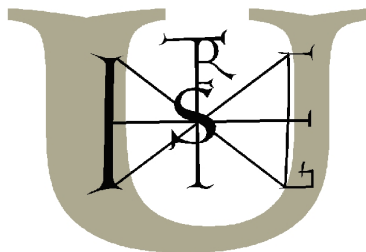


Szent István Egyetem
Környezettudományi Doktori Iskola



Hazai térképi és más talajadatok nemzetközi digitális térképekbe
és adatbázisokba illesztésének megalapozása

Doktori (Ph. D.) értekezés tézisei

Waltner István

2013

A doktori iskola

megnevezése: Környezettudományi Doktori Iskola

tudományága: Környezettudományok

vezetője: Dr. Heltai György
egyetemi tanár, az MTA doktora
SZIE, Mezőgazdaság- és Környezettudományi Kar,
Környezettudományi Intézet, Kémia és Biokémia Tanszék

Témavezető: Csákiné Dr. Michéli Erika
egyetemi tanár, az MTA doktora
SZIE, Mezőgazdaság- és Környezettudományi Kar,
Környezettudományi Intézet, Talajtani és Agrokémiai Tanszék

.....
Az iskolavezető jóváhagyása

.....
A témavezető jóváhagyása

1. A MUNKA ELŐZMÉNYEI, A KITŰZÖTT CÉLOK

Tekintettel arra, hogy mind globálisan, mind országos szinten csökkennek a lehetőségek új, egységes módszertanú talaj-felvételezések kivitelezésére, így egyre nagyobb figyelem összpontosul a már meglévő adatok minél hatékonyabb megőrzésére és felhasználására. Mivel ezek az archív adatok gyakran eltérő módszertan, osztályozási rendszer alapján készültek, szükségessé vált valamilyen közös nevezőre hozásuk, harmonizációjuk.

A Világ Talaj Referenciabázis-t (World Reference Base for Soil Resources, WRB) a Nemzetközi Talajtani Unió 1998-ban elfogadta, mint hivatalos talaj korrelációs rendszerét. A rendszer a talajosztályozás diagnosztikai szemléletű megközelítésén alapszik, határozott definíciókkal és mennyiségileg kifejezhető feltételekkel (IUSS Working Group WRB 2007). Ezzel ellentétben számos nemzeti osztályozási rendszer – beleértve a magyart is – eltérő elveket követ, így nem feltétlen rendelkeznek a WRB eredeti követelményeinek maradéktalanul megfelelő adatokkal.

A hazai archív talajadatok felhasználásával és digitális adatbázisokba rendezésével kapcsolatosan az elmúlt tíz évben számos publikáció született, ám a mai napig nem jött létre egy egységes, országos, harmonizált adatbázis, amely összhangban lenne a talajtan jelenleg elfogadott nemzetközi nevezéktanával, azaz a WRB-vel.

Kutatómunkám főbb célkitűzései az alábbiak voltak:

1. Felmérni a hazánkra legnagyobb területi és tematikus lefedettséggel rendelkező, archív adatokra épülő talajtani adatbázisok adatrendszerét és megfeleltetési lehetőségeiket a WRB-vel.
2. Korrelációs módszertan részletes kidolgozása a három legnagyobb országos lefedettséggel rendelkező adatbázisra a hazai talajadatok megfeleltetésére a WRB-vel.
3. A módszertan kidolgozása a korrelált adatok térbeli kiterjesztésére egy mintaterületen a digitális talajtérképezés eszközeivel.
4. Az alkalmazott módszerek kiértékelése, az eredmények ellenőrzése.
5. Módszertan fejlesztése a hazai archív talajadatok országos szinten alkalmazható integrációjára.

2. ANYAG ÉS MÓDSZER

Felhasznált anyagok

Mivel munkám elsődleges célja az volt, hogy országosan alkalmazható módszertant dolgozzak ki, így alapvetően három talajtani adatforrásra támaszkodtam:

- A Kreybig-féle Átnézetes Talajismereti Térképezés és a Digitális Kreybig Talajinformációs Rendszer (Kreybig)
- A Géczy-féle térképek adatai (Géczy)
- 1:10 000 méretarányú üzemi talajtérképezés adatai (Üzemi)

Az adatok térbeli vizsgálatára és a digitális térképezési munkához a Gyöngyösi Kistérség területére a fenti három adatforrásból származó, az MTA ATK TAKI által rendelkezésemre bocsátott adatbázisból indultam ki. Tekintettel arra, hogy a Géczy-térképek esetébe nem rendelkeztem pontadatokkal, a kategorikus térképi adatok krigelésben való felhasználása túlságosan leterhelte volna a számítási kapacitást, így végül az adatok elhagyása mellett döntöttem, azaz a végső digitális térképezési munkában csupán a maradék két adatforrás szerepelt tanuló adatsorként.

A digitális térképezés során a vizsgált WRB egységek térbeli elhelyezkedésének meghatározásához a talajadatokon túl domborzati és távérzékelési adatokat is felhasználtam.

A domborzati adatok kiválasztásánál is elsődleges szempont volt azok ingyenes hozzáférhetősége, így végül az SRTM adatok felhasználása mellett döntöttem.

A domborzati adatok mellett műholdas adatállományként a LANDSAT 5 Thematic Mapper(TM) felvételeit használtam fel. Ezek az adatok az előbbiekhöz hasonlóan ingyenesen rendelkezésre állnak. A Landsat TM 7 hullámhossz-sávban készít felvételeket, a hatost kivéve (120x120m) 30x30m-es felbontással:

- | | |
|--|--|
| 1. 0,45-0,52 μm (kék) | 5. 1,55-1,75 μm (közép-infravörös) |
| 2. 0,52-0,60 μm (zöld) | 6. 10,4-12,5 μm (termális-infravörös) |
| 3. 0,63-0,69 μm (vörös) | 7. 2,08-2,35 μm (közép-infravörös) |
| 4. 0,76-0,90 μm (közeli-infravörös) | |

Mivel a felvételek értékeit jelentősen befolyásolják a vegetációs viszonyok, ezért több, eltérő időpontban készült felvételt is felhasználtam, az alábbi dátumokkal:

- | | |
|---------------------------|------------------------------|
| • 2000. október 23. (ősz) | • 2003. március 22. (tavasz) |
| • 2001. február 1. (tél) | • 2007. szeptember 17. (ősz) |
| • 2002. június 23. (nyár) | |

A rendelkezésre álló adatokat figyelembe véve az archív adatokra alapozó modern talaj-felvételezés és térképezés két, alapvetően egymásra épülő szintjén vizsgáltam a WRB-megfeleltetési problémákat:

1. A talajszelvény szintjén
2. A talajtérképek szintjén.

Az első szint adatai többnyire táblázatos formában állnak rendelkezésre, így az alkalmazott módszertan csupán az adatbázis kezelés szintjén eltérő.

Talajszelvény adatok esetében a WRB követelményrendszerét is figyelembe véve kézenfekvő volt az alapjaiban Eberhardt & Waltner (2010) által a leíró jellegű talajadatokra javasolt WRB-algoritmusok módszerének részletes hazai kidolgozása archív adatokra. Ezek kialakításuk után az egyes adatbázisokra viszonylag könnyen alkalmazhatóak, így lehetővé téve a talajszelvény adatok tömeges WRB-harmonizációját.

Alkalmazott módszerek

Korrelációs algoritmusok (szelvény-adatok)

Tekintettel arra, hogy az archív talajadatoknak ma már jelentős része rendelkezésre áll digitális adatbázis formájában, munkám során három ilyen hazai adatbázis felvételezési módszertanából kiindulva vizsgáltam, hogy a bennük tárolt adatok milyen mértékben használhatók fel a nemzetközi harmonizációra. Vizsgálataim fókuszában a WRB állt, mint a talajtani korreláció nemzetközileg elfogadott eszköze.

A rendszer lényeges tulajdonsága, hogy a talajok különböző kategóriákba való elhelyezését konkrét értékekkel meghatározott, ún. diagnosztikai szintek, anyagok és tulajdonságok (diagnosztikai egységek, szakmai gyűjtőneveikön „diagnosztikák”) jelenlétéhez vagy kizásárához köti.

A WRB (IUSS Working Group WRB 2010) két fő szintet különböztet meg:

1. A referencia csoportok szintjén 32 fő talajtípust különít el, egy terepi, illetve laboratóriumi vizsgálatokra épülő kulcs alapján.
2. A minősítők (qualifiers) szintjén az egyes referencia csoportoknál előforduló további jellemzők (összesen 179) közül lehet választani, a megfelelő kritériumok teljesülése esetén. A rendszer megkülönböztet ún. elő- (prefix) és utó- (suffix) minősítőket.

A módszertan alapjait a Michéli és munkatársai (2011) által a WRB-re kidolgozott egyszerűsített algoritmusok képezik. Ezek kiemelték az egyes WRB egységeknél kritikus tulajdonságokat, melyeknek meghatározó szerepük van az adott egység meghatározásakor. Az algoritmusok hazai adaptálásakor a TIM módszertanát vették első sorban figyelembe.

A WRB 65 diagnosztikai szintjéből, tulajdonságából és talajanyagából 28 nem fordul elő hazai viszonyok között. Ezeket vizsgálataim során nem vettem figyelembe, mivel a hazai adatbázisok nem tartalmazzák a szükséges információkat.

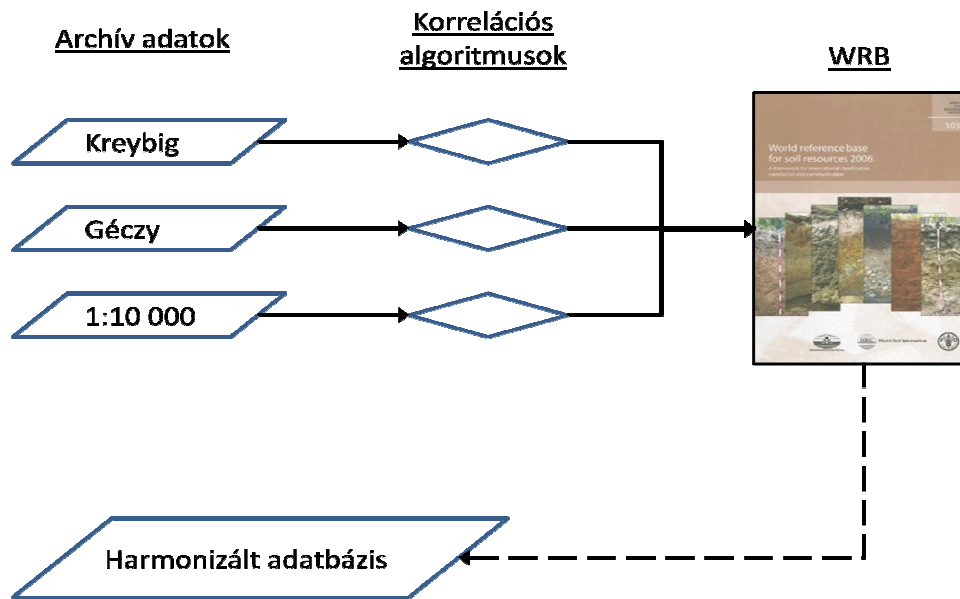
A fennmaradó 28 WRB diagnosztika kritériumaira, valamint 30, többnyire ezekből származtatható WRB minősítőre megfeleltetési módszertant dolgoztam ki mindhárom adatbázisra.

A három adatbázis eltérő felvételezési és mintázási útmutatót, valamint részben eltérő laboratóriumi módszereket használt.

A korrelációs szabályok kialakítása során bizonyos esetekben szakértői tapasztalatokat kellett figyelembe venni, hogy mérlegelni lehessen az egyes archív adatok beilleszthetőségét a mai adatbázisokba. Ennek következtében alkalmanként kénytelen voltam igen korlátozott adatokból származtatni WRB diagnosztikai egységeket („diagnosztikákat”) vagy minősítőket.

A kialakított korrelációs szabályok értékelésénél és felhasználásánál meg kell jegyezni, hogy ezek csupán „legjobb megközelítést” jelentenek a vizsgált hazai talajadatokra és semmiképpen sem tekinthetők egy az egyben megfeleltetésnek.

A kidolgozott algoritmusok módszertani szerepét szemlélteti az 1. ábra.



1. ábra A felállított korrelációs algoritmusok alkalmazásának folyamata

A kidolgozott algoritmusok segítségével példákon keresztül összehasonlítottam a mintaterületre rendelkezésre álló három archív adatbázist. Figyelmem középpontjában elsősorban a WRB egységek térbeli elhelyezkedése volt, illetve annak változása a különböző források alapján.

A térinformatikai adatok megjelenítésére és feldolgozására részben a SAGA geoinformatikai szoftvercsomagot (System for Automated Geoscientific Analyses 2008) használtam. A szoftver egyik előnye, hogy nyílt forráskódú, ingyenesen hozzáférhető, így a dolgozat eredményei könnyebben és jobban reprodukálhatóak. A SAGA másik nagy előnye, hogy számos, jól használható modullal rendelkezik a raszteres alapú, így domborzat- vagy távérzékelési adatok feldolgozására és vizsgálatára.

A harmonizált adatok térbeli kiterjesztésének vizsgálatához, olyan WRB egységekre volt szükség, amelyeket a lehető legtöbb adatbázisból lehet becsülni, valamint a vizsgált területen is rendelkeznek megfelelő térbeli változatossággal ahhoz, hogy a térképezési munka megfelelő színvonalon kiértékelhető lehessen.

A pontadatok térbeli eloszlásának ellenőrzésére a teljes random ponteloszlás módszert (complete spatial randomness, CSR) alkalmaztam.

Digitális térképezés

A segédváltozók előkészítése

Domborzati adatok

A domborzati adatok a digitális talajtérképezés során leggyakrabban alkalmazott környezeti változók, ám nem csupán a „nyers” magassági adatokat, hanem azokból számított további domborzati mutatókat is széles körben alkalmaznak (Dobos & Hengl 2009). Témámhoz

kapcsolódóan tehát logikus volt néhány ilyen mutató alkalmazása. A SAGA programcsomag moduljainak alkalmazásával a magassági adatokból az alábbi paramétereket számítottam:

- lejtőszög
- konvergencia index
- lejtőhajlás
- lefolyási viszonyok
- anyagmérleg
- nedvességi index
- vízhálózattól való függőleges távolság
- felszín érdességi index

Landsat műholdfelvételek

Tekintettel arra, hogy a Landsat 5 TM hét sávja egyszerre túl sok, egymással korreláló adatot hordoz, így főkomponens analízis alkalmazásával a 7 sávot 4 főkomponensre redukáltam.

A fentiek mellett a műholdfelvétélből számított NDVI növényborítási indexet is csatoltam külön réteggként a digitális talajtérképezést szolgáló adatbázishoz.

A fentiek alkalmazásával minden időpontra 5 db raszter réteg jött létre, azaz a Landsat felvételekből összesen $5 \times 5 = 25$ réteg került végül a digitális térképezési fázisba.

Geostatisztika, regresszió krigelés

A digitális térképezés eszköztárának egyik legszélesebb körben használt ága a geostatisztika, azon belül is kitüntetett szerepe van a krigelés különböző alkalmazásainak.

A geostatisztikai módszerek digitális talajtérképezési alkalmazása esetében a leggyakrabban azt jelenti, hogy a vizsgált talajtulajdonság térbeli változását mért adatok alapján, annak térbeli változását leíró, illetve a vele közvetlen vagy közvetett térbeli kapcsolatban álló (általában a mért adatnál könnyebben és olcsóbban hozzáférhető) környezeti változókat figyelembe véve kíséreljük meg leírni valamilyen matematikai modell segítségével. A modell felállítása után az adatok a környezeti változók alapján olyan helyekre is becsülhetőek, ahol az eredeti tulajdonságot nem vizsgálták.

A krigelés, mint interpolációs módszer lényege, hogy a vizsgált tulajdonság értékét egy adott s_0 pontra úgy közelíti, hogy az ismert (s_1, s_2, \dots, s_i) pontokban ismert értékek alapján olyan lineáris interpolációs függvényt keres melynek segítségével (az ismert pontok megfelelő súlyozásával) a becsült pontra vonatkozó variancia a lehető legkisebb.

Számos geostatisztikai módszer alkalmazható környezeti változók térképezésére. Egy rendkívül széles körűen alkalmazható ezek közül a regresszió krigelés (RK), melynek speciális eseteként felfogható a legtöbb egyéb krigelési módszer (Hengl 2011).

A regresszió krigelés lényege, hogy a térbeli eloszlási függvény determinisztikus összetevőjét (trendet) regresszió analízis segítségével meghatározzuk a segédváltozók alapján, majd a determinisztikus részt eltávolítva csak a reziduális hibára számítjuk a variogramot, és a determinisztikus részt utólag illesztjük a krigelt értékekre.

A regresszió krigelést összesen 34 bemeneti segédváltozó adatréteggel végeztem (20 Landsat főkomponens, 5 NDVI, 9 domborzati jellemző), két WRB egységre: a **calcaric anyagra** és a **clayic minősítőre**. Míg előbbi elsősorban a talaj kémiai állapotáról (CaCO_3 tartalom), utóbbi a fizikai féleségről (agyagtartalom) nyújt mezőgazdasági szempontból kiemelkedően fontos információt.

A jelen dolgozathoz alkalmazott regresszió krigelés módszertana javarészt a Hengl (2011) által leírt módszertant veszi alapul, ám azt több ponton módosítottam, illetve kiegészítettem az adatok és a célkitűzések figyelembevételével.

A térinformatikai elemzéseket a SAGA programcsomaggal, míg a statisztikai, geostatistikai vizsgálatokat az R szoftvercsomag (R Development Core Team 2012) moduljaival végeztem.

A felhasznált bemeneti adatokat és felhasználásukat egyszerűsítve a 2. ábra foglalja össze.

A pontos munkamenet az alábbi volt:

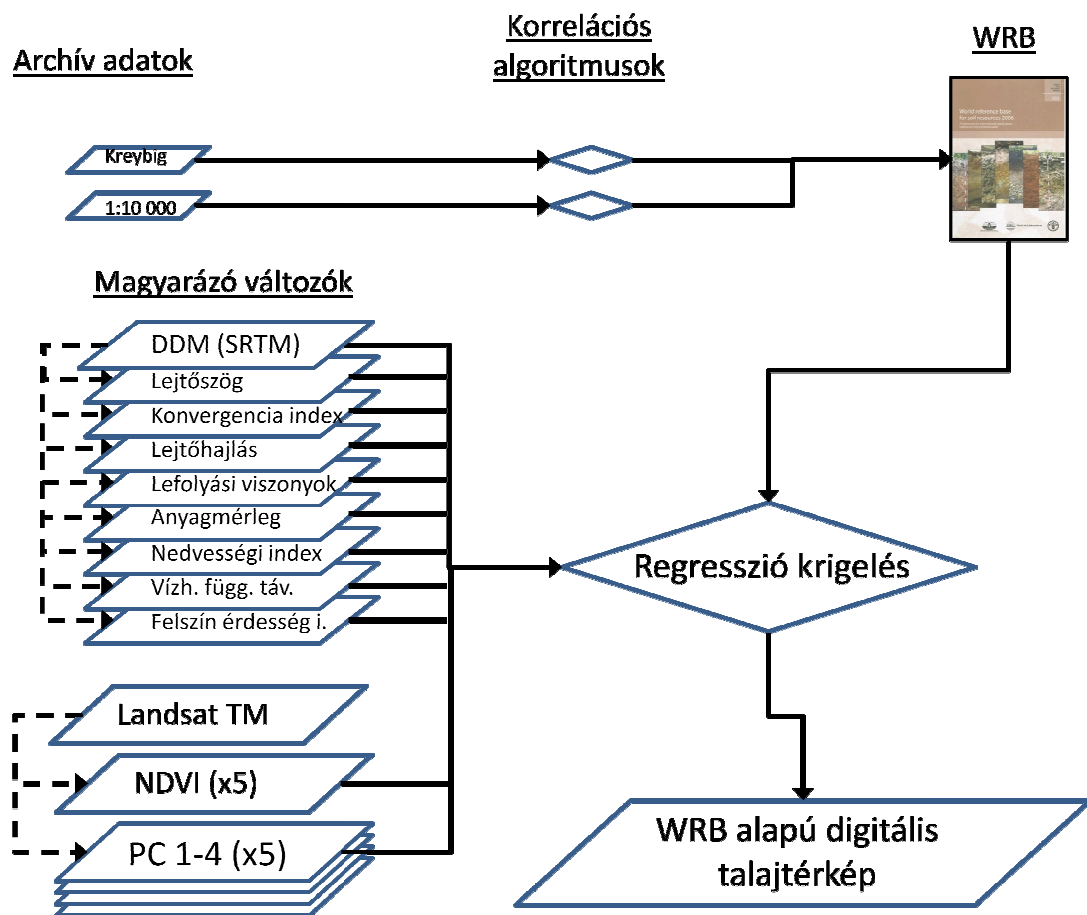
1. A raszter állományok átalakítása egységes, 100 x 100 m cellaméretre.
2. Az adatok szétválasztása tanuló és teszt populációra
3. A felhasznált adatok előzetes statisztikai vizsgálata
4. A segédváltozók multikollinearitásának csökkentése, a változók helyett főkomponensek számításával.
5. Logisztikus regressziós modellt illesztettem a tanuló populációra a főkomponensek felhasználásával mindkét vizsgált WRB egységre.
6. Lépcsős (stepwise) regresszió alkalmazásával csökkentettem a vizsgálandó főkomponensek számát.
7. Variogram modellt illesztettem a reziduális hibákra, valamint az egyszerű krigelés kivitelezéséhez az alapadatokra is.
8. Elvégeztem magát a krigelést mindkét esetben (OK és RK).
9. Az adatokat exportáltam térinformatikai programok által hozzáférhető formátumba.
10. Az eredményeket mind térinformatikai, mind statisztikai módszerekkel kiértékeltem.

Az eredmények ellenőrzése

Az eredmények statisztikai vizsgálatát Baldi és munkatársai (2000) alapján, az eredeti adatokból elkülönített tesztadatok és a hat TIM pont felhasználásával végeztem.

A teszt- és forrásadatok alapján öt különböző hozzátartozási határérték esetén meghatároztam a találati arányok átlagából képzett $Q\alpha$ mutatót, valamint a Pearson-féle korrelációs együtharót.

A TIM esetében a tesztszelvények alacsony száma miatt csupán találati táblázat felállítására volt lehetőség.



2. ábra A digitális térképezés bemeneti paramétereit és felhasználásuk sematikus ábrája

3. EREDMÉNYEK

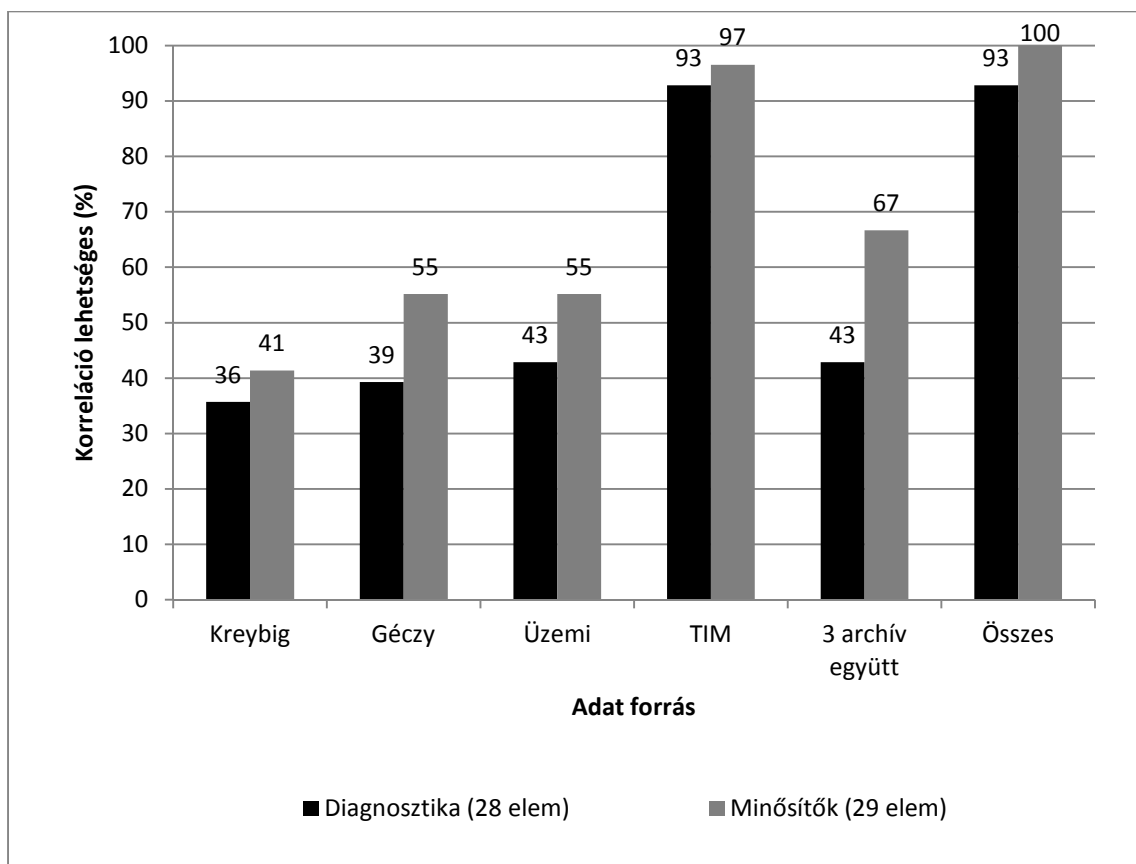
Az archív adatok WRB-megfeleltetése

Munkám során vizsgáltam, hogy a kifejlesztett korrelációs szabályok felhasználásával milyen arányban lehet információt kinyerni az archív adatokból a WRB diagnosztikák, illetve minősítők meghatározására.

Számos esetben a konkrét adattípus nem állt rendelkezésre, ugyanakkor voltak olyan egyéb paraméterek (kategorikus, vagy származtatott adatok), amelyek segítségével – ha egzakt, numerikus módon nem is, de – következtetni lehet az adott tulajdonságra. Az adatokkal kapcsolatos problémák közül a leginkább jellemző a mélységi értékek korlátozott száma volt. Mindhárom „archív” adatbázis adatainak egy részét feltalaj-altalaj felosztásban kezelte, ami a WRB precíz mélységi kritériumainak való megfeleltetést sok esetben megnehezíti vagy kizárja.

Összesítésként a 3. ábra bemutatja, hogy a vizsgált 28 WRB diagnosztikára, illetve 29 minősítőre milyen arányban volt lehetséges a megfeleltetés. A jobb áttekinthetőség kedvéért a felhasznált adatbázisok a felmérések időpontja szerint vannak felsorolva.

Az ábrában megfigyelhető egyfajta időben emelkedő trend a WRB-be beilleszthető adatok mennyiségét tekintve. Ennek valószínűsíthető magyarázata a módszertanok fejlődése. Jól látható továbbá, hogy míg önmagában egyetlen adatbázis sem képes információt szolgáltatni valamennyi Magyarországon várható WRB egységre, kombinációjukkal – főleg a minősítők esetében – lényeges javulás érhető el.



3. ábra Az egyes adatbázisok által elérhető „legjobb” megfelelés a WRB követelményeinek

Az egyes adatbázisokra kidolgozott algoritmusok WRB egységenként

Az alábbiakban bemutatom a Michéli és munkatársai által kidolgozott (2011) egyszerűsített harmonizációs algoritmusok alapján az egyes WRB egységekre kidolgozott döntési folyamatokat mind a három vizsgált adatbázis esetén. Megfelelő bemutatásukhoz elengedhetetlen a WRB adott definíciójának ismertetése, így ezek saját, nem hivatalos fordításai az eredeti dokumentumnak (IUSS Working Group WRB 2007). Minden esetben készült döntési fa is, ám tézisekben csak a calcic szint példáját mutatom be.

Tekintettel arra, hogy az eltérő felvételezési módszertanokból az adatok típusán túl további hibaforrások is következhetnek – különböző kategóriahatárok, laboratóriumi módszerek – ideális esetben szükség lenne az alábbi következő módszerek numerikus megbízhatóságának leírására. Sajnos ez az adatok jellegéből adódóan a jelenlegi hazai körülmények között nem megvalósítható, mivel az adatok időben és módszertanban annyira eltérnek egymástól, hogy egységes, az időbeli változásokat az 1900-as évektől követő monitoring rendszer, vagy egy országos újrafelvételezés hiányában csupán kvalitatív jellegű meghatározásuk lehetséges. Ezen okból valamennyi esetben (több döntési ág esetén többször) nagybetűs, vastagon szedett, zárójelbe zárt „megbízhatósági mutató” jelzi az egyes algoritmusok felhasználhatóságát az alábbiak szerint:

- (MAGAS)** – Az adott döntés nagy valószínűséggel pontosan becsli az adott WRB-egység jelenlétét. Ezen esetekben általában rendelkezésre állnak a szükséges laboratóriumi, illetve terepi adatok, a döntés elsősorban ezekre támaszkodik.
- (KÖZEPES)** – A döntés közepes megbízhatóságú, a vizsgált WRB-egység meghatározásánál kisebb hibák előfordulhatnak, mivel a döntés kisebb-nagyobb mértékben szakmailag megalapozott becslésekre épül.
- (ALACSONY)** – A döntésnél az adatok hiányából adódóan szinte kizárólag becslésekre lehet támaszkodni, a felhasznált kategóriák nem minden esetben mutatják egyértelműen az adott WRB kritériumok teljesülését. Alkalmazásuk fokozott körültekintést igényel!

Az egyes WRB egységeknél csak azokat az archív adatbázisokat soroltam fel, melyeknél rendelkezésre állnak a meghatározáshoz szükséges információk.

Calcic szint (PÉLDA)

WRB:

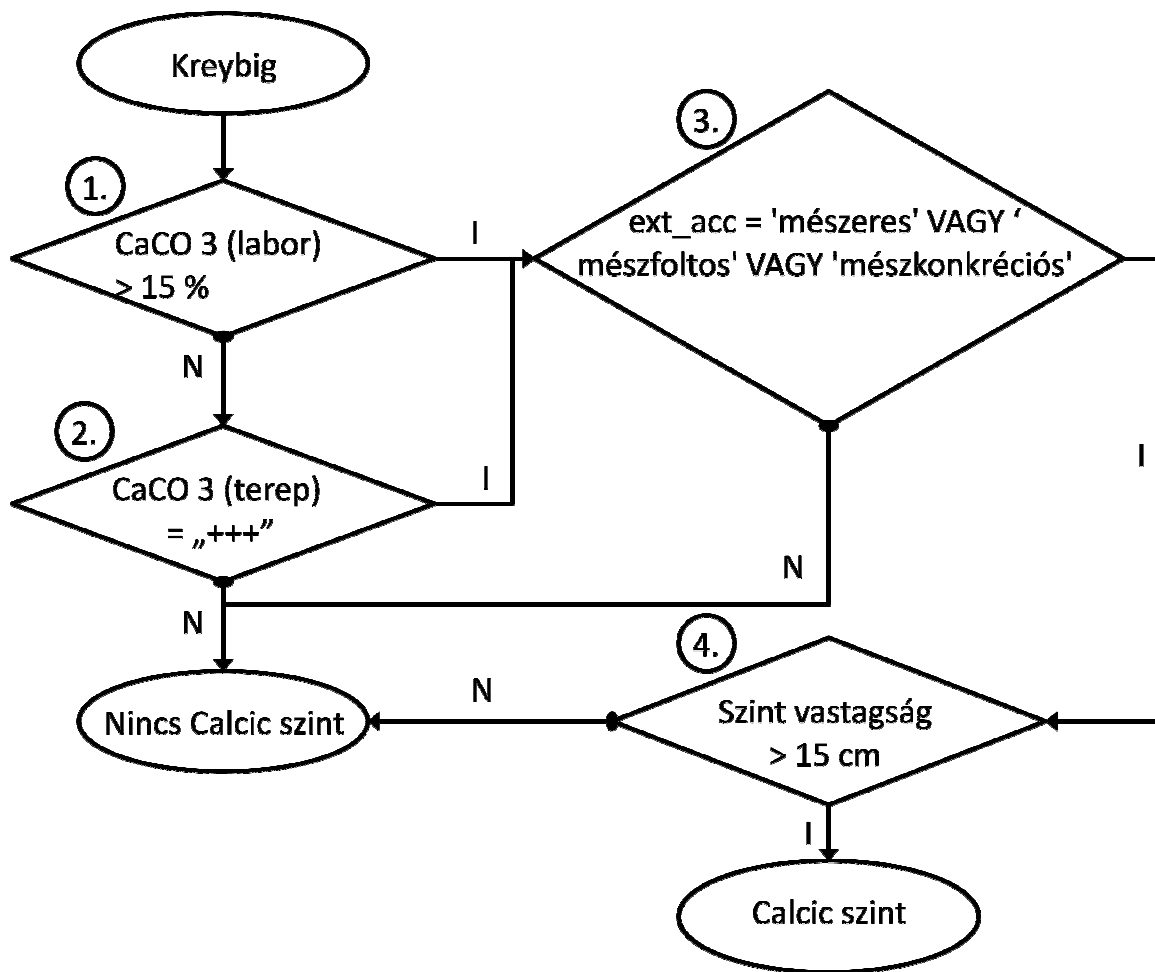
Egy calcic szint rendelkezik:

- 1. a földes részben mért legalább 15 százalékos kalcium karbonát egyenértékkel*
és
- 2. legalább öt (térfogat)százalék másodlagos karbonáttal vagy legalább 5 (abszolút, tömeg) százalékkal több kalcium karbonát egyenértékkel mint az alatta fekvő réteg*
és
- 3. legalább 15 cm vastagsággal.*

Kreybig (PÉLDA)

Tekintettel arra, hogy az eredeti felvételezés során is történtek laboratóriumi mérések, azoknál a szelvényeknél ahol ezek rendelkezésre állnak, a CaCO₃ tartalomról (4.ábra 1. döntési pont) közvetlenül becsülhető a calcic szint jelenléte (**MAGAS**), ám az eredmények felhasználásakor

figyelembe kell venni, hogy az eltérő laboratóriumi módszerekből hibák következhetnek. Amennyiben az adott szelvényhez nem készültek laboratóriumi vizsgálatok akkor általában terepi, HCl csepegtetési módszerrel vizsgált adatok vannak. A „+++” jelölésű szelvényeknél előfordulhat a calcic szint, mivel ez több, mint 8% karbonát jelenlétére utal (4. ábra 2. döntési pont). Ugyanakkor ez a becslés szignifikáns hibát eredményezhet, ha nem kellő körültekintéssel alkalmazzuk, hiszen a kategória az eredeti követelménynél lényegesen alacsonyabb értékeket is magába foglal (**ALACSONY**). A másodlagos karbonátok jelenlétére a mészerek, mészfoltok és mészkonkréciók jelenlétéből következtethetünk (4. ábra 3. döntési pont). A vastagsági kritériumot általában a mintázott talajszint vastagsága alapján, illetve több, a kritériumoknak megfelelő, közvetlenül egymás alatt/felett található szint együttes vastagságából számíthatjuk (4. ábra 4. döntési pont).



4. ábra Calcic szint meghatározása a Kreybig adatok alapján (PÉLDA)

Az előzetes elemzés eredményei

A mintaterületre rendelkezésre álló pontadatok tanuló, illetve tesztpopulációra való szétválasztásánál az adatok tíz százalékát különítettem el. A két részhalmaz mindkét vizsgált WRB egység esetén megőrizte az eredeti eloszlást, így reprezentatívnak tekinthetőek a teljes populációra.

A tanuló adatok eloszlásának vizsgálatára meghatározott átlagos távolság számítása alapján. Ebből megállapítható, hogy az egyes pontok legközelebbi szomszédainak átlagos távolsága 277 m. Ugyanakkor a minimum és maximum közötti különbség igen nagy. Ez többek között következhet a két eltérő méretarányú térképezés pontjainak kombinációjából, valamint a domb- és hegyvidéki, illetve erdős területek hiányos felvételezéséből. Előbbi magyarázhatja a minimum, míg utóbbi a maximum értéket.

A mintaterületre elkülönített tanuló populációra elvégzett térbeli eloszlás (CSR) vizsgálat alapján megállapítható, hogy a pontok körülbelül 150 m-es távolságon belül jellemzően véletlenszerű eloszlást mutatnak, míg e fölött a pontok csoportosulása figyelhető meg. Ennek több, egymással esetenként összefüggő magyarázata is lehetséges. Az üzemi térképezés pontjainak a térképezés méretaránybeli különbségéből adódóan nagyobb a sűrűsége. A terület egyes részein – többnyire domborzati viszonyok, illetve területhasználati, vagy egyéb okokból – a térképezés során nem készült felvételezés, vagy nem megfelelő pontsűrűséggel. Ugyanakkor ezek a területek többnyire ma sincsenek mezőgazdasági művelés alatt, így a talajtérképek felhasználhatósága szempontjából kritikus területekre a pontok eloszlása és sűrűsége megfelelő.

A pontadatokat a segédváltozók raszteres állományaira helyezve lehetővé vált utóbbiak eloszlásán túl a tulajdonságterek vizsgálata is. Erre azért volt szükség, hogy értékelhessem a segédváltozók pontbeli adatai mennyire tekinthetők reprezentatívnak a teljes területre vonatkoztatva. Ehhez ún. back-to-back hisztogramokon ábrázoltam valamennyi segédváltozó eloszlását a pontok és a teljes mintaterület szerint. Míg általánosságban az eloszlások jellege megegyezett, változók többségénél megfigyelhetők olyan területek, melyek többnyire alul- illetve felül reprezentáltak a mintában.

A pontokon és a teljes területen vizsgált eloszlások összehasonlítására Kolmogorov-Szmirnov tesztet végeztem valamennyi segédváltozóra. Az eredmények tükrözik a hisztogramok vizuális vizsgálata alapján levont következtetéseket, szinte valamennyi eloszlásnál kimutatható statisztikailag szignifikáns különbség.

A domborzati információkat hordozó változóknál figyelhető meg elsősorban az eltérés, ez megerősíti a korábbi feltételezést, miszerint a hegy- és dombvidéki területekre az adatbázis nem rendelkezik elegendő információval.

Ez elsősorban azt mutatja, a hazai archív talajadatok digitális térképezési célú felhasználásához a mintaterület alapján a mezőgazdasági térképezések adatain túl várhatóan további, elsősorban erdészeti talajinformációkra is szükség lehet, amennyiben maximalizálni kívánjuk a területi lefedettséget.

Az adatok ugyanakkor felhasználhatóak, figyelembe véve, hogy várható eredményeik az alulreprezentált területeken kevésbé lesznek felhasználhatóak.

A digitális térképezés eredményei

A térképezés eredményei

A krigelés során a calcaric anyag esetén exponenciális, a clayic minősítőnél gauss típusú elméleti variogramot illesztettem az adatokra.

A calcaric anyagra készült térképek vizuális interpretációja alapján látható, hogy a tesztpopuláció calcaric anyagot tartalmazó pontjai többnyire azokra a területekre esnek, ahol a modell magasabb valószínűséggel jósolta jelenlétüket, míg a calcaric anyagot nem tartalmazó szelvények az alacsonyabb valószínűségű területekre esnek az esetek többségében.

Ugyanakkor a valószínűségi értékek általában alacsonyak, ez következhet abból is, hogy a felhasznált segédváltozók között nem szerepelt semmilyen, a talajképző kőzetre információt hordozó réteg.

A varianciát figyelembe véve látható, hogy azokon a területeken, ahol a tanuló populáció pontsűrűsége magasabb volt, a variancia lényegesen alacsonyabb, mint a pontokban szegény, magasabban fekvő, illetve peremterületeken.

Ugyanakkor feltűnő, hogy Domsztló térségében a nagyméretarányú térképezés pontjainak hiánya nem csökkentette szignifikánsan a modell varianciáját, a pontok eredeti eloszlásán ugyanis látható, hogy a Kreybig-féle pontok sűrűsége valamivel magasabb az átlagosnál azon a területen, így az egyes adatbázisok kompenzálhatják egymás hiányosságait.

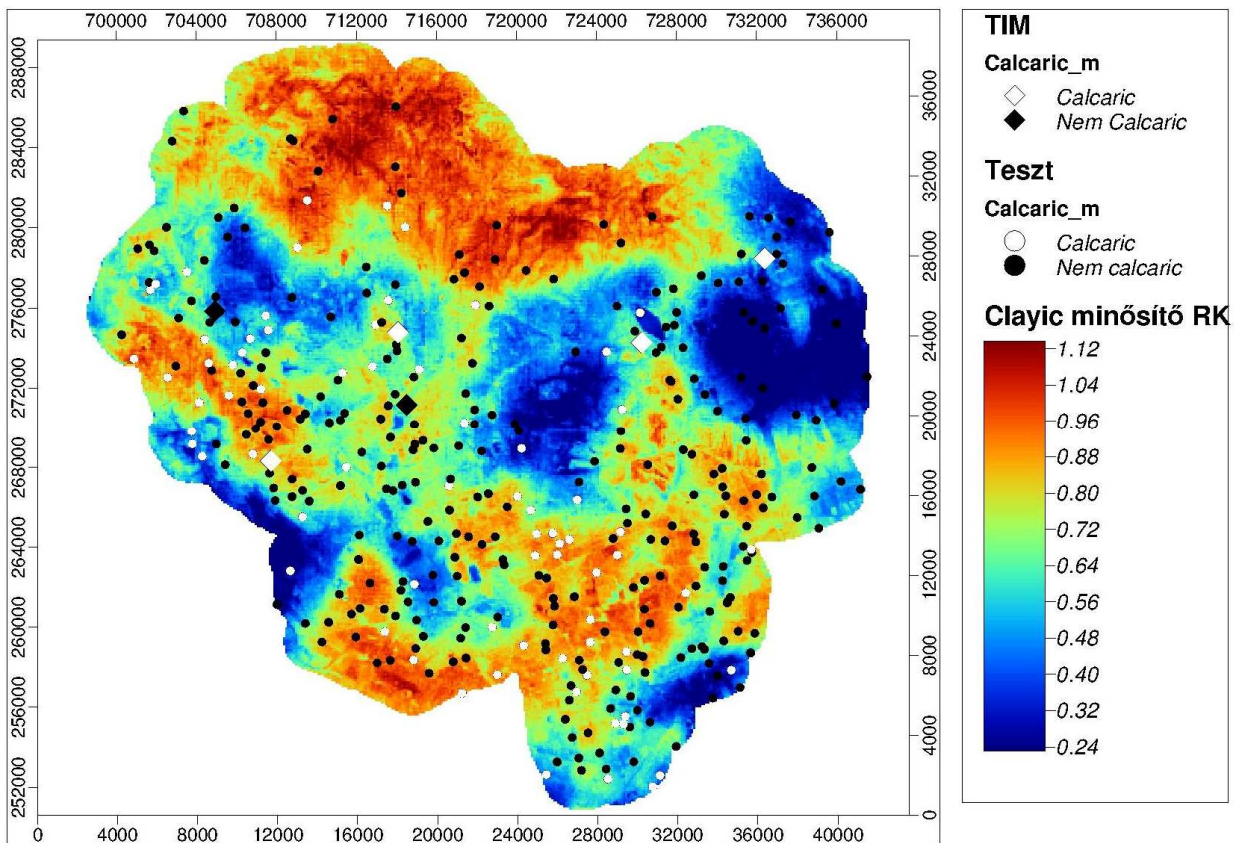
A regresszió krigelés és az általános krigelés eredményeit összevetve megfigyelhető, hogy bár térbeli mintázatuk hasonló képet mutat, előbbi sokkal részletesebb, kevésbé elnagyolt képet ad a területről, ugyanakkor varianciája valamivel magasabb. A variancia értékek térbeli eloszlása lényegében megegyezik, ami várható, hiszen mindkét modell ugyanazon pontok alapján dolgozott.

A clayic minősítő esetén látható, hogy mindkét modell magasabb valószínűségi értékeket adott meg. Figyelembe véve az adatok előzetes vizsgálatát, a magas agyagtartalmú talajok dominanciája alapján várható volt a terület jelentős részén a clayic minősítő jelenléte. Az 5. és 6. ábra regresszió krigelés eredményét, illetve varianciáját mutatja a clayic minősítő esetén

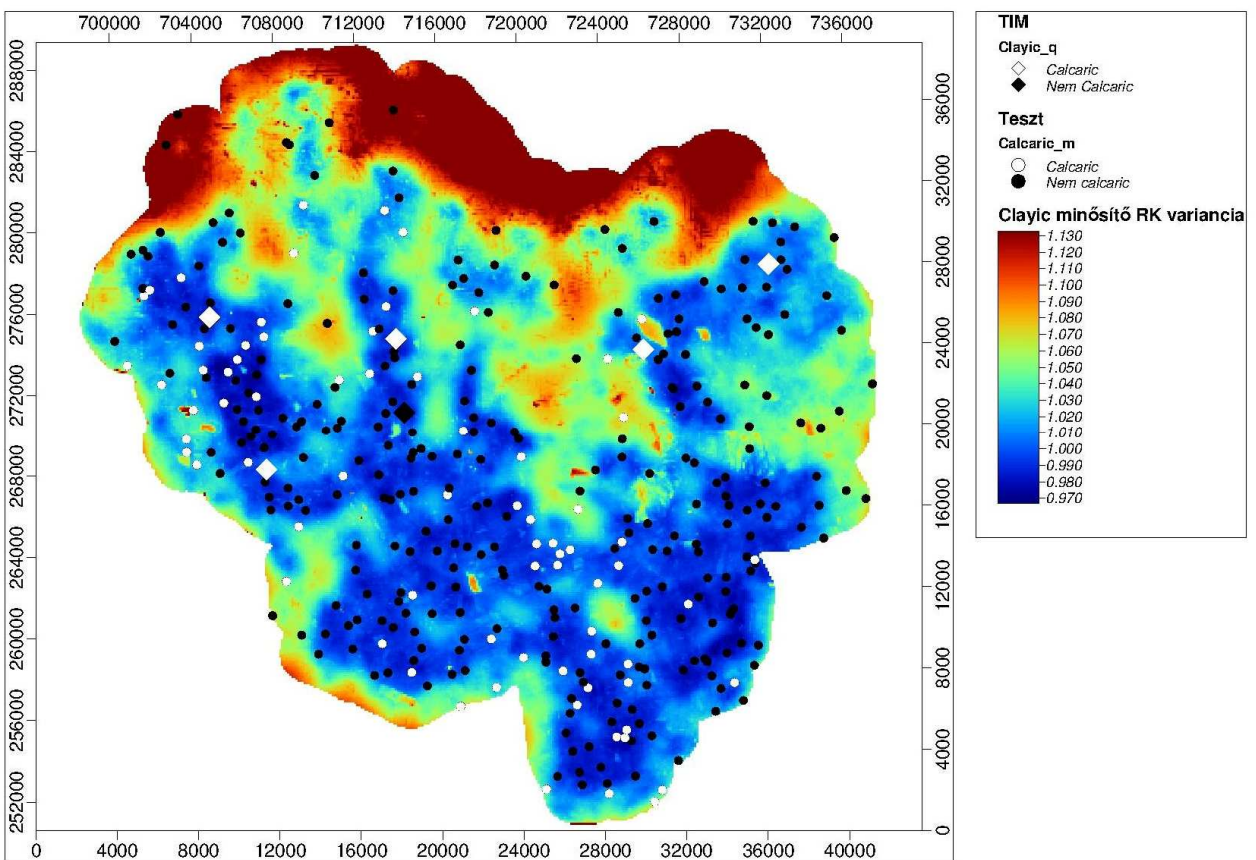
Az eredmények numerikus vizsgálata

Tekintettel arra, hogy a krigelt térképek előfordulási valószínűséget határoznak meg, az eredmények kiértékelésekor az elmosódott (fuzzy) halmazok módszertanához hasonlóan alkalmazhatóak hozzátartozási függvények, melyek a felhasználó igényei szerinti valószínűség felett jósolják egy adott feltétel teljesülését. Mindezek alapján az általános, 50%-os határérték mellett további, szigorúbb hozzátartozási értékeket is meghatároztam (60, 70, 80, 90 %)

A tesztadatok alapján, a modellek első- és másodfajú hibái alapján számolt Q_u értékek alapján látható, hogy a regresszió-krigelés – bár nem szignifikáns mértékben, de – az esetek többségében jobban teljesített, mint az általános, pusztán a minták térbeli elhelyezkedését vizsgáló krigelési eljárás.



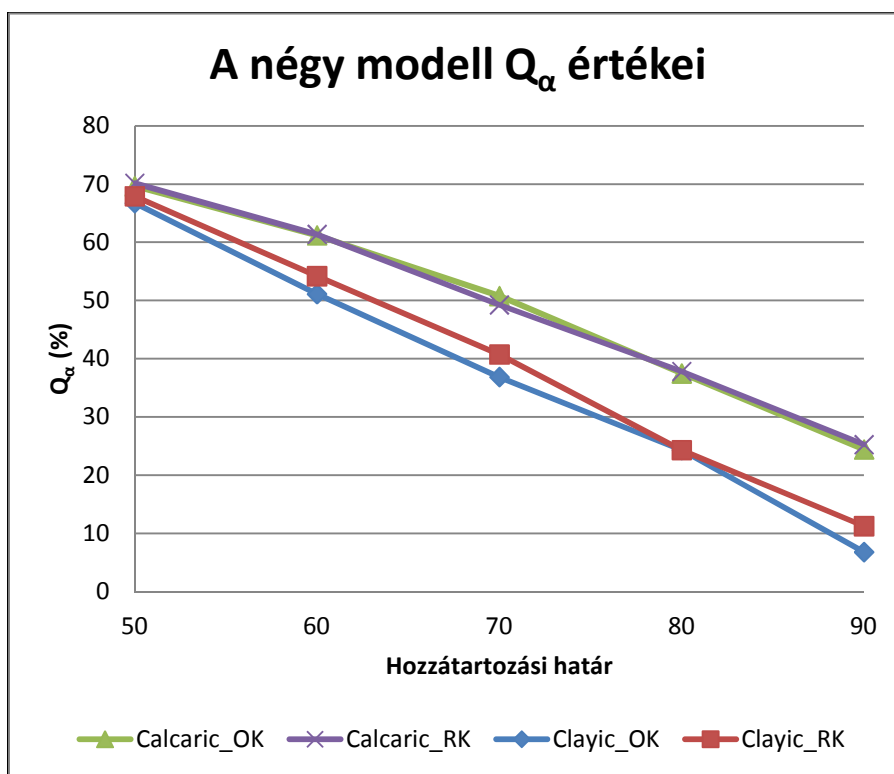
5. ábra A Clayic minősítő térbeli eloszlása regresszió krigelés alapján Gyöngyös térségében



6. ábra A regresszió krigelési modell varianciája a clayic minősítő meghatározásánál

A Q_α értékek változását mutatja be a 7. ábra melyen látható, hogy a calcaric anyag esetén a két módszer szinte azonos találati aránnyal bírt. Ezzel szemben a clayic minősítő esetében a regresszió krigelés szinte valamennyi hozzátartozási szinten láthatóan jobban teljesített.

A Pearson-féle korrelációs együttható megerősíti a Q_α mutatót, a regresszió krigelés mindkét esetben jobban teljesített, bár nem szignifikáns különbséget felmutatva (1. táblázat).



7. ábra A négy modell Q_α értékei a hozzátartozási szintek függvényében

1. táblázat A krigelés eredményeinek korrelációja a tesztpopulációval

Pearson-féle korrelációs együttható		
Calcaric	OK	RK
Teszt adat	0,5402	0,5519
Clayic	OK	RK
Teszt adat	0,4384	0,4771

Tekintettel arra, hogy a TIM szelvényeiből csak hat található a mintaterületen, az eredmények statisztikai kiértékelése nem volt lehetséges. Ugyanakkor megfigyelhető, hogy mindkét modell lényegesen jobban teljesített a clayic minősítő esetében, ez magyarázható a magas agyagtartalmú talajok nagyobb rarányával a területen.

Új tudományos eredmények

1. A rendelkezésre álló adatok alapján nemzetközi szinten is újszerű megfeleltetési módszertant adaptáltam a Kreybig, a Géczy és a Nagyméretarányú országos talajtérképezések által felhalmozott adatokra, így megkönnyítve azok integrációját más nemzetközi, a nemzetközi Talajtani Unió terminológiáját alkalmazó adatbázisokkal.
2. Megállapítottam, hogy az archív talajadatok WRB célú felhasználhatóságak időbeli változása emelkedő trendet mutat. (A WRB-megfelelés szerinti rangsor 1. TIM; 2. Üzemi; 3. Géczy; 4. Kreybig)
3. A módszertan példáján bemutattam, hogy az egyes adatbázisok alkalmazása helyett azok kombinációja jelentősen megnövelheti a harmonizációs munkák várható eredményességét.
4. A módszertan alapján digitális térképezési módszert alkalmaztam a WRB egységeinek közvetlen becslésére eltérő módszertanú és felvételezési idejű archív talajadatbázisok alapján.

4. KÖVETKEZTETÉSEK ÉS JAVASLATOK

Az eredmények rámutatnak, hogy az országosan legnagyobb területi lefedettséggel rendelkező, elsősorban archív adatokra épülő adatbázisok önmagukban nem elégítik ki teljes körűen napjaink adatigényeit.

Megfigyelhető egyfajta emelkedő trend az archív adatok WRB-célú felhasználhatóságában, ahogy közelebb jutunk napjainkhoz. Egyértelműen látszik ugyanakkor, hogy az archív talajadatok kombinált felhasználása jelentősen (esetünkben akár 24 százalékkal is) megnövelheti a harmonizációs munkák várható eredményességét.

Fontos megjegyezni, hogy jelen tanulmány nem tért ki a korlátozottan felhasználható adatok alkalmazásából következő magasabb hibalehetőségekre. Ebből adódóan az adatok felhasználásakor minden esetben figyelembe kell venni az eltérő módszertanokból és felvételi időpontokból következő pontatlanságot. Kutatómunkám eredményei alapján sikerült a hazai három legnagyobb talajtani adatbázis információinak korlátozott WRB-megfeleltetése.

Az elvégzett kutatómunka során bebizonyítottam, hogy az általam alkalmazott módszertan alkalmas lehet a hazai talajtani adatok nemzetközi megfeleltetésére. A megújuló hazai talajosztályozás ugyanakkor várhatóan az eddigieknél nagyobb hangsúlyt helyez majd a konkrét határértékekre a talajok besorolásánál, így várható, hogy bizonyos archív talajadatok esetén szükség lehet majd a jelen dolgozatban leírthoz hasonló módszertan alkalmazására (Michéli et al. 2009).

A digitális térképezési munka eredményeképpen megállapítható, hogy a regresszió krigelés alkalmazása a hagyományos krigelési eljárásnál jobb eredményeket adott a mintaterületen. Ugyanakkor az is kiderült, hogy a megfelelő segédváltozók kiválasztása kritikus lehet az egyes WRB egységek térképezésekor.

Talán a fentieknél is lényegesebb, hogy eredményeim alapján úgy érzem munkám nem csupán lehetőséget teremt a hazai adatok nemzetközi megfeleltetésére – hiszen ilyen irányú törekvések már évtizedek óta folynak hazánkban –, de talán rámutatott arra is, hogy egy hasonló módszertannal nem csak a WRB-be lehetne beintegrálni adatokat, hanem egy országos, valamennyi hazai talajtanos szakember javát szolgáló adatbázisba is.

Mindezt alátámasztani látszik, hogy a jelen dolgozatban bemutatott algoritmusok egy része felhasználásra került az OSIRIS (Observation and Spatial Information System for Soil Conservation – Talajvédelmi felvételezési és talajinformációs rendszer, osiris.helion.hu) rendszer létrehozásánál, amely egy a fentiekben felvázolt rendszer alapjait fekteti le (Szabó et al.2009).

5. FELHASZNÁLT IRODALOM

BALDI P., BRUNAK S., CHAUVIN Y., ANDERSEN C.A.F., NIELSEN H. (2000): Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics Review* Vol. 16, No. 5, 412-424 p.

DOBOS E., HENGL T. (2009): Soil Mapping Applications 461-479. p. In: HENGL T., REUTER H.I. (Szerk.): *Geomorphometry – Concepts, Software, Applications* [Developments in Soil Science, vol. 33] Elsevier 765 p.

EBERHARDT E. & WALTNER I. (2010): Finding a way through the maze – WRB classification with descriptive data. [5–8. p.] In: GILKES, J. R. & PRAKONGKEP, N (szerk.): *Soil Solutions for a Changing World: Proc. 19th World Congress of Soil Science, 1–6 August 2010, Brisbane, Australia 5–8*. International Union of Soil Sciences. Brisbane.

HENGL T. (2011): *A Practical Guide to Geostatistical Mapping, Second Edition*, University of Amsterdam. 291. p.

IUSS WORKING GROUP WRB (2007): *World Reference Base for Soil Resources 2006, update 2007, 2nd ed. World Soil Resources Reports 103*. FAO. Rome.
http://www.fao.org/ag/agl/agll/wrb/doc/wrb2007_corr.pdf, 21.01.2010).

MICHÉLI E., ET AL. (2009) *A hazai talajosztályozás korszerűsítése és nemzetközi megfeleltetése*. Munkabeszámoló. OTKA.

MICHÉLI, E. ET AL. (2011): Deliverable D5 – A soil data base for the 1:1 million scale windows. WP1 and WP2 report of the „e-SOTER – Regional pilot platform as EU contribution to a Global Soil Observing System” Project. EU 7th Framework Programme. Project No. 211578.

R DEVELOPMENT CORE TEAM (2012): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>.

SYSTEM FOR AUTOMATED GEOSCIENTIFIC ANALYSES (SAGA GIS) (2008) by SAGA Development Team

SZABÓ J., MICHÉLI E., PÁSZTOR L., DOBOS E., BAKACSI ZS., DOMBOS M. (2009): Elaboration of the OSIRIS framework for new, cost-effective soil survey and monitoring programs in Hungary. [17.p.] In.: *Bridging the Centuries 1909-2009 Budapest, From the Dokuchaev School to numerical soil classifications, Workshop Programme and Abstracts* (http://www.talaj.hu/BtC/09_Conf_poster/Dokuchaev_abstracts.pdf)

6. AZ ÉRTEKEZÉS TÉMAKÖRÉHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK

Folyóiratokban megjelent közlemények

Láng, V., Fuchs, M., Waltner, I. & Michéli, E., 2013. Taxonomic distance metrics, a tool for soil correlation : As exemplified by the Hungarian Brown Forest Soils and related WRB Reference Soil Groups. *Geoderma* **192** 269-276 p.

Impact factor: 2,318

Láng, V., Fuchs, M., Waltner, I. & Michéli, E., 2013. Taxonomic distance measurements applied for soil correlation *Agrokémia és Talajtan* **59**(1), 57-64. p.

Független idézettség:

Zádorová, T. & Penížek V., 2011. Problems in correlation of Czech national soil classification and World Reference Base 2006. *Geoderma* **167-168** 54-60 p.

Fuhs Márta, Waltner István, Szegi Tamás, Láng Vince és Michéli Erika, 2011: A hazai talajtípusok taxonómiai távolsága a képződésüket meghatározó folyamattársulások alapján. *Agrokémia és Talajtan*. **60** 33-44 p.

Waltner I., Fuchs M., Láng V. és Michéli E., 2012: Hazai archív talajadatok beillesztésének lehetőségei nemzetközi adatbázisokba. *Agrokémia és Talajtan*. **61**(2) 263-276 p.

Konferencia kiadványok

Eberhardt E. and Waltner I., 2010: Finding a way through the maze – WRB classification with descriptive data In: Gilkes RJ, Prakongkep N, editors. Proceedings of the 19th World Congress of Soil Science; Soil Solutions for a Changing World; ISBN 978-0-646-53783-2; Published on DVD; <http://www.iuss.org>; Symposium WG 1.1; The WRB @evolution; 2010 Aug 1-6. Brisbane, Australia, IUSS; 2010, pp. 5-8.

Láng V., Fuchs M., Waltner I. and Michéli E., 2010: Pedometrics applications for correlation of Hungarian soil types with WRB. In: Gilkes RJ, Prakongkep N, editors. Proceedings of the 19th World Congress of Soil Science; Soil Solutions for a Changing World; ISBN 978-0-646-53783-2; Published on DVD; <http://www.iuss.org>; Symposium WG 1.1; The WRB @evolution; 2010 Aug 1-6. Brisbane, Australia, IUSS; 2010, pp. 21-24