Terrain-related Revision of Existing Soil Maps

M. Möller[1], T. Koschitzki[1], K.-J. Hartmann[2]

[1]Geoflux GbR, Lessingstraße 37, D-06114 Halle (Saale)
Telephone: 00493451352244
Fax: 00493452394019
Email: moeller@geoflux.de

[2]State Institute for Geology and Natural Resources Saxony-Anhalt, Köthener Straße 34, 06118 Halle (Saale)
Telephone: 04493455212117
Email: hartmann@lagb.mw.sachsen-anhalt.de

1. Introduction

In Germany, federal geological surveys are responsible for mesoscale soil mapping.In the federal state Saxony-Anhalt, the soil map 1:50,000 results from an integration process of already existing soil maps (Hartmann 2006). These maps were mainly surveyed in the former East Germany where another classification system was valid. The data integration process was subjectively realized by soil surveyors and aimed primarily at the semantic transformation to the current valid classification system of the German Handbook of Soil Mapping (Ad-hoc-AG Boden 2005). While the former soil unit boundaries were generally taken over, the soil attributes were semantically aggregated. That means that the soil units, which originally represented genetically linked soils, were now described by only the dominant soil.

The resulting soil map does not contain any quality information and is therefore labeled as “preliminary” (*in German*: Vorläufige Bodenkarte 1:50,000 or VBK 50). Thus, a simple quality assessment of VBK 50 should be applied in a reproducible manner by the following general conditions:

- Implementation of expert knowledge should be ensured.
- No training or validation information was available for automatic classification approaches.

In this paper, on the example of VBK 50 we present a cost- and time-effective terrain-related revision of mesoscale soil maps. The revision bases on a state-wide available Digital Elevation Model with a resolution of 20 x 20 m (DEM 20) and focuses on the terrain-related soil properties *floodplain membership*, *colluvium membership* and *humus layer thickness*.

2. Methods

The procedure can be distinguished in five steps: First, soil-related terrain attributes were derived which had been proved to be suitable for the classification of the above mentioned target soil properties (section 2.1). Second, terrain attributes were segmented into landform elements and then geometrically overlaid with aggregated VBK 50 units (section 2.2). Third, the resulting landform soil elements (LSE) were statistically analyzed and classified by fuzzy membership functions (section 2.3). Finally, the classified LSE were semantically and geometrically aggregated (section 2.4) as well as assessed regarding their terrain-related plausibility by means of a quality measure (section 2.5).

2.1 Terrain Analysis

The target soil properties *colluvium membership* and *humus layer thickness* are related to the terrain attribute *mass balance index MBI*. The index is calculated by the combination of the terrain attributes *slope*, *vertical distance to channel network* and *profile curvature* (Fig. 1 e). Negative *MBI* values represent areas of net deposition such as depressions, positive *MBI* values represent areas of net erosion such as hill slopes, *MBI* values close to 0 indicate areas with a balance between erosion and deposition such as plain areas (Fig. 1 a, c; Möller et al. 2008).

The floodplainindex *FPI* enables the detection of floodplains (Fig. 1 b, d). They can be characterized by a maximal value of *topographic wetness index TWI*, low slope values and minimal values of *vertical distance to channel network* (Fig. 1 f).

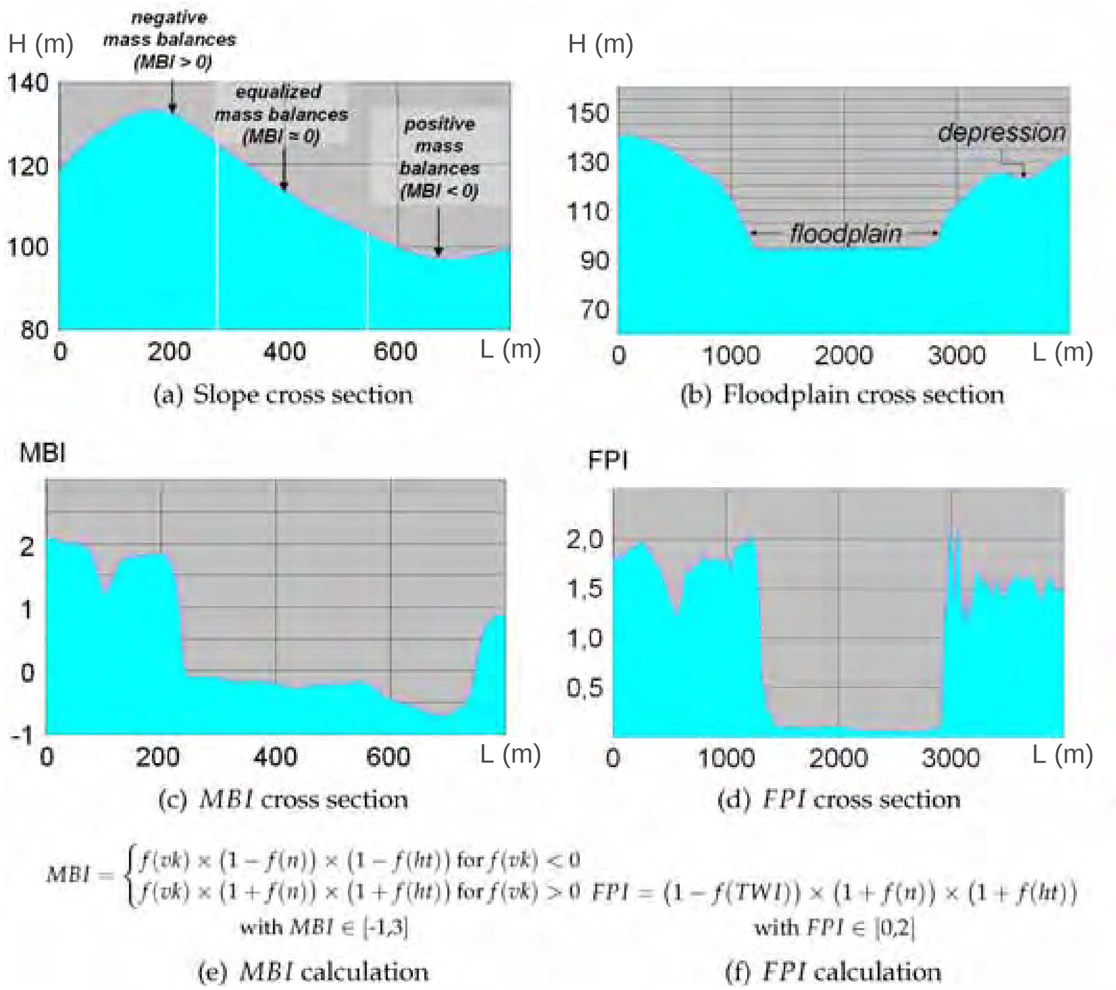


Figure 1. Relations between DEM cross sections (a, b) and value ranges of *MBI* (c) and *FPI* (d) | H, height (m) | L, length of cross section (m) | *vk*, profile curvature | *n*, slope | *ht*, vertical distance to channel network | *TWI*, topographic wetness index

2.2 Segmentation

Segmentation algorithms applied on terrain attributes have been established as an approach for the reproducible delimitation of soil-related landform elements (Möller et al. 2008, Minár and Evans 2008, MacMillan and Shary 2009). In this study, the fractal net evolution algorithm (FNEA) was used which is described in detail by Benz et al. (2004). The FNEA belongs to hierarchical region-growing algorithms starting with raster cell groups (seeds) representing local minima within raster grid and cluster of smallest Euclidean distance within the associated n-dimensional feature space. Those seeds are growing as far as a halting criterion is reached. Halting criteria are defined by the average heterogeneity of resulting segments or landform elements. The segmentation process generates different aggregation levels of discrete landform elements. Each level represents a specific target scale consisting of segments with a comparable heterogeneity. The segmentation results can be influenced by parameters which allow the adaptation of the target segment's heterogeneity and shape.

The overlay of landform elements and aggregated VBK 50 soil units led to landform soil units (LSE). The VBK 50 aggregation was performed semantically (not geometrically). Soil types, representing a horizon-related classification according to soil forming processes, were summarized to soil classes considering similar terrain-related soil forming conditions (Ad-hoc-AG Boden 2005).

2.3 Classification

The classification process corresponds to a data base query applied on LSE mean *MBI* and *FPI* values using fuzzy set theory (Zadeh 1965). The crucial point while defining fuzzy membership functions is the identification of appropriate fuzzy sets. Following Kuo et al. (2009), memberships were derived from *k means* clustering of LSE mean values of each soil class within the R statistics environment (cf. Reimann et al. 2008).

2.4 Assignment und Aggregation

The classification can be considered as a test if LSE terrain properties and the associated soil classes fit together. If so, the original soil unit attributes can be confirmed and taken over. If not, new suitable soil information has to be assigned. Therefore, we created a lookup table containing soil information of all possible spatial neighbors. Applying a Boolean value, all combinations were expert-based judged regarding their plausibility. The actual GIS-based aggregation routine searches in an iterative manner for the best contextual fitting neighbor delivering its soil information whereas every iteration step produces new neighborhood relations. Finally, all classified and assigned LSE are aggregated: On condition that all original VBK 50 boundaries are retained unchanged, all neighbors are merged geometrically and semantically if they have an area smaller than 2.5 ha and belong to the same soil class. This operation should ensure that only scale relevant geometric boundary modifications are taken into account which affect the cartographic presentability for the target scale of 1:50,000.

2.5 Assessment

Fuzzy classification results can be assessed by the best class membership $Z1$ and classification stability $Zrel$. While $Z1$ indicates simply the height of the best class membership (0 = low; 1 = high), $Zrel$ results from the combination of the first and second best class memberships $Z1$ and $Z2$ according to Equation 2.

$$Zrel = \frac{Z1 - Z2}{Z1} \qquad (2)$$

High $Zrel$ values indicate that the first class membership is considerably higher that the second one. The classifications can be considered as *stable*. Low $Zrel$ values stand for *instable* classifications because $Z1$ and $Z2$ values are similar. Finally, a classification can be highly esteemed if $Zrel$ and $Z1$ values are high.

The integral consideration of $Z1$ and $Zrel$ gives a so called plausibility measure PM which is applied separately on confirmed, new classified and not classified LSE.

PM is calculated from the k *means* cluster analysis of $Z1$ and $Zrel$ values, their following cluster related summation $(Z1, mean + Zrel, mean)$ and ranging according to equation (4).

$$PM = \frac{(Z1, mean + Zrel, mean) - (Z1, mean + Zrel, mean)_{\min}}{(Z1, mean + Zrel, mean)_{\max} - (Z1, mean + Zrel, mean)_{\min}} \quad (4)$$

3. Study Area

The approach was applied for the total area of Saxony-Anhalt (20,443 km^2) but we have visualized some results on the example of a study area with heterogeneous soil and relief conditions (Fig. 2). This area corresponds to the official German topographic map 4336 at a scale of 1:25,000 with a size of about 100 km^2 (cf. Möller et al. 2008).

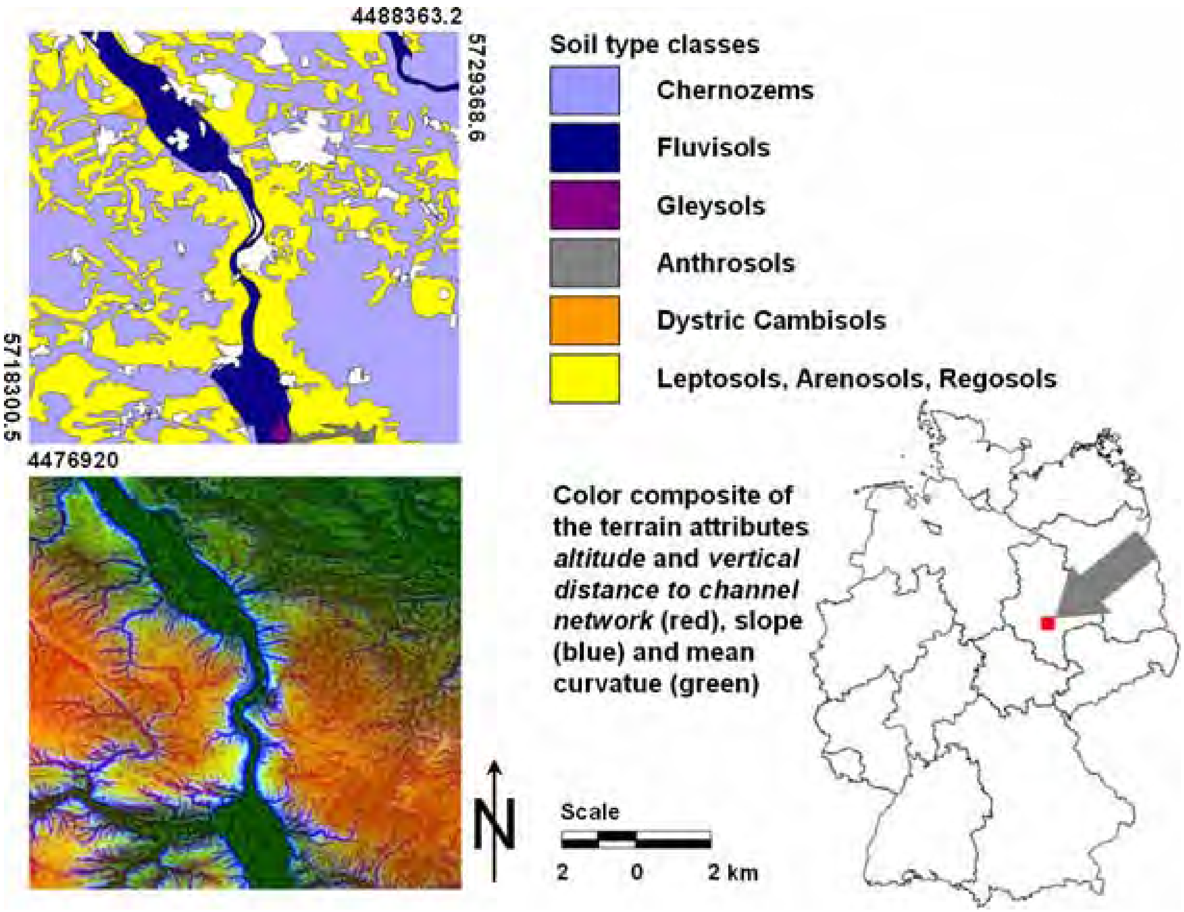


Figure 2. Study area: Soil type classes and color composite of selected terrain attributes.

4. Results

The upper image in Fig. 3 shows the segmentation and classification results on the example of the study area. Related to Saxony-Anhalt, the overlay of segmentation result and VBK 50 led to 469,430 LSE. The original number of soil units was 36,636.

Fig. 4 shows on the example of the soil class *Chernozem* class specific membership functions which result form k *means* cluster analysis. The cluster number was chosen subjectively (here: 10 cluster). The frequency charts of each cluster give a clue of the cluster relevance. The lower and upper cluster quartiles as well as cluster means revealed fuzzy sets for the used membership function types (mft) smaller than and about range (cf. Definiens 2008).

The lower image of Fig. 3 clarifies the effects of the applied aggregation operation: Only scale relevant modifications were made. In Saxony-Anhalt, the aggregation led to a decrease in LSE number from 469,430 to 87,012.

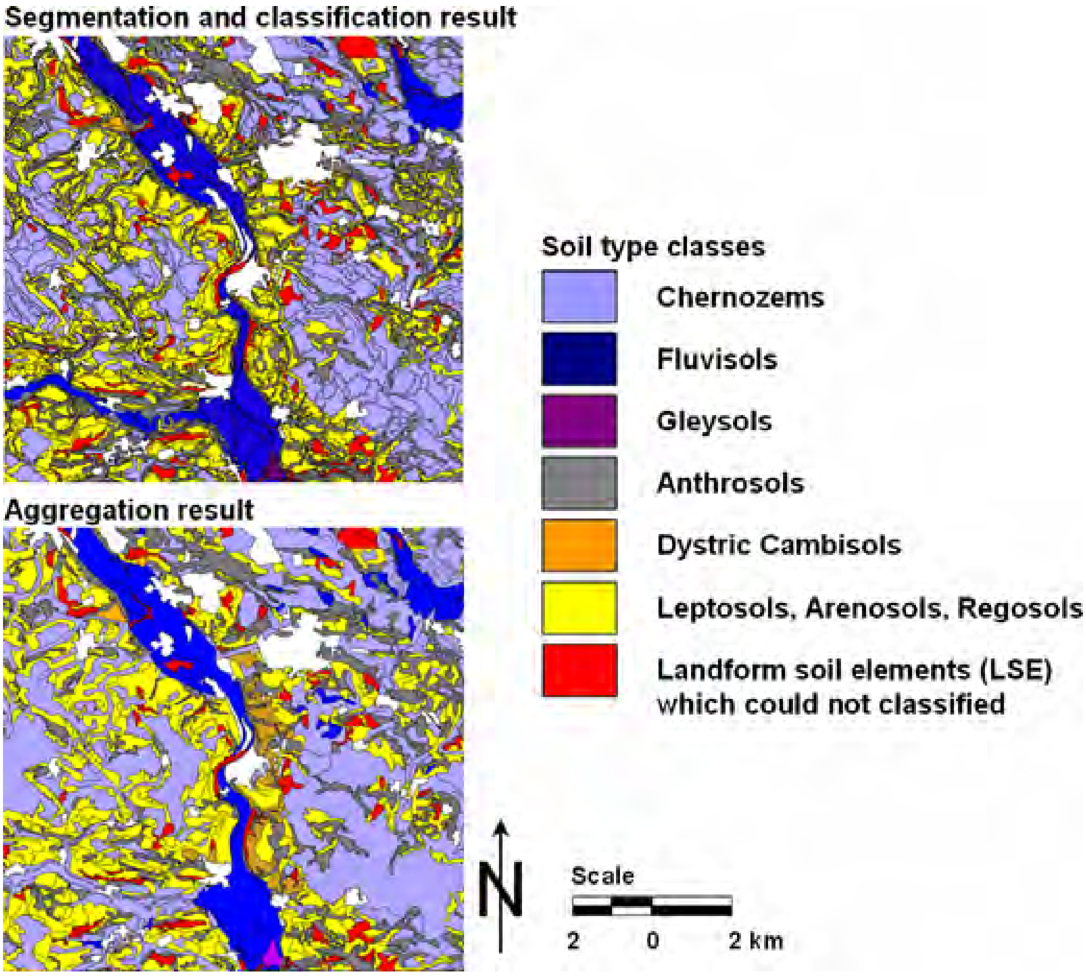
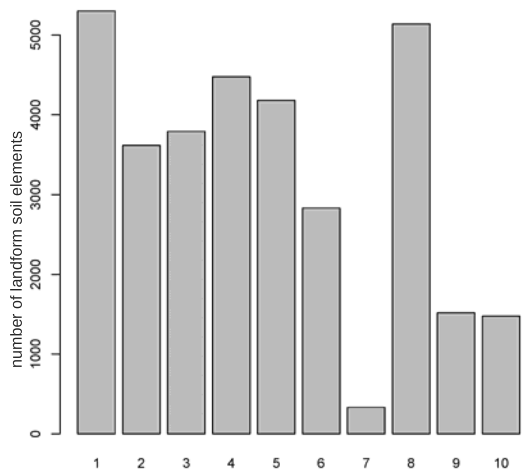


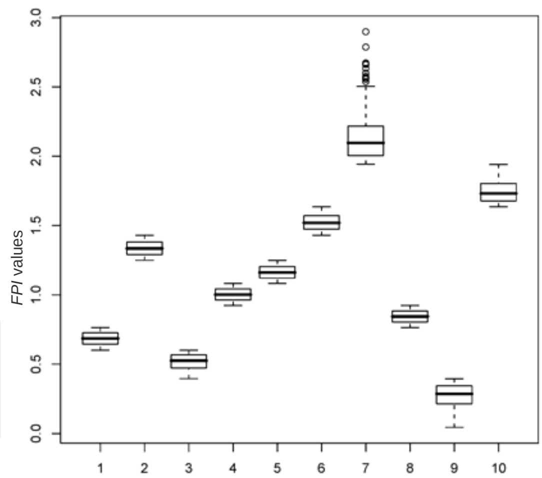
Figure 3. Visualization of segmentation, classification and aggregation results on the example of the study area Könnern.

Fig. 5 uncovers that the proportion of the soil type class *Anthrosol* (Y1) had been increased considerably to the disadvantage of other soil type classes. The information about colluvium's occurrence was hidden in the semantic attributes of the original soil maps which got lost during the semantic transformation process (section 1). Thus, the classification result represents a geometric disaggregation revealing original semantic terrain-related information.

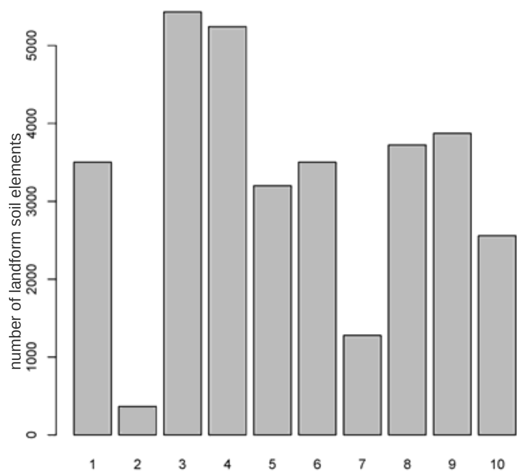
57 % of all LSE could be confirmed or assigned with new soil information (classes A, B-L, D, G, R, S-P, T, Y), 36 % could be classified but not assigned (classes A1, A2, O1, R1, Y1) and 7 % could be neither classified nor assigned (class REST). Finally, Table 1 contains the plausibility measure *PM* for each confirmed or new classified LSE derived from clustered and summarized $Z1$ and $Zrel$ values.



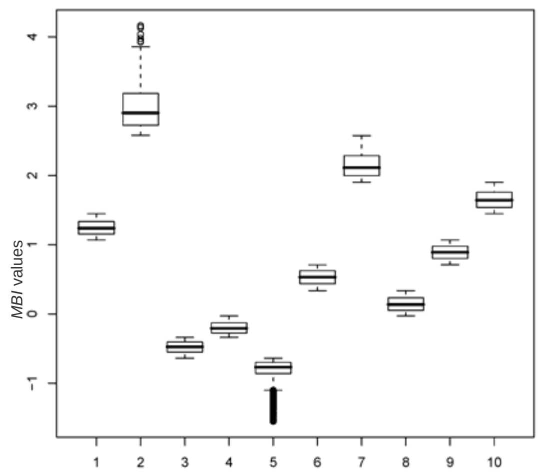
(a) *FPI* cluster barplots



(b) *FPI* cluster boxplots



(c) *MBI* cluster barplots



(d) *MBI* cluster boxplots

- Class *Chernozem*
    - *feature MBI*: *mft* about range *border* 0.1 … 1.3 (used cluster. 4: 1, 6, 8, 9)
- Class *Chernozem to Floodplain*
    - *feature FPI*: *mft* smaller than *border* 0.3 … 0.7 (used cluster: 1, 3, 9)
- Class *Chernozem to Anthrosol*
    - *feature MBI*: *mft* smaller than *border* -0.8 … 0 (used cluster: 3, 4, 5)
- Class *Chernozem to Leptosol, Arenosol or Regosol*
    - *feature MBI*: *mft* smaller than *border* 1.6 … 2.9 (used cluster: 2, 7, 10)

(e) Class definitions

Figure 4. Frequency charts (a, c), Box Whisker plots (b, d) of *FPI* and *MBI* cluster as well as membership functions (e) for the soil class *Chernozem* related to the total area of Saxony-Anhalt.

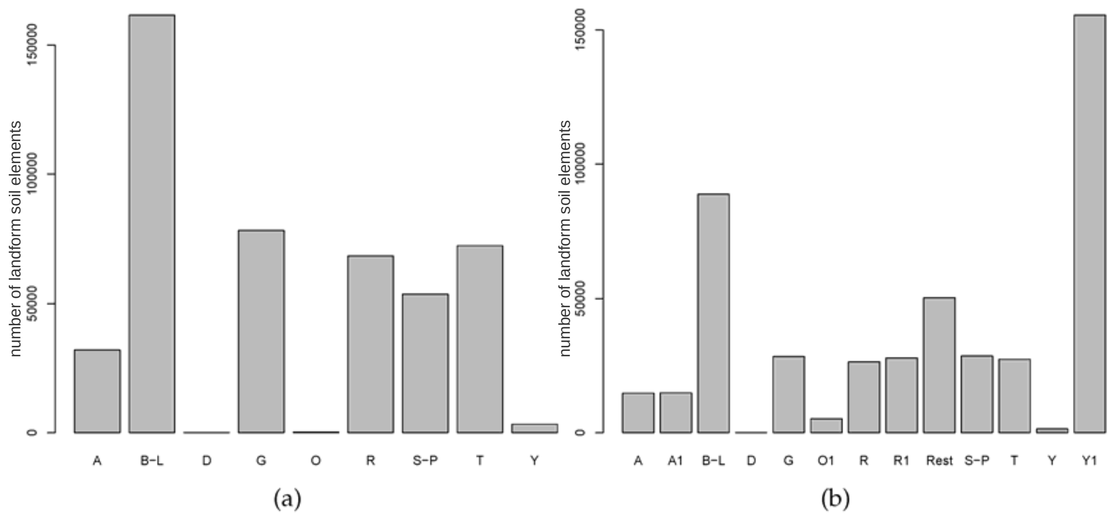


Figure 5. Frequency charts of the original VBK 50 (a) and the modified LSE number (b) (A1, O1, R1 and Y1 = new classified soil classes. REST = LSE could neither classified nor assigned).

$Z1, mean$	$Zrel, mean$	PM
0.01	0.01	0.00
0.88	0.09	0.48
0.34	0.24	0.28
0.86	0.45	0.65
0.47	0.62	0.54
0.90	0.75	0.83
0.62	0.92	0.77
0.29	0.94	0.61
0.85	0.97	0.91
0.99	1.00	1.00

Table 1. Clustered $Z1$ and $Zrel$ mean values and related PM values.

5. Conclusions

We presented an effective algorithm to integrate terrain information into existing mesoscale soil maps. The applied approach bases on the segmentation of terrain attributes into landform elements. On the resulting object data sets a fuzzy classification based on two-dimensional membership functions was carried out. The membership function borders were defined by a preceding *k means* cluster analysis. The classification result was aggregated considering scale and neighborhood relations as well as cartographic readability. All algorithm steps were affected by expert knowledge.

The modified soil units contain additional information to their terrain related plausibility. We are aware that the derived quality measures cannot replace a semantic and geometric validation (cf. MacMillan 2008). However, this approach can help to get an idea about the terrain related accuracy of existing older soil maps which often contain no quality information. Finally, the classification and assessment results can be used for the definition of training areas for automatic classification approaches (cf. Scull et al. 2003, MacMillan 2008) which be subject of further work.

References

Ad-hoc-AG Boden. 2005. *Bodenkundliche Kartieranleitung*. E. Schweizerbart'sche Verlagsbuchhandlung. 5th edition. Hanover. Germany.

Benz U C. Hofmann P. Willhauck G. Lingenfelder I and Heynen M. 2004. Multi-resolution. object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *Journal of Photogrammetry & Remote Sensing*. 58: 239–258.

Definiens. 2008. *Definiens User Guide*. Document Version 7.0.5.968. Definiens AG. München. Germany.

Hartmann K-J. 2006 . Bodeninformationen. In: Feldhaus D and Hartmann K-J (eds). *Bodenbericht Sachsen-Anhalt 2006: Böden und Bodeninformationen in Sachsen-Anhalt*. Landesamt für Geologie und Rohstoffe Sachsen-Anhalt. Halle (Saale). 71–87.

Kuo R J. Chao C M and Liu C Y. 2009. Integration of *k means* algorithm and AprioriSome algorithm for fuzzy sequential pattern mining. *Applied Soft Computing.* 9: 85–92.

Minár J and Evans I S. 2008. Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping. *Geomorphology*. 95: 236–259.

Möller M. Volk M. Friedrich K and 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: 419–430.

MacMillan R A. 2008. Experiences with applied DSM: Protocol. availability. quality and capacity building. In: Hartemink A E. McBratney A and Mendonca-Santos M L (eds). *Digital soil mapping with limited data.* pp. 113–136. Springer. Heidelberg. New York

MacMillan R A and Shary P A. 2009. Landforms and landform elements in geomorphometry. In: Hengl T and Reuter H I (eds). *Geomorphometry: concepts. software. application.* Developments in soil science. volume 33: 227–254. Elsevier: Amsterdam.

Reimann C. Filzmoser P. Garrett R and Dutter R. 2008. *Statistical Data Analysis Explained: Applied Environmental Statistics with R*. Wiley & Sons. Chichester. West Sussex. UK.

Scull P. Franklin J. Chadwick O and McArthur D. 2003. Predictive soil mapping: a review. *Progress in Physical Geography*. 27: 171–197.

Zadeh L. 1965. Fuzzy sets. *Information and Control*. 8: 338–353.