ARTICLE IN PRESS

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

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

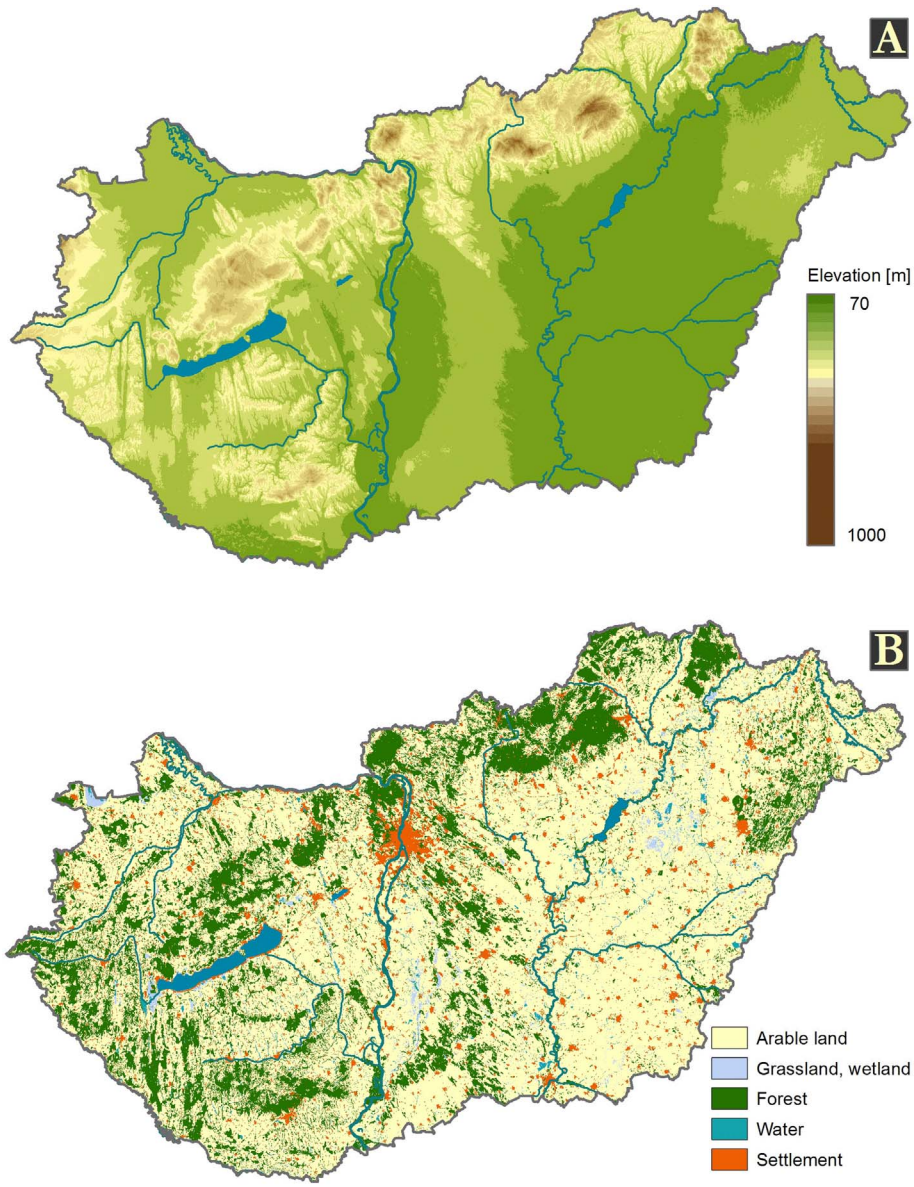


**Fig. 1.** Hungary's general geographical (A) and the related land use (B) conditions. Plains are dominated by arable lands; mountainous regions are mainly characterized by forests.

variables are chosen at each split, and the strongest variable is selected to split the data from this random subset. RF make lots of weak, independent trees, therefore it discerns patterns that otherwise may be disregarded in the cases of few strong trees (Stum et al., 2010). RF is a relatively new method in DSM, it was used to predict topsoil texture classes (Hitziger and Ließ, 2014), soil parent material (Heung et al., 2014), soil organic matter (Wiesmeier et al., 2011) as well as soil types (Brungard et al., 2015; Láng et al., 2016; Stum et al., 2010). Hengl et al. (2015) generated numerous soil property predictions (organic carbon, pH, sand, silt and clay fractions, bulk density, cation-exchange capacity, total nitrogen, and exchangeable acidity, Al content and exchangeable bases) of Africa in 250 m resolution by RF.

ANNs are standard techniques inspired by data processing in biological nervous systems, where different kinds of cells aim at receiving, storing and forwarding information, as well as the outward release of it. ANNs, as supervised learning algorithms require information to extract knowledge that can be used for a subsequent prediction (Zell, 1994). ANNs are current methods in DSM, e.g. predicting soil physical properties (Chang and Islam, 2000; Pachepsky et al., 1996), soil chemical properties (Amini et al., 2005; Patel et al., 2002; Saffari et al., 2009), yield prediction (Dai et al., 2011; Kaul et al., 2005), and soil erosion (Kim and Gilley, 2008). Behrens et al. (2005) mapped soil units in a German sample area and found the predictive power of ANNs considerably high. Bagheri Bodaghabadi et al. (2015) examined ANNs and DEM attributes to predict soil classes.

## 2. Materials and methods

For the mapping procedure a harmonized soil dataset consisting of the detailed description of almost 60,000 soil profiles, describing 41 representative soil-types with spatial reference and a corresponding dataset of 32 spatially exhaustive, ancillary, environmental variables – including legacy soil data – was established covering the whole area of the country.

### 2.1. Soil profile data for agricultural land

Two independent datasets were involved in our study. The Hungarian Soil Information and Monitoring System (SIMS) consists of 1234 observation locations, which have been selected to represent physiographical-soil-ecological units. SIMS contains detailed and up-to-date quantitative soil information about physical and chemical proper-

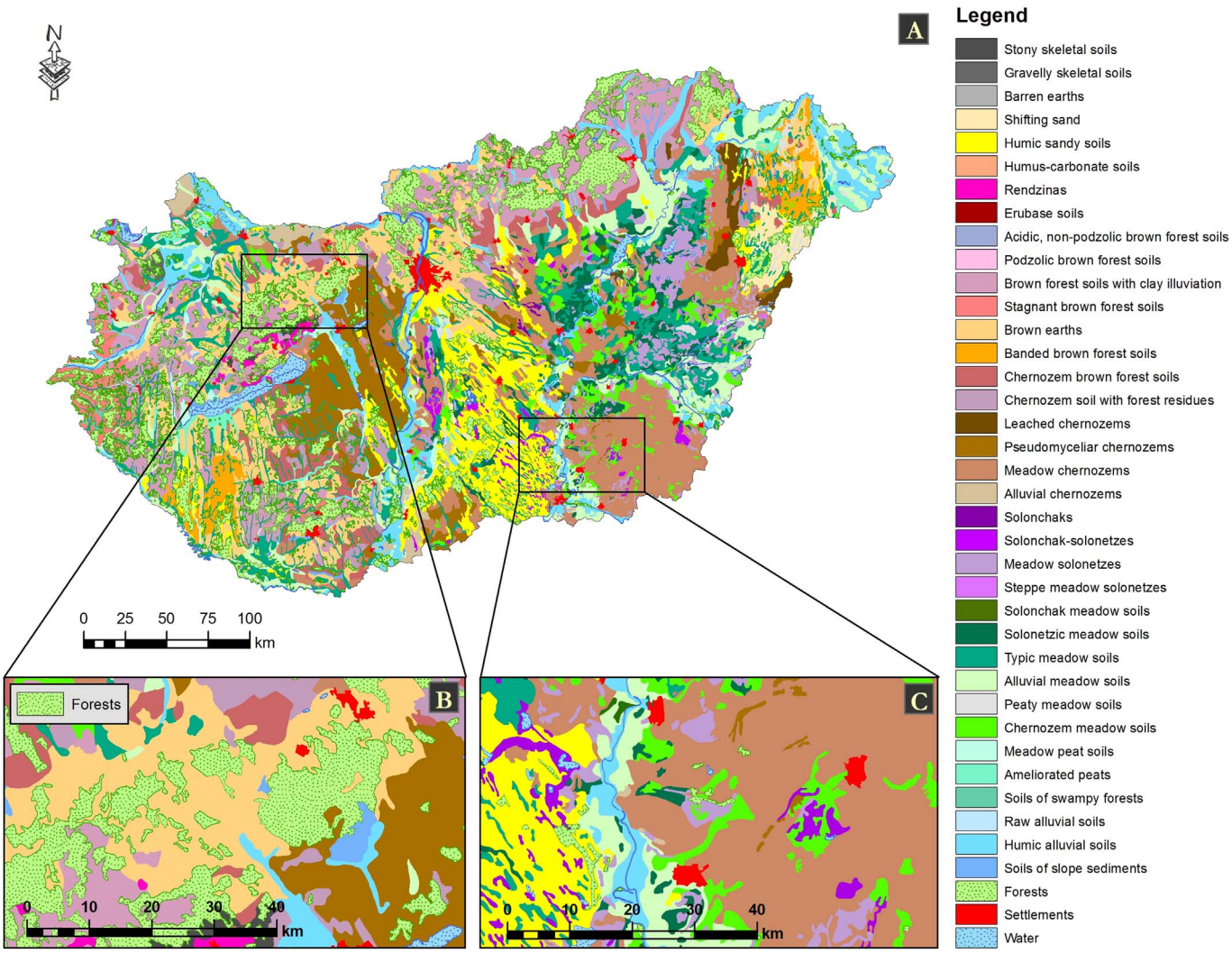


**Fig. 2.** Nationwide legacy soil type maps: 1:200.000 scale genetic soil map (A). Forest is a unique legend element (B); the map was synthetized from larger scale soil maps, which did not represent the country homogeneously, this fact is strongly reflected in its pattern (C).

ties on layer level (Várallyay, 2002). It is considered as a reference dataset which was also heavily relied on during the renewal of the Hungarian Soil Classification System (Michéli et al., 2015).

The Hungarian Detailed Soil Hydrophysical Database (MARTHA; Makó et al., 2010) contains harmonized soil hydrophysical and chemical information collected from various sources. In MARTHA, the soil information is available for 3937 profiles, which can be considered representative for agricultural areas.

### 2.2. Soil profile data for forest land

In Hungary soil data collection in forests is bound most commonly to the procedure of the preparation of management plans. In the past, before the 1990s, forest site data were collected within the framework of site mapping campaigns, however, they had an uneven spatial and temporal distribution, and the preparation of countrywide forest site maps remained unfinished. In the case of forest planning the forest surveying experts follow the instructions described in the "Guidelines for forest management planning" (State Forest Service — ÁESZ, 2004). During the field survey each compartment should have at least one soil profile, however, in cases when more than one typical site types are probably present – on the basis of vegetation differences – each of them has to be sampled by at least one profile. If necessary, between profiles the survey points can be multiplied by core sampling. Besides the in situ examination of profiles, samples are taken from genetic horizons for laboratory analysis. The workflow of in situ and laboratory analysis are carried out according to the national standards (MSZ, 1978a, 1978b). The applied forestry database consists of app. 55,000 data points of forest compartments that were subject to site and soil surveys in the past (profiles with laboratory analysis, profiles with in situ description, core samplings with in situ description). The database contains the following information on the soil profiles: genetic soil type, texture class and rooting depth class.

### 2.3. Harmonization of soil profile datasets

The majority of the soil types in the datasets with differing origins are identical or very similar. However, we had to make some correlation to elaborate a harmonized platform, because of some differences between the two systems as it was mentioned in the Introduction. The Hungarian soil classification system served as a common base for the correlation, where type and sub-type levels are used differently in the two systems. The Hungarian soil classification is a genetic system, based on the appearance and strength of the soil formation processes in the profile. However, the attitude of the descriptions was slightly different, the surveyors tried to involve more details in the description according to the soil's actual usage, even on subtype level in some places. In agricultural areas the properties of management-associated near-surface levels were highlighted (organic matter status, salt quality, etc.). In the forest areas, where the water supply of the tree root zone is

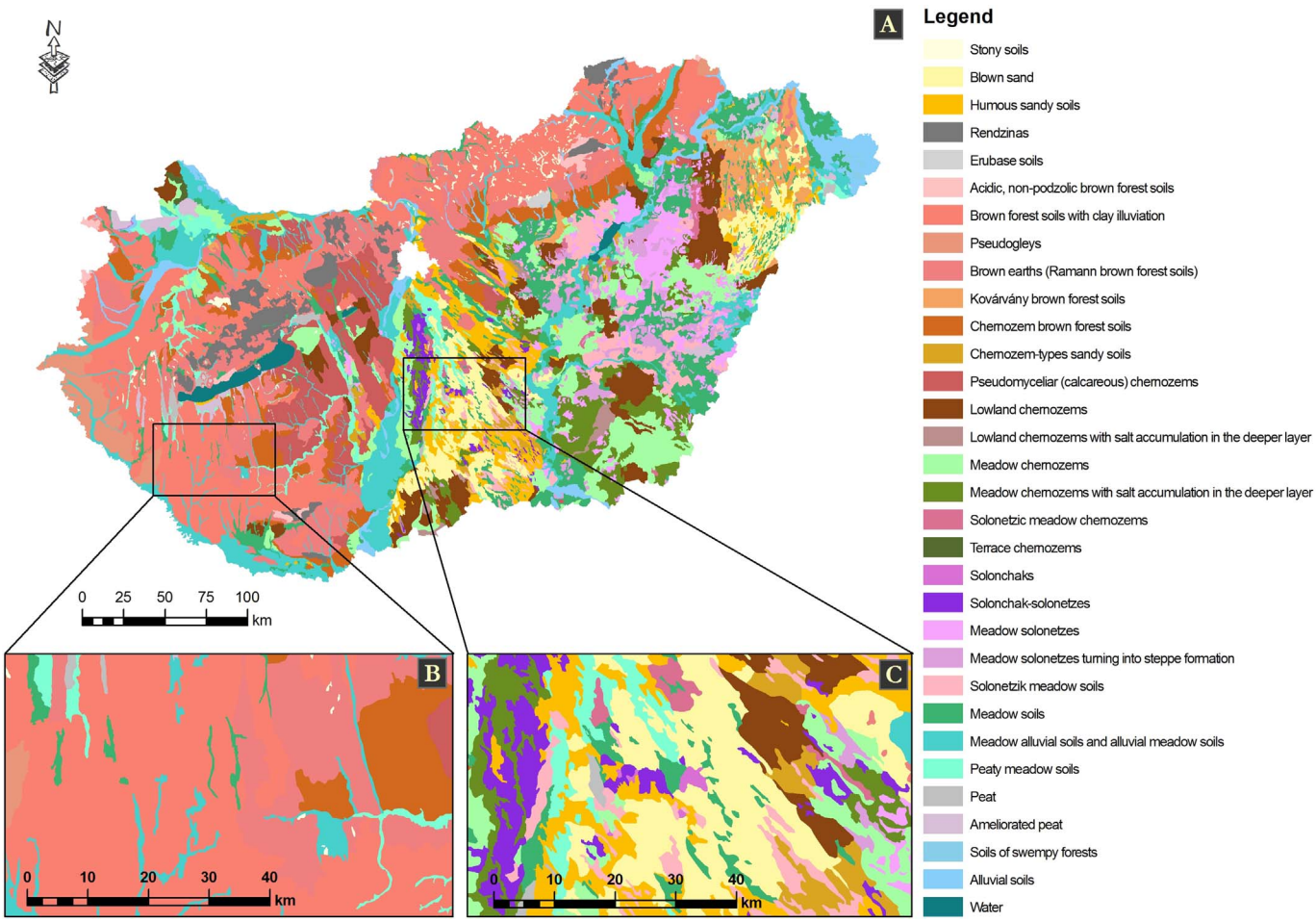


**Fig. 3.** Nationwide legacy soil type maps: Genetic soil type layer of AGROTOPO (A). The representation of areas with different physiography/land use is rather inhomogeneous: hilly regions dominated by forests (B); plains characterized by croplands (C).

essential, the description has been focused on this part. As a consequence, some "local" soil types were also described concerning the physical-chemical properties of the deeper layers. We found such forest soil-variants typically in transient zones, e.g. between the forest and meadow soils, or forest and alluvial soils.

Specifically, the so-called 'cseri' soil was considered as gravelly skeletal soil, due to its characteristics. Meadow and alluvial forest soils were classified into soils of swampy forests. The brown earths on sand have been distinguished from brown earths, even though they are massed together into the same type in the agricultural system. Table 1 lists the harmonized soil types with their main soil type group classification and their population in the two types of data sources. The harmonized soil dataset contains the data of almost 60,000 soil profiles with spatial reference, describing 41 soil types. The majority of the final classes are pure, only 4 of them are aggregated from multiple soil types.

### 2.4. Environmental co-variables

A corresponding dataset of 32 spatially exhaustive, ancillary, environmental variables – including legacy soil data– was established covering the whole area of the country.

**Topography** was based on the EU-DEM (2015), which is one of the most detailed, freely available DEM data sources. Its elevations were captured at 1 arc sec postings (2.78E-4 degrees), the tiles are provided at 25 m resolution. In addition to elevation, further terrain features were calculated from the DEM within SAGA GIS (Conrad et al., 2015). Beside Altitude, its following derivatives were used: Aspect, Channel Network Base Level, Diurnal Anisotropic Heating, Morphometric Features, LS-Factor, Mass Balance Index, Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Index of Ridge Top Flatness (MRRTF), General-, Plan-, and Profile Curvature, SAGA Wetness Index, Slope, Stream Power Index, Real Surface Area, Topographic Position Index, Topographic Wetness Index, Vertical Distance to Channel Network, Distance to Actual Stream Network.

**Lithology** was derived from the Geological Map of Hungary 1:100.000 (Gyalog and Síkhegyi, 2005). In order to simplify the large number of lithology and facies categories, units were correlated with the nomenclature of parent material defined in the FAO Guidelines for soil description (Bakacsi et al., 2014; FAO, 2006).

The level of **groundwater** was taken from the Geological Atlas of Hungary (Pentelényi and Scharek, 2006). The polygon based map displays rather broad interval categories, not continuous depth.

The **climatic** properties of the country were represented by four parameters. Spatial layers of average annual evapotranspiration, average annual precipitation, average annual temperature, and annual evaporation were interpolated, applying the MISH (Meteorological Interpolation based on Surface Homogenized Data Basis, Szentimrey and Bihari, 2007) method for gridding hourly station data. It was developed at the Hungarian Meteorological Service specifically for the interpolation of meteorological data, and is based on the idea that the highest quality interpolation formula can be obtained when certain statistical parameters are known. These parameters are derived by modelling, using long term homogenized data of neighbouring stations.

**Land use** was taken from the CORINE Land Cover 1:50.000 (CLC50; Büttner et al., 2004). CLC50 is a national land cover database

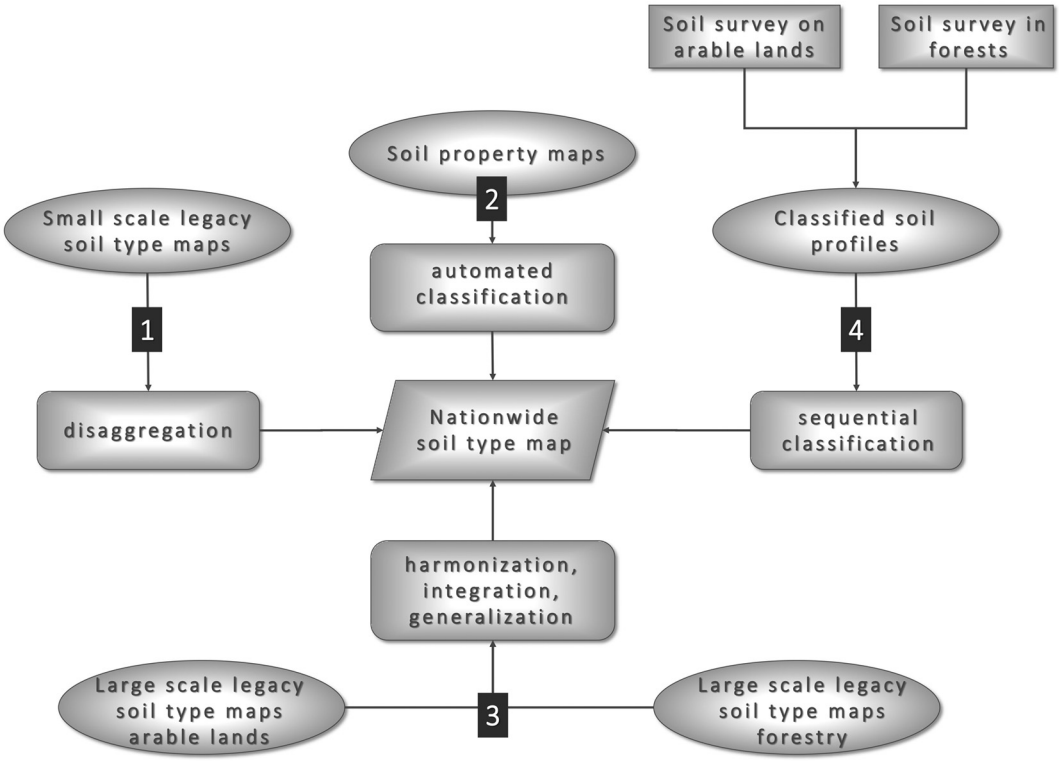


**Fig. 4.** Four approaches for the compilation of the nationwide soil type map: disaggregation of small scale legacy soil type maps (1). Based upon sufficient quantity of relevant soil property maps and their automated classification (2). Digital processing, harmonization and integration of large scale soil maps originating from various sources (3). Sequential classification of unified, classified soil profiles (4).

elaborated on the basis of the CORINE nomenclature of the European Environment Agency (EEA), and adapted to fit the characteristics of Hungary. In order to stratify regions with different land cover, merged categories of CLC50 were used. In the merge we considered that the different soil forming factors and soil types are generally reflected in the land use. The main objective of this approach is to improve the predictive applicability of remotely sensed information (described in the next paragraph).

Earth Observation imagery provides temporally versatile spatial information on the state of **vegetation**, which is strongly correlated with soil properties (Dokuchaev, 1899; Jenny, 1941; Shantz, 1911). MODIS images from two dates (16.03.2012 and 07.09.2013) representing different phases and states of vegetation were chosen for mapping. Red, near-infrared (NIR) bands as well as Normalized Difference Vegetation Index (NDVI) were used from both dates (MODIS 09 products). Furthermore, two NDVI products (MOD13Q1) were involved, which provide information from 16 day periods (03.2012 and 09.2013). Spatial resolution of the images is 250 m (NASA LP DAAC, 2015).

The use of **legacy soil data** supports the applicability of DSM and improves the accuracy of DSM products (Pásztor et al., 2013, 2016). In the present work we also turned to spatial soil information systems, which were elaborated based on legacy soil maps. The Digital Kreybig Soil Information System (DKSIS, Pásztor et al., 2012) is the most detailed spatial dataset related to soils, and covers the whole country. The scale of the original legacy maps is 1:50.000. The categories of the DKSIS physical soil property layer are attributed according to water retention capability, permeability and infiltration rate of soils. The chemical property categories are based on pH, $CaCO_3$ and salt content. AGROTOPO is a national spatial soil information system originally compiled and displayed on 1:100.000 scale topographic map sheets. It consists of ~3.500 soil mapping units (SMUs as polygons) which are characterized by 9 basic soil parameters from which we used genetic soil type. Furthermore, we used the digital version of the genetic soil map produced by MÉM NAK (the predecessor of the National Food Chain Safety Office). The map is thematically the most detailed nationwide representation of the Hungarian soil classification system, comprising the major soil groups, types and subtypes (Kocsis et al., 2015).The basic, but not well documented soil mapping knowledge applied during the process of their compilation was intended to be utilized for the elaboration of the new map. DKSIS represented the large scale spatial delineation of the soil cover, while the two, synthetized soil type maps contributed the regionalization of the genetic features.

However, the applied classification techniques do not require the normality of the variables, the continuous environmental variables with non-normal marginal distribution were transformed using square root and logarithmic transformation, since the same ancillary dataset was further used in some (here not presented) digital soil mapping methods. Nonetheless, several environmental variables remained, where these transformations could not be carried out due to their intrinsic properties. In those cases, we suitably categorized them generating factor variables. For example, Real Surface Area was categorized into 4 classes (10000–10002/10002–10008/10008–10045/10045–12316); Aspect was converted into Eastness and Northness, which were categorized from − 10 till 10 by 0,2 ranges; MRVBF and MRRTF were rounded to integer values. In order to harmonize the (moderately) different spatial resolution of the predictor variables, we resampled them into a common 100 m grid system (in SAGA GIS), which also defines the spatial resolution of the result maps.

### 2.5. Segmentation according to environmental co-variables

Image segmentation is a widely applied method in image analysis. It is a process, which partitionates a digital image into multiple segments, that is to sets of connected pixels. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse (Benz et al., 2004). The purpose of the introduction of object based classification was to

**Table 1**
Harmonized soil types with their main soil type group classification and their population in the two types of data sources.

| Main soil type group | Type code | Unified soil type name | # of observations | | # of aggregated soil types |
|---|---|---|---|---|---|
| | | | Arable land | Forests | |
| Skeletal soils | 110 | Stony skeletal soils | 15 | 167 | 1 |
| | 120 | Gravelly skeletal soils | 7 | 364 | 2 |
| | 130 | Barren earths | 26 | 555 | 1 |
| | 930 | Soils of slope sediments | 145 | 603 | 2 |
| Sand soils | 140 | Shifting sand | 73 | 175 | 1 |
| | 150 | Humic sandy soils | 107 | 18,092 | 1 |
| | 460 | Brown earths on sand | 139 | 5970 | 1 |
| | 470 | Banded brown forest soils | 63 | 2220 | 1 |
| Lithomorphic soils | 310 | Humus-carbonate soils | 0 | 339 | 1 |
| | 320 | Rendzinas | 22 | 1197 | 1 |
| | 330 | Erubase soils | 1 | 21 | 1 |
| | 340 | Rankers | 0 | 486 | 1 |
| Brown forest soils | 410 | Acidic, non-podzolic brown forest soils | 0 | 556 | 1 |
| | 420 | Podzolic brown forest soils | 10 | 333 | 1 |
| | 430 | Brown forest soils with clay illuviation | 289 | 4926 | 1 |
| | 440 | Stagnant brown forest soils | 204 | 1070 | 1 |
| | 450 | Brown earths | 127 | 4519 | 1 |
| | 490 | Brown forest soils with carbonate residues | 8 | 517 | 1 |
| Chernozems | 480 | Chernozem brown forest soils | 172 | 222 | 1 |
| | 510 | Leached chernozems | 21 | 61 | 1 |
| | 520 | Pseudomyceliar chernozems | 312 | 875 | 2 |
| | 530 | Meadow chernozems | 406 | 506 | 1 |
| | 540 | Alluvial chernozems | 38 | 96 | 1 |
| Salt-affected soils | 610 | Solonchaks | 6 | 6 | 1 |
| | 620 | Solonchak-solonetzes | 9 | 5 | 1 |
| | 630 | Meadow solonetzes | 72 | 61 | 1 |
| | 640 | Steppe meadow solonetzes | 23 | 66 | 1 |
| | 650 | Human-induced salt-affected soils | 0 | 4 | 1 |
| | 730 | Solonchak meadow soils | 9 | 34 | 1 |
| | 740 | Solonetzic meadow soils | 52 | 106 | 1 |
| Meadow soils | 710 | Typic meadow soils | 528 | 4626 | 1 |
| | 713 | Meadow soils, salt accumulation in deeper layers | 0 | 109 | 1 |
| | 750 | Alluvial meadow soils | 289 | 815 | 1 |
| | 760 | Peaty meadow soils | 74 | 396 | 1 |
| | 770 | Chernozem meadow soils | 172 | 9 | 1 |
| Peat soils | 810 | Sphagnum peats | 1 | 1 | 1 |
| | 820 | Meadow peat soils | 9 | 948 | 1 |
| | 825 | Ameliorated peats | 6 | 0 | 1 |
| | 910 | Soils of swampy forests | 0 | 1532 | 2 |
| Alluvial soils | 210 | Raw alluvial soils | 12 | 254 | 1 |
| | 220 | Humic alluvial soils | 393 | 1878 | 1 |

delineate areas composed of a set of similar locations (represented by pixels) featured by the applied environmental co-variables, which are different from surrounding areas. The assumption is that these areas are exposed to similar soil forming processes and consequently can be considered as individual soil bodies. From this method we expected (i) on the one hand, faster classifications, since instead of all the pixels, much fewer image objects should be handled, and (ii) on the other hand, we expected more rational classification results, since environmental structure elements (landscape elements) are expressed more strongly in this way than in a pixel based approach.

In the present case instead of a real, remotely sensed multispectral image, a synthetically compiled image, composed of environmental co-variables was subjected to segmentation. We synthesized a geo-referenced TIFF image consisting of the predictor variables as image bands. It was loaded into the eCognition Developer as synthetic image data. A sequence of multi-resolution segmentations was applied onto the “image layers” to delineate homogeneous spatial entities that were used later as objects for classifications based on selected homogeneity criteria, which are a combination of colour (spectral values) and shape properties. The shape criterion was set to 0.3, the compactness criterion to 0.5. The shape and compactness parameters of multi-resolution segmentation algorithm were selected by trial and error. Firstly, we set a 0.1 lag to modify the value of compactness and shape parameter in each step. Secondly, we made a full set of segmented images of a selected part of the country, equally consisting of lowlands and mountainous areas. During the segmentation steps we set the shape parameter to 0.1, then we made segmented images running through the compactness parameter from 0.1 to 1. We continued till the shape parameter reached 1 and vice versa. This way what happened could be followed with segments of environmental co-variables using different parametrization. We found that changing shape and compactness parameters did not cause any major differences in the resulted segmentation patterns. In our opinion it means that the selected environmental co-variables are robust enough in the definition of homogenous areas. Finally, values representing the middle ranges of both parameters were selected.

We also applied different scales for segmentation in order to find the best result for the required spatial resolution. The altering of segmentation scales (Fig. 5) corresponds to different map scales resulting in perfect topology of image objects allowing the reasonable aggregation (upscaling) and disaggregation (downscaling) of soil bodies. A further practical, useful consequence of the segmentation is the significant reduction of the elements used in the classification algorithms, which notably accelerates the computations as opposed to pixel based calculations. Reference data were added as thematic layer. Image segments containing learning sample soil data points served as learning image objects to train the applied image classifiers.

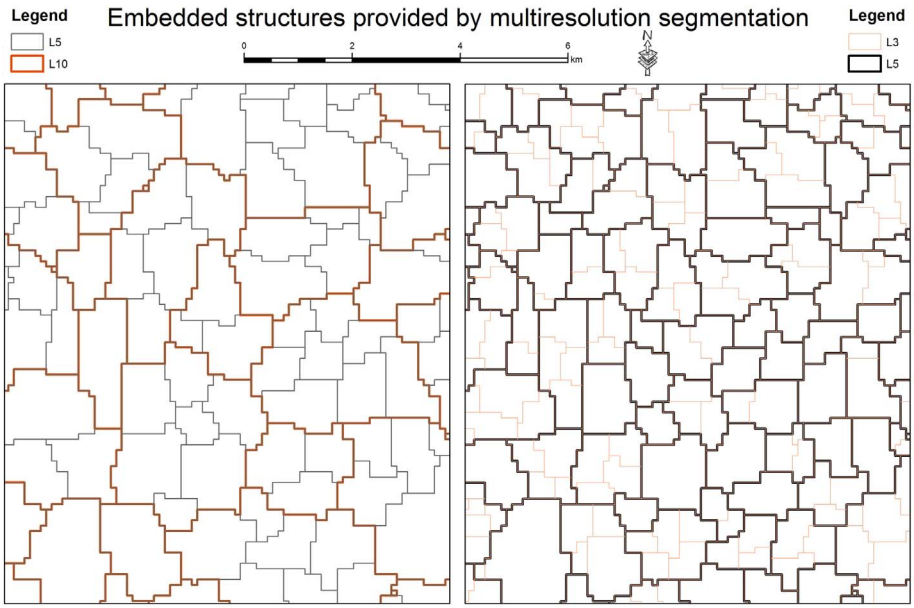


**Fig. 5.** Segmentation results at different scales.

### 2.6. Two-level, multi-step, sequential classification

A sequence of classification processes was applied to establish rule sets to determine soil type instances in the environmental space. Various methods (Classification trees, Random forest, and Neural Network classifiers) were used on two levels and in several, distinct steps.

On the first level main soil type groups (MSTGs) were classified and predicted in multiple steps. Soil types (ST) were targeted only in a second phase within the areas formerly attributed with their respective (containing) MSTG.

We turned to this approximation, since the direct, one-level classification failed to produce results, which could properly represent the soil cover of the country. The uneven spatial distribution of soil types within the country and their unbalanced representation in the unified reference dataset resulted in predictions, which did not even pass mental validation, the Hungarian soil characteristics were not reflected by the results.

This first level (MSTG) classification was carried out in multiple steps. Overall classification resulted acceptable delineation only for selected MSTGs (sand, brown forest and lithomorphic soils). So in the first step areas assigned to any of these three MSTGs were retained, in the next step, classification was carried out only on the excluded areas. Reference profiles belonging to the left out MSTGs were classified again to get the spatial distribution of the other six MSTGs (skeletal, salt-affected, meadow, peat, alluvial soils and chernozems).

As the map of the MSTGs was compiled (Fig. 6), we divided all of the points according to the 9 MSTGs, thus 9 training datasets were formed and 9 soil type maps were compiled for the different, spatially complementary territories of the MSTGs. Salt-affected soils were conspicuously under-represented, therefore they were classified again collectively with meadow soils. This result was added to the other salt-affected soils. Finally, the 9 spatially complementary layers were mosaicked, and the nationwide ST map was compiled.

On both levels, in each step multiple classification models were applied (Table 2). Six models use CART, five are based on RF and one is based on ANN classification. The models with identical classification tools differ either in the inherent parameters of the method or on the segmentation level, on which they are applied. Segmentation was made under different scale factors (L) ranging from 100 to 3 (100, 50, 30, 25, 20, 15, 10, 3). Models (M1-M12) were run at each segmentation level and those of the best performing pairs were selected.

### 2.7. Validation

Two types of validation were carried out. On the one hand, profiles were split into learning and test sets, 20% of the profiles were left out for validation. The validation provided by the test sets was used for the estimation of classification accuracy, which was carried out on both (MSGT and ST) levels and in all steps. The best performing classifier was identified in each phase for each soil type. The 12 models rivalled in the final categorization of segments; the class predicted by the best performer was assigned.

Besides the former data driven validation a trial was done for a certain external validation. A set of digitally processed, large scale legacy soil type maps (presented in the 3rd approach listed in Section 1.1) were also available for a non-systematic comparison (Fig. 7). The predicted raster map was compared to the legacy soil type maps on a pixel by pixel level, which were rasterized with 1 ha resolution to the applied grid system. Since these type of legacy maps were not compiled for forest areas, they cannot say anything about the recent ST estimations predicted for these areas. For forest area validation a collection of independent soil observations was set up. They originate from a variety of sources: monographies, studies, expert's report, reports on forest reservations and forest history etc. The 3253 plots are scattered in 14 out of Hungary's 19 counties and cover all the represented soil types except for salt affected soils.

## 3. Results and discussion

### 3.1. Evaluation of the newly compiled nationwide soil type map

The main final product of our work is a newly compiled nationwide soil type map with harmonized legend and spatially consistent predictivity, which is shown in Fig. 8. A quick visual interpretation reveals some of its advantages. Both the thematic and spatial representation of hilly/mountainous areas is much more detailed than on former national soil maps. Nevertheless, the mosaic-like pattern of lowlands is retained and the large scale geographical structural elements are very well reflected: large sandy soil areas can be identified in the region between the Danube and Tisza rivers and in the Eastern part of the country; the salt affected, meadow and chernozem soils dominating areas of the Great Hungarian Plain are also well identifiable. The extended areas represented by seemingly homogeneous soil cover on former maps are resolved according to their within pedological heterogeneity.

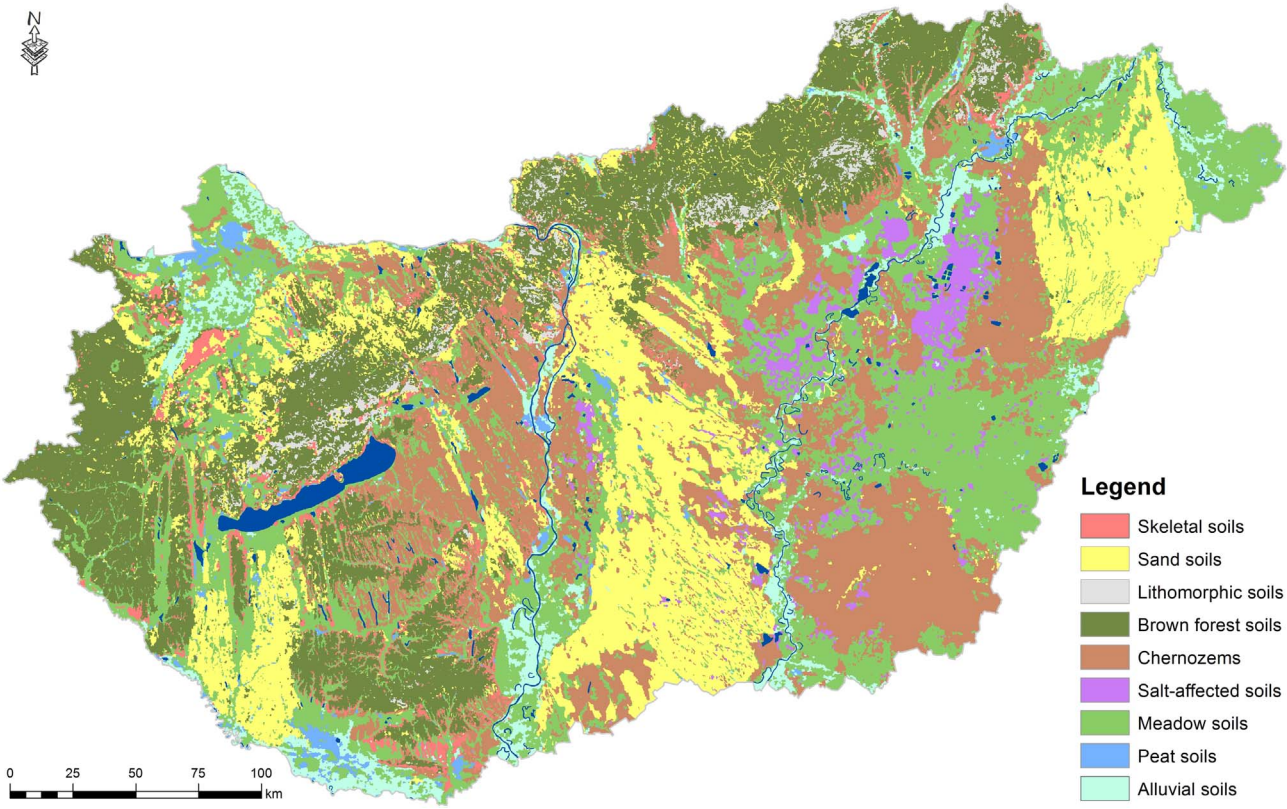


**Fig. 6.** Nationwide MSTG map mosaicked from partial results.

The newly compiled product was compared to the two earlier nationwide soil type maps on a pixel by pixel level. For this purpose, they were rasterized with 1 ha resolution to the applied grid system. The similarity was measured by overall accuracy and overall kappa (Cohen, 1960; Rossiter, 2014). Both measures showed that the maps are rather dissimilar, that is in spite of the overall resemblance, locally they contain divergent local predictions. To see the generality of this behaviour we tested the three main physiographically separable land types (Table 3). According to the results, the greatest differences between the new and the old maps can be found in the mountainous areas, dominated by forest. The level of dissimilarity is smaller in the case of AGROTOPO, which also assigned soil types to the forest dominated, hilly/mountainous regions (even with significantly lower thematic and spatial resolution), as opposed to the 1:200.000 scale genetic soil map, which characterized these regions simply as forests. On the other hand, on lowlands the new maps resemble more this latter, which fact can be attributed to some common elements in their production.

To elucidate these findings, the content and elaboration of the two pre-existing national soil maps should be briefly summarized. Both AGROTOPO and the 1:200.000 scale genetic soil map is a synthetized map, but with different origin and compilation rules. The agro-ecological mapping units of AGROTOPO were delineated for the support of the "Assessment of the agro-ecological potential of Hungary" (Láng, 1983). Creation of soil polygons were mainly based on the map series produced by the national Kreybig soil survey (Kreybig, 1937; Pásztor et al., 2010). Seven soil features (in addition to soil type parent material, soil reaction and carbonate status, soil texture, hydrophysical properties, organic matter resources and soil depth) were assigned to the polygons based on the Kreybig maps and further source materials (Várallyay et al., 1985). Soil type assignment

**Table 2**
Summary table of the applied models. Segmentation was made using different scale factors (L). Models (M1-M12) were run at each segmentation level and those of the best performing pairs are presented.

| Name | Classifier | Segmentation level (L) | No. of classes | Max number of trees/training algorithm | Max. depth of tree/activation func. | Cross validation | Forest accuracy/error function |
|---|---|---|---|---|---|---|---|
| M1 | Decision tree | 20 | 41 | 1 | 100 | 3-fold | – |
| M2 | Decision tree | 15 | 41 | 1 | 50 | 3-fold | – |
| M3 | Decision tree | 3 | 41 | 1 | 100 | 3-fold | – |
| M4 | Decision tree | 3 | 41 | 1 | 50 | 3-fold | – |
| M5 | MLP NN | 20 | 41 | BFGS 74 | Tanh | 3-fold | Entropy |
| M6 | Decision tree | 15 | 41 | 1 | 100 | 3-fold | – |
| M7 | Random forest | 15 | 41 | 500 | – | 3-fold | 0.01 |
| M8 | Random forest | 3 | 41 | 50 | – | 3-fold | 0.001 |
| M9 | Random forest | 3 | 41 | 100 | – | 3-fold | 0.01 |
| M10 | Random forest | 20 | 41 | 100 | – | 3-fold | 0.001 |
| M11 | Random forest | 15 | 41 | 50 | – | 3-fold | 0.05 |
| M12 | Decision tree | 3 | 41 | 1 | 200 | 5-fold | – |

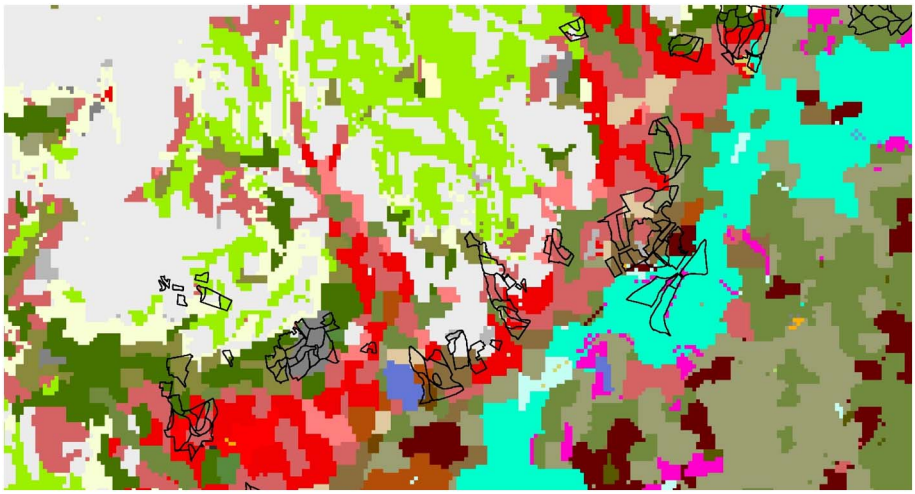

**Fig. 7.** Sporadically available digitized large scale soil type maps were used for external validation. Black lines are soil map delineations with type classification. In the background a part of the newly compiled soil type map.

heavily relied on the available information on soil forming factors with certain interpretation and extrapolation; no direct soil profile data was applied in the process.

The 1:200.000 scale genetic soil map was synthetized from 1:10,000 scale genetic soil maps by counties. These large scale maps were directly prepared to present soil mapping units according to the Hungarian (genetic) soil classification system. They were compiled by soil surveyors based on a detailed soil survey and well-established large-scale soil-landscape models. The integration of county maps was carried out by traditional cartographic methods, applying the available within-county maps, which were compiled (as mentioned earlier) only for the areas of agricultural land. Thus county maps already contained extrapolated areas, where no survey data was available. The national map was put together from the county maps with some cross-border correlation. The mapping intensity and coverage of the counties was not homogeneous, which is strongly reflected in the pattern of the countrywide map. The reliability of the map is the highest on the areas where intensive survey was carried out, and the sequential reconstruction retained the original survey's quality. There is also an initiative for the improvement of this map per se (Kocsis et al., 2015).

The areal representation of soil types predicted by the new map was also compared to the former two national soil type maps. Table 4 summarizes the prediction of the areal extension of various soil types provided by the three distinct maps. Due to the above mentioned differences in their origins and mental models built in their compilations this type of comparison should rather be considered an indication of their similarities and dissimilarities.

The importance of the newly prepared map could actually be evaluated from the practical point of view. This is the first countrywide soil type map that unifies expert inputs and databases from both the agricultural farmlands and forested areas. As a consequence, this map can be equally used for agricultural or forestry oriented purposes providing interoperability between the sectors. Because of the robustness and huge data background, the map is suitable to be involved in nationwide spatial and land use management planning as it first took place in 2015, in the process of the renewal of the National Spatial Plan (2016).

### 3.2. Validation results

Due to the two-level approach, validation was also carried out on both levels. The performance of the classification on MSTG level is presented in Table 5 in the form of confusion matrix and accuracy measures. The average Producer's Accuracy (PA) is 0.5, the average User's Accuracy (UA) is 0.6.

According to PA and UA the weakest predictivity was achieved in the case of salt affected and skeletal soils. The best prediction was provided for sand, brown forest and alluvial soils. As for an overall evaluation of the MSTG level classification results, the overall kappa coefficient shows agreement on the border of moderate and substantial categories (Landis and Koch, 1977).

Performance of the 12 rivalled classification models (M1-M12) on ST level was also evaluated. Table 6 shows the number of profiles in the retained validation set, which were classified correctly by the given model into the various soil types. “N” means the number of observations in the soil type of a given row. “Hit” means the accuracy (in percentage) of the model having the maximum number of correct predictions with regard to the given soil type.

According to the results, M2 and M10 proved to be the best classifiers for most STs, both in the case of nine. The second most frequently best performed model are M5, M9 and M12 with seven STs. However, the number of STs is just one measure of classifier performance. If we also consider the population of various STs represented by the number of observations in the learning dataset the ranking changes slightly. The best classifier for the majority of cases is M12 with 62%, the second one is M5 with 57% and the third one is M6 with 55%. The former ranking reflects the capability of models for the successful identification of wide range of STs even with lowly populated ones. The latter is featuring the applicability of the models for the accurate classification of densely populated STs with spatially extended occurrence. The majority of the models do not show any intention for ST preference, they proved to be the best classifier for diverse set of STs, except for M5, which became the “winner” for four STs, the less populated four of the six Brown forest soil types.

None of the models overperformed 62% accuracy. However, by a proper combination we could finally produce 70% accuracy. It was achieved by rivalling the 12 models in the final categorization of segments by a kind of an ensemble output. Every model has a prediction on what class the given image element belongs to. We ranked these single predictions on the basis of their confidence coming from the validation results. The class predicted by the best performer was assigned to the mapping units. This way we did not stick in one of the prediction models, we kept every model as an extraction of their best performance.

Results of the external validation with legacy type maps are summarized in Table 7. The numbers in the cells indicate areas in hectares. MSTGs are grouped in the table to reveal misclassifications within the same main group more easily. Skeletal and salt-affected soils are more or less correctly classified at least on MSTG level but they are very lowly represented, which is not so surprising, since the legacy

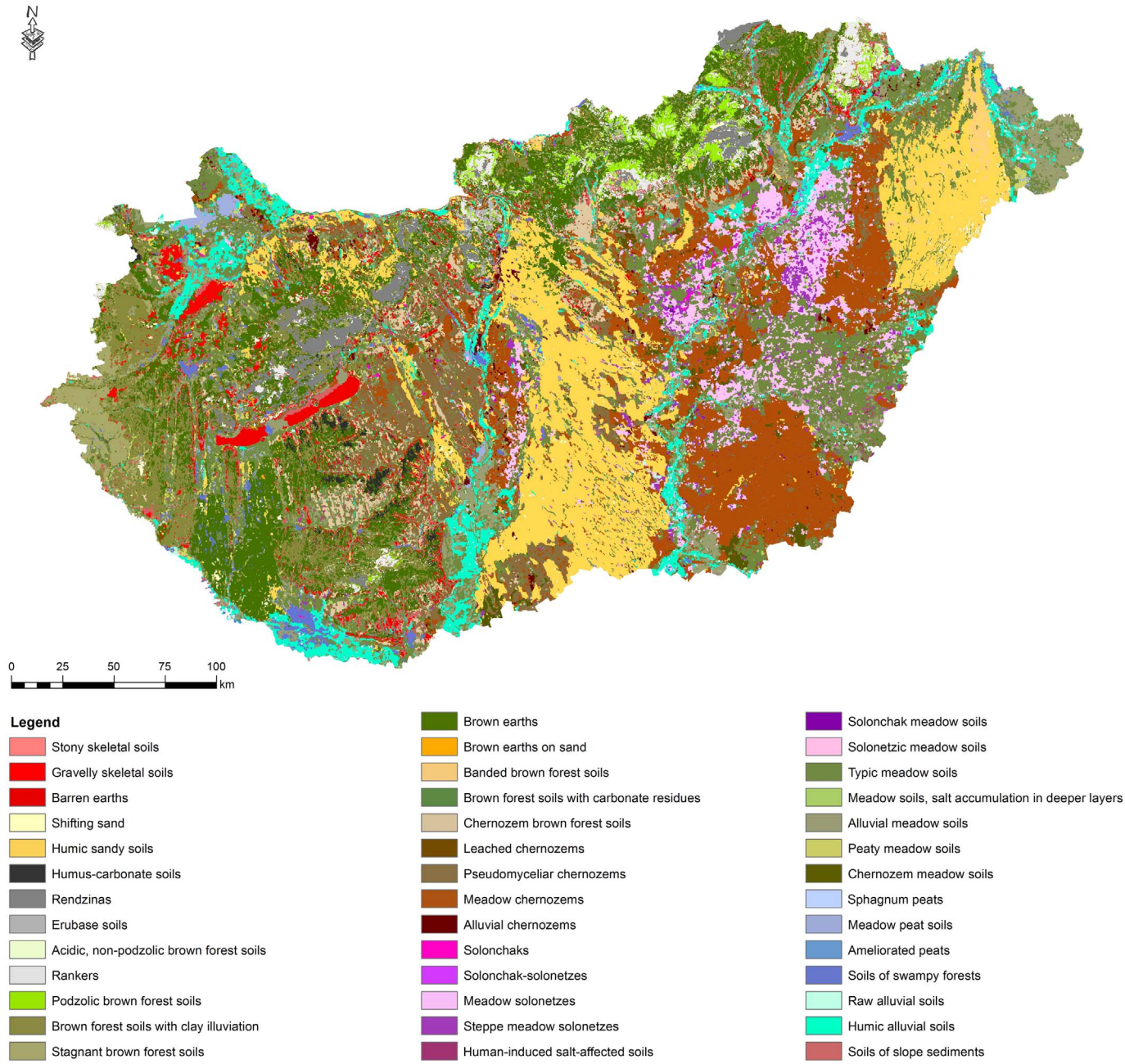


**Fig. 8.** The final product: a unified, national, soil type map with spatially consistent predictive capabilities.

**Table 3**
Comparison of the three nationwide soil type maps. Their similarity is tested for three physiographically separable land types and is expressed by two measures. The smaller the number, the more dissimilar the compared maps are.

| Agreement with the predicted soil type map | | Overall accuracy | Overall kappa |
|---|---|---|---|
| AGROTOPO | Lowlands | 0,31 | 0,24 |
| | Hilly areas | 0,27 | 0,15 |
| | Mountainous areas | 0,29 | 0,18 |
| 1:200.000 genetic soil map | Lowlands | 0,36 | 0,29 |
| | Hilly areas | 0,25 | 0,17 |
| | Mountainous areas | 0,17 | 0,09 |

maps were created for agricultural areas of great farms generally operating on good quality land. Predicted and legacy mapped forest and chernozem soils also coincide with higher confidence, but there are numerous misclassifications. Palpably there are some predicted soil types, which occur in almost every reference class. There are some easily interchangeable ST pairs, like Meadow chernozems (530) and Chernozems meadow (770), in the Hungarian soil classification system (Michéli et al., 2014; Sisák, 2016). They systematically show strong preference for misclassification in our case, too.

Some of the former findings can be partly attributed to the eventuality of the available validation sites and the under-representation of forest areas. As a consequence, this non-systematic comparison could also simply be considered as an indication. The usage of the forest validation data set, which also consists of eventually collected, but point-like observations resulted in an accuracy of 65%, which fairly well approximates the internally achieved 70% overall accuracy.

**Table 4**
Comparison of the areal extension of various soil types in nationwide soil type maps.

| Type code | Soil type name | Area (ha) | | |
|---|---|---|---|---|
| | | AGROTOPO | 1:200.000 genetic soil map | Predicted soil type map |
| 110 | Stony skeletal soils | | 10,738 | 6253 |
| 120 | Gravelly skeletal soils | | 11,836 | 76,235 |
| 130 | Barren earths | 53,828 | 4822 | 163,227 |
| 140 | Shifting sand | 372,435 | 99,075 | 15,553 |
| 150 | Humic sandy soils | 423,743 | 640,561 | 1,273,573 |
| 210 | Raw alluvial soils | 253,195 | 3335 | 59,998 |
| 220 | Humic alluvial soils | | 554,508 | 414,024 |
| 310 | Humus-carbonate soils | | 2598 | 50,201 |
| 320 | Rendzinas | 246,842 | 32,305 | 206,046 |
| 330 | Erubase soils | 14,813 | 442 | 21,715 |
| 340 | Rankers | | | 152,626 |
| 410 | Acidic, non-podzolic brown forest soils | 39,259 | 113 | 38,041 |
| 420 | Podzolic brown forest soils | | 316 | 111,024 |
| 430 | Brown forest soils with clay illuviation | 1,483,198 | 813,400 | 593,451 |
| 440 | Stagnant brown forest soils | 168,367 | 112,028 | 153,778 |
| 450 | Brown earths | 868,462 | 798,665 | 1,039,247 |
| 460 | Brown earths on sand | | | 13,746 |
| 470 | Banded brown forest soils | 191,658 | 162,316 | 59,023 |
| 480 | Chernozem brown forest soils | 435,555 | 291,055 | 217,367 |
| 490 | Brown forest soils with carbonate residues | | | 65,305 |
| 510 | Leached chernozems | | 77,269 | 39,356 |
| 520 | Pseudomyceliar chernozems | 943,472 | 483,017 | 476,746 |
| 530 | Meadow chernozems | 1,015,093 | 991,973 | 1,052,983 |
| 540 | Alluvial chernozems | 9271 | 114,985 | 51,108 |
| 610 | Solonchaks | 4670 | 8257 | 26,081 |
| 620 | Solonchak-solonetzes | 65,859 | 57,664 | 8672 |
| 630 | Meadow solonetzes | 275,110 | 256,851 | 177,705 |
| 640 | Steppe meadow solonetzes | 212,658 | 7608 | 104,933 |
| 650 | Human-induced salt-affected soils | | | 12,172 |
| 710 | Typic meadow soils | 779,730 | 749,357 | 1,410,316 |
| 713 | Meadow soils, salt accumulation in deeper layers | | | 6111 |
| 730 | Solonchak meadow soils | | 21,197 | 5831 |
| 740 | Solonetzic meadow soils | 242,065 | 256,738 | 186,887 |
| 750 | Alluvial meadow soils | 772,865 | 586,934 | 520,324 |
| 760 | Peaty meadow soils | 179,580 | 67,529 | 67,469 |
| 770 | Chernozem meadow soils | | 231,631 | 86,747 |
| 810 | Sphagnum peats | | | 646 |
| 820 | Meadow peat soils | 41,613 | 45,312 | 42,039 |
| 825 | Ameliorated peats | 90,685 | 65,671 | 5764 |
| 910 | Soils of swampy forests | 7984 | 3595 | 66,300 |
| 930 | Soils of slope sediments | | 52,684 | 148,184 |
| 990 | Chernozem soils with forest | | 188,028 | |

The integration of legacy soil data and machine learning methods is used rather widely internationally for the elaboration of improved soil maps. The authors generally evaluate their product as improved spatial soil information, but not necessarily the final product (Bulmer et al., 2016; Kempen et al., 2012). The suggestion is that further effort should be made to extend the applied databases by adding further legacy data or incorporating new soil observations (e.g.: Hengl et al., 2015; Sulaeman et al., 2013) using more and at the same time more informative covariates as well as more sophisticated statistical models (Hengl et al., 2014). In accordance with these trends, we do not

**Table 5**
Accuracy of classification on Main Soil Type Groups (MSTG) achieved by best performing classifiers.

| | | # of observed profiles | | | | | | | | | Sum | User's Accuracy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Chernozems | Brown forest s. | Sand s. | Lithomorphic s. | Peat s. | Alluvial s. | Meadow s. | Salt-affected s. | Skeletal s. | | |
| # of predicted profiles | Chernozems | 246 | 12 | 147 | 6 | 5 | 16 | 80 | 18 | 12 | 542 | 0,45 |
| | Brown forest s. | 26 | 2054 | 280 | 258 | 31 | 6 | 68 | | 207 | 2930 | 0,70 |
| | Sand s. | 139 | 141 | 4492 | 19 | 59 | 12 | 408 | 3 | 34 | 5307 | 0,85 |
| | Lithomorphic s. | | 102 | 8 | 112 | | | 1 | | 20 | 243 | 0,46 |
| | Peat s. | 2 | 6 | 28 | | 232 | 9 | 17 | | | 294 | 0,79 |
| | Alluvial s. | 17 | 4 | 28 | 1 | 49 | 349 | 81 | 2 | 8 | 539 | 0,65 |
| | Meadow s. | 94 | 101 | 338 | 4 | 115 | 111 | 719 | 50 | 17 | 1549 | 0,46 |
| | Salt-affected s. | 4 | | 7 | | 1 | | 19 | 19 | | 50 | 0,38 |
| | Skeletal s. | 13 | 98 | 34 | 13 | 6 | 4 | 11 | | 78 | 257 | 0,30 |
| Sum | | 541 | 2518 | 5362 | 413 | 498 | 507 | 1404 | 92 | 376 | 11,711 | |
| Producer's accuracy | | 0,45 | 0,82 | 0,84 | 0,27 | 0,47 | 0,69 | 0,51 | 0,21 | 0,21 | | |
| Overall kappa | | 0,6 | | | | | | | | | | |
| Overall accuracy | | 0,7 | | | | | | | | | | |

**Table 6**
Performance of the 12 rivalled models on ST level. The best result for each ST is set in bold. "N" means the number of observations in the soil type of a given row. "Hit" means the accuracy (in percentage) of the model having the maximum number of correct predictions with regard to the given soil type.

| Soil type | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 | M9 | M10 | M11 | M12 | Max | N | Hit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stony skeletal soils | 13 | 0 | 3 | 1 | 4 | 5 | 3 | 0 | 12 | **20** | 12 | 2 | **20** | 37 | **54%** |
| Gravelly skeletal soils | **62** | 50 | 3 | 0 | 11 | 12 | 0 | 1 | 56 | 36 | 36 | 31 | **62** | 75 | **83%** |
| Barren earths | 35 | 1 | 0 | 1 | 11 | 9 | 2 | 2 | 7 | **53** | 27 | 14 | **53** | 117 | **45%** |
| Shifting sand | 30 | 32 | 22 | 25 | 34 | 39 | 24 | 22 | **41** | 25 | 26 | 40 | **41** | 50 | **82%** |
| Humic sandy soils | 1803 | 2972 | 2786 | 2786 | 3012 | 2997 | 2758 | 2753 | 2119 | 1925 | 1730 | **3122** | **3122** | 3640 | **86%** |
| Raw alluvial soils | 33 | 35 | 37 | 41 | 42 | **44** | 38 | 39 | 29 | 37 | 29 | 37 | **44** | 54 | **81%** |
| Humic alluvial soils | 221 | 287 | 307 | 306 | 318 | 314 | 289 | 283 | 207 | 254 | 199 | **322** | **322** | 455 | **71%** |
| Humus-carbonate soils | 42 | 38 | 25 | 23 | 26 | 27 | 24 | 24 | 37 | **51** | 44 | 12 | **51** | 68 | **75%** |
| Rendzinas | 164 | **194** | 188 | 183 | 187 | 183 | 184 | 182 | 144 | 161 | 153 | 95 | **194** | 244 | **80%** |
| Erubase soils | **5** | 3 | 4 | 1 | **5** | 0 | 1 | 0 | 4 | 4 | **5** | 0 | **5** | 5 | **100%** |
| Rankers | 59 | **81** | 71 | 75 | 58 | 62 | 70 | 79 | 48 | 50 | 41 | 44 | **81** | 98 | **83%** |
| Acidic, non-podzolic brown forest soils | 75 | 13 | 67 | 67 | **92** | 91 | 17 | 13 | 84 | 83 | 73 | 66 | **92** | 112 | **82%** |
| Podzolic brown forest soils | 51 | 22 | 44 | 41 | **65** | 62 | 25 | 20 | 50 | 54 | 52 | 51 | **65** | 69 | **94%** |
| Brown forest soils with clay illuviation | 499 | 517 | 635 | 462 | **755** | 556 | 603 | 427 | 477 | 500 | 390 | 721 | **755** | 1044 | **72%** |
| Stagnant brown forest soils | 139 | 130 | 157 | 137 | 178 | 156 | 153 | 132 | 154 | 157 | 120 | **207** | **207** | 255 | **81%** |
| Brown earths | 525 | 539 | 661 | 742 | 950 | 1044 | 637 | 712 | 582 | 810 | 593 | **1286** | **1286** | 2124 | **61%** |
| Brown earths on sand | 18 | 4 | 2 | 10 | 3 | 7 | 3 | 5 | **19** | 5 | 4 | 11 | **19** | 28 | **68%** |
| Banded brown forest soils | 342 | 4 | 277 | 281 | 155 | 151 | 47 | 52 | 299 | **398** | 333 | 85 | **398** | 457 | **87%** |
| Chernozem brown forest soils | 35 | 17 | **59** | 57 | 46 | 47 | 37 | 34 | 25 | 42 | 29 | 21 | **59** | 79 | **75%** |
| Brown forest soils with carbonate residues | 33 | 15 | 47 | 48 | **71** | 61 | 50 | 43 | 43 | 46 | 32 | 12 | **71** | 106 | **67%** |
| Leached chernozems | 12 | **14** | 8 | 5 | 11 | 10 | 5 | 8 | 12 | 9 | 8 | 3 | **14** | 17 | **82%** |
| Pseudomyceliar chernozems | 77 | **114** | 58 | 54 | 63 | 59 | 58 | 50 | 60 | 71 | 56 | 103 | **114** | 238 | **48%** |
| Meadow chernozems | 35 | 84 | 39 | 41 | 55 | 58 | 39 | 38 | 23 | 29 | 39 | **96** | **96** | 183 | **52%** |
| Alluvial chernozems | 17 | 4 | 6 | 6 | 4 | 6 | 1 | 3 | 12 | **19** | 10 | 9 | **19** | 27 | **70%** |
| Solonchaks | **3** | **3** | 2 | 2 | 0 | 2 | **3** | **3** | 0 | 0 | 1 | 0 | **3** | 3 | **100%** |
| Solonchak-solonetzes | **3** | 0 | **3** | 2 | 1 | 0 | 0 | 0 | **3** | 0 | **3** | 0 | **3** | 3 | **100%** |
| Meadow solonetzes | 16 | 14 | **22** | 18 | 16 | **22** | 21 | 16 | 16 | 20 | 11 | 10 | **22** | 27 | **81%** |
| Steppe meadow solonetzes | 11 | 12 | 11 | 12 | 14 | 15 | **17** | **17** | 11 | 11 | 9 | 7 | **17** | 18 | **94%** |
| Human-induced salt-affected soils | 0 | 0 | 0 | 0 | 0 | 0 | **1** | **1** | **1** | **1** | 0 | 0 | **1** | 1 | **100%** |
| Typic meadow soils | 112 | 28 | 82 | 115 | 136 | 176 | 59 | 89 | 11 | 61 | 38 | **440** | **440** | 1031 | **43%** |
| Meadow soils, salt accumulation in deeper layers | 10 | 0 | 2 | 3 | 6 | 5 | 0 | 5 | 9 | 10 | **11** | 7 | **11** | 22 | **50%** |
| Solonchak meadow soils | **6** | 0 | 3 | 3 | 2 | 4 | 2 | 3 | 0 | 5 | 1 | 0 | **6** | 9 | **67%** |
| Solonetzic meadow soils | 8 | 0 | 3 | 1 | 8 | 8 | 2 | 0 | 7 | **14** | 6 | 9 | **14** | 32 | **44%** |
| Alluvial meadow soils | 90 | 38 | 74 | 53 | 88 | 61 | 61 | 43 | **95** | 88 | 89 | 84 | **95** | 221 | **43%** |
| Peaty meadow soils | **43** | 17 | 23 | 20 | 32 | 32 | 22 | 15 | 34 | **43** | 34 | 29 | **43** | 95 | **45%** |
| Chernozem meadow soils | **16** | 0 | 3 | 2 | 3 | 0 | 0 | 2 | 13 | 2 | 5 | 8 | **16** | 37 | **43%** |
| Sphagnum peats | 0 | 0 | 0 | **1** | **1** | 0 | 0 | 0 | 0 | **1** | **1** | 0 | **1** | 1 | **100%** |
| Meadow peat soils | 157 | **175** | 39 | 37 | 30 | 25 | 36 | 38 | 170 | 139 | 27 | 141 | **175** | 192 | **91%** |
| Ameliorated peats | 0 | **2** | 0 | 1 | 0 | 0 | **2** | **2** | **2** | 0 | **2** | 0 | **2** | 2 | **100%** |
| Soils of swampy forests | 90 | **189** | 123 | 121 | 133 | 137 | 118 | 117 | 88 | 146 | 104 | 109 | **189** | 307 | **62%** |
| Soils of slope sediments | 18 | **44** | 7 | 7 | 8 | 11 | 5 | 6 | 12 | 34 | 21 | 16 | **44** | 150 | **29%** |
| All | 4908 | 5692 | 5903 | 5791 | 6634 | 6498 | 5417 | 5279 | 5016 | 5414 | 4404 | 7250 | **8272** | 11,733 | |
| Model perf. | **42%** | **49%** | **50%** | **49%** | **57%** | **55%** | **46%** | **45%** | **43%** | **46%** | **38%** | **62%** | **71%** | | |

consider our map as the perfect and ultimate product, however, we suggest that numerous improvements were achieved by its compilation.

## 4. Conclusions

A unified, national soil type map with spatially consistent predictive capabilities was compiled applying traditional and newly tested DSM classification methods: segmentation of a synthesized image consisting of predictor variables and multi-phase, sequential classification by Classification and Regression Trees, Random Forests and Artificial Neural Networks. Object based classification using spatial-thematic segments was applied to produce more map-like products and accelerate computations. Classifications in the phase space of the co-variables were carried out on two levels (main soil type group and soil type) to achieve better results. Performance of classifiers was continuously assessed and applied for the identification of best performing predictions, which were combined for the production of the final map.

We do not consider our map as the ultimate product, it could and should be refined and improved in a number of ways. The workflow inherently makes it possible to keep the map easily updated or refined if new qualified data becomes available. However, there are other opportunities for its upgrading. Partly independently of the presented approach, further attributes of profiles contained by forestry databases, namely soil texture and soil depth classes have also been mapped (not discussed in details in the present paper but with plans for a forthcoming publication). Texture and depth predictions are also planned to be used to (i) verify and (ii) fine-tune the recent soil type map in order to obtain a coherent soil map series. Their mutual consideration can help to avoid certain inconsistencies (e.g. between soil type and texture class at the same location), which can easily occur if maps are developed independently. The joint mapping of various soil features can be imagined as a successive approximation process, where one map will assist in the compilation of the other, which can then be further used for the next modelling of the former. One of our recent challenging tasks is the investigation of the convergence of the aforementioned process.

Last but not least, the present work produced a soil type map according to the traditional soil classification system. This should be attributed to the shortage of geo-referenced profile description based on WRB and/or Hungarian renewed classification, not to the intention of the authors. In the case of sufficient number of observations with WRB and/or Hungarian renewed classification, the work could and should be repeated to produce a brand new Hungarian soil type map based on recent and surveyed data and with a legend using WRB and/or Hungarian renewed classification.

**Table 7**

| Predicted soil type | | Type according to external validation dataset | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Skeletal s. | | | | Sand s. | | Lithomorphic s. | | | Brown forest s. | | | | Chernozems | | | | Salt–affected s. | | | | | Meadow s. | | | | Peat s. | | | Alluvial s. | |
| | | 110 | 120 | 130 | 930 | 140 | 150 | 310 | 320 | 330 | 410 | 430 | 440 | 490 | 480 | 520 | 530 | 540 | 610 | 620 | 630 | 730 | 740 | 710 | 750 | 760 | 770 | 820 | 825 | 910 | 210 | 220 |
| Skeletal s. | 110 | 0 | 0 | 35 | 9 | 0 | 0 | 0 | 0 | 9 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 120 | 0 | 10 | 3 | 44 | 0 | 5 | 0 | 0 | 3 | 0 | 51 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 2 | 7 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 130 | 15 | 7 | 82 | 257 | 0 | 10 | 5 | 33 | 4 | 0 | 526 | 6 | 1 | 612 | 37 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 38 | 7 | 0 | 2 | 0 | 1 | 0 | 0 | 6 |
| | 930 | 0 | 1 | 22 | 235 | 0 | 6 | 0 | 31 | 7 | 0 | 885 | 1 | 12 | 518 | 17 | 36 | 0 | 0 | 0 | 1 | 3 | 3 | 81 | 17 | 18 | 7 | 9 | 15 | 3 | 0 | 33 |
| Sand s. | 140 | 0 | 0 | 0 | 0 | 0 | 83 | 0 | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 150 | 0 | 0 | 0 | 43 | 102 | 5925 | 0 | 0 | 0 | 0 | 119 | 25 | 0 | 14 | 60 | 591 | 39 | 0 | 3 | 3 | 41 | 9 | 238 | 359 | 126 | 224 | 0 | 0 | 0 | 0 | 45 |
| Lithom.s. | 310 | 3 | 0 | 8 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 111 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 320 | 117 | 0 | 42 | 88 | 0 | 0 | 12 | 61 | 51 | 3 | 248 | 0 | 0 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| | 330 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Brown forest s. | 410 | 0 | 0 | 8 | 14 | 0 | 0 | 0 | 0 | 39 | 0 | 214 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 430 | 0 | 10 | 98 | 220 | 0 | 16 | 6 | 0 | 328 | 0 | 1651 | 84 | 0 | 175 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 108 | 22 | 2 | 0 | 9 | 13 | 0 | 2 | 180 |
| | 440 | 0 | 2 | 3 | 39 | 0 | 31 | 0 | 0 | 0 | 0 | 105 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 12 | 0 | 6 | 0 | 4 | 0 | 21 |
| | 490 | 0 | 0 | 0 | 39 | 1 | 7 | 0 | 0 | 0 | 0 | 508 | 0 | 0 | 7 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 10 |
| Chernozems | 480 | 0 | 0 | 15 | 266 | 1 | 36 | 0 | 2 | 118 | 0 | 1062 | 21 | 0 | 1827 | 216 | 461 | 0 | 0 | 0 | 21 | 0 | 0 | 163 | 139 | 30 | 124 | 0 | 1 | 6 | 15 | 75 |
| | 520 | 0 | 0 | 0 | 128 | 2 | 269 | 0 | 0 | 0 | 0 | 315 | 0 | 0 | 390 | 754 | 1168 | 0 | 0 | 0 | 50 | 25 | 87 | 277 | 16 | 16 | 390 | 0 | 0 | 0 | 0 | 38 |
| | 530 | 0 | 0 | 0 | 115 | 1 | 6 | 0 | 0 | 16 | 0 | 520 | 0 | 0 | 1516 | 421 | 3705 | 0 | 0 | 25 | 80 | 374 | 693 | 713 | 141 | 1 | 1150 | 0 | 0 | 0 | 0 | 153 |
| | 540 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 10 | 2 | 90 | 0 | 0 | 9 | 71 | 36 | 39 | 148 | 6 | 0 | 85 | 0 | 0 | 0 | 0 | 64 |
| Salt–affected s. | 610 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 1 | 0 | 36 | 0 | 0 | 0 | 0 | 8 | 17 | 140 | 34 | 0 | 13 | 0 | 0 | 0 | 0 | 67 |
| | 620 | 0 | 0 | 0 | 0 | 0 | 124 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 |
| | 630 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 67 | 0 | 0 | 0 | 210 | 24 | 529 | 736 | 17 | 0 | 101 | 0 | 0 | 0 | 0 | 24 |
| | 730 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 740 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 119 | 7 | 0 | 10 | 164 | 107 | 792 | 864 | 18 | 0 | 78 | 0 | 0 | 0 | 0 | 31 |
| Meadow s. | 710 | 87 | 15 | 23 | 564 | 0 | 300 | 20 | 47 | 70 | 0 | 936 | 54 | 7 | 830 | 60 | 2234 | 3 | 94 | 256 | 817 | 165 | 2365 | 6661 | 757 | 504 | 956 | 50 | 226 | 33 | 21 | 798 |
| | 750 | 14 | 7 | 23 | 372 | 0 | 65 | 1 | 21 | 50 | 0 | 652 | 28 | 0 | 538 | 0 | 506 | 0 | 0 | 0 | 45 | 40 | 514 | 1384 | 441 | 221 | 243 | 9 | 70 | 6 | 1 | 454 |
| | 760 | 0 | 0 | 6 | 85 | 0 | 59 | 0 | 1 | 39 | 0 | 13 | 0 | 0 | 34 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 5 | 56 | 27 | 33 | 0 | 56 | 42 | 0 | 0 | 55 |
| | 770 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 38 | 0 | 0 | 37 | 0 | 312 | 0 | 0 | 0 | 9 | 0 | 51 | 171 | 0 | 0 | 181 | 0 | 0 | 0 | 0 | 0 |
| Peat s. | 820 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 86 | 0 | 0 | 0 | 0 | 18 | 0 | 68 | 0 | 0 | 0 | 0 | 9 | 0 | 2 | 21 |
| | 825 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 910 | 20 | 5 | 8 | 152 | 0 | 4 | 0 | 31 | 28 | 0 | 33 | 1 | 0 | 22 | 0 | 112 | 0 | 0 | 0 | 1 | 76 | 105 | 801 | 129 | 124 | 80 | 3 | 11 | 3 | 0 | 15 |
| Alluv. s. | 210 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 21 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 22 |
| | 220 | 1 | 0 | 0 | 67 | 0 | 26 | 0 | 0 | 11 | 0 | 135 | 0 | 0 | 175 | 0 | 65 | 0 | 0 | 0 | 0 | 0 | 69 | 241 | 289 | 0 | 9 | 0 | 0 | 0 | 0 | 377 |

**Results of the external validation. The numbers in the cells display areas in ha. MSTGs are indicated with identical colours to depict misclassifications within the same main group.**

## Acknowledgement

Our work was supported by the Hungarian National Scientific Research Foundation (OTKA) under Grant K105167 and AGRAR-KLÍMA.2 VKSZ_12-1-2013-0034.

## References

AGRAGis, 2016. AGRATéR [WWW Document]. http://agrater.hu/.
Amini, M., Abbaspour, K.C., Khademi, H., Fathianpour, N., Afyuni, M., Schulin, R., 2005. Neural network models to predict cation exchange capacity in arid regions of Iran. Eur. J. Soil Sci. 56, 551–559. http://dx.doi.org/10.1111/j.1365-2389.2005.0698.x.
Andrew, M.E., Wulder, M.a., Nelson, T.A., Coops, N.C., 2015. Spatial data, analysis approaches, and information needs for spatial ecosystem service assessments: a review. GISci. Remote Sens. 52, 344–373. http://dx.doi.org/10.1080/15481603.2015.1033809.
Bagheri Bodaghabadi, M., Martínez-Casasnovas, J.A., Salehi, M.H., Mohammadi, J., Esfandiarpoor Borujeni, I., Toomanian, N., Gandomkar, A., 2015. Digital soil mapping using artificial neural networks and terrain-related attributes. Pedosphere 25, 580–591.
Bakacsi, Z., Laborczi, A., Szabó, J., Takács, K., Pásztor, L., 2014. Proposed correlation between the legend of the 1:100.000 scale geological map and the FAO code system for soil parent material. [In Hungarian]. Agrokém. Talajt. 63, 189–202.
Behrens, T., Scholten, T., 2006. A comparison of data-mining techniques in predictive soil mapping. Dev. Soil Sci. 31, 353–364. http://dx.doi.org/10.1016/S0166-2481(06)31025-2.
Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. J. Plant Nutr. Soil Sci. 168, 21–33. http://dx.doi.org/10.1002/jpln.200421414.
Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Remote Sens. 58, 239–258. http://dx.doi.org/10.1016/j.isprsjprs.2003.10.002.
Bou Kheir, R., Bøcher, P.K., Greve, M.B., Greve, M.H., 2010. The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data. Hydrol. Earth Syst. Sci. 14, 847–857. http://dx.doi.org/10.5194/hess-14-847-2010.
Brauch, H.G., Spring, Ú.O., Mesjasz, C., Grin, J., Kameri-Mbote, P., Chourou, B., Dunay, P., Birkmann, J., 2011. Coping with Global Environmental Change, Disasters and Security, Security. Springer, Berlin. http://dx.doi.org/10.1007/978-3-642-17776-7.
Breiman, L., Cutler, A., 2009. Random Forests homepage [WWW Document]. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees, The Wadsworth Statisticsprobability Series. Wadsworth, Belmont, California.
Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: history and future directions. Geoderma 264, 256–274. http://dx.doi.org/10.1016/j.geoderma.2015.05.017.
Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239, 68–83. http://dx.doi.org/10.1016/j.geoderma.2014.09.019.
Bulmer, C., Schmidt, M.G., Heung, B., Scarpone, C., Zhang, J., Filatow, D., Finvers, M., Berch, S., Smith, S., 2016. Improved soil mapping in British Columbia, Canada, with legacy soil data and random forest. In: Zhang, G.-L., Brus, D., Liu, F., Song, X.-D., Lagacherie, P. (Eds.), Digital Soil Mapping Across Paradigms, Scales and Boundaries. Springer Singapore, Singapore, pp. 291–303.
Büttner, G., Maucha, G., Bíró, M., Kosztra, B., Pataki, R., Petrik, O., 2004. National land cover database at scale 1:50000 in Hungary. EARSeL eProceedings 3, 323–330.
Chang, D., Islam, S., 2000. Estimation of soil physical properties using remote sensing and artificial neural network. Remote Sens. Environ. 74, 534–544. http://dx.doi.org/10.1016/S0034-4257(00)00144-9.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. http://dx.doi.org/10.1177/001316446002000104.
Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. http://dx.doi.org/10.5194/gmd-8-1991-2015.
Dai, X., Huo, Z., Wang, H., 2011. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. F. Crop. Res. 121, 441–449. http://dx.doi.org/10.1016/j.fcr.2011.01.016.
Dobos, E., Vadnai, P., Michéli, E., Láng, V., Fuchs, M., Seres, A., 2011. Új generációs nemzetközi talajtérképek készítése, az e-SOTER módszertan. In: Az Elmélet És a Gyakorlat Találkozása a Térinformatikában II, pp. 53–60.
Dobos, E., Seres, A., Vadnai, P., Michéli, E., Fuchs, M., Láng, V., Bertóti, R.D., Kovács, K., 2013. Soil parent material delineation using MODIS and SRTM data. Hungarian Geogr. Bull. 62, 133–156.
Dobos, E., Vadnai, P., Pásztor, L., Micheli, E., Kovács, K., Bertóti, R.D., 2016. A WRB based harmonized digital soil map of the Carpathian-basin. In: Geophysical Research Abstracts, (p. EGU2016-12592).
Dokuchaev, V.V., 1883. The Russian Chernozem [In Russian]. St. Petersburg, Russia.
Dokuchaev, V.V., 1899. A Contribution to the Theory of Natural Zones: Horizontal and Vertical Soil Zones [In Russian]. Mayor's Office Press, St. Petersburg.
EU-DEM, 2015. Digital Elevation Model Over Europe [WWW Document]. http://www.eea.europa.eu/data-and-maps/data/eu-dem.
FAO, 2006. Guidelines for Soil Description. FAO, Rome. http://dx.doi.org/10.1017/S0014479706384906.
Giasson, E., Sarmento, E.C., Weber, E., Flores, C.A., Hasenack, H., 2011. Decision trees for digital soil mapping on subtropical basaltic steeplands. Sci. Agric. (Piracicaba, Braz.) 68, 167–174. http://dx.doi.org/10.1590/S0103-90162011000200006.
Golden, M., Michéli, E., Ditzler, C., Eswaran, H., Owens, P., Zhang, G., McBratney, A., Hempel, J., Montanarella, L., Schad, P., 2010. Time for a Universal soil classification system. In: 19th World Congress of Soil Science, Soil Solutions for a Changing World, pp. 48–51.
Greve, M.H., Kheir, R.B., Greve, M.B., Bøcher, P.K., 2012. Quantifying the ability of environmental parameters to predict soil texture fractions using regression-tree model with GIS and LIDAR data: the case study of Denmark. Ecol. Indic. 18, 1–10. http://dx.doi.org/10.1016/j.ecolind.2011.10.006.
Gyalog, L., Síkhegyi, F., 2005. Geological Map of Hungary, 1:100.000. Geological Institute of Hungary, Budapest [In Hungarian]. (WWW Document). http://loczy.mfgi.hu/fdt100/.
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398. http://dx.doi.org/10.1016/j.geoderma.2004.06.007.
Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1 km—global soil information based on automated mapping. PLoS One 9, e105992. http://dx.doi.org/10.1371/journal.pone.0105992.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10. http://dx.doi.org/10.1371/journal.pone.0125814.
Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. Geoderma 214–215, 141–154. http://dx.doi.org/10.1016/j.geoderma.2013.09.016.
Hiederer, R., Michéli, E., Durrant, T., 2011. Evaluation of BioSoil Demonstration Project — Soil Data Analysis. EUR 24729 EN.
Hitziger, M., Ließ, M., 2014. Comparison of three supervised learning methods for digital soil mapping: application to a complex terrain in the Ecuadorian Andes. Appl. Environ. Soil Sci. 2014. http://dx.doi.org/10.1155/2014/809495.
Jeney, I., Jassó, F., 1983. Magyarország genetikus talajtérképe (méretarány: 1:200.000) [Genetic soil type map of Hungary]. Kartográfiai Vállalat, Budapest.
Jenny, H., 1941. Factors of soil formation, a system of quantitative pedology. In: Soil Science. NY: Dover Publications, New York. http://dx.doi.org/10.1097/00010694-194111000-00009.
Kaul, M., Hill, R.L., Walthall, C., 2005. Artificial neural networks for corn and soybean yield prediction. Agric. Syst. 85, 1–18. http://dx.doi.org/10.1016/j.agsy.2004.07.009.
Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G.B.M., de Vries, F., 2012. Efficiency comparison of conventional and digital soil mapping for updating soil maps. Soil Sci. Soc. Am. J. 76, 2097–2115. http://dx.doi.org/10.2136/sssaj2011.0424.
Kim, M., Gilley, J.E., 2008. Artificial neural network estimation of soil erosion and nutrient concentrations in runoff from land application areas. Comput. Electron. Agric. 64, 268–275. http://dx.doi.org/10.1016/j.compag.2008.05.021.
Kocsis, M., Berényi Üveges, J., Várszegi, G., Sisák, I., 2015. The genetic soil map of Hungary and the evaluation of the soil categories. [In Hungarian]. Agrokém. Talajt. 64, 53–72.
Kreybig, L., 1937. Magyar Királyi Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere (The survey, analytical and mapping method of the Hungarian Royal Institute of Geology). Magy. Királyi Földtani Intézet Évkönyve 31, 147–244.
Lacarce, E., Le Bas, C., Cousin, J.L., Pesty, B., Toutain, B., Houston Durrant, T., Montanarella, L., 2009. Data management for monitoring forest soils in Europe for the Biosoil project. Soil Use Manag. 25, 57–65. http://dx.doi.org/10.1111/j.1475-2743.2009.00194.x.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174. http://dx.doi.org/10.2307/2529310.
Láng, I., 1983. The Agro-ecological Potential of Hungarian Agriculture in 2000. Mezőgazdasági Kiadó, Budapest.
Láng, V., Fuchs, M., Szegi, T., Csorba, Á., Michéli, E., 2016. Deriving World Reference Base reference soil groups from the prospective global soil map product — a case study on major soil types of Africa. Geoderma 263, 226–233. http://dx.doi.org/10.1016/j.geoderma.2015.07.005.
Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sens. Environ. 90, 331–336. http://dx.doi.org/10.1016/j.rse.2004.01.007.
Makó, A., Tóth, B., Hernádi, H., Farkas, C., Marth, P., 2010. Introduction of the Hungarian detailed soil hydrophysical database (MARTHA) and its use to test external pedotransfer functions. Agrokém. Talajt. 59, 29–38.
McBratney, A., Mendonça Santos, M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. http://dx.doi.org/10.1016/S0016-7061(03)00223-4.
Michéli, E., 2011. Expression of soil-forming processes in soil classifications based on diagnostic principles. [In Hungarian]. Agrokém. Talajt. 60, 17–32.
Michéli, E., Fuchs, M., Hegymegi, P., Stefanovits, P., 2006. Classification of the major soils of Hungary and their correlation with the World Reference Base for soil resources (WRB). Agrokém. Talajt. 55, 19–28.
Michéli, E., Fuchs, M., Láng, V., Szegi, T., Szabóné Kele, G., 2014. Methods for modernizing the elements and structure of the Hungarian soil classification system. Agrokém. Talajt. 63, 69–78.
Michéli, E., Fuchs, M., Láng, V., Szegi, T., Dobos, E., Szabóné Kele, G., 2015. Javaslat

talajosztályozási rendszerünk megújítására: alapelvek, módszerek, alapegységek. Agrokém. Talajt. 64, 285–297.

Miller, B.A., Schaetzl, R.J., 2014. The historical role of base maps in soil geography. Geoderma 230–231, 329–339. http://dx.doi.org/10.1016/j.geoderma.2014.04.020.

Minasny, B., Malone, B., McBratney, A.B., 2012. Digital Soil Assessments and Beyond. Taylor and Francis Group, London.

Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modelling. Int. J. Geogr. Inf. Sci. 16, 533–549. http://dx.doi.org/10.1080/13658810210138715.

MSZ, 1978a. Hungarian National Standards – MSZ-08-0206-1, Analysis of Soil Chemical Properties. (General approaches).

MSZ, 1978b. Hungarian National Standards – MSZ-08-0206-2, Analysis of Soil Chemical Properties. (Laboratory analysis).

NAGis, 2016. National Adaptation Geo-information System (NAGiS) [WWW Document]. http://nagis.hu/en.

NASA LP DAAC, 2015. MODIS 09 (16.03.2012, 07.09.2013) and MODIS13Q (03.2012, 09.2013) Products. USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (WWW Document). http://modis.gsfc.nasa.gov/data.

National Spatial Plan, 2016. Areas With Excellent Productivity [WWW Document]. http://www.terport.hu/webfm_send/4211.

Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213, 385–399. http://dx.doi.org/10.1016/j.geoderma.2013.08.024.

Pachepsky, Y.a., Timlin, D., Varallyay, G., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60, 727–733. http://dx.doi.org/10.2136/sssaj1996.03615995006000030007x.

Pásztor, L., Szabó, J., Bakacsi, Z., 2010. Digital processing and upgrading of legacy data collected during the 1:25 000 scale Kreybig soil survey. Acta Geod. Geophys. Hungarica 45, 127–136. http://dx.doi.org/10.1556/AGeod.45.2010.1.18.

Pásztor, L., Szabó, J., Bakacsi, Z., Matus, J., Laborczi, A., 2012. Compilation of 1 : 50,000 scale digital soil maps for Hungary based on the digital Kreybig soil information system. J. Maps 8, 215–219. http://dx.doi.org/10.1080/17445647.2012.705517.

Pásztor, L., Bakacsi, Z., Laborczi, A., Szabó, J., 2013. Downscaling of categorical soil maps with the aid of auxiliary spatial soil information and data mining methods. [In Hungarian]. Agrokém. Talajt. 62, 205–218. http://dx.doi.org/10.1556/Agrokem.62.2013.2.3.

Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Dobos, E., Illés, G., Bakacsi, Z., Szabó, J., 2015. Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools. Hungarian Geogr. Bull. 64, 49–64.

Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Bakacsi, Z., Szabó, J., 2016. Variations for the implementation of SCORPAN's "S". In: Zhang, G.-L., Brus, D.J., Liu, F., Song, X.-D., Lagacherie, P. (Eds.), Digital Soil Mapping Across Paradigms, Scales and Boundaries. Springer Science + Business Media, Singapore, pp. 331–342.

Patel, R.M., Prasher, S.O., God, P.K., Bassi, R., 2002. Soil salinity prediction using artificial neural networks. J. Am. Water Resour. Assoc. 38, 91–100. http://dx.doi.org/10.1111/j.1752-1688.2002.tb01537.x.

Pentelényi, A., Scharek, P., 2006. A talajvízszint mélysége a felszín alatt, 1:500.000. (Groundwater level map of Hungary). [WWW Document]. http://map.mfgi.hu/tvz_251020/.

Rossiter, D.G., 2014. Technical note: statistical methods for accuracy assessment of classified thematic maps. Tech. Rep. ITC.

Saffari, M., Yasrebi, J., Sarikhani, F., Gazni, R., Moazallahi, M., Fathi, H., Emadi, M., 2009. Evaluation of artificial neural network models for prediction of spatial variability of some soil chemical properties. Res. J. Biol. Sci. 4, 815–820.

Scull, P., Franklin, J., Chadwick, O.A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecol. Model. 181, 1–15. http://dx.doi.org/10.1016/j.ecolmodel.2004.06.036.

Shantz, H.L., 1911. Natural Vegetation as an Indicator of the Capabilities of Land for Crop Production in the Great Plains Area. US Dep. Agric. Bur. Plant Ind. Bull. 201, 100.

Sisák, I., 2016. Mai tanulságok a talajosztályozás 1961 és 1999 közötti változásaiból. Agrokém. Talajt. 65, 161–168.

State Forest Service — ÁESZ, 2004. Guidelines for Forest Management Planning.

Stefanovits, P., 1963. Magyarország talajai. Akadémiai Kiadó, Budapest.

Stefanovits, P., 1972. Talajtan. Mezőgazda Kiadó, Budapest.

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: Digital Soil Mapping — Progress in Soil Science 2: Bridging Research, Environmental Application, and Operation. Springer Science + Business Media B.V, Dordrecht, pp. 179–190.

Sulaeman, Y., Minasny, B., McBratney, A.B., Sarwani, M., Sutandi, A., 2013. Harmonizing legacy soil data for digital soil mapping in Indonesia. Geoderma 192, 77–85. http://dx.doi.org/10.1016/j.geoderma.2012.08.005.

Szabolcs, I., 1966. A genetikus üzemi talajtérképezés módszerkönyve. OMMI, Budapest.

Szentimrey, T., Bihari, Z., 2007. Mathematical background of the spatial interpolation methods and the software MISH (meteorological interpolation based on surface homogenized data basis). In: Proceedings From the Conference on Spatial Interpolation in Climatology and Meteorology. Budapest, pp. 17–27.

Szodfridt, I., 1993. The Discipline of Forest Landsites (in Hungarian: Erdészeti termőhelyismeret-tan). Mezőgazda Kiadó, Budapest.

Várallyay, G., 2002. Soil survey and soil monitoring in Hungary. Eur. Soil Bur. Res. Rep. 9, 139–149.

Várallyay, G., Szűcs, L., Murányi, A., Rajkai, K., Zilahy, P., 1979. Map of soil factors determining the agro-ecological potential of Hungary (1:100000). I. [In Hungarian]. Agrokém. Talajt. 28, 363–384.

Várallyay, G., Szűcs, L., Zilahy, P., Rajkai, K., Murányi, A., 1985. Soil factors determining the agro-ecological potential of Hungary. Agrokém. Talajt. 34, 90–94.

Verheye, W.H., 2009. Land Use, Land Cover and Soil Sciences – Vol. I – Land Use, Land Cover and Soil Sciences. EOLSS Publishers/UNESCO, Oxford.

Webster, R., 2015. Technological developments for spatial prediction of soil properties , and Danie Krige's influence on it. J. South. Afr. Inst. Min. Metall. 115, 165–172.

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. Plant Soil 340, 7–24. http://dx.doi.org/10.1007/s11104-010-0425-z.

Zell, A., 1994. Simulation Neuronaler Netze. Oldenbourg, München.