Article

Topographic control on soil function evaluation - a case study from South Tyrol

Fabian Ernst Gruber [1],*,†,‡ (ID) **, Elisabeth Schaber** [1],‡**, Jasmin Baruck** [1],‡ **and Clemens Geitner** [1],

[1]   University of Innsbruck, Institute of Geography; geographie@uibk.ac.at
*   Correspondence: fabian.gruber@ages.at
†   Current address: Austrian Agency for Health and Food Security
‡   These authors contributed equally to this work.

1

---

2   **1. Introduction**

3   Information on soil, a, at least from a human time perspective, non-renewable ressource, is
4   of increasing importance given erosion, soil degradation and soil sealing. It is necessary to know
5   where and where not certain practises are applicable and to adjust land-use planning appropriately.
6   Accordingly, soil function evaluation is an invaluable tool for the future.

7

8   [1] and further literature

9

10   In this study, we present the soil evaluation tool *Soil Evaluation for Planning Procedures*
11   *(SEPP)* and investigate topographic and parent material control of the different soil functions by
12   applying a cross-validated machine learning approach based on availible soil pit information in the
13   Oltradige/Überetsch region of the Autonomous Province Bolzano - South Tyrol.
14   The main aim of the study is

15   **2. Materials and Methods**

16   *2.1. Study area and soil data*

17   *2.2. SEPP - Soil Evaluation for Planning Procedures*

18   The software SEPP currently computes a soil function evaluation based on soil pit descriptions.
19   It requires that the pit descriptions are performed following the Austrian Soil classification [2,3] and
20   related mapping manuals. The minimum soil profile site characteristics are local slope, thickness
21   of organic horizons, soil depth, groundwater table, soil parent material, soil type, humus form,
22   altitudinal zone, moisture level and land use.For each horizon, the minimum characteristics necessary
23   for computing the soil function are the master horizon designation, depth, pH value, proportion
24   of the dominant soil structure type and class membership with regard to carbonate content, soil
25   texture, organic content, abundance of rock fragments, bulk density, soil structure. These class
26   attributes can be substituted by exact values if available. The soil functions, for which 15 different
27   potentials are computed, are *habitat for living organisms* (specifically the potential as habitat for
28   drought-tolerant species, moisture tolerant species, soil organisms and crops), *infiltration and drainage*
29   *regulation* (minimum, average and heavy precipitation retention capacity as well as groundwater
30   reformation rate), *natural soil fertility* as well as *filter and buffer for pollutants* (heavy metal, organic,

acidifying and water-soluble). The result is a grade between 1 and 5 for each soil function potential, with 1 signifying a high potential and 5 a low one.

Potential as a habitat for drought-tolerant species

Both this potential and the following potential as a habitat for moisture-tolerant species are performed based on modifications of the approaches decribed by [4] and [5]. The evaluation of a soil's potential as a habitat for drought-tolerant species is based on the parameters land use, soil type and available field capacity. While the first two parameters are applied to distinguish especially suited (ruderal locations and corresponding soil types) or unsuited (mire deposits and soil types commonly found on these) sites, the latter is used to grade those soil profile sites showing the remaining landuse and soil type combinations.

Potential as a habitat for moisture-tolerant species

This potential is evaluated similarly to the the one for drought-tolerant species, in that specific soil types, e.g. Gleysols, are attributed specific grades. In addition, the depth of the groundwater table is used to distinguish sites with high potential, and the available field capacity is used to differentiate even further.

Habitat for soil organisms

This potential is evaluated according to [6] with some minor adaptions. In this framework, a number of species groups are used as indicators for the composition of soil life, with emphasis on earth worms (Lumbricidae) as they are influential on soil structure and bioturbation. This method is based on the relationship between soil organism communities and a number of abiotic soil parameters. Specifically, one of 14 possible soil organism communities, which are the basis for the grade awarded to a site, is attributed to a site according to a classification tree applying the parameters pH, moisture level, land use and soil texture.

Potential for agricultural production

The assessment of the potential for agricultural production is performed according to the method proposed in the framework TUSEC-IP [5] by an accumulative rating of five criteria. The criteria *general conditions of the profile site* is rated based on soil depth, topsoil aggregate structure and topsoil as well as subsoil bulk density. While the criteria *water supply* is based on available field capacity and the depth of the groundwater table, the grade for *air supply* is derived from air capacity and for *nutrient supply* the alkaline cation exchange capacity is regarded. The *climate* criteria is derived from the mean annual temperature of the growing season if available, or else replaced by proxy values such as mean annual temperature or altitudinal zone. The combination of the grades of the individual criteria for agricultural production leads to an overall grade that is then adjusted for the slope gradient of the location.

Average and minimum precipitation retention capacity

Following a modified version of the procedures presented by [7] and [4], this potential is assessed by combining the permeability coefficient (using either the average value of the soil profile or the minimum value) with the water storage capacity. For more or less planar areas, the water storage capacity is regarded as the sum of the usable field capacity and the air capacity, whereas for steeper slopes only the former parameter is used. Additionally, permeability coefficient and water storage capacity is considered only for soil horizons not linked to contact with groundwater or stagnant water.

Retention capacity for heavy precipitation events

The evaluation of this soil potential is a modfied version of the scheme proposed by [5]. It differs from the agerage precipitation retention capacity by being based on the assumption that flooding hazards are greatest when soils are already saturated with water, and therefore only the air storage capacity is considered for retention. This retention volume is then compared to the design rainfall event under consideration of the infiltration rate.

Quality of groundwater reformation

Also following [5], this potential is assessed using the same parameters as the precipitation capacity but considers the assumption that very quick infiltration leads to an increase of pollutants in the groundwater. In the same reasoning, soil types linked to groundwater or locations with high groundwater table are given poorer grades.

Potential for providing nutrients for plants

Adhering to [8], the assessment of this potential uses the parameter alkaline cation exchange capacity. As this is only a coarse approximation, this potential is not differentiated into five, but only three classes (poor, average and high potential).

Potential as a CO2 sink

By applying a modified version of the rating proposed by [9], selected land uses, especially forests, are awarded good grades, whereas other landuses are graded based on the amount of organic matter, summed up over all soil horizons.

Potential for retention of heavy metals

In this assessment, the ability to bind cadmium is used as a proxy for other heavy metals. Based on modifications of the procedures proposed by [10] and [4],in a first step this ability is evaluated for different pH-values for sandy soils with little organic content, and later adjusted with regard to organic matter content and soil texture as a proxy for clay content.

Potential for transforming organic contaminants

As organic pollutants are generally transformed by soil organisms, this potential can be essentially assessed by rating the living conditions for soil micro-organisms. Consequently, the parameters which contribute to the rating procedure based on [7] are topsoil organic matter content, topsoil clay content and the average topsoil pH-value. In a first step, microbial activity is estimated based on humus form and pH-value, and then the potential for transformation is further differentiated based on organic matter and clay content.

Potential as filter and buffer for organic contaminants

As the evaluation of a representative contaminant (such as cadmium for heavy metals) is not feasible for organic contaminants due to their variety, this potential is assessed by estimating a mean binding capacity for organic pollutants using organic matter and clay content for the fine material contained within a soil profile.

Potential for retention of water-soluble contaminants

For the assessment of the potential for retention of water-soluble pollutants, with emphasis on nitrate, the yearly seepage rate is calculated based on mean precipitation, mean evaporation and an estimate of surface run-off derived from soil texture. The grade for this potential is rewarded based on the annual exchange rate of soil water by comparing seepage volume with field capacity.

113 Potential as buffer for acidic contaminants

114 The potential buffer capacity is evaluated according to [4] by considering the alkaline cation
115 exchange capacity and the carbonate content of the mineral horizons on the one hand, and estimating
116 the buffer capacity of the organic layer from its humus form and thickness on the other hand.

117 *2.3. The predictor variable set:terrain parameters and landform classification*

118 Terrain parameters, i.e. variables derived from performing calculations on digital terrain models
119 (DTM) have long played an important role in modeling parameters or processes associated with
120 geomorphology, soil, natural hazards, or habitat modeling, to just name a few scientific fields.
121 An overview of terrain parameters, their calculation, and application can be found for instance
122 in [11,12]. While a wide variety of different terrain parameters are available, the presented study
123 concentrates on simple local terrain parameters as well as an automated landform classification
124 algorithm. For the former group of parameters, slope and a variety of curvature measures (planar,
125 profile, cross-sectional, and tangential) were computed at varying window sizes using the GRASS GIS
126 7 [13] module r.param.scale. With regard to landforms, [14] showed in their comparison of automated
127 landform classifications and the topographic description of soil pit sites,that the r.geomorphon
128 algorithm by [15] is a valuable tool for use in soil survey. It uses pattern recognition based on
129 line-of-sight calculations to classify each grid cell of a digital terrain model as one of 10 possible
130 landforms (Figure 1). These landform classifications were performed considering modifications of the
131 two parameters search window size and flatness threshold. While the former parameter is responsible
132 for how many surrounding grid cells are considered when performing the line-of-sight calculations,
133 thus influencing the scale at which the landforms are computed, the latter represents the slope angles
134 up to which a grid cell is regarded to be flat. [14] showed that investigating the result of varying
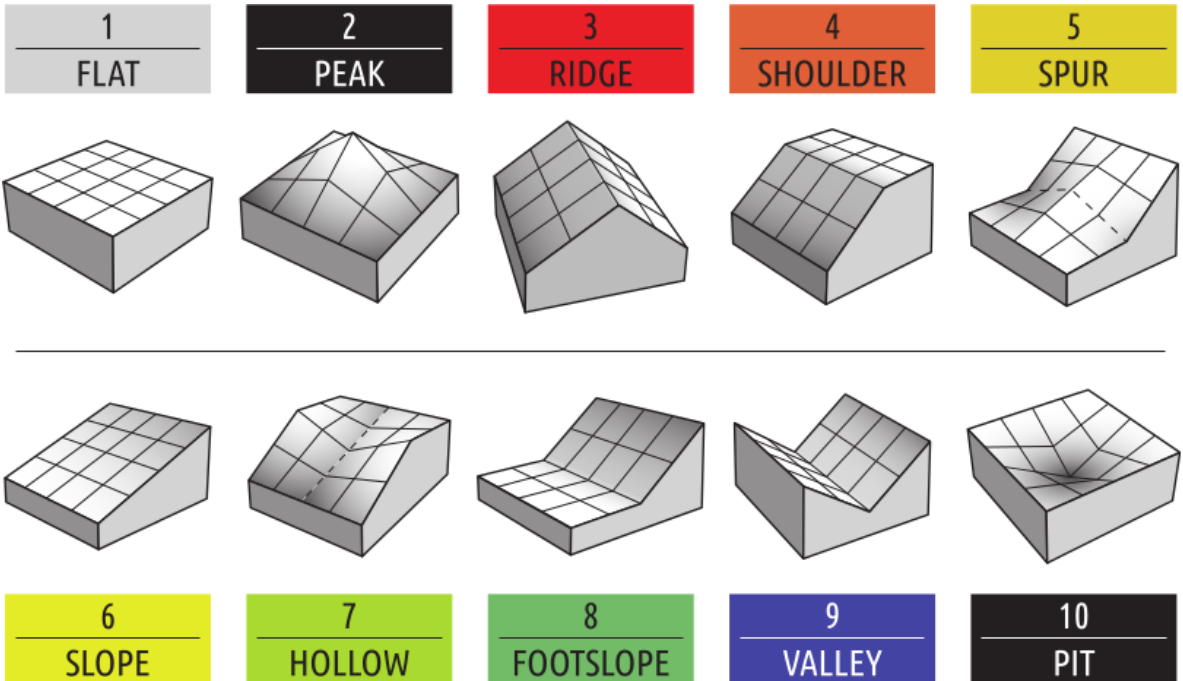flatness thresholds is of interest especially in an Alpine environment.

Figure 1. The r.geomorphon algorithm classifies every grid cell of a given digital terrain model as one
of these 10 landforms based on line-of-sight calculations (figure based on [15])

135

2.4. Statistical learning and variable selection procedure

Support vector machine (SVM) classification is a statistical learning approach first described by [16]. Its classification algorithm computes a hyper-plane that best separates different classes by concentrating on those data points of both classes (support vectors) which are within a certain margin of the hyperplane. In this process, SVM classification also allows some points to be on the *wrong* side of the hyperplane, a behavior that can be addressed with the tuning parameter c. By applying a radial kernel this linear binary classification can be extended to the non-linear case. A majority vote system then further allows application of SVM-classification for multiple classes. In the presented study, the SVM-algorithm as implemented in the package *e1071* [17] of the open source statistical computing environment R [18] was used.

A forward step-wise feature selection procedure based on 10-fold cross-validation using SVM classification was applied to select the best parameter or parameter combination for each soil function or potential. Cross-validation (5-fold) was also applied to the entire feature selection procedure, meaning that after partitioning the data set into five sets, one part is set aside for validation to be used after the feature selection procedure has been carried out with the four remaining parts used for the model fitting. The usefulness of additional parameters in the model were evaluated using the *one-standard-error-rule* [19], meaning that the number of explanatory variables was seen sufficient if the improvement through adding more explanatory variables did not exceed one standard error of the predictions. As the selection process at times revealed a number of parameter combinations leading to comparable cross-validated accuracies as well as showing a wide range of these values, the different parameter combinations were then compared by performing the 10-fold CV 100 times, each time with 10 different, random partitions. The median values and distribution of accuracy values were compared, as were the confusion matrices based on the whole data set, where the final modeled soil function classes were attributed using the majority vote of the 100 predictions. Additionally, the models resulting from different parameter combinations were used to predict the soil function potentials for the entire study area and visualized as maps to examine how they performed spatially.

2.5. Post-feature selection workflow

After the feature selection procedure, the classes, consisting of soil pits with the same grade for a given potential, were analyzed with regard to the chosen terrain parameters or landform classification. This was performed with the original classes, but also with the projected classes in order to investigate the effect of the different parameters on the output of the model based on the support vector machine classification. An example of such a box plot analysis of the distribution of terrain parameters is given in Figure 4. This analysis also gives insight into certain aspects explaining user's and producer's reliability as derived from evaluation of the confusion matrix based on the 100 prediction runs. User's accuracy gives the percentage of correctly classified members of a class with regard to the total number of members in the respective reference class, which in this study is the classification performed by the SEPP tool. Producer's reliability on the other hand relates the correctly classified members of a given class to the total number of members predicted by the model. If a landform classification was chosen as best suited to distinguish between soil profile site locations of different grades for a potential, bar plots instead of box plots were applied to investigate and visualize the model output. For each soil function grade, the distribution of the 10 landforms was plotted in order to conceive how the dominance of different landforms changed for the different grades, and how the SVM model attributed the projected grades to different landforms. Figure 5 shows such a plot for the soil function of transforming organic contaminants.

3. Results

The soil function grades for each of the 15 potentials was calculated for each of the 108 soil profile pits in the study area with the SEPP application. Figure 2 shows the distribution of these grades.

A first evaluation of the feature selection procedure shows that mostly 2 parameters are sufficient,
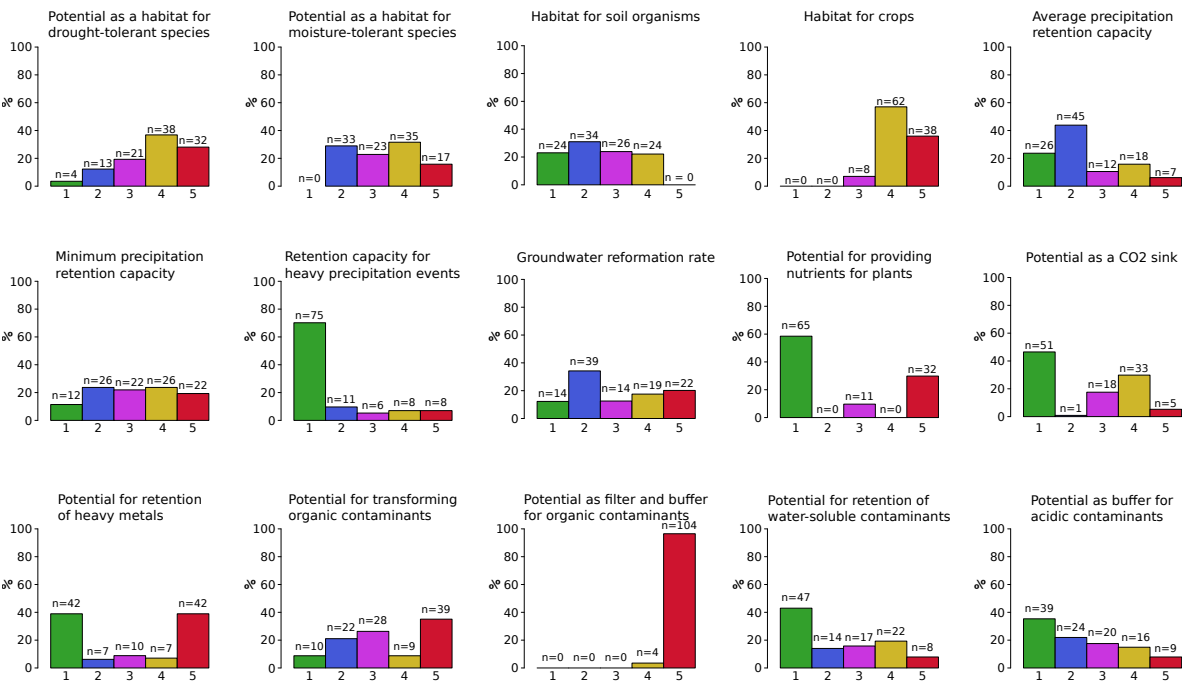

Figure 2. Barplots representing the distribution of the soil function grades for the various analysed potentials.

that is that there is no increase in cross-validated prediction accuracy by adding more predictors. In many cases, these parameter combinations consist of one to three local terrain parameters, sometimes complemented with a landform classification. parameter.

Table 1. Accuracy values (%) and parameters for the different potentials

potential	parameter	res	windowsize	multiacc	testacc	med2	test2
habitat for drought-tolerant species	landforms	10	100				
	slope	50	350	50.0	50.0	43.5	57.4
habitat for moisture-tolerant species	long. curvature	2.5	7.5				
	landforms	10	70	53.7	53.7	52.3	66.6
habitat for soil organisms	cross-sec. curvature	50	350				
	slope	50	350				
	convexity	50	150	59.3	62.0	53.7	80.6
agricultural production	slope	2.5	7.5	86.1	86.1	84.3	85.2
average precipitation retention capacity	cross-sec. curvature	2.5	47.5				
	profile curvature	10	150	50.9	54.6	45.4	62.0
minimum precipitation retention capacity	plan curvature	2.5	72.5				
	minimal curvature	50	150	41.6	46.3	29.6	66.6
retention capacity for heavy precipitation events	long. curvature	10	150	73.1	73.1	73.1	80.6
groundwater reformation rate	cross-sec. curvature	2.5	57.5				
	profile curvature	10	50	47.2	49.1	39.8	62.0
providing nutrients for plants	minimal curvature	2.5	22.5	71.3	72.2	71.3	87.0
CO2 sink	minimal curvature	2.5	27.5	61.1	61.1	66.7	73.1
retention of heavy metals	landforms	10	70	63.9	63.9	63.0	68.5
transforming organic contaminants	landforms	10	500				
	maximal curvature	10	70	46.3	50.0	44.4	52.8
filter and buffer for organic contaminants	-	-	-	-	-	-	-
retention of water-soluble contaminants	slope	2.5	27.5				
	minimal curvature	2.5	7.5	53.7	59.3	52.8	65.7
buffer for acidic contaminants	plan curvature	2.5	12.5				
	slope	2.5	12.5	48.1	50.0	38.9	65.7

3.1. Potential as a habitat for drought-tolerant species

Figure 2 shows that of the 108 soil profile sites in the study area, 38 fall into class 4 (35%) and 32 into class 5 regarding the potential as a habitat for drought-tolerant species. The intermediate

class 3 contains 21 soil profiles whereas the high potential classes 1 and 2 are attribute to only 4 and 13 sites, respectively. As the predictor set neither contains land use nor soil type, the SVM classification essentially attempts to model the different classes of available field capacity. In the majority of the feature selection runs a landform map based on a flatness threshold between 3 and 5°,a spatial resolution of 10 m and a search radius of 100 m was chosen as the first predictive feature. The landform flat is dominant amongst the profile sites with a graded potential of 5, which is accordingly connected to minimal curvature values around 0. The landform slope is most common for profiles with a potential of 4, whereas spurs and hollow can present profile locations with a potential score of 2 and, as expected, have increasingly negative minimum curvature values. A support vector classifier using these landforms and slope at a low DTM resolution as predictor variables, results in a median cross-validated prediction accuracy of 50%, where the most common error is that a large number of sites are mistakenly classified as having grade 4. Nevertheless, the general implications of the feature selection are plausible, as flat areas can be expected to have higher field capacity values than sloping regions with negative curvature values.

3.2. Potential as a habitat for moisture-tolerant species

None of the soil profile sites in the study area is awarded the best grade (1) for its potential as a habitat for moisture-tolerant species. As seen in Figure 2, the intermediate classes (two to four) are quite evenly distributed with 33, 23 and 35 members, respectively. The class with the poorest potential consists of 17 soil profile sites. Given the similarity in soil parameters and profile site characteristics used for the evaluation of this potential and the potential as a habitat for drought-tolerant species, a very similar landform classification is chosen in the feature selection procedure (Table 1), the only difference being a slightly tighter search window of 70 m. This feature is complemented by the local terrain parameter longitudinal curvature to achieve the best median cross-validated accuracy of 53.7% with a SVM-model of this potential. The model predictions show that while the SVM-classifer associates good grades with curvature values around zero, soil pits with the worst potential for moisture tolerant species can be found at locations with negative longitudinal curvature. This trend is also visible in the landform distribution, were the landform flat, and, to a lesser degree, foot slopes are characteristic for soil pits with higher potentials. This combinations seems reasonable, given the potential hydrological situation on these landforms.

3.3. Habitat for soil organisms

With the exception of the class with grade 5 (low potential as a habitat for soil organisms), which does not occur, the four other grades are distributed relatively evenly amongst the soil pits in the study area. Figure 3A) shows the spatial distribution of the profiles sites and their grades in the study area. The class with the good grade of 2 is the most common with 34 members. The feature selection procedure distinguished three local terrain parameters as being most useful in separating the soil profile sites with different grades with regard to their potential as a habitat for soil organisms, representing the parameters slope, convexity, and cross-sectional curvature. Lower convexity values characterize those soil profile sites best suited for soil organisms. Similarly, high slope values are helpful in separating members the intermediate class with grade 3 from the remaining 3 classes, based on the general trend that the two best grades are more closely associated with lower slope angles than the grades 3 and 4. An analysis of the cross-sectional curvature values shows that the the class with grade 4 tends to have more members related to slightly positive curvature values, compared to the other classes. A SVM-classifier trained with the three aforementioned terrain parameters leads to a median accuracy rate of 59.3 % which is comparably high compared to other potentials which also have their grades distributed over more than three classes. Figure 3B) shows how this prediction turns out spatially for the study area.
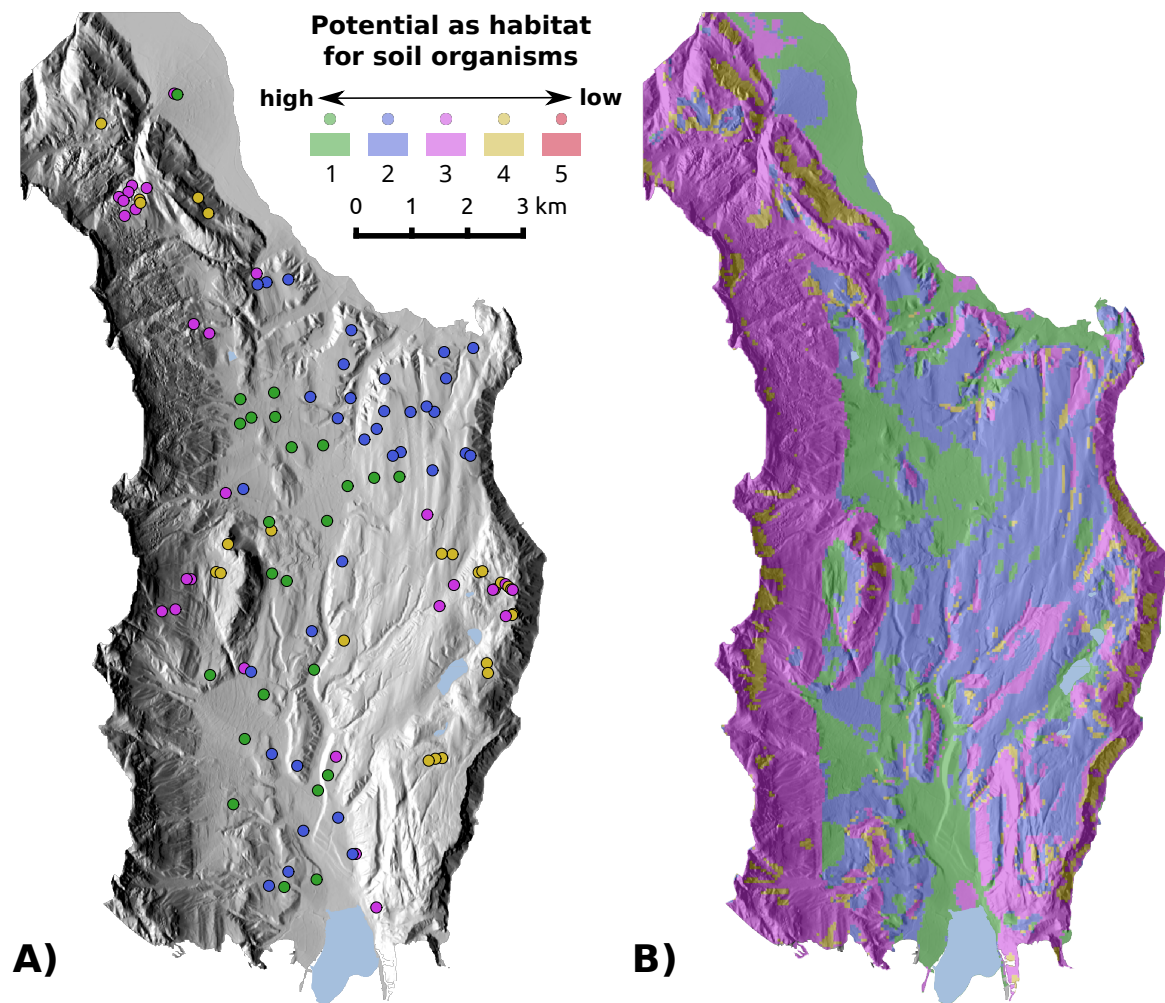
Figure 3. A) shows the study area along with the location of the soil profile sites and their grade regarding their potential as a habitat for soil organisms. B) shows the same potential as predicted by a SVM-classifier based on the terrain parameters slope, convexity, and cross-sectional curvature.

3.4. Potential for agricultural production

The soil pit locations within the study area exhibit a very low diversity with regards to the grades for representing a habitat for crops. Only eight of the locations receive the intermediate grade 3, whereas the remaining soil pit sites are rewarded the poorer grades 4 (n=62) and 5 (n=38). The feature selection procedure leads to a model which incorporates only the local terrain parameter slope based on the high resolution DTM and leads to a median cross-validated accuracy of 86 %. Grade 3 is not depicted in the model output, as the soil pits belonging to this class are misclassified as being part of the class with grade 4, and, to a lesser degree, grade 5.
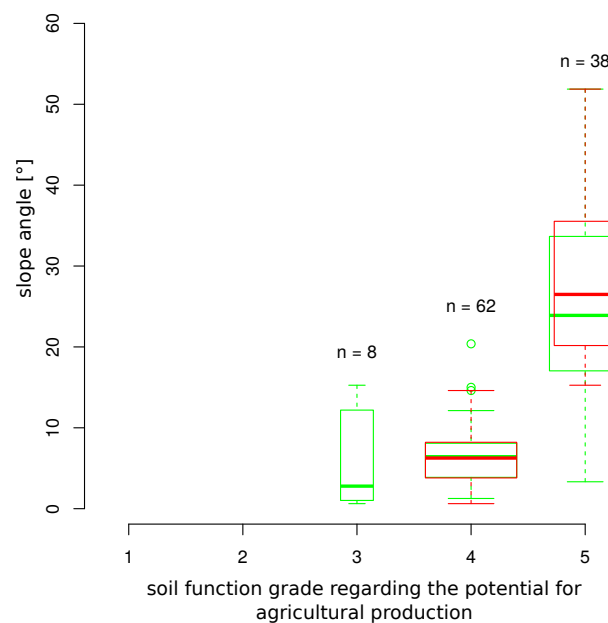
Figure 4. Box plots representing the distribution of the slope angles of the soil profile sites classified according to the original grades (green) and the predicted grades (red).

Analysis of the distribution of the slope values of the predicted as well as original classes (Figure 4) shows that the SVM classifier applied a threshold of 15° to separate the locations with the grades 4 and 5, which is a direct result of this exact slope threshold applied by the algorithm in SEPP which calculates the grades for the potential for agricultural production. The consequent difference in slope of these classes apparently overrules any other potentially present differences with regard to other terrain parameters influencing this potential, and leads to the non-representation of the class with grade 3 in the model output. This issue is a consequence of the specifics of the study area, which can be roughly divided into the valley floors reserved for agriculture and the forested steep slopes. This is further complicated by the dominance of the classes with poor grades, which cannot be solely attributed to the slope threshold, but also the generally rather shallow and skeleton-rich soils encountered.

3.5. Minimum and average precipitation retention capacity

While the grades for minimum precipitation retention capacity are relatively even distributed over the study area, with only the best class having significantly less members than the other classes, the distribution of the average precipitation retention capacity shows a skew towards the classes with higher potential (Figure 2). Grade 2 has the most members (46), constituting almost half of the soil profiles. The feature selection procedure for both soil potentials shows that terrain parameters describing various curvature are best suited to model the difference with regard to this potentials. Surprisingly, the specific curvatures differed for the two potentials. For the average precipitation retention, high resolution cross-sectional curvature was combined with profile curvature at medium (10 m) resolution to achieve a median cross-validated accuracy of 50.9 %. The model output predicts four out of the five possible grade classes, with the intermediate grade 3 missing. The predictions lead to an even larger dominance of grade 2 than in the original data. While the 15 soil profiles points with grade 1 which were attributed grade 2 by the SVM classifier seem acceptable, the 14 soil profiles with grade 4 but classified as having grade 2 are of more concern with regard to the predictive power of the model. The box plot analysis shows that for the model fitting, the SVM classifier concentrated on the outliers of grades 4 and 5, especially with regard to cross-sectional curvature, which also leads to a low producer's reliability for these classes.

For minimum precipitation retention capacity, the best suited parameters were found to be high resolution planar curvature together with minimal curvature based on the low resolution DTM (50 m),

i.e. a more regional scale topography. Combined by a SVM classifier, these predictor variables lead to median cross validated accuracy of just 41.6 %. The confusion matrix shows misclassification between all classes, indicating that not only are the soil profiles distributed evenly amongst the grade classes, but this is also the case with regard to topography, leading to a low prediction accuracy.

3.6. Retention capacity for heavy precipitation events

The capacity for retention of heavy precipitation events is graded as very high by SEPP for 75 soil profiles in the study area, whereas the other grades have relatively low membership numbers, distributed more or less evenly over the remaining classes. Feature selection identifies longitudinal curvature as helpful to seperate the soil profiles with grade 4 from the rest of the profiles, as members of these class show higher, positive curvatures. These does lead to a median accuracy of 73.1 %, but also results in a large number of misclassifications to grade 1, and only a limited number of soil profiles (4) correctly attributed with grade 4. The other classes are not considered in the model output due to the dominance of grade 1 on concave terrain and the assignment of convex areas such as ridges to grade 4.

3.7. Quality of groundwater reformation

The potential for groundwater reformation, as evaluated by the SEPP tool, shows almost 40 % of the soil profiles have a relatively high quality with grade 2, but also 41 profiles belonging to the classes with grades 4 and 5. Accordingly, the SVM algorithm employs cross-sectional and profile curvature to generally divide the soil profiles into the classes 2,4, and 5. This classification leads to a median cross-validated accuracy of 47.5 %. Higher, and, in the majority, positive cross-sectional curvatures are attributed to the poorer classes, where curvature values surrounding zero are linked to grade 2, a class which incorporates a large proportion of soil profiles originally given the grades 1 and 3. While this sort of misclassification may seem acceptable when seeking a general trend with regard to the quality of groundwater reformation, the still substantial number of grade 4 and 5 soils predicted to have grade 2 by the SVM classifier indicates a strong influence of other factors beside topography.

3.8. Potential for providing nutrients for plants

With regard to their potential for providing plant nutrients, 90 % of the soil profiles belong to one of the extreme grades 1 or 5, whereas the intermediate grade 3 has 11 members. Compared to the predictive performance for other soil potentials, this somewhat bimodal distribution leads to a relatively high cross validated accuracy of 71.3 % when using the local terrain parameter minimal curvature at a high DTM resolution of 2.5 m as the sole predictor variable. It is however important to consider that compared to other potentials, the SEPP tool provides only three possible classes for this the nutrient-providing potential. Due to the predominance of the more extreme grades, the model output predicts only these two grade classes, with the majority of the members of intermediate class being attributed to grade 1. In this case, the classifier links the soil profiles with low potential to negative minimal curvature values, while the high potential soils are generally characterized by minimal values not far below zero.

3.9. Potential as a CO2 sink

Of the 108 soil profiles evaluated in the study area with the SEPP tool, 51 were given the grade 1 for their potential as a CO2 sink. The class with the second mos members is grade 3 (33 profiles), followed by the intermediate grade 3 (18 profiles). Almost the same local terrain parameter was chosen by the feature selection procedure as for the plant nutrient potential, with minimal curvature leading to a prediction accuracy of 61.1 %. When the model result, which only leads to two classes, 1 and 4, being predicted for the study area, it is important to acknowledge the SVM classifier is essentially modeling forest landuse for profiles sites with the grade 1. Consequently, areas with less distinct curvature, i.e. values surrounding zero, are classified as having low potential as a CO2 sink, whereas the slopping

regions surrounding the paleovalley, which are in fact mostly covered by forest, are attributed a high potential by the classifier. The majority of the misclassifications are connected to the grade 4, which has a low producer's reliability of 24 %.

3.10. Potential for retention of heavy metals

The distribution of this potential for the soil profiles can be characterized as bimodal, with the more extreme grades 1 and 1 each having 42 members, whereas the remaining 24 profiles are distribute over the 3 intermediate classes, with grade 3 being the largest class containing 10 profile sites. For the model best suited to correctly predict the potential for as many soils as possible, a landform classification based on the 10 m DTM and a search window of 70 m surpassed the other predictor variables with a median classification accuracy of 63.9 %. However, due to the dominance of two distinctly different classes over the remaining classes, the resulting model only predicts these two dominant grades. Providing the model with further predictor variables did not improve the number of correctly predicted classes. In this model, the highest potential for retention of heavy metals is linked to the landform flat, whereas all locations with a different automated landform class were attributed the lowest potential. The members of the three intermediate grades are almost equally divided among the dominant classes without a clear trend.

3.11. Potential for transforming organic contaminants

The class representing the lowest potential with regard to transforming organic contaminants is the most common with 39 member soil profiles. This class, which is characterized by low pH-values and/or low organic matter and clay content, can be predominantly found on ridges formed of Rhyolite outcrops or steeper slopes. Accordingly, the feature selection procedure produces minimal curvature as an important terrain parameter and links grade 5 soils to areas with higher, positive maximal curvature values. Additionally, a landform classification based on a search window of 500 m, thus representing regional-scale topography, is implemented in the SVM classifier and identifies the landform slope as being closely correlated with a low potential for transforming organic contaminants, whereas the landform class flat is mainly attributed to grade 2, which is the class with the highest potential predicted by the SVM classifier (Figure 5). Together, a model fed with these two parameter leads to a median cross-validated correct classification rate of 46.3 %. It predicts the grades 2, 3, and 5, with the majority of class 1 soils being misclassified as grade 3, and all but two of the grade 4 soils as incorporated into the class with grade 5.

3.12. Potential as filter and buffer for organic contaminants

The distribution of this potential over the study area is very one-sided, with all but four soil profiles evaluated as belonging to the class with the poorest potential. As a consequence, modeling this potential is not very productive, as the classifiers will simply classify the four soils with grade 4 as grade 5, which still leads to the high correct classification rate of 96 %.

3.13. Potential for retention of water-soluble contaminants

The most dominant grade for potentially retaining water-soluble contaminants such as nitrate is grade 1, with 47 soil profiles. Grades 2 and 3 are similar with 14 and 17 member, followed by a slight peak in member ship for grade 4 with 22 profile sites. As the climatic framework is the same for the study area, this potential is basically evaluted using soil texture. A SVM-classifier applying high resolution slope and minimum curvature leads to a median cross validated accuracy of 53.7. Due to its dominance in the original data, the predicted class grade 1, which the model links to very low slope angles and minimum curvature values not far away from zero, incorporates a large number of soil profiles with different grades, leading to a very high user's accuracy of 96 %, but also a producer's reliability of only 60 %. The model predicts all classes except for grade 2, however grades 3 and 5 are projected for only two soils each, with grade 3 producing a user's accuracy of only 12 %. The
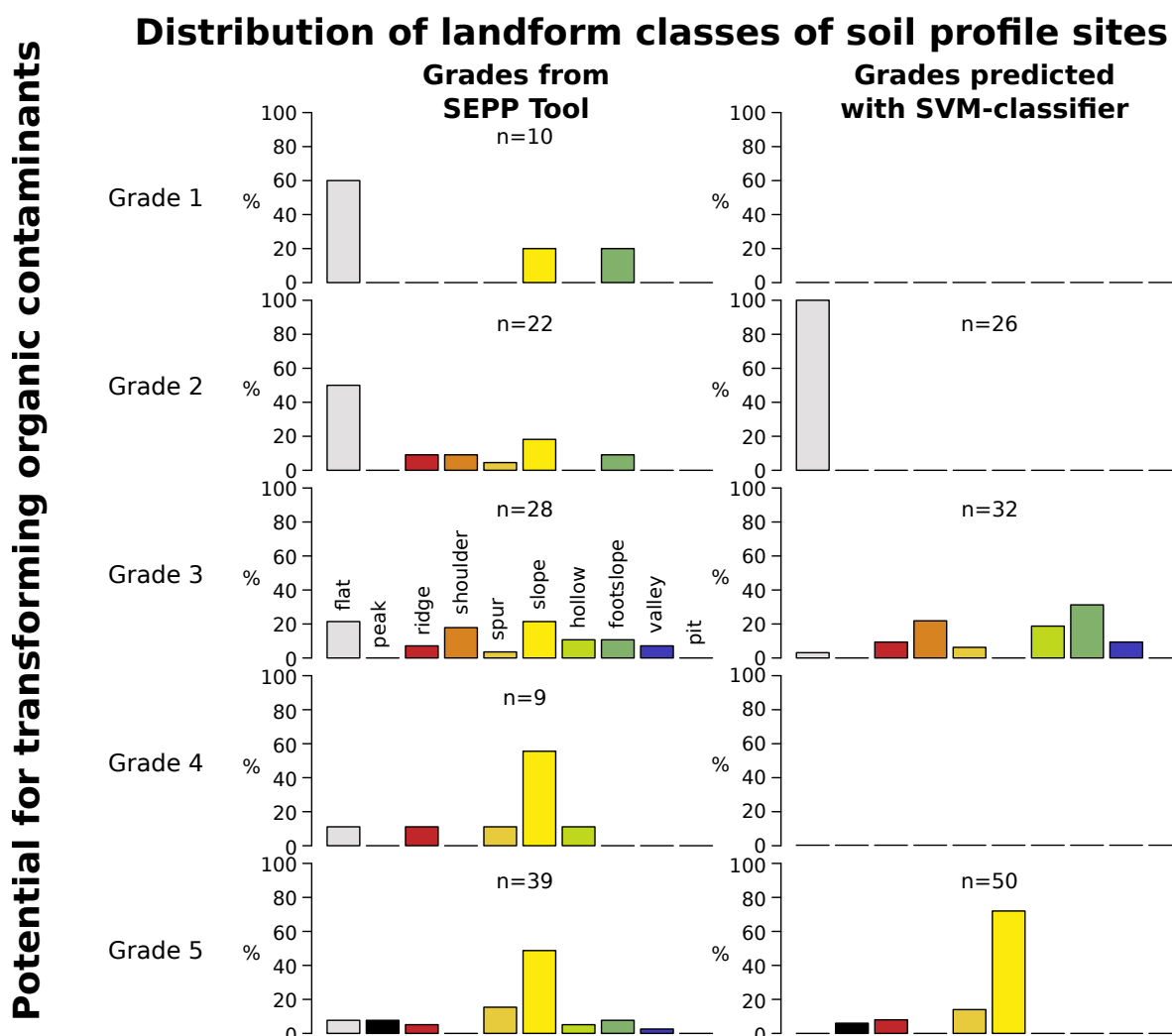
Figure 5. Barplots representing the distribution of the landform classes for the different grades regarding the potential to transform organic contaminants. The left column shows the distribution for the original set of grades as evaluated by the SEPP tool, where the right column shows the distribution of the grades as awarded by the SVM-classifier based on landforms and maximal curvature.

reason for this can be identified in the box plot analysis, which shows that for these 2 smaller classes the classifying algorithm concentrates on outliers in order to produce significant differences to the distribution of the relevant terrain parameters of the other classes.

3.14. Potential as buffer for acidic contaminants

Soils with the highest potential are the most numerous in the study area, constituting 36 % of all profile sites. With decreasing potential, the membership numbers also decrease, with only 9 soils attributed with the lowest potential by the SEPP tool. Planar curvature and slope, both computed with the high resolution (2.5 m) DTM and local-scale window size of 12.5 m, result in a model with a median accuracy of 48.1 %. A major drawback of this model is that with regard to the result based on the majority vote of 100 model runs, it fails to reproduce any members of the classes with grade 2 and 5. While the small sample size may be an issue for grade 5, 83 % of the soils originally with grade 2 are fitted into the predicted grade 1, as these classes share a very similar terrain parameter distribution for both slope and curvature. Grades 3 and 4 are distinguishable by increasingly higher slope and also planar curvature values.

4. Discussion

Some of the algorithms implemented in SEPP, such as the one responsible for calculating the grades for the potential for agricultural production, apply thresholds directly related terrain parameters such as slope. As a consequence, these cut-off values are immediately detected by the SVM classifier and therefore used in its model. This may have the negative effect that the less dominant influence of other topographic factors or landforms on the calculation of this potential is overprinted by the effect of this threshold value and cannot even me mitigated by the addition of further terrain parameters to the model.

The analysis of the SVM classifier predicting average precipitation retention capacity highlights the problem of outliers in the distribution.

A large part of the models lead to separation of the study area into two topographically different regions. On the one hand only gently sloping area of the paleovalley, and on the other the steeper slopes to the east and west. This is also well in line with the different land uses, as the former represents the area which is used for agricultural purposes, while the latter is mainly covered with forest. The potential as a CO_2 sink shows similar issues!

Using less classes, as is the case for the potential to provide nutrients for plants, may be more appropriate when searching for general trends with regard to the influence of topography on the evaluation of soil functions.

Prediction of the potentials proved best for the soil functions....

Problem with bimodal distributions leads to the prediction of only a limited number of classes.

sampling

Important for model evaluation to see if the class where misclassifications occure is "far" away from the correct class.

SVM classification was preferred to other statistical models, for instance logistic regression, for its smooth predictions [20] when applied as a spatial predictor.SVM are used as an exploratory tool here. Analysis of predictions with more predictor variables has shown a decrease in cross validated error due to overfitting with increasing number of explanatory variables.

5. Conclusions

XXX

Abbreviations

The following abbreviations are used in this manuscript:

SVM	Support Vector Machine
DTM	Digital Terrain Model
CV	Cross validation

1. Haslmayr, H.P.; Geitner, C.; Sutor, G.; Knoll, A.; Baumgarten, A. Soil function evaluation in Austria — Development, concepts and examples. *Geoderma* **2016**, *264*, 379 – 387. Soil mapping, classification, and modelling: history and future directions, doi:https://doi.org/10.1016/j.geoderma.2015.09.023.

2. Nestroy, O.; Danneberg, O.; Englisch, M.; Geßl, A.; Hager, H.; Herzberger, E.; Kilian, W.; Nelhiebel, P.; Pecina, E.; Pehamberger, A.; Schneider, W.; Wagner, J. Systematische Gliederung der Böden Österreichischs (Österreichische Bodensystematik 2000). *Mitt. Österr. Bodenkdl. Ges.* **2000**, *60*.

3. Nestroy, O.; Aust, G.; Blum, W.; Englisch, M.; Hager, H.; Herzberger, E.; Kilian, W.; Nelhiebel, P.; G. Ortner and, E.P.; und J. Wagner, A.P.W.S. Systematische Gliederung der Böden Österreichs. Österreichische Bodensystematik 2000 in der revidierten Fassung von 2011. *Mitt. Österr. Bodenkdl. Ges.* **2011**, *79*.

4. BayGLA and BayLfU. Das Schutzgut Boden in der Planung - Bewertung natührlicher Bodenfunktionen und Umsetzung in Planungs- und Genehmigungsverfahren. Technical report, Bayerisches Geologisches Landesamt, München und Bayerisches Landesamt für Umweltschutz, Augsburg, Augsburg, 2003.

5. Lehmann, A.; David, S.; Stahr, K. *TUSEC - Bilingual-Edition: Eine Methode zur Bewertung natürlicher und anthropogener Böden (Deutsche Fassung) Technique for Soil Evaluation and Categorization for Natural and Anthropogenic Soils (English Version)*; Vol. 86, *Hohenheimer Bodenkundliche Hefte*, Universität Hohenheim - Institut für Bodenkunde und Standortslehre, 2008.

6. Beylich, A.; Höper, H.; Ruf, A.; Wilke, B.M. Bewertung des Bodens als Lebensraum für Bodenorganismen im Rahmen von Planunsprozessen. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* **2005**, *107*.

7. Lehle, M.; Bley, J.and Mayer, E.; Veit-Meya, R.; Vogl, W. Bewertung von Böden nach ihrer Leistungsfähigkeit. Luft, Boden, Abfall 31, Landesamt für Umwelt, Messungen und Naturschutz Baden-Würtemberg, 1995.

8. Müller, U.; Waldeck, A. Auswertungsmethoden im Bodenschutz - Dokumentation zur Methode des Niedersächsischen Bodeinformationssystems (NIBIS). GeoBerichte 19, Landesamt für Bergbau, Energie und Geologie, 2011.

9. Gerstenberg, J.H.; Smettan, U. Erstellung von Karten zur Bewertung der Bodenfunktionen - Umsetzung der im Gutachten von Lahmeyer aufgeführten Verfahren in Flächendaten. Technical report, Senatsverwaltung für Stadtentwicklung Berling, 2005.

10. AG Boden. *Methodendokumentation Bodenkunde - Auswertungsmethoden zur Beurteilung der Empfindlichkeit und Belastbarkeit von Böden*; G, Bundesanstalt für Geowissenschaften und Rohstoffe und Staatliche Geologische Dienste in der Bundesrepublik Deutschland: Stuttgart, Germany, 2000.

11. Gallant, J.C.; Wilson, J.P., Terrain Analysis - Principles and Applications; John Wiley & Sons, Inc., 2000; chapter Primary Topographic Attributes, pp. 51–85.

12. Olaya, V. Chapter 6: Basic Land-Surface Parameters. In *Geomorphometry Concepts, Software, Applications*; Hengl, T.; Reuter, H.I., Eds.; Elsevier, 2009; Vol. 33, *Developments in Soil Science*, pp. 141–169. doi:http://dx.doi.org/10.1016/S0166-2481(08)00006-8.

13. GRASS Development Team. *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.0*. Open Source Geospatial Foundation, 2016.

14. Gruber, F.E.; Baruck, J.; Geitner, C. Algorithms vs. surveyors: A comparison of automated landform delineations and surveyed topographic positions from soil mapping in an Alpine environment. *Geoderma* **2017**, *308*, 9–25. doi:http://dx.doi.org/10.1016/j.geoderma.2017.08.017.

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

16. Cortes, C.; Vapnik, V. Support-vector networks. *Machine Learning* **1995**, *20*, 273–297. doi:10.1007/BF00994018.

17. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*, 2014. R package version 1.6-4.

18. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

19. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer-Verlag New York, 2013. doi:10.1007/978-1-4614-7138-7.

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