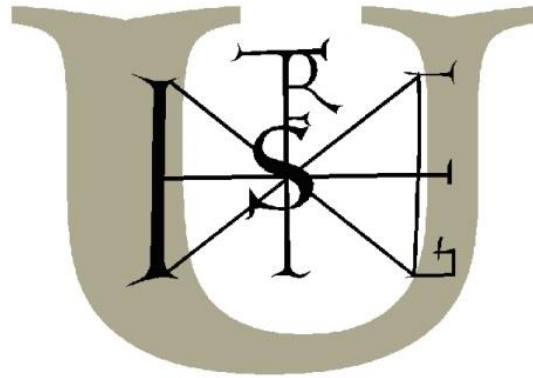


Szent István Egyetem  
Környezettudományi Doktori Iskola



A hazai talajvédelmet és a nemzetközi megfeleltetést  
szolgáló adatrendszer fejlesztése

Doktori (Ph.D.) értekezés

Láng Vince

Gödöllő

2013

**A doktori iskola megnevezése:** Környezettudományi

**tudományága:** Talajtan, agrokémia, környezeti kémia

**vezetője:** Csákiné Dr. Michéli Erika  
egyetemi tanár  
Szent István Egyetem  
Mezőgazdaság- és Környezettudományi Kar  
Környezettudományi Intézet  
Talajtani és Agrokémiai Tanszék

**témavezető:** Csákiné Dr. Michéli Erika  
egyetemi tanár  
Szent István Egyetem  
Mezőgazdaság- és Környezettudományi Kar  
Környezettudományi Intézet  
Talajtani és Agrokémiai Tanszék

.....  
Az iskolavezető jóváhagyása

.....  
A témavezető jóváhagyása

---

## A munka előzményei, és a kitűzött célok

---

A klímaváltozás, a népességnövekedés nehezen előre jelezhető terheket ró mind a bolygóra, mind a lakosságra. A fenntartható fejlődésnek, a modern talajkímélő technológiák, a környezetben bekövetkező negatív változások időbeni előrejelzése lehet a kulcsa. Ezekhez elengedhetetlen, hogy környezetünkről, a rendelkezésre álló információt összegyűjtsük, rendszerezzük, majd ezeket sikeresen alkalmazzuk a tervezésben, gyakorlatban. A talajra vonatkozó információ jelentős szerephez kell, hogy jusson ezekben a munkákban. Várhatóan azonban az információ nagy része nem új felvételezésből fog származni, annak költség és időigénye miatt. Az adatigény kielégítésére az archív adatokban lévő potenciált kell kihasználni és azokat kell összegyűjteni, harmonizálni. A nemzetközi trendek, Európai uniós direktívák és harmonizálási projektek, mint az INSPIRE direktíva (Európa Tanács, 2007), az eSOTER EU FP7-s program (eSOTER, 2008a,b), a GSSoil eContentplus program (Feiden, 2012), a Nemzetközi Talajtani Unió „Universal Soil Classification System” munkacsoportjának megalakulása, mind ezt célozzák kielégíteni. A fent említett projektek mindegyike magyar részvétellel, közreműködéssel kerültek kivitelezésre.

E trendeknek megfelelően elengedhetetlen az ország archív adatállományának összegyűjtése, harmonizálása, rendszerbe foglalása, annak érdekében, hogy az Európai Uniós igényeket, napjaink technológiai igényeit, társtudományokat, és természetesen a hazai talajtani kutatást szolgáljuk, azokat megfelelő adatokkal lássuk el.

Hazánkban jelenleg még kifejlesztésre vár egy olyan adatstruktúra modell, mely képes a több forrásból származó általános talajadatokat egy rendszerben tárolni és azokat a nyilvánosság számára elérhetővé tenni. Ennek létrehozása mindenképpen fontos lenne, hogy a kisebb műhelyekben tárolt, talajvédelmi tervekben szereplő adatokat és a nagy, állami támogatással létrejött adatbázisokat, talajinformációs rendszereket egy a szakma, a döntéshozók, és az állampolgárok számára elérhető és érthető rendszerben harmonizált módon tároljuk és interpretáljuk.

A korábban doktori munkatervemben megjelöltek és a megszerzett tapasztalatok alapján munkámban a következő célkitűzéseket fogalmaztam meg:

1. A rendelkezésre álló nagy mennyiségű, értékes archív talajadat mentésére szolgáló minősítési, szűrési módszer, modern adatstruktúrába konvertálás és integrálás megalapozása.
2. Adatstruktúra modell létrehozása, mely alkalmas archív, és több forrásból származó adatok tárolására, úgy, hogy az szolgálja az adatfeldolgozást, illetve a nemzetközi harmonizálhatóságot.
3. A Szent István Egyetem Talajtani és Agrokémiai Tanszékének koordinálásában megújítás és tesztelés alatt lévő hazai talajosztályozás alacsonyabb szintjeinek fejlesztése, számszerűsítése, matematikai és statisztikai alapon archív adatok felhasználásával.
4. Az elkészült adatbázis struktúra és az abban tárolt adatok alapján származtatott tematikus állományok tesztelése modern, digitális térképezési célokra.

---

## Anyag

---

A munka, melynek jelentős része az adatok harmonizálását, illetve azok új struktúrába illesztését tűzte ki célul, több hazai és nemzetközi rendszeren és adatbázison alapult:

### **A magyar genetikai és talajföldrajzi osztályozási rendszer**

A magyar genetikai és talajföldrajzi osztályozási rendszer adta a talajosztályozásra vonatkozó adatharmonizációs munkáknak az alapját (Szabolcs, 1966; Stetanovits, 1981). A munka során a rendszer elemeit a WRB rendszerére korreláltuk.

### **A megújítás alatt álló magyar osztályozási rendszer**

A megújítás alatt álló magyar osztályozási rendszer osztályozó kulcsa az archív adatok kovertálásánál kapott jelentős szerepet, majd az automatizált osztályozó algoritmussal létrehozott adatbázis szolgáltatta az osztályozási rendszer alacsonyabb szintjeinek fejlesztéséhez használt adatokat (Michéli és mtsai, 2013a).

### **Világ Talaj Referenciabázis (WRB)**

A WRB-t a hazai osztályozási rendszer(ek) nemzetközi megfeleltetéséhez használtam. A hazai típusokat erre a talajosztályozási rendszerre korreláltam a nemzetközi harmonizálás érdekében (IUSS Working group, 2006).

### **FAO Útmutató a talajok leírásához**

Az adatok harmonizálásához a hazai talajosztályozási rendszer részeként bemutatott „Javaslat helyszíni talajfelvételezés általános módszertanára” (Szabóné Kele, 2013a) mellett ez a kiadvány szolgáltatta az alapot (FAO, 2006).

### **Talajvédelmi Információs és Monitoring Rendszer (TIM)**

A rendszer előnye, hogy egységes módszer alapján készült el és a laboratóriumi vizsgálatok is limitált számú akkreditált laboratóriumban kerültek kivitelezésre szabvány módszerek segítségével. Ezek alapján hazánkban a legegységesebb és legrészletesebb adatbázisnak tekinthető (TIM, 1995). A megújítás alatt álló magyar osztályozási rendszer vizsgálata és a WRB-vel való korrelálhatósági vizsgálatok magyar vonatkozású adatai a TIM-re épültek.

### **Világ Talajainak Emisszió-potenciál Nyilvántartása (WISE)**

Az adatbázis 149 ország több mint 10000 talajszelvényére vonatkozóan tárol adatokat. Ezek mindegyike a WRB szerinti osztályozási információt is tartalmaz a laboratóriumi és leíró adatok mellett. A WRB-vel való korrelálhatósági vizsgálatok alapadatait a WISE 3.1 szolgáltatta (Batjes, 2008).

### **Az Amerikai Egyesült Államok Talajvédelmi szolgálatának pontadatbázisa (NASIS) (USDA-NRCS NASIS point database)**

Az adatbázist elsősorban a strukturális felépítésének, előnyeinek, hátrányainak vizsgálatára használtam. Emellett mivel az adatbázis alkalmas genetikai szintenként több ugyanolyan morfológiai bélyeg tárolására is, azt is tanulmányoztam, hogy egyes morfológiai bélyegeket milyen részletesen rögzítenek az amerikai talajvédelmi szolgálat talajtani szakemberei ([http1](http://)).

## **Python programnyelv (3.x verzió)**

Az adatok ellenőrzésének, illetve az adatok harmonizálásának automatizálásához a Python programnyelvet választottam. Elsősorban könnyű átláthatóságának (olvashatóságának) és a nagy adattömegekkel való gyors és hatékony munkavégzésének köszönhetően (<http2>)

## **Microsoft Office Access**

Az adatbázis szerkezet felépítését MS Access használatával végeztem el. A szoftvercsomag mellett elsősorban a könnyű, kezdő felhasználók számára is tanulható kezelhetőség miatt döntöttem. Emellett lehetőség van nyílt forráskódú, ingyenes szoftverekkel is megnyitni az Access alapértelmezett formátumot, így az mindenki számára elérhető. Amennyiben az adatbázis eléri az Access-ben való kezelhetőség határait segédprogramokkal könnyen exportálható MySQL vagy egyéb formátumba.

---

## **Módszer**

---

### **Minőségellenőrzési módszerek**

A minőségellenőrzéshez tesztadatbázisként a TIM szolgált, az ellenőrző algoritmusokat (funkciók) pedig Python környezetben írtam. A különböző minőségellenőrzési algoritmusok közül határértékek, inkonzisztens értékek és adatkapcsolat ellenőrzését végeztem el egyes talajparamétereken, paraméter csoportokon. A program megírásánál figyelembe vettem a nemzetközi adatbázisokban használt funkciókat. Emellett a TIM és a hazai osztályozási és terepi útmutató sajátosságaira épülő algoritmusokat is létrehoztam.

### **Leíró adatok harmonizálásának módszere**

Az adatok ellenőrzése mellett fontos azok harmonizálása egy leíró és/vagy osztályozási rendszerre. Erre jó alapot szolgáltat a megújítás alatt lévő hazai osztályozási rendszer és az ahhoz kapcsolódó terepi útmutató, illetve a nemzetközi harmonizálásra készült WRB illetve FAO Útmutató a talajok leírásához. Az archív leíró adatokat, morfológiai jellemzőket a javasolt útmutató alapján viszonylag könnyen lehet konvertálni az új rendszerbe. A javasolt útmutató nagymértékben harmonizál a nemzetközi rendszerekkel is, így az már könnyen korrelálható a nemzetközi adatbázisokkal, kutatásokkal.

### **Laboratóriumi adatok harmonizálásának módszere**

A laboratóriumi módszerek, illetve az azok közötti konvertálást megkönnyítő faktorok leírása az adatbázis fontos része, ezért azok integrálását ott tárgyalom.

### **Talajosztályozásra vonatkozó adatok korrelálásának módszerei**

A harmonizálási munka egyik legproblémásabb része a talajosztályozási rendszerek közötti harmonizálás. A legtöbb esetben nem létezik egy egy arányú korrelálás rendszerek csoportjai között. Minden egyes pedon egyenkénti újraosztályozása rendkívül idő és emberi erőforrás igényes, így szükséges az újraosztályozás automatizálása. Ez mind a javasolt osztályozási rendszerre való korrelálás, mind a WRB-re való korrelálás esetében igaz. Michéli és munkatársai (2011) és Waltner és munkatársai (2012) egyaránt magyar adatbázisokra, így a TIM-re, a Kreybig adatbázisra és továbbiakra dolgozott ki olyan egyszerűsített

algoritmusrendszert, mely lehetővé teszi a WRB diagnosztikus szintek, anyagok és tulajdonságok származtatását, majd azok alapján az újraosztályozást.

Ehhez hasonló osztályozórendszert hoztam létre Python környezetben a javasolt magyar osztályozási rendszer alapján. A diagnosztikus rendszer lehetővé teszi a valós osztályozással nagymértékben megegyező eredményt adó automatikus osztályozást, az osztályozókulcs programnyelvre fordításával. A program lehetővé teszi az archív adatok gyors és hatékony újraosztályozását így azok korrelálását a modern rendszerre.

A WRB esetén egy ilyen algoritmusrendszer felépítésével próbálkozott német talajadatok felhasználásával Eberhardt és Waltner (2010). A WRB rendszere azonban annyira bonyolult és adatigényes, hogy a hazai adatbázisok ezt nem tudják kielégíteni. Alternatív megoldást kínál az úgynevezett taxonómiai távolság számítás, amely lehetővé teszi adott csoportok matematikai távolságának, azaz hasonlóságának vizsgálatát választott változók mentén. A változók, mind valós laboratóriumi adatok, bináris jellemzők, elméleti a talajosztályozási rendszer definícióit követő súlyozott értékek lehetnek. Mind a jelenleg használatban lévő, mind a javasolt osztályozási rendszer esetében megvizsgáltam a módszer használhatóságát korrelálási feladatokra. A módszer nem ismeretlen a talajtudományban, Hole és Hironaka (1960) nevéhez fűződik az első kutatás a témában, melyet a múlt század derekán számos kutatás követett (Bidwell és Hole, 1964a,b; Sarkar és mtsai., 1966; McBratney és mtsai., 2000), elsősorban helyi adatokra alapozva, korlátozott kiterjeszhetőséggel. A kétezres évek elején Minasny és McBratney (2007, 2010) az ausztrál osztályozásra, majd a WRB rendszerére alkalmazta az egyes egységek jelentésbeli tartalmának összehasonlítására. Az ausztrál talajok esetében számításaikat valós talajadatokkal végezték, a WRB esetében pedig 21 elkülönítő tulajdonság alapján kódolták a referencia csoportokat. A kódolás alapja a tulajdonságok jelenléte (1) vagy hiánya (0). Ezek az elkülönítő tulajdonságok jellemző talajképző tényezőket, illetve folyamatokat reprezentáltak. Az eredményül kapott mátrix alapján, elvégezhető a csoportok matematikai hasonlítása. A módszer az ún. „konceptió alapú” taxonómiai távolság számítási módszer nevet kapta.

#### A hazai barna erdőtalajok WRB referencia csoportokkal való korrelálásának folyamata taxonómiai távolságszámítás módszerével

A korrelálás során két módszert teszteltük:

1. konceptió alapú, ami azt jelenti, hogy a kiválasztott jellemző elkülönítő tulajdonságok mentén lettek kódolva a kiválasztott talajtípusok
2. centroid alapú, mely során mért, laboratóriumi adatok alapján történt a csoportok hasonlítása.

A módszereket a hazai barna erdőtalajok (BET) mintáján mutatom be. A BET típusokat választott WRB referencia csoportokkal korreláltuk a fent említett két módszer alapján. Az eredményeket pedig a korábban a két rendszer közötti harmonizálást tárgyaló szakirodalom eredményeivel hasonlítottuk össze (Michéli és mtsai, 2006).

#### *Konceptió alapú módszer*

A taxonómiai távolság számítás alapja egy tulajdonságmátrix, ahol a 7 BET típus és a választott 13 WRB referencia csoport a jellemző elkülönítő tulajdonságok mentén van kódolva. A jellemző elkülönítő tulajdonságokat a WRB-ban (2006) megfogalmazott úgynevezett egyszerűsített kulcs alapján választottuk ki. Ezek a tulajdonságok az ott megfogalmazott folyamatokat, folyamatársulásokat reprezentálják. A WRB kulcs egyes

jellemzőket alacsonyabb szinten kezel, mint a hazai osztályozás, például a kovárányosodást. Ezen tulajdonságok megjelenésének reprezentálására a WRB egyszerűsített kulcsban nem megjelenő elkülönítő tulajdonságokat kiegészítettük azokkal, amelyek a hazai osztályozásban nélkülözhetetlenek a típusok elkülönítéséhez.

A 19 talajtípust a 13 elkülönítő tulajdonság mentén kódoltuk. A kódolás során a Minasny és munkatársai (2010) által bevezetett korábban tárgyalt kódolást (a tulajdonság jelen van/nincs jelen) tovább finomítva bevezettünk egy köztes értéket, így a kódok jelentése a 1. táblázat szerint módosult. A kódolás után kaptuk meg a távolságszámításhoz használt tulajdonságmátrixot.

1. táblázat A talajtípusok kódolására szolgáló értékek és azok definíciói

Érték	Definíció
0	Elkülönítő tulajdonság nincs jelen
0,5	Elkülönítő tulajdonság jelenléte lehetséges
1	Elkülönítő tulajdonság definíció szerint kritérium

*Centroid alapú módszer*

A módszer során a fentiekben tárgyalt elkülönítő tulajdonságokat megpróbáltuk valamely mért paraméterrel/paraméterekkel, vagy azokból származtatott értékkel leírni. Az így meghatározott 9 paraméterre a meglévő és korábban tárgyalt adatbázisok alapján (TIM, WISE 3.1), úgynevezett centroid értékeket számoltunk. A számítás során minden egyes talajszelvény meghatározott mélységére súlyozott átlag segítségével kiszámoltuk a centroid értékeket, majd azok átlagát véve kaptuk meg a típus centroidját.

*Taxonómiai távolság számítása*

A talajtípusok közötti távolság mind a koncepció, mind a centroid alapú módszer esetében Mahalanobiszi távolság számítással számítottuk az R szoftver (R Development Core Team, 2009), HDMD (McFerrin, 2013) csomagjának segítségével.

$$d_{ij} = ((\mathbf{x}_i - \mathbf{x}_j)^t S^{-1} (\mathbf{x}_i - \mathbf{x}_j))^{1/2},$$

ahol  $d_{ij}$  a  $D$  távolságmátrix eleme, ami  $c \times c$  méretű,  $c$  a talajtípusok mennyisége.  $S$  reprezentálja a kovariancia mátrixot,  $d_{ij}$  pedig a taxonómiai távolság  $i$  és  $j$  talajtípus között. Az  $x$  a változók vektorának jelölésére szolgál.

A hazai talajtípusok korrelálása a javasolt hazai osztályozás típusaira

A javasolt hazai talajosztályozási rendszer osztályozó kulcsa alapján Python környezetben létrehozott algoritmus sorozattal származtattam a javasolt típusokat. Az osztályozó algoritmus a laboratóriumi adatoknak ad prioritást, azok nagyobb megbízhatósága miatt (akkreditált laboratóriumi méréseknek köszönhetően). Az osztályozó algoritmus nagymértékben követi a definíciókat, köszönhetően a jól definiált, egyszerű diagnosztikus rendszernek.

## **Modern adatstruktúra modell létrehozása**

Az ingyenesen elérhető adatbázisok használata és vizsgálata után, az amerikai NASIS adatbázis és az „európai” talaj adatbázis struktúráknak megfelelő WISE adatbázis előnyeinek integrálásával egy hatékony, a kor kívánalmainak megfelelő adatstruktúra alapjait fektettem le MS Access program használatával. A rendszer alkalmas a különböző forrásból származó adatok tárolására és könnyen lekérdezhető gyakorlatlan felhasználó számára is.

## **A javasolt hazai osztályozási rendszer alacsonyabb szintjeinek fejlesztése, számszerűsítéséhez alkalmazott statisztikai módszerek**

A célkitűzésben megfogalmazottak teljesítésének előfeltétele volt az osztályozó algoritmus létrehozása. Az osztályozó algoritmus segítségével javasolt típusokba sorolt TIM felhasználásával adatbázis szinten volt lehetőség megvizsgálni az új osztályokon belüli elsődleges elkülönítő tulajdonságokat. Ehhez több matematikai módszer nyújtottak segítséget:

### Silhouette analízis

Silhouette analízis (Rousseeuw, 1987) során a klasztereződés erősségét vizsgáljuk, illetve, hogy egy egyed milyen erős tagsággal rendelkezik adott klaszterben. A Silhouette analízis az alábbiak szerint kerül levezetésre:

$$SW_i = (b_i - a_i) / \max(a_i, b_i);$$

ahol  $a$  az átlagos távolság  $i$  pont és az  $i$  pont klaszterének többi egyede között, míg  $b$  az átlagos távolság  $i$  pont és a többi klaszter egyedei között, így minden esetben:  $-1 < SW_i < 1$

Az  $SW_i$  az alábbiak szerint értelmezhető:

0,71 – 1,00 erős kapcsolat (jó klasztereződés)

0,51 – 0,70 mérsékelt kapcsolat

0,26 – 0,50 gyenge kapcsolat

$\leq 0,25$  nincs kapcsolat (a klasztereződés nem valós)

### Főkomponens analízis

Főkomponens analízis (Pearson, 1901) során a változók kovariancia struktúráját a legkevesebb számú lineáris kombinációval írjuk, le miközben a lehető legkevesebb adatot veszítjük el. Az első főkomponens az a lineáris kombináció (egyenes), amely mentén az adatok (pontfelhő) szóródása a legnagyobb. Az erre merőleges irányok mentén határozzuk meg a további főkomponenseket, melyek száma maximálisan a változók száma lehet.



## k-közép klaszterezés

Klaszterezéshez az egyszerű k-közép klaszterezést használtam (MacQueen, 1967). A módszer a felhasználó által rögzített számú klaszterekbe rendezi az egyedeket egyszerű euklideszi távolság alapján. A módszernek több levezetése is van a használt szoftvercsomag alapján a klaszterezés az alábbiak szerint történik (Hartigan és Wong, 1979):

$$SS(k) = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_{kj})^2$$

ahol  $k$  a klaszter,  $x_{ij}$  a  $j$  változó értéke az  $i$  mintapont esetén és  $\bar{x}_{kj}$  a  $j$  változó átlaga a  $k$  klaszteren belül.

## Felhasznált adatok

A TIM adatbázisból a létrehozott osztályozó algoritmus alapján Agyegbemosódásos talajoknak sorolt talajszelvényeken teszteltem a fent részletezett módszereket.

## **Talajtérképek létrehozása modern terepi felvételezési módszerek és tematikus talajtérképek használatával**

A célkitűzésben megfogalmazottak motivációja a digitális talajtérképezési prioritások megváltozása, mely a talajtípus térképektől a tematikus térképek irányába való elmozdulást jelenti. Napjaink legnagyobb térképezési projektjei úgymint az African Soil Information Service (African Soil Information Service, 2013) és a GlobalSoilMap (GlobalSoilMap, 2013) elsősorban a tematikus talajparaméterek térképezésre fókuszál. Egyik projektnek sem célja talajtípus térkép létrehozása. A talajtípusok nagyobb információtartalma viszont vitathatatlan a tematikus információkkal szemben. A GSM specifikációja alapján kíséreltem meg talajosztályozási információt származtatni. Különböző módszereket vizsgáltam meg annak érdekében, hogy a specifikációk alapján javasolt magyar talajtípusokat származtassak a modellezett paraméterekből.

A specifikációk és a TIM alapján létrehoztam a javasolt magyar típusok centroidjait és a taxonómiai távolság számítás segítségével teszteltem, hogy a paraméterek alapján származtatható-e talajtípus információ. A taxonómiai távolság számítás során a Mahalanobis távolságot alkalmaztam. A vizsgálatok során 250 szelvényt random módszer segítségével kiválasztottam a tréning állománytól, ezek a szelvények szolgáltak a módszer validálására. A 250 szelvény a minta 22 százalékát jelentette.

A taxonómia távolság számítás mellett az összes szelvény segítségével random forest (Breiman, 2001) módszerrel megvizsgáltam, hogy döntési fák segítségével milyen pontossággal osztályozhatók az egyes szelvények. A random forest lényege, hogy több, előre meghatározott számú osztályozó (döntési) fát hozunk létre, majd a vizsgált megfigyeléseket ezek alapján osztályozzuk. Minden egyes fa eredményét (az ismeretlen osztályú megfigyelés osztályát) rögzítjük és a végső osztály az lesz, mely a legtöbb esetben előfordul.

---

## Eredmények

---

### Minőségellenőrzés

A minőségellenőrzés a laboratóriumi adatokat illetve azok leíró paraméterekkel való kapcsolatát vizsgálja. A program egyenként ellenőrzi a szelvények minden egyes szintjét adott paraméterre kidolgozott módszer alapján. A határérték ellenőrzése során a valós minimális és maximális értékekkel rendelkező paraméterek, - így mindegyik, mely százalékos értéket fejez ki – esetében az alsó érték minden esetben 0, míg a felső határ 100 volt. A nem százalékkal kifejezett értékek esetében, amennyiben a nemzetközi harmonizált adatbázisok tartalmaztak erre vonatkozó értéket, azokat vettem alapul. Amennyiben sem az irodalom, sem a mértékegység folytán nem állt rendelkezésre határérték, a TIM-et mint az ország teljes területét reprezentáló adatbázist vettem alapul és az abban paraméterenként előforduló maximális érték 20%-al növelt mennyiségét vettem a paraméter maximális értékének. Ezek a határértékek a minőségellenőrzés sztenderdizálása esetén könnyen módosíthatóak.

A program a hibajavítás mellett egy segédfájlt is létrehoz, melyben minden egyes funkció mellett megtalálható, hogy mely talajszelvények mely szintjei kerültek javításra. A segédfájl nagy szerepet kap az elvégzett módosítások nyomon követésében, mivel azok visszakereshetősége fontos követelmény. A program könnyen módosítható, oly módon, hogy a hibásnak vélt adatokat ne törölje, csak a segédfájlt hozza létre, így a felhasználónak lehetősége nyílik az alapján az eredeti adatbázisban visszakeresni az immáron csak jelölt adatokat és a hiba okát megvizsgálni.

### Leíró adatok harmonizálása

A harmonizálás alapját a TIM módszerkönyv (TIM, 1995), a javasolt útmutató és a FAO Útmutató a talajok leírásához (FAO, 2006) adta. Az adatokat az előzőekben bemutatott funkciókhoz hasonló algoritmusokkal korreláltam. A programnyelv viszonylagos bonyolultsága és hossza miatt az elvégzett harmonizálási lépéseket táblázatos formában ismertetem, ahol feltüntettem a rekord típusát, eredeti és harmonizált jelölését. A korrelálás mellett bizonyos, egyéb paraméterekből származtatható jellemzőket is hozzáadtam a leíráshoz, az eredeti leírást így bővítve, a táblázat ezeket az algoritmusokat is tartalmazza. Minden paramétert ékezetek nélküli kódokkal, rövidítésekkel helyettesítettem, adatkezelési és tárolási okokból, ezek pontos jelentését az adatbázis metaadatokat tartalmazó táblákban tárolja.

A leíró adatok harmonizálása során több esetben problémát jelentett a javasolt útmutatóban az adott paraméterre vonatkozó definíció hiánya. Ilyen esetekben a FAO Útmutató a talajok leírásához jó alapot szolgáltatott, azonban a hazai talajosztályozási rendszer fejlesztése során a definíciókat ezekre a paraméterekre is ki kell terjeszteni, oly módon, hogy azok harmonizálhatók legyenek a nemzetközi rendszerrel. Több esetben jelentett gondot a direkt konvertálás, vagy azért mert a hazai leírás, vagy azért mert a nemzetközi útmutató részletezi jobban az adott paramétert. Megfontolandó a FAO szerinti hierarchikus leírás bevezetése. Erre jó példa a területhasználatra vonatkozó leírás, ahol a FAO az első szinten megkülönböztet 9 típust, ezek egyike a szántóföldi mezőgazdaság. A hazai szántó definíció ebbe a legfelső kategóriába illik bele, azonban annál több (vagy eltérő) információt hordoz. A második szinten a szántóföldi mezőgazdaság esetében egynyári, évelő, fás szárú vegetációt különböztet meg. A hazai szántó kategória ezek közül elsősorban az egynyári és évelő kategóriába tartozhat bele.

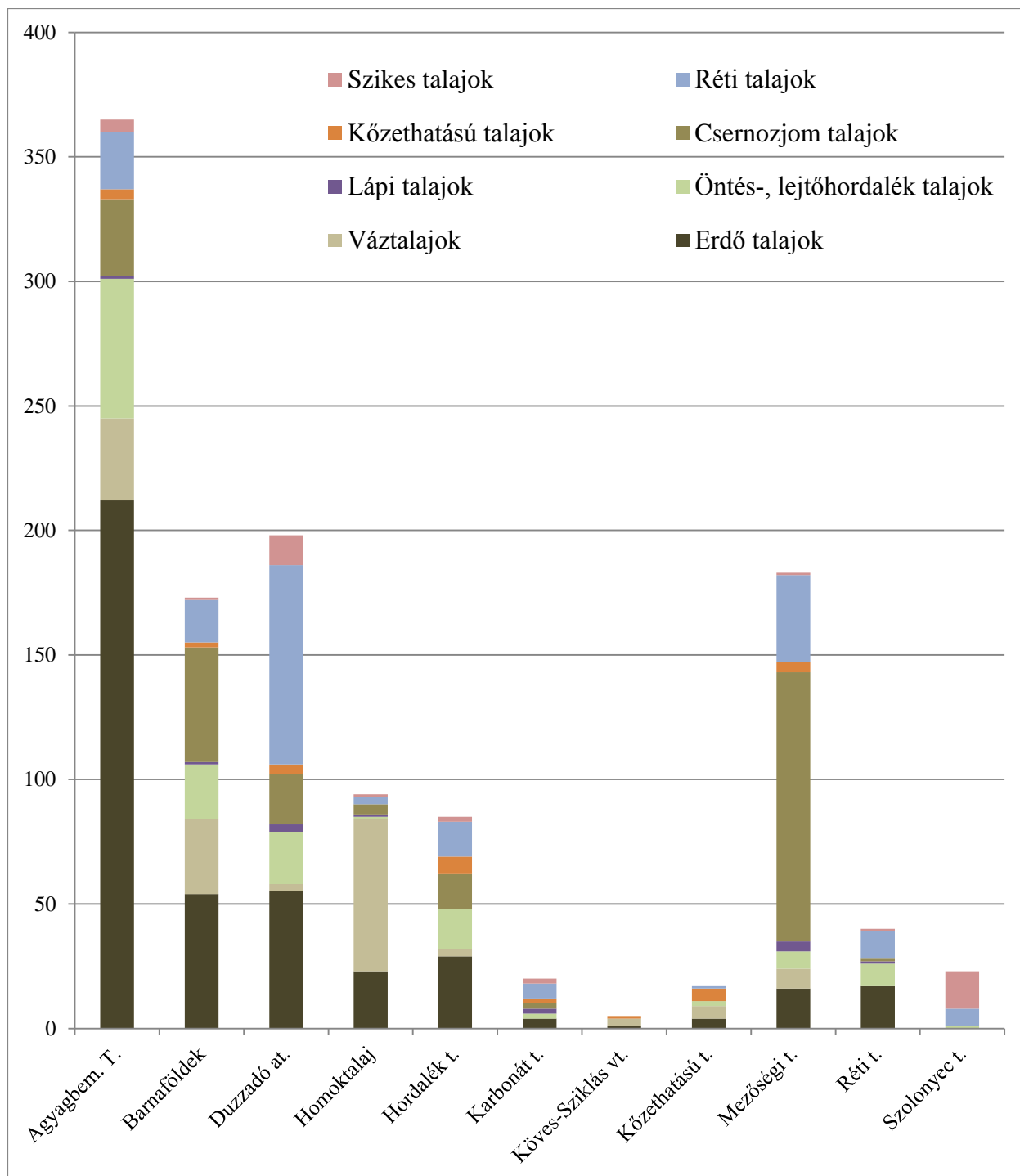
A talajok szerkezeti elemeinek esetében a hazai definíció sok esetben hiányos a TIM módszerkönyv 13 típust részletez, majd a függelékben már 19 lehetséges típust találunk. A definíciók erősen hiányosak, így a konvertálás sok esetben a terepi tapasztalatokra a felvételezők módszerére kell, hogy támaszkodjon. A szerkezetnél is, mint több más morfológiai bélyeg esetében a FAO megkülönböztet típust, arra vonatkozó méretet, fokot, mennyiséget, melyek plusz információt szolgáltatnak, így integrálásuk az új rendszerben fontos lenne. Emellett az adatbázisban azok külön rendezése, így az adatok használatának könnyítése is szükségszerű.

### **Az archív adatok konvertálása a javasolt magyar osztályozás típusaira**

A javasolt osztályozó kulcs a TIM-ben tárolt adatok alapján jó hatásfokkal, a definíciókat jól követve programozható. A javasolt kulcsban meghatározott struktúrát követve épült fel az osztályozó kulcs is. Bizonyos javasolt típusok programozása – így az Antropogén talajok – azonban a rendelkezésre álló információk hiánya (egy elérhető magyar adatbázis sem tartalmaz jelenleg erre vonatkozó adatokat) miatt nem kódolható. Ezek a típusok nagy valószínűséggel egyébként sem jelennének meg a TIM adatbázisában. A program a korábban is használt Python nyelvben íródott, és az adatellenőrzési, korreláló programmal összekapcsolható, annak kimenetét (is) olvassa. A bemeneti adatoknál már a korrelált új rendszer szerinti morfológiai leírásokra épít.

Az osztályozó algoritmust a TIM adatbázisán teszteltem (1. ábra). A teljes adatbázis (>1200 szelvény) osztályozását egy erősebb munkaállomáson (4Gb memória, Intel Core i5-tel egyenértékű processzor) a memória terheltségétől függően 40-60 másodperc alatt végzi el. Összekapcsolva az adatellenőrző és korreláló résszel a teljes TIM adatbázis körülbelül 2-3 perc alatt áthalad a teljes ellenőrzésen és az újraosztályozáson. A program több fájl kimentésére alkalmas. Egyik opció az egyedi szelvényazonosítók és az újraosztályozás eredményének fájlba mentése. A másik alternatíva a teljes adatszerkezet fájlba mentése, az új és a régi osztályok feltüntetésével. Emellett a program módosítható oly módon, hogy az adatok kimentése egy MS Access vagy OpenOffice Base kapcsolt táblás adatstruktúrába is beolvasható SQL kódoláson keresztül.

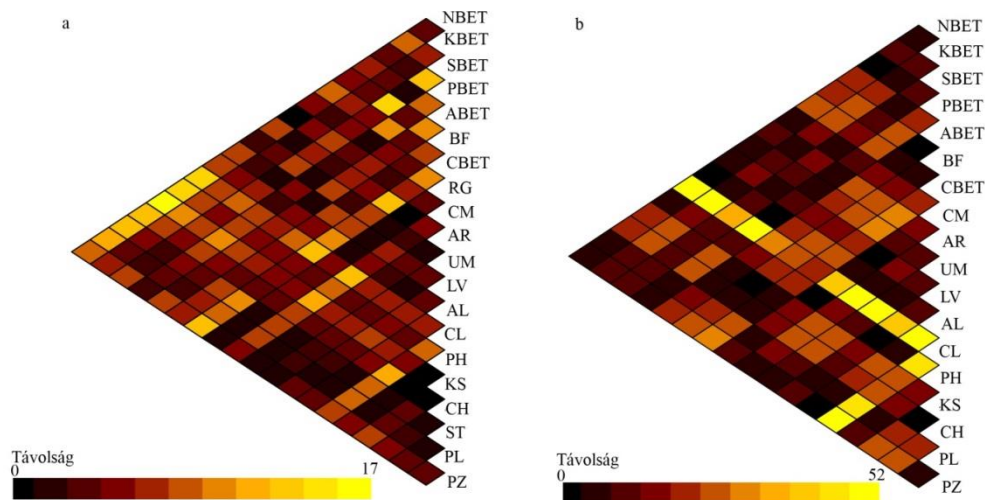
Az osztályozó programba lehetőség van egy olyan algoritmust beépíteni, mely a paraméterek rendelkezése állása és az osztályozott típus alapján egy valószínűségi értéket ad. Oly módon, hogy kiinduló megbízhatósági értéket (pl.: 100) minden egyes esetben csökkenti, ha a paraméter hiányzik, de kulcsfontosságú lenne az osztályba soroláshoz.



1. ábra. A létrehozott osztályozó algoritmus alapján a TIM talajszelvényeinek besorolása a javasolt magyar osztályozás típusaiba, feltüntetve a jelenlegi magyar osztályozást

### Taxonómiai távolságszámítás alapú korrelálás eredményei

A magyar típusok és a WRB referenciacsoportok közötti távolságok külön a koncepció és a centroid alapú módszer szerint, a könnyebb vizuális értelmezhetőség érdekében az R szoftverben található „squash” (Eklund, 2011) csomag segítségével ábrázolva hőterképen. a 2. ábrán látható.



2. ábra. Számított távolságmátrix hőtérképen ábrázolva, koncepció alapon (a) és centroid alapon (b)

A 2. számú táblázatban a távolságmátrixok egy kivonata található, ahol a magyar típusok és a hozzájuk legközelebb eső 3 referenciacsoport található, előbb a koncepció, majd a centroid alapú módszer szerint. A negyedik oszlopban az eredmények értékelésénél alpműként használt munkában (Michéli és mtsai., 2006) leírt, az adott magyar típusokkal korreláló referenciacsoportok találhatóak sorrendiségét mellőzve.

2. táblázat. BET típusok és WRB referencia csoportok közötti lehetséges korrelációk a különböző taxonómiai távolságszámítási módszerek és korábbi kutatások alapján

BET típusok	3 legközelebbi WRB RSG, koncepció alapú számítás alapján	3 legközelebbi WRB RSG, centroid alapú számítás alapján	Korreláló RSG-k (Michéli és mtsai. 2006)
Csernozjom BET	Chernozem, Phaeozem, Kastanozem	Kastanozem, Chernozem, Luvisol	Chernozem, Kastanozem, Phaeozem
Barnaföld	Umbrisol, Cambisol, Chernozem	Kastanozem, Chernozem, Luvisol	Cambisol
Agyagbemosódásos BET	Luvisol, Planosol, Stagnosol	Kastanozem, Chernozem, Cambisol	Luvisol, Alisol
Pangóvízes BET	Luvisol, Stagnosol, Planosol	Cambisol, Luvisols, Planosol	Luvisol, Stagnosol
Kovárványos BET	Arenosol, Luvisol, Regosol	Arenosol, Alisol, Umbrisol	Luvisol, Arenosol
Podzolos BET	Umbrisol, Regosol, Podzol	Alisol, Podzol, Arenosol	Luvisol, Umbrisol, Alisol
Savanyú nem podzolos BET	Cambisol, Umbrisol, Regosol	Alisol, Planosol, Umbrisol	Cambisol, Umbrisol, Alisol

A számítások alapján a csernozjom barna erdőtalajokhoz a WRB sztyeppe talajai találhatóak a legközelebb, mely jól korrelál a tapasztalati eredményekkel. A Luvisolok megjelenése, mint harmadik legközelebbi referencia csoport a centroid alapú távolság számításnál elsősorban annak köszönhető, hogy a csernozjom barna erdőtalaj definíciója nem részletezi az agyagmozgás mértékét illetve a CaCO<sub>3</sub> megjelenési mélységét.

A barnaföldek esetében előre ki kell emelni, hogy a centroid alapú módszer a kevésbé fejlett talajokat nehezebben „azonosítja”, sokkal jobban működik, azon csoportok esetében, amelyeknek valamilyen speciális, mért értékekben jól elkülönülő tulajdonsága van. A TIM barnaföldekből több szelvényt tartalmaz, mint bármely más barna erdőtalajból, az agyagbemosódásos barna erdőtalajokat leszámítva. A koncepció alapú eredményeknél a Cambisolok megjelenése a három legközelebbi csoport között jól korrelál a korábbi eredményekkel. A centroid alapú korrelálás jól mutatja, hogy a típus nem definiált numerikusan. A szelvényeket egyenként megvizsgálva, majdnem minden releváns referencia csoportot megtalálunk köztük.

Agyagbemosódásos barna erdőtalajok esetében a várakozásnak megfelelően a Luvisols referencia csoport a legközelebbi, azonban a centroid alapú eredmények ezt nem reflektálják. Itt a sztyeppe talajok és a Cambisol található, mint legközelebbi csoportok ismét bizonyítva, hogy a definíciók számszerűsítése milyen fontos eleme az osztályozásnak.

Pangóvizes barna erdőtalajok esetében mind a két módszer esetében megjelenik a Luvisol referencia csoport. A koncepció alapú módszer esetében a Stagnosol és a Planosol megjelenése is megfelel a korábbi kutatási eredményeknek. A centroid alapú vizsgálatoknál a Stagnosol csoport nem szerepelt a számításokban, mivel a WISE 3.1 nem tartalmaz erre a referencia csoportra vonatkozó adatokat.

A kovárványos barna erdőtalajok esetében jól látható az a korábban említett tény, hogy speciális karakterisztikával rendelkező talajok könnyebben azonosíthatók a módszerrel. Mind a két módszer esetében az Arenosol referencia csoport került a legközelebb a típushoz. Az Umbrisolok megjelenése a centroid alapú módszernél azzal magyarázható, hogy az alacsony pH és bázistelítettség jóval nagyobb jelentőséget kap a típusok elkülönítésénél, mint a szervesanyagtartalom, köszönhetően annak, hogy e két paraméter, és főleg a bázistelítettség mentén nagyobb a típusok szórása.

A legtöbb hazai podzolos barna erdőtalaj nem teljesíti a WRB spodic diagnosztikus szint kritériumait, ennek ellenére a talajok karakterisztikája és a spodic szint lehetséges előfordulása közel hozta őket a Podzol referencia csoporthoz mindkét módszer esetében. A többi eredmény jól korrelál a korábbi eredményekkel illetve a tapasztalatokkal, elsősorban a bázistelítettség és alacsony pH értékeknek köszönhetően.

Savanyú nem podzolos barna erdőtalajok esetében is meglehetősen jól korrelálnak az eredmények a korábbi kutatásokkal, egyedül a Planosol referencia csoport megjelenése az, ami váratlan a korábbi munkákat figyelembe véve. A centroid értékeket megvizsgálva, azonban látható, hogy a két csoport a szelvényen belüli agyagnövekedést leszámítva jól korrelál.

Összességében megállapítható, hogy a két módszer a megfelelő változók megválasztásával alkalmas lehet talajosztályozási rendszerek korrelálására. A változóknak megfelelően kell reprezentálnia a kiválasztott talajtípusokat. Koncepció alapú módszer esetén a talajok jellemző talajképző tényezőit, folyamatait, míg centroid alapon a legfontosabb elkülönítő tulajdonságokat. A centroid alapú számítások fontos alapja az ellenőrzött, megbízható adatbázis. A kutatás során sok esetben találtunk félreosztályozott szelvényeket a WISE adatbázisban, ami arra enged következtetni, hogy a centroidok számításánál a szelvények paramétereiben a szórás nagyobb volt, mint egy teljesen megbízható adatbázis esetében. Ellenőrzött adatbázis alapján készült WRB centroidok jól használhatók egy másik rendszer típusainak WRB-re való korrelálására. Az eredmények alapján megállapítható, hogy bár a 3 módszer sok esetben egybevágtott, a módszerek együttes alkalmazása ajánlott, ily módon kiegészítve egymást és hozzájárulva a pontosabb eredmények eléréséhez. A korrelálási eredmények mellett, a kutatás rámutatott a genetikus szemléletű rendszerek hiányosságaira, a pontos definíciók hiányára.

## A javasolt hazai osztályozás egyes egységeinek taxonómiai távolság alapú korrelálása WRB referencia csoportokra

Az előbb ismertetett módszerek közül a centroid alapú taxonómiai távolság számítás az osztályozó algoritmus alapján újraosztályozott TIM adatbázison is elvégeztem. A korábban használt centroidokat meghagyva, 5 „új” magyar talajtípust vizsgáltam a korábban definiált referencia csoportokhoz viszonyítva. A centroidok a barna erdőtalajok alapján lettek definiálva, így egyes talajtípusokat, úgymint a Mezőségi talajok tulajdonságait nem teljes mértékben reprezentálva. A távolságszámítást a korábban is alkalmazott módszerrel végeztem el. Az értelmezhetőség érdekében ebben az esetben is kivonatoltam az eredményeket, a 3. táblázatban. Az egyes talajtípusok mellett a 3 legközelebbi referenciacsoport található, sorrendben.

3. táblázat. Javasolt hazai típusok és a centroid alapú távolságszámítás alapján legközelebb eső 3 WRB referenciacsoport

Javasolt hazai típusok	Legközelebbi RSG	2. RSG	3. RSG
Agyagbemosódásos talajok	Luvisol	Cambisol	Chernozem
Barnaföldek	Chernozem	Kastanozem	Luvisol
Homoktalajok	Arenosol	Luvisol	Cambisol
Karbonát talajok	Calcisol	Kastanozem	Chernozem
Mezőségi talajok	Calcisol	Kastanozem	Chernozem

Az eredmények alapján elmondható, hogy az Agyagbemosódásos talajok az új osztályozás alapján - egybevágóan a szakvéleményekkel - egyértelműen a Luvisol referencia csoporthoz vannak a legközelebb. A Homoktalajokhoz a számítás ebben az esetben is az Arenosol referencia csoportot eredményezte a legközelebbinek, mely szintén megfelel a korábbi kutatásoknak. A Karbonát talajok, melyek a felszínhez képest 50 centiméteren belül több mint 40% kalcium-karbonátot tartalmaznak, a Calcisol referencia csoporthoz kerültek a legközelebb, ami a két csoport tulajdonságainak ismeretében a valóságnak megfelelő. A Mezőségi talajok esetében a Calcisolok megjelenése, mint legközelebbi talajtípus némileg meglepő, ugyanakkor a centroid értékeket megvizsgálva egyértelművé válik, hogy a nemzetközi adatbázisban a sztyeppe talajok a  $\text{CaCO}_3$  megjelenést tekintve jóval kilúgzottabbak, míg a hazai Mezőségi talajoknál már átlagosan 47 cm-nél megtaláljuk az úgynevezett „calcic” szintet. A nemzetközi adatbázis (WISE) alapján számított Kastanozem „calcic” szint megjelenése ugyanakkor meglepően mély (véltetően félreosztályozott szelvények), elsősorban ez eredményezte a hazai Mezőségi talajok Calcisolhoz való közelségét.

A Barnaföldek esetében, mely gyengén fejlett talajokat reprezentál, így a korábbiakban említettek alapján a pontos elkülönítésük nehéz a módszerrel, a Cambisolok megjelenése az első helyek esetén volt várható. Ezzel szemben azok csupán a negyedik legközelebbi referencia csoport. A centroidokat megvizsgálva, jól látható, hogy a WISE adatbázisban található Cambisolok a hazai Barnaföldeknél jelentősen savanyúbbak, kilúgzottabbak, ennek megfelelően nem is kerülhettek közelebb a Cambisolokhoz, mint a magasabb pH-val és  $\text{CaCO}_3$  tartalommal jellemzett más referencia csoportokhoz. Az eredmények értelmezésekor azt is figyelembe kell venni, hogy a centroidok a Barna erdőtalajokra lettek definiálva. Az eredmények nagymértékben javíthatóak, olyan paraméterek bevezetésével, melyek jobban reprezentálják a korábbi vizsgálatokban nem szereplő hazai talajtípusokat. Emellett azt is fontos kiemelni, hogy környezeti paraméterek, nagyban befolyásolják a paramétereket, így egyes típusok (elsősorban a gyengén fejlett típusok) esetében javasolt a nemzetközi adatbázisból felhasznált szelvények csökkentése a vizsgált talajtípusok környezeti

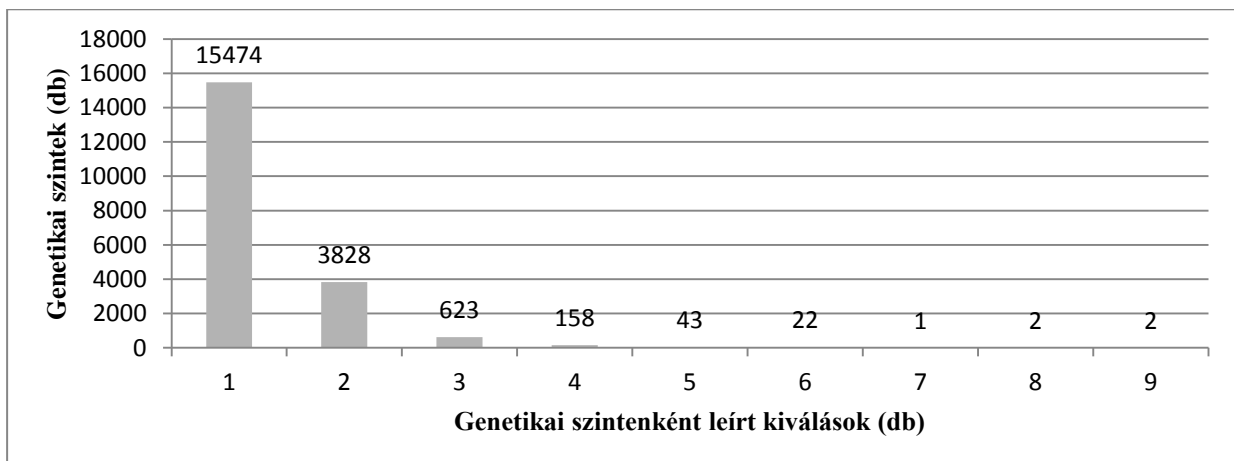
paramétereinek megfelelően. A hazai Barnaföldek hasonlítása, trópusi, kilúgzott Cambisolokhoz nyilvánvalóan nem a prekoncepciónak megfelelő eredményt hozza, míg ha a Kárpát-medence környezetéből válogatjuk le a nemzetközi szelvényeket, akkor az nagyobb eséllyel ad a valósághoz közeli eredményt.

### **Modern adatstruktúra modell létrehozása**

Az adatstruktúra létrehozásánál elsősorban az anyag részben említett adatbázisokkal való harmonizálási, minőségellenőrzési munka tapasztalataira, az amerikai adatbázis morfológiai tábláinak elemzésére, továbbá a javasolt Terepi útmutató és a FAO terepi útmutató javaslataira támaszkodtam.

#### A morfológiai adatok rögzítése

A vizsgálatok során megvizsgáltam, hogy egyes szintek esetében bizonyos tulajdonságcsoportok, hányszor kerültek leírásra. Példaként a 3. ábrán a genetikai szintekben található különböző kiválások megjelenésének száma szerepel. A kiválások különbözősége, mind anyagban, mind színben, mind méretben lehetséges.



3. ábra. A NASIS adatbázisban egy genetikai szinthez leírt kiválások mennyiségének eloszlása

A fenti példához hasonló módon készült el a többi morfológiai tulajdonság elemzése, majd az alapján definiáltam a morfológiai adatokat tároló táblákat.

#### A laboratóriumi adatok harmonizálása a nemzetközi módszerek alapján

Az adatok harmonizálásánál szükséges az átszámítási faktorok tárolása, a hivatkozás pontos megjelölésével. Az erre vonatkozó adatokat egy úgynevezett *Korreláló faktor* táblában tároltam, mely kapcsolati rendszerén keresztül összeköttetésben van a mért adatokkal. A tulajdonságra vonatkozó lekérdezésben egy egyszerű kifejezéssel és feltétellel végeredményként már a nemzetközi módszernek megfelelő értékre korrelált eredményt kapjuk.

#### A végleges adatstruktúra modell

A fentiekben kifejtett fő jellemzők alapján készült el az adatbázis modell, amely a TIM adatokkal lett feltöltve tesztelés céljából. A struktúra összesen 38 táblát tartalmaz, melyek alkalmasak a helyszín, a genetikai szintek morfológiai jellemzőinek és a rétegek/szintek laboratóriumi értékeinek tárolására.

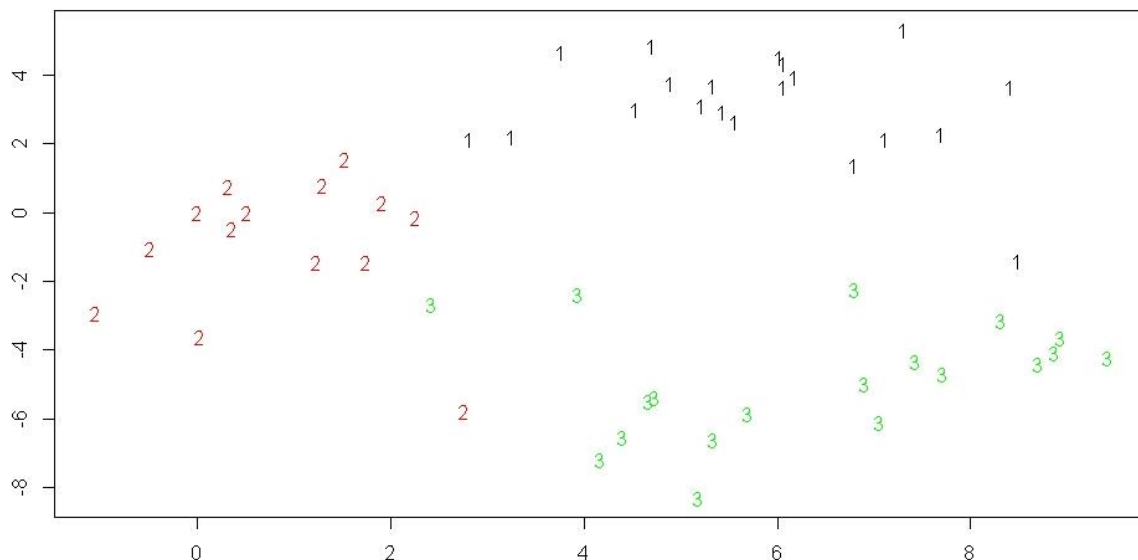


## A javasolt hazai osztályozási rendszer alacsonyabb szintjeinek fejlesztése, számszerűsítése

Az automatizált osztályozó algoritmus létrehozása után lehetőség nyílt a javasolt típusok adatbázis szintű vizsgálatára, a típuson belüli fontosabb elkülönítő paraméterek vizsgálatára. Ehhez az újraosztályozott TIM adatbázist használtam fel. A vizsgálatokat a legnagyobb számú és adott típusnál releváns paraméterekre terjesztettem ki. A módszereket a javasolt osztályozási rendszer Agyagbemosódásos talaj típusának példáján teszteltem. Az alábbiakban 2 példán szemléltetve az eredményeket.

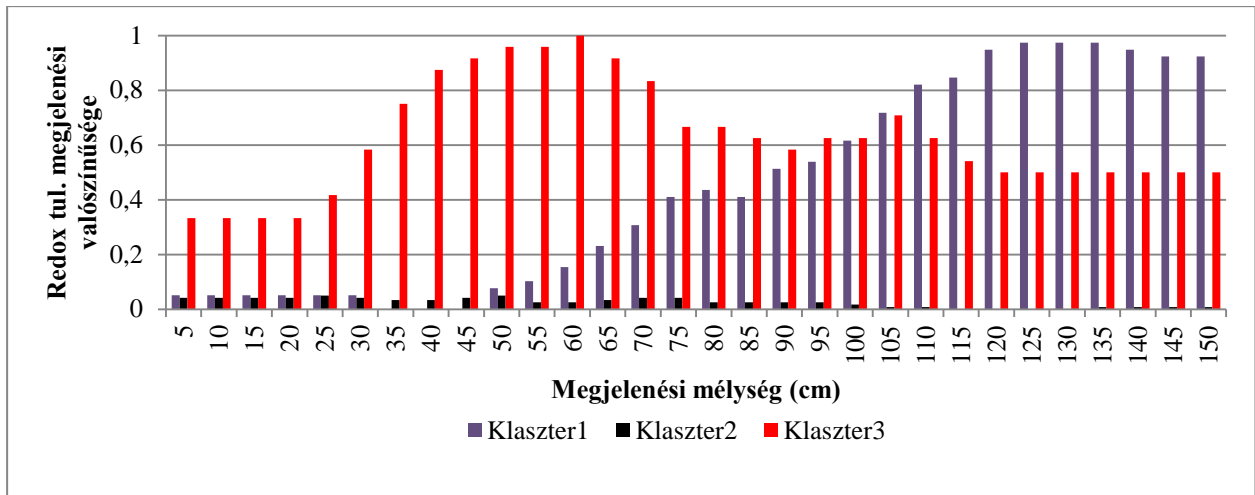
### Agyagbemosódásos talajok vizsgálata redox tulajdonságok megjelenése alapján

A redox tulajdonságokra vonatkozó morfológia bélyegek megjelenése illetve megjelenési mélysége fontos elkülönítő bélyeg lehet az osztályozás alsóbb szintjein. Silhouette analízist követően, egyszerű k-közép klaszterezést elvégezve, látható a populáció elkülönülése 3 csoportba (4. ábra). Természetesen a folytonos adatok miatt előfordulnak, egyedek, melyek matematikailag 2 csoportba ugyanakkora valószínűséggel tartozhatnak, azonban a megközelítés alapján 3 csoport nagy valószínűség mellett elkülöníthető.



4. ábra. Az Agyagbemosódásos talajok klasztereződése 3 jól elkülöníthető klaszterbe (1-3) két fő komponens mentén ábrázolva

A klasztereket külön ábrázolva és a megjelenés valószínűségét átlagolva (a redox tulajdonság korábban származtatott bináris érték. Az átlag valószínűség maximális értéke, így 1.) megállapítható a 3 mélység tartomány is, amely alapján az elkülönítés történt (5. ábra). Az egyes klaszter esetében egy méter alatt nagy valószínűséggel már megjelennek a bélyegek, míg 115 cm alatt szinte biztosan redox tulajdonságokat találunk. A 2. klaszter esetében a redox tulajdonságok megjelenésének hiánya, vagy azok minimális valószínűsége a felszíni szintekben tapasztalható. A 3. számú klaszter esetében 40 és 80 cm között találunk egy maximumot a megjelenés tekintetében, ez nagy valószínűséggel a jelenlegi osztályozás pangóvízes talajainak klasztereződése. Ezek alapján megállapítható, hogy a Barna erdőtalajokat a redox tulajdonságok alapján matematikai alapon 3 mélység alapján érdemes elkülöníteni: felszín közelben, 35-50 cm-es mélység környékén, és egy méter alatt.

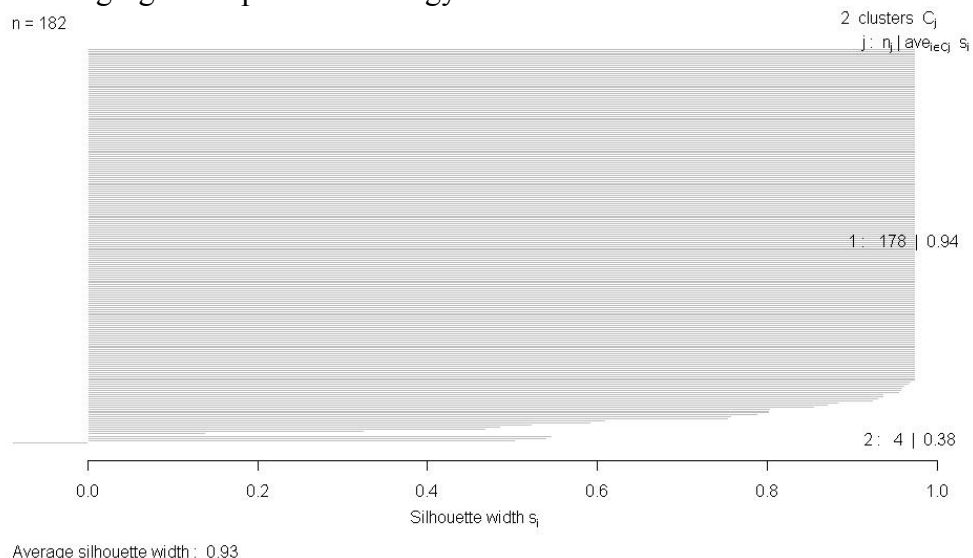


5. ábra. Az Agyagbemosódásos talajok 3 klasztere a redox tulajdonságok megjelenésének valószínűsége alapján, egyes mélységekben

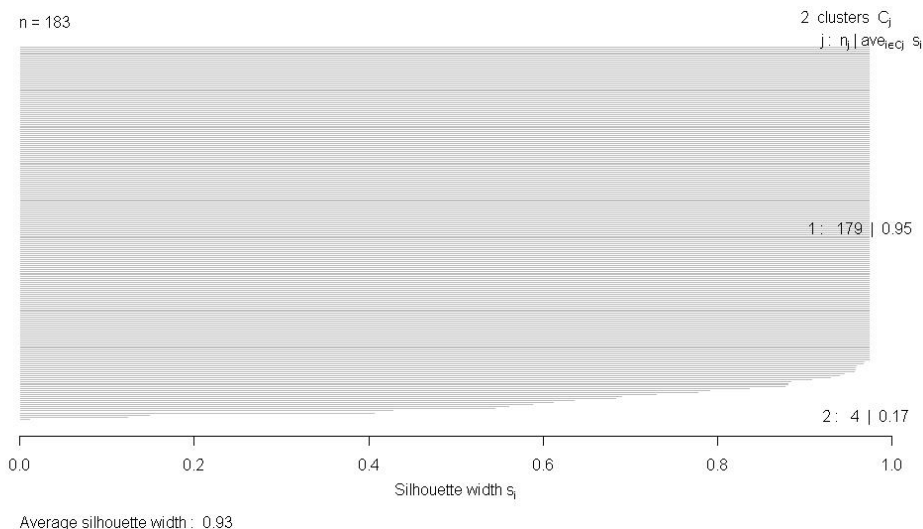
### Agyagbemosódásos talajok vizsgálata durva vázrészek mennyisége alapján

A durva vázrész tartalom vizsgálatok során az előzőekben alkalmazott, rétegeken elvégzett vizsgálatok mellett oly módon is elvégeztem a vizsgálatokat, hogy a bemeneti adatokat előzetesen harmadfokú polinomiális egyenletekké alakítottam és azok koefficienseit használtam változókként. A vizsgálatokat mind a réteg alapú adatokon, mind a kumulatív görbék esetén elvégeztem. Az előzetes vizsgálatok alapján a tulajdonság jó lehetőséget biztosított a két bemeneti állomány használhatóságának összehasonlítására, mivel jól megfigyelhető a szelvények különbözősége, illetve azok csoportokba rendeződése.

A Silhoutte vizsgálatokat elvégezve a két állományon hasonló adatokat kaptam (6. és 7. ábra) mindkét állomány esetében két klaszter megjelenésével. A görbék esetében a 2. klaszter nagyobb sziluett szélességgel (6. ábra), de még így is 0.5 alatti értékkel, mely azt mutatja, hogy az egyedek tagsága a csoportban nem egyértelmű.

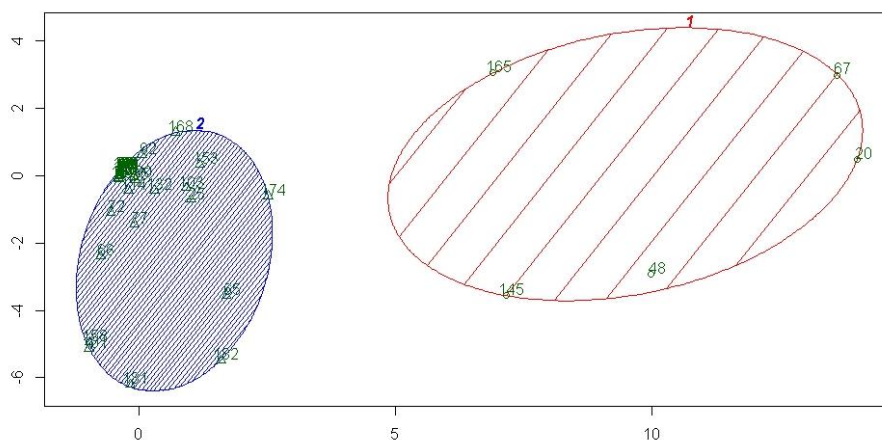


6. ábra. Silhouette analízis a polinomiális görbékkel jellemzett Agyagbemosódásos talajok szelvények durva vázrész tartalmán



7. ábra. Silhouette analízis a rétegekkel jellemzett Agyagbemosódásos talajszelvények durva vázrész tartalmán

A klaszteranalízis eredményeként két jól elkülönülő csoport látható (8. ábra). Nagy számban vannak a durva vázrészt nem tartalmazó egyedek, illetve ebbe a csoportba tartoznak a kevés durva vázrészt tartalmazók. Ettől a csoporttól elkülönülve található azok a szelvények, amelyekben a vázrész mennyisége az adatbázis alapján visszakeresve egyes rétegekben eléri vagy meghaladja a 30 %-ot, mélységtől függetlenül. Ezek alapján javasolt a durva vázrészek, mint alsóbb szinten elkülönítő tulajdonság felvétele az agyagbemosódásos talajok kritériumai közé.



8. ábra. Agyagbemosódásos talajok klaszteranalízisének eredménye, a durva vázrészek mennyiségét vizsgálva, illesztett polinomiális görbék alapján. A két fő komponens, melyek mentén az eredmény ábrázolásra került a változatosság 90 %-át írja le

A 30 %-os durva vázrész mennyiség matematikailag határértéknek bizonyult a klasztereződés szempontjából, ekkora mintaszám mellett azonban annak vizsgálata és megállapítása, hogy egyéb tulajdonságokat (vízgazdálkodás, művelhetőség) mely durva vázrész mennyiség befolyásol, nagyobb súlyt kell, hogy kapjon az osztályozási rendszerben a tényleges határértékek megállapításánál.

A vizsgálatok alapján elmondható, hogy matematikai módszerek használatával lehetőség van adatbázisok segítségével a talajosztályozási egységeken belüli, az elkülönítéshez jelentős tulajdonságok definiálására, illetve azok vizsgálatára mélység szerinti megjelenés alapján. A vizsgálatok több esetben harmonizáltak a korábban is elkülönített típusok tulajdonságaival

(kovárványos barna erdőtalajok homoktartalom alapján, vagy a pangóvízes talajok elkülönülése redox tulajdonságok alapján). A további, vizsgálatok feltárhatnak további olyan tulajdonságokat, melyeket érdemes a javasolt hazai osztályozás alacsonyabb szintjein kiemelni, így a vizsgálatokat javasolom kiterjeszteni a többi típusra is.

## **Talajtérképek létrehozására modern terepi felvételezési módszerek és tematikus talajtérképek használatával**

### Osztályozás a Global Soil Map specifikáció alapján taxonómiai távolság számítással

A Köves-Sziklás típust adathiány miatt a további vizsgálatokba nem tudtam bevonni, mivel a centroid számításhoz nem állt rendelkezésre megfelelő mennyiségű adat.

Az eredmények alapján elmondható, hogy a Homoktalajokat leszámítva a GSM specifikációt használva az új magyar osztályozás típusainak származtatása talajtérkép létrehozására taxonómiai távolság számítással nagy hibát eredményez. Az átlagos pontosság (taxonómiai távolság szerint legközelebbi) 250 validáló szelvény esetében nem éri el a 30 %-ot.

### Osztályozás a Global Soil Map specifikáció alapján random forest módszerrel

Az átlagos megbízhatóság a módszer esetében meghaladja a 75 %-ot (4. táblázat). Ugyanakkor egyes típusok esetében ez meglehetősen alacsony (50 % körüli). A Szolonyec és a Karbonát talajok félreosztályozása elsősorban a típust meghatározó fő paraméterek hiányával magyarázható. A Közethatású talajok esetében a félreosztályozás alapja az, hogy a fő meghatározó paraméterek, melyek a fa „tetején” válnak szét, nem a durva vázrész mennyiségére vonatkoznak. Emellett problémát jelenhet az is, hogy a Közethatású talajok szelvényének száma limitált az adatbázisban (14). Összességében azonban elmondható, hogy a módszer alkalmas lehet a GSM specifikáció alapú térképek osztályozására a javasolt hazai talajtípusok alapján. Emellett kiemelendő, hogy a modellbe környezeti paraméterek, úgymint domborzat, vegetáció, elhelyezkedés, talajképző kőzet nem kerültek bevonásra. Vélhetően a környezeti paraméterek bevonásával a módszer pontossága tovább javítható.

4. táblázat. Global Soil Map specifikáció alapú random forest számítás megbízhatósága

Talajtípus	Megbízhatóság
Agyagbemosódásos talaj	79%
Barnaföld	71%
Duzzadó agyagtalaj	88%
Homoktalaj	82%
Hordaléktalaj	62%
Karbonát talaj	54%
Közethatású talaj	58%
Mezőségi talaj	69%
Réti talaj	76%
Szolonyec talaj	60%
Átlag	75%

---

## Következtetések és a javaslatok

---

A nemzetközi trendek, Európai uniós elvárások és a hazai talajadat iránti igény kielégítésére elengedhetetlen az ország archív adatállományának összegyűjtése, minőségi vizsgálata, harmonizálása, rendszerbe foglalása. Hazánk a gazdag adatállománynak és a képzett talajtani szakembereinek köszönhetően többek között mezőgazdasági technológiák fejlesztésével, a legújabb kutatások gyakorlati munkában való átültetésével tűnhet ki nemzetközi szinten. Ennek eléréséhez azonban szükségszerű a hazai talajtanos szervek, kutató központok, egyetemi és egyéb műhelyek összefogása, egy közös cél érdekében való együttes fellépése. Ennek első lépése lehetne egy olyan egységes adatbázis létrehozása, mely alapját képezheti a kutatásoknak, döntéshozásnak és adatszolgáltatóként működik mind az állampolgárok, mind az egyéb tudományos területek kutatói számára.

Egy ilyen rendszer létrehozását vizsgáltam meg és hoztam létre elemeit, melyek alkalmasak az egy rendszerbe integrálásra. Vizsgálataim alapján megállapítottam, hogy az adatharmonizálás és adatellenőrzés automatizálható, így mind az archív adatok integrálása, mind az új, egy esetleges sztandardizált terepi felvételezési programból származó adatok ellenőrzése, egy struktúrára rendezése gyorsan és költséghatékonyan elvégezhető. Javaslom egy ilyen rendszer szabványosítását, egy alap minőségellenőrzési algoritmus sorozat lefektetését, mely alapja lehet a hazai talajadatok minőségellenőrzésének.

Létrehoztam egy olyan adatstruktúra modellt, amely alkalmas lehet a jelenlegi, hazánkban alkalmazott és vizsgált struktúráknál több, részletesebb adat tárolására, illetve a több adatforrásból származó adatok integrálására, azok harmonizálására. Az adatstruktúra a leíró adatok harmonizálása és a metaadatok tárolása mellett tárolja a laboratóriumi adatok harmonizálására szolgáló faktorokat, lehetővé téve a könnyebb átszámítást, a talajtanban és a módszerek diverzitásában nem jártas felhasználó számára is. Egy hazai harmonizált adatbázis struktúrájának véglegesítése esetén javaslom ezen adatok integrálását.

Az archív adatok korrelálásának programozása mellett létrehoztam a javasolt új magyar osztályozás típusaira való konvertálást szolgáló osztályozó algoritmust, amely amellelt, hogy alapját képezheti az osztályozás alacsonyabb szintjeinek fejlesztésének, illetve az archív adatok osztályozásával az ország újratérképezésének, lehetőséget nyújt az archív adatok konvertálására, így azok mentésére.

Az újraosztályozással lehetőség nyílt a javasolt típusok vizsgálatára, az osztályozás alacsonyabb szintjeinek fejlesztésére. Egyszerű statisztikai módszerekkel, Silhouette- és klaszteranalízissel megvizsgáltam a típuson belüli fontos elkülönítő tulajdonságok. Ezek alapján javaslom a hazai osztályozás fejlesztésénél a matematikai módszerek alkalmazását. A módszer segítségével feltárhatók a fontos tulajdonságok, illetve azok mélység szerinti klasztereződése, azonban a matematikai módszerek csupán segítségként kell, hogy szolgáljanak, az osztályozás gyakorlati alkalmazhatósága, annak egységessége kell, hogy vezesse a határértékek, számszerűsített definíciók létrehozását.

A taxonómiai távolság számítás sikeres alkalmazásával talajosztályozási rendszerek közötti harmonizációra egy olyan, más tudományágakban is alkalmazott módszert használtam, mely alkalmas lehet archív adataink korrelálásra, illetve a javasolt osztályozási rendszer és a WRB, vagy a jelenlegi osztályozás kapcsolatának vizsgálatára is. A módszer alkalmazása a további harmonizálási és osztályozási rendszer-fejlesztési munkákban javasolt, azok kiterjesztésével további talajtípusokra.

Rámutattam annak lehetőségére, hogy matematikai módszerekkel talajosztályozási információ, pontosabban a javasolt talajtípusok származtathatók tematikus, GlobalSoilMap specifikációjú adatokból. A hazai adatgazdagságnak köszönhetően a projektben használt módszerek mellett a térképek valószínűleg nagyobb megbízhatósággal származtathatók, mint más, adathiányos területeken. Ennek köszönhetően a hazai talajtípus térképek aktualizálása a

vizsgált módszerekkel elvégezhető. A módszer alapja az archív, ellenőrzött, vagy újonnan felvételezett talajadatok megléte. Ennek megvalósíthatóságának, így a hazai talajtérképi állomány megújításának vizsgálatát javaslom.

A dolgozatomban felvázolt lépések alapján egy sztandardizált hazai rendszer létrehozásával, nemcsak a hazai talajtanos közösség nyerne, hanem a közösség megítélése is javulna a társtudományok kutatói szemében (sztandardizáld módszerek és eljárások létrehozásának igénye). A hazai talajtanos közösség potenciáljával lehetőség nyílna egy olyan rendszer létrehozására, mely példaként szolgálhatna a világban és hazánk visszakerülhetne a korábban elfoglalt, elismert pozíciójába a talajtan világtérképén.

---

## Új tudományos eredmények

---

1. Létrehoztam egy adatharmonizáló algoritmust és egy minőségellenőrzési rendszert, továbbá egy osztályozó algoritmust, melyek alkalmasak a hazai TIM nomenklátúra alapján készült archív adatok nemzetközi megfeleltetésére, illetve a javasolt hazai talajosztályozási rendszer egységeire konvertálni azokat.
2. Alkalmaztam a taxonómiai számítás módszerét különböző talajosztályozási rendszerek típusainak korrelálására, mind koncepció, mind centroid alapon.
3. Matematikai módszerek használatával kidolgoztam a hazai osztályozási rendszer alacsonyabb osztályozási szintjeinek fejlesztéséhez használható eszközt.
4. Matematikai módszerekkel réteg alapú talajinformációkból sikeresen származtattam a javasolt hazai osztályozás típusait, ezzel megalapozva egy, a javasolt osztályozási rendszer típusain alapuló új hazai talajtérkép létrehozását.

---

## Az értekezés témaköréhez kapcsolódó publikációk

---

### FOLYÓIRATCIKKEK

#### 1. Tudományos folyóiratokban megjelent (közlésre elfogadott), lektorált, teljes szövegű tudományos közlemény

##### 1.1. Idegen nyelvű, impakt faktoros folyóiratban (WEB OF SCIENCE szerint)

###### 1.1.2. Külföldi kiadású

LÁNG, V., FUCHS, M., WATNER, I. & MICHÉLI, E., 2013. Taxonomic distance metrics, a tool for soil correlation. *Geoderma* **192**, 269-276.  
<http://dx.doi.org/10.1016/j.geoderma.2012.07.023>  
Impact factor: 2,318

##### 1.2. Idegen nyelvű, nem impakt faktoros folyóiratban

###### 1.2.1. Hazai kiadású

LÁNG, V., FUCHS, M., WALTNER, I. & MICHÉLI, E., 2010. Taxonomic distance measurements applied for soil correlation. *Agrokémia és Talajtan*, **59**, 1. 57-64.

###### *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.

MICHÉLI, E., SZABARI, SZ., LÁNG, V., WALTNER, I., DOBOS, E. 2009: Applying diagnostic categories of the World Reference Base for Soil Resources (WRB) for identifying and delineating risk areas of salinization and sodification, Proceedings of the VIII. Alps-Adria Scientific Workshop, Neum, Bosnia-Herzegovina, 27 April – 2 May, 2009. *Cereal Research Communications*, Vol. **37** No. 3. 399-402. pp

### 1.3. Magyar nyelvű, nem impakt faktoros hazai folyóiratban

FUCHS M., WALTNER I., SZEGI T., LÁNG V. & MICHÉLI E., 2011. A hazai talajtípusok taxonómiai távolsága a képződésüket meghatározó folyamattársulások alapján. *Agrokémia és Talajtan*, **60**, 1. 33-44.

WALTNER I., FUCHS M., MICHÉLI E., LÁNG V., 2012. Hazai archív talajadatok beillesztésének lehetőségei nemzetközi adatbázisokba. *Agrokémia és Talajtan*, **61**, 2. 263-76.

## KONGRESSZUSI KIADVÁNYOK

### 4. Kongresszusi kiadványokban megjelent közlemények (nyomtatott formában v. elektronikus adathordozón – kizárólag az ISBN, ISSN vagy más, hitelesített kiadványaira vonatkozóan)

#### 4.1. Teljes szövegű közlemény, alkalmi (nem periodika jellegű) kongresszusi kiadványban, idegen nyelven, lektorált formában megjelentetve:

LÁNG, V., FUCHS, M., WALTNER, I. & MICHÉLI, E., 2010. Pedometrics applications for correlation of Hungarian soil types with WRB. In: Gilkes R.J., Prakongkep N. (eds.), Proceedings of the 19<sup>th</sup> 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.

#### 4.2. Teljes szövegű közlemény, alkalmi (nem periodika jellegű) kongresszusi kiadványban, magyar nyelven, lektorált formában megjelentetve

FUCHS, M., SZÖCS, A., LÁSZLÓ, P., LÁNG, V., & MICHÉLI, E., 2008. A Bodroghöz vízhatás alatt álló talajainak osztályozási problémái. Talajvédelem Különszám. Talajtani vándorgyűlés, Nyíregyháza, 2008. május 28-29. Talajvédelmi Alapítvány, Bessenyei György Könyvkiadó, Nyíregyháza, 595-601. p. ISBN 978-963-9909-03-8

---

## Felhasznált irodalom

---

AFSIS (2013): African Soil Information Service projekt honlap , 2013, November, (<http://www.africasoils.net/about/rationale>)

BATJES N.H.(2008): ISRIC-WISE Harmonized Global Soil Profile Dataset (Ver. 3.1). Report 2008/2, ISRIC – World Soil Information, Wageningen, The Netherlands.

BIDWELL O. W., HOLE F. D. (1964a): Numerical taxonomy and soil classification. *Soil Science* 97. 58–62.

BIDWELL O. W., HOLE F. D., (1964b): An experiment in the numerical classification of some Kansas soils. *Soil Sci. Soc. Amer. Proc.* 26. 263–268.

BREIMAN L. (2001): Random Forests, *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

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.
- EKLUND A. (2011): Color-based plots for multivariate visualization (squash) package for R version 1.0.1
- EURÓPAI TANÁCS (2007): Az Európai Parlament és a Tanács 2007/2/EK irányelve (2007. március 14.) az Európai Közösségen belüli térinformációs infrastruktúra (INSPIRE) kialakításáról. Az Európai Unió Hivatalos Lapja, L108/1
- e-SOTER (2008a): e-SOTER projekt honlap. <http://www.esoter.net/?q=category/homepage/welcome> (2013.11.02)
- e-SOTER (2008b): Description of Work (DOW), Proposal No. 211578, 19-Jul-2008. [http://www.esoter.net/sites/default/files/files/DoW\\_e-SOTER\\_19jul.pdf](http://www.esoter.net/sites/default/files/files/DoW_e-SOTER_19jul.pdf)
- FAO (2006): Guidelines for soil description. 4th edition. FAO, Rome.
- FEIDEN K. ÉS MUNKACSOPORT VEZETŐK (2012): ECP-2008-GEO-318004 GS Soil, „Assessment and strategic development if INSPIRE compliant geodata services for European soil data” Final report, 03. June 2012 ([http://www.gssoil-portal.eu/Best\\_Practice/GS\\_SOIL\\_D1.5.3\\_3\\_annual\\_public\\_report.pdf](http://www.gssoil-portal.eu/Best_Practice/GS_SOIL_D1.5.3_3_annual_public_report.pdf))
- GLOBALSOILMAP.NET (2011): Specifications GlobalSoilMap.net products Version 2.1, Report 1. July, 2011 ([http://www.globalsoilmap.net/system/files/GlobalSoilMap\\_net\\_specifications\\_v2\\_0\\_edited\\_draft\\_Sept\\_2011\\_RAM\\_V12.pdf](http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_draft_Sept_2011_RAM_V12.pdf))
- HARTIGAN J. A., WONG M. A. (1979): A K-means clustering algorithm. *Applied Statistics* 28, 100–108.
- HOLE F. D., HIRONAKA M. (1960): An experiment in ordination of some soil profiles. *Soil Sci. Soc. Amer. Proc.* 24. 309–312.
- [http1: http://ncsslslabdatamart.sc.egov.usda.gov/](http://ncsslslabdatamart.sc.egov.usda.gov/)
- [http2: http://en.wikipedia.org/wiki/Python\\_\(programming\\_language\)](http://en.wikipedia.org/wiki/Python_(programming_language))
- IUSS WORKING GROUP WRB (2006): World Reference Base for Soil Resources 2006. World Soil Resources Reports, No. 103. FAO. Rome.
- MACQUEEN J. B. (1967): Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press. pp.281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.
- MCBRATNEY A. B., ODEH I. O. A., BISHOP T. F. A., DUNBAR M. S., SHATAR T. M. (2000): An overview of pedometric techniques for use in soil survey. *Geoderma*. 97. 293–327.
- MCFERRIN L. (2013): Statistical Analysis Tools for High Dimension Molecular Data (HDMD) package for R version 1.2
- MICHÉLI E., LÁNG V., FUCHS M., WALTNER I., SZEGI T., DOBOS E., SERES A., VADNAI P., VAN ENGELEN V., DIJKSHOORN K., DAROUSSIN J., EBERHARDT E., SCHULER U., ZADOROVA T., KOZAK J., HANNAM J., HALLETT S., ZHANG G., YUGUO Z., BALAGHI R., MOUSSADEK R. (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
- MICHÉLI E., FUCHS M., HEGYMEGI P., STEFANOVITS P. (2006): Classification of the Major Soils of Hungary and their Correlation with the World Reference Base for Soil Resources (WRB). *Agrokémia és Talajtan* 55 (1) 19–28. p.
- MICHÉLI E., FUCHS M., LÁNG V., SZABÓNÉ KELE G. (2013a): Alapelvek, osztályozó kulcs. Vitaanyag a Magyar Talajtani társaság 2013. június 20-i ülésére. [www.talaj.hu/magyar/szakosztalyok/Talajgenetika](http://www.talaj.hu/magyar/szakosztalyok/Talajgenetika)

- MINASNY B., MCBRATNEY A. B. (2007): Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* **142**. 285–293.
- MINASNY B., MCBRATNEY A., HARTEMINK A.E. (2010): Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma*, 155(3-4), 132-139. p.
- PEARSON K. (1901): On Lines and Planes of Closest Fit to Systems of Points in Space (PDF). *Philosophical Magazine* 2 (11): 559–572.
- R DEVELOPMENT CORE TEAM (2009): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- ROUSSEEUW P. J. (1987): Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics* 20: 53–65. doi:10.1016/0377-0427 (87) 90125-7.
- SARKAR P. K., BIDWELL O. W., MARCUS L. F. (1966): Selection of characteristics for numerical classification of soils. *Soil Sci. Soc. Amer. Proc.* **30**. 269–272.
- STEFANOVITS P. (1981): Talajtan. Mezőgazdasági Kiadó. Budapest
- SZABOLCS I. (szerk.) (1966): A genetikus üzemi talajtérképezés módszerkönyve. OMMI. Budapest.
- SZABÓNÉ KELE, G. (2013a): Javaslat helyszíni talajfelvételezés általános módszertanára. Vitaanyag a Magyar Talajtani társaság 2013. június 20-i ülésére. [www.talaj.hu/magyar/szakosztalyok/Talajgenetika](http://www.talaj.hu/magyar/szakosztalyok/Talajgenetika)
- TIM (TALAJVÉDELMI INFORMÁCIÓS ÉS MONITORING RENDSZER) (1995): Módszertan. Földművelésügyi Minisztérium Növényvédelmi és Agrár-környezetgazdálkodási Főosztály, Budapest.
- WALTNER I., FUCHS M., MICHELI E., LÁNG V., (2012): Hazai archív talajadatok beillesztésének lehetőségei nemzetközi adatbázisokba. *Agrokémia és Talajtan*, 61, 2. 263-76.