DOI: 10.15201/hungeobull.64.1.5 Hungarian Geographical Bulletin **64** 2015 (1) 49–64.

Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools

László PÁSZTOR[1], Annamária LABORCZI[1], Katalin TAKÁCS[1], Gábor SZATMÁRI[2], Endre DOBOS[3], Gábor ILLÉS[4], Zsófia BAKACSI[1] and József SZABÓ[1]

Abstract

Due to former soil surveys and mapping activities significant amount of soil information has accumulated in Hungary. Present soil data requirements are mainly fulfilled with these available datasets either by their direct usage or after certain specific and generally fortuitous, thematic and/or spatial inference. Due to the more and more frequently emerging discrepancies between the available and the expected data, there might be notable imperfection as for the accuracy and reliability of the delivered products. With a recently started project we would like to significantly extend the potential, how soil information requirements could be satisfied in Hungary. We started to compile digital soil maps, which fulfil optimally the national and international demands from points of view of thematic, spatial and temporal accuracy. In addition to the auxiliary, spatial data themes related to soil forming factors and/or to indicative environmental elements we heavily lean on the various national soil databases. The set of the applied digital soil mapping techniques is gradually broadened incorporating and eventually integrating geostatistical, data mining and GIS tools. Regression kriging has been used for the spatial inference of certain quantitative data, like particle size distribution components, rootable depth and organic matter content. Classification and regression trees were applied for the understanding of the soil-landscape models involved in existing soil maps, and for the post-formalization of survey/compilation rules. The relationships identified and expressed in decision rules made the compilation of spatially refined category-type soil maps (like genetic soil type and soil productivity maps) possible with the aid of high resolution environmental auxiliary variables. In our paper, we give a short introduction to soil mapping and information management concentrating on the driving forces for the renewal of soil spatial data infrastructure provided by the framework of Digital Soil Mapping. The first results of DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary) project are presented in the form of brand new national and regional soil maps.

Keywords: classification and regression trees, digital soil mapping, regression kriging, spatial soil information

[1] Institute for Soil Science and Agricultural Chemistry, Centre for Agricultural Research, H-1022 Budapest, Herman Ottó út 15. E-mails: pasztor.laszlo@agrar.mta.hu, laborczi.annamaria@agrar.mta.hu, takacs.katalin@agrar.mta.hu, bakacsi.zsofia@agrar.mta.hu, szabo.jozsef@agrar.mta.hu

[2] Department of Physical Geography and Geoinformatics, University of Szeged, H-6722 Szeged, Egyetem u. 2–6. E-mail: szatmari.gabor.88@gmail.com

[3] Department of Physical Geography and Environmental Sciences, University of Miskolc, H-3515 Miskolc-Egyetemváros, E-mail: ecodobos@uni-miskolc.hu

[4] Forest Research Institute, National Agricultural Research and Innovation Centre, H-9600 Sárvár, Várkerület 30/a. E-mail: illesg@erti.hu

Introduction

Demands on spatial soil information

Demands on soil related information have been significant worldwide and are still increasing (Bullock, P. 1999; Mermut, A.R. and Eswaran, H. 2000; Tóth, G. *et al.* 2008; Sanchez, P.A. *et al.* 2009; Baumgardner, M.F. 2011). Recent requests often do not refer to primary or even secondary soil properties, but to various processes, functions, services and/or systems related to soils (Omuto, C. *et al.* 2013).

Soil maps were typically used for a long time to satisfy these needs. Due to the relatively high costs of new data collection and the spreading of Geographic Information technology, Spatial Soil Information Systems (SSISs) and Digital Soil Mapping (DSM), these approaches have taken over the role of traditional soil maps in the field of data service. Nevertheless, legacy soil data are still heavily relied on, as they include an abundance of information exploitable by proper methodology in GIS/SSIS/DSM environment. Not only the degree but also the nature of current needs for soil information has changed. Traditionally focus was on the agricultural functions of soils, which was also reflected in the methodology of data collection and mapping.

Recently the information related to additional soil functions is becoming identically important (Blum, W.E.H. 2005; Panagos, P. *et al.* 2012). This types of information requirement generally cannot be fulfilled with new data collections, at least not on such a level as in the frame of traditional soil surveys (Montanarella, L. 2010). As a consequence of these issues the framework of spatial soil information service has also altered significantly *(Figure 1).*

Main issues of soil mapping

The goal of soil mapping is to reveal and visualize the spatial relationships of the thematic knowledge related to soil cover. Soil maps are thematic maps, where theme is determined by some specific information related to soils. This can be a primary or secondary (derived) soil property or class as well as any knowledge characterizing functions, processes or services of soils (Pásztor, L. *et al.* 2014).

The greatest and inevitable challenge of the compilation of soil maps is the regionalization of the local knowledge, its spatial inference (Várallyay, Gy. 2012). Reconnaissance of specific soil properties is carried out by sampling, which provides definitely point-like information. To create maps, the data related

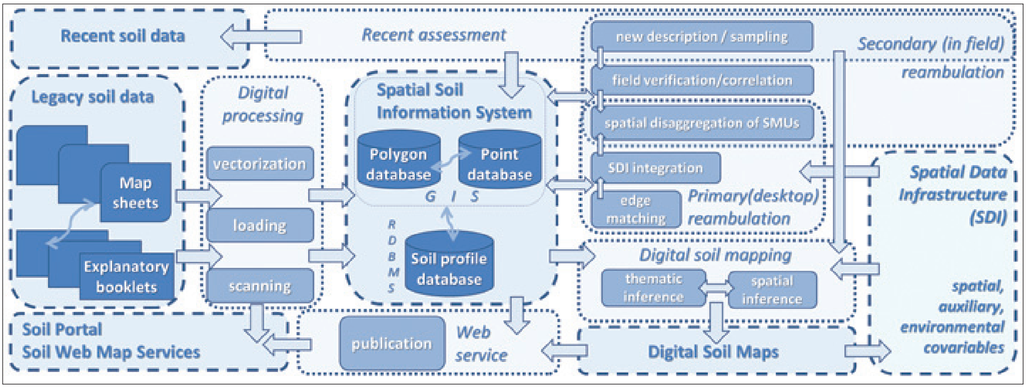

Fig. 1. Framework of spatial soil information services. Dashed line: information sources; dotted line: data flow/transportation between various elements

to locations should be spatially inferred using appropriate methods. From a certain point of view, the development of soil mapping is the conscious expansion of the repository of these methods: from mental space usage, along (base)map delimitation based on soil-landscape models till the various (mechanical, geometrical, geostatistical) interpolation methods and further until the introduction of ancillary, environmental, spatial data as auxiliary co-variables related to various components of soil forming processes.

Sampling based mapping is inherently predictive, the value or class of the mapped variable can only be estimated at unvisited locations (Gessler, P.E. *et al.* 1995; Scull, P. *et al.* 2003). Spatial prediction can be carried out (i) taking exclusively the mapped variable into consideration based on its spatial features; (ii) also based on the mapped variable, but the constraints of spatial validity are provided by further spatial, ancillary information; (iii) in every predicted locations supported by environmental, auxiliary co-variables (McKenzie, N.J. and Ryan, P.J. 1999).

Basically there are two, virtually conflicting but rather complementary conceptions for the description of the spatial heterogeneity of soils (Heuvelink, G.B.M. and Webster, R. 2001). One builds on similarity and is basically object based. It represents the soil cover with soil patches, which are either homogeneous mapping units or aggregates with estimated composition. The map realization of this concept is the traditional crisp soil map. According to the inherent model of these maps, at the given spatial resolution the soil properties within the mapping units are either homogeneous or heterogeneous but in cartographically unmappable way; and there is discontinuity in the mapped soil feature at the borders (Dobos, E. and Hengl, T. 2009; Szabó, J. *et al.* 2011).

The other approach emphasizes the continuous spatial variation of soil properties. The mapped soil property is predicted in cells and the spatial resolution is determined by the cell size (Mark, D.M. and Csillag, F. 1989). Raster data models of GIS provide ideal framework for this representation. It should be remarked, there are also compromised approaches between the two concepts (like certain fuzzy methods suitable for soil mapping; McBratney, A.B. and Odeh, I.O.A. 1997).

Each soil map can be characterized by three basic aspects which are more or less inter-related. A map displays a theme, a regionalized soil (or more generally soil related) property expressed by either quantitatively or qualitatively using categories (thematic issues). The map is compiled for a geographic region in a predefined scale, with some spatial resolution (geometrical issues). Finally the map has an overall and also spatially variable accuracy, purity, reliability (uncertainty issues). A demand on at least a tiny change in any of these issues theoretically induces the compilation of a new map with the required parameters. In traditional soil mapping the creation of a new map was troublesome and laborious. As a consequence robust maps were elaborated and rather the demands were fitted to the available map products.

Formation of digital soil mapping

A soil map is an object specific spatial model of the soil cover, whose compilation is dominated by the consideration of soil forming processes (Böhner, J. *et al.* 2002). There have been significant and essentially concurrent changes concerning three central elements of this definition. The growing and spread of digital soil mapping in the last decade can be attributed to the effects of these changes (Dobos, E. *et al.* 2006; Lagacherie, P. *et al.* 2007; Lagacherie, P. 2008; Boettinger, J.L. *et al.* 2010). Spatial and at the same time digital (that is GIS conform) information related to various segments of soil formation processes has become available in more and more quantity, with better and better spatial resolution and on lower and lower costs.

Mathematical (geo)statistical and data mining methods have been developed, which are efficiently applicable in the lack of deterministic models for the quantification of the some-times really complex and indirect relationships between soil features and the formerly mentioned, so called environmental auxiliary co-variables. Originally these methods were elaborated for the treatment of substantially different specialties, but they proved to be well adaptable in soil mapping, too.

Along the globalization processes the significant inhomogeneity in the knowledge of the world's soil cover has become evident. In one hand this has induced the compilation of relatively reliable soil maps based on limited soil data on the majority of the world, this way achieving at least a minimal coverage of these regions with spatial soil information. On the other hand it has initiated the elaboration of the principles of unification. The former surveys and mappings were carried out on national level based on independent methodologies, which caused disturbing effects in the mapping of the geographically continuously varying soil cover along administrative borders showing artificial disrupt changes.

The framework of DSM (McBratney, A.B. *et al.* 2003; Lagacherie, P. and McBratney, A.B. 2007; Hartemink, A.E. *et al.* 2008) involves spatial inference of the information collected at sampled points based on ancillary environmental variables related to soil forming processes *(Figure 2)*. DSM is formalized by the so called SCORPAN equation:

$$S_{property\ or\ class} = f\,(S, C, O, R, P, A, N),$$

where on the left side $\boldsymbol{S}_{property\ or\ class}$ is the (either numerical or categorical) mapped soil feature, while on the right side the predictive soil forming factors are ***C***limate, ***O***rganisms, ***R***elief, ***P***arent material, ***A***ge and ***G***eographic position. An original but well-established feature of the SCORPAN approach as opposed to Jenny's formula of soil formation (Jenny, H. 1941), that it also takes further ***S***oil related spatial information into consideration in the spatial prediction of a given soil variable. The most commonly used spatial auxiliary data layers are terrain attributes

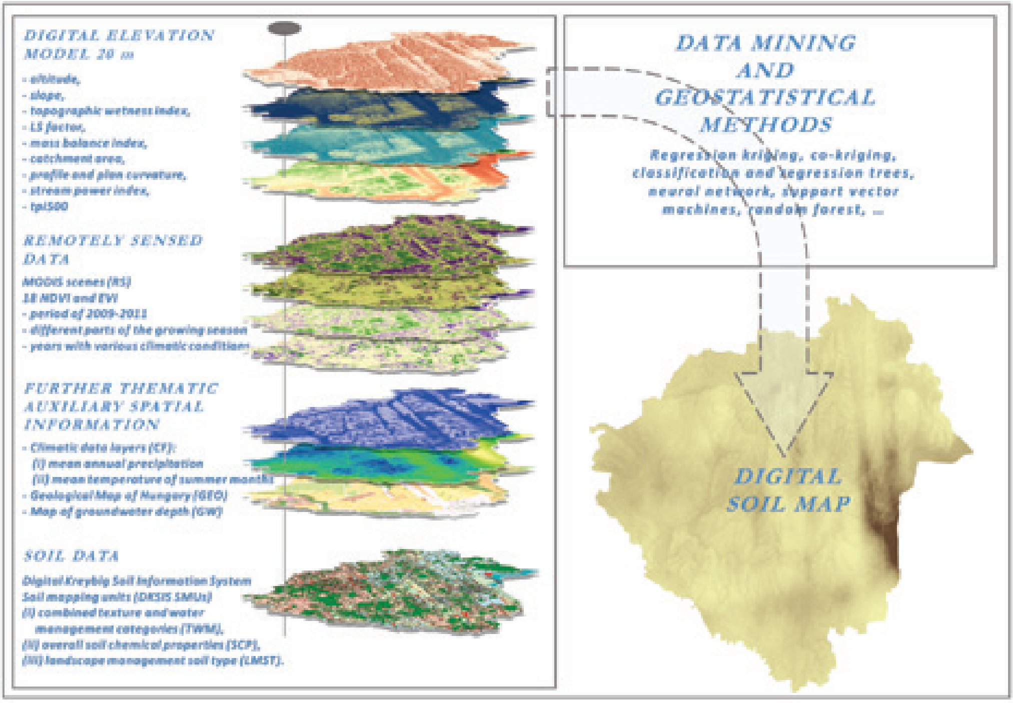
DIGITAL ELEVATION MODEL 20 m
- altitude,
- slope,
- topographic wetness index,
- LS factor,
- mass balance index,
- catchment area,
- profile and plan curvature,
- stream power index,
- tpi500
REMOTELY SENSED DATA
MODIS scenes (RS)
18 NDVI and EVI
- period of 2009-2011
- different parts of the growing season
- years with various climatic conditions
FURTHER THEMATIC AUXILIARY SPATIAL INFORMATION
- Climatic data layers (CF):
(i) mean annual precipitation
(ii) mean temperature of summer months
- Geological Map of Hungary (GEO)
- Map of groundwater depth (GW)
SOIL DATA
Digital Kreybig Soil Information System
Soil mapping units (DKSIS SMUs)
(i) combined texture and water management categories (TWM),
(ii) overall soil chemical properties (SCP),
(iii) landscape management soil type (LMST).
DATA MINING AND GEOSTATISTICAL METHODS
Regression kriging, co-kriging, classification and regression trees, neural network, support vector machines, random forest, ...
DIGITAL SOIL MAP

Fig. 2. Concept of digital soil mapping

derived from digital elevation models, and spectral reflectance bands from satellite imagery. Furthermore *f* refers to a specific function relating to mapped property with the actually used predictors, which may be realized in various forms.

Predictive mapping, taking exclusively the mapped variable into consideration, is numerically realized by spatial interpolation, which is supported by Tobler's First Law of Geography (Tobler, W. 1970). It states "Everything is related to everything else, but near things are more related than distant things" and it is basically an analogous formulation of the concept of spatial autocorrelation. The main feature of these types of methods, they operate in the geographical space. The branch of various interpolation methods dominated by stochastic modelling of environmental features is geostatistics.

If prediction is supported by environmental auxiliary co-variables, the quantification of the relationship between the mapped soil parameter and the ancillary data is the main challenge. Generalized classification, that is data mining methods proved to be suitable for the solution of these types of tasks. These methods investigate essentially the feature space, analysing its structure, thus unfolding the hidden and/or complex relationships. Regression and classification trees, random forests, neural networks, Bayesian belief network, support vector machines and some more techniques were successfully tested.

There are also compromised approaches between the concepts which concentrate purely on geographical or feature space. The two most widely used are co-kriging and regression kriging. In co-kriging a more densely sampled ancillary parameter supports the interpolation as opposed to ordinary kriging. In regression kriging the variation of the mapped variable is subdivided into two parts: the trend is estimated by MLRA and the residual of the explained part is then kriged (Hengl, T. *et al.* 2004).

Due to the simultaneous richness of spatial inference methods and the potentially available auxiliary environmental information (Grunwald, S. 2009; Hengl, T. 2009; Mulder, V.L. *et al.* 2011), there is a high versatility of possible approaches for the compilation of a given soil (related) map.

The framework of digital soil mapping also provides opportunity for the elaboration of goal specific soil maps, since the parameters characterizing the map product (thematic, resolution, accuracy, reliability etc.) may be predefined. The activity of DSM goes beyond mapping purely primary and secondary soil properties, the regionalization of further levels of soil related features (processes, functions and services) is also targeted (Minasny, B. *et al.* 2012).

Spatial soil information in Hungary

Hungary has long traditions in soil survey and mapping. Large amount of soil information is available in various dimensions and generally presented in maps, serving different purposes as to spatial and/or thematic aspects (Várallyay, Gy. 2012). Increasing proportion of soil related data has been digitally processed and organized into various spatial soil information systems (Pásztor, L. *et al.* 2013a).

The existing maps, data and systems served the society for many years, however the available data are no longer fully satisfactory for the recent needs of policy making. There were numerous initiatives for the digital processing, completion, improvement and integration of the existing soil datasets.

Presently soil data requirements are fulfilled with the recently available datasets either by their direct usage or after certain specific and generally fortuitous, thematic and/or spatial inference (Szabó, J. *et al.* 2007; Dobos, E. *et al.* 2010; Szatmári, G. *et al.* 2013; Sisák, I. and Benő, A. 2014; Waltner, I. *et al.* 2014). Due to the frequent discrepancies between the available and the expected data, notable imperfection may occur in the accuracy and reliability of the delivered products.

A recently started project (DOSoReMI.hu: Digital, Optimized, Soil Related Maps and

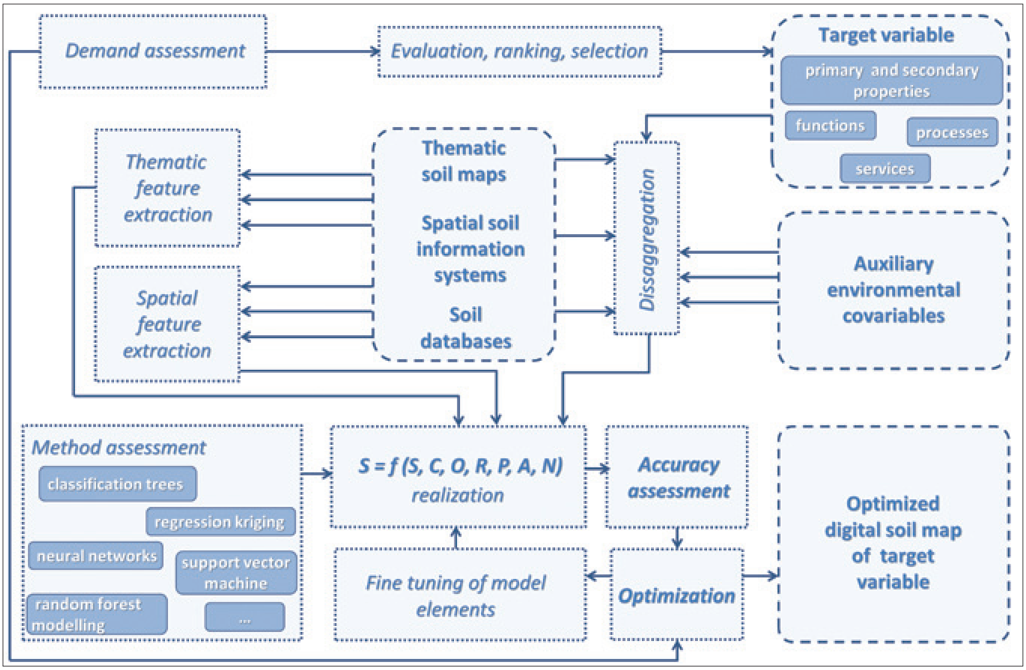
Demand assessment
Evaluation, ranking, selection
Target variable
primary and secondary properties
functions
processes
services
Thematic feature extraction
Thematic soil maps
Spatial soil information systems
Soil databases
Disaggregation
Spatial feature extraction
Auxiliary environmental covariables
Method assessment
classification trees
regression kriging
neural networks
support vector machine
random forest modelling
...
S = f (S, C, O, R, P, A, N) realization
Accuracy assessment
Fine tuning of model elements
Optimization
Optimized digital soil map of target variable

Fig. 3. Framework of the DOSoReMi.hu project

Information in Hungary; *Figure 3*) aims to significantly extend the potential how soil information requirements could be satisfied in Hungary.

In the frame of our project hitherto we have carried out spatial and thematic data mining of a significant amount of soil related information available in the form of legacy soil data as well as digital databases and spatial soil information systems (Pásztor, L. *et al.* 2013b, 2014).

In the course of the analyses auxiliary, spatial data themes related to soil forming factors as well as indicative environmental elements are relied on. Our objective is to compile digital soil related maps that optimally fulfil the national and international demands from points of view of thematic, spatial and temporal accuracy. In the following we shortly present some developments achieved so far in the frame of our activities.

Materials and methods

Digital mapping of soil properties in Zala County

Impact assessment of the forecasted climate change and the analysis of the possibilities of the adaptation in the agriculture and forestry can be supported by scenario based land management modelling, whose results can be incorporated in spatial planning. This framework requires adequate and spatially detailed knowledge of the soil cover. For the satisfaction of these demands in Zala County (3,784 km^2; Hungary), the soil conditions of the agricultural areas were digitally mapped based on the most detailed, available recent and legacy soil data. The agri-environmental conditions were characterized according to the 1:10,000 scale genetic soil mapping methodology and the category system applied in the Hungarian soil-agricultural chemistry practice. The factors constraining the fertility of soils were featured according to the biophysical criteria system elaborated for the delimitation of naturally handicapped areas in the EU. Production related soil functions were regionalized incorporating agro-meteorological modelling.

Various soil related information were mapped in three distinct sets: (i) basic soil properties determining agri-environmental conditions (soil type according to the Hungarian genetic classification, rootable depth, sand and clay content for the 1st and 2nd soil layers, pH, OM and carbonate content for the plough layer); (ii) biophysical criteria of natural handicaps defined by common European system and (iii) agro-meteorologically modelled yield values for different crops, meteorological and management scenarios. The applied method(s) for the spatial inference of specific themes was/were suitably selected: regression and classification trees for categorical data, indicator kriging for probabilistic management of criterion information; and typically regression kriging for quantitative data.

The appropriate derivatives of a 20 m digital elevation model were used in the analysis. Multitemporal MODIS products were selected from the period of 2009–2011 representing different parts of the growing season and years with various climatic conditions. Additionally two climatic data layers (mean annual precipitation and mean temperature of summer months), the 1:100,000 Geological Map of Hungary (Gyalog, L. and Síkhegyi, F. 2005) and the map of groundwater depth prepared by Water Research Institute (VITUKI, 2005) were used as auxiliary environmental co-variables.

Disaggregating category type soil maps

Numerous formerly elaborated thematic soil maps are not available in Hungary in the recently required scale. The original maps were compiled (i) in analogue environment and (ii) applying hardly identifiable soil-landscape models and unrecorded rules, so their reproducibility is problematic. Their theme, however, represents a widely used, embedded information source, which is expected to be (re)produced in larger scales. Various possibilities were studied for the solution of the problem. Decision trees proved to be adequate data mining technique to improve the spatial resolution of category-type soil maps disaggregating their soil mapping units (SMUs).

The agro-ecological units in the AGROTOPO (1994) database, compiled as a result of a substantial scientific synthesizing work (Várallyay, Gy. *et al.* 1985), were elaborated dominantly on the basis of mapping units originating from Kreybig soil maps (Kreybig, L. 1937), applying appropriate spatial and thematic generalization. Consequently, the Kreybig pattern contains significant and potentially utilizable information on the heterogeneity of these agroecological units, as do the elevation models characterizing the relief features.

Digital Kreybig Soil Information System is a countywide SSIS, which synthesizes the full soil information collected and processed during the Kreybig survey (Pásztor, L. *et al.* 2010, 2012). The readiness of AGROTOPO and DKSIS spatial soil information systems together with appropriate Digital Elevation Models and further environmental ancillary data available for the whole country has huge potential, which can be exploited in an integrated manner for the disaggregation of the thematic soil layers stored exclusively by AGROTOPO. The new maps display the same thematic but with increased spatial resolution and accuracy.

Compilation of country-wide physical soil property maps

The increasing demands on spatial soil information in order to support environmental related and land use management decisions vigorously concern physical soil properties, which also played important role in traditional soil mapping. Physical soil properties are directly related to water-holding capacity and nutrient supply, they affect water infiltration, runoff, and movement within the soil (Várallyay, Gy. 2011; Tóth, B. *et al.* 2014; Farkas, Cs. *et al.* 2014).

Soils can be characterized by different physical soil parameters; one of the most widely used is particle size distribution (Rajkai, K. and Kabos, S. 1999; Nemes, A. *et al.* 2011). Particles according to their size are categorized as clay, silt or sand. The size intervals are defined by national or international textural classification systems. The relative mass percentages of sand, silt, and clay in the soil constitute textural classes, which are also specified miscellaneously in various national and/or specialty systems. The most commonly used is the classification system of the United States Department of Agriculture (USDA 1987, Shirazi, M.A. and Boersma, L. 1984). Soil texture information classified according to USDA system is essential input data in (agri-)meteorological and hydrological modelling (Vereecken, H. *et al.* 1989; Saxton, K.E. and Rawls, W.J. 2006), which are also widely used in Hungary (Kozma, Zs. 2012; Ács, F. *et al.* 2014; Fodor, N. *et al.* 2014).

Our work for producing the very first texture class map according to USDA classification for Hungary has been recently presented in detail by Laborczi, A. *et al.* (2015). In addition to texture, particle size fractions (clay, silt and sand content) are also important in themselves. They are mandatory variables of the GlobalSoilMap data structure according to its Specifications (2014) as well as main indicators used by biophysical criteria to define natural constraints for agriculture in Europe (Van Orshoven, J. *et al.* 2013). Firstly clay, silt and sand content were independently predicted spatially using regression kriging with a predefined, 150 meter spatial resolution. Reference data have originated from the Hungarian Soil Information and Monitoring System. Auxiliary spatial information was represented by digital elevation model and its derived components, geological, climatic, landuse maps and last but not least the physical property SMU layer of DKSIS.

The applied geostatistical and data mining tools

In the framework of DSM numerous digital mapping methods have been elaborated that apply on of or integrate geostatistical and data mining tools (Goovaerts, P. 2000; Hengl, T. 2009; Moran, C.J. and Bui, E.N. 2002; Lagacherie, P. *et al.* 2007; Boettinger, J.L. *et al.* 2010; Hartemink, A.E. *et al.* 2008). Here only three of them are shortly discussed, which were used in the works presented in this paper.

Regression kriging (RK) is a spatial prediction technique, which jointly employs correlation with auxiliary maps and spatial correlation. It is widely used for the spatial inference of quantitative soil properties (e.g. Hengl, T. *et al.* 2004; Illés, G. *et al.* 2011; Szatmári, G. and Barta, K. 2013). Similarly to other DSM methods, RK is based on the application of auxiliary environmental variables (derivatives of digital elevation model (DEM), remotely sensed images, etc.), which can be widely interpreted, that is spatial information on independent soil features can also be involved. In RK firstly correlation of the environmental factors and the predicted variable is determined by MLRA. Then kriging of the residuals provides the stochastic factor which is added to the regression result thus producing the final map. Essentially, RK respects the fact; neither environmental correlation nor geostatistical interpolation alone is able to account for the whole spatial variation that is to produce map products with satisfactory accuracy. They can be used as complementary spatial inference approaches where one can improve the other's drawbacks.

Indicator kriging (IK) is a nonparametric interpolation method without any assumption on concerning the distribution of the modelled variables providing estimation of probability. Based on these features IK is a useful tool for the spatial inference of categorical variables. Regionalization of specific simple and/or simplified secondary and functional soil (related) data can be supported by IK.

IK was heavily based on in the process of the delineation of areas affected by natural constrains in Hungary defined by common European biophysical criteria related to soil. The elaborated European system consists of detailed definitions, justification and associated critical limits or threshold values for each biophysical criterion. The fulfilment of a specific criterion had to be regionalized, that is the final product should have to be a binary map displaying yes/no categories. Decisions carried out on soil profile resulted in binary (indicator) form, which had to be spatially extended. As a consequence, indicator kriging proved to be proper approach. It provided probability (spatial) distribution maps, indicating the probability of fulfilling the criteria within the block used for the calculation.

Classification and regression trees (CART) are also widely applied in Digital Soil Mapping too (Moran, C.J. and Bui, E.N. 2002; Scull, P. *et al.* 2005; Bou Kheir, R. *et al.* 2010; Giasson, E. *et al.* 2011; Greve, M.H. *et al.* 2012), due to their manifold advantages.

CART is one of the most successful, widespread and efficient data mining techniques for supervised classification learning, which builds complex relations by a sequence of simple decisions. The tree is built from a training database by recursive method. The decision rules can be easily interpreted; they start with a single node, and then look for the binary distinction which gives the most information on the classification. At each node the conditions (based on homogeneity indices) split the reference data into two child nodes. Each of the resulting new nodes is taken and the process is repeated continuing the recursion until reaching certain predefined stopping criterion. CART is easy to interpret and discuss, when a mix of continuous and categorical type environmental parameters are used as predictors, furthermore, they have excellent predictive capabilities (Breiman, L. 2001; Lawrence, R. *et al.* 2004; Henderson, B.L. *et al.* 2005).

CART can be applied for the understanding of the soil-landscape models involved in existing soil maps, and for the post-formalization of survey/compilation rules. The relationships identified and expressed in decision rules make the creation of spatially refined maps possible with the aid of high resolution environmental auxiliary variables. Among these co-variables, a special role could be played by larger scale soil information with diverse attributes.

Results

Digital mapping of soil properties in Zala County

Some results of our activities for the mapping of soil properties in Zala County are discussed in recent papers. Szatmári, G. *et al.* (2013) debate in details the experiences gathered during the application of RK in spatial inference of quantitative soil properties. The two main findings, which we also want to emphasize here, referred to the application of ancillary data. In one hand, usage of various co-variables may result in remarkable differences of the final map. On the other hand inclusion of spatial soil data significantly improves the performance of RK. Illés, G. *et al.* (2014) have presented the elaboration of a unified large scale soil type map according to the Hungarian genetic soil classification system. The reference soil data originated from various legacy datasets with differing density provided for areas with different land use. Some further, formerly unpublished soil property maps compiled for the agricultural areas of the country are presented in *Figure 4*.

The results of the agro-meteorological modelling for the regionalization of production related soil functions are scheduled to be published soon.

Disaggregating category type soil maps

The disaggregation of categorical soil maps with the aid of auxiliary spatial soil information was successfully applied in cases with different thematic and spatial extent. Some results have been recently presented in detail by Pásztor, L. *et al.* (2013b). The most useful product has been the spatially refined

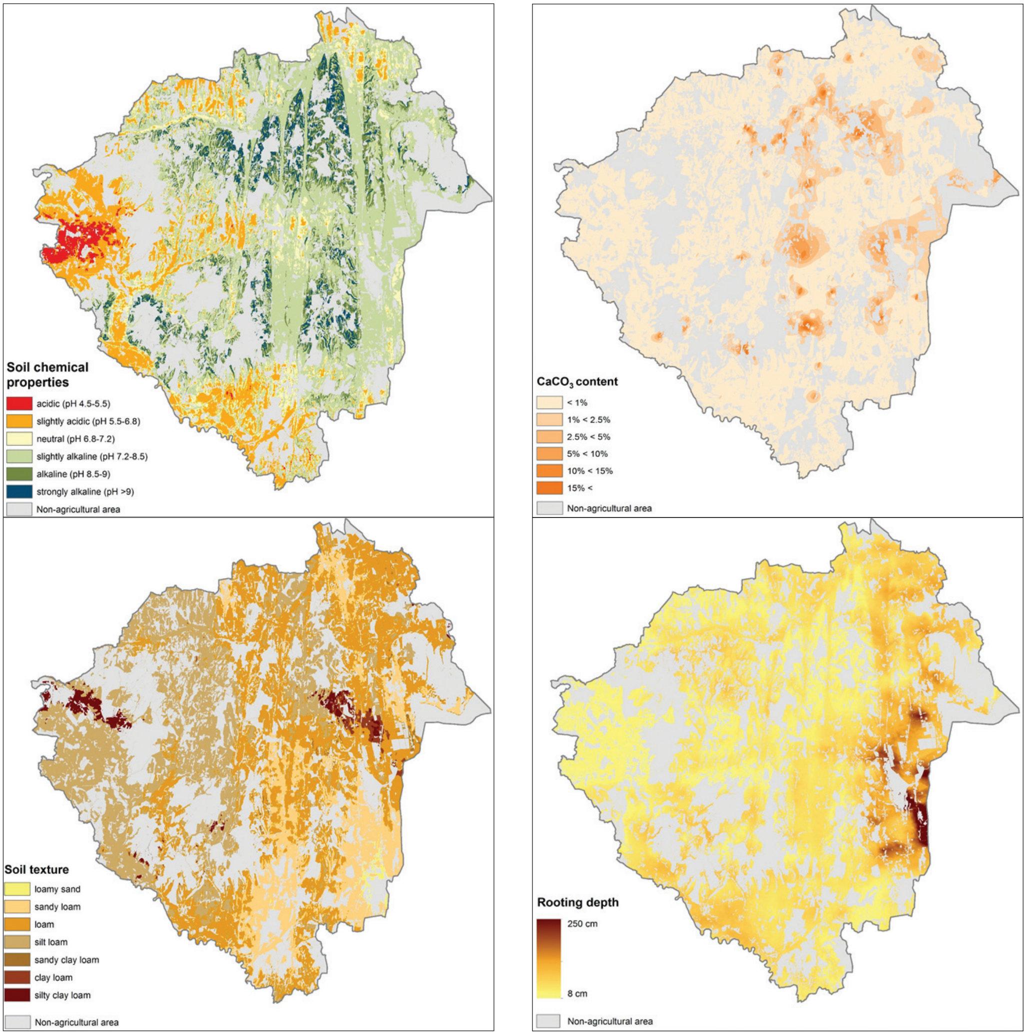

Fig. 4. Basic soil property maps of Zala County

nationwide soil productivity map, which was operationally used for the delineation of Areas with Excellent Productivity in the framework of the National Regional Development Plan.

The other challenge has been the characterization of the soil cover in terms of genetic soil types at a scale of 1:50,000–1:25,000, which is required due to various purposes, like the digital implementation of large-scale mapping methodology designed for irrigation planning or spatial planning activities. For the fulfilling of these demands the genetic soil type layer of AGROTOPO was disaggregated for pilot areas based on DKSIS, environmental auxiliary variables and using decision trees.

The downscaled soil type map according to the Hungarian soil classification system compiled for the Danube–Tisza Interfluve is presented in *Figure 5*. The map was compiled with the aid of decision trees using

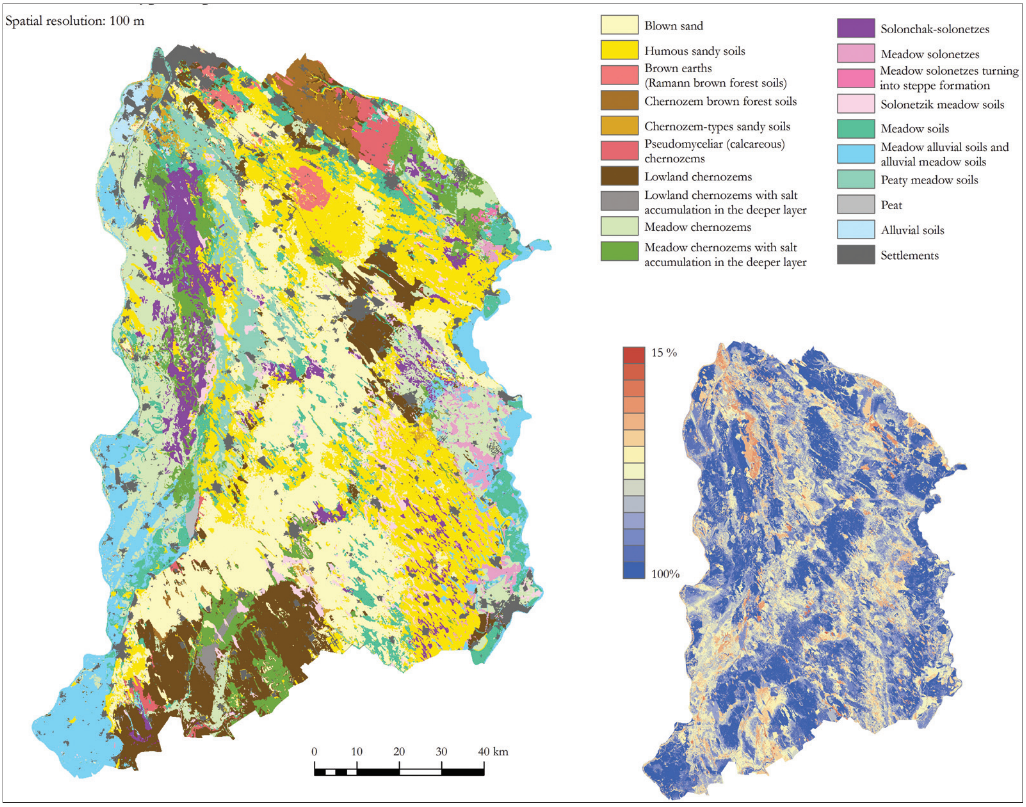
Spatial resolution: 100 m
Blown sand
Humous sandy soils
Brown earths (Ramann brown forest soils)
Chernozem brown forest soils
Chernozem-types sandy soils
Pseudomyceliar (calcareous) chernozems
Lowland chernozems
Lowland chernozems with salt accumulation in the deeper layer
Meadow chernozems
Meadow chernozems with salt accumulation in the deeper layer
Solonchak-solonetzes
Meadow solonetzes
Meadow solonetzes turning into steppe formation
Solonetzik meadow soils
Meadow soils
Meadow alluvial soils and alluvial meadow soils
Peaty meadow soils
Peat
Alluvial soils
Settlements
15 %
100%
0
10
20
30
40 km

Fig. 5. Predicted soil type map for the Danube–Tisza Interfluve (on the left) and estimated reliability of the spatial prediction (on the right)

non-soil environmental auxiliary variables together with the SMUs of DKSIS. Virtual reference point sets were created by multiple point sampling. One point per km^2 was randomized in the geographical space, representing virtual sampling locations.

Conditional generalization was applied, prescribing a minimum spacing of 100 meters (equal to the cell size applied in spatial modelling) between the generated points. Randomized points closer than 100 meters to SMU borders were eliminated to avoid transition zones between neighbouring soil types. The values of the dependent (predicted) and independent (predictor) variables were identified at the randomized locations and their records were used in data mining classification. The rules established during the building of the classification tree were applied to the spatial layers as operations providing soil type prediction for the whole area of interest. The randomization process was repeated 100 times providing 100 classification results for each cell. The final categorization was done by maximum-likelihood decision; the most frequent class was attributed to the cell as most likely soil type. The vagueness of the classification is also inferred by the occurrence value of the most frequent class. *Figure 5* contains an inset map, which displays the reliability of the spatial prediction expressed this way.

Compilation of country-wide physical soil property maps

Maps of particle size classes (clay, silt and sand content) have been multi-mapped, that is the same target variable has been predicted in various ways. Either the method or the predictor variables have been selected in differing way. The purpose has been to identify the best performing constellation. The performance of the various approaches can be identified by proper validation and in our case it has been measured by three validation parameters: mean error (ME); mean absolute error (MAE) and root mean square error (RMSE).

Validation was based on particle size distribution data provided by the Hungarian Detailed Soil Hydrophysical Database (MARTHA, Makó, A. *et al.* 2010). MARTHA has been developed to collect information on measured soil hydraulic and physical characteristics in Hungary.

Recently this is the largest and most detailed national hydrophysical database. 780 records were used having both georeferenced location and topsoil particle size distribution data entries. The results for two independent approaches (both using RK, but with different ancillary dataset) are presented in *Table 1*.

Table 1. Validation of particle size fraction maps by various parameters for two different approaches

Para-meters	RK1			RK2		
	clay	silt	sand	clay	silt	sand
ME	-2.77	0.05	2.72	-2.50	-0.13	2.62
MAE	7.63	11.20	13.49	7.72	11.03	13.36
RMSE	10.42	15.28	18.38	10.50	15.10	18.19

According to the figures of the table, the two independent approaches result in slightly different maps as for their performance. The final map products of sand and clay fraction for the uppermost (0–5 cm) soil layer are presented in *Figure 6*.

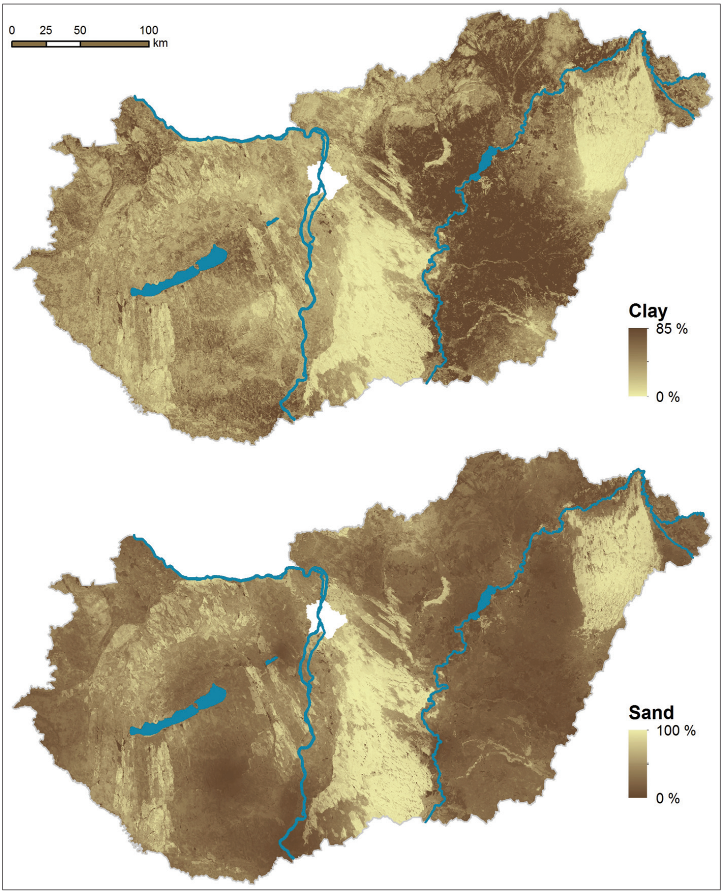

Fig. 6. Countrywide maps of two main particle size fractions (sand, clay) for the uppermost soil layer (0–5 cm)

Conclusions

In the frame of DOSoReMI.hu project a significant amount of experiences have been accumulated so far concerning the compilation of novel and renewed, goal oriented, digital soil maps using geostatistical and data mining tools either at national or regional level.

By the aid of selected and optimized geostatistical, data mining and GIS tools some basic soil properties were multi-mapped and optimized, but it is also planned to conduct the spatial extension of certain more sophisticated pedological variables featuring the state, processes (including degradation), functions and services of soils.

It is hoped to achieve further progress in the performance by expanding the pool of environmental co-variables applied and by testing additional methods (random forests, neural networks, support vector machines, etc.). The environmental correlation used in RK expressed by MLRA is also suggested to be substituted by further, knowledge based data mining methods for improving the modelling of the complex relationship between a specific soil variable and its affecting/determining/indicating factors. The first planned step is the substitution of MLRA with Regression Tree Analysis to generalize the linear model between the predicted and the explanatory environmental variables.

The country-wide physical soil property maps elaborated according GlobalSoilMap specifications will represent the first Hungarian contribution to the GlobalSoilMap.net project (Minasny, B. and McBratney, A.B. 2010). Further GSM.net conform map products are also under development in the frame of DOSoReMI.hu to contribute to the worldwide activities. Based on more extended data infrastructure, it is suggested that national initiatives could produce more accurate and reliable products.

The application of an inset map for the expression of the inherent vagueness of spatial prediction is a rather recently introduced, but we propose its general usage for displaying the results of digital maps elaborated using geo-mathematical methods and/or environmental models.

***Acknowledgements*:** Our work has been supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167). Authors thank J. Matus for her indispensable contribution.

REFERENCES

Ács, F., Gyöngyösi, A.Z., Breuer, H., Horváth, Á., Mona, T. and Rajkai, K. 2014. Sensitivity of WRF-simulated planetary boundary layer height to land cover and soil changes. *Meteorologische Zeitschrift*, PrePub DOI 10.1127/0941-2948/2014/0544

AGROTOPO 1994. *AGROTOPO database of RISSAC*. Budapest, RISSAC HAS, http://maps.rissac.hu/agrotopo_en

Baumgardner, M.F. 2011. Soil databases. In *Handbook of Soil Sciences: Resource Management and Environmental Impacts*. Eds.: Huang, P.M., Li, Y. and Sumner, M.E. Boca Raton, CRC Press, 21–35.

Blum, W.E.H. 2005. Functions of soil for society and the environment. *Reviews in Environmental Science and Biotechnology* 4. 75–79.

Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E. and Kienast-Brown, S. Eds. 2010. *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Heidelberg, Springer, 473 p.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A. and Selige, T. 2002. Soil regionalisation by means of terrain analysis and process parameterisation. In *Soil Classification 2001*. Eds.: Michéli, E., Nachtergaele, F. and Montanarella, L. EUR 20398 EN. The European Soil Bureau, Joint Research Centre, Ispra, 213–222.

Bou Kheir, R., Bøcher, P.K., Greve, M.B. and 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. *Hydrology and Earth System Sciences* 14. 847–857.

Breiman, L. 2001. Decision-tree forests. *Machine Learning* 45. (1): 5–32.

Bullock, P. 1999. Soil resources of Europe – An overview. In *Soil Resources of Europe*. Eds.: Bullock, P., Jones, R.J.A. and Montanarella, L. European Soil Bureau Research Report 6. Luxembourg, Office for Official Publications of the European Communities, 15–25.

Dobos, E. and Hengl, T. 2009. Soil mapping applications. In *Geomorphometry – Concepts, Software, Applications*. Eds.: Hengl, T. and Reuter, H.I. Development in Soil Science series 33. 461–479.

Dobos, E., Bialkó, T., Micheli, E. and Kobza, J. 2010. Legacy soil data harmonization and database development. In *Digital Soil Mapping Bridging Research Environmental Application, and Operation*: Eds.: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E. and Kienast-Brown, S. Heidelberg, Springer, 309–323.

Dobos, E., Carré, F., Hengl, T., Reuter, H.I. and Tóth, G. Eds.. 2006. *Digital soil mapping as a support to production of functional maps*. EUR 22123 EN. Luxembourg, Office for Official Publications of the European Communities, 68 p.

Farkas, Cs., Gelybó, Gy., Bakacsi, Zs., Horel, Á., Hagyó, A., Dobor, L., Kása, I. and Tóth, E. 2014. Impact of expected climate change on soil water regime under different vegetation conditions. *Biologia*, Manuscript, accepted for publication.

Fodor, N., Pásztor, L. and Németh, T. 2014. Coupling the 4M crop model with national geo-databases for assessing the effects of climate change on agro-ecological characteristics of Hungary. *International Journal of Digital Earth* 7. (5): 391–410.

Gessler, P.E., Moore, I.D., McKenzie, N.J. and Ryan, P.J. 1995. Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems* 9. (4): 421–432.

Giasson, E. Sarmento, E.C., Weber, E., Flores, C.A. and Hasenack, H. 2011. Decision trees for digital soil mapping on subtropical basaltic steeplands. *Scienta Agricola* 68. (2): 167–174.

Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228. (1–2), 113–129.

Greve, M.H., Kheir, R.B., Greve, M.B. and 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. *Ecological Indicators* 18. 1–10.

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modelling approaches. *Geoderma* 152. 195–207.

Gyalog, L. and Síkhegyi, F. Eds. 2005. Magyarország geológiai térképe, 1:100 000 (Geological Map of Hungary, 1:100 000), Budapest, Geological Institute of Hungary, Digital version: http://loczy.mfgi.hu/fdt100/

Hartemink, A.E., Mcbratney, A.B. and Mendonça-Santos, M.De L. Eds. 2008. *Digital Soil Mapping with Limited Data*. Springer, The Netherlands, 445 p.

Henderson, B.L., Bui, E.N., Moran, C.J. and Simon, D.A.P. 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma* 124. (3–4): 383–398.

Hengl, T., Heuvelink, G. and Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 122. (1–2), 75–93.

Hengl,T. 2009. *A Practical Guide to Geostatistical Mapping*. Amsterdam, University of Amsterdam, 291 p.

Heuvelink, G.B.M. and Webster, R. 2001. Modelling soil variation: past, present, future. *Geoderma* 100. 269–301.

Illés, G., Kovács, G. and Heil, B. 2011. Comparing and evaluating digital soil mapping methods in a Hungarian forest reserve. *Canadian Journal of Soil Science* 91. (4): 615–626.

Illés, G., Kovács, G., Laborczi A. and Pásztor L. 2014. Zala megye egységes talajtípus adatbázisának összeállítása (Developing a unified soil type database for Zala County using classification algorithms). *Erdészettudományi Közlemények*. (in press)

Jenny, H. 1941. *Factors of Soil Formation*. New York, McGraw-Hill, 281 p.

Kozma, Zs., Derts, Zs., Kardos, M. and Koncsos, L. 2012. A mezőgazdasági termelés mint ökoszisztéma-szolgáltatás értéke: hidrológiai modellhez kapcsolt számítási módszertan (The value of agricultural crops as an ecosystem service: calculation methodology connected to a hydrological Model) *Tájökológiai Lapok* 10. (1): 55–69.

Kreybig, L. 1937. A Magyar Királyi Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere (The survey, analytical and mapping method of the Hungarian Royal Institute of Geology). *Magyar Királyi Földtani Intézet Évkönyve* 31. 147–244.

Laborczi, A., Szatmári, G., Takács, K. and Pásztor, L. 2015. Topsoil texture class map of Hungary compiled using classification trees. *Journal of Maps* (submitted paper)

Lagacherie P. 2008. Digital soil mapping: A state of art. In *Digital Soil Mapping with Limited Data*. Eds.: Hartemink, A.E., Mcbratney, A.B. and Mendonça-Santos, M.De L. Dordrecht, Springer, 3–14.

Lagacherie, P. and Mcbratney, A.B. 2007. Spatial Soil Information Systems and Spatial Soil Inference Systems: perspectives for digital soil mapping. In *Digital soil mapping: an introductory perspective.* Eds.: Lagacherie, P., Mcbratney, A.B. and Voltz, M., Amsterdam, Elsevier, 3–22.

Lagacherie, P., Mcbratney, A.B. and Voltz, M. Eds. 2007. *Digital soil mapping: an introductory perspective.* Amsterdam, Elsevier, 658 p.

Lawrence, R., Bunn, A., Powell, S. and Zambon, M. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of the Environment* 90. 331–336.

Makó A., Tóth, B., Hernádi, H., Farkas, Cs. and Marth, P. 2010. Introduction of the Hungarian Detailed Soil Hydrophysical Database (MARTHA) and its use to test external pedotransfer functions. *Agrokémia és Talajtan* 59. (1): 29–38.

Mark, D.M. and Csillag, F. 1989. The nature of boundaries on 'area-class' maps. *Cartographica* 26. (1): 65–78.

McBratney, A.B. and Odeh, I.O.A. 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma* 77. 85–113.

McBratney, A.B., Mendonça-Santos, M.L. and Minasny, B. 2003. On digital soil mapping. *Geoderma* 117. 3–52.

McKenzie, N.J. and Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89. (1–2): 67–94.

Mermut, A.R. and Eswaran, H. 2000. Some major developments in soil science since the mid-1960s. *Geoderma* 100. 403–426.

Minasny, B. and McBratney, A.B. 2010. Methodologies for Global Soil Mapping, Dordrecht, Springer Netherlands, 429–436.

Minasny, B., Malone, B.P. and McBratney, A.B. Eds. 2012. *Digital Soil Assessments and Beyond*. London, Taylor and Francis Group, 466 p.

Montanarella, L. 2010. Need for interpreted soil information for policy making. In *19th World Congress of Soil Science, Soil Solutions for a Changing World*, 1–6 August 2010, Brisbane, Australia. Published on DVD.

Moran, C.J. and Bui, E.N. 2002. Spatial data mining for enhanced soil map modelling. *International Journal of Geographic Information Science* 16. 533–549.

Mulder, V.L., de Bruin, S., Schaepman, M.E. and Mayr, T.R. 2011. The use of remote sensing in soil and terrain mapping. – A review. *Geoderma* 162. 1–19.

Nemes, A., Pachepsky, Y.A. and Timlin, D.J. 2011. Toward improving global estimates of field soil water capacity. *Soil Science Society of America Journal* 75. (3): 807–812.

Omuto, C., Nachtergaele, F. and Rojas, R.V. 2013. *State of the Art Report on Global and Regional Soil Information: Where are we? Where to go?* Global Soil Partnership Technical Report. Rome, FAO, 69 p.

Panagos, P., van Liedekerke, M., Jones, A. and Montanarella, L. 2012. European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy* 29. 329–338.

Pásztor, L., Bakacsi, Zs., Laborczi, A. and Szabó, J. 2013b. Kategória típusú talajtérképek térbeli felbontásának javítása kiegészítő talajtani adatok és adatbányászati módszerek segítségével (Downscaling of categorical soil maps with the aid of auxiliary spatial soil information and data mining methods). *Agrokémia és Talajtan* 62. (1): 205–218.

Pásztor, L., Dobos, E., Szatmári, G., Laborczi, A., Takács, K., Bakacsi, Zs. and Szabó, J. 2014. Application of legacy soil data in digital soil mapping for the elaboration of novel, countrywide maps of soil conditions. *Agrokémia és Talajtan* 63. (1): 79–88.

Pásztor, L., Szabó, J. and Bakacsi, Zs. 2010. Digital processing and upgrading of legacy data collected during the 1:25 000 scale Kreybig soil survey. *Acta Geodaetica et Geophysica Hungarica* 45. 127–136.

Pásztor, L., Szabó, J., Bakacsi, Zs. and Laborczi, A. 2013a. Elaboration and applications of spatial soil information systems and digital soil mapping at the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. *Geocarto International* 28. (1): 13–27.

Pásztor, L., Szabó, J., Bakacsi, Zs., Matus, J. and Laborczi, A. 2012. Compilation of 1:50 000 scale digital soil maps for Hungary based on the digital Kreybig soil information system. *Journal of Maps* 8. (3): 215–219.

Rajkai, K. and Kabos, S. 1999. A talaj víztartóképesség-függvény (pF-görbe) talajtulajdonságok alapján történő becslésének továbbfejlesztése (Estimation of Soil Water Retention Characteristics (pF Curves) From Other Soil Properties. *Agrokémia és Talajtan* 48. (1–2): 15–32.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A. and Zhang, G.L. 2009. Digital soil map of the world. *Science* 325. 680–681.

Saxton, K.E. and Rawls, W.J. 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal* 70. (5): 1569–1578.

Scull, P., Franklin, J. and Chadwick, O.A. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling* 181. 1–15.

Scull, P., Franklin, J., Chadwick, O.A. and McArthur, D. 2003. Predictive soil mapping: a review. *Progress in Physical Geography* 27. (2): 171–197.

Shirazi, M.A. and Boersma, L. 1984. A Unifying Quantitative Analysis of Soil Texture. *Soil Science Society of America Journal* 48. (1): 142–147.

Sisák, I. and Benő, A. 2014. Probability-based harmonization of digital maps to produce conceptual soil maps. *Agrokémia és Talajtan* 63.(1): 89–98.

Specifications Version 1 GlobalSoilMap.net products, Release 2.1, (2014) <http://www.globalsoilmap.net/specifications>.

Szabó, J., Pásztor, L. and Bakacsi, Zs. 2011. Demand, feasibility and construction stages of a national spatial soil information system. *Agrokémia és Talajtan* 60. (Suppl.), 149–160.

Szabó, J., Pásztor, L., Bakacsi, Zs., László, P. and Laborczi, A. 2007. A Kreybig Digitális Talajinformációs Rendszer alkalmazása térségi szintű földhasználati kérdések megoldásában (Application of the Kreybig Digital Soil Information System to solve land use problems at regional level). *Agrokémia és Talajtan* 56. (1): 5–20.

Szatmári, G. and Barta, K. 2013. Csernozjom talajok szervesanyag-tartalmának digitális térképezése erózióval veszélyeztetett mezőföldi területen (Digital mapping of the organic matter content of chernozem soils on an area endangered by erosion in the Mezőföld region). *Agrokémia és Talajtan* 62. (1): 47–60.

Szatmári, G., Laborczi, A., Illés, G. and Pásztor, L. 2013. A talajok szervesanyag-készletének nagyléptékű térképezése regresszió krigeléssel Zala megye példáján (Large-scale mapping of soil organic matter content by regression kriging in Zala County). *Agrokémia és Talajtan* 62. (2): 219–234.

Tobler,W. 1970. A computer movie simulating urban growth in the Detroit region, *Economic Geography* 46. (2): 234–240.

Tóth, B., Makó, A. and Tóth, G. 2014. Role of soil properties in water retention characteristics of main Hungarian soil types. *Journal of Central European Agriculture* 15. (2): 137–153.

Tóth, G., Montanarella, L., Stolbovoy, V., Máté, F., Bódis, K., Jones, A., Panagos, P. and van Liedekerke, M. 2008. *Soils of the European Union*. EUR 23439 EN. Luxembourg, Office for Official Publications of the European Communities, 85 p.

USDA 1987. *Soil Mechanics Level I. Module 3 – USDA Textural Soil Classification Study Guide.* National Employee Development Staff, Soil Conservation Service, United States Department of Agriculture, USA, 232 p.

Van Orshoven, J., Terres, J-M. and Tóth. T. 2013. *Updated common bio-physical criteria to define natural constraints for agriculture in Europe*. Definition and scientific justification for the common criteria. Technical Factsheets. Luxembourg: Office for Official Publications of the European Communities 2012. Scientific and Technical Research, EUR 25203 EN. Joint Research Centre, Institute for Environment and Sustainability.

Várallyay, Gy. 2011. Water storage capacity of Hungarian soils. *Agrokémia és Talajtan* 60. (Suppl.), 7–26.

Várallyay, Gy. 2012. Talajtérképezés, talajtani adatbázisok. *Agrokémia és Talajtan* 61. (Suppl.), 249–267.

Várallyay, Gy., Szűcs, L., Zilahy, P., Rajkai, K. and Murányi, A. 1985. Soil factors determining the agro-ecological potential of Hungary. *Agrokémia és Talajtan* 34. (Suppl.), 90–94.

Vereecken, H., Maes, J., Feyen, J. and Darius, P. 1989. Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. *Soil Science* 148. 389–403.

Waltner, I., Micheli, E., Fuchs, M., Láng, V., Pásztor, L., Bakacsi, Zs., Laborczi, A. and Szabó, J. 2014. Digital mapping of selected WRB units based on vast and diverse legacy data. In *Global Soil Map: Basis of the Global Spatial Soil Information System*. Eds.: Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A. and McBratney, A.B. London, Taylor and Francis Group, 313–318.