Target-specific digital soil mapping supporting spatial planning in Hungary

Katalin Takács
takacs.katalin@rissac.hu

Gábor Szatmári
szatmari@rissac.hu

Zsófia Bakacsi
zsofi@rissac.hu

Annamária Laborczi
laborczi@rissac.hu

József Szabó
james@rissac.hu

Tibor Tóth
tibor@rissac.hu

László Pásztor
pasztor@rissac.hu

Institute for Soil Sciences and Agricultural Chemistry,
Centre for Agricultural Research, Hungarian Academy of Sciences
Herman Ottó 15,
H-1022 Budapest, Hungary

Abstract

In Hungary a significant amount of soil data is available in different databases or soil information systems, however there are frequent discrepancies between the available and the expected soil information. The tasks of spatial planning – like delineation of Areas with Natural Constraints or Areas with Excellent Productivity, irrigation strategy – increasingly demand unusual or more complex information about soils, which cannot be fully satisfied by formerly elaborated spatial soil information systems. The soil data of the previous surveys should be reinterpreted and reprocessed to meet the demands of regional planning activities. Digital soil mapping (DSM) integrates geostatistical, data mining and GIS (geographic information system) tools. Applying DSM methods makes the elaboration of target-specific soil maps with improved and specific thematic, spatial and temporal accuracy as opposed to former, more general soil maps.

Keywords: target-specific soil maps, digital soil mapping, regression kriging, soil information systems, Areas with Natural Constraints

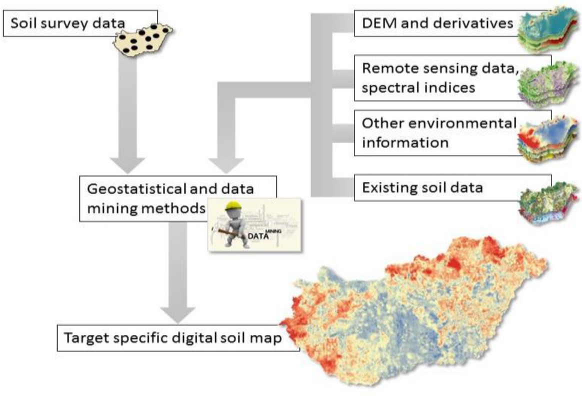
1 Introduction

The national level spatial planning in Hungary requires adequate, preferably timely detailed spatial knowledge about soil cover. Numerous databases and soil information systems were created in Hungary by different purposes, methods, spatial coverage and resolution [28], but the former soil maps cannot fully satisfy the recent needs of spatial planning and policy making. More and more frequently the requests do not refer to primary or secondary soil properties, but to more complex or derivated soil information, functions (e.g. filtering, buffering), processes (e.g. degradation), services (e.g. provisioning, regulating) [21]. The gap between the available and expected soil data can be filled by the reinterpretation and reprocessing of existing soil survey data or by the fusion of former soil information [19].

GIS (geographic information system) provides proper environment for combined processing of data originating from different sources and for object oriented spatial interpretation.

Digital soil mapping (DSM) is the creation of spatial soil information systems by the coupling of observational soil data with environmental data through qualitative relationships [18]. Applying suitable DSM technique leads to target-specific soil maps with appropriate thematic, spatial and temporal accuracy. Soil survey data originate from point like observations, their spatial support is usually soil profiles. For mapping this point type information should be spatially extended by a properly chosen process. DSM methods use spatially exhaustive, environmental auxiliary variables related to soil forming factors for the spatial inference [17]. These variables should be in direct or indirect relation with the target soil property and should provide full coverage for the target area [18]. The soil-landscape relation can be modelled by geostatistical and data-mining methods and the result is a target-specific soil map (Figure 1). The spatial accuracy and the reliability of the prediction can also be estimated, both locally and globally.

Figure 1: The concept of digital soil mapping.



It is a recent task in Hungary to designate Areas with Natural Constraints according to the common European regulation. An objective, science based common criteria system for the EU member states was compiled [22], which includes a total of eight criteria referring to climate, soil and topography. The soil criteria consist of further 11 sub-criteria, which require information on many basic and derivative soil properties (Table 1). The criteria, which refer to a basic soil property – like rooting depth or pH – can be easily mapped by the data of the most commonly used soil information systems. But other criteria, which are not directly observed in Hungary or refer to a more complex soil property, need more data sources or the reprocessing of soil survey data.

In our paper we present three examples how the requirements of delineation of Areas with Natural Constraint according to the common European biophysical criteria were fulfilled by specific DSM products based on the reinterpretation of former soil survey information.

Table 1: Soil criteria and thresholds for delineation of Areas with Natural Constraints according to EU regulation

Criterion	Definition	Severe threshold
Limited soil drainage	Areas which are water logged for a significant duration of the year	Wet 80cm > 6 months, or 40cm > 11 months
		Poorly or very poorly drained
		Gleyic color pattern within 40cm
Unfavorable texture and stoniness	Relative abundance of clay, silt, sand, organic matter (weight %) and coarse material (volumetric %) fractions	≥ 15% of topsoil volume is coarse material, rock outcrop, boulder
		Texture class in half or more (cumulatively) of the 100 cm soil surface is sand, loamy sand
		Topsoil texture class is heavy clay (≥ 60% clay)
		Organic soil (organic matter ≥ 30%) of at least 40cm
		Topsoil contains 30% or more clay and there are vertic properties within 100cm of the soil surface
Shallow rooting depth	Depth (cm) from soil surface to coherent hard rock or hard pan	Rooting depth ≤ 30cm
Poor chemical properties	Presence of salts, exchangeable sodium, excessive acidity	Salinity ≥ 4 dS/m in topsoil
		Sodicity ≥ 6 ESP in half or more of the 100 cm surface layer
		Soil acidity topsoil pH (H_20) ≤ 5

Source: [22]

2 Materials and methods

2.1 Soil data

In Hungary a large amount of soil information is available because of the long tradition is soil survey and soil mapping [28]. Digital Kreybig Soil Information System (DKSIS; [23]) was compiled based on the most detailed nationwide soil survey [14] and it covers the whole area of Hungary. DKSIS consists of two types of soil information. Soil mapping units are defined by the physical and chemical properties of the rooting zone, but it is just a robust categorization. Soil profile dataset contains many measured records about the physical and chemical soil properties on layer level. Detailed profile descriptions are available for about 22.000 sampling sites, which is spatially transferred for further 250.000 locations. The original mapping concept and survey strategy is discussed in details in [23].

The Hungarian Soil Information and Monitoring System (SIMS) consists of 1.234 observation locations, which have been selected to represent the main pedological characteristics. SIMS contains detailed and up-to-date quantitative soil information about physical and chemical properties on layer level [27].

The Hungarian Detailed Soil Hydrophysical Database (MARTHA) contains harmonized soil hydrophysical and chemical information collected from numerous data sources. In MARTHA, the soil information is available for 3937 profiles, but they are representative mainly for the cultivated area [15].

2.1.1 Sandiness

The first criterion, which was estimated in this study is related to sand content. The threshold defined for unfavorable texture due to significant amount of sand fraction requires that the texture class in half or more of the 100 cm soil surface is sand or loamy sand. So based on the FAO (Food and Agriculture Organization) texture triangle [6] the calculation of this criterion requires the knowledge of the amount of sand and silt particle fraction of the 100 cm soil surface' [22].

The data of SIMS were used. SIMS contains measurements about seven particle size fractions, from which the rate of sand, silt and clay can be calculated. Based on FAO particle size classes, the limit between silt and sand is 0.063 mm [6], but in SIMS it is 0.05 mm (Table 2).

Table 2: Measured particle sizes in SIMS.

particle size class	size [mm]
clay	< 0.002
silt	0.002-0.005
	0.005-0.01
	0.01-0.02
	0.02-0.05
sand	0.05-0.2
	0.2-2.0

SIMS data were converted into FAO particle size classes by log linear interpolation based on the cumulative frequency distribution of the particle sizes [20]. It was tested on the converted data if the criterion of the sandiness is met in a layer. The layer level results were summarized for the upper 100 cm of soil surface.

2.1.2 Vertic properties

Another criterion, which has required non-existing spatial characterization of soils, is in connection with vertic properties [22]. Vertic properties emerge in the form of wedge-shaped soil aggregates with a longitudinal axis, slickensides and shrink-swell cracks, which are typical for soils with high clay content [13]. The definition of biophysical criterion refers to the presence of vertic properties within 100 cm of the soil surface [22].

Vertic properties cannot be directly measured, they can be observed only in field. Soil profile descriptions in DKSIS have indication in the form of notes, like shrink-swell cracks or slickensides. Therefore a binary parameter was created for the occurrence of vertic property: if the profile description includes notes about vertic properties, the value of the parameter is 1, in all other cases it was set to 0. The result of the mapping, due to the applied spatial interpolation method, is a continuous map, with probability values for vertic property between 0 and 1. The presence of vertic property was considered to be verified above 66% probability.

2.1.3 Salinity

The third target criterion in this study refers to soil salinity. Soil salinity is determined by measuring the electrical conductivity (EC) of a solution extracted from a water-saturated soil paste [22]. The prediction of salinity requires the knowledge of EC in topsoil.

None of the Hungarian soil information systems contains data on directly measured EC, therefore this parameter should have been estimated by other basic soil properties. EC can be calculated by the liquid limit and the total salt content of the soil [7]. MARTHA was used as data source in case of EC estimation. MARTHA was created primarily not for mapping purposes, therefore the spatial coverage is not enough consistent for the whole country and not so ideal for countrywide prediction. Therefore the data of MARTHA was completed by another soil information system to achieve sufficient spatial coverage for the spatial prediction. DKSIS provides data on points where the salt content (and EC as well) is certainly zero. By the aid of these auxiliary points the spatial extension of saline soils with higher EC can be refined.

2.2 Auxiliary data

For the mapping of soil properties different auxiliary environmental variables were used, and further covariables were also derived. These variables characterize the soil forming factors, which determine the predicted soil properties.

The most commonly used covariables characterize the terrain. In this study we used the relevant part of EU-DEM. EU-DEM is a digital elevation model (DEM) for Europe, with 25 m spatial resolution [1]. Based on this DEM numerous derivatives were calculated using SAGA GIS software [2]. These derivatives provide information not only on the pure terrain properties, but also on other environmental parameters. Flow line curvature, general curvature, real surface area and vector ruggedness measure characterize the morphometry. Relative slope position index and topographic position index are in connection with topographic situation. Channel network base level, mass balance index, MRVBF (multiresolution index of valley bottom flatness), MRRTF (multi-resolution index of the ridge top flatness), stream power index, vertical distance to channel network are in connection with the hydrological and run-off properties of the area. Diurnal anisotropic heating and SAGA wetness index are in relation with microclimate.

Satellite remote sensing provides direct information on land cover, land use, vegetation condition and bare soils. For the modelling the satellite images of Moderate-Resolution Imaging Spectroradiometer (MODIS) sensor (on the board of Terra and Aqua satellite) were used, which have a 250 m spatial resolution. The applied MODIS images were acquired in spring and autumn (16.03.2012 and 07.09.2013) in line with plant phenology. The data of the red and the near infra-red (NIR) bands were used and NDVI (Normalized Difference Vegetation Index; [24]) was calculated, because these bands and the calculated index are in strong relationship with the state of vegetation and biomass, which reflect certain soil features.

Climatic conditions can be more generally characterized by the long term average of yearly mean temperature, precipitation, actual evaporation and evapotranspiration [26].

Further environmental variables were also applied as national 1:50.000 CORINE Land Cover map (CLC50; [3]) and the 1:100.000 Geological Map of Hungary (FDT100; [8]). The land cover and the geological dataset contain numerous category, from which object specific groups were aggregated, 8 and 14 respectively.

Existing spatial knowledge on soil properties can be also used as environmental covariable, so physical and chemical soil property maps generated from the soil mapping units of DKSIS were also involved in the modelling.

A set of 70 environmental variables was compiled. For the mapping of each target soil property the suitable covariates were selected. The auxiliary dataset needed some preprocessing, before the spatial inference. The maps of covariables were unified to 100 m spatial resolution. The DEM and derivatives as well as the satellite images were resampled, the point like meteorological data were interpolated and the vector data were rasterized to a common, 100 m resolution grid system. The categorical data (land cover, geology and soil property maps) were used in indicator form.

2.3 Spatial inference

The target-specific maps were created by intentionally selected DSM methods. The final set of environmental covariables in the case of each target variable has consisted of at least 40 layers. To reduce the number of predictor variables and to avoid their multicollinearity, principal component analysis (PCA) was carried out at first. In the further analysis the first principal components, which explain together the 99% of the variance, were used.

The spatial extension of soil data was performed by regression kriging (RK), which is widely used in digital soil mapping [4, 5, 10, 12, 16]. RK is a spatial prediction technique that combines the regression of the dependent variable on auxiliary variables with kriging of the regression residuals [11]. So at first, the target soil property was modelled by multiple linear regression analysis (MLRA) with stepwise selection method (5% significance level) of the environmental variables. The interpolation of residuals (the differences between the predicted and observed values) was carried out by ordinary kriging. The result of the estimation is the sum of the regression model and the interpolated residuals. RK was carried out in SAGA GIS environment.

The overall accuracy of the predicted maps was checked by Leave One Out Cross Validation (LOOCV; [25]). By LOOCV the estimation of the target soil property is carried out n-1 times, leaving out each time one of the samples. Then the predicted and the measured values of the left-out sample are compared.

Table 3: Validation methods used in this study.

Validation methods	Calculation
mean error	$\mathrm{ME} = \frac{1}{n} \cdot \sum_{i=1}^{n} [\hat{z}(s_i) - z(s_i)]$
mean absolute error	$\mathrm{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} \lvert \hat{z}(s_i) - z(s_i) \rvert$
root mean square error	$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} [\hat{z}(s_i) - z(s_i)]^2}$
root mean normalized square error	$\mathrm{RMNSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left[\frac{\hat{z}(s_i) - z(s_i)}{\hat{\sigma}_i}\right]^2}$

$\hat{z}(s_i)$ and $z(s_i)$ are the estimated and measured values at the s_i control point and $\hat{\sigma}_i$ is the prediction variance

The estimation of the overall accuracy was tested by the following parameters: mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and root mean normalized square error (RMNSE) (Table 3).

The expected value of the ME and RMNSE are 0 and 1, respectively. MAE and RMSE refer to the accuracy of the estimation, the lower the value of the MAE and the RMSE, the better is the prediction accuracy [9].

The validation of the probability map is carried out by slightly differently. In this case the final map was validated by a coincidence matrix.

The spatial accuracy of the maps can be characterized by the prediction variance [10], which expresses the uncertainty of the prediction. The prediction variance maps are provided for all map results.

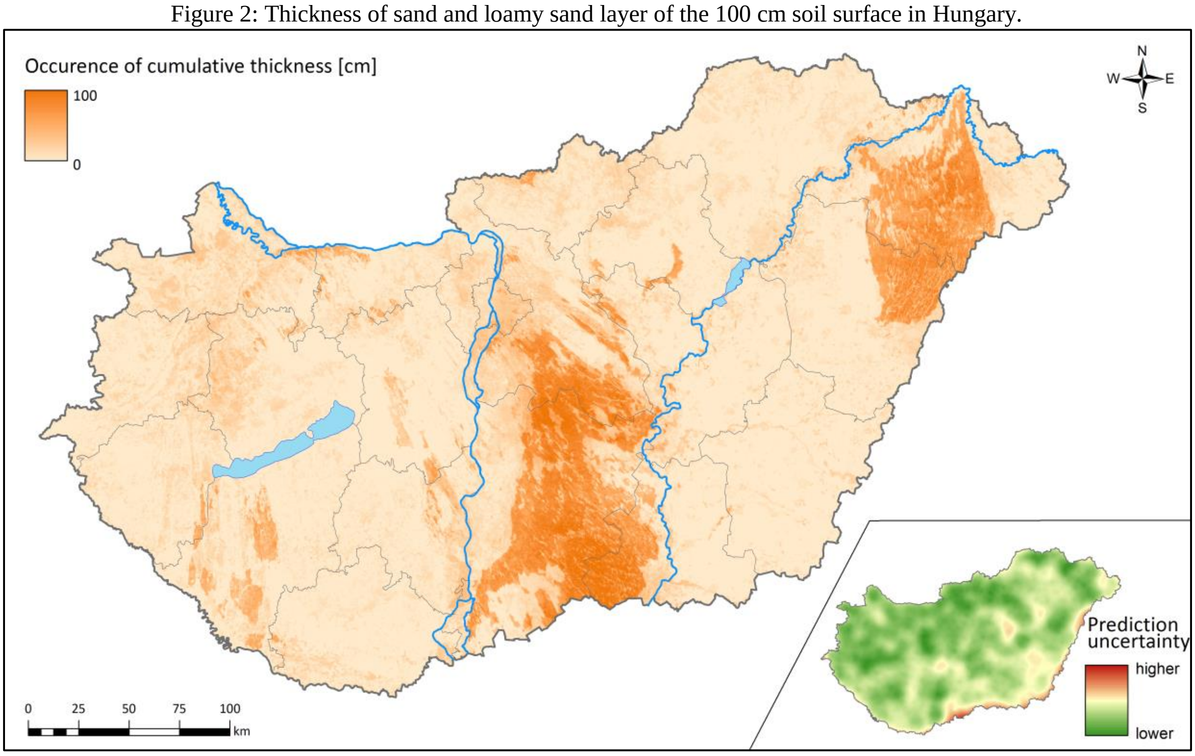
3 Results and Discussion

3.1 Compilation of the map of unfavorable texture due to significant amount of sand fraction

Soil texture properties are affected mainly by topography, climate, vegetation and parent material. For the spatial inference the following environmental auxiliary variables were selected by the stepwise method of the MLRA:

- DEM and derivatives: elevation, slope, multiresolution ridge top flatness, multi resolution valley bottom flatness
- climatological data: yearly mean precipitation, actual evaporation, evapotranspiration
- satellite data: NDVI, spring-NIR

Figure 2: Thickness of sand and loamy sand layer of the 100 cm soil surface in Hungary.



- geological data: blown sand
- land cover: arable land, forest
- physical soil property map: poor water retention and very high permeability, and infiltration rate

The predicted sandiness criterion map is displayed in Figure 2. The areas with coarser particle size are well-delineated and coincided with the location of the main sand ridges in Hungary. The result of the LOOCV is summarized in Table 4. The MAE of the prediction is 12 cm.

Table 4: Results of the LOOCV validation according to the prediction of sandiness.

Validation method	Validation result
ME	-0.003
MAE	12.478
RMSE	19.853
RMNSE	1.094

3.2 Compilation of the map of vertic properties

In this case the target soil property was the probability of the occurrence of vertic properties. The stepwise regression method provides possibility for the evaluation of the environmental auxiliary variables selected for the modelling. The covariables were the followings:

- DEM and derivatives: aspect, topographic wetness index, SAGA wetness index, multiresolution ridge top flatness, multi resolution valley bottom flatness, diurnal anisotropic heating
- climatological data: yearly mean temperature, precipitation, actual evaporation, evapotranspiration
- satellite data: NDVI, NIR-band, red-band
- physical soil property map: good and high water retention, saline soils, peaty soils

The spatial extension of the probability of presence of vertic properties was carried out by RK. The predicted vertic property criterion map is displayed in Figure 3, where the mapped probability values were converted to percentage. The occurrence of vertic properties has coincided with lowlands covered with fluvial sediment and where the inland excess water inundation is frequent in Hungary.

By the validation the predicted probability values (0–1) and the observed parameters (0 or 1) were compared; the predicted values above 0.66 were taken into consideration as 1, under 0.66 as 0. This value was chosen, because using 66% as a limit is generally accepted in the methodology of defining natural constraints [22]. The LOOCV results referring to vertic property map are shown in Table 5. The overall accuracy of the map is rather good, because misclassification occurs only in case of 3%.

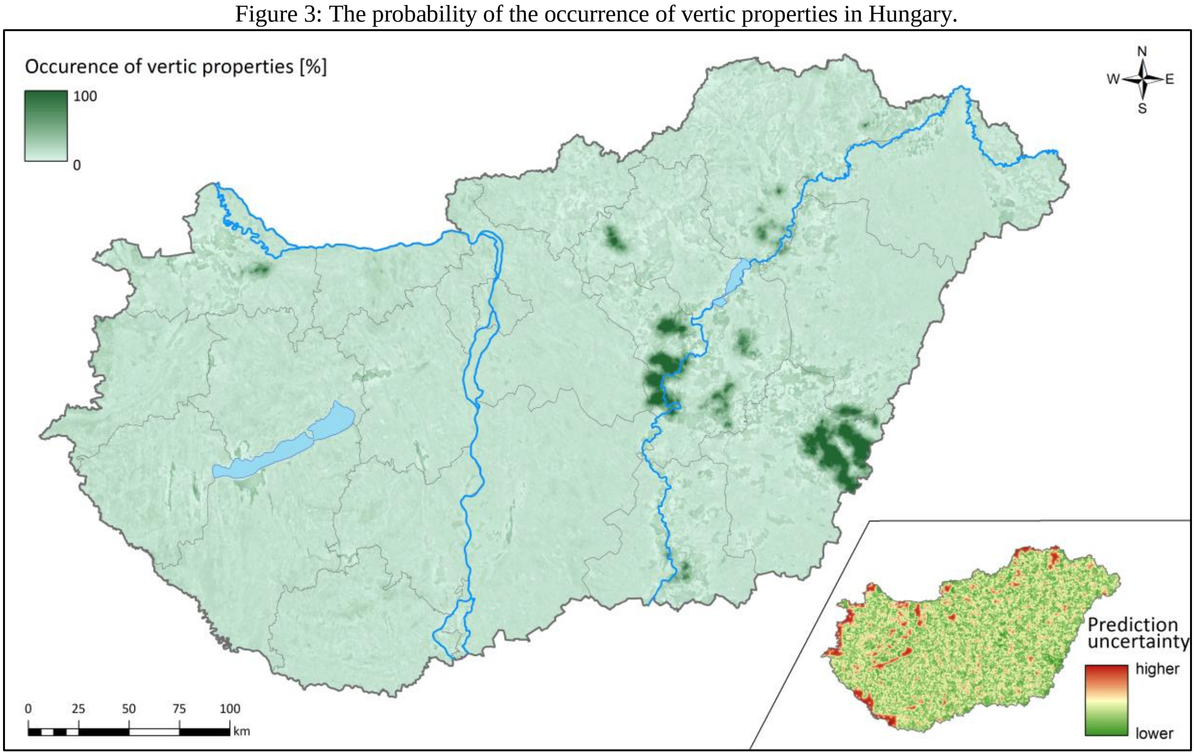
Table 5: Result of the validation in case of vertic property.

Fulfilled criterion		predicted	
		Yes (1)	No (0)
observed	Yes (1)	34.8%	1.8%
	No (0)	1.2%	62.2%

3.3 Compilation of the map of salinity in topsoil

The target soil property was the EC of soils. While EC and salt content is affected by topography, climate, vegetation and parent material, therefore in its spatial inference the following

Figure 3: The probability of the occurrence of vertic properties in Hungary.

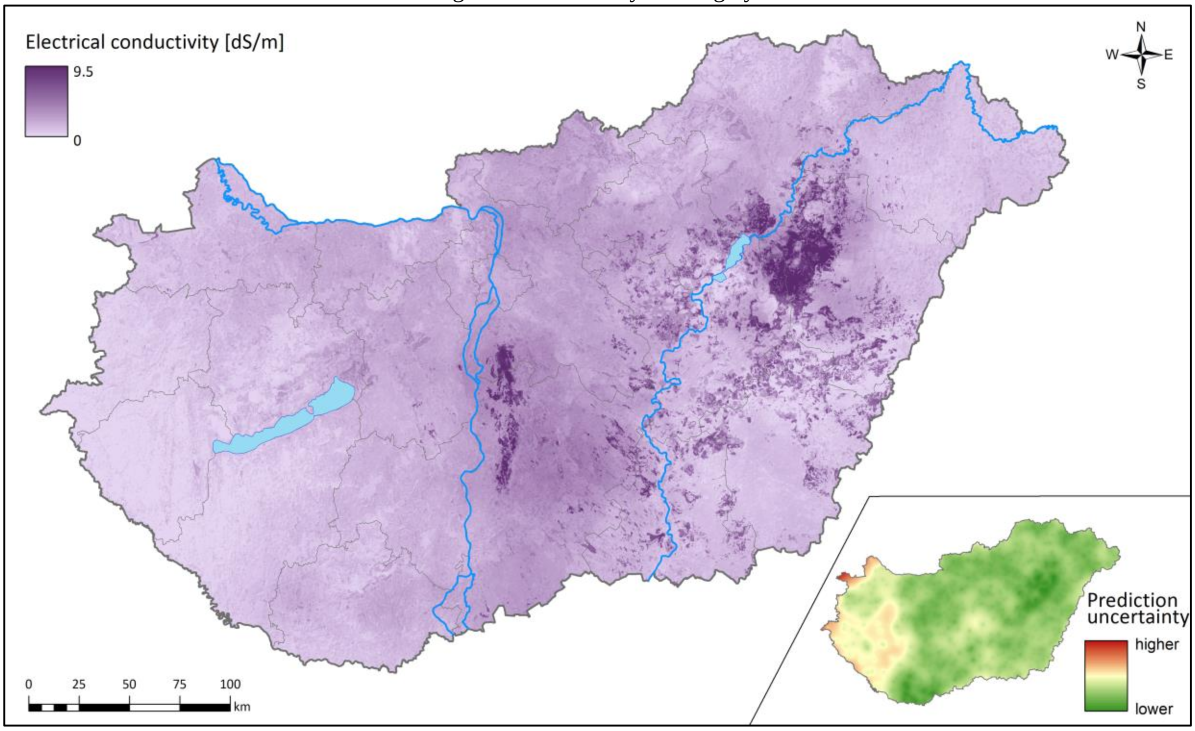


environmental auxiliary variables were used by the stepwise method of the MLRA:

- DEM and derivatives: SAGA wetness index, multiresolution ridge top flatness, multi resolution valley bottom flatness
- climatological data: yearly mean precipitation, actual evaporation
- satellite data: autumn-NDVI
- land cover: grassland, sparse vegetation
- physical soil property map: saline soils
- chemical soil property map: saline soils, neutral and calcic soils

The predicted salinity criterion map is displayed in Figure 4. The areas with higher EC and higher salt content are in good agreement with the areas covered by saline soils in Hungary.

Figure 4: Soil salinity in Hungary.



The result of the LOOCV are summarized in Table 6. Based on the ME result, the model somewhat overestimates. RMNSE is also close to its expected value. Based on MAE and RMSE the overall accuracy of the prediction is acceptable. MAE and RMSE has relative low values, they are lower than the 10% of the value set.

Table 6: Results of the LOOCV validation according to the prediction of salinity.

Validation method	Validation result
ME	0.010
MAE	0.850
RMSE	1.430
RMNSE	0.935

4 Conclusion

The task was to produce reliable, nationwide maps on specific soil features, which were not mapped or expressed spatially formerly. In this study the elaboration of three new object-oriented maps is presented, which were created by regression kriging. To fulfil the data demand of the presented maps the reinterpretation of former soil survey information and the integration of more datasources were needed. The compiled new maps can satisfy the needs of designation of Areas with Natural Constraints. Our approach can be also applied by other tasks of spatial planning – like designation of Areas with Excellent Productivity, support of irrigation strategy and flood risk assessment –, if the target soil property is clearly defined.

References

[1] A. Bashfeld and A. Keim. Continent-wide DEM Creation for the European Union. *34th International Symposium on Remote Sensing of Environment*, 10–15th April 2011, Sydney, Australia, 2011.

[2] J. Böhner, K. R. McCloy and J. Strobl editors. *SAGA – Analysis and Modelling Applications.* Göttinger Geographische Abhandlungen, 115, 2006.

[3] Gy. Büttner, B. Kosztra, G. Maucha and R. Pataki. *Implementation and achievements of CLC2006.* European Environment Agency, Technical Report – Revised Final Draft, 2012.

[4] W. de Carvalho Junior, P. Lagacherie, C. da Silva Chagas, B Calderano Filho and S. B. Bhering. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma,* 232–234: 479–486, 2014.

[5] E. Dobos, T. Bialkó, E. Micheli and J. Kobza. Legacy soil data harmonization and database development. In. J. L. Boettinger, D. W. Howell, A. C. Moore, A. E. Hartemink and S. Kienast-Brown editors. *Digital Soil Mapping Bridging Research Environmental Application, and Operation*, pages 309–323. Springer Science+Business Media B.V., 2010.

[6] FAO. *Guidelines for soil profile description* (4th ed.). Rome, Food and Agriculture Organization of the United Nations, 2006.

[7] Gy. Filep and M. J. K. Wafi. A talajoldat sókoncentrációjának és a talaj nátriumtelítettségének (ESP) számítása a telítési oldat jellemzőiből. *Agrokémia és Talajtan*, 42(3–4):245–256, 1993.

[8] L. Gyalog and F. Síkhegyi editors. *Magyarország földtani térképe, M=1:100 000.* A Magyar Állami Földtani Intézet kiadványa, Budapest, 2005.

[9] T. Hengl. *A practical guide to geostatistical mapping.* University of Amsterdam, 2009.

[10] T. Hengl, G. B. M. Heuvelink and A. Stein. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma,* 120:75–93, 2004.

[11] T. Hengl, G. B. M. Heuvelink and D. G. Rossiter. About regression kriging: from equations to case studies. *Comput & Geosci*, 33:1301–1315, 2007.

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

[13] IUSS Working Group WRB. *World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps.* World Soil Resources Reports No. 106. FAO, Rome, 2014.

[14] L. Kreybig. A Magyar Királyi Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere. *Magyar Királyi Földtani Intézet Évkönyve*, 31:145–244, 1937.

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

[16] A. Marchetti, C. Piccini, R. Francaviglia, S. Santucci and I. Chiuchiarelli. Estimating soil organic matter content by regression kriging. In. J. L. Boettinger, D. W. Howell, A. C. Moore, A. E. Hartemink and S. Kienast-Brown editors. *Digital Soil Mapping, Bridging Research, Environmental Application, and Operation,* pages 241–254. Springer Science+Business Media B.V., 2010.

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

[18] A. B. McBratney, B. Minasny, R. A. MacMillen and F. Carre. Digital soil mapping. In. P. M. Huang, Y. Li and M. E. Sumner editors. *Handbook of soil sciences: properties, processes.* CRC Press, USA, 2011.

[19] L. Montanarella. Need for interpreted soil information for policy making. *19th World Congress of Soil Science, Soil Solutions for a Changing World*, 1–6 August 2010, Brisbane, Australia, 12–15. (Published on DVD), 2010.

[20] A. Nemes, J. H. M. Wösten, A. Lilly and J. H. Oude Voshaar. Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. *Geoderma,* 90:187–202, 1999.

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

[22] J. van Orshoven, J.-M. Terres and T. Tóth editors. *Updated common biophysical criteria to define natural constraints for agriculture in Europe.* EUR 26638 EN, JRC, 2014.

[23] L. Pásztor, J. Szabó and Zs. Bakacsi Digital processing and upgrading of legacy soil data collected during the 1:25000 scale Kreybig soil survey. *Acta Geodaetica et Geophysica Hungarica,* 45(1):127–136, 2010.

[24] J. W. Rouse, R. H. Haas, J.A. Schell and D.W. Deering Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS Symposium,* NASA SP-351 I:309–317, 1973.

[25] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of Royal Statistical Society: Series B.* 36:111–147, 1974.

[26] T. Szentimrey and Z. Bihari. Application of MISH method for gridding of SPI series. *BALWOIS 2010*, Ohrid, Republic of Macedonia, 25–29th May 2010.

[27] Gy. Várallyay. *Soil survey and soil monitoring in Hungary*. European Soil Bureau, Research Report No.9, 2002.

[28] Gy. Várallyay. Talajtérképezés, talajtani adatbázisok. *Agrokémia és Talajtan*, 61(Suppl.):249–268, 2012.