Geoderma xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Geoderma

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

Compilation of a national soil-type map for Hungary by sequential classification methods

László Pásztor[a], Annamária Laborczi[a,*], Zsófia Bakacsi[a], József Szabó[a], Gábor Illés[b]

[a] Department of Environmental Informatics, Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, H-1022 Budapest, Herman Ottó út 15, Hungary

[b] Forest Research Institute, National Agricultural Research and Innovation Centre, H-9600 Sárvár, Várkerület 30/A, Hungary

ARTICLE INFO

Keywords:
Soil type map
Digital soil mapping
Data mining
Multi stage classification
Segmentation

ABSTRACT

Traditionally in Hungary the soil cover under agricultural and forestry management is typically characterized independently and just approximately identically. Soil data collection is carried out and the databases of soil features are managed irrespectively. As a consequence, nationwide soil maps cannot be considered homogeneously predictive for soils of croplands and forests, plains and hilly/mountainous regions. In order to compile a national soil type map with harmonized legend as well as with spatially relatively homogeneous predictive power and accuracy, the authors unified their resources. Soil profile data originating from the two sources (agriculture and forestry) were cleaned up and harmonized according to a common soil type classification. Various methods were tested for the compilation of the target map: segmentation of a synthesized image consisting of the predictor variables, multi stage classification by Classification and Regression Trees, Random Forests and Artificial Neural Networks. Evaluation of the results showed that the object based, multi-level mapping approach performs significantly better than the simple classification techniques. A combination of best performing classifiers, when each classifier's vote on the same object is weighted according to its confidence in the voted class, led to the final product: a unified, national, soil type map with spatially consistent predictive capabilities.

1. Introduction

Land use requirements are becoming more and more complex, and demand for rational land use is increasing (Verheye, 2009). In the meantime, more and more environmental conflicts and risks are emerging (Brauch et al., 2011). These driving forces make spatial and land management planning more and more important, which are expected to result in increasingly reliable plans. Planning, in turn, requires accurate, coherent and quality spatial data (Andrew et al., 2015). The goal of soil mapping is to reveal and visualize the spatial relationships of the thematic knowledge related to soil cover (Brevik et al., 2016). Soil maps are thematic maps, where theme is determined by some specific information related to soils (Miller and Schaetzl, 2014). This can be a primary or secondary (derived) soil property or class as well as any knowledge characterizing functions, processes or services of soils (Minasny et al., 2012). Traditionally, spatial knowledge on soils is mostly summarized in the form of soil type maps based on an appropriate classification system (Brevik et al., 2016). Generally, these maps are simply called soil maps, which in fact reflect their importance. On the other hand, according to Webster (2015) the emphasis on classification may cause certain constraints and the availability of various soil property, and functional soil maps (Hengl et al., 2015; McBratney et al., 2003; Scull et al., 2005) is increasing rapidly. Historically, soil mapping was based on soil typology and soil types have strong didactic significance. Soil type maps have been created on different levels and according to different classification systems. The actual applicability of the recently elaborated system (new national systems, World Reference Base, USDA Soil Taxonomy, Universal Soil Classification, Golden et al., 2010) for mapping greatly depends on the availability of the profile description in the given system due to their inherent differences, and the difficulties in their accurate correlation opportunities (Michéli et al., 2006).

1.1. Soil classification and soil type maps

The Hungarian Soil Classification System is based on the genetic approach of Dokuchaev (1883). It considers soil forming as a genetic process (pedogenesis), in which geographic conditions are substantial (Stefanovits, 1963, 1972; Szabolcs, 1966; Várallyay et al., 1979). In the last few decades, due to the development of soil science and infor-

* Corresponding author.
E-mail addresses: pasztor@rissac.hu (L. Pásztor), laborczi@rissac.hu (A. Laborczi), zsofi@rissac.hu (Z. Bakacsi), james@rissac.hu (J. Szabó), illesg@erti.hu (G. Illés).

http://dx.doi.org/10.1016/j.geoderma.2017.04.018
Received 18 March 2016; Received in revised form 10 January 2017; Accepted 18 April 2017
0016-7061/

matics, as the social and global demands have been changing, the diagnostic approach of soil classification systems has come to the forefront (USDA ST, WRB, Universal Soil Classification). WRB is widely used by Hungarian experts, moreover, the Hungarian soil classification system is nowadays undergoing a modernization process including the adoption of diagnostic categories (Michéli, 2011; Michéli et al., 2014, 2015). A WRB based harmonized digital soil map and database for the Danube Basin of the Danube-region (embracing Hungary) has been compiled very recently by Dobos et al. (2016), whose legend stopped on RSG level and its spatial resolution is 463 m. The applied, slightly modified e-SOTER methodology used automated classification algorithms and soil diagnostic property maps as regionalized qualifiers, which were elaborated based on proper reference data (Dobos et al., 2011, 2013). It did not use classified soil profiles, which is reasonable, since too few surveys were carried out according to the standardized nomenclature of WRB. The most extended, harmonized surveying campaign was made within the framework of the BIOSOIL project (Lacarce et al., 2009). In the BIOSOIL survey the ICP Level I and Level II monitoring plots were used to conduct a uniformly detailed soil survey using the WRB nomenclature and description rules (Hiederer et al., 2011). For Hungary 78 Level I and 4 Level II monitoring plots were assigned. The low number of spatially representative soil observations with sufficient diagnostic description and/or which are classified according to the renewed systems did not reach the level, where a high resolution, nationwide soil type map could be targeted with a legend according either to WRB or the renewed Hungarian system.

Due to the shortage of WRB compatible data, the traditional Hungarian classification system was used in the present mapping process. The system sorts soils into main soil types such as skeletal soils, lithomorphic soils, brown forest soils, chernozems, salt-affected soils, meadow soils, alluvial and deluvial soils, and peat soils. There are some differences between the soil classification used in forests and on arable lands. The soils of agricultural and forest areas have been surveyed independently, the former was carried out by agricultural experts, while the latter was conducted by foresters. Forest classification makes several differences between some soils to fit soil utilization better (Szodfridt, 1993). Forestry classification includes a gravelly skeletal soil with a native forest cover of turkey oak (*Quercus cerris* L.) and pedunculate oak (*Quercus robur* L.) called ‘cseri’ soil. It also differentiates between brown earths on loess and sand. The Hungarian name of the latter (‘rust-red brown forest soil’) indicates Fe-oxides. This type of soil has poorer fertility than classic brown earths. Forestry classification also differentiates between meadow forest soils and alluvial forest soils, which differ in the characteristic humus form. Agricultural types were supplemented with some typical forest soils to cover the whole range of soil types.

Due to the general geographical conditions of Hungary physiography and land use is strongly related (Fig. 1). Great plains, characteristically with fertile soils, are dominated by arable lands while hilly and mountainous regions are characterized by forests. Traditionally, soil cover under agricultural and forestry management is typically characterized independently. Soil maps with forestry origin never went beyond the areas characterized by forest land use. On the other hand, countrywide soil (type) maps were compiled with full national coverage based on soil data originating purely from out-of-forest areas. These national maps characterized forest dominated, hilly/mountainous regions either simply as forest (1:200.000 scale genetic soil map; Jeney and Jassó, 1983; Fig. 2), or with significantly lower thematic and spatial resolution (AGROTOPO; Fig. 3) as opposed to plains with arable lands. Consequently, there has not been a real nationwide soil type map – which is consistent regardless of land use – with harmonized legend and spatially homogeneous predictive power and accuracy for soils of croplands and forests. However, tasks of national spatial planning and basement of agricultural adaptation strategies (AGRAGis, 2016; NAGis, 2016) have increasingly required the availability of such a map product. For the support of these demands the compilation of a unified, national, soil type map with spatially consistent predictive capabilities was targeted by the present work by testing and applying suitable digital mapping approaches.

Four approaches were taken into consideration for the compilation of the targeted map product (Fig. 4).

1. A trial was made for the disaggregation of the above mentioned, national, small scale, legacy soil type maps (Pásztor et al., 2015). On plains the approach performed sufficiently, but the significantly low predictivity of the source maps within forests could not be improved by this technique.
2. The second approach was successfully applied by Dobos et al. (2016) for the compilation of the WRB RSG level, digital soil map of the Danube Basin. But it did not prove to be feasible in the case of soil type level Hungarian classification, since the numerous soil properties, necessary for the classification, are not available in map form. Their compilation would require much more resources than it has been available for the recent “single” mapping.
3. Soil type maps according to the traditional Hungarian soil classification were compiled for the areas of agricultural land across the country at a scale of 1:10.000 in the 1960s, '70s and '80s. Theoretically, their digital processing, harmonization and integration could provide a further opportunity. Nevertheless, there are also some shortcomings. (i) In forestry the delineation of mapping units on large scale maps is not based on pedological boundaries, but soil types are assigned to forest parcels. (ii) These large scale legacy soil maps were not produced comprehensively neither on arable lands nor in forest. (iii) Only a part of them has been digitally processed in the last few decades, consequently their availability has been stressfully limited for the present initiative.
4. Finally, there is the possibility to return to the original survey data by the application of sufficient number and properly classified soil profiles originating from the two branches. Appropriate classification techniques together with high resolution, thematically diverse environmental auxiliary information can be used for the spatial inference of the collected legacy data. Due to the shortcomings and disadvantages of the former three methods, we turned to this approach and elaborated a unified, national soil-type map for Hungary by integrated, object-based and multi stage classification methods.

1.2. Numerical classification in digital soil mapping

Data-mining methods (e.g.: Classification and Regression Trees - CART, Random Forest - RF, Artificial Neural Networks - ANNs) aim at extracting hidden information from a data set, in order to make (either spatial) predictions. In Digital Soil Mapping (DSM), data mining methods can reveal relationships between soil features (properties or classes) to be mapped (dependent variable), and the available environmental data (independent variables) related to the soil-forming factors (Behrens and Scholten, 2006).

CART is a non-parametric, recursive partitioning method with excellent predictive capabilities (Breiman et al., 1984). It is simple to understand and interpret, when both continuous and categorical environmental predictors are available (Henderson et al., 2005; Lawrence et al., 2004). CART models are currently and prevalently applied in DSM in order to compile soil type maps (Giasson et al., 2011; Scull et al., 2005), disaggregated categorical soil maps (Moran and Bui, 2002; Nauman and Thompson, 2014; Pásztor et al., 2013), the prediction of particle size distributions (Greve et al., 2012), or geographic distribution of hydromorphic organic landscapes (Bou Kheir et al., 2010).

RF (Breiman and Cutler, 2009) is based on the CART method, growing many classification trees. For each tree, the training data set is randomly split to a subset, which grows the tree, with the remaining data serving for testing or validation. Randomly selected predictor