Elaboration of novel, countrywide maps for the satisfaction of recent demands on spatial, soil related information in Hungary

L. Pásztor, J. Szabó, Zs. Bakacsi & A. Laborczi
Institute for Soil Science and Agricultural Chemistry, Centre for Agricultural Research, Budapest, Hungary

E. Dobos
University of Miskolc, Miskolc, Hungary

G. Illés
Hungarian Forest Research Institute, Budapest, Hungary

G. Szatmári
University of Szeged, Szeged, Hungary

ABSTRACT: The main objective of the DOSoReMI.hu (Digital, Optimized, Soil Related Maps and Information in Hungary) project is to significantly extend the potential, how demands on spatial soil related information could be satisfied in Hungary. Although a great amount of soil information is available due to former mappings and surveys, there are more and more frequently emerging discrepancies between the available and the expected data. The gaps are planned to be filled with optimized DSM products heavily based on legacy soil data, which still represent a valuable treasure of soil information at the present time. Our paper presents the first pilot results achieved for Zala County, Hungary: the identified effects of DSM components on the accuracy of a specific output map; together with some modest proposals for their applicability in the optimization of the whole mapping process.

1 1 INTRODUCTION

There is a heap of evidences that demands on soil related information have been significant worldwide and it is still increasing (Bullock, 1999; Mermut & Eswaran, 2000; Tóth et al., 2008, Sanchez, et al., 2009; Baumgardner, 2011). Soil maps were typically used for long time to satisfy these requests. 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 et al., 2002). The demands often do not refer to primary or even secondary soil properties but to various processes, functions, services and/or systems related to soils (Omuto et al., 2013). Due to the relatively high costs of new data collection and by the spread of GI technology, spatial soil information systems (SSIS) and digital soil mapping (DSM) took the role of traditional soil maps in the field of data service. Legacy soil data is still heavily relied on, since they contain a wealth of information that can be exploited by proper methodology in GIS/SSIS/DSM environment.

Not only the degree of current needs for soil information has changed but also its nature. Traditionally the agricultural functions of soils were focused on, which was also reflected in the methodology of data collection and mapping. Recently the multifuntionality of soils is getting to gain more ground (Blum, 2005); consequently information related to additional functions of soils becomes identically important. The new types of information requirements however cannot be fulfilled generally with new data collections at least not on such a level as it was done in the frame of traditional soil surveys (Montanarella, 2010).

In Hungary, 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 (Pásztor et al. 2013). 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 (DOSoReMI.hu; Digital, Optimized, Soil Related Maps and Information in Hungary) we would like to significantly extend the potential, how countrywide soil information requirements could be satisfied in Hungary. Primarily the national demands are intended to be treated, which in one hand cannot be considered independent of the GSM.net objectives and on the other hand similar or identical specifications are planned to be applied.

In the frame of our project (Fig. 1) we plan the execution of spatial and thematic data mining of significant amount of soil related information available in the form of legacy soil data as well as digital databases and spatial soil information systems. We plan to compile digital soil related maps of certain pedological variables featuring the state, processes, functions and services of soils, which fulfill optimally the national and international demands from points of view of thematic, spatial and temporal accuracy.

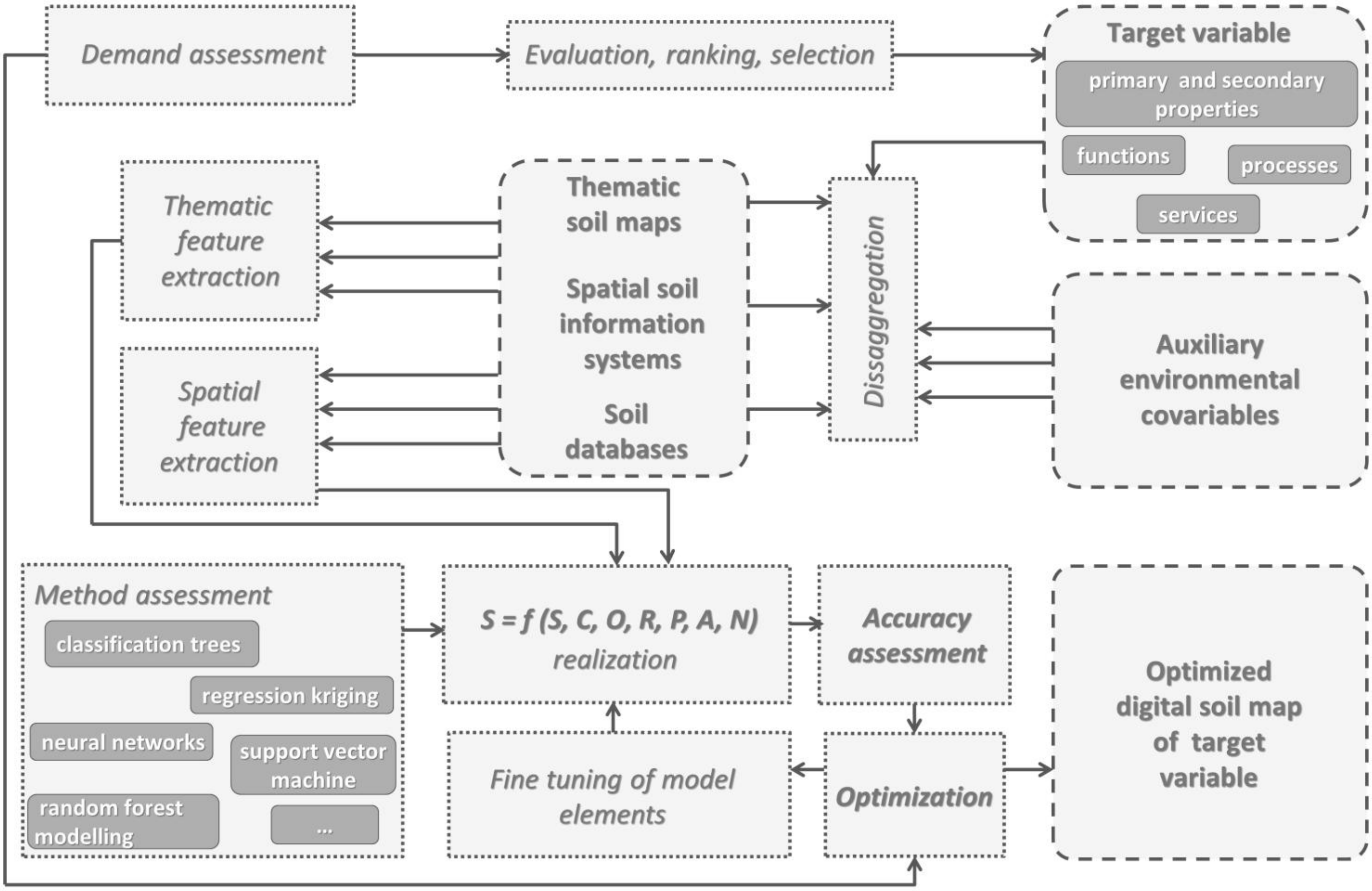

Figure 1. Framework of the DOSoReMi.hu project

Due to the simultaneous richness of available Hungarian legacy soil data (Várallyay, 2002), spatial inference methods and auxiliary environmental information (Grunwald, 2009; Hengl 2009; Mulder et al., 2011), there is a high versatility of possible approaches for the compilation of a given soil (related) map. This suggests the opportunity of optimization. For the creation of an object specific soil (related) map with predefined parameters (resolution, accuracy, reliability etc.) one might intend to identify the optimum set of soil data, method and auxiliary covariables optimized for the resources (data costs, computation requirements etc.). Prior to the management of countrywide challenges we started comprehensive analysis of the effects of the various DSM components on the accuracy of the output maps on pilot areas.

In the case of the mapping of a specific soil property, various, spatial, environmental correlation methods have been tested with varying set of reference and training soil data, while auxiliary covariables have been also changed miscellaneously. We have studied the effects of changes in the components (and in their parameters) on the output maps. Various test data sets and measures have been used for the evaluation of accuracy.

2 MATERIALS AND METHODS

2.1 *The pilot area*

Zala County (3.784 km²) is one of the nineteen counties of Hungary, with variable physiographical conditions (Fig. 2) characterized by varied landscape of hills and valleys. The present surface of the county developed through long lasting complex geological processes showing strongly transitional features. The county's climate is moderated by the effect of the relative proximity to the Alps. Due to the climate and the great variety in the landscape, the vegetation cover is also diversified: the most significant element is the continuous forests covering the hillsides. Due to the relatively cool and moist climate and undulating terrain, the county has a dense network of watercourses. Its largest river is the Zala which is encompassed by drained swamps along its way to the lake Balaton, Central-Europe' largest lake.

Almost all Hungarian soil datasets provide information on the county's soil cover, so their applicability can be also tested. Potentially nationwide auxiliary spatial datasets has been also available for the county. In addition to detailed DEM and vast remotely sensed information, thematic maps on climatic factors, geology and groundwater are available for the area.

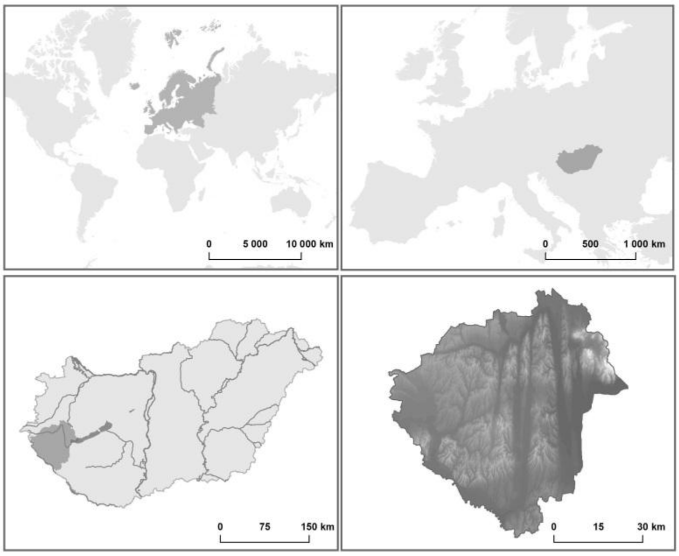

Figure 2. The situation of Zala county in the South-Western part of Hungary and its topography

2.2 *Soil data*

However concerning a specific soil attribute the situation is not necessarily promising. Spatial distribution of SOM, either as a primary soil property or as an notable indicator of various soil related features (functions, services), has a great importance. Prior to our case study the following opportunities were available:

• SMUs of the AGROTOPO countrywide spatial soil information system (1:100,000) with categorized SOM stock information given in t/ha units (Fig. 3).

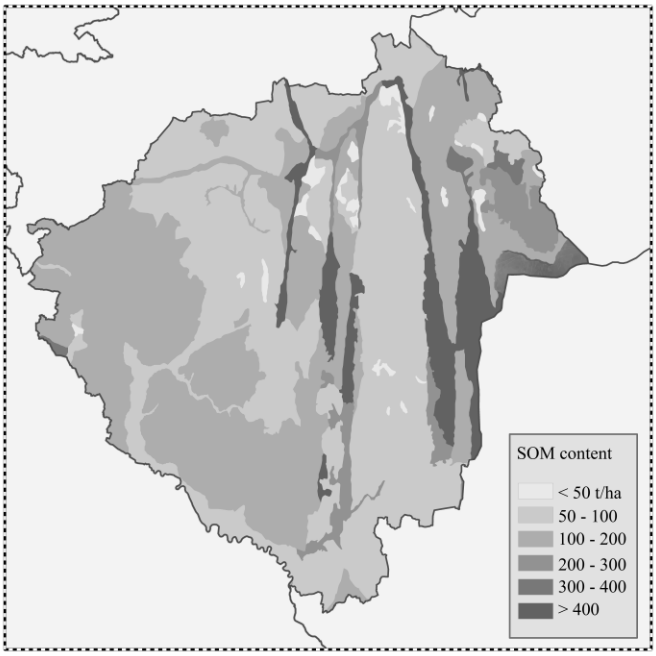

Figure 3. SOM map of Zala county according to the AGROTOPO database

• There is applicable spatial information on organic matter resource provided by the 1:10,000 scale genetic soil maps. These maps however do not cover totally even the agricultural areas of the country; are partly processed digitally and are not managed properly.

• The Hungarian Soil Information and Monitoring System (SIMS; Várallyay, 2009) provides up-to-date and reliable data on SOM, but there are only 59 profiles within the country. On the other hand SIMS locations were definitely not selected to be spatially representative.

• In the frame of the Soil Fertility Monitoring System (SFMS; Várallyay, 1994), which was established to provide a soil and agronomy database for rational soil management and plant nutrition, numerous soil characteristics were measured in the topsoil (0-30cm soil layer or the ploughed horizon) of about 100,000 agricultural plots. At the present merely fragments are available of this valuable legacy data set both thematically and spatially, but they can be suitable used for verification purposes.

• The Digital Kreybig Soil Information System (DKSIS; Pásztor et al., 2010, 2012) simultaneously contains two types of geometrical datasets. Soil mapping units (SMU) are characterized by three attributes: (i) combined texture and water management categories (TWM), (ii) overall soil chemical properties (SCP), (iii) and a so called landscape management soil type (LMST). The SMUs were delineated based on overall chemical and physical soil properties of the soil root zone. Detailed soil properties were determined and measured in soil profiles. DKSIS is available for the whole area of Hungary with representative profile description for about 22,000 plots. This profile information is transferred for further locations, which sums up in approximately 250,000 plots. Since the Kreybig survey (Kreybig, 1937) did not regionalize SOM data, such type of information has been available only at profile level.

2.3 *Auxiliary data*

Digital elevation models are available for the whole area of Hungary with various spatial resolutions from 5 m. In the Zala pilot the performance of the 20 m DEM was tested. In this way smaller area (roughly 1/25 that of the country) is offset as compared the 100 m DEM planned to be used countrywide from computational point of view. Altitude, slope, topographic wetness index, LS factor, mass balance index, catchment area, profile and plan curvature, stream power index and tpi500 derivatives were used in the analysis.

Informative, low-cost remotely sensed data (RS) were available in the form of multitemporal MODIS products. Luckily, MODIS scenes entirely cover the county, thus no merging of neighboring images was necessary. 18 NDVI and EVI were selected from the period of 2009-2011representing different parts of the growing season and years with various climatic conditions.

There is further thematic auxiliary spatial information available for the whole area of the country, which was tested in the Zala pilot. Two climatic data layers (CF; mean annual precipitation and mean temperature of summer months) were used in addition to the 1:100.000 Geological Map of Hungary (MFGI, 2011) and the map of groundwater depth prepared by Water Research Institute (VITUKI, 2005).

2.4 *Spatial inference of soil related information*

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 modeled 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. Due to extent limits, in this paper merely the detailed analysis of the elaboration of SOM maps with regression kriging using varying set of auxiliary variables is discussed.

2.5 *Data preprocessing*

Soil organic matter content measured in the topmost layer of 1789 DKSIS soil profiles was regionalized. Calibration (80%) and validation (20%) subsets were randomly selected. SOM data were cleared, standardized and logit-transformed in order to better fit the applied geostatistical demands. Categorical data of DKSIS SMUs (TWM, SCP and LMST) were also used in the form of indicator variables. Covariables were unified both in spatial resolution, resampled to 20 m raster as well as in value range, transformed to 0-255 range according to Hengl (2009). The principal components of predictor maps were used in each case to reduce their multicollinearity.

3 RESULTS AND DISCUSSION

A series of regression kriging with variable set of auxiliary variables was executed. The tested models are listed in Table 1.

Table 1. The composition of the tested models

	DDM	RS	CF	TWM	SCP	LMST
A	x	-	-	-	-	-
B	x	x	-	-	-	-
C	x	-	x	-	-	-
D	x	x	x	-	-	-
E	x	-	x	x	-	-
F	x	-	x	-	x	-
G	x	-	x	x	x	-
H	x	-	x	-	-	x
I	x	x	x	-	x	-
J	x	x	x	x	-	-
K	x	x	x	x	x	-
L	x	x	x	-	-	x

The first evaluation of the models was established on the results of the multiple regression (Table 2). The performance of multiple regression increased significantly by the involvement of soil layers. Spatial pattern provided by SMUs improves the prediction even if their delineation is based on differing soil features.

Table 2. Performance of the multiple regressions

Maps	R2	Std. Error	SSR	SSE	MSE	IVM*
A	0.17	0.48	82.4	407.1	0.23	3
B	0.11	0.50	52.2	437.3	0.25	6
C	0.18	0.47	89.4	400.1	0.22	6
D	0.20	0.47	98.3	391.2	0.22	8
E	0.34	0.43	166.7	322.8	0.18	9
F	0.26	0.45	128.7	360.8	0.20	7
G	0.35	0.42	173.1	316.4	0.18	12
H	0.35	0.42	170.9	318.6	0.18	10
I	0.27	0.45	133.8	355.6	0.20	9
J	0.35	0.42	169.8	319.7	0.18	11
K	0.36	0.42	176.3	313.2	0.18	14
L	0.36	0.42	173.9	315.6	0.18	12
GEO**	0.36	0.42	173.9	315.7	0.18	13
GW***	0.35	0.42	170.9	318.6	0.18	12

* Independent Variables in the Model after PCA and stepwise selection of the variables on 5% significance level.

** K model supplemented with geology

** K model supplemented with groundwater

The difference in the improvement can be attributed to the differences in the number of categories used in TWM, SCP and LMST respectively. Since the inclusion of geology and groundwater did not performed substantively better, in spite of the numerous indicator variables introduced by their lots of categories, they were omitted from the further analysis.

After the execution of regression kriging, the resulted maps were compared along various features (Table 3). The range of the predicted variable generally proved to be lower than that of the calibration data set, that is except for the *H* and *L* model, the maps are significantly smoothed, which mostly evidences in the case of *B* model, where the spatially highly downscaled RS data is compensated only by the DDM derivatives.

Table 3. Statistical properties of resulted maps and those of the base training soil data set

Maps	Min	Max	Range	Mean	Std. Dev.
A	0.67	22.82	22.15	2.51	1.39
B	0.79	16.14	15.35	2.50	1.33
C	0.66	22.55	21.89	2.51	1.40
D	0.70	21.01	20.32	2.56	1.40
E	0.69	20.36	19.67	2.63	1.63
F	0.70	19.42	18.72	2.59	1.56
G	0.69	20.22	19.53	2.61	1.63
H	0.70	32.35	31.65	2.63	1.92
I	0.78	19.13	18.35	2.58	1.53
J	0.71	21.57	20.86	2.67	1.62
K	0.71	21.77	21.06	2.62	1.64
L	0.73	34.77	34.04	2.73	1.93
Data	0.43	29.34	28.91	2.70	2.96

Fig. 4 displays the map results. The overall spatial pattern of them is similar and fairly consistent with small scale map of AGROTOPO presented in Fig. 3. Their detailedness is however conspicuously different. Map of *B* model is strongly smoothed. Entering of soil layers at model *E* heavily increases the contrasts on the maps.

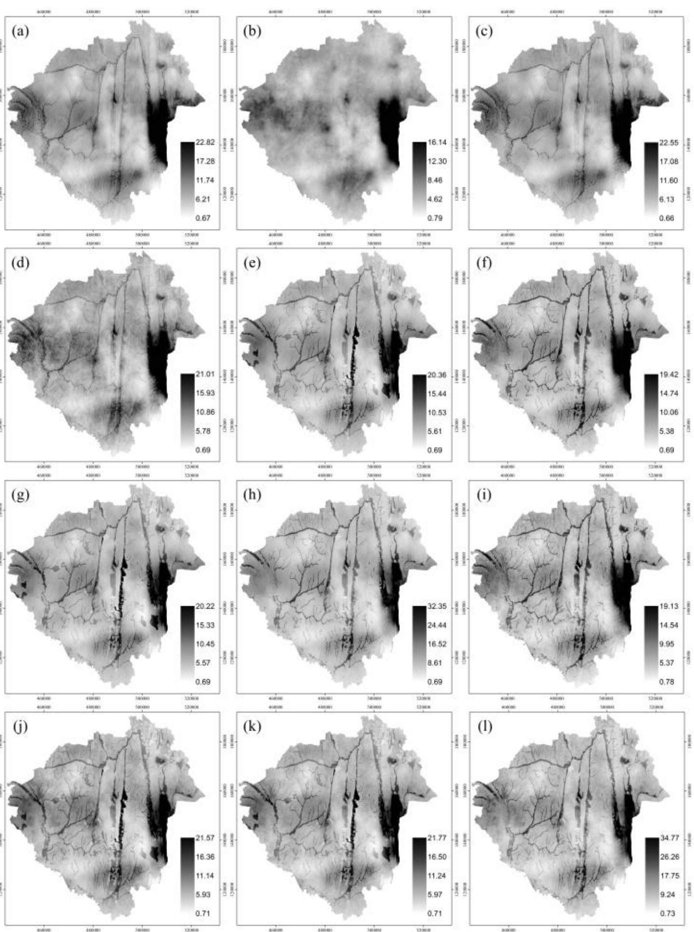

Figure 4. SOM maps created by regression kriging with variable set of auxiliary variables. The letters in the upper left corner refer to the models listed in Table 1.

The maps were validated by the aid of two different data sets. According to the independently selected 20% DKSIS profiles, the predicted maps consistently underestimate, while validation with SFMS suggests overestimation with an almost double value. These differences can be attributed in one hand to the differing sampling and measurements methods applied and on the other hand the time-shift between the two data collection.

Table 4. Validation of the resulted maps based on test DKSIS data set

Maps	ME	MAE	MSE	RMSE	RMNSE	RI
A	-0.43	0.86	6.73	2.59	1.21	5%
B	-0.46	0.88	7.43	2.73	1.26	-
C	-0.43	0.87	6.70	2.59	1.21	5%
D	-0.45	0.88	7.21	2.69	1.22	2%
E	-0.29	0.75	4.69	2.17	1.11	21%
F	-0.37	0.84	6.04	2.46	1.14	10%
G	-0.29	0.75	4.69	2.17	1.11	21%
H	-0.40	0.83	6.12	2.47	1.09	9%
I	-0.41	0.88	6.76	2.60	1.12	5%
J	-0.31	0.77	5.19	2.28	1.11	16%
K	-0.31	0.78	5.20	2.28	1.11	16%
L	-0.42	0.85	6.68	2.59	1.10	5%

Table 5. Validation of the resulted maps based on the independent SFMS data set .

Maps	ME	MAE	MSE	RMSE	RI
A	0.75	0.96	2.24	1.50	1%
B	0.70	0.91	2.18	1.48	2%
C	0.75	0.96	2.27	1.51	-
D	0.77	0.96	2.04	1.43	5%
E	0.76	0.93	1.65	1.29	15%
F	0.84	1.01	2.04	1.43	5%
G	0.75	0.92	1.63	1.28	15%
H	0.82	0.99	1.87	1.37	9%
I	0.81	0.97	1.87	1.37	9%
J	0.77	0.92	1.60	1.26	16%
K	0.73	0.89	1.55	1.24	17%
L	0.87	1.01	1.87	1.37	9%

Relative improvement of the model is compared to the worst *B* and *C* model respectively. In the case of DKSIS validation *E* and *G* models provide the most, while for SFMS verification *K* and *J* slightly overcome the following *E* and *G* models. The worst results turn up for the "soil free” models. An interesting outcome is the consistently better performance of TWM as compared to SCP or LMST with the same auxiliary variables.

4 CONCLUSIONS

According to our results the following conclusions can be drawn, which will be relied on in our countrywide mapping activities:

Inclusion of spatial soil data significantly improves the performance of RK for the compilation of SOM maps for the topsoil as compared to the application of pure environmental covariables. Maps based on models including soil layers are less smoothed and display stronger contrasts.

Usage of various soil covariables, even provided by the same SSIS, in a model with the same auxiliary variables results in remarkable differences of the final map. Joint involvement of more than one soil layer does not necessarily improves the performance.

Various evaluations of the result maps are not necessarily consistent. Consequently, either a nationally uniform measure should be defined (which still can be different for various soil parameters) to efficiently compare the various possible results, or the identification of the target soil (related) variable should be supplemented (on metadata level) by details of the specific expectations (e.g.: its extreme values should be kept/minimal smoothing is accepted).

5 ACKNOWLEDGEMENT

Our work has been supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167) and the Bolyai Research Grant Program.

6 REFERENCES

Baumgardner, M.F. 2011. Soil Databases. In Huang P.M., Li Y. & Sumner M.E. (eds), *Handbook of Soil Sciences: Resource Management and Environmental Impacts*: 21-35. Boca Raton: CRC Press.

Blum, W.E.H. 2005. Functions of Soil for Society and the Environment. *Reviews in Environmental Science and Biotechnology* 4: 75-79.

Böhner, J, Köthe, R., Conrad, O., Gross, J., Ringeler, A. & Selige, T. 2002. Soil Regionalisation by Means of Terrain Analysis and Process Parameterisation. In Micheli, E., Nachtergaele, F. & Montanarella, L. (eds), *Soil Classification 2001*. Ispra: The European Soil Bureau, Joint Research Centre, EUR 20398 EN.

Bullock, P. 1999. Soil Resources of Europe – An Overview. In Bullock, P., Jones R.J.A. & Montanarella, L. (eds), *Soil Resources of Europe*. Luxembourg: European Soil Bureau Research Report 6, Office for Official Publications of the European Communities.

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

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

Kreybig, L. 1937. *Soil survey, analysis and mapping methodology of the Royal Hungarian Geological Institute (in Hungarian)*. In M. Kir. Földtani Intézet Évkönyve, 31, 147–244. Budapest.

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

MFGI. 2013. Geological map of Hungary at a scale of 1:200,000. Magyar Földtani és Geofizikai Intézet http://loczy.mfgi.hu/atlasz200/.

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.

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

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

Pásztor, L., Szabó, J., & 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., Matus, J. & 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.

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

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

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

Várallyay, Gy. 1994. Soil data-base for long-term field experiments and sustainable land use. *Agrokémia és Talajtan* 43: 269-290.

Várallyay, Gy. 2002. Soil survey and soil monitoring in Hungary. In European Soil Bureau - Research Report 9, ESB, Ispra, 139–149.

Várallyay, Gy. 2009. Soil conditions in Hungary based on the data from the Soil Conservation Information and Monitoring System (SIMS). In Juhász, I. (ed.) Budapest: Földművelésügyi és Vidékfejlesztési Minisztérium.