

dc_16_10

MTA Doktori Pályázat
Doktori értekezés

A cönológiai adatbázisok alkalmazása a vegetációkutatásban:
módszerfejlesztések és esettanulmányok

Botta-Dukát Zoltán

Vácrátót, 2010

Tartalomjegyzék

1. Bevezetés.....	5
1.1. A dolgozat szerkezete.....	11
2. CoenoDAT Referencia Adatbázis.....	12
2.1. Bevezetés.....	12
2.2. Az adatbázis létrehozására kidolgozott módszer.....	12
2.2.1. Az új felvételek készítése	12
2.2.2. Az archív felvételek kiválasztása	14
2.2.3. Fajlisták és társulási lista	15
2.3. Az adatbázis jelenlegi helyzete.....	15
3. A cönológiai adatbázisok felhasználásának lehetőségei és korlátai	17
3.1. Preferenciális mintavétel – elméleti megfontolások	18
3.1.1. Alapsokaság és minta	18
3.1.2. Random mintavétel.....	19
3.1.3. Mit jelent a preferenciális mintavétel a cönológiában?	20
3.1.4. A statisztikai tesztek formális és szokványos feltételei.....	20
3.1.5. Záró megjegyzések	22
3.2. A preferenciális és a random mintavétel összehasonlítása a gyakorlatban	22
3.2.1. Anyag és módszer	23
3.2.2. Eredmények	24
3.2.3. Diskusszió.....	28
3.3. Következtetések	28
4. Zajsűrés	30
4.1. Anyag és módszer.....	30
4.1.1. Zajsűrés metrikus ordinációval.....	30
4.1.2. Elemzett adatok	31
4.1.3. A vizsgált klasszifikációk	33
4.1.4. A klasszifikációk „jóságának” mérése	33
4.2. Eredmények.....	33
4.3. Értékelés	34
5. Fidelitás	36
5.1. Fidelitás mérőszámok: egy felvételtcsoport.....	36
5.1.1. A fidelitás mérőszámok összehasonlítása.....	40
5.2. Páros összehasonlítások helyett átfogó mérőszámok	46
5.3. A fajok válogatóképessége	48
6. A klasszifikációk értékelése	55
6.1. A klasszifikációk értékelésére használt módszerek áttekintése.....	55
6.1.1. A csoportosítás mennyire tükrözi a távolságmátrixban rejlő információt?.....	56
6.1.2. A csoportok összekötöttsége, tömörsége és elkülönülése.....	58

6.1.3. Az eredmények robosztussága.....	61
6.1.4. Az eredmények stabilitása	62
6.1.5. Az eredmények repetitivitása	63
6.1.6. A csoportok értelmezhetősége	65
6.1.7. A csoportosítás prediktív ereje	68
6.1.8. Összehasonlítás random adatokkal.....	68
6.1.9. A klasszifikációk értékelésére szolgáló módszerek csoportosításának egyéb szempontjai.....	69
6.2. Az értékelő módszerek alkalmasságának összehasonlítása.....	71
6.2.1. Anyag és módszer	71
6.2.2. Eredmények	73
6.2.3. Diszkusszió.....	74
6.2.4. Következtetések, javaslatok	86
6.3. Új módszer a csoportok repetitivitásának vizsgálatára	87
6.3.1. A módszer bemutatása	87
6.3.2. Példák a módszer alkalmazására.....	88
6.3.3. Következtetések	92
7. Fajkészlet becslése	94
7.1. Beals simítás	95
7.2. Transzformáció bináris skálára.....	96
7.3. Anyag és módszer.....	97
7.4. Eredmények.....	97
7.5. Diszkusszió.....	99
8. Niche szélesség.....	100
8.1. A nicheszélesség mérésére javasolt béta diverzitás indexek áttekintése.....	101
8.2. A távolság alapú béta-diverzitás indexek hátrányai.....	102
8.3. A fajkészlet méretének hatása a béta-diverzitásra	103
8.4. A nicheszélesség robosztus becslése	107
8.5. A módszerválasztás hatása az eredményekre.....	107
8.6. Következtetések	108
9. Mocsárrétek klasszifikációja.....	109
9.1. Bevezetés.....	109
9.2. Anyag és módszer.....	110
9.2.1. Vegetációadatok.....	110
9.2.2. Klímaadatok.....	110
9.2.3. Adatelemzés	111
9.3. Eredmények és megvitatásuk	111
9.3.1. A felvételek csoportosítása	111
9.3.2. Ordináció és a vegetáció kapcsolata a makroklimával	128
10. Nicheszélesség és az area méret kapcsolata.....	133
10.1. Anyag és módszer.....	133

10.2. Eredmények.....	134
10.3. Diskusszió.....	136
11. Összegzés és kitekintés	138
12. Köszönetnyilvánítás.....	140
13. Irodalomjegyzék	141
14. Az értekezés témakörében megjelent publikációk	158
15. Mellékletek.....	159
1. melléklet: A CoenoDat adatbázis számára gyűjtött új cönológiai felvételek fejlécében szereplő adatok.....	159
2. Melléklet: A khi-négyzet statisztika és az R^2 közti kapcsolat	162
3. melléklet: A módosított Rand-index	165

1. Bevezetés

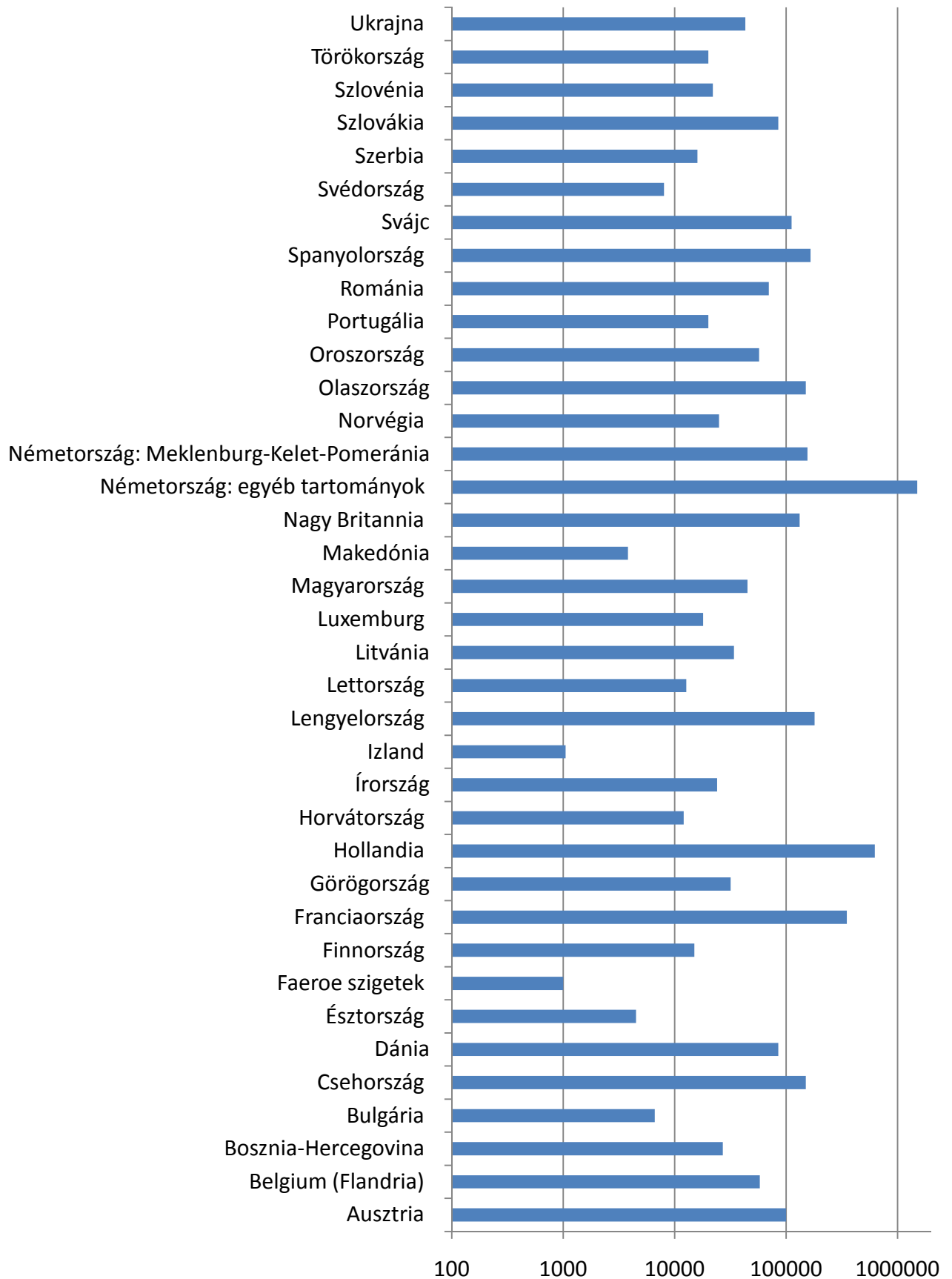
Egy-egy tudományterület fellendülése az új elméletek, vagy az új adatforrások megjelenéséhez kötődik. Az új elmélet magyarázatot adhat a korábban is ismert, de a régebbi elméletek kereteibe nem illeszkedő adatokra (például az általános relativitáselmélet alapján jól magyarázható a Merkúr pályájának eltérése a korábbi elméletek alapján várttól), de ugyanakkor újabb adatok gyűjtésére is inspirál. A hatás azonban nem csak egyirányú. Hubble felfedezése előtt a táguló világegyetem csak Einstein kozmológiai modelljeinek egy lehetséges, de valószínűtlennek tartott megoldása volt. A vörös-eltolódás felfedezésével, amit a rádiócsillagászat megjelenése tett lehetővé, viszont minden kozmológiai elmélet központi elemévé vált.

A fenti – szándékosan igen távoli – példában az új adatforrások megjelenése az adatgyűjtés módszerének megváltozását és a hozzáférhető adatok körének ebből következő kibővülését jelentette. A vegetációtudományban az adatgyűjtés módjában az elmúlt 100 évben nem volt ilyen drámai változás. Már a 20. század elején készültek cönológiai felvételek, amelyek a mintaterület fajlistáját és a fajok tömegességét rögzítették, és mai napig is ez a leggyakrabban használt adattípus. A módszer részletei – mintavételi egység mérete, becslési skála stb. – ugyan változhattak, de alapvetően ma is ugyanolyan szerkezetű adatokat gyűjtünk.

Jelentős fejlődés zajlott le viszont az adatok kezelésének módjában, és ezzel új adatforrás jelent meg: a vegetációs adatbázisok. A felvételek kézi átrendezése csak kisszámú felvétel kezelését tette lehetővé. Sok esetben a felvételek inkább csak a levont következtetések illusztrációi a cikkekben, és nem az információk elsődleges forrásai. A szintetizáló munkák (pl. Borhidi 1963, 1965) is a fajok konstanciáin és nem az egyedi felvételeken alapultak. Az elemzésre használható számítógépes programok az 1960-as évektől indultak fejlődésnek, de csak az 1980-as évektől, a személyi számítógépek megjelenésével váltak széles körben alkalmazott eszközzé a vegetációkutatásban (Mucina & van der Maarel 1989). Kezdetben még az elemezhető adatsor mérete korlátozott volt, így előfordult, hogy az elemzés objektumai nem az egyes felvételek, hanem a korábban elkülönített szüntaxonok voltak (pl. Török et al. 1989). A személyi számítógépek kapacitásának gyors bővülésével azonban a feldolgozó programok egyre kevésbé korlátozták a feldolgozható adatok mennyiségét. Ma már egy több száz felvételtől álló adatsor elemzésénél lényegesen nagyobb munka az adatbevitel, mint magának az elemzésnek a végrehajtása. Ezért érdemes az adatbevitelt csak egyszer elvégezni, a digitalizált adatokat adatbázisokban tárolni és az elemzéseknél elsősorban az adatbázisokban tárolt adatokból dolgozni. Az adatbázisokban tárolt felvételek lehetővé teszik, hogy több száz vagy esetleg több ezer felvétel alapján vizsgáljuk meg az aktuális problémát, esetleg távoli vidékek adatait is bevonva.

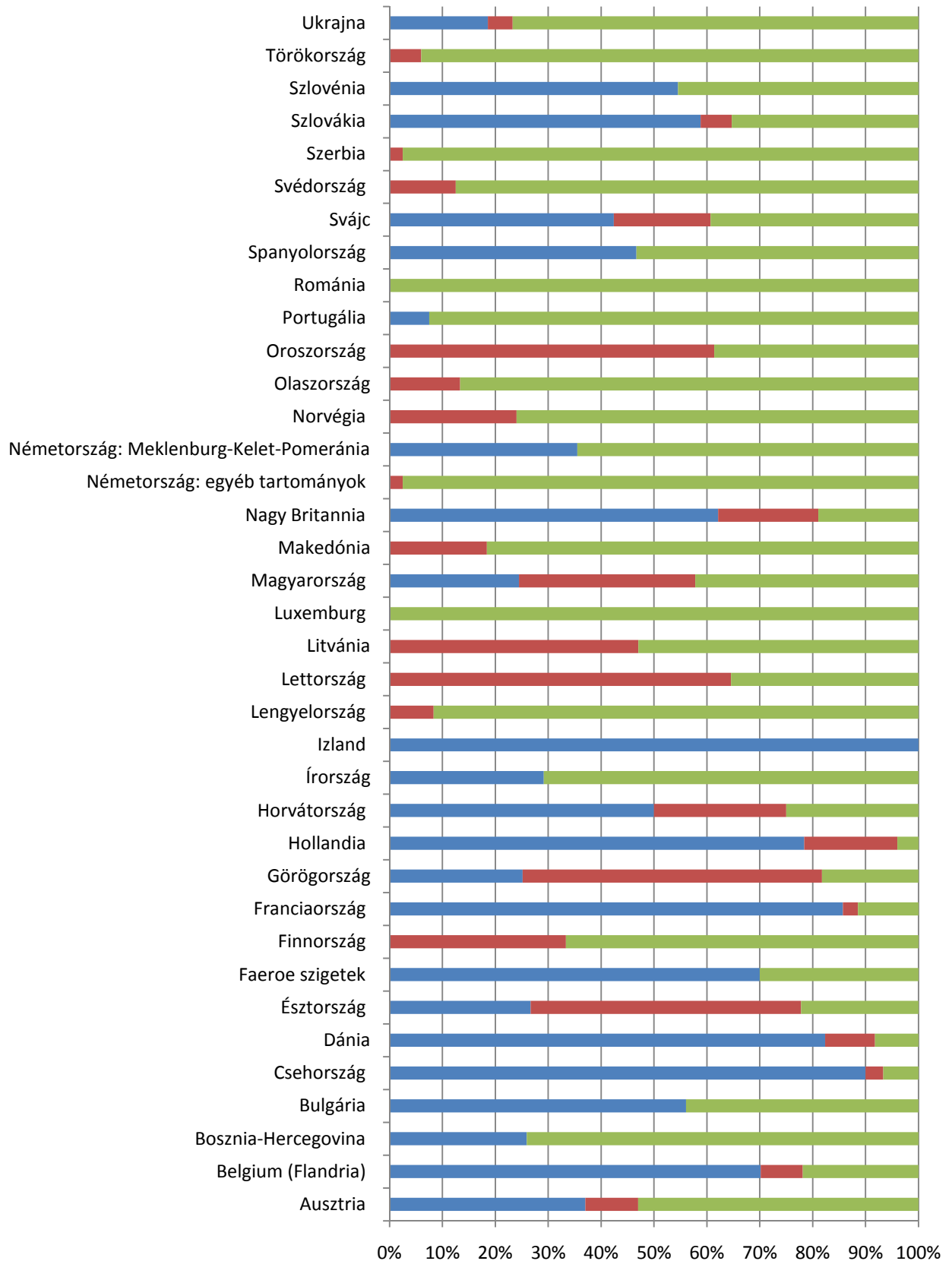
Schaminée és munkatársai (2009) becslése szerint Európában több mint 4,3 millió vegetációs felvétel készült, amelyből legalább 1,8 millió már elektronikus adatbázisban is hozzáférhető. A legtöbb felvétel Németországban, Hollandiában és Franciaországban készült,

dc_16_10



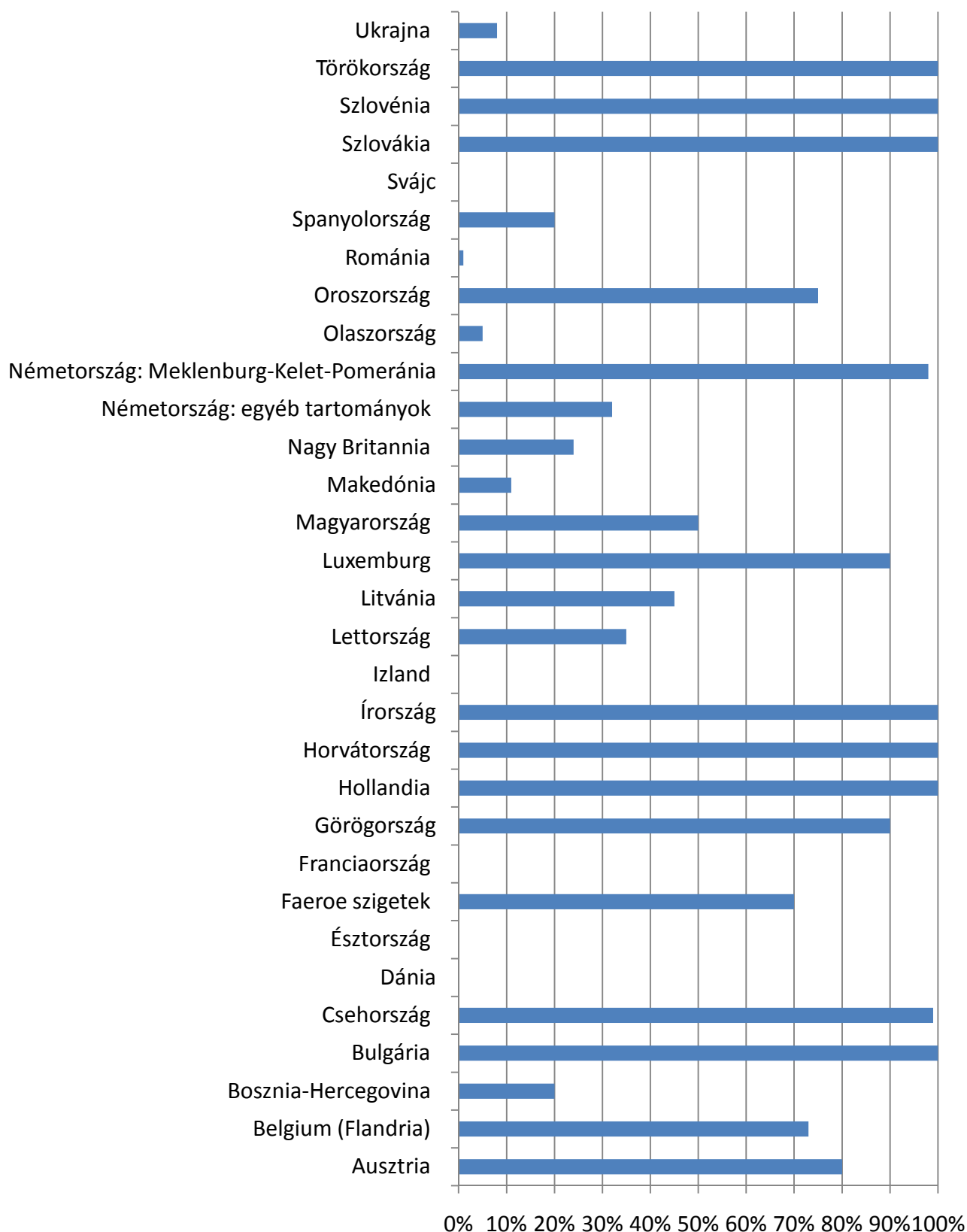
1. ábra: A cönológiai felvételek száma országonként (illetve régióként) Schaminée és munkatársai (2009) bebecslése alapján

dc_16_10



2. ábra: A központi adatbázisban tárolt (kék oszlop), a más adatbázisokban megtalálható (piros oszlop) és a nem digitalizált (zöld oszlop) felvételek aránya az európai országokban

dc_16_10



3. ábra: A TurboVeg formátumban tárolt felvételek aránya a különböző európai országokban

de több mint 100 ezerre becsülhető a felvételek száma Lengyelországban, Spanyolországban, Csehországban, Olaszországban, Nagy Britanniában, Svájcban és Ausztriában is (1. ábra). Az adatbázisban tárolt felvételek aránya országonként erősen változó (2. ábra). A legnagyobb cönológiai adatbázisok Hollandiában, Franciaországban, Csehországban és Nagy Britanniában vannak. A felvételek száma viszonylag alacsony azokban az országokban (Skandinávia, egyes balkáni és kelet-európai országok), ahol a Braun-Blanquet cönológiai iskolának nincsenek jelentős hagyományai. A felvételek többségét ugyanis ennek az irányzatnak a követői készítették. Ez alól kivétel Nagy Britannia, ahol az adatok a jelentős része a Countryside Survey (Smart et al. 2003) eredménye. Emellett több országban is jelentős adatforrások az erdőfelmérések, erdő leltárok.

A korábbi becslések (Rodwell 1995, Ewald 2001) mind a rendelkezésre álló felvételek számát, mind az adatbázisok méretét lényegesen alacsonyabbra becsülték. Ennek oka, hogy az adatbázisok építése több országban csak abban az időben vagy később indult el, és a felvételek számának pontos becslése adatbázisok hiányában nagyon nehéz.

A nemzeti adatbázisok kialakításához (1. táblázat) jelentős lökést adott a Nemzetközi Vegetációtudományi Társaság (IAVS) European Vegetation Survey (EVS) nevű munkacsoportjának megalakulása 1992-ben (Mucina et al. 1993, Rodwell et al. 1995). A munkacsoport célja az európai vegetáció egységes szemléletű leírása, amelyből először az európai vegetáció asszociációosztály (Mucina 1997a), majd a részletesebb, asszociációcsoport szintű (Rodwell et al. 2002) áttekintése készült el. Ezek az áttekintések még az egyes országok vegetációját áttekintő művek szintézisei, mert a teljes európai vegetációt lefedő adatbázis még ma sem áll rendelkezésre. Ugyanakkor, a munkacsoport tevékenységének köszönhetően születtek adatbázisokon alapuló országos (pl. Chytrý 2007, 2009, Jarolínek & Šibík 2008) és regionális (Berg et al. 2004) vegetáció-áttekintések, illetve egy-egy vegetációtípusra vonatkozó európai léptékű szintézisek (pl. Zuidhoff et al. 1995). Az EVS 1994-es találkozóján a TurboVeg programot (Hennekens & Schaminée 2001) választotta a vegetációs adatok kezelésének nemzetközi standardjává. Ennek nyomán a program használata a legtöbb európai országban meghonosodott (3. ábra), megkönnyítve a nemzetközi adatcserét. Jelenleg az európai adatbázisokban tárolt felvételek kb. 60%-át ilyen rendszerű adatbázisokban találjuk (Schaminée et al. 2009). Részben az erős cönológiai hagyományoknak, részben a munkacsoport tevékenységének köszönhetően az európai adatbázisok tágabb kitekintésben is a legnagyobbak közé tartoznak. Nagyobb adatbázisok vannak még az USA-ban (www.vegbank.org), Új-Zélandon és Dél-Afrikában (Mucina et al. 2000).

Az EVS munkacsoport célja egy európai léptékű, modern szüntaxonómiai szintézis, de az adatbázisokban tárolt felvételeket nem csak ilyen célra lehet felhasználni, hanem más ökológiai kutatásoknak is hasznos eszközei lehetnek (Bekker et al. 2007). Felhasználhatók például a klímaváltozás hatásának detektálására (Kienast et al. 1998, Duckworth et al. 2000, Lenoir et al. 2008), fajok termőhely-preferenciáinak vizsgálatára (Hölzel 2003, Coudun & Gégout 2007) és ehhez kapcsolódóan a nicheszélesség becslésére (lásd a 8. fejezetben), a diszperzál-limitáltság (Ozinga et al. 2004, Ozinga, Schaminée, et al. 2005, Ozinga, Hennekens,

dc_16_10

et al. 2005), vagy társulások előzőnölhetősége (Chytrý et al. 2008, 2009) vizsgálatára, hosszú távú változások detektálására (Ruprecht & Botta-Dukát 1999, Sýkora et al. 2002, Holeksa & Wozniak 2005) vagy a környezeti tényezők diverzitásra gyakorolt hatásának leírására (Chytrý et al. 2003, Ewald 2003, Rédei et al. 2003).

1. táblázat: A fontosabb nemzeti és regionális adatbázisok Európában

Ország	Forrás
Ausztria	http://vegedat.vinca.at
Belgium	http://www.inbo.be/content/page.asp?pid=BIO_NT_vlavedat
Csehország	Chytrý & Rafajová (2003); http://www.sci.muni.cz/botany/vegsci/dbase.php?lang=en
Dánia	http://www.danveg.dk , http://www.naturdata.dk
Franciaország	SOPHY adatbázis: Brisse és munkatársai (1995); http://jupiter.u-3mrs.fr/~msc41www EcoPlant adatbázis: Gégout és munkatársai (2005); http://efdp.nancy-engref.inra.fr/bd/ecoplant.htm Francia Nemzeti Erdőleltár (Inventaire Forestier National, IFN): http://www.ifn.fr/spip
Hollandia	http://www.synbiosys.alterra.nl/lvd
Horvátország	Stančić (2008)
Írország	http://nationalvegetationdatabase.biodiversityireland.ie
Magyarország	Lájer és munkatársai (2008)
Németország	Ewald (1995), Berg & Dengler (2004) http://www.floraweb.de/vegetation/aufnahmen.html http://geobot.botanik.uni-greifswald.de/portal/vegetation
Skandináv és balti államok	Dengler és munkatársai (2006)
Spanyolország	Font & Ninot (1995), Font és munkatársai (1998); http://biodiver.bio.ub.es/biocat/homepage.html http://biodiver.bio.ub.es/vegana
Spanyolország és Portugália	http://www.sivim.info/sivi
Svájc	Wohlgemuth (1992)
Szlovákia	Hrivnak et al. (2003), Hegedúsová (2007), Janišova & Škodova (2007), Šibikova és munkatársai (2009); http://www.ibot.sav.sk/cdf
Ukrajna	Solomakha (1996)
Nagy Britannia	http://www.jncc.gov.uk/page-4259 , http://www.countrysidesurvey.org.uk

1.1. A dolgozat szerkezete

A dolgozat először bemutatja a hazai nemzeti cönológiai adatbázis fejlesztésére tett eddigi erőfeszítéseket és az ezen a téren elért eddigi eredményeket (2. fejezet). Az adatbázis nem csak a dolgozatban szereplő módszerfejlesztések és esettanulmányok alapját jelenti, de kiépítésében személyesen is jelentős szerepem volt. A harmadik fejezet azt vizsgálja meg, hogy a döntő részben szubjektív mintavétellel gyűjtött adatok mennyire használhatók ökológiai hipotézisek statisztikai tesztelésére.

A számítógépek kapacitásának növekedése megteremtette a hardver feltételeket a nagy mennyiségű cönológiai adat feldolgozásához és a cönológiai adatbázisokban rendelkezésre is állnak az elemzéshez az adatok. A feldolgozott felvételek számának emelkedése azonban a feldolgozás során használható módszerek továbbfejlesztését is szükségessé tette. A dolgozat 4-8. fejezete a feldolgozó módszerek fejlesztésében elvégzett munkámat mutatja be.

A cönológiai adatok felhasználásának leggyakoribb formája a numerikus szüntaxonómiai elemzés, amelynek eszköze a numerikus klasszifikáció. A vegetációs adatok mindig zajosak, így a numerikus klasszifikáció hatékonyságát erősen növelheti az előzetes zajszűrés, amelyre korábban másokhoz hasonlóan a metrikus ordinációt javasoltuk (Botta-Dukát et al. 2005), amelynek hatékonyságát elsőként teszteltem (4. fejezet). Több száz – esetleg több ezer – felvétel klasszifikációjakor már nincs lehetőség az eredmények részletes áttekintésére, a csoportok értelmezése a karakterfajokon alapul, amelynek objektív eszköze a fidelitás kiszámítása. Az 5. fejezet áttekinti és értékeli az erre a célra javasolt módszereket, amelyek egy részét szerzőtársaimmal dolgoztuk ki (Botta-Dukát & Borhidi 1999, Chytrý et al. 2002). A numerikus klasszifikáció számos módszere (beleértve a távolságfüggvényeket és az összevonási algoritmust) közötti választás gyakran a kutató döntésén múlik. Bár szemben a szubjektív csoportosítással ezek a döntések jól dokumentáltak, így az eredmények reprodukálhatók, szükség lenne a módszerválasztást segítő objektív kritériumokra. A 6. fejezet ezeket a kritériumokat tekinti át, illetve alkalmazhatóságukat vizsgálja a vegetációs adatok elemzésében. Ugyanezek a módszerek egy másik központi kérdés, az optimális csoportszám megállapítására is használhatók.

A következő két fejezet a cönológiai felvételek fajelőfordulási-adatainak két ökológiai alkalmazásával – fajkészlet becslése (7. fejezet) és nicheszélesség vizsgálata (8. fejezet) – kapcsolatos módszertani kérdéseket vizsgál. Végül az utolsó két fejezet két esettanulmányt mutat be: a közép-európai mocsárrétek változatosságának vizsgálatát (9. fejezet) és a hazai fajok cönológiai adatbázisból becsült nicheszélessége és areájuk mérete közötti kapcsolat elemzését (10. fejezet).

A dolgozatban bemutatott eredmények közül több társszerzőimmal közös munka eredménye, bár a dolgozatba csak azokat az eredményeket vettem fel amelyek elérésében döntő szerepem volt. Az ilyen eredmények ismertetésekor többes szám első személyű igét használok, míg az egyedül elért eredményekről egyes szám első személyben írok.

2. CoenoDAT Referencia Adatbázis

2.1. Bevezetés

Habár a cönológiának nagy tradíciója van Magyarországon (Soó 1964, Borhidi & Sánta 1999, Borhidi 2003), az 1980-as, 90-es években lemaradásba kerültünk ezen a területen (Mucina et al. 1993). Miközben Európa szerte elindultak a cönológiai adatbázis-építő programok, Magyarországon 2002-ig nem volt működő cönológiai adatbázis.

2002-ben a „Magyarország természetes növényzeti örökségének felmérése és összehasonlító értékelése” program (Bartha et al. 2002) keretében indulhatott meg a CoenoDAT Referencia Adatbázis kiépítése. A projekt tervezésekor – a már működő külföldi adatbázisok példáiból – láttuk, hogy ha csak a már meglévő – publikált vagy kéziratoss, de hozzáférhető – felvételeket gyűjtjük össze, az adatbázis nem fogja jól reprezentálni a hazai vegetációt: egyes területek és szüntaxonok alul-, mások felülreprezentáltak lesznek (v.ö. Knollová et al. 2005). Ezért jelentős számú új felvétel összegyűjtése mellett döntöttünk, és csak a legfontosabb archív (=korábban készült) felvételek digitalizálását terveztük. Az adatgyűjtéshez egy részletes módszertani útmutatót dolgoztunk ki. Az újonnan készült felvételeknél számos kiegészítő információ begyűjtését is előírtuk, míg az archív felvételeknél csak a legfontosabbak megléte volt elvárás. Az egyes szüntaxonok alul-, illetve túlreprezentálásnak elkerülésére megterveztük, hogy társulásonként hány felvételt kellene tartalmaznia az adatbázisnak, tekintettel a társulás elterjedtségére, fajgazdagságára és belső változatosságára. Az adatgyűjtés megszervezésére minden társuláshoz az illető társulás jól ismerő szakemberek közül egy koordinátort kértünk fel.

Az alábbiakban először bemutatom az adatbázis építésére kidolgozott módszert, majd annak megvalósulását és az adatbázis jelenlegi állapotát.

2.2. Az adatbázis létrehozására kidolgozott módszer

2.2.1. Az új felvételek készítése

A felvételek helyének kiválasztása során elvárás volt, hogy az alábbi nagy tájak (Marosi & Somogyi 1990) közül mindegyikben készüljön felvétel, ahol a társulás előfordul:

1. Alföld
2. Kisalföld
3. Nyugat-magyarországi-peremvidék
4. Dunántúli-dombság
5. Dunántúli-középhegység
6. Észak-magyarországi- középhegység

Az egyes nagytájokban készített felvételek számát a társulás nagytájon belüli állományainak számára és változatosságára való tekintettel kell megállapítani. A felvételeknek, az előre megszabott felvételszám jelentette korlátokon belül, tükrözni kell a társulás belső változatosságát. A belső változatosságba beleértendő bármilyen különbség (a degradáció okozta különbségek is) a vegetáció struktúrájában, fiziognómiájában, fajösszetételében vagy dominancia viszonyaiban, ha az állomány még egyértelműen az adott társulásba sorolható.

Az állományon belül a felvételi területet úgy kell kijelölni, hogy annak növényzete homogén legyen (Lájer 2002). A mintaterület lehetőleg négyzet alakú legyen, de ettől el lehet térni, ha a négyzet alakú homogén folt nem jelölhető ki. A felvétel helyét az adatlaphoz csatolt fénymásolt 1: 25 000 méretarányú térképen meg kell jelölni.

A mintavétel időpontját úgy kell meghatározni, hogy az előforduló fajok egyértelműen felismerhetők legyenek. Amennyiben az összes faj meghatározásához több időpontban is szükséges felvételezni, a mintavételi egységet egyértelműen meg kell a terepen jelölni. Az alábbi időpontok tájékoztató jellegűek a felvételezés optimális időpontját illetően:

- nyílt szárazgyepek: március (-április), és május-június (július), esetenként szeptember.
- lomboserdők: április-május, és június-július
- láp-, mocsár- és kaszálórétek: május, és esetenként június-július
- mocsarak, lápok és egyéb vizes élőhelyek: június-július, és esetenként augusztus-szeptember.
- iszapnövényzet: augusztus-szeptember
- egyéb vegetációtípusok: június-július.

Ha ugyanabban a mintavételi egységben egy éven belül több borításbecslés is készül, minden fajnál a legnagyobb értéket kell a felvételben feltüntetni.

A mintavételi egység mérete legalább akkora legyen, mint a minimum area. Az alábbi adatok tájékoztató jellegűek, a konkrét értékekről a társulás-koordinátorok döntenek:

- hínár társulások, semlyék növényzet: 1 m²
- forráslápok, iszapnövényzet, sziklahasadék növényzet: 4-9 m²
- oligo- és mezotróf lápok, sós mocsarak, fajszegény nyílt gyepek: 9-16 m²
- szárazgyepek: 4-25 m²
- nádasok, magassásosok, magaskórós vegetáció és rétek: 25-49 (-100) m²
- cserjések és vágásnövényzet: 25-100 m²
- erdők: 400 m².

Ajánlás: Amennyire lehetséges egy asszociációcsoporton belül azonos mintavételi egység méretet kell használni, amiről a társulás-koordinátoroknak kell megállapodni.

A felvételezés során a fajok borításán túl a felvételezés során gyűjtendő adatok listáját Mucina és munkatársai (2000) javaslatait figyelembe véve állítottuk össze (lásd 1. melléklet). A listában szereplő adatok egy részének felmérése kötelező, a többi csak ajánlott.

A felvételben a mintavételi egységben előforduló összes hajtásos növényfajnak szerepelnie kell. Ajánlott a mohák és zuzmók faji szintű meghatározása is, különösen, ha jelentős borítást érnek el. Ha megfelelő szakértő hiányában (különösen a zuzmóknál) a fajlista nem teljes, akkor a moha/zuzmó adatokat a megjegyzés rovatban kell megadni. Kötelező a mohák faji szintű határozása és borításuk becslése a következő vegetációtípusokban: *Montio-Cardaminetea* Br.-Bl. & Tx. 1943, *Caricion lasiocarpae* Van den Berghen ap. Lebrun & al. 1949, *Caricion fuscae* Koch 1926 em. Klika 1934, *Eriophoro vaginati-Sphagnetum recurvi* Hueck 1925, *Genisto pilosae-Festucetum ovinae* Simon 1970, *Festuco ovinae-Nardetum* Dostál 1933, *Luzulo albidae-Callunetum* (l. Horv. 1931) Soó 1971, *Asplenietea trichomanis* (Br.-Bl. in Meyer & Br.-Bl. 1934) Oberd. 1977, *Arabidopsidion*

thalianae Passarge 1964, *Corynephorion canescentis* Klika 1931, *Achilleo ochroleucae-Corynephorium* (Hargitai 1940) Borhidi 1996, *Galio veri-Holoschoenetum vulgaris* (Hargitai 1940) Borhidi 1996, *Bassio laniflorae-Brometum tectorum* (Soó 1938) Borhidi 1996, *Minuartio-Festucetum pseudodalmaticae* (Mikyska 1933) Klika 1938, *Festuco pallentis-Brometum pannonicum* Zólyomi 1958, *Alnetalia glutinosae* Tx. 1937, *Alnenion glutinosae-incanae* Oberd. 1953, *Tilio platyphyllae-Acerion pseudoplatani* Klika 1955, *Seslerio hungaricae-Fagetum* Zólyomi 1967, *Tilio-Sorbetum* Zólyomi & Jakucs (1957) 1967, *Luzulo-Fagion* Lohm. & Tx. in Tx. 1954, *Genisto germanicae-Quercion* Neuhäusl & Neuhäuslová-Novotná 1967, *Castaneo-Quercion* Soó 1962 em. 1971, *Erico-Pinion* Br.-Bl. in Br.-Bl. et al. 1939, *Bazzanio-Abietetum* Ellenberg & Klötzli 1972, *Pino-Quercion* Medwecka-Kornas & al. 1959.

A borítás becslés során a 7-fokozatú, klasszikus Braun-Blanquet skálával azonos, vagy annál finomabb felbontású skálák használhatók.

Ha a fajnak csak fejletlen egyede (csíranövénye) van jelen a felvételen, és adott körülmények között annak kifejlődése nem is várható, ezt a fajnév elé tett „Cs” betűvel kell jelölni. Fásszárú fajok borítását szintenként kell megadni, a cserjeméretet el nem érő egyedeket a gyepszintnél feltüntetve.

Állományonként egy dokumentumfotót is mellékelni kell a felvételekhez.

2.2.2. Az archív felvételek kiválasztása

Az archív felvételek közül azok kerülhetnek be az adatbázisba, amelyek teljes fajlistát és minden fajhoz borításbecslést tartalmaznak, és valamint a fejléc adatok egy szűkített köre (lásd 1. melléklet) a tabellából, vagy a kiegészítő információkból meghatározható.

Ha a társulást Magyarországon írták le, akkor a típusfelvétel szerepeljen a kiválasztott felvételek között. Ha külföldön írták le a társulást, akkor az első olyan forrásból, amely említi a társulást és a fenti kívánalmaknak megfelelő felvétel(ek)e)t tartalmaz, legyen legalább egy felvétel kiválasztva.

A társulás koordinátorának a kiválasztott archív felvételhez kapcsolódó összes megállapítható fejléc adatot össze kell gyűjtenie és a felvétel adatlapján meg kell adnia.

Az archív felvételek kiválasztásakor tekintettel kell arra is lenni, hogy lehetőség szerint minél kevésbé térjen el a mintaterületek mérete az új felvételeknél alkalmazottól.

A megjegyzés rovatban meg kell adni a felvételezett állomány jelenlegi állapotára vonatkozó információt, az alábbi kategóriákat használva:

- Szinte változatlan
- Az állomány fennmaradt, de jelentősen megváltozott (a változást le kell írni). A közelben azonban található az eredetihez hasonló állomány.
- Az állomány fennmaradt, de jelentősen megváltozott (a változást le kell írni), és a közelben sem található az eredetihez hasonló állomány.
- Az állomány megsemmisült, de a közelben található az eredetihez hasonló állomány.
- Az állomány megsemmisült, és a közelben sem található az eredetihez hasonló

állomány.

Szintén a megjegyzés rovatban röviden indokolni kell a felvétel kiválasztását (pl. „Ez a felvétel a társulás holotípusa.” vagy „Tipikus dunántúli állomány.”)

A felvétel helyét – azzal a pontossággal, amellyel a rendelkezésre álló adatokból az megállapítható – meg kell jelölni 1: 25 000-es méretarányú térképen. Ha a felvétel helye bizonytalan, azt a területet kell bejelölni, ami valószínűleg tartalmazza a felvétel helyét, és azon belül meg kell jelölni a felvétel legvalószínűbb helyét.

2.2.3. Fajlisták és társuláslista

Az adatbázis a FLÓRA adatbázis 1.2 (Horváth et al. 1995) fajlistáján alapul, mert ez *sensu lato* fajokat és aggregátumokat is tartalmaz azoknál a nemzetségeknél, amelyekben a fajok elhatárolása bizonytalan (pl. *Rubus* spp.), vagy a fajok terepen nehezen határozhatók (pl. *Quercus* spp.). A fajlistában minden taxont besorolt Dobolyi (2002) az alábbi három kategória egyikébe:

- **aa:** a *sensu stricto* taxon azonosítása szükséges, mert a durvább szintű azonosítás esetén cönológiai információk veszhetnek el,
- **a:** elegendő ezen a szinten azonosítani a taxont,
- **b:** nem szükséges (de megengedett) az ilyen finom szintű azonosítás

A mohák taxonlistája Erzerberg és Papp (2004) munkáját követi, a szüntaxonok elnevezésénél Borhidi (2003) rendszerét használtuk.

2.3. Az adatbázis jelenlegi helyzete

Az adatbázis-építést nem sikerült teljesen a terveknek megfelelően megvalósítani. Több asszociációhoz nem sikerült koordinátort találnunk, mert nem volt olyan szakember, aki elég jól ismerte volna ezeket. A párhuzamosan folyó flóra- és vegetációfelmérés nagyon lekötötte a terepbotanikus szakma erejét, így nem tudtunk annyi új felvételt elkészíttetni, mint amennyit terveztünk. Ezért a tervezettől eltérően elkezdtük a publikált felvételek szisztematikus digitalizálását. Így a tervezett 7200 helyett az adatbázis több mint 9200 felvételt tartalmaz, de a vegetációtípusok reprezentáltsága nem olyan egyenletes, mint terveztük (2. táblázat). Esetenként (pl. *Utricularietea intermedio-minoris*, *Pulsatillo-Pinetea*) a terepmunka már befejeződött, de a felvételek beépítése az adatbázisba még nem történt meg. Az adatbázis fejlesztése „Magyarország természetes növényzeti örökségének felmérése és összehasonlító értékelése” program befejeződése óta is tovább folyik, igaz kisebb intenzitással. A tervezett felvételszámtól való eltérésnek esetenként objektív okai is vannak: egyes társulások kipusztultak Magyarország területéről (pl. *Sphagno tenelli-Rhynchosporietum albae*), de szerepelnek az adatbázisban olyan társulások felvételei is, amelyek hazai előfordulása a tervezéskor még nem volt ismert (pl. *Glycerietum nemoralis-plicatae*), vagy a tudományra nézve is újak (pl. *Astero pannonic-Schoenetum nigricantis* Lájér 2006).

Az adatbázis tervezettől való eltérések ellenére is már jelenlegi formájában is jól használható kutatási eszköz, de további folyamatos bővítését is tervezzük.

2. táblázat: Összesítő adatok a tervezett és az adatbázisban szereplő felvételszámokról asszociációosztályok szintjén. A hiányzó és a tervezetthez képest többlet felvételeket asszociációként vettük számba, így egy osztályon belül egyszerre lehet hiány (mert bizonyos asszociációkból a tervezettnél kevesebb felvétel van az adatbázisban) és többlet (mert más asszociációkból a tervezettnél több felvétel gyűlt össze).

Asszociációosztály	Tervezett felvételszám	Felvételek száma jelenleg az adatbázisban	Hiány (társulás szinten)	Többlet (társulás szinten)
<i>Lemnetea</i>	125	310	14	199
<i>Utricularietea intermedio-minoris</i>	10	0	10	0
<i>Charetea fragilis</i>	10	16	5	11
<i>Potametea</i>	165	321	73	231
<i>Litorelletea uniflorae</i>	5	18	0	13
<i>Isoeto-Nanojuncetea</i>	95	339	50	294
<i>Phragmitetea australis</i>	735	825	389	479
<i>Montio-Cardaminetea</i>	40	65	10	35
<i>Scheuchzerio-Caricetea fuscae</i>	340	175	224	59
<i>Oxycocco-Sphagnetea</i>	10	8	2	0
<i>Puccinellio-Salicornietea</i>	607	725	202	320
<i>Molinio-Arrhenatheretea</i>	927	736	400	209
<i>Calluno-Ulicetea</i>	115	70	63	18
<i>Asplenietea trichomanis</i>	35	26	23	14
<i>Sedo-Scleranthetea</i>	50	42	20	12
<i>Koelerio-Corynephoretea</i>	30	32	10	12
<i>Festucetea vaginatae</i>	220	393	79	252
<i>Festuco-Brometea</i>	935	1199	387	651
<i>Trifolio-Geranietea sanguinei</i>	80	5	75	0
<i>Stellarietea mediae</i>	0	353	0	353
<i>Artemisietea vulgaris</i>	0	76	0	76
<i>Oryzetea sativae</i>	0	18	0	18
<i>Bidentetea tripartiti</i>	0	107	0	107
<i>Galio-Urticetea</i>	0	26	0	26
<i>Polygono arenastri-Poetea annuae</i>	0	107	0	107
<i>Epilobietea angustifolii</i>	0	39	0	39
<i>Rhamno-Prunetea</i>	195	222	94	121
<i>Salicetea purpureae</i>	95	170	0	75
<i>Alnetea glutinosae</i>	150	211	71	132
<i>Quercu-Fagetea</i>	1230	1518	512	800
<i>Quercetea pubescentis-petraeae</i>	930	1015	206	291
<i>Erico-Pinetea</i>	20	9	11	0
<i>Pulsatillo-Pinetea</i>	10	0	10	0
<i>Vaccinio-Piceetea</i>	35	31	19	15
<i>Robinietea</i>	0	0	0	0
összesen	7199	9207	2959	4969

3. A cönológiai adatbázisok felhasználásának lehetőségei és korlátai

A vegetációs adatbázisok nem kizárólag, de jelenleg döntő többségében preferenciális – Nyugat- és Közép-Európában a Braun-Blanquet cönológiai iskola szabályai szerint készült – felvételekből állnak. Felmerülhet a kétség, hogy alkalmasak-e ezek az adatbázisok a valós vegetáció reprezentálására, illetve, hogy lehetnek-e ezek a nem random mintavétellel gyűjtött adatok statisztikai elemzés tárgyai.

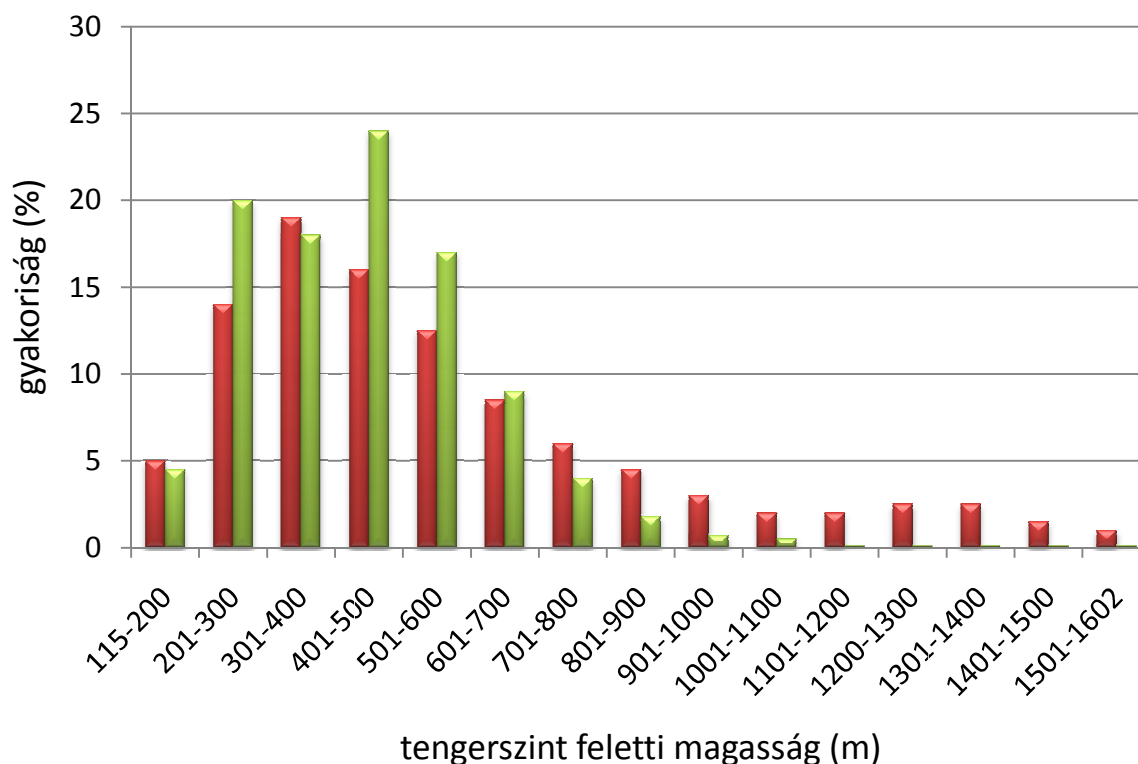
Az első kérdésre a válasz, hogy részben igen, és még ha nem is tökéletes, de a rendelkezésünkre álló legjobb eszközök a durva léptékű vegetációs vizsgálatokban.

Az, hogy mit tekintünk megfelelő mintavételnek, attól is függ, hogy hogyan definiáljuk a vizsgálatunk tárgyát, a vegetációt (Chiarucci 2007). A Braun-Blanquet iskola felfogása szerint a vegetáció jól elkülönülő egységekből áll, alapegysége az asszociáció a fajjal analóg módon leírható és tipizálható természetes egység (Pignatti et al. 1997, a faj és az asszociáció közötti különbségekről lásd Mirkin 1989). Ezért, hasonlóan az idiotaxonómiához, ahol a faj leírása a típusegyeden alapul, a növényyszociológusok is a tipikus állományok felvételezésére törekedtek (Westhoff & van der Maarel 1973). Ezért hiányoznak, vagy legalábbis alulreprezentáltak az „atipikus”, „jellegtelen”, „átmeneti” vegetációtípusok az adatbázisban. Roleček és munkatársai (2007) szerint random mintavételből származó adatokkal lehetne és kellene pótolni ezt a hiányt.

A preferenciális mintavétel következtében a vegetációtípusok gyakorisága a mintában nem azonos az előfordulási gyakoriságukkal a természetben (4. ábra). Ugyanakkor, mivel a preferenciális mintavételezés során a felmérő igyekszik a területen előforduló minden típust megmintázni ("catch as much variation as possible"; Diekmann et al. 2007), a preferenciális minta több ritka típust tartalmaz, és így jobban reprezentálja a vegetáció változatosságát, mint az azonos méretű random minta (pl. Økland 2007, Roleček et al. 2007). Roleček és munkatársai (2007) megfogalmazása szerint ökológiailag reprezentatívabb a preferenciális mintavétel, mert jobban reprezentálja a vegetáció változatosságát. Elvileg a rétegzett random mintavétel megoldást jelentene arra a problémára, hogy a random mintavétel erősen alulbecsli a változatosságot, de nehéz a rétegzéshez a megfelelő változókat megtalálni, amelyek a vegetáció szempontjából is relevánsak, és amelyekről megfelelő felbontású adatok is vannak (Lepš & Šmilauer 2007). A rétegzett random mintavétel kivitelezése nagy területen bonyolultabb, mint a preferenciális mintavételé, ezért ritka, de nem lehetetlen (Grabherr et al. 2003).

A második kérdést illetően, hogy lehet-e statisztikai tesztekben a preferenciális mintavétellel gyűjtött adatokat használni, megoszlanak a vélemények. Lájér (2007) szerint a válasz egyértelműen nem, mert nem teljesül az adatok függetlensége. Mások viszont (pl. Roleček et al. 2007, Lepš & Šmilauer 2007) sokkal megengedőbbek ebben a kérdésben, bár elismerik, hogy szigorú matematikai értelemben a függetlenség nem teljesül. vitatják, hogy ebből az következik, hogy egyáltalán nem lehet statisztikai tesztek alkalmazni. Magam is az utóbbi állásponttal értek egyet. Úgy érzem ezt a fontos és vitatott kérdést érdemes

részletesebben is körüljárni, ezért a következőkben először elméleti alapon próbálom véleményemet alátámasztani, majd egy szerzőtársaimmal készített esettanulmányt mutatok be, amely szemlélteti a random és preferenciális mintavétel közti különbségeket az eredményekben.



4. ábra: A cönológiai felvételek tengerszint feletti magasság szerinti megoszlása a Cseh Nemzeti Cönológiai Adatbázisban (piros oszlopok), és Csehország területének megoszlása tengerszint feletti magasság szerint (zöld oszlopok). Az ország legnagyobb részét borító sík- és dombvidéki területek alul-, míg a hegyvidékek felülreprezentáltak az adatbázisban. (Forrás: Chytrý & Rafajová 2003)

3.1. Preferenciális mintavétel – elméleti megfontolások

3.1.1. Alapsokaság és minta

Statistikai elemzést akkor használunk, ha úgy akarunk egy halmazról (halmazokról) megállapításokat tenni, hogy nem vizsgáljuk meg minden elemét (minden elemüket). A halmazt, amiről a statisztikai elemzés alapján kijelentéseket teszünk, alapsokaságnak vagy (statisztikai) populációnak nevezzük¹. Az alapsokaság minden elemének megmérése sokszor lehetetlen, mivel a halmaz mérete végtelen (pl. egy kísérletet elvileg végtelen sokszor megismételhetünk, ezért ebben az esetben az alapsokaság egy végtelen méretű halmaz: a lehetséges kísérletek halmaza). Véges, de nagyméretű alapsokaság (pl. az összes növényegyed egy réten) esetén elvileg lehetséges, de gyakorlatilag alig kivitelezhető – és

¹ Az angol nyelvű szakirodalomban általában a 'population', ritkábban az 'universe' kifejezés használatos (Zar 1999). Mivel a populáció kifejezés a biológiában más értelemben használatos, érdemes a magyar nyelvű szövegekben – így ebben az értekezésben is – az alapsokaság kifejezést használni.

felesleges is – az alapsokaság összes elemének megmérése. Ezért csak az alapsokaság egy részhalmazának – a mintának – az elemeit mérjük meg (Zar 1999), majd a kapott eredményeket kiterjesztjük az egész alapsokaságra.

Bármilyen statisztikai módszereket alkalmazó vizsgálat első lépése az alapsokaság meghatározása. Az alapsokaság tetszőleges halmaz lehet, de még a mintavétel előtt pontosan definiálni kell. A pontos definíció elengedhetetlen, mert a statisztikai elemzés eredményei az alapsokaságra és csak az alapsokaságra lesznek érvényesek.

Ennek ellenére a publikációkban a gyakran nem definiálják elég pontosan a vizsgált alapsokaságot. Például munkatársaimmal a sziklagyeppek diverzitását vizsgáló cikkünkben (Rédei et al. 2003) a "nyílt dolomit sziklagyep" és a "*Seseli leucospermi-Festucetum pallentis*" elnevezéseket, mint ugyanannak az alapsokaságnak a szinonim elnevezését használtuk. A nyílt dolomit sziklagyep az alapkőzet és a növényzet összborítása alapján definiálható. A *Seseli leucospermi-Festucetum pallentis* társulás viszont a fajösszetétele alapján azonosítható, figyelmen kívül hagyva az abiotikus körülményeket, így az alapkőzetet is (a dolomit mellett ritkán mészkövön is előfordul; Kun & Ittész 1995). Ebben a speciális esetben az asszociációnév (*Seseli leucospermi-Festucetum pallentis*) tágabb kategóriát jelöl, mint a hozzá kapcsolódó élőhelynév (nyílt dolomit sziklagyep), de általában ennek a fordítottja igaz. Végezzünk el egy gondolat kísérletet! A cseres-kocsánytalan-tölgyesek aljnövényzetét akarjuk vizsgálni. A cönológiai szakirodalomból (pl. Borhidi 2003) tudjuk, hogy a cönológusok ezt a vegetációtípust *Quercetum petraeae-cerris*-nek hívják. Ha ezután leválogatjuk a cönológiai adatbázisból az ebbe a társulásba tartozó felvételeket, akkor az azok elemzésével kapott eredmények nem lesznek érvényesek a cseres-kocsánytalan-tölgyes élőhelyre, mert a felvételek szinte mindegyike idős állományokban készült, míg az élőhelybe az ettől eltérő aljnövényzetű (Csontos 1996) korábbi regenerációs fázisok is beletartoznak.

3.1.2. Random mintavétel

Az alapsokaság definiálása után a következő lépés a mintavétel, vagyis az alapsokaság elemei közül azoknak a kiválasztása, amelyek részei a mintának. A standard statisztikai módszerek csak akkor adnak érvényes eredményt, ha:

- (1) az alapsokaság minden eleme azonos valószínűséggel kerül kiválasztásra,
- (2) az alapsokaság egyik elemének kiválasztása nem befolyásolja a többi elem kiválasztását,
- (3) a mintabeli értékek független valószínűségi változók.

Ha ezek a feltételek sérülnek, a minta jellemzői (pl. átlag) nem lesznek torzítatlan becslései az alapsokaság jellemzőinek. Míg az alapsokaság jellemzői fix értékek, a mintából becsült értékek valószínűségi változók. A standard statisztikai eljárásokban ezeknek a becsült értékeknek (amelyek közé tartoznak a próbastatisztikák is) az eloszlását a mintaelemek eloszlásából csak akkor lehet kiszámítani, ha a mintában az értékek független valószínűségi változók.

A szakirodalom a random mintavételt (azaz a minta elemeinek random kiválasztását) annak érdekében javasolja, hogy a fenti feltételek közül az első kettő teljesüljön. Ha az

alapsokaságbeli értékek autokorrelálatlanok, a második feltétel teljesülése maga után vonja a harmadik feltétel teljesülését. Ha az alapsokaságbeli értékek autokorreláltak – például a közeli rokon fajok a közös leszármazás miatt hasonlóbbak, mint két véletlenül választott faj (filogenetikai autokorreláció), vagy a felvételek hasonlósága nő, ahogy csökken köztük a térbeli távolság (térbeli autokorreláció), állandó kvadrátos felvételeknél a közelebbi időpontokból származó adatok hasonlóbbak (időbeli autokorreláció) – a random mintavétel miatt teljesül ugyan a második feltétel is, de a mintabeli értékek mégsem lesznek független valószínűségi változók. Ebben az esetben – ami a vegetációs adatok esetében nagyon gyakori (Økland 2007, Wilson 2007) – olyan statisztikai eljárásokat kell alkalmazni (pl. térbeli statisztikák, független filogenetikai kontrasztok, kevert modellek), amelyek az autokorrelációt is figyelembe veszik. Mivel az autokorreláció figyelembevétele nem érinti érvelésünk lényegét, a továbbiakban feltételezzük annak hiányát és az ilyenkor alkalmazható standard statisztikai eljárásokra koncentrálunk.

Lájer (2007) szerint a preferenciális mintavétel esetén nem teljesül a második feltétel, az adatok autokorreláltak lesznek, ezért nem használhatók a statisztikai elemzésben. Lepš & Šmilauer (2007) rámutatott, hogy sokkal jellemzőbb az első feltétel sérülése, mert vannak olyan foltok, amelyeket a felvételező elkerül, vagyis nem minden hely kerül egyenlő eséllyel megmintázásra.

Vannak ugyan olyan eljárások (pl. véletlen koordináták használata) amelyekkel a vegetáció mintavételezése során is megvalósítható a random mintavétel, de a terepmunka során a vegetációkutatók ritkán alkalmazzák ezeket; általában formális szabályok nélkül választják ki a felvételek helyét akkor is, ha nem preferenciális, hanem „random” mintavételt végeznek.

3.1.3. Mit jelent a preferenciális mintavétel a cönológiában?

Elméletben a preferenciális mintavétel azt jelenti, hogy csak a legtipikusabb állományokban készülnek felvételek. Személyes tapasztalataink², a szakirodalom és a hazai cönológusok szóbeli közlései alapján ez az alapsokaságnak a cönológia szabályai szerinti definiálását jelenti – azaz a vizsgált alapsokaság nem tartalmazza a „nem tipikus”, „nem homogén”, „átmeneti”, fajszegény, zavart stb. állományokat (Westhoff & van der Maarel 1973, Kent & Coker 1992, Dierschke 1994). A továbbiakban a mintavétel az így leszűkített alapsokaságból már válogatás nélkül történik.

3.1.4. A statisztikai tesztek formális és szokványos feltételei

A statisztikában a tesztek alkalmazásának szigorú formális kritériumai vannak. Például a varianciaanalízis (ANOVA) alkalmazásának feltétele a hibák függetlensége, normális eloszlása, additivitása és azonos szórása (homoscedaszticitás) (pl. Zar 1999). Ha ezek a feltételek nem teljesülnek, az elsőfajú hiba becslése torzított lesz. Tisztán matematikai

² Itt nem csak a saját tapasztalataimra utalok, hanem a témakörben megjelent cikkünk (Botta-Dukát et al. 2007) szerzőinek tapasztalataira

szempontból nincs különbség a kis és nagymértékű torzítás között. A gyakorlatban viszont általában csak azt mérlegeljük, hogy az elsőfajú hiba kisebb-e mint az előre definiált szignifikancia szint, így a kismértékű torzítás a következtetéseinket általában nem befolyásolja, ezért megengedhető.

Ezért a gyakorlatban a formális feltételek helyett kevésbé szigorúakat használunk. Ezeket a kutatási gyakorlatban általánosan elfogadott, a kutatói közösség hagyományán alapuló, kevésbé szigorú feltételeket nevezzük *szokványos* vagy *konvencionális* feltételeknek. Például az ANOVA kis torzítással becsli az első fajú hiba valószínűségét, ha a hiba eloszlása szimmetrikus, de nem normális (pl. annál lapultabb vagy csúcsosabb). Ezért, általánosan elfogadott az ANOVA használata akkor is, ha a hibák eloszlása kis mértékben eltér a normális eloszlástól.

Egyetértek Lájér Konráddal (2007) abban, hogy a cönológiai mintavétel nem teljesíti a random mintavétel formális kritériumait. Nem értek viszont az ebből levont következtetésével, hogy a statisztikai módszerek nem használhatók a cönológiai adatok elemzésére. Lepš & Šmilauer (2007) megfogalmazása szerint, Lájér (2007) érvelése szerint „ha nem teljesülnek teljes mértékben a próba feltételei, akkor **biztosan hibás lesz** az eredmény”, míg szerintük (és szerintem is) ilyenkor csak „**nem biztos, hogy hibátlan lesz** az eredmény.”

A statisztikai elemzések két típusát különböztethetjük meg: adatfeltárás és hipotézistesztesztelés (pl. Hallgren et al. 1999). Kétféle elemzés célja különböző, de gyakran ugyanazokat a módszereket használják mindkét célra. Mivel az adatfeltárás során az elsőfajú hiba becsült valószínűségét csak leíró statisztikaként használjuk, legalább erre a célra akkor is használhatók a cönológiai felvételekből számolt statisztikai tesztek, ha a függetlenség nem teljesül. Erre a célra sem használhatók azonban akkor, ha a helyszínek nem véletlen kiválasztása szisztematikus torzítást okoz (Lepš & Šmilauer 2007).

Természetesen a kevésbé szigorú feltételek alkalmazása a mintavétel során kizárja a statisztikai módszerek automatikus alkalmazását. A kutatónak minden esetben alaposan mérlegelnie kell, az alkalmazott módszer mely formális kritériumai sérülnek, ennek milyen következményei lehetnek (pl. mennyire robusztus a módszer ebben a tekintetben), és ezt figyelembe véve kell értelmeznie az eredményeket, tartózkodva azok túlintertpretálásától.

Mik azok a szokványos feltételek, amiket figyelembe kell venni, amikor preferenciális mintavétellel gyűjtött felvételeket használunk vegetációtípusok összehasonlítására (pl. Rédei et al. 2003)?

- a) Csak olyan változókat használjunk akár függő- akár független változóként, amelyekre nézve a mintavétel neutrális, vagyis amelyek értéke nem befolyásolja a felvételek kiválasztását (Lepš & Šmilauer 2007).
- b) A felvételeknek elsősorban az abiotikus környezetet és/vagy a növényföldrajzi hatásokat kell tükrözni és nem a lokális zavarásokat (kivéve, ha a zavarás jól definiálható és általánosan jellemző a vegetációtípusra; pl. kaszálás, elárasztás)
- c) A mintavételi pontok kiválasztása minden vegetációtípusban azonos módon, egy megismételhető algoritmus szerint történjen. Ez az egyik kulcspontja a vitának

Lájerrel (2007), mivel a cönológiában használt algoritmus nem formalizált. A felvételező először egy általános képet alkot magában a vizsgált terület vegetációjáról, majd összehasonlítja ezt a korábbi ismereteivel a felismert társulásokról. Ezután kiválasztja a „legtipikusabb” állományokat, amelyek reprezentálják a felismert társulást, figyelembe véve a helyi eltéréseket is a társulás tipikus összetételétől. Habár ez az algoritmus nem teljesen független alkalmazójának személyétől, megtanulható és hatékonyan alkalmazható. Így a természettudományok általános „objektivitási” kritériuma nem teljesül ugyan teljes mértékben, de a megismételhetőség – legalábbis a gyakorlott cönológusok esetén (Kent & Coker 1992, Mucina 1997b) – teljesül.

- d) Az összes felvétel készüljön azonos módszerrel, beleértve a mintavételi egység méretét, a borításbecslés pontosságát, az alkalmazott taxonlistát, stb.

3.1.5. Záró megjegyzések

Véleményem szerint, habár a cönológiában alkalmazott mintavétel nem teljesíti a statisztikai elemzés formális követelményeit, a cönológiai felvételek használhatók statisztikai elemzésekben, ha a formális feltételek teljesülnek. Sajnos a biológusoknak szóló statisztikai kézikönyvek (pl. Sokal & Rohlf 1981, Zar 1999) csak a formális kritériumokat tárgyalják, a szokványos kritériumokról sem ezekben, sem a cönológiai szakirodalomban nem olvashatunk.

A legfontosabb ilyen kritériumok az interpretálás szabályai. Mi az alapsokaság, amit a cönológiai felvételekből álló minta reprezentál, és amire az eredmények vonatkoznak? Úgy gondolom ez nem egy egyszerű kérdés, amire létezik általános érvényű válasz. Ha a kutató adatbázisból vagy irodalomból származó felvételeket használ, mindig alaposan át kell gondolnia mik lehettek a felvételek készítőjének a preferenciái. Ezt általában könnyen meg tudjuk tenni, de a következtetéseinket nem tudjuk ellenőrizni. A következő részben egy olyan esettanulmányt mutatok be, ahol lehetőség volt erre az ellenőrzésre.

3.2. A preferenciális és a random mintavétel összehasonlítása a gyakorlatban

Általánosan elfogadott, hogy a Zürich-Montpellier cönológiai iskola tagjai a felvételek készítésekor előnyben részesítik a fajgazdag állományokat (v.ö. Chytrý 2001), különösen azokat, amelyek gazdagok karakter (specialista, ritka, szűk elterjedésű vagy speciális növényföldrajzi hatást jelző) fajokban, és elkerülik az átmeneti és degradált állományokat (Westhoff & van der Maarel 1973). A degradáció jele lehet a fajszám csökkenése és a ruderalis stratégiájú fajok (gyomok) jelenléte. Ebben az esettanulmányban olyan jellemzők szempontjából hasonlítjuk össze a random és a preferenciális mintavételt, amelyeket a cönológusok figyelembe vesznek a mintavétel során: fajgazdagság, az életformák, a flóraelemek és a fajok cönológiai preferenciáinak megoszlása. Azt várjuk, hogy a preferenciális mintavétel túlbecsli a fajgazdagságot, a specialista, a szűk elterjedésű vagy a speciális növényföldrajzi hatást jelző fajok arányát, és alulbecsli a gyomok, a generalista és a

nagy areájú fajok arányát.

A vizsgálatok célja gyakran nem egy terület vegetációjának jellemzése, hanem különböző területek összehasonlítása. Statisztikai szempontból az előbbi becslés, míg az utóbbi hipotézis vizsgálat. A hipotézisek tesztelése során a preferenciális és a random mintavétel között a becsléskor tapasztalt különbségek csökkenhetnek, de fel is nagyíthatnak. Ezért érdemes a becslés mellett a hipotézis tesztelés eredménye tekintetében is összehasonlítani a kétféle mintavételt. Mivel a hagyományos cönológia inkább a tipikus állományok felvételezésére törekszik, mint a variáció dokumentálására, azt várjuk, hogy ha ugyanannak a vegetációtípusnak két különböző helyen készült felvételeit hasonlítjuk össze, akkor a preferenciális mintavétel túlhangsúlyozza a vegetációtípus általános jellegzetességeit, és következésképpen alulbecsli a helyek közötti különbséget.

3.2.1. Anyag és módszer

3.2.1.1. Vizsgált terület és mintavétel

A vizsgálathoz a Duna-Tisza köze homokbuckásai közül választottunk két mintaterület: Csévharasztot és Fülöpházát. Mindkét terület természetes vegetációja homoki erdőssztyepp: nyílt homoki gyepek, üde rétek és nyáras-borókás foltok mozaikja. A régió belülről északnyugatról (Csévharaszt) dél-kelet (Fülöpháza) felé haladva nő az ariditás, és ennek következtében megváltozik az erdőssztyepp szerkezete is: lecsökken a fásszárúakkal borított terület aránya. A nyílt homoki gyepek szerkezete is megváltozik: Csévharaszton gazdagabb évelő, xero-mezikus erdőssztyepp fajokban, míg a kontinentális vagy szubmediterrán elterjedésű egyéves fajok aránya Fülöpházán magasabb (Kovács-Láng et al. 2000).

Mindkét területen 50 ha-os (400x1250m) mintaterület lett kijelölve az összehasonlító vizsgálatokhoz. A területen belül a nyílt homoki gyepekben (*Festucetalia vaginatae* Soó 1957) párhuzamosan készültek preferenciális felvételek a „klasszikus” Braun-Blanquet módszertan szerint (30 felvétel), és nem-preferenciális felvételek 40-50 random kiválasztott ponton, amelyek közül random választottunk ki 30-t ehhez a vizsgálathoz. A kvadrátok mérete mindkét mintavételi elrendezésben 4 x 4 m volt. A preferenciális felvételeket mindkét területen Kovácsné Láng Edit készítette, így a területek közötti különbségeket nem okozhatta a felvételező eltérő preferenciája.

3.2.1.2. Az összehasonlítás során használt jellemzők

A teljes fajösszetétel összehasonlítása mellett, vizsgáltuk azoknak a fajcsoportoknak a gyakoriságát, amelyeket várhatóan preferálnak, vagy éppen elkerülnek a cönológusok.

A fajokat földrajzi elterjedésük alapján (Horváth et al. 1995) négy csoportba soroltuk:

- **kis areájú fajok** = endemikus és szubendemikus fajok, amelyek areája a Kárpát-medencére és az azzal szomszédos területekre korlátozódik
- **szubmediterrán fajok**
- **kontinentális fajok**

- **nagy areájú fajok** = kozmopolita, cirkumpoláris, eurázsiai stb. flóraelemek, amelyeknek a tág area miatt növényföldrajzi szempontból nem karakterisztikusak

Élőhely preferenciájuk alapján – Borhidi (1995) részletesebb besorolását leegyszerűsítve – négy csoportba soroltuk a fajokat:

- **homoki fajok** = a meszes, nyílt homoki gyepek specialistái
- **erdőssztyepp fajok** = a homoki erdőssztyepp xero-mezikus fajai
- **generalisták** = a szárazgyepek generalista fajai
- **gyomok** = emberi zavarást jelző fajok

Életforma tekintetében csak két kategóriát különböztettünk meg: az **egyéveseket** (ide sorolva kétéves fajokat is) és az **évelőket**.

3.2.1.3. Statisztikai elemzés

A terepi adatgyűjtésben több felvételező vett részt, ami eltérő borítás-bebecslésekhez vezethet (Sykes et al. 1983, Kennedy & Addison 1987, Hahn & Scheuring 2003). Ennek a hatásnak a kiküszöbölésére csak a fajok jelenlét/hiány adatait használtuk fel.

A területek és a mintavételi módok közötti különbséget a fajösszetételben távolság-alapú redundancia-elemzéssel (Legendre & Anderson 1999) vizsgáltuk, a Jaccard hasonlóság komplementerének négyzetgyökét használva távolságfüggvényként (Podani 1997, Legendre & Legendre 1998). Annak eldöntésére, hogy a két terület közötti különbséget befolyásolja-e a mintavétel módja, távolság-alapú parciális redundancia elemzéssel teszteltük a két faktor (a mintavétel módja és a terület) közötti interakciót. Az első fajú hibát permutációs teszttel állapítottuk meg. Az elemzéshez a CANOCO 4.52 programot használtuk.

A fajgazdagság és az előző fejezetben felsorolt fajcsoportok csoportrészesedésének elemzéséhez nem-paraméteres egyváltozós statisztikai módszereket használtunk. Egy mintavételi helyen belül a kétféle mintavétellel kapott eloszlások összehasonlítására Szmirnov tesztet használtunk (Sprent & Smeeton 2001). A Szmirnov teszt előnye, hogy a teljes eloszlásokra, nem csak azok valamelyik jellemzőjére (pl. középérték, szóródás) vonatkozó null-hipotézist tesztel. Hátránya, hogy gyengébb, mint az eloszlás valamelyik kiválasztott jellemzőjére vonatkozó tesztek (pl. Wilcoxon-Mann-Whitney teszt) (Sprent & Smeeton 2001). Ezért a mediánokat is összehasonlítottuk Wilcoxon-Mann-Whitney teszttel. Ezt követően azt vizsgáltuk, hogy a két mintavételi helyet összehasonlításakor az eredmények mennyire függnak a mintavétel módjától. A mintavételi helyek összehasonlítására Wilcoxon-Mann-Whitney tesztet használtunk, mert ezt a tesztet (vagy kiterjesztését több csoportra, a Kruskal-Wallis tesztet) gyakran alkalmazzák a vegetációkutatók is (pl. Csecserits & Rédei 2001, Kröel-Dulay et al. 2004), ha az adatok nem normális eloszlásúak. A nem-paraméteres tesztek kiszámítására az SPSS 10 statisztikai programot használtuk.

3.2.2. Eredmények

A távolság-alapú redundancia analízisben szignifikáns interakciót találtunk a mintaterületek és a mintavételi módszer között ($F=1.50$, $p=0.003$), mivel a random és a

preferenciális mintavétellel gyűjtött adatsor szignifikánsan eltért Fülöpházán ($F=1.61$, $p=0.0059$), de nem volt köztük szignifikáns különbség Csévharaszton ($F=1.29$, $p=0.0524$). Részletesebben megvizsgálva a fülöpházi adatokat, két faj frekvenciáiban találtunk jelentősebb különbséget: a *Cerastium semidecandrum* gyakoribb, a *Salsola kali* ritkább a preferenciális mintában. Ha ezt a két fajt kihagyjuk az elemzésből, a két mintavételi módszer közötti különbség Fülöpházán sem szignifikáns ($F=1.07$, $p=0.3034$).

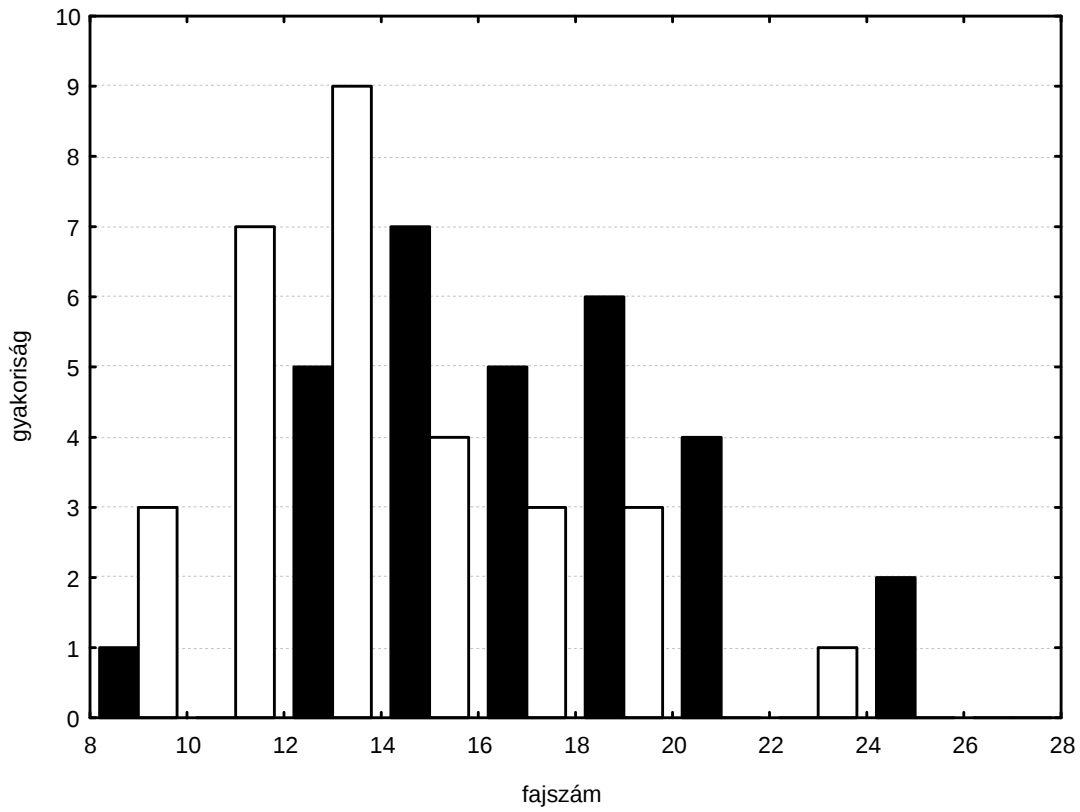
3. táblázat: A fajgazdagság és a fajcsoportok csoportrészesedéseinek átlagos értékei a két területen és az első fajú hiba a Wilcoxon-Mann-Whitney tesztben a két terület összehasonlításakor. (1) átlagok Csévharaszton, (2) átlagok Fülöpházán, (3) az első fajú hiba valószínűsége (az 5%-nál kisebb értékeket vastag betűvel kiemelttem).

ípreferenciális mintavétel		preferenciális mintavétel			random mintavétel		
		(1)	(2)	(3)	(1)	(2)	(3)
fajgazdagság		14.3	15.9	0.071	17.5	16.1	0.177
életforma	egyévesek	29.9%	39.4%	0.037	34.2%	43.1%	0.016
	évelők	70.1%	60.6%	0.037	65.8%	56.9%	0.016
flóraelem	kis areájú	27.1%	25.5%	0.424	23.9%	23.6%	0.988
	szubmediterrán	20.6%	27.7%	0.001	19.0%	24.3%	0.030
	kontinentális	40.0%	39.5%	0.947	36.1%	42.0%	0.010
	nagy areájú	28.7%	23.6%	0.101	35.6%	27.3%	0.003
élőhely	homoki fajok	59.1%	61.2%	0.584	48.3%	66.2%	<0.001
preferencia	erdőssztyepp fajok	8.9%	1.3%	<0.001	15.6%	2.1%	<0.001
	generalisták	29.2%	34.6%	0.078	32.1%	28.4%	0.625
	gyomok	2.8%	2.9%	0.290	4.8%	3.3%	0.718

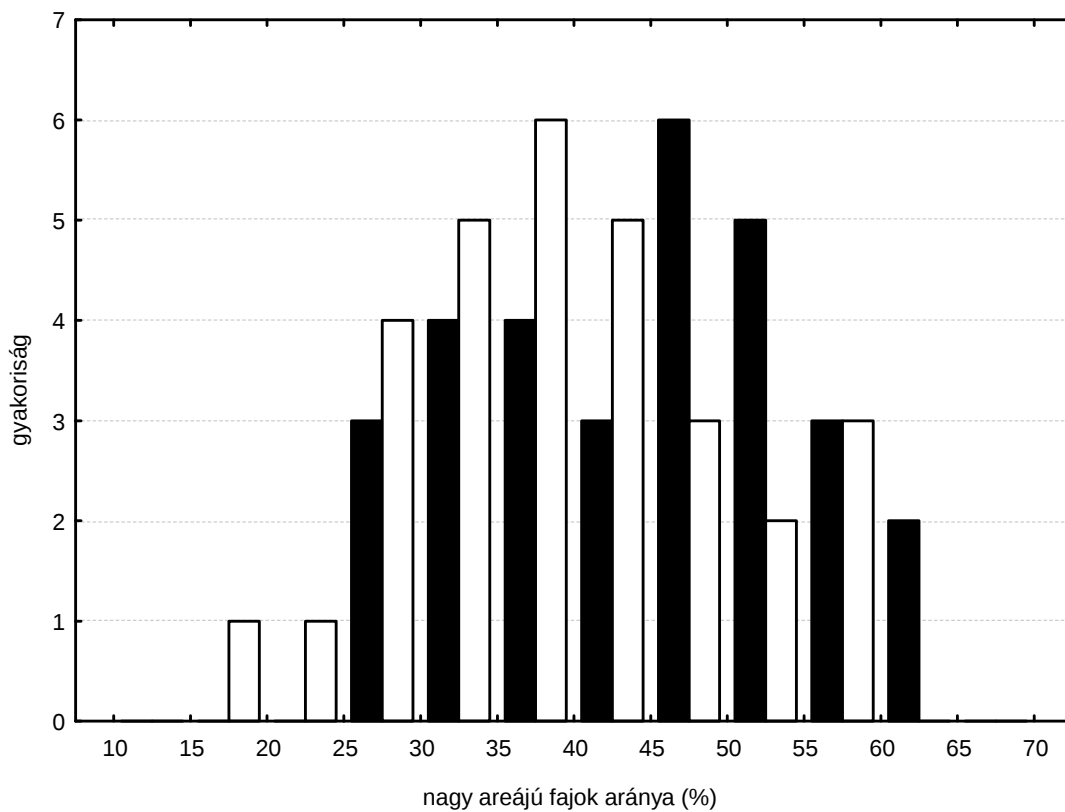
4. táblázat: Az elsőfajú hiba valószínűsége a két mintavételi módszer összehasonlításakor. Az 5%-nál kisebb értékeket vastag betűvel kiemelttem.

		Csévharaszt		Fülöpháza	
		Szmirnov teszt	WMW teszt	Szmirnov teszt	WMW teszt
fajgazdagság		0.003	<0.001	0.980	0.970
életforma	egyévesek	0.549	0.201	0.353	0.403
	évelők	0.549	0.201	0.353	0.403
flóraelem	kis areájú	0.320	0.120	0.688	0.496
	szubmediterrán	0.995	0.524	0.198	0.077
	kontinentális	0.321	0.157	0.525	0.318
	nagy areájú	0.116	0.027	0.334	0.086
élőhely	homoki fajok	0.060	0.003	0.371	0.314
preferencia	erdőssztyepp fajok	0.059	0.079	0.175	0.137
	generalisták	0.683	0.322	0.341	0.096
	gyomok	0.166	0.087	0.998	0.787

dc_16_10

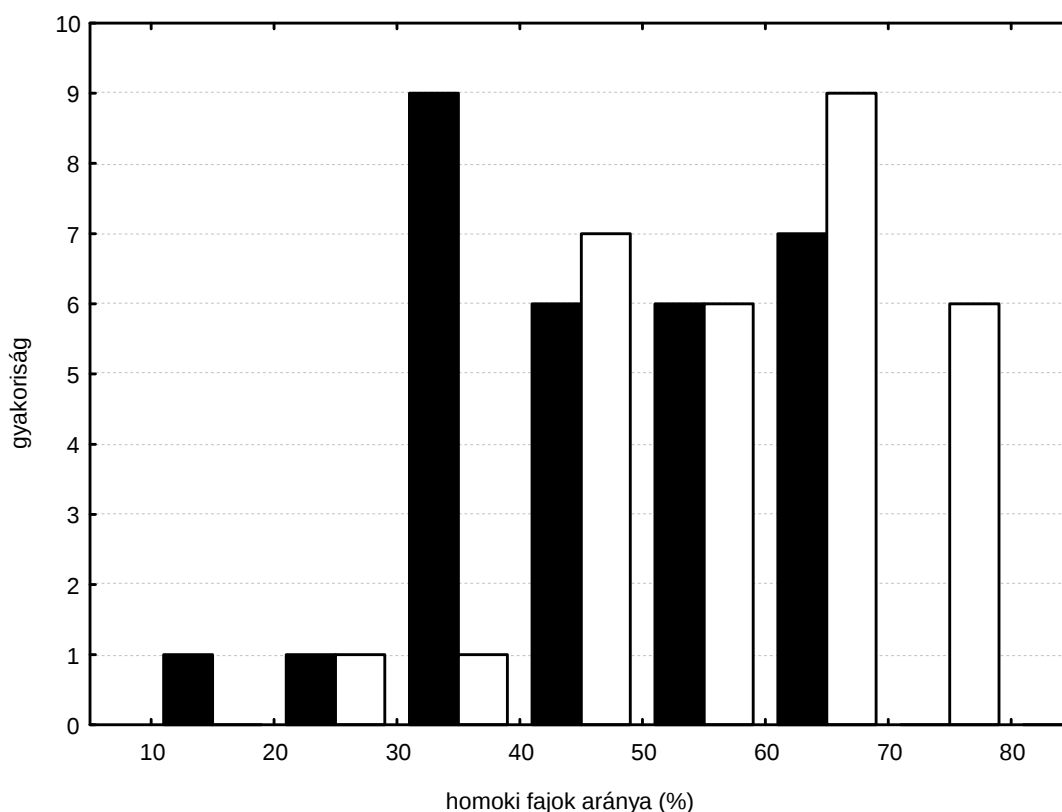


5. ábra: A fajszámok gyakoriságeloszlása a preferenciális (üres oszlopok) és random (fekete oszlopok) mintában Csévharasztban



6. ábra: A nagy areájú fajok csoportrészesedésének gyakoriságeloszlása a preferenciális (üres oszlopok) és random (fekete oszlopok) mintában Csévharasztban

dc_16_10



7. ábra: A homoki fajok csoportrészesedésének gyakoriságeloszlása a preferenciális (üres oszlopok) és random (fekete oszlopok) mintában Csévharaszton

A fajgazdagság és a fajcsoportok csoportrészesedése tekintetében a mintavételi módszerek összehasonlítása a két mintaterületen eltérő eredményt adott. Fülöpházán nem volt egyik vizsgált jellemző tekintetében sem különbség a két mintavétel között (4. táblázat). Csévharaszton viszont (4. táblázat) a preferenciális mintavételben szignifikánsan alacsonyabb volt a fajszám (5. ábra, 3. táblázat) és a nagy areájú fajok aránya (6. ábra, 3. táblázat), és szignifikánsan magasabb a homoki fajké (7. ábra, 3. táblázat).

A preferenciális mintavétel alapján nincs szignifikáns különbség a két mintaterület nagy areájú fajok arányában, míg a random mintavétellel kimutatható, hogy ezek a fajok gyakoribbak Csévharaszton (3. táblázat). Habár a kétféle módszerrel kapott becslések egy mintaterületen belül nem különböznek szignifikánsan, a kétféle mintavétel eltérő eredményt ad, ha a kontinentális fajok csoportrészesedését hasonlítjuk össze a két területen. A random mintavétel alapján ezen fajok csoportrészesedése Fülöpházán nagyobb, míg a preferenciális mintavétel alapján a két terület közötti különbség nem mutatható ki (3. táblázat). A meglepő eredmény magyarázata, hogy a preferenciális mintavétel Fülöpházán alul-, míg Csévharaszton felül-becsülte a kontinentális fajok csoportrészesedését. A preferenciális mintavétel alapján nincs különbség a két terület között a homoki fajok csoportrészesedésében, míg a random mintavétel szerint arányuk Fülöpházán nagyobb (3. táblázat).

3.2.3. Diszkusszió

Az elemzés során egy mintaterületen belül a vártnál kevesebb különbséget találtunk a kétféle mintavétel között. A fajösszetétel szignifikáns különbségét Fülöpházán mindössze két faj okozta. A mintavétel időpontjában mindkét faj felismerése nehéz, mert a *Cerastium* már elszáradt, a *Salsola* pedig még csíranövény. Így a különbséget inkább a felvételezők eltérő tapasztaltsága, mint a felvételi helyek kiválasztásának eltérő módszere okozta.

A származtatott változóknál nem volt szignifikáns különbség a mintavételi módszerek között Fülöpházán, és Csévharaszton is csak három változó esetén találtunk szignifikáns különbséget.

A nyílt homoki gyepeknek mind a két mintaterületen viszonylag nagy, nem vagy csak enyhén zavart állományai vannak. Az állományok ellenállóképessége a gyomokkal és az özönnövényekkel szemben jelentős. Gyenge vagy közepes zavarás után a terület nem gyomosodik el, hanem a társulás „belső pionír” fajai szaporodnak el. Emiatt az atipikus vagy zavart foltok, amelyeket a preferenciális mintavételt végző cönológus elkerülne, ritkák. Azonban a két mintaterület nem teljesen egyforma ebben a tekintetben: Fülöpházán, ahol a nagy foltokat alkotó nyílt gyepp az erdőssztyepp domináns komponense, az erős aszály és a taposás okozta zavarások ellenére is egységes, egyértelműen *Festucetalia vaginatae* jellegű gyepeket találunk. Csévharaszton viszont a gyepp erősen fragmentált, nagy kiterjedésben találunk nyáras-borókás foltokat és tájidegen akáctelepítéseket. A fragmentáltság miatt jelentős a szegélyhatás, ami megnöveli a társulásidegen – pl. üde erdőssztyepp vagy gyom – fajok arányát a homoki gyeppben (Kovács-Láng et al. 2000). A tipikus foltok keresése során azokat a helyeket ahol ezek a fajok nagy tömegben fordulnak elő, elkerüli a felvételező.

Első pillantásra meglepő, hogy a fajgazdagság nagyobb a random mintában. A magyarázat az lehet, hogy a nyílt homoki gyepp egy fajszegény közösség, amelynek a fajkészlete is kicsi. A magasabb fajszám fajgazdagabb társulások (homoki sztyepp, buckaközi zárt gyepek, erdőszegélyek) felé átmenetet mutató állományokra jellemző, illetve degradációt jelez. Ezért a preferenciális mintavételből kimaradnak ezek a fajgazdagabb foltok. Más társulásokban, ahol a degradáció a fajszám csökkenésével jár, valószínűleg ezzel ellentétes eredményt kapnánk (v.ö. Diekmann et al. 2007, Hédli 2007).

A preferenciális mintavétel nem felnagyította, ahogy azt vártuk, hanem lecsökkentette a különbségeket a két mintaterület között. Azonban hangsúlyozni kell, hogy mind a két területen ugyanazt a vegetációtípust mintázták meg. Véleményünk szerint, ha két olyan társulást hasonlítottunk volna össze, amelyek növényföldrajzi helyzetükből (is) adódóan különböznek, a preferenciális mintavétel felnagyítaná a különbségeket, mert a felvételező keresné a növényföldrajzi szituációt jól jelző fajokban gazdag helyeket.

3.3. Következtetések

A fejezet első részében, elméleti megfontolások alapján arra a következtetésre jutottunk, hogy a Braun-Blanquet módszerrel végzett preferenciális felvételezéssel kapott adatsorok is alkalmasak hipotézisek statisztikai tesztelésére. Az alapsokaság azonban, amire az eredmények vonatkoznak más, mint a preferenciális mintavétel esetén. A fejezet második

részében egy konkrét szituációban mutattuk be ezt a különbséget.

Habár lehetne egy listát összeállítani arról, hogy milyen állományokat preferál és milyeneket kerül el egy cönológus a preferenciális mintavétel során, mégsem lehet általános érvényű előrejelzést adni arról, hogy milyen különbségek lesznek az ugyanarról az objektumról preferenciális és random mintavétellel gyűjtött adatok között. Ennek oka az, hogy a cönológus preferenciáit erősen befolyásolják a prekonceptiói a vizsgált vegetációtípusokról. Például, ha ugyanazon a helyen élő különböző társulásokat akarja összehasonlítani, akkor önkéntelenül is hangsúlyozni fogja a köztük levő különbségeket. Ha viszont ugyanak a társulásnak a különböző helyeken előforduló állományait felvételezi, a hasonlóságokat fogja hangsúlyozni. Ezért, ha el akarjuk kerülni a félreértelmezéseket, a preferenciális adatsor elemzése előtt – amennyire lehetséges – meg kell ismernünk a felvételező prekonceptióit. Nagy adatbázisoknál, ahol sok különböző felvételezőtől származó adatot együtt elemzünk, lehetetlen minden egyes felvételező, minden adatforrás esetén az előzmények figyelembe vétele. Szerencsére a különböző felvételezők eltérő prekonceptiói sok esetben kioltják egymás hatását. Ennek ellenére, mindig visszafogottan kell értékelni az eredményeket, megvizsgálva, hogy azt mennyire okozhatta a preferenciális mintavétel.

4. Zajszerűs

A numerikus klasszifikáció célja olyan felvételtcsoportok kialakítása, amelyben a hasonló környezeti körülmények miatt, a növényzet hasonló. Azonban két növényzeti felvétel akkor is jelentősen különbözhet, ha a környezeti feltételek – beleértve a releváns múltbeli feltételeket is – teljesen azonosak. Ezeket a különbségeket, amiket a terjedés és megtelepedés sztochasztikus jellege okoz az elemzés szempontjából zajnak tekinthetjük (Gauch 1982). Gauch (1982) definíciója szerint a zaj: “variation of single species’ abundances not coordinated with other species’ variation”. A metrikus ordinációs eljárásokban azokhoz a tengelyekhez tartozik nagy sajátérték, amelyek több faj gyakoriságának szimultán változását írják le, míg a csak zajt tartalmazó tengelyek sajátértéke kicsi (Gauch 1982). Így a numerikus klasszifikáció bemenő adataként a fajok abundanciái helyett a magas sajátértékű tengelyek menti koordinátákat használva várhatóan megbízhatóbb eredményt kapunk. Ezt az eljárást Equihua (1990) javasolta elsőként, majd tőle függetlenül munkatársaimmal több esettanulmányban (Botta-Dukát et al. 2005, Illyés et al. 2007, 2009) sikeresen alkalmaztuk.

A jelen vizsgálat célja annak megmutatása, hogy a zajszerűs valóban növeli a klasszifikáció hatékonyságát.

4.1. Anyag és módszer

4.1.1. Zajszerűs metrikus ordinációval

Habár Gauch (1982) cikkében csak a főkomponens analízist (PCA), a korrespondencia-analízist (CA) és a DCA-t (detrended correspondence analysis) említi, mint zajszerűsre alkalmas módszereket, érvelése bármely metrikus ordinációs módszerre igaz. Ezért a továbbiakban a metrikus sokdimenziós skálázást fogom ebben a fejezetben használni, amely a többi módszernél sokkal flexibilisebb, mert tetszőleges, az adott probléma szempontjából legmegfelelőbbnek tartott távolságfüggvénnyel kombinálható (Legendre & Anderson 1999). Eredetileg metrikus távolságok elemzésére fejlesztették ki ezt a módszert, de egy megfelelő konstans hozzáadva a távolságmátrix nem-diagonális elemeihez, nem-metrikus távolságok elemzésére is alkalmas, anélkül, hogy negatív sajátértékeket kapnánk (Legendre & Legendre 1998). Az ebben a fejezetben bemutatott példákban Jaccard különbözőséget használtam.

A zajszerűs szempontjából központi kérdés, hogy hány ordinációs tengely tartalmaz érdemi információt. Ha túl kevés tengelyt választunk ki a további elemzéshez, akkor nem csak a zajt távolítjuk el, de érdemi információ is elveszhet. Ha viszont túl magas a további elemzéseknél figyelembe vett tengelyek száma, akkor a zajszerűs nem lesz hatékony. A főkomponens analízis esetén számos módszert javasoltak az értelmezendő tengelyek számának megállapítására, amelyekről Peres-Neto és munkatársai (2005) cikke jó áttekintést ad. Szimulált adatokon végzett összehasonlító vizsgálataik alapján (Peres-Neto et al. 2005) hat módszer alkalmazását javasolják. Ezek közül háromnak a PCA-hoz szorosan kapcsolódó előfeltételei (normalitás és linearitás) vannak, amelyek a metrikus sokdimenziós skálázás esetén nem teljesülnek, ezért itt ezek nem alkalmazhatók. A másik három ajánlott módszer

randomizált adatmátrixok elemzésén alapul. A vegetációs adatok megfelelő randomizálása azonban nem triviális feladat. Az értékek összekeverése egy változó belül - amit Peres-Neto és munkatársai (Peres-Neto et al. 2005) javasolnak – egyes felvételekben irreálisan magas összborítást (összegyedszámot), illetve „üres” felvételeket egyaránt eredményezhet. Miklós and Podani (2004) kidolgoztak egy hatékony algoritmust a bináris adatok randomizálására, amely megőrzi az eredeti sor- és oszlopösszegeket (azaz a fajok frekvenciáit és a felvételek fajszerkezetét), de nem ismerik hasonló megoldást a borítás vagy egyedszám adatokra.

Ezért a törött-pálca eloszláson alapuló szabály (Jackson 1993, Legendre & Legendre 1998) alkalmazását javaslom. Ennek alapgondolata, hogy ha csak zaj van az adatokban (vagyis ha random adatokat elemzünk), a megmagyarázott variáció véletlenszerűen oszlik meg a tengelyek között. Ebben az esetben az j -dik tengely által megmagyarázott variáció arányának várhatóértéke ugyanannyi, mint az j -dik leghosszabb pálcadarab hosszának várhatóértéke, ha egy egységnyi hosszúságú pálcát véletlenszerűen törünk annyi darabra, mint ahány tengelyünk van. Ez a várhatóérték az alábbi képlettel számolható ki:

$$L_j = \frac{1}{p} \sum_{k=j}^p \frac{1}{k}$$

ahol : p = a tengelyek száma

A j -dik ordinációs tengelyt akkor tekintünk fontosnak és vonunk be a további elemzésbe, ha az általa megmagyarázott variáció aránya nagyobb a random esetben vártnál, azaz L_j -nél.

4.1.2. Elemzett adatok

Az elemzések során terepi és szimulált adatokat használtam. A szimulált adatoknál mindkét mintavételi séma (lásd később) esetén 50 ismétlésben végeztem el a mintavételt és azt követően a vizsgálatot.

A szimulált adatok előállításánál minden szimulációban 200 faj válaszgörbéit generáltam random paraméterű béta függvényeket (Minchin 1987) használva egy gradiens (cönoklin) mentén, majd a gradiens 5 pontján, egymástól egyenlő távolságra mintát vettem. Kétféle mintavételi sémát alkalmaztam: minden pontban 20 felvétel készült (egyenlő csoportméret), illetve a felvételek száma 10 és 30 között változott. A szimuláció paramétereit az 5. táblázat tartalmazza. A klasszifikációs eljárásoktól azt várjuk, hogy a gradiens különböző pontjairól származó felvételek külön csoportokba kerüljenek. Előzetes vizsgálatok alapján a mintavételi pontok távolságát úgy állítottam be, hogy zajsűrítés nélkül is jó, de nem tökéletes eredményeket adjanak a klasszifikációs módszerek. A szimuláció során a Friedley és munkatársai (2007) által közölt R script módosított változatát használtam.

A felhasznált 142 terepi felvétel az Északi-középhegység öt különböző sziklagyep-társulásából származik, amelyek öt különböző alapkőzethez kötődnek (6. táblázat). A klasszifikációs eljárásoktól azt várjuk, hogy az alapkőzetnek megfelelően sorolják öt csoportba a felvételeket.

dc_16_10

5. táblázat: A szimulációban használt paraméterek

	azonos mintaméret	egyenlőtlen mintaméret
A gradiens hossza	500 (0-500)	
A válaszgörbe maximum értéke	lognormál eloszlású véletlenszám (átlag=2, szórás=1)	
A fajok optimumának pozíciója	egyenletes eloszlású véletlenszám -1000 és 1500 között	
Nicheszélesség	egyenletes eloszlású véletlenszám 250 és 2500 között	
A válaszgörbe bal oldali ferdeségét beállító paraméter	egyenletes eloszlású véletlenszám 0.1 és 4 között	
A válaszgörbe jobb oldali ferdeségét beállító paraméter	egyenletes eloszlású véletlenszám 0.1 és 4 között	
Mintavételi pontok pozíciója a gradiens mentén	50, 150, 250, 350, 450	
Mintaméret	20 felvétel minden pontban	10, 15, 20, 25, 30 felvétel

6. táblázat: A felhasznált sziklagyepi adatsor felvételeinek legfontosabb adatai. A kvadrát méret egységesen 4 x 4 m volt.

Alapkőzet	Társulásnév	Felvételek helye	Felvételek száma	Forrás
dolomit	Seseli leucospermi-Festucetum pallentis Zólyomi (1936)1958	Budai-hegység	20	Zólyomi in (Török & Zólyomi 1998)
mészkö	Campanulo-Festucetum pallentis Zólyomi1958	Bükk	25	Zólyomi in (Török & Zólyomi 1998)
homokkő	Cleistogeni-Festucetum pallentis Csiky 2002	Nógrád-Heves-Borsodi dombság	28	(Csíky 2002) (10 felvétel) és Rédei publikálatlan adatok (18 felvétel)
andezit	Minuartio-Festucetum pseudodalmaticae (Mikyška 1933) Klika 1938	Visegrádi-hegység	30	Horánszky in (Török et al. 1994)
riolit	Minuartio-Festucetum pseudodalmaticae (Mikyška 1933) Klika 1938	Zemplén-hegység	39	(Simon 1977)

4.1.3. A vizsgált klasszifikációk

A különböző klasszifikációs módszerek eltérő mértékben lehetnek érzékenyek az adatok zajosságára, emiatt a zajszűrés hatása is eltérő lehet a különböző módszerek esetén. Ezért a vizsgálat során négy hierarchikus (átlagos lánc /UPGMA/, teljes lánc, béta-flexibilis / $\beta = -0.25$ /; Podani 1997) és egy nem hierarchikus (pam = partitioning around medoid; Kaufman & Rousseeuw 1990) klasszifikációs eljárást használtam. A zajszűrés nélküli klasszifikáció ebben a vizsgálatban azt jelenti, hogy a klasszifikációs eljárások inputja a Jaccard különbözőségek mátrixa volt. Ugyanezt a távolságmátrixot használtam a zajszűrés során a metrikus sokdimenziós skálázás inputjaként is. Ebben az esetben a klasszifikációs eljárások inputja a felvételek közötti euklideszi távolságok mátrixa a korábban leírt módon kiválasztott fontos tengelyek menti koordinátákból számolva. Mivel a Ward módszer csak az euklideszi távolsággal használható, ezt az algoritmust csak a zajszűrés után használtam.

4.1.4. A klasszifikációk „jóságának” mérése

A klasszifikációk kiértékelése során a kapott csoportosítást az előzetesen várttal hasonlítottam össze. A szimulált adatoknál a várt csoportosítás a felvételek gradiens menti pozícióinak megfelelő 5 csoport. A sziklagyepi adatoknál az 5 alapkőzet elkülönülését vártam. A cönológiai szakirodalom ugyan egy társulásba sorolja a riolit és andezit sziklagyepet, de korábbi vizsgálatunkban (Rédei et al. 2003) jelentős különbségeket találtunk a két alapkőzeten kifejlődött gyepek diverzitás mintázataiban, ami alapján joggal várható, hogy fajösszetételükben is elkülönüljenek.

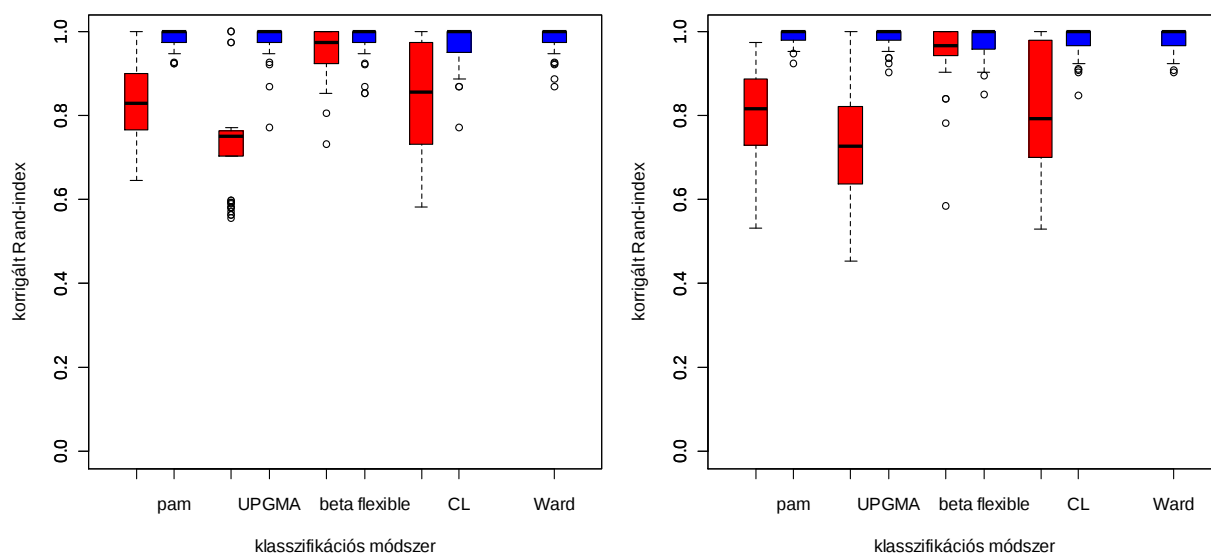
A várt és a megfigyelt csoportosítás közötti egyezést a Rand-index Hubert and Arabie (1985) által módosított változatával mértem (részletesen lásd a 3. mellékletben). Az index maximuma 1, ami azt jelzi, hogy a két csoportosítás teljesen megegyezik. Két véletlen csoportosítás hasonlóságának várhatóértéke nulla.

Az összehasonlításhoz a dendrogramokat 5 csoportból álló partícióvá vágtam szét (5 várt csoport volt mindkét vizsgált adatsorban). A nem-hierarchikus klasszifikációnál is 5-re állítottam be a csoportszámot. A szimulált adatoknál az egyes klasszifikációs módszerek hatékonyságát zajszűréssel és anélkül Wilcoxon páros teszttel, a különböző algoritmusokkal készített csoportosításokat Friedman nem-paraméteres ANOVA-val hasonlítottam össze (Zar 1999).

4.2. Eredmények

A szimulált adatok elemzésekor a zajszűrés – mintavételi elrendezéstől függetlenül – szignifikánsan (az elsőfajú hiba értékei 10^{-3} és 10^{-10} között voltak) javította a klasszifikációk jóságát (8. ábra). A béta-flexibilis algoritmussal kapott klasszifikációk „jóságát” befolyásolta legkevésbé a zajszűrés, vagyis ez a módszer a legkevésbé érzékeny a zajos adatokra. Zajszűrés nélkül jelentős különbség van a „jóság”-ban az algoritmusok között ($p = 10^{-12}$ - 10^{-13}), illetve egy algoritmuson belül az ismétlések között. A zajszűrés mindkét variációt jelentősen csökkentette. Azonos csoportméret esetén a zajszűrés után az algoritmusok jósága közötti különbség nem szignifikáns ($p = 0,13$). Eltérő mintaméret esetén a klasszifikációs módszerek

teljesítménye közötti különbség a csökkenés ellenére is szignifikáns maradt ($p = 0,025$).



a) egyenlő csoportméretek

b) eltérő csoportméretek

8. ábra: A zajszűrés hatása a klasszifikációk jóságára (szimulált adatok). A piros oszlopok a zajszűrés nélkül, a kék oszlopok a zajszűrés után készült klasszifikációk Rand-indexszel mért jóságainak eloszlását mutatják

7. táblázat: A különböző klasszifikációs eljárásokkal kapott csoportosítások hasonlósága (korrigált Rand-index) az alapkőzet alapján várt csoportokhoz a sziklagyepi adatsor esetén.

Klasszifikációs eljárás	Zajszűrés nélkül	Zajszűrés után
partitioning around medoid (pam)	0.8392	0.9587
UPGMA	0.8321	0.7348
béta flexibilis	0.7259	1.000
teljes lánc (CL)	0.6131	0.7282
Ward		1.000

A sziklagyepi adatsor esetén a vizsgált négy klasszifikációs módszer közül háromnál a zajszűrés növelte a kapott csoportosítás hasonlóságát az alapkőzet alapján várthoz képest (7. táblázat). Zajszűréssel kombinálva a béta-flexibilis és a Ward módszer a várttal teljesen megegyező, a *pam* módszer ahhoz nagyon hasonló csoportosítást eredményezett.

4.3. Értékelés

A metrikus sokdimenziós skálázás segítségével végzett zajszűrés mind szimulált, mind terepi adatok esetén növelte a klasszifikációk átlagos jóságát. Néha előfordulhat ugyan, hogy a zajszűréssel rosszabb eredményt kapunk, mint nélküle, de belső (internal) jósági mérőszámok (lásd 6. fejezet) alapján ezek az esetek felismerhetők.

A klasszifikáció jóságának növelésén túl a bemutatott módszernek két további előnye

van: (1) lehetővé teszi olyan klasszifikációs algoritmusok (pl. Ward módszer) alkalmazását, amelyek csak euklideszi távolsággal együtt használhatók, és (2) csökkenti a klasszifikációs módszerek közötti különbségeket. Szimulált adatok elemzése alapján elmondható, hogy nincs olyan klasszifikációs algoritmus, amely minden szituációban felülmúlná a többi: teljesítményük a csoportok méretétől és alakjától függ (Podani 1997). Ha az adataink ezen tulajdonságait nem ismerjük (és a klasszifikáció előtt erről legfeljebb sejtéseink vannak, de általában még az sem), akkor az összevonási algoritmus kiválasztása az elemző szubjektív döntése. A zajszűrés ennek a szubjektív döntésnek a súlyát csökkenti.

A vegetációs adatok zajszűrésére az itt vizsgálaton kívül csak két alternatív megoldást találtam a szakirodalomban, amelyek alkalmazhatósága a metrikus sokdimenziós skálázáson alapuló módszerénél lényegesen korlátozottabb. Tóthmérész (1993) a fajok abundanciáinak nem-lineáris simítását javasolta. Sajnos ez a módszer csak akkor alkalmazható, ha a felvételek nagyjából egyenletesen helyezkednek el egy előre ismert gradiens mentén. A Beals simítás (Beals 1984) szintén használható zajszűrésre (McCune 1994), de nem világos, hogy a simított adatokra milyen távolságfüggvényeket kellene alkalmazni.

Habár Gauch (1982) cikkét sokan idézik (a Google Scholar keresőben 113 idézetet találtam), és valószínűleg a vegetációkutatók többsége tisztában van adataink zajosságával, a metrikus ordináción alapuló zajszűrést alig alkalmazzák a klasszifikációkban. Ennek egyik lehetséges oka, hogy Gauch (1982) világosan bemutatta ugyan, hogy a metrikus ordinációk első néhány tengelye összegzi az ökológiai szempontból releváns információkat, de nem mutatta be a módszer alkalmazhatóságát a klasszifikációban. Az itt bemutatott vizsgálatok ezt a hiányt igyekeztek pótolni.

5. Fidelitás

A vegetációs adatok bármilyen csoportosítása esetén feltehető az kérdés, hogy mely fajokat találunk egy csoportban a véletlenül vártnál gyakrabban, nagyobb egyedsszámmal, nagyobb borítással, vagy éppen mely fajok kerülnek el a csoportot. A véletlen alapján várt gyakoriságtól való eltérést méri a fajok fidelitása. A fidelitáshoz szorosan kötődő karakterfaj-konceptió központi szerepet játszik a Braun-Blanquet-i cönológiai iskola módszertanában (Braun-Blanquet 1928, Westhoff & van der Maarel 1973), a fidelitás alkalmazásának elterjedését azonban gátolta, hogy az objektív mérésére vonatkozó javaslatok (pl. Goodall 1953) nem váltak széles körben ismertté. A fidelitás indexeket bemutató és összehasonlító cikkünk (Chytrý et al. 2002) és a Juice program (Tichý 2002) nyomán mára elterjedt a fidelitás alkalmazása a klasszifikációval kapott csoportok jelentésének feltárására. Újabb fejlemény ezen a területen a páros összehasonlítások helyett az átfogó (overall) mérőszámok alkalmazása, és ehhez kapcsolódóan a fajok megkülönböztető képességének (separation power) mérése, amely szorosan kapcsolódik a klasszifikációk jóságának méréséhez, amivel az 6. fejezet foglalkozik.

5.1. Fidelitás mérőszámok: egy felvételszoport

A legegyszerűbben azt vizsgálhatjuk, hogy egy faj mennyire preferál (vagy éppen mennyire kerül el) egy kiválasztott felvételszoportot. A felvételeket ilyenkor az alábbi 2 x 2-es kontingencia-táblába rendezhetjük:

	a felvétel a csoportba tartozik	a felvétel nem tartozik a csoportba	Összesen
a faj előfordul a felvételen	n_{11}	n_{10}	$n_{1.}$
a faj nem fordul elő a felvételen	n_{01}	n_{00}	$n_{0.}$
összesen	$n_{.1}$	$n_{.0}$	$n_{..}$

A fidelitás mérőszámok azt számszerűsítik, hogy a táblázat belső celláinak értéke mennyire tér el a véletlen esetben várttól, ha a csoport mérete és a faj előfordulási gyakorisága rögzített. A véletlen esetben várható gyakoriságokat az alábbi táblázat mutatja be:

	a felvétel a csoportba tartozik	a felvétel nem tartozik a csoportba	Összesen
a faj előfordul a felvételen	$e_{11} = \frac{n_{.1}n_{1.}}{n_{..}}$	$e_{01} = \frac{n_{.0}n_{1.}}{n_{..}}$	$n_{1.}$
a faj nem fordul elő a felvételen	$e_{01} = \frac{n_{.1}n_{0.}}{n_{..}}$	$e_{00} = \frac{n_{.0}n_{0.}}{n_{..}}$	$n_{0.}$
összesen	$n_{.1}$	$n_{.0}$	$n_{..}$

Másképp megfogalmazva a kérdést, tesztelhetjük, hogy a faj jelenléte és a felvétel csoportba sorolása független-e egymástól. Minél szorosabb köztük az összefüggés, annál nagyobb a fidelitás értéke.

A nominális változók függetlenségének tesztelésére leggyakrabban a khi-négyzet statisztikát használják (Sokal & Rohlf 1981):

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Ennek alternatívája lehet a G-statisztika (Sokal & Rohlf 1981, Botta-Dukát & Borhidi 1999):

$$G = 2 \sum_i \sum_j n_{ij} \ln \frac{n_{ij}}{e_{ij}}$$

Függetlenség esetén mindkét próbastatisztika megközelítőleg 1 szabadsági fokú khi-négyzet eloszlást követ. Az illeszkedés az elméleti eloszláshoz jobb, ha folytonossági korrekciókat (Yates-, illetve Williamson-korrekció) alkalmazunk (Sokal & Rohlf 1981). A korrekciók csak kis mértékben befolyásolják az eredményeket, például a Yates-korrekció esetén a ritka fajok fidelitása relatíve kisebb lesz (Chytrý et al. 2002). Mivel a fidelitás vizsgálatok során célunk nem a függetlenség tesztelése, hanem inkább az összefüggés erősségének mérése, aminek nem feltétele az illeszkedés az elméleti eloszláshoz, a korrekciókat nem szokták alkalmazni.

A kontingencia-táblázat négy belső mezője nem független egymástól, ezért elegendő az n_{11} érték eltérését tesztelni a véletlen esetén várt értéktől. Ennek egyik módja a Fisher-féle egzakt teszt. Chytrý és munkatársai (Chytrý et al. 2002) a teszt egyoldali változatának alkalmazását javasolták, amely azt méri, hogy mennyire valószínű, hogy véletlenül a megfigyelttel azonos vagy annál nagyobb n_{11} értéket kapunk. Ha a táblázat marginálisáiban szereplő értékek rögzítettek, akkor véletlen csoportosítás esetén az n_{11} értékek hipergeometrikus eloszlásúak, ami alapján egy konkrét $n_{11}=k$ érték valószínűsége:

$$P_k = \frac{\binom{n_{1.}}{k} \binom{n_{0.}}{n_{1.}-k}}{\binom{n_{..}}{n_{1.}}} = \frac{n_{1.}! n_{0.}! n_{1.}! n_{0.}!}{k! (n_{1.}-k)! (n_{1.}-k)! (n_{0.}-n_{1.}+k)! n_{..}!}$$

Az elsőfajú hiba valószínűsége, ezeknek az értékeknek az összege:

$$P = \sum_{k=n_{11}}^{\min\{n_{1.}, n_{.1}\}} P_k$$

Mivel a kapott értékek nagyon kis számok, és a kisebb elsőfajú hiba jelent nagyobb fidelitást, Chytrý és munkatársai (Chytrý et al. 2002) a $-\log_{10}P$ -t javasolták, mint fidelitás mérőszámot.

Bruehlheide (1995, 2000) azt javasolta, hogy az n_{11} és e_{11} értékek közötti különbséget fejezzük ki szórás egységekben:

$$u = \frac{n_{11} - e_{11}}{\sigma}$$

Brulheide (1995, 2000) eredetileg azt feltételezte, hogy a faj előfordulási gyakorisága rögzített, de a felvételtcsoport mérete nem. Ebben az esetben, ha a faj előfordulása és a felvételek csoportosítása függetlenek, az n_{11} változó binomiális eloszlású. Az ez alapján számolt szórás (σ_{binB}) alkalmazva aszimmetrikus fidelitás mérőszámhoz jutunk (Chytrý et al. 2002), amely a faj hűségét méri a felvételtcsoporthoz (Juhász-Nagy 1964, Botta-Dukát & Borhidi 1999). Akkor kapunk a kölcsönös hűséget mérő mennyiséget, ha feltételezzük, hogy mind a csoport mérete, mind a faj előfordulási gyakorisága rögzített (Chytrý et al. 2002). Ilyenkor, ha a faj előfordulása független a csoportosítástól, n_{11} hipergeometrikus eloszlású, így a szórás:

$$\sigma_{hyp} = \sqrt{\frac{n_{1.}n_{0.}n_{.1}n_{.0}}{n_{..}^2(n_{..} - 1)}}$$

és a fidelitás mérőszáma:

$$u_{hyp} = \frac{n_{11} - e_{11}}{\sqrt{\frac{n_{1.}n_{0.}n_{.1}n_{.0}}{n_{..}^2(n_{..} - 1)}}}$$

Az u_{hyp} és a χ^2 között függvénykapcsolat van:

$$\chi^2 = u_{hyp}^2 \frac{n_{..}}{n_{..} - 1}$$

Az u_{hyp} statisztika előnye, hogy nem csak a fidelitás erősségét, de annak előjelét is mutatja. Értéke pozitív, ha a faj preferálja a felvételtcsoportot és negatív, ha elkerüli (Chytrý et al. 2002). Ha a faj előfordulása és a felvételtcsoport független egymástól, az u_{hyp} megközelítőleg standard normális eloszlású valószínűségi változó.

Az eddig bemutatott fidelitás mérőszámok lehetséges maximális értéke az adatbázis méretének függvénye, ezért nem alkalmasak adatbázisok közötti összehasonlításokra. Ezt a problémát küszöböli ki a Φ -koefficiens:

$$\Phi = \frac{u_{hyp}}{\sqrt{n_{..} - 1}} = \frac{n_{..}n_{11} - n_{1.}n_{.1}}{\sqrt{n_{1.}n_{0.}n_{.1}n_{.0}}} = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1.}n_{0.}n_{.1}n_{.0}}}$$

amely a khi-négyzet statisztikából is származtatható:

$$\Phi^2 = \frac{\chi^2}{N}$$

A Φ -koefficiens értéke megegyezik a faj előfordulását és a felvételnél a vizsgált csoportba tartozását leíró bináris változók közötti lineáris korrelációval. Ez lehetővé teszi, hogy általánosítsuk tetszőleges abundancia adatokra, hiszen a lineáris korreláció akkor is számolható, ha a faj abundanciáját nem bináris skálán mérjük (De Cáceres & Legendre 2009).

A Φ elméleti maximuma az adatbázis méretétől függetlenül 1. Viszont az adott gyakoriságú faj által elérhető maximum a csoport méretétől függ, akkor 1, ha $n_{.1}=n_{1.}$. Ez félrevezető lehet, amikor azt vizsgáljuk, hogy a faj melyik csoporthoz mennyire kötődik. Ennek elkerülésére Tichý és Chytrý (2006) egy korrigált érték alkalmazását javasolta. Ebben az esetben nem elegendő a vizsgált felvételtcsoport és a többi felvétel megkülönböztetése,

ezért olyan kontingencia-táblázatot kell felírni, amelynek 2 sora és annyi oszlopa van, mint ahány csoportunk van (K):

	1. csoport	2. csoport	...	j. csoport	...	K. csoport	összesen
a faj előfordul	n_{11}	n_{12}	...	n_{1j}	...	n_{1K}	$n_{1.}$
a faj hiányzik	n_{01}	n_{02}	...	n_{0j}	...	n_{0K}	$n_{0.}$
összesen	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.K}$	$n_{..}$

Felhasználva a táblázat jelöléseit a korrigált Φ koefficiens ($\Phi^{(g)}$) a következő módon definiálhatjuk (Tichý & Chytrý 2006, De Cáceres & Legendre 2009):

$$\Phi_j^{(g)} = \frac{n_{..} n_{1j}^{(g)} - n_{1.}^{(g)} n_{.j}^{(g)}}{\sqrt{n_{1.}^{(g)} n_{.j}^{(g)} n_{0.}^{(g)} (n_{..} - n_{.j}^{(g)})}}$$

ahol:

$$n_{.j}^{(g)} = \frac{n_{..}}{K}$$

$$n_{1j}^{(g)} = n_{.j}^{(g)} \frac{n_{1j}}{n_{.j}} = \frac{n_{..} n_{1j}}{K n_{.j}}$$

$$n_{1.}^{(g)} = n_{.j}^{(g)} \sum_{k=1}^K \frac{n_{1k}}{n_{.k}}$$

A Dufrene és Legendre (1997) a fidelitás számszerűsítésére egyedszám adatok alapján az IndVal indexet javasolta:

$$IndVal_j^{(ind,g)} = \frac{\bar{X}_j}{\sum_k \bar{X}_k} \frac{n_{1j}}{n_{.j}}$$

ahol: $IndVal_j^{(ind,g)}$ = az i faj fidelitása a j csoporthoz

$\bar{X}_k$ = a faj átlagos borítása a k csoportban

Az átlagos borítások alkalmazásával a csoportok eltérő méretéből adódó hatásokat kiküszöböli ki (Dufrene & Legendre 1997, De Cáceres & Legendre 2009). A felső indexben az „ind” az egyedszámokon alapuló számításra, a „g” a csoportméret hatásának kiküszöbölésére utal. Bináris adatok esetén azonban a csoportok méretét figyelmen kívül hagyó képletet javasoltak:

$$IndVal^{(pa)} = \frac{n_{11} n_{11}}{n_{1.} n_{.1}}$$

A csoportméretek hatását kiküszöbölő változat bináris adatokra a következő (De Cáceres & Legendre 2009):

$$IndVal_j^{(pa,g)} = \frac{n_{1j}/n_{.j}}{\sum_{k=1}^K n_{1k}/n_{.k}} \frac{n_{1j}}{n_{.j}}$$

Ellentétben a többi bemutatott fidelitás indexszel, az $IndVal^{(pa)}$ és az $IndVal^{(pa,g)}$ értéke nem akkor nulla, ha a faj előfordulásainak száma a csoportban megfelel a véletlen alapján vártnak

($n_{11}=e_{11}$), hanem akkor, ha a faj nem fordul elő a csoportba tartozó felvételek egyikében sem. Ennek a tulajdonságnak a kiküszöbölésére Podani és Csányi (2010) a következő módosított képletet javasolták:

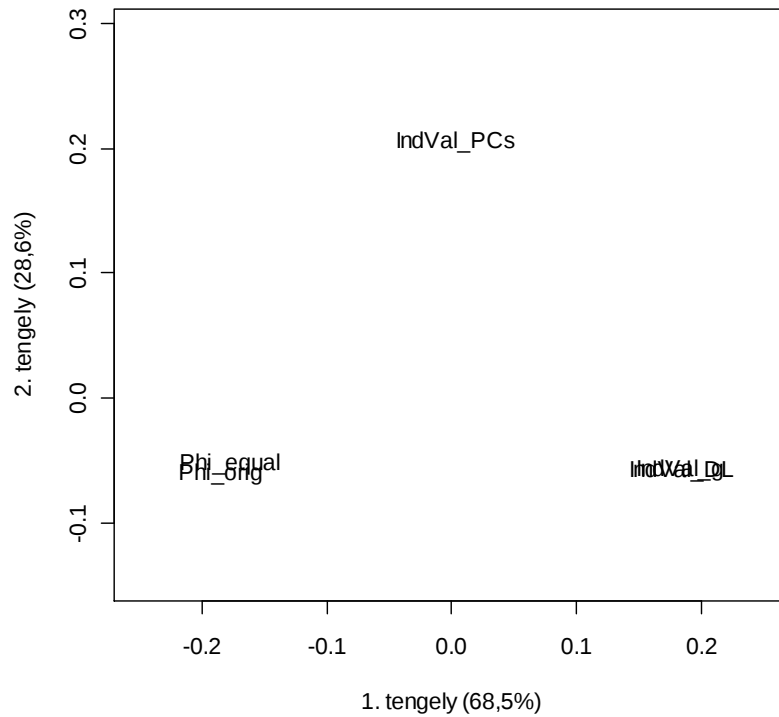
$$IndVal_j^{(pa,PCS)} = \frac{\frac{n_{1j}}{n_{.j}} - \frac{1}{K-1} \sum_{k \neq j} \frac{n_{1k}}{n_{.k}}}{\max_k \left\{ \frac{n_{1k}}{n_{.k}} \right\}}$$

Ha a faj előfordulása és a felvételek csoportba tartozása független, a χ^2 , G és u_{hyp} értékeknek az eloszlása nevezetes eloszlásokkal (1 szabadsági fokú khi-négyzet eloszlás, illetve standard normális eloszlás) közelíthető. Ez alapján megmondható, hogy mik azok az értékek, amelyeket véletlenül csak nagyon kis valószínűséggel kaphatunk, vagyis amelyeket érdemes interpretálni. A Φ és az $IndVal$ mérőszámok esetében azonban nem tudunk ilyen nevezetes eloszlást megadni. Az eredmények túlinterepretálásának elkerülésére ilyenkor is érdemes statisztikai tesztet végezni: amely lehet randomizációs teszt, amikor a felvételek csoporttagságait randomizáljuk, a csoportméreteket nem változtatva (Dufrene & Legendre 1997, De Cáceres & Legendre 2009), vagy lehet Fisher-féle egzakt teszt (Tichý & Chytrý 2006). De Cáceres és Legendre (2009) kimutatták, hogy a Φ és az $IndVal^{(pa)}$ esetén a fidelitás index helyett próbastatisztikaként használható a faj előfordulásainak száma a vizsgált csoportban (n_{1j}). Ebben az esetben a randomizációs teszt a Fisher-féle egzakt teszt közelítő megoldása, elegendő számú randomizáció esetén a két módszerrel kapott első fajú hibák közötti különbség elhanyagolható. Tichý és Chytrý (2006) a csoportmérettel korrigált fidelitás indexek ($\Phi^{(g)}$ és $IndVal^{(pa,g)}$) használatakor is a Fisher-féle egzakt teszt alkalmazását javasolja, míg De Cáceres és Legendre (2009) szerint a csoportméret-korrektíót nem csak a fidelitás mérésekor, hanem a tesztelésnél is alkalmazni kell.

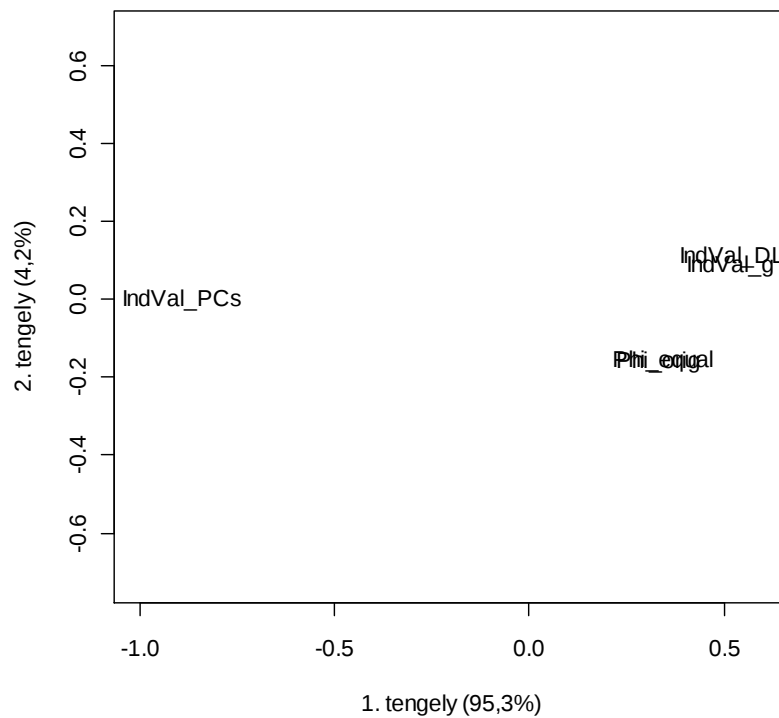
5.1.1. A fidelitás mérőszámok összehasonlítása

Chytrý és munkatársai (2002) vizsgálataiból tudjuk, hogy az $IndVal$ kivételével a kölcsönös hűséget mérő szimmetrikus mérőszámok nagyon hasonló eredményeket adnak. Ezért ezek közül csak egyet, a leggyakrabban használt Φ -t, választottam ki, és ezt hasonlítottam össze a csoportméretet korrigáló változatával, illetve az $IndVal$ különböző verzióival. Az összehasonlításhoz kiszámítottam a fajok fidelitását az alapközetekhez az 4.1.2. fejezetben bemutatott sziklagyepi adatsoron, majd a fidelitás értékek között Kendall rangkorrelációt számoltam. A korrelációk grafikus megjelenítéséhez elkészítettem a módszerek ordinációját metrikus sokdimenziós skálázással az 1-rangkorrelációt használva különbözőségként. Mivel az eredeti adatsorban a csoportok mérete között nincsenek nagy különbségek, az elemzést megismételtem egy módosított adatsoron is kitörölve az eredeti adatsorból 15 dolomit-sziklagyepi felvételt, így 5-re csökkentve a csoport méretét. Mivel legtöbbször csak a legmagasabb pozitív fidelitású fajokat interpretálják, elvégeztem a vizsgálatot csak azokra a fajokra is, amelyek legalább egy index szerint legalább egy csoportnál a 20 legfidélisebb faj között vannak.

dc_16_10

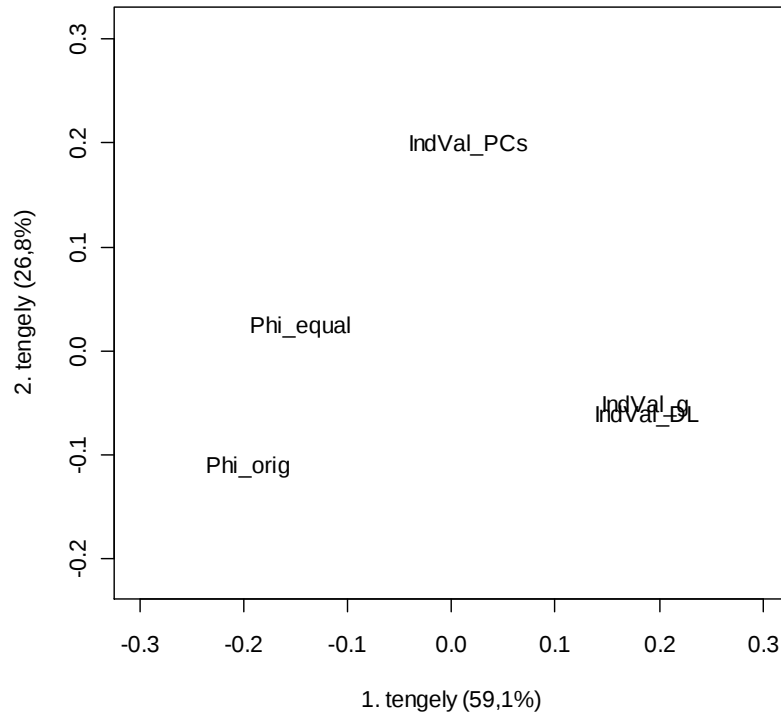


9. ábra: A fidelitás indexek ordinációja metrikus sokdimenziós skálázással az eredeti sziklagyepi adatsorra számolt értékek alapján (összes faj). Távolságfüggvény: 1 – Kendall rangkorreláció

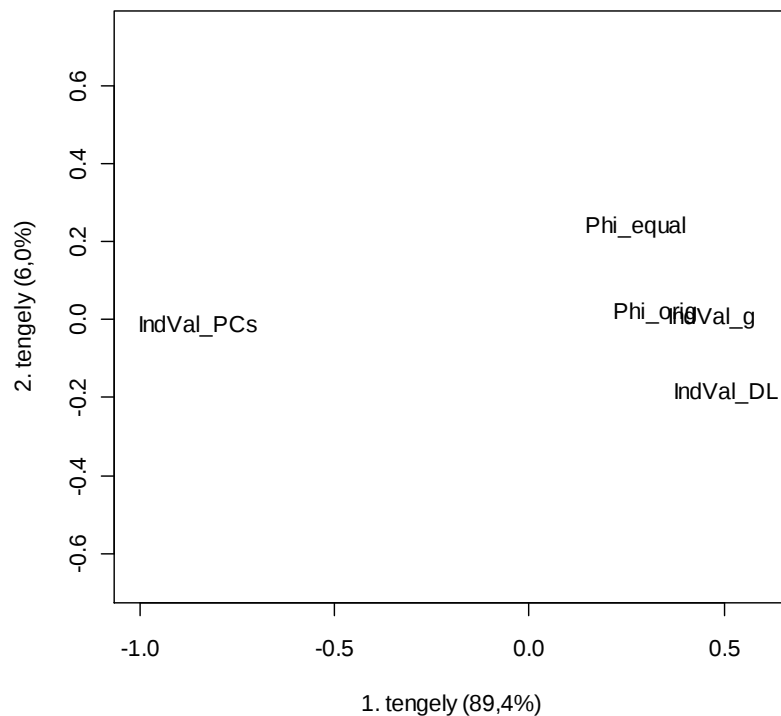


10. ábra: A fidelitás indexek ordinációja metrikus sokdimenziós skálázással az eredeti sziklagyepi adatsorra számolt értékek alapján (csak azokra a fajokra, amelyek legalább egy csoportban a 20 legmagasabb fidelitású faj között vannak). Távolságfüggvény: 1 – Kendall rangkorreláció

dc_16_10



11. ábra: A fidelitás indexek ordinációja metrikus sokdimenziós skálázással a módosított sziklagyepi adatsorra számolt értékek alapján (összes faj). Távolságfüggvény: 1 – Kendall rangkorreláció



12. ábra: A fidelitás indexek ordinációja metrikus sokdimenziós skálázással a módosított sziklagyepi adatsorra számolt értékek alapján (csak azokra a fajokra, amelyek legalább egy csoportban a 20 legmagasabb fidelitású faj között vannak). Távolságfüggvény: 1 – Kendall rangkorreláció

8. táblázat: A fajok különböző képletekkel számolt fidelitásai közötti Kendall rangkorreláció a sziklagyepi adatbázisban

	$IndVal^{(pa)}$	$IndVal^{(pa,g)}$	$IndVal^{(pa,PCs)}$	Φ
$IndVal^{(pa,g)}$	0.986			
$IndVal^{(pa,PCs)}$	0.680	0.681		
Φ	0.628	0.627	0.670	
$\Phi^{(g)}$	0.633	0.635	0.680	0.894

9. táblázat: A fajok különböző képletekkel számolt fidelitásai közötti Kendall rangkorreláció a sziklagyepi adatbázisban (csak azokra a fajokra, amelyek legalább egy csoportban a 20 legmagasabb fidelitású faj egyike)

	$IndVal^{(pa)}$	$IndVal^{(pa,g)}$	$IndVal^{(pa,PCs)}$	Φ
$IndVal^{(pa,g)}$	0.912			
$IndVal^{(pa,PCs)}$	-0.370	-0.361		
Φ	0.715	0.722	-0.117	
$\Phi^{(g)}$	0.710	0.738	-0.135	0.905

10. táblázat: A fajok különböző képletekkel számolt fidelitásai közötti Kendall rangkorreláció a módosított sziklagyepi adatbázisban

	$IndVal^{(pa)}$	$IndVal^{(pa,g)}$	$IndVal^{(pa,PCs)}$	Φ
$IndVal^{(pa,g)}$	0.965			
$IndVal^{(pa,PCs)}$	0.682	0.686		
Φ	0.605	0.600	0.625	
$\Phi^{(g)}$	0.627	0.640	0.690	0.735

11. táblázat: A fajok különböző képletekkel számolt fidelitásai közötti Kendall rangkorreláció a módosított sziklagyepi adatbázisban (csak azokra a fajokra, amelyek legalább egy csoportban a 20 legmagasabb fidelitású faj egyike)

	$IndVal^{(pa)}$	$IndVal^{(pa,g)}$	$IndVal^{(pa,PCs)}$	Φ
$IndVal^{(pa,g)}$	0.782			
$IndVal^{(pa,PCs)}$	-0.302	-0.251		
Φ	0.716	0.646	-0.074	
$\Phi^{(g)}$	0.575	0.718	-0.017	0.709

Az eredeti adatbázisban az összes fajt figyelembe véve a vizsgált 5 index három, egymástól nagyjából egyforma távolságban lévő csoportot alkot (9. ábra, 8. táblázat): az első csoportba a Φ és a $\Phi^{(g)}$ tartozik, a másodikba az $IndVal^{(pa,PCs)}$, a harmadikba az $IndVal$ másik két változata. Csak a legmagasabb fidelitású fajokat vizsgálva az $IndVal^{(pa,PCs)}$ elkülönülése a többi indextől még nagyobb mértékű, míg a másik négy index közötti különbségek lecsökkennek (10. ábra, 9. táblázat).

A várakozásoknak megfelelően, a módosított adatsorban, ahol a csoportok mérete közötti különbség nagyobb, megnő a csoportméret-korrektció hatása. Az $IndVal^{(pa)}$ és az $IndVal^{(pa,g)}$ közötti különbség az összes fajt figyelembe véve nagyon kicsi (11. ábra, 10. táblázat), aminek az oka az, hogy mindkét index értéke nulla, ha a faj nem fordul elő a csoportban. Csak a legfidélisebb fajokat vizsgálva ez a különbség nagyjából ugyanakkora, mint a Φ és a $\Phi^{(g)}$ közötti különbség (12. ábra, 11. táblázat). Az $IndVal^{(pa,PCs)}$ itt is erősen elkülönül a többi indextől, különösen a legfidélisebb fajok esetén.

Ahhoz, hogy megértsük ezeknek a különbségeknek a hátterét, érdemes néhány kiválasztott fajt, amelyeknél a különböző indexek nagyon eltérő eredményeket adnak részletesebben is megvizsgálni (13. táblázat). Az $IndVal^{(pa,PCs)}$ értéke mindig 1, ha a faj csak a vizsgált csoportban fordul elő, függetlenül attól, hogy mennyire gyakori a vizsgált csoportban (lásd *Carex liparicarpos* és *Globularia punctata*; 13. táblázat). Így fordulhat elő, hogy a teljes adatsorban csak egy felvételben előforduló *Carex liparicarpos* fidelitása a dolomithoz nagyobb, mint a mind az öt dolomit-sziklagyepi felvételben előforduló *Helianthemum canum*-é, mert az utóbbi egy mészkősziklagyepi felvételben is előfordul. Ez alapján úgy tűnik, hogy az $IndVal^{(pa,PCs)}$ nem a kölcsönös hűséget méri, hanem a faj hűségét a vegetációtípushoz (Juhász-Nagy 1964, Botta-Dukát & Borhidi 1999, Chytrý et al. 2002). Ezt megerősíti a Chytrý és munkatársai (Chytrý et al. 2002) által javasolt teszt adatsorok elemzése (12. táblázat).

A negatív fidelitás esetén az $IndVal$ különböző változatai nem érzékenyek a faj gyakoriságára (v.ö. *Alyssum alyssoides* és *Euphorbia cyparissias* fidelitása; 13. táblázat), az $IndVal^{(pa)}$ és az $IndVal^{(pa,g)}$ nem jelzi a negatív fidelitást, az $IndVal^{(pa,PCs)}$ korrekten jelzi annak előjelét, de nem érzékeny a faj frekvenciájára. Velük szemben a Φ és a $\Phi^{(g)}$ korrekten jelzi, hogy egy gyakoribb faj hiánya valószínűtlenebb, így nagyobb negatív fidelitást jelent. A csoportméret korrektio hatása elsősorban a gyakoribb fajoknál jelentkezik, mind a Φ , mind az $IndVal$ esetén, utóbbinál csak a pozitív fidelitásoknál (13. táblázat).

12. táblázat: A Φ és az $IndVal^{(pa,PCs)}$ összehasonlítása három teszt adatsor esetén (Chytrý et al. 2002, 4. táblázat nyomán), amelyek a cönológiai hűség különböző formáit reprezentálják.

	I.	II.	III.
Kölcsönös fidelitás	Magas	Közepes	Közepes
A faj fidelitása a vegetációtípushoz	Magas	Magas	Alacsony
A vegetációtípus fidelitása a fajhoz	Magas	Alacsony	Magas
felvételek száma ($n_{..}$)	100	100	100
felvételek száma a vizsgált csoportban ($n_{.1}$)	10	50	10
a faj előfordulásainak száma ($n_{1.}$)	10	10	50
a faj előfordulásainak száma a vizsgált csoportban (n_{11})	10	10	10
Φ	1	0.33	0.33
$IndVal^{(pa,PCs)}$	1	1	0.56

13. táblázat: Néhány kiválasztott faj előfordulási gyakorisága alapkőzetenként és fidelitása a dolomithoz a módosított sziklagyepi adatbázisban

Faj	a faj előfordulásainak száma						a faj fidelitása a dolomithoz				
	dolomit (5 felv.)	mészkeő (25 felv.)	homokkeő (28 felv.)	andezit (30 felv.)	riolit (39 felv.)	Σ	IndVal ^(pa)	IndVal ^(pa,g)	IndVal ^(pa,PCs)	Φ	$\Phi^{(g)}$
<i>Carex liparicarpos</i>	1	0	0	0	0	1	0,200	0,200	1,000	0,440	0,408
<i>Globularia punctata</i>	5	0	0	0	0	5	1,000	1,000	1,000	1,000	1,000
<i>Helianthemum canum</i>	5	1	0	0	0	6	0,833	0,962	0,990	0,909	0,976
<i>Alyssum alysoides</i>	0	1	0	1	1	3	0	0	-0,619	-0,031	-0,071
<i>Euphorbia cyparissias</i>	0	0	22	27	26	75	0	0	-0,653	-0,243	-0,471
<i>Potentilla arenaria</i>	5	8	21	12	32	78	0.064	0.304	0.427	0.160	0.360

14. táblázat: Három példafaj előfordulási gyakorisága alapkőzetenként a sziklagyepi adatsorban és fidelitásuk a savanyú alapkőzetekhez külön-külön és összevontan. A zárójelben szereplő fidelitás-értékeknel a Fisher-féle egzakt teszt nem szignifikáns.

faj	a faj előfordulásainak száma						$\Phi^{(g)}$		
	dolomit (20 felv.)	mészkeő (25 felv.)	homokkeő (28 felv.)	andezit (30 felv.)	riolit (39 felv.)	Σ	andezit	riolit	andezit+riolit
<i>Potentilla argentea</i>	0	0	2	9	6	17	(0,318)	(0,080)	0,290
<i>Inula hirta</i>	0	0	0	6	4	10	(0,293)	(0,088)	0,280
<i>Viscaria vulgaris</i>	0	0	0	6	4	10	(0,293)	(0,088)	0,280

5.2. Páros összehasonlítások helyett átfogó mérőszámok

Ha a fajok fidelitását minden egyes felvételőcsoporthoz külön-külön teszteljük, az ahhoz hasonló, mint ha az egytényezős ANOVA helyett kétmintás t-próbákat csinálnánk minden párosításban: mindkét esetben megnő az elsőfajú hiba elkövetésének valószínűsége. Ennek illusztrálására egy szimulációt készítettem, amelyben a vizsgált faj a 100 felvételtől 20-ban fordult elő, a felvételeket pedig 5 egyforma méretű csoportba soroltam. A faj előfordulását minden szimulációs lépésben randomizáltam, majd Fisher-féle egzakt tesztekkel vizsgáltam kötődését az 5 felvételőcsoporthoz. A szignifikanciaszint 5% volt, vagyis bár a faj előfordulása független volt a csoportoktól, az esetek 5%-ban várhatunk szignifikáns kötődést valamelyik csoporthoz. Ehhez képest az esetek 11,4%-ban kaptam szignifikáns eredményt.

A másik probléma a felvételőcsoportok egyenkénti vizsgálatánál, hogy a generalistább fajok esetleg nem fidelisek egyetlen csoporthoz sem, de fidelisek a csoportok egy részéhez, ha azokat összevonjuk. A 14. táblázat három olyan fajt mutat be, amelynek a fidelitása nem szignifikáns sem a riolithoz, sem az andezithez, de ha összevonjuk a két savanyú alapkőzetet, fidelitásuk az így kapott csoporthoz már szignifikáns. De Cáceres és munkatársai (2010) a Barro Colorado szigeti őserdő fájának élőhelykötődését vizsgálva 44 olyan fajt talált, amely szignifikánsan kötődött egy élőhelyhez. Ha viszont az élőhelykombinációkat is megvizsgálták a szignifikáns élőhelykötődésű fajok száma 64-re nőtt.

Az utóbbi problémát felismerve Tsiripidis és munkatársai (2009) azt javasolták, hogy a fidelitás vizsgálatát több hierarchiaszinten kellene elvégezni. A szüntaxonómiai hierarchia vagy egy hierarchikus klasszifikáció magasabb szintjein megjelenő összevont csoportok azonban nem feltétlenül azokat a csoportokat vonják össze, amelyeket az adott faj preferál. A szüntaxonómiai rendszerben egymástól nagyon távoli csoportoknak – például száraz gyepeknek és száraz erdőknek (Festuco-Bormetea és Quercetea pubescenti-petraeae fajok; Molnár & Rédei 1995) – is lehetnek közös karakterfajai.

De Cáceres és munkatársai (2010) a hierarchiából adódó merevséget feloldandó azt javasolták, hogy az összes lehetséges összevont csoportra ki kellene számolni a fidelitást, és a legnagyobb fidelitású összevont csoportot tesztelni. Ennek a megközelítésnek több hátránya van. Egyrészt csoportok számának növekedésével gyorsan nő az elvégzendő tesztek száma; De Cáceres és munkatársai (2010) szerint K csoport esetén $2^K - 1$ fidelitás értéket kell kiszámolni, azaz 8 csoport esetén 511-t a páronkénti megközelítésnél használt 8-cal szemben³. Ahhoz, hogy az elsőfajú hiba valószínűsége ne nőjön meg, a randomizációs teszt minden egyes lépésében meg kell keresni a legnagyobb fidelitású összevont csoportot, vagyis minden randomizációs lépésben ugyancsak $2^K - 1$ fidelitást kellene kiszámolni. A másik problémát a 15. táblázat szemlélteti: a módszer alapján a faj preferál egy olyan csoportot, amelyben az előfordulási gyakorisága némileg kisebb, mint a teljes adatsorban.

³ Ez csak az IndVal esetén igaz, amely az összes felvételre együtt is kiszámítható (kérdéses, hogy van-e értelme), ha ezt az esetet nem számoljuk, akkor $2^K - 2$ fidelitást kell kiszámolnunk. A szükséges tesztek száma a felére csökkenthető, ha kihasználjuk, hogy a faj fidelitása abszolútértékben ugyanakkora egy csoporthoz és annak komplementeréhez. Így „csak” $2^{K-1} - 1$ fidelitást kell kiszámítani, azaz 8 csoport esetén 255-t.

15. táblázat: Egy példa arra, amikor a De Cáceres és munkatársai (2010) által javasolt módszer a várttal ellentétes eredményt ad. A táblázat 140 felvétel megoszlását mutatja a vizsgált faj előfordulása és a három felvételcsoport (A, B, C) alapján. Zárójelben a faj csoporton belüli előfordulási gyakorisága látható. A Φ akkor lesz maximális, ha a B és C csoportot összevonjuk ($\Phi_{B+C}=0,724$; $\Phi_C=0,710$). Ez alapján a faj a B és a C csoportot preferálja, pedig előfordulási gyakorisága a B csoportban a teljes adatsoron belüli gyakoriságnál kicsit kisebb.

	A	B	C	Σ
a faj előfordul	5 (11.1%)	3 (60%)	80 (88.9%)	88 (62.9%)
a faj hiányzik	40	2	10	52
Σ	45	5	90	140

A felvételek kétféle csoportosításának (a faj jelenléte/hiánya is egy csoportosítás) függetlensége nem csak 2-2 csoport esetén tesztelhető: a khi-négyzet és a G-statisztika tetszőleges méretű kontingencia-táblára kiszámítható (pl. Sokal & Rohlf 1981), sőt a Fisher-féle egzakt teszt is általánosítható 2x2-esnél nagyobb kontingencia-táblákra (Sprent & Smeeton 2001). Meglepő módon a fidelitás vizsgálatára ezeket a tesztek tudomásom szerint eddig nem alkalmazták, csak egy korábbi cikkünkben (Botta-Dukát et al. 2005) használtuk a G-tesztet a faj válogató képességének (lásd a következő fejezetben) mérésére. A khi-négyzet és a G-tesztekben az elsőfajú hiba elméleti eloszlás alapján számolt valószínűsége csak akkor megbízható, ha nincsenek olyan cellák, amelyekben nagyon alacsony a várt esetek száma. A csoportok számának növekedésével előbb-utóbb ezek a cellák törvényszerűen megjelennek, ezért érdemes az elsőfajú hiba valószínűségét randomizációs teszttel megállapítani.

A Φ koefficiens is használható a nagyobb kontingencia-táblák esetén is, ha a χ^2 statisztikából kiindulva definiáljuk:

$$\Phi = \sqrt{\frac{\chi^2}{n_{..}}}$$

Az így definiált Φ -koefficiens előjele mindig pozitív, ezért definiáltuk eltérően a 2x2-es kontingencia-tábla esetén, ahol az előjelnek van értelme. Φ koefficiens értéke más módon is meghatározható: ha a faj előfordulását/hiányát tekintjük függő, és a felvételek csoportosítását kategoriális független változónak, akkor a lineáris modell determinációs koefficiense (R^2) és a Φ koefficiens négyzete egyenlő (bizonyítását lásd a 2. mellékletben). A Φ és a korrelációs koefficiens közötti kapcsolatra 2 csoport esetén már De Cáceres és Legendre (2009) is rámutatott, de ez a kapcsolat több csoport esetén is fennáll. Ez nem csak matematikai érdekesség; Φ koefficiens csak a faj jelenléte/hiánya alapján számolható, míg a lineáris modell alapján az R^2 borítás vagy egyedszám adatokra is számolható. A lineáris modell előfeltevései (pl. normalitás) általában nem teljesülnek a növényzeti adatokra, ezért az elsőfajú hiba valószínűségének becslésére randomizációs tesztet érdemes használni. Az R^2 és a Φ közötti kapcsolat miatt a továbbiakban csak az előbbit fogom használni, mert az

általánosabb (nem bináris abundancia adatokra is használható).

A χ^2 - és G statisztika vagy az R^2 alkalmas a faj abundanciája⁴ és a csoportosítás közötti összefüggés erősségének mérésére, az ezeket használó randomizációs tesztek, vagy a Fisher-féle egzakt teszt pedig arra, hogy eldöntsük, hogy kaphatnánk-e véletlenül is ennyire szoros összefüggést. Nem informálnak viszont arról, hogy mely csoportokat preferálja a faj és melyeket kerüli el. Bináris adatok esetén ennek megállapítására a Freeman-Tukey eltérések ('deviates') használhatók (Sokal & Rohlf 1981, Legendre & Legendre 1998). A j -dik csoporthoz tartozó eltérés:

$$d_j = \sqrt{n_{1j}} + \sqrt{n_{1j} + 1} - \sqrt{4 \frac{n_{1.} n_{.j}}{n_{..}} + 1}$$

Sokal és Rohlf (1981) szerint az eltérés akkor tekinthető α %-os szinten szignifikánsnak, ha abszolútértéke nagyobb, mint $\sqrt{\frac{K-1}{2K}} \chi^2_{1,\alpha}$, ahol $\chi^2_{1,\alpha}$ a 1 szabadsági fokú khi-négyzet eloszlás α %-os szignifikancia szinthez tartozó kritikus értéke. Neu és munkatársai (1974, cit. Legendre & Legendre 1998) csak akkor javasolják a teszt elvégzését, ha $5 \leq \frac{n_{1.} n_{.j}}{n_{..}} \leq n_{..} - 5$, mert egyébként az alkalmazott közelítés pontatlan. Így a felvétels csoportokat – szemben De Cáceres és munkatársai (2010) módszerével – három kategóriába sorolhatjuk: a faj preferálja a csoportot, elkerüli azt, vagy a véletlen esetben várt gyakorisággal fordul benne elő.

5.3. A fajok válogatóképessége⁵

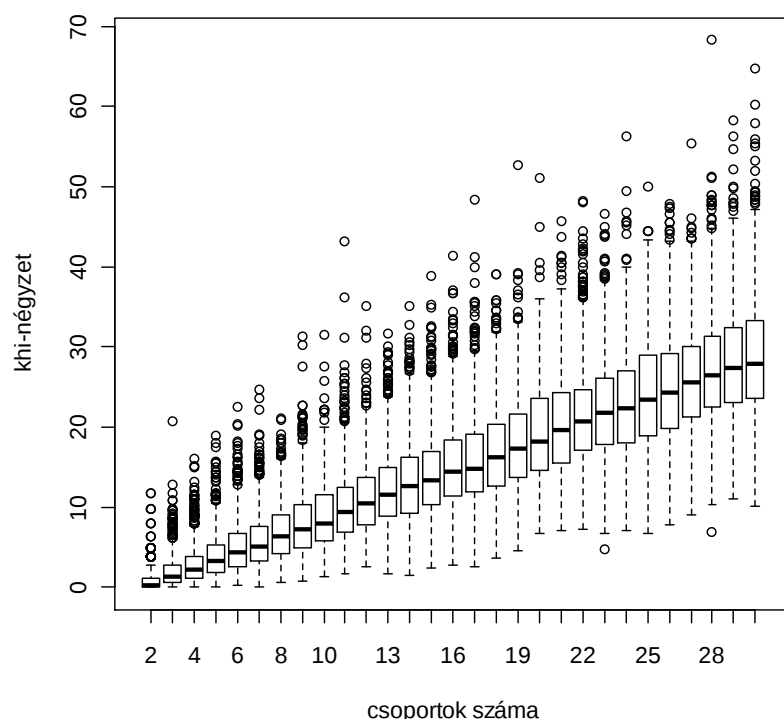
A fidelitás vizsgálatokhoz kapcsolódóan feltehetjük a következő kérdéseket is: A felvételek több alternatív csoportosítása közül melyik alapján mennyire jósolható meg a faj előfordulása (abundanciája)? A faj jelenléte (tömegessége) alapján mennyire jósolható meg, hogy a felvétel melyik csoportba tartozik? Ezekre a kérdésekre a fajok válogatóképességének (separation power) a mérésével adhatunk választ. Egy faj válogatóképessége annál nagyobb, minél inkább megjósolható a faj jelenléte a felvétel csoportja alapján, illetve minél inkább megjósolható a felvétels csoport a faj jelenlétéből/hiányából.

Az előző fejezetben bemutatott átfogó mérőszámok (χ^2 -, G -statisztika, R^2) alkalmasak a fajok válogatóképességének mérésére, de értékük függ a csoportok számától (13. ábra). Ez nem jelent problémát, ha azt vizsgáljuk, hogy egy csoportosítást leginkább mely fajok magyaráznak. Ha viszont különböző csoportosításokat hasonlítunk össze, azok gyakran a csoportok számában is különböznek. Ilyenkor korrigálni kell az értékeket, hogy a csoportok számának hatását eltávolítsuk.

Ha a faj előfordulása és a felvételek csoportosítása független egymástól, a χ^2 és a G -statisztika megközelítőleg $K-1$ szabadsági fokú χ^2 eloszlást követ. Az ilyen eloszlású valószínű-

⁴ A faj jelenléte/hiánya egy bináris skálájú abundancia mérőszámnak tekinthető.

⁵ Az angol „separation power” kifejezés tükörfordítása helyett szemléletesebbnek érzem ezt a megnevezést.



13. ábra: A khi-négyzet statisztika eloszlása különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételben fordul elő).

ségi változók várhatóértéke $K-1$, szórása $\sqrt{2K-2}$. Ez alapján javasoltuk korábban (Botta-Dukát et al. 2005) a G-statisztika értékeinek standardizálását:

$$G_{stand} = \frac{G - (K - 1)}{\sqrt{2K - 2}}$$

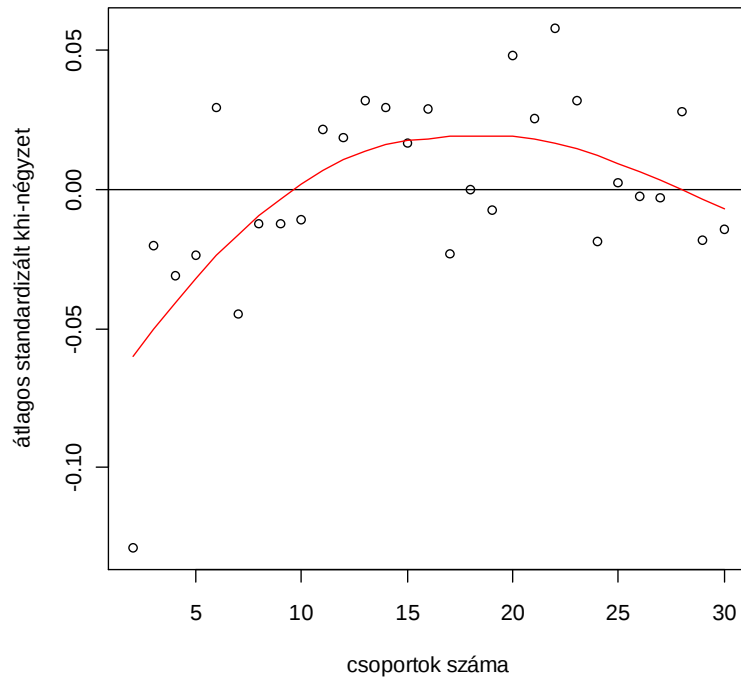
Hasonló standardizálás alkalmazható a χ^2 -statisztika esetén is:

$$\chi^2_{stand} = \frac{\chi^2 - (K - 1)}{\sqrt{2K - 2}}$$

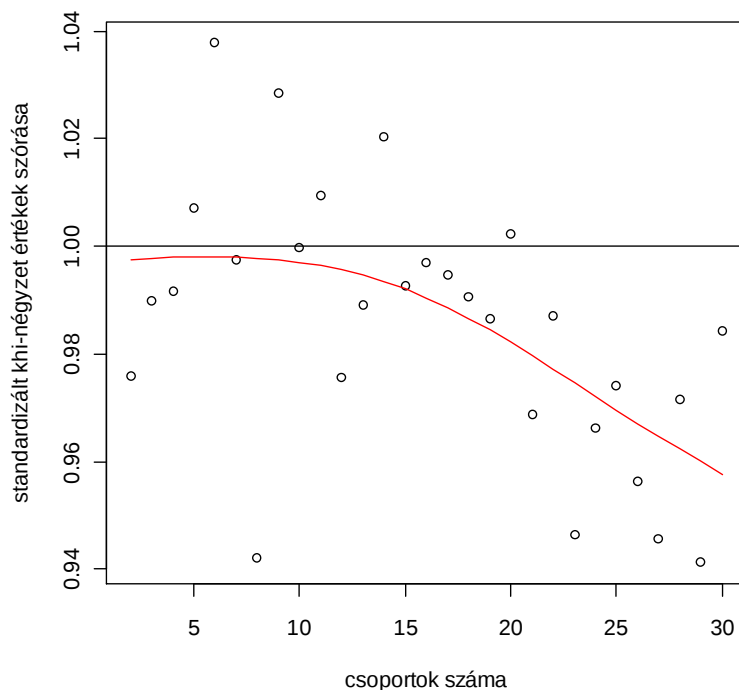
A χ^2 -eloszlás csak akkor illeszkedik jól a valódi eloszláshoz, ha a várt gyakoriság a kontingencia-tábla egyik cellájában sem túl kicsi. Emiatt ritka fajoknál ez a standardizálás nem alkalmazható. A csoportok számának emelkedésével csökken a csoportok mérete, így egyre gyakrabban jelennek meg a kis várt gyakoriságok. Emiatt az elméleti eloszlás illeszkedése a csoportszám növekedésével romlik. Ezért a standardizált értékek sem teljesen függetlenek a csoportok számától: átlaguk emelkedik (14. ábra), szórásuk csökken (15. ábra) a csoportszám emelkedésével.

A függetlenség esetén várható átlagot és szórást a csoporttagságok permutációjával készült random csoportosításokból is becsülhetjük. Ebben az esetben nincs trendje sem a standardizált értékek átlagának (16. ábra), sem a szórásuknak (17. ábra), mindkettő a várt elméleti értékek körül ingadozik. Az eredmények stabilitásához azonban elegendő permutáció alapján kell a standardizálásnál használt átlagot és szórást kiszámolni (18. ábra). A szükséges permutációk számára nem lehet általános receptet adni, minden konkrét esetben érdemes ellenőrizni, mennyire stabilak a standardizált értékek.

dc_16_10

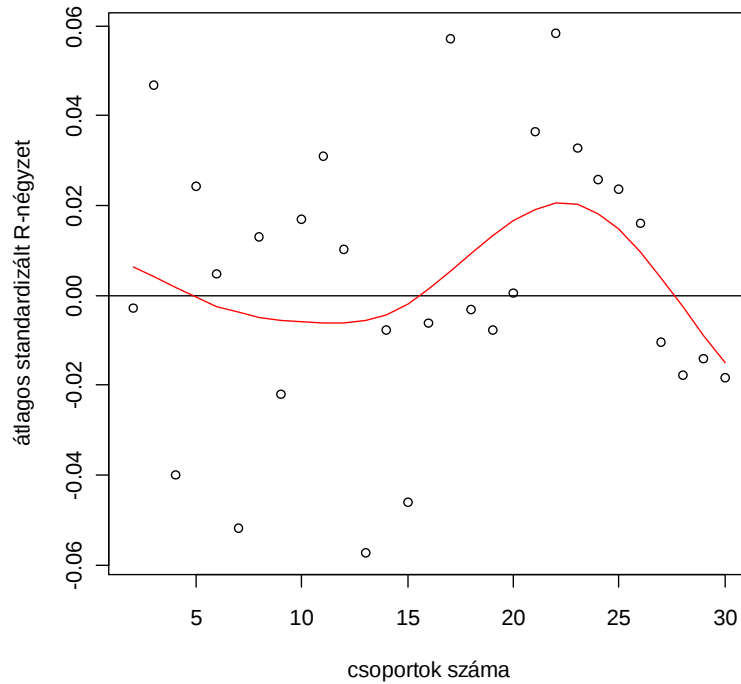


14. ábra: Az elméleti értékekkel standardizált khi-négyzet értékek átlaga (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételen fordul elő). A vízszintes vonal a várt elméleti értéket jelöli. A piros görbe az általánosított additív modellel illesztett trendvonal.

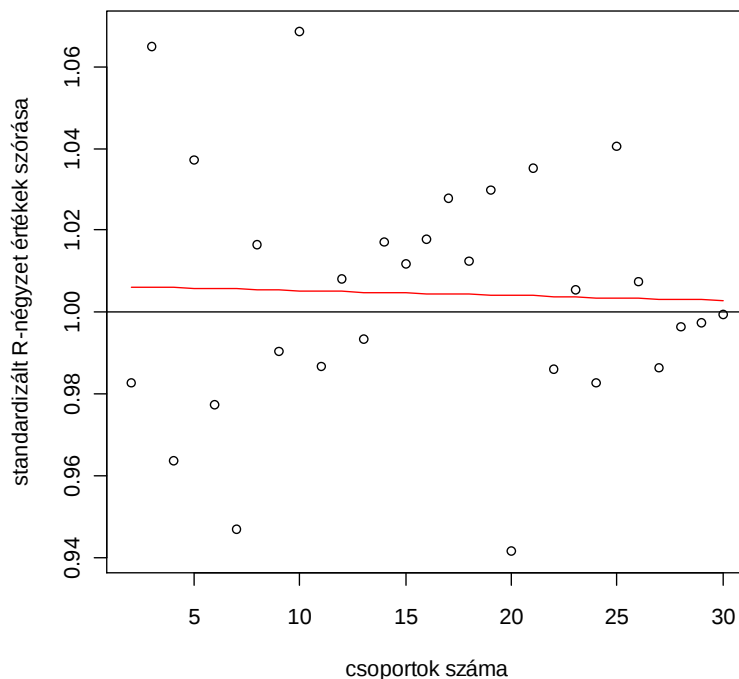


15. ábra: Az elméleti értékekkel standardizált khi-négyzet értékek átlaga (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételen fordul elő). A vízszintes vonal a várt elméleti értéket jelöli. A piros görbe az általánosított additív modellel illesztett trendvonal.

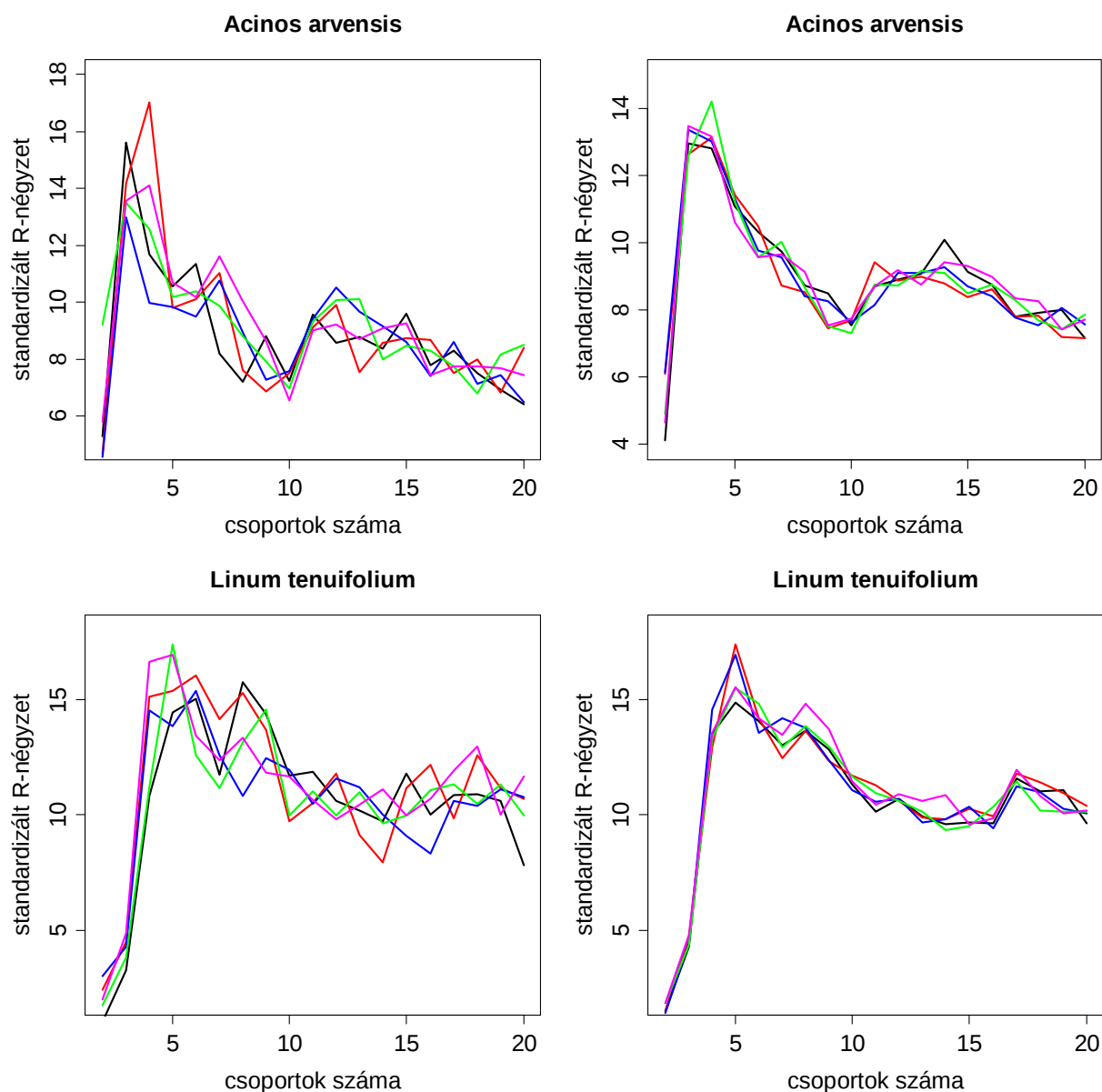
dc_16_10



16. ábra A permutációval becsült (500 permutáció) értékekkel standardizált R^2 értékek átlaga (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételben fordul elő). A vízszintes vonal a várt elméleti értéket jelöli. A piros görbe az általánosított additív modellel illesztett trendvonal.



17. ábra: A permutációval becsült (500 permutáció) értékekkel standardizált R^2 értékek átlaga (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételben fordul elő). A vízszintes vonal a várt elméleti értéket jelöli. A piros görbe az általánosított additív modellel illesztett trendvonal.

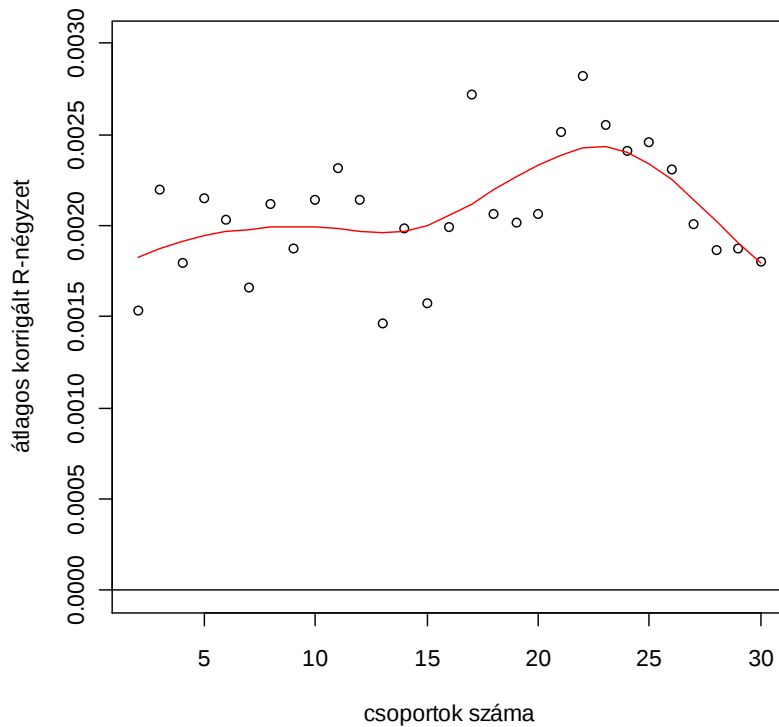


18. ábra: A várhatóérték és a szórás becslésénél használt permutációk számának hatása az eredmények stabilitására. A sziklagyepi adatsor zajsűrés (lásd 5.1.4 fejezet) utáni Ward módszerrel készített hierarchikus klasszifikációját elvágva alakítottam ki a csoportokat. Mindkét faj esetén 5 ismétlésben vizsgáltam a fajok válogatóképességét különböző csoportszámoknál. A bal oldali oszlopban ismétlésenként 100, a jobboldaliban ismétlésenként 500 permutáció alapján számoltam a standardizálásnál használt átlagot és szórást.

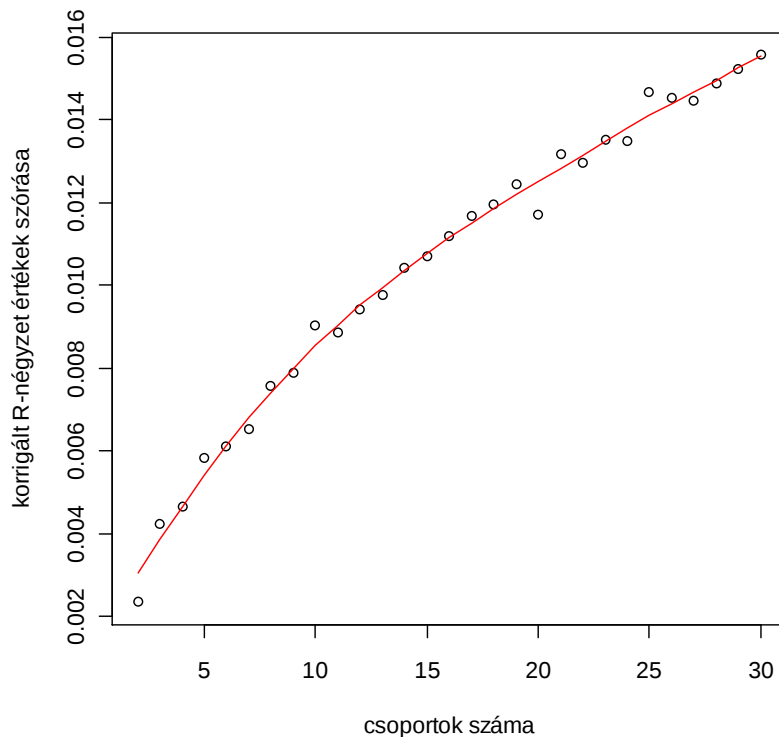
Az R^2 értékek növekedése a paraméterek számával jól ismert a lineáris modelleknél. Esetünkben a paraméterek száma a csoportok száma – 1. A különböző számú paramétert tartalmazó lineáris modellek összehasonlítására a korrigált (adjusted) R^2 -t javasolják:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n_{..} - 1}{n_{..} - (K - 1)}$$

dc_16_10



19. ábra: A korrigált (adjusted) R^2 értékek átlaga (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételen fordul elő). A piros görbe az általánosított additív modellel illesztett trendvonal.



20. ábra: A korrigált R^2 értékek szórása (100 ismétlésből) különböző csoportszámok esetén, ha a faj jelenléte és a csoportosítás független egymástól (felvételek száma: 500, a faj összesen 50 felvételen fordul elő). A piros görbe az általánosított additív modellel illesztett trendvonal.

Hátránya, hogy értéke akkor is pozitív (igaz nagyon kis érték), ha a faj előfordulása független a csoportosítástól (19. ábra) és szórása nem állandó, hanem nő a csoportszám növekedésével (20. ábra).

Dufrene & Legendre (1997) az egyes csoportokra kiszámolt *IndVal* értékek maximumát javasolták a faj válogatóképességének jellemzésére. Ennek hátránya az eddig bemutatott módszerekkel szemben, hogy előnyben részesíti a csak egy csoporthoz hű fajokat, a több csoportot is preferáló fajokkal szemben. De Cáceres és munkatársai (2010) javasolt csoportok összevonásán alapuló módszerrel számolt átfogó (overall) fidelitás is használható a faj válogatóképességének mérésére, de magas csoportszám esetén számításigénye extrém magas lehet. Mindkét esetben szükség van a standardizálásra a permutációk alapján becsült átlagot és szórást felhasználva.

6. A klasszifikációk értékelése⁶

A numerikus klasszifikáció során a kutatónak számos döntést kell hoznia – Milyen adat-transzformációt, távolságfüggvényt, klasszifikációs algoritmust használjon? Hány csoportot különböztessen meg? Minden csoportot értelmezzen, vagy vannak műtermék-csoportok is? – amelyek befolyásolják a kapott eredményeket. A lehetőségek száma óriási, még akkor is, ha a kézikönyvekben szereplő elméleti megfontolások alapján a konkrét esetben szűkíthető a szóba jöhető módszerek köre (Podani 1997). Jó lenne valamilyen objektív módszer, ami segít az eredmények értékelésében, a „legjobb” módszer kiválasztásában. Bár történtek próbálkozások a módszerek általános érvényű összehasonlítására (pl. Belbin & McDonald 1993), ezek alapján csak az általában jó és az általában kevésbé jó módszerek különíthetők el. Nincs egyetlen, minden adatsorra optimális módszer (Kaufman & Rousseeuw 1990), ezért érdemes mindig a konkrét adatsorra legjobb módszert megkeresni.

A fejezetben először áttekintést adok a klasszifikációk értékelése során használható módszerekről. Ez után szimulált adatok alapján vizsgálom, hogy az egyszerű (randomizációt, illetve újramintavételezést nem alkalmazó) indexek – köztük azok, amelyeket munkatársaimmal (Botta-Dukát et al. 2005, Tichý et al. 2010) javasoltunk – mennyire alkalmasak a legjobb csoportosítás kiválasztására, illetve az optimális csoportszám megállapítására. Végül a csoportok repetitivitásának vizsgálatára kidolgozott új módszert mutatom be.

6.1. A klasszifikációk értékelésére használt módszerek áttekintése

Mit is értünk „jó” klasszifikáció alatt? A válasz részben attól függ, hogy egy csoport, egy partíció vagy egy hierarchikus osztályzás „jóságát” akarjuk megmérni (Sousa & Tendeiro 2005). Hagyjuk most figyelmen kívül a hierarchikus klasszifikációkat, hiszen ezek is felfoghatók partíciók sorozataként, és a gyakorlatban is rendszerint a dendrogramot kiválasztott szinten elvágva kapott partíciót értelmezzük. A partíciók jósága kapcsán általában ugyanazokat a szempontokat lehet vizsgálni, mint az egyes csoportok esetén (Handl et al. 2005 nyomán, módosítva):

1. a csoportok mennyire tükrözik a távolságmátrixban levő információt
2. a csoportok tömörsége (compactness), összekötöttsége (connectedness) és elkülönülése (separation)
3. az eredmények robusztussága
 - a. az elemzés során hozott döntésekkel szemben
 - b. a figyelembevett változókkal szemben
4. az eredmények stabilitása az adatok kismértékű megváltozásakor
5. ismétlődő csoportok ugyanabból az alapsokaságból vett párhuzamos minták esetén (repetitivitás)
6. a csoportok értelmezhetősége

⁶ Az angol nyelvű szakirodalomban használt „evaluation”, illetve „validation” kifejezések magyar megfelelőjeként javaslom az értékelést.

7. a csoportok prediktív ereje az osztályozás során nem használt változókra
8. a csoportosítás összehasonlítása a random adatok csoportjaival

Az egyes szempontok számszerűsítésére használt módszereknek nagyon nagy irodalma van. Milligan és Cooper (1985) csak az első két szempont vizsgálatára javasolt módszerek közül több mint harmincat ismertet, hozzátéve, hogy áttekintésük valószínűleg nem teljes. Az alábbi áttekintés sem az, és egy ponton szándékosan nem törekedtem a teljességre: nem tárgyalom azokat az indexeket, amelyek az euklideszi távolságokat vagy az eltérésnégyzet-összegeket használják. Így a gyakran alkalmazott Calinski & Harabasz (1974) indexet sem, amely Milligan és Cooper (1985) összehasonlító elemzésében a legjobbnak bizonyult. Ennek oka, hogy a vegetációs adatok feldolgozására általában nem euklideszi távolságot használunk, elsősorban azért, mert a domináns fajokat túlhangsúlyozza. Nem volna logikus, hogy egy másik távolságfüggvény alapján készült osztályozás jóságát a pontok euklideszi távolságai alapján értékeljük. Az eltérésnégyzet-összeg tulajdonképpen az objektumok négyzetes euklideszi távolságainak összege osztva az objektumok számával (Legendre & Legendre 1998), vagyis az euklideszi távolságról mondottak ezekre az indexekre is igazak. Megfelelő transzformáció után az euklideszi távolság ugyan jól használható lehet vegetációs adatokra is (Legendre & Gallagher 2001, Pinke et al. 2010), illetve a metrikus ordináció tengelyeit használva változóként szintén euklideszi távolságot használhatunk, az eredeti távolságfüggvény helyett, de az általánosabban használható módszerek választéka is elég széles ahhoz, hogy eltekinthessünk a csak speciális esetekben használható módszerektől. Szintén szándékosan nem foglalkozom a lágy (fuzzy) osztályozásokra alkalmas módszerekkel (pl. Pal & Bezdek 1995), mert a lágy osztályozást ritkán alkalmazzák a vegetációkutatók, habár De Cáceres és munkatársai (2009, 2010) szerint jól használhatók lennének a nagy adatbázisokon alapuló szakértői rendszerek kiépítésére.

6.1.1. A csoportosítás mennyire tükrözi a távolságmátrixban rejlő információt?

A sokváltozós elemzés minden lépésében kiemeljük az adott szempontból fontos jellemzőit az adatsornak, és elhanyagoljuk a kevésbé fontosakat. Az első szempont szerint egy klasszifikáció akkor jó, ha a minél jobban leképezi a távolságmátrixban meglevő információt. Ez akkor teljesül, ha az eredeti távolságok, és csoportosítás „ideális” távolságmátrixa (két csoport távolsága 0, ha azonos csoportba vannak, és 1, ha különbözőben) minél jobban korrelál.

1.1. Pont-biszeriális korreláció⁷ (point-biserial correlation) (Milligan & Cooper 1985)

A távolság értéke és pozíciója közötti kapcsolatot mérhetjük a lineáris korrelációval az eredeti távolságmátrix és a csoportosítás alapján kapott távolságmátrix között, amelyben két felvétel távolsága 0, ha azonos csoportba tartoznak, és 1, ha különbözőbe.

⁷ A tesztek értékelésével foglalkozó magyar nyelvű szakirodalomba ezzen a néven került be, a – sok más terület mellett – ott is használható képlet (lásd pl. Rózsa et al. 2006)

1.2 *Tau* (Rohlf 1974, Milligan & Cooper 1985)

Az előző index egy változatának tekinthető, ahol nem a lineáris korrelációt, hanem a Kendall-rangkorrelációt számítjuk ki a két mátrix között.

1.3. *Baker & Hubert* (1975, Milligan & Cooper 1985)

Az összes lehetséges módon párba állítjuk a csoporton belüli és a csoportok közötti távolságokat. C = azoknak a pároknak a száma, ahol a csoporton belüli távolság kisebb, mint a csoportok közötti, D = azoknak a pároknak a száma, ahol a csoporton belüli hasonlóság a nagyobb.

$$G_2 = \frac{C - D}{C + D}$$

Az előző két index variánsának is tekinthető: a két távolságmátrix közötti korrelációt ebben az esetben a Goodman-Kruskall gammával (Goodman & Kruskal 1954) mérjük.

1.4. G_+ (Rohlf 1974, Milligan & Cooper 1985)

Az előző mérőszámmal rokon, de csak azokat a párokat veszi figyelembe, ahol a csoporton belüli hasonlóság nagyobb, mint a csoportok közötti:

$$G_+ = \frac{2D}{n_w(n_w - 1)}$$

ahol:

D = az összes lehetséges módon párba állítva a csoporton belüli és csoportok közötti távolságokat, azoknak a száma, ahol a csoporton belüli távolság a nagyobb

n_w = a csoporton belüli távolságok száma

Minél kisebb az index értéke, annál jobb a csoportosítás.

1.5. *Hubert & Levine* (1976, Dimitriadou et al. 2002)

Az eredeti távolságok és a csoportosítás „ideális” távolságmátrixa közötti rangkorreláció maximális, ha egy csoporton belüli távolság mindig kisebb, mint egy csoportok közötti. Ha kis távolságok vannak a csoportokon belül, akkor az összegük is kicsi. Természetesen az, hogy mi számít kicsinek, attól függ, hogy mekkora értékek szerepelnek a távolságmátrixban, és hány csoporton belüli távolság van. Ezért a csoporton belüli távolságok összegét érdemes az alábbi módon standardizálni:

$$C = \frac{D_w - \min_r\{D\}}{\max_r\{D\} - \min_r\{D\}}$$

ahol:

D_w = a csoporton belüli távolságok összege

$\min_r\{D\}$ = az r darab legkisebb távolság összege

$\max_r\{D\}$ = az r darab legnagyobb távolság összege

r = a csoporton belüli távolságok száma

Az index értéke 0 és 1 között változik, minél kisebb az érték, annál jobb a csoportosítás.

1.6. *McClain – Rao* (1975)

Ha a csoporton belüli távolságok kicsik a csoportok közötti távolságokhoz képest, a csoportok közötti és a csoporton belüli átlagos távolságok hányadosa is kicsi lesz:

$$MR = \frac{\sum_i \sum_j I_{ij} d_{ij} / (N^2 - \sum_k n_k^2)}{\sum_i \sum_j (1 - I_{ij}) d_{ij} / \sum_k n_k (n_k - 1)}$$

ahol:

d_{ij} = az i és j felvétel távolsága

$I_{ij} = \begin{cases} 1, & \text{ha az } i \text{ és } j \text{ felvétel azonos csoportba tartozik} \\ 0, & \text{égyébként} \end{cases}$

n_k = a k csoport mérete

Minél kisebb az index értéke, annál jobb a csoportosítás.

Vendramin és munkatársai (2010) szerint, ha az optimális csoportszám megállapítására használjuk, akkor érdemes az alábbi módon transzformálni:

$$MR^*(K) = abs \left(\frac{MR(K-1) - MR(K)}{MR(K) - MR(K+1)} \right)$$

ahol :

K = a csoportok száma

A transzformált érték 2 csoport esetén nem számolható ki, ezért alkalmazását nem javaslom.

1.7. PARTANA (Aho et al. 2008)

Az előző indexel rokon, de az átlagos távolságok helyett az átlagos hasonlóságok hányadosa. Két felvétel távolsága (d_{ij}) a következő képlettel alakítható hasonlósággá (s_{ij}):

$$s_{ij} = 1 - \frac{d_{ij}}{d_{max}}$$

ahol: d_{max} a távolságfüggvény elméleti maximuma, vagy ha az végtelen, akkor az adatsorban szereplő legnagyobb távolság

Mivel távolságok helyett hasonlóságot használunk, a nagyobb értékek jelzik a jobb csoportosítást.

1.8. Parciális Mantel-teszt, parciális távolság alapú RDA (Anderson & Clements 2000)

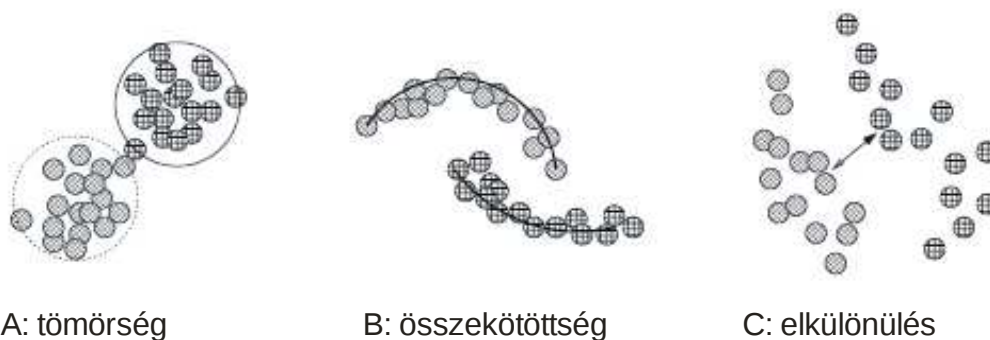
Ha nem csak azt szeretnénk tudni, hogy melyik csoportosítás illeszkedik legjobban a távolságmátrixhoz, hanem azt is, hogy a két csoportosítás közötti különbség szignifikáns-e, akkor parciális Mantel-tesztet vagy parciális távolság alapú redundancia-analízist használhatunk.

6.1.2. A csoportok összekötöttsége, tömörsége és elkülönülése

A csoportok *tömörsége* (compactness) a kicsi csoporton belüli különbségeket jelenti, az *összekötöttség* azt jelenti, hogy az egymáshoz közeli felvételek azonos csoportba tartoznak, míg a csoportok *elkülönülése* akkor maximális, ha a csoportok közötti távolságok maximálisak (21. ábra). A klasszifikációs algoritmusok is ezen tulajdonságok valamelyikét⁸ optimalizálják (például a k -közép és az átlagos lánc módszerek a csoportok tömörségét

⁸ A módszerek többsége csak a három szempont egyikét veszi figyelembe. Kivételnek tekinthető Podani (1989) globális optimalizáló algoritmus, amely a csoportok tömörségét és elkülönülését egyszerre veszi tekintetbe.

maximalizálják, az egyszerű lánc viszont a csoportok összekötöttségén alapul). A csoportszám növekedésével még random csoportosítás esetén is nő a csoportok tömörsége, viszont csökken az elkülönülésük. Ezért a módszerek többsége valamilyen formában kombinálja ezeket a szempontokat, annak érdekében, hogy a csoportszám növelésével maximumon (vagy minimumon) átmenő görbét kapjunk. Handl és munkatársai (2005) alternatív megoldásként javasolták, hogy egyszerre több szempontot is vizsgáljunk, és alkalmazzuk a Pareto kritériumot: az A partíció akkor jobb a B partíciónál, ha minden szempontból legalább olyan jó, és legalább egy szempontból jobb. Ennek a megközelítésnek az előnye, hogy nem tüntetünk ki semmilyen alakú csoportot eleve a jóság mérésére használt index kiválasztásával, hátránya, hogy gyakran nem lehet két csoportosítás közül a jobbat kiválasztani. Ezért a továbbiakban csak a szempontokat kombináló, és így a csoportosításokat egyértelműbben rangsoroló indexekkel foglalkozom.



21. ábra: A csoportok három jellemző tulajdonságának szemléltetése (Handl et al. 2005 nyomán)

2.1. Dunn (Dunn 1974, cit. Xu & Wunsch 2009)

A két legközelebbi csoport távolságának és a legkevésbé kompakt csoport heterogenitásának a hányadosa:

$$D = \frac{\min_i \min_j D_{ij}}{\max_i H_i}$$

Dunn (Dunn 1974, cit. Xu & Wunsch 2009) eredeti javaslata szerint két csoport távolsága a legközelebbi pontjaik távolsága, a csoport heterogenitása (átmérője), pedig a csoporton belüli legnagyobb távolság ($D_{min-max}$):

$$D_{ij} = \min_{k \in C_i} \min_{l \in C_j} d_{kl}$$

$$H_i = \max_{k \in C_i} \max_{l \in C_i} d_{kl}$$

ahol:

C_i = az i csoport

d_{kl} = a k és l felvétel távolsága

Az így definiált index azonban érzékeny a kilógó adatokra (Xu & Wunsch 2009), ezért a csoportok átmérője és két csoport távolsága másképp is definiálható (Pal & Biswas

1997, Bezdek & Pal 1998). Például a csoport heterogenitása mérhető a csoporton belüli átlagos távolsággal, két csoport távolsága pedig a pontjaik átlagos távolságával ($D_{\text{avg-avg}}$):

$$D_{ij} = \frac{\sum_{k \in C_i} \sum_{l \in C_j} d_{kl}}{n_i n_j}$$

$$H_i = \frac{\sum_{k \in C_i} \sum_{l \in C_i} d_{kl}}{n_i(n_i - 1)}$$

ahol:

n_i = az i csoportba tartozó felvételek száma

2.2. Módosított Davis-Bouldin index

A Dunn index a legkisebb elkülönülés és a legkevésbé tömör csoport alapján értékeli a partíciót. A Davis-Bouldin index esetén viszont minden egyes csoportnak kiszámítjuk a jóságát, majd ezeket az értékeket átlagoljuk:

$$DB = \frac{1}{K} \sum_i \max_{j \neq i} \frac{H_i + H_j}{D_{ij}}$$

Az eredeti verzióban (Davies & Bouldin 1979) a csoportok heterogenitását a csoporton belüli eltérésnégyzet-összeggel, a csoportok távolságát pedig centroidjaik távolságával mérik. Itt is alkalmazhatók azonban ezekre a mennyiségekre más definíciók is. A továbbiakban az index módosított verzióját használom, ahol a csoport heterogenitását a csoporton belüli átlagos távolsággal, két csoport távolsága pedig a pontjaik átlagos távolságával mérjük. Minél kisebb az index értéke, annál jobb a csoportosítás.

2.3 Popma (Popma et al. 1983)

Popma és munkatársai (Popma et al. 1983) a módosított Davis-Bouldin indexhez hasonló indexet javasolták:

$$PT_u = \frac{1}{K} \sum_i \frac{H_i}{\max_{j \neq i} D_{ij}}$$

Ez a súlyozatlan index minden csoportot azonos súllyal vesz figyelembe. Annak érdekében, hogy ne a kisméretű, kompakt és jól elkülönülő csoportokat tartalmazó csoportosításokat részesítse előnyben, Popma és munkatársai (Popma et al. 1983) javasolták a csoportmérettel súlyozott átlag alkalmazását:

$$PT_w = \frac{1}{N} \sum_i \frac{n_i H_i}{\max_{j \neq i} D_{ij}}$$

ahol:

n_i = az i csoportba tartozó felvételek száma

N = az elemzett felvételek száma

Minél kisebb az index értéke, annál jobb a csoportosítás.

2.4. Átlagos sziluett (Kaufman & Rousseeuw 1990)

Az előző módszerektől eltérően nem közvetlenül a csoportok tömörségét és

elkülönülését hasonlítjuk össze, hanem az egyes felvételek sziluettjét (S_i) számoljuk ki, amely azt mutatja meg, hogy mennyivel hasonlít jobban a saját csoportjához (tömörség), mint a hozzá legközelebbi csoporthoz (elkülönülés):

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

ahol:

a_i = az i felvétel átlagos távolsága a vele azonos csoportba tartozó felvételtől

b_i = az i felvétel átlagos távolsága a hozzá legközelebbi csoport tagjaitól, azaz attól a csoporttól, amelynek tagjai átlagosan legközelebb vannak hozzá

Az egyes felvételek sziluettjei alapján kiszűrhetők az átmeneti helyzetű vagy rossz csoportba sorolt felvételek, míg az átlagos sziluett a csoport elkülönülését, illetve a teljes klasszifikáció jóságát méri.

2.5. Helyesen klasszifikált felvételek aránya (Feoli et al. 2006)

Egy felvételt helyesen klasszifikáltnak tekinthetünk, ha átlagos távolsága a saját csoportjának tagjaitól kisebb, mint átlagos távolsága bármely más csoport tagjaitól, vagyis a sziluettje pozitív. Hasonlóan a sziluetthez, ez az index is az egyedi felvételek szintjén méri a csoport tömörségét és az elkülönülését, majd utólag összesíti a csoportra, illetve a teljes osztályzásra.

2.6. Feoli (Feoli et al. 2009)

A felvételek hasonlóságai alapján kiszámítható a felvétels csoportok hasonlósága is. Bár többféle definíció is elképzelhető, a legcélszerűbbnek az látszik, ha a két csoport hasonlósága az elemeik átlagos hasonlósága. Ideális esetben a felvétels csoportok hasonlóságmátrixában a főátlón kívüli elemek értéke nulla (minden csoport maximálisan különbözik egymástól). Ilyenkor a hasonlóságmátrix minden sajátértéke ugyanakkora, a sajátértékek egyenletessége maximális. Ebből kiindulva javasolták Feoli és munkatársai a sajátértékek egyenletességét a csoportosítás jóságának mérésére:

$$L = \frac{\ln \Lambda - \sum \lambda_i \ln \lambda_i}{\ln k}$$

ahol:

λ_i = az i -dik sajátérték

$$\Lambda = \sum \lambda_i$$

6.1.3. Az eredmények robosztussága

Ha az eredmények *robosztusak a módszerválasztással szemben*, vagyis ha ugyanazt az adatsort többféle módszerrel elemezve ugyanazokat a csoportokat kapjuk, akkor joggal gondolhatjuk, hogy azok az adatok szerkezetéből adódnak és nem a konkrét elemzőmódszer által előállított műtermékek. Ez a logika alkalmazható az optimális csoportszám megállapítására is: legjobb az a csoportszám, ahol a párhuzamos elemzések eredményei a

leg hasonlóbba (Chiang & Mirkin 2010). A különböző módszerekkel előállított eredményeket összehasonlítva (ennek lehetőségeiről lásd Podani 1997) megvizsgálhatjuk, hogy melyik döntésünk mennyire befolyásolta az eredményeket (pl. Virágh 1987). A különböző módszerekkel kapott eredményekből konszenzus klasszifikációt készítve (Podani 1997) kiemelhetjük a csoportosítások módszer-független jellemzőit.

A filogenetikában gyakran alkalmazott módszer a szekvencia adatok újramintavételezése bootstrap módszerrel (Felsenstein 1985), amivel az eredményeknek *figyelembevett változókkal szembeni robosztusságát* vizsgálják. A bootstrap mintavétel során az eredeti (nukleinsav vagy fehérje) szekvenciából visszatevéssel vesznek mintát, így előállítva az eredetivel azonos hosszúságú pszeudoszekvenciákat. Minél hasonlóbba a pszeudoszekvenciák elemzésével kapott eredmények, annál kevésbé érzékeny a módszer a kiválasztott változókra. Vegetációs adatok elemzésekor a módszer azt jelenti, hogy az eredeti fajlistából visszatevéssel választanánk ki fajokat, vagyis bizonyos fajok kimaradhatnak, mások viszont többször is szerepelhetnek. A filogenetikai elemzésekben az elvileg megvizsgálható szekvenciáknak csak egy nagyon kis része kerül bele ténylegesen az elemzésbe, ezért nagyon fontos, hogy a változók kiválasztásával szemben robosztus eredményeket kapjunk. A vegetációs vizsgálatban viszont általában az összes lehetséges változó – a felvételekben előforduló összes növényfaj – szerepel, ezért a „jósnak” ez az aspektusa kevésbé fontos. A génexpressziós (microarray) vizsgálatok esetén *figyelembevett változókkal szembeni robosztusság* mérésére Datta és Datta (2003) az egy-egy változó kihagyása által okozott átlagos változás kiszámítását javasolta. Ez vegetációs adatok esetén egy-egy faj kihagyását jelenti. Az ilyen robosztusságnak már van jelentősége a vegetációs vizsgálatokban, mert a taxonómiai bizonytalanságok miatt (v.ö. Jansen 2009, Jansen & Dengler benyújtva) fontos, hogy az eredmények ne legyenek érzékenyek egy-egy faj esetleg hibás adataira.

6.1.4. Az eredmények stabilitása

Az *eredmények stabilitásának vizsgálata*kor kis mértékben megváltoztatjuk a kiindulási adatokat, és megvizsgáljuk, hogy mennyire változott meg az eredmény. Az adatok megváltoztatásának leggyakoribb módja az objektumok – és nem a változók, mint a robosztusság esetén – visszatevés nélküli újramintavételezése (bootstrap módszer), de történhet egyes felvételek kihagyásával (a legtöbb osztályozó módszer szempontjából a bootstrap egyenértékű ezzel), random adatpontok vagy random zaj hozzáadásával is (Hennig 2007). Közös jellemzője ezeknek a módszereknek, hogy az eredeti és a módosított adatsor nem független egymástól, mivel a módosított adatsor felvételeinek legalább egy része az eredeti adatsornak is eleme. Az értékelés során ezeknek a mindkét adatsorban előforduló elemeknek a két adatsor alapján kapott csoportosítását hasonlítják össze. A módszer előnye, hogy a módosítás random módon sokszor elvégezhető, így a filogenetikában alkalmazott bootstrap technikához hasonlóan százalékos megbízhatóság-értékeket kaphatunk. Vizsgálhatjuk a teljes csoportosítás stabilitását (pl. Levine & Domany 2001), vagy az egyes csoportokét (Bertrand & Bel Mufti 2006, Hennig 2007), de felhasználható a módszer az

optimális csoportszám megállapítására is (Guidi et al. 2009). Az eredeti és a módosított adatsor közti átfedés miatt, a csoportok stabilitása mindig nagyobb, mint annak a valószínűsége, hogy független mintákból ugyanazokat a csoportokat kapjuk, ezért ha lehetőség van rá, érdemes inkább az utóbbi szempontot vizsgálni (Lange et al. 2004).

Pillar (1999) meglehetősen bonyolult módszere a jó elválás és a stabilitás követelményét kombinálja: a csoportok hasonlóságát a bootstrap minta csoportjaihoz összehasonlítja a csoportok hasonlóságával a legközelebbi csoportból vett bootstrap mintához.

6.1.5. Az eredmények repetitivitása

A klasszifikációs eljárásokat szokás sokváltozós adatfeltáró módszernek tekinteni (Podani 1997). Az ebben az értelemben vett adatfeltárás feladata: „lényegkiemelés, a látens struktúrák felismerése, láthatóvá tétele, vagy egyszerűen csak a biológiai mintázatok leírása (deskripció) és tömör összefoglalása, megmagyarázása” (Podani 1997, 6. oldal). A statisztikában az adatfeltárást azonban sokszor ennél szűkebb értelemben használják, azokra a vizsgálatokra, amelyek célja tesztelhető hipotézisek előállítása⁹. Az így nyert hipotéziseket aztán független mintán tesztelhetjük.

A klasszifikációs módszerek nem feltételezik, hogy adatsorunk egy nagyobb alapsokaságból vett minta, és általában a kutatók sem így tekintenek az adataikra. Ezért, bár az eredmények repetitivitása, mint elvárás régen megjelent (McIntyre & Blashfield 1980), viszonylag ritkán alkalmazzák a csoportosítások értékelésére. Pedig talán az egyik legfontosabb szempont, hiszen a cél általános érvényű, nem csak a vizsgált adatsorra érvényes csoportosítások felállítása. McIntyre and Blashfield (1980) megfogalmazása szerint: *"If a cluster solution is repeatedly discovered across different samples from the same general population, it is plausible to conclude that this solution has some generality. Certainly, a cluster solution which is not stable is unlikely to have general utility."*

A 22. ábra egy egyszerű példán mutatja be a módszer lényegét. Ugyanabból az alapsokaságból vett két párhuzamos mintát k-közép módszerrel 2 csoportba osztva nagyon eltérő eredményeket kapunk, vagyis ezek az osztályozások nem tükrözik jól az alapsokaság szerkezetét. Megismételve az elemzést négy csoporttal, mindkét mintában hasonló csoportokat kapunk, ezért feltételezhetjük, hogy ezek a csoportok az alapsokaságra is jellemzők.

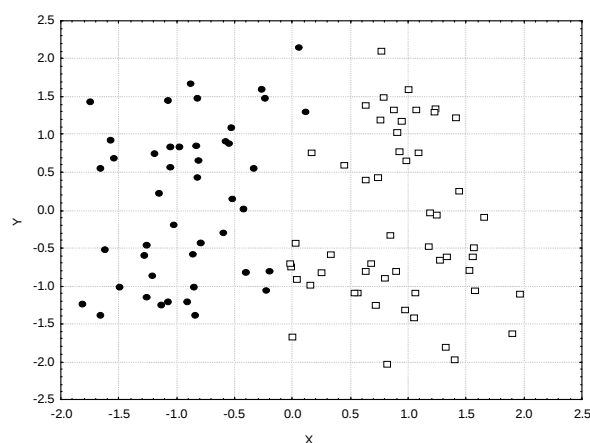
Ha egy csoport több mintában is megjelenik, ez még nem feltétlenül jelenti, hogy az alapsokaságban is van egy ennek megfelelő elkülönült csoport (Krieger & Green 1999). Beauchaine & Beauchaine (2002) példája szerint, ha egy normális eloszlású változóból – ilyen például az IQ – párhuzamos mintákat veszünk és azokat két részre osztjuk, akkor mindig kapunk alacsony és magas átlagú csoportot. Ez azonban nem jelenti azt, hogy az alapsokaságban is létezik két elkülönülő csoport, például az alacsony és magas IQ-jú emberek csoportja. Vagyis a repetitivitás hiánya bizonyíték arra, hogy a csoportosítás nem

⁹ Az ebben az értelemben vett sokváltozós adatfeltárás ökológiai alkalmazásának szép példája Hallgren és munkatársainak (1999) cikke

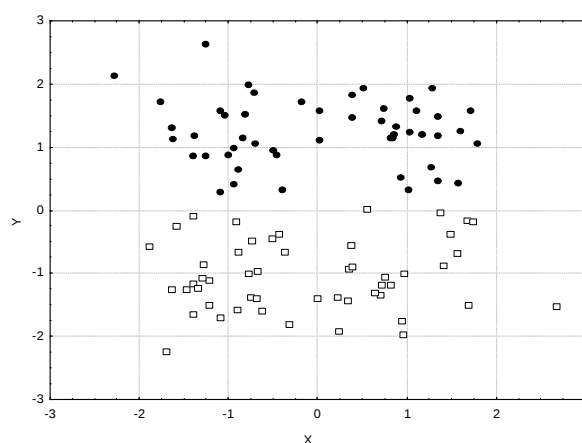
jó, de meglete nem garantálja, hogy „természetes” csoportokat találtunk.

A repetitivitást vizsgáló módszerek elterjedésének egyik gátja a párhuzamos minták hiánya volt (Legendre & Legendre 1998), de a nagy adatbázisok esetén ez már nem jelent problémát. A túlmintavételezés (oversampling) kiküszöbölésére nem az adatbázisban szereplő összes felvételt elemezzük, hanem rétegzett mintát veszünk abból (Knollová et al. 2005). Általában semmi akadály, hogy két párhuzamos, nem átfedő mintát vegyünk.

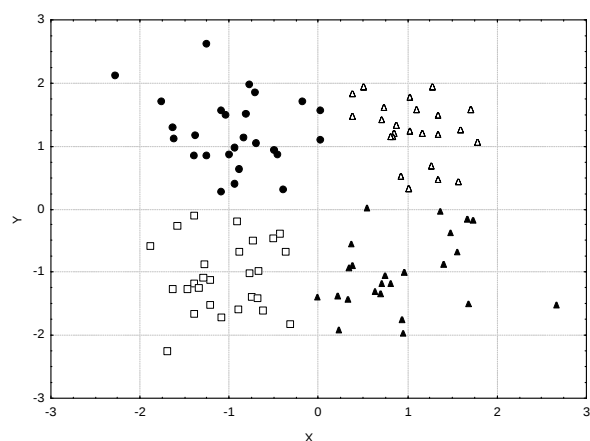
A legtöbb módszer a következő sémát követi (Breckenridge 1989, Lange et al. 2004): az egyik mintát klasszifikáljuk, majd valamilyen osztályozási szabály alapján (lásd később) a második minta elemeit is besoroljuk ezekbe a csoportokba. Ezután a második mintát is klasszifikáljuk ugyanazzal a módszerrel, mint az első mintát. Minél nagyobb a második minta elemeinek kétféle csoportosítása közötti hasonlóság, annál jobb a klasszifikáció.



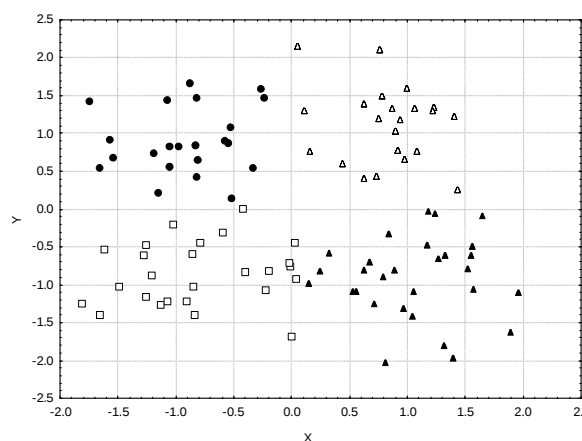
a) első minta, két csoport



b) második minta, két csoport



c) első minta, négy csoport



d) második minta, négy csoport

22. ábra: Ugyanabból az alapsokaságból vett két független minta csoportosítása k-közép módszerrel kettő és négy csoportba. A tengelyek a két vizsgált változó (x és y), a különböző csoportokat eltérő alakú pontokkal jelöltem. Ha két csoportba osztjuk a pontokat, a két párhuzamos minta nagyon különböző csoportosításhoz vezet. Négy csoport esetén viszont jó egyezés figyelhető meg a párhuzamos mintákban kapott csoportok között.

A módszer gyenge pontja, hogy megfelelő klasszifikációs szabályt kell találni. McIntyre & Blashfield (1980) eredeti módszere szerint a felvételt abba a csoportba kell besorolni,

amelynek centroidjához legközelebb van (legközelebbi centroid szabály). Breckenridge (1989) szerint jobb eredményt kapunk, ha abba a csoportba soroljuk, amelybe a hozzá leginkább hasonló felvétel tartozik (legközelebbi szomszéd szabály). Vegyük észre, hogy az osztályozási szabályok rokonságban vannak a hierarchikus klasszifikáció során alkalmazott összevonási szabályokkal: a legközelebbi centroid szabály a centroid (UPGMC) módszerrel, a legközelebbi szomszéd módszer az egyszerű lánc módszerrel rokon. Ezt kihasználva más összevonási algoritmusok – például teljes lánc vagy átlagos lánc (UPGMA) – alapján is definiálhatnánk klasszifikációs szabályokat. A felvételek besorolása *a priori* csoportokba a cönológiai szakértői rendszerekben is jelentkező feladat (pl. Kočí et al. 2003, De Cáceres et al. 2010), ezért kifejezetten a vegetációs adatokra kidolgozott klasszifikációs szabályok is léteznek (pl. Tichý 2005, Tongeren et al. 2008). Nagyon sokféle szabály közül választhatunk tehát, és sajnos Lange és munkatársai (2004) eredményei szerint a választásunk nagymértékben befolyásolhatja az eredményt. Ezért egy olyan új módszert dolgoztam ki (Botta-Dukát 2008a), amely a csoportok jelentésén alapul (részletesen lásd a 6.3 fejezetben), ezért nem alkalmaz klasszifikációs szabályt. Míg a korábbi módszerek csak a teljes partíció értékelésére használhatók, ez az új módszer az egyes csoportokat külön-külön értékeli.

6.1.6. A csoportok értelmezhetősége

Az adatok klasszifikációja nem cél, hanem eszköz az adataink struktúrájának jobb megértéséhez. Ennek érdekében a kapott csoportokat értelmezzük, elsősorban a klasszifikációban is szereplő változók alapján, de ha lehetséges, akkor további, a csoportosítással feltételezhetően kapcsolatban álló változók (esetünkben általában a környezeti háttértényezők) alapján is. Ebben a részben a csoportosításban is felhasznált változók alapján történő értelmezhetőséggel foglalkozom, a külső változók alkalmazásának lehetőségeit a következő fejezet tárgyalja.

Azt, hogy hogyan értelmezzük a csoportokat, elsősorban a szakma szabályai, hagyományai szabják meg. A cönológiában gyökerező, durva léptékű vegetációtanulmányokban az értékelés (és gyakran maga az elemzés is), elsősorban a fajok konstanciái, illetve az ebből számolt fidelitás alapján történik, a fajok tömegessége emellett másodlagos, legfeljebb a magas borítást elérő, domináns fajok jelennek meg az értékelésben. Ennek megfelelően, a csoportok értelmezhetőségét mérő módszerek többsége is általában a csoporton belüli előfordulási gyakoriságokon, vagyis végső soron bináris adatokon alapul. A fidelitás koncepció, illetve ehhez kapcsolódóan a fajok válogatóképessége azonban kiterjeszthető a borítás adatokra is (lásd az 5.3 fejezetben), így a bemutatott indexek egy része a borítás adatokra is kiszámolható.

Vezessük be a következő jelöléseket:

F_{ij} = az i faj előfordulásainak száma a j csoportban

$$F_{i.} = \sum_j F_{ij}$$

$$F_{.j} = \sum_i F_{ij}$$

$$F_{..} = \sum_i \sum_j F_{ij}$$

n_j = a j csoportba tartozó felvételek száma

N = a vizsgálatban szereplő felvételek száma

K = a csoportok száma

S = a fajok száma

5.1 ISAMIC (Aho et al. 2008)

A csoport értelmezését megkönnyíti, ha a fajok vagy nagy, vagy nagyon alacsony konstanciával fordulnak elő benne, vagyis alacsony a közepes konstanciájú fajok aránya. Ez az index ezt az elvárást számszerűsíti:

$$I = \frac{\frac{2}{K} \sum_i \sum_j \left| \frac{F_{ij}}{n_j} - 0,5 \right|}{S}$$

5.2. Morisita index (Aho et al. 2008)

A csoportok értelmezése annál könnyebb, minél inkább eltérőek a fajok konstanciái a csoportokban. Aho és munkatársai (2008) javaslata szerint két felvétel konstanciáinak összehasonlítására a Morisita index használható¹⁰:

$$M_{jk} = 1 - \frac{2 \sum_i F_{ij} F_{ik}}{\frac{F_{.k}}{F_{.j}} \sum_i F_{ij}^2 + \frac{F_{.j}}{F_{.k}} \sum_i F_{ik}^2}$$

A partíció jósága a páronkénti különbségek átlagával mérhető.

5.3. Relatív divergencia (Feoli & Lausi 1980)

Feoli és Lausi (1980) szerint a bináris faj x felvétel mátrixot átrendezhetjük egy olyan kontingencia-táblázattá, amelynek sorai a fajok, oszlopai a csoportok és celláiban az adott faj, adott csoportban való előfordulásainak száma szerepel. Annál jobb a csoportosítás, minél inkább különböznek egymástól az oszlopokban szereplő gyakoriságeloszlások. A gyakoriságeloszlások különbözőségének mérésére három információelméleti mérőszámot is javasoltak, de a köztük levő függvénykapcsolat miatt csak a relatív divergenciát mutatom be:

$$D = \frac{\sum_i \sum_j F_{ij} \ln \frac{F_{..} F_{ij}}{F_{i.} F_{.j}}}{2 \sum_i \sum_j F_{ij} \ln \frac{F_{ij}}{F_{i.} F_{.j}}}$$

5.4. A fajok átlagos válogatóképessége

A csoportosítás annál jobb, minél több faj használható a csoportok jó elkülönítésére. Egyik lehetőség ennek számszerűsítésére az 5.3 fejezetben bemutatott válogatóképesség értékek átlagának használata.

A mocsárrétek rendszerezéséről szóló cikkünkben (Botta-Dukát et al. 2005) az elméleti várhatóértékkel és szórással korrigált G-statisztika alkalmazását javasoltuk, a

¹⁰ Aho és munkatársai (2008) az átlagos hasonlóság alkalmazását javasolták, de komplementerét, az átlagos különbözőséget szemléletesebbnek tartom.

csoportosítás élessége (crispness¹¹) néven. Annak érdekében, hogy a kiszámított értékek eloszlása minél jobban illeszkedjen az elméleti értékhez, az számítást csak a felvételek legalább 10%-ban előforduló fajokra végeztük el.

Az elméleti értékek helyett a standardizálásra a randomizációból számolt átlag és szórás is használható, illetve a fajok válogatóképessége a korrigált (adjusted) R^2 értékkel is mérhető.

5.5. Átlagos fidelitás, illetve a fidelitás értékek átlagos elsőfajú hibája

Dufrene & Legendre (1997) eredetileg a szignifikáns fidelitás (esetükben IndVal) értékek összegét javasolta a csoportosítás jóságának mérőszámaként. Podani és Csányi (2010) módosító javaslata szerint, érdemes az összegzésnél az összes faj fidelitását figyelembe venni, de az eredményből vonjuk le a random klasszifikációkra kapott értékek átlagát, és osszuk el azok szórásával. Aho és munkatársai (Aho et al. 2008) viszont az egyes fajokra számolt elsőfajú hibák átlagolását javasolták. Mindhárom eljárásban közös, hogy a fidelitás átlagos erősségét mérik.

5.6. A szignifikáns fidelitások száma

Egymástól függetlenül Aho és munkatársai (Aho et al. 2008) és kutatócsoportunk (Tichý et al. 2010) is javasolta ennek a mérőszámnak az alkalmazását. Aho és munkatársai javaslatában az IndVal értékek (Dufrene & Legendre 1997) alapján, randomizációval állapítjuk meg, hogy mely fidelitás értékek szignifikánsak egy előre eldöntött szinten (cikkükben ez 5%). Ennek a módszernek az előnye, hogy borítás értékeket is figyelembe vehetünk; hátrány a randomizáció miatti nagyobb számításigény.

Az általunk javasolt módszer (OptimClass1; Tichý et al. 2010) az elsőfajú hibát Fisher-féle egzakt teszttel állapítja meg, így csak bináris adatokra használható, viszont nincs szükség hosszadalmas randomizációra. Ha az optimális csoportszám megállapítása a cél, az eredmény függ az alkalmazott szignifikancia-szinttől (alacsonyabb szignifikancia-szint esetén magasabb csoportszám lesz az optimális), ezért érdemes többféle szignifikancia szintet kipróbálni.

Ugyanebben a cikkben (Tichý et al. 2010) javasoltunk egy második mérőszámot is (OptimClass2): azoknak a csoportoknak az arányát, amelyeknek legalább k szignifikáns karakterfaja van, ahol a k egy a felhasználó által beállítandó paraméter. Az OptimClass2 háttérében az a gondolat van, hogy ha már elég sok karakterfaja van egy csoportnak, újabbnak már nem növelik a csoportosítás jóságát. Vagyis fontosabb, hogy minden csoportnak legyen kellő számú karakterfaja, mint, hogy összesen sok karakterfaj legyen (ami úgy is megvalósulhat, hogy csak egy csoportnak van sok karakterfaja).

5.7. A fajok átlagos nicheszélessége (Casado et al. 1997)

A klasszifikáció során kapott csoportokat tekinthetjük különböző termőhelyeknek, és ez alapján kiszámolhatjuk a fajok nicheszélességét. Ha a csoportok jól elkülönülő

¹¹ Utólag úgy gondolom, szerencsésebb lett volna a sharpness kifejezést használnunk.

élőhelyeknek feleltethetők meg, a fajok nicheszélessége kicsi lesz. Casado és munkatársai (1997) a faj nicheszélességét a Shannon-entrópiával mérték:

$$W_i = - \sum_j \frac{F_{ij}}{n_j} \ln \frac{F_{ij}}{n_j}$$

Mivel az entrópia elméleti maximuma függ a csoportok számától, javasolták, hogy a fajok entrópiáinak az előfordulások számával súlyozott átlagát osszuk el az elméleti maximummal:

$$W = \frac{\sum F_i \cdot W_i}{F \cdot \ln K}$$

6.1.7. A csoportosítás prediktív ereje

A vegetációs adatok elemzésekor kapott csoportoktól elvárjuk, hogy ha ismerjük, hogy a felvétel melyik csoportba tartozik, akkor ezzel ne csak arról kapjunk információt, hogy milyen növényfajok előfordulása várható ott, de közvetve az élőhelyi háttértényezőkről is. Vagyis a jó csoportosítás alapján olyan változók tekintetében is tehetünk predikciókat, amelyeket nem használtunk fel a csoportok kialakításakor. Mivel ebben az esetben a vizsgált változó(k) és a csoportosítás elvben függetlenek is lehetnek, használhatjuk a klasszikus statisztikai módszereket is, beleértve a függetlenségi hipotézis tesztelését. Általában nem egy-egy környezeti változót vizsgálunk, hanem egyszerre többet, ezért a leggyakrabban alkalmazott módszer a diszkriminancia-analízis (pl. Gerdol et al. 1985, Hakes 1994).

Ebbe a csoportba sorolható Picard és Franc (2003) munkája is, akik a fajok csoportosításának jóságát azzal mérik, hogy az erdődinamikai modellekben a fajok helyett fajcsoportokat használva mennyire változnak meg az eredmények.

6.1.8. Összehasonlítás random adatokkal

Az eddig tárgyalt szempontok azt vizsgálták, hogy melyik legjobb csoportosítás az adott adatsor esetén. A klasszifikációs eljárások azonban random adatok esetén is létrehozhatnak csoportokat (Podani 1997). Ezért feltehető az a kérdés is, hogy vannak-e egyáltalán csoportok az adatainkban (Tan et al. 2005), és ha igen, mely csoportok tekinthetők a véletlen csoportoknál jobbaknak (Strauss 1982)?

Az első kérdésre az adatok klasszifikációja nélkül is válaszolhatunk. Próbá statisztikaként elsősorban a legközelebbi szomszédok távolságán alapuló jellemzők, például a Hopkins statisztika javasolható (Bock 1996, Tan et al. 2005). A kritikus pont a megfelelő null-modell megtalálása. A klasszifikációs szakirodalomban (pl. Bock 1996) általában javasolt modellek – egymástól független, egyenletes vagy unimodális eloszlású változók – nem alkalmasak a vegetációs adatokra, ahol egy reális null-modellben a változók összege egy felvételen belül (bináris adatoknál a fajszám, borítás adatoknál az összborítás) nem lehet tetszőleges érték. Bináris adatoknál Strauss (1982) a sor és oszlopösszegeket is megőrző randomizálást javasol, kivéve, ha feltételezhető, hogy a felvételek fajgazdagságának különbségei csak a véletlenből adódnak és nem az adott lokalitás jellemzőiből. Az utóbbi esetben elegendő a fajok előfordulásainak számát megőrizni a randomizáció során. Borítás

adatok esetében közel sem ennyire egyértelmű, hogy mi a megfelelő random modell.

Összehasonlíthatjuk a teljes klasszifikáció jóságát a random klasszifikációk jóságával. Erre leginkább a csoportok tömörségét, összekötöttségét és elválását vizsgáló indexek használhatók. Nem csak a bemutatott több szempontot kombináló indexek, hanem csak egy szempontot figyelembe vevők is. Ez azonban inkább csak elméleti lehetőség, gyakorlati alkalmazására nem láttam még példát. Nem alkalmasak azonban a klasszifikáció jóságának vizsgálatára az *a priori* csoportok elkülönülésének elemzésére szolgáló módszerek, például a diszkriminancia-analízis vagy az ANOSIM (Clarke 1993). Utóbbi hibás alkalmazására Aho és munkatársainak (2008) tanulmánya egy friss példa.

Nagyobb gyakorlati jelentősége van a dendrogramokban az egyes szétválásokat tesztelő eljárásoknak (Strauss 1982, Hunter & McCoy 2004). A korábbi módszereket továbbfejlesztve Hunter és McCoy (2004) a következő eljárást javasolta:

1. számos random adatsort állítunk elő
2. a random adatokat ugyanazzal a hierarchikus módszerrel csoportosítjuk, mint a terepi adatokat
3. a terepi adatok dendrogramjában az adott összevonáshoz tartozó fúziós szintet a megfelelő fúziós szint null-eloszlásához viszonyítjuk
4. ha a terepi fúziós szint szignifikánsan magasabb a random esetben vártnál, azaz az elválás szignifikáns, a továbbiakban a két részmátrixot külön randomizáljuk.

6.1.9. A klasszifikációk értékelésére szolgáló módszerek csoportosításának egyéb szempontjai

Az eddigiekben a klasszifikációk értékelésére használt módszereket az általuk figyelembevett szempont (kritérium) szerint csoportosítva mutattam be. Érdekes azonban röviden kitérni az egyéb csoportosítási lehetőségekre is.

A leggyakrabban említett csoportosítás (pl. Feoli et al. 1981, Gordon 1999, Aho et al. 2008) a figyelembevett adatok szempontjából osztja két részre a módszereket: a belső (internal) kritériumok olyan változókat vesznek figyelembe, amelyeket a csoportok kialakításánál is felhasználtunk, míg a külső (external) kritériumok olyan információt is felhasználnak, amit a csoportosításnál nem vettünk figyelembe. A bemutatott módszerek többsége az első kategóriába tartozik, kivételt a csoportosítás prediktív erejét mérő módszerek jelentenek. Szintén a külső kritériumokhoz sorolható, amikor a szimulált adatok elemzésekor a kapott csoportok az ismert helyes csoportosításhoz hasonlítjuk (lásd pl. a zajsűrűségről szóló fejezetben).

Az optimális csoportszám megállapítására használható módszereket Gordon (1999) két csoportba sorolja: a *globális* módszerek a teljes klasszifikáció jóságát vizsgálják, míg a *lokális* módszerek azt tesztelik, hogy érdemes-e egy adott csoportot két részre osztani. A bemutatott módszerek döntő többsége globális kritérium. A lokális kritériumra példa Hunter és McCoy (2004) által javasolt eljárás.

16. táblázat: Az egyszerű jósági indexek néhány tulajdonsága

	Geometriai / Nem-geometriai	Mire értelmezhető? O = egy objektumra Cs = egy csoportra K = a teljes klasszifikációra	Maximum vagy minimum jelzi az optimális csoportosítást?	Az elméletben lehetséges legjobb érték
Pont-biszeriális korreláció	geometriai	K	maximum	1
Tau	geometriai	K	maximum	1
Baker & Hubert	geometriai	K	maximum	1
G ₊	geometriai	K	minimum	0
Hubert & Levine	geometriai	K	minimum	0
McClaine & Rao	geometriai	K	minimum	0
PARTANA	geometriai	K	maximum	∞
Dunn _{min-max}	geometriai	K	maximum	∞
Dunn _{avg-avg}	geometriai	K	maximum	∞
módosított Davis & Bouldin	geometriai	Cs, K	minimum	0
Popma (súlyozatlan)	geometriai	Cs, K	minimum	0
Popma (súlyozott)	geometriai	Cs, K	minimum	0
Átlagos sziluett	geometriai	O, Cs, K	maximum	1
Helyesen klasszifikált felvételek aránya	geometriai	O, Cs, K	maximum	1
Feoli	geometriai	K	maximum	1
ISAMIC	nem-geometriai	Cs, K	maximum	1
Morisita	nem-geometriai	Cs, K	maximum	1
Relatív divergencia	nem-geometriai	K	maximum	∞
Crispness	nem-geometriai	K	maximum	-
átlagos R_{adj}^2	nem-geometriai	K	maximum	1
fidelitás értékek átlaga	nem-geometriai	Cs, K	maximum	-
fidelitás értékek átlagos elsőfajú hibája	nem-geometriai	Cs, K	minimum	0
OptimClass1	nem-geometriai	Cs, K	maximum	-
átlagos nicheszélesség	nem-geometriai	K	minimum	0

Praktikus szempontból érdemes megkülönböztetni a klasszifikáció valamilyen tulajdonságát számszerűsítő egyszerű indexeket, és a több párhuzamos csoportosítás összehasonlításával az eredmények stabilitását vagy repetitívitasát vizsgáló módszereket. Kisebb számításigényük miatt a nagy adatbázisok elemzésekor inkább az egyszerű indexek alkalmazása célszerű. Ezek egy része csak a teljes partícióra értelmezhető, mások azonban kiszámolhatók az egyes csoportokra is, és a partícióra jellemző érték, a csoportok értékeinek (súlyozott) átlaga. Utóbbiak alapján nem csak a teljes osztályozás, hanem az egyes csoportok tömörsége és elkülönültsége is értékelhető. A sziluett értéke pedig az egyes felvételekre is kiszámolható, így megkereshetők a csoportokat összekötő, vagy éppen félreklasszifikált felvételek (16. táblázat).

Aho és munkatársai (2008) nyomán megkülönböztethetünk geometriai és nem-geometriai értékelő módszereket. Az előbbiek a csoportosítás során használt távolság (esetleg hasonlóság) mátrixot felhasználják az értékelés során is. Ezért ezek a módszerek nem használhatók a különböző távolságfüggvények közti választáshoz. Ha az a kérdésünk, hogy melyik távolságfüggvényt használva kapunk jobb csoportosítást (pl. Tichý et al. 2010) akkor nem geometriai értékelő módszert kell választanunk.

6.2. Az értékelő módszerek alkalmasságának összehasonlítása

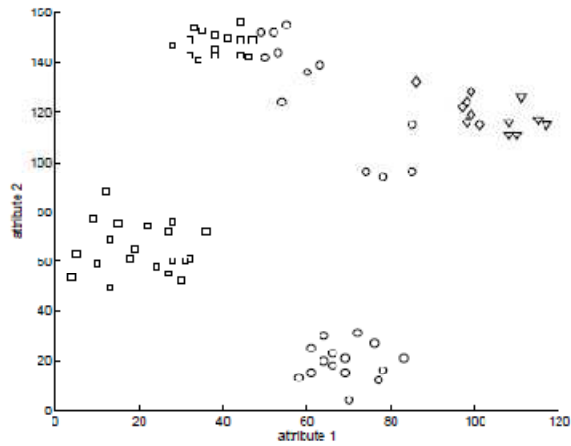
Mint már említettem, a nagy adatbázisok elemzésekor, kisebb számításigényük miatt, érdemes az egyszerű indexeket választani. Az alábbiakban 23 ilyen index tulajdonságait értékelem szimulált adatok alapján.

6.2.1. Anyag és módszer

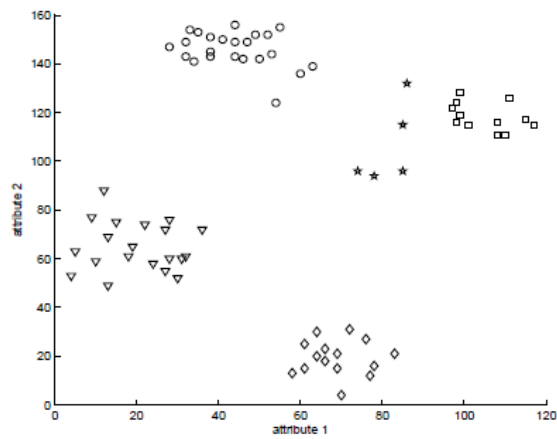
A klasszifikációkat értékelő módszerek tesztelésében úttörő szerepe volt Milligan (1981, Milligan & Cooper 1985) munkáinak, ezért érdemes az általa használt módszerekből kiindulni. Milligan és Cooper (1985) – és nyomukban más szerzők is (pl. Pal & Biswas 1997, Halkidi et al. 2002) – olyan szimulált adatokat vizsgáltak, ahol a csoportok száma előre ismert, és akkor tekintették jónak az értékelő módszert, ha a beállított csoportszámot jelezte optimálisnak. Ennek a látszólag kézenfekvő megoldásnak több hátránya is van. A legfontosabb, hogy ha az elméleti és az optimálisnak jelzett csoportszám megegyezik, az még nem bizonyítja, hogy az optimálisnak jelzett csoportosítás megegyezik az elméletivel (Dimitriadou et al. 2002), sőt az is lehetséges, hogy egy kicsit eltérő csoportszám esetén az elméletihez jobban hasonlító csoportosítást kapunk (Vendramin et al. 2010) (23. ábra). Az elméletivel megegyező optimális csoportszám aránya azért sem jó, mert nem jelzi, hogy mekkora a torzítás, amikor az értékelő módszer hibás eredményt ad. Az elméleti és az optimálisnak jelzett csoportszám különbsége – ahogy azt Milligan és Cooper (1985) javasolta – nem alkalmas erre, mert ugyanakkora eltérés, nagyon eltérő torzítást jelezhet (Vendramin et al. 2010) (23. ábra).

Vendramin és munkatársai (2010) szerint jobb kiindulópont a Milligan (1981) által használt módszer, amelynek alapgondolata, hogy az ismert szerkezetű mesterséges adatoknál a csoportosítások jóságát mérhetjük a hasonlóságukkal a valós csoportosításhoz.

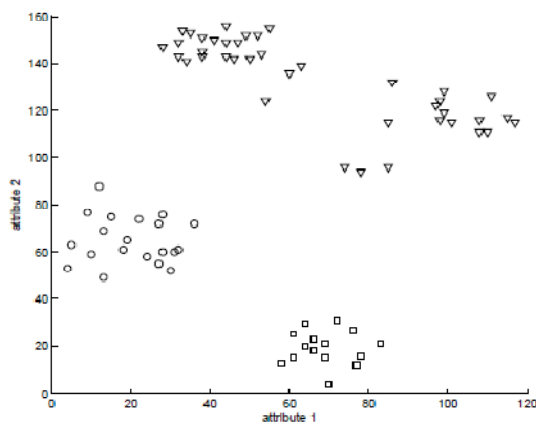
dc_16_10



a)



b)



c)

23. ábra: Egy kétváltozós mesterséges adatsor (Ruspini 1970) különböző klasszifikációinak eredményei. Az adatsor négy csoportot tartalmaz és az (a) esetben is négy csoportot kaptunk, mégis láthatóan természetesebb a 3 csoportot tartalmazó (b) vagy az 5 csoportos (c) klasszifikáció. Legjobbnak az utóbbi tekinthető, mert a (b) esetben a két valós csoport összeolvasztása információvesztést jelent (Vendramin et al. 2010 nyomán).

Az a jó belső adatokon alapuló index, amely minél inkább reprodukálni tudja ennek az összehasonlításnak az eredményét, vagyis amely erősen korrelál a hasonlóság értékekkel. A kapott csoportosítások és valós csoportosítás összehasonlítására többféle index létezik a szakirodalomban (Podani 1986), amelyek közül az egyik leggyakrabban használtat, a módosított Rand-indexet használtam (3. melléklet).

Milligan (1981) eredeti módszere szerint az összehasonlításhoz az ismert a valós csoportszámmal megegyező csoportszámú partíciókat kell használni, de Vendramin és munkatársai (2010) rámutattak, hogy ezzel túlságosan leszűkítjük a vizsgált partíciók körét, ezért különböző csoportszámok alkalmazását javasolták.

A mesterséges adatokat a 4.1.2 fejezetben leírt módon állítottam elő, így a valódi csoportok száma 5 volt. Minden adatsort négyféle klasszifikációs módszerrel – átlagos lánc (UPGMA), béta flexibilis ($\beta=-0.25$), teljes lánc (Podani 1997) és *pam* (partitioning around medoid; Kaufman & Rousseeuw 1990) – elemeztem, valamennyi esetben a Jaccard-különbsőséget használva távolságként. A nem-hierarchikus klasszifikációban a csoportszámot 2 és 10 között változtattam, a hierarchikus osztályozások dendrogramjait pedig úgy vágtam el, hogy 2-10 csoportot kapjak, így végül egy-egy adatsornak 36 különböző osztályozása állt rendelkezésemre. Ezeket hasonlítottam össze a valós csoportosítással, illetve ezekre számoltam ki a klasszifikációt értékelő indexeket. Ezután lineáris korrelációt számoltam a klasszifikációt értékelő index értéke, és a korrigált Rand-index között. A jobb összehasonlíthatóság kedvéért, azoknál az indexeknél, amelyeknek a kisebb értéke jelzi a jobb klasszifikációt, a kapott korrelációt mínusz eggyel megszoroztam. Vendramin és munkatársai (2010) javaslatát követve 50-szer megismételtem a szimulációt, így minden indexet 50 korrelációs érték eloszlása jellemez.

Bár a vizsgált indexek nem annak eldöntésére szolgálnak, hogy vannak-e egyáltalán csoportok az adatokon belül, mégis érdemes megvizsgálni jelzik-e és ha igen, hogyan a csoportstruktúra teljes hiányát. Ehhez a 4.1.2 fejezetben bemutatott sziklagyepi adatsorból állítottam elő 50 random adatsort. A randomizálás során Miklós és Podani (2004) módszerét használva összekevertem a bináris adatmátrixban az előfordulás és hiány értékeket, változatlanul hagyva a felvételek fajszerát és a fajok gyakoriságát. Ezután a felvételek Jaccard különbsősége alapján béta-flexibilis algoritmussal ($\beta=-0.25$) hierarchikus osztályozásokat készítettem, amelyeket 2-20 csoportból álló partíciókká alakítottam. Vizsgáltam, hogy hogyan alakul az indexek értéke a csoportszám függvényében, illetve, hogy a vizsgált tartományon belül melyik csoportszámot jelzik a legjobbnak.

A számításokat az R 2.9.2 programmal (R Development Core Team 2010) végeztem, felhasználva a *vegan* (Oksanen et al. (2010), *cluster* (Maechler et al. 2005) és *clues* (Chang et al. 2009) csomagokat.

6.2.2. Eredmények

Az azonos és eltérő csoportméretek esetén kapott eredmények között nem volt egyik index esetén sem lényeges eltérés. A hét index közül, amelyek azt vizsgálják, hogy mennyire torzul a távolságmátrixban rejlő információ, csak három adott jó eredményt: a Baker &

Hubert, a Hubert & Levine és a McClaine & Rao (24. ábra). A többi index esetén a korreláció a korrigált Rand-indexszel meglepő módon gyakran negatív. Csoportstruktúra hiányában Baker & Hubert index értéke a csoportszámmal nő, míg a Hubert & Levine és a McClaine & Rao indexé csökken (27. ábra). A Hubert & Levine index esetén az optimálisnak talált csoportszám ilyen esetben változó, ezért ez az index nem jelzi a csoportstruktúra hiányát. Ezzel szemben a Baker & Hubert és a McClaine & Rao indexeknél csoportstruktúra hiányában szinte mindig a megvizsgált legmagasabb csoportszám bizonyult a legjobbnak.

A csoportok tömörségét és elválását vizsgáló indexek közül a Popma-index súlyozott verziója korrelált legerősebben a korrigált Rand-indexszel, de nem sokkal maradt el tőle súlyozatlan változat sem (25. ábra). Kisebb, de még mindig döntően pozitív korrelációkat mutatott a Dunn_{min-max} és a helyesen klasszifikált felvételek aránya. A csoportstruktúra hiányát szinte csak a súlyozott Popma-index jelezte egyértelműen, bár az átlagos sziluett nagyon alacsony értékei alapján is gyanakodhatunk ugyanerre (27. ábra).

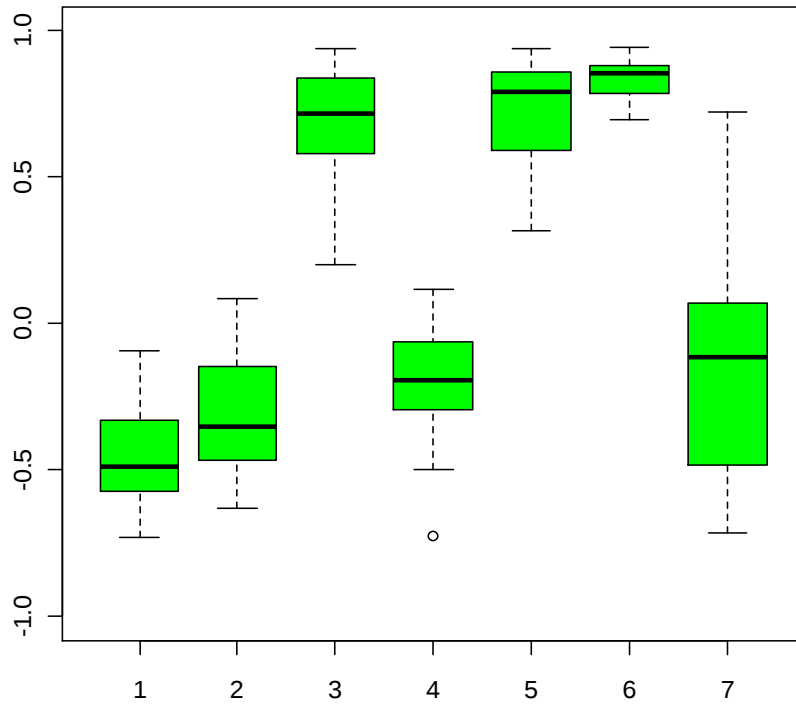
A csoportok értelmezhetőségét számszerűsítő indexek közül a Rand-indexszel való korreláció alapján jónak bizonyult a relatív divergencia, az átlagos korrigált R^2 és az átlagos nicheszélesség. A karakterfajok számánál meglepő módon csak akkor kaptunk elfogadható korrelációt, ha a szignifikanciaszintet viszonylag magasra (10^{-3}) választottuk (26. ábra). A csoportstruktúra hiányát jól jelzi a relatív divergencia és az átlagos nicheszélesség trendje (mindkét esetben egyértelműen a legnagyobb megvizsgált csoportszám bizonyult a legjobbnak), valamint a karakterfajok száma (OptimClass1), különösen, ha alacsonyabb szignifikancia szinteket választunk (29. ábra).

6.2.3. Diszkusszió

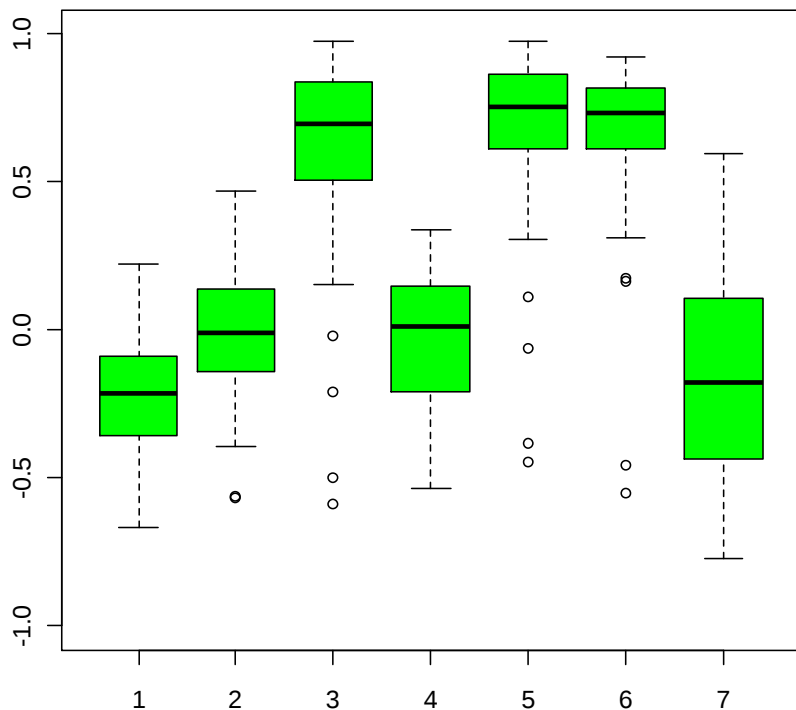
Habár számos vizsgált index kifejezetten ökológiai indíttatású, az ökológiai szakirodalomban nem találtam példát a klasszifikációt értékelő indexek tesztelésére. Aho és munkatársai (2008) vizsgálták a különböző indexek közti korrelációt, ebből azonban nem lehet következtetéseket levonni arra, hogy mely indexek a jól használhatók.

A más tudományterületeken – elsősorban a pszichológiában – elvégzett korábbi összehasonlító vizsgálatok (pl. Milligan & Cooper 1985, Vendramin et al. 2010) mesterséges adatai általában kevés dimenziós, euklideszi térben jól elkülönülő csoportokat alkotó pontfelhők. A vegetációs adatok viszont sokdimenziósak, gyakran bináris skálájúak, és még ha arány skálájúak is, nem euklideszi térbeli pontok. Ezért az ott kapott eredmények nem vehetők át előzetes ellenőrzés nélkül. Ahogy Vendramin és munkatársai (2010) cikkük zárszavában nagyon korrekten leszögezik: *„As a word of caution, it is worth remarking that the above results and conclusions hold for a particular collection of data sets. Since such a collection is reasonably representative of a particular class, namely, data with volumetric clusters following normal distributions, it seems legitimate to believe that similar results are likely to be observed for other data sets of this class. However, nothing can be presumed about data sets that do not fall within this class, at least not before new experiments involving such data are performed.”*

dc_16_10



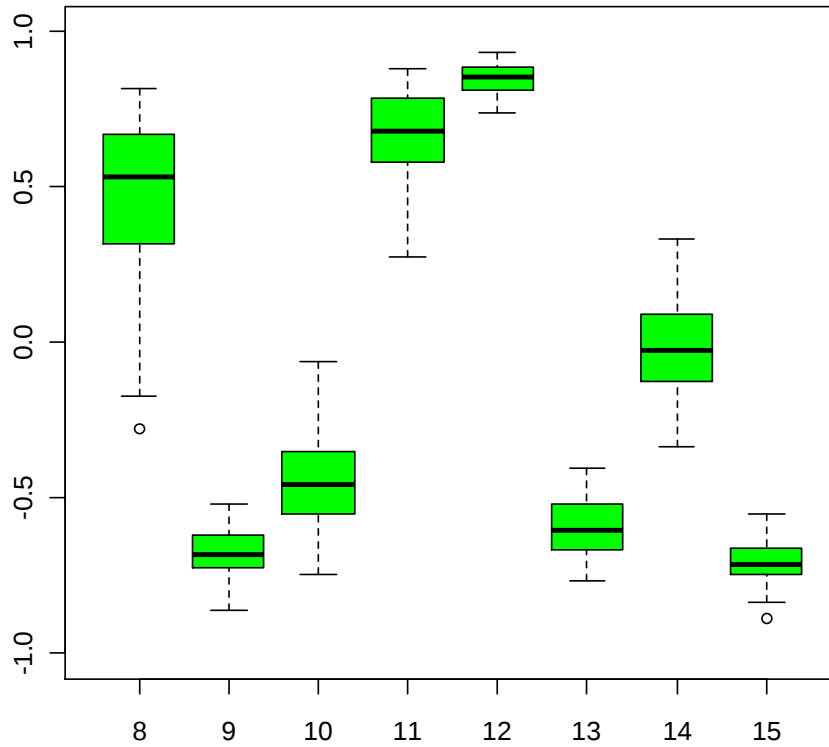
a)



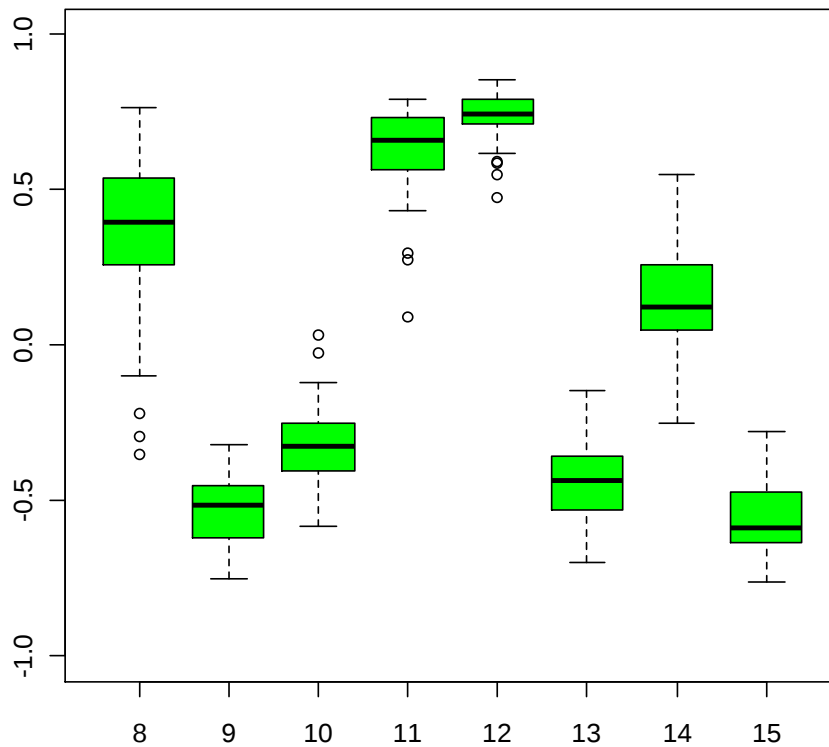
b)

24. ábra: A távolságmátrix és a csoportosítás közötti kapcsolat szorosságát mérő indexek tesztelésének eredménye. a) egyenlő csoportméret, b) eltérő csoportméret. Jelmagyarázat: 1. Pont-biszeriális korreláció, 2. Tau, 3. Baker & Hubert, 4. G_+ , 5. Hubert & Levine, 6. McClaine & Rao, 7. PARTANA

dc_16_10



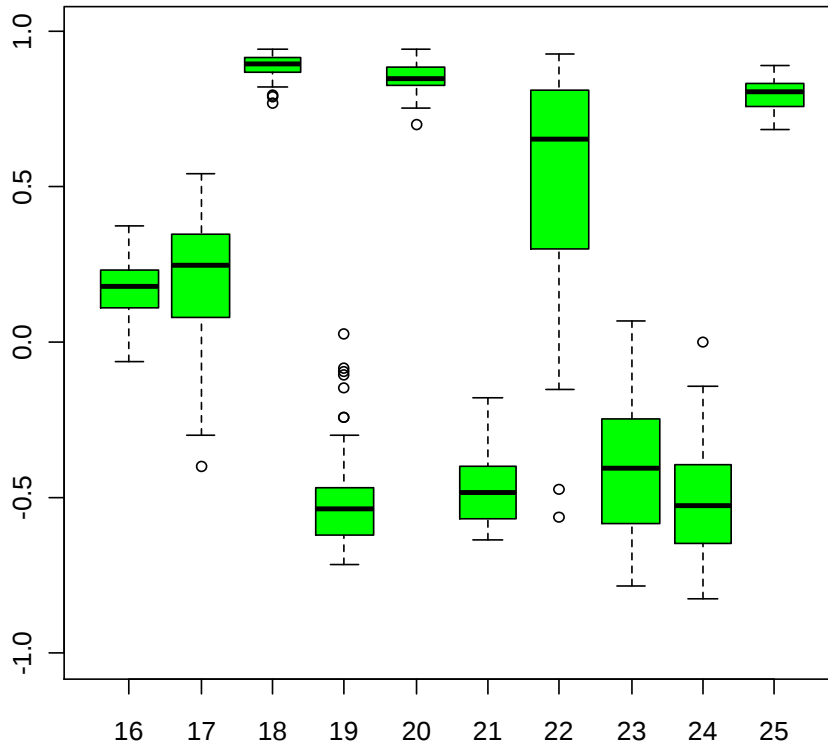
a)



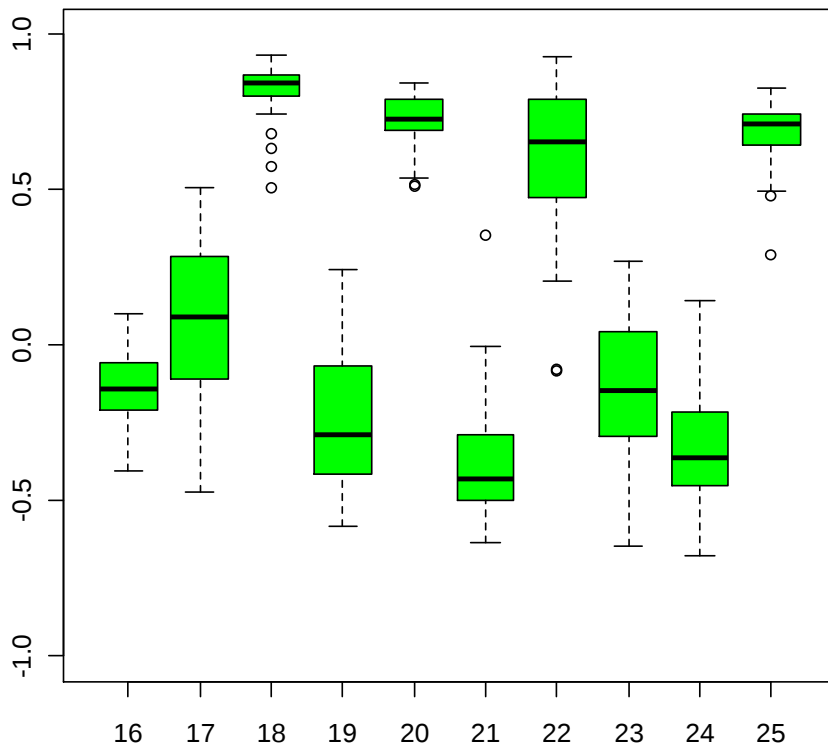
b)

25. ábra: A csoportok tömörségét és elválását mérő indexek tesztelésének eredménye. a) egyenlő csoportméret, b) eltérő csoportméret. Jelmagyarázat: 8. $Dunn_{min-max}$, 9. $Dunn_{avg-avg}$, 10. módosított Davis & Bouldin, 11. Popma (súlyozatlan), 12. Popma (súlyozott), 13. átlagos sziluett, 14. helyesen klasszifikált felvételek aránya, 15. Feoli

dc_16_10



a)

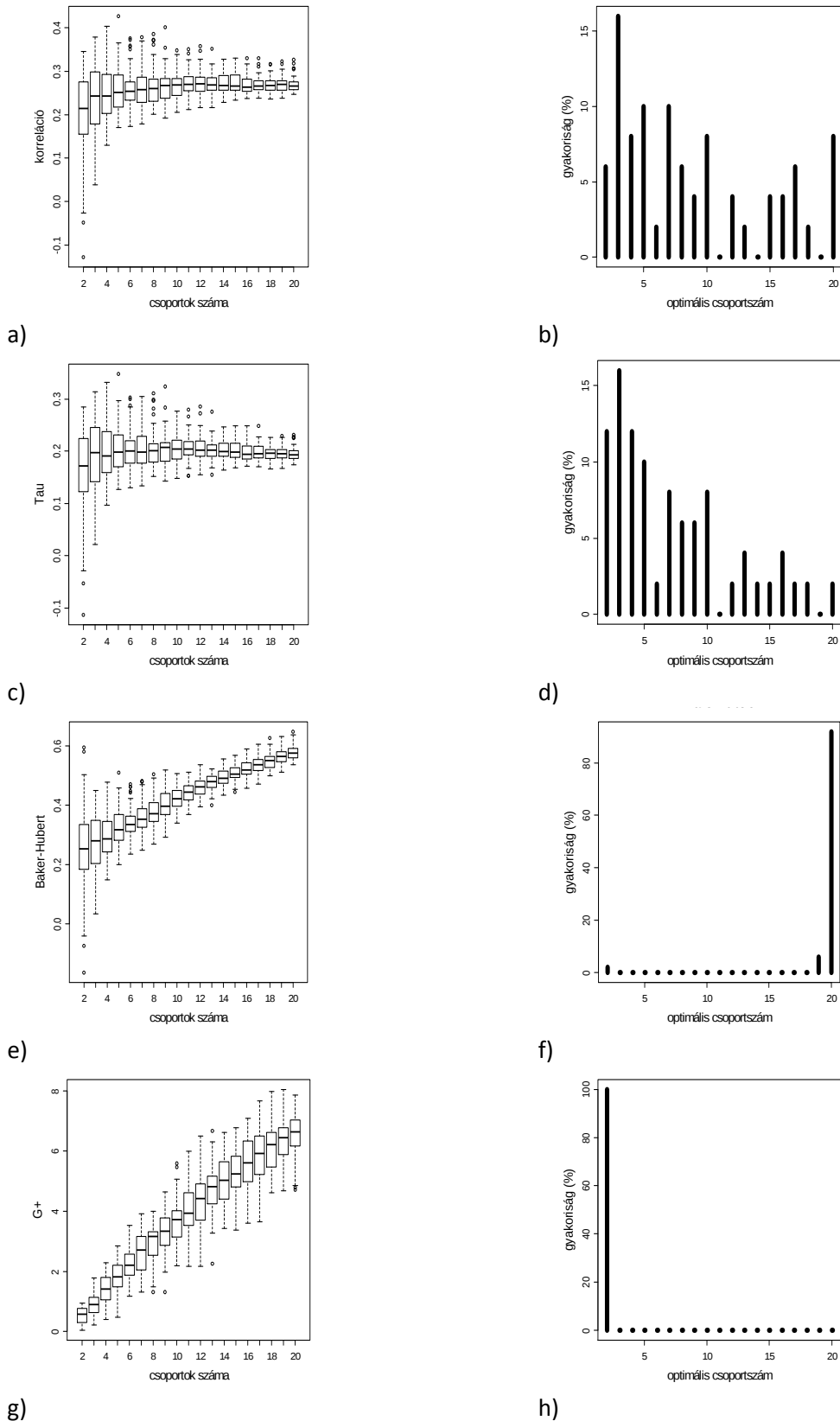


26. ábra: A csoportok értelmezhetőségét mérő indexek tesztelésének eredménye. a) egyenlő csoportméret, b) eltérő csoportméret. Jelmagyarázat: 16. ISAMIC, 17. Morisita, 18. relatív divergencia, 19. Crispness, 20. átlagos R_{adj}^2 , 21. fidelitás értékek átlagos elsőfajú hibája, 22. OptimClass1 ($p=10^{-3}$), 23. OptimClass1 ($p=10^{-6}$), 24. OptimClass1 ($p=10^{-9}$), 25. átlagos nicheszélesség

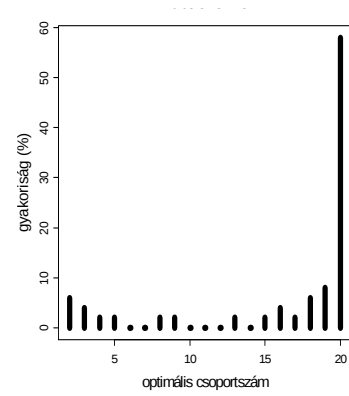
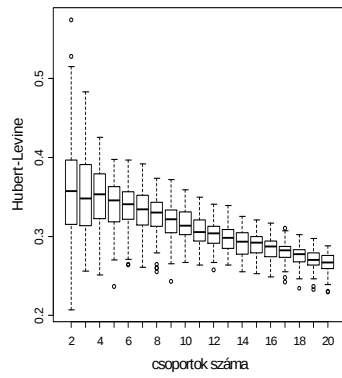
dc_16_10

17. táblázat: A vizsgált indexek értékei az 50 vizsgált mesterséges adatsor egyikénél a valódi csoportosítás, és egy két csoportból álló partíció esetén. Mivel az indexek egy részénél a kisebb, másoknál viszont a nagyobb érték jelez jobb csoportosítást, zölddel jelöltem azokat az indexeket, amelyek jobbnak jelezték a valódi csoportosítást a másiknál, és pirossal azokat, amelyek rosszabbnak.

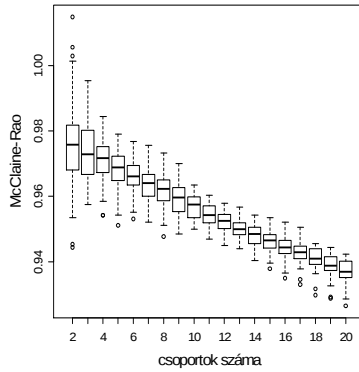
	valódi csoportosítás	2 csoport (1. és 2. csoport, illetve 3-5. csoport összevonva, RHA=0.204)
1. Pont-biszeriális korreláció	0.546	0.662
2. Tau	0.482	0.528
3. Baker & Hubert	0.865	0.799
4. G_+	0.567	0.096
5. Hubert & Levine	0.058	0.110
6. McClaine & Rao	0.582	0.647
7. PARTANA	1.267	1.314
8. Dunn _{min-max}	0.267	0.481
9. Dunn _{avg-avg}	0.902	1.531
10. módosított Davis & Bouldin	1.556	1.200
11. Popma (súlyozatlan)	0.441	0.600
12. Popma (súlyozott)	0.441	0.632
13. Átlagos sziluett	0.191	0.348
14. Helyesen klasszifikált felvételek aránya	0.970	0.940
15. Feoli	0.633	0.788
16. ISAMIC	1.014	0.934
17. Morisita	0.144	0.601
18. Relatív divergencia	0.014	0.008
19. Crispness	16320.30	15207.47
20. átlagos R_{adj}^2	0.355	0.211
21. fidelitás értékek átlagos elsőfajú hibája	0.636	0.600
22. OptimClass1 ($p=10^{-3}$)	62	45
23. OptimClass1 ($p=10^{-6}$)	28	27
24. OptimClass1 ($p=10^{-9}$)	16	16
25. átlagos nicheszélesség	2.783	6.593



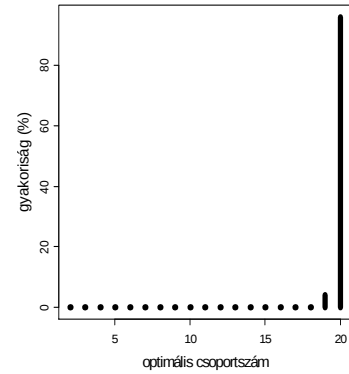
27. ábra: A távolságmátrix és a csoportosítás közötti kapcsolat szorosságát mérő indexek – pont-biszeriális korreláció (a,b), Tau (c,d), Baker & Hubert (e,f), G_+ (g,h), Hubert & Levine (i,j), McClaine & Rao (k,l), és (m,n) PARTANA – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja



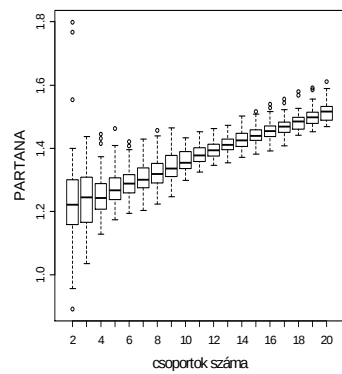
i)



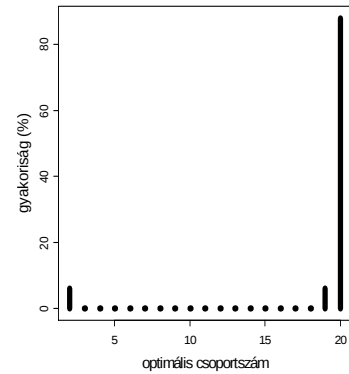
j)



k)



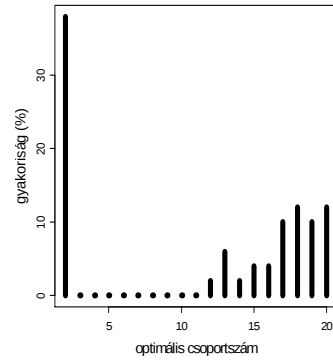
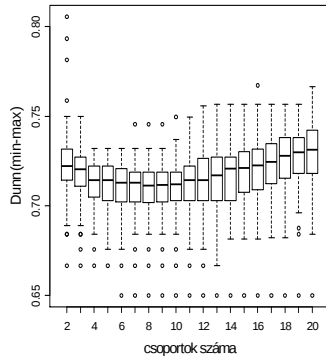
l)



m)

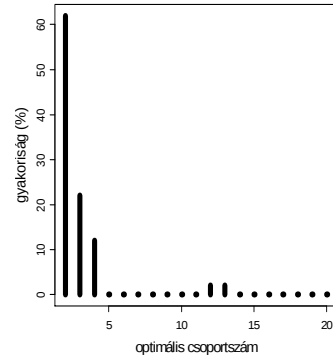
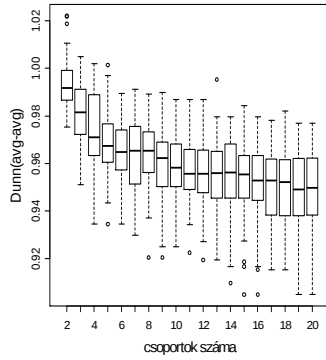
n)

27. ábra: (folytatás) A távolságmátrix és a csoportosítás közötti kapcsolatot mérő indexek – pont-biszeriális korreláció (a,b), Tau (c,d), Baker & Hubert (e,f), G_+ (g,h), Hubert & Levine (i,j), McClaine & Rao (k,l), és (m,n) PARTANA – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja



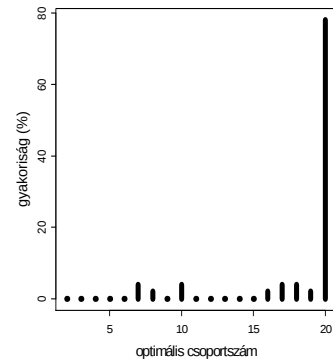
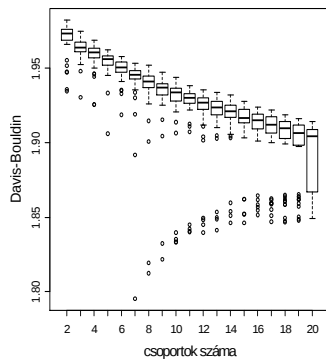
a)

b)



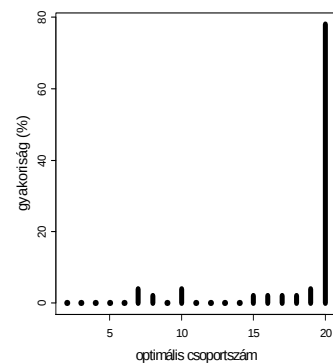
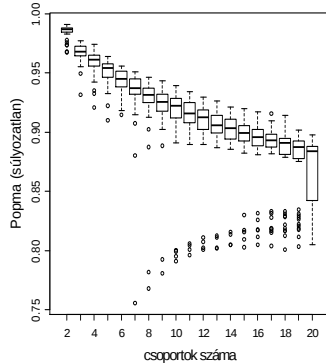
c)

d)



e)

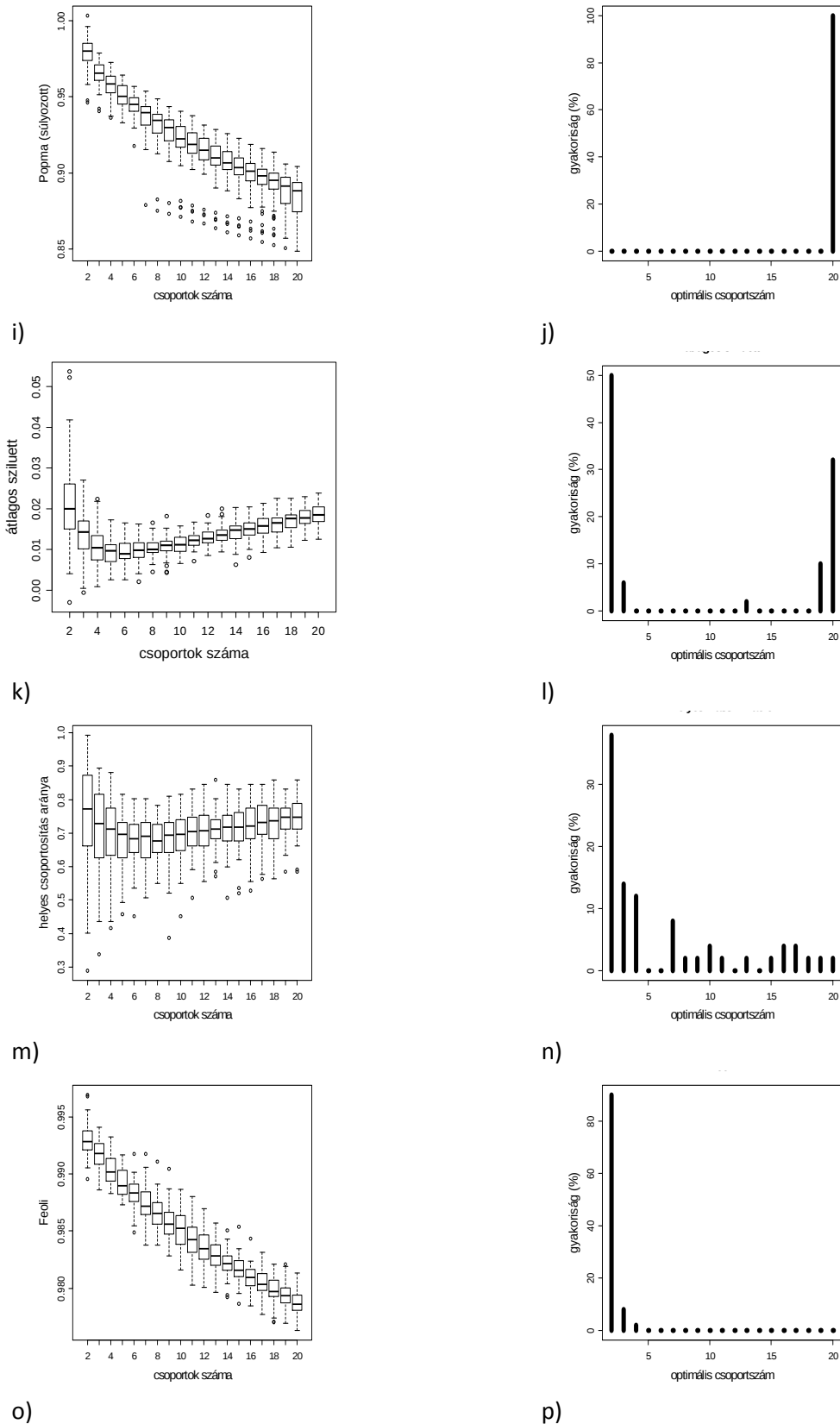
f)



g)

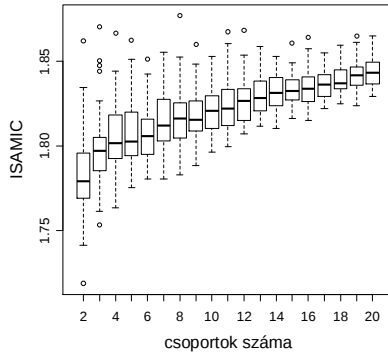
h)

28. ábra: A csoportok tömörségét és elválását mérő indexek – $Dunn_{min-max}$ (a,b), $Dunn_{avg-avg}$ (c,d), módosított Davis & Bouldin (e,f), Popma (súlyozatlan) (g,h), Popma (súlyozott) (i,j), átlagos sziluett (k,l), helyesen klasszifikált felvételek aránya (m,n), Feoli (o,p) – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja

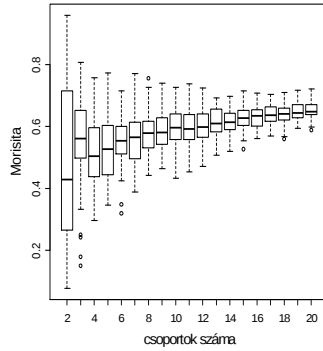


28. ábra: (folytatás) A csoportok tömörségét és elválását mérő indexek – $Dunn_{min-max}$ (a,b), $Dunn_{avg-avg}$ (c,d), módosított Davis & Bouldin (e,f), Popma (súlyozatlan) (g,h), Popma (súlyozott) (i,j), átlagos sziluett (k,l), helyesen klasszifikált felvételek aránya (m,n), Feoli (o,p) – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja

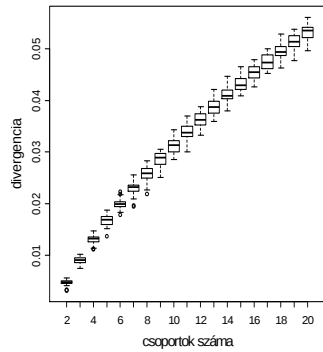
dc_16_10



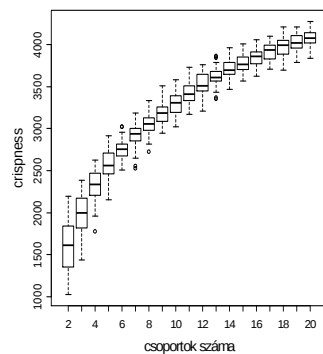
a)



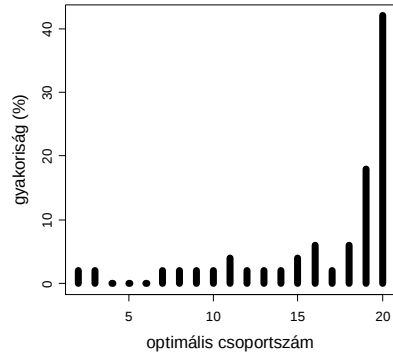
c)



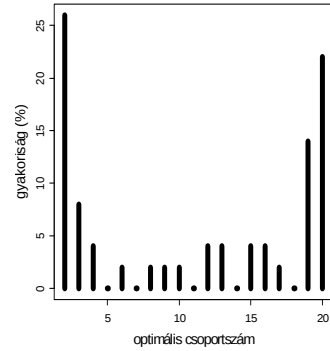
e)



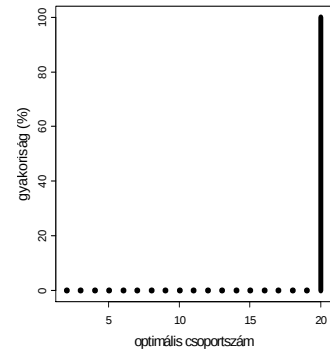
g)



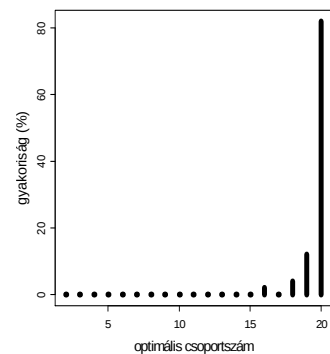
b)



d)

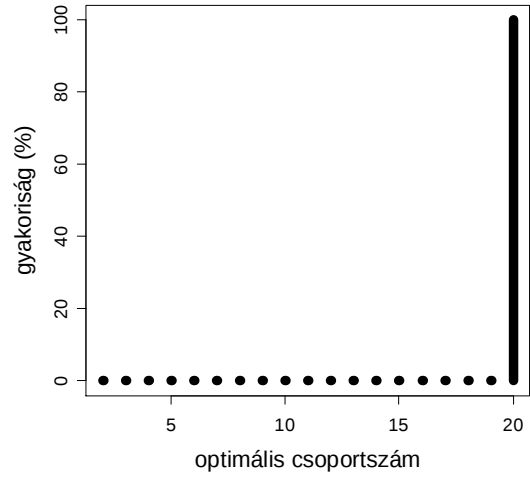
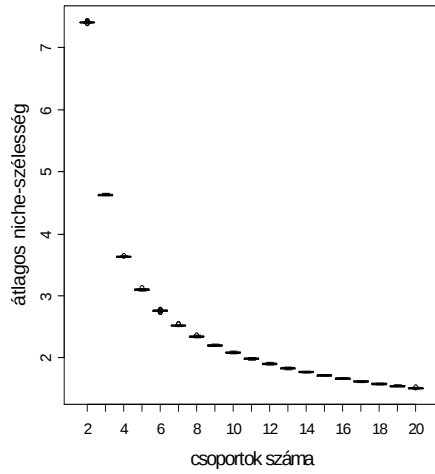


f)

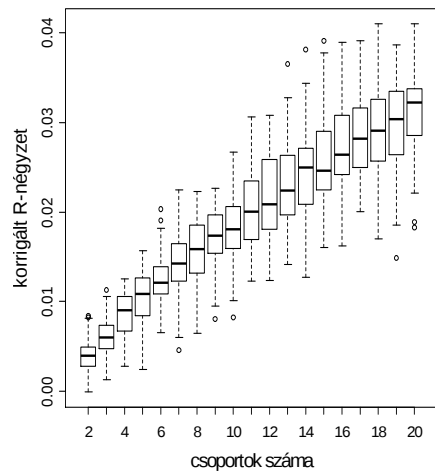


h)

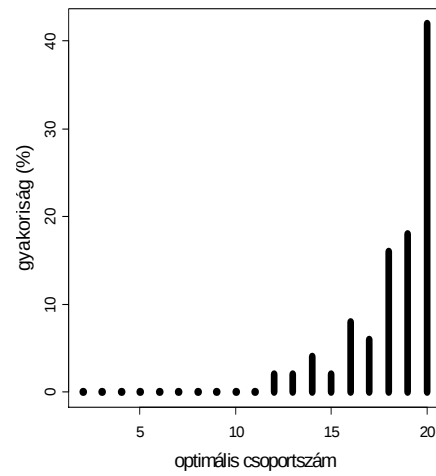
29. ábra: A csoportok értelmezhetőségét mérő indexek –ISAMIC (a,b), Morisita (c,d), relatív divergencia (e,f), Crispness (g,h), átlagos nicheszélesség (i,j), átlagos R_{adj}^2 (k,l), fidelitás értékek átlagos elsőfajú hibája (m,n), OptimClass1 ($p=10^{-3}$) (o,p), OptimClass1 ($p=10^{-6}$) (q,r), OptimClass1 ($p=10^{-9}$) (s,t) – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja



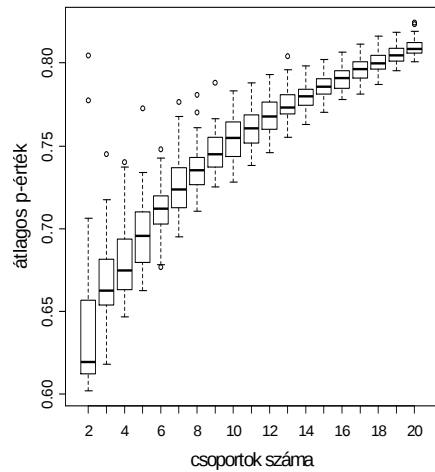
i)



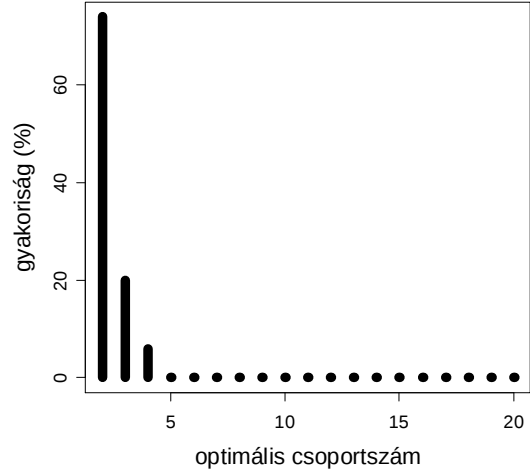
j)



k)



l)

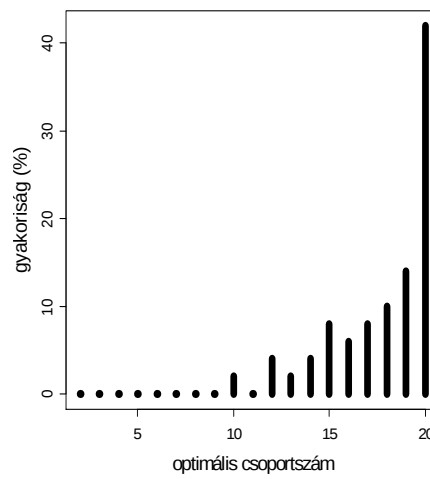
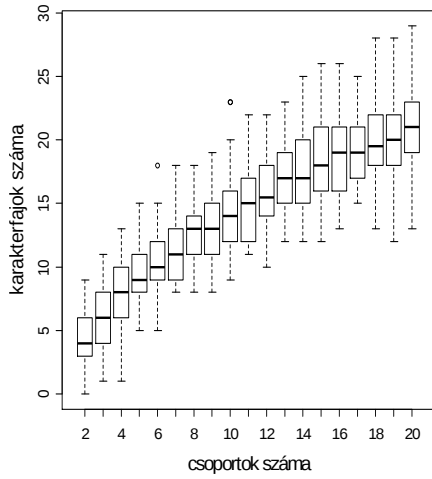


m)

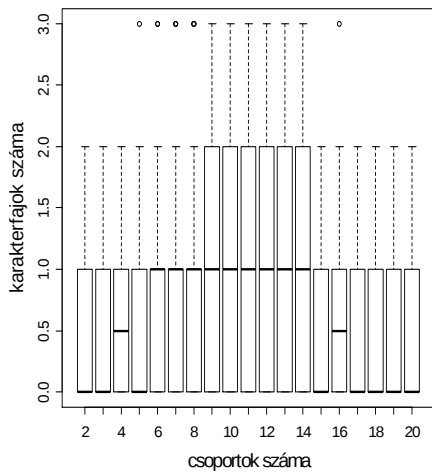
n)

29. ábra: A csoportok értelmezhetőségét mérő indexek –ISAMIC (a,b), Morisita (c,d), relatív divergencia (e,f), Crispness (g,h), átlagos nicheszélesség (i,j), átlagos R_{adj}^2 (k,l), fidelitás értékek átlagos elsőfajú hibája (m,n), OptimClass1 ($p=10^{-3}$) (o,p), OptimClass1 ($p=10^{-6}$) (q,r), OptimClass1 ($p=10^{-9}$) (s,t) – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja

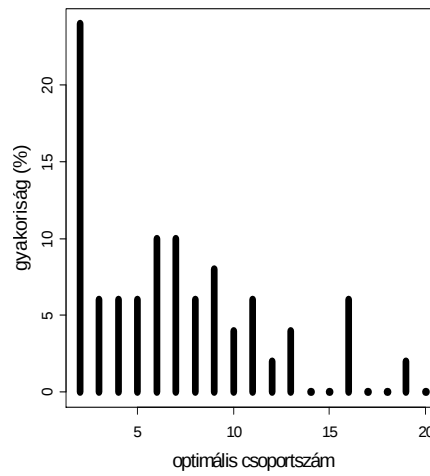
dc_16_10



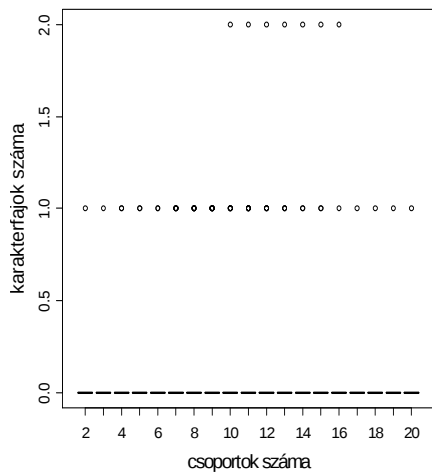
m)



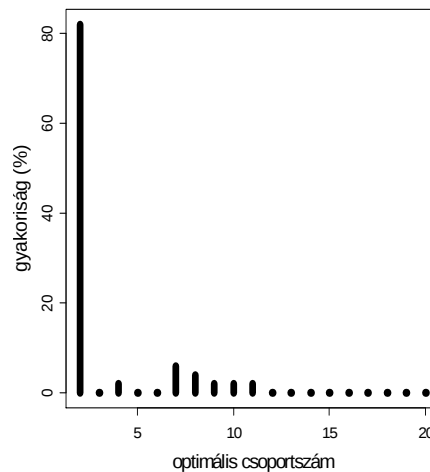
n)



o)



p)



q)

r)

29. ábra: A csoportok értelmezhetőségét mérő indexek –ISAMIC (a,b), Morisita (c,d), relatív divergencia (e,f), Crispness (g,h), átlagos nicheszélesség (i,j), átlagos R^2_{adj} (k,l), fidelitás értékek átlagos elsőfajú hibája (m,n), OptimClass1 ($p=10^{-3}$) (o,p), OptimClass1 ($p=10^{-6}$) (q,r), OptimClass1 ($p=10^{-9}$) (s,t) – viselkedése csoportstruktúra hiányában. A baloldali oszlop az indexek értékének változását mutatja a csoportszám függvényében, a jobboldali pedig az optimálisnak talált csoportszámok hisztogramja

A kapott eredmények azt mutatják, hogy a kételkedés helyénvaló volt. Míg a korábbi vizsgálatokban a pont-biszeriális korreláció és az átlagos sziluett a legjobb indexek között volt, addig ebben a vizsgálatban kifejezetten rossz eredményt adtak.

Meglepően sok esetben kaptunk negatív korrelációt a módosított Rand-indexszel, ami azt jelenti, hogy az indexek egy része a valós csoportosítástól jobban eltérő partíció(ka)t jobbnak értékelt, mint attól kevésbé eltérőket. Ennek okait keresve részletesebben is megvizsgáltam az 50 szimulált adatsor egyikét (a megállapítások valószínűleg a többire is igazak). Azt a meglepő tényt tapasztaltam, hogy számos index a valódi csoportosításnál jobbnak jelez egy olyat, ahol csak két csoport van, az eredeti csoportok közül az 1-2 és a 3-5 össze van vonva (17. táblázat). Az ezt a hibát elkövető indexek mindegyike rosszul szerepelt a korrelációs vizsgálatokban is. A problémát az okozhatta, hogy ugyan elvileg a gradiens mentén szomszédos csoportok között a különbség egyforma, de a kis különbség miatt egyes csoportok nem különültek el elég jól.

A geometriai indexek esetében felmerülhet, hogy a Jaccard helyett más távolságfüggvényt választva a csoportok elkülönülése jobb lett volna, és így némileg eltérő eredményeket kaphattunk volna.

A szomszédos csoportok fajkészletei közötti jelentős átfedésnek tulajdonítható, hogy a fajok fidelitásán alapuló indexek nem működtek jól. Amikor a faj preferenciáját vizsgálva egy-egy csoportot az összes többihez viszonyítunk, akkor ilyen esetben nem lesznek elég markánsak a különbségek. Sokkal jobb eredményt adott a korrigált R^2 (szintén bináris adatokból számolva!), mert a faj válogatóképességét nem páros összehasonlításokkal méri.

6.2.4. Következtetések, javaslatok

1. A bemutatott összehasonlításokban jó eredményt elért indexek bátran alkalmazhatók, de az itt sikertelenül szereplők sem feltétlenül használhatatlanok, különösen ha élesebben elváló fajkészletű csoportok várhatók.
2. Az indexek ugyan nem annak tesztelésére szolgálnak, hogy vannak-e egyáltalán csoportok, de trendjük (esetenként értékük) jelzi az adatstruktúra hiányát. Érdemes ezekre a jelekre odafigyelni, és nem egyszerűen az index hibás viselkedésének tekinteni őket.
3. Érdemes ugyanarra a problémára több indexet is kipróbálni, lehetőleg a három nagy csoport mindegyikéből.
4. Ha távolságfüggvények összehasonlítása a cél, csak nem-geometriai indexek használhatók.
5. Amennyiben a rendelkezésre álló számítási kapacitás és az adatsor mérete megengedi, érdemes az egyszerű indexek mellett, a csoportosítás repetitívitasát és stabilitását is vizsgálni.
6. További hasonló, mesterséges adatokat használó összehasonlításokra lenne szükség, ahol a csoportok elkülönülése nagyobb, elhelyezkedésük a gradiens mentén nem egyenletes és esetleg nem csak egy környezeti gradiens van, hogy feltárjuk, hogyan befolyásolják ezek a beállítások a kapott eredményeket.

6.3. Új módszer a csoportok repetitivitásának vizsgálatára

A klasszifikáció célja, hogy ne csak az adott mintára jellemző, hanem a vizsgált alapsokaságra általánosan érvényes, csoportokat találjunk. Az ilyen csoportok várhatóan repetitívek, azaz ha ugyanabból az alapsokaságból párhuzamosan több mintát veszünk, és azokat ugyanolyan módon elemezzük, minden esetben megjelennek ezek a csoportok.

A repetitivitás vizsgálatára mások által javasolt módszereket röviden bemutattuk a 6.1.5 fejezetben. Közös gyengéi minden eddigi módszernek, hogy ki kell választanunk egy klasszifikációs szabályt, és hogy csak a teljes partíció jóságáról adnak felvilágosítást. A lehetséges klasszifikációs szabályok száma igen nagy, és elvi alapon nem lehet közülük választani, ugyanakkor a választás erősen befolyásolja a későbbi eredményeket. Ezért a következőkben egy olyan saját fejlesztésű módszert mutatok be, amely nem használ klasszifikációs szabályt; helyette a csoportok jelentésének összehasonlításán alapul. Előnye továbbá a korábbi módszerekkel szemben, hogy ilyen módon az egyes csoportok repetitivitása külön is vizsgálható, és megkülönböztethetők a valós¹² és a műtermék csoportok.

6.3.1. A módszer bemutatása

Először a nem-hierarchikus csoportosítások vizsgálatának módszerét mutatom be, ez után foglalkozom azzal, hogy hogyan alkalmazható ez a hierarchikus klasszifikációk esetén, partíciók sorozatának tekintve azokat.

A korábbi módszerektől eltérően, az új módszer nem csoportosítások összehasonlításán, hanem a csoportok jelentésén alapul. A 22. ábrán látható példában a csoport jelentése a két változó alkotta koordináta-rendszerben elfoglalt helye volt.

A vegetációs adatok esetén a csoportok értékelése gyakran a fajok jelenléte/hiánya alapján történik, ezért a továbbiakban a bináris adatokkal foglalkozom. Bináris adatok esetén a csoportok centroidja a változók (fajok) relatív frekvenciáinak vektora, ami általában egyben a csoport értelmezésének az alapja is.

Két relatív gyakoriság (arány) összehasonlítására Zar (1999) a Z statisztikát ajánlja, amely közelítőleg standard normális eloszlású, ha a null-hipotézis igaz:

$$Z = \frac{|q_1 - q_2|}{\sqrt{\frac{Q(1-Q)}{n_1} + \frac{Q(1-Q)}{n_2}}}$$

ahol:

q_1, q_2 a két relatív gyakoriság

n_1, n_2 a két csoport mérete (felvételek száma)

$$Q = \frac{n_1 q_1 + n_2 q_2}{n_1 + n_2}$$

Ha két felvétels csoportot akarunk összehasonlítani, akkor először kiszámítjuk minden fajra a Z

¹² Az angol nyelvű szakirodalom a „valid” kifejezést használja, amit jobb híján valósnak vagy valódinak magyaráztam.

statisztikát és az elsőfajú hiba valószínűségét, majd az utóbbiakat egyetlen értékbe sűrítjük a Fisher-féle omnibusz teszttel (Sokal & Rohlf 1981):

$$\chi^2 = -2 \sum \ln p_i$$

ahol:

p_i = az elsőfajú hiba valószínűsége i faj relatív gyakoriságainak összehasonlításakor.

Ha a két felvételtcsoport azonos alapsokaságból vett minta, és a relatív gyakoriságok különbségét csak a véletlen okozza, akkor a kiszámított χ^2 érték megközelítőleg khi-négyszet eloszlást követ, amelynek szabadsági foka a fajsám kétszerese. Így a khi-négyszet eloszlás kiválasztott szignifikanciaszinthez tartozó kritikus értéke lehet a határérték, ami alapján megkülönböztetjük a hasonló és különböző jelentésű csoportokat. Mivel azonban az alkalmazott próbák feltételei általában nem teljesülnek maradéktalanul, ehelyett egy másik megoldást ajánlok: χ^2 értéket kiszámítjuk a két vizsgált adatsorra, és azokat a csoportokat tekintjük hasonlóknak, amelyeknél a χ^2 érték kisebb ennél. Vagyis két csoport akkor hasonló, ha kevésbé különbözik, mint a két adatsor.

Egy csoportot akkor tekintünk valódinak, ha egy és csak egy olyan csoport van a másik adatsorban, amely hasonlít hozzá. Ha nincs egyáltalán ilyen csoport a másik adatsorban, akkor a csoport csak a mintára jellemző és nem az alapsokaságra. Ha viszont több ilyen csoport is van, akkor valószínűleg egy valódi csoport önkényes felosztásával állunk szemben.

A hierarchikus klasszifikáció eredménye felfogható növekvő csoportszámú partíciók sorozataként is. Az előzőekben bemutatott eljárás a dendrogram minden szintjén elvégezhető, kezdve két csoporttal, majd folytatva három, négy stb. csoport esetén, mindaddig, amíg a valódinak minősített csoportok száma nem kezd csökkenni a csoportok önkényes feldarabolása miatt. Ha az értékeléshez egyetlen szintet akarunk kiválasztani, akkor a dendrogramot azon a szinten kell kiértékelni, ahol a valós csoportok száma maximális.

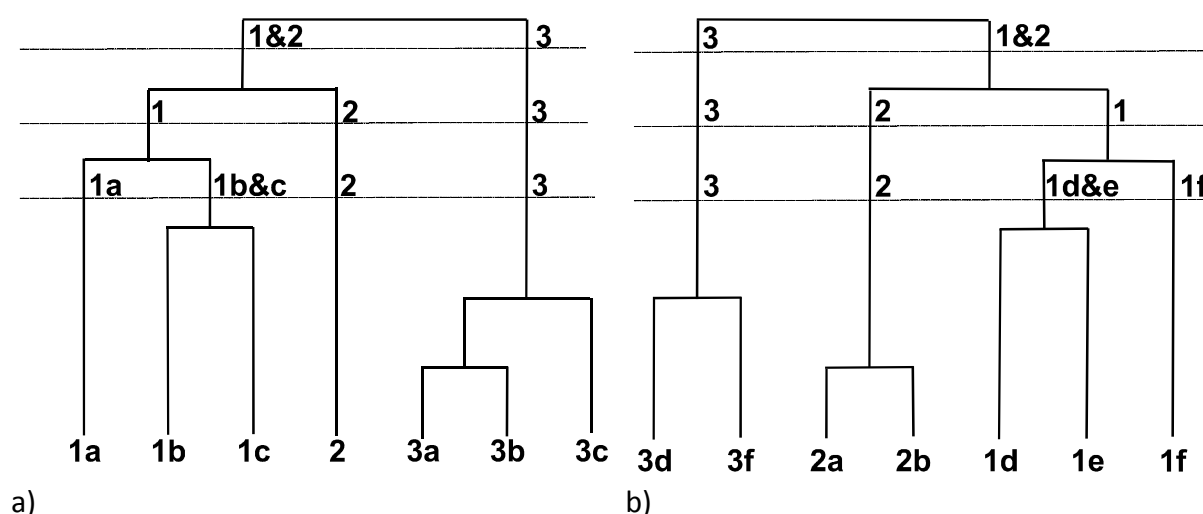
6.3.2. Példák a módszer alkalmazására

A módszer alkalmazását elsőként egy mesterséges adatsoron szemléltetem, amely 300 felvételtől áll és 200 faj jelenlét/hiány adatait tartalmazza. Az 300 felvétel három egyforma méretű, jól elkülönülő, egymástól egyformán különböző csoportba sorolható. (Az adatsor a honlapomról letölthető: www.botanika.hu/bdz/validaton_sampledata).

Az elemzés során az adatokat véletlenszerűen két egyenlő részre osztottam (A és B adatsor)¹³, és elkészítettem mindkét adatsor dendrogramját Sörensen különbözőségeken alapján átlagos lánc (UPGMA) összevonási algoritmust használva (30. ábra). Ha a dendrogramokat úgy vágjuk el, hogy három csoportot kapjunk, az A adatsor minden csoportja egy és csak egy csoportba hasonlít a B adatsorban, és viszont. Vagyis, ahogy azt vártuk is, ezen a szinten minden csoport valósnak bizonyult. Magasabb és alacsonyabb szinteken nem várunk ilyen teljes egyezést. Ennek ellenére ha csak két csoportot különböztetünk meg,

¹³ A megjelent cikkekben a „train” és „test” elnevezéseket használtam a két adatsorra, ahogy az más hasonló munkákban is szokásos. Mivel azonban itt a két adatsor szerepe szimmetrikus, helyesebbnek érzem az ezt kifejező A és B adatsor elnevezés használatát.

akkor mindkettő valósan bizonyul, mert véletlenül mindkét dendrogram az 1-es és 2-es csoportot vonja össze. Habár ez az összevonás mindkét adatsorban megfigyelhető, mégsem az alapsokaság, csak a két minta jellemzője (lásd a Diskusszió fejezetben). Ha négy csoportot különböztetünk meg, akkor mindkét adatsorban az 1-es csoportot vágjuk önkényesen két részre (30. ábra). Az A adatsor 1a csoportja két csoporthoz is hasonlít a B adatsorban: az 1d&e és az 1f csoporthoz. Az 1b&c csoport szintén ugyanehhez a két csoporthoz hasonlít. Ilyen mintázat, vagyis amikor egy csoport a másik adatsor több csoportjához is hasonlít, a valós csoportok önkényes felosztását jelzi. Így ezen a szinten már csak két valós csoportot találtunk, szemben a korábbi hárommal. A valós csoportok számának csökkenése azt jelzi, hogy nem érdemes tovább folytatni az elemzést, mert a további felosztások valószínűleg a valós csoportok feldarabolásához vezetnek.



30. ábra: A példa adatsorok dendrogramjainak felső része (Sörensen különbség, UPGMA algoritmus). A vízszintes vonalak mentén elvágva a dendrogramokat két, három, illetve négy csoportból álló partíciókat kapunk. A hasonló csoportokat a két dendrogramon azonos számmal jelöltük, míg a számok utáni betűk a valós csoportok önkényes elvágásával kapott részcsoporthoz jelölik. Ezek a részcsoporthoz hasonlóak az összes olyan részcsoporthoz a másik adatsorban, amelynek a száma ugyanaz, például az 1a hasonlít az 1d, 1e és 1f részcsoporthoz. Az & jel a (rész)csoporthoz összevonását jelzi.

A következőkben két valós adatsor elemzését mutatom be, kiemelve a csoportok repetitívásával kapcsolatos eredményeket. Az adatok összegyűjtése és az eredmények értékelése társszerzőim feladata volt, ezért itt ezekre csak röviden térek ki.

Az első esettanulmány a félszáraz gyepek vizsgálata volt egy Németországtól Erdélyig húzódó kontinentalitási gradiens mentén (Illyés et al. 2007). A felvételek nagyobb részt cönológiai adatbázisokból származtak (Jandt 1999, Valachovič 1999, Chytrý & Rafajová 2003), kisebb részben ehhez az elemzéshez lettek digitalizálva. A felvételek kiválasztásának szempontjai a következők voltak: a mintavételi egység mérete legalább 4 és legfeljebb 100 m² lehet, a *Festuco-Brometea* osztályba tartozik, és vagy a *Brachypodium pinnatum*, vagy a *Bromus erectus* borítása legalább 25%. Erre a bonyolult kiválasztási szabályra azért volt

szükség, mert a félszáraz gyepek (*Brometalia erecti*) lehatárolása országonként eltérő. Így végül Németország, Ausztria, Csehország, Szlovákia, Magyarország és Románia területéről összesen 2926 darab, a fenti kritériumoknak megfelelő felvételt gyűjtöttünk össze. Elkerülendő az intenzíven vizsgált területek túltreprezentálását, földrajzilag rétegzett újrámintavételezést alkalmaztunk (Knollová et al. 2005). A kilógó felvételek kizárása után a két párhuzamos adatsor 422 – 422 felvételből állt. A feldolgozás során bináris adatokat használtunk; a felvételek különbözőségét a Sörensen különbözőséggel mértük. A 4. fejezetben leírt zajsűrés után, Ward módszerrel klasszifikáltuk a felvételeket.

A dendrogramokat 11 csoportnál elvágva kaptuk az optimális klasszifikációt, ahol 6 csoport volt valós, amelyek jelentése a következő:

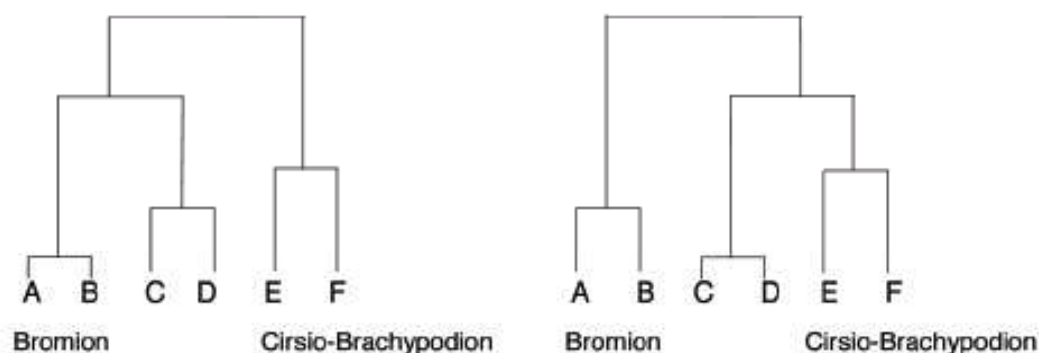
- A. Szubatlantikus területek *Brachypodium pinnatum* dominálta félszáraz gyepei
- B. Szubatlantikus területek *Bromus erectus* vagy *Brachypodium pinnatum* dominálta, cserjésedő gyepei
- C. Üde talajokon kialakuló, *Bromus erectus* dominálta félszáraz gyepek
- D. A Fehér Kárpátok, főleg *Bromus erectus* dominálta gyepei
- E. Meszes alapkőzetten kialakuló, nyílt gyepek, elsősorban Csehországban
- F. Kárpát-medencei *Brachypodium*-os gyepek

A mindkét adatsorban egyértelműen azonosítható, valós csoportoknak általában több karakterfajuk és szűkebb földrajzi elterjedésük volt, mint a nem valós csoportoknak.

A csoportok majdnem fele nem bizonyult valósnak, és a felvételek nagyjából fele tartozik a valós csoportokba. Vagyis ha nem vizsgálnánk a csoportok ismétlődését, jóval több csoportot kellene megpróbálnunk értelmezni és beilleszteni szüntaxonómiai rendszerbe. Ezen csoportok többsége nem különül jól el a többi csoporttól, kevés karakterfajuk van, így nehezen jellemezhetők. De még ha jól jellemezhető is egy ilyen csoport, akkor is problémás a beillesztése a szüntaxonómiai rendszerbe, amivel szemben fontos elvárás a stabilitás, ezek a csoportok viszont valószínűleg nem jelennének meg egy másik vizsgálatban. Ezzel természetesen nem azt akarom mondani, hogy csak valós csoportba tartozó felvételeket, azaz csak a felvételek felét kellene besorolni valamilyen a szüntaxonba. A nem ismétlődő, főleg generalistákból álló csoportok, mint alapasszociáció (basal community) (Kopecký & Hejný 1978) vagy központi szüntaxon (Dierschke 1981) hozzárendelhetők egy asszociációfeletti egységhez, esetünkben valószínűleg¹⁴ a *Brometalia erecti* rendhez.

A valós csoportok összekapcsolódási sorrendje a két adatsorban nem teljesen azonos (31. ábra), valószínűleg azért mert három egyenrangú csoport van, amit a dichotomikus elágazásrendszer nem tud jól leképezni. A három nagyobb csoport közül kettő egyértelműen megfeleltethető a korábban leírt asszociációcsoportoknak, a harmadik viszont új.

¹⁴ A kérdés megnyugtató tisztázásához ezeket a felvételeket össze kellene hasonlítani, nem a *Brometalia erecti* rendbe tartozó felvételekkel is.



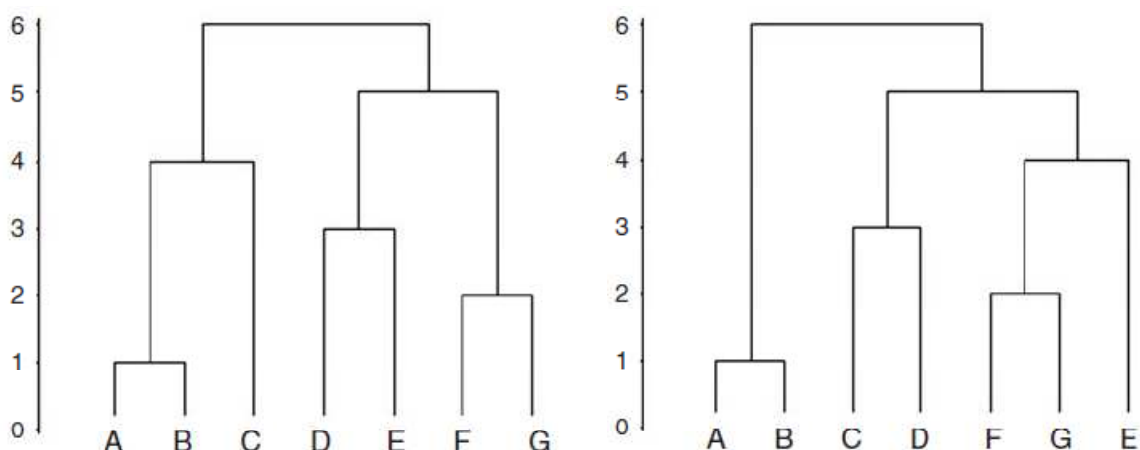
31. ábra: A közép-európai félszárazgyepi adatsor valós csoportjainak dendrogramjai a két párhuzamos adatsorban. Egyértelműen elválik egymástól az A és B csoportból álló *Bromion erecti* asszociációcsoport, és az E és F csoportból álló *Cirsio-Brachypodion* asszociációcsoport. A C és D csoport valószínűleg egy harmadik asszociációcsoport a renden belül.

A második esettanulmány szintén félszáraz gyepeket vizsgált, de szűkebb földrajzi régióban, csak Magyarország területén. Az elemzéshez részben az előző esettanulmányban is szereplő felvételeket használtuk, de Bauer Norbert és Illyés Eszter számos új felvételt is gyűjtött ehhez az elemzéshez, amelyek helyének kiválasztásához felhasználták a MÉTA adatbázis (Molnár et al. 2007, 2008) H4 (=Félszáraz irtásrétek, száraz magaskórósok és erdőssztyeprétek) élőhelyekre vonatkozó adatait is. A két vizsgált adatsor 287, illetve 290 felvételből állt. Az adatfeldolgozás az előző esettanulmánynál leírtak szerint történt ebben az esetben is.

A csoportok optimális száma 10 volt, amelyek közül hét bizonyult valósnak. Ezek a következők voltak:

- A. tipikus *Euphorbio pannonicae-Brachypodietum* gyepek
- B. *Euphorbio pannonicae-Brachypodietum* társulás *Bromus erectus*-os altípusa
- C. *Euphorbio pannonicae-Brachypodietum* átmeneti állományai Érd közeléből
- D. *Polygalo majoris-Brachypodietum* gyepek a Bükk hegységben
- E. *Sanguisorbo minoris-Brometum erecti* gyepek a Dunántúli-középhegységben
- F. *Trifolio medii-Brachypodium pinnatum* gyepek az Északi-középhegységben
- G. Regenerálódó vagy az üde rétek felé átmenetet alkotó állományok

A valós csoportok összekapcsolódását megvizsgálva nem találunk egyértelmű hierarchiát (32. ábra), bár a kontinentálisabb klímájú alföldi területek, löszön kialakult vastag termőrétegű talajaira jellemző csoportok (A és B csoport) mindkét dendrogramon magas szinten elkülönül a csapadékosabb és hűvösebb hegyvidéki területekre jellemző típusoktól (E, F és G csoport). A két bizonytalan helyzetű csoport (C és D) nagyon szűk elterjedésű, valószínűleg a helyi jellegzetességeket tükrözi. A nem valós csoportok is valószínűleg ilyen lokális, a helyi körülményeket tükröző típusok lehetnek, amelyeket – a G csoportba sorolt felvételekkel együtt – nem érdemes külön asszociációként leírni, hanem csak asszociációcsoport szinten besorolhatók a *Cirsio-Brachypodion* asszociációcsoportba.



32. ábra: A hazai fészarázgyepi adatsor valós csoportjainak dendrogramjai a két párhuzamos adatsorban. Egyértelműen elválnak egymástól a száraz területeken, lösz alapkőzetén kialakult mély talajon élő A és B csoport, és a csapadékosabb hegyvidéki területeken, általában sekélyebb talajon élő E, F és G csoport.

6.3.3. Következtetések

A bemutatott példákban látszik, hogy a csoportok repetitívitasának vizsgálata számos esetben segít elkerülni azt, hogy csak a mintára jellemző csoportokat is értékeljünk. Ugyanakkor a mesterséges adatok elemzése arra is rámutatott, hogy nem tudjuk elkerülni az összes ilyen műterméket. Ott az első és a második csoport összevonása mindkét mintában megjelent, annak ellenére, hogy az alapsokaságban mindhárom csoport egyforma távolságban volt egymástól. Ez a probléma a módszertől függetlenül mindig jelentkezik, ha a csoportok repetitívitasát vizsgáljuk, mert ezek a módszerek nem az eredmények pontosságát („accuracy”) mérik (McIntyre & Blashfield 1980). A repetitívitas azt mutatja meg, hogy az eredmények mennyire ismétlődnek a párhuzamos mintákban, míg a pontosság azt méri, hogy a mintából kapott eredmények mennyire egyezik meg az alapsokaság tényleges struktúrájával. Utóbbi csak akkor mérhető, ha az alapsokaság valódi szerkezete ismert, de ebben az esetben nem lenne szükség a klasszifikációs eljárásokra. Ezért a pontosságot csak módszertani vizsgálatokban szokták mérni, ahol ismert szerkezetű mesterséges adatsorok elemzésével vizsgálják a különböző módszerek tulajdonságait (például Podani 1997).

Szerencsére a két tulajdonság nem független, a pontos klasszifikáció általában stabil is, de ez nem igaz fordítva. Lehetségesek ismétlődő csoportok, amelyek ennek ellenére sem az alapsokaság szerkezetét tükrözik. Ezért hibás gondolatmenet volna egy klasszifikáció repetitívitasából annak pontosságára következtetni (McIntyre & Blashfield 1980). Helyes viszont az ellenkező irányú gondolatmenet: a csoport nem lehet pontos, ha nem repetitív. Vagyis a módszer célja nem a csoportok igazolása, hanem a nem ismétlődő, ezért valószínűleg műtermék csoportok kiszűrése (v.ö. Popper falszifikációs kritériumával; Popper 1997).

Elméletileg minden valós csoport egyben repetitív is. Lehetséges azonban, hogy egy ritka csoport tagjai – pusztán a kis gyakoriságuk miatt – nem kerülnek bele mindkét adatsorba, így a csoport nem lesz ismétlődő. Ennek a valószínűsége az elemzett felvételek

dc_16_10

számának növekedésével csökken, de teljesen nem küszöbölhető ki.

A repetitivitás vizsgálatára szolgáló korábbi módszerek csak a teljes klasszifikáció elemzésére, vagy az optimális csoportszám megállapítására alkalmasak, de nem mondanak semmit az egyes csoportok ismétlődéséről. Az új módszer ezt is lehetővé teszi, és a valós adatok elemzését bemutató esettanulmányokban látható volt, hogy ez mennyire hasznos. A módszer másik előnye, hogy nincs szükség klasszifikációs szabályra, amelynek kiválasztása általában nem egyértelmű, és amely jelentősen befolyásolja az eredményeket.

7. Fajkészlet becslése

Az ökológiai szakirodalomban általánosan elfogadott tény, hogy a közösségek fajösszetételét részben a rendelkezésre álló fajkészlet, és az abból való válogatás szabályai, a társulási szabályok (assembly rules) együttesen határozzák meg (Oborny & Botta-Dukát 2007). A fajkészlet szerepét a társulások összetételének kialakításában számos példa demonstrálja, például Mack (2003) számos érdekes példát mutat be a történeti okokból hiányzó fajcsoportok hatásáról.

A fajkészlet hatására vonatkozó vizsgálatok egyik korlátja a fajkészlet becslésére szolgáló általánosan használható, objektív módszerek hiánya (Eriksson & Ehrlén 1992, Dupré 2000). Számos módszer van, amely az azonos közösségben készült felvételek alapján becslést ad a fajkészlet méretére. Ezekről a módszerekről Palmer (1990) és Colwell & Coddington (1994) cikkei jó áttekintést adnak. Ezek a módszerek azonban nem mondanak semmit arról, hogy melyek azok a fajok a fajkészletnek, amelyeket nem sikerült egyetlen felvételben sem megfigyelnünk, csak ezeknek a fajoknak a számára adnak becslést.

Pärtel és munkatársai (1996) a fajok környezeti igényét jelző Ellenberg-féle indikátorértékeket használták annak megállapítására, hogy adott termőhely mely fajok számára lehet alkalmas, és ezeket tekintették a fajkészlet részének. Dupré (2000) szerint a fajok cönológiai karakter szerinti besorolása jobb eredményt ad, mint az indikátorszámok, mert a fajok környezeti igényének részletesebb (habár implicit) leírását tartalmazza. Mindkét megközelítés hátránya, hogy a fajok szubjektív besorolásán alapul, ami csak a közép-európai növényfajokra áll rendelkezésre, másutt általában nem. Ha a fajok előfordulását limitáló környezeti tényezők egyértelműen azonosíthatók – például a talaj sótartalma (Grace & Pugsek 1997) –, akkor a fajok realizált ökológiai toleranciája (Pásztor & Mészéna 2007) és a környezeti tényezők mért értékei alapján is becsülhető a fajkészlet. A gyakorlatban ez azonban csak ritkán kivitelezhető, mert a releváns limitáló faktorok azonosítása és mérése nehéz, és a fajok ökológiai toleranciájáról sincs általában elegendő adat.

A probléma megoldására Ewald (2002) a Beals simítás alkalmazását javasolta. Ez az eredetileg a sokváltozós elemzések előtti adattranzformációra javasolt módszer (Beals 1984, McCune 1994), az eredeti bináris adatokat a fajok együttes előfordulási valószínűségei alapján számolt 0 és 1 közötti értékekkel helyettesíti, amelyek azt fejezik ki, hogy a termőhely mennyire alkalmas a faj számára. A módszert viszonylag ritkán alkalmazzák (pl. Chytrý et al. 2003), aminek egyik oka lehet, hogy a szakirodalomban nem található egyértelmű megoldás arra, hogy a folytonos értékeket hogyan alakítsuk át bináris skálává (1 = a fajkészlet eleme, 0 = nem tartozik bele a fajkészletbe).

Ebben a fejezetben ezzel az átskálázással kapcsolatos módszertani vizsgálatok kerülnek bemutatásra. Ezt megelőzően azonban először röviden összefoglalom a vizsgálat előzményeit, azaz bemutatom a Beals simítás lényegét, és összefoglalom a bináris skálára történő tranzformációval kapcsolatos problémákat és az eddigi megoldási javaslatokat. Az elvégzendő vizsgálatok pontosabb meghatározására csak ezután kerülhet sor.

7.1. Beals simítás

A Beals simítás során a faj jelenlétének feltételes valószínűségeit becsüljük a többi előforduló faj jelenlétében, majd az így kapott értékeket átlagoljuk (De Cáceres & Legendre 2008):

$$b_{ik} = \frac{1}{S_k - 1} \sum_{j \neq i} f_{jk} \frac{C_{ij}}{P_j}$$

ahol:

b_{ik} = az i faj simított értéke a k felvételben

S_k = a k felvétel fajsza

$f_{jk} = 1$, ha a j faj előfordul a k felvételben, 0 ha nem

C_{ij} = az i és a j felvétel együttes előfordulásainak száma a referencia adatbázisban

P_j = a j faj előfordulásainak száma a referencia adatbázisban

Az itt bemutatott képlet eltér a sokváltozós módszerek alkalmazása előtti simításra használható eredeti képlettől, amely szerint:

$$b_{ik}^* = \frac{1}{S_k} \sum_j f_{jk} \frac{C_{ij}}{P_j}$$

A módosítást az teszi szükségessé, hogy ha a kapott értékeket annak jellemzésére használjuk, hogy a termőhely a faj számára mennyire alkalmas, és a számolás során figyelembe vesszük, hogy a faj előfordul-e a vizsgált helyen, akkor a becslés torzított lesz (Münzbergová & Herben 2004, Oksanen et al. 2010).

A simított értékek megbízhatósága attól függ, hogy a fajok előfordulási (P_j) és együtt-előfordulási (C_{ij}) gyakoriságára vonatkozó adatok mennyire megbízhatóak. Az eredeti módszer szerint ezeket ugyanabból az adathalmazból számoljuk, mint amelyiknek az elemeire a simított értékeket megállapítjuk. Megbízhatóbb eredményt kapunk azonban, ha ezeknek a mennyiségeknek a becslését egy nagyobb adatbázisból (referencia adatbázis) végezzük (Münzbergová & Herben 2004). Ilyenkor az eredmények értékelésénél tekintettel kell lennünk arra, hogy a simított érték arról tájékoztat, hogy az adott faj számára a termőhely mennyire megfelelő, és nem veszi figyelembe azt, hogy a faj az adott régióban előfordul-e. Vagyis ezzel a módszerrel az élőhelyi fajkészletről (Belyea & Lancaster 1999, Oborny 2000) kapunk információt. Ha vannak más forrásból (például regionális flóraművekből) információink a geográfiai fajkészletről, akkor elkészíthetjük a két halmaz metszetét, a fajkínálatot (Oborny & Botta-Dukát 2007). Ha a fajok élőhelypreferenciáiban regionális különbségek vannak, akkor érdemes az országos adatbázis helyett szűkebb, regionális adatbázist használni referenciaként.

7.2. Transzformáció bináris skálára

A bináris skálára való átalakítás után kapott eredményeket összehasonlítva a valósággal négy kimenetel lehetséges, amelyeket az alábbi táblázat szemléltet:

		a faj eleme a fajkészletnek	
		igen	nem
a transzformáció	1	helyes döntés	másod fajú hiba
után kapott érték	0	első fajú hiba	helyes döntés

A táblázat emlékeztet a statisztikai döntések esetén felírható hasonló táblázatra és az elnevezéseket (első és másodfajú hiba) is onnan vettem át, ügyelve arra, hogy az elnevezések összhangban legyenek a Dupré (2000) által használt elnevezésekkel (error type I és error type II). A statisztikában használatos elnevezésekkel akkor lesznek összhangban ezek a nevek, ha nullhipotézisünk az, hogy a faj eleme a fajkészletnek. Sajnos a bemutatott analógia statisztikai döntésekkel nem segít abban, hogy a folytonos értékeket bináris adatokká alakítsuk.

A legegyszerűbb megoldás, hogy egy határérték alatti értékeket nullává, az vele egyenlő vagy nála nagyobb értékeket eggyé alakítjuk. Ewald (2002) bemutatta, hogy különböző határértékeket használva mennyire különböző eredményeket kapunk. Ha a határérték túl magas megnő az első fajú hiba valószínűsége, ha viszont túl alacsony, akkor a másodfajú hibát követjük el gyakran. Hasonló problémával találkozhatunk a fajok elterjedési területének modellezésekor is, ahol a modellezés eredménye általában a becsült előfordulási valószínűség (Guisan & Zimmermann 2000), de előfordulás/hiány adatokat szeretnénk a térképen megjeleníteni. A probléma megoldására számos módszer dolgozott ki, amelyekről Liu és munkatársainak (2005) cikke ad jó áttekintést. Sajnos ezek a módszerek nem adaptálhatók a fajkészlet megállapításának problémájára, mert a földrajzi elterjedés modellezése esetén a cél olyan modell létrehozása, amely minél pontosabban írja le, mind a megfigyelt előfordulásokat, mind a megfigyelt hiányokat (Liu et al. 2005). Ezzel szemben a most vizsgált esetben a faj hiányából semmilyen következtetést nem vonhatunk el arra nézve, hogy eleme-e a fajkészletnek.

Münzbergová és Herben (2004) rámutattak, hogy az egyetlen reális kiindulópontunk az lehet, hogy ritka kivételektől eltekintve az előforduló fajok elemei a fajkészletnek. A ritka kivételek oka lehet hibás határozás, vagy az, hogy a termőhely korábban alkalmas volt a faj számára és egyedei a megváltozott körülmények között is még túlélnek, annak ellenére, hogy a termőhely már nem alkalmas a faj számára. Vagyis olyan határértéket érdemes választani, ahol az előforduló fajokra (vagy legalábbis azok döntő többségére) a transzformáció után 1-es értéket kapunk.

Münzbergová és Herben (2004) megfigyelése szerint a ritka fajoknál a megfigyelt előfordulásokhoz tartozó simított értékek (b_{ik}) eloszlása eltolódik az alacsonyabb értékek felé. Cikkük nem foglalkozik a bináris skálára való átalakítás problémájával, helyette arra

adnak megoldást, hogy hogyan tehetők összehasonlíthatóvá a különböző gyakoriságú fajokra kapott simított értékek. Az eloszlások megfigyelt eltolódása azonban azt is jelenti, hogy a ritka fajokra optimális határértéket használva minden fajra, a gyakori fajoknál megnő a másodfajú hiba valószínűsége. A gyakori fajokra optimális határérték viszont magas elsőfajú hibát eredményez a ritka fajok esetén. Vagyis érdemes fajonként eltérő határértéket választani.

Münzbergová és Herben (2004) 22 folt fajlistájának elemzése alapján vonta le a következtetéseit, ezért úgy gondoltam érdemes a megfigyelt előfordulásokhoz tartozó simított értékek eloszlása és a faj gyakorisága közötti kapcsolatot egy nagyobb adatbázison is tesztelni. Mivel esetükben minden folt ugyanabba a vegetációtípusba tartozott, fel sem merült, hogy a felvételek fajszerkezetének van-e esetleg valamilyen szisztematikus hatása a simított értékekre. De Cáceres és Legendre (2008) vizsgálatában viszont a fajszerkezet felvételekből álló adatbázisban gyenge a korreláció a simított értékek és a fajok előfordulási valószínűsége között. A vizsgálat másik célja annak ellenőrzése volt, hogy a Beals simítás alkalmas-e a fajkészlet becslésére fajszerkezet közösségeiben is.

7.3. Anyag és módszer

A vizsgálatok során a CoenoDat adatbázisból (Lájer et al. 2008) származó 8437 cönológiai felvételt használtam, amelyben összesen 1854 faj fordult elő. A TurboVeg adatbázisban (Hennekens & Schaminée 2001) tárolt adatok előkészítése, a nehezen megkülönböztethető fajok összevonása a Juice programban (Tichý 2002) történt. A statisztikai elemzések az R statisztikai programmal készültek (R Development Core Team 2010), a Beals simításhoz a *vegan* (Oksanen et al. 2010) csomagot használtam.

A fajok frekvenciáinak hatását vizsgálva a fajonként kiszámoltam és a faj gyakoriságának függvényében ábrázoltam az előfordulásokhoz tartozó simított értékek eloszlásának jellemzőit: minimum, nem kilógó értékek¹⁵ minimuma, első kvartilis, medián, harmadik kvartilis, nem kilógó értékek maximuma, maximum.

A fajszerkezet hatásának vizsgálatához fajonként kiszámoltam a felvétel fajszerkezet és az előfordulásokhoz tartozó simított értékek közötti Kendall rangkorrelációt.

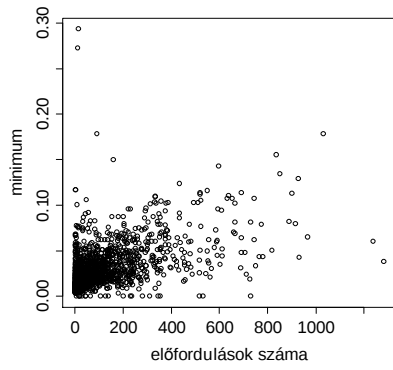
7.4. Eredmények

A simított értékek eloszlásának eltolódása a kisebb értékek felé a ritkább fajok esetén az eloszlást leíró valamennyi jellemző alapján jól látható (33. ábra), bár a minimum értékek esetén a kapcsolat kevésbé szembetűnő.

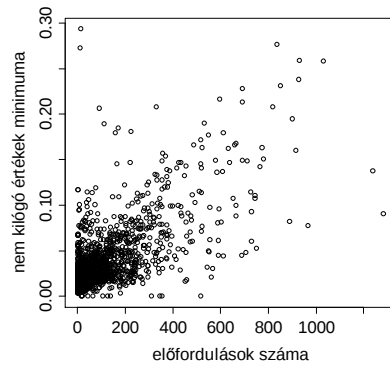
A legalább 10 felvételben előforduló 1372 faj 50,7% esetén nincs szignifikáns korreláció az előfordulásokhoz tartozó simított értékek és a felvételek fajszerkezet között; a fajok 39,3%-ánál találtam szignifikáns negatív, 10%-ánál szignifikáns pozitív korrelációt. A korrelációs koefficiensek eloszlása közel szimmetrikus (34. ábra).

¹⁵ A statisztikai szakirodalomban általában az alsó kvartiliséknél több mint az interkvartilis terjedelem másfélszeresével kisebb, vagy a felső kvartiliséknél több mint az interkvartilis terjedelem másfélszeresével nagyobb értékeket tekintik kilógónak.

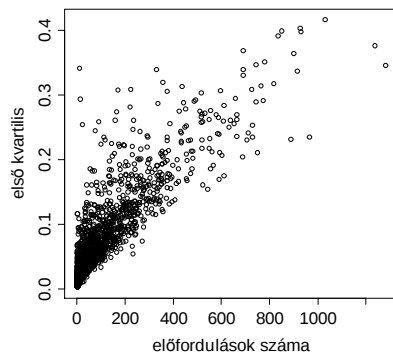
dc_16_10



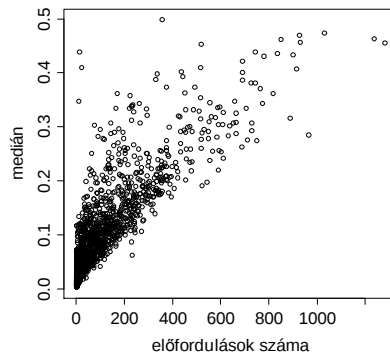
a)



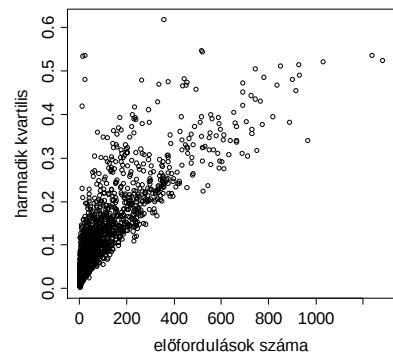
b)



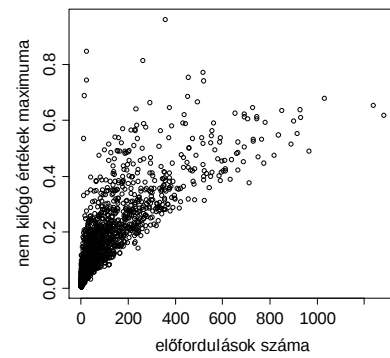
c)



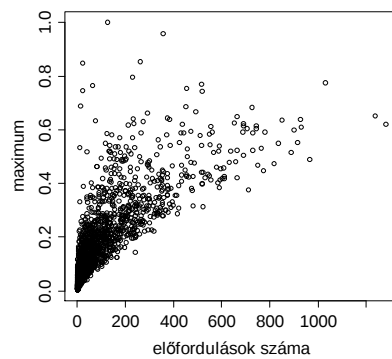
d)



e)

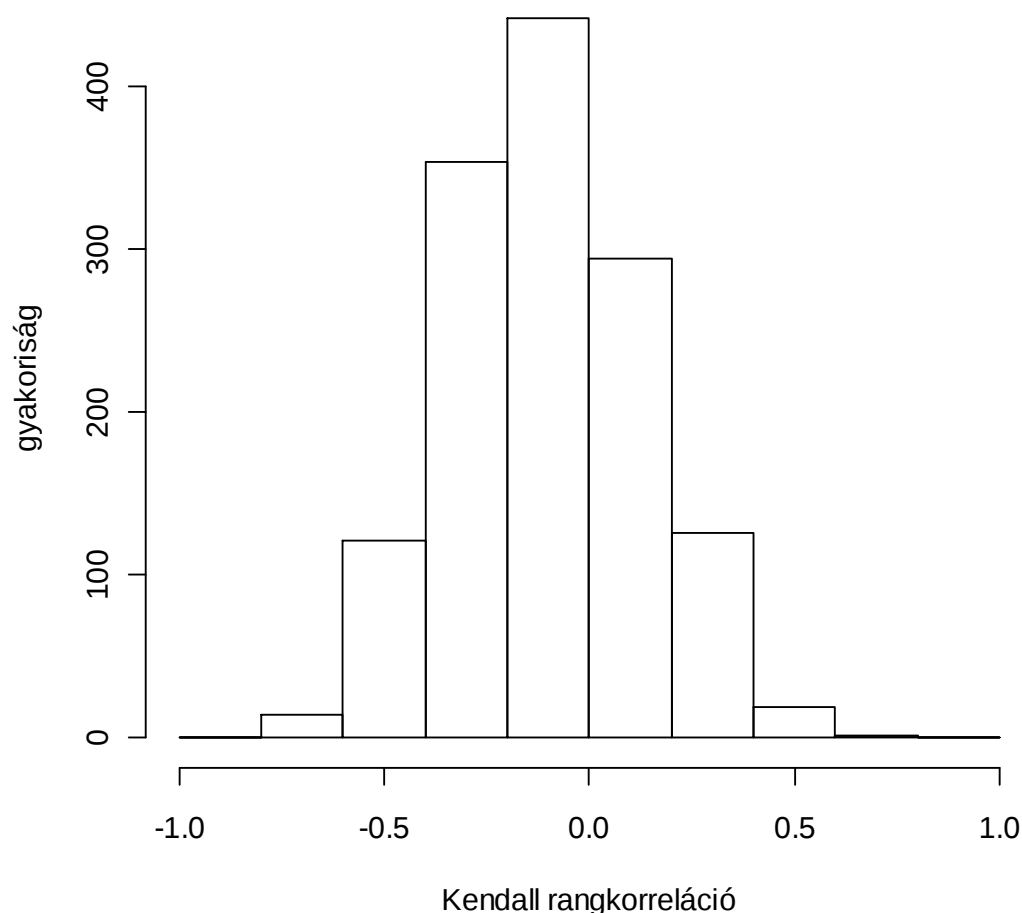


f)



g)

33. ábra: A faj előfordulásaihoz tartozó Beals simított értékek eloszlásának jellemzői a faj előfordulási gyakoriságának függvényében



34. ábra: Az előfordulásokhoz tartozó Beals simított értékek és a felvétel fajsza közötti fajon belüli Kendall rangkorrelációs koefficiensek eloszlása

7.5. *Diszkusszió*

A nagy adatbázison végzett elemzés megerősítette Münzbergová és Herben (2004) eredményeit, hogy fajonként eltérő határértéket kell alkalmazni a bináris skálára való átalakításkor. Ha az a célunk, hogy minden előforduló fajt a fajkészlet részének tekintsünk, akkor a faj előfordulásaihoz tartozó simított értékek minimumát kellene határértéknek választani. Azonban Münzbergová és Herben (2004) is rámutattak, hogy esetenként egy vagy néhány kilógó érték miatt a minimum nagyon alacsony lehet. Ezeknek a kilógó értékeknek a hátterében állhat adatbázis-hiba, vagy olyan szituáció, amikor a korábbi állapot alkalmas volt a fajnak, a jelenlegi már nem, de a faj néhány egyede még képes túlélni. Ha a minimumot használjuk határértéknek, akkor a kilógó értékekkel rendelkező fajoknál nagyon magas lehet a másodfajú hiba valószínűsége. Ezért célszerű a kilógó értékek kizárása utáni minimumot használni.

Egy fajon belül az előfordulásokhoz tartozó simított értékek és a felvételek fajsza között nem találtam általános érvényű kapcsolatot.

8. Niche szélesség

Habár a „specialista” és „generalista” fajok széles körben használt fogalmak, mégis a „specialistaság”-nak elfogadott mérőszáma. Emiatt a fajok kategorizálása általában szubjektív módon történik, jellemzően két kategóriát használva: specialisták és generalisták. Egy folytonos skálájú, objektív mérőszám sokkal részletesebb képet adna a fajok viselkedéséről, és az így nyert információkat számos ökológiai kérdés vizsgálatában fel lehetne használni.

Ha rendelkezünk környezeti adatokkal azokról a helyekről ahol a faj előfordul, megbecsülhetjük a nicheszélességet. Ez azonban nem használható a „specialistaság” egyértelmű mérőszámaként, mert egy faj niche-e lehet keskeny az egyik környezeti tényező tekintetében, de széles, ha egy másik változót vizsgálunk.

A vegetációs adatbázisokban rejlő információk kiaknázása megoldást kínál erre a problémára. A Fridley és munkatársai (2007) által javasolt módszer alap gondolata egyszerű: a generalista faj sokféle környezetben, változatos vegetációtípusokban fordul elő, míg a specialista faj csak szűk környezeti tartományban él, ezért kevésbé változatosak azok a felvételek, amelyekben megtaláljuk. Vagyis a faj „specialistaság”-ának megmérésehez azt kell megvizsgálni, hogy mennyire különböző összetételűek azok a felvételek, amelyekben előfordul, vagyis mekkora a béta diverzitás. Az így kapott értéket egy minden releváns környezeti tényezőt együttesen figyelembe vevő (realizált) nicheszélességnek is tekinthetjük (és a továbbiakban annak is hívom), annak ellenére, hogy a számítások során nem használunk környezeti adatokat.

Mivel a szakirodalomban számos béta-diverzitás indexet találhatunk (lásd pl. Koleff et al. 2003, Tuomisto 2010 áttekintő cikkeit), nem triviális a kérdés melyik a legalkalmasabb a nicheszélesség mérésére. Fridley és munkatársai (2007) a gamma diverzitás alfa- és béta-diverzitás komponensre történő additív felosztását használják (Veech et al. 2002), ahol a béta diverzitás a felvételekben előforduló fajok számának (összfajszám, gamma diverzitás) és az egyes felvételek átlagos fajszámának (alfa diverzitás) a különbsége. Zeleny (2009) azonban kimutatta, hogy az így kapott becslés torzított, felülbecsli a nicheszélességet a nagy fajkészlettel rendelkező termőhelyek fajainál. Ezért a Whittaker-féle béta diverzitás (Whittaker 1960) használatát javasolta, amely kiküszöböli ezt a problémát, ha a lokális fajgazdagság és a fajkészlet mérete között lineáris kapcsolat van. Manthey és Fridley (2009) egy másik torzítási lehetőségre hívták fel a figyelmet: ha a faj általában fajszegény közösségekben fordul elő, de néha előfordul fajgazdag helyeken is, vagyis ha a fajszámok eloszlása erősen jobbra ferde, túlbecsülhetjük a nicheszélességét, különösen a Whittaker féle béta diverzitást használva. Négy különböző béta diverzitás indexeket (lásd később) összehasonlítva a kétféle torzítás szempontjából, arra a következtetésre jutottak, hogy ha a fajkészlet és a lokális fajszám között lineáris összefüggés van, akkor az átlagos Jaccard különbözőség, vagy a többszörös (multiple) Simpson hasonlóság (Baselga et al. 2007) komplementere adja a legjobb eredményt. Nem találtak viszont ideális megoldást, ha a fajkészlet mérete és a lokális fajszám között nem-lineáris összefüggés van.

A fejezet további részében először bemutatom a nicheszélesség mérésére javasolt négy béta diverzitás függvényt. Ezután a távolság alapú béta diverzitás képletek egy kedvezőtlen tulajdonságával foglalkozom: a nagy nicheszélesség esetén alulbecslik azt. Az ezután következő fejezetben újragondolom Zeleny (2009) érvelését, és ez alapján új módszert javaslok a nicheszélesség mérésére, amelynek eredménye akkor is független a fajkészlet méretétől, ha a lokális fajszám és a fajkészlet mérete közötti kapcsolat nem-lineáris. A következő fejezetben a fajszám erősen jobbra ferde eloszlásából adódó felülbecslés problémájának megoldásával foglalkozom, végül a CoenoDat adatbázis elemzésével megvizsgálom, hogy a módszerválasztás során hozott döntéseink közül melyiknek mekkora hatása van a végeredményre.

8.1. A nicheszélesség mérésére javasolt béta diverzitás indexek áttekintése

Az alfa-, béta- és gamma-diverzitás fogalmát Whittaker (1960) vezette be: az alfa-diverzitás egy lokalitás (egy felvétel) diverzitása, a gamma-diverzitás egy terület diverzitása, míg a béta-diverzitás a területen belüli lokalitások közötti különbségeket számszerűsíti. Az alfa- és gamma-diverzitás mérésére ugyanazok a diverzitás függvények használhatók. A szakirodalomban számos ilyen függvény található (Magurran 2004), amelyek a ritka fajok súlyozásában különböznek egymástól (Tóthmérész 1995, Jost 2007). A nicheszélesség becslése bináris adatokon alapul, ezért az alfa- és a gamma-diverzitás mérésére a fajszám használható.

Míg az alfa- és a gamma-diverzitás esetén a bináris adatok esetén egyértelmű, hogy a diverzitás mérésére a fajszám szolgál, a béta-diverzitás számszerűsítésére ebben az esetben is számos lehetőség kínálkozik (Koleff et al. 2003, Tuomisto 2010). Ezek egy része azon az elképzelésen alapul, hogy a gamma-diverzitást (vagyis a terület össz fajszáma) az alfa-diverzitás (vagyis az átlagos lokális fajszám) és a béta-diverzitás értékét kombinálva kapjuk meg. Az egyik lehetőség, hogy a gamma-diverzitás (γ) az alfa- és a béta-diverzitás szorzata: $\gamma = \alpha\beta$, amiből adódik a béta-diverzitás kiszámítási módja: $\beta = \gamma/\alpha$. Ezt az összefüggést először Whittaker (1960) javasolta a béta-diverzitás mérésére, és a szakirodalomban Whittaker-féle béta-diverzitásként ismert, habár Whittaker ugyanabban a cikkben még három másik módszert is javasolt a béta-diverzitás kiszámítására. Jost (2007) bebizonyította, hogy csak a gamma-diverzitás multiplikatív felbontása vezet független alfa- és béta-diverzitás komponensre.

Lande (1996), valamint Veech és munkatársai (2002) a gamma-diverzitás additív felbontása – $\gamma = \alpha + \beta$ és $\beta = \gamma - \alpha$ – mellett érveltek, mert így mind a három diverzitás mértékegysége ugyanaz, esetünkben fajszám. Tuomisto (2010) szerint viszont bármilyen diverzitás mértékegysége a típusok effektív száma (effective number of types) kellene, hogy legyen: az alfa- és a gamma-diverzitásé a fajok effektív száma, a béta-diverzitásé a felvételek effektív száma. A fajok effektív száma azt a fajszámot jelenti, amelynél a diverzitás ugyanekkora lenne, ha a fajok abundanciái azonosak lennének (Jost 2006, 2007), funkcionális vagy filogenetikai diverzitás esetén kiegészítve azzal, hogy „és a fajok maximálisan

különböznie” (Ricotta & Szeidl 2009). A felvételek effektív száma a teljesen különböző felvételek azon száma, amely a megfigyelttel azonos béta-diverzitást eredményezne (Jost 2007). Emiatt Tuomisto (2010) szerint csak a Whittaker-féle béta diverzitás tekinthető valódi diverzitásnak, mert annak a mértékegysége a típusok effektív száma.

A béta-diverzitás a felvételek fajösszetétele közötti különbséget számszerűsíti, ezért a szakirodalomban számos olyan mérőszám ismert, amely ebből a jelentésből indul ki és nem a gamma-diverzitás felbontásán alapul (már Whittaker 1960 cikkében is találunk ilyeneket). Két felvétel fajösszetétele közötti különbség mérésére számos távolságfüggvény ismert (Podani 1997, Legendre & Legendre 1998). Több felvétel esetén ezek átlagát használhatjuk; Manthey és Fridley (2009) például az átlagos Jaccard különbséget használta a nicheszélesség mérésekor.

A távolságmátrix elemei természetesen nem függetlenek egymástól, és bár véleményem szerint addig ez nem okoz problémát, amíg nem végzünk az átlagra vonatkozó statisztikai próbát, a nem független értékek átlagolásából adódó „probléma” kiküszöbölésére Diserud és Ødegaard (2007), valamint Baselga és munkatársai (2007) a többszörös (multiple-site) hasonlóságok használatát javasolta. A többszörös hasonlóságok komplementere (1-hasonlóság), a többszörös különbség szintén felhasználható a felvételek közötti különbségek számszerűsítésére. Manthey és Fridley (2009) a többszörös Simpson hasonlóságot (Baselga et al. 2007) használta, mivel rögzített felvételszám esetén a többszörös Sørensen hasonlóság (Diserud & Ødegaard 2007) a Whittaker féle béta-diverzitás függvénye.

Két felvételre a Simpson hasonlóság:

$$S_{Sim} = \frac{a}{\min(b, c) + a}$$

ahol:

a = a két felvétel közös fajainak száma,

b = csak az első felvételben előforduló fajok száma

c = csak a második felvételben előforduló fajok száma

Baselga és munkatársai (2007) ezt több felvételre a következő módon általánosították:

$$MS_{Sim} = \frac{(\sum_i S_i) - S_T}{[\sum_{i < j} (\min(S_i, S_j) - a_{ij})] - [(\sum_i S_i) - S_T]}$$

ahol:

S_i = az i felvétel fajszáma

S_T = az össz fajszám

a_{ij} = az i és a j felvételek közös fajainak száma

8.2. A távolság alapú béta-diverzitás indexek hátrányai

A távolság alapú béta-diverzitás indexek két nagy csoportba sorolhatók: átlagos páronkénti távolságok és többszörös (multiple) különbség indexek. Az első csoportba tartozó indexek kiszámításakor bármilyen, ökológiai problémák vizsgálatára alkalmas távolságfüggvény használható (ezekről magyar nyelven Podani 1997 könyve ad jó

áttekintést), nem csak a Manthey and Fridley (2009) által használt Jaccard különbség. A bináris különbségképletek legtöbbje 0 és 1 közötti értékeket vehet fel, és a különbség maximális, ha nincs közös faj a két összehasonlított fajlistának. Vagyis ha két felvétel termőhelye már annyira különböző, hogy – a vizsgált fajon kívül – nincs közös fajuk, a fajlisták különbsége nem informál a termőhelyi különbségekről.

Első pillantásra ez egy ritka esetnek és ezért elhanyagolható problémának tűnik, de a valóságban nem az. Például a CoenoDat adatbázis 8437 felvételét megvizsgálva a 784 olyan faj közül, ami legalább 50 felvételben előfordul, 774 fajnál találtam olyan felvételpárt, amelyben a vizsgált faj volt az egyetlen közös faj.

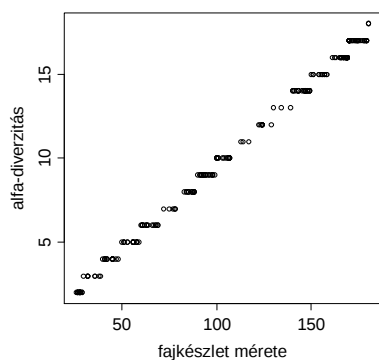
Az, hogy a felvételek különbsége egy határon túl már nem nő, annak ellenére, hogy a termőhelyek különbsége igen, az ordinációkban a „patkó-jelenség” megjelenéséhez vezet (Podani 1997). Ennek kiküszöbölésére az egyik megoldás, hogy azoknál a felvételpároknál, amelyeknek nincs közös faj, megkeresik azt a legrövidebb útvonalat a két felvétel között, amelynél a szomszédos felvételeknek van közös faj, és a különbséget ennek az útvonalnak a hosszával helyettesítik (shortest path adjustment: Williamson 1978, Bradfield & Kenkel 1987, extended dissimilarity: De'ath 1999). Elvileg az így korrigált értékek átlaga is használható lenne béta-diverzitásként, de ilyen alkalmazásukról nincs tudomásom.

A páros különbségek kapcsán bemutatott probléma egy általánosabb probléma speciális esete. Általában is igaz, hogy a fajkompozíciós különbségeknek (így a többszörös különbségeknek is) van egy véges elvi maximuma, ami felett már érzéketlenek a környezeti különbségekre.

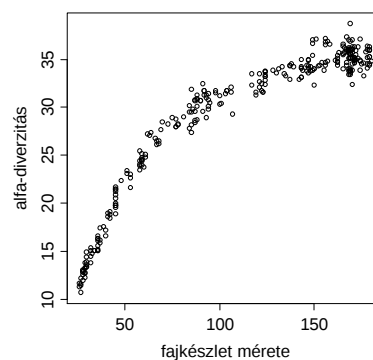
8.3. A fajkészlet méretének hatása a béta-diverzitásra

Az, hogy a lokális fajgazdagságot jelentősen befolyásolhatja a fajkészlet mérete, evidenciának számít az ökológiában (Zobel 1997). Az összefüggés alakja, amely lehet lineáris és telítődési görbe jellegű (35.a. b. ábra), ami alapján megkülönböztethetünk telített és telítetlen közösségeket. Az elnevezések abból a feltételezésből erednek, hogy a telítetlen közösségekben az összefüggés azért lineáris, mert a fajgazdagságot nem a lokális interakciók, hanem csak a fajkészlet határozza meg, míg a telített közösségekben a lokális interakcióké a döntő szerep (Cornell & Lawton 1992). A későbbi szimulációs vizsgálatok (pl. Herben 2000, Bartha & Ittész 2001, Hillebrand 2005, Fox 2006) azonban kimutatták, hogy az összefüggés alakjából nem lehet következtetni a fajgazdagságot meghatározó mechanizmusokra: mivel az akkor is lehet nemlineáris, ha egyáltalán nincsenek lokális interakciók a fajok között (Fox 2006), míg más esetben az erős kompetíció ellenére lineáris marad (Hillebrand 2005). A továbbiakban a telített/telítetlen kifejezéspárt csak a görbe lineáris/nemlineáris alakjának szinonímájaként használom, anélkül, hogy a lokális fajgazdagságot meghatározó mechanizmusra utalnék vele.

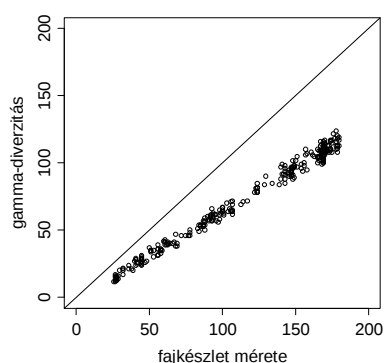
Zeleny (2009) szerint telítetlen közösségek esetén a Whittaker-féle béta diverzitást alkalmazva kiszűrhető a fajkészlet méretének torzító hatása. Az erre vonatkozó levezetésben szinoním fogalomként használja a fajkészlet méretét és a gamma diverzitást, holott ezek jelentése csak hasonló, de nem azonos. A fajkészlet mérete alatt (ahogy azt Zeleny is helye-



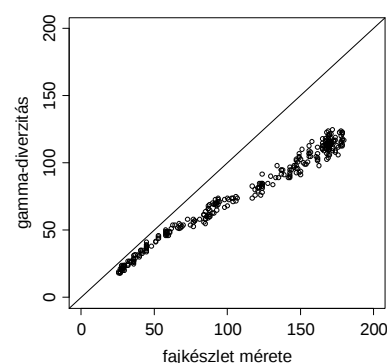
a)



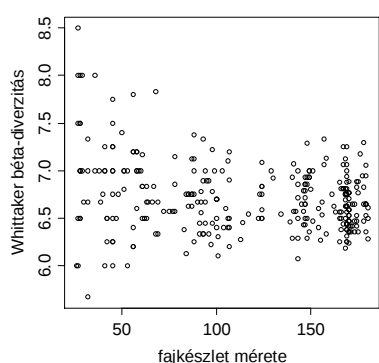
b)



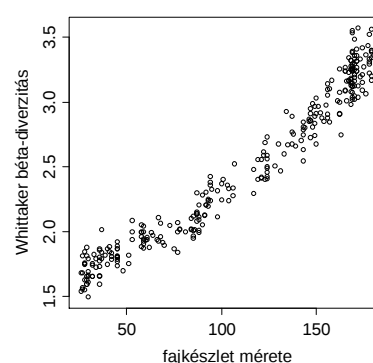
c)



d)



e)



f)

35. ábra: A fajkészlet mérete és az alfa-diverzitás (a,b), a gamma-diverzitás (c,d) és a Whittaker-féle béta-diverzitás (e,f) közötti összefüggés szimulált telített (bal oldali oszlop) és telítetlen (jobb oldali oszlop)

sen írja cikkében a 11. oldalon) azoknak a fajoknak a számát értjük, amelyek az adott helyen előfordulhatnak, míg a gamma diverzitás azoknak a fajoknak a száma, amelyek a vizsgált felvételek közül legalább egyben előfordulnak. Ha minden felvétel azonos környezetben készül, akkor a felvételszám emelkedésével a gamma diverzitás egyre jobban közelít a lokális fajkészlet méretéhez, de csak igen magas felvételszámnál éri azt el (v.ö. Palmer 1990). Zeleny (2009) a képletekben gamma-diverzitást használ, de azok szöveges értelmezésekor már a fajkészlet méretéről beszél. A két fogalom közti jelentése közötti különbség miatt nem mondhatjuk azt, hogy elméleti levezetése igazolta, hogy a Whittaker-féle béta-diverzitás használatával kiküszöbölhető a fajkészlet méretének torzító hatása. Ez nem jelenti azt, hogy

az állítás nem igaz, csak annyit, hogy a bizonyítás nem állja meg a helyét.

Zeleny (2009) azonban szimulált és valós adatok elemzésével is demonstrálta állítása valódiságát. Megismételve szimulációit megállapítható, hogy az általa vizsgált telítetlen közösségekben a gamma-diverzitás lineáris függvénye a fajkészlet méretének (35.c. ábra), míg a telített közösségekben ez az összefüggés nem lineáris (35.d. ábra). Jelöljük a továbbiakban S -sel a fajkészlet méretét, α -val a lokális fajgazdagságot és γ -val a gamma-diverzitást. Telítetlen közösségekben adott nicheszélesség esetén $\alpha=k_1S$ és $\gamma=k_2S$, így a Whittaker-féle béta-diverzitás értéke S -től független konstans: $\beta_w=\gamma/\alpha=k_1S/k_2S=k_1/k_2=k$. Ha azonban az alfa- és/vagy a gamma-diverzitás bármilyen más alakú függvénye a fajkészlet méretének – $\alpha=f(S)$ és $\gamma=h(S)$ – akkor a béta-diverzitás is S függvényében fog változni: $\beta_w=\gamma/\alpha=f(S)/h(S)$ (35.e.,f. ábra). Hangsúlyozandó, hogy a béta-diverzitás csak akkor lesz független a fajkészlet méretétől, ha a tengelymetszet értéke mindkét lineáris összefüggés esetében nulla. Habár a fajkészlet meghatározásának nehézségei miatt kevés az erre vonatkozó adat, ez a feltétel valós adatokra valószínűleg csak ritkán teljesül maradéktalanul.

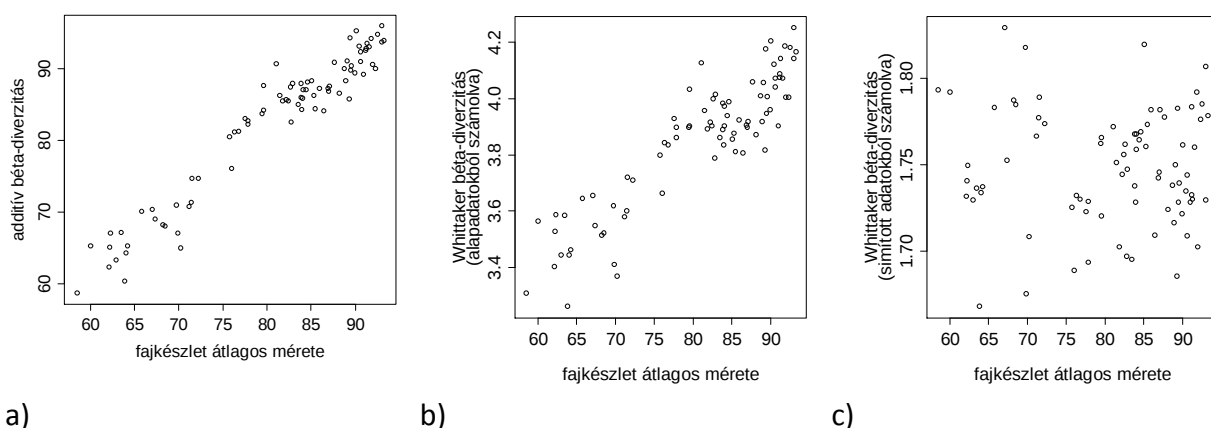
Ennek ellenére Zeleny (2009) vizsgálatában a Whittaker-féle béta-diverzitás használata számottevően csökkentette a fajkészlet méretének torzító hatását a valós adatok vizsgálatakor is. A fajkészlet mérete részben az alfa- részben a gamma-diverzitáson keresztül hat a béta diverzitásra. Jost (2007) kimutatta, hogy a Whittaker-féle béta diverzitás értéke független a lokális fajgazdagságtól, míg az additív béta diverzitásé (absolute effective species turnover sensu Tuomisto 2010) nem. Vagyis a Whittaker-féle béta-diverzitás alkalmazása mindig csökkenti a fajkészlet méretének torzító hatását kiiktatva az egyik utat, amin keresztül az hat, de teljesen csak a fenti szigorú feltételek teljesülése esetén szünteti meg.

A fajkészlet méretének hatása úgy küszöbölhető ki teljesen, ha a megfigyelt fajlisták helyett a lokális fajkészletet használjuk a Whittaker-féle béta-diverzitás kiszámításakor. Ebben az esetben az alfa-diverzitás a lokális fajkészlet mérete, míg a gamma-diverzitás azoknak a fajoknak a száma amelyek a környezeti változók vizsgált tartományának legalább egy pontján a fajkészlet részei. A gamma-diverzitás multiplikatív felbontása, azaz a Whittaker-féle béta-diverzitás használata, garantálja, hogy a béta diverzitás minden esetben független lesz az alfa-diverzitástól (Jost 2007), vagyis a lokális fajkészlet méretétől. Az így kiszámolt nicheszélesség további előnye, hogy a kapott értékek könnyen értelmezhetők. Az elméleti minimum 1, ami azt jelenti, hogy a fajkészlet azonos a vizsgált faj minden előfordulási helyén, vagyis a faj extrém specialista. Ha a faj nicheszélessége 2, akkor az azt jelenti, hogy épp annyira generalista, mint egy olyan faj, amely két olyan termőhelyen képes előfordulni, amelyek fajkészlete – a vizsgált fajtól eltekintve – nem fed át.

A lokális fajkészlet becslésének módszerével az 7. fejezet foglalkozott részletesen, így arra itt nem térek ki. Mivel a béta-diverzitás értékét befolyásolja az is, hogy hány felvételre számoltuk ki, a Fridley és munkatársai (2007) által javasolt módon kell eljárunk: minden faj esetében ugyanannyi felvételt választunk random, majd ezekre számolunk béta-diverzitást. A véletlen hatását csökkentendő a random kiválasztást sokszor megismételjük és a kapott eredményeket átlagoljuk. Azokra a fajokra, amelyek kevesebb felvételben fordulnak elő, mint az előre megállapított mintaméret, az elemzést nem tudjuk elvégezni. A mintaméret

megállapításakor arra kell tekintettel lennünk, hogy a nagyobb mintaméret pontosabb becslést ad, de ugyanakkor a fajok szűkebb körére végezhető el a számolás.

Annak illusztrálására, hogy az ilyen módon számolt nicheszélesség független a fajkészlet méretétől, szimulált cönoklin adatokat használtam, ahol minden faj nicheszélessége azonos (a vizsgált gradiens hosszának a fele), de a fajkészlet mérete erősen változik a gradiens mentén. A Fridley és munkatársai (2007) által használt scenáriók közül a legegyszerűbbet választottam, ahol a fajok válaszgörbéi szimmetrikusak és a felvételek egyenletesen vannak elosztva a gradiens mentén, mert az volt a célom, hogy a nicheszélesség mérőszámok viselkedését optimális mintavételi körülmények között teszteljem. Zeleny (2009) javaslatait követve kissé megváltoztattam a szimuláció paramétereit: a fajszámot 50-ről 300-ra, a felvételek számát 500-ról 3000-re emeltem, és a fajok előfordulási optimumait nem egyenletesen szórtam ki a gradiens mentén, hogy a fajkészlet méretében minél jelentősebb különbségek legyenek. A gradiens egy pontján a fajkészlet mérete azoknak a fajoknak a száma, amelyek előfordulási valószínűsége nem nulla. Ezeknek az értékeknek az átlaga azokra a pontokra, ahol a faj előfordulhat, szerepel az ábrákon a fajkészlet átlagos méreteként. A szimulációt és a számításokat az R programban végeztem el (R Development Core Team 2010).



- a) b) c)
36. ábra A becsült nicheszélesség és a fajkészlet átlagos méretének kapcsolata szimulált adatokban, ahol a valós nicheszélesség minden fajra azonos. A szimuláció során a Fridley és munkatársai (2007) által javasolt algoritmust használtam, így a fajkészlet mérete és a lokális fajgazdagság közötti kapcsolat nem-lineáris
- (a) A Fridley és munkatársai (2007) által javasolt módon additív béta-diverzitást számolva a becsült nicheszélességek és a fajkészlet mérete között nagyon erős pozitív korreláció van ($r=0.967$, $p<0.1\%$).
- (b) A Zeleny (2009) által javasolt módon Whittaker-féle béta-diverzitást használva a korreláció gyengébb, de még mindig nagyon erős ($r=0.899$, $p<0.1\%$).
- (c) Az itt javasolt módon becsülve a nicheszélességet az független a fajkészlet méretétől ($r=-0.035$, $p=75.5\%$).

A 36. ábrán látható, hogy Manthey és Fridley (2009) eredményeivel összhangban, a telítetlen közösségek esetén mind az additív béta-diverzitás, mind az alapadatokból számolt Whittaker-féle béta diverzitás pozitívan korrelál fajkészlet átlagos méretével. A fentiekben

javasolt új algoritmussal becsült nicheszélesség viszont függetlennek bizonyult a fajkészlet méretétől.

8.4. A nicheszélesség robosztus becslése

Manthey és Fridley (2009) legfontosabb ellenérve a Whittaker-féle béta-diverzitás alkalmazásával szemben, hogy egyes fajok nicheszélességét ez a módszer alaposan túlbecsli. Példájuk a *Taxodium ascendens*, amely általában a fajszegény mocsárerdőkben él, de néha előfordul a fajgazdag ártéri erdőkben is. Akár egyetlen felvétel a fajgazdag élőhelyről jelentősen megnöveli a gamma-diverzitást, de alig változtatja meg az átlagos lokális fajszámot, így a Whittaker-féle béta-diverzitás értéke nagyon érzékeny lesz arra, hogy az az egy felvétel bekerült-e a mintába. Más szavakkal azt is mondhatnánk, hogy a Whittaker-féle béta-diverzitás nem elég robosztus. Manthey és Fridley (2009) szerint ez a probléma leginkább akkor jelentkezik, ha egy fajszegény élőhelyre jellemző faj néha előfordul a szomszédos fajgazdag élőhelyeken is. Azonban a nem robosztus becslés elvileg bármely faj esetén problémát okozhat.

Ennek szemléltetésére nézzünk egy egyszerű példát! Tegyük fel, hogy a vizsgált faj az A élőhelyre jellemző, de ritkán előfordul a B élőhelyen is. A fajkészlet mérete (az alfa-diverzitás) a két élőhelyen azonos, de a vizsgált fajtól eltekintve nem fed át. A nicheszélesség becsléséhez random választunk N olyan felvételt, amelyben a faj előfordul, és kiszámítjuk a Whittaker-féle béta-diverzitást erre a mintára. Ha az összes felvétel az A élőhelyről származik (ez lehetséges, hiszen a faj a másik élőhelyen csak ritkán fordul elő), akkor a béta-diverzitás értéke 1 lesz, hiszen a fajkészlet minden felvételben azonos. Ha viszont egy felvétel a B élőhelyről származik, és csak a maradék $N-1$ felvétel származik az A élőhelyről, a béta-diverzitás értéke (a vizsgált fajt kizárva a számításból) kettő lesz. A különbség még nagyobb, ha a B élőhely fajkészlete nagyobb, mint az A élőhelyé, viszont valamivel kisebb, ha a két fajkészlet átfed. A CoenoDat adatbázison végzett nicheszélesség becslések esetén (8.5. fejezet) a kapott értékek 1,59 és 4,41 között voltak, vagyis a példabeli értékek közötti különbség jelentős.

A megoldás a kilógó felvételek kizárása az elemzésből. A kilógó értékek kiszűrésére a sokváltozós adatokból McCune és Mefford (1999) javasolt egy távolságokon alapuló algoritmust. Először, minden felvételnek kiszámítjuk az átlagos távolságát a többi felvételtől. Kilógónak azokat a felvételeket tekintjük, ahol ez az érték legalább kétszórásnyival nagyobb, mint az átlagos érték. Az eljárás tetszőleges távolságfüggvénnyel elvégezhető. Ebben az esetben a becsült fajkészlet alapján számolt bináris euklideszi távolság használatát javaslom. Előnye az általánosan használt bináris különbözőségekkal (pl. Jaccard, Sørensen) szemben, hogy a felvételek fajszámát is figyelembe veszi.

8.5. A módszerválasztás hatása az eredményekre

A különböző módszerek összehasonlításához a CoenoDat adatbázis (2. fejezet) 8437 felvételét használtam. Az elemzésekhez az adatbázisból vett minták mérete 50 felvétel volt, így a legalább 50 felvételben előforduló 710 növényfaj nicheszélességét becsültem négy

különböző módon: additív béta diverzitással, Whittaker-féle béta-diverzitással az alapadatokból számolva, Whittaker-féle béta-diverzitással a becsült fajkészletből számolva és Whittaker-féle béta-diverzitással a becsült fajkészletből számolva, kizárva a kilógó felvételeket. A négy módszerrel kapott értékek között lineáris korrelációt számoltam.

18. táblázat: A hazai fajokra a CoenoDat adatbázis 8437 felvétele alapján négy különböző módon becsült nicheszélesség közötti korrelációk

	Additív béta- diverzitás	Whittaker-féle béta-diverzitás (alapadatok)	Whittaker-féle béta-diverzitás (fajkészlet)
Additív béta-diverzitás	1		
Whittaker-féle béta-diverzitás (alapadatok)	0.413	1	
Whittaker-féle béta-diverzitás (fajkészlet)	0.324	0.864	1
Whittaker-féle béta-diverzitás (fajkészlet, kilógó felvételek kizárva)	0.378	0.873	0.925

A legfontosabb döntésnek a béta-diverzitás típusa bizonyult, az additív béta-diverzitással kapott értékek erősen eltérnek a másik három módszerrel kapottaktól (18. táblázat). A Beals simítással becsült fajkészletet használata az alapadatok helyett kisebb, de még mindig nem elhanyagolható különbséget okozott. A kilógó felvételek kizárásának hatása összességében csekély, azonban jelentős lehet egy-egy faj esetében. Például a kilógó felvételek kizárása nyomán a *Chara fragilis* becsült nicheszélessége 3,39-ről 1,92-re csökkent. (Összehasonlításként a vizsgált fajok nicheszélességei 1.59 és 4.41 között változtak.) A *Chara fragilis* elsősorban különböző hínárközösségekben fordul elő, de megtalálható egy láperdei (*Aegopodio – Alnetum*) felvételen is, amelynek fajkészlete a hínártársulásokétól nagyon eltérő és azokénál jóval nagyobb is. Ennek a kilógó felvételnek a figyelembe vétele nagyon megnövelte a nicheszélességre kapott becslést, hasonlóan a Manthey and Fridley (2009) által bemutatott példához.

8.6. Következtetések

Fridley és munkatársai (2007) zseniális ötlete a nicheszélesség becslésére a cönológiai adatbázisokban rejlő információkból hamar népszerűvé vált az ökológusok közt, és máris számos alkalmazása született: például a nicheszélesség és az elterjedési terület mérete (Lambdon 2008) vagy a ritkaság (Rannap et al. 2009) közötti kapcsolat vizsgálatától a bolygatás és fragmentáció okozta változások elemzéséig (Devictor et al. 2008, Devictor & Robert 2009, Clavero & Brotons 2010).

Sajnos az eredeti módszer (Fridley et al. 2007) – amit az alkalmazások mindegyikében használtak – torzított becslést ad, amit a Zeleny (2009) által javasolt módosítás is csak bizonyos esetekben küszöböl ki. Ebben a fejezetben bemutatam egy algoritmust, ami

torzítatlan becslésre vezet és robusztus a kilógó felvételekkel szemben. Az eredeti és az új módszerrel kapott eredmények között jelentős, akár az eredmények interpretációját is befolyásoló különbségek lehetnek, különösen ha egyes fajok viselkedését is értelmezzük.

9. Mocsárrétek klasszifikációja

9.1. Bevezetés

A 20. század folyamán a cönológiai tanulmányok eredményeképpen jelentős mennyiségű ismeret halmozódott fel a növényzet változatosságáról, amelynek egyik leképeződése a vegetáció osztályozása. A tanulmányok többsége azonban kisebb területekre vonatkozik, és sokáig a szintetizáló munkák is döntő többsége is megmaradt a nemzeti keretek között (Mucina et al. 1993). A European Vegetation Survey nevű nemzetközi munkacsoport fő célkitűzése egy európai léptékű szintézis megalkotása, amelynek első lépése az országos léptékű szintézisek összeszerkesztése volt (Mucina 1997a, Rodwell et al. 2002). Pusztán az irodalmak feldolgozásával azonban nem lehet teljesen feloldani a különböző rendszerek közötti ellenmondásokat és különbségeket. Ehhez nagyobb területekről származó eredeti felvételek újraelemzésére is szükség van (Bruehlheide & Chytrý 2000).

A nehezen összeegyeztethető nemzeti vegetációosztályozási rendszerekre jó példa a közép-európai síkságok és dombvidékek nedves rétjeinek osztályozása. Míg a szubóceánikus típusú nedves rétek osztályozása többé-kevésbé stabil, és olyan nemzetközileg elfogadott asszociációcsoportokat használ, mint például *Arrhenatherion elatioris*, *Polygono-Trisetion*, *Calthion palustris* és *Molinion coeruleae* (Ellmauer & Mucina 1993, Oberdorfer 1993, Dierschke 1995, Pott 1995, Kučera & Šumberová 2001, Schubert et al. 2001, Stanová & Valachovič 2002, Burkart et al. 2004, Havlová et al. 2004), a szubóceánikus-szubkontinentális elterjedésű típusok esetén nem jött létre ilyen konszenzus. Ezeket a réteket általában a következő asszociációcsoportokba sorolják: *Agrostion albae*, *Alopecurion pratensis*, *Cnidion venosi*, *Deschampsion cespitosae* és *Veronico longifoliae-Lysimachion vulgaris*. Még ha a használt név meg is egyezik a különböző szerzőknél, az asszociációcsoportok pontos jelentése országonként és szerzőnként változó. Ritka fajok élőhelyeiként (Hölzel 2003) ezek a gyepek jelentős természetvédelmi értéket képviselnek, ugyanakkor a megváltozó tájhasználat (Joyce & Wade 1999, Šeffer & Stanová 1999) és a terjedő inváziós fajok (Botta-Dukát 2008b) erősen veszélyeztetik a fennmaradásukat. Ezért az élőhely szerepel a Natura2000 területek kijelölésére szolgáló Európai Unió Élőhely Direktíva élőhely listáján: „6440 Alluvial meadows of river valleys of the *Cnidion dubii*” néven (Šeffer et al. 2008). Az Európai Unió szintű jogi szabályozás még inkább időszzerűvé tette egy közép-európai léptékű osztályozás létrehozását.

A fejezet célja egy olyan esettanulmány (Botta-Dukát et al. 2005) bemutatása, amely (1) feltárja a sík- és dombvidéki nedves rétek fajkompozíciós változatosságát egy kontinentalitási gradiens mentén Csehországtól, Kelet-Ausztrián, Szlovákián és Magyarországon át, Északkelet-Horvátországig; (2) a felvételek elemzésével azonosítja a

főbb típusokat; és (3) bemutatja ezek szüntaxonómiai értelmezését. Az esettanulmány egyben jó példa a cönológiai adatbázisok egyik tipikus alkalmazására. Az eredmények az alap kutatáson túl a gyakorlat igényeit is szolgálják, például az itt bemutatott cikkben (Botta-Dukát et al. 2005) kidolgozott osztályozás megjelent a Natura2000-es kezelési tervekben is (Šeffner et al. 2008).

9.2. Anyag és módszer

9.2.1. Vegetációadatok

Részben nemzeti adatbázisokból (pl. Valachovič 1999, Chytrý & Rafajová 2003), részben az irodalomból gyűjtöttük össze az alföldi nedves rétek felvételeit egy Közép-Európán keresztül húzódó északnyugat-délkeleti gradiens mentén. Azokat a Csehországban, Kelet-Ausztriában, Szlovákiában, Magyarországon és Észak-Horvátországban, 350 m-es tengerszint feletti magasság alatti területeken készült felvételeket választottuk ki, amelyeket a készítőik a Molinietalia rendbe soroltak. Kizártuk az elemzésből a szokatlanul kicsi (4 m²-nél kisebb) vagy nagy (100 m²-nél nagyobb) felvételeket, és azokat, amelyek 10°-nál nagyobb lejtésű területen készültek. Az egyes területek túlreprezentáltságából adódó esetleges problémák kiküszöbölésére rétegzett újramintavételezést alkalmaztunk (Knollová et al. 2005). Amennyiben ugyanarról a lokalitásról (esetünkben a legközelebbi település nevét tekintve lokalitásnak) egy asszociációból vagy szubasszociációból több felvétel is rendelkezésre állt, azokból random választottunk egyet az elemzéshez. Így összesen 387 felvételt választottunk ki, 178-at Csehországból, 11-et Kelet-Ausztriából, 115-öt Szlovákiából, 76-ot Magyarországról és 7-et Észak-Horvátországból. A szórványosan előforduló moha adatokat töröltük, mert a legtöbb felvételben a mohák nem voltak meghatározva, és ebben a vegetációtípusban egyébként sem játszanak meghatározó szerepet. A hajtásos növények elnevezéseit Ehrendorfer (1973) fajlistája alapján egységesítettük, kivéve a *Taraxacum* sect. *Ruderalia* esetében, ami a listában nem szereplő, de általunk használt név.

9.2.2. Klímaadatok

A felvételek helyeinek klímára vonatkozó alapadatait (éves átlagos csapadék, évi, januári és júliusi középhőmérséklet) a korábbi Csehszlovákia (Vesecký et al. 1958) és Magyarország térképeiről olvastuk le. A csapadék hasznosulása a hőmérséklet emelkedésével romlik, ezért önmagában az éves csapadék nem nyújt elegendő információt a területek humiditásáról. Ezért kiszámítottuk a de Martone féle ariditás/humiditás indexet (Tuhkanen 1980):

$$H = \frac{P}{T + 10}$$

ahol:

P = az éves átlagos csapadék

T = az évi középhőmérséklet.

A kontinentalitás nem csak az alacsonyabb humidításban, hanem a nagyobb éves hőingásban is jelentkezik. Ennek számszerűsítésére kiszámítottuk a januári és a júliusi átlaghőmérséklet

különbségét.

9.2.3. Adatelemzés

A felvételek többségében a borítás adatok a 7 fokozatú Braun-Blanquet skálán, vagy más hasonló borítás intervallumokat megadó skálán voltak feltüntetve. Ezeket az értékeket az intervallum közepének megfelelő borítással helyettesítettük, majd négyzetgyök transzformáltuk a borításokat, így csökkentve a domináns fajok súlyát az elemzésben (van der Maarel 1979).

Faith és munkatársai (1987) összehasonlító vizsgálatának eredményei alapján a felvételek különbözőségének mérésére a relatív Manhattan-távolságot, vagyis a relatív borításokból számolt Manhattan-távolságot használtuk. Ennek előnye, hogy figyelmen kívül hagyja az összborítás különbségeit (amik legalább részben a skálatranszformáció következményei), és nem súlyozza túl a domináns fajokat. A főbb növényzeti gradiensek feltárására a távolságmátrix alapján főkoordináta elemzést (PCoA; Legendre & Legendre 1998) végeztünk. Az értelmezés során az első három tengelyt használtuk, amelyek összesen a variáció 23,8%-t magyarázzák. Az értelmezéshez az ordinációs ábrákon CanoDraw program (ter Braak & Šmilauer 2002) segítségével utólag („passive projection”) elhelyeztük a fajokat és az éghajlati változókat.

A felvételek klasszifikációja előtt az 4. fejezetben leírt módon zajszűrést végeztünk. A törött pálca modellel való összehasonlítás alapján 13 szignifikáns ordinációs tengely volt, melyek együttesen a teljes variáció 47,5%-t magyarázzák. A klasszifikációt ezen tengelyek koordinátái alapján, Ward-módszerrel készítettük el. Az ordináció és a klasszifikáció a SYN-TAX 2000 programcsomaggal (Podani 1994) készült. Az optimális csoportszám megállapításához kiszámítottuk a legalább 10 felvételben előforduló fajok válogatóképességét az elméleti várhatóértékkel és szórással korrigált G-statisztikát használva (lásd 5.3 fejezet), majd ezeket az értékeket átlagolva kaptuk meg a csoportosítás élességét (crispness), amelynek csúcsai jelzik az optimális csoportszámokat.

A csoportok jellemzésére kiszámoltuk a fajok fidelitás értékeit. A fidelitás mérésére az u_{hyp} statisztikát használtuk (Chytrý et al. 2002). Az 5,3-nál nagyobb fidelitású fajokat tekintettük karakterfajnak. A határértéket úgy választottuk ki, hogy végül összesen 100 karakterfajt kapjunk. A fidelitások kiszámításához a Juice 6.1 programot (Tichý 2002) használtuk.

A felvételcsoportok éghajlati adatait Kruskal-Wallis teszttel hasonlítottuk össze, és amennyiben szignifikáns eredményt kaptunk, a páros összehasonlításokat Dunn post-hoc teszttel (Zar 1999: 195-200, 223-224) végeztük el.

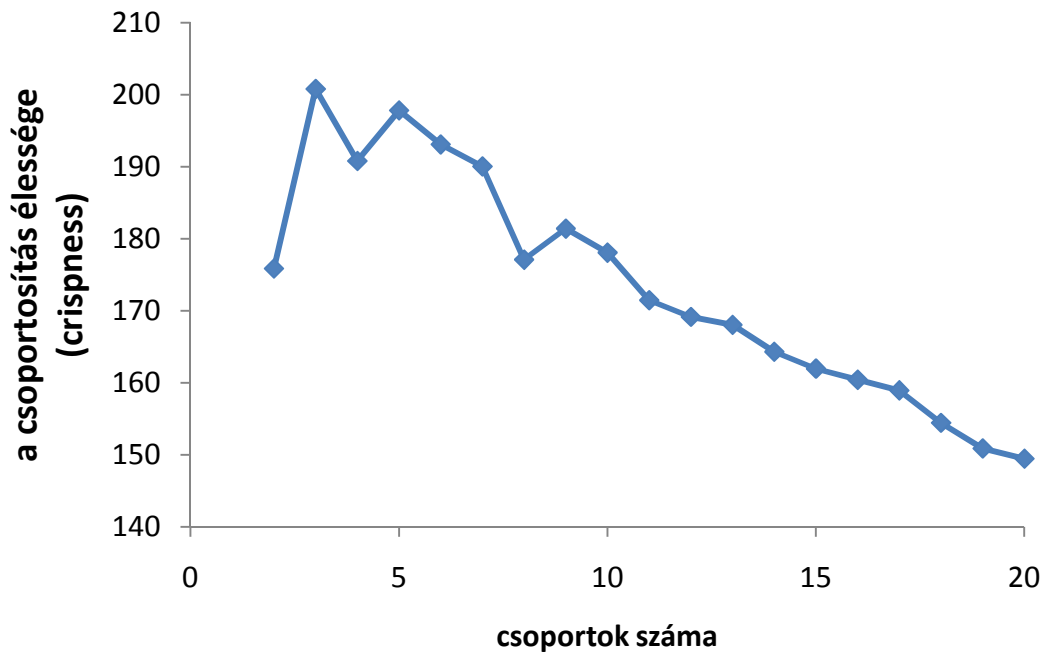
9.3. Eredmények és megvitatásuk

9.3.1. A felvételek csoportosítása

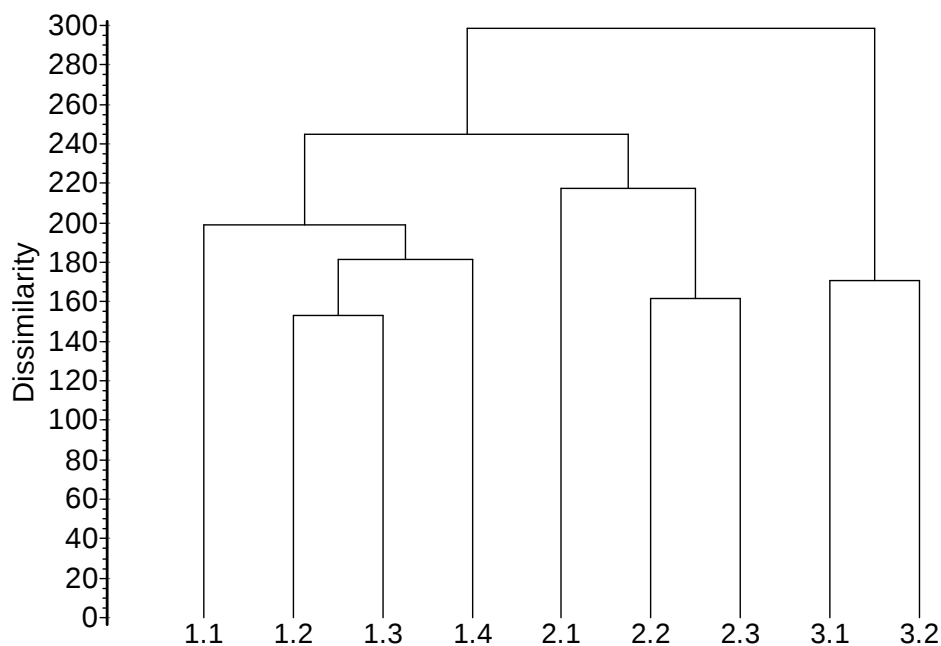
A klasszifikáció élessége (crispness) alapján (37. ábra) az optimális csoportszám három. Ez a három csoport a korábbi szüntaxonómiai rendszerek asszociációcsoportjainak

dc_16_10

feleltethetők meg. További csúcsok figyelhetők meg öt és kilenc csoportnál, amelyek a finomabb felosztások optimális csoportszámai. Ezek a felvételcsoportok a szüntaxonómiai rendszerben az asszociáció-alcsoportoknak (suballiance), illetve az asszociációknak feleltethetők meg.



37. ábra: A klasszifikáció élességének (crispness) változása a csoportszám függvényében. A görbe maximuma 3 csoportnál, és csúcsai 5 és 9 csoportnál jelzik az optimális csoportszámokat



38. ábra: A zajszűrés után, Ward módszerrel készült dendrogram felső része. A csoportokon belüli részleteket a jobb áttekinthetőség kedvéért nem ábrázoljuk. A jelölések a csoportoknak a szövegben és a táblázatokban használt számai

9.3.1.1. Durvább csoportosítás: asszociációcsoportok

A három csoport közül kettő egyértelműen megfeleltethető a szüntaxonómiában nemzetközileg elfogadott asszociációcsoportoknak (Rybníček et al. 1984, Ellmauer & Mucina 1993, Oberdorfer 1993, Dierschke 1995, Moravec et al. 1995, Kučera & Šumberová 2001, Schubert et al. 2001, Stanová & Valachovič 2002, Borhidi 2003): *Calthion* (2. csoport), illetve *Molinion* (3. csoport) (38. ábra).

A *Calthion* réteket magaskórós kétszikűek dominálják, a graminoid típusú fajok alárendelt szerepet játszanak. Általában kisebb patakok árterein fordulnak elő, ahol a talaj egész évben nedves, és a vízellátottság is jó (Hájek & Hájková 2004). Elterjedésük súlypontja a montán-szubmontán területekre esik, az alföldi területeken csak ritkán fordulnak elő. Közép-Európán belül elterjedési mintázatuk atlantikus-szubatlantikus: gyakoriak Németországban és Csehországban (Oberdorfer 1993, Kučera & Šumberová 2001, Schubert et al. 2001, Burkart et al. 2004, Havlová et al. 2004), de Magyarországon már ritkák (Borhidi 2003).

A *Molinia arundinacea* vagy *M. coerulea* dominálta *Molinion* rétek tápanyagszegény, gyakran tőzeges, talajon fordulnak elő. Termőhelyükön a talajvíz szintje ingadozó: tavasszal magas, de nyáron lesüllyed és kiszárad a talaj. Közép-Európában általánosan elterjedtek. Míg tőlünk nyugatra az alföldektől a submontán területekig bárhol találkozhatunk velük (Oberdorfer 1993, Kučera & Šumberová 2001, Burkart et al. 2004), addig Magyarországon elsősorban az alföldi területeken fordulnak elő, leggyakrabban tőzeges talajon, ahol természetes körülmények között üde láprétekkel alkottak vegetációkomplexeket (Kovács 1962, Borhidi 2003).

Az első csoportba tartozó rétek jellemzően a nagy folyók alföldi árterein fordulnak elő. Ezeket a területeket ugyan rendszeresen elönti a víz, de szemben a *Calthion* rétek termőhelyével nyáron a talajuk kiszárad, és a kontinentális klíma miatt a légkör páratartalma is alacsony. A termőhelyükön a talajvíz dinamikája hasonló a kékperjés rétekéhez, de különböznek azoktól a talaj magasabb tápanyagtartalmában. Ezeket a gyepeket a különböző szerzők különböző asszociációcsoportokba sorolták: *Agrostion albae* (Soó 1941), *Alopecurion pratensis* (Passarge 1964), *Cnidion venosi* (Balátová-Tulácková 1966, 1969), *Deschampsion cespitosae* (Horvatić 1930) and *Veronico longifoliae-Lysimachion vulgaris* (Balátová-Tulácková 1981). Az ugyanazokat a neveket időnként országonként némileg eltérő értelemben használták, míg egyes neveket csak a közép-európai országok egy részében használták. Az *Alopecurion pratensis* asszociációcsoportot egyes szerzők földrajzilag elkülönülő, Közép-Európa északi és nyugati, szubóceánikus klímájú területére jellemző vegetációtípusnak tartják, míg a többi elnevezést a Pannon régió rétjei számára tartják fenn (Moravec et al. 1995, Kučera & Šumberová 2001, Stanová & Valachovič 2002). A *Veronico longifoliae-Lysimachion vulgaris* csoportot eredetileg a kontinentális területek magaskórós kétszikűekben gazdag, felhagyott állományainak elkülönítésére javasolták, de az ilyen állományok ritkák és fajösszetételük nagyon hasonló a rendszeresen kaszált rétekéhez (Ellmauer & Mucina 1993, Šumberová 1997, Kučera & Šumberová 2001). A *Cnidion* asszociációcsoport Balátová-Tulácková (1969) eredeti leírása szerint Közép- és Kelet-Európa

kontinentális fajokat tartalmazó ártéri rétjeit tartalmazza. Szerinte a csoport areája Dél-Oroszországtól Ukrajnán, Lengyelországon és a Kárpát-medencén át Németország keleti részéig, illetve Délnyugat-Németországban a Rajna völgyéig terjed. Ebben az értelmezésben a *Cnidion* asszociációcsoport átfed az Északkelet-Horvátországból leírt *Deschampsion cespitosae* asszociációcsoporttal (Horvatić 1930). Ellmauer & Mucina (1993) viszont a *Deschampsion*-t illír, míg a *Cnidion*-t pannon elterjedésű asszociációcsoportnak tekinti, annak ellenére, hogy eredetileg a *Deschampsion* csoportot Horvátország pannóniai és nem illír flóratartományából írták le, és az eredeti leírásban szereplő felvételek nem tartalmazzak illír fajokat.

Habár az első csoport heterogénebb, mint a másik kettő, belső struktúrája egyiket sem támasztja alá a korábbi szakirodalomban megjelenő felosztásoknak. A csoport karakterfajai között - szemben a *Calthion* és *Molinion* csoporttal - több kontinentális areájú fajt találunk (19. táblázat). Azonban a leginkább kontinentális elterjedési területtel rendelkező fajok (pl. *Allium angulosum*, *Carex melanostachya*, *C. praecox*, *Cnidium dubium*, *Galium boreale*, *Scutellaria hastifolia*, *Veronica longifolia* and *Viola pumila*) csak az ide tartozó felvételek egy részében fordulnak elő (elsősorban az 1.1 és 1.2 csoportban; 20. táblázat), sokkal gyakoribbak azok az állományok, amelyekből ezek a fajok hiányoznak.

Az 1. csoport belső szerkezete alapján azt javasoltuk (Botta-Dukát et al. 2005), hogy tekintsük a csoportot egyetlen asszociációcsoportnak, amire a legrégebbi érvényes nevet – *Deschampsion cespitosae* Horvatić 1930 – kell használni. Az *Agrostion albae* (Soó 1941) elnevezés ennek nómenklaturai szinonímja, míg a többi elnevezés (*Alopecurion pratensis*, *Cnidion venosae*, and *Veronico longifoliae-Lysimachion vulgaris*) szüntaxonómiai szinonímnak tekinthető.

19. táblázat: A három asszociációcsoport karakterfajainak megoszlása a kontinentalitási gradiens menti elterjedésük szerint. A táblázatban fajszaámok, és zárójelben az összes diagnosztikus fajon beüli arány szerepel. A fajok besorolása Rothmaler és munkatársai (1990) munkája alapján történt: a kategóriákat a leginkább óceánikus elterjedési típustól /oz/ a leginkább kontinentálisig /k/ rendeztük sorba. Diagnosztikusnak azokat a fajokat tekintettük, amelyek uhyp értéke a 3 csoport esetén 5-nél nagyobb.

	<i>Deschampsion</i>	<i>Calthion</i>	<i>Molinion</i>
oz	–	–	4 (15%)
(oz)	4 (33%)	6 (33%)	14 (52%)
suboz	–	4 (22%)	2 (7%)
(suboz)	1 (8%)	3 (17%)	3 (11%)
(subk)	1 (8%)	–	–
subk	1 (8%)	–	1 (4%)
(k)	1 (8%)	–	1 (4%)
k	–	–	–
indifferens fajok	4 (33%)	5 (28%)	2 (7%)

9.3.1.2. Finomabb csoportosítás: asszociációk

Ha kilenc csoportot különítünk el, akkor azok nagyjából megfeleltethetők a hagyományos cönológiai leírások asszociációinak. Azonban, mivel elemzésünk a fajösszetételén alapult, ezek a csoportok tágabbak a gyakran a domináns faj alapján elkülönített asszociációknál. Mivel célunk a durva földrajzi léptékű variáció leírása volt, nem törekedtünk a részletes összehasonlításra a meglevő szüntaxonómiai rendszerekkel, illetve a nómenklatúrai kérdések tisztázására. A kapcsolatot az elkülönített csoportok és a hagyományos szüntaxonok között a 20. táblázat mutatja be.

1.1 csoport: Egész évben jó vízellátottságú területek (általában árterek) mocsárrétjei

Az ide tartozó állományok domináns fajai különböző füvek (sások), leggyakrabban az *Alopecurus pratensis*, de helyenként a *Deschampsia cespitosa*, *Agrostis stolonifera*, *Poa palustris* vagy *Carex vulpina*. A fideális fajok listája (21. táblázat) alapján talaja nedvesebb, mint – a szintén *Alopecurus pratensis* dominálta – 1.3. csoporté. A magassásosok fajainak magas fidelitás értéke azt jelzi, hogy a talajvíz szintje még nyáron is viszonylag magas. Ilyen termőhely a folyók árterének mélyebb részein és mélyedéseiben tud kialakulni. Ez a vegetációtípus elterjedt Délkelet-Csehországban, Dél-Szlovákiában, és Magyarországon. Horvátország Dráva menti területein szintén előfordul. Ha csak 5 csoportot különböztetünk meg, ez a csoport elkülönül a többi *Deschampsion* gyeptől.

1.2. csoport: Tavasszal elárasztott, de nyáron erősen kiszáradó területek mocsárrétjei

A csoportnak nincs egyetlen jellemző domináns faja, számos faj – *Alopecurus pratensis*, *Poa pratensis*, *Carex praecox* és *Serratula tinctoria* – érhet el magas borítást ezekben a gyepekben. A csoportot jól elkülönül a többi csoporttól. Jellemző fajai között (21. táblázat) számos kontinentális elterjedésű, mint például a *Cnidium dubium*, *Carex praecox*, *Viola pumila*, *Veronica longifolia*, *Inula salicina* és *Carex melanostachya*. Tavasszal néhány hétre az árvíz rendszerint elönti ezeket a gyepeket, de nyárra a száraz, kontinentális jellegű klíma miatt a talajvíz szintje 1 méteres mélységbe, vagy még mélyebbre süllyed (Balátová-Tulácková 1968). Évente egyszer-kétszer kaszálják, de nem legeltetik ezeket a gyepeket (Šeffler & Stanová 1999). Ez a vegetációtípus elsősorban a nagy folyók árterein fordul elő Délkelet-Morvaországban, Kelet-Ausztriában, valamint Nyugat- és Dél-Szlovákiában (Vicherek 1962, Balátová-Tulácková 1966, 1969, Balátová-Tulácková & Hübl 1974, Šeffler & Stanová 1999, Vicherek et al. 2000, Zlinská & Kubalová 2001). Az elemzett adatok között nem volt ide tartozó magyarországi felvétel, de a Steták (2005) által Gemencről jelzett „száraz *Alopecuretum*” és az *Alopecuretum pratensis* Tisza ártéréről leírt *Carex praecox* fáciése (Tímár 1953) nagyon közel áll ehhez a típushoz.

1.3. csoport: Átmenet a kaszálórét (Arrhenatherion) felé szubóceánikus klímájú területen

Az ebbe a csoportba tartozó rétek leggyakoribb uralkodó faja az *Alopecurus pratensis*, de ritkábban domináns lehet a *Holcus lanatus*, *Poa pratensis*, *Festuca pratensis*, *Sanguisorba officinalis*, és *Cirsium canum* is. Az üde rétek (*Arrhenatherion*) fajainak magas fidelitás érté-

20. táblázat: A hagyományos asszociációnevek és az itt elkülönített csoportok közötti kapcsolat. Csak a gyakori neveket és a legfontosabb irodalmakat tüntettük fel a táblázatban

	Csehország	Szlovákia és Kelet-Ausztria	Magyarország és Észak-Horvátország
1.1 csoport	Agrostion: <i>Alopecuretum pratensis</i>, <i>Deschampsietum croato-pannonicum</i> (Vicherek 1960) Cnidion: <i>Lathyro palustris-Gratioletum</i> (Vicherek et al. 2000)	Alopecurion: <i>Alopecuretum pratensis</i>(Řehořek 1969, Špáníková 1971) Deschampsion: <i>Cirsio cani-Deschampsietum cespitosae</i> (Řehořek 1969) Molinion (Zahradníková-Rošetzká 1965, Bosáčková 1971) Cnidion: <i>Cnidio-Violetum pumilae</i> (Maťašová 1987, Šeffer & Stanová 1999), <i>Lathyro palustris-Gratioletum</i>, <i>Gratiolo-Caricetum suzae</i> (Balátová-Tuláčková & Hübl 1974)	Agrostion: <i>Carici vulpinae-Alopecuretum pratensis</i> (Kovács & Máthé 1967), <i>Agrostietum albae</i> (Simon 1960) Cnidion: <i>Veronica longifolia-Euphorbia lucida</i> (Balátová-Tuláčková & Knežević 1975), <i>Gratiola officinalis-Carex praecox-suzae</i> (Balátová-Tuláčková & Knežević 1975), <i>Serratulo-Plantaginetum altissimae</i> (Balátová-Tuláčková & Knežević 1975)
1.2 csoport	Molinion: <i>Cnidium venosum-Jacea vulgaris ssp. angustifolia</i> ass. (Vicherek 1960), <i>Serratulo-Festucetum commutatae</i> (Blažková 1993) Cnidion: <i>Gratiolo-Caricetum suzae</i>, <i>Cnidio venosi-Jaceetum angustifoliae</i>, <i>Cnidio-Violetum pumilae</i> (Vicherek et al. 2000)	Cnidion: <i>Gratiolo-Caricetum suzae</i>(Balátová-Tuláčková 1968, Balátová-Tuláčková & Hübl 1974, Šeffer & Stanová 1999, Zlinská & Kubalová 2001), <i>Cnidio-Violetum pumilae</i> (Balátová-Tuláčková & Hübl 1974) Molinion: <i>Serratulo-Festucetum commutatae</i> (Zlinská & Kubalová 2001)	Agrostion: <i>Festucetum pratensis anthoxanthetosum</i> (Jeanplong 1960)
1.3 csoport	Alopecurion: <i>Alopecuretum pratensis</i> (Kovář 1981, Balátová-Tuláčková 1997a, Duchoslav 1997), <i>Holcetum lanati</i>(Kovář 1981, Duchoslav 1997) Calthion: <i>Angelico-Cirsietum oleracei</i>	Deschampsion: <i>Cirsio cani-Deschampsietum cespitosae</i> (Řehořek 1969)	–

	(Kovář 1981, Neuhäusl & Neuhäuslová 1989), <i>Scirpo-Cirsietum cani</i> (Neuhäusl & Neuhäuslová 1989, Balátová-Tuláčková 1997a) Cnidion: <i>Pseudolysimachio-Alopecuretum</i> (Blažková 1993) Molinion: <i>Serratulo-Festucetum commutatae</i> (Kovář 1981), <i>Sanguisorbo-Festucetum commutatae</i> (Duchoslav 1997)		
1.4 csoport		Deschampsion: <i>Cirsio cani-Deschampsietum cespitosae</i> (Špániková 1971)	Agrostion: <i>Cirsio cani-Festucetum pratensis</i> (Siroki 1956, 1958, 1962), <i>Carici vulpinae-Alopecuretum pratensis</i> (Bodrogekői 1962), <i>Agrostietum albae</i> (Kovács 1955, Siroki 1958)
2.1 csoport	Calthion: <i>Angelico-Cirsietum oleracei</i> (Balátová-Tuláčková 1981, Duchoslav 1997), <i>Cirsietum rivularis</i> (Balátová-Tuláčková 1997a, Duchoslav 1997), <i>Polygono-Cirsietum palustris</i> (Balátová-Tuláčková 1981), <i>Scirpo-Cirsietum cani</i> (Balátová-Tuláčková 1981, 1997a), <i>Caricetum cespitosae</i> (Balátová-Tuláčková 1981), <i>Scirpetum sylvatici</i> (Kovář 1981)	Calthion: <i>Cirsietum rivularis</i> (Balátová-Tuláčková 1968, Špániková 1971), <i>Scirpetum sylvatici</i> (Špániková 1971), <i>Filipendulo-Menthetum longifoliae</i> (Zlinská 1989) Arrhenatherion: <i>Holcetum lanati</i> (Špániková 1971, Kováčová 1976)	–
2.2 csoport	Calthion: <i>Filipendulo-Geraniatum</i>	Calthion: <i>Filipendulo-Geraniatum palustris</i> (Klika)	–

	<i>palustris</i> (Balátová-Tuláčková 1979, 1997b), <i>Lysimachio vulgaris-Filipenduletum</i> (Balátová-Tuláčková 1997b, Kolbek et al. 1999)	1958, Balátová-Tuláčková 1968)	
2.3 csoport	Calthion: <i>Scirpetum sylvatici</i> (Kolbek et al. 1999, Hajková 2000)	Calthion: <i>Angelico-Cirsietum oleracei</i> (Sitášová 1995), <i>Scirpetum sylvatici</i> (Sitášová 1999)	–
3.1 csoport	Molinion: <i>Junco-Molinietum caeruleae</i> (Balátová-Tuláčková 1997b)	Molinion (Zahradníková-Rošetzká 1965, Špáníková 1971, Bosáčková 1975)	Molinion: <i>Succiso-Molinietum</i> , <i>Junco-Molinietum</i> , <i>Molinio-Salicetum rosmarinifoliae</i> (Kovács 1962)
3.2 csoport	Molinion: <i>Molinietum caeruleae</i> (Duchoslav 1997), <i>Gentiano pneumonanthis-Molinietum litoralis</i> (Balátová-Tuláčková 1993)	Molinion: <i>Selino-Molinietum caeruleae</i> (Balátová-Tuláčková 1968, Bosáčková 1975), <i>Silaetum pratensis</i> (Balátová-Tuláčková 1968) Calthion: <i>Cirsietum rivularis</i> (Bosáčková 1975)	Molinion: <i>Succiso-Molinietum</i> (Kovács 1962)

21. táblázat: A kilenc csoport fidelis fajainak szinoptikus táblázata. A táblázat bal oldali része a fajok frekvenciáit mutatja százalékban, a jobb oldali része pedig a fidelitás értékeit. A csoportokon belül a fajokat fidelitás szerint csökkenő sorrendben tüntettük fel. A száz legmagasabb fidelitású fajt tekintettük karakterfajnak. A karakterfajok után feltüntettük a táblázatban a többi olyan fajt, amelynek a frekvenciája a teljes adatbázisban meghaladta a 10%-t.

	Frekvencia (%)									Fidelitás (u_{hyp})								
	<i>Deschampsion</i>				<i>Calthion</i>			<i>Molinion</i>		<i>Deschampsion</i>				<i>Calthion</i>			<i>Molinion</i>	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
1.1. Egész évben jó vízellátottságú területek (általában árterek) mocsárrétjei																		
<i>Poa palustris</i>	54	13	2	–	14	–	8	–	5	10	–	–	–	–	–	–	–	–
<i>Galium palustre</i> agg.	68	23	5	15	26	9	25	8	18	8.6	–	–	–	–	–	–	–	–
<i>Iris pseudacorus</i>	29	3	–	–	2	–	–	2	–	8.6	–	–	–	–	–	–	–	–
<i>Gratiola officinalis</i>	34	30	–	15	1	–	–	2	–	7.2	4.2	–	1.2	–	–	–	–	–
<i>Rumex crispus</i>	44	40	18	12	6	9	–	–	–	6.8	4.0	–	–	–	–	–	–	–
<i>Carex vulpina</i>	41	13	9	23	12	–	–	2	3	6.8	–	–	1.5	–	–	–	–	–
<i>Mentha pulegium</i>	19	3	–	–	–	–	–	3	–	6.7	–	–	–	–	–	–	–	–
<i>Eleocharis palustris</i> agg.	34	20	4	8	4	–	–	13	5	6.0	1.6	–	–	–	–	–	–	–
1.2. Tavasszal elárasztott, de nyáron erősen kiszáradó területek mocsárrétjei																		
<i>Cnidium dubium</i>	10	67	2	4	2	–	–	–	5	–	12	–	–	–	–	–	–	–
<i>Viola pumila</i>	–	40	2	–	–	–	–	–	3	–	11	–	–	–	–	–	–	–
<i>Carex praecox</i>	22	73	9	12	4	–	–	–	5	2.4	11	–	–	–	–	–	–	–
<i>Inula salicina</i>	2	53	2	8	–	4	–	3	26	–	9.1	–	–	–	–	–	–	4.0
<i>Veronica arvensis</i>	–	30	4	–	1	–	–	–	–	–	8.8	–	–	–	–	–	–	–
<i>Serratula tinctoria</i>	31	97	11	4	5	–	–	53	44	–	8.8	–	–	–	–	–	4.8	2.3
<i>Carex melanostachya</i>	15	37	–	–	–	–	–	–	–	3.8	8.1	–	–	–	–	–	–	–
<i>Clematis integrifolia</i>	–	20	–	4	–	–	–	–	–	–	7.8	–	–	–	–	–	–	–
<i>Valerianella locusta</i>	–	20	2	–	–	–	–	–	–	–	7.8	–	–	–	–	–	–	–

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
<i>Veronica longifolia</i>	17	43	5	–	2	–	–	–	3	3.0	7.8	–	–	–	–	–	–	–
<i>Agropyron repens</i>	19	57	23	23	–	–	–	2	–	1.6	7.6	2.6	1.7	–	–	–	–	–
<i>Glechoma hederacea</i>	15	73	39	19	20	9	17	2	–	–	7.5	3.7	–	–	–	–	–	–
<i>Scutellaria hastifolia</i>	5	23	–	–	–	–	–	–	–	1.3	7.4	–	–	–	–	–	–	–
<i>Lythrum virgatum</i>	10	33	–	12	–	–	–	–	3	1.9	7.2	–	1.5	–	–	–	–	–
<i>Vicia tetrasperma</i>	5	27	–	8	1	–	–	–	–	–	7.0	–	1.2	–	–	–	–	–
<i>Myosotis ramosissima</i>	–	17	2	–	–	–	–	–	–	–	7.0	–	–	–	–	–	–	–
<i>Allium angulosum</i>	15	40	4	–	–	–	–	12	3	2.2	6.7	–	–	–	–	–	1.1	–
<i>Potentilla reptans</i>	44	67	14	23	6	–	–	12	21	4.8	6.5	–	–	–	–	–	–	–
<i>Euphorbia esula</i>	–	17	4	–	–	–	–	–	–	–	6.3	1.0	–	–	–	–	–	–
<i>Calamagrostis epigejos</i>	–	20	–	4	–	–	–	–	8	–	6.3	–	–	–	–	–	–	2.1
<i>Rumex thyrsiflorus</i>	–	13	2	–	–	–	–	–	–	–	6.1	–	–	–	–	–	–	–
<i>Vicia angustifolia</i>	–	10	–	–	–	–	–	–	–	–	6.0	–	–	–	–	–	–	–
<i>Iris sibirica</i>	12	30	–	–	1	–	–	2	10	2.2	6.0	–	–	–	–	–	–	1.3
<i>Festuca rupicola</i>	–	20	5	–	–	–	–	–	5	–	5.9	1.2	–	–	–	–	–	–
<i>Cirsium arvense</i>	7	43	28	15	5	–	–	–	8	–	5.7	4.3	–	–	–	–	–	–
<i>Symphytum officinale</i> agg.	41	67	37	15	26	9	–	3	5	3.1	5.5	2.3	–	–	–	–	–	–
1.3. Átmenet a kaszálórétek (<i>Arrhenatherion</i>) felé szubóceánikus klímájú területen																		
<i>Geranium pratense</i>	3	3	60	–	19	–	–	3	3	–	–	11	–	1.2	–	–	–	–
<i>Trisetum flavescens</i>	–	–	42	–	10	–	–	–	3	–	–	9.8	–	–	–	–	–	–
<i>Cerastium holosteoides</i>	5	40	81	38	41	4	–	8	28	–	1.1	8.7	–	2.1	–	–	–	–
<i>Dactylis glomerata</i>	5	10	63	12	19	17	–	12	46	–	–	7.8	–	–	–	–	–	3.6
<i>Campanula patula</i>	8	27	40	–	7	–	–	5	3	–	2.6	7.2	–	–	–	–	–	–
<i>Veronica chamaedrys</i> agg.	2	17	46	4	25	9	8	–	13	–	–	6.7	–	2.5	–	–	–	–

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
<i>Galium mollugo</i> agg.	15	40	54	15	14	9	8	13	13	–	2.6	6.6	–	–	–	–	–	–
<i>Pimpinella major</i>	5	3	28	–	9	–	–	3	3	–	–	6.2	–	–	–	–	–	–
<i>Anthriscus sylvestris</i>	–	–	18	–	4	4	8	–	–	–	–	5.8	–	–	–	–	–	–
<i>Poa pratensis</i> agg.	32	97	88	77	53	35	25	12	64	–	5.0	5.7	2.6	–	–	–	–	1.5
<i>Vicia sepium</i>	–	3	26	–	16	4	–	–	–	–	–	5.7	–	3.1	–	–	–	–
1.4. Átmenet a kaszálórét (Arrhenatherion) felé szubkontinentális klímájú területen																		
<i>Medicago lupulina</i>	2	–	5	50	4	–	–	5	10	–	–	–	8.9	–	–	–	–	–
<i>Bromus commutatus</i>	–	–	–	12	–	–	–	–	–	–	–	–	6.5	–	–	–	–	–
<i>Daucus carota</i>	7	13	11	54	4	–	–	28	10	–	–	–	6.2	–	–	–	3.7	–
2.1. Kaszált Calthion gyepek																		
<i>Myosotis palustris</i> agg.	15	7	11	–	56	–	58	7	10	–	–	–	–	9.0	–	3.4	–	–
<i>Holcus lanatus</i>	14	3	67	15	83	13	–	33	44	–	–	4.3	–	8.6	–	–	–	–
<i>Alchemilla vulgaris</i> agg.	–	–	19	–	36	13	–	2	3	–	–	2.0	–	7.6	–	–	–	–
<i>Cirsium rivulare</i>	–	–	2	–	35	–	8	3	31	–	–	–	–	7.4	–	–	–	4.0
<i>Anthoxanthum odoratum</i>	–	17	46	–	59	–	–	17	44	–	–	3.3	–	7.2	–	–	–	2.4
<i>Caltha palustris</i>	25	3	2	15	54	48	58	10	21	–	–	–	–	6.8	2.6	2.7	–	–
<i>Angelica sylvestris</i>	2	–	16	8	46	39	25	12	18	–	–	–	–	6.7	2.5	–	–	–
<i>Juncus effusus</i>	14	–	5	4	33	4	25	10	3	–	–	–	–	6.2	–	1.3	–	–
<i>Lotus uliginosus</i>	–	–	5	–	23	26	–	3	–	–	–	–	–	5.9	3.4	–	–	–
<i>Cardamine pratensis</i> agg.	41	67	39	15	62	39	8	2	5	1.1	3.9	–	–	5.8	–	–	–	–
<i>Dactylorhiza majalis</i>	–	–	4	–	17	4	–	–	5	–	–	–	–	5.8	–	–	–	–
<i>Equisetum palustre</i>	7	–	9	19	52	70	50	23	31	–	–	–	–	5.7	4.8	1.8	–	–
<i>Lychnis flos-cuculi</i>	61	77	63	62	77	4	25	2	38	1.9	3.1	2.2	1.2	5.4	–	–	–	–

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
2.2. <i>Filipendula ulmaria</i> dominálta felhagyott <i>Calthion</i> gyepek																		
<i>Geranium palustre</i>	–	–	2	–	11	57	–	–	–	–	–	–	–	2.2	11	–	–	–
<i>Anemone nemorosa</i>	–	–	4	–	–	26	–	–	–	–	–	–	–	–	8.3	–	–	–
<i>Chaerophyllum hirsutum</i>	–	–	–	–	–	17	–	–	–	–	–	–	–	–	8.0	–	–	–
<i>Crepis paludosa</i>	–	–	–	–	15	48	–	–	10	–	–	–	–	3.1	7.9	–	–	–
<i>Filipendula ulmaria</i>	14	43	28	–	56	96	58	–	31	–	1.4	–	–	5.2	6.8	2.0	–	–
<i>Cirsium oleraceum</i>	–	–	19	–	30	57	25	3	3	–	–	1.3	–	4.6	6.1	1.1	–	–
<i>Galium aparine</i>	–	7	2	–	9	30	8	–	–	–	–	–	–	1.9	6.0	–	–	–
2.3. <i>Scirpus sylvaticus</i> dominálta felhagyott <i>Calthion</i> gyepek																		
<i>Urtica dioica</i>	10	7	14	4	11	30	67	–	–	–	–	–	–	–	3.2	6.4	–	–
3.1 and 3.2. csoport: <i>Molinion</i> gyepek																		
<i>Achillea asplenifolia</i>	10	–	–	–	–	–	–	33	5	–	–	–	–	–	–	–	8.5	–
<i>Festuca pseudovina</i>	–	3	–	4	–	–	–	22	3	–	–	–	–	–	–	–	7.4	–
<i>Nardus stricta</i>	–	–	–	–	–	–	–	18	3	–	–	–	–	–	–	–	7.4	–
<i>Gentiana pneumonanthe</i>	3	–	–	–	1	–	–	25	8	–	–	–	–	–	–	–	7.3	–
<i>Hieracium lactucella</i>	–	–	–	–	–	–	–	12	–	–	–	–	–	–	–	–	6.2	–
<i>Centaurium littorale</i> ssp. <i>uliginosum</i>	–	–	–	–	–	–	–	10	–	–	–	–	–	–	–	–	5.8	–
<i>Tetragonolobus maritimus</i>	3	–	–	–	–	–	–	22	21	–	–	–	–	–	–	–	5.6	4.1
<i>Briza media</i>	–	3	7	–	26	–	–	28	74	–	–	–	–	1.9	–	–	2.1	9.4
<i>Trifolium montanum</i>	–	3	–	–	1	–	–	–	28	–	–	–	–	–	–	–	–	9.1
<i>Dianthus superbus</i>	2	–	–	–	1	–	–	–	26	–	–	–	–	–	–	–	–	8.6
<i>Carex davalliana</i>	–	–	–	–	–	–	–	13	31	–	–	–	–	–	–	–	3.1	7.6
<i>Galium verum</i> agg.	5	37	23	15	7	9	–	35	69	–	1.9	–	–	–	–	–	2.5	7.4

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
<i>Filipendula vulgaris</i>	–	17	9	–	2	–	–	–	33	–	2.4	–	–	–	–	–	–	7.2
<i>Galium boreale</i>	8	63	42	–	6	–	–	23	72	–	5.1	3.3	–	–	–	–	–	7.2
<i>Carex hostiana</i>	–	–	–	–	–	–	–	10	26	–	–	–	–	–	–	–	2.5	7.1
<i>Carex hartmanii</i>	–	–	–	–	–	–	–	3	18	–	–	–	–	–	–	–	–	6.8
<i>Gymnadenia conopsea</i>	–	–	–	–	–	–	–	2	15	–	–	–	–	–	–	–	–	6.7
<i>Valeriana dioica</i>	5	–	–	–	19	13	–	17	46	–	–	–	–	1.8	–	–	1.0	6.6
<i>Carex tomentosa</i>	7	10	2	–	4	–	–	7	31	–	–	–	–	–	–	–	–	6.1
<i>Danthonia decumbens</i>	–	–	–	–	2	–	–	13	26	–	–	–	–	–	–	–	3.1	6.1
<i>Selinum carvifolia</i>	2	13	19	–	9	4	–	15	44	–	–	1.6	–	–	–	–	–	6.0
<i>Molinia arundinacea</i> et <i>caerulea</i>	5	3	–	–	11	9	–	100	87	–	–	–	–	–	–	–	13	8.6
<i>Succisa pratensis</i>	–	7	9	–	11	–	–	77	72	–	–	–	–	–	–	–	11	7.6
<i>Potentilla erecta</i>	–	–	2	4	14	–	–	52	69	–	–	–	–	–	–	–	7.2	8.6
<i>Carex panicea</i>	10	10	5	27	42	4	–	72	87	–	–	–	–	1.7	–	–	6.7	7.4
<i>Leontodon hispidus</i>	5	17	9	12	2	–	–	45	49	–	–	–	–	–	–	–	6.4	5.7
<i>Linum catharticum</i>	2	3	–	–	1	–	–	30	41	–	–	–	–	–	–	–	5.8	7.0
<i>Carex flacca</i>	–	–	–	–	5	–	–	28	41	–	–	–	–	–	–	–	5.4	7.0
Több csoport közös karakterfajai																		
<i>Taraxacum officinale</i> agg.	37	77	79	92	14	–	8	15	18	–	4.7	7.2	6.1	–	–	–	–	–
<i>Rumex acetosa</i>	20	53	89	31	86	22	–	15	49	–	–	6.6	–	7.5	–	–	–	–
<i>Scirpus sylvaticus</i>	3	–	9	–	58	52	100	3	5	–	–	–	–	9.1	3.7	6.8	–	–
További legalább 10%-s frekvenciájú fajok																		
<i>Ranunculus acris</i>	32	80	82	77	84	17	–	78	95	–	1.4	2.4	–	3.3	–	–	1.7	3.7
<i>Ranunculus repens</i>	88	67	65	62	77	26	25	33	36	4.9	–	–	–	3.5	–	–	–	–

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
<i>Alopecurus pratensis</i>	83	97	88	69	63	48	67	3	10	4.3	4.5	5.0	1.3	1.1	–	–	–	–
<i>Sanguisorba officinalis</i>	19	77	82	8	64	22	25	60	62	–	2.8	4.9	–	2.4	–	–	1.3	1.2
<i>Festuca pratensis</i> agg.	36	50	79	85	65	13	–	28	64	–	–	4.4	3.4	2.7	–	–	–	1.6
<i>Lathyrus pratensis</i>	20	93	68	8	73	70	25	18	54	–	5.0	3.1	–	4.7	2.0	–	–	–
<i>Lysimachia nummularia</i>	59	80	61	19	70	13	8	17	36	2.0	3.7	2.3	–	4.6	–	–	–	–
<i>Deschampsia cespitosa</i>	32	37	60	12	56	13	8	43	82	–	–	2.4	–	2.2	–	–	–	4.9
<i>Poa trivialis</i>	44	23	68	50	60	39	75	3	10	–	–	4.6	1.0	4.0	–	2.4	–	–
<i>Plantago lanceolata</i>	25	80	51	27	36	–	–	47	64	–	4.6	1.7	–	–	–	–	1.0	3.2
<i>Centaurea jacea</i> agg.	22	57		42	25	–	–	58	74	–	2.3	–	–	–	–	–	3.7	5.1
<i>Achillea millefolium</i> agg.	33	50		58	32	4	8	28	59	–	1.5	4.4	2.3	–	–	–	–	3.0
<i>Cirsium canum</i>	63	23	51	42	38	–	–	35	51	–	–	2.9	–	–	–	–	–	2.4
<i>Trifolium pratense</i>	27	30	39	65	33	–	8	25	41	–	–	1.2	3.8	–	–	–	–	1.3
<i>Prunella vulgaris</i>	15	37	26	19	26	9	–	55	54	–	–	–	–	–	–	–	4.5	3.4
<i>Vicia cracca</i>	27	57	40	23	19	30	8	7	54	–	3.6	2.2	–	–	–	–	–	3.7
<i>Festuca rubra</i> agg.	5	17	53	–	51	17	–	13	49	–	–	4.4	–	5.0	–	–	–	3.0
<i>Ranunculus auricomus</i> agg.	19	57	49	4	37	70	–	2	13	–	3.6	3.8	–	2.0	4.5	–	–	–
<i>Leucanthemum vulgare</i> agg.	24	57	40	23	27	–	–	13	46	–	3.7	2.3	–	–	–	–	–	2.7
<i>Agrostis stolonifera</i>	49	33	12	38	28	–	8	32	18	4.1	–	–	1.3	–	–	–	–	–
<i>Carex gracilis</i>	37	37	16	8	25	43	25	22	28	2.1	1.4	–	–	–	2.0	–	–	–
<i>Carex hirta</i>	19	57	28	50	40	–	8	3	21	–	4.0	–	2.9	3.2	–	–	–	–
<i>Lysimachia vulgaris</i>	34	13	7	–	21	57	42	12	56	2.0	–	–	–	–	3.8	1.5	–	5.0
<i>Lythrum salicaria</i>	37	33	5	–	26	17	42	30	23	2.6	1.3	–	–	–	–	1.5	1.2	–
<i>Lotus corniculatus</i>	8	33	14	31	12	–	–	42	38	–	1.7	–	1.3	–	–	–	4.3	2.8
<i>Equisetum arvense</i>	10	33	16	15	20	22	25	10	36	–	2.1	–	–	–	–	–	–	2.9

	Frekvencia (%)									Fidelitás (u_{hyp})								
	Deschampsion				Calthion			Molinion		Deschampsion				Calthion			Molinion	
Csoport	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2
Felvételek száma	59	30	57	26	81	23	12	60	39	59	30	57	26	81	23	12	60	39
<i>Trifolium repens</i>	14	17	32	46	28	–	–	7	3	–	–	2.8	3.8	2.6	–	–	–	–
<i>Colchicum autumnale</i>	7	43	40	8	14	9	–	–	31	–	3.9	5.0	–	–	–	–	–	2.3
<i>Galium uliginosum</i>	–	–	5	–	30	48	25	12	46	–	–	–	–	3.4	4.0	–	–	5.1
<i>Trifolium hybridum</i>	36	30	18	15	23	9	–	–	–	4.2	2.0	–	–	1.8	–	–	–	–
<i>Carex acutiformis</i>	7	13	4	12	17	43	–	22	15	–	–	–	–	–	4.1	–	1.7	–
<i>Phragmites australis</i>	5	–	7	–	16	35	17	30	21	–	–	–	–	–	2.9	–	3.7	1.1
<i>Phalaris arundinacea</i>	32	27	14	4	17	4	17	–	–	4.5	2.1	–	–	1.1	–	–	–	–
<i>Luzula campestris</i> agg.	–	7	19	4	30	–	–	7	21	–	–	1.6	–	5.0	–	–	–	1.5
<i>Rhinanthus minor</i>	12	7	11	8	14	–	–	15	31	–	–	–	–	–	–	–	–	3.6
<i>Stellaria graminea</i>	14	37	18	4	15	9	–	–	8	–	4.3	1.3	–	–	–	–	–	–
<i>Carex nigra</i> agg.	2	–	2	4	27	9	–	10	36	–	–	–	–	4.6	–	–	–	4.8
<i>Juncus articulatus</i>	8	3	–	4	11	–	–	27	18	–	–	–	–	–	–	–	4.6	1.7

kei, és a magassásosokra (*Magnocaricion*) jellemző fajok hiánya egyaránt azt jelzi, hogy az elárasztás kevésbé érinti ezeket a gyepeket, mint az előző két csoportba tartozókat. Legfeljebb rövid ideig borítja őket a víz, vagy esetleg egyáltalán nem kapnak elárasztást. Ebben a tekintetben hasonlítanak az 1.4. csoporthoz. Az onnan hiányzó szárazság-érzékeny fajok (pl. *Lysimachia nummularia*, *Chaerophyllum aromaticum*, *Polygonum bistorta*) jelenléte viszont azt jelzi, hogy az ide tartozó állományok talaja nyáron sem szárad ki erősen. Ez a különbség a két csoport (1.3 és 1.4) között a makroklimatikus gradiens menti elhelyezkedésükkel magyarázható: az 1.3 csoportba tartozó rétek elterjedtek Csehországban és Szlovákiában, de – a nyugati határszél kivételével (Jeanplong 1960) – hiányoznak Magyarországról, ahol viszont az 1.4 csoportba tartozó állományok gyakoriak.

1.4. csoport: Átmenet a kaszálórétek (*Arrhenatherion*) felé szubkontinentális klímájú területen

Az ebbe a csoportba tartozó réteken általában domináns a *Festuca pratensis*, de helyenként más fűvek (pl. *Alopecurus pratensis*, *Deschampsia cespitosa*, *Poa pratensis*) helyettesíthetők. A csoport magas fidelitású fajainak száma igen alacsony (21. táblázat); közülük számos faj (pl. *Daucus carota*, *Trifolium pratense*, *Pastinaca sativa*) az üde kaszálórétekre jellemző. Ugyanakkor, nagy frekvenciával fordulnak elő itt a nedves rétek gyakori fajtái, amelyek az üde kaszálóréteken hiányoznak, vagy ritkák. Ez azt mutatja, hogy az ide tartozó rétek átmeneti helyzetben vannak az ártéri nedves rétek (1.1 és 1.2 csoport) és az üde kaszálórétek (*Arrhenatherion*) között: termőhelyük az előbbieknél szárazabb, az utóbbiaknál nedvesebb. Általában az ártereken kívül fordulnak elő, vagy ha az árteren, akkor annak magasabb, ritkábban és rövidebb időre elárasztott részét foglalják el. Évente egyszer kaszált, általában nem trágyázott gyepek (Ružičková 1971). Elterjedtek Magyarországon és Szlovákia déli részén, de északnyugat felé haladva megritkulnak, fokozatosan az előző (1.3) csoportba tartozó gyepek veszik át a helyüket. Hozzá kell azonban tenni, hogy ennek a típusnak az elterjedtségét Csehországban az adatsorunk alulbecsli, mert az ide tartozó állományok egy részét az *Arrhenatherion* asszociációcsoportba sorolták be a felvételezők, és így ezek nem kerültek be az elemzésünkbe.

2.1. csoport: Kaszált *Calthion* gyepek

Ebbe a csoportba elsősorban Csehországból és Szlovákiából származó felvételek kerültek. Az ide sorolt gyepeknek számos domináns faja van. A fidelis fajok között több higrofil fajt is találunk: pl. *Caltha palustris*, *Myosotis palustris* agg., *Scirpus sylvaticus*, and *Cirsium rivulare*. Általában folyamatosan nedves, tápanyagban gazdag talajon alakulnak ki ilyen állományok (Hájek & Hájková 2004). Az alföldi nagy folyók árterén ritkán fordulnak elő, elsősorban azokon a helyeken, ahol a víz lefolyását vízzáró réteg gátolja (Duchoslav 1997). A csoport elterjedésének centruma a humid klímájú hegyvidéki területekre esik, ahol általában források közelében, vagy kis patakok árterén fordul elő (Rybníček et al. 1984). A hasonló fajösszetételű 2.2. és 2.3. csoporttal szemben az ide tartozó állományokat rendszeresen kaszálják és a felvételek fajszáma magasabb. Ha csak 5 csoportot különítenénk el (második optimum a crispness elemzés szerint), ez a csoport akkor is elkülönülne a másik két *Calthion*

csoporttól. Ez a szétválás összhangban van a *Calthion* aszociációcsoport hagyományos felosztásával két alcsoportra: a kaszált, fajgazdag állományokból álló *Calthenion* és a felhagyott, fajszegényebb gyepeket tartalmazó *Filipendulenion* alcsoportra.

2.2. csoport: *Filipendula ulmaria* dominálta felhagyott *Calthion* gyepek

Ezeknek a magaskórós társulásoknak a jellemező domináns faja a *Filipendula ulmaria*, de egyes állományokban a *Geranium palustre*, *Carex acutiformis* vagy a *C. cespitosa* is magas borítást érhet el. A társulás karakterfajai magas produktivitású nedves rétek fajai (*Geranium palustre*, *Crepis paludosa*, *Cirsium oleraceum*) és árnyéktűrő fajok (*Anemone nemorosa*, *Chaerophyllum hirsutum*). Ez a vegetációtípus általában a *Calthion* gyepek rendszeres kaszálásának felhagyása után alakul ki, főként kalcium- és tápanyag-gazdag talajokon (Hájek & Hájková 2004). Előfordulása inkább a hegy- és dombvidéki területeken jellemző (Rybníček et al. 1984, Kučera & Šumberová 2001). Alacsonyabb tengerszint feletti magasságon csak a szubóceánikus klímájú területeken jelenik meg, főként Észak-Csehországban.

2.3. csoport: *Scirpus sylvaticus* dominálta felhagyott *Calthion* gyepek

Az összes ebbe a csoportba tartozó felvétel domináns faja, a magas növésű, graminoid növekedési formájú erdei káka (*Scirpus sylvaticus*). A domináns faj erős kompetíciós képessége miatt ezt a vegetációtípust alacsony fajszám jellemzi, és csak kevés diagnosztikus faja van (21. táblázat). Elemzésünkben, a *Scirpus sylvaticus* mellett csak az *Urtica dioica* ért el magas fidelitást, ami egyértelműen az állományok eutrofizálódását és ruderalizálódását jelzi. Az erdei kákás állományok a kaszálás felhagyása után alakulnak ki a mészben szegény, de humuszban gazdag, vízzel átitatott, glejesedő talajokon (Hájek & Hájková 2004). Az ilyen élőhelyek, és ennek megfelelően az ide tartozó állományok, ritkák a nagy, alföldi folyók árterein, elterjedtek viszont a hegyvidéki területeken Csehországban és Szlovákiában (Špáníková 1982, Rybníček et al. 1984).

3.1 & 3.2 csoport *Molinion* gyepek

Ha kilenc csoportot különböztetünk meg a klaszteranalízisben, a 3. csoport kettéválik (38. ábra), azonban a két csoport fajösszetétele nagyon hasonló. Az alapvető különbség, hogy a 3.1 csoport fajszegényebb, a felvételek fajszáma 25 körül van, szemben a másik csoport 40 körüli felvételenkénti fajszámával. A szüntaxonómiai rendszerek általában a talaj kémhatása és kalcium tartalma alapján osztják két csoportra a kékperjés réteket (Kovács 1962). Elemzésünkben ez a két csoport nem különült el, ez azonban nem jelenti azt, hogy eredményeink megcáfolják ennek a csoportosításnak a jogosságát. A savanyú talajokra jellemző típusból ugyanis csak néhány felvétel került be az elemzésbe, mert ezek inkább magasabb tengerszint feletti magasságokban fordulnak elő.

Az ide tartozó gyepekben mindig dominánsak a kékperje fajok (*Molinia arundinacea* vagy *M. coerulea*), de számos karakterfajuk alapján is jól elkülönülnek a többi gyeptípustól (21. táblázat). A karakterfajok egy része a kékperjés rétekre specialista faja (pl. *Gentiana pneumonanthe*, *Succisa pratensis*), de gyakoriak közöttük a oligotróf gyepek (pl. *Danthonia decumbens*, *Nardus stricta*, *Potentilla erecta*) és az üde láprétek (pl. *Carex davalliana*, *C.*

hostiana) fajai is. Ezek a gyepek az ártereken kívül helyezkednek el. A hagyományos gazdálkodás során nem trágyázták, és évente egyszer, esetleg csak két évente kaszálták ezeket (Ellmauer & Mucina 1993, Ellenberg 1996, Kučera & Šumberová 2001). A kékperjés rétek elterjedtek Szlovákia és Magyarország alföldi területein, de igen ritkák a hasonló területeken – és ebből kifolyólag az elemzett felvételek között is – Csehországban, ahol a *Molinion* típusú gyepek inkább magasabb tengerszint feletti magasságon, ásványi talajon fordulnak elő.

9.3.2. Ordináció és a vegetáció kapcsolata a makroklímával

Az ordinációs térben az első tengely mentén a *Molinion* csoportok (3.1 és 3.2) különülnek el a többi csoporttól (39. ábra). A felhagyott (kaszálatlan) *Calthion* gyepek (2.2 és 2.3 csoport) jól elválnak – főleg a második tengely mentén – a *Deschampsion* felvételektől, míg a kaszált *Calthion* gyepek (2.1 csoport) közelebb vannak azokhoz.

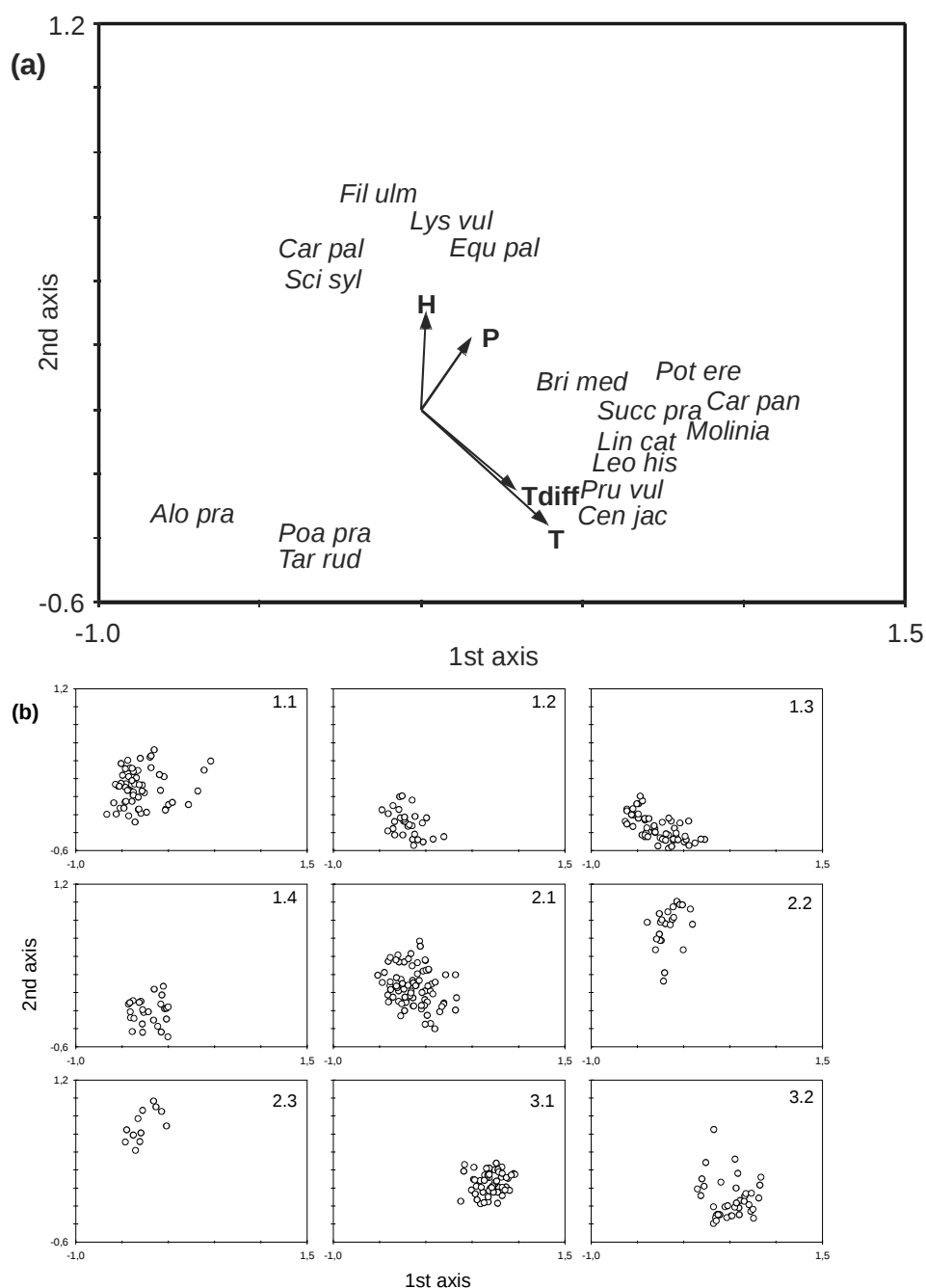
A *Molinion* gyepek látszólag a magas éves középhőmérsékletű és magas éves hőingású helyekhez kötődnek (39. ábra), azonban ez a mintázat valószínűleg műtermék, ami annak következtében alakult ki, hogy kékperjés rétek ritkák Csehország sík- és dombvidéki részén, mert ott inkább nagyobb magasságokban fordulnak elő, gyakoriak viszont a kontinentálisabb klímájú Magyarországon és Dél-Szlovákiában. A valóságban azonban ezeken a helyeken is általában a makroklímánál hűvösebb mezoklímájú helyeken fordulnak elő (Kovács 1957). A *Calthion* gyepek a hűvösebb, csapadékosabb, kiegyenlítettebb (szub)óceáni klímához kötődnek (39. ábra). A *Deschampsion* gyepek viszont Közép-Európa aridabb területeire jellemzők: talajuk tavasszal az áradásoknak köszönhetően vizes, de nyáron kiszárad.

Az éghajlati tényezők az első két ordinációs tengely menti koordináták variációjából csak kis részt magyaráznak (22. táblázat). Azoknak a fajoknak a listája, amelyek előfordulását az első vagy a második ordinációs tengely jól magyarázza (39. ábra) erősen átfed a három asszociációcsoport karakterfajainak listájával (21. táblázat). Az első tengely az ásványi talajoktól (bal oldal) a tőzeges talajokig és lápokig nyúló talajtani gradienssel azonosítható, míg a második tengely a talaj nyári nedvességtartalmával hozható összefüggésbe: az ordinációs diagram alsó részén található a nyáron kiszáradó talajú állományok, a felső részén azok, amelyek talaja nyáron is nedves.

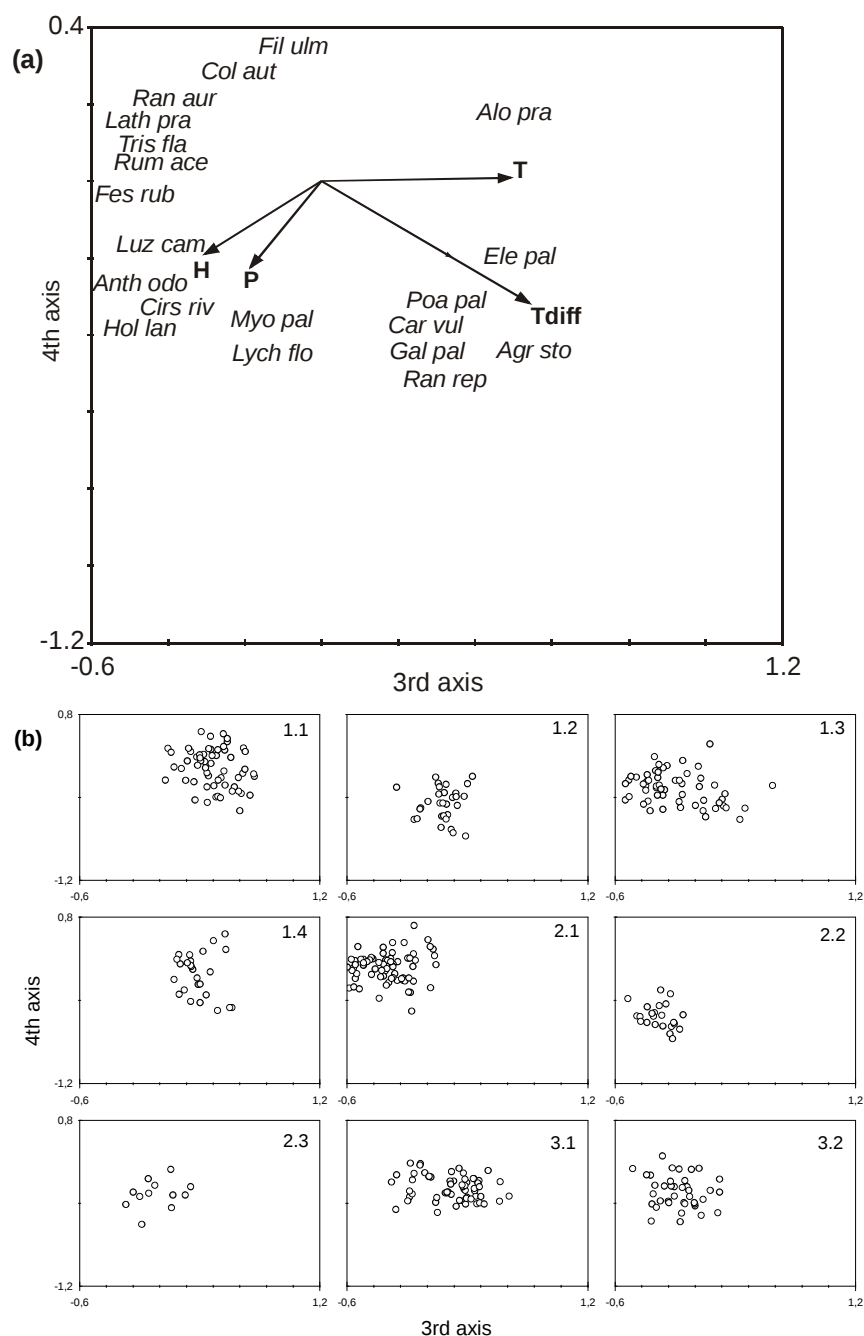
22. táblázat: Az éghajlati tényezők és az ordinációs tengelyek közötti lineáris korreláció. ***
p < 0.001, ** p < 0.01, NS: p > 0.05.

	1. tengely	2. tengely	3. tengely	4. tengely
éves középhőmérséklet (°C)	0.37***	-0.22***	0.60***	0.02 ^{NS}
éves hőingás (°C)	0.28***	-0.10 ^{NS}	0.60***	-0.31***
éves csapadék (mm)	0.16**	0.17**	-0.24***	-0.23***
humiditás (de Martonne index)	0.03 ^{NS}	0.22***	-0.38***	-0.20***

dc_16_10

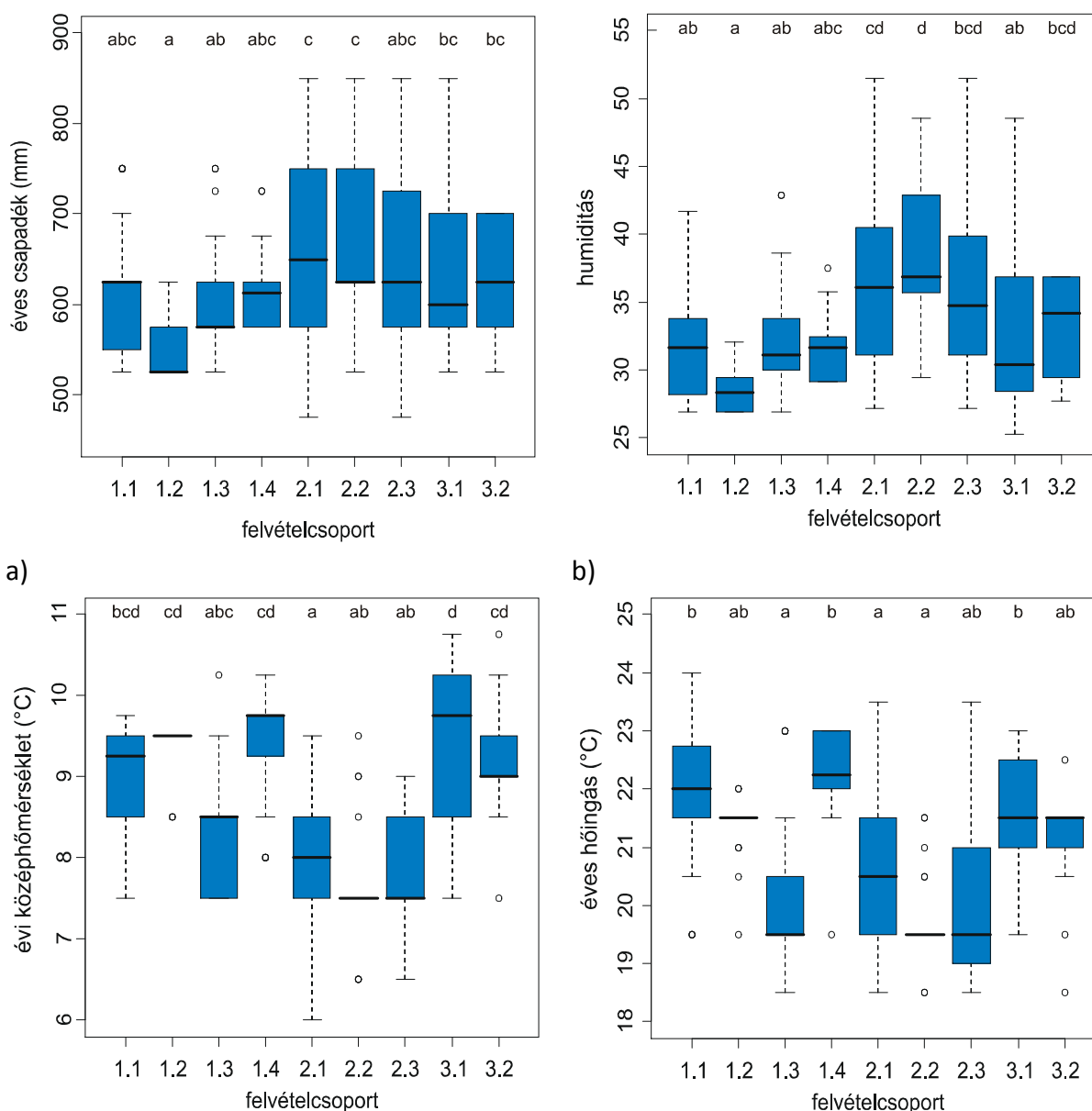


39. ábra: A felvételek ordinációja relatív Manhattan távolság alapján főkoordináta elemzéssel: az első és második tengely. (a) Az éghajlati változók (T = éves középhőmérséklet, Tdiff = éves hőingás, P = éves csapadék, H = humiditás) és a fajok (Alo pra = *Alopecurus pratensis*, Bri med = *Briza media*, Car pal = *Carex pallescens*, Car pan = *Carex panicea*, Cen jac = *Centaurea jacea*, Equ pal = *Equisetum palustrre*, Fil ulm = *Filipendula ulmaria*, Leo his = *Leontodon hispidus*, Lin cat = *Linum catharticum*, Lys vul = *Lysimachia vulgaris*, Molinia = *Molinia* spp., Pru vul = *Prunella vulgaris*, Poa pra = *Poa pratensis*, Pot ere = *Potentilla erecta*, Sci syl = *Scirpus sylvaticus*, Succ pra = *Succisa pratensis*, Tar rud = *Taraxacum* sect. *Ruderalia*) utólagos rávetítése az ordinációs térre. Csak azokat a fajokat jelenítettük meg, amelyek előfordulását a két tengely legjobban magyarázza. (b) A csoportok elválása az ordinációs térben.



40. ábra: A felvételek ordinációja relatív Manhattan távolság alapján főkoordináta elemzéssel: az harmadik és negyedik tengely. (a) Az éghajlati változók (T = éves középhőmérséklet, Tdiff = éves hőingás, P = éves csapadék, H = humiditás) és a fajok (Agr sto = *Agrostis stolonifera*, Alo pra = *Alopecurus pratensis*, Anth odo = *Anthoxanthum odoratum*, Car vul = *Carex vulpina*, Cirs riv = *Cirsium rivulare*, Col aut = *Colchicum autumnale*, Ele pal = *Eleocharis palustris* agg., Fes rub = *Festuca rubra*, Fil ulm = *Filipendula ulmaria*, Gal pal = *Galium palustre*, Hol lan = *Holcus lanatus*, Lath pra = *Lathyrus pratensis*, Luz cam = *Luzula campestris*, Lych flo = *Lychnis flos-cuculi*, Myo pal = *Myosotis palustris*, Poa pal = *Poa palustris*, Ran aur = *Ranunculus auricomus* agg., Ran rep = *Ranunculus repens*, Rum ace = *Rumex acetosa*, Tris fla = *Trisetum flavescens*) utólagos rávetítése az ordinációs térre. Csak azokat a fajokat jelenítettük meg, amelyek előfordulását a két tengely legjobban magyarázza. (b) A csoportok elválása az ordinációs térben.

dc_16_10



c) d)
41. ábra: A felvételszoportok makroklíma adatainak összehasonlítása. Az azonos betűkkel megjelölt csoportok a Dunn post hoc teszt alapján nem különböznek szignifikánsan ($\alpha=5\%$ -os szinten)

Az első két tengelytől eltérően, a harmadik tengely erősen korrelál a makroklímával, a pozitívan az éves középhőmérséklettel és az éves hőingással, és negatívan a csapadékkal és a humiditással (22. táblázat, 40. ábra). Ez alapján a harmadik tengely kontinentalitási gradiensként értelmezhető. Ez a gradiens fontos a Deschampsion asszociációcsoporton belül: a szubóceáni klímájú területek Deschampsion gyepeire (1.3 csoport) jellemzők az Arrhenatherion asszociációcsoportba tartozó, mezikus gyepek generalista fajai (pl. *Festuca rubra*, *Trisetum flavescens*), amelyek többsége a kontinentális klímájú Pannon flóratartományban ritka. A csoport számos faja közönséges Magyarországon is (pl. *Anthoxanthum odoratum*, *Luzula campestris*, *Holcus lanatus*), de nem jellemző előfordulásuk az ártéri gyepekben.

A csoportok klímáját Kruskal-Wallis tesztel összehasonlítva (41. ábra) megállapítható, hogy a de Martonne féle humiditás index jobban jellemzi a vízellátottságot, mint a csapadék mennyisége. A Calthion csoportok – a kis méretű 2.3 csoport kivételével – szignifikánsan humidabb területeken fordulnak elő, mint a Deschampsion asszociációcsoportba tartozó csoportok. Az éves középhőmérséklet tekintetében két nagy csoport különböztethető meg: a Calthion gyepek (2.1, 2.2 és 2.3 csoport) és a szubóceánikus területek *Deschampsion* gyepei a hűvösebb, míg a többi *Deschampsion* csoport és az alföldi Molinion gyepek a melegebb területekre jellemzőek. Az éves hőingás is az első csoportban alacsonyabb és a másodikban magasabb, ami azt jelenti, hogy a szubatlantikus Calthion asszociációcsoportnak és a kontinentális Deschampsion asszociációcsoport leginkább szubóceánikus alcsoportjának termőhelye nem különbözik a kontinentalitás hőmérsékleti komponense szempontjából, de eltér a humiditásuk.

10. Nicheszélesség és az area méret kapcsolata

A nicheszélesség vizsgálatára kidolgozott egyszerű és objektív módszer (8. fejezet) lehetővé teszi, hogy eddig nehezen tesztelhető hipotéziseket is megvizsgáljunk. Ilyen az a hipotézis is, hogy a generalista fajok, mivel kevésbé válogatnak a termőhelyben, nagyobb areájúak. Számos a hipotézist bizonyító példát lehet hozni, de találhatunk ellenpéldákat is. Vagyis a feltételezett összefüggés csak tendencia jellegű, statisztikusan érvényesülő szabály, aminek a meglétét az alábbiakban a magyar flóra fajain tesztelem.

Ha a hipotézis igaz, akkor is várható, hogy vannak kis areájú generalista és nagy areájú specialista fajok (Rabinowitz 1981). Érdekes, kérdés hogy vajon ezek a fajok egyformán gyakoriak minden élőhelyen, vagy egyes közösségekben gyakrabban, másutt ritkábban fordulnak elő?

10.1. Anyag és módszer

A hazai flóra gyakoribb fajainak nicheszélességét a 8. fejezetben bemutatott módszerrel becsültem a CoenoDat adatbázisban (lásd 2. fejezet) szereplő felvételek alapján. A számítást a legalább 50 felvételen előforduló fajokra végeztem el.

A fajok areájának méretét a Flóra adatbázisban szereplő flóraelem besorolás (Rédei & Horváth 1995) alapján 7 fokozatú ordinális skálán adtam meg. A skála összeállításánál figyelembe vettem Rédei és Horváth (1995) által megadott areaméret értékeket, de helyenként módosítottam azokat, annak érdekében, hogy az egyes kategóriákba tartozó fajok számában kisebbek legyenek a különbségek (23. táblázat). Az areaméret és a nicheszélesség közötti kapcsolatot Kendall-féle rang korrelációval teszteltem (Zar 1999).

A kis areájú generalista és a nagy areájú specialista fajok leválogatásakor közepes areaméretűnek tekintettem a kontinentális és európai flóraelemeket, kis areájúnak az ennél kisebb, nagy areájúnak az ennél nagyobb elterjedési területű fajokat. Specialistának tekintettem azokat a fajokat, amelyek nicheszélessége az átlagosnál kisebb, nagy areájúnak azokat, amelyeké az átlagosnál nagyobb.

A fajokat élőhelypreferenciáik alapján – a Flóra adatbázisban szereplő cönoszisztematikai besorolásokra támaszkodva – az alábbi 12 csoportba soroltam:

1. társulásközömbös fajok
2. hínárfajok
3. mocsári és lápi fajok
4. mocsárrétek, láprétek és kaszálórétek fajtái (röviden réti fajok)
5. szikes fajok
6. szárazgyepi fajok
7. gyomok
8. általános erdei fajok
9. ligeterők (Salicetea) és láperdők (Alnetea) fajtái
10. üde erdei fajok
11. savanyú erdei fajok

12. száraz erdei fajok

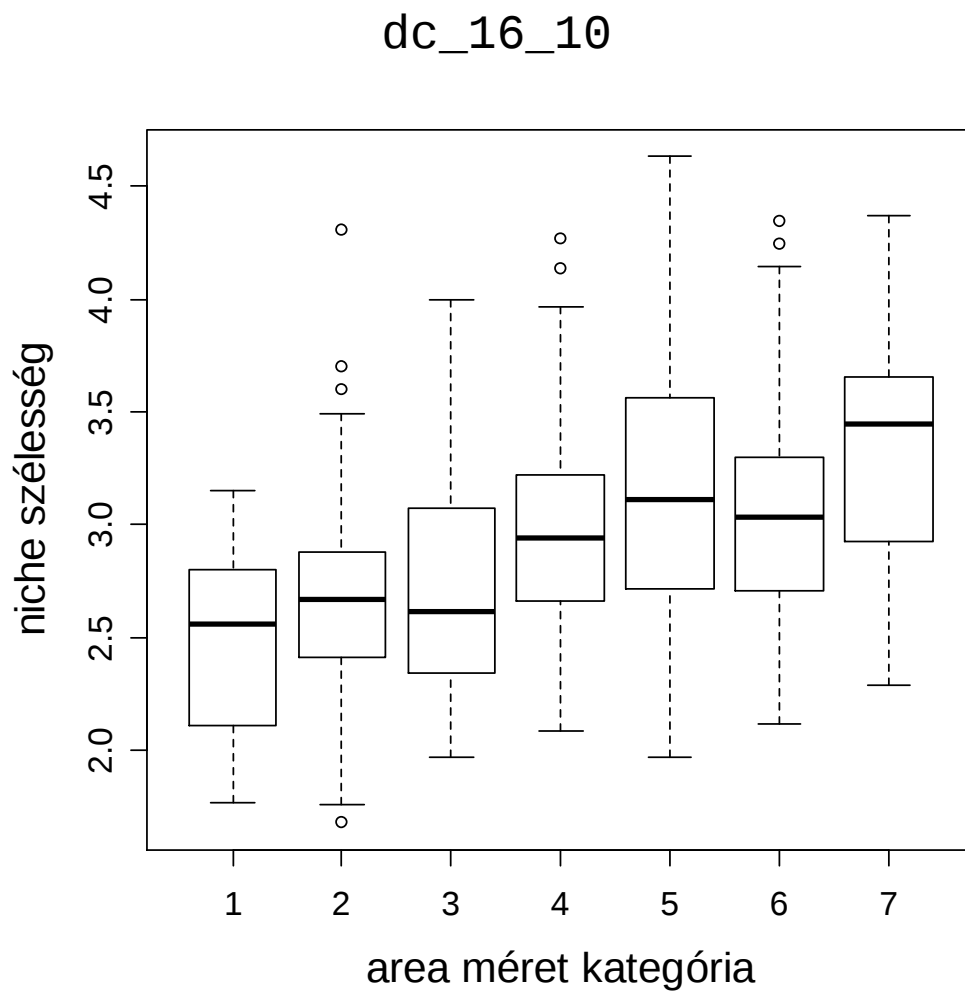
Az élőhelypreferencia kategóriák várt gyakoriságát a kis areájú specialisták és a nagy arájú generalisták csoportjában loglineáris modellel számoltam ki (Sokal & Rohlf 1981), rögzítve az élőhelypreferencia-kategóriák, valamint az areaméret és nicheszélesség kombinált kategóriák marginális gyakoriságait.

23. táblázat: A flóraelemek és a hozzájuk rendelt areaméret kategóriák.

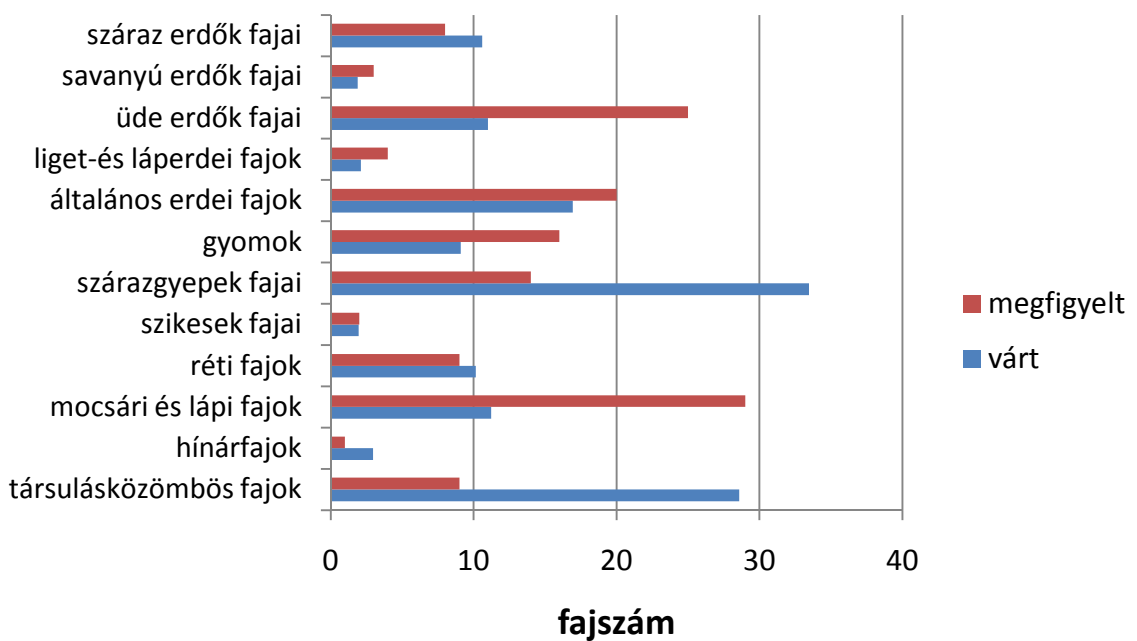
Flóraelem kategória (Rédei & Horváth 1995)	Area méret kategória Rédei és Horváth (1995) szerint	A dolgozatban használt 7 fokozatú area méret kategória	A dolgozatban használt 3 fokozatú area méret kategória
lokális endemizmusok (END)	1	1	kicsi
pannóniai endemizmusok (PAN)	1		
kárpáti endemizmusok (CAR)	2		
illír (ILL)	2		
alpin-balkáni (ALB)	3	2	
balkáni (BAL)	3		
középeurópai-alpin (CEA)	3		
közép-európai (CEU)	3		
pannon-balkáni (PaB)	3		
pontus-szubmediterrán (PoM)	3		
pontusi (PON)	3		
pontus-pannon (PoP)	3		
keleti-szubmediterrán (SMO)	3		
turáni (TUR)	3		
havasi (alpin, ALP)	4	3	
(szub)atlanti- szubmediterrán (AsM)	4		
szarmata (SAR)	4		
szubatlanti (SAT)	4		
szubmediterrán (SME)	4		
kontinentális (CON)	4	4	közepes
európai (EUR)	4		
eurázsiai (EUA)	5	5	nagy
cirkumpoláris (CIR)	6	6	
kozmozopolita (KOZ)	6	7	

10.2. Eredmények

Az elvégzett elemzés igazolta a várt pozitív összefüggést a nicheszélesség és az areaméret között (Kendall tau = 0.278, $p < 0.1\%$), bár az egyes areaméret kategóriákon belül

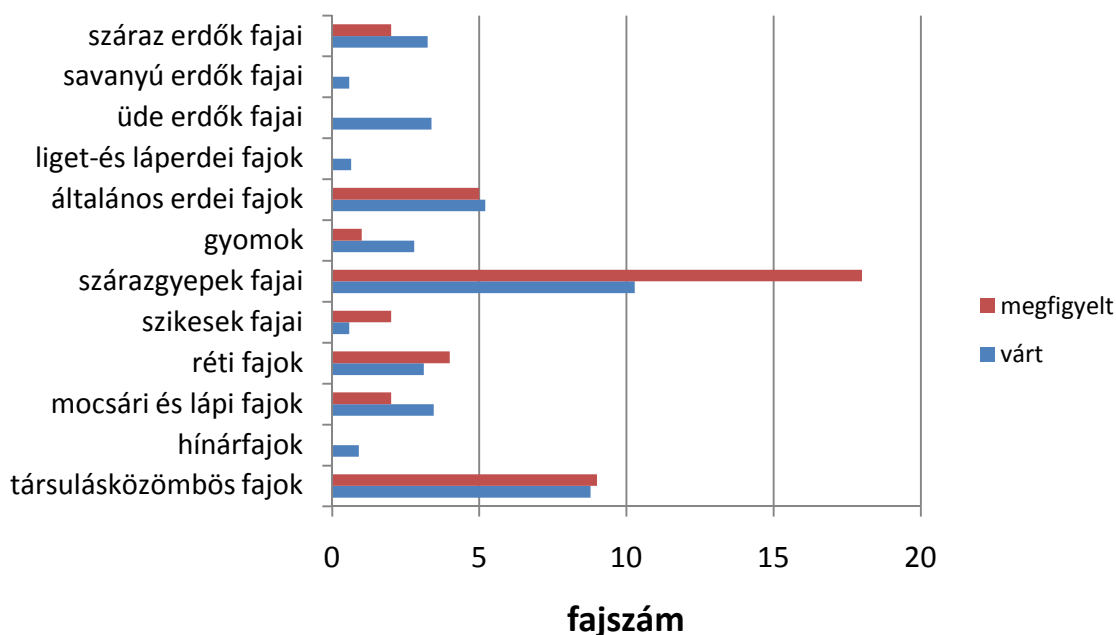


42. ábra: A hazai flóra gyakoribb (a CoenoDat adatbázis legalább 50 felvételében előforduló fajainak areamérete (a kategóriák jelentését lásd a 23. táblázatban) és nicheszélessége közötti kapcsolat.



43. ábra: A különböző élőhelypreferenciák várt és megfigyelt gyakorisága a nagy areájú specialisták között.

dc_16_10



44. ábra: A különböző élőhelypreferenciák várt és megfigyelt gyakorisága a kis areájú generalisták között.

erősen szórnak a nicheszélesség értékek (42. ábra).

A nagy areájú specialisták között a vártnál jóval nagyobb számban találunk mocsári és lápi fajokat, üde erdei fajokat és gyomokat (43. ábra), míg a szárazgyepi és társulásközömbös fajok száma a vártnál kisebb. A kis areájú generalisták között viszont a vártnál jóval több szárazgyepi fajt találunk, míg az üde erdei fajok teljesen hiányoznak (44. ábra).

10.3. Diszkusszió

Az elvégzett elemzés igazolta a várakozásunkat, hogy a generalistább fajok areája nagyobb. Ez jól magyarázható azzal, hogy ezek a fajok könnyebben alkalmazkodnak a különböző környezeti körülményekhez, így nagy területen találhatnak a számukra megfelelő élőhelyet. Egy specialista faj is lehet azonban tág elterjedésű, ha a számára alkalmas speciális élőhely elég nagy földrajzi területen előfordul. Ez a helyzet például a mocsári és lápi fajoknál, amelyek csak megfelelő vízellátás, trofitás és pH viszonyok mellett fordulnak elő, de ezek a termőhelyek az északi mérsékelt övben általánosan elterjedtek. A rendszeres emberi bolygatáshoz alkalmazkodott gyomok életfeltételeit világszerte megteremti az emberi tevékenység. A leginkább meglepő az üde erdei fajok magas száma a nagy areájú specialisták között. Az ide fajok többsége (a 25 fajból 18) eurázsiai flóraelem. Ebben az esetben a specializálódást a zárt erdők fényszegény körülményeihez való alkalmazkodás jelenti és ezek a körülmények Eurázsia nagy területein megtalálhatók, mégha az erdők uralkodó fafaja területenként eltérő is.

Míg a tág areájú specialisták élőhelypreferenciáit elméleti megfontolások alapján – legalább részben – előzetesen is megjósolhattuk, sokkal nehezebb lenne előrejelezni a kis areájú, generalista szárazgyepi fajok magas számát. A jelenség hátterében valószínűleg az áll,

hogy a Kárpát-medencében szárazgyepek nagy kiterjedésben és változatos termőhelyi körülmények (pl. különböző alapkőzetek) között fordulnak elő, ami lehetővé teszi, hogy sok generalista szárazgyepi fajunk legyen. Ezek között magas a pontusi és szubmediterrán fajok aránya. Érdeemes lenne filogeográfiai módszerekkel megvizsgálni ezeknek a fajoknak az elterjedéstörténetét: valószínűleg a balkáni és Fekete-tenger környéki jégkorszaki refúgiumokból rajzottak szét és a refúgiumokban lezajló speciáció is feltételezhető.

Az eredmények földrajzi érvényességi területe is érdekes kérdés. A specialista fajok kisebb várható areamérete valószínűleg általános érvényű. Szintén nagy területen érvényesnek gondolom a gyomok és a mocsári, lápi fajok magas arányát nagy areájú specialisták között. A szárazgyepi fajok nagy száma a kis areájú generalisták között viszont valószínűleg csak a Kárpát-medencére és a szomszédos területekre érvényesek, és más eredményeket kapnánk, ha megismételnénk a vizsgálatokat a nyugat-európai lomboserdő-zónában.

11. Összegzés és kitekintés

A cönológiai adatbázisok a jövőben várhatóan fontos eszközei lesznek az ökológiai kutatásoknak. Ehhez azonban ki kell dolgozni a megfelelő elemzőmódszereket is. A dolgozat nagyobbik része az ezen a téren elért eredményeimről szól. A fejlesztések egy része a korábban, kisebb adatsorokon is használt numerikus klasszifikációs eljárásokhoz kapcsolódik. Ezeket azt tette szükségessé, hogy ha sok felvételt elemzünk, akkor nehéz az eredmények minden részletét áttekinteni, szükség van az eredmények értékelését segítő módszerekre. Más módszerek olyan kérdésekhez kapcsolódnak (fajkészlet és nicheszélesség becslése), amelyek numerikus vizsgálatára nagy adatbázisok hiányában korábban nem volt lehetőség.

A módszerfejlesztések nem érnek véget az elméletileg jól működő módszerek kitalálásával. Azok alkalmasságát tesztelni kell és ha ugyanarra a problémára több módszer is létezik érdemes azok hatékonyságát is összehasonlítani. A dolgozatban nagy figyelmet fordítottam erre a két utóbbi lépésre. A módszerek tesztelése során a terepi adatok mellett szimulált adatokat is használtam, amelyeknél a várt eredmény előre ismert.

Az adatbázisok használatát bemutató két esettanulmány a dolgozat kisebbik részét teszi ki. Ennek oka, hogy az alkalmazások feltétele a megfelelő méretű adatbázis és a már kidolgozott módszerek. Az adatbázisok kiépülésével (lásd 1. és 2. fejezet) és a módszertani kérdések tisztázásával az ilyen tanulmányok száma várhatóan emelkedni fog mind a nemzetközi, mind a hazai szakirodalomban.

A cönológiai adatbázisok kiépülésével párhuzamosan több országban is új vegetációmonográfiák készültek (pl. Chytrý 2007, 2009, Janišová et al. 2007) és hasonló készítését Magyarországra is tervezzük (Csiky et al. 2008). A klasszifikációhoz kapcsolódó módszertani fejlesztések elsősorban ezekben a munkákban, illetve az ezeket megalapozó tanulmányokban (pl. Botta-Dukát et al. 2005, Illyés et al. 2007, 2009) hasznosulnak.

A fajkészlet becslése a társulási szabályok vizsgálatában játszhat a jövőben fontos szerepet, de emellett felhasználható a védett, ritka fajok potenciális termőhelyeinek megkeresésében is (pl. Münzbergová & Herben 2004), ezzel növelve a restaurációs beavatkozások hatékonyságát.

Számos érdekes és egyelőre kihasználatlan lehetőséget rejt a cönológiai adatbázisok összekapcsolása a növényi tulajdonság (trait) adatbázisokkal (pl. Klotz et al. 2002). Ezek közül talán a legfontosabb, hogy igen nagy mintákon vizsgálható, hogy az együtt élő fajok tulajdonságainak eloszlása milyen irányban tér el a random esetben várttól (pl. Fukami et al. 2005, Grime 2006, Mouillot, Dumay, et al. 2007, Mouillot, Mason, et al. 2007). Szintén érdekes kérdés, hogy milyen tulajdonságokban térnek el a specialista és a generalista fajok. Az ilyen hazai vizsgálatoknak jelenleg a legnagyobb korlátja a teljes hazai flórára vonatkozó sok tulajdonságot tartalmazó adatbázis hiánya (Csecserits et al. 2009), mivel a nyugat-európai adatbázisokból számos nálunk gyakori faj is hiányzik.

A cönológiai adatbázisokban található irodalmi felvételekhez általában nincsenek környezeti háttéradatok, de az újabb felvételekhez már egyre több esetben ilyen adatokat is gyűjtenek, ami lehetővé teszi a fajok előfordulását meghatározó környezeti tényezők

dc_16_10

vizsgálatát is (pl. Pinke et al. 2010).

Egy teljes lista összeállítása reménytelen feladat lenne, de remélhetőleg a fenti kiragadott példák is jól szemléltetik, hogy mennyire sokféle területen használhatók fel a kutatásban a cönológiai adatbázisok.

12. Köszönetnyilvánítás

Köszönöm szerzőtársaimnak – Bagi Istvánnak, Bauer Norbertnek, Borhidi Attilának, Milan Chytrýnek, Csiky Jánosnak, Dobolyi Konstantinnak, Garadnai Jánosnak, Hahn Istvánnak, Milan Hájeknek, Ondrej Hájeknek, Petra Hajkovának, Marcela Havlovának, Jason Holtnak, Horváth Ferencnek, Illyés Eszternek, Ute Jandtnak, Monika Janišovának, Kertész Miklósnak, Kovács J. Attilának, Kovács-Láng Editnek, Lájér Konrádnak, Rédei Tamásnak, Iveta Škodovának, Stephen S. Talbotnak, Lubomír Tichýnek, Szmorad Ferencnek, és Wolfgang Willnernek – a közös munka során tőlük kapott segítségért.

A dolgozatban szereplő cikkek elkészítése során számos hasznos tanácsot kaptunk Bartha Sándortól, Helge Bruelheide-től, Miquel De Cácerestől, Jürgen Denglertől, Philip Dixontól, Jörg Ewaldtól, Fekete Gábortól, Norbert Hölzeltől, Kalapos Tibortól, Meelis Pärteltől, Podani Jánostól, Somodi Imeldától és Katka Šumberovától.

A dolgozat témaválasztását inspirálta a CoenoDAT adatbázis létrehozása, amelynek hiányában a vizsgálatok egy részének elvégzése nem is lett volna lehetséges. Ezúton is köszönöm a „Magyarország természetes növényzeti örökségének felmérése és összehasonlító értékelése” NKTH program anyagi támogatását, valamint a cönológiai adatgyűjtésben és az adatok digitalizálásában résztvevő valamennyi kolléga munkáját, amely nélkül a cönológiai adatbázis nem jöhetett volna létre.

Köszönöm munkatársaimnak az MTA ÖBKI Növényökológiai Osztályán azt a kellemes és stimuláló munkahelyi légkört, ami segítségemre volt eddigi munkámban és lehetővé tette a dolgozat elkészítését.

Végül, de nem utolsósorban köszönöm családomnak a támogatásukat és a türelmüket a dolgozatírás hónapjaiban.

13. Irodalomjegyzék

- Aho, K., Roberts, D.W. & Weaver, T. 2008. Using geometric and non-geometric internal evaluators to compare eight vegetation classification methods. *Journal of Vegetation Science* **19**: 549-562.
- Anderson, M.J. & Clements, A. 2000. Resolving environmental disputes: a statistical method for choosing among competing cluster models. *Ecological Applications* **10**: 1341-1355.
- Baker, F.B. & Hubert, L.J. 1975. Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association* **70**: 31-38.
- Balátová-Tuláčková, E. 1966. Synökologické Charakteristik der südmährischen Überschwemmungswiesen. *Rozpr. Čs. Akad. Věd, ser. math.-natur.* **76**: 1-41.
- Balátová-Tuláčková, E. 1968. Grundwasserganglinien und Wiesengesellschaften (Vergleichende Studie der Wiesen aus Südmähren und der Südwestslowakei). *Acta Sci. Nat. Brno* **2**: 1-37.
- Balátová-Tuláčková, E. 1969. Beitrag zur Kenntnis der tschechoslowakischen Cnidion venosi-Wiesen. *Vegetatio* **17**: 200-207.
- Balátová-Tuláčková, E. 1979. Synökologické Verhältnisse der Filipendula ulmaria-Gesellschaften NW-Böhmens. *Folia Geobotanica & Phytotaxonomica* **14**: 225-258.
- Balátová-Tuláčková, E. 1981. Phytozoölogische und synökologische Charakteristik der Feuchtwiesen NW Böhmens. *Rozpr. Čs. Akad. Věd, ser. math.-natur.* **91**: 1-90.
- Balátová-Tuláčková, E. 1993. Das Gentiano pneumonanthis-Molinietum litoralis Ilijanič 1968 in Süd-Mähren und der Slowakei. *Tuexenia* **13**: 193-201.
- Balátová-Tuláčková, E. 1997a. Wiesengesellschaften im südlichen Teil der Talebene Hornomoravský úval (Mittelmähren). *Preslia* **69**: 253-282.
- Balátová-Tuláčková, E. 1997. Feuchtwiesen- und Hochstaudengesellschaften des Landschaftsschutzgebietes Lužické hory und der angrenzenden Randgebiete (Nordböhmen). *Verh. Zool.-Bot. Ges. Österr.* **134**: 233-304.
- Balátová-Tuláčková, E. & Hübl, E. 1974. Über die Phragmitetea- und Molinietales-Gesellschaften in der Thaya-, March- und Donau-Aue Österreichs. *Phytocoenologia* **1**: 263-305.
- Balátová-Tuláčková, E. & Knežević, M. 1975. Beitrag zur Kenntnis der Überschwemmungswiesen in der Drava- und Karašica-Aue (Nord-Jugoslawien). *Acta Botanica Croatica* **34**: 63-80.
- Bartha, D., Király, G. & Molnár, Z. 2002. A botanikus szakma nagy terve: Magyarország természetes növényzeti örökségének felmérése és összehasonlító értékelése. In: *Magyar botanikai kutatások az ezredfordulón. Tanulmányok Borhidi Attila 70. születésnapja tiszteletére* (szerk. Salamon-Albert, É.). PTE Növénytani Tanszék, Pécs, o. 309-342.
- Bartha, S. & Ittész, P. 2001. Local richness - species pool ratio: a consequence of the species-area relationship. *Folia Geobotanica & Phytotaxonomica* **36**: 9-23.
- Baselga, A., Jiménez-Valverde, A. & Niccolini, G. 2007. A multiple-site similarity measure independent of richness. *Biology Letters* **3**: 642-645.
- Beals, E. 1984. Bray-Curtis-ordination: an effective strategy for analysis of multivariate ecological data. *Adv. Ecol. Res.* **14**: 1-55.
- Beauchaine, T.P. & Beauchaine, R.J. 2002. A Comparison of Maximum Covariance and K-Means Cluster Analysis in Classifying Cases Into Known Taxon Groups. *Psychological Methods* **7**: 245-261.
- Bekker, R.M., van der Maarel, E., Bruelheide, H. & Woods, K. 2007. Long-term datasets: From descriptive to predictive data using ecoinformatics. *Journal of Vegetation*

- Science* **18**: 457-462.
- Belbin, L. & McDonald, C. 1993. Comparing three classification strategies for use in ecology. *Journal of Vegetation Science* **4**: 341-348.
- Belyea, L.R. & Lancaster, J. 1999. Assembly rules within a contingent ecology. *Oikos* **86**: 402-416.
- Berg, C. & Dengler, J. 2004. Von der Datenbank zur Regionalmonografie: Erfahrungen aus dem Projekt „Die Pflanzengesellschaften Mecklenburg-Vorpommerns und ihre Gefährdung“. *Ber. Reinhold-Tüxen-Ges.* **16**: 29-56.
- Berg, C., Dengler, J., Abdank, A. & Isermann, M. (Szerk.). 2004. *Die Pflanzengesellschaften Mecklenburg-Vorpommerns und ihre Gefährdung*. Wiessdorn Verlag, Jena.
- Bertrand, P. & Bel Mufti, G. 2006. Loevinger's measures of rule quality for assessing cluster stability. *Computational Statistics & Data Analysis* **50**.
- Bezdek, J. & Pal, N. 1998. Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics* **28**: 301-315.
- Blažková, D. 1993. Vegetace polabských nivních luk Kelšnice u Mělníka (Vegetation of the Elbe floodplain meadows in Kelšnice near Mělník). *Muz. Souč., ser. natur.* **7**: 35-64.
- Bock, H.H. 1996. Probabilistic models in cluster analysis. *Computational Statistics & Data Analysis* **23**: 5-28.
- Bodrogközi, G. 1962. Die Vegetation des Theiss-Wellenraumes. I. Zönologische und ökologische Untersuchungen in der Gegend von Tokaj. *Acta Biologica Szeged* **8**: 3-44.
- Borhidi, A. 1963. Die zönologie des Fagion illyricum Verbandes I. Allgemeiner Teil. *Acta Botanica Academiae Scientiarum Hungaricae* **9**: 259-297.
- Borhidi, A. 1965. Die zönologie des Fagion illyricum Verbandes II. Systematischer Teil. *Acta Botanica Academiae Scientiarum Hungaricae* **11**: 53-102.
- Borhidi, A. 1995. Social behaviour types, the naturalness and relative ecological indicator values of the higher plants in the Hungarian Flora. *Acta Botanica Hungarica* **39**: 97-181.
- Borhidi, A. 2003. *Magyarország növénytársulásai (Plant communities of Hungary)*. Akadémiai Kiadó, Budapest.
- Borhidi, A. & Sánta, A. 1999. *Vörös könyv Magyarország növénytársulásairól*. Természetbúvár Alapítvány Kiadó, Budapest.
- Bosáčková, E. 1971. Príspevok k ochrane Podunajských slatín (A contribution to conservation of the Danube fens). *Čs. Ochr. Prír.* **11**: 195-210.
- Bosáčková, E. 1975. Rastlinné spoločenstvá slatinných lúk na Záhorskej nížine (Plant communities of fen meadows in Záhorská Lowland). *Čs. Ochr. Prír.* **15**: 173-273.
- Botta-Dukát, Z. 2008a. Validation of hierarchical classifications by splitting dataset. *Acta Botanica Hungarica* **50**: 73-80.
- Botta-Dukát, Z. 2008b. Invasion of alien species to Hungarian (semi-)natural habitats. *Acta Botanica Hungarica* **50**: 219-227.
- Botta-Dukát, Z. & Borhidi, A. 1999. New objective method for calculating fidelity. Example: The Illyrian beechwoods. *Annali di Botanica (Roma)* **57**: 73-90.
- Botta-Dukát, Z., Chytrý, M., Hajková, P. & Havlová, M. 2005. Vegetation of lowland wet meadows along a climatic continentality gradient in Central Europe. *Preslia* **77**: 89-111.
- Botta-Dukát, Z., Kovács-Láng, E., Rédei, T., Kertész, M. & Garadnai, J. 2007. Statistical and biological consequences of preferential sampling in phytosociology: theoretical considerations and a case study. *Folia Geobotanica* **42**: 141-152.
- ter Braak, C.F.J. & Šmilauer, P. 2002. *CANOCO reference manual and CanoDraw for Windows user's guide. Software for Canonical Community Ordination (version 4.5)*. Biometris, Wageningen & České Budějovice.

- Bradfield, G.E. & Kenkel, N.C. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. *Ecology* **68**: 750-753.
- Braun-Blanquet, J. 1928. *Pflanzensoziologie. Grundzüge der Vegetationskunde*. Springer, Berlin.
- Breckenridge, J.N. 1989. Replicating cluster-analysis – methods, consistency, and validity. *Multivariate Behavioral Research* **24**: 147-161.
- Brisse, H., de Ruffray, P., Grandjouan, G. & Hoff, M. 1995. The Phytosociological Database “SOPHY” Part 1: Calibration of indicator plants, Part II: Socio-ecological classification of the relevés. *Annali di Botanica (Roma)* **53**: 177-223.
- Bruelheide, H. 1995. Die Grünlandgesellschaften des Harzes und ihre Standortbedingungen. Mit einem Beitrag zum Gliederungsprinzip auf der Basis von statistisch ermittelten Artengruppen. *Diss. Bot.* **244**: 1-338.
- Bruelheide, H. 2000. A new measure of fidelity and its application to defining species groups. *Journal of Vegetation Science* **11**: 167-178.
- Bruelheide, H. & Chytrý, M. 2000. Towards unification of the national vegetation classifications: a comparison of two methods for the analysis of large data sets. *Journal of Vegetation Science* **11**: 295-306.
- Burkart, M., Dierschke, H., Hölzel, N., Nowak, B. & Fartmann, T. 2004. *Synopsis der Pflanzengesellschaften Deutschlands. Heft 9. Molinio-Arrhenatheretea (E1). Kulturgrasland und verwandte Vegetationstypen. Teil 2: Molinietales. Futter- und Streuwiesen feucht-nasser Standorte und Klassenübersicht Molinio-Arrhenatheretea*. Floristisch-soziologische Arbeitsgemeinschaft, Göttingen.
- Calinski, R.B. & Harabasz, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics* **3**: 1-27.
- Casado, M.A., Ramírez-Sanz, L., Castro, I., de Miguel, J.M. & de Pablo, C.L. 1997. An objective method for partitioning dendrograms based on entropy parameters. *Plant Ecology* **131**: 193-197.
- Chang, F., Carey, V., Qiu, W., Zamar, R.H., Lazarus, R. & Wang, X. 2009. *clues: Clustering Method Based on Local Shrinking*. . R package version 0.3.2. <http://CRAN.R-project.org/package=clues>
- Chiang, M. & Mirkin, B. 2010. Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. *Journal of Classification* **27**: 3-40.
- Chiarucci, A. 2007. To sample or not to sample? That is the question ... for the vegetation scientist. *Folia Geobotanica* **42**: 209-216.
- Chytrý, M., Maskell, L.C., Pino, J., Pyšek, P., Vilà, M., Font, X. & Smart, S.M. 2008. Habitat invasions by alien plants: a quantitative comparison among Mediterranean, subcontinental and oceanic regions of Europe. *Journal of Applied Ecology* **45**: 448-458.
- Chytrý, M., Pyšek, P., Wild, J., Pino, J., Maskell, L.C. & Vilà, M. 2009. European map of alien plant invasions based on the quantitative assessment across habitats. *Diversity and Distributions* **15**: 98-107.
- Chytrý, M. & Rafajová, M. 2003. Czech National Phytosociological Database: basic statistics of the available vegetation-plot data. *Preslia* **75**: 1-15.
- Chytrý, M., Tichý, L. & Roleček, J. 2003. Local and Regional Patterns of Species Richness in Central European Vegetation Types along the pH/Calcium gradient. *Folia Geobot. Phytotax.* **38**: 429-442.
- Chytrý, M. 2001. Phytosociological data give biased estimates of species richness. *Journal of Vegetation Science* **12**: 439-444.
- Chytrý, M. (Szerk.). 2007. *Vegetace České republiky 1. Travinná a keříčková vegetace*.

- Vegetation of the Czech Republic 1. Grassland and heathland vegetation*. Academia, Praha.
- Chytrý, M. (Szerk.). 2009. *Vegetace České republiky 2. Ruderální, plevelová, skalní a suťová vegetace (Vegetation of the Czech Republic 1. Ruderal, weed, rock and scree vegetation)*. Academia, Praha.
- Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* **13**: 79-90.
- Clarke, K.R. 1993. Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* **18**: 117-143.
- Clavero, M. & Brotons, L. 2010. Functional homogenization of bird communities along habitat gradients: accounting for niche multidimensionality. *Global Ecology and Biogeography* **19**: 684-696.
- Colwell, R.K. & Coddington, J.A. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. Roy. Soc. London B* **345**: 101-118.
- Cornell, H.V. & Lawton, J.H. 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology* **61**: 1-12.
- Coudun, C. & Gégout, J. 2007. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. *Journal of Vegetation Science* **18**: 517-524.
- Csecserits, A. & Rédei, T. 2001. Secondary succession on sandy old-fields in Hungary. *Applied Vegetation Science* **4**: 63-74.
- Csecserits, A., Szabó, R. & Czúcz, B. 2009. Növényi tulajdonságok, tulajdonság-adatbázisok és ezek felhasználása az ökológiai kutatásokban. *Botanikai Közlemények* **96**: 31-47.
- Csiky, J., Babai, D., Botta-Dukát, Z., Horváth, F., Lájér, K. & Mesterházy, A. 2008. New synthesis of the Hungarian vegetation I. Freshwater aquatic vegetation of Hungary. Using phytosociological data to address ecological questions. In: *17th International Workshop of European Vegetation Survey. Brno, 1-5. May 2008*.
- Csíky, J. 2002. *A Nógrád-Gömöri bazaltvidék flórája és vegetációja* (PhD értekezés).
- Csontos, P. 1996. *Az aljnövényzet változásai cseres-tölgyes erdők regenerációs szukcessziójában*. Scientia Kiadó, Budapest.
- Datta, S. & Datta, S. 2003. Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics* **19**: 459-466.
- Davies, D. & Bouldin, D. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**: 224-227.
- De Cáceres, M., Font, X. & Oliva, F. 2010. The management of numerical vegetation classifications with fuzzy clustering methods. *Journal of Vegetation Science* **in press**.
- De Cáceres, M., Font, X., Vicente, P. & Oliva, F. 2009. Numerical reproduction of traditional classifications and automatic vegetation identification. *Journal of Vegetation Science* **20**: 620-628.
- De Cáceres, M. & Legendre, P. 2008. Beals smoothing revisited. *Oecologia* **156**: 657-669.
- De Cáceres, M. & Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* **90**: 3566-3574.
- De Cáceres, M., Legendre, P. & Moretti, M. 2010. Improving indicator species analysis by combining groups of sites. *Oikos*. <http://dx.doi.org/10.1111/j.1600-0706.2010.18334.x>
- De'ath, G. 1999. Extended dissimilarity: a method of robust estimation of ecological distances from high beta diversity data. *Plant Ecology* **144**: 191-199.
- Dengler, J., Rūsiņa, S., Boch, S., Bruun, H.H., Diekmann, M., Dierßen, K., Dolnik, C., Dupre, C., Golub, V.B., Grytnes, J., Helm, A., Ingerpuu, N., Löbel, S., Pärtel, M., Rašomavičius, V., Tyler, G., Znamenskiy, S.R. & Zobel, M. 2006. Working group on

- dry grasslands in the Nordic and Baltic region – Outline of the project and first results for the class Festuco-Brometea. *Annali di Botanica (Roma) nuova series* **6**: 1-28.
- Devictor, V., Julliard, R., Clavel, J., Jiguet, F., Lee, A. & Couvet, D. 2008. Functional biotic homogenization of bird communities in disturbed landscapes. *Global Ecology and Biogeography* **17**: 252-261.
- Devictor, V. & Robert, A. 2009. Measuring community responses to large-scale disturbance in conservation biogeography. *Diversity and Distributions* **15**: 122-130.
- Diekmann, M., Kühne, A. & Isermann, M. 2007. Random vs non-random sampling: effects on patterns of species abundance, species richness and vegetation-environment relationships. *Folia Geobotanica* **42**: 179-190.
- Dierschke, H. 1981. Zur syntaxonomischen Bewertung schwach gekennzeichnete Pflanzengesellschaften. In: *Syntaxonomie* (szerk. Dierschke, H.). J. Cramer, Vaduz, FL., o. 109-122.
- Dierschke, H. 1994. *Pflanzensoziologie*. Verlag Eugen Ulmer, Stuttgart.
- Dierschke, H. 1995. Syntaxonomical survey of Molinio-Arrhenatheretea in Central Europe. *Colloques Phytosociologiques* **23**: 387-395.
- Dimitriadou, E., Dolnicar, S. & Weingessel, A. 2002. An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets. *Psychometrika* **67**: 137-160.
- Diserud, O.H. & Ødegaard, F. 2007. A multiple-site similarity measure. *Biology Letters* **3**: 20-22.
- Dobolyi, K. 2002. Útmutató a növényökológiai felvételekben használandó taxonlistához.
- Duchoslav, M. 1997. The present state of meadow vegetation (Molinio-Arrhenatheretea) in the Morava river floodplain (Hornomoravský úval area). *Zpr. Čes. Bot. Společ., Mater.* **15**.
- Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. 2000. Vegetation gradients in Atlantic Europe: the use of existing phytosociological data in preliminary investigations on the potential effects of climate change on British vegetation. *Global Ecology and Biogeography* **9**: 187-199.
- Dufrene, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* **67**: 345-366.
- Dunn, J.C. 1974. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics* **4**: 95 - 104.
- Dupré, C. 2000. How to determine a regional species pool: a study in two Swedish regions. *Oikos* **89**: 128-136.
- Ehrendorfer, F. 1973. *Liste der Gefäßpflanzen Mitteleuropas*. 2. edn. Gustav Fischer Verlag, Stuttgart.
- Ellenberg, H. 1996. *Vegetation Mitteleuropas mit den Alpen in ökologischer, dynamischer und historischer Sicht*. 5. edn. Verlag Eugen Ulmer, Stuttgart.
- Ellmauer, T. & Mucina, L. 1993. Molinio-Arrhenatheretea. In: *Die Pflanzengesellschaften Österreichs. Teil I. Anthropogene Vegetation* (szerk. Mucina, L., Grabherr, G. & Ellmauer, T.). Gustav Fischer Verlag, Jena, o. 297-401.
- Equihua, M. 1990. Fuzzy clustering of ecological data. *Journal of Ecology* **78**: 519-534.
- Eriksson, O. & Ehrlén, J. 1992. Seed and microsite limitation of recruitment in plant populations. *Oecologia* **91**: 360-364.
- Erzberger, P. & Papp, B. 2004. Annotated checklist of Hungarian bryophytes. *Studia Botanica Hungarica* **35**: 91-149.
- Ewald, J. 1995. Eine vegetationskundliche Datenbank bayerischer Bergwälder. *Hoppea* **56**: 453-465.
- Ewald, J. 2001. Der Beitrag pflanzensoziologischer Datenbanken zur

- vegetationsökologischen Forschung. *Ber. Reinhold-Tüxen-Ges.* **13**: 53-69.
- Ewald, J. 2002. A probabilistic approach to estimating species pools from large compositional matrices. *Journal of Vegetation Science* **13**: 191-198.
- Ewald, J. 2003. The calcareous riddle: Why are there so many calciphytic species in the Central European flora? *Folia Geobotanica* **38**: 357-366.
- Faith, D.P., Minchin, P.R. & Belbin, L. 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**: 57-68.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.
- Feoli, E., Ferro, G. & Ganis, P. 2006. Validation of phytosociological classifications based on a fuzzy set approach. *Community Ecology* **7**: 99-108.
- Feoli, E., Gallizia-Vuerich, L., Ganis, P. & Woldu, Z. 2009. A classificatory approach integrating fuzzy set theory and permutation techniques for land cover analysis: a case study on a degrading area of the Rift Valley (Ethiopia). *Community Ecology* **10**: 53-64.
- Feoli, E., Lagonegro, M. & Biondini, F. 1981. Strategies in syntaxonomy: a discussion of two classifications of grasslands of Friuli (Italy). In: *Syntaxonomie* (szerk. Dierschke, H.), J. Cramer, Vaduz, FL.
- Feoli, E. & Lausi, D. 1980. Hierarchical levels in syntaxonomy based on information functions. *Vegetatio* **42**: 113-115.
- Font, X. & Ninot, J. 1995. A regional project for drawing up inventories of flora and vegetation in Catalonia (Spain). *Annali di Botanica (Roma)* **53**: 99-105.
- Font, X., Soriano, I. & Vigo, J. 1998. El mostratge fitocenològic a Catalunya. *Acta Bot. Barcin.* **45**: 501-515.
- Fox, J.W. 2006. Predicting local-regional richness relationship using island biogeography models. *Oikos* **113**: 376-382.
- Fridley, J.D., Vandermast, D.B., Kuppinger, D.M., Manthey, M. & Peet, R.K. 2007. Co-occurrence based assessment of habitat generalists and specialists: a new approach for the measurement of niche width. *Journal of Ecology* **95**: 707-722.
- Fukami, T., Martijn Bezemer, T., Mortimer, S.R. & van der Putten, W.H. 2005. Species divergence and trait convergence in experimental plant community assembly. *Ecology Letters* **8**: 1283-1290.
- Gauch, H.G. 1982. Noise reduction by eigenvector ordinations. *Ecology* **63**: 1643-1649.
- Gégout, J., Coudun, C., Bailly, G. & Jabiol, B. 2005. EcoPlant: A forest site database linking floristic data with soil and climate variables. *Journal of Vegetation Science* **16**: 257-260.
- Gerdol, R., Ferrari, C. & Piccoli, F. 1985. Correlation between soil characters and forest types: a study in multiple discriminant analysis. *Plant Ecology* **60**: 49-56.
- Goodall, D.W. 1953. Objective methods for the classification of vegetation II. Fidelity and indicator value. *Australian Journal of Botany* **1**: 446-456.
- Goodman, L.A. & Kruskal, W.H. 1954. Measures of association for cross classifications. *Journal of the American Statistical Association* **49**: 732-764.
- Gordon, A.D. 1999. *Classification*. 2. edn. Chapman & Hall/CRC, Boca Raton.
- Grabherr, G., Reiter, K. & Willner, W. 2003. Towards objectivity in vegetation classification: the example of the Austrian forests. *Plant Ecology* **169**: 21-34.
- Grace, J.B. & Pugsek, B.H. 1997. A Structural Equation Model of Plant Species Richness and Its Application to a Coastal Wetland. *The American Naturalist* **149**: 436.
- Grime, J.P. 2006. Trait convergence and trait divergence in herbaceous plant communities: Mechanisms and consequences. *Journal of Vegetation Science* **17**: 255-260.
- Guidi, L., Ibanez, F., Calcagno, V. & Beaugrand, G. 2009. A new procedure to optimize the

- selection of groups in a classification tree: Applications for ecological data. *Ecological Modelling* **220**: 451-461.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147-186.
- Hahn, I. & Scheuring, I. 2003. The effect of measurement scales on estimating vegetation cover: a computer experiment. *Community Ecology* **4**: 29-33.
- Hájek, M. & Hájková, P. 2004. Environmental determinants of variation in Czech Calthion wet meadows: a synthesis of phytosociological data. *Phytocoenologia* **34**: 33-54.
- Hajková, P. 2000. Rostlinná společenstva mokřadních luk, potočních rákosin a porostů vysokých ostřic v Hostýnských vrších (Plant communities of wet meadows, alluvial reeds and tall-sedge stands in the Hostýnské Mts.). *Sborn. Přír. Kl. Uh. Hradiště* **5**: 7-51.
- Hakes, W. 1994. On the predictive power of numerical and Braun-Blanquet classification: an example from beechwoods. *Journal of Vegetation Science* **5**: 153-160.
- Halkidi, M., Batistakis, Y. & Vazirgiannis, M. 2002. Clustering validity checking methods: part II. *SIGMOD Rec.* **31**: 19-27.
- Hallgren, E., Palmer, M.W. & Milberg, P. 1999. Data diving with cross-validation: an investigation of broad-scale gradients in Swedish weed communities. *Journal of Ecology* **87**: 1037-1051.
- Handl, J., Knowles, J. & Kell, D.B. 2005. Computational cluster validation in post-genomic data analysis. *Bioinformatics* **21**: 3201-3212.
- Havlová, M., Chytrý, M. & Tichý, L. 2004. Diversity of hay meadows in the Czech Republic: major types and environmental gradients. *Phytocoenologia* **34**: 551-567.
- Hédl, R. 2007. Is sampling subjectivity a distorting factor in surveys for vegetation diversity? *Folia Geobotanica* **42**: 191-198.
- Hegedúšová, K. 2007. Centrálna databáza fytocenologických zápisov (CDF) na Slovensku. *Bull. Slov. Bot. Spoločn.* **29**: 124-129.
- Hennekens, S.M. & Schaminée, J.H. 2001. TURBOVEG, a comprehensive data base management system for vegetation data. *Journal of Vegetation Science* **12**: 589-591.
- Hennig, C. 2007. Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis* **52**: 258-271.
- Herben, T. 2000. Correlation between richness per unit area and the species pool cannot be used to demonstrate the species pool effect. *Journal of Vegetation Science* **11**: 123-126.
- Hillebrand, H. 2005. Regression of local on regional diversity to not reflect the importance of local interactions or saturation of local diversity. *Oikos* **110**: 195-198.
- Holeksa, J. & Wozniak, G. 2005. Biased vegetation patterns and detection of vegetation changes using phytosociological databases. A case study in the forest of the Babia Góra National Park (the West Carpathians, Poland). *Phytocoenologia* **35**: 1-18.
- Horváth, F., Dobolyi, Z.K., Morschhauser, T., Lőkös, L., Karas, L. & Szerdahelyi, T. (Szerk.). 1995. *FLÓRA adatbázis 1.2*. MTA-ÖBKI & MTM Növénytára, Vácrátót.
- Horvatić, S. 1930. Soziologische Einheiten der Niederungswiesen in Kroatien und Slavonien. *Acta Bot. Inst. Bot. Univ. Zagreb* **5**: 57-118.
- Hölzel, N. 2003. Re-assessing the ecology of rare flood-meadow violets (*Viola elatior*, *V. pumila* and *V. persicifolia*) with large phytosociological data sets. *Folia Geobotanica* **38**: 281-298.
- Hrivnák, R., Ujházy, K., Chytrý, M. & Valachovič, M. 2003. The database of the Western Carpathian forest vegetation. *Thaiszia* **13**: 89-95.
- Hubert, L. & Arabie, P. 1985. Comparing partitions. *Journal of Classification* **2**: 193-218.
- Hubert, L.J. & Levin, J.R. 1976. A general statistical framework for assessing categorical

- clustering in free recall. *Psychological Bulletin* **83**: 1072-1080.
- Hunter, J.C. & McCoy, R.A. 2004. Applying randomization tests to cluster analyses. *Journal of Vegetation Science* **15**: 135-138.
- Illyés, E., Bauer, N. & Botta-Dukát, Z. 2009. Classification of semi-dry grassland vegetation in Hungary. *Preslia* **81**: 239-260.
- Illyés, E., Chytrý, M., Botta-Dukát, Z., Jandt, U., Škodová, I., Janišová, M., Willner, W. & Hájek, O. 2007. Semi-dry grasslands along a climatic gradient across Central Europe: Vegetation classification with validation. *Journal of Vegetation Science* **18**: 835-846.
- Jackson, D.A. 1993. Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**.
- Jandt, U. 1999. Kalkmagerrasen am Südharzrand und im Kyffhäuser. Gliederung im überregionalen Kontext, Verbreitung, Standortsverhältnisse und Flora. *Dissertationes Botanicae* **322**: 1-246.
- Janišová, M., Hajková, P., Hegedúšová, K., Hrivnák, R., Kliment, J., Micháľková, D., Ružičková, H., Řezníčková, M., Tichý, L., Škodová, I., Uhliarová, E., Ujházy, K. & Zaliberová, M. 2007. *Travinnobylinná vegetácia Slovenska – elektronický expertný systém na identifikáciu syntaxónov. Grassland vegetation of Slovak Republic – Electronic expert system for indentification of syntaxa*. Botanický ústav SAV, Bratislava.
- Janišová, M. & Škodová, I. 2007. Phytosociological database of Slovak grassland vegetation. *Annali di Botanica (Roma) nuova series* **7**: 19-26.
- Jansen, F. 2009. *vegdata: Vegetation Data Access and Evaluation*. 0.1.1. <http://geobot.botanik.uni-greifswald.de/portal>
- Jansen, F. & Dengler, J. Plant names in vegetation databases – a neglected source of bias. *Journal of Vegetation Science*.
- Jarolímek, I. & Šibík, J. (Szerk.). 2008. *Diagnostic, constant and dominant species of the higher vegetation units of Slovakia*. Veda, Bratislava.
- Jeanplong, J. 1960. Vázlatok a Rába határvidéki árterének rétjeiről (Outline of meadows in the floodplain of Rába river). *Botanikai Közlemények* **48**: 289-299.
- Jost, L. 2006. Entropy and diversity. *Oikos* **113**: 363-375.
- Jost, L. 2007. Partitioning diversity into independent alpha and beta components. *Ecology* **88**: 2427-2439.
- Joyce, C.B. & Wade, P.M. 1999. *European wet grasslands: biodiversity, management and restoration*. Wiley, Chichester.
- Juhász-Nagy, P. 1964. Some theoretical models of cenological fidelity I. *Acta Biologica Debrecina* **3**: 33-43.
- Kaufman, L. & Rousseeuw, P.J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Kennedy, K.A. & Addison, P.A. 1987. Some consideration for the use of visual estimates of plant cover in biomonitoring. *Journal of Ecology* **75**: 151-157.
- Kent, M. & Coker, P. 1992. *Vegetation description and analysis. A practical approach*. John Wiley & Sons, Chichester.
- Kienast, F., Wildi, O. & Brzeziecki, B. 1998. Potential impacts of climate change on species richness in mountainforests – An ecological risk assessment. *Biological Conservation* **83**: 291-305.
- Klika, J. 1958. K fytocenologii rašelinných a slatinných společenstev na Záhorské nížině (On phytosociology of mire and fen communities in Záhorská Lowland). *Biol. Pr.* **4**: 5-35.
- Klotz, S., Kühn, I. & Durka, W. 2002. BIOLFLOR—Eine Datenbank zu biologisch-ökologischen Merkmalen der Gefäßpflanzen in Deutschland. *Schriftenreihe für Vegetationskunde* **38**: 1-334.

- Knollová, I., Chytrý, M., Tichý, L. & Hájek, O. 2005. Stratified resampling of phytosociological databases: some strategies for obtaining more representative data sets for classification studies. *Journal of Vegetation Science* **16**: 479-486.
- Kočí, M., Chytrý, M. & Tichý, L. 2003. Formalized reproduction of an expert-based phytosociological classification: A case study of subalpine tall-forb vegetation. *Journal of Vegetation Science* **14**: 601-610.
- Kolbek, J., Blažková, D., Břízová, E., Ložek, V., Rybníčková, E., Rybníček, K. & Rydlo, J. 1999. *Vegetace Chráněné krajinné oblasti Křivoklátsko. 1. Vývoj krajiny a vegetace, vodní, pobřežní a luční společenstva (Vegetation of the Protected Landscape Area Křivoklátsko. 1. Development of landscape and vegetation, aquatic, coastal and meadow communities)*. Agentura ochrany přírody a krajiny ČR & Botanický ústav AV ČR, Praha.
- Koleff, P., Gaston, K.J. & Lennon, J.J. 2003. Measuring beta diversity for presence-absence data. *Journal of Animal Ecology* **72**: 367-382.
- Kopecký, K. & Hejný, S. 1978. Die Anwendung einer 'deduktiven Methode syntaxonomischer Klassifikation' bei der Bearbeitung der strassenbegleitenden Pflanzengesellschaften Nordostböhmens. *Vegetatio* **36**: 43-51.
- Kováčová, M. 1976. Asociácia Holcetum lanati (Issler 1936) em. Passarge 1964 v údolí rieky Krupinice (Holcetum lanati (Issler 1936) em. Passarge 1964 association in the Krupinica River valley). *Biológia* **31**: 795-800.
- Kovács, M. 1962. *Die Moorzweiden Ungarns*. Akadémiai Kiadó, Budapest.
- Kovács, M. 1955. A Gödöllő-Máriabesenyő környéki rétek botanikai felvételezése, ökológiai és gazdasági szempontok figyelembevételével (Botanical survey of meadows in Gödöllő and Máriabesenyő surroundings, from ecologic and economic viewpoints). *Agrártudományi Egyetem Agronómiai Kar Kiadványai* **1**: 1-24.
- Kovács, M. 1957. Magyarország láprétejeinek ökológiai viszonyai (Talaj- és mikroklíma-viszonyok) (Ecological properties of Hungarian rich fens and Molinia meadows /Soils and microclimate/). *MTA Biol. Csop. Közlem.* **1**: 387-456.
- Kovács, M. & Máthé, I. 1967. Die Vegetation des Inundationsgebietes der Ipoly. I. Zöologische Untersuchungen. *Acta Bot. Acad. Sci. Hung.* **13**: 133-168.
- Kovács-Láng, E., Kröel-Dulay, G., Kertész, M., Fekete, G., Bartha, S., Mika, J., Dobi-Wantuch, I., Rédei, T., Rajkai, K. & Hahn, I. 2000. Changes in the composition of sand grasslands along a climatic gradient in Hungary and implications for climate change. *Phytocoenologia* **30**: 385-407.
- Kovář, P. 1981. The grassland communities of the southeastern basin of the Labe River. 1. Syntaxonomy. *Folia Geobotanica & Phytotaxonomica* **16**: 1-44.
- Krieger, A. & Green, P. 1999. A cautionary note on using internal cross validation to select the number of clusters. *Psychometrika* **64**: 341-353.
- Kröel-Dulay, G., Ódor, P., Peters, D.P.C. & Hochstrasser, T. 2004. Distribution of plant species at a biome transition zone in New Mexico. *Journal of Vegetation Science* **15**: 531-538.
- Kučera, T. & Šumberová, K. 2001. Louky a pastviny (Meadows and pastures). In: *Katalog biotopů České republiky (Habitat catalogue of the Czech Republic)* (szerk. Chytrý, M., Kučera, T. & Kočí, M.). Agentura ochrany přírody a krajiny ČR, Praha, o. 109-125.
- Kun, A. & Ittész, P. 1995. A Seseli leucospermum W. et K. és a nyílt dolomitsziklagyep (Seseli leucospermo-Festucetum pallentis) előfordulása szarmata mészkövön. *Botanikai Közlemények* **82**: 27-34.
- Lájer, K. 2007. Statistical tests as inappropriate tools for data analysis performed on non-random samples of plant communities. *Folia Geobotanica* **42**: 115-122.
- Lájer, K., Botta-Dukát, Z., Csiky, J., Horváth, F., Szmorad, F., Bagi, I., Dobolyi, K., Hahn, I.,

- Kovács, J.A. & Rédei, T. 2008. Hungarian Phytosociological database (COENODATREF): sampling methodology, nomenclature and its actual stage. *Annali di Botanica (Roma) nuova series* **7**: 197-201.
- Lájer, K. 2002. Gondolatok a növényi társadalmak természetéről. In: *Magyar botanikai kutatások az ezredfordulón – tanulmányok Borhidi Attila 70. születésnapja tiszteletére* (szerk. Salamon-Albert, É.). PTE Növénytani Tanszék és MTA Pécsi Akadémiai Bizottság, Pécs, o. 463-483.
- Lambdon, P.W. 2008. Why is habitat breadth correlated strongly with range size? Trends amongst the alien and native floras of Mediterranean islands. *Journal of Biogeography* **35**: 1095-1105.
- Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* **76**: 5-13.
- Lange, T., Roth, V., Braun, M.L. & Buhmann, J.M. 2004. Stability-Based Validation of Clustering Solutions. *Neural Computation* **16**: 1299-1323.
- Legendre, P. & Anderson, J.M. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**: 1-24.
- Legendre, P. & Legendre, L. 1998. *Numerical Ecology*. 2. edn. Elsevier Science BV, Amsterdam.
- Legendre, P. & Gallagher, E. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271-280.
- Lenoir, J., Gégout, J., Marquet, P.A., de Ruffray, P. & Brisse, H. 2008. A significant upward shift in plant species optimum elevation during the 20th century. *Science* **320**: 1768-1771.
- Lepš, J. & Šmilauer, P. 2007. Subjectively sampled vegetation data: Don't throw out the baby with the bath water. *Folia Geobotanica* **42**: 169-178.
- Levine, E. & Domany, E. 2001. Resampling Method for Unsupervised Estimation of Cluster Validity. *Neural Computation* **13**: 2573-2593.
- Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**: 385-393.
- van der Maarel, E. 1979. Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* **38**: 97-114.
- Mack, R. 2003. Phylogenetic Constraint, Absent Life Forms, and Preadapted Alien Plants: A Prescription for Biological Invasions. *International Journal of Plant Sciences* **164**: S185-S196.
- Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. 2005. *Cluster Analysis Basics and Extensions*.
- Magurran, A.E. 2004. *Measuring biological diversity*. Blackwell, Oxford.
- Manthey, M. & Fridley, J.D. 2009. Beta diversity metrics and the estimation of niche width via species co-occurrence data: reply to Zeleny. *Journal of Ecology* **97**: 18-22.
- Marosi, S. & Somogyi, S. 1990. *Magyarország kistájainak katasztere*. MTA Földrajztudományi Kutató Intézet, Budapest.
- Maťašová, Z. 1987. Rastlinné spoločenstvá vlhkých lúk pravobrežnej depresie rieky Latorice (Bešanský polder) (Plant communities of wet meadow in the right-hand depression of the Latorica River (Bešanský polder)). *Biológia* **42**: 497-507.
- McClain, J.O. & Rao, V.R. 1975. CLUSTISZ: A program to test for the quality of clustering of a set of objects. *Journal of Marketing Research* **12**: 456-460.
- McCune, B. 1994. Improving community ordination with the Beals smoothing function. *Ecoscience* **1**: 82-86.
- McCune, B. & Mefford, M.J. 1999. *PC-ORD: Multivariate Analysis of Ecological Data* 4.17.

- User's Guide*. MjM Software, Gleneden Beach, Oregon.
- McIntyre, R.M. & Blashfield, R.K. 1980. A Nearest-Centroid Technique For Evaluating The Minimum-Variance Clustering Procedure. *Multivariate Behavioral Research* **15**: 225 - 238.
- Miklós, I. & Podani, J. 2004. Randomization of presence-absence matrices: comments and new algorithms. *Ecology* **85**: 86-92.
- Milligan, G.W. 1981. A Monte Carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika* **46**: 187-199.
- Milligan, G.W. & Cooper, M.C. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**: 159-179.
- Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* **71**: 145-156.
- Mirkin, B.M. 1989. Plant taxonomy and syntaxonomy: a comparative analysis. *Vegetatio* **82**: 35-40.
- Molnár, Z., Bartha, S., Seregélyes, T., Illyés, E., Botta-Dukát, Z., Tímár, G., Horváth, F., Révész, A., Kun, A., Bölöni, J., Bíró, M., Bodoncz, L., Deák, J.Á., Fogarasi, P., Horváth, A., Isépy, I., Karas, L., Kecskés, F., Molnár, C., Ortmann-né Ajkai, A. & Rév, S. 2007. A grid-based, satellite-image supported, multi-attributed vegetation mapping method (MÉTA). *Folia Geobotanica* **42**: 225-247.
- Molnár, Z., Bíró, M., Bölöni, J. & Horváth, F. 2008. Distribution of the (semi-)natural habitats in Hungary I. Marshes and grasslands. *Acta Botanica Hungarica* **50**: 59-106.
- Molnár, Z. & Rédei, T. 1995. A hazai edényes flóra cönoszisztematikai besorolásának Soó-i rendszere. In: *FLÓRA adatbázis 1.2* (szerk. Horváth, F., Dobolyi, Z.K., Morschhauser, T., Lőkös, L., Karas, L. & Szerdahelyi, T.). MTA-ÖBKI & MTM Növénytára, Vácrátót.
- Moravec, J., Balátová-Tulácková, E., Blažková, D., Hadač, E., Hejný, S., Husák, Š., Jeník, J., Kolbek, J., Krahulec, F., Kropáč, Z., Neuhausl, R., Rybníček, K., Řehořek, V. & Vicherek, J. 1995. Rostlinná společenstva České republiky a jejich ohrožení (Red list of plant communities of the Czech Republic and their endangerment). Ed. 2. *Severočes. Přír. suppl.* **1995**: 1-206.
- Mouillot, D., Dumay, O. & Tomasini, J.A. 2007. Limiting similarity, niche filtering and functional diversity in coastal lagoon fish communities. *Estuarine, Coastal and Shelf Science* **71**: 443-456.
- Mouillot, D., Mason, N. & Wilson, J. 2007. Is the abundance of species determined by their functional traits? A new method with a test using plant communities. *Oecologia* **152**: 729-737.
- Mucina, L. 1997a. Conspectus of classes of European vegetation. *Folia Geobotanica & Phytotaxonomica* **32**: 117-172.
- Mucina, L. 1997b. Classification of vegetation: Past, present and future. *Journal of Vegetation Science* **8**: 751-760.
- Mucina, L., Bredenkamp, G.J., Hoare, D.B. & McDonald, D.J. 2000. A national vegetation database for South Africa. *S. Afr. J. Sci.* **96**: 497-498.
- Mucina, L. & van der Maarel, E. 1989. Twenty years of numerical syntaxonomy. *Vegetatio* **81**: 1-15.
- Mucina, L., Rodwell, J.S., Schaminée, J.H.J. & Dierschke, H. 1993. European Vegetation Survey: current state of some national programmes. *Journal of Vegetation Science* **4**: 429-439.
- Mucina, L., Schaminée, J.H. & Rodwell, J.S. 2000. Common data standards for recording relevés in field survey for vegetation classification. *Journal of Vegetation Science* **11**: 769-772.

- Münzbergová, Z. & Herben, T. 2004. Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos* **105**: 408-414.
- Neu, C.W., Byers, C.R. & Peek, J.M. 1974. A technique for analysis of utilization-availability data. *Journal of Wildlife Management* **38**: 541-545.
- Neuhäusl, R. & Neuhäuslová, Z. 1989. Polopřirozená travinná a vysokobylinná vegetace Železných hor (Semi-natural grassland and tall-forb vegetation of the Železné Mts.). *Stud. ČSAV* **1989**: 1-200.
- Oberdorfer, E. (Szerk.). 1993. *Süddeutsche Pflanzengesellschaften. Teil 3. Wirtschaftswiesen und Unkrautgesellschaften*. 3. edn. Gustav Fischer Verlag, Jena.
- Oborny, B. 2000. Játék határokkal – társulási szabályok az ökológiai közösségekben. In: *Vegetáció és dinamizmus* (szerk. Virágh, K. & Kun, A.). MTA ÖBKI, Vácrátót, o. 79-96.
- Oborny, B. & Botta-Dukát, Z. 2007. Fajtöbbség, fajhiány. In: *Ökológia* (szerk. Pásztor, E. & Oborny, B.). Nemzeti Tankönyvkiadó, Budapest, o. 333-336.
- Økland, R. 2007. Wise use of statistical tools in ecological field studies. *Folia Geobotanica* **42**: 123-140.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.G., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. 2010. *vegan: Community Ecology Package*. 1.17-0. <http://CRAN.R-project.org/package=vegan>
- Ozinga, W.A., Bekker, R.M., Schaminée, J.H. & van Groenendael, J.M. 2004. Dispersal potential in plant communities depends on environmental conditions. *Journal of Ecology* **92**: 767-777.
- Ozinga, W.A., Schaminée, J.H.J., Bekker, R.M., Bonn, S., Poschlod, P., Tackenberg, O., Bakker, J. & Groenendael, J.M.V. 2005. Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *Oikos* **108**: 555-561.
- Ozinga, W., Hennekens, S., Schaminée, J., Bekker, R., Prinzing, A., Bonn, S., Poschlod, P., Tackenberg, O., Thompson, K., Bakker, J. & van Groenendael, J. 2005. Assessing the relative importance of dispersal in plant communities using an ecoinformatics approach. *Folia Geobotanica* **40**: 53-67.
- Pal, N.R. & Bezdek, J.C. 1995. On cluster validity for fuzzy c-means model. *IEEE Transactions on Fuzzy Systems* **3**: 370-379.
- Pal, N.R. & Biswas, J. 1997. Cluster validation using graph theoretic concepts. *Pattern Recognition* **30**: 847-857.
- Palmer, M.W. 1990. The Estimation of Species Richness by Extrapolation. *Ecology* **71**: 1195-1198.
- Pärtel, M., Zobel, M., Zobel, K. & van der Maarel, E. 1996. The species pool and its relation to species richness – evidence from Estonian plant communities. *Oikos* **75**: 111-116.
- Passarge, H. 1964. *Pflanzengesellschaften des nordostdeutschen Flachlandes I*. Gustav Fischer Verlag, Jena.
- Pásztor, E. & Meszéna, G. 2007. Versengés és együttélés. In: *Ökológia* (szerk. Pásztor, E. & Oborny, B.). Nemzeti Tankönyvkiadó, Budapest.
- Peres-Neto, P.R., Jackson, D.A. & Somers, K.M. 2005. How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis* **49**: 974-997.
- Picard, N. & Franc, A. 2003. Are ecological groups of species optimal for forest dynamics modelling? *Ecological Modelling* **163**: 175-186.
- Pignatti, S., Dominici, E. & Pietrosanti, S. 1997. European Vegetation Survey – from the methodological discussion to the first approximation. *Annali di Botanica (Roma)* **45**: 5-16.

- Pillar, V.D. 1999. How sharp are classifications? *Ecology* **80**: 2508–2516.
- Pinke, G., Pál, R. & Botta-Dukát, Z. 2010. Effects of environmental factors on weed species composition of cereal and stubble fields in western Hungary. *Central European Journal of Biology* **5**: 283–292.
- Podani, J. 1986. Comparison of partitions in vegetation studies. *Abstracta Botanica* **10**: 235–290.
- Podani, J. 1989. A method for generating consensus partitions and its application to community classification. *Coenoses* **4**: 1–10.
- Podani, J. 1994. *Multivariate Data Analysis in Ecology and Systematics*. Ecological Computations Series. SPB, The Netherlands.
- Podani, J. 1997. *Bevezetés a többváltozós biológiai adatfeltárás rejtelmeibe*. Scientia Kiadó, Budapest.
- Podani, J. & Csányi, B. 2010. Detecting indicator species: Some extensions of the IndVal measure. *Ecological Indicators* **10**: 1119–1124.
- Popma, J., Mucina, L., van Tongeren, O. & van der Maarel, E. 1983. On the determinants of optimal levels in phytosociological classification. *Vegetatio* **52**: 65–75.
- Popper, K. 1997. *A tudományos kutatás logikája*. Európa Könyvkiadó, Budapest.
- Pott, R. 1995. *Die Pflanzengesellschaften Deutschlands*. 2. edn. Verlag Eugen Ulmer, Stuttgart.
- R Development Core Team. 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2.9.2. <http://www.R-project.org>
- Rabinowitz, D. 1981. Seven forms of rarity. In: *The biological aspects of rare plant conservation* (szerk. Synge, H.). John Wiley, Chichester, UK, o. 205–217.
- Rand, W.M. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**: 846–850.
- Rannap, R., Lõhmus, A. & Briggs, L. 2009. Niche position, but not niche breadth, differs in two coexisting amphibians having contrasting trends in Europe. *Diversity and Distributions* **15**: 692–700.
- Rédei, T. & Horváth, F. 1995. A flóraelemek kategória-rendszere. In: *FLÓRA adatbázis 1.2* (szerk. Horváth, F., Dobolyi, Z.K., Morschhauser, T., Lőkös, L., Karas, L. & Szerdahelyi, T.). MTA-ÖBKI & MTM Növénytára, Vácrátót.
- Rédei, T., Botta-Dukát, Z., Csiky, J., Kun, A. & Tóth, T. 2003. On the possible role of local effects on the species richness of acidic and calcareous rock grasslands in northern Hungary. *Folia Geobotanica* **38**: 453–467.
- Řehořek, V. 1969. *Údolné lúky nížinného stupňa v povodí Slanej (Meadows of the valleys in the lowland belt in the Slaná watershed)* (PhD thesis).
- Ricotta, C. & Szeidl, L. 2009. Diversity partitioning of Rao's quadratic entropy. *Theoretical Population Biology* **76**: 299–302.
- Rodwell, J.S. 1995. The European Vegetation Survey questionnaire: an overview of phytosociological data, vegetation survey programmes and databases in Europe. *Annali di Botanica (Roma)* **53**: 87–98.
- Rodwell, J.S., Pignatti, S., Mucina, L. & Schaminée, J.H. 1995. European Vegetation Survey: update on progress. *Journal of Vegetation Science* **6**: 759–762.
- Rodwell, J.S., Schaminée, J.H.J., Mucina, L., Pignatti, S., Dring, J. & Moss, D. 2002. *The Diversity of European Vegetation. An overview of Phytosociological alliances and their relationships to EUNIS habitats*. EC-LNV, Wageningen, NL.
- Rohlf, F.J. 1974. Methods of comparing classifications. *Annual Review of Ecology and Systematics* **5**: 101–113.
- Roleček, J., Chytrý, M., Hájek, M., Lvončík, S. & Tichý, L. 2007. Sampling design in large-

- scale vegetation studies: Do not sacrifice ecological thinking to statistical purism! *Folia Geobotanica* **42**: 199-208.
- Rothmaler, W., Bässler, M., Jäger, E.J. & Werner, K. 1990. *Exkursionsflora von Deutschland*. 8. edn. Gustav Fischer Verlag, Jena.
- Rózsa, S., Nagybányai Nagy, O. & Oláh, A. (Szerk.). 2006. *A pszichológiai mérés alapjai. Elmélet, módszer és gyakorlati alkalmazás*. Bölcsész Konzorcium, Budapest. <http://mek.niif.hu/05500/05536/05536.pdf>
- Ruprecht, E. & Botta-Dukát, Z. 1999. Long-term vegetation textural changes of three fen communities near Cluj-Napoca (Romania). *Acta Botanica Hungarica* **42**: 263-281.
- Ruspini, E.H. 1970. Numerical methods for fuzzy clustering. *Information Science* **2**: 319-350.
- Ružičková, H. 1971. Rastlinné spoločenstvá lúk a slatín v povodí Čiernej vody (Východoslovenská nížina) (Plant communities of meadows and fens in the Čierna voda watershed (East Slovakian Lowland)). *Biol. Pr.* **17**: 1-133.
- Rybníček, K., Balátová-Tuláčková, E. & Neuhäusl, R. 1984. Přehled rostlinných společenstev rašelinišť a mokřadních luk Československa (Survey of plant communities of mires and wet meadows of Czechoslovakia). *Stud. ČSAV* **1984**: 1-123.
- Schaminée, J.H.J., Hennekens, S.M., Chytrý, M. & Rodwell, J.S. 2009. Vegetation-plot data and databases in Europe: an overview. *Preslia* **81**: 173-185.
- Schubert, R., Hilbig, W. & Klotz, S. 2001. *Bestimmungsbuch der Pflanzengesellschaften Deutschlands*. Spektrum Akademischer Verlag, Heidelberg.
- Šeffer, J., Janák, M. & Šefferová Stanová, V. 2008. *Management models for habitats in Natura 2000 Sites. 6440 Alluvial meadows of river valleys of the Cnidion dubii*. European Commission. http://ec.europa.eu/environment/nature/natura2000/management/habitats/pdf/6440_Alluvial_meadows.pdf
- Šeffer, J. & Stanová, V. (Szerk.). 1999. *Morava river floodplain meadows – importance, restoration and management*. DAPHNE – Centre for Applied Ecology, Bratislava.
- Šibíková, I., Šibík, J., Jarolímek, I. & Kliment, J. 2009. Šibíková I., Šibík J., Jarolímek I. & Kliment J. (2009): Current knowledge and phytosociological data on the high-altitude vegetation in the Western Carpathians – a review. – *Biologia* **64**: 215–224. *Biológia* **64**: 215-224.
- Simon, T. 1960. Die Vegetation der Moore in den Naturschutzgebieten des nördlichen Alföld. *Acta Bot. Acad. Sci. Hung.* **6**: 107-137.
- Simon, T. 1977. *Vegetationsuntersuchungen im Zempléner Gebirge*. Akadémiai Kiadó, Budapest.
- Siroki, Z. 1956. A Debreceni Mezőgazdasági Akadémia nyulasi legelőjének botanikai összetétele (Vegetation of pastures of Agricultural College of Debrecen near Nyulas). *Debreceni Mezőgazdasági Akadémia Évkönyve* **1956**: 183-205.
- Siroki, Z. 1958. Egy nyírségi reliktumterület monografikus cönológiai feldolgozása (A monographic phytosociological study of a relict area in the Nyírség). *Debreceni Mezőgazdasági Akadémia Évkönyve* **1958**: 109-141.
- Siroki, Z. 1962. Vegetációtanulmányok a debreceni löszháton. (Vegetation studies on loess ridges near Debrecen). *Debreceni Mezőgazdasági Akadémia Évkönyve* **1962**: 435-461.
- Sitášová, E. 1995. Vegetačné pomery Prírodnej rezervácie Kráľova studňa (Vegetation characteristics of Kráľova studňa Nature Reserve). *Natura Carpat.* **36**: 35-38.
- Sitášová, E. 1999. Botanická charakteristika močiaru pri Žabianoch (Botanical characteristics of the wetland near Žabiany). *Natura Carpat.* **40**: 57-72.
- Smart, S.M., Clarke, R.T., van de Poll, H.M., Robertson, E.J., Shield, E.R., Bunce, R.G.H. & Maskell, L.C. 2003. National-scale vegetation change across Britain; an analysis of sample-based surveillance data from the Countryside Surveys of 1990 and 1998.

- Journal of Environmental Management* **67**: 239-254.
- Sokal, R.R. & Rohlf, F.J. 1981. *Biometry. The Principles and Practice of Statistics in Biological Research*. 2. edn. Freeman, New York.
- Solomakha, V.A. 1996. Features of the creation of National Phytocoenoteca of the Ukraine. *Ukr. Phytosoc. Collect., Ser. A* **3**: 3-5.
- Soó, R. 1941. A magyar (pannóniai) flóratartomány növényközvetkezőinek áttekintése. *Magyar Biológiai Kutatóintézet Munkái* **13**: 498-511.
- Soó, R. 1964. *A magyar flóra és vegetáció rendszertani-növényföldrajzi kézikönyve I-VI*. Akadémiai Kiadó, Budapest.
- Sousa, F. & Tendeiro, J. 2005. A Validation Methodology in Hierarchical Clustering. In: *Proceedings of the Conference: Applied Stochastic Models and Data Analysis, Brest, France, May, 17–20, 2005* (szerk. Janssen, J. & Lenca, P.). ENST Bretagne, Brest.
- Špániková, A. 1971. Fytocenologická štúdia lúk juhozápadnej časti Košickej kotliny (Phytosociological study of meadows in southwestern part of Košická valley). *Biol. Pr.* **17**: 1-103.
- Špániková, A. 1982. Pflanzengesellschaften mit der Art *Scirpus sylvaticus* in der Slowakei. *Biológia* **37**: 503-512.
- Sprenst, P. & Smeeton, N.C. 2001. *Applied nonparametric statistical methods*. Chapman & Hall/CRC, Boca Raton.
- Stančić, Z. 2008. Phytosociological database of non-forest vegetation of Croatia. In: *17th International Workshop European Vegetation Survey. Using phytosociological data to address ecological questions. 1–5 May 2008, Masaryk University, Brno, Czech Republic. Abstracts and excursion guides* (szerk. Chytrý, M.). Masaryk University, Brno, o. 128.
- Stanová, V. & Valachovič, M. (Szerk.). 2002. *Katalóg biotopov Slovenska (Habitat catalogue of Slovakia)*. Daphne – Inštitút aplikovanej ekológie, Bratislava.
- Steták, D. 2005. A Duna-Dráva Nemzeti Park Gemenci Tájegysége mocsári és mocsárréti növénytársulásairól. *Botanikai Közlemények* **92**: 119-157.
- Strauss, R.E. 1982. Statistical significance of species clusters in association analysis. *Ecology* **63**: 634-639.
- Šumberová, K. 1997. Současný stav vegetace svazu *Veronico longifoliae*-*Lysimachion vulgaris* na jižní Moravě (The present state of vegetation of the *Veronico longifoliae*-*Lysimachion vulgaris* alliance in South Moravia). *Zpr. Čes. Bot. Společ., Mater.* **15**: 177-189.
- Sykes, J.M., Horrill, A.D. & Mountford, M.D. 1983. Use of visual cover assessments as quantitative estimators of some British woodland taxa. *Journal of Ecology* **71**: 437-450.
- Sýkora, K.V., Kalwij, J.M. & Keizer, P. 2002. Phytosociological and floristic evaluation of a 15-year ecological management of roadside verges in the Netherlands. *Preslia* **74**: 421-436.
- Tan, P., Steinbach, M. & Kumar, V. 2005. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc.
- Tichý, L. 2002. JUICE, software for vegetation classification. *Journal of Vegetation Science* **13**: 451-453.
- Tichý, L., Chytrý, M., Hájek, M., Talbot, S.S. & Botta-Dukát, Z. 2010. OptimClass: Using species-to-cluster fidelity to determine the optimal partition in classification of ecological communities. *Journal of Vegetation Science* **21**: 287-299.
- Tichý, L. 2005. New similarity indices for the assignment of relevés to the vegetation units of an existing phytosociological classification. *Plant Ecology* **179**: 67-72.
- Tichý, L. & Chytrý, M. 2006. Statistical determination of diagnostic species for site groups of

- unequal size. *Journal of Vegetation Science* **17**: 809-818.
- Tímár, L. 1953. A Tiszamente Szolnok-Szeged közti szakaszának növényföldrajza. *Földrajzi Értesítő* **2**: 87-113.
- Tongeren, O., Gremmen, N. & Hennekens, S. 2008. Assignment of relevés to pre-defined classes by supervised clustering of plant communities using a new composite index. *Journal of Vegetation Science* **19**: 525-536.
- Tóthmérész, B. 1993. Noise elimination in gradient analysis. *Abstracta Botanica* **17**: 155-158.
- Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. *Journal of Vegetation Science* **6**: 283-290.
- Török, K., Horánszky, A. & Kósa, G. 1994. Long-term changes of species composition in an andesite grassland community of the Visegrád Mts., Hungary. *Abstracta Botanica* **18**.
- Török, K., Podani, J. & Borhidi, A. 1989. Numerical revision of Fagion illyricum alliance. *Vegetatio* **81**: 169-180.
- Török, K. & Zólyomi, B. 1998. A Kárpát-medence öt sziklagyeptársulásának szüntaxonómiai revíziója (Syntaxonomical revision on five rocky grassland communities of the Carpathian Basin). In: *Sziklagyepek szünbotanikai kutatása* (szerk. Csontos, P.). Scientia Kiadó, Budapest, o. 109-132.
- Tsiripidis, I., Bergmeier, E., Fotiadis, G. & Dimopoulos, P. 2009. A new algorithm for the determination of differential taxa. *Journal of Vegetation Science* **20**: 233-240.
- Tuhkanen, S. 1980. Climatic parameters and indices in plant geography. *Acta Phytogeogr. Suec.* **67**: 1-105.
- Tuomisto, H. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* **33**: 2-22.
- Valachovič, M. 1999. Centrálna databáza fytocenologických zápisov (CDF) na Slovensku (Central database of phytosociological relevés (CDF) in Slovakia). In: *Zborník 7. zjazdu SBS, Hrabušice, Podlesok*. o. 218-220.
- Veech, J.A., Summerville, K.S., Crist, T.O. & Gering, J.C. 2002. The additive partitioning of species diversity: recent revival of an old idea. *Oikos* **99**: 3-9.
- Vendramin, L., Campello, R.J.G.B. & Hruschka, E.R. 2010. Relative clustering validity criteria: A comparative overview. *Statistical Analysis and Data Mining* **3**: 209-235.
- Vesecký, A., Petrovič, Š., Briedoň, V. & Karský, V. 1958. *Atlas podnebí Československé republiky (Climate atlas of the Czechoslovak Republic)*. Ústřední správa geodesie a kartografie, Praha.
- Vicherek, J. 1960. *Typologicko-ekologická studie lučních společenstev v dolním Podyjí (Typological-ecological study of meadow communities in the lower Dyje valley)* (PhD thesis).
- Vicherek, J. 1962. Typy fytocenů aluviální nivy dolního Podyjí se zvláštním zaměřením na společenstva luční (Types of phytocoenoses in the lower Dyje floodplain with special emphasis on meadow communities). *Folia Fac. Sci. Nat. Univ. Purk. Brun., Biologia* **3/5**: 1-113.
- Vicherek, J., Antonín, V., Danihelka, J., Grulich, V., Gruna, B., Hradílek, Z., Řehořek, V., Šumberová, K., Vampola, P. & Vágner, A. 2000. *Flóra a vegetace na soutoku Moravy a Dyje (Flora and vegetation in the confluence area of the Morava and Dyje Rivers)*. Masaryk University, Brno.
- Virágh, K. 1987. The effect of herbicides on vegetation dynamics: Comparison of classifications. *Abstracta Botanica* **11**: 53-70.
- Warrens, M. 2008. On the Equivalence of Cohen's Kappa and the Hubert-Arabie Adjusted Rand Index. *Journal of Classification* **25**: 177-183.
- Westhoff, V. & van der Maarel, E. 1973. The Braun-Blanquet approach. In: *Ordination and*

- classification of plant communities* (szerk. Whittaker, R.H.). W. Junk, The Hague, o. 617-737.
- Whittaker, R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs* **30**: 279-338.
- Williamson, M.H. 1978. The ordination of incidence data. *Journal of Ecology* **66**: 911-920.
- Wilson, J.B. 2007. Priorities in statistics, the sensitive feet of elephants, and don't transform data. *Folia Geobotanica* **42**: 161-167.
- Wohlgemuth, T. 1992. Die vegetationskundliche Datenbank. *Schweiz. Z. Forstw.* **143**: 22-36.
- Xu, R. & Wunsch, D.C. 2009. *Clustering*. John Wiley & Sons, Hoboken, New Jersey.
- Zahradníková-Rošetzká, K. 1965. Geobotanická charakteristika slatinných lúk a pasienkov (Molinion Koch 1926) na Žitnom ostrove (Geobotanical characteristics of fen meadows and pastures (Molinion Koch 1926) on Žitný Island). *Biol. Pr.* **11**: 5-45.
- Zar, J.H. 1999. *Biostatistical analysis*. 4. edn. Prentice and Hall, Upper Saddle River, NY, US.
- Zeleny, D. 2009. Co-occurrence based assessment of species habitat specialization is affected by the size of species pool: reply to Fridley et al. (2007). *Journal of Ecology* **97**: 10-17.
- Zlinská, J. 1989. Filipendulo ulmariae-Menthetum longifoliae eine neue Assoziation für die West Karpaten. *Biológia* **44**: 837-847.
- Zlinská, J. & Kubalová, S. 2001. Flora a vegetácia lokality Gamota na Podunajskej nížine a ochrana jej biodiverzity (Flora and vegetation of Gamota locality in Podunajská Lowland and its biodiversity conservation). *Acta Environm. Univ. Comen.* **11**: 1-120.
- Zobel, M. 1997. The relative of species pools in determining plant species richness: an alternative explanation of species coexistence? *Trends in Ecology & Evolution* **12**: 266-269.
- Zuidhoff, A., Schaminée, J.H. & Rodwell, J.S. 1995. The Cynosurion cristati Tx. 1947 of central, southern and western Europe: a tentative overview based on an analysis of individual relevés. *Annali di Botanica (Roma)* **53**: 25-46.

14. Az értekezés témakörében megjelent publikációk

Impakt faktoros közlemények:

- Illyés E, Bauer N, Botta-Dukát Z 2009. Classification of semi-dry grassland vegetation in Hungary. *Preslia* **81**: 239-260.
- Tichý L, Chytrý M, Hájek M, Talbot S, Botta-Dukát Z 2010. OptimClass: Using species-to-cluster fidelity to determine the optimal partition in classification of ecological communities. *Journal of Vegetation Science* **21**: 287-299.
- Illyés E, Chytrý M, Botta-Dukát Z, Jandt U, Skodova I, Janisova M, Willner W, Hajek O 2007. Semi-dry grasslands along a climatic gradient across Central Europe: Vegetation classification with validation. *Journal of Vegetation Science* **18**: 835-846. (2007)
- Chytrý M, Tichý L, Holt J, Botta-Dukát Z 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* **13**: 79-90.
- Botta-Dukát Z, Chytrý M, Hajkova P, Havlova M 2005. Vegetation of lowland wet meadows along a climatic continentality gradient in Central Europe. *Preslia* **77**: 89-111. (2005)
- Botta-Dukát Z, Kovács-Láng E, Rédei T, Kertész M, Garadnai J 2007. Statistical and biological consequences of preferential sampling in phytosociology: Theoretical considerations and a case study. *Folia Geobotanica* **42**: 141-152.

Nem impakt faktoros cikkek:

- Botta-Dukát, Z & Borhidi, A 1999. New objective method for calculating fidelity. Example: The Illyrian beechwoods. *Annali di Botanica (Roma)* **57**: 73-90.
- Botta-Dukát Z 2004. A magyarországi mocsárrétek cönológiai irodalmának áttekítése és szüntaxonómiai revíziója. *Kanitzia* **12**: 43-73.
- Botta-Dukát Z 2008. Validation of hierarchical classifications by splitting dataset. *Acta Botanica Hungarica* **50**: 73-80.
- Lájer, K, Botta-Dukát, Z, Csiky, J, Horváth, F, Szmorad, F, Bagi, I, Dobolyi, K, Hahn, I, Kovács, J A & Rédei, T 2008. Hungarian Phytosociological database (COENODATREF): sampling methodology, nomenclature and its actual stage. *Annali di Botanica (Roma) nuova series* **7**: 197-201.

15. Mellékletek

1. melléklet: A CoenoDat adatbázis számára gyűjtött új cönológiai felvételek fejlécében szereplő adatok

Az aláhúzott adatok megadása az új felvételeknél kötelező, a többi ajánlott. Archiv felvételeknél csak a *-gal jelölt mezők megadása kötelező.

A felvétel alapadatai

A felvétel egyedi azonosítója

*Társulásnév

Az élőhely mmÁ-NÉR kódja

*A felvételező neve

*A felvétel időpontja

*A mintavételi egység mérete (m²)

*A mintavételi egység alakja

*Az alkalmazott borítás becslési skála

Az állomány mérete (m² vagy ha)

Megye

*Település

Földrajzi hely

Földrajzi koordináták

UTM koordináta

Közép-Európai Flóratérképezési program hálójának koordinátája

Borítás adatok (akkor is százalékban kell megadni, ha a becslés más skálán történik)

Összborítás

Lomkoronaszint (E3) borítása

Kiemelkedő fák (E3δ) borítása

Felső lombkorona-szint (E3γ) magassága¹⁶

Középső lombkorona-szint (E3β) magassága

Alsó lombkorona-szint (E3α) magassága

Cserjeszint (E2) borítása

Felső cserjeszint (E2β) borítása

Alsó cserjeszint (E2α) borítása

Gyepszint (E1) borítása

Felső gyepszint (E1γ) borítása

Középső gyepszint (E1β) borítása

Alsó gyepszint (E1α) borítása

Hínárnövényzet emergens szintjének (E1e) borítása

¹⁶ Ha csak egy lomkoronaszintet lehet megkülönböztetni, akkor az a felső lombkoronaszint, ha kettőt akkor azok a felső- és középső lombkoronaszintnek tekintendők.

Hínárnövényzet úszó szintjének (E1n) borítása

Hínárnövényzet alámerült szintjének (E1s) borítása

Kriptogámszint (E0) borítása (csak a talajon vagy kövön)

Mohaszint (E0m) borítása (csak a talajon vagy kövön)

Zúzmószint (E0l) borítása (csak a talajon vagy kövön)

Algaszint (E0al) borítása (csak a talajon vagy vízben megfigyelhető makroszkópikus algák)

Avar borítása

Kövek borítása

Nyílt vízfelszín borítása

Csupasz talaj borítása

Magasság adatok

Lombkorona-szint (E3) magassága

Kiemelkedő fák (E3δ) magassága

Felső lombkorona-szint (E3γ) magassága

Középső lombkorona-szint (E3γ) magassága

Alsó lombkorona-szint (E3γ) magassága

Cserjeszint (E2) magassága

Felső cserjeszint (E2β) magassága

Alsó cserjeszint (E2α) magassága

E1 = gyepszint magassága

E1γ felső gyepszint magassága

E1β középső gyepszint magassága

E1α alsó gyepszint magassága

Hínárnövényzet emergens szintjének (E1e) magassága

Hínárnövényzet úszó szintjének (E1n) magassága

Hínárnövényzet alámerült szintjének (E1s) magassága

Kriptogámszint (E0) magassága

Mohaszint (E0m) magassága

Zúzmószint (E0l) magassága

Algaszint (E0al) magassága

Abiotikus háttér adatok:

Tengerszint feletti magasság (50 m-es pontossággal)

Kitettség (É, ÉK, K, DK, D, DNY, NY, ÉNY)

Lejtőszög (től-ig tartomány is megadható. Felbontás: sík, 1-5° között 1°, 5° felett 5°)

Távolság a vízparttól (m; csak vízi társulásoknál)

Vízmélység (dm)

Élőhely jellemzők (természetesség, veszélyeztető tényezők stb.)

Geomorfológia

Alapkőzet

Fizikai talajféleség (homok, homokos vályog, vályog, agyagos vályog, agyag)

dc_16_10

Genetikai talaj típus

Mikroklíma megfigyelések

Vízdinamika

Kezelés

Vegetáció komplex, szomszédos vegetációtípusok

Vegetáció dinamika (pl. parlag eredetű, gyomosodás, inváziós fajok terjedése, foltosság stb.)

Megjegyzések (határozási bizonytalanságok, mintavétel, stb.)

2. Melléklet: A khi-négyzet statisztika és az R^2 közti kapcsolat

Az alábbiakban annak a bizonyítása látható, hogy a faj előfordulása/hiánya és K darab felvételtcsoport alapján felírható $2 \times K$ méretű kontingenciatáblára kiszámolt Φ^2 , és ugyanezen adatokra felírt lineáris modelben, ahol a faj előfordulása a függő változó és a felvételtcsoport a kategóriális független változó, az R^2 értéke azonos.

Vezessük be a következő jelöléseket:

n_{1j} = a j . csoportba tartozó felvételek száma, amelyben a faj előfordul

n_{0j} = a j . csoportba tartozó felvételek száma, amelyben a faj nem fordul elő

$$n_{.j} = n_{1j} + n_{0j}$$

$$n_{1.} = \sum_j n_{1j}$$

$$n_{0.} = \sum_j n_{0j}$$

$$n_{..} = n_{1.} + n_{0.} = \sum_j n_{.j}$$

Definíció szerint:

$$\Phi = \sqrt{\frac{\chi^2}{n_{..}}}$$

amiből következik, hogy:

$$\Phi^2 = \frac{\chi^2}{n_{..}}$$

ezért a

$$\frac{\chi^2}{n_{..}} = R^2$$

egyenlőséget fogom bebizonyítani.

A determinációs koefficiens (R^2) kiszámításához ki kell számolnunk a teljes (SQ_T) és a csoporton belüli eltérésnégyzet-összeget (SQ_E). Mivel csak nulla és egy értéket vehet fel a függő változó, a két eltérésnégyzetösszeg az alábbi speciális formában írható fel:

$$SQ_T = n_{1.} \left(1 - \frac{n_{1.}}{n_{..}}\right)^2 + n_{0.} \left(\frac{n_{1.}}{n_{..}}\right)^2 = \frac{n_{1.} n_{0.}^2}{n_{..}^2} + \frac{n_{0.} n_{1.}^2}{n_{..}^2} = \frac{n_{0.} n_{1.}}{n_{..}^2} (n_{0.} + n_{1.}) = \frac{n_{0.} n_{1.}}{n_{..}}$$

$$SQ_E = \sum_j n_{1j} \left(1 - \frac{n_{1j}}{n_{.j}}\right)^2 + n_{0j} \left(\frac{n_{1j}}{n_{.j}}\right)^2 = \sum_j \frac{n_{0j} n_{1j}}{n_{.j}}$$

Ez alapján a determinációs koefficiens képlete ebben az esetben:

$$R^2 = 1 - \frac{SQ_E}{SQ_T} = 1 - \frac{n_{..}}{n_{0.}n_{1.}} \sum_j \frac{n_{0j}n_{1j}}{n_{.j}}$$

A kontingencia táblára a χ^2 statisztika a következő módon számolható:

$$\chi^2 = \sum_j \frac{(n_{1j} - e_{1j})^2}{e_{1j}} + \sum_j \frac{(n_{0j} - e_{0j})^2}{e_{0j}}$$

ahol:

$$e_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$$

Kihasználhatjuk, hogy:

$$n_{0j} = n_{.j} - n_{1j}$$

$$e_{0j} = n_{.j} - e_{1j}$$

$$(n_{0j} - e_{0j})^2 = (n_{1j} - e_{1j})^2$$

Behelyettesítve a fenti képletbe:

$$\chi^2 = \sum_j \left(\frac{(n_{1j} - e_{1j})^2}{e_{1j}} + \frac{(n_{1j} - e_{1j})^2}{n_{.j} - e_{1j}} \right) = \sum_j \frac{(n_{.j} - e_{1j})(n_{1j} - e_{1j})^2 + e_{1j}(n_{1j} - e_{1j})^2}{e_{1j}(n_{.j} - e_{1j})}$$

$$\chi^2 = \sum_j \frac{n_{.j}(n_{1j} - e_{1j})^2}{e_{1j}e_{0j}}$$

Helyettesítsük be a nevezőben szereplő várt értékeket!

$$\chi^2 = \sum_j n_{.j}(n_{1j} - e_{1j})^2 \frac{n_{..}}{n_{1.}n_{.j}} \frac{n_{..}}{n_{0.}n_{.j}} = \frac{n_{..}^2}{n_{1.}n_{0.}} \sum_j \frac{(n_{1j} - e_{1j})^2}{n_{.j}}$$

Ezután helyettesítsük be a számlálóban szereplő várt értéket is és rendezzük át az egyenletet!

$$\chi^2 = \frac{n_{..}^2}{n_{1.}n_{0.}} \sum_j \frac{1}{n_{.j}} \left(n_{1j} - \frac{n_{1.}n_{.j}}{n_{..}} \right)^2 = \frac{n_{..}^2}{n_{1.}n_{0.}} \sum_j \frac{(n_{..}n_{1j} - n_{1.}n_{.j})^2}{n_{..}^2 n_{.j}}$$

$$\chi^2 = \frac{1}{n_{1.}n_{0.}} \sum_j \frac{n_{..}^2 n_{1j}^2 + n_{1.}^2 n_{.j}^2 - 2n_{..}n_{1j}n_{1.}n_{.j}}{n_{.j}}$$

$$\chi^2 = \frac{1}{n_{1.}n_{0.}} \left[n_{..}^2 \left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) + n_{1.}^2 \left(\sum_j n_{.j} \right) - 2n_{..}n_{1.} \left(\sum_j n_{1j} \right) \right]$$

$$\chi^2 = \frac{1}{n_{1.}n_{0.}} \left[n_{..}^2 \left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) + n_{1.}^2 n_{..} - 2n_{..}n_{1.}^2 \right] = \frac{1}{n_{1.}n_{0.}} \left[n_{..}^2 \left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) - n_{..}n_{1.}^2 \right]$$

$$\chi^2 = n_{..} \left(\frac{n_{..}}{n_{1.}n_{0.}} \left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) - \frac{n_{1.}^2}{n_{0.}n_{1.}} \right)$$

$$\frac{\chi^2}{n_{..}} = \frac{n_{..}}{n_{1.}n_{0.}} \left[\left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) - \frac{n_{1.}^2}{n_{..}} \right] = \frac{n_{..}}{n_{1.}n_{0.}} \left[\left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) - \frac{n_{1.}(n_{..} - n_{0.})}{n_{..}} \right]$$

$$\frac{\chi^2}{n_{..}} = \frac{n_{..}}{n_{1.}n_{0.}} \left[\left(\sum_j \frac{n_{1j}^2}{n_{.j}} \right) - n_{1.} + \frac{n_{1.}n_{0.}}{n_{..}} \right] = 1 - \frac{n_{..}}{n_{1.}n_{0.}} \sum_j n_{1j} - \frac{n_{1j}^2}{n_{.j}}$$

$$\frac{\chi^2}{n_{..}} = 1 - \frac{n_{..}}{n_{1.}n_{0.}} \sum_j \frac{n_{1j}n_{.j} - n_{1j}^2}{n_{.j}} = 1 - \frac{n_{..}}{n_{1.}n_{0.}} \sum_j \frac{n_{1j}(n_{.j} - n_{1j})}{n_{.j}}$$

$$\frac{\chi^2}{n_{..}} = 1 - \frac{n_{..}}{n_{1.}n_{0.}} \sum_j \frac{n_{1j}n_{0j}}{n_{.j}} = R^2$$

3. melléklet: A módosított Rand-index

Az eredeti Rand-index (Rand 1971) a helyesen klasszifikált felvételpárok számát hasonlítja az összes felvételpárhoz:

$$R = \frac{a + d}{N}$$

ahol:

a = azoknak a felvételpároknak a száma, amelyek mindkét partícióban azonos csoportba tartoznak

d = azoknak a felvételpároknak a száma, amelyek mindkét partícióban különböző csoportba tartoznak

N = a felvétel párok száma

Az eredeti Rand index hátránya, hogy két random partíció hasonlóságának várhatóértéke a csoportszám függvényében változik. Ezt a problémát küszöböli ki a Hubert and Arabie (1985) által javasolt módosított Rand-index. A korrekció az alábbi általános képleten alapul (Warrens 2008):

$$R_{adj} = \frac{R - R_{exp}}{R_{max} - R_{exp}}$$

ahol:

R_{adj} = a korrigált index

R = az eredeti (korrigálatlan) index

R_{exp} = az index várhatóértéke két random partíció összehasonlításakor

R_{max} = az index elméleti maxima, a Rand index esetén 1

Ha a csoportméretek a két partícióban rögzítettek (Hubert & Arabie 1985, Warrens 2008):

$$R_{exp} = 1 + \frac{2(a+b)(a+c)}{N^2} - \frac{(a+b) + (a+c)}{N}$$

ahol:

a = azoknak a felvételpároknak a száma, amelyek mindkét csoportosításban azonos csoportba tartoznak

b = azoknak a felvételpároknak a száma, amelyek azonos csoportban vannak a várt csoportosításban, de különböző csoportba kerültek a kiértékelt klasszifikációban

c = azoknak a felvételpároknak a száma, amelyek különböző csoportban vannak a várt csoportosításban, de azonos csoportba kerültek a kiértékelt klasszifikációban

d = azoknak a felvételpároknak a száma, amelyek mindkét csoportosításban különböző csoportba tartoznak

N = a felvétel párok száma, $N=a+b+c+d$

Behelyettesítve és átrendezve a képletet:

$$R_{HA} = \frac{2(ad-bc)}{(a+c)(c+d) + (a+b)(b+d)}$$