

Kassai Piroska¹ – Sisák István¹

Az 1:100 000-es felszíngeológiai térkép, mint környezeti segédváltozó felhasználásának lehetőségei a Balaton vízgyűjtő digitális talajtérképezésében

The possible use of the 1:100 000 scale surface geology map as environmental covariate in the digital soil mapping of the watershed area of Lake Balaton

piroska.kassai@gmail.com

¹Pannon Egyetem Georgikon Kar, 8360 Keszthely, Deák Ferenc u. 16.

Összefoglalás

A talajképző kőzet, mint környezeti segédváltozó a talaj modellezési és térképezési projektek során kevésbé használt paraméter. Az ilyen irányú munkák csupán negyedében használnak geológiai információt, holott a talajképző kőzet egyike a 6 fontos talajképző tényezőnek. A meglévő talajtérképek a hegy- és dombvidékekről csak alig szolgáltatnak információt, míg a földtani térképek általában ezeken a területeken a legjobban kidolgozottak, így fontos információt nyújthatnak a digitális talajtérképezéshez. Jelen munka a Magyar Állami Földtani Intézet által kiadott 1:100 000-es felszíngeológiai térkép, mint környezeti segédváltozó felhasználásának lehetőségét vizsgálja a Balaton vízgyűjtő digitális talajtérképezésében. A földtani térképek jellemzően kategóriaváltozóként kerülnek felhasználásra az elemzések során, de a talajtani célú felhasználásához szükséges a geológiai egységek újraosztályozása, melyhez számos korábbi munka adhat iránymutatást. A legmodernebb szemlélet szerint az egyes kőzetek geokémiai tulajdonságait, azon belül főként SiO₂ tartalmukat javasolt figyelembe venni és szerint csoportosítani az egyes kőzeteket. Ennek megfelelően készítettük el a felszíngeológiai térkép újraosztályozott változatát, amely jelen munkában, mint talajképző kőzet térkép került felhasználásra talajtulajdonságok modellezésében. A modellezéshez az ún. véletlen erdő (random forest) módszert használtuk, amely során lehetőség nyílt az egyes független (magyarázó) változók modellben betöltött szerepének kiértékelésére is. Független változóként az Agrokémiai Információs és Irányítási Rendszer (AIIR adatbázis) mintavételi pontjain (12396 db) mért talajtulajdonságokat (pH, kötöttség) adtuk meg. Független változóként domborzati és klíma adatok kerültek felhasználásra a geológiai térkép mellett. Habár a vizsgált talajtulajdonságok esetében a geológia a legfontosabb független változók között szerepelt, csak kis mértékben tudta növelni az R² értékeket, ami feltehetően arra vezethető vissza, hogy a talaj mintavételezési pontok elsősorban a geológiailag sokkal homogénebb sík-és dombvidéki területre esnek.

1. Bevezetés

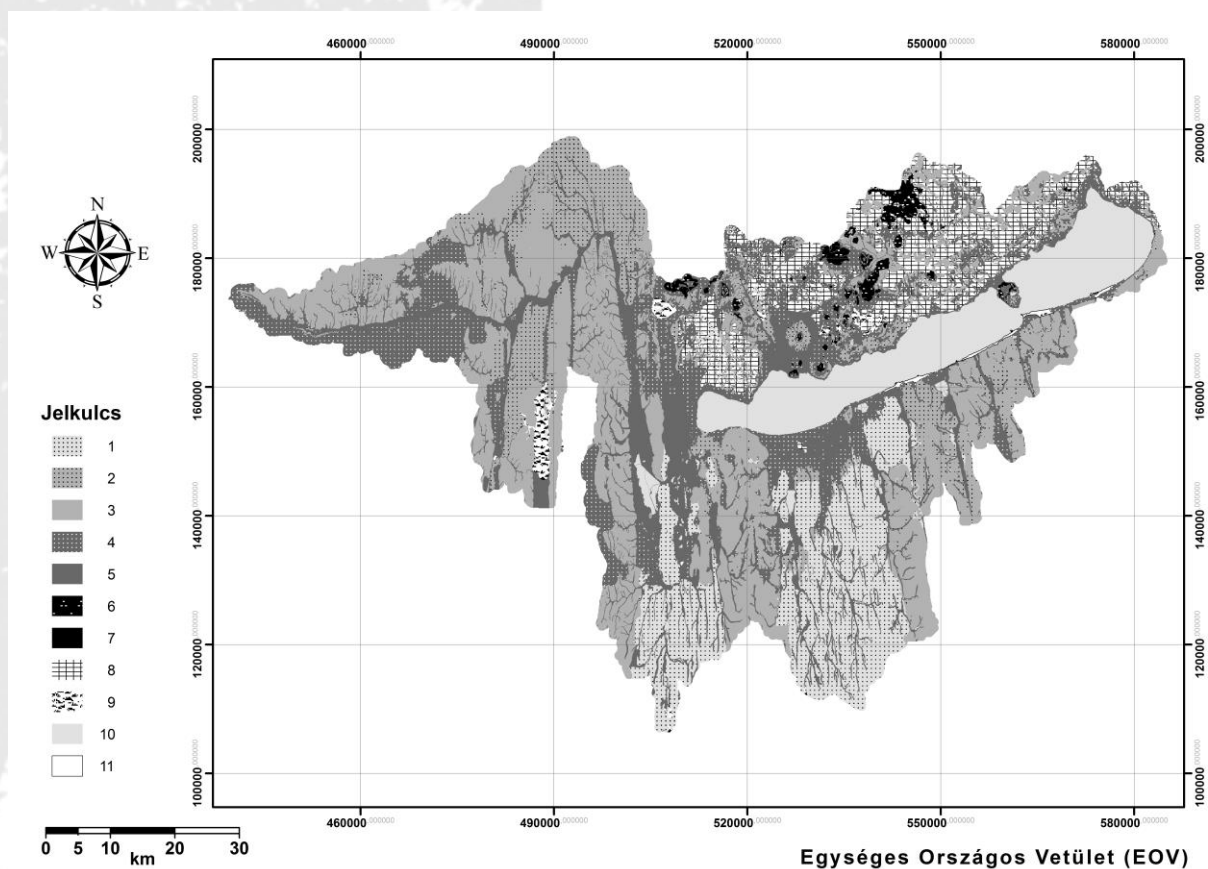
Az elmúlt néhány évtizedben a digitális információ és a gépi tanulás térhódításának köszönhetően a talajtérképezés újszerű megközelítést nyert. A digitális talajtérképezés elve, hogy a korábban szerzett talajinformációk és talajtérképek integrálva különböző környezeti segédváltozókkal a korábbiaknál részletesebb talajtérképeket eredményezhetnek. Jenny (1941) talajképződésre felállított alapegyenletéből kiindulva McBratney (2003) alkotta meg az ún. SCORPAN modellt, amely a digitális talajtérképezési projektek kiindulópontja. Eszerint:

S (talajparaméter vagy talajtípus) = $f(S, C, O, R, P, A, N)$,

ahol a függő változó (célváltozó) egy adott talajparaméter vagy talajtípus, a független (magyarázó) változók pedig ún. környezeti segédváltozók: C=klíma, O=biológiai tényező, R=domborzat, P=talajképző kőzet, A=idő tényező, N=földrajzi elhelyezkedés. A digitális talajtérképezés egyik legnagyobb előnye, hogy a modellek

kiterjesztésével releváns információ állítható elő olyan területekre is, ahol korábban egyáltalán nem készült talajtérkép. Ehhez csupán a meglévő információk digitalizálására és részletes elemzésére van szükség.

A Balaton vízgyűjtőjéről korábban nem készült részletes (1:10 000-es), teljes területet lefedő talajtérkép és a már meglévő országos léptékű talajtérképek (1:100 000-1:500 000) sem eléggé pontosak. Emellett a legtöbb talajtérkép csak a mezőgazdasági területekről szolgáltat információt, a hegy- és dombvidékek területe csak elnagyoltan szerepel a térképeken. Léteznek azonban olyan adatbázisok, amelyek segítségével információt gyűjthetünk azokról a területekről is, ahonnan kevesebb talajtani adattal rendelkezünk. A Magyar Állami Földtani Intézet (MÁFI) 1:100 000-es felszíngeológiai térképe részletgazdagságában sok helyen (különösen a hegy- és dombvidékeken) egyértelműen felülmúlja a nagyobb méretarányú talajtérképeket is. Ezt felhasználva a geológia és a talaj kapcsolatának számszerűsítésével egy igen értékes segédinformációhoz juthatunk a Balaton vízgyűjtő digitális talajtérképezéséhez. Egy korábbi munkában (Kassai et al., 2017) már bemutattuk, hogy a földtani térképen szereplő egységek geokémiai szempontú átcsoportosításával egy olyan talajképző kőzet térkép hozható létre, amely a korábbiaknál részletesebb és talajtani szempontból is releváns információt szolgáltat.



1. ábra: A Balaton vízgyűjtő talajképző kőzet térképe

Jelölések: 1: eolikus homokok, 2: folyóvízi homok-kavics üledékek, homokkő, 3: löszös képződmények, 4: aleuritok, 5: agyagos képződmények, 6: bazalt, 7: bauxitos agyag/agyagos bauxit, 8: karbonátos kőzetek, 9: szerves üledékek, 10: vízfelület, 11: antropogén képződmények

A térkép készítéséhez az AIIR adatbázis Balaton vízgyűjtő területén található 12396 pontján mért 8 féle talajparaméter (kötöttség, humusz, pH, CaCO₃, Mn, Mg, Zn, Cu) értékeit vetettünk össze a MÁFI 1:100 000-es felszíngeológiai térkép általunk létrehozott kategóriáival. Ehhez az eredeti földtani térkép 143 geológiai egységét

Gray és munkatársai (2016) módszere alapján a kőzetek SiO₂ tartalma szerint osztályoztuk. Kilenc litológiai kategóriát hoztunk létre és igazoltuk, hogy ezek a kategóriák a rajtuk kialakult talajok tulajdonságai szempontjából is lényegesen eltérnek egymástól. Megállapítottuk, hogy a kőzetek SiO₂ tartalma és számos talajtulajdonság között inverz lineáris kapcsolat van. Ezt a 9 kategóriából létrehozott talajképző kőzet térképet (1. ábra) használjuk fel jelen munkánkban, ahol egy komplex modellezés során egyéb környezeti segédváltozókat (domborzat, klíma) is elemezve megpróbáljuk kideríteni, hogy ténylegesen mekkora a geológia szerepe a talajtulajdonságok értékeinek alakulásában és mennyivel javítja a modellt az új geokémiai szemléletű földtani információ bevonása.

2. Anyag és módszerek

2.1. Statisztikai módszerek

A modellépítés során az ún. véletlen erdő (random forest) módszert alkalmaztuk. A módszer lényege, hogy előre meghatározott számú döntési fa előrejelzéseit kombináljuk, ahol mindegyik fa véletlen vektorok egy független halmazának értékei alapján kerül létrehozásra. A véletlen vektor létrehozásának módja, hogy a döntési fa minden csúcsának vágásához az összes rendelkezésre álló független változó közül csak előre meghatározott számú véletlenszerűen kiválasztott változókat vizsgálunk (Breiman, 2001). A modellépítés során lehetőség nyílik a vágáshoz használt változók optimális számának meghatározására. A módszer regresszióra és klasszifikációra egyaránt használható, attól függően, hogy függő változóként milyen típusú változót használunk. Esetünkben mindkét függő (célváltozó) folytonos változó volt, így két regressziós modellt hoztunk létre. A független változók típusára a modellben nincsen kikötés, így használhatunk folytonos és kategória típusú változókat is. Jelen esetben a független változók között 7 folytonos és 1 kategória típusú változó szerepelt.

Az elkészített modellek felhasználhatók további predikcióra, amely lehetővé teszi, hogy - a megadott környezeti segédváltozók ismeretében - olyan területek talajtani adatait becsüljük ahol korábban nem történt talajtani felmérés.

A becslést a munkánk során úgy teszteltük, hogy a 8 db független változóból és 1-1 célváltozóból álló adatbázist tréning és teszt adatsorra bontottuk. A tréning adatbázis az adatok véletlenszerűen kiválasztott 80%-ából hoztuk létre, a becslés tesztelésére pedig a maradék 20%-ot használtuk. A modellt a tréning adatbázisra építettük fel majd ennek segítségével becsültük a teszt adatbázis értékeit. Ezt követően a teszt adatbázis valós és becsült értékeinek kapcsolatát korrelációs számítással mértük.

Mivel a módszer lehetőséget nyújt az egyes változók modellben betöltött szerepének kiértékelésére is, meg tudtuk vizsgálni a geológia szerepét a két talajtulajdonság alakulásában. Ezt többféle megközelítésben vizsgáltuk: egyrészt a determinációs együttható (R²) értékének csökkenését másrészt a becslési hiba értékének növekedését vizsgáltuk a geológiai információ elhagyása esetén. Továbbá megvizsgáltuk a geológia szerepét a többi környezeti segédváltozó viszonyában is.

Az adatok térinformatikai előkészítését az ArcGIS 10.0., a statisztikai feldolgozást az R 3.4.1. segítségével végeztük el.

2.2. Célváltozók

Az elemzés során külön modelleket hoztunk létre a két különböző célváltozó, a talajok kötöttsége valamint pH értékeire nézve. A célváltozóként felhasznált talajadatok az Agrokémiai Információs és Irányítási Rendszer (AIIR) adatbázisának Balaton vízgyűjtő területére eső pontjain mért értékek voltak. Összesen 12396 adatpont értékeit elemeztük. Az adatbázisról Kocsis és munkatársai (2014) összefoglaló munkát készítettek, melyben a mérések körülményeiről részletes információ érhető el.

2.3. Környezeti segédváltozók

A modell építés során a geológia mellett klíma és domborzati adatokat használtunk. Mivel az elemzés egyik célja az egyes független (magyarázó) változók fontosságának megállapítása volt ezért törekedtünk arra, hogy csak egymással nem korreláló változókat használjunk. fel. Ha ugyanis egymással korreláló változókat használunk (pl. klímaadatok esetében), akkor nem lehet szétválasztani a magyarázó változók hatásait így a változók fontosságának vizsgálatakor téves következtetésekre juthatunk (Gregorutti et al., 2017). Ebből következik, hogy elemzésünk során limitált volt a bevonható független változók száma, így értelemszerűen nem a lehető legerősebb modellt vizsgáltuk. A modellépítés során összesen 8 db független (magyarázó) változót vontunk be, amelyeket az alábbiakban mutatunk be.

2.3.1. Geológiai adatok

A modellben használt geológiai információt a MÁFI 1:100 000-es felszíngeológiai térképének geokémiai szempontú reklassifikációjával hoztuk létre. Az új talajtani célú klasszifikációt Gray és munkatársai (2016) módszeréből kiindulva a Balaton vízgyűjtő terület földtani adottságaihoz igazítva készítettük el. A 9 kategória térbeli elhelyezkedése az 1. ábrán látható. Az 1.-8. kategória a kőzetek átlagos SiO₂ tartalma alapján csökkenő sorrendben lett felállítva. A 9. kategória a szerves üledékeket jelöli, melynél a SiO₂ tartalom nem játszott szerepet (1. táblázat). A kategóriák felállításának módjáról egy korábbi munkánkban részletes információ érhető el (Kassai et al., 2017). A geológiai tényezőt a modell építés során kategóriaváltozóként használtuk fel.

1. táblázat: geológiai kategóriák a kőzetek SiO₂ tartalma alapján

Kategória	Átlagos SiO ₂ %	Kőzettípus
1	80-90%	eolikus kvarchomok
2	70-80%	folyóvízi homok, kavics, proluviális, deluviális homok, homokkő
3	65%	löss, homokos lösz
4	60%	aleurit (folyóvízi, tavi, lejtőüledék, deluviális üledék)
5	55%	agyag (folyóvízi, tavi, lejtőüledék, deluviális üledék)
6	50%	bazalt
7	10%	bauxit, bauxitos agyag, agyagos bauxit
8	1-5%	mészkö, dolomit, márga
9	-	szerves üledékek (tőzeg)

2.3.2. Domborzati adatok

A modell illesztéséhez a Földmérési és Távérzékelési Intézet digitális domborzatmodelljéből származtatott 90X90 pixelméretű magassági raszterből levezetett változókat használtunk. ArcGIS programcsomag segítségével a domborzatmodell simítása után kiszámítottuk a lejtés, a kitettség és a felszín görbületet leíró paramétereket. Ezt követően kiszámoltuk az ún topográfiai pozíciós index (TPI) értékeit (Jenness, 2006) ami az adott terület magassági értékeit egy adott nagyságú és alakú környezet (jelen munka esetében 2000 méter sugarú kör) értékeihez viszonyítja. Ez a változó a geomorfológiai helyzet leírását szolgálja, mivel pozitív értékei magaslatot, negatív értékei völgyet, nulla körüli értékei sík területet jeleznek. Az 5 különböző domborzati paraméter (magasság, lejtés, kitettség, görbület, TPI) folytonos változóként került felhasználásra az elemzés során.

2.3.3. Klíma adatok

Munkánk során a nyílt hozzáférésű WorldClim adatbázis adataival dolgoztunk. Az adatbázis ún. bioklimatikus változói közül az éves átlagos csapadék valamint az éves átlagos hőmérséklet 1 km²-es pixelméretű raszterre számolt adatait, mint folytonos változókat használtuk. A klíma adatok az 1970-2000-es évek közötti értékek átlagait mutatják (Fick et al., 2017).

3. Eredmények és értékelésük

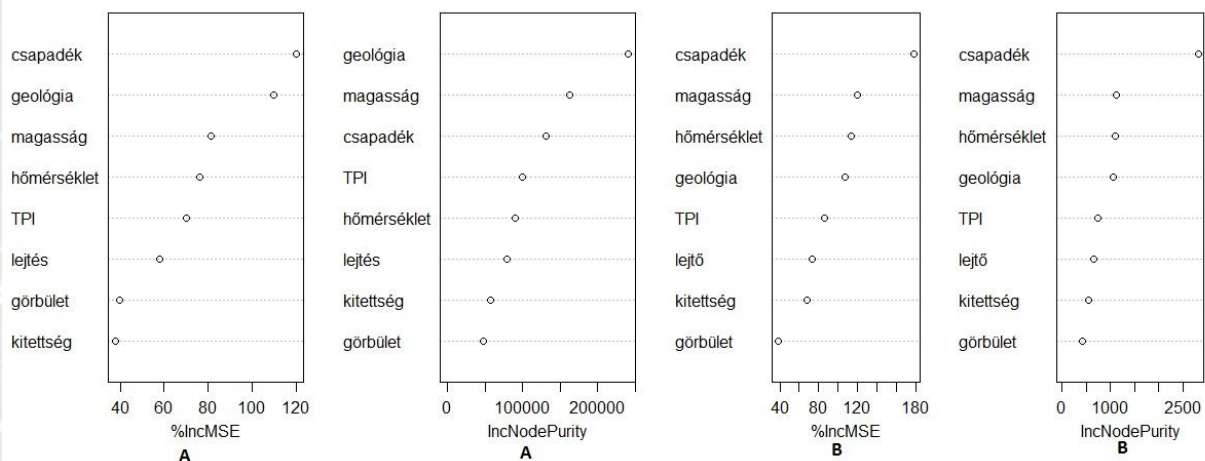
Az elemzés során először két regressziós modellt építettünk fel a talajokban mért pH értékeire valamint a kötöttségi értékekre nézve. A modelleket 500 db regressziós fából, minden fa esetében négy véletlenszerűen kiválasztott magyarázó változó segítségével hoztuk létre. Ezt követően a geológia, mint független változó kizárásával újabb két modellt építettünk abból a célból, hogy a geológia szerepét megvizsgálhassuk a determinációs együttható valamint a becslés hibájának alakulásában. A négy modell összefoglalása a 2. táblázatban látható.

2. táblázat: a négy regressziós modell eredményei

	Determinációs együttható (modell)	Becslési hiba	Korreláció vizsgálat (becsült-valós)
kötöttség	R²	RMSE	R²
geológia felhasználásával	0.68	5,6	0.67
geológia elhagyásával	0.64	6,1	0.63
pH			
geológia felhasználásával	0.64	0.53	0.65
geológia elhagyásával	0.62	0.56	0.63

Az első két modell (ahol a geológiát is figyelembe vettük) eredményei azt mutatják, hogy a kötöttségi értékek variációját 68%-ban magyarázzák a környezeti segédváltozók, a pH értékek esetében ez az érték 64%. Ha a geológiát nem vesszük figyelembe mindkét érték valamelyest csökken, bár a szakirodalmi adatokhoz képest lényegesen kisebb mértékben. Az átlagos négyzetes hiba négyzetgyöke (becslési hiba) bár növekedett, de szintén nem tapasztalható igazán nagy eltérés. Ugyanez elmondható a tesztadatsor becsült és valós értékeinek összehasonlítása esetén is.

A véletlen erdő módszer azonban lehetővé teszi, hogy az összes változó fontosságát egy lépésben is vizsgáljuk, így egymáshoz képest is látható az egyes változók szerepe. Ennek a vizsgálatnak kétféle megközelítése létezik: az egyik az átlagos négyzetes hiba növekedését (IncMSE) figyeli az egyes változók kihagyása esetén, majd ez alapján rangsorol. A másik megközelítés a döntési fák csomópontjaiban mért homogenitás növekedését (IncNodePurity) nézi a vágások során a reziduális négyzetösszeg eltérése alapján. Ennek eredményei a 2. ábrán láthatóak. Ez alapján jól látszik, hogy különösen a kötöttség esetén a geológiának igen fontos szerepe van, de a pH esetében is a geológia a negyedik legfontosabb változó.



2. ábra: Az egyes független változók fontossága A: kötöttség, B: pH esetében az átlagos négyzetes hiba (IncMSE) és a reziduális négyzetösszeg eltérései (IncNodePurity) alapján.

4. Következtetések, javaslatok

A fenti elemzések eredményeiből nehéz általános következtetéseket levonni, hiszen egy adott modellben, az adatok jellemzői (változók típusa, téradat típusa, mintavételezés eloszlása) erősen befolyásolhatja az egyes változók szerepét. A geológia esetében ez azt jelenti, hogy hiába egyértelműen erős az összefüggés számos talajtulajdonság és a geológiai kategóriák között, de mivel a talajadatok legnagyobb része geológiailag homogénebb területről származik (pl. szántók), így az erre az adatbázisra felépített modellben a geológiai hatás kevésbé tud érvényesülni. Ezt erősíti a SiO₂ tartalmat figyelembe vevő kategorizálásunk is, ami talajtani szempontból racionális, de eltüntet bizonyos különbségeket, és hasonlóknak mutatja azokat a geológiai egységeket, amelyek főleg szántókon fordulnak elő. Ez az oka annak, hogy bár látszik, hogy fontos változóról van szó, igazán nagy különbséget nem tapasztalunk, ha kihagyjuk a modellekből. Más lenne természetesen a helyzet, ha azokon a területeken is lenne több talajmintavételi pont, ahol a MÁFI térkép igen részletes és a geológia sokszínűbb (hegy- és dombvidékek). Emellett valószínű, hogy egy országos szintű elemzésnél, a geológia lényegesen nagyobb jelentőségű lehet, mint környezeti segédváltozó, mivel az ország egésze sokkal diverzebb földtani szempontból, mint a Balaton vízgyűjtő szántó területei. Ez a fajta léptékfüggőség az, ami miatt Gray és munkatársai (2016) az Ausztrál kontinensre átfogó vizsgálatukban nagyobb jelentőséget tulajdonítanak a geológiának, mint ami az esetünkben egy sokkal kisebb területen kimutatható. Mindazonáltal a geológia nem elhanyagolható jelentőségű a Balaton vízgyűjtő terület mezőgazdasági területein sem, így a további munka során felhasználásra kerül részletes talajtulajdonság és talajtípus térképek előállításához.

5. Köszönetnyilvánítás

A publikáció és poszter elkészítését az EFOP-3.6.3-VEKOP-16-2017-00008 és az EFOP-3.6.1-16-2016-00015 számú projekt támogatta. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.

6. Felhasznált irodalom

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

Fick, S.E. - R.J. Hijmans (2017): Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. doi:10.1002/joc.5086

Gray, J.M. - Bishop, T.F.A. - Wilford, J.R. (2016): Lithology and soil relationships for soil modelling and mapping. *Catena*, 147, 429–440.

Gregorutti, B. - Michel B. - Saint-Pierre, P. (2017): Correlation and variable importance in random forests. *Statistics and Computing*, 27/3, 659-678.

Gyalog L. - Síkbegyí F. [szerk.] (2000): Magyarország földtani térképe, M=1:100 000. Magyar Állami Földtani Intézet kiadványa. Budapest.

Jenness, J. (2006): Topographic Position Index extension for ArcView 3.x,v.1.3a. Jenness Enterprises. [online:www.jennessent.com/arcview/tpi.htm]

Jenny, H. (1941): Factors of soil formation. McGraw-Hill Book Company, New York.

Kassai, P. - Kocsis, M. - Sisák, I. (2017): A talajképző kőzet SiO₂ tartalmának hatása egyes talajtulajdonságokra a Balaton vízgyűjtőjén. In: Absztraktkötet XXIII. Ifjúsági Tudományos Fórum, Keszthely, 2017. május 26.

Kocsis, M. - Tóth, G. - Berényi Üveges, J. - Makó, A. (2014): Az Agrokémiai Irányítási és Információs Rendszer (AIIR) adatbázis talajtani adatainak bemutatása és térbeli reprezentativitás-vizsgálata. *Agrokémia és Talajtan*, 63, 223-248.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B. (2003): On digital soil mapping. *Geoderma* 117, 3-52.

LIX. GEORGIKON NAPOK

59th Georgikon Scientific Conference