

## **Kategória típusú talajtérképek térbeli felbontásának javítása kiegészítő talajtani adatok és adatbányászati módszerek segítségével**

PÁSZTOR László, BAKACSI Zsófia, LABORCZI Annamária és SZABÓ József

MTA Agrártudományi Kutatóközpont Talajtani és Agrokémiai Intézet  
(MTA ATK TAKI), Budapest

### **Bevezetés**

A talajok különböző statikus és dinamikus jellemzőire, illetve a környezeti folyamatokban betöltött szerepére vonatkozó információk iránt a társadalmi igények folyamatosan bővülnek (BULLOCK, 1999; MERMUT & ESWARAN, 2000; TÓTH et al., 2008; BAUMGARDNER, 2011). A korábban gyűjtött, térképezések, felvételezések által szolgáltatott információk hosszú időn keresztül jól szolgálták az elvárásokat, alapvetően annak köszönhetően, hogy döntően ez utóbbiak határozták meg gyűjtésük célját. A talaj multifunkcionalitásának széleskörű felismerése (BLUM, 2005) azonban éppen az adatgyűjtésre fordítható erőforrások beszűkülésével egy időben következett be. Az aktuálisan rendelkezésre álló, illetve a felhasználók által specifikusan megkívánt információk pedig nem okvetlenül, sőt egyre ritkábban fedik egymást. 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). Emiatt számos esetben a döntéshozók jelenlegi igényeinek kielégítése sem történhet meg megfelelő hatékonysággal. A probléma megoldása érdekében számos próbálkozás született a létező talajtani információk kiegészítésére, pontosítására, harmonizációjára, korreláltatására és integrálására.

A hazai térképi alapú (mostanra szinte kizárólagosan) digitális talajtani adatigények aktuális kiszolgálására több lehetőség adódik. Az ideális megoldás – legalábbis a felhasználó szempontjából – minden felmerülő probléma esetén egyedi, specifikus, térben és tematikusan is nagy részletességű adatgyűjtésen alapuló térképezés lenne. Erre azonban talán soha nem lesz lehetőség. Marad a korlátozott adatgyűjtés, avagy jellemzően annak teljes hiánya. Ilyenkor a rendelkezésre álló, archív adatokra tudunk támaszkodni.

Jelentős mennyiségű, korábbi adatokon alapuló digitális talajinformáció áll Magyarországon is készen a talajtani adatigények legalább részleges kielégítésére köszönhetően, hogy az 1980-as évektől kezdődően a térképi alapú talajtani információ-

ók jelentős része került digitális feldolgozásra és épült be különböző térbeli talajinformációs rendszerekbe (TTIR; PÁSZTOR et al., 2013a). Számos előnyük ellenére figyelemmel kell lenni azonban arra, hogy ezen adatbázisok eredeti adatrendszere nem a mai igények kielégítésére jött létre és semmi esetre sem tekinthetők omnipotensnek. Jól hasznosíthatóknak viszont igen, hiszen jelentős tudás és tapasztalat halmozódott fel bennük, amelyek megfelelő kérdésfeltevással és módszerek alkalmazásával belőlük kinyerhetők, megnyitva a továbbfejlesztésük lehetőségeit.

Gyakran felmerülő probléma, hogy különböző léptékekben nem áll rendelkezésre azonos tematikus információ. Léteznek például genetikus talajtérképek, illetve a talajértékelésre vonatkozó információt hordozó térképek országos és üzemi léptékben, de a kettő közötti térbeli felbontásban nem. Márpedig számos esetben szükséges az üzemi térképek által lefedett területeknél jóval nagyobb kiterjedésben (megyékre, de akár az egész országra is) az országos térképek által nyújtott térbeli felbontást meghaladó térbeli információ. Jelen munkában ezen problémára, azaz bizonyos, kategória típusú tematikus talajtérképek térbeli felbontásának újraosztályozással történő javítására (leskálázására) mutatunk megoldást.

## Anyag és módszer

### *Talajtérképi és talajtérképezési háttér*

Az Agrotopográfiai térképek szintetizáló munka eredményeként születtek. Ez döntően a Kreybig-féle talajfelvételezés adataira alapozva és a Kreybig térképek talajfoltjainak térbeli és tematikus generalizálásával történt (VÁRALLYAY et al., 1979, 1980). A generalizálás összevonással és elhanyagolással, ennek megfelelően információvesztéssel jár. Az így született agroökológiai egységek értelemszerűen inhomogének, amelyek összetételére vonatkozóan a Kreybig mintázat (PÁSZTOR et al., 2001) jelentős információtartalommal bír. Az agroökológiai egységeket másrészről az akkor segédeszközülnél választott 1:100.000-es topográfiai térképek által szolgáltatott domborzathoz illesztették. Nagyobb léptékű, jobb felbontású topográfiai alap alkalmazása szintén hatékony eszközzel szolgálhat a foltokon belüli heterogenitás elemzéséhez. Az Agrotopográfiai térképek az AGROTOPO, a Kreybig térképezés alap- és feldolgozott adatai pedig a DKTIR (Digitális Kreybig Talajinformációs Rendszer) adatbázisban az ország teljes területére digitális elemzésre alkalmas formában rendelkezésre állnak:

(<http://mta-taki.hu/hu/osztalyok/kornyezetinformatikai-osztaly/agrotopo/>;

<http://mta-taki.hu/hu/osztalyok/kornyezetinformatikai-osztaly/dktir/>;

<http://maps.rissac.hu/agrotopo/>).

A talajtérkép a talajtakaró specifikus térbeli modellje, melynek megalkotása a talajképző folyamatok szem előtt tartásával történik. Ezen definíció egyes elemeit illetően a digitális forradalom során jelentős és lényegében egyidejű változások következtek be, amelyek hatásának köszönhető a digitális talajtérképezés kialakulása, megerősödése, majd elterjedése az utóbbi évtizedben (MCBRATNEY et al., 2003; DOBOS et al., 2006; LAGACHERIE et al., 2006; BOETTINGER et al., 2010; ILLÉS et al.,

2011a; SZATMÁRI & BARTA 2013). A talajképző folyamatok egyes szegmenseire közvetve vagy közvetlenül vonatkozó térinformatikai (térbeli és egyben digitális) információk egyre nagyobb mennyiségben, térbeli felbontásban és olcsóbban váltak elérhetővé. Az így elérhető ún. környezeti segédinformációk és a talajok egyes jellemzői közötti, néha nagyon bonyolult és áttételes kapcsolatok számszerűsítésére determinisztikus modellek híján is hatékonyan alkalmazható matematikai (geo-)statisztikai és adatbányászati eszközök jelentek meg. Ezeket eredendően más szakterületek problémáinak kezelésére fejlesztették ki, de digitális talajterképezési feladatokban is jól adaptálhatónak bizonyultak.

#### *Döntési fák és talajterképezési alkalmazásuk*

A döntési fák olyan adatbányászati módszer családot képviselnek, amelynek alapötlete, hogy bonyolult összefüggéseket egyszerű döntések sorozatából építenek fel (mintegy sorba fejtve azokat). A fát egy ún. tanító adatbázisból rekurzívan állítjuk elő. Az osztályozó fák mindig a lehető legnagyobb homogén osztályt választják le valamely, jól megválasztott, jellemzően entrópia jellegű mérték (pl.: Gini, Shannon index) szerint. A döntési fákon a kezdőfeltételtől (a gyökérből) egy besorolásig (levélbe) vezető út mentén a feltételeket összeolvasva könnyen értelmezhető szabályokat kapunk.

A döntési fák számos előnnyel rendelkeznek, többek közt:

- Kategória és numerikus típusú változókra egyaránt alkalmazhatók.
- A változók közötti nem-linearitás kapcsolatok esetén is alkalmazhatók.
- Automatikusan felismerik a lényegtelen változókat és így a fák teljesítménye zaj jelenlétében sem romlik.
- Automatikusan felismerik a kollinearitást.
- Nagyméretű adathalmazokra is hatékonyan felépíthetők.
- A fák szerkezete invariáns a független (numerikus) változók monoton transzformációjára.

A döntési fákból nyert döntési szabályhalmazok egyértelműek, így osztályozásra használhatók. Viszont több feltétel sor is vezethet azonos döntéshez, azaz a fa több levele is reprezentálhatja ugyanazt az osztályt.

A döntési, avagy osztályozási fákat hatékonyan alkalmazták a világ különböző részein talajtípusok előfordulásának térbeli modellezésére, azaz hagyományos értelemben vett talajterképek előállítására (Ausztrália: MORAN & BUI, 2002; Brazília: GIASSON et al., 2011; Dánia: BOU KHEIR et al., 2010; Magyarország: ILLÉS et al., 2011b; USA: SCULL et al., 2005). Ezen munkáknál, jellemzően, viszonylag kevés feltárásban elvégzett osztályozás eredményének térbeli kiterjesztése történt meg a helyi fiziográfiai viszonyokhoz illeszkedő talaj-táj modellek szerinti környezeti segédváltozók segítségével.

A döntési fák felhasználhatók a már létező talajterképekben foglalt talaj-táj modellek megértésére, a felvételezési, szerkesztési szabályok utólagos formalizálására is (BUI & MORAN, 2001; HARING et al., 2012). Az így feltárt és döntési szabályokba foglalt összefüggések pedig nagy felbontású környezeti segédváltozók segítségével térbelileg finomított térképek előállítását teszik lehetővé. Ezen segédváltozók között

speciális szerepet töltenek be a nagyobb térbeli felbontású, de eltérő tematikájú talajtani információk. Magyarországon a teljes Kreybig térképi archívum (KREYBIG, 1937) térinformatikai feldolgozása (PÁSZTOR et al., 2012) ezen a téren hatalmas potenciált hozott létre. Mindezen lehetőségek együttes kihasználásával tettünk kísérletet kategória típusú tematikus talajtérképek leskálázására. Jelen dolgozatban két példát mutatunk a vázolt koncepció konkrét alkalmazására:

(i) Az Országos Területrendezési Terv megújítása szándékai szerint már a kezdetektől olyan területi lehatárolásokat tűz ki célul, amelyek a korábbiaknál harmonikusabb átmenetet jelentenek a nagyobb felbontású megyei és települési szint felé. A részletesebb szintű rendezési tervek kidolgozásához a Digitális Kreybig Talajinformációs Rendszer (DKTIR) használatát korábbi munkánkban ismertettük (SZABÓ et al., 2007). A DKTIR országos kiépítése természetesen sugallta a gondolatot, hogy a Kiváló Termőképességű Területek országos léptékű lehatárolásához jól alkalmazható eszközt adhat a kezünkbe megfelelő módszer alkalmazása esetén. Ez utóbbi pedig az AGROTOPO talajértékszám kategóriáinak leskálázása döntési fák segítségével. (Mivel a talajértékszám ordinális skálán jellemzi a talajok termőképességét, numerikus változóként is kísérletet tettünk térbeli leskálázására, jelesül regressziós krigelés segítségével, mely módszerrel részletesebben foglalkozunk ugyanezen kötetbe benyújtott másik cikkünkben [SZATMÁRI et al., 2013].)

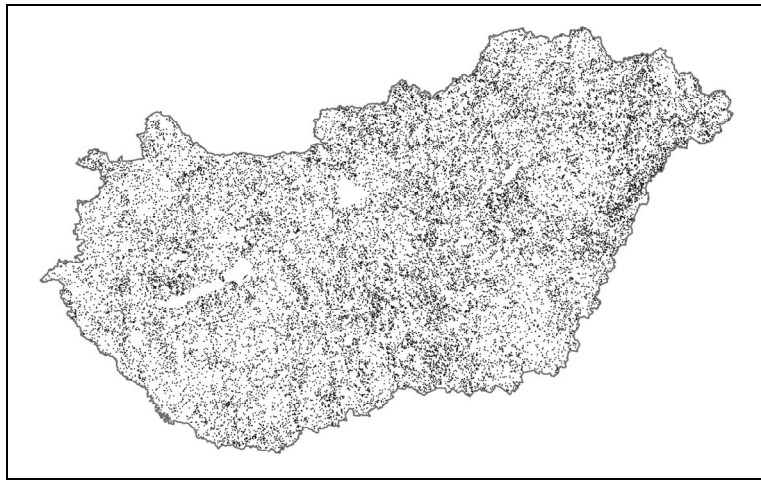
(ii) SZABOLCS és munkatársai (1968, 1969a) az öntözés tervezését szolgáló térképezési módszertant dolgoztak ki és hajtottak végre 1:100.000-es léptékben Szolnok, Hajdú-Bihar, Békés és Csongrád megyék területére. Nagyobb léptékű térképezési módszertant is kidolgozták 1:25.000-es léptékben, részletesen ismertetve az ahhoz szükséges alaptérképek elkészítését (SZABOLCS et al., 1969b), de magát a térképezést csak korlátozottan, mintajelleggel végezték el. Az öntözés jelentősége ma sem vitatható; tervezési munkálataihoz ideális eszközt nyújthat a korábban kidolgozott módszertan digitális implementálása és a részletes felbontású térképezés elvégzése további területekre. Egy későbbi cikkünkben részletesebben foglalkozunk majd ezzel a kérdéskörrel. Jelen kontextusban a lényeges elem, hogy az öntözés talajtani lehetőségeinek regionalizálásához a módszertan épít a talajtakaró genetikus osztályozására, ami 1:25.000-es léptékben nem áll általánosságban rendelkezésre hazánkban. Ennek áthidalására kísérletet tettünk az AGROTOPO genetikus típus fedvényének leskálázására egy alföldi belvízöblözet területére a DKTIR és természetesen egyéb környezeti segédváltozók, illetve osztályozó fák felhasználásával.

#### *Talajértékszám kategória térkép leskálázása az ország teljes területére*

A feladat elvégzéséhez szükséges segédadatok kiválasztása a talajképződési tényezők szem előtt tartásával, illetve az ezekre vonatkozó elérhető térképi alapú információk elérhetősége alapján történt. Segédváltozóként az ún. SRTM digitális domborzatmodellt (RABUS et al., 2003) és a DKTIR talajtérképi egységeit használtuk. A DDM-ből a következő elsődleges, illetve másodlagos paramétereket származtattuk: tengerszint feletti magasság, LS faktor (WISCHMEIER & SMITH, 1978), topográfiai pozíció index (500 m, illetve 2000 m-rel paraméterezve) és relatív magasság. A talajokat öt szempont alapján jellemeztük. Vízgazdálkodási, kémiai tulaj-

donságaik, sekély termőréteg vastagságuk, illetve tájtermesztési besorolásuk a DKTIR talajfoltjainak attribútumai. A szervesanyag-tartalomra vonatkozóan a DKTIR talajszelvény adatbázisának felhasználásával készítettünk regresszió krigeléssel – a SZATMÁRI és munkatársai (2013) által bemutatott módon – országos szerves anyag térképet.

Az elemzésekhez 100 méteres cellaméretet használtunk. A vektoros állományokat az így definiált referencia rácsra vonatkoztatva alakítottuk raszterekké. Az AGROTOPO talajértékszám kategóriáiban megjelenő modellezési szabályokat



1. ábra

A virtuális mintavételi pontok térbeli eloszlása

mintegy 30.000, véletlenszerűen, de bizonyos peremfeltételek szerint elszórt virtuális mintavételi pontban (1. ábra) próbáltuk értelmezni a segédváltozók alapján. Kitétel volt, hogy az AGROTOPO minden térképi egységébe legalább egy mintavételi pont essen, illetve, hogy két szomszédos pont távolsága legalább 200 méter legyen. A döntési fa algoritmusban a talajértékszám kategória jelentette az osztályozandó, függő változót, a környezeti segédadatok pedig a független változók halmazát. Az osztályozást különböző paraméterezéssel végeztük el.

#### *Genetikus típus térkép leskálázása a Mezőberényi belvízöblözet területére*

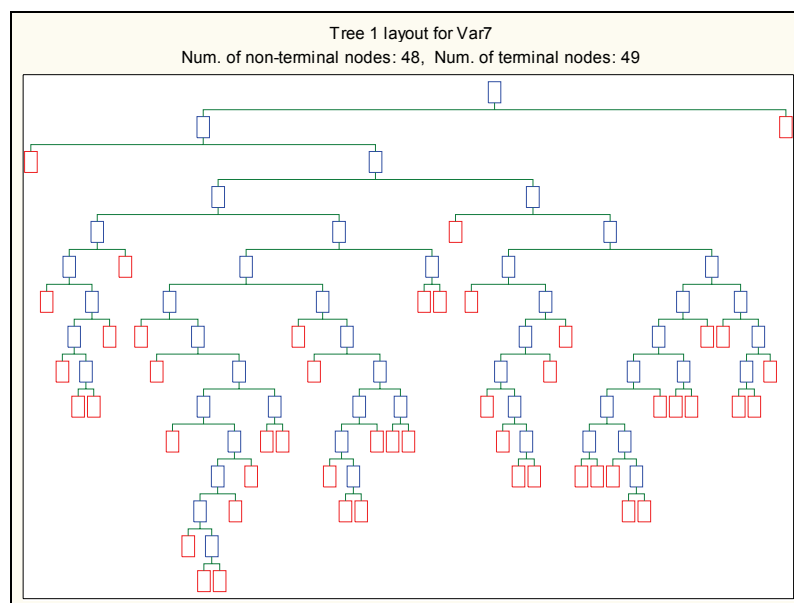
A feladat elvégzéséhez szükséges segédadatok kiválasztása a talajképződési tényezők szem előtt tartásával, illetve az ezekre vonatkozó elérhető térképi alapú információk elérhetősége alapján történt. Segédadatként 20 méteres digitális domborzatmodellt és a DKTIR talajtérképi egységeit használtuk. A DDM-ből a következő elsődleges, illetve másodlagos paramétereket származtattuk: tengerszint feletti magasság, sík-görbület, csatorna-hálózat alap szint (channel network base level), relatív magasság. A talajokat két szempont, vízgazdálkodási és kémiai tulajdonságaik alapján jellemeztük.

Az elemzésekhez a DDM 20 méteres cellaméretét használtunk. A vektoros állományokat az így definiált referencia rácsra vonatkoztatva alakítottuk raszterekké. Az AGROTOPO genetikus típusaiban megjelenő modellezési szabályokat mintegy 5.000, véletlenszerűen kijelölt pontban próbáltuk értelmezni a segédváltozók alapján. Peremfeltétel volt, hogy az AGROTOPO genetikus foltjaiba legalább öt mintavételi pont essen, illetve, hogy két szomszédos pont távolsága legalább 40 méter legyen. A döntési fa algoritmusban a genetikus típus jelentette a függő, cél-változót, a környezeti segédadatok pedig a független változók halmazát. Az osztályozást különböző paraméterezéssel végeztük el.

A domborzati paraméterek származtatását mindkét esetben SAGA (System for Automated Geoscientific Analysis) GIS környezetben végeztük el. Az osztályozó fák előállításához a STATISTICA szoftver (StatSoft, Inc.) adatbányászati modulját használtuk.

### Eredmények

A genetikus típus térkép Mezőberényi belvízöblözet területére történő leskálázásánál vizsgált osztályozások közül a legjobb eredményt a 2. ábrán látható, 49 levelet tartalmazó rekurzió szolgáltatta. Az előállított döntési fa az egyes talajtípusok előfordulását több levélen is megjósolja. Az 1. táblázat példaképpen az öt levélen megjelenő réti csernozjomok becslésére előállított szabálysorokat mutatja be.



2. ábra

A mintaterületi genetikai talajtérkép leskálázásához előállított döntési fa szerkezete

## 1. táblázat

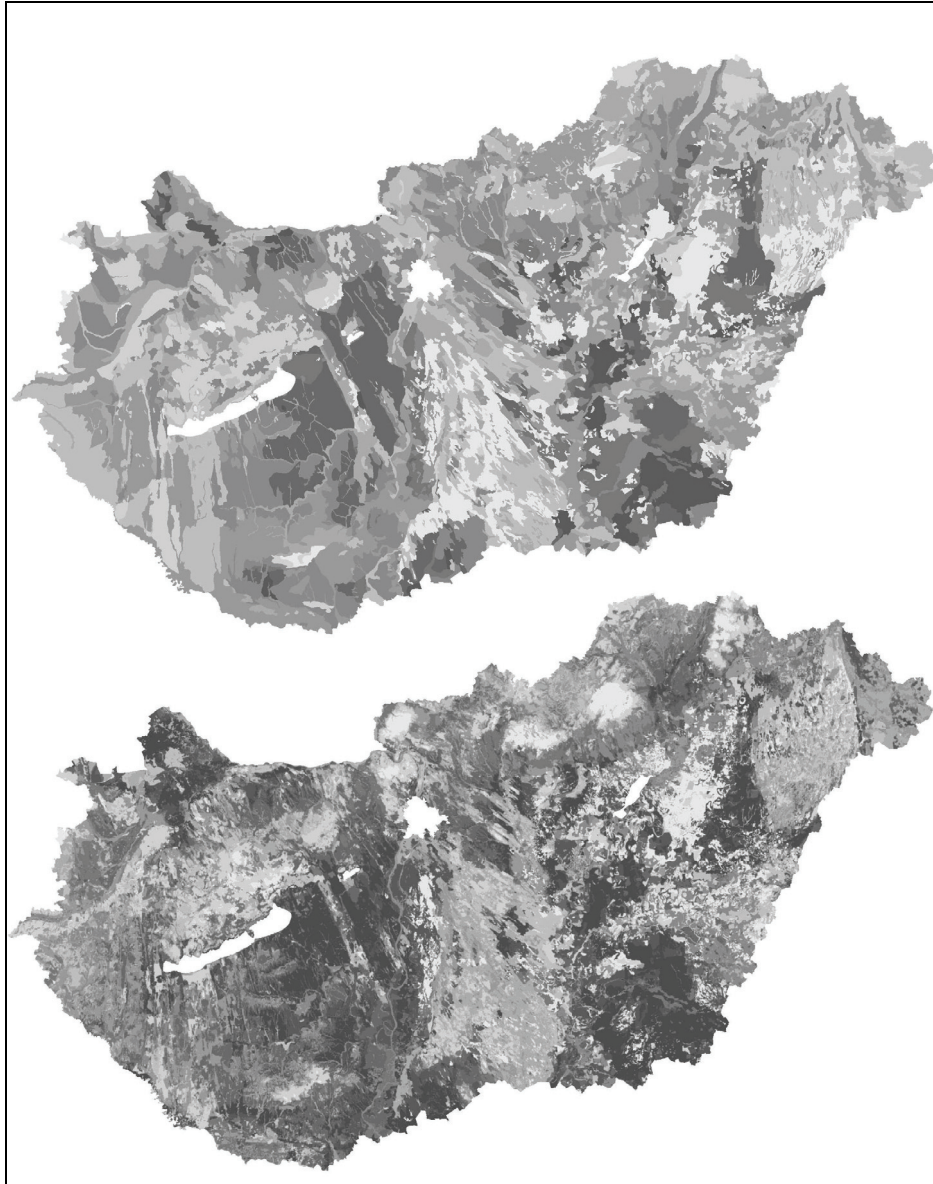
Réti csernozjom talajok térbeli előfordulásának egyes becslései és azok feltétel rendszere

(1) Becslési valószínűség (%)	(2) Feltétel
100	(sík-görbület $\leq 50,57$ ) és (DKTIR kémia $\diamond 6$ ) és (DKTIR kémia $\diamond 1$ ) és (DKTIR $\diamond 66$ és DKTIR $\diamond 77$ és DKTIR $\diamond 88$ ) és (DKTIR fizika $\diamond 9$ és DKTIR $\diamond 99$ ) és (channel_ntw $> 8399,97$ ) és (DKTIR fizika $\diamond 2$ ) és (channel_ntw $> 8596,93$ )
37	(sík-görbület $\leq 50,57$ ) és (DKTIR kémia $\diamond 6$ ) és (DKTIR kémia $\diamond 1$ ) és (DKTIR $\diamond 66$ és DKTIR $\diamond 77$ és DKTIR $\diamond 88$ ) és (DKTIR fizika $\diamond 9$ és DKTIR $\diamond 99$ ) és (channel_ntw $> 8399,97$ ) és (DKTIR fizika = 2)
29	(sík-görbület $\leq 50,57$ ) és (DKTIR kémia $\diamond 6$ ) és (DKTIR kémia = 1) és (DKTIR fizika = 2 vagy DKTIR fizika = 3 vagy DKTIR fizika = 5) és (vert_dist_t $> 234$ )
19	(sík-görbület $\leq 50,57$ ) és (DKTIR kémia $\diamond 6$ ) és (DKTIR kémia = 1) és (DKTIR fizika = 2 vagy DKTIR fizika = 3 vagy DKTIR fizika = 5) és (vert_dist_t $\leq 234,01$ ) és (DKTIR fizika = 5)
18	(sík-görbület $\leq 50,57$ ) és (DKTIR kémia $\diamond 6$ ) és (DKTIR kémia = 1) és (DKTIR fizika $\diamond 2$ és DKTIR fizika $\diamond 3$ és DKTIR fizika $\diamond 5$ ) és (vert_dist_t $\leq 186,59$ ) és (TFM $> 8692$ ) és (channel_ntw $\leq 8606,8$ )

*Megjegyzés: DKTIR (Digitális Kreybig Talajinformációs Rendszer) térképezési egységeinek kód értelmezése: 66: Időszakosan vízállásos, vízjárta területek, 77: Erdők, 88: Tavak, nádasok, folyóvizek, 99: Települések. DKTIR fizikai talajtulajdonságok: 2: Közepes vízvezető képességű, a vizet erősen tartó talajok, 3: Gyenge vízvezető képességű, a vizet erősen tartó, erősebben repedező talajok, 5: Igen nagy vízvezető képességű, gyengén víztartó talajok, 9: Szikes talajok. DKTIR kémiai talajtulajdonságok: 1: Túlnyomóan semleges vagy gyengén lúgos, mésszel telített talajok, 6: Szántóföldi művelésre alkalmatlan szikes talajok, mésszel nem javíthatók. channel\_ntw: csatorna hálózat alap szint; vert\_dist\_t: relatív magasság a csatorna hálózat alap szintre vonatkoztatva*

A talajértékszám kategória térkép az ország teljes területére történő leskálázásánál vizsgált osztályozások közül a legjobb eredményt a 2. ábrán láthatóhoz hasonló, de hely hiányában be nem mutatott, 97 levelet tartalmazó rekurzió szolgáltatva.

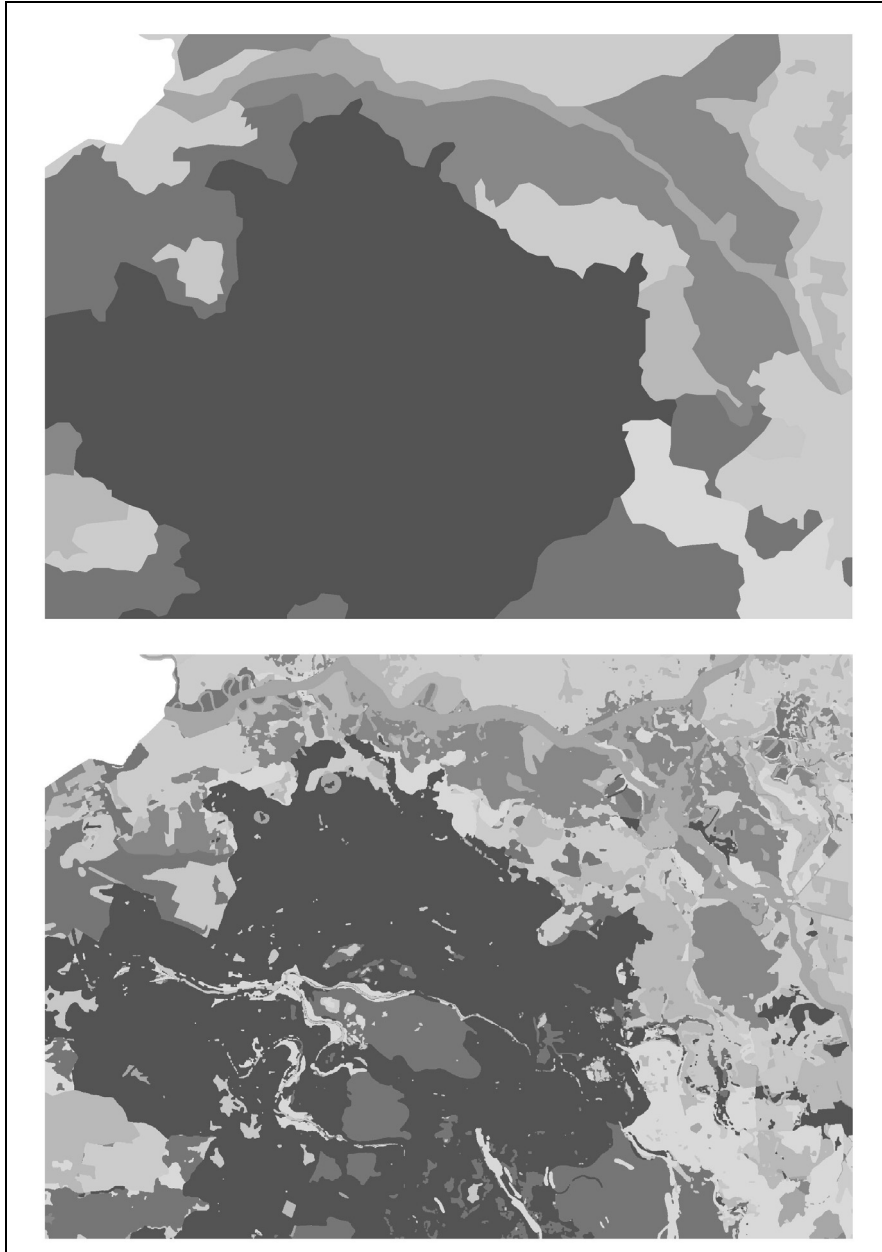
Az osztályozás eredményeit alkalmaztuk a tanuláshoz használt, teljes területi fedettséget nyújtó, nagy térbeli felbontású fedvényekre. A pixel alapú kategorizálás eredményeképpen állnak elő a célul kitűzött leskálázott térképek. A 3. és 4. ábra megpróbálja szemléltetni azt a térbeli felbontásban történt drasztikus javulást, amit a visszaosztályozás utáni térképek szolgáltatnak.



3. ábra

Az eredeti és a leskálázott, országos talajértékszám térkép





*4. ábra*

Az eredeti és a leskálázott, mintaterületi genetikai talajtérkép

## Összefoglalás

Kategória típusú talajterképek térbeli felbontásának javítására (leskálázására) adatbányászati módszerként osztályozó fákat alkalmaztunk. A feladat elvégzését az indokolta, hogy bizonyos tematikus talajterképek nem állnak rendelkezésre a megkívánt léptékben. A számos előnnyel bíró döntési fák felhasználhatók többek között létező talajterképekben foglalt talaj-táj modellek megértésére, a felvételezési, szerkesztési szabályok utólagos formalizálására. A feltárt, döntési szabályokba foglalt összefüggések pedig nagy felbontású környezeti segédváltozók segítségével térbelileg finomított térképek előállítását teszik lehetővé. Ezen segédváltozók között speciális szerepet töltenek be a nagyobb térbeli felbontású, de eltérő tematikájú talajtani információk. A szintetizáló munka eredményeként született Agrotopográfiai térképek agroökológiai egységei döntően a Kreybig-féle talajfelvételezés adataira alapozva, a Kreybig térképek talajfoltjainak térbeli és tematikus generalizálásával jöttek létre. A Kreybig mintázat jelentős információtartalommal bír ezek heterogenitására vonatkozóan, csakúgy mint a talajképződésben jelentős szerepet betöltő domborzati viszonyokat nagy léptékben jellemző domborzat modellek.

Az ország teljes területére elérhető AGROTOPO és DKTIR (Digitális Kreybig Talajinformációs Rendszer) adatbázisok, illetve digitális domborzat modellek révén hatalmas potenciál áll rendelkezésre, mely lehetőségek integrált kiaknázásával számos kategória típusú tematikus talajterkép leskálázására nyílik lehetőség. Ezek közül kettőt mutattunk be részletesebben:

(i) Az Országos Területrendezési Terv Kiváló Termőképességű Területeinek országos léptékű lehatárolásához az AGROTOPO talajértékszám kategória térképét skáláztuk le döntési fák segítségével.

(ii) Az öntözés tervezését szolgáló nagyléptékű térképezési módszertan digitális implementálásának feltétele a talajok genetikus osztályozásának ismerete 1:25.000-es léptékben. Egy belvízöblözet területére az AGROTOPO genetikus típus fedvényét skáláztuk le a DKTIR és természetesen egyéb környezeti segédváltozók, illetve osztályozó fák felhasználásával.

A felhasznált módszert a későbbiekben is alkalmazni kívánjuk hasonló feladatok megoldására, például az országos digitális talajterképezési projektben (DOSoReMI.hu; PÁSZTOR et al., 2013b), a levont következtetéseket és a tanulságokat pedig hasznosítani. Tovább lépést a környezeti segédadatok körének bővítésével, illetve a döntési fák egy további szintjének, az ún. „random forest” (véletlen erdő) módszercsalád alkalmazásával tervezünk.

Munkánkat az K105167 OTKA pályázat és a Bolyai János Kutatási Ösztöndíj program támogatta.

**Kulcsszavak:** adatbányászat, digitális talajterképezés, DKTIR, döntési fák, környezeti segédváltozók

### 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.
- 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. et al., (Eds.) 2010. *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Springer. Dordrecht.
- BOU KHEIR, R. et al., 2010. The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data. *Hydrol. Earth Syst. Sci.* **14**. 847–857.
- BUI, E. N. & MORAN, C. J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*. **103**. 79–94.
- 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. Luxembourg.
- 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. Luxembourg.
- HÄRING, T. et al., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*. **185–186**. 37–47.
- GIASSON, E. et al., 2011. Decision trees for digital soil mapping on subtropical basaltic steplands. *Sci. agric. (Piracicaba, Braz.)* [online]. **68**. (2) 167–174.
- ILLÉS, G., KOVÁCS, G. & HEIL, B., 2011a. Comparing and evaluating digital soil mapping methods in a Hungarian forest reserve. *Can. J. Soil Sci.* **91**. (4) 615–626.
- ILLÉS G., KOVÁCS G. & HEIL B., 2011b. Nagyfelbontású digitális talaj térképezés a Vaskereszt Erdőrezervátumban. *Erdészeti Közlemények*. **1**. 29–43.
- KREYBIG L., 1937. A Magyar Királyi Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere. *M. Kir. Földtani Intézet Évkönyve*. 31. 147–244.
- LAGACHERIE, P., MCBRATNEY, A. B. & VOLTZ, M. (Eds.), 2006. *Digital Soil Mapping An Introductory Perspective*. Elsevier B.V. Amsterdam.
- MCBRATNEY, A. B., MENDONÇA SANTOS, M. L. & MINASNY, B., 2003. On digital soil mapping. *Geoderma*. **117**. 3–52.
- MERMUT, A. R. & ESWARAN, H., 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. *Proc. 19<sup>th</sup> World Congress of Soil Science, Brisbane, Australia, Aug 2010*. (DVD)
- MORAN, C. J. & BUI, E. N., 2002. Spatial data mining for enhanced soil map modelling. *International Journal of Geographic Information Science*. **16**. 533–549.
- PÁSZTOR L. et al., 2001. 1:25.000-es méretarányú talajtani-földrajzi mintázat az ország egyes területein a Kreybig Digitális Talajinformációs Rendszer alapján. In: *A földrajz eredményei az új évezred küszöbén*. (Szerk.: DORMÁNY G. et al.). CD-ROM (ISBN 963 482 544 3) Szegedi Tudományegyetem Természeti Földrajzi Tanszék.
- PÁSZTOR, L. et al., 2012. Compilation of 1:50,000 scale digital soil maps for Hungary based on the Digital Kreybig Soil Information System. *J. Maps*. **8**. (3) 215–219.

- 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**. (1) 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: *GlobalSoilMap* (Ed.: ARROUAYS, D.). CRC Press/Balkema (megjelenés alatt).
- RABUS, B. et al., 2003. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. *Photogramm. Rem. Sens.* **57**. 241–262.
- SCULL, P., FRANKLIN, J. & CHADWICK, O. A., 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modeling*. **181**. 1–15.
- SZABÓ J. et al., 2007. A Kreybig Digitális Talajinformációs Rendszer alkalmazása területi szintű földhasználati kérdések megoldásában. *Agrokémia és Talajtan*. **56**. 5–20.
- SZABOLCS I., DARAB K. & VÁRALLYAY GY., 1968. A tiszai öntözőrendszerek és a Magyar Alföld talajainak termékenysége. I. Az öntözés talajtani lehetőségei és feltételei Szolnok, Hajdú-Bihar, Békés és Csongrád megyék területén. *Agrokémia és Talajtan*. **17**. 453–464.
- SZABOLCS I., DARAB K. & VÁRALLYAY GY., 1969a. A tiszai öntözőrendszerek és a Magyar Alföld talajainak termékenysége. II. A talajvíz „kritikus” mélysége a kiskörei öntözőrendszer által érintett területeken. *Agrokémia és Talajtan*. **18**. 211–220.
- SZABOLCS I., DARAB K. & VÁRALLYAY GY., 1969b. A tiszai öntözőrendszerek és a Magyar Alföld talajainak termékenysége. III. Az öntözés lehetőségeit és feltételeit ábrázoló 1:25000-es léptékű térképek készítésének módszerei. *Agrokémia és Talajtan*. **18**. 221–234.
- SZATMÁRI G. & BARTA K., 2013. Csernozjom talajok szervesanyag-tartalmának digitális térképezése erózióval veszélyeztetett mezőföldi területen. *Agrokémia és Talajtan*. **62**. 47–60.
- SZATMÁRI G. et al., 2013. A talajok szervesanyag-készletének nagyléptékű térképezése regresszió krigeléssel Zala megye példáján. *Agrokémia és Talajtan*. **62**. 219–234.
- TÓTH, G. et al., 2008. *Soils of the European Union*. EUR 23439 EN, Office for Official Publications of the European Communities, Luxembourg.
- VÁRALLYAY GY. et al., 1979. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100.000 méretarányú térképe I. *Agrokémia és Talajtan*. **28**. 363–384.
- VÁRALLYAY GY. et al., 1980. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100.000 méretarányú térképe II. *Agrokémia és Talajtan*. **29**. 35–76.
- WISCHMEIER, W. H. & SMITH, D. D., 1978. *Predicting Rainfall Erosion Losses: A Guide to Conservation Planning*. Agric. Handbook No. 537. US Govern. Printing Office. Washington, D. C.

*Érkezett: 2013. szeptember 30.*

## Downscaling of categorical soil maps with the aid of auxiliary spatial soil information and data mining methods

L. PÁSZTOR, Z. BAKACSI, A. LABORCZI and J. SZABÓ

Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research,  
Hungarian Academy of Sciences, Budapest

### Summary

Decision trees were used as a data mining technique to improve the spatial resolution of category-type soil maps (thematic downscaling). The approach was justified by the fact that certain thematic soil maps are not available on the required scale. Decision trees, which have numerous advantages, 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 is played by larger scale soil information with diverse attributes. The agro-ecological units in the AGROTOPO database, compiled as a result of synthesising work, were elaborated mainly on the basis of mapping units from Kreybig soil maps, after spatial and thematic generalization. The Kreybig pattern contains significant information on the heterogeneity of these agro-ecological units, as do the elevation models characterizing the relief circumstances that play a significant role in soil formation processes.

The availability of AGROTOPO and DKISIS (Digital Kreybig Soil Information System) spatial soil information systems and Digital Elevation Models for the whole country has huge potential, which can be exploited in an integrated manner for the downscaling of various, category-type thematic soil maps. Two such possibilities are presented in the present paper:

(i) The soil productivity map included in the AGROTOPO database was downscaled with the aid of decision trees for the delineation of Areas with Excellent Productivity in the framework of the National Regional Development Plan.

(ii) The characterization of the soil cover in terms of genetic soil types at a scale of 1:25,000 is a requirement for the digital implementation of the large-scale mapping methodology designed for irrigation planning. The genetic soil type layer of AGROTOPO was downscaled for the pilot area based on DKISIS, environmental auxiliary variables and decision trees.

The further application of these methods is planned for solving similar problems, for example in the nationwide digital soil mapping project (DOSoReMI.hu; PÁSZTOR et al., 2013b), while the lessons and experience gained will also be exploited. It is also hoped to achieve progress by expanding the pool of environmental co-variables applied and by testing a further level of decision trees, known as the 'random forest' method.

*Table 1.* Predictions of the spatial occurrence of meadow chernozem soils and their decision rule sequences. (1) Prediction probability (%). (2) Condition. *Remarks:* DKISIS (Digital Kreybig Soil Information System) *mapping unit codes:* 66: Areas periodically inundated with water, 77: Forests, 88: Water bodies, 99: Settlements. *DKISIS physical soil properties:* 2: Soils with moderate water conductivity and high water retention,

3: Cracked soils with poor water conductivity and high water retention, 5: Soils with high water conductivity and poor water retention, 9: Salt affected soils. *DKSIS chemical soil properties*: 1: Dominantly neutral or weakly alkaline soils, saturated with lime, 6: Salt affected soils unsuitable for arable use, not ameliorable with lime. *channel\_ntw*: channel network basic level; *vert\_dist\_t*: relative height compared with the channel network basic level.

*Fig. 1.* Spatial distribution of the virtual sampling points.

*Fig. 2.* Decision tree structure elaborated for the downscaling of the genetic soil map of the sample area.

*Fig. 3.* Original and downscaled countrywide soil productivity maps.

*Fig. 4.* Original and downscaled genetic soil maps for the pilot area.