Geoderma 185–186 (2012) 37–47

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils

Tim Häring [a,b,*], Elke Dietz [c], Sebastian Osenstetter [c], Thomas Koschitzki [d], Boris Schröder [b,e]

[a] BASF SE, Environmental Fate-Modelling, APD/EF, Speyerer Str. 2, 67117 Limburgerhof, Germany
[b] Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350 Freising, Germany
[c] Bavarian State Institute of Forestry, Hans-Carl-von-Carlowitz-Platz 1, 85354 Freising, Germany
[d] Geoflux GbR, Schleiermacherstraße 15, 06114 Halle (Saale), Germany
[e] University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

ARTICLE INFO

Article history:
Received 27 April 2011
Received in revised form 8 March 2012
Accepted 7 April 2012
Available online 7 May 2012

ABSTRACT

Detailed knowledge on the spatial distribution of soils is crucial for environmental monitoring, management, and modeling. However soil maps with a finite number of discrete soil map units are often the only available information about soils. Depending on the map scale or the detailing of the map legend this information could be too imprecise. We present a method for the spatial disaggregation of map units, namely the refinement of complex soil map units in which two or more soil types are aggregated. Our aim is to draw new boundaries inside the map polygons to represent a single soil type and no longer a mixture of several soil types. The basic idea for our method is the functional relationship between soil types and topographic position as formulated in the concept of the catena. We use a comprehensive soil profile database and topographic attributes derived from a 10 m digital elevation model as input data for the classification of soil types with random forest models. We grouped all complex map units which have the same combination of soil types. Each group of map units is modeled separately. For prediction of the soil types we stratified the soil map into these groups and apply a specific random forest model only to the associated map units. In order to get reliable results we define a threshold for the predicted probabilities at 0.7 to assign a specific soil type. In areas where the probability is below 0.7 for every possible soil type we assign a new class "indifferent" because the model only makes unspecific classification there. Our results show a significant spatial refinement of the original soil polygons. Validation of our predictions was estimated on 1812 independent soil profiles which were collected subsequent to prediction in the field. Field validation gave an overall accuracy of 70%. Map units, in which shallow soils were grouped together with deep soils could be separated best. Also histosols could be predicted successful. Highest error rate were found in map units, in which Gleysoils were grouped together with deep soils or Anthrosols. To check for validity of our results we open the black box random forest model by calculating the variable importance for each predictor variable and plotting response surfaces. We found good confirmations of our hypotheses, that topography has a significant influence on the spatial arrangement of soil types and that these relationships can be used for disaggregation.

1. Introduction

Knowledge on the spatial distribution of soils and soil attributes is crucial for many tasks in environmental management, monitoring, and modeling. In forestry, for instance, high-resolution spatial information on soils is required in order to conduct sustainable management of forests which concerns the site-specific environmental conditions in relation to species-specific requirements (e.g. Falk and Mellert, 2011; Thwaites and Slater, 2000). Up to now, soil maps representing categorical soil units in finite number of map entities were the most common source of spatial soil information in environmental authorities (Hartemink et al., 2010). In order to allot soil properties to the soil map, it is common to assign several representative soil profile data to the different map units (Ad-hoc-AG Boden, 2005; Legros, 2006; Soil Atlas of Europe, 2005). Soil properties can then be derived from soil polygon maps by calculating area-weighted or non-weighted averages across the different soil profiles in each map unit.

Discretizing soils into several soil units is a challenge for the mapper, since the spatial distribution of soils and their associated properties can change significantly within short distances and due to the

* Corresponding author at: Landscape Ecology, Technische Universität München, Emil-Ramann-Strasse 6, 85350 Freising, Germany. Tel.: +49 621 60 27587; fax: +49 621 60 27945.
E-mail addresses: tim.haering@basf.com (T. Häring), elke.dietz@lwf.bayern.de (E. Dietz), sebastian.osenstetter@lwf.bayern.de (S. Osenstetter), info@geoflux.de (T. Koschitzki), boris.schroeder@tum.de (B. Schröder).

0016-7061/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2012.04.001

continuous nature of soil (Heuvelink and Webster, 2001; Webster and Becket, 1968). In nature soils do not occur as discrete bodies with sharp boundaries (Odgers et al., 2011a). Therefore, mapping soils as categorical map units can be criticized from an ontological point of view because it contradicts this situation (e.g. Burrough and Frank, 1995). Nevertheless, it has been proven to be practical and effective, because it allows for structuring our knowledge by classification (Legros, 2006; Webster and Becket, 1968). A common approach in soil mapping is the aggregation of several soil types with different soil properties into one single complex soil map unit depending on the specific mapping scale on the one hand and the small scale heterogeneity of soils across the landscape on the other (Ad-hoc-AG Boden, 2005; IUSS Working Group WRB, 2010; Soil Atlas of Europe, 2005).

The aim of this study is to present an approach for disaggregating complex soil map units. Even though the construction of complex map units is comprehensible from a mapper's point of view, it poses a question regarding site-specific forest management or land evaluation where aggregated soil map units may cause problems. To assign soil physical or soil chemical properties to map units, usually several representative soil profiles have to be selected from an existing soil data base or by analyzing soil material in the laboratory obtained from a soil pit. If there are different soil types combined in one complex map unit, it may cause unrealistic results when calculating areal weighted or non-weighted means. As an example consider following map unit (in which the originally German soil types were translated into the international WRB system):

> *Soil complex with small scale variation of Stagnosols and Leptosols. Very to extremely blocky-stony, sandy-loamy periglacial detritus of amphibolites, diorites, and gabbros.*

Stagnosols are characterized by periodically stagnating surface water leading to mottled color pattern or bleaching due to anaerobic conditions. They develop on a wide variety of unconsolidated materials and can be found in flat or gently sloping terrain positions (IUSS Working Group WRB, 2007). In contrast, Leptosols are very shallow soils and extremely gravelly and/or stony. They can be found on exposed landscape positions with strongly dissected topography (IUSS Working Group WRB, 2007). Clearly, the properties of those two soil types are very different, e.g. with respect to their suitability for planting tree species or to their vulnerability regarding windfall. A typical Stagnosol belonging to this map unit has an available water capacity (AWC) of 167 mm/m^2, an impermeable layer at a depth of 35 cm and a fraction of coarse fragments of 21%. In contrast, a typical Leptosol in the same map unit has an AWC of 45 mm/m^2, no impermeable layer and a fraction of coarse fragments of 76%. Clearly this causes problems in calculating the mean values for these attributes from e.g. ten Stagnosols and seven Leptosols, as the resulting calculated mean will express neither the characteristic properties of Stagnosols nor Leptosols correctly.

Scale issues in soil science either with respect to transforming soil information to finer scales (downscaling, disaggregation) or to coarser scales (upscaling, aggregation) have been addressed in literature (Carre et al., 2008; Heuvelink and Pebesma, 1999; Odgers et al., 2011a, 2011b; Panagos et al., 2011). According to Cheng (2008), downscaling is the process of estimating values for smaller scales without observation of the values available in surrounding locations. Soil distribution at one scale is therefore used to estimate the distribution at another scale.

In order to transform soil information to another scale, spatial prediction techniques have been applied in digital soil mapping literature (Grunwald, 2010). Beside some theoretical considerations for aggregation and disaggregation of soil information, McBratney (1998) proposed three approaches for disaggregation of polygon soil maps: transfer functions, fractal analysis, and pycnophylactic splines. De Bruin et al. (1999) used stepwise image interpretation and inductive learning to formalize soil–landscape relationships. Terrain objects, which were delineated from aerial photographs, were connected with location-specific soil sample data. Bui and Moran (2001) apply decision trees for disaggregation and extrapolation of fluvial facies units into unmapped areas.

In areas where no soil profiles were available and no detailed information on where in the landscape a specific soil type of a complex map unit is located, several studies proposed clustering methods for spatial predictions. Bui and Moran (2001) use *k*-means clustering to classify soils with Landsat MSS bands, slope position and relief as predictor variables. Yang et al. (2011) used fuzzy clustering to quantify soil–landscape relationships on a 1:20.000 soil map in Canada. The extracted knowledge was used for refined soil mapping using the Soil Land Inference Model SoLIM. Similarly, Smith et al. (2010) disaggregated soil maps in the Canadian province of British Columbia using terrain attributes, landform classes, and ecological subzones as predictor variables for fuzzy classification rules.

However, in cases where representative soil profiles as training data were available, supervised classification is an alternative method for spatial prediction. The benefit from supervised classification is its ability to estimate prediction accuracy and the identification of clearly described map units or subunits. Thereby it is possible to follow the traditional top-down approach in soil mapping to divide an existing map unit in more homogeneous sub-units and leave the former boundaries of map polygons unchanged. This is in contrast to the aforementioned studies which result in completely disaggregated soil maps.

However, there are situations in which dissolving is not intended. Kempen et al. (2009) presented an approach to update the existing 1:50.000 Dutch soil map. This was motivated due to an area-wide transformation of peat soils to other soil types.

Similar to Kempen et al. (2009), we do not alter the boundaries of soil polygons in our study. Even though polygons of soil maps cannot be viewed as 100% correct, soil maps serve as a basis for several applications. We aim to disaggregate not the entire mapped area, but only complex map units. Therefore, the existing methods are not useful for our purpose.

Moreover, regarding the number of different classes it is unfeasible to estimate one model for our entire study area. If the number of classes becomes very high—as in our case 104 groups of map units—one model needs to be either very complex or it is not able to discriminate between all single classes (cf. Bui and Moran, 2001; Kempen et al., 2009). Therefore, we developed an approach applying comparably simple but class-specific models for the delineation of sub-areas. Almost all studies in which categorical map units were disaggregated considered a smaller number of classes (i.e. less than 10 classes: Behrens et al., 2010; Brus et al., 2008; Sun et al., 2011; 10 to 20 classes: Debella-Gilo and Etzelmüller, 2009; Hengl et al., 2007; Kempen et al., 2009; Moonjun et al., 2010; more than 20 classes: Grinand et al., 2008; Smith et al., 2010; Stum et al., 2010). Only Smith et al. (2010) predicted more than 100 classes, however not in a single model but with knowledge-based fuzzy classification rules for every class separately.

Because many map units do not occur on the entire map but only on small subareas a stratification of the study-area to get parsimonious models is also favorable in our case.

In Bavaria, traditional soil maps were the main source of soil information. Even though 9924 soil profiles are available in Bavaria for modeling purposes, it was not possible to generate soil maps using classical spatial interpolation techniques. In many cases, several sampling points were located on representative landscape positions within few hundreds of meters on a catena. Therefore, we encountered a high density of profiles in some areas, whereas in others the sampling is rather sparse. Moreover, soil properties data are available only for a small subset of the sample (e.g. soil chemical properties are available for only 11% of the profiles). For the majority of samples, we only got information on the soil type, coordinates, succession of soil horizons, and texture. To generate high-resolution soil data we needed to strike a new path.

In this study, we present a method to disaggregate soil map units, especially suitable for complex map units in which two or more different soil types were combined. Thereby we follow the traditional top-down approach in soil mapping to divide an existing map unit in more homogeneous sub-units. We use decision tree-based models to quantify the relationship between soil types and topography and use these models to predict the single soil types within the complex map unit.

2. Materials and methods

2.1. Study area

The study area is the forest area of the German federal state Bavaria in the south-east of Germany which is covered by the 1:25.000 soil map (Figure 1). Bavaria has an area of ca. 70 550 km^2 and is characterized by diverse physical-geographic conditions. It measures approximately 366 km in N–S direction and 352 km in E–W direction and has a long altitudinal gradient from Kehl am Main (102 m asl) to Germany's highest mountain in the south (Zugspitze 2962 m asl). The climate is cool humid with a mean annual temperature decreasing from 10.3 °C at lower elevations to −4 °C at the summits and annual precipitation ranging from 483 mm up to 2800 mm. Due to a high geological diversity (from crystalline basement rocks, volcanic rocks, different triassic sedimentary rock, large areas of limestones, tertiary molasse to quaternary fluvial, glacial, and aeolian deposits), Bavaria is also characterized by a rich mosaic of soil types.

2.2. Soil profile database

We established a soil profile database as our main information about soil types for statistical modeling. Our aim was to merge all soil profiles within forests which were available in the forest and environmental administration in the state of Bavaria, inside as well as outside of the mapped area (Figure 2). Since our focus lies on the forest area, we only took soil profiles into account which were located in forests.

We ended up with 9924 soil profiles consisting of 93 different soil types according to the German soil classification system. When using the terms *soil type* and *complex soil map unit*, we refer to the German soil classification system (Ad-hoc-AG Boden, 2005). A soil type is characterized by a specific sequence of soil horizons influenced by soil forming processes. Every soil map unit contains information about the soil type and its parent material. A soil type in the German soil classification system is similar to the reference soil groups in the World Reference Base for Soil Resources (IUSS Working Group WRB, 2007). In cases in which there are two or more soil types associated into one single map unit, we use the term *complex soil map unit*. Within the WRB system, there is a similar approach for the description of complex units (IUSS Working Group WRB, 2010). Throughout this paper, the German soil types were translated into WRB names. All profiles were attributed with geographic coordinates to join the points with topographic observations.

2.3. Soil map

We used the official soil map of Bavaria (Bavarian Environment Agency, *Übersichtsbodenkarte ÜBK25*, http://www.lfu.bayern.de). It is

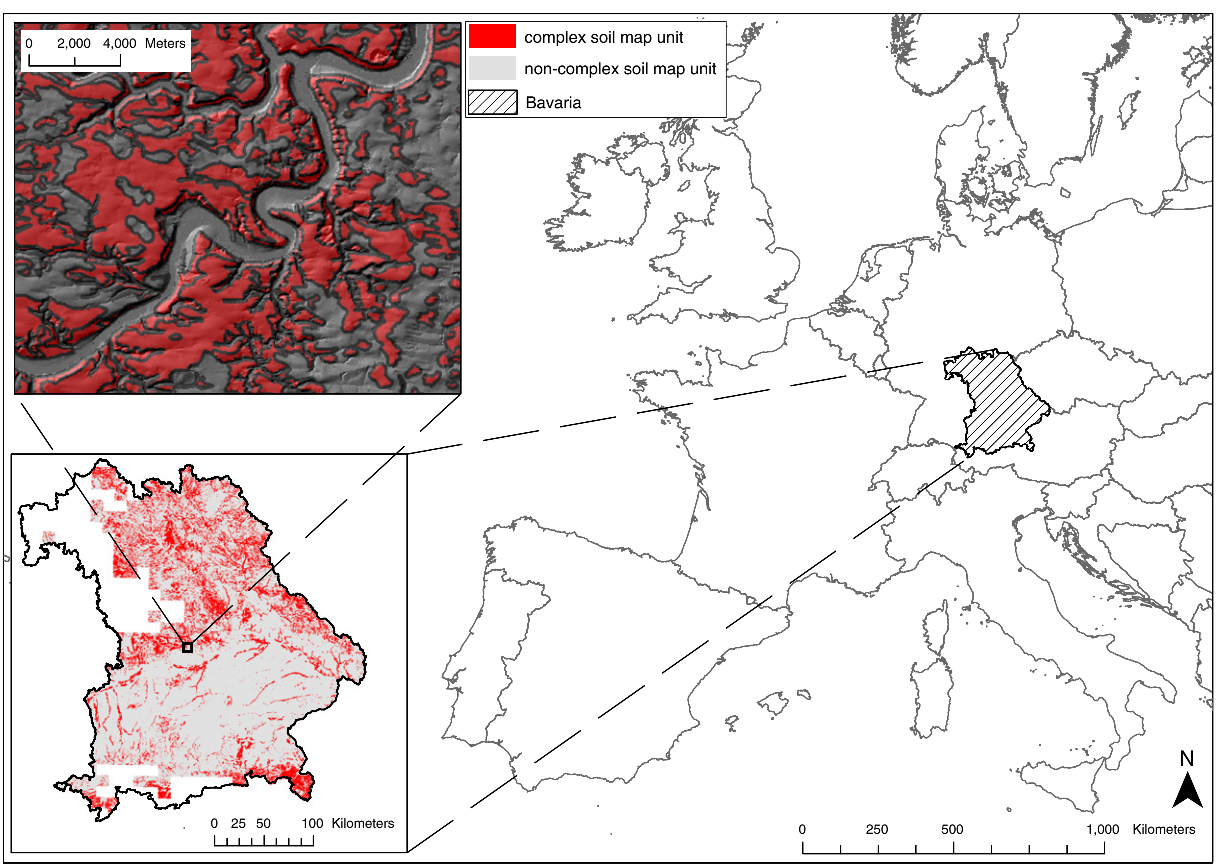

Fig. 1. Complex and non-complex soil map units of the study-area and location in Germany and Europe.

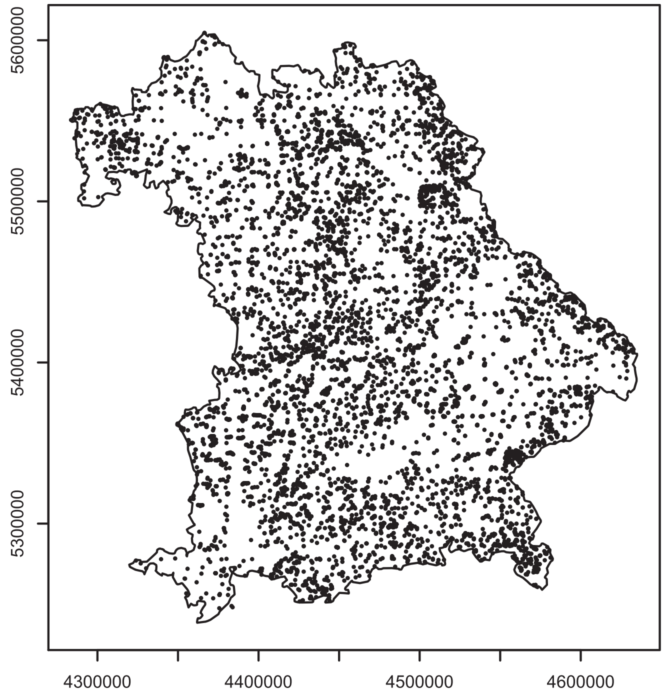

Fig. 2. Distribution of 9924 soil profiles located in forested areas.

mapped with a scale of 1:25.000 and covers 64 071 km^2 which is about 80% of the Bavarian State territory (south-east Germany, Figure 1). The map contains more than 700 different map units, i.e. a clearly defined entity in the map legend, from which more than 450 are complex map units containing more than one single soil type. It accounts for the official German guideline for soil mapping (Ad-hoc-AG Boden, 2005). Soil mapping in Germany follows the concept of substrate-systematic mapping, which means that every soil map unit contains information about the soil type and its parent material.

2.4. Topographic covariates

We use a digital elevation model (DEM) with a cell size of 10 m as basis for the delineation of topographic attributes. It was constructed using airborne laser-scan data by the Bavarian Topographical Survey and has a vertical accuracy of 0.3 m and a positional accuracy of approximately 1 m. Errors and anthropogenic elements like roads or settlements in the DEM were eliminated by the topographical survey before terrain attributes were derived.

We derived a set of 23 terrain attributes. Besides classical local terrain attributes calculated with a 3 × 3 moving window (e.g. slope gradient, curvatures), we derived complex secondary terrain attributes (Pike et al., 2009; Wilson and Gallant, 2000). In addition, we used different window sizes (8 × 8, 15 × 15) to analyze the effect of scales and spatial context (Grinand et al., 2008; Smith et al., 2006). To select the most important variables and to remove highly correlated variables, we applied the feature selection method ReliefF (Kira and Rendell, 1992; results not shown). ReliefF measures the usefulness of terrain attributes based on their ability to distinguish between very similar soil profiles belonging to different soil types. We found a high importance of secondary terrain attributes. Slope gradient (calculated on a 3 × 3 window) was the only important local terrain attribute. Finally, we got seven topographic attributes which were used for modeling (Table 1). All terrain parameter were calculated in SAGA-GIS (Saga Development Team, 2011).

Table 1
Terrain attributes used as topographic covariates for statistical modeling. All parameters were calculated in SAGA-GIS.

Terrain attribute	Definition
Topographical wetness index (twi)	SAGA Wetness Index, implemented in SAGA GIS (Böhner et al., 2002)
relative height (hut)	vertical distance to channel network
floodplain index (fpi)	Indicates flat areas with high flow accumulation and low relative elevation (1 + slope gradient) * (1 - twi) * (1 + relative height)
modified floodplain index (fpi2)	Indicates areas with high flow accumulation and low relative elevation (1 + relative height) * (1 – twi)
mass balance index (mbi)	Indicates areas of erosion and deposition (Möller et al., 2008) (plan-curvature + profile-curvature) * (1 + slope)
Slope gradient	according to Zevenbergen and Thorne (1987)
mid-slope position	The higher the relative vertical distance to the mid slope in valley or crest directions the higher this value. (Böhner and Antonic, 2009) |2 * normalized.height – 1|

2.5. Data preparation

First, all georeferenced soil profiles were attributed with the seven topographic attributes in order to establish a dataset for subsequent modeling. We interpolate our profiles with the terrain parameter using the bilinear interpolation method. Due to the fact that the coordinates of our soil profiles were mainly measured with GPS, we had to deal with a certain degree of spatial uncertainty, because the forest canopy blocks and reflects the satellite signal causing multipath effects and signal losses that lower the accuracy (Mauro et al., 2011). Smoothing values by interpolation attenuates this problem.

Secondly, we grouped the 250 complex soil map units that needed to be disaggregated, according to their soil types. In cases in which two or more map units had the same combination of soil types, they were grouped together, e.g. Calcaric Cambisols and Umbric Leptosols developed on dolomite on the one hand and on limestone on the other. Finally, we got 104 different groups of map units with the same combination of soil types. 89 groups consist of only two different soil types, the remaining 15 groups consist of three different soil types. The number of profiles which were assigned to the 104 groups ranged from 35 to 2668 (median = 339.0, mean = 638.4).

2.6. Soil landscape relationship

Topography is one of the elementary soil forming factors. The influence of relief on the spatial distribution of soils especially on field to landscape scale was first formulated in the catena concept (Milne, 1935). Numerous studies in soil landscape modeling and digital soil mapping used topographic attributes as spatial covariates (see for an overview Behrens et al., 2010; Deumlich et al., 2010; McBratney et al., 2003; Möller et al., 2008). Our method relies on these relationships. We hypothesize that it is possible to separate different soil types within a complex map unit by quantifying the functional relationship between soil type and topography by means of statistical modeling. We expect that we can derive a specific topographic fingerprint for each soil type by investigating the distributions of several topographic attributes for different soil types respectively. If the fingerprint of a specific soil type is different from that of an accompanying soil type, we are able to draw new boundaries inside a soil map polygon.

Fig. 3 shows boxplots of the aforementioned Stagnosols and Leptosols for three topographic attributes.

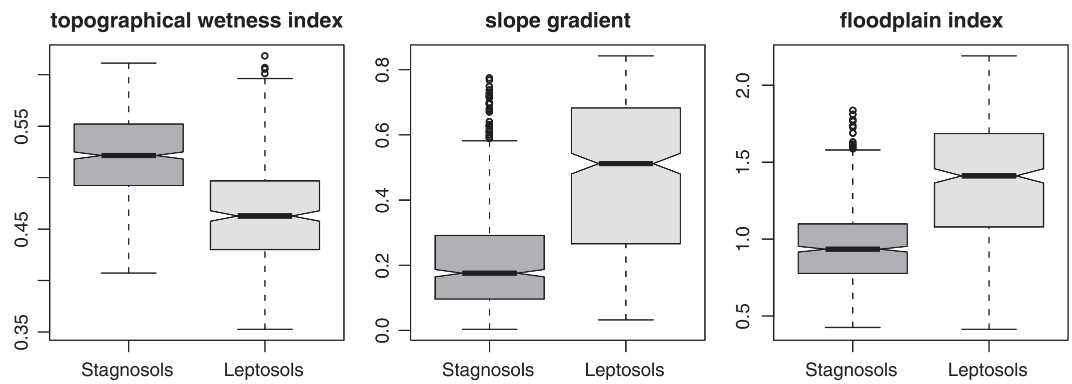

Fig. 3. Boxplots of topographical wetness index, slope gradient, and floodplain index for Stagnosols and Leptosols of our soil profile database. The values for the topographic attributes were transformed. The soil types show a significant difference regarding topographical gradients. We use these differences to classify the soil types with random forest.

The plots show significant differences in the appearance of these soil types on topographic gradients. These differences can be used for classification by decision tree-based classification models.

There are several complex map units in our soil map in which other parameters than topography might have a significant influence on the spatial arrangement of and the differentiation between different soil types like geological, chemical, or hydrological conditions. For example a map unit containing different cambisols developed on tertiary marl or sandstones or a map unit containing calcaric and dystric Gleysols is much more influenced by geology than by topography. Different geologic conditions may also be reflected by topography, but not at all times. In order to produce meaningful results, we selected only those complex map units that can be separated with topographical information according to our expert-knowledge. From the 450 existing complex map units, we selected 250 units for spatial disaggregation with our approach presented here. The area of these 250 complex units is 14776 km^2, which is 23% of the entire soil map. Reduced to the forested area, these 250 map units cover 6150 km^2 (i.e. 30% of the soil map under forest).

2.7. Statistical modeling

Classification of our soil types was performed with random forest (Breiman, 2001). Random forest is an ensemble method in which many different classification trees are combined to produce a more stable and accurate classification compared to a single decision tree (Bauer and Kohavi, 1999; Breiman, 1996; Dietterich, 2000). Each tree is built on a bootstrap sample of the given data. To form the ensemble, the different trees are combined using bagging (bootstrap aggregating). The resulting “forest” is a “random” forest because at each split only a random subset of the candidate predictors is considered for the binary partition (Elith and Graham, 2009). This de-correlates the trees, improves the variance reduction and finally leads to more accurate predictions (Bühlmann and Yu, 2002; Strobl et al., 2009). The predictions of each single tree are combined using a majority vote to get a final ensemble prediction. In recent years, random forests have been widely used in digital soil mapping (e.g. Roecker et al., 2010; Stum et al., 2010).

Widely used decision trees like Breiman et al.'s CART (1984) or Quinlan's C5 (http://www.rulequest.com/, 1993) were built on recursive partitioning and impurity reduction. Entropy measures, like the Gini Index or the Shannon Index, are used to quantify the impurity in each node (Hastie et al., 2009; Strobl et al., 2009). When working with environmental data and in particular topographical data as covariates for statistical modeling, we always have to concern multicollinearity (Graham, 2003; Hengl and MacMillan, 2009). Strobl et al. (2007, 2008) showed that CART-based random forest implementations (like the R-package randomForest, Liaw and Wiener, 2002) are biased when predictor variables are correlated or measured on different scales. Therefore, we used a random forest implementation applying conditional inference trees as base learners, which has been proven to be unbiased (Hothorn et al., 2006; Strobl et al., 2010). The splitting in recursive partitioning in conditional inference trees is based on significance tests of independence between any of the predictors and the response. Such a framework is implemented with the functions ctree() and cforest() in the package party in R (Hothorn et al., 2006; Müller et al., 2009; R Development Core Team, 2010).

Typically, predictions of classification models like random forest are response classes. The predictions are made on a majority vote using the predicted probabilities for the present soil types, i.e. the class with the highest probability is assigned (Strobl et al., 2009). We do not use the predicted classes in our study, but an estimate of the conditional class probabilities. We defined a probability threshold at $P>0.7$ to allow for unspecified classifications in the model assigning a specific soil type in the prediction only if its probability exceeds 0.7. All areas with a maximum probability for any soil type below 0.7 were classified as “indifferent”.

Validation of the random forest models was performed using the out-of-bag error. The predictive performance of the model is calculated on those observations which were not included in the learning sample for a specific decision tree, i.e. those observations which were not part of the bootstrap sample of the original data set. Using those out-of-bag observations, we have independent test samples for computing the prediction accuracy. It could be shown that the out-of-bag error is a conservative estimate (Strobl et al., 2009).

In order to detect the dependencies between predictor and dependent variables and to select the relevant predictors, one can calculate variable importance measures. The extraction of important predictors is calculated on the permutation accuracy importance measure. This measure is estimated by randomly permuting the values of a particular variable. By comparing the prediction accuracy before and after permuting a variable we get a measure of variable importance. For plausibility check in this study we use the permutation importance in the party package because it is a reliable measure even in cases with correlated predictors (Strobl et al., 2010).

Since there are some soil types in our profile database which were very frequent (such as Cambisols, Gleysols, or Luvisols) and others that are rather scarce (e.g. Histosols or soil types with stagnic properties), we often had the problem of extremely unbalanced datasets for modeling. So at least one of the soil types constituted only a very small minority of the data which may cause a limited classification performance (Japkowicz and Stephen, 2002). Therefore, we implemented an if-then condition in our modeling framework: if the number of observations for one soil type in our database is greater than a proportion of 2:1 to the number of profiles of the other soil type in the grouped map unit, then we take a random sample of the former to enforce a proportion of 2:1. The proportion of 2:1 is a compromise between having a more balanced dataset on the one hand and using

Table 2
Classification accuracy for unbalanced and balanced dataset.

			Full data					Reduced data				
Map unit	ST1[1]	ST2	n ST1[2]	n ST2	Acc[3]	TP1[4]	TP2	n ST1	n ST2	Acc	TP1	TP2
70	BB[5]	RQ[6]	2000	35	0.998	1	0	75	35	0.7	0.9	0.26
3	BB	BB/CF[7]	2000	155	0.93	0.99	0.03	310	155	0.77	0.85	0.61
92	SS[8]	GG[9]	741	113	0.9	0.97	0.41	226	113	0.83	0.93	0.62
103	SS	FF[10]	741	85	0.95	0.98	0.67	170	85	0.91	0.94	0.85

[1]soil type1, [2] number of observations for soil type 1, [3] Accuracy (correctly classified instances), [4] TP1 = true positive (fraction of soil type 1, that is actually classified as soil type 1), [5] Cambisol, [6] Anthrosol, [7] Calcaric Cambisol, [8] Stagnosol, [9] Gleysol, [10] Leptosols.

sufficient samples on the other (Japkowicz and Stephen, 2002). The influence of an unbalanced dataset on model performance is presented in Table 2.

Even though the overall model performance calculated on the entire data for the unbalanced dataset is better than for the reduced dataset, the problem lies in the ability of the model to discriminate between the two soil types. The evaluation of the different classes separately is performed with true positive measure (TP). TP is the fraction of a predicted class which is actually this class. Fitting a model on a highly unbalanced dataset which only predicts the over-represented class one gets a high classification accuracy but TP = 0 for the under-represented class (see map unit 70 in Table 2).

The entire modeling framework is illustrated in Fig. 4.

2.8. Field validation

In addition to the statistical model evaluation using the out-of-bag error, we estimated model performance in addition on field validation

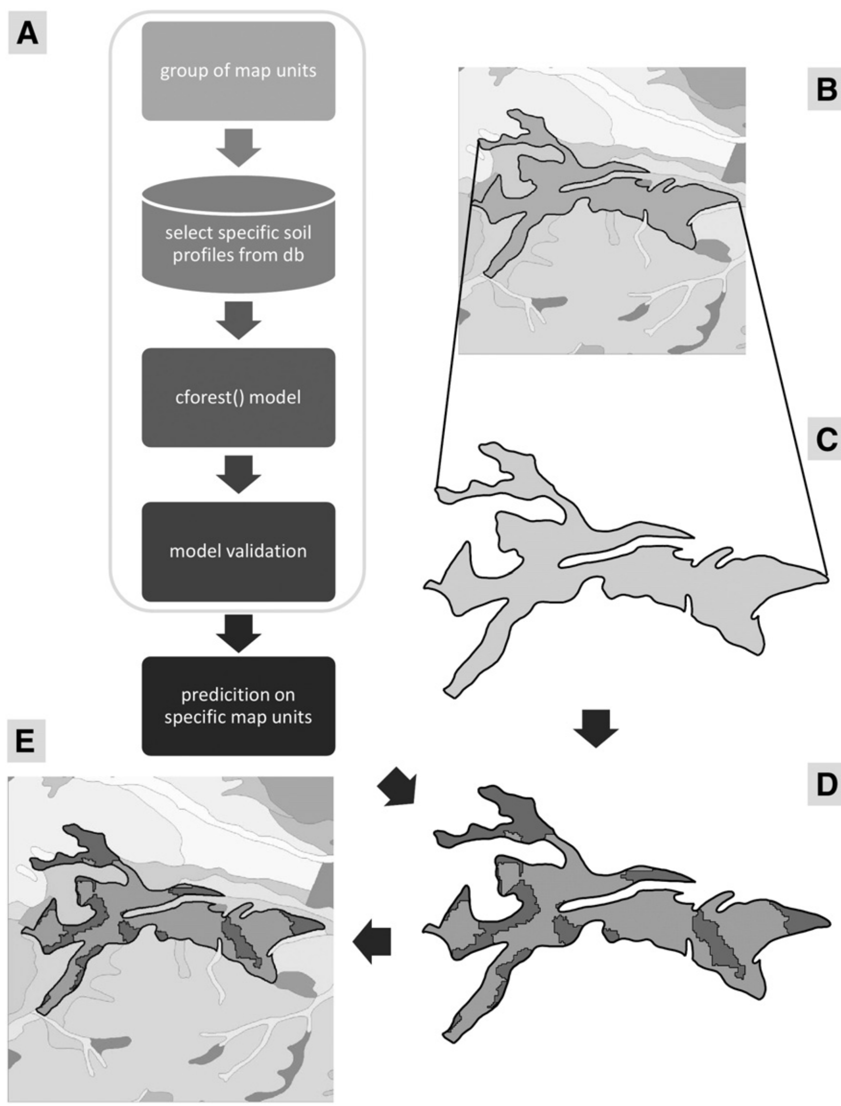

Fig. 4. Illustration of our modeling framework for spatial disaggregation of complex soil map units. We disaggregate only those map units which can be separated with terrain information. The map units were grouped according to their combination of soil types. For classification of soil types we select specific soil samples from a profile database. Statistical modeling is performed with random forest. Model validation is estimated with the out-of-bag error (A). The soil map (B) will be stratified for prediction. We predict soil types only in those areas that belong to a specific group of complex map units (C, D). To generate the final map we merge the single disaggregated parts to one soil map (E).

data as suggested by Brus et al. (2011). Subsequent to modeling, soil experts described augered soil profiles on defined locations. The distribution of sampling locations followed a stratified random sampling design. First, the study area was subdivided based on an official physiographic classification of Bavaria (Wittmann, 1991). Within these areas, sampling locations were randomly distributed over the entire predicted area. In sum we got 1820 validation points.

We grouped the validation data according to the occurrence of soil types in different map units to investigate which soil types could be separated successfully.

3. Results and discussion

3.1. Statistical Modeling

Random forest models are estimated for each of the 104 grouped map units. We calculated 500 trees in every random forest model (the "ntree" argument in cforest). Random forest models with 1000 trees did not improve the performance (results not shown). The models were stable mostly with less than 200 trees. The number of randomly selected variables as candidates at each split (the "mtry" argument in cforest) was three as recommended by Hastie et al. (2009) (mtry = square root of number of predictor variable).

The statistical validation of the models based on the out-of-bag error showed misclassification rates ranging from 0.09 to 0.55 with a median of 0.31. Compared to other digital-soil-class-mapping studies this are quite good results (cf., e.g., Hengl et al., 2007; Kempen et al., 2009; Lemercier et al., 2012; Stum et al., 2010).

3.2. Prediction

After fitting the 104 random forest models, we applied these models on the corresponding regions of the soil map in order to predict the occurrence probability of each soil type that is present in the map unit.

Fig. 5 shows six examples of the final disaggregated soil map. Finally, 57% of the area have been predicted as a specific soil type ($p > 0.7$), whereas 43% have been predicted as 'indifferent'.

3.3. External validation

Estimation of model performance on 1820 field validation points gave an overall accuracy of 70% (1246 correct classified, 540 incorrect, 8 not usable due to erroneous profile descriptions).

The predictive performance depended on the number of available profiles.

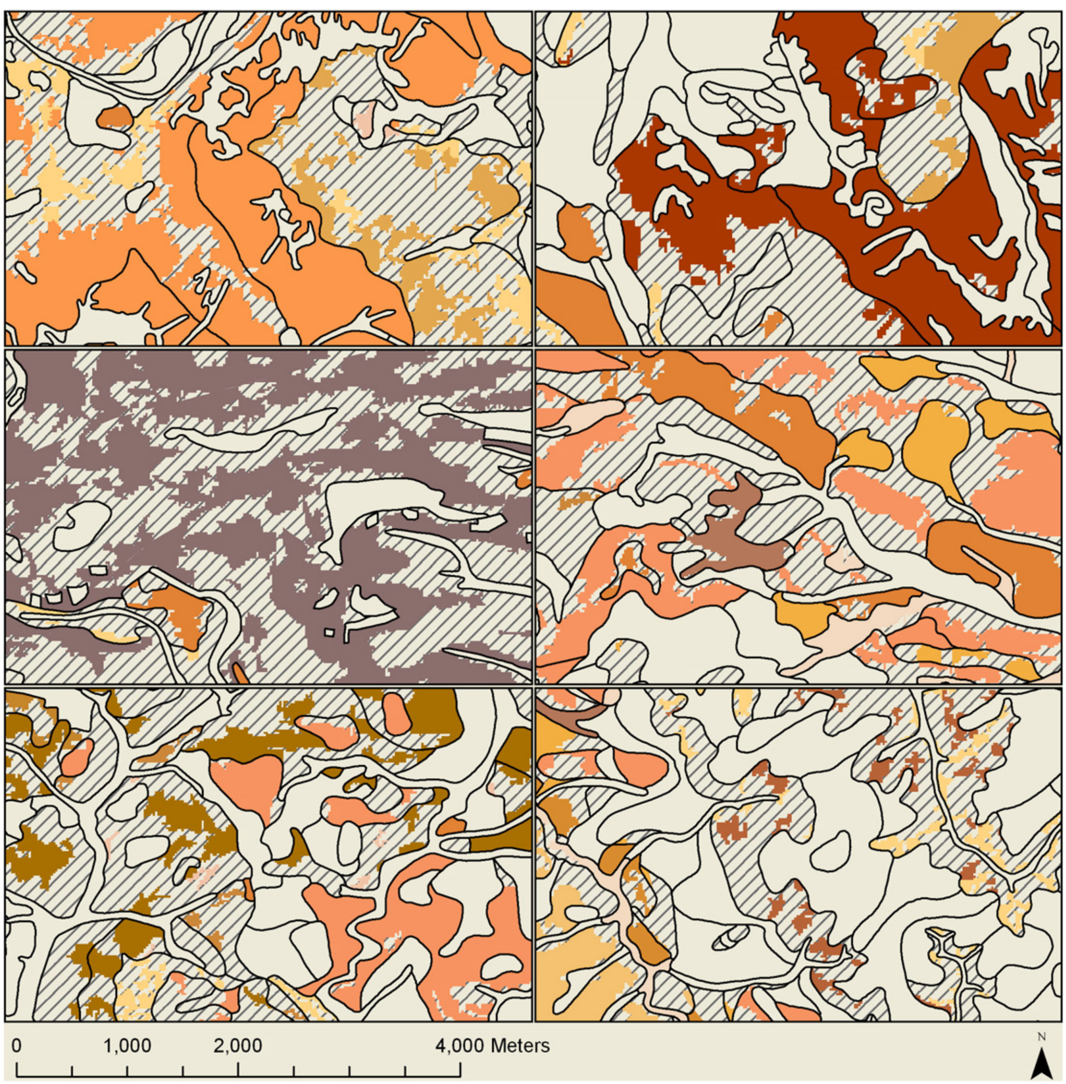

Fig. 5. Six cutouts of our resulting disaggregated soil map. The polygons of the original soil map were shown in solid black lines. Map units which were not considered during downscaling (non-complex map units) were colored in white. Areas which were classified as indifferent (predicted probability for every soil type < 0.7) were hatched.

In Fig. 6 we plot the accuracy calculated on the field validation data over the number of profiles for a given group of map units. We plot only those groups of map units which have more than 15 validation points in order to have reliable accuracy values. Accuracy increases roughly with the number of calibration data. If the number of data points exceeds 350, this correlation disappears and the accuracy values becomes more scattered. However, accuracy values always exceed 0.7. There are two outliers in the plot. One with an accuracy of 0.44 and 1184 calibration profiles which is a map unit in which Stagnosols and Cambisols with stagnic properties were grouped. And a second outlier with an accuracy of 0.5 and 447 calibration profiles, which is a map units in which Gleysols, deep soils and Stagnosols were grouped. Obviously, these two groups of soil types are too similar regarding their topographical properties that they couldn't be separated well.

These findings are very promising and confirm our approach. It seems that even better results are possible, if more profiles are available.

Fig. 7 shows a fluctuation plot, which is a graphical representation of contingency table. The extent of a graph is proportional to count. On the right hand side of the plot the numbers of correct and incorrect predicted validation points are listed.

12 out of 14 groups have more true predictions than false predictions. For these 12 groups, we can conclude that separation between groups of soil types is possible, however with different success, because prediction accuracy differs between the groups. Groups with better reliability are those in which soils are highly influenced by topographic characteristics, which was also reported by Debella-Gilo and Etzelmüller (2009). A high proportion of true predictions can be found in groups which differ in the profile depth (deep soils vs. shallow soil, initial soils vs. shallow soils) due to the strong dependency of profile depth and terrain position. Also Histosols could be predicted very successfully, since there is a strong influence of water availability on their development which is in turn mainly controlled by topography. Our results confirm findings of Seibert et al. (2007) in Swedish forest soils, who could show a strong dependency of Histosols with topography.

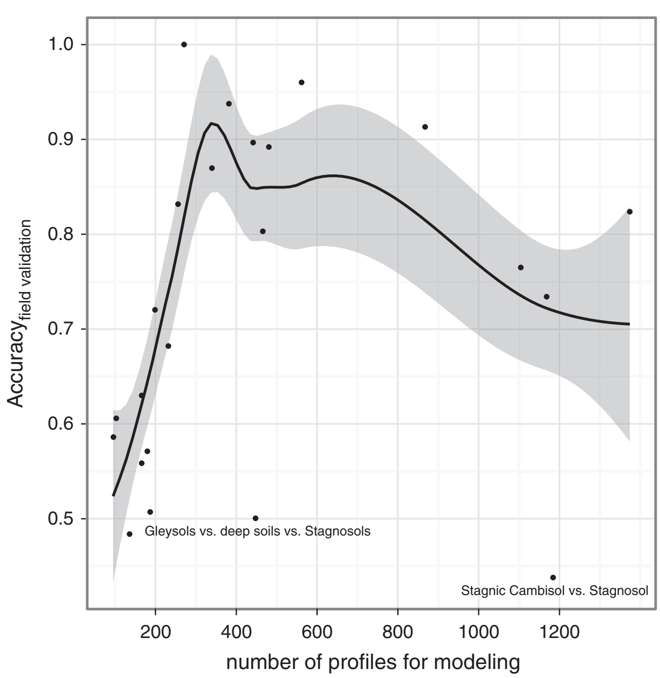

Fig. 6. Classification accuracy calculated on field-validation data plotted over the number of profiles used for training. There is a relationship between model performance and the number of profiles available for calibration. If the number of data points exceeds 350 this correlation disappears and the accuracy values becomes more scattered. However, accuracy values always exceed 0.7.

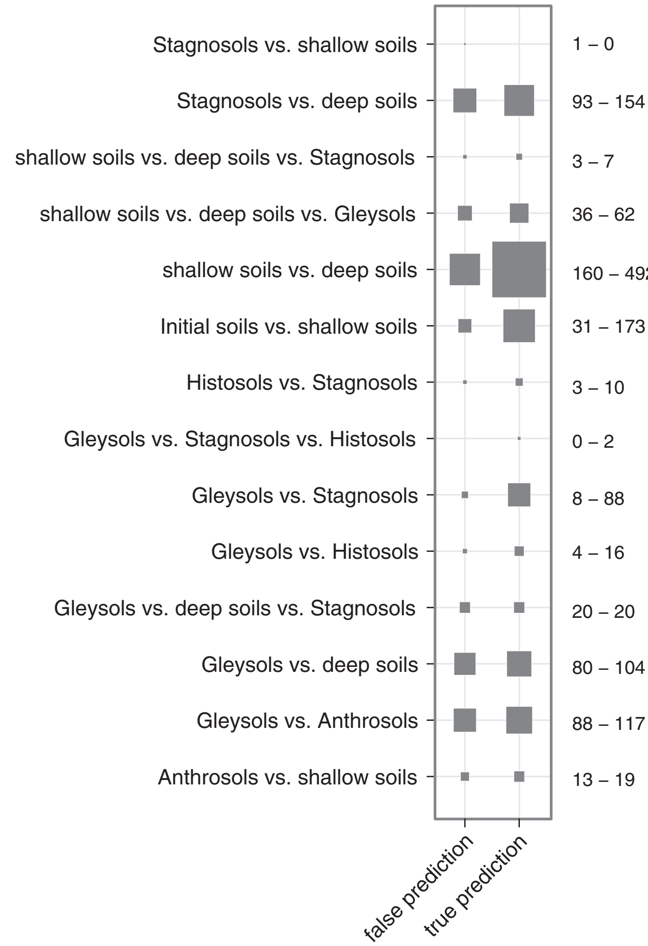

Fig. 7. Fluctuation plot of true and false predicted groups of map units. Numbers on the right hand side indicates the count of true and false. The size of the figures is proportional to the count of true and false which is displayed on the right hand side.

Except from Histosols and Stagnosols map units containing Gleysols could be separated the worst. Mainly because those map units are located in flat valley bottoms where no variability in terrain exists. Therefore, discrimination based on terrain attributes becomes extremely difficult. Similar to Histosols, the discrimination of Gleysols vs. Stagnosols is strongly influenced by the availability of groundwater and surface water respectively, which depends on topography and could therefore be executed successfully.

With the introduction of a threshold at $P>0.7$ for the prediction of soil types we are able to generate results with high accuracy as shown with the field validation data. On the other hand, 43% of the predicted area classified as "indifferent" is not optimal for our purpose. To capture this problem we might reduce the threshold and thereby minimize the indifferent area. The threshold can be reduced until the indifferent class disappears completely and prediction is made on the highest class probability ($P>0.5$). Prediction in such a way is done in many studies (e.g. Behrens et al., 2010; Debella-Gilo and Etzelmüller, 2009; Grinand et al., 2008). However, prediction performance has then to be estimated in more detail on additional validation data which were not available at the moment. Response surface plots (Figure 8) indicate less accurate predictions at smaller probabilities.

3.4. Plausibility check

Finally we calculated the variable importance for each predictor variable in all 104 random forest models and visualized response surfaces for every soil type of the models. This procedure provides

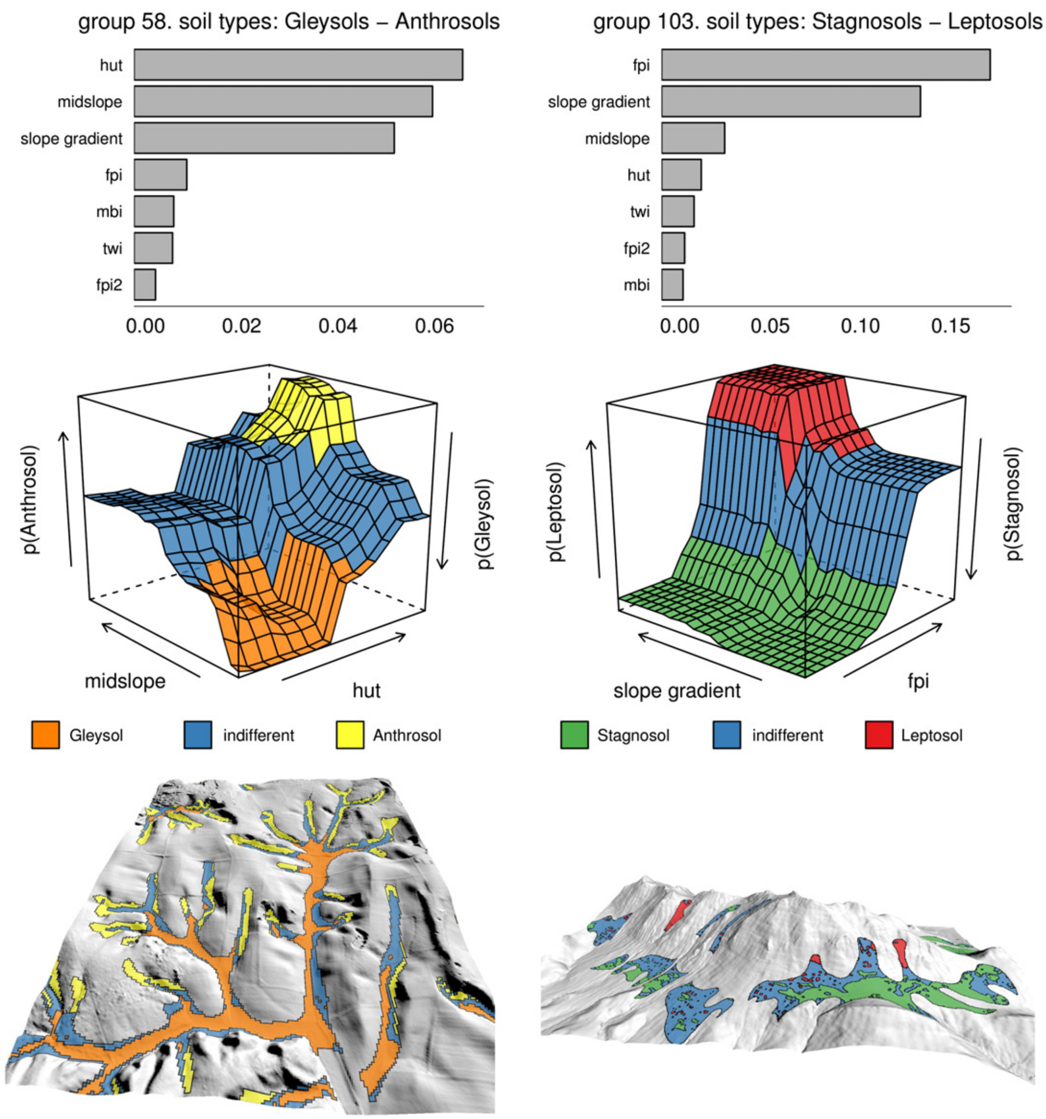

Fig. 8. Variable importance plot (top), response surfaces (middle), and exemplary cutout of the soil map for two randomly selected groups of soil units. In the response surface plots the probability of a specific soil type is plotted over the two most important predictor variables in the model. Coloring of the response surfaces indicates our probability thresholds ($P>0.7$, $0.7>P>0.3$, $P<0.3$). The area of the response surface between the thresholds indicates unspecified predictions ("indifferent"). The plots were used to validate our models. We checked if soil types are located where we expected them following our expert knowledge. The plots confirm our expectations, e.g. shallow Leptosols in exposed position where erosion occur (group 103), Anthrosols at footslopes where soil material is accumulated, and Gleysols in area with low vertical distance to channels (group 58). "fpi" = flood plain index; "fpi2" = modified flood plain index; "hut" = relative height; "twi" = topographic wetness index; "mbi" = mass balance index.

reasonable insights on how the soil types depend on the topographic attributes, which were the relevant attributes in the model, and where specific thresholds can be found. Response surface plots are an effective tool to display model behavior and thresholds in more than one dimension (Elith and Graham, 2009; Lintz et al., 2011).

During this procedure we also checked whether our soil units were predicted in those landscape positions where expert knowledge would expect them. For example, a map unit consisting of Cambisols and Leptosols we would expect to find Leptosols with shallow soil depth on exposed positions with high slope angle and mass balance index where erosion occur. Cambisols, on the other hand, should be located in flat areas with a low mass balance index, i.e. those areas where the development of Cambisols is not disturbed by erosion processes

Fig. 8 shows the variable importance, response surfaces for the two most important variables in the model, as well as an exemplary cutout of the soil map for that particular group for two specific groups of soil units.

All plots reveal meaningful dependencies between soil types and topography. In group 103, Stagnosols and Leptosols were aggregated. The plots show a high importance of flood plain index (fpi) and slope gradient (see Figure 3). The remaining parameters have only marginal influence on the model. The response surfaces show high probability for Leptosols for high value of flood plain index and slope gradient, i.e. exposed terrain positions such as steep backslopes. Stagnosols are influenced mainly by the flood plain index. Slope gradient has no effect on Stagnosols in group 103. These dependencies can also be identified in the map cutout.

Gleysols and Anthrosols, which were aggregated in group 58, could be discriminated mostly by the vertical distance above a channel (hut), the midslope position, and slope gradient. Anthrosols can be found at footslopes where eroded material is accumulated (high midslope). On the other hand, Gleysols have their highest probability in areas with low vertical distances to channels. In the map cutout, Gleysols were located in the flat and inner areas of the original map polygon. Anthrosols could be found at the bottom of slope gradients

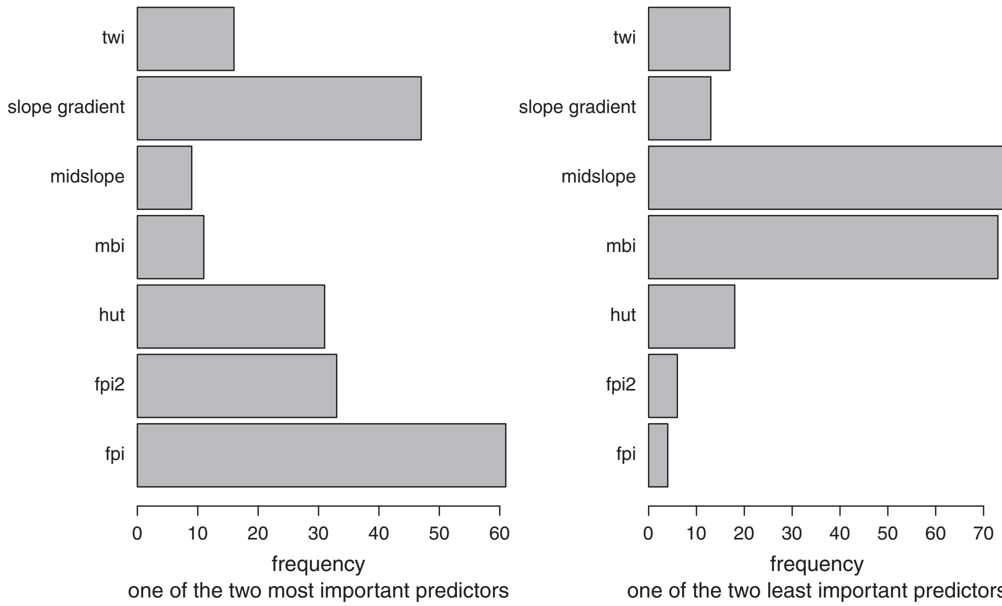

Fig. 9. The overall variable importance estimated over all 104 random forest models. We counted how often a predictor variable is one of the two most and the two least important predictors in all models.

where colluvium has been deposited. The area between these two landscape positions is classified as indifferent.

3.5. Global variable importance

Lastly we evaluated the variable importance of all predictor variables over all 104 random forests. Therefore, we counted how often a variable is one of the two most important predictors in a model and how often a variable is one of the two least important predictors. These frequencies are plotted in Fig. 9.

The plots show patterns which complements one another. Flood plain index (fpi), slope gradient, relative height (hut), and modified flood plain index (fpi2) are frequently one of the two most important predictors in a random forest. Midslope position and mass balance index (mbi) are only selected nine and eleven times respectively as one of the two most important predictors in all models. However, these two predictors are very frequently one of the two least important predictors and the remaining five are all less than 20 times part of this group. We found no preference of using either local parameter (slope gradient, mass balance index) or regional parameters (all the remaining) as important variable. This suggests that both small scale variations as well as landscape scale patterns provide important information in our approach.

4. Conclusions

High-resolution spatial information of soils and soil properties are essential for many application areas in environmental sciences. Soil maps provide the main information on soils. We demonstrate a method for the spatial disaggregation of existing soil maps for providing soil information on higher resolution. Our focus lies on soil map units in which two or more different soil types were aggregated into one map unit.

We found a significant influence of topography on the spatial arrangement of soil types. By comparing different soil types we found a characteristic topographical fingerprint for each soil type. These topographical differences were quantified with unbiased random forest models.

Future work will focus on the selection and assignment of soil profile data with representative soil physical and soil chemical properties to the disaggregated and thereby newly generated map units. Soil property maps will be generated by calculating mean values for each map unit. In those areas, in which the models predict the new "indifferent class", a mean value of the former entire map units is assigned, as it was before disaggregation.

Acknowledgements

The authors acknowledge funding by the Bavarian State Ministry of Nutrition, Agriculture, and Forestry. This work is part of project "KLIP4—Maps for the future" at the Bavarian State Institute of Forestry. We thank the Bavarian Environment Agency and the Bavarian Topographical Survey for providing the data. We also thank the editor and two anonymous reviewers for providing helpful comments. BS acknowledges funding by the German Science Foundation (grant no. SCHR1000/3-1, SCHR1000/6-1).

References

Ad-hoc-AG Boden, 2005. Bodenkundliche Kartieranleitung KA 5. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.
Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning 36 (1–2), 105–139.
Behrens, T., Zhu, A.X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155 (3–4), 175–185.
Böhner, J., Antonic, O., 2009. Land-surface parameters specific to topo-climatology. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Developments in Soil Science, vol. 33. Elsevier, pp. 195–226.
Böhner, J., Köthe, R., Conrad, O., Ringeler, A., Selige, T., 2002. Soil regionalisation by means of terrain analysis and process parameterisation. In: Micheli, E., Nachtergaele, F., Montanarella, L. (Eds.), Soil Classification 2001: EUR 20398 EN. European Soil Bureau, Research Report No. 7, Luxembourg, pp. 213–222.
Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Montery, CA.
Brus, D.J., Bogaert, P., Heuvelink, G.B.M., 2008. Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information. European Journal of Soil Science 59, 166–177.
Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62, 394–407.
Bühlmann, P., Yu, B., 2002. Analyzing Bagging. The Annals of Statistics 30 (4), 927–961.
Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103 (1–2), 79–94.

Burrough, P.A., Frank, A.U., 1995. Concepts and paradigms in spatial information: are current geographical information systems truly generic? International Journal of Geographical Information Systems 9 (2), 101–116.
Carre, F., Reuter, H.I., Daroussin, J., Scheurer, O., 2008. From a Large to a Small Scale Soil Map: Top-Down Against Bottom-Up Approaches. Application to the Aisne Soil Map (France). In: Hartemink, A.E., McBratney, A.B., Mendonça Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, Dordrecht, pp. 203–212.
Cheng, Q., 2008. Modeling local scaling properties for multiscale mapping. Vadose Zone Journal 7 (2), 525–532.
de Bruin, S., Wielemaker, W.G., Molenaar, M., 1999. Formalisation of soil–landscape knowledge through interactive hierarchical disaggregation. Geoderma 91 (1–2), 151–172.
Debella-Gilo, M., Etzelmüller, B., 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: examples from Vestfold County, Norway. Catena 77, 8–18.
Deumlich, D., Schmidt, R., Sommer, M., 2010. A multiscale soil–landform relationship in the glacial-drift area based on digital terrain analysis and soil attributes. Journal of Plant Nutrition and Soil Science 173 (6), 843–851.
Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning 40 (2), 139–157.
Elith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32 (1), 66–77.
Falk, W., Mellert, K., 2011. Species distribution model as a tool for forest management planning under climate change: risk evaluation of Abies alba in Bavaria. Journal of Vegetation Science 22 (4), 621–634.
Graham, M.H., 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84 (11), 2809–2815.
Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143 (1–2), 180–190.
Grunwald, S., 2010. Current State of Digital Soil Mapping and what is next. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, Netherlands, pp. 3–12.
Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R.A., Minasny, B., Montanarella, L., Mendonça Santos, M.L., Sanchez, P., Walsh, M., Zhang, G.L., 2010. GlobalSoilMap.net—a new digital soil map of the world. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 423–428.
Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
Hengl, T., MacMillan, R.A., 2009. Geomorphometry— Key to Landscape Mapping and Modelling. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Elsevier, pp. 433–460.
Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma 140 (4), 417–427.
Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modeling. Geoderma 89, 47–65.
Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present and future. Geoderma 100, 269–301.
Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. Journal of Computational and Graphical Statistics 15 (3), 651–674.
IUSS Working Group WRB, 2007. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome.
IUSS Working Group WRB, 2010. Addendum to the World Reference Base for Soil Resources: guidelines for constructing small-scale map legends using the World Reference Base for Soil Resources. http://www.fao.org/fileadmin/templates/nr/images/resources/pdf_documents/WRB_Legend.pdf (last access 21.02.2011).
Japkowicz, N., Stephen, S., 2002. The class imbalance problem: a systematic study. Intelligent Data Analysis 6 (5), 429–449.
Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151, 311–326.
Kira, K., Rendell, L.A., 1992. A Practical Approach to Feature Selection. In: Sleeman, D., Edwards, P. (Eds.), Proc. Intern. Conf. On Machine Learning (Aberdeen, July 1992). Morgan Kaufmann, pp. 249–256.
Legros, J.-P., 2006. Mapping of the Soil. Science Publishers, Enfield.
Lemercier, B., Lacoste, M., Loum, M., Walter, C., 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. Geoderma 171–172, 75–84.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2 (3), 12–22.
Lintz, H.E., McCune, B., Gray, A.N., McCulloh, K.A., 2011. Quantifying ecological thresholds from response surfaces. Ecological Modelling 222 (3), 427–436.
Mauro, F., Valbuena, R., Manzanera, J.A., Garcia-Abril, A., 2011. Influence of Global Navigation Satellite System errors in positioning inventory plots for tree-height distribution studies. Canadian Journal of Forest Research 41 (1), 11–23.
McBratney, A.B., 1998. Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems 50 (1), 51–62.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.
Milne, G., 1935. Some suggested units of classification and mapping particularly for East African soils. Soil Research 4, 193–198.
Möller, M., Volk, M., Friedrich, K., Lymburner, L., 2008. Placing soil-genesis and transport processes into a landscape context: a multiscale terrain-analysis approach. Journal of Plant Nutrition and Soil Science 171 (3), 419–430.
Moonjun, R., Farshad, A., Shresha, D.P., Vaiphase, C., 2010. Artificial Neural Network and Decision Tree in Predictive Soil Mapping of Hoi Num Rin Sub-Watershed. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 151–163.
Müller, D., Schröder, B., Müller, J., 2009. Modelling habitat selection of the cryptic Hazel Grouse *Bonasia bonasia* in a montane forest. Journal of Ornithology 150, 719–732.
Odgers, N.P., McBratney, A.B., Minasny, B., 2011a. Bottom-up digital soil mapping. I. Soil layer classes. Geoderma 163, 38–44.
Odgers, N.P., McBratney, A.B., Minasny, B., 2011b. Bottom-up digital soil mapping. II. Soil series classes. Geoderma 163, 30–37.
Panagos, P., Van Liedekerke, M., Montanarella, L., 2011. Multi-scale European Soil Information System (MEUSIS): a multi-scale method to derive soil indicators. Computational Geosciences 15 (3), 463–475.
Pike, R.J., Evans, I.S., Hengl, T., 2009. Geomorphometry: A Brief Guide. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry. Concepts, Software, Applications. Elsevier, pp. 3–30.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco.
R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0. http://www.R-project.org.
Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., Blinn, C., 2010. A Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 369–384.
Saga Development Team, 2011. System for Automated Geoscientific Analyses (SAGA GIS). http://www.saga-gis.org/ (last access 20.04.2011).
Seibert, J., Stendahl, J., Sorensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141 (1–2), 139–148.
Smith, M.P., Zhu, A.X., Burt, J.E., Stiles, C., 2006. The effects of DEM resolution and neighborhood size on digital soil survey. Geoderma 137 (1–2), 58–69.
Smith, S., Bulmer, C., Flager, E., Frank, G., Filatow, D., 2010. Digital soil mapping at multiple scales in British Columbia, Canada. Program and Abstracts, 4th Global Workshop on Digital Soil Mapping, 24–26 May 2010, Rome, Italy, p. 17.
Soil Atlas of Europe, 2005. European Soil Bureau Network European Commission. 128 pp. Office for Official Publications of the European Communities, L-2995 Luxembourg. http://eusoils.jrc.ec.europa.eu/projects/soil_atlas/index.html (last access 05.09.2011).
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8 (1), 25.
Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9 (1), 307.
Strobl, C., Malley, J., Tutz, G., 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. Psychological Methods 14 (4), 323–348.
Strobl, C., Hothorn, T., Zeileis, A., 2010. Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package. The R Journal 1 (2), 14–17.
Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random Forests Applied as a Soil Spatial Predictive Model in Arid Utah. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Bridging Reasearch, Environmental Application, and Operation. Springer, pp. 179–190.
Sun, X.L., Zhao, Y.G., Zhang, G.L., Wu, S.C., Man, Y.B., Wong, M.H., 2011. Application of a Digital Soil Mapping Method in Producing Soil Orders on Mountain Areas of Hong Kong Based on Legacy Soil Data. Pedosphere 21 (3), 339–350.
Thwaites, R.N., Slater, B.K., 2000. Soil–landscape resource assessment for plantations - a conceptual framework towards an explicit multi-scale approach. Forest Ecology and Management 138 (1–3), 123–138.
Webster, R., Becket, P.H.T., 1968. Quality and Usefulness of Soil Maps. Nature 219, 680–681.
Wilson, J.P., Gallant, J.C., 2000. Terrain Analysis: Principles and Applications. Wiley, New York.
Wittmann, O., 1991. Standortkundliche Landschaftsgliederung von Bayern. GLA-Fachberichte, 5. Geologisches Landesamt, München.
Yang, L., Jiao, Y., Fahmy, S., Zhu, A.-X., Hann, S., Burt, J.E., Qi, F., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. Soil Science Society of America Journal 75 (3), 1044–1053.
Zevenbergen, L.W., Thorne, C.R., 1987. Quantitaitive Analysis of Land Surface Topography. Earth Surface Processes and Landforms 12, 47–56.