

## **A talajok szervesanyag-készletének nagyléptékű térképezése regresszió krigeléssel Zala megye példáján**

<sup>1</sup>SZATMÁRI Gábor, <sup>2</sup>LABORCZI Annamária, <sup>3</sup>ILLÉS Gábor és <sup>2</sup>PÁSZTOR László

<sup>1</sup>Szegedi Tudományegyetem Természeti Földrajzi és Geoinformatikai Tanszék, Szeged, <sup>2</sup>MTA Agrártudományi Kutatóközpont Talajtani és Agrokémiai Intézet, Budapest és <sup>3</sup>Erdészeti Tudományos Intézet, Budapest

### **Bevezetés**

A talajok állapotára, folyamataira, funkcióira vonatkozó információk iránti igények mind hazai, mind nemzetközi szinten számottevőek és folyamatosan bővülnek (BULLOCK, 1999; MERMUT & ESWARAN, 2000; TÓTH et al., 2008; SANCHEZ et al., 2009; BAUMGARDNER, 2011). Magyarországon jelenleg a talajokkal kapcsolatos információigényeket a rendelkezésünkre álló adatbázisok segítségével elégítjük ki, melyek döntő részét korábbi talajfelvételezésekből nyert tematikus térképek, illetve feltárt talajszelvények helyszíni és laboratóriumi jegyzőkönyveinek digitalizált térbeli adatbázisai jelentik. Az aktuálisan rendelkezésünkre álló, illetve a felhasználók által specifikusan megkívánt információk azonban nem feltétlenül, sőt egyre ritkábban fedik egymást (PÁSZTOR et al., 2013a). A korábbi kiterjedt adatgyűjtés, felvételezés, térképezés célja, az annak alapján elvégzett munka, illetve az ezek eredményeképpen született adatok direkt módon nem feltétlenül alkalmazhatók egy adott talajtani információkat igénylő problémakör kapcsán (MONTANARELLA, 2010). Többek közt ilyen problémakör lehet az archív mintavételi pontokban mért talajtulajdonságok adataiból történő térkép szerkesztése, mely adatokból a felvételezések idején nem készültek tematikus térképek.

A legideálisabb megoldást természetesen a kérdéses terület(ek) újbóli felvételezése, térképezése jelentené, azonban újabb, nagyobb területekre kiterjedő, hagyományos térképezési munkákra a közeljövőben nemigen látszik lehetőség ma Magyarországon (SZABÓ et al., 2005). Ezért a fenti problémára a digitális talajtérképezés nyújthat idő- és költségtakarékos megoldást. A földrajzi információs rendszerek (GIS), a Globális Helymeghatározó Rendszer (GPS), a távérzékelés, a geostatistika fejlődésének, illetve az olcsó, egyre nagyobb mennyiségben és nem utolsósorban növekvő tér- és időbeli felbontásban egyre könnyebben hozzáférhető, ún. környezeti segédváltozóknak köszönhetően számos új perspektíva mutatkozik a talajtulajdonságok digitális térképezése terén (MCBRATNEY et al., 2003; DOBOS et al., 2006). A regresszió krigelés tipikusan olyan térbeli becslő eljárás, mely a térképe-

---

*Postai cím:* SZATMÁRI GÁBOR, Szegedi Tudományegyetem, Természeti Földrajzi és Geoinformatikai Tanszék, 6720 Szeged, Egyetem u. 2.  
*E-mail:* szatmari.gabor.88@gmail.com

zendő talajtulajdonság pontszerű kvantitatív adatai, illetve a környezeti segédváltozók figyelembe vétele mellett ad becslést az adott változó térbeli eloszlásáról (ODEH et al., 1995; HENGL, 2009).

Jelen munkánkban Zala megye digitális szervesanyag-tartalom térképét kívántuk elkészíteni a rendelkezésünkre álló Digitális Kreybig Talajinformációs Rendszer (DKTIR) adataira, illetve környezeti segédváltozókra alapozva. A feltalaj humusz-tartalmának térbeli becslését regresszió krigeléssel végeztük el, mely az utóbbi években egyre népszerűbbé vált hazánk digitális talajtérképezési gyakorlatában (DOBOS et al., 2007; ILLÉS et al., 2011; SZATMÁRI, 2013; SZATMÁRI & BARTA, 2013). Dolgozatunkban változó számú és változó talajképződési tényezőkhöz kapcsolódó környezeti segédváltozót használtunk fel. Munkánk fő célja, hogy a regresszió krigelés modelljébe vont segédváltozó kombinációk minőségi hatását vizsgáljuk a becslő eljárás alapját jelentő többszörös lineáris regresszió modellre, illetve a becsült térkép pontosságára vonatkozóan. Továbbá célunk olyan szempontok megfogalmazása, melyek segítik a regresszió krigelésen alapuló térbeli becslések során a potenciális környezeti segédváltozók kiválasztását. A levont következtetéseket az országos digitális talajtérképezési projektben (DOSoReMI.hu; PÁSZTOR et al., 2013b) kívánjuk hasznosítani.

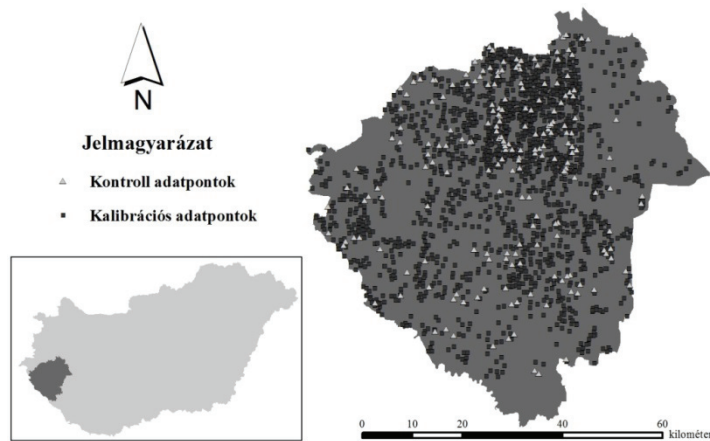
Jelen dolgozatunkban nem térünk ki részletesen a regresszió krigelés elméleti koncepcióira, illetve a becslési modell lépésről lépésre történő felépítésére; erről magyarul, részletesen SZATMÁRI és BARTA (2013) munkájában olvashatunk.

## Vizsgálati anyag és módszer

### *A DKTIR szelvények szervesanyag-tartalom adatai Zala megyére*

Zala megye (területe: 3.784 km<sup>2</sup>) változatos természetföldrajzi adottságokkal rendelkezik, melynek kialakításában hosszú időn át tartó komplex geológiai folyamatok játszottak közre. A talajképző kőzet szénsavas meszet nem tartalmazó periglaciális vályog, illetve pannon agyagos üledék, melyeket kelet felé fokozatosan vékony lösztakaró vált fel. A megye területét É-D-i irányú széles völgyek, illetve azokra közel merőleges elhelyezkedésű kisebb völgyhálózatok differenciálják. A megye éghajlatát alapvetően meghatározza az Alpok közelsége. A viszonylag hűvös és csapadékos klíma, illetve a hullámos felszín következtében a megye nagyon gazdag vízfolyásokban. A megye nyugati és délnyugati területeit pszeudoglejes barna erdőtalajok borítják, melyek a relatíve sok csapadék, az alacsony evaporáció, illetve a savanyú talajképző kőzet hatására alakultak ki. Kelet felé fokozatosan felváltják őket az agyagbemosódásos barna erdőtalajok, melyek mellett Ramann-féle barna erdőtalajokat is találunk. A széles völgytalpakban a réti talajok mellett megjelennek a tözezes láptalajok; a Kis-Balaton nem vízzel fedett területein síkláp talajokat találunk (STEFANOVITS et al., 1999; DÖVÉNYI, 2010).

A Zala megye feltalajára (0–30 cm) vonatkozó szervesanyag-tartalom adatokat a Digitális Kreybig Talajinformációs Rendszer (DKTIR) talajszelvény adataiból válogattuk le. Összesen csaknem kétezer darab adatpont állt a rendelkezésünkre,



1. ábra

A Digitális Kreybig Talajinformációs Rendszer (DKTIR) talajszelvényeinek elhelyezkedése Zala megye területén

amelyekben a humusztartalom tömegszázalékban van kifejezve. A Kreybig-féle Átnézetes Talajismereti Térképsorozathoz feltárt talajszelvények térbeli elhelyezkedését az 1. ábra szemlélteti. A szelvények eloszlása alapján elmondható, hogy a zalai mintaterület déli, illetve északkeleti területein szemmel láthatóan alacsonyabb sűrűséggel találunk talajszelvényeket, melynek oka az ezen területeket borító kiterjedt erdővel magyarázható. A kiindulási talajszelvény adatsort két részre osztottuk: egyrészt kialakítottunk egy kalibrációs adatponti halmazt (1. ábra), mely a konkrét térbeli becslések során lett felhasználva (az összes pontszerű adat 80%-a), illetőleg egy véletlenszerűen leválogatott kontroll adatponti halmazt (1. ábra), mely mindvégig független maradt a térbeli becslési eljárástól (az összes pontszerű adat 20%-a). Utóbbival, mint független mérési adathalmazzal kívántuk értékelni a térbeli becsléseinket. A kontroll adatpontok kvázi az egyes részterületekre jellemző sűrűséggel fedik le a mintaterületet.

A kalibrációs adatsoron feltáró alapadat-elemzést végeztünk. Boxplot segítségével eltávolítottuk a kiugró, illetve extrém értékeket, a valószínűség-eloszlás normalitását analitikus módszerrel (Kolmogorov-Smirnov próba Lilliefors-féle kiterjesztéssel) vizsgáltuk. A próba eredménye szerint a humusztartalom valószínűségi eloszlása nem modellezhető normál eloszlással, ezért logaritmikus adattranszformációt (HENGL, 2009) végeztünk. A transzformációt követően az adatsor normál eloszlást mutatott.

#### *A környezeti segédváltozók*

A szervesanyag-tartalom térbeli becsléséhez a szakirodalomban leggyakrabban használt környezeti segédváltozókat alkalmaztuk (MCBRATNEY et al., 2000, 2003; HENGL, 2009). Felhasználtuk Zala megye digitális domborzatmodelljét, melyből a

következő elsődleges, illetve másodlagos paramétereket származtattuk: tengerszint feletti magasság, lejtőhajlás, sík- és profilgörbület, összegyülekezési raszter, topográfiai nedvesség index, LS faktor (WISCHMEIER & SMITH, 1978), ún. „mass balance index”, vízfolyás energia index, topográfiai pozíció index. Felhasználásra került 18 darab, a 2009 és 2011 közötti időszakban készült MODIS (Moderate Resolution Imaging Spectroradiometer) műholdképekből származtatott NDVI (Normalized Difference Vegetation Index), illetve EVI (Enhanced Vegetation Index) vegetációs index állomány, melyek geometriai felbontása 250 méter. A 18 db index állománnyal az említett időszak – változatos klíma feltételei mellett adódó – vegetáció fejlődési periódusait kívántuk jellemezni. Segédváltozóként további két klímamutatót (évi átlagos csapadékösszeget, illetve a nyári hónapok középhőmérséklete) is alkalmaztunk. A fenti segédadatok mindegyike hozzárendelhető egy-egy talajképződési tényezőhöz; a digitális domborzatmodell elsődleges és másodlagos derivátumai a domborzati tényezőhöz, a vegetációs index állományok a biológiai tényezőhöz, míg a klímamutatók az éghajlati tényezőhöz rendelhetők.

A fentiek mellett segédadatként használtuk fel a DKTIR talajtérképi egységeit, név szerint a talajok textúra és vízgazdálkodási tulajdonságait (DKTIR-F), kémiai tulajdonságait (DKTIR-K) és tájtermesztési besorolását (DKTIR-T). Vizsgálataink során a talajképződési tényezők, illetve a DKTIR talajtérképi egységeinek különböző kombinációinak a hatását vizsgáltuk a többszörös lineáris regresszióra, illetve a térbeli kiterjesztés pontosságára vonatkozóan (1. táblázat). A kialakított segédadat kombinációk regresszió analízise során az egyes talajképző tényezőkhöz rendelhető összes segédadatot felhasználtuk. Az egyes kombinációk kialakításakor kiemelt

1. táblázat

A felhasznált környezeti segédváltozók kombinációi az egyes térbeli becslési modellekben

(1) Modellek	(2) Környezeti segédváltozók					
	(3) Domborzati tényező	(4) Biológiai tényező	(5) Éghajlati tényező	(6) Talajtani információk		
				DKTIR-F	DKTIR-K	DKTIR-T
A	X	–	–	–	–	–
B	–	X	X	–	–	–
C	X	X	X	–	–	–
D	X	X	X	–	X	–
E	X	X	X	X	–	–
F	X	X	X	–	–	X
G	X	X	X	X	X	–
H	X	–	X	–	–	–
I	X	–	X	X	–	–
J	X	–	X	X	X	–
K	X	–	X	–	X	–
L	X	–	X	–	–	X

Megjegyzés: DKTIR-F: DKTIR talajok textúra és vízgazdálkodási tulajdonságai, DKTIR-K: talajok kémiai tulajdonságai, DKTIR-T: DKTIR talajok tájtermesztési besorolása

figyelmet fordítottunk a talajtani segédinformációkra, hogy segítségükkel megvizsgálhassuk azok konkrét hatását a térbeli becslések pontosságára.

A felhasznált környezeti változók térbeli felbontása eltérő volt (20 métertől 250 méterig), ezért az elemzésekhez egy közös cellaméretet kellett választani. Ennek értékét HENGL (2006) munkája alapján határoztuk meg. Az eredményül kapott 20 méteres felbontást alkalmaztuk a térbeli becsléshez. A vektoros állományokat az így definiált referencia rácsra vonatkoztatva alakítottuk raszterekké, a durvább felbontású képi állományokat pedig az erre illeszkedő 20 m-es rácsra alakítottuk át.

A regressziós vizsgálatokat megelőzően a kvantitatív típusú környezeti segédadatok fedvényeinek pixelértékeit 0–255 skálára transzformáltuk (HENGL, 2009). Az egyes segédváltozó kombinációkhoz (1. táblázat) tartozó kvantitatív változókon főkomponens analízist (PCA) végeztünk, mellyel célunk volt a multikollinearitás csökkentése a többszörös lineáris regresszió analízis során. A kvalitatív segédváltozókat (DKTIR talajtérképi egységeit) indikátor változókká alakítottuk. A regressziós vizsgálatok során független változóként a PCA-k eredményeként kapott főkomponenseket, illetve az indikátor változókat használtuk fel.

#### *A térbeli becslés regresszió krigeléssel*

A regresszió krigelés megoldása a következőképp írható fel (HENGL, 2009):

$$\hat{z}(s_0) = \mathbf{q}_0^T \cdot \boldsymbol{\beta}_{\text{GLS}} + \boldsymbol{\lambda}_0^T \cdot (\mathbf{z} - \mathbf{q} \cdot \boldsymbol{\beta}_{\text{GLS}}) \quad (1)$$

*ahol:*  $\hat{z}(s_0)$  regresszió krigeléssel adott becslés az  $s_0$  pontban;  $\mathbf{q}_0$  a független változók vektora az  $s_0$  pontban;  $\boldsymbol{\beta}_{\text{GLS}}$  a regressziós koeficiens vektora,  $\boldsymbol{\lambda}_0$  a krigelési súlyok vektora az  $s_0$  pontban;  $\mathbf{z}$  a mért szervesanyag-tartalom értékek vektora és  $\mathbf{q}$  mátrix tartalmazza az egyes független változók értékeit a mérési pontokban.

A becslési algoritmus alapját jelentő többszörös lineáris regresszió analízist a szervesanyag-tartalom kalibrációs adatsorán (mint függő változón), illetve az egyes segédadat kombinációkon (mint független változókon) végeztük el. Az analízisek során mindvégig 5%-os szignifikancia szintet alkalmaztunk. Az eredményül kapott regressziós modellekből minden esetben származtattuk a reziduumokat, melyeken feltáró variográfiai vizsgálatokat végeztünk. A kiszámított tapasztalati félvariogramokhoz elméleti félvariogram modelleket illesztettünk szemi-automatikus megoldással. A kapott variogram modellek alapján kerültek meghatározásra a mintaterület egyes pontjaihoz rendelhető krigelési súlyok vektorai. A regressziós modellek, illetve a reziduumokon alapuló krigelési becslő eljárás eredőjeként kaptuk meg Zala megyére vonatkozóan a humusztartalom térbeli becsléseit regresszió krigeléssel.

Vizsgálataink során a szervesanyag-tartalom és az egyes segédadat kombinációk közötti többszörös lineáris regresszió analízist, illetve a regresszió krigelésen alapuló térbeli becslést SAGA (System for Automated Geoscientific Analysis) GIS környezetben végeztük el. A regresszió modellekből származtatott reziduumok statisztikai elemzéséhez, illetve feltáró variográfiai vizsgálatához az R szoftvert használtuk (R DEVELOPMENT CORE TEAM, 2009).

### A térbeli becslés értékelése

A regresszió krigeléssel elkészített térbeli becsléseinket a DKTIR talajszelvényeiből előzetesen leválogatott, a becslési eljárástól független kontroll adatokkal értékeltük (1. ábra). A kontroll pontokban mért szervesanyag-tartalom értékeket összevetettük az ugyanazon pontokra becsült humusztartalom értékekkel. A becslések validációját a következő paraméterekre alapozva végeztük el:

$$ME = \frac{1}{n} \cdot \sum_{i=1}^n [\hat{z}(s_i) - z(s_i)] \quad (2)$$

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |\hat{z}(s_i) - z(s_i)| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n [\hat{z}(s_i) - z(s_i)]^2} \quad (4)$$

ahol:  $\hat{z}(s_i)$  és  $z(s_i)$  a becsült és a mért értékek az  $s_i$  kontroll pontban. Az ME (Mean Error) az átlagos hibát, az MAE (Mean Absolute Error) a hibák abszolút értékének az átlagát, míg az RMSE (Root Mean Square Error) az átlagos négyzetgyök hibát fejezi ki.

Az egyes térbeli becslések pontosságának (ahol is a pontosságot az RMSE számított értékén keresztül értelmezzük SIMBAHAN és munkatársai (2006) munkája alapján) a relatív növekedését az RI (Relative Improvement) paraméter segítségével számszerűsítettük (SIMBAHAN et al., 2006):

$$RI_i(\%) = 100 \cdot \left( \frac{RMSE_{\max} - RMSE_i}{RMSE_{\max}} \right) \quad (5)$$

ahol:  $RI_i(\%)$  az  $i$ -edik térbeli becslés pontosságának a relatív növekedése a viszonyítási alaphoz választott becsléshez képest,  $RMSE_{\max}$  és  $RMSE_i$  a viszonyítási alaphoz választott, illetve az  $i$ -edik térbeli kiterjesztés RMSE értéke. Vizsgálataink során a legmagasabb RMSE értékkel rendelkező becslést vettük a viszonyítási alaphoz, hogy a számított  $RI_i(\%)$  értéke minden esetben pozitív szám legyen (SIMBAHAN et al., 2006).

### Vizsgálati eredmények

A regresszió krigelésen alapuló térbeli becslések során az 1. táblázatban bemutatott segédadat kombinációkat használtuk fel. Vizsgálataink során külön hangsúlyt fektettünk arra, hogy az egyes talajképző tényezők és a rendelkezésünkre álló szervesanyag-tartalom adatok közötti kapcsolatot regresszió analízis segítségével jellemezzük a zalai mintaterületen. Ezek figyelembe vételével alakítottuk ki az „A”, „B”, „C” és „H” segédadat kombinációkat (1. táblázat). Az „A” modell a humusztartalom és a topográfiai tényező közötti kapcsolatot hivatott vizsgálni, a „B” modell a humusztartalom és a biológiai, illetve éghajlati tényezők közötti kapcsolatot

számszerűsíti. A „C” modell az előző két modell együttese, míg a „H” modell kialakítására azért volt szükség, hogy a biológiai tényezőhöz rendelt MODIS műholdképek térbeli becslésre gyakorolt hatását vizsgáljuk. A bemutatott modellek kizárólag kvantitatív segédváltozókat foglaltak magukba. A „D”, „E”, „F”, „G”, „I”, „J”, „K” és „L” modellek különböző kombinációkban kvalitatív segédadatokat is tartalmaznak, melyek a DKTIR talajtérképi egységeihez kapcsolódnak (1. táblázat).

2. táblázat

A térbeli becslések során felhasznált többszörös lineáris regresszió modellek és azok jellemzői

(1) Regresszió modellek	(2) Regresszió modellek jellemzői *				
	R <sup>2</sup> (%)	Std. Error	SSR	SSE	MSE
A	16,8	0,477	82,4	407,0	0,228
B	10,7	0,495	52,1	437,3	0,245
C	20,1	0,469	98,3	391,2	0,220
D	27,4	0,447	133,8	355,6	0,200
E	34,7	0,424	169,8	319,7	0,180
F	35,5	0,422	173,9	315,6	0,178
G	36,0	0,420	176,3	313,2	0,177
H	18,3	0,474	89,4	400,1	0,225
I	34,1	0,426	166,7	322,8	0,182
J	35,4	0,422	173,1	316,4	0,178
K	26,3	0,450	128,7	360,8	0,203
L	34,9	0,423	170,9	318,6	0,179

Megjegyzés: \* R<sup>2</sup>: determinációs koefficiens; Std. Error: standard hiba; SSR: regressziós négyzetösszeg; SSE: hibából eredő négyzetösszeg; MSE: átlagos hibanégyzet

A szervesanyag-tartalom és a különböző segédadat kombinációkra illesztett regressziós modellek jellemzőit a 2. táblázat foglalja össze. A legalacsonyabb R<sup>2</sup> értékkel a „B” regresszió modell rendelkezik, mely modellben a független változók a biológiai, illetve az éghajlati tényezők voltak. A „B” modell determinációs koefficiensét meghaladja az „A” modellé, ahol is az „A” független változói a domborzathoz, mint talajképző tényezőhöz kapcsolódtak. Fontos megjegyezni, hogy míg a „B” modell a szervesanyag-tartalom térbeli változékonyságának valamivel több mint 10%-át volt képes leírni, addig az „A” közel 16%-ot volt képes determinálni. Az összes kvantitatív segédadat figyelembe vétele mellett („C” regresszió modell) az illesztett függvény a humusztartalom változékonyságának az egyötödét írta le. A „H” modell, mely mellőzi a biológiai tényezőt, a teljes varianciának több mint 18%-át volt képes determinálni. A 2. táblázatban megfigyelhető, hogy a kvalitatív típusú, talajtani segédinformációk bevonása esetén relatíve jelentősen emelkedtek a regresszió modellek determinációs koefficiens értékei. A legmagasabb R<sup>2</sup> értékkel a „G” modell rendelkezik, mely az összes általunk használt talajképződési tényezőt és a DKTIR két talajtérképi egységének (DKTIR-F, DKTIR-K) fedvényét használja

segédváltozóként. A „G” modell 36%-át írja le a humusz térbeli változékonyságának. Az utóbbi mellett relatíve magas  $R^2$  értékkel rendelkezik az „E”, „F”, „I”, „J” és „L” modell is. Az 1. és 2. táblázat alapján elmondható, hogy a legmagasabb determinációs koefficienssel rendelkező regresszió modellekben a DKTIR-F, illetve DKTIR-T segédváltozók szerepelnek, mint talajtani segédváltozók.

A szervesanyag-tartalom regresszió krigeléssel előállított térbeli kiterjesztéseiből térképeket szerkesztettünk. A térképek pixelértékeire, illetve a kalibrációs adatponti értékekre számolt összefoglaló statisztikát a 3. táblázat tartalmazza. A térké-

### 3. táblázat

A regresszió krigeléssel előállított térképek és a kalibrációs adatok (KA) összefoglaló statisztikája

(1) Térképek	(2) Statisztikai mutatók			
	Minimum	Maximum	(3) Átlag	(4) Szórás
A	0,7	22,8	2,51	1,393
B	0,8	16,1	2,50	1,329
C	0,7	21,0	2,56	1,400
D	0,8	19,1	2,57	1,531
E	0,7	21,6	2,67	1,616
F	0,7	34,8	2,73	1,925
G	0,7	21,8	2,62	1,635
H	0,7	22,6	2,51	1,404
I	0,7	20,4	2,63	1,626
J	0,7	20,2	2,61	1,632
K	0,7	19,4	2,58	1,560
L	0,7	32,4	2,63	1,918
KA	0,4	29,3	2,71	2,955

pek minimuma magasabb, mint a becslés alapját jelentő pontszerű adatok minimum értéke, míg a maximumok alacsonyabbak, mint a kalibrációs adatok esetén; utóbbi alól kivételt képez az „F” és „L” térkép, melynek oka a többszörös lineáris regresszió függvény extrapolációra való alkalmazása a magasabb humusztartalom értékek irányában (ez esetben fontos megjegyeznünk, hogy a regresszió függvények extrapolációra nem alkalmazhatók). A legalacsonyabb szórással a „B” jelű térkép rendelkezik, vagyis e térkép esetén volt a legnagyobb a krigelésekre jellemző simító hatás (smoothing effect) (GOOVAERTS, 1997). Relatív magas szórással rendelkeznek az „F”, „L”, „G”, „J”, „I” és „E” térképek, de az értékük alacsonyabb, mint a kalibráló adatokra számolt szórás.

A kontroll pontokban mért szervesanyag-tartalom értékek és az ugyanazon pontokra becsült humusztartalom értékek összevetéséből származtatott validációs paraméterek értékét a 4. táblázat mutatja be. A ME (Mean Error), MAE (Mean Absolute Error) és RMSE (Root Mean Square Error) szerint a legpontosabb becslést az „I” és a „J” térképek adták, mert a várt értéktől való eltérés ezen térképek esetén



volt a legkisebb. Az említett két térkép közül az „I” jelű rendelkezik a legkisebb MAE értékkel, így ezen térképet tekinthetjük a kontroll adatok alapján a legpontosabb térbeli becslésnek. Az „I” becslés esetén segédváltozó volt a topográfiai és az éghajlati tényező, illetve a DKTIR talajok textúra és vízgazdálkodási tulajdonságai, míg a „J” térkép esetén a fentiek mellett a DKTIR talajok kémiai tulajdonságaira vonatkozó fedvény is segédváltozóként szerepelt. A validációs mutatók alapján a legpontosabb térbeli kiterjesztést a „B” jelű térkép adta, mely a legmagasabb paraméterértékekkel rendelkezik. Ez indokolta, hogy az egyes térbeli becslések pontosságának a relatív növekedését a „B” térképhez viszonyítsuk. A „B” térképhez képest az „I” és a „J” becslések 21%-kal adtak pontosabb becslést, azokat követik az „E” és „G” jelű becslések, melyek 16%-kal pontosabb becslést adtak, mint a „B” térkép.

A validációs mutatók segítségével kapott sorrend is rávilágít arra, hogy az eróziót és akkumulációt befolyásoló domborzati adottságok mellett a talajtakaró fejlődésében az egyik legfontosabb szerepet a fizikai tulajdonságok kapják. A talaj fizikai félesége döntően befolyással van a talaj vízháztartására, a beszívargásra, a kilúgozás intenzitására, a nedvességtartalomra és ezeken keresztül a humuszképződés folyamatára. A talajkémiai tulajdonságok (pH, mészellátottság) segédváltozóként való bevonása („J”) már nem javította a becslés pontosságát, mely okát a terület viszonylagos homogenitásában kereshetjük. Ugyanakkor lényegesen nagyobb hibával terhelt térképeket kapunk, ha az „I” és a „J” modellek segédadatai mellett a

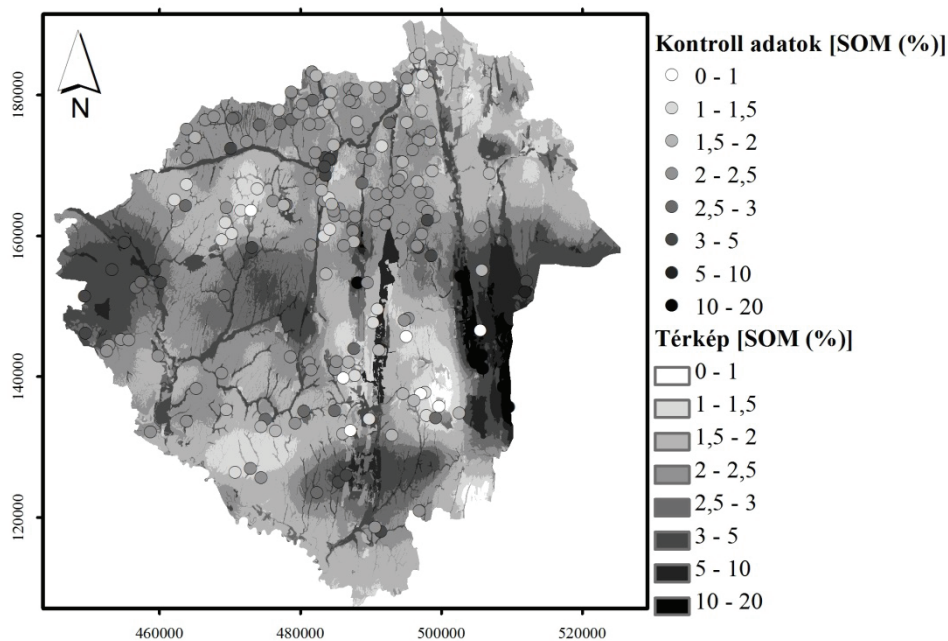
#### 4. táblázat

Regresszió krigeléssel előállított szervesanyag-tartalom térképek validációja a kontrollnak választott Digitális Kreybig Talajinformációs Rendszer szelvény adataival

(1) Térképek	(2) Validációs paraméterek			
	ME	MAE	RMSE	RI <sub>i</sub> (%)
A	-0,431	0,863	2,594	5
B	-0,457	0,883	2,727	–
C	-0,453	0,883	2,686	2
D	-0,406	0,883	2,599	5
E	-0,309	0,768	2,278	16
F	-0,424	0,852	2,585	5
G	-0,311	0,779	2,280	16
H	-0,432	0,865	2,589	5
I	-0,288	0,747	2,166	21
J	-0,286	0,754	2,166	21
K	-0,374	0,840	2,457	10
L	-0,397	0,827	2,473	9

*Megjegyzés:* ME (Mean Error): átlagos hiba; MAE (Mean Absolute Error): hibák abszolút értékének átlaga; RMSE (Root Mean Square Error): átlagos négyzetgyök hiba; RI<sub>i</sub> (%): Relative Improvement): az i-edik térbeli becslés pontosságának a relatív növekedése a viszonyítási alapnak választott becsléshez képest

biológiai tényezőhöz rendelhető 250 m-es térbeli felbontású MODIS vegetációs index állományokat is bevonjuk a regresszió krigelésbe („E” és „G” jelű térkép). Itt két potenciális oka lehet a becslési hiba növekedésének: 1. a MODIS képek 250 m-es felbontásából származó nagyságrendi léptékváltás, illetve 2. a vegetációs index állományok és a szervesanyag-tartalom közötti „gyenge” statisztikai kapcsolatot. Az „I” jelű térképet, illetve a kontroll pontokban mért szervesanyag-tartalom értékeket a 2. ábra mutatja be. Az ábra alapján elmondható, hogy a térkép déli, illetve észak-



2. ábra

Regresszió krigeléssel előállított legpontosabb feltalaj szervesanyag-tartalom térkép Zala megye területére. *Megjegyzés:* SOM(%): szervesanyag-tartalom %-ban kifejezve

keleti területein kevésbé pontos becslést kaptunk, melynek oka, hogy ezen területek – mint korábban említettük – a mintavételezések szempontjából alulreprezentáltak voltak. Az alulreprezentáltság negatív hatással van a térbeli kiterjesztések pontosságára, hisz regresszió krigelés esetén „elvárjuk”, hogy a mintavételi pontok egységesen fedjék le nemcsak a kutatási terület, hanem az általa prezentált tulajdonság teret is (HENGL, 2009). Utóbbi miatt figyelhető meg, hogy a lápos területek bizonyos részein alulbecsültük, míg az erdővel borított, magasabb térszínű területeken felülbecsültük a humusztartalmat. A tulajdonság tér közbenső részén, ahol is kellő számú minta állt rendelkezésre a regresszió függvény illesztéséhez a becslésünk jól közelítette a szervesanyag-tartalmat.

### Következtetések

A bemutatott vizsgálati eredmények alapján a következő konklúziókat vonhatjuk le, mely tapasztalatokat hosszú távon kívánjuk alkalmazni, átültetve azokat az országos digitális talajtérképezési projektbe (DOSoReMI.hu).

A talajtani segédinformációk felhasználása jelentősen emelte a regresszió krigeléssel, mint térbeli becslő eljárással származtatott térképek pontosságát. Ugyanakkor megjegyzendő, hogy egyszerre több talajtani információ modellbe építése nem feltétlen vonja maga után az elkészült becslés pontosságának a növekedését, ezért törekedni kell a releváns talajtani segédváltozók kiválasztására.

Zala megye példáján megfigyelhető volt, hogy a MODIS műholdképekből származtatott vegetációs index állományok modellbe vonása során csökkent a térképek pontossága, melynek két oka lehet: 1. az index állományok jelentősen alacsonyabb geometriai felbontása (250 méter), illetve 2. a szervesanyag-tartalom és az állományok közötti „gyenge” statisztikai kapcsolat.

A fentiek alapján általánosan megfogalmazható, hogy regresszió krigelésen alapuló térbeli becslés esetén a térképezendő (talajtani) célváltozóhoz rendelendő potenciális környezeti segédváltozók kiválasztása során tekintettel kell lenni a célváltozó térbeli változékonyságára (feltáró variográfiai vizsgálatokon keresztül), illetve az azt kialakító háttér folyamat(ok)ra, a talajképződési tényezőkre, illetve a releváns térbeli talajinformációkra és nem utolsósorban a térképezés léptékére.

Megjegyzendő, hogy a ME és a MAE kvázi azonos sorrendet állított fel a validáció során az elkészült térképek között, azonban számítási mechanizmusából kifolyólag a MAE szigorúbban vizsgálja a várt értéktől való eltérést, ezért inkonzisztens eredmények esetén célszerűnek látjuk a MAE értékét előnyben részesíteni.

### Összefoglalás

Dolgozatunkban Zala megye feltalajainak szervesanyag-tartalmát kívántuk digitálisan térképezni regresszió krigeléssel a rendelkezésünkre álló Digitális Kreybig Talajinformációs Rendszer (DKTIR) adataira, illetve környezeti segédváltozókra alapozva. A térbeli kiterjesztések során különböző kombinációkban használtuk fel a talajképződési tényezőket, illetve DKTIR talajtérképi egységeit. Munkánk célja volt, hogy a regresszió krigelés modelljébe vont segédváltozó kombinációk minőségi hatását vizsgáljuk a becslő eljárás alapját jelentő többszörös lineáris regresszió modellre, illetve a becslés pontosságára vonatkozóan.

A szervesanyag-tartalom térbeli kiterjesztéséhez szükséges segédváltozókat a szakirodalom alapján választottuk ki. Segédadatként használtuk fel Zala megye digitális domborzatmodelljét, a 2009 és 2011 között készült MODIS műholdképekből származtatott vegetációs index állományokat, két klímaméter fedvényét, illetve a DKTIR talajtérképi egységeit. A regresszió krigeléssel becslés humusztartalom térképeket a DKTIR talajszelvény adataiból előzetesen leválogatott, a becslési eljárástól független kontroll adatpontokkal értékeltük. A validációhoz származtatuk a ME (Mean Error), a MAE (Mean Absolute Error), az RMSE (Root Mean

Square Error), illetve az  $RI_i(\%)$  (Relative Improvement) paraméterek értékeit, ahol utóbbi az egyes térképek pontosságának relatív növekedését fejezi ki egy viszonyítási alaphoz választott térképhez képest.

A vizsgálati eredmények alapján a térbeli talajinformációk segédadatként történő felhasználása jelentősen növelte a regresszió modellek determinációs koefficiensének értékét, illetve a becsült humusztérképek pontosságát. A talajtani segédinformációkat is figyelembe vevő regresszió modellek  $R^2$  értékei – két eset kivételével – jóval meghaladták a 30%-ot, vagyis a szervesanyag-tartalom térbeli változékonyságának több mint egyharmadát voltak képesek determinálni. A validációs mutatók alapján azon térképek bizonyultak pontosabbnak, melyekben a DKTIR talajok textúra és vízgazdálkodási tulajdonságait (DKTIR-F) használtuk fel talajtani segédváltozóként. A legalacsonyabb MAE értékkel (0,747) a domborzati és éghajlati talajképző tényezőket, illetve a DKTIR-F talajtérképi egységet segédváltozóként alkalmazó humusztérkép rendelkezett, ezen térkép esetén az  $RI_i(\%)$  paraméter értéke 21%-nak adódott. A mutatók alapján ezen térkép adta a legpontosabb becslést a mintaterület szervesanyag-tartalmára, hisz a felhasznált segédváltozókon keresztül figyelembe veszi a mintaterület szervesanyag-tartalmát alapvetően befolyásoló eróziót és akkumulációt, illetve a talajok fizikai féleségét, mely utóbbi hatással van a vízháztartásra, a beszivárgásra, a kilúgozásra és ezeken keresztül a humuszképződés folyamatára. A biológiai talajképző tényezőt reprezentáló MODIS vegetációs index állományok esetében megfigyelhető volt, hogy segédadatként történő alkalmazásuk esetén kevésbé pontos becsléseket kaptunk összevetve az ezen segédadatokat mellőző becslésekkel.

Munkánkat a K105167 számú OTKA, illetve a TÁMOP-4.2.2.A-11/1/KONV-2012-0013. pályázatok támogatják.

**Kulcsszavak:** szervesanyag-tartalom, regresszió krigeles, környezeti segédváltozó, digitális talajtérképezés, validáció

### Irodalom

- BAUMGARDNER, M. F., 2011. Soil databases. In: Handbook of Soil Sciences: Resource Management and Environmental Impacts (Eds.: HUANG, P. M., LI, Y. & SUMNER, M. E.) 21–35. CRC Press. Boca Raton.
- BULLOCK, P., 1999. Soil resources of Europe – An overview. In: Soil Resources of Europe (Eds.: BULLOCK, P., JONES, R. J. A. & MONTANARELLA, L.) European Soil Bureau Research Report 6. 15–25. Office for Official Publications of the European Communities. Luxemburg.
- DOBOS, E., MICHÉLI, E. & 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. & VOLTZ, M.) 487–495. Elsevier B.V. Amsterdam.

- DOBOS, E. et al., (Eds.), 2006. Digital Soil Mapping as a Support to Production of Functional Maps. EUR 22123 EN. Office for Official Publications of the European Communities. Luxemburg.
- DÖVÉNYI Z. (szerk.), 2010. Magyarország kistájainak katasztere. MTA Földrajztudományi Kutatóintézet. Budapest.
- GOOVAERTS, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press. New York.
- HENGL, T., 2006. Finding the right pixel size. *Computers & Geosciences*. **32**. 1283–1298.
- HENGL, T., 2009. A Practical Guide to Geostatistical Mapping. University of Amsterdam. Amsterdam.
- ILLÉS G., KOVÁCS G. & HEIL B., 2011. Nagyfelbontású digitális talajtérképezés a Vas-kereszt erdőrezervátumban. *Erdészettudományi Közlemények*. **1**. 29–43.
- MCBRATNEY, A. B., MENDONÇA SANTOS, M. L. & MINASNY, B., 2003. On digital soil mapping. *Geoderma*. **117**. 3–52.
- MCBRATNEY, A. B. et al., 2000. An overview of pedimetric techniques for use in soil survey. *Geoderma*. **97**. 293–327.
- MERMUT, A. R. & ESWARAN, H. S., 2000. Some major developments in soil science since the mid-1960s. *Geoderma*. **100**. 403–426.
- MONTANARELLA, L., 2010. Need for interpreted soil information for policy making. In: 19<sup>th</sup> World Congress of Soil Science, Soil Solutions for a Changing World. Brisbane, Australia, 1–6 August 2010. (DVD)
- ODEH, I. O. A., MCBRATNEY, A. B. & CHITTLEBOROUGH, D. J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*. **67**. 215–226.
- PÁSZTOR, L. et al., 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**. 13–27.
- PÁSZTOR, L. et al., 2013b. Elaboration of novel, countrywide maps for the satisfaction of recent demands on spatial, soil related information in Hungary. In: *Global Soil Map*. (Ed.: ARROUAYS, D.) CRC Press/Balkema (megjelenés alatt).
- R DEVELOPMENT CORE TEAM, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. (<http://www.R-project.org>). Vienna, Austria.
- SANCHEZ, P. A. et al., 2009. Digital soil map of the world. *Science*. **325**. 680–681.
- SIMBAHAN, G. C. et al., 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. *Geoderma*. **132**. 471–489.
- STEFANOVITS P., FILEP GY. & FÜLEKY GY., 1999. *Talajtan*. Mezőgazda Kiadó. Budapest.
- SZABÓ J., PÁSZTOR L. & BAKACSI ZS., 2005. Egy országos, átnézetes, térbeli talajinformációs rendszer kiépítésének igénye, lehetősége és lépései. *Agrokémia és Talajtan*. **54**. 41–58.
- SZATMÁRI, G., 2013. High-resolution mapping of soil organic matter content based on regression kriging in a study area endangered by erosion in Hungary. In: XVI. Congress of Hungarian Geomathematics and V. Congress of Croatian and Hungarian Geomathematics, Mórahalom, 30 May–1 June 2013. (CD)

- SZATMÁRI G. & BARTA K., 2013. Csernozjom talajok szervesanyag-tartalmának digitális térképezése erózióval veszélyeztetett mezőföldi területen. *Agrokémia és Talajtan*. **62**. 47–60.
- TÓTH, G. et al., 2008. *Soils of the European Union*. Office for Official Publications of the European Communities. Luxemburg.
- WISCHMEIER, W. H. & SMITH, D. D., 1978. *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*. Agriculture Handbook No. 537. US Government Printing Office. Washington, D. C.

*Érkezett: 2013. szeptember 24.*

## Large-scale mapping of soil organic matter content by regression kriging in Zala County

<sup>1</sup>G. SZATMÁRI, <sup>2</sup>A. LABORCZI, <sup>3</sup>G. ILLÉS and <sup>2</sup>L. PÁSZTOR

<sup>1</sup>Department of Geography and Geoinformatics, University of Szeged, Szeged

<sup>2</sup>Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest and

<sup>3</sup>Forest Research Institute, Budapest

### Summary

The organic matter content (SOM) of topsoil in Zala County was digitally mapped using regression kriging based on data from the Digital Kreybig Soil Information System (DKSIS) and environmental auxiliary variables. Various combinations of soil-forming factors and DKSIS mapping units were used during the spatial inference process. The objective was to analyse the effects of these combinations on multiple linear regression, which is the core of the applied prediction method, as well as on the accuracy of the estimated maps.

The auxiliary variables were selected with the help of state-of-the-art literature: the digital elevation model of Zala County; the NDVI and EVI layers of MODIS images from the period of 2009–2011; two climatic parameters; and the soil mapping units of DKSIS. The SOM maps estimated by regression kriging were evaluated using a control data set previously selected from the DKSIS profile database.

The ME (Mean Error), MAE (Mean Absolute Error), RMSE (Root Mean Square Error) and RI<sub>i</sub>(%) (Relative Improvement) parameters were computed for the validation, the last of which expresses the relative improvement in the accuracy of the resulting maps compared to a reference map.

The results showed that the inclusion of spatial soil data significantly increased the determination coefficients of the regression models, as well as the accuracy of the topsoil SOM maps. With two exceptions, the R<sup>2</sup> values of regression models that used the auxiliary information were significantly higher than 30%; i.e. they determined more than one-third of the spatial variability of topsoil SOM. Based on the validation indices, maps using DKSIS soil physical mapping units proved to be more accurate. The lowest MAE value (0.747) was obtained for the map where relief and climatic soil-forming factors were used together with DKSIS soil physical mapping units; the RI<sub>i</sub>(%) parameter for this map was 21%. The indices showed that this map provided the best prediction for the organic matter content of the area. This can be attributed to the fact that, due to the inclusion of auxiliary variables, it took into consideration not only the erosion and accumulation processes influencing SOM content, but also soil hydrophysical properties, which have a strong indirect effect on the process of humus formation. Predictions made using the biological soil-forming factor represented by MODIS vegetation index layers were worse than those which excluded them.

*Table 1.* Combinations of environmental auxiliary variables used in the spatial prediction models. (1) Models. (2) Environmental auxiliary variables. (3) Relief factor. (4) Biological factor. (5) Climatic factor. (6) Soil science information. DKTIR-F: tex-

ture and water management traits of DKTIR soils; DKTIR-K: chemical properties of the soils; DKTIR-T: so-called landscape management soil type (LMST).

*Table 2.* Multiple regression models used in the spatial predictions, and their characteristics. (1) Regression models. (2) Characteristics of the regression models. *Remarks:*  $R^2$ : coefficient of determination; Std. Error: standard error; SSR: regression sum of squares; SSE: sum of squared error; MSE: mean squared error.

*Table 3.* Summary statistics for the maps created by regression kriging and for the calibration data set (KA). (1) Maps. (2) Statistical indices. (3) Mean. (4) Deviation.

*Table 4.* Validation of topsoil SOM maps created by regression kriging using the control DKSIS data set. (1) Maps. (2) Validation parameters. *Remarks:* ME: Mean Error; MAE: Mean Absolute Error; RMSE: Root Mean Square Error;  $RI_i$  (%): Relative Improvement.

*Fig. 1.* Spatial distribution of the soil profiles in the Digital Kreybig Soil Information System (DKSIS) in Zala County. *Legend:* Control data points; calibration data points.

*Fig. 2.* The best topsoil SOM map created by regression kriging for the area of Zala County. *Remarks:* SOM (%): organic matter content as a %. *Legend:* Control data. Map.