CONTENT

Studies

Júlia Göndöcs, Hajnalka Breuer, Ákos Horváth, Ferenc Ács and *Kálmán Rajkai:* Numerical study of the effect of soil texture and land use distribution on the convective precipitation 3

Csaba Centeri, Zoltán Szalai, Gergely Jakab, Károly Barta, Andrea Farsang, Szilárd Szabó and *Zsolt Bíró:* Soil erodibility calculations based on different particle size distribution measurements 17

Judit Szabó, Gergely Jakab and *Boglárka Szabó:* Spatial and temporal heterogeneity of runoff and soil loss dynamics under simulated rainfall 25

Gábor Szatmári, Károly Barta and *László Pásztor:* An application of a spatial simulated annealing sampling optimization algorithm to support digital soil mapping 35

László Pásztor, Annamária Laborczi, Katalin Takács, Gábor Szatmári, Endre Dobos, Gábor Illés, Zsófia Bakacsi and *József Szabó:* Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools 49

István Sisák, Mihály Kocsis, András Benő and *Gábor Várszegi:* Method development to extract spatial association structure from soil polygon maps 65

Literature

H. Kérdő, K. – Schweitzer, F. (eds): AQUINCUM Ancient landscape – ancient town *(Sándor Gulyás)* 79

Churski, P. ed.: The social and economic growth vs. the emergence of economic growth and stagnation areas *(Tibor Tiner)* 81

DOI: 10.15201/hungeobull.64.1.1 Hungarian Geographical Bulletin **64** 2015 (1) 3–15.

Numerical study of the effect of soil texture and land use distribution on the convective precipitation

Júlia GÖNDÖCS[1], Hajnalka BREUER[1*], Ákos HORVÁTH[2], Ferenc ÁCS[1] and Kálmán RAJKAI[3]

Abstract

In this study the Weather Research Model is used to analyse the sensitivity of convection to soil texture and land use distribution based on a heavy precipitation event. Both characteristics affect the latent heat flux and the near surface temperature distribution which are related to buoyancy. The model defaults Food and Agriculture Organization (FAO) soil texture and USGS (United States Geological Survey) land use have been replaced with more accurate databases in Hungary: soil texture based on the Digital Kreybig Soil Information System (DKSIS), land use based on the COoRdination of INformation on the Environment (CORINE). Regarding to soil texture the main changes are related on one hand to clay loam diversification to silty clay, loam, silty loam and sandy loam affecting area over 40 percent of Hungary, and on the other hand reclassification of sandy loam to sand. The difference between USGS and CORINE land use is sporadic, but significant. It is found that the diurnal latent heat flux is the highest at 12 UTC, at this peak the spatial average difference in latent heat flux is +6.5 W/m^2 and –4.3 W/m^2 with respect to soil texture and land use change, while the absolute differences range from –70 W/m^2 to +70 W/m^2 in all cases. As a result temperature at 2 m on average increased by 0.1 °C during soil texture and decreased by 0.15 °C during land use database comparison; the absolute differences are a magnitude higher. When comparing simulations regarding temperature at 2 m over main soil types and main land use categories results indicate –3 °C to +0.5 °C difference. It is found that the modification of both the soil texture and the land use have sometimes a compensating effect on latent heat flux and temperature change. Decrease in latent heat flux results an increase in buoyancy affecting convective precipitation. The formation of precipitation is also affected by large scale advection, therefore, no systematic changes can be seen on daily precipitation distribution. In spite of this, results indicate shifts in precipitation bands with about 30 km, and formation of new storm cells. Locally the replacement of soil texture and land use information to a more accurate one produced ±8 mm/day differences in precipitation.

Keywords: convective precipitation, soil texture, land use, numerical weather prediction

[1] Eötvös Loránd University, Department of Meteorology, H-1117 Budapest, Pázmány P. sétány 1/a. E-mails: kisbucedli@caesar.elte.hu, bhajni@nimbus.elte.hu, acs@caesar.elte.hu
[2] Hungarian Meteorological Service, H-8600 Siófok, Vitorlás u. 17. E-mail: horvath.a@met.hu
[3] Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, H-1022 Budapest, Herman Ottó u. 15. E-mail: krajkai@mail.iif.hu
[*] corresponding author

Introduction

Land surface characteristics are playing an important role in the exchange of energy, water vapour and moment with the lower atmosphere. These exchange processes make the topic relevant also from a meteorological point of view and became increasingly important with the improvement of numerical weather prediction systems. The upward flux of water vapour – essentially the evapotranspiration – is affected by the vegetation and the available soil moisture. Relationships between the lower atmosphere and the soil moisture (Dickinson, R.E. 1984; Pielke, R.A. and Avissar, R. 1990), soil texture (Ek, M. and Cuenca, R.H. 1994; Alapaty, K. *et al.* 1997), soil parameters (Mölders, N. 2005; Breuer, H. *et al*. 2012), land surface heterogeneity (Avissar, R. and Liu, Y. 1996; Pielke, R.A. 2001) and vegetation (Pielke, R.A. *et al.* 1997; Adegoke, J.O. *et al.* 2007) have been extensively analysed in several aspects. Dry soils increase the sensible heat flux responsible for creating updrafts, while wet soils add moisture to the boundary layer (lower atmosphere) through evapotranspiration. Depending on the atmospheric conditions, the added moisture can either make the atmosphere more favourable for storms or more stable. These differences affect thermally driven local atmospheric circulations (Hong, X. *et al.* 1995) and convective precipitation formation (Teuling, A.J. *et al.* 2009).

The feedbacks between soil moisture and precipitation are controversial even in the case of measurements. Some studies have found that higher soil moisture increases the possibility of thunderstorm development by raising the convective available potential energy (CAPE) but the increased water vapour barely affects the convective inhibition (CIN) of the atmosphere (Pielke, R.A. and Zeng, Z. 1989; Eltahir, E.A. 1998). While others concluded that the increased vapour amount decreases the atmospheric temperatures creating greater inhibition (Taylor, C.M. and Ellis, R.J. 2006) resulting less precipitation. Taylor, C.M. *et al.* (2012) described even a higher precipitation formation over dry soils in the Sahel region. Depending on the geographical region, the effect of soil moisture can even negate the effect of atmospheric lower temperatures and higher moisture content. However, it has to be noted that through advection the precipitation occurrence is higher over dry soils when the wind transports additional atmospheric moisture from areas with greater evapotranspiration (deAngelis, A. *et al.* 2010).

Since soil moisture directly affects the water vapour flux to the atmosphere, and it depends on soil texture and land use, the spatial heterogeneity of soil texture and land use can significantly effects the precipitation. However, the joint analysis of both surface characteristics is rare. On climatological scale, the land cover change can affect the atmospheric circulation, especially in South-East Asia, North America and Europe, causing statistically significant changes in the regional distribution of temperature and precipitation without changing globally averaged temperature or rainfall (Chase, T.N. *et al.* 1996, 2000). Changes in either land use (Collow, T.W. *et al*. 2014) or soil texture (Khodayar, S. and Schädler, G. 2013) toward more realistic spatial distribution result more accurate surface heat fluxes and near surface temperatures. In turn, this has an effect on CAPE and convective precipitation formation.

It was shown that a more accurate soil texture distribution can improve the convective precipitation forecast (Breuer, H. 2012) and the effect of land use change has a notable effect as well (Drüszler, Á. 2011). To be able to improve model simulations, it is needed to examine the effect of the employment of different land use and soil texture distribution databases. In this preliminary study using WRF (Weather Research Forecast) model both the land use and the soil texture distribution are replaced with more accurate ones than the commonly used on the global scale. The aim is to assess the magnitude of the effects caused by both of the changes, and also to make a comparison to each other. Simulations are created for a single precipitation event, and are analysed discussing the surface characteristics/precipitation relationship.

Free convection

The atmospheric buoyancy is responsible for free convection, formed by the sun's shortwave radiation, which heats the land surface. The nearby atmospheric layers are warmed mostly by sensible heat flux. The warmer, ascending air is also controlled by the humidity and the thermal stratification of the atmosphere. When the ascending air is warmer than its surroundings the atmosphere is dynamically unstable (Horváth, Á. 2007). When the rising air contains enough moisture, condensation occurs, releasing latent heat of vaporization, lifting air mass higher.

There are different measures to estimate the instability of the atmosphere, like K-index, CAPE or CIN. CAPE (J/kg) is regarded as an indicator of the potential intensity of deep convection, and it is strongly controlled by the properties of the planetary boundary layer. CAPE is calculated from the temperature difference between the ascending air particle and its environment at each height level (e.g. model vertical levels) going from LFC (Level of Free convection) to EL (Equilibrium Level). LFC is the level, where the air particle becomes warmer than its surroundings for the first time, due to latent heat release. In the lower part of the atmosphere, under the LFC, the energy is negative as the particle requires this energy, named as the CIN. The CIN's value shows the atmospheric stability by giving the energy which must be overcome by the air particle to result convection. Neither the CAPE nor the CIN is a measure of possible precipitation, rather they express the potential possibility of free, non-forced (e.g. without the ascent forcing cold front) convective cloud forming if sufficient moisture is present in the atmosphere.

Model and its settings

The calculations were made with the WRF 3.4.1 model, developed by NCEP (National Centre for Environmental Prediction) and NCAR (National Centre of Atmospheric Research) (Skamarock, W.C. *et al*. 2008). This numerical weather prediction model system is a limited area, mesoscale, non-hydrostatical model, which is freely available on the internet, and for this study it was run on the Atlasz cluster of the Eötvös Loránd University. Spatial and temporal distribution settings can be varied in a wide range, the horizontal grid scale can be 1,000 km scaling down to 1 km. But with such fine resolution as a few kilometres, a nesting technique application is needed in the model area. This means the usage of several encompassing model domains with decreasing horizontal grid size in each nest.

In addition to calculating the hydro-thermodynamic equations governing the dynamics of the atmosphere, sub-grid processes (e.g. radiation-transmission, cumulus cloud convection, cloud microphysics, planetary boundary layer processes, soil-atmosphere interactions) were also calculated. The model uses terrain following, hydrostatic pressure vertical coordinate system and a staggered Lambert conformal horizontal grid.

The used nesting technique had an external model area with a 9 km horizontal resolution covering the Carpathian Basin, and a nested domain of 3 km covering Hungary (45.3°–49.8°N, 15.6°–23.6°E). For the simulations, 34 vertical levels were defined. The simulations were run by making changes to the WRF static data, such as soil texture and land use. The model reads the static data as binary files, which are used to create the model area. Four simulations were made: one using the original settings (FAO) soil texture and USGS land use (reference), the next by changing the soil type data (DKSIS) inside Hungary, another by modifying the land use cover (CORINE) and in the fourth applying both of the changed datasets (DKSIS&CORINE).

Meteorological initial and boundary conditions for the simulations were taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) model which has a horizontal resolution of 15' (approx. 25 km). From the available model levels only the lowest 12 standard pressure levels were selected for initialization, boundary conditions were updated in every 3 hours. These

files contain the horizontal and vertical wind components, specific and relative humidity, dew point, geopotential height, surface temperature and pressure. Soil moisture and temperature is available in four layers.

Surface and soil data

Figure 1 shows the soil types over the model area based on the FAO database. The DKSIS database (Szabó, J. *et al.* 2000; Pásztor, L. *et al.* 2010) *(Figure 2)* over Hungary was also used in performing simulations, while outside the country no changes were made in the FAO distribution. In both cases the dominant soil texture of the model area is loam and its variants.

The most prominent difference is the appearance of sand in the Danube–Tisza Interfluve, and at the eastern part of the country. Most of the clay loam disappears, resulting in an about 45 percent reduction when changing from FAO to DKSIS. In the new distribution, silt and its variants appear to be scattered over Hungary. The used land use database was the USGS *(Figure 3)* and the CORINE 2000 (European Environmental Agency, 2002) *(Figure 4)*. The previous one was determined using AVHRR measurements in 1992–1993, while the latter one by using Landsat-7 imagery in 2000.

In the WRF model the USGS land use is available at a 0.5°, while the CORINE was implemented at a 30″ horizontal resolution. According to USGS, two-thirds of the investigated area is "dryland cropland and pasture" and this suffers the greatest change, around 10 percent, when the USGS is replaced by CORINE. Areas occupied by deciduous broadleaf forests appear mostly on mountain ridges. The previously scattered cropland/woodland areas disappear almost entirely within the

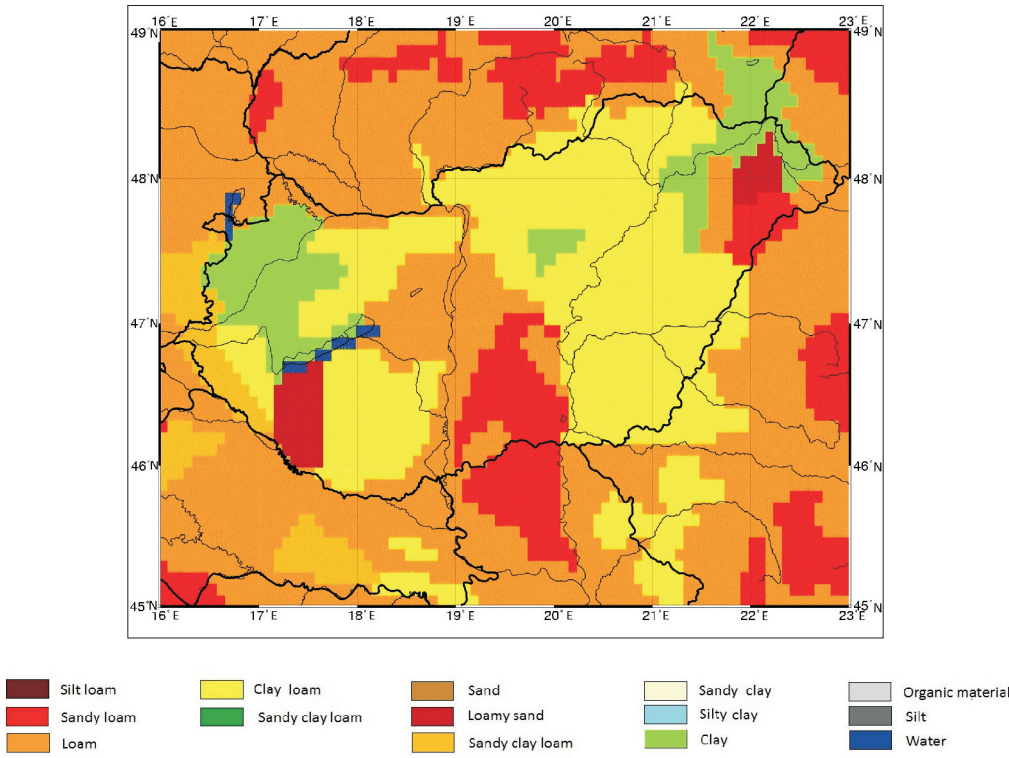


Fig. 1. Soil type over the model area based on FAO database

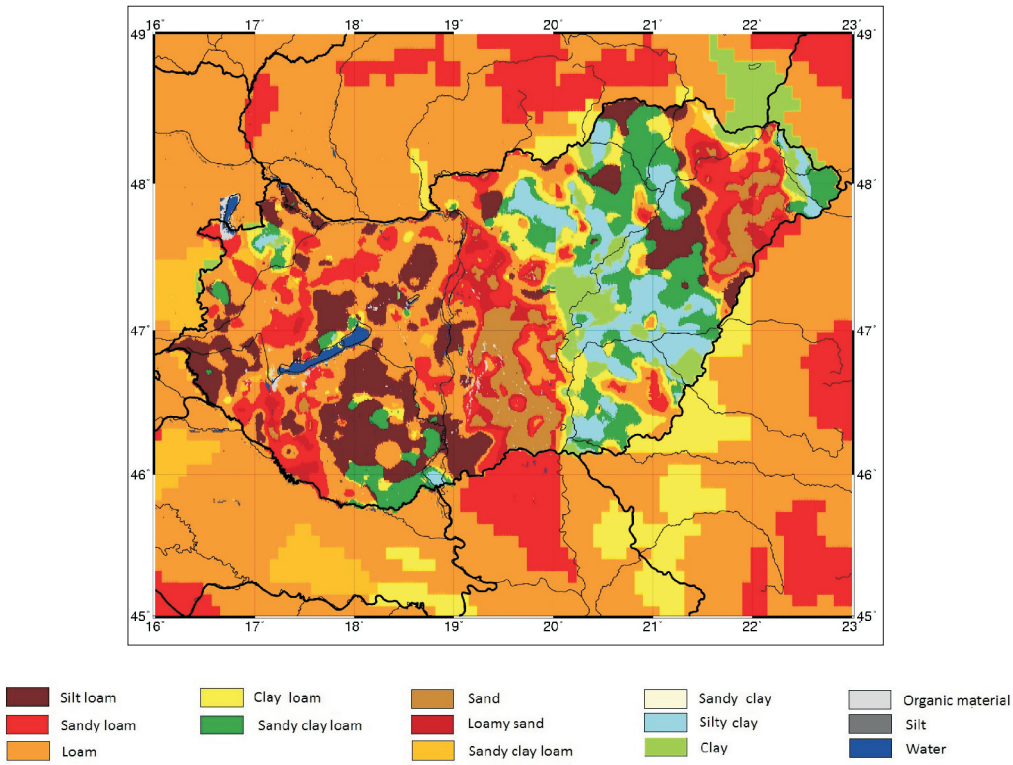


Fig. 2. Soil type over the model area based on DKSIS database in Hungary. Outside the country FAO database is used

country borders. The grass- and shrublands are more apparent in *Figure 4*. The ratio of urban and built-in land areas increased as well as the area of larger cities, such as Budapest and Bratislava, has become more visible. The size of water-covered areas is constant.

Weather

The convective precipitation was examined for 20 August 2006, when all weather conditions were favourable for its formation. The weather in Europe had been influenced by a large, well-developed cyclone, whose cold front has reached Hungary. The convective instability was further increased by the 25 m/s wind and the cold advection at the 500 hPa level. Above the Carpathian Basin a jet stream was flowing. The thunderstorm line has reached the Carpathian Basin in the late afternoon. The cold front reached Hungary at 16 UTC (Coordinated Universal Time) and has left it by 00 UTC on 21 August. Precipitation occurred in the northern part of the country, and local showers occurred Southeast. The majority of the precipitation occurred in the Northwest, the highest measured precipitation in Hungary was 17 mm at Kapuvár. The maximum temperature in Hungary was between 28 °C (North-Northwest) and 34 °C (Southeast) (Horváth, Á. 2006).

Results: temperature, latent heat flux

By changing soil texture map without changing meteorological (initial and boundary) conditions, the water holding capacity will also change, which defines the rate of evapotranspiration. Land use change results changes in the minimum stomatal resist-

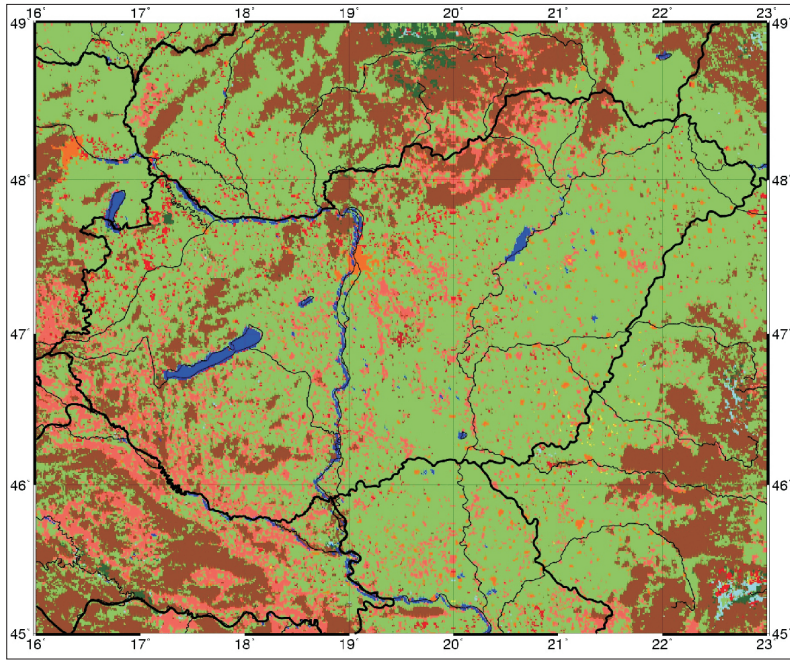


Fig. 3. Land use cover over Hungary based on USGS database

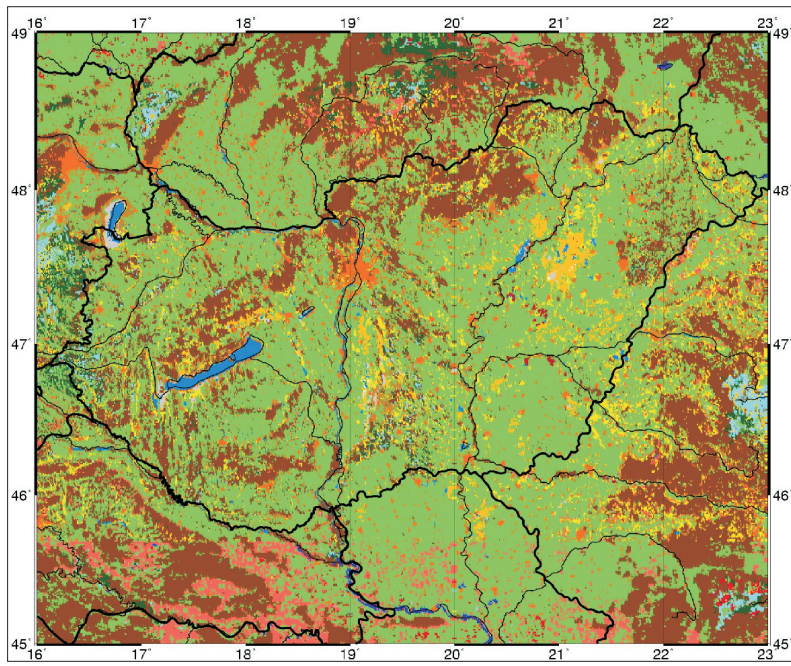


Fig. 4. Land use cover over Hungary based on CORINE database

ance also affecting the evapotranspiration. Furthermore changes in surface affect the albedo. These effects have an impact on near surface temperatures and atmospheric moisture content.

Both temperature and latent heat flux reach their daily maximum in the early afternoon, first the latent heat, then approximately an hour later the temperature caused by turbulent mixing. *Figure 5* shows latent heat flux (LH) and 2 m temperature (T2) differences between simulations obtained at 12 UTC, when the T2 changed from 28 °C to 34 °C over flat terrain, and the LH was between 200 W/m² and 360 W/m².

After changing soil type, the LH increased by 6.5 W/m² in spatial average *(Figure 5,* a*).* The T2 changed inversely with an average of 0.12 °C *(Figure 5,* b*).* Daytime average change was +3.48 W/m² and 0.1 °C, respectively. At the Danube–Tisza Interfluve, at the Small Plain (Kisalföld) and at Nyírség area the latent heat values increased. The cause of this is that the hydraulic properties of soils altered resulting in a decreased ability to hold moisture and therefore prompting an intensified evaporation. At the Kisalföld, clay has been modified to sandy loam or loam, while at Kiskunság sandy loam altered to sand. Over these areas the LH had increased by 20–40 W/m². In opposition to this, there was an observed decrease of 10 W/m² at Körös region due to more moisture being able to remain in the soil. Evaporation distracts heat from its surroundings leading to cooling, as it was seen at Kiskunság where the T2 change was approximately 1°C.

Areas outside Hungary, where the soil texture is unchanged, differences were caused by advection and especially North to Hungary this had an effect on cumulus cloud formation over the mountains resulting great differences in LH.

After switching the land use cover *(Figure 5,* c, d*),* warming of the model area by an average of 0.15 °C, and an approximate of 8 W/m² decrease in the LH was observed (daytime averages: +0.12 °C, –4.3 W/m²). Following the modification North to Lake Balaton, the broadleaf forest coverage increased, which caused a decrease in evaporation and in LH, mainly causing the increasing of the minimum stomatal resistance from 40 s/m to 100 s/m and the decrease of radiation stress function coefficient from 100 to 30. These areas show a decrease in the LH as indicated by red and orange dots.

At the Danube–Tisza Interfluve the T2 rose since the albedo decreased and the minimum stomatal resistance became stronger. The effect of the built-in areas reached its maximum at the early evening, when the T2 was 2.5–3 °C higher and the LH was 60–70 W/m² lower than the reference. The modification of soil types affected the results more than changes in the land use cover especially over areas where sandy texture replaced loam or clay. All four simulations show that the location of the changes moved slightly to the East because of the daylong westerly winds at the Hungarian Great Plain (Alföld), while these relocations shifted south at the northern parts of the country due to strong northerly winds.

Model used 12 different soil texture and 16 land use categories appearing in Hungary during calculations. In order to make the comparison, simpler groups of land use and soil texture with similar physical properties were created *(Table 1.* and *2*).

In each case, the altered land use was compared to the reference, shown in *Figures 6.* and *7.* By changing sand and its variants, a warming was caused (average 0.5 °C, max. 2 °C) before sunrise and after the arrival of the front. The initial amount of soil moisture during the simulations remained the same, thus by replacing e.g. loam with sand the available soil moisture increased. Due to the increment of water resources, a more intensive evaporation was caused and therefore a slight (approx. 0.4 °C average) cooling during the day.

Table 1. Created groups of soil type categories

Sand	Loam	Clay
Sand	Silt loam	Clay
Sandy clay loam	Loam	Silty clay
	Loamy sand	Clay loam
	Sandy loam	Sandy clay loam

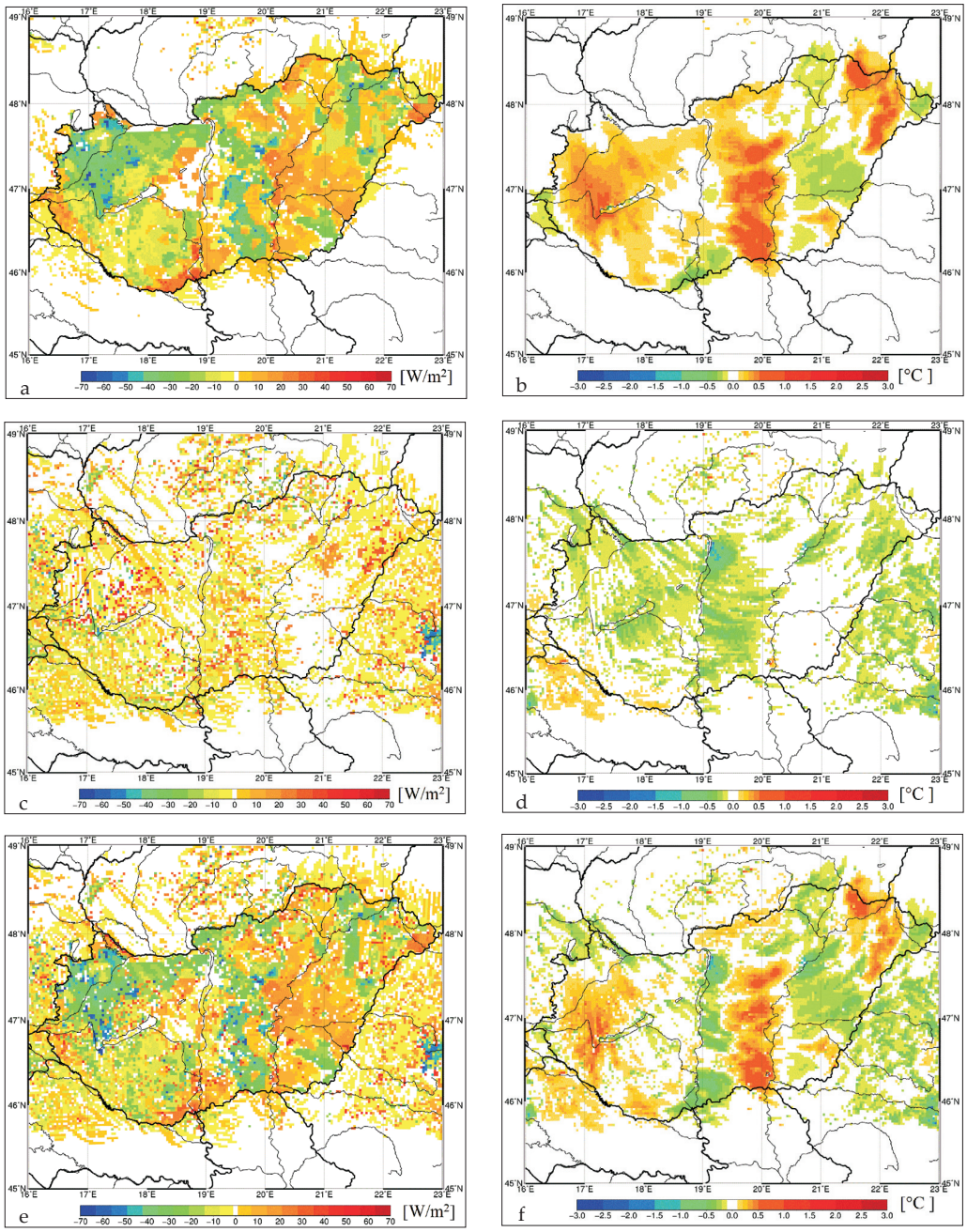


Fig. 5. Latent heat flux and 2 m temperature difference between REF-DKSIS (a, b), REF-CORINE (c, d) and REF-DKSIS&CORINE (e, f) simulations on 20th August 2006 at 12 UTC

the entire day, there was less intensive evaporation over the appearing grass vegetation type resulting in surface air warming with maximum at approximately 2 °C. In the case of the forested areas a difference of 0.5–1 °C can be observed.

Table 2. Created groups of land use categories

Urban, built-up area	Woodland	Shrubland	Grassland
Urban and built-up land	Deciduous broadleaf forest Evergreen needle leaf forest Mixed forest Wooded wetland	Shrubland Mixed shrubland/ grassland	Dryland, cropland and pasture Irrigated cropland and pasture Cropland/grassland mosaic Cropland/woodland mosaic Grassland

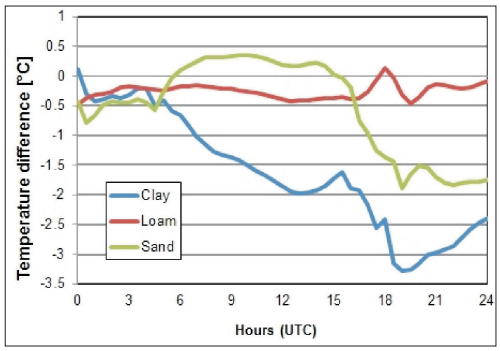


Fig. 6. 2 m temperature difference between reference and DKSIS in the case of combined soil textures on 20th August 2006

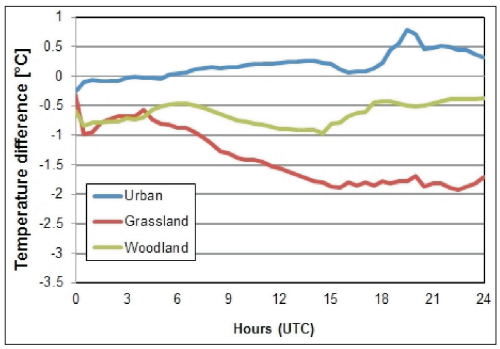


Fig. 7. 2 m temperature difference between reference and CORINE in the case of combined vegetation types on 20th August 2006

The steep change at around 16 UTC denotes the arrival of the cold front. At the loam categories no significant temperature difference was observed (avg. –0.3 °C), because the spatial ratio and distribution of this type remained approximately constant. The clay type had the biggest impact on the temperature with an average 3 °C warming. During

Convective Available Potential Energy (CAPE)

Convective Available Potential Energy is used to estimate the instability of the atmosphere through temperature and humidity of the atmosphere, which is partly controlled by the land surface. *Figure 8* shows the CAPE values at 12 UTC. The calculated values were 0–750/2,000 J/kg. The drier and colder air mass preceding the front had a CAPE of 0 J/kg (stable stratification, less possibility of storm). As the figures show, the CAPE has increased for DKSIS simulation, while the area by stable stratified air mass has decreased for CORINE.

Prior to the front's passage through the country, the DKSIS values were by 60–80 J/kg greater than the reference, while the CORINE's values remained below the reference by 120–160 J/kg. In the case of CORINE, the changes are consistent throughout the entire country, contrary to DKSIS, which concentrates on the sites of modification similarly to *Figure 5*, b. Due to wind the difference formed in band shapes. In the case of the DKSIS&CORINE simulation the absolute values of the changes did not exceed the maxima of the previous simulation differences, which was 240 J/kg. At some areas soil texture had a more significant effect on CAPE. At some parts of the model area, the processes of the two simulations amplified each other's effect.

The convective inhibition changed parallel as CAPE, the increasing latent heat resulting greater values of CIN. At around noon, the average values of CIN reached in absolute terms 60 J/kg. At the western part of the model area the inhibition was smaller, thus less CAPE was enough to form convective cloudi-

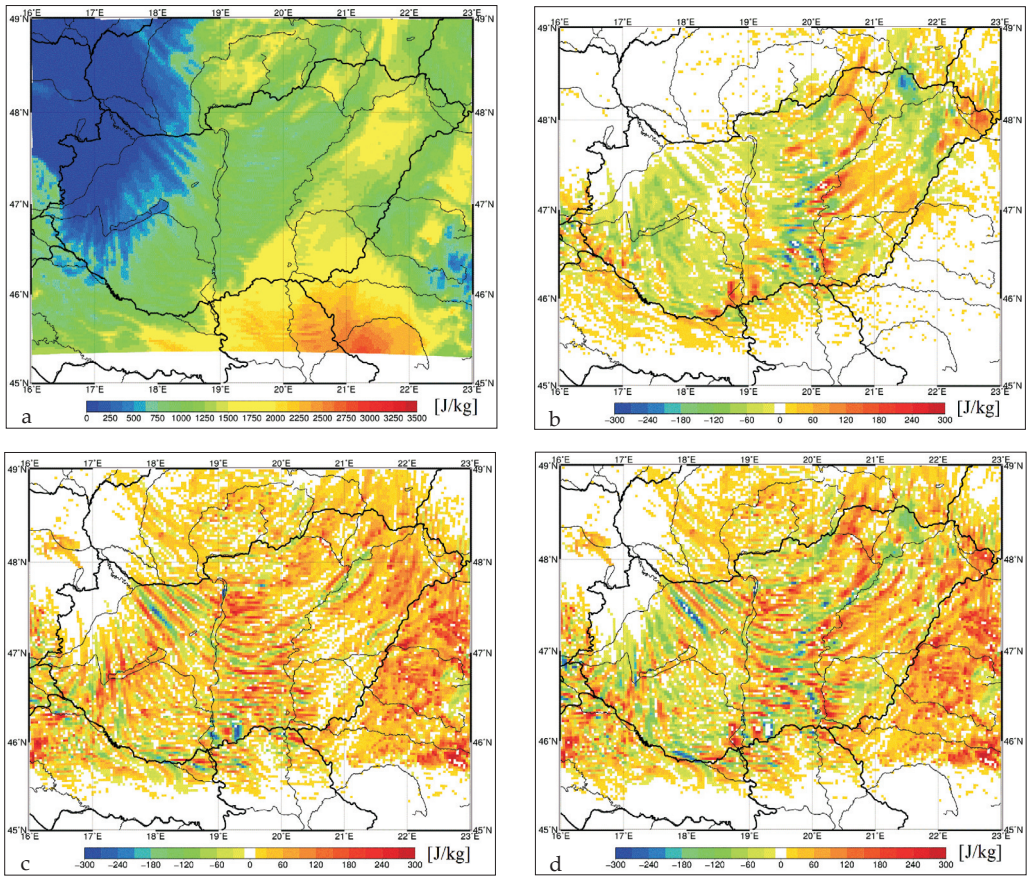


Fig. 8. Convective available potential energy CAPE (J/kg) for reference run (a), CAPE difference between REF-DKSIS (b), CAPE difference between REF-CORINE (c), and CAPE difference between REF-DKSIS&CORINE (d) at 12 UTC on 20th August 2006

ness. On the other part of the country the modification resulted greater CIN in each simulation. In case of breakthrough of CIN, stronger thunderstorms and precipitation may occur, when the CAPE's value is sufficiently large. That process has been observed at the central area, while CIN increased by 20–40 J/kg.

Precipitation *Figure 9* (a) shows the 24-hour accumulated precipitation. In the western parts of the country, there is a smaller deviation between the simulation and the actual measurements (not shown), however, East of the Danube these differences became larger, and translocated cells can be observed by the Northeast border. Southern part of the Alföld measurements showed the development of local convective systems, which didn't appear on the simulation even though the amount of CAPE would have allowed the formation.

24-hour accumulated precipitation was 8.75 mm on average-based on the reference simulation. The DKSIS simulation positively deviated from that value by 0.30 percent (*Figure 9*, b), while the DKSIS&CORINE (*Figure 9*, d) predicted a value 1.56 percent lower. The modification of land use cover (CORINE) caused a temperature increase resulting in higher vertical velocities leading to an overall 1.70 percent increase of precipitation (*Figure 9*, c) compared to the reference. Though this difference is neg-

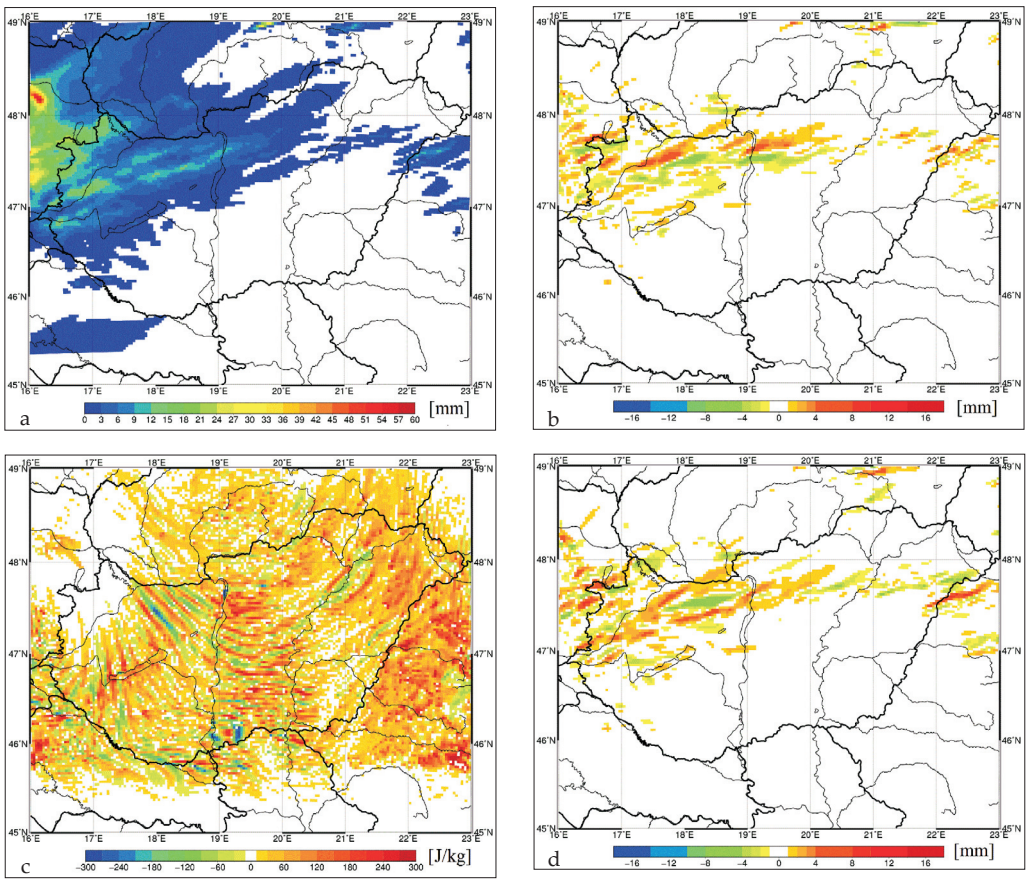


Fig. 9. 24-hour accumulated precipitation (PREC) reference run (a), PREC difference between REF-DKSIS (b), PREC difference between REF-CORINE (c) and PREC difference between REF-DKSIS&CORINE (d) on 20th August 2006

ligible, the migration of the precipitation band was faster with 30 minutes, leading to an East-west directional shift. Similar shift was noticeable in case of soil texture modification but it was in the North–South direction.

As it could be seen on accumulated precipitation differences, even if the total precipitation does not change significantly in the whole domain, locally ±8 mm/day changes can be seen which reaches 50 percent difference. These are usually related to a shift in precipitation cells with about 30 km; in most cases new precipitation cells appeared which weren't present in the reference run (e.g. northern border of the Alföld in *Figure 9,* d*).*

Conclusions

This study examined the effect of the land use cover and soil texture distribution change on convective precipitation, and on alterations in state variables affecting that. In the four simulations, the boundary conditions of the model area were modified. The evaluated data relates to the Hungarian land area over a 24-hour time period. In regard to the examined variables, the differences, in comparison to the reference, were larger in the case of the DKSIS than for the CORINE simulation.

At the middle of the day, the absolute values of the differences in latent heat for the DKSIS

were mostly close to 25–50 W/m^2, while for the change to CORINE they were around 15–25 W/m^2. Furthermore it has to be noted, that the DKSIS simulation was characterized by both a latent heat flux decrease and increase, while the CORINE simulation showed only decrease in latent heat flux over the model area. Simultaneously, the latent heat excess, a decrease in temperature was observed, resulting in an average of 0.5 °C temperature difference between the two simulations.

The combined effect of soil texture and land use change negated each other over some areas, but where the soil became more saturated the effect of soil texture change was dominant. For CAPE values a similar tendency could be observed. The CAPE values were 50–60 J/kg greater in the case of the soil texture modification than for the altered land use cover due to the latent heat difference. Considering precipitation there were no significant alterations in the 24-hour accumulated precipitation over the whole model area. However, the location of storm cells with the highest intensity shifted toward south in case of soil texture change, and an East–West horizontal tilt and shift was observable in case of land use change. As a result locally ±8 mm/day precipitation differences occurred. Also mostly in the case of land use change, new cells formed as a result of higher temperatures.

It can be stated that even though this weather event and the related precipitation was mainly related to cold front passage, the effect of switching to a more realistic soil texture and land use distribution is not negligible, not only for near surface temperatures but also for precipitation forecast. To determine whether the realistic distribution results more accurate forecast further analyses, using radar precipitation verification are needed.

Acknowledgements: The work was supported by OTKA (Hungarian Scientific Research Found) under contract number K 81432.

REFERENCES

Adegoke, J.O., Pielke, R. and Carleton, A.M. 2007. Observational and modelling studies of the impacts of agriculture related land use change on planetary boundary layer processes in the central U.S. *Agricultural and Forest Meteorology* 142. 203–215.

Alapaty, K., Raman, S. and Niyogi, D. 1997. Uncertainty in the specification of surface characteristics: a study of prediction errors in the boundary layer. Boundary–Layer *Meteorology* 82. 475–502.

Asharaf, S., Dobler, A. and Ahrens, B. 2011. Soil moisture initialization effects in the Indian monsoon system. *Advances in Science and Research* 6. 161–165. doi:10.5194/asr-6-161-2011.

Avissar, R. and Liu, Y. 1996. A three-dimensional numerical study of shallow convective clouds and precipitation induced by land-surface forcing. *Journal of Geophysical Research* 101. 7499–7518.

Breuer, H. 2012. *A talaj hidrofizikai tulajdonságainak hatása a konvektív csapadékra és a vízmérleg egyes összetevőire: meteorológiai és klimatológiai vizsgálatok Magyarországon* (The effect of soil hydro-physical properties on convective precipitation and on components of the water budget: meteorological and climatological investigations in Hungary). PhD Thesis, Budapest, Eötvös Loránd University, 117 p.

Breuer, H., Ács, F., Laza, B., Horváth, Á., Matyasovszky, I. and Rajkai, K. 2012. Sensitivity of MM5-simulated planetary boundary layer height to soil dataset: Comparison of soil and atmospheric effects. *Theoretical and Applied Climatology* 109. 577–590.

Chase, T.N., Pielke, R.A., Kittel, T.G.F., Nemani, R. and Running, S.W. 1996. Sensitivity of a general circulation model to global changes in leaf area index. *Journal of Geophysical Research* 101. 7393–7408.

Chase, T.N., Pielke, R.A., Kittel, T.G.F., Nemani, R. and Running, S.W. 2000. Simulated impacts of historical land cover changes on global climate. *Climate Dynamics* 16. 93–105.

Collow, T. W., Robock, A. and Wu, W. 2014. Influences of soil moisture and vegetation on convective precipitation forecasts over the United States Great Plains. *Journal of Geophysical Research: Atmospheres* 119. 9338–9358, doi:10.1002/2014JD021454.

DeAngelis, A., Dominguez, F., Fan, Y., Robock, A., Kustu, M. D. and Robinson, D. 2010. Observational evidence of enhanced precipitation due to irrigation over the Great Plains of the United States. *Journal of Geophysical Research* 115. D15115, doi:10.1029/2010JD013892.

Dickinson, R.E. 1984. Modelling evapotranspiration for three dimensional global climate models. In *Climate Processes and Climate Sensitivity* Eds.: Hansen, J.E. and Takahashi, T. Geophysical Monograph Series 29. Washington, DC, AGU, 58–72.

Drüszler, Á. 2011. *A 20. századi felszínborítás-változás meteorológiai hatásai Magyarországon* (Meteorological effects of the land cover types changes during the 20th century in Hungary). PhD Thesis, Sopron, University of West Hungary, 137 p.

Ek, M. and Cuenca, R.H. 1994. Variation in soil parameters: implications for modelling surface fluxes and atmospheric boundary-layer development. *Boundary–Layer Meteorology* 70. 369–383.

Eltahir, E.A. 1998. A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research* 34. (4): 765–776, doi:10.1029/97WR03499.

European Environment Agency 2002. *Corine Land Cover 2000 (CLC2000) seamless vector data.* Available at: http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-seamless-vector-database.

Hong, X., Leach, M.J. and Raman, S. 1995. A sensitivity study of convective cloud formation by vegetation forcing with different atmospheric conditions. *Journal of Applied Meteorology* 34. 2008–2028.

Horváth, Á. 2006. A 2006. augusztus 20-i budapesti vihar időjárási háttere (Meteorological background of the storm of 20th August, 2006 in Budapest). *Légkör* 51. (4): 24–27.

Horváth, Á. 2007. *Atmospheric convection*. Budapest, Hungarian Meteorological Service, 64 p.

Khodayar, S. and Schädler, G. 2013. The impact of soil moisture variability on seasonal convective precipitation simulations. Part II: sensitivity to land-surface models and prescribed soil type distributions. *Meteorologische Zeitschrift* 22. (4): 507–526.

Mölders, N. 2005. Plant – and soil – parameter – caused uncertainty of predicted surface fluxes. *Monthly Weather Review* 133. 3498–3516.

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.

Pielke, R. A. and Zeng, X. 1989. Influence on severe storm development of irrigated land. *National Weather Digest* 14. (2): 16–17.

Pielke, R.A. 2001. Influence of spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics* 39. 151–177.

Pielke, R.A. and Avissar, R. 1990. Influence of landscape structure on local and regional climate. *Landscape Ecology* 4. 133–155.

Pielke, R.A., Lee, T.J., Copeland, J.H., Eastman, J.L., Ziegler, C.L. and Finley, C.A. 1997. Use of USGS-provided data to improve weather and climate simulations. *Ecological Applications* 7. 3–21.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.-Y., Wang, W. and Powers, J.G. 2008. *A Description of the Advanced Research WRF* Version 3 NCAR/TN–475+STR, June 2008. NCAR Technical Note.

Szabó, J., Pásztor, L., Bakacsi, Zs., Zágoni, B. and Csökli, G. 2000. Kreybig Digitális Talajinformatikai Rendszer (Előzmények, térinformatikai megalapozás). (Kreybig Digital Soil Information System) (Preliminaries, GIS establishment). *Agrokémia és Talajtan* 49. 265–276.

Taylor, C.M. and Ellis, R.J. 2006. Satellite detection of soil moisture impacts on convection at the mesoscale. *Geophysical Research Letters* 33. L03404, doi:10.1029/2005GL025252.

Taylor, C.M., de Jeu, R.A.M., Guichard, F., Harris, P.P. and Dorigo, W.A. 2012. Afternoon rain more likely over drier soils. *Nature* 489. 423–426., doi:10.1038/nature11377.

Teuling, A.J., Uijlenhoet, R., van den Hurk, B. and Seneviratne, S.I. 2009. Parameter sensitivity in LSMs: an analysis using stochastic soil moisture models and ELDAS soil parameters. *Journal of Hydrometeorology* 10. 751–765.

EUGEO
BUDAPEST
2015

CONVERGENCES AND DIVERGENCES OF GEOGRAPHY IN EUROPE

Fifth EUGEO Congress on the Geography of Europe

30 August – 2 September 2015

BUDAPEST, Eötvös Loránd University

The programme will include keynote lectures, oral sessions, poster sessions, thematic panels, scientific and social events. Researchers and experts from all over the world are invited to submit proposals to: info@eugeo2015.com

The aim is to facilitate the exchange of new ideas and to offer opportunities for networking within an informal atmosphere. The Congress venues are centrally located and well connected by public transport.

PhD students and early career researchers are encouraged to actively participate and will be offered a reduced fee.

HUNGARIAN GEOGRAPHICAL SOCIETY

EUGEO

Further information and online registration:

www.eugeo2015.com – info@eugeo2015.com

DOI: 10.15201/hungeobull.64.1.2 Hungarian Geographical Bulletin **64** 2015 (1) 17–23.

Soil erodibility calculations based on different particle size distribution measurements

Csaba CENTERI[1], Zoltán SZALAI[2], Gergely JAKAB[2], Károly BARTA[3], Andrea FARSANG[3], Szilárd SZABÓ[4] and Zsolt BÍRÓ[5]

Abstract

In this study we focused on the factors affecting final outputs of the USLE (Universal Soil Loss Equation) model. In doing so, we conducted soil particle size measurements in different institutions (University of Debrecen, University of Szeged and Geographical Institute, Research Centre for Astronomy and Earth Sciences of the Hungarian Academy of Sciences) with a variety of methodologies (laser, aerometer and pipette methods) on various soil materials (sandy, loamy and clay). Statistical analyses of the eight examined soil samples have been shown some significant and some non-significant differences among the particle size measurements. This paper is aimed at i) to ascertain whether these significant differences in particle size measurements cause significant differences in soil erodibility calculations; and ii) to assess the amount of soil loss calculated by these K factors. The results suggest that regardless of the relatively small percentage between the smallest and the greatest K factor values, the amount of soil loss can be fairly high, especially when erosion occurs on a longer or steeper slope. In the present case, when we compare simulations results, the amount of soil loss is more important than the difference in percentage between the minimum and maximum values. Because the percentage of the difference can remain the same between the simulations, while the amount of soil loss increases way beyond soil loss tolerance limits.

Keywords: methods of particle size measurement, soil erodibility, USLE

Introduction

There has been a great deal of discussion about soils and their role in food production. Perhaps most importantly, soil is the main natural element from where the majority of the food for the human population originates. This topic becomes especially prescient because numerous scientists have declared that soil is a finite resource (Ángyán,

[1] Department of Nature Conservation and Landscape Ecology, Institute of Environmental and Landscape Management, Szent István University, H-2100 Gödöllő, Páter K. u. 1. E-mail: centeri.csaba@kti.szie.hu
[2] Geographical Institute, Research Centre for Astronomy and Earth Sciences, HAS, H-1112 Budapest, Budaörsi út 45. E-mails: szalai.zoltan@csfk.mta.hu, jakab.gergely@csfk.mta.hu
[3] Department of Physical Geography and Geoinfor-matics, University of Szeged, H-6720 Szeged, Egyetem u. 2. E-mails: farsang@geo.u-szeged.hu, barta@geo.u-szeged.hu
[4] Department of Physical Geography and Geoinfor-mation Systems, Debrecen University, H-4032 Debrecen, Egyetem tér 1. E-mail: szabo.szilard@science.unideb.hu
[5] Institute for Wildlife Conservation, Szent István University, H-2100, Gödöllő, Páter K. u. 1. E-mail: bzsolti@ns.vvt.gau.hu

J. 2001; Centeri, Cs. 2002; Centeri, Cs. *et al.* 2009, 2011, 2012; Madarász, B. *et al*. 2012). Therefore, understanding soil erosion in a more efficient and comprehensive way has a furthermost importance.

Soil stands in the focal point of soil erosion researches whose aims are primarily to protect this valuable resource (Kertész, Á. 1993; Szilassi, P. *et al.* 2006; Bádonyi, K. *et al*. 2008; Barczi, A. and Joó, K. 2009; Madarász, B. *et al*. 2011). When we are examining soil, it is done so from various points of view (Merinó, A. *et al.* 2004; Barczi, A. *et al.* 2009; Pető, Á. 2011; Fonseca, F. et al. 2012; Pető, Á. 2013; Kondrlová, E. *et al.* 2013). Soil erosion modelling is a useful tool for predicting potential amounts of soil loss (Rojas, R. *et al.* 2008; Heng, B.C.P. *et al.* 2011; Pradhan, B. *et al.* 2011). Soil erosion models must be examined *in situ* to obtain as much appropriate data as possible (Centeri, Cs. 2002; Centeri, Cs. *et. al.* 2009, 2011, 2012). Any additional data and research related to the increase of reliability of the models are most welcomed by model users (Madarász, B. *et al.* 2012). Soil particle size distribution is measured by various authors for various purposes (Su, Y.Z. *et al.* 2004). In the present case, the soil erodibility factor is analysed based on the liability of measuring an important input parameter, namely, the particle size distribution.

In the field of soil science, there has recently been a growing number of physically-based soil erosion models created and their application is rapidly increasing. As the input need of such physical models is much larger than those of the empirical models, any research investigating the reliability of factors affecting the final outputs of a model is valuable.

This research illustrates many effects of particle size measurements methods on soil erodibility factors of the USLE (Universal Soil Loss Equation) model. As particle size distribution is an important parameter for all other soil erosion models, these data can be used for other models as well (Giovannini, G. 2001).

Data and methods

Eight soil samples were chosen from seven different Hungarian locations of various soils (*Figure 1*). The samples represent a wide palette of soil textures and soil structures. In some cases there were no significant aggregating effects among the coarse particles. Other samples had higher clay contents with additional inorganic and humus colloids that resulted in more resistant aggregates (i.e. samples from the BOR, GFH and GAH).

Three institutions participated in the measurements and three methods were used. The codification of all information and basic geographical and other relevant parameters of the environment of the sample sites are available in *Table 1*.

Table 1. Codification of samples, sample sites and participating institutes

Code	Name of the participating institute
S	University of Szeged
D	University of Debrecen
F	Geographical Institute, RCAES HAS
Code	**Sample site information**
BOR	Börzsöny Mountains, mountain top
GAH	Gyöngyöstarján (Mátra Mountains)*
GFH	Gyöngyöstarján (Mátra Mountains)**
SZG	Szentgyörgyvár (Zala Hills)
TUR	Tura (Lowlands of Hatvan) ***
KMA	Kiskunmajsa (sandy lowland)
FES	Dabas (sandy lowland)
GAL	Galgahévíz (Lowlands of Hatvan) ***
Code	**Method of measurement**
A	Areometer
L	Laser method
P	Pipette method
P1	Pipette method, laboratory staff No. 1. (D)
P2	Pipette method, laboratory staff No. 2. (D)
Code	**Replicates**
1	Replicate 1
2	Replicate 2

*Lower third, **upper third of the slope.
***Along the Galga Stream

Measurements with the Laser Particle Sizer Analysette 22 MicroTech method

Sample preparation was carried out without OM (organic matter) takeout using sodium

Fig. 1. Origin of the eight soil samples from seven locations, in Hungary

pyrophosphate in order to disperse the aggregates into elemental particles. 20 g of air dried soil was dispersed in 25 ml (0.5n) sodium pyrophosphate for 24 hours.

The suspension was leached through a 500 μm sieve and measured in a diffractometer Laser Particle Sizer Analysette 22 (Fritsch GmbH Germany). The measuring range of the used unit (MicroTec) was 0.1–670 μm. The coarse fractions (>500 μm) were determined by sieving. The measuring unit of "Analysette 22" contains a helium-neon laser below 5 mW and a wavelength of 655 nm. A Fourier lens then gathered the diffracted beams onto the detector.

The apparatus uses the Mie-theory (Mie, G. 1908) to calculate grain-sizes from the intensity of the diffracted laser light. The results were classified into 102 size classes. One measurement was an average of 180 scans of the sample therefore no repetitions were applied.

Determination of particle size distribution with the Köhn-pipette method

Measurements were carried out according to Buzás, I. (1993), using the Hungarian patent of particle size distribution (MSZ-08-0205-1978). The method needs soil sample preparation (i.e. organic matter removed with H_2O_2, sieved with Ø = 0.2 mm mesh size). A mortar was applied with water and continuous rubbing.

The finest fractions were poured into a sedimentation vessel. This procedure was repeated until there were no fine particles in the mortar in which the whole sample was then washed into the vessel.

The suspension was filled up to 1,000 ml with distilled water and 10 ml 0.2 M sodium-oxalate was added to prevent coagulation. The settling time was calculated at 10 cm below the surface. Finally, after the finest (<0.001 mm) fraction had settled, the pipetted

samples were dried at 105 °C to determine their weight. Soils' particle size classes were expressed in percentage.

Determination of particle size distribution with the Aerometer method

This method is based on Stokes' law. Suspension is made from a 20–60 g sample. The moisture of the original sample is determined with gravimetry. To prevent coagulation, 0.5–1 g sodium-pyrophosphate is added to the suspension and then it is filled to 1,000 cm^3 with distilled water. The density of soil suspension measured at 30 s intervals for 24 hours by an aerometer (MSZ 14043/3: 1979; Buzás, I. 1993).

Calculation of soil erodibility values

Soil erodibility has been calculated with the following equation according to Schwertmann, U. *et al.* (1987):

$$K = 2.77 \cdot M^{1.14} \cdot 10^{-6} \cdot (12–OS) + 0.043 \cdot (A–2) + + 0.033 \cdot (4–D)$$

where M = (particle fraction between 0.063 mm and 0.002 mm [%] + particle fraction between 0.1 mm and 0.063 mm [%]) × (particle fraction between 0.063 mm and 0.002 mm [%] + particle fraction between 2.0 mm and 0.063 mm [%]) OS is the percentage content of organic substance (if $OS > 4\%$, $OS = 4\%$); A = aggregate category; D = category of permeability. In this case, $A = 2$ (soil aggregates are between 1–2 mm) and $D = 3$ (infiltration rate is between 10–40 $cm \cdot day^{-1}$) (Schwertmann, U. *et al.* 1987).

Parametrization of the USLE model

We used USLE model to check whether the soil erodibility values calculated with the measured particle size distributions in different institutions with different methodologies have an effect on the amount of soil loss. The following parameters were in the calculation: R factor = 1,300 (MJ mm ha^{-1} h^{-1} y^{-1}), $LS = 3.5$, $C = 0.5$ and $P = 1$.

Research findings

Results of K factor calculations with USLE methodology based on the particle size distribution measurements from 3 institutions (University of Debrecen, University of Szeged and Geographical Institute, Research Centre for Astronomy and Earth Sciences of the Hungarian Academy of Sciences), using 3 methods (laser, pipette and aerometer). The resulting K factor calculations are shown in *Figure 2*.

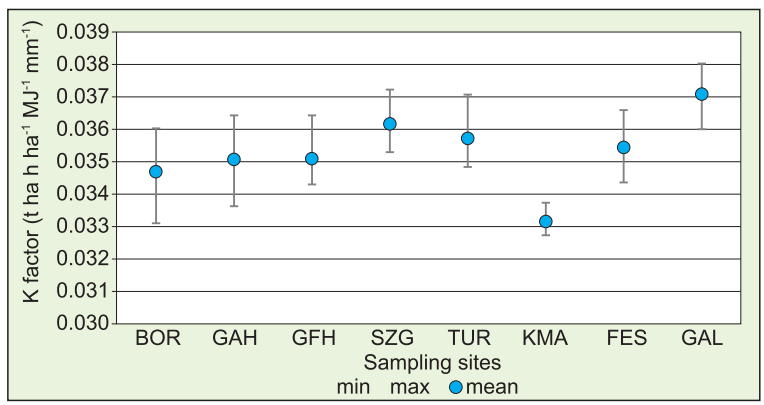


Fig. 2. Results of K factor calculations with USLE methodology including all 3 applied methods (laser, pipette and aerometer)

The calculated *K* factors (*Figure 2*) were used to calculate the amount of soil loss with the USLE model. The results of these calculations are in *Table 2*.

Based on the maximum and minimum values of soil loss calculations, the difference between these two values have been expressed in *Table 3* below. This *Table* shows the differences where the basis was the minimum value, so the percentage is expressing the difference of the maximum value compared to the minimum value (i.e. 6.1% means that the max. value is 6.1% higher than the min. value). The statistical ana-lyses proved that there were no differences in the measurements of the particle size distribution in case of KMA.

The differences between the amounts of soil loss calculated with the measured particle size classes resulted in very small (0.4%) difference bet-ween the smallest and the greatest amount of soil loss. The highest difference of the measured values was 6.1 percent, which can also be regarded as fairly low.

However, if we take into account the soil loss and not the percentage. We have to state that the amount of soil loss with the given parameterization is quite great, exceeding 70 t^{-1} ha^{-1} y^{-1}.

In the case, soil loss simulations on longer or steeper slopes, the difference between the smallest and the greatest amount of soil loss can grow to threefold. Therefore, this is a factor that must be considered as a tremendous increase in the amount of soil loss.

Conclusion

The analyses of the effects of particle size measurements methods proved that there can be considerable differences among the calculated soil losses if we use different par-

Table 2. Amount of soil losses calculated with the different K factors in using the results of the particle size distributions measured with different methods

Site code	Values	Soil loss, t^{-1} ha^{-1} y^{-1}	Site code	Values	Soil loss, t^{-1} ha^{-1} y^{-1}
BOR	Minimum	76.2	TUR	Minimum	80.2
	Maximum	81.0		Maximum	83.4
	Mean	78.9		Mean	81.3
GAH	Minimum	77.4	KMA	Minimum	75.4
	Maximum	81.9		Maximum	75.8
	Mean	79.8		Mean	75.4
GFH	Minimum	78.9	FES	Minimum	79.1
	Maximum	81.9		Maximum	82.2
	Mean	79.9		Mean	80.6
SZG	Minimum	81.2	GAL	Minimum	82.8
	Maximum	83.7		Maximum	85.5
	Mean	82.3		Mean	84.4

Table 3. Differences in the amount of soil losses calculated with the different K factors by using the results of the particle size distributions measured with different methods

Site code	$Values_{max}$ vs.$Values_{min}$, %	Site code	$Values_{max}$ vs.$Values_{min}$, %
BOR	6.1	TUR	3.9
GAH	5.7	KMA	0.4
GFH	3.7	FES	3.9
SZG	3.0	GAL	3.2

ticle size measurement methods to assess the soil erodibility factor and use these factors in the USLE model to calculate the amount of soil losses.

We therefore conclude that, the method of particle size measurement do have an effect on soil erodibility factors and thus, also on the amount of the calculated soil losses, regardless of the fact that in this study there were no analyses of significance on the soil erodibility and soil loss calculations.

REFERENCES

Ángyán, J. 2001. *Az európai agrármodell, a magyar útkeresés és a környezetgazdálkodás* (The European agrarian-model, Hungarian road finding and agri-environmental management). Budapest, Agroinform Kiadóház, 308 p.

Bádonyi, K., Madarász, B., Kertész, Á. and Csepinszky, B. 2008. Talajművelési módok és a talajerózió kapcsolatának vizsgálata zalai mintaterületen (Study of the relationship between tillage methods and soil erosion on an experimental site in Zala County). *Földrajzi Értesítő / Hungarian Geographical Bulletin* 57. (1–2): 147–167.

Barczi, A. and Joó, K. 2009. The role of kurgans in the Palaeopedological and Palaeoecological reconstruction of the Hungarian Great Plain. *Zeitschrift für Geomorphologie Supplementbaende* 53. (1): 131–137.

Barczi, A., Golyeva, A.A. and Pető, Á. 2009. Palaeoenvironmental reconstruction of Hungarian kurgans on the basis of the examination of paleosoils and phytolith analysis. *Quaternary International* 193. (1–2): 49–60.

Buzás, I. 1993. A talajfizikai, vízgazdálkodási és ásványtani vizsgálata (Physical, hidrological and mineralogy analyses of soils). In *Talaj- és agrokémiai vizsgálati módszerkönyv 1.* Ed.: Buzás, I., Budapest, Inda 4231 Publishing, 37–42.

Centeri, Cs. 2002. A talajerodálhatóság terepi mérése és hatása a talajvédő vetésforgó kiválasztására (Measuring soil erodibility on the field and its effects on soil protecting crop rotation). *Növénytermelés* 51. (2): 211–222.

Centeri, Cs., Akác, A. and Jakab, G. 2012. Land Use Change and Soil Degradation in a Nature Protected Area of East-Central Europe. In *Land Use: Planning, Regulations, and Environment.* Eds.: Aubrecht, C., Sergio Freire, S. and Steinnocherpp, K. New York, NOVA Science Publisher, 211–241.

Centeri, Cs., Herczeg, E., Vona, M., Balázs, K. and Penksza, K. 2009. **T**he effects of land-use change on plant-soil-erosion relations, Nyereg Hill, Hungary. *Journal of Plant Nutrition and Soil Science* 172. (4): 586–592.

Centeri, Cs., Kristóf, D., Evelpidou, N., Vassilopoulos, A., Giotitsas, I. and Varvarigos, G. 2011. Soil erosion risk and sediment transport within Paros Island, Greece. In *Soil Erosion: Causes, Processes and Effects.* Ed.: Fournier, A.J., New York, NOVA Science Publisher, 219–234.

Fonseca, F., de Figueiredo, T. and Bompastor Ramos, M.A. 2012. Carbon storage in the Mediterranean upland shrub communities of Montesinho Natural Park, northeast of Portugal. *Agroforestry Systems* 86. (3): 463–475.

Giovannini, G., Vallejo, R., Lucchesi, S., Bautista, S., Ciompi, S. and Llovet, J. 2001. Effects of land use and eventual fire on soil erodibility in dry Mediterranean conditions. *Forest Ecology and Management* 147. (1): 15–23.

Heng, B.C.P., Sander, G.C., Armstrong, A., Quinton, J.N., Chandler, J.H. and Scott, C.F. 2011. Modelling the dynamics of soil erosion and size-selective sediment transport over non-uniform topography in flume-scale experiments. *Water Resources Research* 47. 1–11.

Kertész, Á. 1993. Application of GIS methods in soil erosion modelling. *Computers, Environment and Urban Systems* 17. 233–238.

Kondrlová, E., Igaz, D., Horák, J. and Halászová, K. 2013. Soil texture analysis by optical method laboratory experiment on sample preparation prior to analysis. In *Water resources. Forest, marine and ocean ecosystems.* Proceedings of the 13th International Multidisciplinary Scientific Geoconference SGEM, Sofia, 677–683.

Madarász, B., Bádonyi, K., Csepinszky, B., Mika, J. and Kertész, Á. 2011. Conservation tillage for rational water management and soil conservation. *Hungarian Geographical Bulletin* 60. (2): 117–133.

Madarász, B., Jakab, G., Szalai, Z. and Juhos, K. 2012. Lézeres szemcseösszetétel elemzés néhány előkészítő eljárásának vizsgálata nagy szervesanyag-tartalmú talajokon (Examination of sample preparation methods for the laser grain size analysis of soils with high organic matter content). *Agrokémia és Talajtan* 61. 381–398.

Merino, A., Pérez-Batallón, P. and Macias, F. 2004. Responses of soil organic matter and greenhouse gas fluxes to soil management and land use changes in a humid temperate region of southern Europe. *Soil Biology and Chemistry* 36. (6): 917–925.

Mie, G. 1908. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen., Leipzig. *Annalen der. Physik*. 330. pp. 377–445. MSZ 14043/3: 1979. MSZ-08-0205-1978.

Pető, Á. 2011. Hazai talajszelvények fitolit morfotípus-diverzitása (Morphotype diversity of phytoliths in Hungarian soil profiles). *Agrokémia és Talajtan* 60. (1): 45–64.

Pető, Á. 2013. Studying modern soil profiles of different landscape zones in Hungary: an attempt to establish a soil-phytolith identification key. *Quaternary International* 287. 149–161.

Pradhan, B., Chaudhari, A., Adinarayana, J. and Buchroithner, M.F. 2011. Soil erosion assessment and its correlation with landslide events using remote sensing data and GIS: a case study at Penang Island, Malaysia. *Environmental Monitoring and Assessment* 184. (2): 715–727.

Rojas, R., Velleux, M., Julien, P.Y. and Johnson, B.E. 2008. Grid scale effects on watershed soil erosion models. *Journal of Hydrologic Engineering* 13. 793–802.

Schwertmann, U., Vogl, W., Kainz, M., Auerswald, K. and Martin, M. 1987. *Bodenerosion durch Wasser*. Stuttgart, Ulmer, 64 p.

Su, Y.Z., Zhao, H.L., Zhao, W.Z. and Zhang, T.H. 2004. Fractal features of soil particle size distribution and the implication for indicating desertification. *Geoderma* 122. (1): 43–49.

Szilassi, P., Jordan, G., van Rompaey, A. and Csillag, G. 2006. Impacts of historical land use changes on erosion and agricultural soil properties in the Kali Basin at Lake Balaton, Hungary. *Catena* 68. (3): 96–108.

Changing Ethnic Patterns of the Carpatho–Pannonian Area from the Late 15th until the Early 21st Century

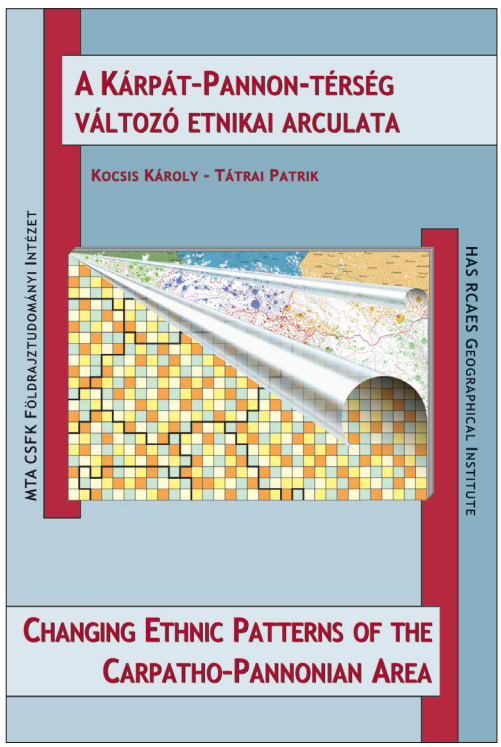
Edited by: Károly KOCSIS and Patrik TÁTRAI

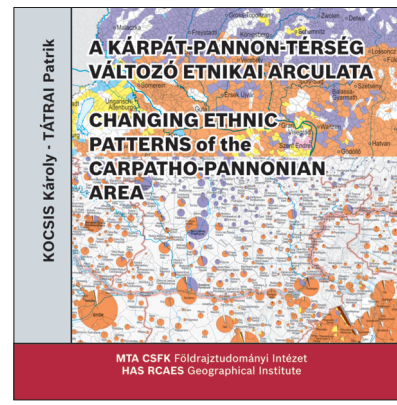
Hungarian Academy of Sciences, Research Centre for Astronomy and Earth Sciences
Budapest, 2013.

This is a collection of maps that visually introduces the changing ethnic patterns of the ethnically, religiously, culturally unique and diverse Carpathian Basin and its neighbourhood, the Carpatho-Pannonian area.

The Hungarian and English volume consist of three structural units. On the main map, pie charts depict the ethnic structure of the settlements in proportion to the population based on census data et the millennium. In the supplementary maps, changes of the ethnic structure can be seen at nine dates (in 1495, 1784, 1880, 1910, 1930, 1941, 1960, 1990 and 2001). The third unit of the work is the accompanying text, which outlines the ethnic trends of the past five hundred years in the studied area.

The antecedent of this publication is the „series of ethnic maps" published by the Geographical Research Institute of the Hungarian Academy of Sciences from the middle of the 1990's, which displayed each of the regions of the Carpathian Basin (in order of publication: Transylvania, Slovakia, Transcarpathia, Pannonian Croatia, Vojvodina, Transmura Region, Burgenland, Hungary). This work represents, on the one hand, the updated and revised version of these areas, and, on the other hand, regions beyond the Carpathian Basin not included on previous maps. Thus, the reader can browse ethnic data of some thirty thousand settlements in different maps.





Price: EUR 12.00
Order: Geographical Institute RCAES HAS Library
H-1112 Budapest, Budaörsi út 45.
E-mail: magyar.arpad@csfk.mta.hu

DOI: 10.15201/hungeobull.64.1.3 Hungarian Geographical Bulletin **64** 2015 (1) 25–34.

Spatial and temporal heterogeneity of runoff and soil loss dynamics under simulated rainfall

Judit SZABÓ[1], Gergely JAKAB[2] and Boglárka SZABÓ[3]

Abstract

The factors affecting soil erosion processes are complex and various, comprises two phases: detachment and transport by water. Previous studies indicated that initial moisture content, slope and soil crusts are playing an important role in soil erosion. The primary objectives of this study were to examine the sediment concentration and aggregate size distribution of the washed sediment. Aims were also to create different season specifically modelled situations in order to check runoff rates on bare soils under heavy rainfall. The experiments were conducted with a laboratory-scale rainfall simulator using a 1/2 HH 40 WSQ fulljet nozzle on eutric calcaric Cambisol loamic. Altogether, 72 soil loss samples were collected (6 separate precipitations, 3 time periods, 4 particle size fractions). The experiments indicated that the runoff rate was not increased by the presence of soil crusts, and even less sediment occurs on crusted surfaces. This sediment contained smaller fractions compared to recently tilled surface. The sediment concentration increased with the slope angle, but the runoff rates probably depend rather on the micro-morphology and initial moisture content of the surface. The main erosion process is the raindrop erosion after inland inundation and drought in gentle slopes, while the intermediate period of the precipitation is the most erosive. In general, the ratio of the macro aggregates in soil losses decrease and the ratio of the smaller fractions increase with the time during a precipitation event. Changing climate conditions are shown to have an effect on agricultural production through the temporal and spatial distribution of the erosion rates.

Keywords: soil erosion, rainfall simulation, runoff, aggregate size

Introduction

The soil loss by erosion is a widespread problem in agricultural areas. Soil erosion processes are affected by complex and various factors, including two phases: detachment and transport by water. A laboratory-scale rainfall simulator is an ideal tool for examining both phases of soil erosion on arable soils since most of the influential factors can be simulated and examined by its help. The advantages of the laboratory scale rainfall simulators are the

[1] Department of Environmental and Landscape Geography, Eötvös Loránd University, Faculty of Science. H-1117 Budapest, Pázmány Péter sétány 1/C. E-mail: szabojuditalexandra@gmail.com
[2] Geographical Institute, Research Centre for Astronomy and Earth Sciences, HAS. Budapest, H-1112 Budapest, Budaörsi út 45. E-mail: jakabg@mtafki.hu
[3] Department of Nature Conservation and Landscape Ecology, Szent István University, Faculty of Agricultural and Environmental Sciences. H-2100, Gödöllő, Páter K. u. 1. E-mail: bogi87@gmail.com

followings: the effect of the soil heterogeneity is negligible, easy to concentrate on one or two factors and this is a fast method (Grismer, M.E. 2010). According to Meyer, L.D. (1965) "The use of rainfall simulators generally provides a more rapid, efficient, controlled and adaptable tool than natural rainfall."

The first rainfall simulator in Hungary was designed by Kazó, B. (1966) in order to study infiltration, while Kerényi, A. (1986) concentrated on the role of initial erosion. Experiments were conducted to determine the erodibility ("factor K" in USLE equitation) of different Hungarian soils using a field scale rainfall simulator (Centeri, Cs. and Császár, A. 2003; Centeri, Cs. and Pataki, R. 2003; Jakab, G. and Szalai, Z. 2005; Kertész, Á. and Centeri, Cs. 2006; Centeri Cs. *et al.* 2011).

Aggregates are groups of soil particles that are bound to each other. Their pattern – soil structure – has an influence on the physical and chemical processes of soils. Aggregate stability is one of the most important properties, which indicate soil resistivity against external effects as raindrop impacted aggregate breakdown (Kerényi, A. 1986; Le Bissonnais, Y. *et al*. 1989). The aggregate breakdown process is also connected with crust formation (West L.T. *et al*. 1992) and thus effects erosion rates.

Erosion studies usually use the measurement of the sediment concentrations, runoff rates and aggregate stability in order to examine the effect of the slope, initial moisture content, rainfall intensity, effect of the crust or the surface roughness on erosion rates (Jin, K. *et al*. 2008; Defersha, M.B. and Melesse, A.M. 2012).

The literature on simulated soil erosion experiments suggests several approaches. Defersha, M.B. and Melesse, A.M. (2012) examined the effect of the initial moisture content and slope steepness on erosion, whereas Le Bissonnais,Y. *et al*. (1989) studied the aggregate breakdown mechanism and soil crusting on pre-wetted and air-dry soils. Jomaa, S., *et al.* (2012) concentrated on initial moisture contents and on the effect of surficial rock fragments during rain splash erosion. Gómez, J.A., and Nearing, M.A. (2005) analysed the natural surface roughness while the effects of the impervious surface runoff were studied by Pappas, E.A. *et al.* (2008). Both provided different points of view regarding surface roughness. Mohammad, A. and Adam, M. (2010) concentrated on the effect of vegetation and land use. However, the major problem with this kind of applications is that various simulators and methodologies exist. All the essential variables that indicate sediment dynamics are to be taken into consideration, especially sediment concentrations, sediment yield and transportability of soil particles (Defersha, M.B. and Melesse, A.M. (2012).

Nearing, M.A. *et al.* (2005) modelled different situations to present how the soil responses to potential climate change. Among others Nearing, M.A. *et al.* (2005) pointed out on the climate change impact of runoff and erosion, as the increasing rainfall intensity and rainfall amount together "will have greater impact on runoff and erosion than changes in rainfall amount alone." Recently, the meteorological conditions in Hungary got more and more inordinate. The probability of drought occurrence is estimated to increase; the wettest months are April and May while the driest are July and August (Bartholy, J. *et al.* 2014). Extreme rainfalls occur more often therefore agricultural areas will be potentially endangered by water erosion in a much wider range. This risk alternates during one year according to the different seasons.

The primary objectives of this study were to examine the sediment concentration and aggregate size distribution of the soil loss and to create different season-specific modelled situations in order to check the runoff rates. Three approaches were in the focus of the experiments on bare soil under heavy rainfall:

(1) Sedimentary crust formed after a precipitation event (West, L.T. *et al*. 1992). The effect of this crust on sediment concentration was examined in case of two different slope steepnesses (5% and 12%) by applying

two simulated rainfalls within following two days.

(2) Two extreme soil moisture contents. The role of inland inundation and the drought were studied related to the changing climate conditions.

(3) Periods of each precipitation were studied beside the seasonal aspect. Aims were to compare the runoff dynamics and aggregate size distribution of the soil loss on different surfaces.

Rainfall simulation

Three fundamental criteria are commonly considered in designing a rainfall simulator (Hall, M.J. 1970), namely,

(1) the control of application rates in both time and space,

(2) the reproduction of drop-size distributions observed in different intensities of natural rainfall at the corresponding application rates,

(3) the reproduction of the terminal velocities of drops in natural rainfall.

Grismer, M.E. (2010) summarized the rainfall simulation methodology, the simulator types, the erosion models and the rainfall characteristics. Several authors reported many types of laboratory scale rainfall simulators those can be used for research (e.g. Le Bissonnais, Y. *et al.* 1989; Gómez, J.A. and Nearing, M.A. 2005; Pappas, E.A. *et al.* 2008; Aksoy, H. *et al.* 2012; Defersha, M.B. and Melesse, A.M. 2012). According to Grismer, M. E (2010) the 80 percent of the simulators (both field and laboratory) are nozzle type simulators.

In this study, a laboratory rainfall simulation procedure was developed and utilized to examine aggregate size distribution of the soil loss and the runoff rates during the precipitation on different surfaces, but using the same soil. Our laboratory scale rainfall simulator is situated in Eötvös Loránd University, Faculty of Science, Budapest. The first stage of the simulator was designed by Zámbó and Weidinger (Zámbó, L. and Weidinger, T. 2006) (*Photo 1*).

For the first time, it had only an individual (pin) drop-former system 9 m above the monolith. In these days 1/2 HH 40 WSQ fulljet nozzle, 1/2 HH 50 WSQ fulljet nozzle can be used, too. The soil sample flume is 0.5 m × 1.0 m × 0.2 m (0.1 m^3) and its steepness is adjustable (0–40%). There are four taps on the bottom of the flume, so the leached water can be collected as well. The soft water from the plumbing runs through a pressure regulator system therefore no water tank is needed during the simulations.

Photo 1. The rainfall simulator: the rainfall simulator viewed from the front (a), drop-former system (b), the examined soil viewed from above with falling droplets (c), and the saturated soil (d)

Methodology

Rain simulation

This paper presents data of six treatment combinations: recently tilled (T) and crusty soil surface (C) on two different slope steepness (5; 12), inland inundation (II) and drought (D) soil conditions on 2 percent slope steepness (*Table 1*). For each treatment the same 20 cm thick soil sample was packed into the flume overlaying a geotextile. 5T, 12T 2D treatments were applied on initially dry soil conditions, 5C, 12C treatments were applied on field capacity water content soil and 2II treatment was applied on the three weeks saturated soil.

The distributed soil was taken in Ceglédbercel, Hungary (N47.249765°, E19.678761°, 150 m a.s.l.). The mean annual temperature in the studied area is 10.8 °C and the annual precipitation is around 600 mm (Dövényi, Z. 2010). The eroded eutric calcaric Cambisol loamic has 18.8 percent of $CaCO_3$, the total organic carbon (TOC) content is 1 percent, and the pH is 7.5.

Drop forming nozzle system was chosen to examine the effect of intensive rainfall under different seasonal situations. Later, the same nozzle system can be used during field experiments in order to compare the results. The experiments were conducted in the laboratory using a 1/2 HH 40 WSQ fulljet nozzle which is widely used in rainfall simulation studies (Strauss, P. *et al*. 2000; Armstrong, Q. and Quinton, J.N. 2009). Since the cone basis of this nozzle was four times larger than the size of the monolith only one nozzle was applied.

The simulated rainfall characteristics depend on the nozzle type used and the pressure applied. According to the measurements of Strauss, P. *et al.* (2000) the kinetic energy of the rainfall simulated by the nozzle 1/2 HH 40 WSQ is 17 kJ m^{-2} mm^{-1} at 20 kPa. This value would correspond to approximately 65 percent of the kinetic energy of natural rainfall with the intensity of 50 mm h^{-1}. *Figure 1* shows

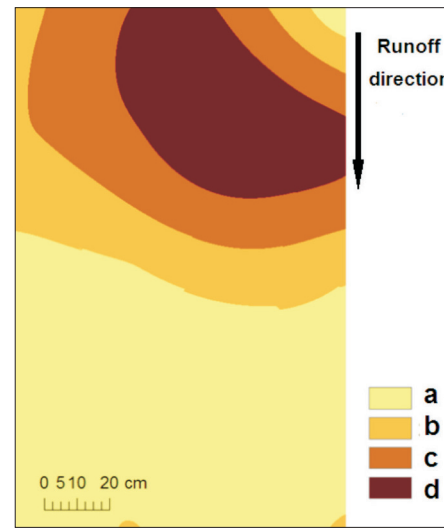


Fig. 1. Spatial distribution of simulated rainfall in mm after 20 minutes precipitation. The average intensity is 80 mm/h. a = 26–28 mm; b = 28–30 mm; c = 30–32 mm; d = 32–34 mm

Table 1. Details of the six treatments

Simulation code	Slope steepness %	Surface	Time min′ sec″	Energy kJ m^{-2} mm^{-1}	CU %	Median drop size mm	Aim of the treatment
5T	5	Recently tilled	42′11″	17	98	1.95	Bare soils in autumn and spring
5C	5	Crusty	32′43′′	17	97	1.95	Effect of the crust
12T	12	Recently tilled	33′34″	17	94	1.95	Bare soils in autumn and spring
12C	12	Crusty	8′45″	17	97	1.95	Effect of the crust
2II	2	After inland inundation	22′25″	17	93	1.95	Extreme situation (summer)
2D	2	After drought simulation	29′22″	17	96	1.95	Extreme situation (summer)

the spatial distribution of simulated rainfall in mm after 20 minutes precipitation. The average intensity is 80 mm h^{-1}. Simulated rainfall intensity was held constantly for these tests. The Christiansen's uniformity coefficient (CU) (Christiansen, J.E. 1942) which determine the uniformity of a sprinkler system was over 90 percent in case of all the six simulations, hence, the rainfall can be considered both spatially and temporary uniform.

Each runoff event was divided into three temporal phases, therefore the eroded sediment was collected in three periods (I, II, III) at three litres of the runoff intervals during the precipitation. Aims were to detect the temporal changes in aggregate sizes. Each three litre runoff phase soil loss was collected through sieve series with the following openings: 1 mm, 250 μm and 50 μm to a bucket. Therefore, by measuring the mass of fractions, the scale of the different aggregates is obtained directly. The capacity limitation of the sieves was equal to the sediment amount of three litres of water. During this study *a* (2%), *b* (5%) and *c* (12%) slope steepness were applied which represented:

a) flat plain where the inland inundation took place,

b) the average steepness of the agricultural areas in Hungary,

c) the suggested steepest slope in arable land areas.

Altogether 72 sample were collected (6 treatments × 3 time periods × 4 aggregate size fractions). The samples were used to calculate sediment concentration. Four sample repetitions of the untreated soil were also separated by this sieve system in prior to the measurements using the wet sieving method of Kemper, D.W. and Rosenau, R.C. (1986) as a control.

The time was recorded after every 1 litre collected runoff. In the experiment 12C only three litres of runoff were collected (one period of the precipitation was represented by one litre runoff) because the high amount of the sediment on the sieves. In the experiment 12T the time and litre data had to be corrected subsequently because of sieve sealing. Weights of dried soil losses were recorded at the nearest 0.01 g.

Results and discussion

The changes of necessary time for 1 litre runoff are presented on *Figure 2.* There are two points in the zero line, the first point represents the time when runoff was started and the second means the time of surface ponding without any runoff. Runoff started after a twice longer period in case of 2D when the drought was simulated, because of the formation of big rifts according to the arid period, and because these needed to be infilled first. Shortest time was needed for the runoff in case of 2II, when the inland inundation was simulated. Runoff starts almost immediately, which means significantly 10 minutes differences (one fourth of the total time) *(Table 1).*

To compare the 5T-5C and 12T-12C cases, the runoff started earlier from the crusty surfaces. This was due to the bigger initial moisture content on the crusty surfaces compared to the tilled surfaces.

Three different runoff periods are separated on *Figure 2*. As first stage the curves are positioned in the zero line, without any runoff yet. The curves are different, because of the changing runoff rates and thus infiltration rates till the runoff of the third litre suspension. The ratio of the runoff and infiltration is constant during the third period (after the runoff of the third litre). The curves turn straight, which means that runoff and infiltration are in balance. The equations on *Figure 2* show the third runoff periods of the 5T, 5C and 2II, 2D treatments. The curves of the 5T and 5C have the same steepness, which means that crust evaluated this way is not characterized by any influences on runoff rates except the length of the time period before the runoff changed. The curves of 12T, 12 C, 2II and 2D treatments are steeper than 5T and 5C curves, therefore a higher infiltration rate is presumed in the latter cases. High runoff rate from 2 percent slope are assumed to be related to the degraded soil structure, but more experiments are needed in order to prove this phenomenon. The results of 12T were corrected and at 12C the duration of the precipitation was too short.

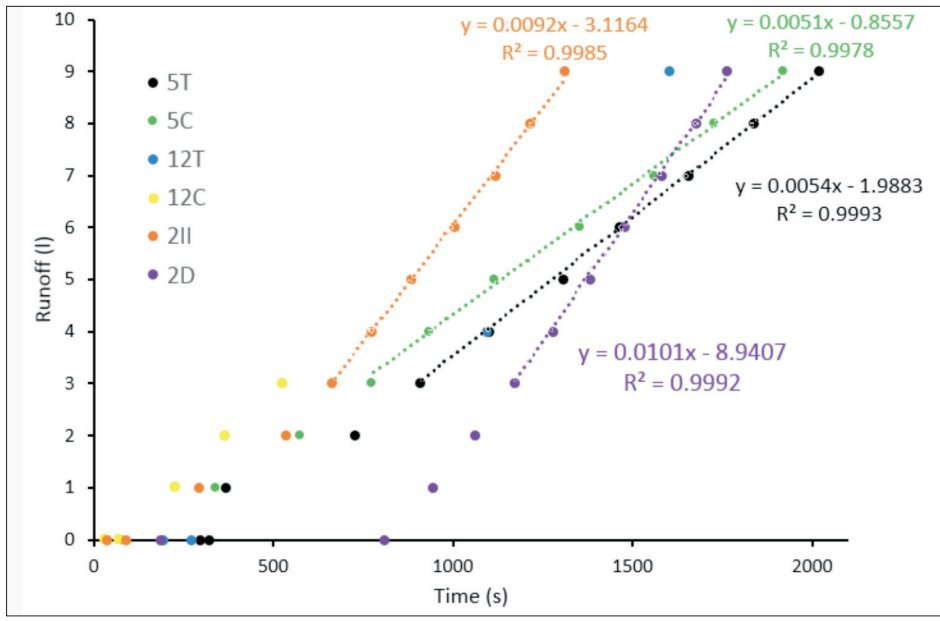


Fig. 2. The changes of necessary time for 1 litre runoff. There are two points in the zero line. The first point represents the time of surface ponding. There is no any runoff yet. Second point means the time when runoff has started. The equations refer to the 5T, 5C and 2II, 2D treatments after the third litre of runoff.

By this means the observed results were not examined in more detail.

Figure 3 represents the average sediment concentration increase with the slope angle. Runoff on steeper slopes contains greater amount of soil particles compared to gentle slopes. Runoff rates are not influenced by slope angles and thus do not influence the velocity of runoff. Defersha, M.B. and Melesse, A.M. (2012) got the same results in three different soil types with two different moisture contents and under three different rainfall intensities. On the other hand, Fox, D.M. *et al.* (1997) reported contradicting results on the slope dependence of the infiltration and crust formation on runoff rates. They assume that this contradiction may be connected to the different micro-morphology of the surfaces.

Table 2 summarizes the sediment concentration changes during the precipitation. The concentration of the sediment is higher at the tilled surface except the third period of the treatment 12C where the steep slope increased the concentration of the sediment, but more data is needed to determine the highest concentration which has effect on the mean values in *Figure 3*. During the precipitation in case of 12T fourfold measure of the sediment was washed down compared to the case of the 5T in the first period. In the second period, the differences decreased threefold and to the end of the precipitation increased a little bit more than fourfold again. During the three periods of the precipitation, the rates in case of 5C and 12C treatments are bigger, 5.0, 3.5 and 5.0-fold respectively.

Almost the same density sediment was washed down from the surfaces of 2II and 2D precipitation and there was no significant difference between the periods of the precipitation. It was one order less dense than the others *(Table 2, Figure 3)*, therefore, we can state that extreme initial moisture content has no effect on sediment concentration changes. Wetting and drying cycles have influence on the soil structure (Bodner, G. *et*

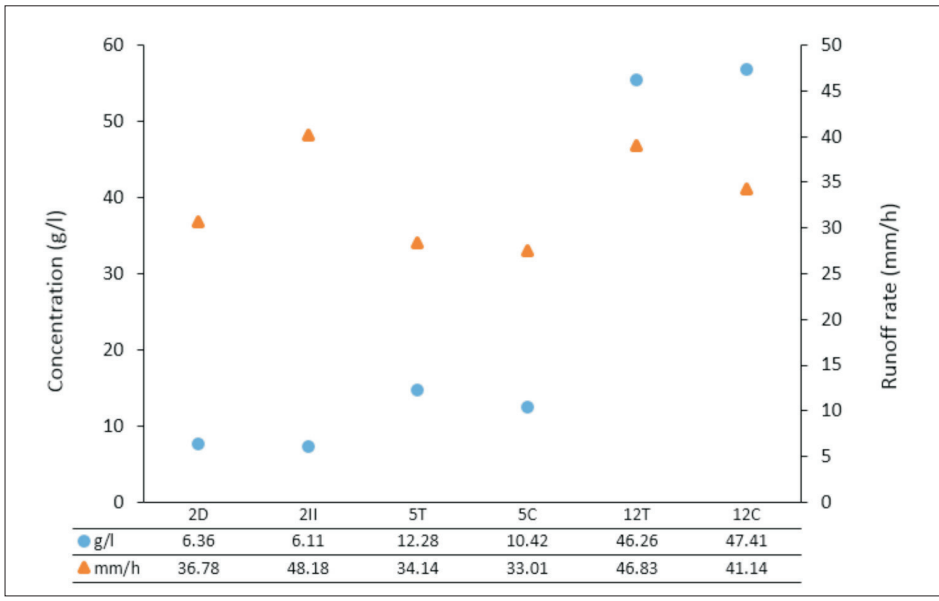


Fig. 3. The sediment concentrations and runoff rates of the six treatments

al. 2013) therefore erosion affects differently in extreme moisture content soils. The treatment of 2D and 2II had the lowest soil content in the runoff because the structure of the soil and aggregate stability was very weak. Aggregates were easy to detach to elementary particles, and due to the gentle slope, the runoff was able to transport only finer particles. Initial erosion and raindrop impact are supposed to have more significant influence on particle redistribution in these two cases. The lowest sediment concentration and highest runoff rate were observed for 2B.

Areas covered by inundation are generally flat or have only gentle slopes hence the main risk there is not the runoff, rather the structure degradation. The fraction content and sediment concentration are almost the same due to the degraded structure. The sediment concentration trend is the same in all cases during the precipitation. The second period is the most erodible, except the 12C when the sediment concentration increased (*Table 2*).

Table 3 summarizes the results of all (72) soil loss samples and the 4 sample sets of the original soil. Washed sediment lacked macro aggregates larger than 1 mm. On the average, the difference is 38 percent between the ratio of the aggregates >1 mm in original soil sample and the soil loss. This fraction is underrepresented

Table 2. Sediment concentration changes during the precipitation in the six treatments

Periods	5T	5C	12T	12C	2II	2D
	g l^{-1}					
Period I	10.95	7.46	43.88	35.19	5.89	6.65
Period II	15.19	12.09	49.93	43.00	6.16	6.16
Period III	10.69	11.70	45.79	64.03	6.29	6.26
Whole precipitation	12.28	10.42	46.26	47.41	6.11	6.35

Table 3. The aggregate size distribution results of all the 72 sediment samples, and the 4 sample of the original soil

Size, μm	Untreated soil	5T	5C	12T	12C	2II	2D
	%						
Period I							
>50	4.81	32.56	29.59	14.08	36.54	52.70	70.16
50–250	24.63	19.84	47.88	52.50	29.87	43.07	19.91
250–1,000	32.00	46.96	22.49	32.77	31.74	3.93	9.63
1,000<	38.57	0.64	0.04	0.66	1.85	0.29	0.30
Period II							
>50	4.81	29.43	51.12	12.46	33.86	79.88	72.87
50–250	24.63	44.58	35.13	49.22	36.72	14.01	17.65
250–1,000	32.00	25.81	13.53	37.52	28.26	5.73	9.31
1,000<	38.57	0.18	0.22	0.79	1.16	0.38	0.16
Period III							
>50	4.81	38.01	49.97	13.30	28.99	80.06	72.38
50–250	24.63	42.75	37.69	49.91	40.93	13.47	18.79
250–1,000	32.00	18.64	12.19	35.96	28.44	6.10	8.62
1,000<	38.57	0.61	0.14	0.84	1.64	0.37	0.21

in the sediment with less than 1 percent except for 12C. During the rain, there were no trends present, but the mass of the soil loss in this fraction was under 1 g (therefore the dynamics were not relevant). The lack of this fraction was probably connected to the raindrop impact and partly to macro aggregates broke down to micro aggregates (50–250 μm) and elementary particles (<50 μm) due to slaking (as it was also reported by Le Bissonnais, Y. 1996).

Generally, the rate of the soil loss of recently tilled surfaces (5T and 12T) was larger in the 250–1,000 μm fraction than in the soil loss of crusty surfaces (5C and 12C). The proportion of this fraction was also larger at 12 percent slope steepness with the exception of the I. period of 5T when this ratio is the largest. The proportion of 250–1,000 μm was under 10 percent in case of 2II and 2D.

The fraction of 50–250 μm showed the largest proportions in the soil loss compared to the original soil sample. The same trend was presented in the 250–1,000 μm fraction, where the tilled surface and the slope increased the proportion of this fraction (with the exception of the I. period of 5T, where the dominant fraction is the 250–1,000 μm). Treatment 2II and 2D showed decreasing trend during the precipitation.

Soil loss was larger than in the case of the original soil, concerning the fraction of <50 μm at all proportions. The largest ratio was found in treatment 2II, whilst the last period of the precipitation. This fraction was characteristic of treatment 2II and 2D has with over 50 percent rate.

To conclude, the runoff was faster from crusty or wetter surfaces (5C, 12C, 2II), however, larger aggregates were eroded from recently tilled surfaces (5T, 12T). The runoff of 2D started late but it was fast. The aggregate size distribution changed in the eroded sediment, which depended on the time frame of the treatment.

Conclusion

Laboratory scale rainfall simulator was used to examine soil erosion and runoff under six different conditions represented by different seasonal situations. Our results showed that the same soil sample under the same precipitation was eroded totally different. Crusting had no definite role in infiltration mitigation; moreover, we measured ambiguous data regarding soil loss reduction.

In general, the ratio of the macro aggregates decreased and the ratio of the micro aggre-

gates and clay fraction increased in the sediment during the precipitation. Larger amount of sediment was transported from steeper slopes by runoff, but larger aggregates were washed down from the tilled surface. The most erodible fractions play an important role in nutrient supply of agricultural areas thus the erosion protection is relevant. Changing climate conditions have even more effect on agricultural production through the temporal and spatial distribution of the erosion rates. It is necessary to understand the erosion processes under different conditions.

The next stage of our research is to perform more experiments at the laboratory by using other soil samples and to compare and verify the results live on the field. Further studies are planned concerning the elementary particles of the aggregate fraction, the organic matter content and the clay mineral composition of the sediment. Future aims are to find the main reason of surficial variability, i.e. to detect the differences among the seasonal erosion characteristics under heavy rainfall situations.

REFERENCES

Aksoy, H., Erdem Unal, N., Cokgor, S., Gediklia, A., Yoonb, J., Kocaa, K., Incia, S.B. and Erisc, E. 2012. A rainfall simulator for laboratory-scale assessment of rainfall-runoff-sediment transport processes over a two-dimensional flume. *Catena* 98. 63–72.

Armstrong, Q. and Quinton, J.N. 2009. Pumped rainfall simulators: the impact of rain pulses on sediment concentration and size. *Earth Surface Processes and Landforms* 34. (9): 1310–1314.

Bartholy, J., Pongrácz, R. and Pieczka, I. 2014. How the climate will change in this century? *Hungarian Geographical Bulletin* 63. (1): 55–67.

Bodner, G., Scholl P. and Kaul, H.P. 2013. Field quantification of wetting–drying cycles to predict temporal changes of soil pore size distribution. *Soil and Tillage Research* 133. 1–9.

Centeri, Cs. and Császár, A. 2003. A talajképződés és az erózió által kiváltott talajpusztulás kapcsolata a Tihanyi-félsziget példáján (The connection of soil formation and erosion induced soil loss in the Tihany Peninsula). *Tájökológiai Lapok* 1. (1): 81–85.

Centeri, Cs. and Pataki, R. 2003. Hazai talajerodálhatósági értékek meghatározásának fontossága a talajveszteség tolerancia értékek tükrében (Importance of determining Hungarian soil erodibility values in connection with the soil loss tolerance values). *Tájökológiai Lapok* 1. (2): 181–192.

Centeri, Cs., Jakab, G., Szalai, Z., Madarász, B., Sisák, I., Csepinszky, B. and Bíró, Zs. 2011. Rainfall simulation studies in Hungary. In *Soil Erosion: Causes, Processes and Effects*. Ed.: Fournier, A.J. New York, NOVA Science Publisher, 177–217.

Christiansen, J.E. 1942. Irrigation by Sprinkler. *Agriculture Experimental Station Bulletin* 37. 1–124.

Defersha, M.B. and Melesse, A.M. 2012. Effect of rainfall intensity, slope and antecedent moisture content on sediment concentration and sediment enrichment ratio. *Catena* 90. 47–52.

Dövényi, Z. (ed.) 2010. *Magyarország kistájainak katasztere* (Cadastre of natural micro-regions of Hungary). 2. átdolgozott és bővített kiadás. Budapest, MTA Földrajztudományi Kutatóintézet, 876 p.

Fox, D.M., Bryan, R.B. and Price, A.G. 1997. The influence of slope angle on final infiltration rate for interrill conditions. *Geoderma* 80. (1–2): 181–194.

Gómez, J.A. and Nearing, M.A. 2005. Runoff and sediment losses from rough and smooth soil surfaces in a laboratory experiment. *Catena* 59. 253–266.

Grismer, M.E. 2010. *Rainfall Simulation Studies – A Review of Designs, Performance and Erosion Measurement Variability.* TSC Rainsim workshop. 110 p.

Hall, M. J. 1970. A critique of methods of simulating rainfall. *Water Resources Research* 6. (4): 1104–1113.

Jakab, G. and Szalai, Z. 2005. Barnaföld erózióérzékenységének vizsgálata esőztetéssel a Tetves-patak vízgyűjtőjén (Brown soill erodibility measurements in the Tetves Stream catchment using rainfall simulator). *Tájökológiai Lapok* 3. (1): 177–189.

Jin, K., Cornelis, W.M., Gabriels, D., Schiettecatte, W., De Neve, S., Lu, J., Buysse, T., Wu, H., Cai, D., Jin, J. and Harmann, R. 2008. Soil management effects on runoff and soil loss from field rainfall simulation. *Catena* 75. (2): 191–199.

Jomaa, S., Barry, D.A., Brovelli, A., Heng, B.C.P., Sander, G.C., Parlange, J.-Y. and Rose, C.W. 2012. Rain splash soil erosion estimation in the presence of rock fragments. *Catena* 92. 38–48.

Kazó, B. 1966. A talajok vízgazdálkodási tulajdonságainak meghatározása mesterséges esőztető készülékkel (Assessment of water management properties of soils with rainfall simulator device). *Agrokémia és Talajtan* 15. (2): 239–252.

Kemper, D.W. and Rosenau, R.C. (1986) Aggregate stability and aggregate size distribution. In: Klute, A. (Ed.), Methods of Soil Analysis Part 1. ASA-SSSA, Madison, WI, 425–442.

Kerényi, A. 1986. Az iniciális erózió laboratóriumi vizsgálata homokon és szerkezetes talajokon (Laboratory simulation study on the initial erosion of sand and soils with well developed structure). *Agrokémia és Talajtan* 35. 18–38.

Kertész, Á. and Centeri, Cs. 2006. Hungary. In *Soil erosion in Europe.* Eds. Boardman, J. and Poesen, J. Chichester, John Wiley & Sons Ltd. 139–154.

Le Bissonnais, Y. 1996. Aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. *European Journal of Soil Science* 47. 425–437.

Le Bissonnais, Y., Bruand, A. and Jamagne, M. 1989. Laboratory experimental study of soil crusting: relation between aggregate breakdown mechanism and crust structure. *Catena* 16. 377–392.

Meyer, L.D. 1965. Simulator of rainfall for soil erosion research. *Transactions of the ASAE* 8. (1): 63–65.

Mohammad, A.G. and Adam, M.A. 2010. The impact of vegetative cover type on runoff and soil erosion under different land uses. *Catena* 81. 97–103.

Nearing, M.A., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M.H., Nunes, J.P., Renschler, C.S., Souchére, V. and Van Oost, K. 2005. Modelling response of soil erosion and runoff to changes in precipitation and cover. *Catena* 61. (2): 131–154.

Pappas, E.A., Smith, D.R., Huang, C.W.D. and Shuster Bonta, J.V. 2008. Impervious surface impacts to runoff and sediment discharge under laboratory rainfall simulation. *Catena* 72. (1): 146–152.

Strauss, P., Pitty, J., Pfeffer, M. and Mentler, A. 2000. Rainfall simulation for outdoor experiments. In *Current Research Methods to Assess the Environmental Fate of Pesticides.* Eds. Jamet, P. and Cornejo, J. Idaho Falls, USA, INRA Editions, 329–333.

West, L.T., Chiang, S.C. and Norton, L.D. 1992. The morphology of surface crusts. In *Soil Crusting: Chemical and Physical Processes.* Advanced Soil Science. Eds Sumner, M.E. and Stewart, B.A. Boca Raton, Lewis Publisher, 73–92.

Zámbó L. Weidinger T. 2006. Karsztkorróziós talajhatás néhány tényezőjének vizsgálata esőszimulációs kísérletek alapján (Investigations of karst corrosional soil effects based on rainfall simulator experiment). In *Táj, környezet és társadalom*. *Ünnepi tanulmányok Keveiné Bárány Ilona professzor asszony tiszteletére*. Eds.: Kiss, A., Mezősi, G. and Sümeghy, Z. Szeged, SZTE Éghajlattani és Tájföldrajzi Tanszék – Természeti Földrajzi és Geoinformatikai Tanszék, 757–765.

DOI: 10.15201/hungeobull.64.1.4 Hungarian Geographical Bulletin **64** 2015 (1) 35–48.

An application of a spatial simulated annealing sampling optimization algorithm to support digital soil mapping

Gábor SZATMÁRI[1], Károly BARTA[1] and László PÁSZTOR[2]

Abstract

Spatial simulated annealing (SSA) was applied to optimize the sampling configuration for soil organic matter mapping through various sampling scenarios in a Hungarian study site. Prediction-error variance of regression kriging was applied as quality measure in the optimization procedures. Requisites of SSA come from a legacy soil dataset and from spatial auxiliary information. Four scenarios were set to represent the major capabilities of SSA. Scenario 1 and 2 represented completely new sampling designs to optimize with predefined constraints. In scenario 1, number of new observations was the constraint, whilst in scenario 2, it was the value of the quality measure. In both scenarios, areas inaccessible for sampling (roads, farms etc.) were also taken into account. Scenario 3 and 4 represented complementary sampling configurations to optimize taking the previously collected samples into consideration. In scenario 3, the constraint was the number of new observations, whilst in scenario 4, it was the value of the quality measure. In both cases, two types of previously collected sampling design were simulated, a regular and a clustered configuration. The resulted designs were evaluated by Kolmogorov–Smirnov test, nearest neighbour distribution function and empty space function. In cases of scenario 1 and 3, the results showed that, all of the optimized sampling configurations cover properly both geographic and feature space, respectively. In cases of scenario 2 and 4, the resulted calibration curves can be used to determine the sample size for a given quality measure value. Furthermore, we could determine the minimal sample size for a given scenario, which has to be collected to represent properly both geographic and feature space. In conclusion, SSA is a valuable tool to optimize the sampling design considering a lot of constraints.

Keywords: spatial simulated annealing, sampling optimization, geostatistics, regression kriging prediction-error variance, digital soil mapping

Introduction

Digital soil mapping (DSM) aims at spatial prediction of soil properties by combining soil observation at points with auxiliary information, such as contained in digital elevation models, remote sensing images and climate data records (McBratney, A.B. *et al.* 2003; Heuvelink, G.B.M. *et al.* 2007). Hence, the direct observations of the soil are im-

[1] Department of Physical Geography and Geoinfor-matics, Faculty of Science and Informatics, University of Szeged. H-6722 Szeged, Egyetem u. 2., E-mails: szatmari.gabor.88@gmail.com, barta@geo.u-szeged.hu
[2] Institute for Soil Science and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences. H-1022 Budapest, Herman Ottó u. 15. E-mail: pasztor@rissac.hu

portant for two main reasons (Heuvelink, G.B.M. *et al.* 2007):
– they are used to characterize the relationship between the soil property of interest and the auxiliary information,
– they are used to improve the predictions based on the auxiliary information, by spatial interpolation of the differences between the observations and predictions.

Regression kriging (RK) (also termed universal kriging or kriging with external drift, see Hengl, T. *et al.* 2007) illustrates well that twofold application of the soil observations. Spatial prediction method of RK combines a regression of the target pedological variable on covariates with kriging of the regression residuals. Nevertheless RK assumes that, the sampling points represent properly both geographic and feature space (Hengl, T. 2009), where the latter is defined by the covariates.

Extensive work has been done on sampling strategy optimization for DSM over the past decades to satisfy the topical demands, which were suggested by soil surveyors, pedometricans, end-users, and so forth. These demands can be e.g. the expectation of the accuracy and/or uncertainty of the prediction(s), taking auxiliary information into account, optimization of the sampling design for more than one soil variable, taking previously collected samples into account, consideration of any kind of constraints, such as the number of the new observations, inaccessible areas for sampling, budget and/or accuracy constraints. One of the optimization algorithms is spatial simulated annealing (SSA) (van Groenigen, J.W. and Stein, A. 1998) that has been frequently applied in soil surveys to optimize the sampling design using the RK prediction-error variance (RKV) as optimization criterion (Brus, D.J. and Heuvelink, G.B.M. 2007; Heuvelink, G.B.M. *et al.* 2007; Baume, O.P. *et al.* 2011; Melles, S.J. *et al.* 2011; Szatmári, G. 2014). SSA with RKV is sporadically able to satisfy the above mentioned demands.

The main aim of this paper is to present and test the SSA sampling optimization algorithm through various sampling scenarios in a Hungarian study site. The scenarios were set to represent the major capabilities of SSA and to cover a major part of soil sampling issues. In all scenarios, the goal was to optimize the sampling design for soil organic matter (SOM) mapping considering some constraints (e.g. number of new observations, inaccessible areas for sampling, previously collected samples). The resulted sampling configurations were evaluated by various statistical and point pattern analysis tools, in order to examine how they cover both the geographic and feature space.

Theoretical backgrounds

Some thoughts on (spatial) soil sampling for digital mapping

Sampling concerns selection of a subset of individuals from a population to estimate the characteristics of the whole population; where these characteristics could be the total or mean parameter value for a random field, values at unvisited sites or location of target(s) (Wang, J.-F. *et al.* 2012).

In case of DSM, the main aim for a given pedological variable is to estimate its values at unsampled locations. For this purpose, various statistical models (i.e. spatial prediction methods) have been widely used, where we assume that the models and the "real world" are compatible. Furthermore, this implies that the sampling is representative for the whole population. According to Bárdossy, Gy. (1997), the sampling is said to be representative (from a statistical viewpoint) for a population, if it reflects the characteristics of the population the best.

On other hand, we do not know exhaustively the whole population, just only a small part of it (provided by the samples). How can we decide that, the sampling is representative for the whole population? If we know the components of the given statistical model, we can set a "quasi optimal state" through the sampling strategy, where we can assume that, the collected samples are representative for the whole population.

Therefore, the statistical inferences are compatible with the "real world". The setting of the sampling strategy can be regarded as an optimization problem.

As we will see in the next subsection, the RK spatial prediction method assumes that, the variation of the soil property of interest can be modelled as a sum of a deterministic (which is based on the covariates) and a stochastic (which is based on the variogram or covariance function) components. Therefore, if we describe properly, through the sampling design, both the feature (which is defined by the covariates) and geographic space, we can assume that, the statistical inference (i.e. map of the soil property of interest) represent the real situation. It can be regarded as an optimization problem, where we need an optimization algorithm and an optimization criterion. As we will see in the next subsections, SSA will be this algorithm and RKV will be this criterion.

Regression Kriging (RK) spatial prediction method

In the last decade, RK has been more and more popular in DSM (Hengl, T. *et al.* 2004; Dobos, E. *et al.* 2007; Hengl, T. *et al.* 2007; Minasny, B. and McBratney, A.B. 2007; Illés, G. *et al.* 2011; Szatmári, G. and Barta, K. 2013; Pásztor, L. *et al.* 2014), as well as in SSA sampling optimization procedure using its prediction-error variance as optimization criterion (Brus, D.J. and Heuvelink, G.B.M. 2007; Heuvelink, G.B.M. *et al.* 2007; Baume, O.P. *et al.* 2011; Melles, S.J. *et al.* 2011; Szatmári, G. 2014). RK assumes that, the deterministic component of the target soil variable is accounted for by the regression model, whilst the model residuals represent the spatially varying but dependent stochastic component, as well as both components can be modelled separately and simultaneously. The estimation for Z variable at an unvisited location s_0 is given by

$$Z(_{s0}) = q_0^T \cdot \beta + \lambda_0^T \cdot (z - q \cdot \beta), \quad (1)$$

where β is the vector of the regression coefficients, q_0 is the vector of the covariates at the unvisited location, λ_0 is the vector of the kriging weights, z is the vector of the observations and q is the matrix of covariates at the sampling locations. Its prediction-error variance at s_0 is given by

$$\sigma^2(s_0) = c(0) - c_0^T \cdot C^{-1} \cdot c_0 + (q_0 - q^T \cdot C^{-1} \cdot c_0)^T \cdot \cdot (q^T \cdot C^{-1} \cdot q)^{-1} \cdot (q_0 - q^T \cdot C^{-1} \cdot c_0), \quad (2)$$

where $c(0)$ is the variance of the residuals, c_0 is the vector of covariances between the residuals at the observed and unvisited locations and C is the variance-covariance matrix of the residuals. RKV is independent from the observed values (see Eq. [2]), so it can be calculated before the actual sampling takes place, which can be considered as a beneficial property in point of costs and time. Furthermore, it incorporates both the prediction error variance of the residuals (first two terms on the right-hand side of Eq. [2]) and the estimation error variance of the trend (third term on the right-hand side of Eq. [2]), which endeavour SSA algorithm to optimize the sampling design both in geographic and feature space (Heuvelink, G.B.M. *et al.* 2007). However, it mainly depends on, how the two types of error variance contribute to RKV.

Spatial simulated annealing (SSA) sampling optimization algorithm

In brief, SSA is an iterative, combinatorial, model-based sampling optimization algorithm in which a sequence of combinations is generated by deriving a new combination from slightly and randomly changing the previous combination (van Groenigen, J.W. *et al.* 1999). When a new combination is generated, the quality measure (in this study the spatially averaged RKV) is calculated and compared with the quality measure value of the previous combination (van Groenigen, J.W. *et al.* 1999; Brus, D.J. and Heuvelink, G.B.M. 2007). The Metropolis criterion defines the probability that, either accepts the new combination as a basis for the further computation, or rejects it and the previous combination stays as a basis further (van Groenigen, J.W. *et al.* 1999):

$$P(C_i \rightarrow C_{i+1}) = 1, \quad \text{if } \Phi(C_{i+1}) \leq \Phi(C_i)$$
$$P(C_i \rightarrow C_{i+1}) = \exp\left(\frac{\Phi(C_i) - \Phi(C_{i+1})}{c}\right), \quad \text{if } \Phi(C_{i+1}) > \Phi(C_i) \tag{3}$$

where C_i and C_{i+1} are the previous and the new combination, c is the positive control parameter (so-called "system temperature", which is lowered as optimization progresses) and $\Phi(\cdot)$ is the quality measure (so-called "fitness or objective function").

For a given soil variable, SSA (using RKV as optimization criterion) requires that the structure of the regression model and the variogram or covariance function of the residuals are known (Heuvelink, G.B.M. *et al.* 2007), which is one of the main drawbacks of the method. On other hand, the algorithm is able to take inaccessible areas and/or previously collected samples into account.

Material and methods

Study site and legacy soil data

The study site (approx. 17 km^2) is located in the central part of Hungary, in the Mezőföld region, near village Előszállás (*Figure 1*). The area of interest is mainly covered by Haplic Chernozems and Kastanozems with significant secondary carbonates. Calcisols and Regosols are found on the eroded steeper slopes, where the top-horizon is too thin for Mollic or it is completely missing. Colluvic material can be found at the bottom of the slopes, where Phaeozems or Regosols were formed. The study site can be characterized mainly by arable lands sown with winter wheat, maize and sunflower.

The available legacy soil data was collected at the end of the 1980s in the framework of the National Land Evaluation Programme. The dataset incorporates 117 topsoil (0–30 cm) observations from the area of interest. Various pedological variables were quantified during the fieldwork and laboratory analyses. In this study, the soil organic matter (SOM) was chosen as target pedological variable to optimize the sampling design for various scenarios. Exploratory data analysis was performed on SOM data to remove the outliers, to calculate summary statistics and to test the normality of the SOM probability distribution. The analysis has shown that, the probability distribution of SOM is close to normal. The summary statistics of SOM are presented in *Table 1.*

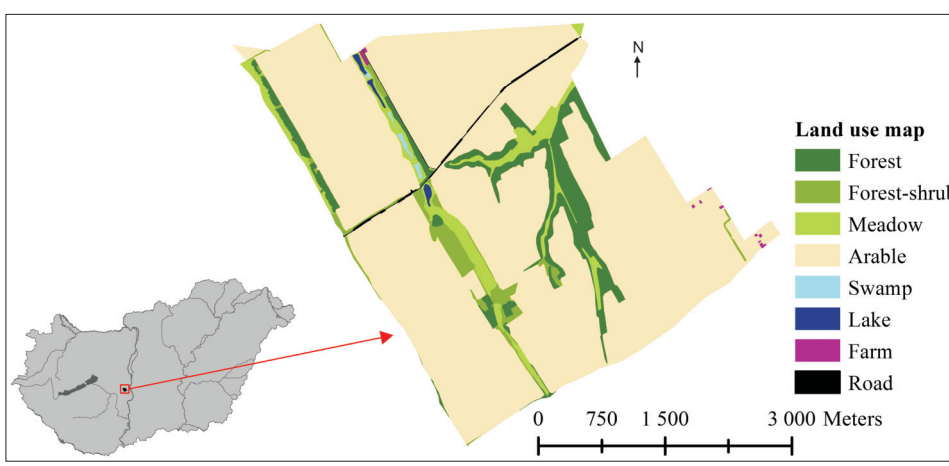


Fig. 1. The location of the study area in Hungary and its land use map

Table 1. Summary statistics of soil organic matter (SOM) computed from the legacy soil dataset without outliers

Variable	Mean	Median	Minimum	Maximum	Standard deviation	Skewness
	value					
SOM, %	2.90	2.95	1.51	4.44	0.56	–0.28

Auxiliary information from the study site

Spatially exhaustive auxiliary information were derived from digital elevation model (DEM) (with 20 meters resolution) and from land use (LU) map of the study area, since Szatmári, G. and Barta, K. (2012, 2013) pointed out that the spatial distribution/variability of SOM mainly depends on the topography and the LU at the area of interest. The following morphometric parameters were derived from DEM: altitude, slope (in percent), slope length, aspect, profile and planar curvature, LS factor (Wischmeier, W.H. and Smith, D.D. 1978), topographic wetness index, vertical distance to channel network and potential incoming solar radiation (direct and diffuse). LU map was derived from the products of the official aerial photography campaign of Hungary, taken in 2005.

In contrast with the morphometric parameters, LU type is a categorical variable. For the sake of the application of RK each LU type was converted into indicator variables. Raster maps were generated for each LU types with value domain showing 1 at the locations of the given LU type and showing 0 for all other locations. These raster maps were resampled for 20 meters.

Principal component (PC) analysis was performed on the auxiliary data and the resulted PCs were used as covariates in the further analysis. It is a crucial step, since the PCs are orthogonal and independent; hence they satisfy the requirements of the multiple linear regression analysis and their application decreases the multi-collinearity effect.

Settings of spatial simulated annealing and sampling scenarios

The requirements of SSA (using RKV as optimization criterion) are the structure of the regression model and the variogram or covariance function of residuals of the model. These requisites were generated from the legacy soil dataset and from the covariates, respectively. Multiple linear regression analysis was performed to characterize the relationship between SOM and covariate data, using a "stepwise" selection method and a significance level of 0.05. In the next step, the residuals were derived from the resulted regression model and exploratory variography was performed on them. The experimental variograms were calculated and the spatial structure was modelled with a theoretical variogram model. The fitted variogram and regression model were used along the optimization process provided by SSA to calculate (using Eq. [2]) the quality measure (i.e. spatially averaged RKV).

There are some land use types (swamp, lake, farm and road), which are out of the scope of soil mapping, so we excluded them from the optimization process as inaccessible areas for sampling.

The initial "system temperature" for SSA was chosen such that the average increase acceptance probability was 0.8 and the "system cooling" was exponentially. Furthermore, a stopping criterion was defined to rein up the simulation when the quality measure did not improve in many tries. The stopping criterion value was set 200.

The sampling scenarios were set to represent the major capabilities of SSA and to cover a major part of soil sampling issues. The following four scenarios were set to optimize the sampling design for SOM mapping:

– Scenario 1 (Sc1): Completely new sampling strategy with fixed number of new observations,
– Scenario 2 (Sc2): Completely new sampling strategy to achieve a predefined quality measure value,

– Scenario 3 (Sc3): Complementary sampling with fixed number of new observations to supplement the previously collected samples,
– Scenario 4 (Sc4): Complementary sampling to supplement the previously collected samples and to achieve a predefined quality measure value.

Two types of previously collected sampling configuration were applied as complementary sampling scenarios (Sc3 and Sc4):
– Regular design, where the sampling points located at the nodes of a square grid,
– Clustered design, where the sampling points showed a clustered pattern in the geographic space.

In case of Sc1, the number of new observations was set 120, which is commensurable with the sample size of the legacy soil dataset. In Sc3 and Sc4, the previously collected sample size was set 35, which were following regular and clustered design, respectively. In case of Sc3, the fixed number of new observation was set 50. In cases of Sc2 and Sc4, the main aim was to create a so-called calibration curve. This calibration curve can be used to determine the sample size for a given quality measure value and vice versa. To calculate this curve, the sample size was systematically increased and the quality measure value of the optimized configuration was calculated. In next step, the quality measure values were plotted as a function of the sample size.

Evaluation of the optimized sampling designs

The optimized sampling designs were evaluated by various statistical and point pattern analysis tools. Kolmogorov–Smirnov (K–S) test was applied to examine for a given covariate, if its distribution from the optimized design is equal to the distribution from the complete area of interest. Based on the test results we can examine how the sampling configurations cover the feature space created by the covariates.

The nearest neighbour distances distribution functions *G(r)* and the empty space functions *F(r)* were calculated, based on the sampling designs, to explore the type of interaction between the sampling points and to examine how they cover the geographic space. The *G(r)* function measures the distribution of the distances from an arbitrary sampling point to its nearest sampling point, while the *F(r)* function measures the distribution of all distances from an arbitrary point of the plane to its nearest sampling point (Bivand, R.S. *et al.* 2008). In case of *F(r)*, the grid nodes of the planned prediction locations were applied to measure the so-called empty space distances. It gives direct information on the kriging neighbourhood.

Results and discussion

Regression and variogram models

The determination coefficient of the resulted regression model was 0.41, which means that the model explains more than 40 percent of the total variability of SOM and the remaining approx. 60 percent have to be modelled stochastically. Five covariates were selected into the model by the "stepwise" method. The observed significance level, which was calculated for the model, was practically zero.

The regression residuals were derived and the experimental variograms (directional and omnidirectional) were calculated to model their spatial continuity. The directional variograms showed an isotropic spatial structure, which structure was approached by a spherical variogram model type. *Table 2* summarizes the parameters of the fitted isotropic variogram model.

Table 2. Parameters of the fitted isotropic variogram model for soil organic matter (SOM) residuals

Variable	Model type	Nugget	Partial sill	Sill	Nugget/Sill, %	Range, m
SOM residuals	Spherical	0.04	0.12	0.16	25.00	1,420

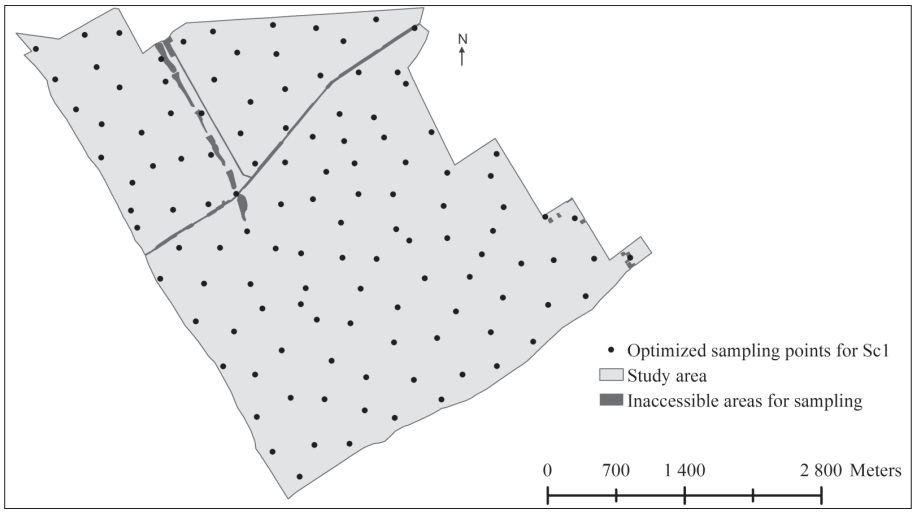


Fig. 2. The optimized sampling design for scenario 1

Optimized sampling designs for Sc1–Sc2 and their performance

The optimized sampling configuration for Sc1 is presented in *Figure 2*, which sampling design shows a "quasi" regular point pattern. *Figure 3* presents the calibration curve for Sc2 (denoted with solid line), as well as the nugget variance of SOM residuals (denoted with dashed line), where the latter is constant, because this part of variance cannot be modelled (Webster, R. and Oliver, M.A. 2007). The so-called "nugget effect" arises from measurement errors and/or small-scale heterogeneity (Goovaerts, P. 1999; Geiger, J. 2006; Webster, R. and Oliver, M.A. 2007). It also means that the value of the spatially averaged RKV cannot be less than this nugget variance (see Eq. [2]). Hence, the calibration curve converges to the nugget variance, if the sample size is infinitely large (see *Figure 3*).

The calculated calibration curve for Sc2 can be used to determine the sample size for a given spatially averaged RKV value expected to be achieved for the SOM map. In a practical point of view, this kind of calibration curve is a useful tool to estimate the sample size considering the predefined RKV value (expected to be achieved for the map) and/or the sampling budget's constraints. For example, if the soil surveyors want to achieve 0.08 [%]^2 value of spatially averaged RKV for the SOM map, then the sample size, using this calibration curve (*Figure 3*) is 98. On other hand, if the budget allows to collect 42 number of soil samples and the question is "What is the expectation of the spatially averaged RKV for the SOM map?", then, using the calibration curve (*Figure 3*), the expectation is 0.1 [%]^2.

The observed significance levels of K–S test for Sc1 and Sc2 are presented by *Table 3*. The null hypothesis was that, the two distributions were drawn from the same distribution.

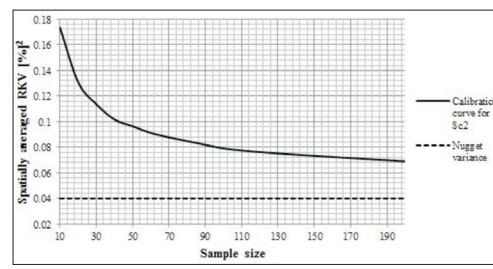


Fig. 3. The calibration curve for scenario 2 and the nugget variance. RKV = regression kriging prediction-error variance

The applied significance level was 0.05. In *Table 3* the values of the observed significance level *were bolded*, where the null hypothesis was accepted. In case of Sc1, the null hypothesis was accepted for all covariates, which means that, the optimized sampling design for Sc1 covers properly the feature space. In case of Sc2, we examined for a given sample size that, how the optimized sampling configuration covers the feature space. As we can see, 60 is the minimal sample size, which is needed to cover properly the feature space (see *Table 3*). Based on this, samples with less than 60 observations are not suitable to describe the trend function, as well as the spatial distribution of SOM.

The observed *F(r)* and *G(r)* functions gave almost the same results for Sc1 and Sc2, thanks to the relatively large range of the variogram (see *Table 2*). We can, however, state that, the optimized sampling configurations covered properly the geographic space, because the *r* value for *F(r) = 1* was lower than the variogram range, respectively. As a consequence, there was no any planned prediction location, which did not have any kriging neighbours. Furthermore, there is an inhibition (i.e. competition) between the sampling points, which follows from that, the $G_{obs}(r)$ function is below the theoretical distribution of complete spatial randomness (e.g. in *Figure 4*, a), whilst the $F_{obs}(r)$ function is above the theoretical distribution of complete spatial randomness (e.g. in *Figure 4*, b).

As a consequence, it causes a quasi-regular point pattern, respectively (as we can also see in *Figure 2*). *Figure 4* presents the observed *G(r)* and *F(r)* functions of the optimized sampling design for Sc1 (the calculated *G(r)* and *F(r)* functions for Sc2 were omitted, because they gave a similar results as in case of Sc1, due to the large range of the variogram).

Optimized sampling designs for Sc3–Sc4 and their performance

The optimized sampling configurations for Sc3 regular and Sc3 clustered are presented in *Figure 5. Figure 6* presents the calculated calibration curves for Sc4 regular (denoted with solid line) and Sc4 clustered (denoted with dashed line), as well as the nugget variance of the fitted variogram model (denoted with dotted line). Both calibration curves converge to the nugget variance, if the sample size is infinitely large (see *Figure 6).*

The calculated calibration curves for Sc4 regular and Sc4 clustered can be used to determine the sample size for a given spatially averaged RKV value and vice versa. For example, if the soil surveyors want to achieve

Table 3. The values of the observed significance level of Kolmogorov-Smirnov test calculated for Scenario 1 and 2.

Sample size	Covariates*				
	SPC1	SPC2	SPC3	SPC4	SPC5
10	0.017	0.000	*0.060*	0.006	0.035
20	*0.240*	0.013	0.021	0.042	*0.173*
30	*0.240*	0.035	*0.153*	0.006	0.013
40	*0.454*	*0.172*	*0.617*	0.017	*0.173*
50	*0.454*	0.035	*0.617*	*0.095*	*0.172*
60	*0.454*	*0.082*	*0.617*	*0.194*	*0.082*
70	*0.734*	*0.329*	*0.617*	*0.194*	*0.173*
80	*0.734*	*0.173*	*0.905*	*0.358*	*0.082*
90	*0.734*	*0.329*	*0.617*	*0.194*	*0.560*
100	*0.454*	*0.173*	*0.617*	*0.358*	*0.329*
110	*0.954*	*0.329*	*0.905*	*0.194*	*0.329*
120	*0.954*	*0.560*	*0.905*	*0.358*	*0.082*
150	*0.734*	*0.560*	*0.617*	*0.193*	*0.173*
200	*0.734*	*0.560*	*0.617*	*0.841*	*0.173*

*The observed significance levels are in italics, where the null hypothesis was accepted at 0.05 significance level

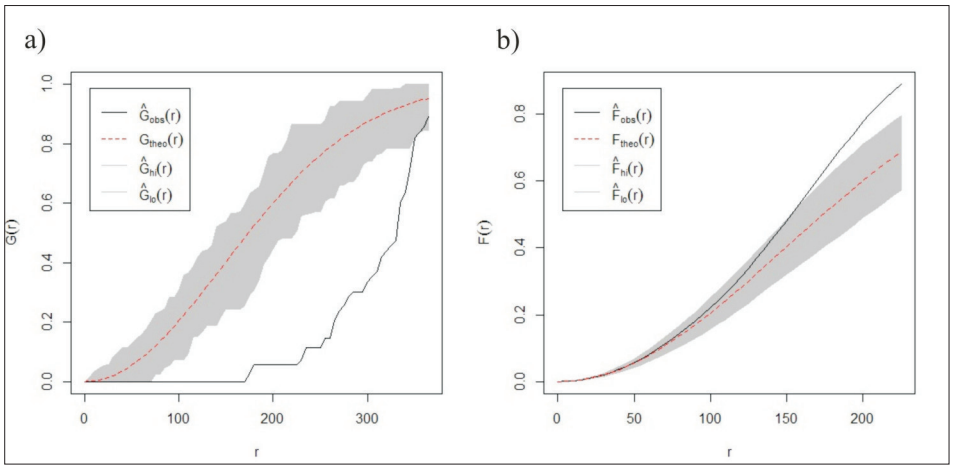


Fig. 4. The observed $G_{obs}(r)$ nearest neighbour distances distribution (a) and $F_{obs}(r)$ empty space function (b) for scenario 1. Abbreviations inside the legend: theo = theoretical distribution of complete spatial randomness; hi = upper envelope of theo; lo = lower envelope of theo

0.08 [%]² value of spatially averaged RKV for the SOM map, when the previously collected sampling design is regular, then the number of new observations, using the calibration curve (*Figure 6*), is 64. On other hand, when the previously collected sampling design is clustered the number of new observations, using the corresponding calibration curve (*Figure 6*) is 84. The large difference between them can be attributed to the follows: when the previously collected sampling design was clustered, the existing samples concentrated only on a small part of the complete area of interest (see *Figure 5*, b), which yielded higher RKV values, as well as caused a poor coverage both in geographic and feature space. On other hand, the existing regular sampling design covered more properly the geographic space (see *Figure 5*, a).

The observed significance levels of K–S tests for Sc3 regular and Sc4 regular are presented in *Table 4*, whilst the observed significance levels for Sc3 clustered and Sc4 clustered are presented in *Table 5*. The null hypothesis and the applied significance level were the same as in case of Sc1 and Sc2. In *Table 4* and *5*, the values of the observed significance level *were bolded*, where the null hypothesis was accepted.

In both cases of Sc3 regular and Sc3 clustered, the null hypothesis was accepted for all covariates, which means that, the optimized sampling designs for Sc3 regular and Sc3 clustered cover properly the feature space. In cases of Sc4 regular and Sc4 clustered, we examined for a given sample size, how the optimized sampling configuration covers the feature space. As we can see in *Table 4* and *5*, 40 is the minimal sample size, which is needed to cover properly the feature space. Based on this, samples with less than 40 observations are not suitable to describe the trend function, as well as the spatial distribution of SOM.

The observed *F(r)* and *G(r)* functions for the previously collected sampling designs are presented in *Figure 7*. In case of clustered design, the *r* value for *F(r) = 1* is higher than the variogram range (which means that, there are some planned prediction locations, which do not have any kriging neighbours), whilst in case of regular design, this *r* value is lower than the variogram range. They support the ascertainment, the clustered design does not cover properly the geographic space, whilst the regular design does (see *Figure 7*).

In cases of Sc3 regular and Sc4 regular, the *F(r)* and *G(r)* functions gave "quasi" the same

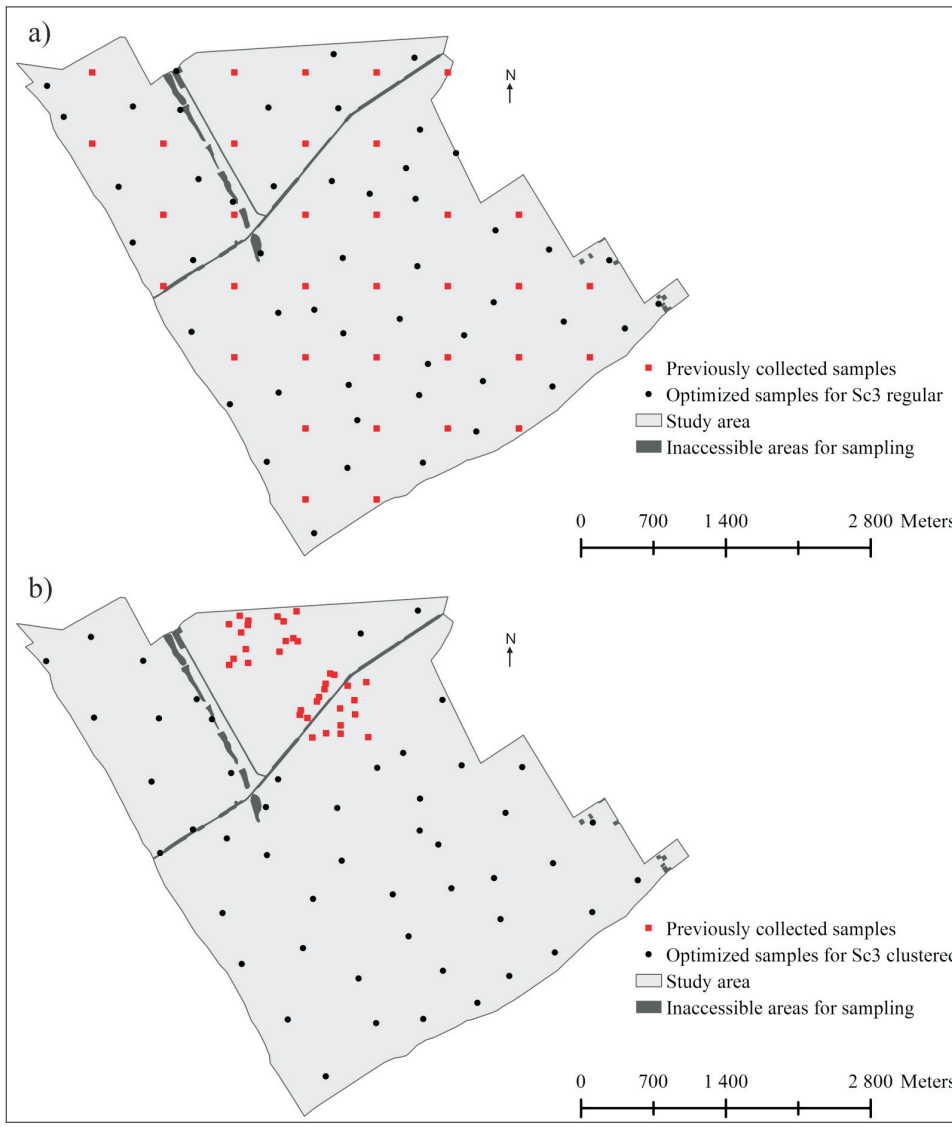


Fig. 5. The optimized sampling designs for scenario 3 regular (a), and scenario 3 clustered (b)

results, thanks to the relatively large range of the variogram model. However, we can state that, the optimized sampling configurations for Sc3 regular and Sc4 regular covered properly the geographic space, so there was no any planned prediction location, which did not have any kriging neighbours. There is an inhibition (i.e. competition) between the sampling points, which causes a quasi-regular point pattern. In case of Sc3 clustered, the optimized sampling design covers properly the geographic space. On other hand, the calculated *F(r)* and *G(r)* functions show a transition between the regular and clustered point pattern types.

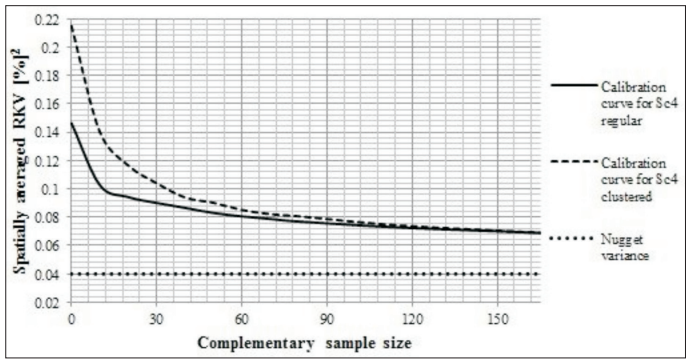


Fig. 6. The calibration curves for scenario 4 regular, scenario 4 clustered and the nugget variance. RKV = regression kriging prediction-error variance

Table 4. The values of the observed significance level of Kolmogorov-Smirnov test calculated for Scenario 3 regular and Scenario 4 regular

Complementary sample size	Covariates*				
	SPC1	SPC2	SPC3	SPC4	SPC5
0	< 0.05				
10	*0.454*	0.035	*0.617*	*0.095*	0.013
20	*0.734*	0.035	*0.617*	*0.358*	*0.082*
30	*0.954*	0.035	*0.617*	*0.194*	*0.329*
40	*0.734*	*0.173*	*0.617*	*0.095*	*0.173*
50	*0.734*	*0.173*	*0.334*	*0.358*	*0.173*
60	*0.954*	*0.329*	*0.617*	*0.095*	*0.329*
70	*0.734*	*0.329*	*0.617*	*0.358*	*0.560*
80	*0.954*	*0.329*	*0.905*	*0.591*	*0.329*
115	*0.954*	*0.560*	*0.905*	*0.358*	*0.560*
165	*0.734*	*0.560*	*0.617*	*0.841*	*0.173*

*The observed significance levels are in italics, where the null hypothesis was accepted at 0.05 significance level

Table 5. The values of the observed significance level of Kolmogorov-Smirnov test calculated for Scenario 3 clustered and Scenario 4 clustered

Complementary sample size	Covariates*				
	SPC1	SPC2	SPC3	SPC4	SPC5
0	< 0.05				
10	0.046	0.000	*0.153*	0.017	*0.082*
20	*0.240*	0.005	*0.617*	*0.095*	*0.082*
30	*0.240*	0.035	*0.334*	0.006	*0.082*
40	*0.954*	*0.329*	*0.617*	*0.095*	*0.560*
50	*0.954*	*0.173*	*0.617*	*0.095*	*0.173*
60	*0.734*	*0.329*	*0.905*	*0.194*	*0.560*
70	*0.954*	*0.329*	*0.905*	*0.194*	*0.560*
80	*0.734*	*0.173*	*0.617*	*0.358*	*0.173*
115	*0.954*	*0.329*	*0.905*	*0.591*	*0.329*
165	*0.734*	*0.560*	*0.617*	*0.841*	*0.173*

*The observed significance levels are in italics, where the null hypothesis was accepted at 0.05 significance level

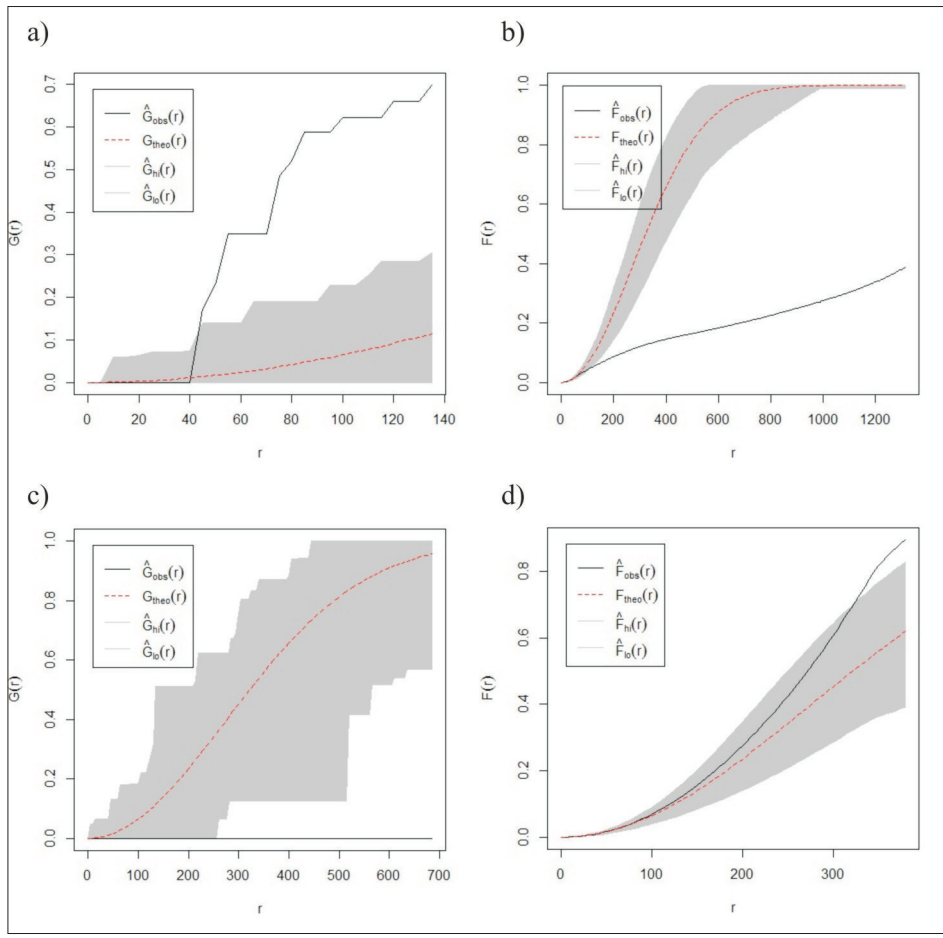


Fig. 7. The observed nearest neighbour distances distribution $G_{obs}(r)$ and empty space $F_{obs}(r)$ function for the previously collected samples in clustered (a–b) and regular (c–d) designs. Abbreviations inside the legend: see *Fig. 4.*

Some thoughts on RKV and SSA

About RKV, we have to notice that, its value(s) mainly characterizes the spatial prediction model rather than the local accuracy of the prediction(s); since it is independent from the observed values (Deutsch, C.V. and Journel, A.G. 1998; Goovaerts, P. 1999; Geiger, J. 2006). We have to consider this fact when we want to use directly its values. However, RKV is a fully suitable measure to compare alternative sampling configuration and to optimize the sampling design for DSM, which follows from its definition (see Eq. [2]).

As we mentioned, the optimized sampling designs for Sc1, Sc2, Sc3 regular and Sc4 regular showed a quasi-regular point pattern. It means that, the variogram model had the dominant influence along these optimization procedures rather than the structure of the regression model, according to Heuvelink, G.B.M. *et al.* (2007). It can be explained by that, the area of interest is fairly homogeneous in point of topography and land use, in other words it has a small "niche" in the feature space, according to Hengl, T. *et al.* (2003). The study site belongs to the Sárbogárd Loess Plateau and only two loess-valleys slice up the area of interest. On

other hand, approx. 85% of the total area is arable (Szatmári, G. and Barta, K. 2012).

There are some limits of SSA using RKV as optimization criterion, e.g. the optimization of sampling design for more than one soil variable. However, it seems to be solved by Szatmári, G. (2014). Another drawback of SSA algorithm is the calculation time, which is lingering. In this study, the elapsed time for a sampling design simulation can take a few hours up to a day. It depends on the settings of the SSA algorithm (initial "system temperature", number of iterations, "cooling" scheme, stopping criterion, etc.), the number of new observations, the size and complexity of the area of interest, the resolution of auxiliary data, the size of matrices for the quality measure calculation (see Eq. [2]), and so forth. We found that, if the maximum of the kriging neighbourhood is restricted to a finite number of observations (according to Webster, R. and Oliver, M.A. 2007, it was set 25, which number of observations is reasonable in point of kriging), then the calculation time decreased significantly.

Conclusions

As it was illustrated by the scenarios, SSA (using RKV as optimization criterion) is a valuable algorithm to optimize soil sampling strategy considering a lot of constraints and demands, which were suggested by soil surveyors, pedometricans and end-users (e.g. the number of new observations, predefined quality measure value (i.e. RKV), as well as taking auxiliary information, previously collected samples and inaccessible areas into account).

RKV is a suitable optimization criterion, because it incorporates the error variance of the trend, as well as the estimation error variance of the residuals, which endeavour SSA to optimize the sampling design both in geographic and feature space. As a consequence, the optimized design absolutely accommodates to the requirements of the RK spatial prediction technique. Therefore, we can assume that the statistical inference (i.e. map of the soil property of interest) is compatible with "the real world". Another beneficial property of RKV is that, it can be calculated before the actual sampling takes place, which can be important in a viewpoint of costs and time. Nevertheless we have to keep in mind that, RKV is independent from the observed values.

The so-called calibration curve can be used to determine the sample size for a given quality measure value and vice versa. As a consequence, this kind of calibration curve is a useful tool to estimate the sample size considering the predefined quality measure value (which is expected to be achieved for the map) and/or the sampling budget's constraints.

***Acknowledgements*:** Our work has been supported by the Hungarian National Scientific Research Foundation (OTKA, Grant No. K105167) and by the TÁMOP-4.1.1.C-12/1/KONV-2012-0012 project (ZENFE).

REREFENCES

Baume, O.P., Gebhardt, A., Gebhardt, C., Heuvelink, G.B.M. and Pilz, J. 2011. Network optimization algorithms and scenarios in the context of automatic mapping. *Computers and Geosciences* 37. 289–294.

Bárdossy, Gy. 1997. Geomatematikai kérdések geológus szemmel (Questions of Geomathematics from the point of view of a geologist). *Magyar Geofizika* 38. (2): 124–141.

Bivand, R.S., Pebesma, E.J. and Gómez-Rubio, V. 2008. *Applied Spatial Data Analysis with R*. New York, Springer, 375 p.

Brus, D.J. and Heuvelink, G.B.M. 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138. 86–95.

Deutsch, C.V. and Journel, A.G. 1998. *GSLIB: Geostatistical Software Library and User's Guide* (2nd Ed.). New York, Oxford University Press, 369 p.

Dobos, E., Michéli, E. and Montanarella, L. 2007. The population of a 500-m resolution soil organic matter spatial information system for Hungary. In *Developments in Soil Science*, Vol. 31. Eds.: Lagacherie, P., McBratney, A.B. and Voltz, M. Amsterdam, Elsevier B.V. 487–495.

Geiger, J. 2006. *Geostatisztika* (Geostatistics). Szeged, University of Szeged, 77 p.

Goovaerts, P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89. 1–45.

Hengl, T. 2009. *A Practical Guide to Geostatistical Mapping*. 2nd Ed. Amsterdam, University of Amsterdam, 291 p.

Hengl, T., Heuvelink, G.B.M. and Rossiter, D.G. 2007. About regression-kriging: from equations to case studies. *Computers and Geosciences* 33. 1301–1315.

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

Hengl, T., Rossiter, D.G. and Stein, A. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research* 41. 1403–1422.

Heuvelink, G.B.M., Brus, D.J. and de Gruijter, J.J. 2007. Optimization of sample configurations for digital mapping of soil properties with universal kriging. In *Developments in Soil Science*, Vol. 31. Eds.: Lagacherie, P., McBratney, A.B. and Voltz, M. Amsterdam, Elsevier B.V. 137–151.

Illés, G., Kovács, G. and Heil, B. 2011. Nagyfelbontású digitális talajtérképezés a Vaskereszt erdőrezervátumban (High resolution digital soil mapping in the Vaskereszt forest reserve). *Erdészettudományi Közlemények* 1. 29–43.

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

Melles, S.J., Heuvelink, G.B.M., Twenhöfel, C.J.W., van Dijk, A., Hiemstra, P.H., Baume, O. and Stöhlker, U. 2011. Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computer and Geosciences* 37. 280–288.

Minasny, B. and McBratney, A.B. 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma* 140. 324–336.

Pásztor, L., Szabó, J., Bakacsi, Zs., Laborczi, A., Dobos, E., Illés, G. and Szatmári, G. 2014. Elaboration of novel, countrywide maps for the satisfaction of recent demands on spatial, soil related information in Hungary. In *Global Soil Map: Basis of the Global Spatial Soil Information System*. Eds.: Arrouays, D. *et al.* London, Taylor &Francis Group, 207–212.

Szatmári, G. 2014. Optimization of sampling configuration by spatial simulated annealing for mapping soil variables. In *6th Croatian–Hungarian and 17th Hungarian Geomathematical Congress: "Geomathematics – from theory to practice"*. Eds.: Cvetković, M., Novak Zelenika, K. and Geiger, J., Zagreb. Croatian Geological Society, 105–111.

Szatmári, G. and Barta, K. 2012. Az erózió, az erózió-veszélyeztetettség és a területhasznosítás kapcsolata mezőföldi területeten (Relationship between water erosion, potential erosion and land use on an area in the Mezőföld region). *Agrokémia és Talajtan* 61. (1): 41–56.

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.

van Groenigen, J.W. and Stein, A. 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* 27. 1078–1086.

van Groenigen, J.W., Siderius, W. and Stein, A. 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87. 239–259.

Wang, J.-F., Stein, A., Gao, B.-B. and Ge, Y. 2012. A review of spatial sampling. *Spatial Statistics* 2. 1–14.

Webster, R. and Oliver, M.A. 2007. *Geostatistics for Environmental Scientists* 2nd Ed. Chichester, Wiley, 330 p.

Wischmeier, W.H. and Smith, D.D. 1978. *Predicting rainfall erosion losses: A guide to conservation planning*. Washington D.C., U.S. Government Printing Office, 58 p.

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

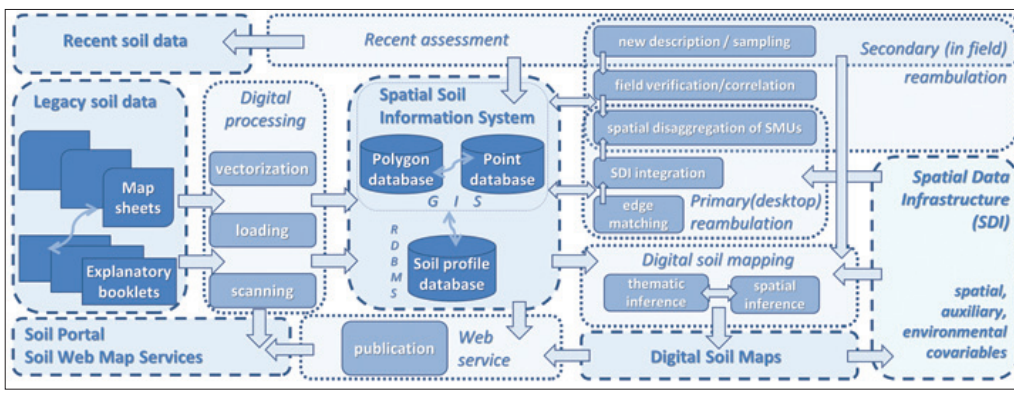


Fig. 1. Framework of spatial soil information services. Dashed line: information sources; dotted line: data flow/transportation between various elements

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

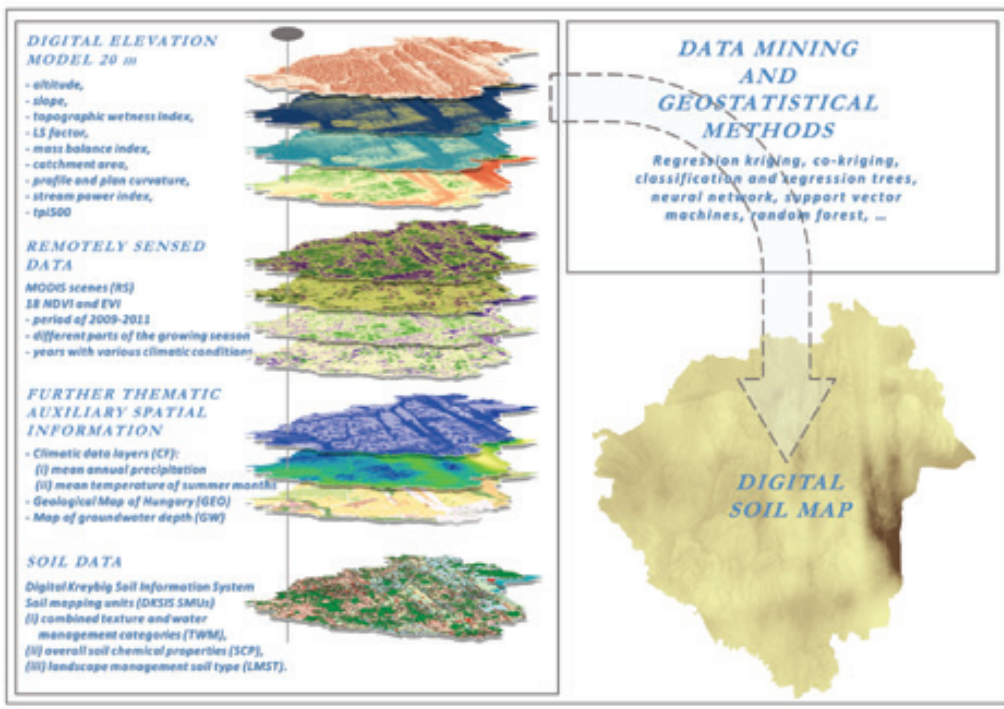


Fig. 2. Concept of digital soil mapping

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

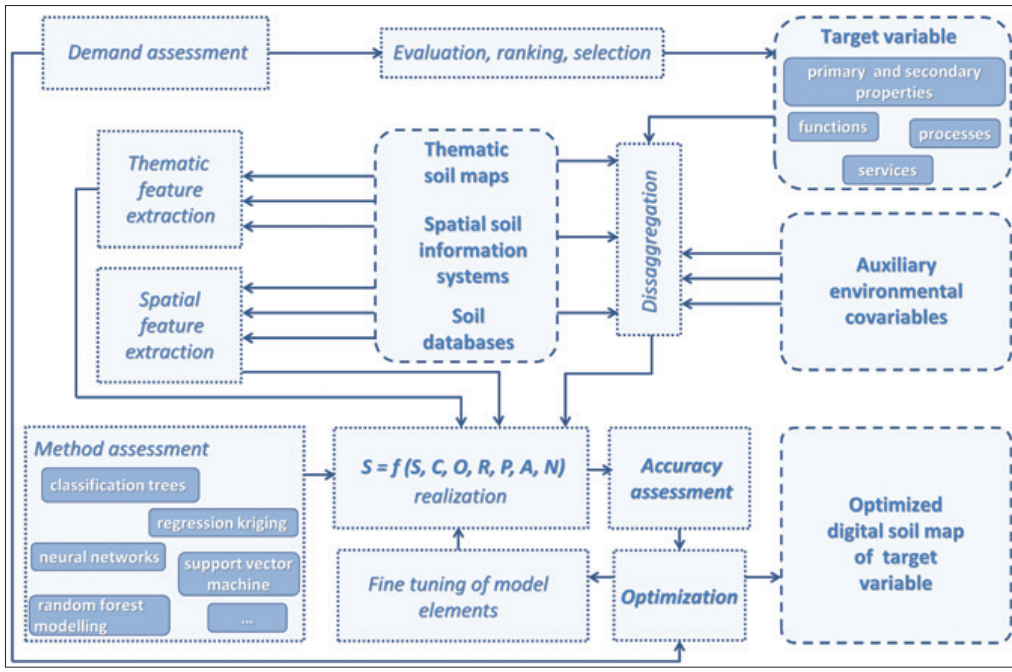


Fig. 3. Framework of the DOSoReMi.hu project

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 km²; 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 agro-ecological 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 tradi-

tional 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

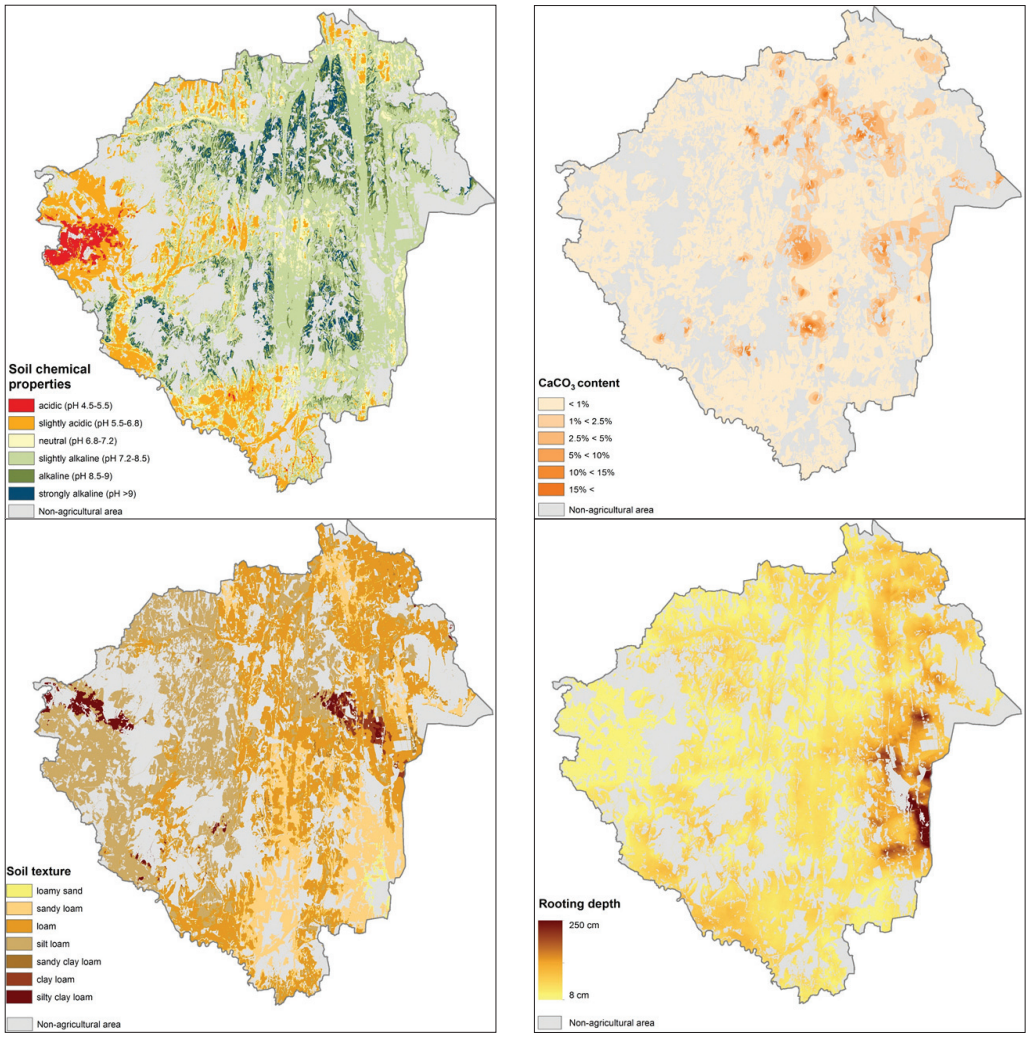


Fig. 4. Basic soil property maps of Zala County

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

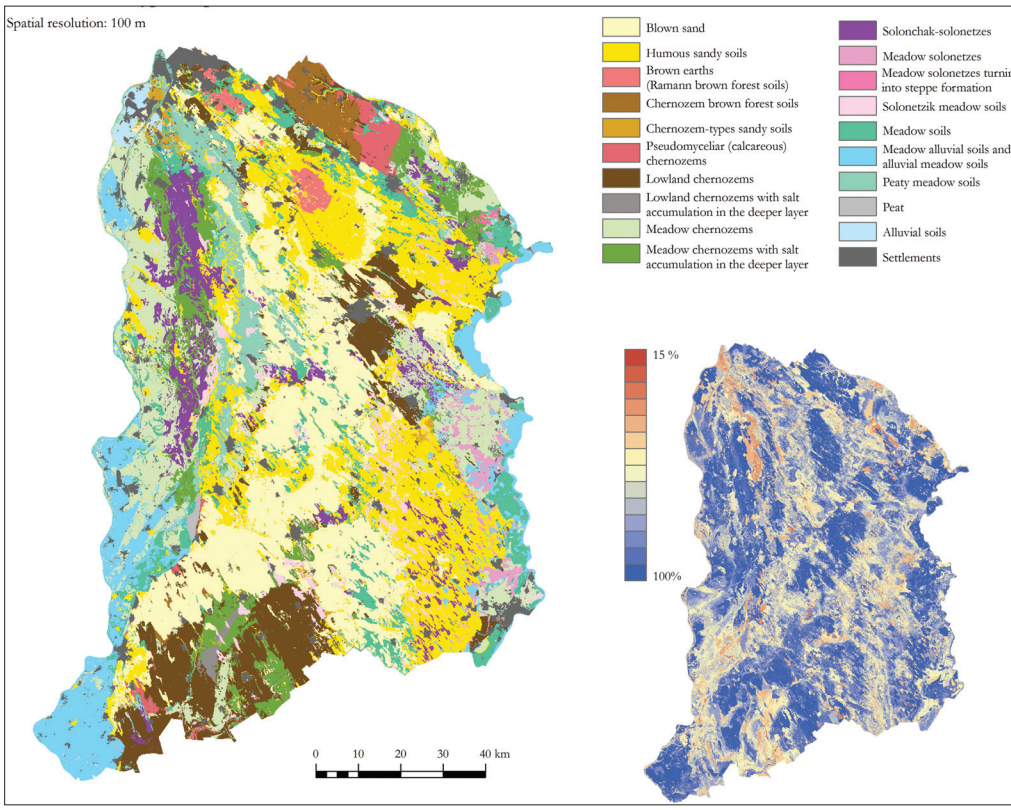


Fig. 5. Predicted soil type map for the Danube–Tisza Interfluve (on the left) and estimated reliability of the spatial prediction (on the right)

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

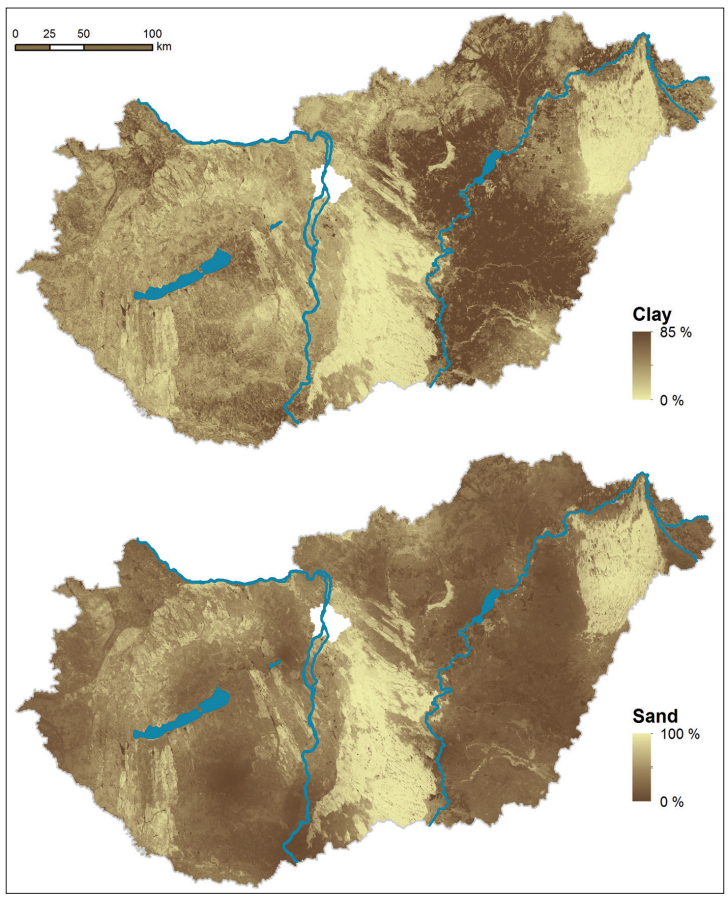
Maps of particle size classes (clay, silt and sand content) have been multi-mapped, that is the same target variable has been predicted in various ways. Either the method or the predictor variables have been selected in differing way. The purpose has been to identify the best performing constellation. 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

Parameters	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.

DOI: 10.15201/hungeobull.64.1.6 Hungarian Geographical Bulletin **64** 2015 (1) 65–78.

Method development to extract spatial association structure from soil polygon maps

István SISÁK[1], Mihály KOCSIS[1], András BENŐ[1] and Gábor VÁRSZEGI[2]

Abstract

Existing soil information systems contain mainly qualitative data on soilscapes, however, quantitative data would be necessary to more effectively guide digital soil mapping efforts. Detailed analysis of small scale overview maps offers the most appropriate way to delineate soilscapes where they are available. In our study, the genetic soil map of Hungary have been used which displays the most complete representation of the Hungarian Soil Classification System. Our goal was to analyse spatial association structure based on the boundary segments between soil polygons. We transformed the polygons into lines. The features of each line segment were the names (or codes) of the soil polygons on both sides. After omission soils with low representation (less than three polygons) and boundaries beside state border, forests and cities, 69 soil units were retained. We calculated a similarity matrix among soil types based on logarithm of ratios between existing segment lengths and theoretical segment lengths. The theoretical lengths were calculated with a Chi-squared calculation by using sums of lengths in rows and columns in the 69 × 69 matrix. The similarity matrix was converted into dissimilarity matrix to distinguish between complete dissimilarity (missing values) and complete similarity (main diagonal). Dissimilarity matrix was clustered and represented in a form of dendrogram both in original form and after dimension reduction with multidimensional scaling method. Our method has resulted a promising approach for delineating soilscapes in presence of overview soil maps. The study resulted fuzzy soilscapes with broad transition zones. The method could be refined by using variable sized moving window method and by combining boundary data with terrain, geology etc.

Keywords soilscape quantification, genetic soil map of Hungary, boundary segment based Chi-squared calculation, hierarchical clustering, multidimensional scaling

Introduction

Since the work of Dokuchaev, the axiom of the soil science is that soil forming factors (climate, geology, hydrology, biota, elevation, time and humans) and their specific interaction determine soil formation and soil properties. Jenny, H. (1941) suggested that these complex relationships should be described with mathematical formulas thus, qualitative and quantitative

[1] Department of Plant Production and Soil Science, Georgikon Faculty, University of Pannonia, H-8360 Keszthely, Deák F. u. 16. E-mails: talajtan@georgikon.hu, kocsis.mihaly@2010.georgikon.hu, beno.andras@gmail.com
[2] Department of Agro-environment Coordination, Directorate of Plant and Soil Protection and Agro-environmental issues, National Food Chain Safety Office H-1024 Budapest, Keleti K. u. 24.
E-mail: varszegig@nebih.gov.hu

soil properties will be predictable. McBratney, A.B. *et al.* (2003) gave an overview on digital soil mapping (DSM) which is Jenny's idea put into practice with help of GIS software and geostatistical analysis.

There is a tremendous complexity of soil associations in some landscapes and this requires segmentation of landscapes into soilscapes as a basis for digital soil mapping (McBratney, A.B. *et al*. 1991; Lagarcherie, P. *et al*. 2001; Schmidt, K. *et al*. 2010). Soilscape is a term introduced by Buol, S.W. *et al*. (1973) and conceptually extended by Hole, F.D. (1978) in the context of pedology. According to Lagarcherie, P. *et al.* (2001) soilscape is a landscape unit including a limited number of soil classes that are geographically distributed according to an identifiable pattern. Very often, mapping soilscapes from soil forming factor maps is more realistic than mapping soil classes. The primary task in mapping larger areas should be to account for these spatial soil-association patterns as a basis to segment landscapes (Schmidt, K. *et al*. 2010).

McSweeny, K. *et al*. (1994) proposed to set up a hierarchical multistage strategy to explain the variability of soils and soil properties in space. The second stage of the proposed method was a geomorphometric characterization of the landscape from digital terrain models, which provides (i) a land surface representation to which other data are referenced and (ii) a division of the land surface into areas that correspond with soil patterns. The recently adapted hierarchical approach to define soilscapes follows the World Soils and Terrain Digital Database (SOTER) methodology (ISRIC, 1993). SOTER has become widely evaluated in European and broader context (Dobos, E. *et al*. 2001, 2005, 2010). However, these terrain-based approaches are more appropriate for finer scales as they mainly focus on deriving terrain facets instead of deriving larger homogeneous geomorphological or pedological regions (Schmidt, K. *et al*. 2010).

Existing soil information systems store data on association of soil bodies within soilscapes or soil series in relation tables. This description is strictly qualitative (Finke, P. *et al*. 2001). Efforts have been made to better define the objects resulting from these groupings (Hewitt, A.E. 1993) and to define the criteria used in their construction (Hudson, B.D. 1990). Recent findings provide more and more quantitative results on how soil bodies are associated (Behrens, T. *et al*. 2009; Hewitt, A.E. *et al*. 2010; Schmidt, K. *et al*. 2010). The latest nationwide digital soil mapping projects in New Zealand (Hewitt, A.E. *et al.* 2010) or Ireland (Creamer, R. *et al.* 2014) adapt strong soilscape-based approach.

In spite of the recent trend (Scull, P. *et al*. 2005) that predictive soil models shift from research to operational phase, Grinand, C. *et al*. (2008) observed that soil class prediction accuracy can only be approximated correctly if test samples are collected at a certain distance from the training samples when predicting unvisited areas.

However, digital soil mapping approaches which utilize soil information from existing (usually small or medium scale) soil maps and field observations perform much better than pure theoretical constructions (Mendonça-Santos, M.D.L. *et al*. 2008). Soil maps are physical representations of the mental models of the mappers on how soil forming factors interact (Bui, E. 2004). They provide us a path through the almost infinite number of theoretically possible combinations to the most probable outcome. In countries where small or medium scale soil maps exist their statistical analysis may help to define homogenous soil regions or soilscapes and representative areas for detailed soil surveys (Behrens, T. *et al*. 2009; Schmidt, K. *et al*. 2010).

The aim of our study was to evaluate an existing nationwide soil map of Hungary and to define soil association rules which then can be used to delineate soil regions or soilscapes. We evaluated boundary line segments of neighbouring polygons and we were using Chi-squared method, hierarchical classification and multidimensional scaling in the analysis.

Materials and methods

The genetic soil map of Hungary and the conversion of its categories into WRB categories

There was a nationwide campaign in Hungary in the 1970's and 1980's to renew the old land evaluation system based on detailed new soil maps. The genetic soil map (MÉM-NAK, 1983) was released as a part of the preparation phase for the fine-scale soil mapping. The purpose of the 1:200,000 scale map was to gather all the available information and to give orientation for the field work before the detailed soil surveys. The latest field guide for soil mapping and an official version of the Hungarian Soil Classification System (HSCS) was published (Horváth, B. *et al.* 1987) as part of the project and it served as a compulsory tool for field surveyors. Soil classification system did not change much between 1983 (release of the genetic map) and 1989 (release of the field guide). Slight changes were introduced but basic concepts and categories stayed intact. The genetic soil map is the most complete display of the HSCS and also contains data on parent material, texture and chemical reaction but does not show soil data for the area of forests and larger towns. We completed and improved the digital version (AIR, 2013) of the genetic soil map of Hungary. We used only soil classes of HSCS (soil types, sub-types) in our analysis and did not use other data.

In *Table 1* we provide an approximate conversion between HSCS soil units of the genetic soil map (MÉM-NAK, 1983) based on the work of Horváth, B. *et al.* (1989) and the IUSS Working Group WRB (2007). We should state that clear one-to-one conversion is not possible at all because of the different soil investigation methods, different limit values of the individual properties and partly because of the different concepts. We still decided to use this conversion since one of the declared primary objectives of the WRB is to serve as "common language" between national soil classification systems.

Despite limitations, approximate conversion is possible (Michéli, E. *et al.* 2006; Krasilnikov, P. *et al*. 2009). We applied the following procedure:

1. We considered the basic concepts of the Reference Soil Groups (RSGs) and their qualifiers and specifiers and we used them to express similar concepts in the HSCS without strict investigations of the detailed definitions and limits.

2. Whenever the HSCS expressed properties which were not part of the specifier set of the given RSG, we used similar specifiers from other RSGs but we added them in *italics*.

3. If the Hungarian concept was not included in the WRB concepts, we added a short explanation in *italics*.

Codes are also an easy way to identify soil units in the figures and tables. We decided to provide approximate categories of an earlier version of the WRB (IUSS Working Group WRB, 2007) because this has been well known in the soil science community. Newly introduced changes (IUSS Working Group WRB, 2014) may not be well established beyond experts in soil classification.

The HSCS contains 99 individual units either as soil types (e.g. 10 Lithic Leptosol) or sub-types (e.g. 31 Haplic Regosol, Calcaric). The code of the soil types can be divided by ten without remainder (see *Table 1*). The codes of the sub-types contain numbers in the place of the last digit other than zero. The MÉM-NAK (1983) soil map displays 81 different soil units. However, some of them are represented only by three or less polygons and those were excluded from our analysis. On this way, 69 soil units were retained and converted into approximate WRB units *(Table 1).*

Data analysis

In the first step we determined the length of each line segment between the soil category polygons (soil types or sub-types). The boundary lines at the state border or in the neighbourhood of forests, lakes or towns were not considered since only one of the

Table 1. Approximate conversion of the units in the Hungarian Soil Classification System into WRB categories

Soil unit codes[1]	Soil unit names in the Hungarian Soil Classification System[1] (in Hungarian in the brackets)	Approximate equivalent in the WRB classification[2]
10	Stony, rocky skeletal soil (Köves sziklás váztalaj)	(Nudi-)Lithic Leptosol
20	Gravelly skeletal soil (Kavicsos váztalaj)	Hyperskeletic Leptosol
31	Calcareous earthy barren (Karbonátos földes kopár)	Haplic Regosol, Calcaric
41	Calcareous blown sand (Karbonátos futóhomok)	Protic Arenosol, Aridic, Calcaric
42	Non-calcareous blown sand (Nem karbonátos futóhomok)	Protic Arenosol, *not calcareous*
45	„Kovárvány" blown sand (Kovárványos futóhomok)	Lamellic Arenosol
51	Calcareous humic sand (Karbonátos humuszos homok)	Haplic Arenosol, Calcaric
52	Non-calcareous humic sand (Nem karbonátos humuszos homok)	Haplic Arenosol, *not calcareous*
53	Calcareous multilayered humic sand (Karbonátos több rétegű humuszos homok)	Haplic Arenosol, Calcaric *with buried A horizon(s)*
54	Non-calcareous multilayered humic sand (Nem karbonátos több rétegű humuszos homok)	Haplic Arenosol *not calcareous with buried A horizon(s)*
60	Humus-carbonate soil (Humuszkarbonát talaj)	Haplic Regosol, Humic, Calcaric
71	Black rendzina (Fekete rendzina)	Rendzic Leptosol *black usually on limestone*
72	Brown rendzina (Barna rendzina)	Rendzic Leptosol *brown usually on dolomite*
112	Nonpodzolic brown forest soil with clay illuviation (Nem podzolos agyag-bemosódásos barna erdőtalaj – BET)	Haplic Luvisol
121	Podzolic pseudogley brown forest soil (Podzolos pszeudoglejes BET)	Albic Stagnic Luvisol, Manganiferric
122	Pseudogley brown forest soil with clay illuviation (Agyagbemosódásos pszeudo-glejes BET)	Stagnic Luvisol
131	Typical Ramann's brown forest soil (Típusos Ramann-féle BET)	Haplic Cambisol, Eutric, Siltic
132	Rust brown Ramann's brown forest soil (Rozsdabarna erdőtalaj)	Brunic Arenosol, Eutric, Chromic
141	Typical brown forest soil with alternating thin layers of clay substance („kovárvány") (Típusos kovárványos BET)	Brunic Lamellic Arenosol
143	„Kovárvány" brown forest soil with clay illuviation (Agyagbemosódásos kovárványos BET)	Lamellic Luvisol
161	Calcareous chernozem brown forest soil (Karbonátos csernozjom BET)	Haplic Cambisol, Humic, Calcaric
162	Non-calcareous chernozem brown forest soil (Nem karbonátos csernozjom BET)	Haplic Cambisol, Humic
171	Calcareous chernozem soils with forest residues (Karbonátos erdőmaradványos csernozjom)	Luvic Phaeozem, Calcaric
172	Non-calcareous chernozem soils with forest residues (Nem karbonátos erdőmaradványos csernozjom)	Luvic Phaeozem, *not calcareous*

Table 1. Continued

Soil unit codes[1]	Soil unit names in the Hungarian Soil Classification System[1] (in Hungarian in the brackets)	Approximate equivalent in the WRB classification[2]
180	Leached chernozem soil (Kilúgzott csernozjom)	Haplic Chernozem, Pachic
191	Typical calcareous chernozem soil (Típusos meszes vagy mészlepedékes csernozjom)	Calcic Chernozem, Pachic, Siltic
192	Lowland calcareous chernozem soils (Alföldi meszes vagy mészlepedékes csernozjom)	(Endosalic) Calcic Chernozem, Pachic
201	Calcareous meadow chernozem soils (Karbonátos réti csernozjom)	*Bathy*gleyic Calcic Chernozem, Pachic
202	Non-calcareous meadow chernozem soil (Nem karbonátos réti csernozjom)	*Bathy*gleyic Chernozem, Pachic, *not calcareous*
203	Meadow chernozem soil, salty in deeper horizons (Mélyben sós réti csernozjom)	Endosalic *Bathy*gleyic Chernozem, Pachic
204	Meadow chernozem soils, solonetz-like in deeper horizons (Mélyben szolonyeces réti csernozjom)	*Bathy*gleyic Chernozem, Pachic *with sodicity in the parent material*
205	Solonetz-like meadow chernozem soil (Szolonyeces réti csernozjom)	*Bathy*gleyic Chernozem, Pachic *with sodicity in the subsurface soil horizon*
211	Calcareous terrace chernozem soil (Karbonátos terasz csernozjom)	Calcic Endofluvic Chernozem
221	Calcareous solonchak soil (Karbonátos szoloncsák)	Calcic Solonchak, Carbonatic
231	Calcareous solonchak-solonetz soil (Karbonátos szoloncsák-szolonyec)	Calcic Salic Solonetz *(Carbonatic)*
232	Calcareous and sulphate-containing solonchak-solonetz soil (Karbonátszulfátos szoloncsák-szolonyec)	Calcic Salic Solonetz *(Carbonatic Sulphatic)*
241	Shallow meadow solonetz soil (Kérges réti szolonyec)	Calcic Solonetz *with an A horizon shallower than 7 cm*
242	Medium meadow solonetz soil (Közepesen mély réti szolonyec)	Calcic Solonetz *with an A horizon between 7 and 20 cm*
243	Deep meadow solonetz soil (Mély réti szolonyec)	Calcic Solonetz *with an A horizon deeper than 20 cm*
251	Medium meadow solonetz turning into steppe formation (Közepesen mély sztyepesedő réti szolonyec)	Mollic Calcic Solonetz
281	Sulphate- or chloride-containing solonchak-like meadow soils (Szulfátos vagy kloridos szoloncsákos réti talaj)	Mollic Gleysol *with salt accumulation in the surface horizon (Sulphatic or Chloridic)*
282	Calcareous solonchak-like meadow soils (Karbonátos szoloncsákos réti talaj)	Mollic Gleysol, Calcaric *with salt accumulation in the surface horizon (Carbonatic)*
291	Solonetz-like meadow soils (Szolonyeces réti talaj)	Mollic Gleysol, (Hypo-)sodic
292	Strongly solonetzized solonetz-like meadow soils (Erősen szolonyeces réti talaj)	Mollic Gleysol, Sodic
301	Calcareous meadow soils (Karbonátos réti talaj)	Calcic Mollic Gleysol
302	Non-calcareous meadow soils (Nem karbonátos réti talaj)	Mollic Gleysol, *not calcareous*

Table 1. Continued

Soil unit codes[1]	Soil unit names in the Hungarian Soil Classification System[1] (in Hungarian in the brackets)	Approximate equivalent in the WRB classification[2]
303	Meadow soils, salty in deeper horizons (Mélyben sós réti talaj)	Endosalic Mollic Gleysol
311	Calcareous alluvial meadow soils (Karbonátos öntés réti talaj)	Fluvic Mollic Gleysol, Calcaric
312	Non-calcareous alluvial meadow soils (Nem karbonátos öntés réti talaj)	Fluvic Mollic Gleysol, *not calcareous*
321	Typical marshy meadow soils (Típusos lápos réti talaj)	Histic Gleysol
331	Calcareous chernozem meadow soils (Karbonátos csernozjom réti talaj)	Calcic Gleyic Chernozem, (Pachic)
332	Non-calcareous chernozem meadow soils (Típusos csernozjom réti talaj)	Gleyic Chernozem, (Pachic)
333	Chernozem meadow soils,salty-like in deep-er layers (Mélyben sós csernozjom réti talaj)	Endosalic Gleyic Chernozem
334	Chernozem meadow soils, solonetz-like in deeper layers (Mélyben szolonyeces csernozjom réti talaj)	Gleyic Chernozem *with sodicity in the subsurface soil horizon*
335	Solonetz-like chernozem meadow soils (Szolonyeces csernozjom réti talaj)	Gleyic Chernozem *with sodicity in the parent material*
350	Peat-bog soils (Rétláp talaj)	Fibric Histosol
360	Drained and cultivated lowmoor fen soils (Lecsapolt és telkesített rétláp talaj)	Histosol, Drainic *in general*
361	Drained peat-bog soil (Lecsapolt tőzegláp)	Hemic Histosol Drainic
362	Drained peaty fen soil (Lecsapolt tőzeges láp)	Hemic Histosol, Drainic *with less than 50 cm deep Histic horizon*
363	Drained fen soil with highly decomposed peaty substance 'kotu' (Lecsapolt kotus láp)	Sapric Histosol, Drainic
364	Cultivated low moor fen soil (Telkesített rétláp)	Hemic Histosol, Drainic *with regulated water level*
370	Soils of marsh and alluvial forests (Mocsári erdőtalaj)	Haplic Gleysol, Dystric (alternative: Acric Gleysol)
381	Calcareous recent alluvial soils (Karbonátos nyers öntés talaj)	Haplic Fluvisol, Calcaric
391	Calcareous humic alluvial soils (Karbonátos humuszos öntés talaj)	Haplic Fluvisol, Humic, Calcaric
392	Non-calcareous humic alluvial soils (Nem karbonátos humuszos öntés talaj)	Haplic Fluvisol, Humic, *not calcareous*
393	Calcareous multilayered humic alluvial soils (Karbonátos, több rétegű humuszos öntés talaj)	Haplic Fluvisol, Humic, Calcaric *with buried A horizon in the upper 150 cm*
394	Non-calcareous multilayered humic alluvial soils (Nem karbonátos több rétegű humuszos öntés talaj)	Haplic Fluvisol, Humic *with buried A horizon in the upper 150 cm, not calcareous*
395	Meadow-like humic alluvial soils (Réti öntés talaj)	Mollic Gleyic Fluvisol
402	Slope deposits of forest soils (Erdőtalaj eredetű lejtőhordalék talajok)	Colluvic Regosol *derived mainly from Luvisols and Cambisols*

[1]According to Horváth, B. *et al.* 1989, [2]According to IUSS Working Group WRB 2007.

neighbouring polygons had soil data. Then we calculated the sum of lengths for each soil category combinations and thus, we got a square matrix with dimensions of 69 by 69.

The values in the main diagonal were dismissed (set to zero) since they represented the same category with slightly different properties (texture or pH). Then we calculated the following theoretical length for each matrix element:

$$L_{ij\text{-}est} = \Sigma L_i \times \Sigma L_j / L_{tot},$$

where $L_{ij\text{-}est}$ = the estimated length for an individual category combination, L_i = the total length of the *i*-th category in the rows of the matrix, L_j = the total length of the *j*-th category in the columns of the matrix, L_{tot} = the total length of all categories (grand total of the matrix).

Then we have calculated the following *P* similarity (neighbourhood) matrix:

$$P_{ij} = \log [(L_{ij} / L_{ij\text{-}est}) \times 100],$$

where L_{ij} = the actual length for an individual category combination. This is the logarithm of the percent ratio between actual and theoretical lengths. Zero values in the main diagonal and missing combinations have no logarithm thus, in this similarity matrix we cannot distinguish between complete similarity (main diagonal) and complete dissimilarity (non-existent combinations). To alleviate this problem, we converted the similarity matrix into $P`_{ij}$ dissimilarity (distance) matrix. All length ratios were less than 100,000 thus, we selected 5 (= log 100,000) as the maximum dissimilarity.

$$P'_{ij} = \begin{cases} 0 & \text{if } P_{ij} = 0 \\ 5 - P_{ij} & \text{if } P_{ij} > 0 \\ 5 & \text{if } P_{ij} = \text{missing} \end{cases}$$

We performed hierarchical cluster analysis with $P`_{ij}$ matrix and presented the results in form of a dendogram. The dimensionality of this matrix is 69 with regard to the soil categories as variables. However, the dissimilarity matrix had several missing combinations and we assumed that the dimensionality can be significantly reduced without much loss of information. We applied the multidimensional scaling procedure to find a simpler and more general structure. Then we applied the hierarchical clustering to the new matrix again and represented the results with another dendogram. We used ArcGIS 10.0 for map data handling and interpretation and SPSS 13.0 for data analysis.

Results and discussion

The frequency distribution of the $P`_{ij}$ distance (dissimilarity) matrix has been shown in *Figure 1* without the values of 5 and 0. The histogram was calculated from the full matrix which means that all values are in duplicate. The distribution is close to the normal. For the half matrix when each combination is considered only once, there are 2,346 possible combinations between 69 soil categories but only 779 of them (33.2%) really exist which means that soil categories can be neighbours of only a subset of other categories which is trivial.

Chi-square statistics are often used for overlaid categorical maps in land use change studies (Pontius, Jr. R.G. 2002). However, the appropriateness of method drew also criticism because mapped area has no clear, statistically independent "case" thus, its error model is flawed (Chrisman, N.R. 1989) and the pixel size or the area of measurement unit will determine the "degree of freedom"

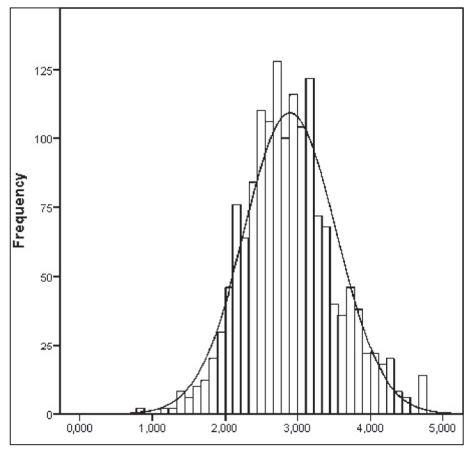


Fig. 1. Data distribution in the dissimilarity matrix

in the test. Similar objections are true for Chi-squared statistics with line segments. However, we did not use the Chi-square calculation in our study to test any significance; we just calculated the P_{ij} matrix elements from segment lengths in a similar way as in Chi-square method without entering into the questioned test calculation.

The resulting dendogram calculated from the first, not simplified distance matrix can be seen in *Figure 2*. After reducing the dimensionality with the PROXSCALE procedure, we got 5 dimensions instead of the previous 69 whereby 7 percent of the information was lost as indicated by the stress-test of the procedure. The second hierarchical clustering with the reduced, five-dimensional matrix has resulted the dendogram shown in *Figure 3*.

There are numerous differences between the two dendograms but generally, the second one has a much more separated structure between the branches than the first one.

The following two soil types are loosely associated with each other and they are rather separated from other categories in the first dendogram (*Figure 2*):

202: Bathygleyic Chernozem, Pachic, *not calcareous,*

301: Calcic Mollic Gleysol.

They lost their separation from other branches, but retained some degree of their association as members of the same group (cluster 3c in *Figure 3*) after dimensionality reduction, however, they were directly associated with other soil categories:

202: Bathygleyic Chernozem, Pachic, *not calcareous,*

363: Sapric Histosol, Drainic,

364: Hemic Histosol, Drainic *with regulated water level* and

301: Calcic Mollic Gleysol,

172: Luvic Phaeozem, *not calcareous.*

The dimensionality reduction may bring forward relationships which explain soil formation processes such as Stagnic Luvisol (112) became associated with Colluvic Regosol *derived mainly from Luvisols and Cambisols* (402) in cluster 5b (*Figure 3*) which association was not so close in the first dendogram (*Figure 2*).

There are very closely related soil categories which, in theory, should express different degree of groundwater influence coupled with strong organic matter accumulation such as Bathygleyic Chernozems (201–205) and Gleyic Chernozems (331–335) as seen in *Figure 3* (clusters 1 and 2c). However, even the latest official field guide (Horváth, B. *et al.* 1989) does not provide enough support to tell them apart in the field. Our analysis points out specific weaknesses in the HSCS which need more precise definitions as part of the necessary future development of the HSCS according to the diagnostic principles (Michéli, E. *et al*. 2006; Krasilnikov, P. *et al*. 2009). *Figure 4* shows the map of soil clusters indicated in *Figure 3*.

There is a clear regional distribution of clusters within the area of the country. The clusters marked with "A" are situated on the Great Plain (South-East part of Hungary) and to lesser extent on the Small Hungarian Plain (North-West part). Most of the clusters marked with "D" are situated on the hilly regions with some remarkable exceptions (D_3b and D_4b) which are associated with sandy regions and large rivers on the Great Hungarian Plain. The lead soil types within the clusters are provided in *Table 2.* At that, we followed the method of Schmidt, *et al.* (2010) instead of trying to characterize the complete soil associations. Further investigation of the association rules and their regional differences can be the objective of future studies. The major soil type gives more than 2/3 of the area within the cluster in five clusters, this ratio is between 1/3 and 2/3 in four clusters and it is below 1/3 in two clusters. The latter two are on lowland where the genetic soil map shows larger pedodiversity.

Close proximity in the dendogram may originate from strong association in one region but in other region this relatedness does not exist sometimes simply because one of the soil categories is not present in the other region. This observation is most striking for the cluster 5a (Mollic Gleysol, *not calcareous* and associated soils). Stagnic Luvisols are included in this cluster (code 121 and 122) and they are

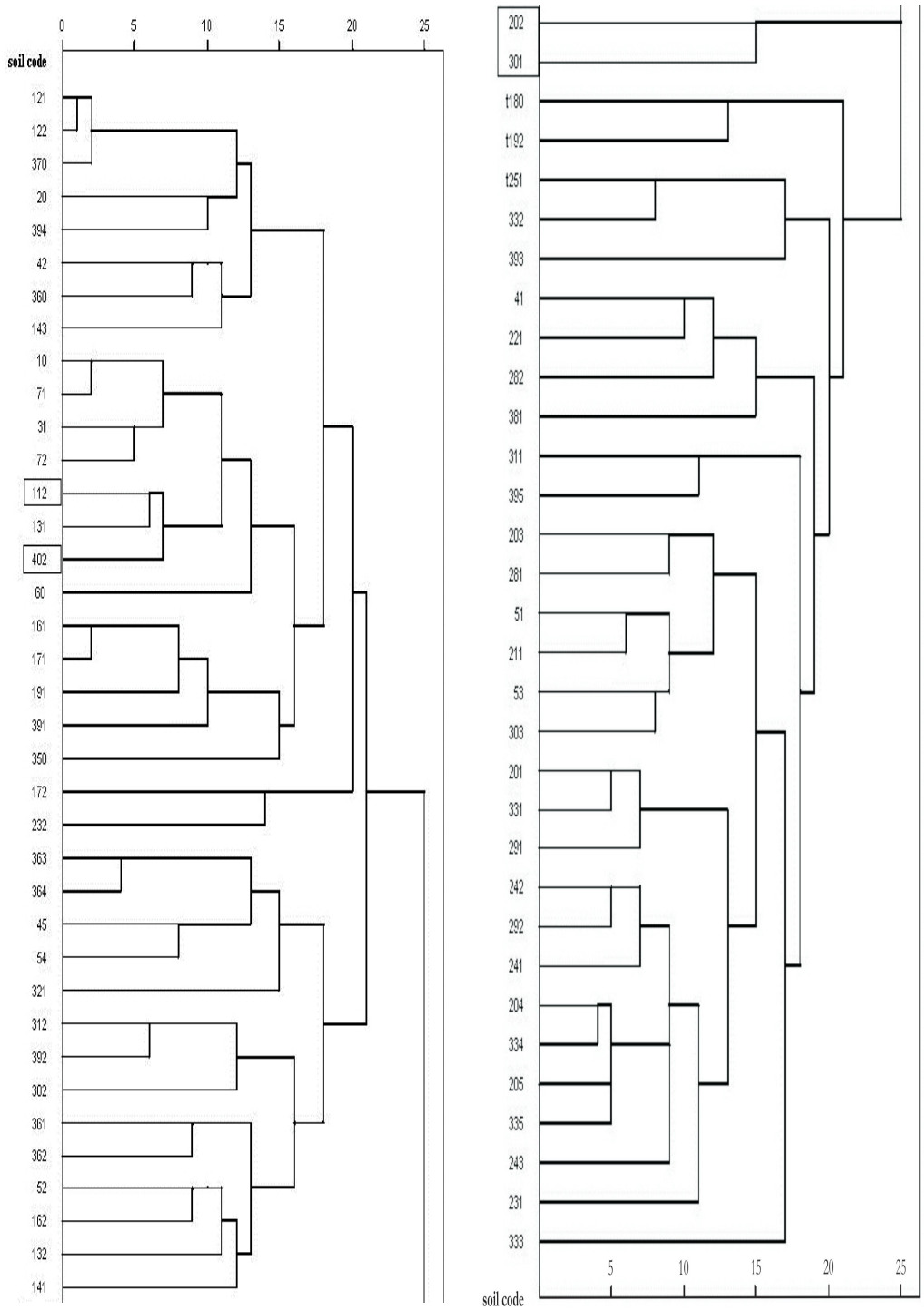


Fig. 2. Dendogram derived by hierarchical clustering from the original dissimilarity matrix

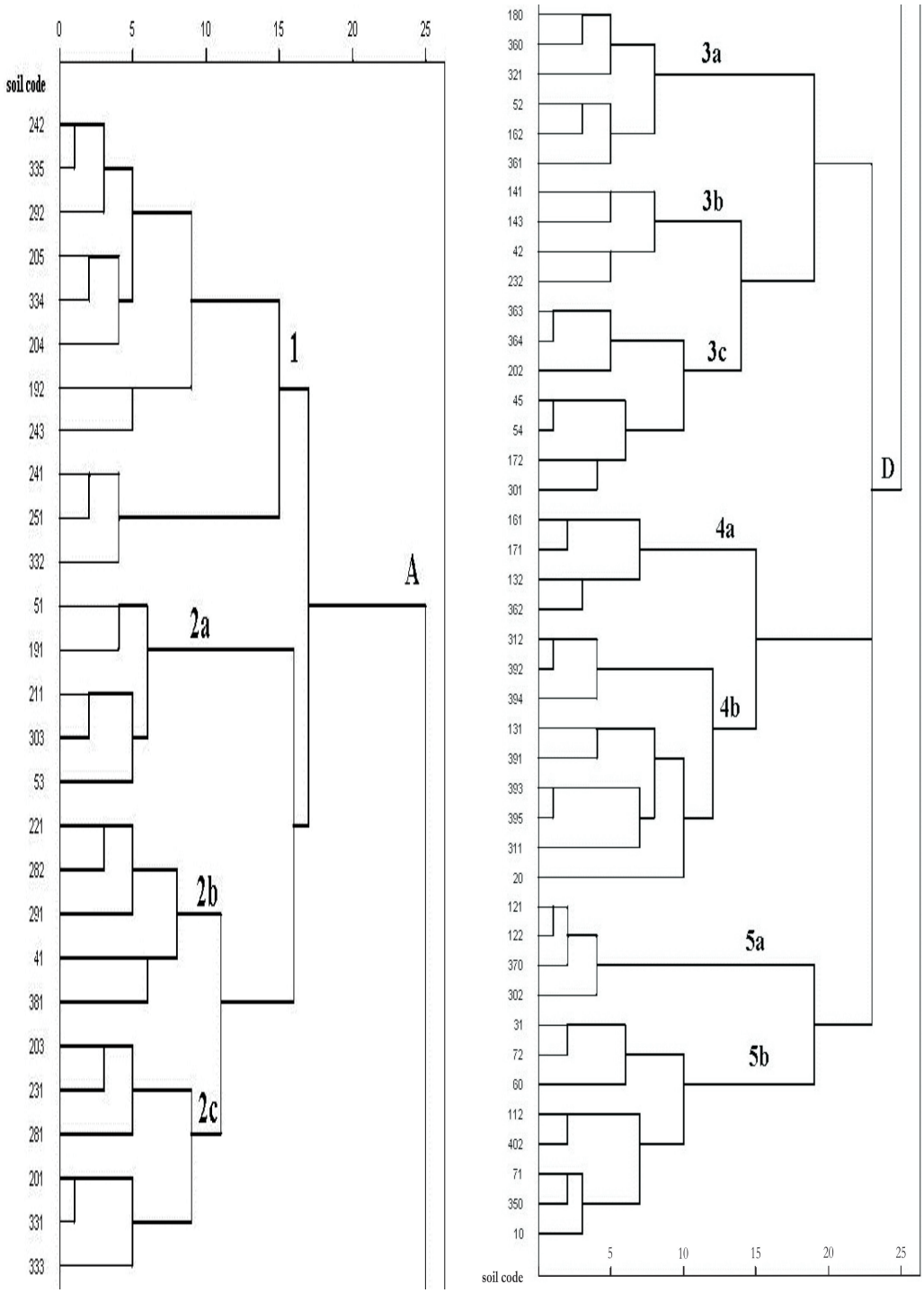


Fig. 3. Dendogram derived by hierarchical clustering from the dissimilarity matrix after dimension reduction

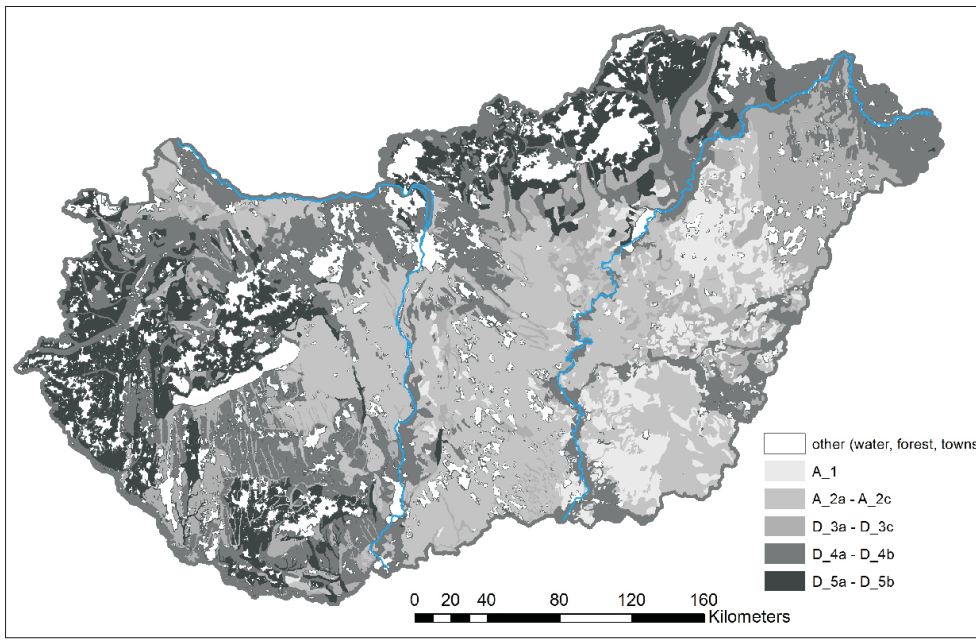


Fig. 4. Map of the soil type clusters

Table 2. Clusters in the dendogram and the major soil type in the cluster

Cluster No. in *Figure 3*	Legend in *Figure 4*	Approximate WRB equivalent of the major soil type in the cluster and its code No.	Area % within the cluster
1	A_1	205: *Bathy*gleyic Chernozem, Pachic *with sodicity in the subsurface soil horizon*	20.3
2a	A_2a–A_2c	51: Haplic Arenosol, Calcaric	45.3
2b		291: Mollic Gleysol, (Hypo-)sodic	68.7
2c		201: *Bathy*gleyic Calcic Chernozem, Pachic	72.1
3a	D_3a–D_3c	162: Haplic Cambisol, Humic	37.7
3b		141: Brunic Lamellic Arenosol	76.5
3c		301: Calcic Mollic Gleysol	71.5
4a	D_4a–D_4b	132: Brunic Arenosol, Eutric, Chromic	55.3
4b		131: Haplic Cambisol, Eutric, Siltic	25.5
5a	D_5a–D_5b	302: Mollic Gleysol, *not calcareous*	56.8
5b		112: Haplic Luvisol	84.7

common near the Western border of Hungary. The cluster presents itself in other parts of the country but Stagnic Luvisols do not.

Soils in a landscape are associated spatially as well as taxonomically (Hole, F.D. 1978). However, spatially associated soils might not be associated taxonomically (Campbell, J.B. and Edmons, W.J. 1984). Thus, a spatial approach seems appropriate to derive soilscapes as a basis for subsequent digital soil-mapping purposes (Schmidt, K. *et al*. 2010).

According to the summarizing works by McBratney, A.B. *et al.* (2003) and Scull, P. *et al.* (2003) tree-based methods are rapidly gaining popularity as means to develop prediction rules that can be rapidly and repeatedly

evaluated. Because of the clear advantages, several authors applied tree-based methods for soil mapping problems (Hengl, T. *et al.* 2007; Grinand, C. *et al*. 2008; Cambule, A.H. *et al.* 2013; Sun, X.L. *et al*. 2011; Häring, T. *et al*. 2012; Pásztor, L. *et al.* 2013). Complex similarity (relatedness) or dissimilarity (distance) matrices and their analysis in tree form are routine procedures in several disciplines such as in psychology (Pecora, L.M. *et al.* 1995) genetics (Yu, J. *et al.* 2005) or in scientometrics (Boyack, K.W. *et al.* 2005). One of the early publications is on representing demographic data (Hartigan, J.A. 1967).

However, there is no evidence in the scientific literature that boundary line segments between soil polygons would have ever been analyzed and spatial association rules would have been extracted as trees from legacy soil maps. Compared to other regionalization studies (Schmidt, K. *et al.* 2010; Lilburne, L.R. *et al.* 2012), we used only boundary segments and soil classes on both sides of the line instead of complex data sets on soil, terrain, geology and other surface properties and analyzed the whole data set instead of subsetting by moving window method with rasterized data (Behrens, T. *et al*. 2009; Schmidt, K. *et al.* 2010). The consequence of our approach is that the region boundaries are rather fuzzy with large mosaicked transition zones around the more homogenous core zones *(Figure 4).* Variable sized moving window method (Behrens, T. *et al*. 2009; Schmidt, K. *et al*. 2010) combined with our boundary line approach may result more homogenous soilscapes. This combination of methods may alleviate the problem of Stagnic Luvisols mentioned above where existing associations in one region were false in another region in spite of the presence of the same cluster simply because one soil class was missing.

Conclusions

There are three nationwide legacy soil maps in Hungary. The first one was published in 1953 at a scale of 1:200,000 (Mattyasovszky, J. *et al.* 1953), the second one (popularly called AGROTOPO) was published between 1983 and 1988 on 1:100,000 sheets (Várallyay, Gy. *et al.* 1979, 1980; MÉM 1983–1988) and the third one (genetic soil map) was compiled by the experts of the agricultural extension agency of the agricultural ministry in 1983 at scale of 1:200,000 (MÉM-NAK 1983). The genetic soil map provides the most complete display of the HSCS thus it is the most appropriate basis for soilscape analysis. Despite its relative completeness, it does not contain all the soil types and sub-types of the HSCS. Further digital soil mapping works are needed since spatial resolution of existing maps are insufficient to the requirements of the policy making (Pásztor, L. *et al.* 2013; Sisák, I. and Benő, A. 2012, 2014).

In conclusion, our method has resulted a promising approach for delineating soilscapes in presence of overview soil maps. We used the method for whole area of Hungary but it has resulted fuzzy soilscapes with broad transition zones. The method could be refined by using variable-sized moving window method and by combining boundary data with terrain, geology etc.

***Acknowledgement*:** Present article was published in the frame of the project TÁMOP-4.2.2.A-11/1/KONV-2012-0064. The project is realized with the support of the European Union, with the co-funding of the European Social Fund. The data analysis was supported by the OTKA K101065 project

REFERENCES

AIR 2013. *Agrár-környezetgazdálkodási Indormációs rendszer* (Information System for the Agri-environmental schemes). Nyilvános térképek (Open access maps). (in Hungarian) http://terkep.air.gov.hu/terkep/nyilvanos/nyilvanos.htm

Behrens, T., Schneider, O., Lösel, G., Scholten, T., Hennings, V., Felix-Henningsen, P. and Hartwich, R. 2009. Analysis on pedodiversity and spatial subset representativity – the German soil map 1:1 000 000. *Journal of Plant Nutrion and Soil Science* 172. 91–100.

Boyack, K.W., Klavans, R. and Börner, K. 2005. Mapping the backbone of science. *Scientometrics* 64. (3): 351–374.

Bui, E. 2004. Soil survey as a knowledge system. *Geoderma* 120. (1–2): 17–26.

Buol, S.W., Hole, F.D. and McCracken, R.J. 1973. *Soil genesis and classification*. Ames, University Press, IA, USA. 404 p.

Cambule, A.H., Rossiter, D.G. and Stoorvogel, J.J. 2013. A methodology for digital soil mapping in poorly-accessible areas. *Geoderma* 192. 341–353.

Campbell, J.B. and Edmonds, W.J. 1984. The missing geographic dimension to soil taxonomy. *Annals of the Association of American Geographers* 74. 83–97.

Chrisman, N.R. 1989. Modelling error in overlaid categorical maps. In *Accuracy of spatial databases.* Eds.: Goodchild, M. and Gopal S. London, Taylor & Francis, 21–34.

Creamer, R., Simo, I., Reidy, B., Carvalho, J., Fealy, R., Hallett, S., Jones, R., Holden, A., Holden, N., Hannam, J., Massey, P., Mayr, T., McDonald, E., O'Rourke, S., Sills, P., Truckell, I., Zawadzka, J. and Schulte R. 2010. *Irish Soil Information SystemSynthesis Report.* No. 130, Johnstown Castle, Environmental Protection Agency, Ireland.

Dobos, E., Daroussin, J. and Montanarella, L. 2005. *An SRTM-based procedure to delineate SOTER Terrain Units on 1:1 and 1:5 million scales.* EUR 21571 EN, Luxembourg, Office for Official Publications of the European Communities. 55 p.

Dobos, E., Daroussin, J. and Montanarella, L. 2010. A quantitative procedure for building physiographic units supporting a global SOTER database. *Hungarian Geographical Bulletin* 59. (2): 181–205.

Dobos, E., Montanarella, L., Nègre, T. and Michéli, E. 2001. A regional scale soil mapping approach using integrated AVHRR and DEM data. *International Journal of Applied Earth Observation and Geoinformation* 3. (1): 30–42.

Finke, P., Hartwich, R., Dudal, R., Ibàñez, J., Jamagne, M., King, D., Montanarella, L. and Yassoglou, N. 2001. *Georeferenced Soil Database for Europe. Manual of procedures. Version 1.1.* European Soil Bureau, Research Report No. 5. EUR 18092 EN. European Communities

Grinand, C., Arrouays, D., Laroche, B. and Martin, M.P. 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma* 143. (1–2): 180–190.

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, Th. and Schröder, B. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185–186. 37–47.

Hartigan, J.A. 1967. Representation of similarity matrices by trees. *Journal of the American Statistical Association* 62. (320.) 1140–1158.

Hengl, T., Toomanian, N., Reuter, H.I. and Malakouti, M.J. 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma* 140. 417–427.

Hewitt, A.E., 1993. Predictive modelling in soil survey. *Soil and Fertilizers* 56. (3): 305–314.

Hewitt, A.E., Barringer, J.R.F., Forrester, G.J. and McNeill, S.J. 2010. Soilscapes basis for digital soil mapping in New Zealand. 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. Dortdrecht, Springer, The Netherlands, 297–308.

Hole, F.D. 1978. An approach to landscape analysis with emphasis on soils. *Geoderma* 21. 1–23.

Horváth, B., Izsó I., Jassó, F., Király, L., Parászka, L. and Várallyay, Gy. 1987. *Guide to conduct nationwide fine-scale soil mapping. Melioráció-öntözés és talajvédelem.* Budapest, Agroinform.

Horváth, B., Izsó, I., Jassó, F., Király, L., Parászka, L. and Szabóné Kele, G. 1989. *Guide to conduct nationwide fine-scale soil mapping. Melioráció-öntözés és tápanyaggazdálkodás.* Budapest, Agroinform.

http://napok.georgikon.hu/cikkadatbazis-2012/cat_view/3-cikkadatbazis/4-2012/10-vii-szekcio-vizgazdalkodas

Hudson, B.D. 1990. Concepts of soil mapping and interpretation. *Soil Survey Horizons* 31. (3): 63–72.

ISRIC 1993. *Global and National Soils and Terrain Digital Databases (SOTER*). Procedures Manual. UNEP-ISSS-ISRIC-FAO. ISRIC. Wageningen, Netherlands. 115 p.

IUSS Working Group WRB 2007. World Reference Base for Soil Resources 2006, first update 2007. *World Soil Resources Reports* No. 103. Rome, FAO.

IUSS Working Group WRB. 2014. 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. Rome, FAO.

Jenny, H. 1941. *Factors of Soil Formation, A System of Quantitative Pedology.* New York, McGraw-Hill, 191 p.

Krasilnikov, P., Arnold, R. and Michéli, E. 2009. Soil Classification of Hungary. In *A handbook of soil terminology, correlation and classification.* Eds.:

Arnold, R., Shoba, S., Krasilnikov, P. and Marti, J.J.I. London, Earthscan, Sterling, VA.

Lagacherie, P., Robbez-Masson, J.M., Nguyen-The, N. and Barthes, J.P. 2001. Mapping of reference area representativity using a mathematical soilscape distance. *Geoderma* 101. 105–118.

Lilburne, L.R., Hewitt, A.E. and Webb, T.W. 2012. Soil and informatics science combine to develop S-map: A new generation soil information system for New Zealand. *Geoderma* 170. 232–238.

Mattyasovszky, J., Görög, L. and Stefanovits, P. 1953. *1:200 000 m.a. mezőgazdasági talajtérkép* (Agricultural soil map at a scale of 1:200 000). Budapest, Tervgazdasági Könyvkiadó.

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

McBratney, A.B., Hart, G.A. and McGarry, D. 1991. The use of region partitioning to improve the representation of geostatistically mapped soil attributes. *Journal of Soil Science* 42. 513–532.

McSweeney, K., Slater, B. K., Hammer, R.D., Bell, J.C., Gessler, P.E., Petersen, G.W. 1994. Towards a new framework for modelling the soil-landscape continuum. In *Factors of Soil Formation: A 50th Anniversary Retrospective.* Eds.: Amundson, R., Harden, J. and Singer, M. SSSA Special Publication 33. 127–145.

MÉM (1983–1988). *Magyarország agrotopográfiai térképe 1:100 000 m.a.* (Agrotopographic map of Hungary 1:100 000). Budapest, MÉM Országos Földügyi és Térképészeti Hivatal, 84 térképlap /mapsheet, 51 x 66 cm.

MÉM-NAK 1983. *Magyarország genetikai talajtérképe 1:200 000 m.a.* (Genetic soil map of Hungary 1:200,000. Budapest, Mezőgazdasági és Élelmezésügyi Minisztérium.

Mendonça-Santos, M.D.L., Santos, H.G., Dart, R.O. and Pares, J.G. 2008. Digital mapping of soil classes in Rio de Janeiro State, Brazil: data, modelling and prediction. In *Digital Soil Mapping with Limited Data.* Eds.: Hartemink, A.E., Mcbratney, A. and Mendonça-Santos, M.L. Dordrecht, Springer, 381–396.

Michéli, E., Fuchs, M., Hegymegi, P. and Stefanovits, P. 2006. Classification of the major soils of Hungary and their correlation with the World Reference Base for Soil Resources (WRB). *Agrokémia és Talajtan* 55. (1): 19–28.

Pásztor, L., Bakacsi, Zs., Laborczi, A. and Szabó, J. 2013. Downscaling of categorical soil maps with the aid of auxiliary spatial soil information and data mining methods. *Agrokémia és Talajtan* 62. 205–218.

Pecora, L.M., Carroll, T.L. and Heagy, J.F. 1995. Statistics for mathematical properties of maps between time series embeddings. *Physical Review E* 52. (4): 3420. p.

Pontius Jr, R.G. 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing* 68. (10): 1041–1050.

Schmidt, K., Behrens, T., Friedrich, K. and Scholten, T. 2010. A method to generate soilscapes from soil maps. *Journal of Plant Nutrition and Soil Science* 173. (2): 163–172.

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.

Sisák, I. and Benő, A. 2012. *Digital publication of the 1:200 000 scale agricultural soil map on the Georgikon Mapserver*. LIV. Georgikon Napok, Keszthely, 2012. okt. 11–12. Papers, 431–436.

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

Sun, X.L., Zhao, Y.G., Zhang, G.L., Wu, S.C., Man, Y.B. and Wong, M.H. 2011. Application of a digital soil mapping method in producing soil orders on mountain areas of Hong Kong based on legacy soil data. *Pedosphere* 21. (3): 339–350.

Várallyay, Gy., Szűcs, L., Murányi, A., Rajkai, K. and Zilahy, P. 1979. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100 000 méretarányú térképe I. (1:100,000 map of the properties determining soil productivity in Hungary I.) *Agrokémia és Talajtan* 28. 363–384.

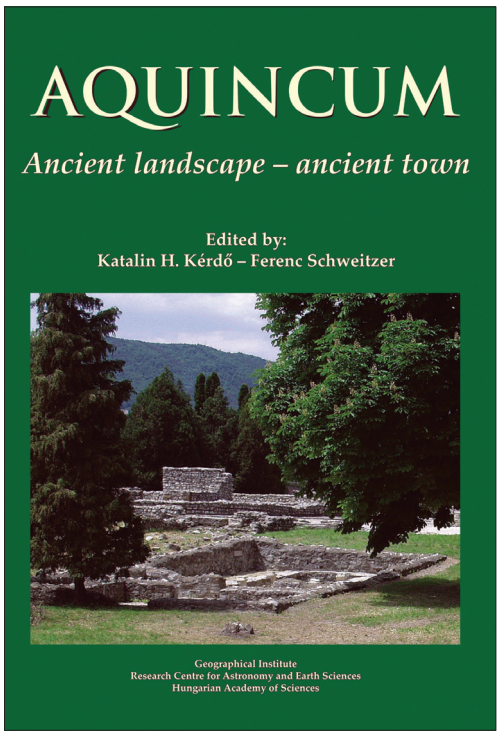
Várallyay, Gy., Szűcs, L., Murányi, A., Rajkai, K. and Zilahy, P. 1980. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100 000 méretarányú térképe II. (1:100,000 map of the properties determining soil productivity in Hungary II.) *Agrokémia és Talajtan* 29. 35–76.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38. (2): 203–208.

LITERATURE

H. Kérdő, K. – Schweitzer, F. (eds): AQUINCUM Ancient landscape – ancient town. Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, Budapest, 2014, 188.p

Historical events, magnificent masterpieces of art, architecture and engineering as well as people linked with the Roman Empire have long fascinated us. Hollywood movies like Ben Hur, the Gladiator or Spartacus, just to mention a few, have helped keeping our fascination alive. No wonder that the Empire, its rich cultural and material heritage appearing all over Europe, and literature on the subject is still popular even after 2000 years.



One of the classical examples of our Roman heritage is found in the heart of the Hungarian capital: the remains of the former Roman town known as Aquincum. The open air museum of the Historical Museum of Budapest is one way to promote this unique heritage. Publications aimed to raise scientific and public awareness of the latest findings on the subject via putting events into a landscape perspective like the present volume are exceptional and as such warmly welcomed.

The present volume, courtesy of a tight collaboration between researchers of the Geographical Institute of the Research Center for Astronomy and Earth Sciences, Hungarian Academy of Sciences and the Historical Museum of Budapest, is a pioneering masterpiece in combining results of archeology, history with those of natural scientific, primarily landscape geomorphological analysis. This way the reader gets firsthand information on which factors and to what extent might have influenced the formation and development of the Roman settlement of Aquincum.

The volume is aimed to elucidate the living conditions and cultural characteristics of the people inhabiting Aquincum, while discussing their interactions with the natural endowments of the landscape touching upon aspects of skillful exploitation and planned transformation. The potential pull-factors influencing initial site choice and further expansion of the settlement are meticulously evaluated taking into account their impact on space and utility management after the establishment of the town. Several exciting questions are lengthily discussed ranging from how raw materials for construction and ornamentation were gained, why the location was chosen and how nature, including floods shaped the everyday lives of the local citizens.

The conundrum of the system of trenches found in several places along the Danube is another intriguing issue tackled. According to the interpretations

presented some of them correspond to artificial canals for sewage disposal and transportation into the nearby Danube. Others might have been channels linking artificial fish ponds around the city. One can instantly understand why the interface of the eastern foothill area of the Buda Hills and the floodplain of the Danube was chosen for settlement besides obvious defensive purposes and despite the potential threat of iterative flooding. The exuberance of cold and warm springs, rocks and clay used for construction and housewares production, well-defended harbors along the river branches and backswamps are just a few among the potential pull factors touched upon in this volume.

Of the nine chapters of this 188 page volume, the first three gives an overview of the inferred natural endowments of the site. Besides the presentation of the morphology, soils, inferred climatic endowments and natural vegetation an attempt to reconstruct initial and later artificially altered hydrography of the area is also made. Based on the complex evaluation of available information seven environmental subtypes are identified to make up the mosaic of the landscape under study ranging from the main channel of the Danube and its side branches through lower and higher floodplains, backswamp areas, terraces, piedmont surfaces as well as karst slopes and horsts.

The next chapters discuss the history and archeology encountered. The initially formed Civil Town occupied the higher terraces and the higher floodplain surrounded by waterlogged areas offering protection. However, with the development of the city and an increase in the population new areas had to be occupied requiring alteration of the hydrography by regulating streams to gain new space. The artificial transformation of the Aranyhegyi Stream into a branch of the Danube and the birth of the Kis (Little) Island hosting the later constructed governor's palace is also highlighted.

Finally, an overview is given on the utilization of natural resources as well as traces reflecting further human-induced environmental transformations. I especially enjoyed reading the chapter dealing with how nature took back the site after abandonment and how these records are preserved in the geology and archeology of the excavated areas.

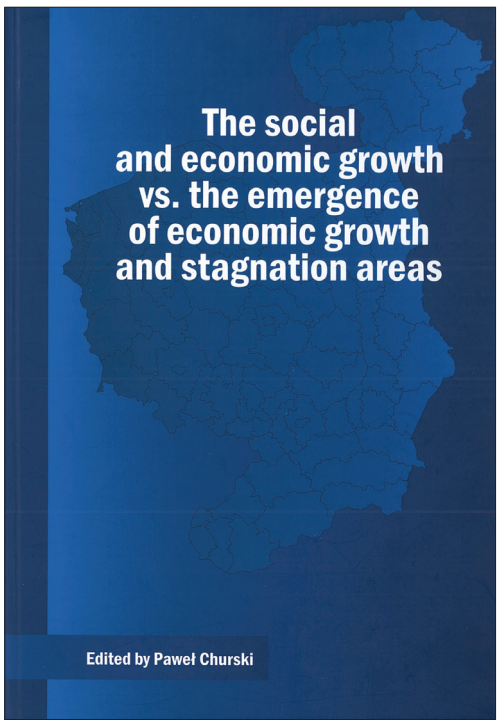
The text reads smoothly and enjoyable for even layman as well. High quality illustrations, excavation photographs, interpretive maps and drawings accompanying the text aid better visualization and easy understanding of the underlying hard science, even for readers outside of the fields of archeology or geography. I found the short list of technical terms at the back of the volume especially useful as a quick reference guide for the terminology. Also the data on excavated feature levels with reference to their photo location in the volume as an appendix is a treat for those looking for utilizing this data in further research. This book may equally count on the interest of professionals and general public alike.

Sándor Gulyás

Churski, P. ed.: The social and economic growth vs. the emergence of economic growth and stagnation areas. Bogucki Wydawnictwo Naukowe, Poznań, 2014, 197 p.

It hardly needs proof that the social and economic growth has been characterized by polarization. This phenomenon results in an emergence of the economic growth and stagnation areas all over East Central Europe. The distribution of these areas and its variability are affected by diverse growth factors which lead to convergence or divergence of the growth rate. This book containing nine chapters attempts to identify and interpret the different consequences of increasing growth diversity on the basis of international research project carried out between 2011 and 2014 in Poland, Lithuania and Slovakia.

The first chapter by Paweł Churski ('The polarization-diffusion model in changes to cohesion policy – consequences for the direction of growth policy') is of introductory and theoretical nature. Its goal is to establish the role of the polarization-diffusion model in the changes to the European Union's cohesion policy. At a time of frequently indicated lack of effectiveness of the compensation model, the polarization-diffusion model is viewed as the basis for the new paradigm in the member states' regional policies conditioning the changed direction of the growth policy intervention. The presented facts ands trends as well as a discussion conducted by the author provide an important basis for conclusion drawn with respect to the results presented in the remaining chapters of this volume.



The main goal of the chapter developed by Anna Borowczak and Michał Dolata ('Distribution of the economic growth and stagnation areas in Poland in 2000–2010') was an analysis of the distribution of the economic growth and stagnation areas and its variability in time, carried out with respect to research into the extent and dynamics of social and economic growth in a total system approach. This distribution was identified for 379 Polish counties with respect to the timeframe of 2000–2010 where the synthetic indicator and data clustering were employed. The adopted spatial arrangements makes it possible to view the polarization processes in the context of the distribution of areas with diverse conditioning of the developmental processes which stem from the contemporary globalization processes as well as the remaining relic divisions of Poland's economic space. The conclusions drawn in this article have been used in the subsequent two chapter to form a content-related entirety.

The article by Robert Perdał and Jan Hauke ('Areas of the economic growth and stagnation in Poland – growth factors') focuses on the issue of identifying and analysing important factors determining the distribution and development of economic growth and stagnation areas in a total system and partial approach with respect to the selected aspects of the process. The procedure of identifying growth factors was based on a canonical correlation analysis and regressive modelling. The research procedure was conducted for all the counties as well as their specific sub-systems. This allows to identify the regularities related to the approach of the territorially-oriented growth policy.

The goal of the chapter prepared by Joanna Dominiak and Barbara Konecka-Szydłowska ('The effect of the crisis on the socio-economic situation of households based on the example of Wielkopolska') is an attempt at identifying the impact of the crisis on the social and economic condition of households. The research was illustrated with the example of Wielkopolska province broken down to the social and economic growth and stagnation areas identified in this region at NUTS 4, on the basis of an analysis of

Poland's total developmental variations. 1,988 representative households were surveyed with respect to changes to different factors (e.g. the unemployment rate among the household members, the income size and structure, running expenses borne by the households, household depths, the ways of spending free time etc.) in the time of crisis. Again, attention was attracted to the regularities resulting from the specificity of the endogenous resources and exogenous conditioning in the economic growth and stagnation areas in a search for the related commonalities and differences.

The series of chapters referring to Poland was closed a paper written by Joanna Dominiak ('The impact of the economic crisis on the business environment service market, based on the example of Wielkopolska'). It includes an analysis of the operations of the business environment service market in Wielkopolska province in the conditions of economic downturn in Poland after 2008. The research was based on the field work in the province and was concerned on 32 business environment institutions. 220 companies rendering commercial business services and 737 entrepreneurs Due to the fact that they are located mainly in large cities in the course of analysis of the social and economic growth, they were included into the growth areas. The questionnaire survey was carried out in the formerly identified growth areas (Poznań, Konin, Kalisz, Leszno and Poznań county). An analysis of the demand part of the business environment takes into consideration the identified stagnation areas also (Słupca, Kalisz and Pleszew counties). On e one hand the research was aimed at analysing the changes to the range of impact of the business environment institutions and the changes to the scope of services rendered by them and their financial condition in the time of crisis. On the other hand the author identifies the changes to the intensity and scope of relations between companies and business environment institutions during the economic crisis.

The second part of the book starts with the article of Pavol Korec ('Lagging regions of Slovakia in the context of their competitiveness'). Its goal is to provoke a discussion on selected issues of the theory and methodology and regional research into competitiveness and the empirical results of the analysis of the competitiveness of Slovak regions with special emphasis placed on the country's less developed regions.

The next chapter written by Gintarė Pociūtė focuses on one of Poland's neighbouring countries ('Trends of imbalances of demographic and socio-economic development in the post-reform period in Lithuania'). The goal of this article was to identify the most significant trends in the irregular social and economic growth in Lithuania in the past two decades. To this end the author analysed the developmental deviation from the average for selected demographic and socio-economic variables. She also identifies lagging regions which are peripheral ones at all.

A long-titled chapter authored by Paweł Churski, Donatas Burneika and Pavol Korec ('Areas of economic growth and areas of stagnation as objects of special intervention under regional policies of the European Union member states: An international comparison') was an attempt for a comparison research for Poland, Lithuania and Slovakia. The goal of this analysis was to identify the developmental differences in the economic growth and stagnation areas and the direction of the regional policy intervention in these regions in these three new EU member states. The research results offer a considerable cognitive benefit resulting from the standardised systematization of the surveyed issues for these countries, overcoming some of the objective limitations related to non-existent comparable data on the subject.

The last chapter of the book called 'Research methodology of spatial variability of socio-economic development at the sub-regional level' developed by four authors, Dovile Krupickaite, Jan Hauke, Barbara Konecka-Szydłowska and Robert Perdał combines two aspects of research. The first one is an analysis of the diversity of social and economic growth in the population and settlement aspect in a supra-national dimension. The other aspect consisted in testing three different methods of identifying the growth and stagnation areas which leads to interesting cognitive conclusions of methodological nature.

A great number of carefully edited coloured and black-and-white maps make more chapters very attractive while demonstrate the distribution of both the growth areas and areas of economic stagnation on NUTS4 level between 2000 and 2010 in Poland, Slovakia and Lithuania.

This comprehensive book is highly recommended for geographers, economist, sociologists and politicians as well as wider audience interested in the problems of regional and economic inequalities of nowadays' East Central Europe.

Tibor Tiner