neural networks (e.g. Nefeslioglu et al., 2008; Pradhan and Lee, 2009) and support vector machines (e.g. Marjanović et al., 2011; Pradhan, 2013). Detailed reviews and comparison of different models can be found, among others, in Guzzetti et al. (1999), Dai et al. (2002), Brenning (2005), Glade and Crozier (2005), Guzzetti (2005), Rossi et al. (2010) and Vorpahl et al. (2012). However, the more flexible machine learning algorithms often tend to overfit (Sect. 5.2), while linear models might not be flexible enough to portray possible nonlinearity in the relationship between landslide occurrence and the explanatory variables (Brenning, 2005; Nefeslioglu et al., 2008; Goetz et al., 2011). In this study the nonlinear generalized additive model (GAM, Hastie and Tibshirani, 1990) was selected for the susceptibility modelling. The GAM is widely used in

ecological modelling (Guisan et al., 2002) and has recently been introduced to landslide susceptibility modelling (Brenning, 2008; Jia et al., 2008; Park and Chi, 2008; Goetz et al., 2011; Vorpahl et al., 2012). It is an extension of the generalized linear models (GLMs, such as logistic regression) that can model nonlinear relationships between response and explanatory variables while being less prone to overfitting in geomorphological modelling than the more flexible methods (Hastie and Tibshirani, 1990; Brenning, 2009; Goetz et al., 2011).

Since the application of the landslide susceptibility map resulting from this study is planned for implementation by municipal authorities, our main focus is to identify different aspects of quality (depending on input data, model performance and provided analysis of prediction uncertainties). Accordingly, the purpose of this study is to review the characteristics that describe quality, to analyse the (statistical) model quality and to assess and visualize the uncertainty of the prediction.

Likewise, this paper presents a review on considerations on quality in statistical landslide susceptibility maps (Sect. 2); background information on the study area and the landslides (Sect. 3); a description of the data sources and the input data preparation (Sect. 4); a detailed section on the methods used assessing the model form uncertainty and prediction uncertainty within a topographically heterogeneous study area (Sect. 5) and a thorough presentation of the results (Sect. 6) followed by a critical discussion (Sect. 7). The paper ends with a short conclusion highlighting the main findings of this study (Sect. 8).

## 2 Considerations on the quality of statistical landslide susceptibility models and resultant maps

As it comes to the application of the landslide susceptibility maps, which is associated with constructing a decisive reality (Egner and Pott, 2010) for the municipality and the land use planners, a detailed and transparent assessment of its quality is necessary. This very broad term of quality can be interpreted in several ways and in several stages of the process of preparing a landslide susceptibility map as described, amongst others, by Carrara (1993), Carrara et al. (1995), Ardizzone et al. (2002), and Guzzetti et al. (2006). In general, good quality refers to input data, a model or result (e.g. susceptibility map) which includes relatively low uncertainties.

Landslide susceptibility modelling is inherently riddled with uncertainties. In engineering and risk assessment a basic distinction between aleatory and epistemic uncertainties is very common (Hoffman and Hammonds, 1994; Oberkampf et al., 2004; Roy and Oberkampf, 2011; Hill et al., 2013). Many processes and conditions leading to landslide occurrence are known and mappable, known but not collectable, or are simply unknown (Carrara et al., 1999). The

imperfect understanding of the complexity of this hazardous phenomenon arising from the missing knowledge (Ardizzone et al., 2002; Kunz et al., 2011) is generally referred to as epistemic uncertainty (Hoffman and Hammonds, 1994; Hora, 1996; Oberkampf et al., 2004). Epistemic uncertainty is often divided into input uncertainty, parametric uncertainty and structural or model form uncertainty (Roy and Oberkampf, 2011; Rougier and Beven, 2013). Naturally, these uncertainties can be reduced by improving knowledge of the subject (Spiegelhalter and Riesch, 2011). Unavoidable uncertainty arising from the natural variability and/or randomness of the natural hazard process can be distinguished as aleatory uncertainty (Rougier and Beven, 2013; Rougier, 2013). Although these distinctions can be applied, the organization of uncertainties is dependent on the study objectives (Hora, 1996; Roy and Oberkampf, 2011).

In the following, we will discuss quality by addressing uncertainty issues associated with statistical modelling regarding input data (parametric uncertainty), model performance (model form uncertainty) and the final susceptibility map. Details on other types of uncertainties which might be more important for other types of susceptibility assessment (e.g. deterministic) or on the propagation of uncertainties can be found e.g. in Hoffman and Hammonds (1994), Draper (1995), Oberkampf et al. (2004), Karam (2005), Helton et al. (2010) and Rougier (2013).

## 2.1 Quality of input data

Achieving a good quality of the final landslide susceptibility map starts with the quality of the input data set for the modelling. Besides the geomorphological relevance, the spatial resolution and accuracy of the geo-environmental as well as the landslide inventory data is important (van Westen et al., 2008). During the preparation of all input data (done by mapping or measuring) subjectivity, the experience of the mapper, measuring errors or imprecision in computer processes are sources of parametric uncertainty (Mosleh, 1986; Ardizzone et al., 2002; Elith et al., 2002; van Westen et al., 2005, 2008). Furthermore, possible incompleteness, not only in a sense of full spatial coverage but also of general availability of important thematic information on (predisposing or preparatory) factors determining the landslide susceptibility influences the quality and makes the susceptibility assessment more difficult (Carrara et al., 1999; Ardizzone et al., 2002; Van Westen et al., 2005). This source of epistemic uncertainty can rarely be reduced and other assessments of the uncertainty have to be selected. The quantitative assessment of parametric uncertainty and incorporation in subsequent hazard and risk assessment is still a challenge (Ardizzone et al., 2002; Oberkampf et al., 2004; Roy and Oberkampf, 2011). Therefore, estimation on the completeness of the landslide inventory and details on the collecting and mapping method giving information on the accuracy and the location of the landslide point/line/polygon (main scarp or entire land-

slide body) are very important. This influences the further usage of the input data set and the feasible interpretation of the susceptibility maps substantially.

## 2.2 Quality of statistical model form

Every model is a simplification of reality. Therefore, while modelling landslide susceptibility, it should be expected that some discrepancies regarding the ability to explain all the processes involved in the phenomenon would naturally occur (Rougier and Beven, 2013). Every modelling result is highly dependent on the input data, the assumptions made to set up the model design and the selection and performance of an appropriate model. These assumptions also define the limitations of the model, its result and the allowed interpretation of it. Quantifying this part of the model form uncertainty is challenging. However, model validation procedures have been developed and are commonly applied to assess the model performance and therefore the model form uncertainty (Roy and Oberkampf, 2011).

The prediction skills (Guzzetti et al., 2006) can be analysed quantitatively with various performance measures and estimation techniques, and qualitatively by considering expert knowledge of a particular study area. Qualitative methods analyse the geomorphic plausibility of the map (Demoulin und Chung, 2007; Bell, 2007). Common quantitative performance measures are success/prediction rate (Chung and Fabbri, 2003, 2008), confusion matrix or error rates (Beguería, 2006a; Brenning, 2005), and cost curves (Frattini et al., 2010). Among the performance estimation techniques and measures, cross-validation using a single hold-out method and the area under receiver operating characteristic curve (AUROC) value based on ROC plots (Beguería, 2006a; Brenning, 2005; Frattini et al., 2010) are usually applied (i.e. Chung and Fabbri, 1999, 2003; Fabbri et al., 2003; Remondo et al., 2003; Brenning, 2005; Beguería, 2006a; Frattini et al., 2010; Rossi et al., 2010).

In single hold-out methods the data set is split in one single training and test sample. The training sample is used to fit the model and the test sample is used to determine the model performance. This results in a single estimate of the performance measure (e.g. one single AUROC value) without providing a measure of precision of the estimator. The estimate depends on the (random) sample used for modelling the susceptibility and testing the model's performance, which may itself have "peculiar" random characteristics that would be different for a different test set. Repeated  $k$ -fold cross-validation solves this problem by using, one after another, different subsets or partitions of the data set as test and training sets, thus effectively using the entire data set for performance estimation (Brenning, 2012a, b). In addition, repeating this procedure for different data partitioning reduces sampling variability and allows for the determination of the precision of the performance estimator (see Sect. 5.3 for details).

Depending on the partitioning method (randomly or spatially) to obtain the folds, these statistical estimation methods can further provide a means for assessing the non-spatial transferability of a model onto a different, independent random sample, and the spatial transferability into a spatially separate area. Spatial transferability refers to the capability of the model to generalize empirical relationships learned on a training data set, and to transfer these relationships to (usually adjacent) regions without major loss in predictive performance (Brenning, 2005; Von Ruetten et al., 2011). This (non-) spatial transferability describes the model form uncertainty besides other performance measures such as the AU-ROC value.

In addition, aleatory uncertainties are of importance in the modelling stage, because even if all parameters were known perfectly, some deviance (or discrepancy) of the model results in nature might still occur due to the natural variability of the process (Elith et al., 2002, Rougier and Beven, 2013).

### 2.3 Quality of final susceptibility map – visualizing prediction uncertainties

One important model uncertainty which can be visualized in a map is the prediction uncertainty arising from using a statistical model. The output of a statistical model for spatial modelling is usually comprised of a single probability value for each unit of the prediction surface (grid cells, slopes or terrain units). These individual probability values represent an estimated conditional mean value of the predicted probability (Hosmer and Lemeshow, 2000). Therefore, there is a prediction uncertainty or possible range, as determined by the standard error of the predicted probabilities, of probability estimates for each unit of the susceptibility map (Guzzetti et al., 2006). The analysis of the standard error of the predicted probabilities and the prediction uncertainty analysis are independent of any class thresholds. Estimating the standard error was analysed and presented in previous research (e.g. Guzzetti et al., 2006; Van den Eeckhaut et al., 2009; Rossi et al., 2010; Sterlacchini et al., 2011). However, the way it affects the appearance of the classified landslide susceptibility map is of interest for planning purposes as considering these prediction uncertainties might result in overlaps of different susceptibility classes. Naturally, the amount of overlapping raster cells and therefore the uncertainty in the final classified map might be dependent on the number of classes and the selected thresholds for the classification.

## 3 Lower Austria and landslide occurrence

The Austrian province of Lower Austria covers a total area of 19 177 km<sup>2</sup>, 15 850 km<sup>2</sup> of which constitute our study area (Fig. 1).

The variation of material and topographic characteristics across Lower Austria can be attributed to the diverse lithology (Petschko et al., 2012). The predominant material types include gravel, sand in the alluvial deposits and fluvial terraces; loess and loam; marl with a high amount of silt in the Molasse Zone and Schlier Zone; sandstone and marlstone of the Rheno-danubian Flysch Zone and Mélange Zone; limestone, dolostone and the igneous rocks of the Austroalpine Unit; and gneisses and granites in the Bohemian Massif (Fig. 1, Table 1).

The median slope angle was used as a simple proxy of the topographic characteristics to highlight the diverse conditions across Lower Austria. The median slope angle ranges from a minimum of < 1° (alluvial deposits) to a maximum of 27° (Austroalpine Unit with dolostone; Table 1).

Land use patterns and variation on precipitation also contribute to the diverse geography of Lower Austria. The land use (40 % agricultural land, 40 % forest, 13 % grassland) is unevenly distributed across the province. Areas south of the Danube and the relatively flat northeast region of the province are predominantly used as agricultural land. The steeper slopes in the south and southwest are mainly forest-covered (coniferous and deciduous forest; Eder et al., 2011). The spatial distribution of the mean annual precipitation rate shows a gradient between the northeast (400–500 mm) and the southwest regions (1600–1700 mm) (Hydrographischer Dienst des Landes Niederösterreich, 2011).

Many different landslide types including rockfalls, earth slides, debris slides and debris flows, can be found in Lower Austria (classified according to Cruden and Varnes, 1996; and Dikau et al., 1996). While rockfall and debris flows typically occur in the southern lithology (Austroalpine Unit), earth- and debris slides occur all over the province with different densities (Table 1), sizes and depths. Only earth- and debris slides which were mapped with one point in the main scarp of each slide were within the scope of this study; recent examples of these landslide types are shown in Fig. 2.

The earth- and debris slides in Lower Austria are mainly triggered by heavy rainfall or rapid snowmelt events (Schwenk, 1992; Schweigl and Hervás, 2009). The Mélange lithology has the maximum landslide density of 5.5 landslides km<sup>-2</sup>. The Rheno-danubian Flysch Zone and the Zone of Molasse with Schlier also have high landslide densities (> 4 landslides km<sup>-2</sup>). The lithological properties and topographic characteristics of these units are highly susceptible to earth- and debris slides (Gottschling, 2006; Wessely, 2006). These densities were calculated from the landslide inventory described in Sect. 4.1.

## 4 Data sources and preparation

The susceptibility modelling in this study was based on an airborne laser scanning digital terrain model (LiDAR-DTM) with a resolution of 1 m × 1 m, acquired between 2006 and

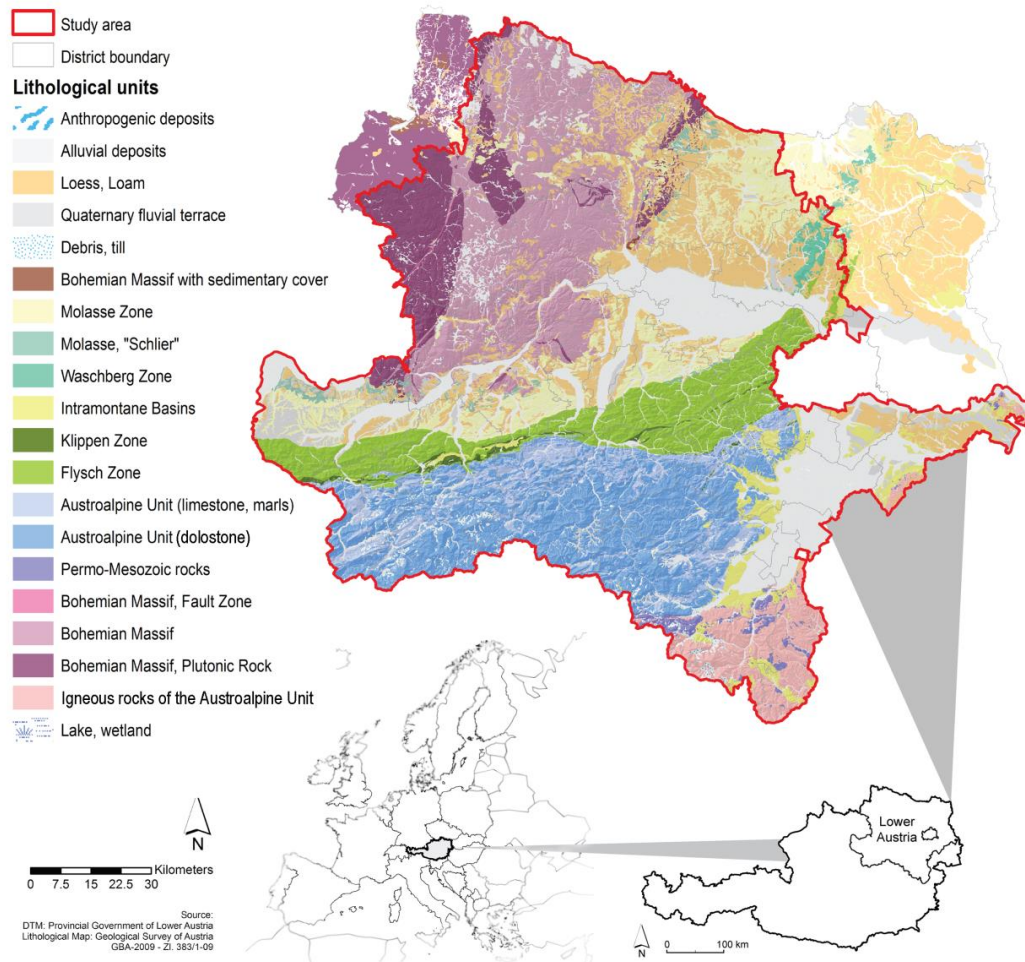
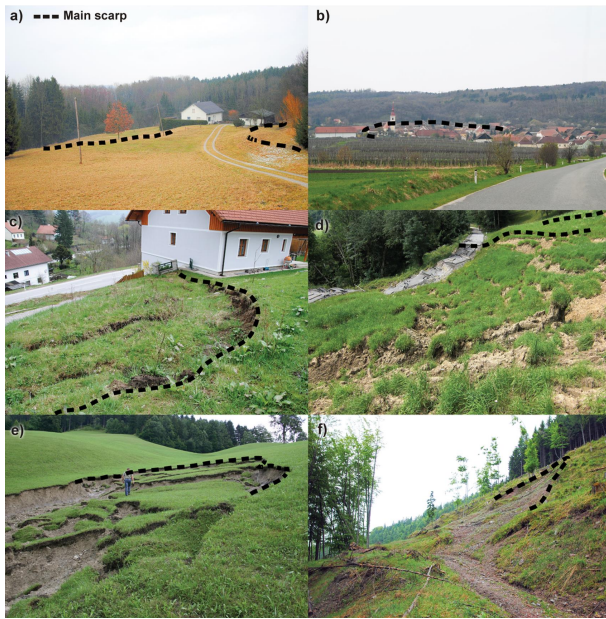


Fig. 1. Location and lithology of Lower Austria.

Table 1. Lithological units of Lower Austria: landslides and topography.

Code	Merged units	Name	Material <sup>+</sup>	Area (km <sup>2</sup> )	No. of slides	Landslide density (No. km <sup>-2</sup> )	Median slope (°)
1	0124	Anthropogenic deposits		2.93	0	0	3.1
2	02297	Alluvial deposits	Gravel, sand,	3739.17	194	0.05	0.9
10	10	Loess, Loam	Sand, loam, loess-loam, drifting sand	2849.22	329	0.12	2.8
24	0124	Quaternary fluvial terrace	Gravel, sand, loess loam,	747.79	222	0.30	2.2
35	35	Debris, till	Debris, till, scree material, rock avalanche material	217.19	177	0.81	16.6
37	3786	Bohemian Massif with sedimentary cover	Sandstone, claystone, conglomerate	30.95	0	0	12.2
39	39	Molasse Zone	Marl, sand, gravel, silt	1462.46	428	0.29	4.7
58	58	Molasse, Schlier	Fine sandy-silty marl in the circum-Alpine Tertiary basins	117.43	501	4.27	4.8
86	3786	Waschberg Zone	Marl, sand, limestone, clay	123.50	52	0.42	5.8
104	104	Intramontane Basins	Sand, gravel, breccia, clay, marl	737.26	291	0.39	5.2
120	120	Mélange Zone, Klippen Zone	Dominantly penninic metasediments and phiolites as well as insignificant Austroalpine elements, Ybbsitzer and Grestener Klippen Zone	73.46	404	5.50	11.9
126	126	Rheno-danubian Flysch Zone	Interbedded sandstone, marlstone to mudstone, marl	1365.96	6281	4.60	12.6
179	179	Austroalpine Unit with limestone, marls and sandstone	Limestone, marl, shale, sandstone, conglomerate	785.85	1636	2.08	20.2
191	191	Austroalpine Unit with dolostone	Dolostone, limestone	2148.57	1419	0.66	27.1
230	230	Permo-Mesozoic rocks (overlying the Austroalpine igneous rocks)	Carbonate rocks, siliciclastics, porphyry (mostly metamorphics)	116.43	88	0.76	19.5
239	239	Igneous rocks of the Austroalpine Unit	Orthogneiss, Paragneiss, Mica-schist, Phyllite	614.35	586	0.95	15.1
251	251 259	Bohemian Massif, Fault Zone	Tectonic fault zone	11.81	0	0	7.0
259	251 259	Bohemian Massif	Paragneiss, mica-schist, phyllite, orthogneiss, Gföhl Gneiss, Granulite	2398.93	227	0.09	6.1
276	276	Bohemian Massif, plutonic rock	Granite, plutonic rock	1606.85	52	0.03	6.5
297	02297	Lake, wetland	Limnic sediments, wetland	38.77	2	0.05	3.5

<sup>+</sup> sorted according to frequency of occurrence, after geological map



**Fig. 2.** Examples of landslides (earth- and debris slides) typical of the respective lithological unit (a) landslide in Molasse Zone at Strengberg, (b) the village “Waitzendorf” located on a landslide in Molasse with Schlier, (c) translational landslide in the Rheno-danubian Flysch Zone at Brand, (d) rotational landslide/flow in the Rheno-danubian Flysch Zone at Stössing, (e) landslide in Austroalpine Unit (Limestone, Marls) at Dippelreith, (f) landslide in Austroalpine Unit (with Dolostone) at Kleinzell. Pictures taken by: (a), (b), (c), (f) Petschko (2010, 2011, 2012); (d) Bertsch (2009); (e) BGR (2006).

2009. LiDAR-DTMs are very useful for mapping landslides and representing the morphology of the study area, even in forest areas (Van den Eeckhaut et al., 2007; Razak et al., 2011).

Considering the detailed resolution we have given in the LiDAR-DTM and the rather coarse resolution given in the soil data (50 m × 50 m) we had to find a compromise for the analysis and output of the susceptibility map. We decided an output resolution of the maps of 10 m × 10 m would be suitable to produce 1 : 25 000 scale maps; we are still able to take advantage of the high resolution of the topographic data but also avoid wrong signals in the modelling from very local variations in the derivatives of the high-resolution DTM (Van Westen et al., 2008). The resampling to a 10 m × 10 m resolution was done by bilinear interpolation. Nevertheless, we are aware that this artificial improvement of the soil data resolution does not increase the data accuracy.

#### 4.1 Response variable – landslide inventory

The landslide inventory originates from previous research where earth- and debris slides were mapped from a high-resolution LiDAR-DTM (Petschko et al., 2010; Glade et al., 2012, Petschko et al., 2013a). This inventory consisted of

points located in each landslide’s main scarp and includes 12 889 earth- and debris slides (Table 1). However, no attributes (landslide type, estimated age, land use, time/date of occurrence etc.) were assigned to each point.

Considering the large study area and the aimed result of a susceptibility map with three classes, a point inventory was preferred over a polygon inventory (Petschko et al., 2013a, c). The decision to map one point per landslide was aimed at increasing mapping effectiveness, avoiding uncertainty related to mapping landslide polygon boundaries, reducing spatial autocorrelation of the case samples (e.g. landslides) and providing equal treatment of small and large landslide samples (Carrara, 1993; Atkinson and Massari, 1998; Van den Eeckhaut et al., 2006; Heckmann et al., 2013; Petschko et al., 2013a). A comparison of modelling landslide susceptibility with sampling either a single point for the main scarp or a random point anywhere in a landslide polygon was conducted by Petschko et al. (2013c). They observed only small differences between the predictive model performances and classified susceptibility maps for the two landslide sampling schemes.

The landslide mapping focussed on distinct and easily detectable morphologies that remain visible after the occurrence of a landslide (McCalpin, 1984). Very old landslides and landslides that were not visible in the available imagery, were not included in the inventory; the full slide with scarp, transportation and accumulation area had to be discernible in the LiDAR-DTM hillshade to be included in the inventory (Petschko et al., 2013a). In areas where human activity on land use was very high, such as agricultural areas, the ability to identify landslides was particularly challenging (Bell et al., 2012). We assume that in agricultural areas fewer landslides were mapped than actually occurred; however, the full extent of the incompleteness of the inventory obtained in the study area remains unknown.

#### 4.2 Explanatory variables

Geomorphically meaningful, explanatory variables available for this study represented terrain conditions, soil properties and potential tectonic activity.

Several terrain attributes were derived from the LiDAR-DTM as proxies for geomorphic and hydrological processes. We used SAGA GIS (Conrad, 2007) to calculate slope angle (°), slope aspect (°, transformed using the sine and cosine representing east versus west and north versus south; Brenning, 2009), overall curvature (all calculated with second-degree polynomial approximation; Zevenbergen and Thornes, 1987), a topographic wetness index (SAGA wetness index; Boehner et al., 2002), catchment height and catchment area calculated with multiple flow direction algorithm (Freeman, 1991; Quinn et al., 1991), as well as convergence indices, calculated with 10 m and 50 m radius, respectively, to represent morphological changes on different

scales. Data preparation was performed within the R package RSAGA (Brenning, 2011).

Van Westen et al. (2008) had discussed the relevance of most of these terrain attributes to landslide susceptibility. The relationship of slope angle with landslide activity is well known from general slope stability literature (e.g. Crozier, 1986). Slope aspect can be used as a proxy for bedding orientation. It may also reflect differences in intensity of solar radiation, which controls the local temperature and evaporation and therefore soil moisture (Van Westen et al., 2008). Curvature represents convex and concave surfaces related to local morphology ( $3 \times 3$  grid cells). The topographic wetness index was used as a proxy for the soil moisture and ground water level (Beven and Kirkby, 1979; Seibert et al., 2007). The position of the landslide on the slope and the distance from the ridge was represented by the variable catchment height. The catchment area was calculated for the sub-catchments and gives a local representation of the contributing area. Convergence indices were calculated to represent the slope morphology on two different scales by using two different window sizes for the calculation (10 and 50 m). Positive values indicate ridges while negative values indicate local depressions.

Information on the soil properties is also very important as they have an effect on the infiltration capacity and water storage in the soil, which ultimately influences the disposition to landslides (Crozier, 1986). From the soil data set of Eder et al. (2011) we extracted four gridded variables representing permeability ( $\text{mm d}^{-1}$ ; average value of the top 20 cm, and minimum value within 100 cm profiles) and void space (%; average value of the top 20 cm, and average of 100 cm profiles), which may be considered as a proxies for the infiltration capacity.

Available tectonic vector data included fault lines and nappe boundaries (Kurz, 2012). The proximity to a fault line might refer to the occurrence of weakened material that has already been strongly tectonically influenced, reworked and mechanically fragmented. This material generally has lower shear strength and may therefore be more prone to landslides (Crozier, 1986). Furthermore, earth- and debris slides might occur with higher density close to nappe boundaries, as these indicate a distinct difference in the material and permeability. For example, many earth- and debris slides have been observed where the nappe boundary of the Austroalpine Unit with limestone overlays the Rheno-danubian Flysch Zone. The higher water permeability of the Austroalpine limestone on top of denser sandstones and marlstones of the Flysch Zone can result in increased soil water availability at the boundary zone and presence of boundary springs (Schnabel, 1985). The fault lines and nappe boundaries were included in the susceptibility analysis by calculating grids representing the Euclidean distance from the respective features.

Land cover data for Lower Austria was also available for this study, but was not included in the analysis. Since the ages of the landslides were not available in the inventory, it would

not be accurate to associate the landslide distribution with present land cover conditions (Petschko et al., 2013b). We simply do not know how land cover had historically changed and influenced all of the landslides in this inventory, which is important to understand the changing conditions of slope stability (Glade, 2003; Beguería, 2006b; Van Den Eeckhaut et al., 2009; Bathurst et al., 2010). For example, areas which are forested today might not have been forested for a long time period in the past, e.g. due to influences of mining ore in the study area (Lettner and Wrška, 2011). Therefore, the landslide susceptibility models were prepared focussing on local terrain conditions (Brabb, 1984), which are relatively static compared to the dynamic nature of land cover (Van Westen et al., 2008).

## 5 Methods

### 5.1 Modelling of heterogeneous areas

The heterogeneity of geotechnical and topographical characteristics over a study area should be considered when modelling landslide susceptibility (Lee et al., 2008; Blahút et al., 2010). Modelling separately in lithology units is one approach to addressing this heterogeneity (Petschko et al., 2012). This approach avoids the use of interaction terms to represent lithology-dependent processes and preparatory factors, and thus facilitates easier interpretation of the models. Fitting the model with individual modelling domains based on the lithology units is expected to improve overall predictive performance by accurately representing the diversity of geotechnical conditions across the study area (Blahút et al., 2010; Petschko et al., 2012; Trigila et al., 2013).

Accordingly, the study area was divided into 16 modelling domains based on a simplified 1 : 200 000 lithological map (Fig. 1; Schnabel, 2002; Bell et al., 2013). The lithological map gives details on the parent material available for soil formation. This determines the geotechnical characteristics on a scale of 1 : 200 000. Data on the parent material was used as a proxy for these characteristics, as no appropriate geotechnical data covering the entire study area was available. Lithology units with no observed landslides were merged with geotechnically similar units to create homogeneous modelling domains (Table 1). The final susceptibility map of the province was obtained by merging the individual susceptibility maps of the 16 domains.

### 5.2 Generalized additive model

Among the currently available methods for landslide susceptibility modelling a generalized additive model (GAM) is a compromise between the nonlinear predictive flexibility of machine learning algorithms and the smooth, yet linear, predictions of GLMs such as logistic regression. The model fit of the GAM can be easily interpretable, unlike most machine learning algorithms (Brenning, 2008; Goetz et al., 2011). In

addition, the GAM is more suitable for representing the non-linear response of slope stability to changing site conditions than the GLM (Goetz et al., 2011).

The basic design of a GAM is to replace the linear function of each co-variate as used in a GLM with an empirically fitted smooth function to “let the data show the appropriate functional form” (Hastie and Tibshirani, 1990). Thus a GAM allows the combination of linear and nonlinear smoothing functions in an additive manner to describe the relationship between explanatory and response variables. In the case of the logistic additive model for binary (presence/absence) response variables, the response variable is not modelled directly, but using the logit of the occurrence probability  $p(\mathbf{x})$  conditional on explanatory variables  $\mathbf{x} = (x_1, \dots, x_m)^T$  (Hastie and Tibshirani, 1990):

$$\begin{aligned} \text{logit}(\mathbf{x}) &= \log(p(\mathbf{x})/(1-p(\mathbf{x}))) \\ &= \beta_0 + \beta_1 f_1(x_1) + \dots + \beta_m f_m(x_m) \end{aligned} \quad (1)$$

where the functions  $f_i$  are nonparametric smoothers. The quantity

$$\text{odds}(\mathbf{x}) = p(\mathbf{x})/(1-p(\mathbf{x})) \quad (2)$$

is referred to as the odds. Thus, the logistic GAM is additive at the logit level, but increases in  $f_i$  have a multiplicative effect on the odds.

We used a combined backward and forward stepwise variable selection based on the Akaike Information Criterion (AIC), which measures the goodness-of-fit while penalizing for model complexity to obtain a parsimonious model (Akaike, 1974). Smaller models help to keep the estimated coefficient standard errors small and prevent the model from overfitting, which occurs when the number of variables in the model is large relative to the number of landslide samples (Hosmer and Lemeshow, 2000). Overfitting refers to an algorithm or model that performs very well on the available training data to which it is fitted, but poorly on future or new test data and therefore produces unreliable predictions (Hand, 1997; Hosmer and Lemeshow, 2000).

Our study design was a case-control study with the mapped landslide samples as cases and randomly selected non-landslide samples as controls. The landslide susceptibility maps were derived for each modelling domain from a model (GAM<sub>1</sub>–GAM<sub>16</sub>) using all landslide samples while the model performance is assessed using cross-validation, i.e. subsamples of landslide cells (Sect. 5.3). The sampling of non-landslide cells for the entire study area was based on a density of 2 % of all grid cells. An equal number of cases and controls (1 : 1) was used for each model fitted; the landslide samples were matched to an equal number of randomly selected non-landslide samples. This gives the sample size in the respective modelling domain.

It was necessary to adjust each model’s raw predictions based on the corresponding sampling rate to account for the general relative landslide susceptibility of each modelling

domain. We adjusted the prediction by considering the sampling rate ( $\tau_0/\tau_1$ ) of each domain, using Eq. (3),

$$\text{odds}^*(\mathbf{x}) = \tau_0/\tau_1 \times \text{odds}(\mathbf{x}), \quad (3)$$

where

$$\begin{aligned} \tau_0 &= \text{number of sampled non-landslide cells} / \\ &\text{total number of non-landslide cells} \end{aligned} \quad (4)$$

and

$$\begin{aligned} \tau_1 &= \text{number of sampled landslide cells} / \\ &\text{total number of landslide cells} \end{aligned} \quad (5)$$

and  $\text{odds}(\mathbf{x})$  is the unadjusted prediction, in our case, based on training a model with an equal number of landslides to non-landslides (1 : 1) in each domain. The  $\text{odds}^*(\mathbf{x})$  gives the adjusted odds (Scott and Wild, 1986; Hosmer and Lemeshow, 2000). As all landslides were used in this step,  $\tau_1 = 1$ . The adjusted probability  $p^*(\mathbf{x})$  was calculated from  $\text{odds}^*(\mathbf{x})$  which was therefore comparable among different modelling domains. GAM modelling was performed with the `gam` package within the open-source statistical software R (Hastie, 2011; R Development Core Team, 2011).

The adjusted probabilities of landslide susceptibility of the merged map were further classified into discrete classes of low, medium and high susceptibility based on thresholds related to the percentage of observed landslides falling within each susceptibility class. This was a result of testing different thresholds and checking them in the field according to the best geomorphic and planning plausibility. A threshold for the low susceptibility class was selected so that 5 % of the observed landslides having the lowest predicted probabilities would fall within this class. An additional 25 % of the observed landslides would fall within the medium susceptibility class, while the high susceptibility class contained the 70 % of the landslides with the highest predicted probabilities. The corresponding class thresholds of 0.00209 and 0.0141 were also applied to classify the maps corresponding to the approximate upper and lower confidence limits calculated for each grid cell at the 95 % confidence level (refer to Sect. 5.5).

### 5.3 Spatial and non-spatial cross-validation

The assembly of the test data for performance assessment can be achieved in three ways: (1) random subsampling, (2) spatial subsampling and (3) temporal partitioning of the landslide data (Chung and Fabbri, 2003). As there was no information on the landslide age or time of occurrence available for temporal partitioning, we focused on testing random and spatial subsampling of cases and controls within each homogeneous modelling domain.

We used non-spatial and spatial  $k$ -fold cross-validation to assess each model’s predictive performance as a measure of



the model's non-spatial and spatial transferability (Kohavi, 1995; Townsend Peterson et al., 2007). With random subsampling traditional non-spatial cross-validation was performed. Furthermore, spatial cross-validation was applied by spatial subsampling based on  $k$ -means clustering classification of point coordinates (Ruß and Brenning, 2010). In  $k$ -fold cross-validation,  $k$  (here:  $k = 5$ ) randomly or spatially selected disjoint subsamples, or folds, are derived. The model was fitted  $k$  times on the combined data of  $k-1$  folds and tested on the data of the remaining fold by applying the fitted model to the test fold and calculating the performance measure (Kohavi, 1995).

In order to obtain results that are independent of a particular partitioning, cross-validation was repeated  $r$  times (here:  $r = 20$ ), and the median and interquartile range of the 20 outcomes was calculated. This resulted in 100 different estimates of the performance measures.

We use the AUROC as a performance measure, which was derived by comparing the sensitivity of a model (proportion of true positives) to the specificity (more precisely, 1-specificity, or false positives rate; Hosmer and Lemeshow, 2000). The AUROC takes values between 0 and 1 where a value of 0.5 would be achieved by pure chance agreement between predictions and observations and a value of 1 represent perfect discrimination (Brenning, 2005; Guzzetti et al., 2006); however, this may also indicate overfitting. Thus, the AUROC measures the model's ability to discriminate landslide and non-landslide cells (Hosmer and Lemeshow, 2000).

Spatial and non-spatial cross-validation was also applied to test the effects of reducing the sample size on the interquartile range of the AUROC values and the thematic consistency within one domain. This was performed within the modelling domain Flysch Zone. The cross-validation was applied 9 times (with 5 folds and 20 repetitions each) while the sample size of the training sample (including landslide and non-landslide samples) is reduced from 12 562 to 6400, 3200, 1600, 800, 400, 200, 100 and to 50 samples. The test sample was determined constant with a sum of 2000 landslide and non-landslide sample cells. Spatial and non-spatial cross-validation was performed with the `sperrorest` package in R (Brenning, 2012b).

#### 5.4 Transferability and thematic consistency indices

The non-spatial and spatial transferability were expressed by the interpretation of the estimated interquartile range (IQR) of the AUROC values resulting from the non-spatial and spatial cross-validation of each modelling domain. The lower the estimated IQR, the better we considered the non-spatial and spatial transferability of the model within the modelling domain. However, sample size differences among modelling domains result in differences in sampling variability of AUROC estimators, which then has an influence on the IQR of AUROC among cross-validation repetitions. In order to account for this contribution to sampling variability and be

able to provide a transferability measure that was comparable among modelling domains, the IQR had to be adjusted according to the sample size. We define a transferability index by adjusting the estimated interquartile range (IQR) for the contribution of test-set estimation to sampling variability. For this purpose, we calculated the approximate standard error (SE) of the AUROC (AUC) estimator on a test set of  $N$  landslide and  $N$  non-landslide samples using the equation presented by Hanley and McNeil (1982):

$$SE = \left( AUC \times (1 - AUC) + (N - 1) \times (Q_1 - AUC)^2 + (N - 1) \times (Q_2 - AUC)^2 \right) / N^2. \quad (6)$$

The quantities  $Q_1$  and  $Q_2$  were calculated from the AUROC (AUC) value as shown by the following equations:

$$Q_1 = AUC / (2 - AUC) \quad (7)$$

$$Q_2 = 2 \times AUC^2 / (1 + AUC). \quad (8)$$

Since the IQR is approximately equal to 1.35 times the corresponding standard deviation under normal distribution, we used the following equation to remove the contribution of SE to the IQR of the AUROC, and refer to this quantity as the transferability index  $T$ :

$$T = \sqrt{IQR^2 - (1.35 \times SE)^2}. \quad (9)$$

When slightly negative terms occurred under the square root due to the approximation used, these values were replaced by a value of zero.

A higher value of  $T$  indicates a greater variation in predictive performance among models fitted to cross-validation training partitions, which can be interpreted as a poorer transferability. We refer to the transferability calculated by spatial (non-spatial) cross-validation as spatial (non-spatial) transferability  $T_{sp}$  ( $T_{nsp}$ ).

Furthermore, the results of the non-spatial and spatial cross-validation provide an estimate on the variable importance and the thematic consistency in each modelling domain. Based on the 100 models fitted on the different cross-validation training sets within each modelling domain, using stepwise variable selection we assess the importance of each variable in predicting landslide susceptibility using its selection frequency (Goetz et al., 2011). This variable-selection frequency can also be interpreted as a proxy for the thematic consistency of the model (Guzzetti et al., 2006). The thematic consistency is an additional measure to the transferability index showing the model's (and the stepwise variable selection's) sensitivity to sample variation. A high thematic consistency is preferable and describes a good quality of the model (Guzzetti et al., 2006). To formalize this concept, we introduce a thematic consistency index that measures the agreement between variable choices among cross-validation repetitions. In analogy to the Gini impurity index

used in classification (Hand, 1997), we define the consistency index  $C$  by

$$C = (p_1(1 - p_1) + \dots + p_m(1 - p_m)) / (0.25 \times m), \quad (10)$$

where  $p_i$  is the proportion of models that include the  $i$ th predictor variable out of  $m$  predictors. The consistency index is calculated for each modelling domain and for spatial ( $C_{sp}$ ) and non-spatial cross-validation ( $C_{nsp}$ ). Good thematic consistency is achieved when each variable has a selection frequency either close to 0% or 100%, resulting in a  $p_m(1 - p_m)$  value near 0. Therefore, a low value of  $C$  indicates a strong consistency among models. Selection frequencies around 50% indicate weak thematic consistency and produce  $p_m(1 - p_m)$  values of up to 0.25 and a maximum  $C$  value of 1.

### 5.5 Spatially varying prediction uncertainties

The basis for the visualization of the landslide susceptibility map is the predicted probability of the occurrence of landslides of each grid cell, which is computed from the predicted logit. These predictions are subject to uncertainty due to sampling variability and model error, which can be expressed by the standard error of model predictions. This standard error further provides a means to determine approximate upper and lower confidence limits for the predicted logit and ultimately the predicted probability. These limits define an interval within which the true logit or probability of sites with given values of the explanatory variables is located with the chosen confidence of, for example, 95% (Hosmer and Lemeshow, 2000). In other words, we have strong confidence that the true probability of landslide occurrence at a given type of location is within the confidence interval, but we would hesitate to claim that the true probability falls within any narrower range of values within the confidence interval.

In this study, confidence interval estimates for the predicted logit and probability are of special interest in order to assess the implications of spatially varying uncertainties for the interpretation of the final classified susceptibility map. Since the available GAM implementation (Hastie, 2011) does not provide standard errors or confidence intervals for “new” objects that are not included in the training sample, we proceed as follows to estimate standard errors for each location in the prediction map. We first compute standard errors on the logit and probability levels for all sample cells. A lookup table is then used to transfer these standard errors to all grid cells of the raster based on the similarity of the values of explanatory variables used by the model. Tolerance thresholds were applied to each explanatory variable to identify suitable observations in the training sample that match any given prediction location and therefore have similar standard errors. Several tolerance thresholds were tested for each modelling domain to maximize the  $R^2$  obtained by comparing, on the training sample, standard errors estimated

by table lookup to the standard errors calculated by the GAM implementation. This results in a raster data set which gives an estimation of the standard error of the predicted logit for each grid cell.

Based on these approximated standard errors we estimate the approximate upper and lower limit of a 95% confidence interval of the predicted logit using a normal approximation. These logit-scale confidence intervals are further converted to the probability level and adjusted based on each modelling domain’s sampling rate (Sect. 5.2). The approximate upper and lower confidence limits and the predicted probability are used to compare the spatially varying uncertainties in a classified landslide susceptibility map. Therefore, the classified susceptibility map is compared to the classified maps of upper and lower confidence limits to assess potential areas and grid cells in which overlaps of the susceptibility classes (high, medium or low) may occur.

Furthermore, the approximate logit-scale standard errors from each model’s predictions are used as relative uncertainty indices of the susceptibility map within each domain. This uncertainty index allows for a more nuanced visualization of prediction uncertainties within each domain, knowing, however, that its interpretation is only applicable within the domain since no adjustment for the domain-specific sampling rate is applied to this index.

To communicate the possible effects of the prediction uncertainty on the classified susceptibility map, a static way of visualization of the results was selected (Elith et al., 2002). Therefore, the classified susceptibility map was presented alongside a map of possible class overlaps considering the upper and lower confidence limits and with a map of the standard errors (unclassified).

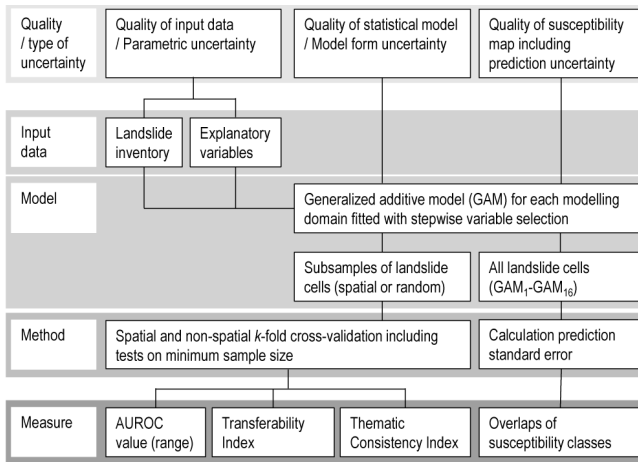
A flow chart of the analysis in this study points out the relation of the different aspects of quality raised in Sect. 2 with the presented methods and measures (Fig. 3).

## 6 Results

### 6.1 Landslide susceptibility map

The three classes of the final landslide susceptibility map classified according to the proportion of landslides included covered 75% (low susceptibility), 19% (medium susceptibility) and 6% (high susceptibility) of the total study area (Fig. 4).

The variable frequency analysis showed that different subsets of the available 15 explanatory variables were included in the GAMs (GAM<sub>1</sub>–GAM<sub>16</sub>) for the different modelling domains (Table 2). The total number of variables used in the models (GAM<sub>1</sub>–GAM<sub>16</sub>) ranged from four variables in the Bohemian Massif with plutonic rock and Waschberg Zone, including Bohemian Massif with sedimentary cover domains (52 landslides each), to 11 variables used in the model for the Flysch Zone (6281 landslides, Table 3). The number of

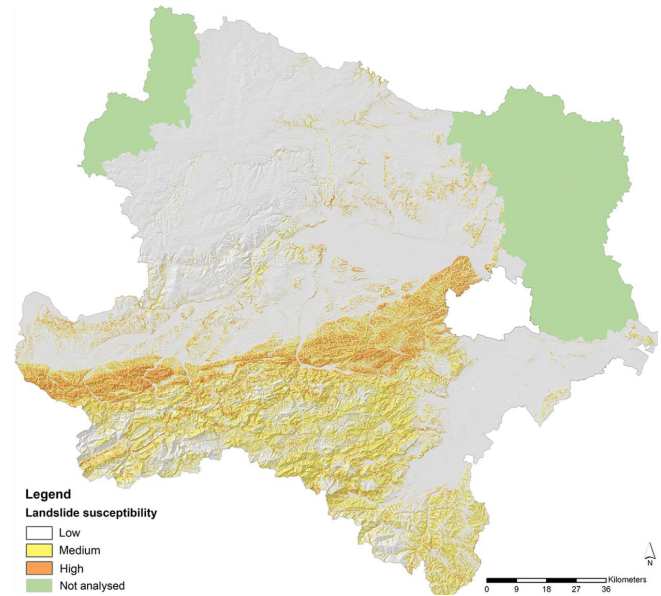


**Fig. 3.** Flow chart presenting the methods and measures applied within this study assessing the quality of input data, quality of statistical model and quality of final susceptibility map.

variables included in the models generally increased with the number of observations in the training sample, which was attributed to the AIC’s penalization based on model complexity relative to sample size. Furthermore, 65 % of the variables were used in a smoothed form (Table 2). However, mainly linear model terms were selected in four modelling domains. A similar overall frequency of nonlinear model terms was obtained in the models fitted within the cross-validation procedures (71 % nonlinear overall) with very similar results for the spatial and non-spatial techniques (refer to Table 2 for details). Two domains (Loess, Loam and Waschberg Zone and Bohemian Massif with sedimentary cover) primarily used linear model terms in spatial and non-spatial cross-validation. Additionally, in the Bohemian Massif with plutonic rock (only in spatial cross-validation) and in the alluvial deposits including lake and wetland (only in non-spatial cross-validation) high proportions of linear terms were observed.

**6.2 Spatial and non-spatial cross-validation**

In general, spatial cross-validation had a larger range of AUROC values than non-spatial cross-validation over all cross-validation repetitions (Fig. 5). However, the median AUROC values were very similar; non-spatial cross-validation had only slightly higher median values. The highest median AUROC values of 0.98 (spatial) and 0.99 (non-spatial cross-validation) were found in the alluvial deposits including lakes and wetland domain. With the exception of the Permo-Mesozoic rocks domain, all median AUROC values are higher than 0.74 (Fig. 5). In this domain the largest differences of median AUROC values between spatial (0.53) and non-spatial cross-validation (0.79) were found. Furthermore, the lowest 1st quartile AUROC value, which was higher performing with non-spatial cross-validation (0.73) than spatial



**Fig. 4.** Resulting classified landslide susceptibility map for Lower Austria. Data source: LiDAR-DTM hillshade – Provincial Government of Lower Austria.

cross-validation (0.35), was found for the Permo-Mesozoic rocks domain. Additionally, this domain showed the highest interquartile range of AUROC values (0.42) with spatial cross-validation and the second highest (0.11) with non-spatial cross-validation.

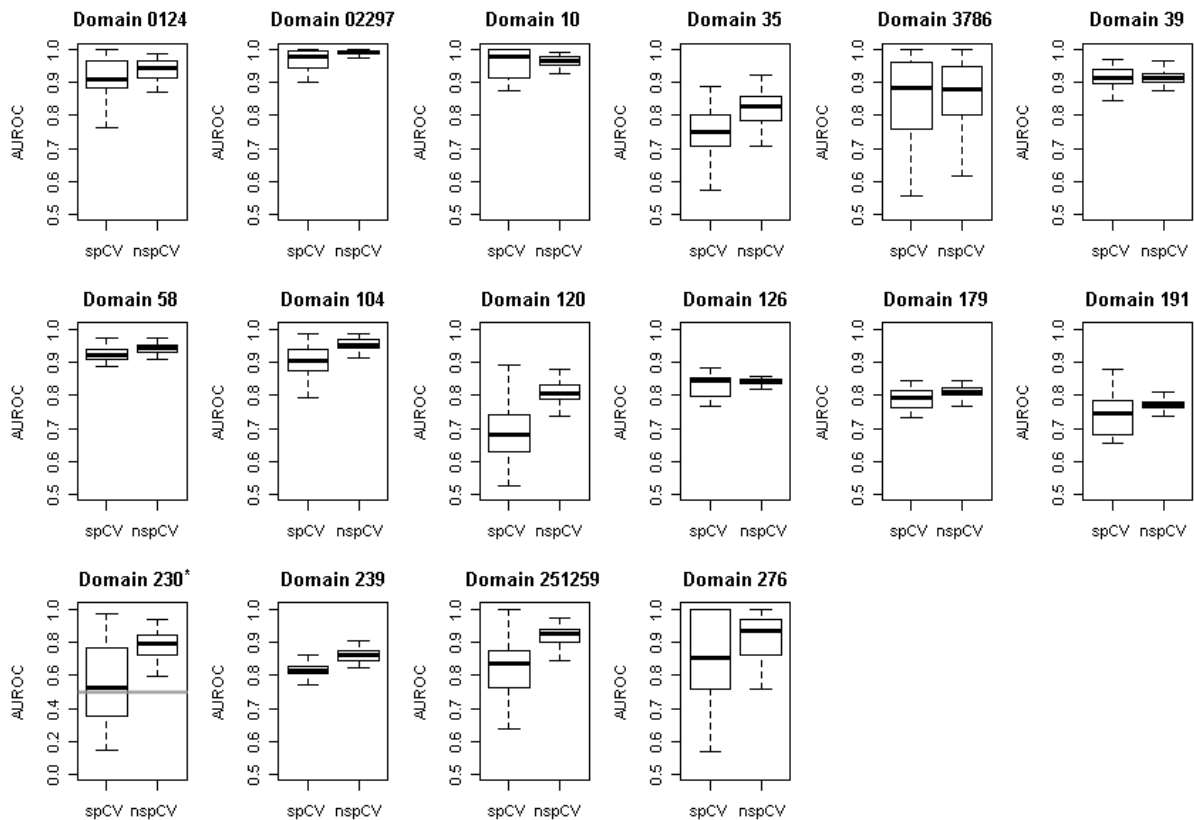
Three domains had very high AUROC values of the 3rd quartile, which ranged from 0.97 to 1 in the spatial and non-spatial cross-validation. These were the loess and loam, the Bohemian Massif with plutonic rock and the alluvial deposits including lakes and wetland domains.

The poorest spatial and non-spatial transferabilities assigned at  $T_{sp} > 0.10$  and  $T_{nsp} > 0.04$  were obtained in the three modelling domains with the smallest sample sizes (Table 3). Transferability tended to be better in domains with larger sample sizes and/or higher landslide densities, but was unrelated to median AUROC. Among the domains with larger sample sizes, the Austroalpine Unit with dolostone stood out with relatively poor spatial transferability ( $T_{sp} = 0.098$ ) in addition to its relatively low median AUROC of 0.75. Spatial transferabilities were best ( $T_{sp} < 0.03$ ) for igneous rocks of the Austroalpine Unit, the Molasse Zone and the Schlier Zone.

Reducing the sample size from using the total number of landslide and non-landslide samples in the case-control sample (12 562) to 50 samples within a modelling domain (Flysch Zone) still resulted in acceptable median AUROC values ( $> 0.76$ ) for all sample sizes. However, the median AUROC values decreased from 0.84 to 0.76 in both the spatial and non-spatial cross-validation (Fig. 6) with very little difference between both approaches. While the AUROC values stay relatively constant until a sample size of 3200,

**Table 2.** Variable frequency for the model using all landslide cells (GAM<sub>1</sub>–GAM<sub>16</sub>) and variable-selection frequency of variables used linearly (*N*) or with a smoothing function (*S*) in spatial (spCV) and non-spatial (nspCV) cross-validation. All values are summarized over all modelling domains.

	Variable frequency (GAM <sub>1</sub> –GAM <sub>16</sub> )		Variable frequency total (GAM <sub>1</sub> –GAM <sub>16</sub> )	Relative variable-selection frequency from cross-validation					
	<i>N</i>	<i>S</i>		All spCV	All nspCV	<i>N</i> spCV	<i>N</i> nspCV	<i>S</i> spCV	<i>S</i> nspCV
Slope angle	2	13	15	91.8	95.7	9.1	17.3	82.8	78.4
Curvature	2	10	12	53.8	58.6	7.1	1.5	46.7	57.1
Topographic wetness index	4	7	11	68.3	61.4	15.9	13.5	52.3	47.9
Catchment height	4	6	10	55.6	65.0	26.3	21.1	29.3	43.9
Convergence Index (10)	3	6	9	48.8	44.1	16.5	19.8	32.3	24.4
Euclidian distance to tectonic lines	4	5	9	53.4	53.3	23.6	18.3	29.8	34.9
Euclidian distance to nappe boundaries	4	5	9	74.2	56.1	24.1	11.8	50.1	44.3
Convergence Index (50)	3	5	8	56.1	64.3	21.3	26.3	34.8	38.0
Void space (0–20 cm)	3	4	7	0.3	0.2	0.0	0.1	0.3	0.1
Catchment area (log)	2	4	6	43.4	45.0	4.9	10.2	38.6	34.8
Permeability (0–20 cm)	3	1	4	36.9	23.7	15.3	8.3	21.6	15.4
East aspect	2	1	3	9.4	6.5	5.2	4.8	4.2	1.8
North aspect	2	0	2	11.8	19.5	8.1	17.9	3.7	1.6
Void space (mean 0–100 cm)	0	2	2	0.0	0.0	0.0	0.0	0.0	0.0
Permeability (min)	0	0	0	1.1	4.9	0.5	0.4	0.6	4.5



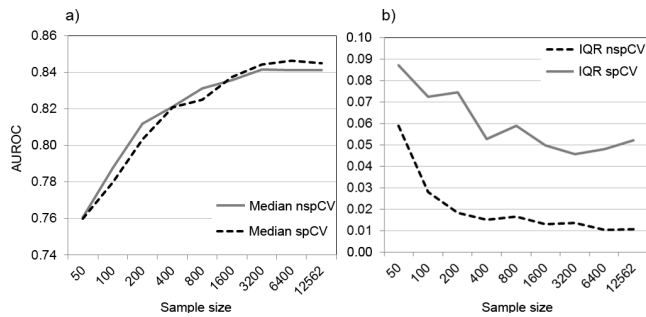
**Fig. 5.** Box plots showing the range of AUROC values resulting from repeated *k*-fold cross-validation with spatial subsampling (spCV) and non-spatial subsampling (nspCV) for each modelling domain. The y axis limits for this domain range from 0 to 1 and a grey line marks the 0.5 value.

they started to decrease more rapidly from there on as the sample size decreases. Despite that, the interquartile ranges were substantially higher with spatial cross-validation, and

generally increased with decreasing sample sizes in both spatial (0.052–0.087) and non-spatial (0.011–0.059) cross-validation. Below a sample size of about 400 (spatial cross-

**Table 3.** Number of selected variables for the models GAM<sub>1</sub>–GAM<sub>16</sub>. Median AUROC values, transferability and thematic consistency of the modelling domains for spatial (spCV) and non-spatial (nspCV) cross-validation. Lower index values  $T_{sp}$ ,  $T_{nsp}$  and  $C_{sp}$ ,  $C_{nsp}$  indicate strong spatial and non-spatial transferability and thematic consistency, respectively. Minimum (Min.), maximum (Max.) and range of standard error (SE) of the prediction on logit level.  $R^2$  obtained by comparing standard errors estimated by the lookup table to the standard errors calculated by the GAM implementation.

Domain	Name	No. of selected variables	Median AUROC spCV	Median AUROC nspCV	Transferability index $T_{sp}$	Transferability index $T_{nsp}$	Consistency index $C_{sp}C_{nsp}$	Consistency index	Min. SE (logit)	Max. SE (logit)	SE range (logit)	$R^2$
239	Igneous rocks of the Austroalpine Unit	7	81.7	86.2	0.000	0.000	0.247	0.151	0.12	1.61	1.49	0.77
58	Molasse, Schlier	6	92.2	94.3	0.015	0.000	0.364	0.191	0.19	2.35	2.16	0.87
39	Molasse Zone	6	91.4	91.4	0.029	0.000	0.418	0.326	0.19	2.63	2.44	0.76
35	Debris, till	7	75	82.7	0.042	0.023	0.352	0.253	0.25	3.61	3.36	0.76
02297	Alluvial deposits including lakes and wetlands	7	97.9	99.3	0.042	0.000	0.339	0.389	0.44	3.12	2.68	0.76
179	Austroalpine Unit with limestone, marls and sandstone	8	79.5	81.2	0.046	0.000	0.176	0.109	0.08	1.63	1.55	0.84
126	Rheno-danubian Flysch Zone	11	84.5	84.1	0.051	0.002	0.099	0.180	0.05	1.62	1.57	0.82
104	Intramontane basins	7	90.8	95.4	0.052	0.000	0.164	0.324	0.25	2.20	1.94	0.79
0124	Quaternary fluvial terrace and anthropogenic deposits	6	90.8	94.3	0.070	0.033	0.363	0.283	0.23	1.69	1.46	0.86
10	Loess, Loam	5	97.8	96.7	0.084	0.010	0.349	0.174	0.27	8.73	8.46	0.90
251259	Bohemian Massif including Fault Zone	6	83.4	92.5	0.091	0.000	0.377	0.251	0.24	1.73	1.49	0.79
120	Mélange Zone, Klippen Zone	8	68.1	80.7	0.098	0.000	0.210	0.081	0.19	1.84	1.65	0.51
191	Austroalpine Unit with dolostone	9	74.7	77	0.098	0.000	0.162	0.068	0.09	1.66	1.57	0.75
3786	Waschberg Zone including Bohemian Massif with sedimentary cover	4	88.4	88	0.170	0.102	0.424	0.235	0.35	4.38	4.03	0.62
276	Bohemian Massif, plutonic rock	4	85.6	93.4	0.208	0.073	0.419	0.323	0.37	16.25	15.88	0.67
230	Permo-Mesozoic rocks (overlying the Austroalpine igneous rocks)	6	52.4	79	0.397	0.045	0.519	0.319	0.27	2.92	2.65	0.57



**Fig. 6.** (a) Median AUROC value and (b) interquartile range of AUROC values (IQR) at reduced sample sizes (numbers given refer to the total number of equally distributed landslide and non-landslide points) in the Flysch Zone (Domain 126) resulting from spatial cross-validation (spCV, grey line) and non-spatial cross-validation (nspCV, black dashed line).

validation) and 200 (non-spatial cross-validation) the interquartile range of the AUROC values sharply increased as the sample size decreased; thus the transferability of the model decreased substantially for sample sizes smaller than 400 (spatial cross-validation) and 200 (non-spatial cross-validation). In addition, the smaller sample sizes led to lower thematic consistency of the model.

### 6.3 Variable importance and thematic consistency

In the comparison of the models fitted within the 16 modelling domains (GAM<sub>1</sub>–GAM<sub>16</sub>), topographic variables were more important than the available variables on soil properties. Out of a maximum of 16 selections, the variable slope angle was selected 15 times within the stepwise variable selection, whereas the minimum permeability value

was not selected at all (Table 2). However, this only shows the result of one specific random sample and variable selection. According to the relative variable-selection frequency resulting from the two cross-validation approaches, the variable importance for predicting the landslide susceptibility also changed distinctly between the modelling domains. All modelling domains (except the domain of the Permo-Mesozoic rocks) had slope angle as the most important variable. It was selected on average in 91.8 and 95.7% of the model repetitions in spatial and non-spatial cross-validation. Another important variable was catchment height, which was selected 55.6% (spatial) and 65% (non-spatial) of the repetitions. In the spatial cross-validation the Euclidian distance to nappe boundaries was the second most important variable as it was selected in 74.2% of the models (non-spatial selection frequency 56.1%, rank 6). The topographic wetness index (68.3% / 61.4%), the convergence index (50 pixel radius; 56.1% / 64.3%) and the curvature (53.8% / 58.6%) were also among the top ranking variables in both cross-validation approaches. Void space (mean 0–100 cm) was not selected in any of the model runs, while void space (0–20 cm) was selected by less than 1% of the models and the minimum permeability was included in 1.1% (spatial cross-validation) to 4.9% (non-spatial cross-validation) of the models on average over all modelling domains.

Overall the thematic consistency was stronger within the non-spatial cross-validation because training sets are less variable when the data are partitioned randomly as opposed to spatially (Table 3). A strong thematic consistency was assigned for a consistency index of  $C_{sp}$ ,  $C_{nsp} < 0.20$  and was found for seven domains in the non-spatial cross-validation but only for four domains in the spatial cross-validation. The strongest thematic consistency was observed in the Flysch Zone ( $C_{sp} = 0.099$ ) in spatial cross-validation and in

the Austroalpine Unit with dolostone ( $C_{\text{nsp}} = 0.068$ ) and the Klippen Zone ( $C_{\text{nsp}} = 0.081$ ) in the non-spatial cross-validation. These domains had a very large sample size and landslide density and also a good spatial or non-spatial transferability. While in spatial cross-validation the domains with the smallest sample size and poorest spatial transferability had the weakest thematic consistency, in non-spatial cross-validation the consistency index was unrelated to sample size and transferability. Among these, the Waschberg Zone (including the Bohemian Massif with sedimentary cover domain) gave very contrasting results. It showed a weak thematic consistency in spatial cross-validation but a medium consistency index in non-spatial cross-validation. The weakest thematic consistency was found for the Permo-Mesozoic rocks ( $C_{\text{sp}} = 0.519$ ) and the alluvial deposits including lakes and wetlands ( $C_{\text{nsp}} = 0.389$ ) domains. In non-spatial cross-validation the thematic consistency was stronger with lower median AUROC values. However, in spatial cross-validation the median AUROC values and the thematic consistency were unrelated.

#### 6.4 Spatially varying prediction uncertainties

The largest range of logit-scale standard errors was obtained for the Bohemian Massif with plutonic rock domain (0.37–16.24; Table 3), whereas the Quaternary fluvial terrace had the minimum range (0.22–1.69). The highest standard error of 16.24 was in the Bohemian Massif with plutonic rock domain, while the median of the highest standard errors over all domains was 2.27. After the transformation from logit-scale to probability scale the typical distribution of the standard error was the following: the range of the standard error of the predicted probability was largest for medium probabilities that were in the medium susceptibility class. The minimum range of the standard errors was typically obtained at the minimum and maximum probability that were contained in the low and high susceptibility class. Nevertheless, the lowest standard errors were computed at the minimum predicted probability.

Within the study area of Lower Austria seven types of susceptibility class uncertainties in the classified landslide susceptibility maps were identified by the analysis of overlaps between the susceptibility classes (Figs. 7 and 8). Most commonly, for about 85 % of the grid cells, there were no overlaps of different susceptibility classes either with the lower or the upper confidence limit meaning that in all maps the susceptibility class was the same. The most common overlaps were found for the low susceptibility class where 6 % of the grid cells were classified with low susceptibility in the predicted probability map but with medium susceptibility in the upper confidence limit map (Fig. 7). Moreover, 2 % of the study area experienced overlaps of the medium class in the predicted probability map to the high susceptibility class in the upper confidence limit map. Even fewer grid cells, 0.03 %

had a range from the low class in the predicted probability map to the high class in the upper confidence limit map.

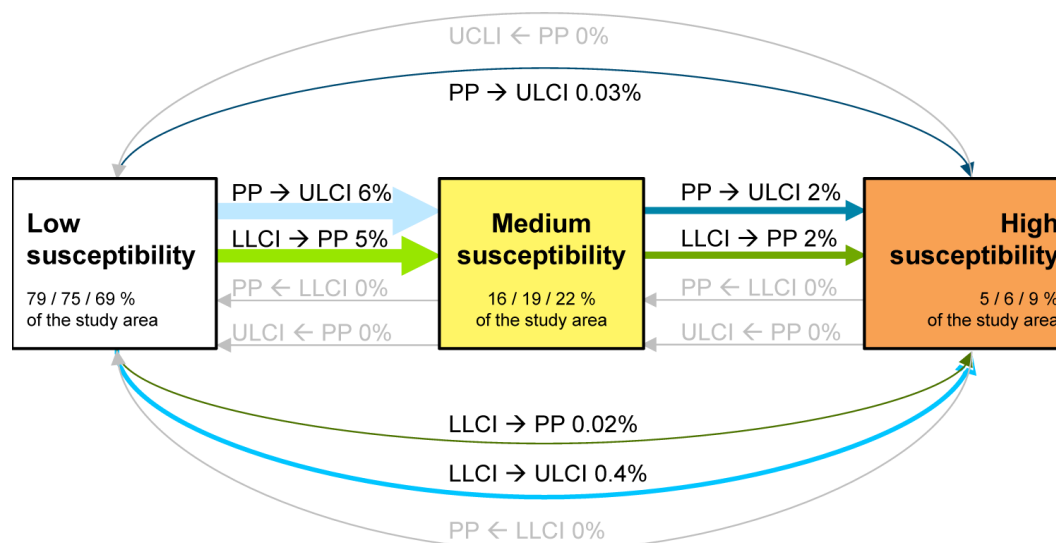
The results of the analysis of spatially varying uncertainties were presented by maps chosen in two exemplary modelling domains with contrasting landslide densities, the Flysch Zone ( $4.6 \text{ km}^{-2}$ ) and the Bohemian Massif ( $0.09 \text{ km}^{-2}$ , Fig. 8). The range of the standard errors of the predicted logit of these domains was very similar. While the standard errors range from 0.05 to 1.6 in the Flysch Zone, the Bohemian Massif had a range from 0.24 to 1.73 (Table 3). With this minimum value the Flysch Zone gives the lowest standard error in the study area. The  $R^2$  resulting from using the lookup table to transfer the standard errors to all grid cells range from 0.51 (in the M elange Zone) to 0.9 (in Loess, Loam; Table 3). In the Flysch Zone the computed  $R^2$  was 0.82 which is only slightly better than in the Bohemian Massif (0.79).

Comparing the classified maps from the Flysch Zone to the Bohemian Massif it can be seen that the grid cells showing an overlap in the Flysch Zone were much more scattered, pixel-wise, and mainly occurred at the boundaries of the susceptibility classes (Fig. 8). In the centre of the high susceptible class no overlaps occurred. However, in the Bohemian Massif the overlaps mainly occurred between the medium and high susceptibility classes and generally covered the entire class or larger areas, instead of single grid cells only. Both maps shared a low frequency of dark blue and dark green grid cells, which showed overlaps from the low class to the high class (either from lower confidence limit to predicted probability or from predicted probability to upper confidence limit). Furthermore, the overlaps of classes were occurring equally in- and outside of permanent settlement areas, which was important when considering the map for future planning purposes (Fig. 8).

## 7 Discussion

### 7.1 Quality of input data

Practical challenges in this study arise from the size of the study area and the intended output map scale of 1 : 25 000. The size of the study area brings along some limitations regarding the availability of data sources that offer a full spatial coverage and a reasonable map scale. This introduces a number of parametric uncertainties into the modelling. Generally a complete, unbiased inventory would be desirable, as for example a full inventory that was mapped directly in the aftermath of a landslide event (single landslide or multiple landslides triggered at the same time) in the area of the susceptibility map (Van Westen et al., 2008). This would allow for the inclusion of land cover or precipitation data in the modelling which might be helpful to learn even more about the landslides in the area and to build scenarios on future landslide susceptibility (or hazard/risk; Begueria, 2006b; Tarolli et al., 2011). However, a complete inventory is rarely



**Fig. 7.** Susceptibility class uncertainty expressed by types of overlaps between susceptibility maps at the lower confidence limit (LLCI) or the upper confidence limit (ULCI) at the 95 % confidence level relative to the class in the predicted probability map (PP). With the susceptibility class the percentage of area in the maps of LLCI, PP and ULCI is given in the box. The arrow thickness is relative to the percentage of affected area. Possible types of overlaps, which did not occur in the study area, are indicated with grey arrows. Next to the type of overlap the percentage of affected area related to the study area is given. In 85 % of the study area, the 95 % confidence limits fall within the same susceptibility class (not indicated here).

available. Particularly for historical inventories, the level and type of completeness is unknown, while it is known that they are generally incomplete (Malamud et al., 2004). Even a substantially complete inventory, which would be useful in statistical modelling as it includes a substantial fraction (random sample) of all landslides at all scales, land use types, lithological units or slope angles, cannot be reached for historical inventories (Malamud et al., 2004). This originates from the observation that landslides and their visibility on aerial photographs or other base maps (e.g. hillshades derived from airborne laser-scanning DTMs) are highly influenced by new landslides, reactivation, erosion, land use type and anthropogenic activities (Bell et al., 2012; Malamud et al., 2004; McCalpin, 1984, Van Westen et al., 2008). Furthermore the mapping and identification of landslides is highly dependent on the experience and knowledge of the investigator (Van Westen et al., 1999; Harp et al., 2002; Ardizzone et al., 2002). If these influences on the completeness are not random they might introduce a bias in the inventory and furthermore in the sampling which results in a model bias or systematic modelling error.

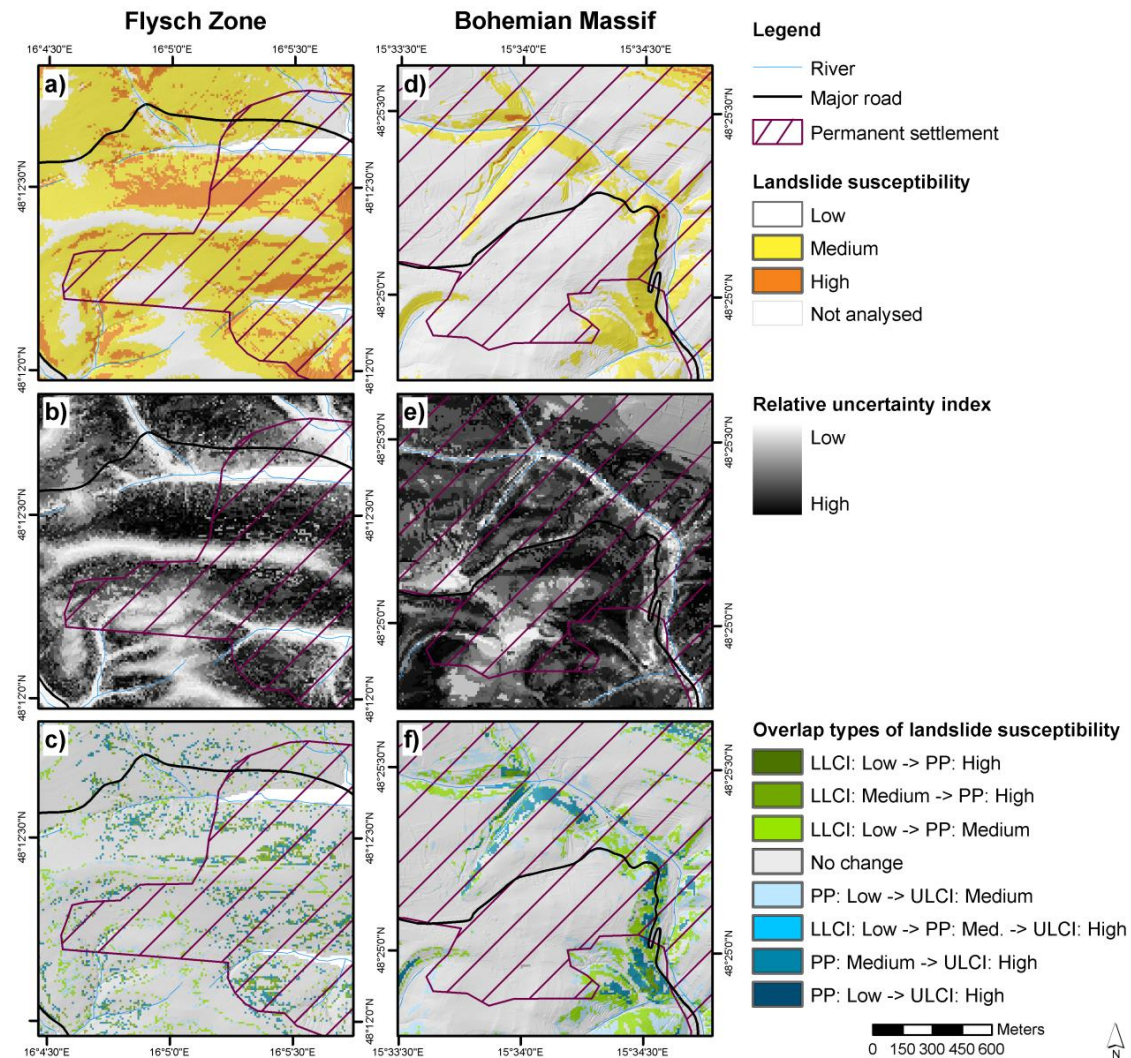
In our study area it is assumed that the inventory is not complete as it originates from recent data sources (not multi-temporal) only and the visibility of landslides in the LiDAR-DTM or orthophoto is influenced by human impact depending on the land use type (Bell et al., 2012). Furthermore, a drawback is that no information on the time of occurrence of the landslide is available. However, the type of incompleteness was not analysed for the entire study area. Therefore,

it is not clear if the missing landslides are missing completely at random or are biased toward absence in certain land uses or lithological units. The implications of an incomplete inventory on the model performance (shown by the AUROC value) were estimated by performing the repeated  $k$ -fold cross-validation using training and test samples. The results show rather high AUROC values for most modelling domains, which indicates that even with an incomplete inventory (training sample) the prediction of landslides of the test sample was successful for most cases. However, sample size is of importance for the model performance. For the discussion of this please see Sect. 7.4.

## 7.2 Quality of statistical model form

### 7.2.1 Study design to meet the heterogeneity of the study area

Observed variable-selection frequencies showed that different explanatory variables were relevant in different domains, which provides additional justification to the decision to model susceptibility in each modelling domain separately. Additionally, not only the different choice of the variables is important but also the way the variables are fitted or smoothed according to the sample in the respective domain. Previous studies showed that within each lithological unit landslides occur at different slope angles (Blahút et al., 2010; Muenchow et al., 2012; Petschko et al., 2012). Similar differences within lithological units or terrain types might be



**Fig. 8.** Landslide susceptibility map (target scale 1 : 25 000) and map uncertainty in an example area. (a), (d) Susceptibility map; (b), (e) relative uncertainty index; (c), (f) susceptibility class uncertainty. Panels (a–c) correspond to an area in the Flysch Zone (very high landslide density), (d–f) are in the Bohemian Massif (very low landslide density). The susceptibility class uncertainty refers to differences between susceptibility maps at the lower confidence limit (LLCI) or the upper confidence limit (ULCI) at the 95 % confidence level relative to the class in the predicted probability map (PP) in panels (a) and (d). Data sources: LiDAR-DTM hillshade, river, major road and settlement – Provincial Government of Lower Austria.

present for other explanatory variables as well, as the geomorphic and geologic characteristics change (Lee et al., 2008). Facing this, our study design gives much more flexibility to represent the characteristics of the study area. Furthermore, it incorporates information on lithology by adjusting the odds of the prediction with the sampling rate of cells in each lithological unit.

However, one may argue that with this approach problems occur at the boundaries of the lithological units. Inaccuracies in the delineation of the lithological map of the area have effects on the model results as the landslides might be assigned incorrect lithological information. This may lead to an underestimation of effect sizes as data from different (true) litho-

logical units would be mixed. Similar mixing effects may occur for quantitative predictor variables as well, for example as a function of positional accuracy for scale and resolution. In regression this effect is known as dilution, which may introduce a bias of estimated regression coefficients toward zero (Frost and Thompson, 2000). As this would also occur using the lithological map as a factor instead of as a mask for partitioning the study area, the boundary inaccuracies are not only a problem in the applied study design.



### 7.2.2 Spatial and non-spatial cross-validation

We assessed the model form uncertainty by deriving a model performance estimate and thematic consistency by spatial and non-spatial cross-validation. Cross-validation estimation of a model's predictive performance is a crucial step in predictive modelling because estimation on the training set is always too optimistic (Hosmer and Lemeshow, 2000; Brenning, 2005). Cross-validation results in bias-reduced performance estimates as the test partitions used in each repetition do not overlap with the training sample (Brenning, 2005). In particular spatial cross-validation is recommendable for spatial data, which may be subject to spatial autocorrelation (Brenning, 2005, 2012a).

The median AUROC values estimated by spatial and non-spatial cross-validation were generally similarly high in this study. However, the median AUROC values and the transferability index clearly showed that non-spatial cross-validation provided a more optimistic or even overoptimistic assessment of the model performance and transferability in contrast to spatial cross-validation. Therefore, spatial and temporal cross-validation should be preferred for performance estimation (Chung and Fabbri, 2008). While spatial performance and transferability do not necessarily reflect a model's predictive performance in the time domain, they provide a more realistic assessment of its ability to generalize from the available data than non-spatial approaches (Brenning, 2005).

The spatially and non-spatially least transferable models in this study were associated with domains that had the smallest sample sizes. The relationship of sample size on predictive abilities has also been shown in other spatial modelling studies (Stockwell and Peterson, 2002; Hjort and Marmion, 2008). However, we believe that the cases of high variation in AUROC values may be also related to the cross-validation sampling variation as indicated by the difference between  $T_{sp}$  and lower  $T_{nsp}$ , and possibly the proportion of stable and unstable terrain in a modelling domain.

Heterogeneity of landslide conditions (e.g. related to topography or land use) in the cross-validation samples is more likely to occur if samples are partitioned spatially, such as the case in spatial cross-validation. Here, similar characteristics of the explanatory variables in both training and test sample are assumed and necessary (Guzzetti et al., 2006). If this assumption is not met by the data (e.g. a rock type or land use class is missing in the test sample) the transfer of the fitted model to the test sample and the estimation of the model performance are difficult (or impossible) (Guzzetti et al., 2006). In our study some model domains might have high contrast between stable (e.g. large flat areas) and unstable (e.g. steep areas) terrain which gives potential for greater variation of sampled terrain conditions; it may be possible that in some samples one terrain condition is overrepresented relative to others. The sampling strategy may be improved further by masking low-lying flat areas that are not typically susceptible to landslides (Van den Eeckhaut et al., 2009; Goetz et al.,

2011). Consequently, the sample may have more potential to capture the differentiating conditions of stable and unstable terrain in an area that is generally susceptible to landslides (e.g. steep hillslopes). This might lead to a smaller variation in the AUROC values.

The high importance of topographic variables in the susceptibility modelling goes along with findings in other studies (Guzzetti et al., 2006; Begueria, 2006b; Van Westen et al., 2008; Guns and Vanacker, 2012). However, it was rather surprising that the soil data on permeability and void space was not selected more often during the stepwise-variable selection. This might be related to the poor resolution of the soil data (50 m × 50 m) and to the usage of topographic data as a proxy for hydrological and soil characteristics (soil moisture and thickness). A strong correlation between the topographic wetness index and soil characteristics was found amongst others by Seibert et al. (2007). Furthermore, the higher resolution of the topographic variables might be of advantage for better describing the local conditions of landslide susceptibility.

Whereas a very strong thematic consistency was generally found for domains with a large sample size and sampling rate, domains with a small sample size and rate showed a high variability in the variable-selection frequencies which gave a weak thematic consistency. Therefore, the weak thematic consistency might also be associated with a poor spatial and non-spatial transferability, both originating from a small sample size and a small sampling rate. This relation was stronger for the spatial cross-validation, while the thematic consistency from non-spatial cross-validation was unrelated.

Generally, the usage of rather static data in the modelling was a necessity resulting from the available landslide inventory. However, this does not give the possibility to include dynamic data on triggers into the modelling or to design scenarios of future landslide susceptibility considering land cover or precipitation change. A clear source of model form uncertainty is the concept of uniformitarianism in modelling. This implies that the predisposing and triggering factors of landslides do not change in future. However, natural variability of landslide triggering mechanisms and also of the climate system might cause future changes. Currently, the influence of climate change on current or future landslide activity is debated. However, no clear evidence on these possible future changes was found in many regions (e.g. Crozier, 2010; Huggel et al., 2012). Additionally, it might be possible that the most important data set explaining the susceptibility might still be missing. This might happen although expert knowledge on geomorphology was applied in selecting geomorphologically relevant data. Moreover, it has to be stated that this type of susceptibility map was designed for the main scarps of landslides. A classified map might cover the possible runout of landslides but by definition does not show how the initiated landslides might move downslope and endanger further areas (Demoulin and Chung, 2007).

### 7.3 Quality of final susceptibility map – visualizing prediction uncertainties

This analysis of prediction uncertainty is an improvement of previous landslide susceptibility studies (Guzzetti et al., 2006; Van den Eeckhaut et al., 2009; Rossi et al., 2010; Sterlacchini et al., 2011), as it not only showed an uncertainty index of the predicted probability on a grid cell basis but additionally provided information on where the susceptibility class uncertainties were located. Some model uncertainties within this method arise from using the lookup table for transferring the prediction standard error to all grid cells as shown by the range of resulting  $R^2$ . This method might be improved or substituted by a function assigning the standard errors to all grid cells.

It was found that in the classified map the majority of grid cells did not change. However, there are differences between the modelling domains where some domains had larger overlaps of different susceptibility classes than others. Special attention should be given to the low susceptibility class. Here, the highest percentage of overlapping classes and underestimation of the susceptibility were detected.

The visualization of these spatially varying uncertainties is of special interest for future land use and development planning usually performed by non-landslide experts. In the aftermath of this study each landslide susceptibility class will be related to, not legally binding, recommendations for the designation of new building areas. Therefore, a misclassification (e.g. low instead of medium susceptibility) might lead to an interpretation by the municipality or landowner that underestimated landslide susceptibility. Knowledge about the susceptibility class overlaps might outline where more caution and detailed investigations are necessary. Additionally, it also shows where no uncertainties are expected, which might help to avoid costs for slope investigations.

There is a need to communicate the research results and their quality with appropriate explanations for the local officials, environmental managers and the public to raise awareness and knowledge about it, which leads to an easier understanding and incorporation of the results into the decision-making process (Knuepfer and Petersen, 2002; Rogers, 2006; Brierley, 2009; Hill et al., 2013). This analysis might aid good acceptance of the landslide susceptibility maps in the local governments, as instead of a fuzzy statement on involved uncertainties, these are clearly shown in a map on grid cell level (Guzzetti et al., 2006; Luoto et al., 2010). Furthermore, the preparation of the susceptibility maps showing the class overlaps contributes to an easier understanding of the possible effects of the prediction uncertainties.

The question if the policy-makers or stakeholders are really interested in knowing more about the uncertainty is discussed conversely. The study of Brugnach et al. (2006) pointed out that the confidence in modelling results is dependent on the way the uncertainties are addressed. Policy-makers were missing more information on the uncertainty

of any model result. Therefore, the modelling results should be presented with a measure of uncertainty or confidence indicator (Brugnach et al., 2006). In habitat suitability modelling the visualization of uncertainty was identified as relevant to inform decision-makers about areas with extreme error, but also about areas which are particularly well modelled (Elith et al., 2002). This openly addresses the uncertainties involved in the maps instead of giving an impression of certainty (Elith et al., 2002). However, interviews of Klimeš and Blahůt (2012) showed that local governments do not want any information on uncertainties.

Nevertheless, the spatially varying prediction uncertainties might have severe consequences on buildings and their inhabitants if an event occurred within the uncertainties of the method used to delineate the hazard zones. The converse discussion shows that more or better communication with the stakeholders or policy-makers (also during the modelling process) is necessary to learn about uncertainties and enlarge confidence into the modelling (Brugnach et al., 2006). However, the way the uncertainties are presented to the stakeholder has to be adapted by the scientist to ensure the success of the communication. The visualization of some aspects of the quality of landslide susceptibility maps, such as the spatially varying prediction uncertainty, can enhance the communication among experts and decision-makers to facilitate informed decisions (Kunz et al., 2011).

Additionally, further aspects of considering and communicating the effects of epistemic uncertainty are still open research fields in susceptibility modelling. A clear assessment of these is necessary to evaluate on their consequences on the susceptibility (or hazard or risk) map.

### 7.4 Considerations on sample size

Summarizing the previously discussed findings some considerations on a minimum sample size might be possible. While the transferability index is less strongly related to sample size or sampling rate, the thematic consistency index shows a stronger relationship to them. Generally, larger sample sizes and sampling rates result in better thematic consistency and transferability of the model. Furthermore, the minimum standard error of the prediction was lower with larger sample size (Table 3).

The effect of a reduced sample size on the median and interquartile range AUROC values was assessed in the Fylsch domain. We found that the median AUROC remained satisfactory high but decreased as sample size decreased, while the interquartile range of the AUROC increased. Even with the smallest sample size the model still achieved a good discrimination between landslide and non-landslide cells according to the median AUROC value. Summarizing the results, a minimum sample size with a sum of around 400 slide and non-slide cells might be recommended for the methods applied in this study. This size leads to an acceptable transferability and thematic consistency of the model in spatial

cross-validation in the analysed modelling domain. However, examples from successfully fitting a susceptibility model with smaller sample sizes (10 landslides with 15 cells each in an area of 177 km<sup>2</sup>; Demoulin and Chung, 2007) give a very contrasting result. This shows that for small sample sizes other modelling methods (statistical or heuristic) might be more suitable. However, some difficulties arise in the modelling with small samples, which might create the need for a larger sample or for selecting a different method. The sample needs to be substantially complete to represent all local terrain conditions (Malamud, et al., 2004). However, increasing the sample size can only be done by enlarging the landslide inventory (e.g. by selecting a larger study area, mapping more landslides). This is challenging, as in some regions no additional data on landslides (or resources for mapping more landslides) might be available due to a general low susceptibility of the area or to missing data.

Giving a recommendation on a minimum sample size is challenging, as this study showed that the general trends found for sample size and sampling rate do not apply for all modelling domains. Therefore, we highlight that the resulting quality estimates (transferability index, consistency index and prediction uncertainty) might additionally be dependent on a combination of the domain size and the landslide density (landslides per km<sup>2</sup>). Also, dependencies on local terrain conditions and their homogeneity in the modelling domain might exist. In this context the provision of the presented quality estimates is recommended, as these clearly show the quality of the results which might be lower compared to studies with larger sample sizes.

Moreover, the geomorphic plausibility of the susceptibility map has to be analysed. Previous studies found that high performance measures do not always guarantee high geomorphic plausibility of the map (Bell, 2007; Trigila et al., 2013). It might be possible that with a smaller sample size the geomorphic plausibility of the map is lower. However, the influence of a small sample size on the geomorphic plausibility of the susceptibility map is unclear. Nevertheless, analysing this was beyond the scope of this study.

## 8 Conclusions

The high quality of landslide susceptibility maps is defined by the low aleatory and epistemic uncertainties involved in the susceptibility modelling. This was analysed in terms of landslide susceptibility model performance and spatially varying prediction uncertainties of the final classified susceptibility map. The analysis gives an overview and some estimates on the epistemic and aleatory uncertainties involved in statistical susceptibility modelling. However, the effect of the propagation of all the single uncertainties on the final map and subsequently to hazard and risk maps has to be analysed further. Considering the present results in high model performance, analysis and visualization of prediction

uncertainties, the applied model and resulting classified landslide susceptibility map are regarded to be of high quality.

The applied study design with modelling in the different domains provides a high flexibility for representing the characteristics of the heterogeneous study area. Opposed to single hold-out validation the repeated *k*-fold cross-validation provides a measure on the precision of the estimator and is independent of the one (random) test sample. The spatial cross-validation gave a more realistic assessment of the model performance and spatial transferability from the available data than non-spatial approaches. This aids as an estimate on the model form uncertainty, which is considered to be low.

A recommendation on an appropriate minimum sample size might be given considering the presented analysis in different modelling domains and the tests on reducing the sample size. According to the results, the larger the sample size the better is the transferability and thematic consistency. Therefore, also the quality of the statistical model form increases with sample size. However, not all modelling domains follow this trend. This might be related to a combined influence of the heterogeneity or homogeneity of the local terrain conditions, the size of the domain and its landslide density. A minimum sample size of around 400 slide and non-slide cells (200 each) might be recommended for the methods applied in this study.

Only rather static data on local terrain conditions could reasonably be included in the analysis given the landslide inventory, with no known landslide age. The stepwise variable selection resulted in a satisfactory thematic consistency. Among the geomorphologically relevant variables, topographic variables were selected with a higher frequency than soil variables. This might be related to the spatial resolution of the respective data. However, this result goes along with comparable studies having topographic variables as the most frequently selected variables. The final landslide susceptibility map gives a good representation of the landslide susceptibility based on topography although the map does not include possible landslide triggers.

Regarding the susceptibility class uncertainties we conclude that the majority of the study area is not affected by class uncertainties. Special attention has to be drawn to possible overlaps of the low and medium susceptibility class in the predicted probability map and the map of the upper confidence limit. A misclassification in the low class might result in an underestimation of the susceptibility. This might have adverse effects on the municipality or landowner if the recommendations for the assignment of building areas might not be restrictive enough.

We discussed that there is a need to assess, minimize and communicate uncertainties involved in susceptibility modelling. The analysis results need to be communicated in an understandable manner to the stakeholder to allow for informed decisions instead of giving an impression of certainty. A possible example was shown by the visualization of the

prediction uncertainty and its effects on the classified landslide susceptibility map.

*Acknowledgements.* This study was conducted within the research project MoNOE – Method development for landslide susceptibility maps for Lower Austria funded by the Provincial Government of Lower Austria. The data on the topography, soil and tectonic lines was provided by courtesy of the Provincial Government of Lower Austria and the geological data by the Geological Survey of Austria (GBA-2009–Zl.383/1-09). The authors are thankful for the support and provision of additional data (landslide archive, simplified geological map and land cover map) of the Geological Survey of Lower Austria and the Department of Spatial Planning and Regional Policy and of the project partners Austrian Institute of Technology and Joanneum Research. We are grateful for the comments of two anonymous referees who helped to substantially improve the quality of this manuscript.

Edited by: P. Reichenbach

## References

- Abteilung Feuerwehr und Zivilschutz, Amt der NÖ Landesregierung: Zusammenfassung Ereignisse 2009, available at: [http://www.no.e.gv.at/Land-Zukunft/Katastrophenschutz/Archiv/zusammenfassung\\_ereignisse\\_2009.wai.html](http://www.no.e.gv.at/Land-Zukunft/Katastrophenschutz/Archiv/zusammenfassung_ereignisse_2009.wai.html), (last access: 9 December 2010), 2010 (in German).
- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P.: Impact of mapping errors on the reliability of landslide hazard maps, *Nat. Hazards Earth Syst. Sci.*, 2, 3–14, doi:10.5194/nhess-2-3-2002, 2002.
- Atkinson, P., Jiskoot, H., Massari, R. and Murray, T.: Generalized linear modelling in geomorphology, *Earth Surf. Proc. Land.*, 23, 1185–1195, 1998.
- Bathurst, J. C., Bovolo, C. I., and Cisneros, F.: Modelling the effect of forest cover on shallow landslides at the river basin scale, *Ecol. Eng.*, 36, 317–327, doi:10.1016/j.ecoleng.2009.05.001, 2010.
- Beguería, S.: Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management, *Nat. Hazards*, 37, 315–329, 2006a.
- Beguería, S.: Changes in land cover and shallow landslide activity: A case study in the Spanish Pyrenees, *Geomorphology*, 74, 196–206, 2006b.
- Bell, R.: Lokale und regionale Gefahren- und Risikoanalyse gravitativer Massenbewegungen an der Schwäbischen Alb, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn. available at: <http://hss.ulb.uni-bonn.de/2007/1107/1107.pdf>, (last access: 20 March 2012), 2007.
- 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, *Geogr. Ann. A*, 94, 135–156, 2012.
- Bell, R., Petschko, H., Proske, H., Leopold, P., Heiss, G., Bauer, C., Goetz, J. N., Granica, K., and Glade, T.: Methodenentwicklung zur Gefährdungsmodellierung von Massenbewegungen in Niederösterreich MoNOE – Vorläufiger Endbericht, Projektbericht, Universität Wien, Wien., 2013.
- Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, *Hydrolog. Sci. Bullet.*, 24, 43–69, 1979.
- Blahút, J., van Westen, C. J., and Sterlacchini, S.: Analysis of landslide inventories for accurate prediction of debris-flow source areas, *Geomorphology*, 119, 36–51, doi:10.1016/j.geomorph.2010.02.017, 2010.
- BMLFUW: Nachhaltig geschützt – Naturgefahrenmanagement im Unwetterjahr 2009, Jahresbericht, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (Lebensministerium), Wien, available at: <http://www.forstnet.at/article/articleview/82437/1/4932>, (last access: 3 December 2010), 2010.
- Boehner, J., Koethe, R., Conrad, O., Gross, J., Ringeler, A., and Selige, T.: Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation, Research Report, European Soil Bureau, Luxembourg, 2002.
- Brabb, E. E.: Innovative approaches to landslide hazard mapping, 4th International Symposium on Landslides, 16–21 September, Toronto, Canada, 307–324, 1984.
- Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, *Nat. Hazards Earth Syst. Sci.*, 5, 853–862, doi:10.5194/nhess-5-853-2005, 2005.
- Brenning, A.: Statistical Geocomputing combining R and SAGA: The Example of Landslide susceptibility Analysis with generalized additive Models, *SAGA – Seconds Out*, 19, 23–32, 2008.
- Brenning, A.: Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection, *Remote Sens. Environ.*, 113, 239–247, 2009.
- Brenning, A.: RSAGA: SAGA Geoprocessing and Terrain Analysis in R, available at: [http://CRAN.R-project.org/package=\\$RSAGA](http://CRAN.R-project.org/package=$RSAGA), (last access: 22 January 2013), 2011.
- 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, A. K., and Leroueil, S., Taylor & Francis, Banff, Alberta, Canada., 789–795, 2012a.
- Brenning, A.: Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package “sperrorest”, *IEEE Int. Geosci. Remote Se.*, 23, 5372–5375, 2012b.
- Brierley, G.: Communicating Geomorphology, *J. of Geography in Higher Educ.*, 33, 3–17, 2009.
- Brugnach, M., Tagg, A., Keil, F., and Lange, W. J.: Uncertainty Matters: Computer Models at the Science-Policy Interface, *Water Resources Management*, 21, 1075–1090, doi:10.1007/s11269-006-9099-y, 2006.
- Carrara, A.: Uncertainty in evaluating landslide hazard and risk, in *Prediction and Perception of Natural Hazards*, 101–109, Kluwer Academic Publishers, the Netherlands, 1993.
- Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: GIS technology in mapping landslide hazard, in *Geographical Information Systems in Assessing Natural Hazards*, edited by: Carrara, A. and Guzzetti, F., Kluwer Academic Publishers, the Netherlands, 135–175, 1995.
- Carrara, A., Guzzetti, F., Cardinali, M., and Reichenbach, P.: Use of GIS technology in the prediction and monitoring of landslide hazard, *Nat. Hazards*, 20, 117–135, 1999.

- Chung, C. J. F. and Fabbri, A. G.: Probabilistic prediction models for landslide hazard mapping, *Photogramm. Eng. Rem. S.*, 65, 1389–1399, 1999.
- Chung, C. J. F. and Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping, *Nat. Hazards*, 30, 451–472, 2003.
- Chung, C. J. and Fabbri, A. G.: Predicting landslides for risk analysis—Spatial models tested by a cross-validation technique, *Geomorphology*, 94, 438–452, 2008.
- Conrad, O.: SAGA – Entwurf, Funktionsumfang und Anwendung eines Systems für Automatisierte Geowissenschaftliche Analysen, PhD thesis, University of Göttingen, Germany, available at: <http://hdl.handle.net/11858/00-1735-0000-0006-B26C-6>, (last access: 27 February 2013), 2007 (in German).
- Crozier, M. J.: Landslides: causes, consequences and environment, Croom Helm, London, Sydney, Dover, New Hampshire, 1986.
- Crozier, M. J.: Deciphering the effect of climate change on landslide activity: A review, *Geomorphology*, 124, 260–267, 2010.
- Cruden, D. M. and Varnes, D. J.: Landslide types and processes, *Landslides, Investigation and Mitigation*, edited by: Turner, A. K. and Schuster, R. L., Transportation Research Board Special Report 247, Washington DC, USA, 36–75, 1996.
- Dai, F. C., Lee, C. F., and Ngai, Y. Y.: Landslide risk assessment and management: an overview, *Eng. Geol.*, 64, 65–87, 2002.
- Damm, A., Eberhard, K., and Patt, A.: Risikowahrnehmung von Erdbeben: Ergebnisse einer empirischen Untersuchung in der Südoststeiermark, Wegener Zentrum für Klima und Globalen Wandel, International Institute for Applied Systems Analysis (IIASA), Graz, Wien, 2010.
- 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, 391–404, 2007.
- Dikau, R., Brunsten, D., Schrott, L., and Ibsen, M. L.: Landslide recognition: identification, movement, and causes, Wiley, Chichester, 1996.
- Draper, D.: Assessment and Propagation of Model Uncertainty, *J. R. Stat. Soc. B*, 57, 45–97, 1995.
- Eder, A., Sotier, B., Klebner, K., Sturmlechner, R., Dorner, J., Markat, G., Schmid, G. and Strauss, P.: Hydrologische Bodenkenndaten der Böden Niederösterreichs (HydroBodNÖ) (Data on hydrological soil characteristics of soils in Lower Austria), unpublished final Report, Bundesamt für Wasserwirtschaft, Institut für Kulturtechnik und Bodenwasserhaushalt; Bundesforschungszentrum für Wald, Institut für Naturgefahren, Petzenkirchen, Innsbruck, 2011 (in German).
- Egner, H. and Pott, A.: Risiko und Raum: Das Angebot einer Beobachtungstheorie, in *Geographische Risikoforschung – Zur Konstruktion verträglichkeit Risiken und Sicherheiten*, edited by: Egner, H. and Pott, A., Franz Steiner Verlag, Stuttgart, 9–35, 2010 (in German).
- Elith, J., Burgman, M. A., and Regan, H. M.: Mapping epistemic uncertainties and vague concepts in predictions of species distribution, *Ecol. Model.*, 157, 313–329, 2002.
- Fabbri, A. G., Chung, C. J. F., Cendrero, A., and Remondo, J.: Is prediction of future landslides possible with a GIS?, *Nat. Hazards*, 30, 487–503, 2003.
- 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, *Eng. Geol.*, 102, 85–98, doi:10.1016/j.enggeo.2008.03.022, 2008.
- Fratini, P., Crosta, G., and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, *Eng. Geol.*, 111, 62–72, 2010.
- Freeman, G. T.: Calculating catchment area with divergent flow based on a regular grid, *Comput. Geosci.*, 17, 413–422, 1991.
- Frost, C. and Thompson, S.: Correcting for regression dilution bias: comparison of methods for a single predictor variable, *J. R. Stat. Soc. Ser. A*, 163, 173–190, 2000.
- Glade, T.: Landslide occurrence as a response to land use change: a review of evidence from New Zealand, *Catena*, 51, 297–314, 2003.
- Glade, T. and Crozier, M. J.: A review of Scale Dependency in Landslide Hazard and Risk Analysis, in *Landslide Hazard and Risk*, edited by: Glade, T., Anderson, M., and Crozier, M. J., John Wiley and Sons, Ltd, Chichester, England, 75–138, 2005.
- Glade, T., Anderson, M. G. and Crozier, M. J.: *Landslide Hazard and Risk*, John Wiley and Sons, Ltd, Chichester, England, 2005.
- Glade, T., Petschko, H., Bell, R., Bauer, C., Granica, K., Heiss, G., Leopold, P., Pomaroli, G., Proske, H., and Schweigl, J.: *Landslide susceptibility maps for Lower Austria – Methods and Challenges*, vol. 1, edited by: Koboltschnig, G., Hübl, J., and Braun, J., International Research Society INTERPRAEVENT, Grenoble, France, 497–508, 2012.
- Goetz, J. N., Guthrie, R. H., and Brenning, A.: Integrating physical and empirical landslide susceptibility models using generalized additive models, *Geomorphology*, 129, 376–386, 2011.
- Gottschling, P.: *Massenbewegungen*, in *Geologie der Bundesländer – Niederösterreich*, Geologische Bundesanstalt, Wien, 335–340, 2006 (in German).
- Guisan, A., Edwards, T. C., and Hastie, T.: Generalized linear and generalized additive models in studies of species distributions: setting the scene, *Ecol. Model.*, 157, 89–100, 2002.
- Guns, M. and Vanacker, V.: Logistic regression applied to natural hazards: rare event logistic regression with replications, *Nat. Hazards Earth Syst. Sci.*, 12, 1937–1947, doi:10.5194/nhess-12-1937-2012, 2012.
- Guzzetti, F.: *Landslide hazard and risk assessment*, Dissertation, Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn, November, 2005.
- 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, 181–216, 1999.
- Guzzetti, F., Cardinali, M., Reichenbach, P., and Carrara, A.: Comparing Landslide Maps: A Case Study in the Upper Tiber River Basin, Central Italy, *Environ. Manage.*, 25, 247–263, 2000.
- Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, *Geomorphology*, 81, 166–184, 2006.
- Hand, D.: *Construction and Assessment of Classification Rules*, Wiley, West Sussex, 1997.
- Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, 29–36, 1982.
- Harp, E. L., Castañeda, M. R., and Held, M. D.: Landslides triggered by Hurricane Mitch in Tegucigalpa, Honduras., US Geological Survey Open-File Report, USGS, 2002.

- Hastie, T.: gam: Generalized Additive Models, available at: [http://CRAN.R-project.org/package=\\$gam](http://CRAN.R-project.org/package=$gam), (last access: 27 February 2013), 2011.
- Hastie, T. and Tibshirani, R.: Generalized additive models, Chapman and Hall/CRC, London, 1990.
- Heckmann, T., Gegg, K., Gegg, A., and Becht, M.: Sample size matters: investigating the effect of sample size on a logistic regression debris flow susceptibility model, *Nat. Hazards Earth Syst. Sci. Discuss.*, 1, 2731–2779, doi:10.5194/nhessd-1-2731-2013, 2013.
- Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Sallaberry, C. J.: Representation of analysis results involving aleatory and epistemic uncertainty, *Int. J. Gen. Syst.*, 39, 605–646, 2010.
- Hervás, J.: Lessons Learnt from Landslide Disasters in Europe, European Commission Joint Research Centre, Institute for Environment and Sustainability, Ispra (VA), Italy, 2003.
- Hill, L. J., Sparks, R. S. J. and Rougier, J. C.: Risk assessment and uncertainty in natural hazards, in *Risk and uncertainty assessment for natural hazards*, edited by: Rougier, J. C., Sparks, R. S. J., and Hill, L. J., 1–18, Cambridge University Press, Cambridge, 2013.
- Hjort, J. and Marmion, M.: Effects of sample size on the accuracy of geomorphological models, *Geomorphology*, 102, 341–350, 2008.
- Hoffman, F. O. and Hammonds, J. S.: Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability, *Risk Analysis*, 14, 707–712, 1994.
- Hora, S. C.: Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management, *Reliability Engineering and System Safety*, 54, 217–223, 1996.
- Hornich, R. and Adelwöhrer, R.: Landslides in Styria in 2009/Hangrutschungsereignisse 2009 in der Steiermark, *Geomechanics and Tunnelling*, 3, 455–461, 2010 (in German).
- Hosmer, D. W. and Lemeshow, S.: Applied logistic regression, Wiley, New York, NY, 2000.
- Huggel, C., Clague, J. J., and Korup, O.: Is climate change responsible for changing landslide activity in high mountains?, *Earth Surf. Process. Landforms*, 37, 77–91, 2012.
- Hydrographischer Dienst des Landes Niederösterreich: Wasserstands-nachrichten und Hochwasserprognosen Niederösterreich, available at: [http://www.noel.gv.at/Externeseiten/wasserstand/wiskiwebpublic/maps\\_N\\_0.htm?entryparakey=\\$N](http://www.noel.gv.at/Externeseiten/wasserstand/wiskiwebpublic/maps_N_0.htm?entryparakey=$N), (last access: 2 March 2011), 2011 (in German).
- Jia, G., Tian, Y., Liu, Y., and Zhang, Y.: A static and dynamic factors-coupled forecasting model of regional rainfall-induced landslides: A case study of Shenzhen, *Sci. China Ser. E*, 51, 164–175, 2008.
- Karam, K. S.: Landslide hazards assessment and uncertainties, PhD Thesis, Massachusetts Institute of Technology, 2005.
- Klimeš, J. and Blahůt, J.: Landslide risk analysis and its application in regional planning: an example from the highlands of the Outer Western Carpathians, Czech Republic, *Nat. Hazards*, 64, 1779–1803, 2012.
- Knuepfer, P. L. and Petersen, J. F.: Geomorphology in the public eye: policy issues, education, and the public, *Geomorphology*, 47, 95–105, 2002.
- Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, 14, 1137–1145, 1995.
- Kunz, M., Grêt-Regamey, A., and Hurni, L.: Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system, *Nat. Hazards*, 59, 1735–1751, 2011.
- Kurz, W.: Erstellung einer digitalen, strukturgeologischen, tektonischen Karte von Niederösterreich (Endbericht), unpublished final report, Karl-Franzens-University Graz, Graz, 2012 (in German).
- Lee, C.-T., Huang, C.-C., Lee, J.-F., Pan, K.-L., Lin, M.-L., and Dong, J.-J.: Statistical approach to earthquake-induced landslide susceptibility, *Eng. Geol.*, 100, 43–58, 2008.
- Lettner, C. and Wrška, T.: Historical Development of the Cultural Landscape at the Northern Border of the Eastern Alps: General Trends and Regional Peculiarities, edited by: Balázs, P. and Konkoly-Gyuró, E., 109–121, University of West Hungary Press, Sopron, Hungary, 2011.
- Luoto, M., Marmion, M., and Hjort, J.: Assessing spatial uncertainty in predictive geomorphological mapping: A multi-modelling approach, *Comput. & Geosci.*, 36, 355–361, 2010.
- Malamud, B. D., Turcotte, D. L., Guzzetti, F., and Reichenbach, P.: Landslide inventories and their statistical properties, *Earth Surf. Proc. Land.*, 29, 687–711, 2004.
- Marjanović, M., Kovačević, M., Bajat, B. and Voženilek, V.: Landslide susceptibility assessment using SVM machine learning algorithm, *Eng. Geol.*, 123, 225–234, 2011.
- McCalpin, J.: Preliminary age classification of landslides for inventory mapping, In: *Proceedings 21st annual Engineering Geology and Soils Engineering Symposium*, 5–6 April, University of Idaho, Moscow, 99–111, 1984.
- Mosleh, A.: Hidden sources of uncertainty: judgment in the collection and analysis of data, *Nucl. Eng. Des.*, 93, 187–198, 1986.
- Muenchow, J., Brenning, A., and Richter M.: Geomorphic process rates of landslides along a humidity gradient in the tropical Andes, *Geomorphology*, 139, 271–284, 2012.
- Nefeslioglu, H. A., Gokceoglu, C., and Sonmez, H.: An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps, *Eng. Geol.*, 97, 171–191, 2008.
- Neuhäuser, B. and Terhorst, B.: Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany), *Geomorphology*, 86, 12–24, 2007.
- Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S.: Challenge problems: uncertainty in system response given uncertain parameters, *Reliability Engineering & System Safety*, 85, 11–19, doi:10.1016/j.res.2004.03.002, 2004.
- Orme, A. R.: Shifting paradigms in geomorphology: the fate of research ideas in an educational context, *Geomorphology*, 47, 325–342, 2002.
- Park, N.-W. and Chi, K.-H.: Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model, *Int. J. Remote Sens.*, 29, 247–264, 2008.
- Petschko, H., Glade, T., Bell, R., Schweigl, J. and Pomaroli, G.: Landslide inventories for regional early warning systems, in: *Proceedings of the International Conference Mountain Risks: Bringing Science to Society*, Firenze, 24–26 November 2010, edited by: Malet, J. P., Glade, T., and Casagli, N., CERG Editions, Strasbourg, 277–282, 2010.
- Petschko, H., Bell, R., Brenning, A., and Glade, T.: Landslide susceptibility modeling with generalized additive models – fac-

- ing the heterogeneity of large regions, in: *Landslides and Engineered Slopes, Protecting Society through Improved Understanding*, Vol. 1, edited by: Eberhardt, E., Froese, C., Turner, A. K., and Leroueil, S., Taylor & Francis, Banff, Alberta, Canada, 769–777, 2012.
- Petschko, H., Bell, R., Glade, T., Leopold, P., and Heiss, G.: Landslide inventories for large regions – mapping effectiveness of visually analyzing LiDAR derivatives, in prep., 2013a.
- 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, *Proceedings of the World Landslide Forum 3*, 2–6 June 2014, Beijing, submitted, 2013b.
- Petschko, H., Bell, R., Leopold, P., Heiss, G., and Glade, T.: Landslide inventories for reliable susceptibility maps in Lower Austria, in: *Landslide Science and Practice. Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning*, edited by: Margottini C., Canuti P., Sassa K., Springer, 281–286, 2013c.
- Pradhan, B.: A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Computers & Geosciences*, 51, 350–365, 2013.
- Pradhan, B. and Lee, S.: Landslide risk analysis using artificial neural network model focusing on different training sites, *Int. J. Phy. Sci.*, 4, 1–15, 2009.
- Quinn, P. F., Beven, K. J., Chevallier, P., and Planchon, O.: The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrol. Process.*, 5, 59–79, 1991.
- R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. available at: <http://www.R-project.org>, (last access: 27 February 2013), 2011.
- Razak, K. A., Straatsma, M. W., Van Westen, C. J., Malet, J.-P., and De Jong, S. M.: Airborne laser scanning of forested landslides characterization: Terrain model quality and visualization, *Geomorphology*, 126, 186–200, 2011.
- Remondo, J., González, A., De Terán, J. R. D., Cendrero, A., Fabri, A., and Chung, C. J. F.: Validation of landslide susceptibility maps; examples and applications from a case study in Northern Spain, *Nat. Hazards*, 30, 437–449, 2003.
- Rogers, K. H.: The real river management challenge: integrating scientists, stakeholders and service agencies, *River Res. Appl.*, 22, 269–280, 2006.
- Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, *Geomorphology*, 114, 129–142, 2010.
- Rougier, J. C.: Quantifying hazard losses, in *Risk and uncertainty assessment for natural hazards*, edited by: Rougier, J. C., Sparks, R. S. J., and Hill, L. J., 19–39, Cambridge University Press, Cambridge., 2013.
- Rougier, J. C. and Beven, K. J.: Model and data limitations: the sources and implications of epistemic uncertainty, in *Risk and uncertainty assessment for natural hazards*, edited by: Rougier, J. C., Sparks, R. S. J., and Hill, L. J., 19–39, Cambridge University Press, Cambridge., 2013.
- Roy, C. J. and Oberkamp, W. L.: A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, *Comput. Methods Appl. M.*, 200, 2131–2144, 2011.
- Ruß, G. and Brenning, A.: Data mining in precision agriculture: management of spatial information, *Lect. Notes in Comput. Sci.*, 6178, 350–359, 2010.
- Schicker, R. and Moon, V.: Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale, *Geomorphology*, 161–162, 40–57, 2012.
- Schnabel, W.: Die geologischen Ursachen der Schäden an der II. Wiener Hochquellenleitung bei Scheibbs, in: *Festschrift der MA 31: 75 Jahre II. Wiener Hochquellenleitung*, edited by: Magistrat der Stadt Wien, MA 31-Wasserwerke, Wien, 1985 (in German).
- Schnabel, W.: *Geologische Karte von Niederösterreich 1 : 200 000*, Geologische Bundesanstalt, Wien, Austria, 2002 (in German).
- Schwab, J. C., Gori, P. L., and Jeer, S.: *Landslide Hazards and Planning*, Planning Advisory Service Report, American Planning Association, Chicago, 2005.
- Schwarz, L. and Tilch, N.: Möglichkeiten und Limitierungen der Regionalisierung mittels Neuronaler Netze am Beispiel einer Rutschungsanfälligkeitskarte für die Region Gasen-Haslau, *Beiträge zum 20. AGIT-Symposium, Angewandte Geoinformatik*, 2–4 July 2008, Salzburg, Austria, 643–648, 2008 (in German).
- 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 at: [http://eusoils.jrc.ec.europa.eu/ESDB\\_Archive/eusoils\\_docs/other/EUR23785EN.pdf](http://eusoils.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/other/EUR23785EN.pdf), (last access: 1 March 2011), 2009.
- Schwenk, H.: Massenbewegungen in Niederösterreich 1953–1990, in: *Jahrbuch der Geologischen Bundesanstalt, Geologische Bundesanstalt*, Wien, 135, 597–660, 1992 (in German).
- Scott, A. J. and Wild, C. J.: Fitting logistic models under case-control or choice based sampling, *J. Roy. Stat. Soc. B*, 48, 170–182, 1986.
- Seibert, J., Stendahl, J., and Sørensen, R.: Topographical influences on soil properties in boreal forests, *Geoderma*, 141, 139–148, 2007.
- Soeters, R. and Van Westen, C. J.: Slope instability recognition, analysis and zonation., in *Landslides, Investigation and Mitigation*, edited by: Turner, A. K. and Schuster, R. L., 129–177, National Academy Press, Washington, USA, 1996.
- Spiegelhalter, D. J. and Riesch, H.: Don't know, can't know: embracing deeper uncertainties when analysing risks, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369, 4730–4750, 2011.
- Sterlacchini, S., Ballabio, C., Blahút, J., Masetti, M., and Sorichetta, A.: Spatial agreement of predicted patterns in landslide susceptibility maps, *Geomorphology*, 125, 51–61, 2011.
- Stockwell, D. R. and Peterson, A. T.: Effects of sample size on accuracy of species distribution models, *Ecol. Model.*, 148, 1–13, 2002.
- Tarolli, P., Borga, M., Chang, K.-T., and Chiang, S.-H.: Modeling shallow landsliding susceptibility by incorporating heavy rainfall statistical properties, *Geomorphology*, 133, 199–211, 2011.
- Townsend Peterson, A., Papeş, M., and Eaton, M.: Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent, *Ecography*, 30, 550–560, 2007.

- Trigila, A., Frattini, P., Casagli, N., Catani, F., Crosta, G., Esposito, C., Iadanza, C., Lagomarsino, D., Mugnozza, G., Segoni, S., Spizzichino, D., Tofani, V., and Lari, S.: Landslide Susceptibility Mapping at National Scale: The Italian Case Study, in *Landslide Science and Practice*, edited by: Margottini, C., Canuti, P., and Sassa, K., 287–295, Springer Berlin Heidelberg, 2013.
- 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, 392–410, 2006.
- Van den Eeckhaut, M., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Van Beek, L. P. H., and Vandekerckhove, L.: Use of LIDAR-derived images for mapping old landslides under forest, *Earth Surf. Proc. Land.*, 32, 754–769, 2007.
- Van den Eeckhaut, M., Moeyersons, J., Nyssen, J., Abraha, A., Poesen, J., Haile, M., and Deckers, J.: Spatial patterns of old, deep-seated landslides: A case-study in the northern Ethiopian highlands, *Geomorphology*, 105, 239–252, 2009.
- Van Westen, C. J., Rengers, N., Terlien, M. T. J., and Soeters, R.: Prediction of the occurrence of slope instability phenomena through GIS-based hazard zonation, *Geologische Rundschau*, 86, 404–414, 1997.
- Van Westen, C. J., Asch, T. W. J., and Soeters, R.: Landslide hazard and risk zonation—why is it still so difficult?, *B. Eng. Geol. Environ.*, 65, 167–184, 2005.
- Van Westen, C. J., Castellanos, E., and Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, *Eng. Geol.*, 102, 112–131, 2008.
- Varnes, D. J.: *Landslide hazard zonation: a review of principles and practice*, United Nations Educational, Scientific and Cultural Organization, Paris, France, 1984.
- Von Ruette, J., Papritz, A., Lehmann, P., Rickli, C., and Or, D.: Spatial statistical modeling of shallow landslides – Validating predictions for different landslide inventories and rainfall events, *Geomorphology*, 133, 11–22, 2011.
- Vorpahl, P., Elsenbeer, H., Märker, M., and Schröder, B.: How can statistical models help to determine driving factors of landslides?, *Ecol. Model.*, 239, 27–39, 2012.
- Wessely, G.: *Geologie der österreichischen Bundesländer – Niederösterreich*, Geologische Bundesanstalt, Wien, 2006 (in German).
- Zevenbergen, L. W. and Thornes, J. B.: Quantitative analysis of land surface topography, *Earth Surf. Proc. Land.*, 12, 47–56, 1987.