A kombinált klaszter- és diszkriminanciaanalízis (CCDA) adatelemző módszer alkalmazása földtudományi feladatok megoldására

Application of the Combined cluster and discriminant analysis (CCDA) data analysis method in solving earth-science tasks

KOVÁCS József

ELTE TTK Általános és Alkalmazott Földtani Tanszék,
Budapest, Pázmány Péter sétány 1/C

Abstract

The grouping of variables/sampling sites/events etc. is a frequent task in modern research. When applied, the question needed to be answered is: How to determine groups with not only similar but homogeneous elements? Combined cluster and discriminant analysis (CCDA) is a new technique that combines two traditional methods to determine the optimal number of homogeneous groups in an objective way. A software applying CCDA was also developed, which can be used under any operating system supporting R (http://cran.r-project.org/). To demonstrate the applicability of the method, several research results are presented from numerous fields of Earth Sciences. The present paper mainly focuses on two examples: 1. the determination of optimal groups of the karst water spring in Budapest, which clustered springs and wells with the same hydrogeological background; and 2. the classification of Lake Balaton's water quality sampling sites into homogeneous groups, which can significantly help the subsequent recalibration of the lake's monitoring network in the future.

Kulcsszavak: optimális, homogén, csoportosítás

1. Bevezetés

A csoportosítás (klasszifikáció) mint a modern kutatásokban általánosan használt módszer alkalmazása esetén gyakran merül fel a kérdés: hogyan lehetne az optimális csoportszámot, valamint nemcsak a hasonló, hanem a legnagyobb homogén csoportokat létrehozni? Ha utóbbi sikerül – az egyéb igen széles körű felhasználási lehetőségek mellett – az eredmény felhasználható a mintavételi pontok számának csökkentésére is, minimális információveszteséggel (gyakorlatban ez információveszteség nélküli). Jelen dolgozat egy olyan új módszert mutat be, ami lehetővé teszi az optimális csoportszám és a homogén csoportok meghatározását. A dolgozat a módszer alkalmazása nyomán kapott eredményeket is prezentál: Budapest termálvizei, egy folyó, egy vizes élőhely és egy sekély tó, a Balaton, monitoring rendszereinek adatain történt számítások nyomán.

2. A módszer

A kombinált klaszter- és diszkriminanciaanalízis (CCDA; [1]) két hagyományos eljárást ötvöző új technika, ami a csoportosítás (klasszifikáció) során felmerülő optimális csoportszám és ahhoz tartozó csoportosítás, továbbá a csoportok között nemcsak hasonló, hanem a legnagyobb homogén csoportok meghatározását célozza. A CCDA olyan esetekben használható, amikor több származási helyről azonos paraméterkörre (azaz több dimenzióra) érhetők el adatok, úgy, hogy minden származási helyről több megfigyelés álljon rendelkezésre. A föld- és környezettudományokban legtöbbször maguk a mintavételi pontok az említett származási helyek, de ilyen „származási helyek” lehetnek időintervallumok, például hónapok, évszakok, vagy akár évtizedek is, amelyekhez a megfigyelések tartoznak.

A CCDA az iteratív módon a vizsgálandó csoportosítások mindegyikénél a lineáris diszkriminanciaanalízis (LDA) által helyesen klasszifikált esetek arányát viszonyítja véletlen beosztásoknál helyesen klasszifikált esetek arányszámához. A csoportosítások létrehozhatók például (hierarchikus)

klaszteranalízis segítségével a mért paraméterek standardizált átlagaira az egyes származási helyeken, de vizsgálhatók szakértők által létrehozott csoportosítások is. Minden egyes vizsgált csoportosításnál a módszer először meghatározza az LDA által helyesen klasszifikált esetek arányát, majd ebből kivonja N (pl. N=100 vagy N=500) véletlen beosztás helyes klasszifikációs arányszámainak 95%-os kvantilisét, ezáltal egy különbségértéket rendel az éppen vizsgált csoportosításhoz. Ez a különbségérték azt adja meg, hogy mennyire jó a vizsgált csoportosítás a véletlenhez képest, illetve ezáltal az egyes csoportosításokhoz rendelt különbségértékek egymással is összevethetők. Az éppen vizsgált csoportosítási lehetőségek közül az tekinthető objektív módon optimálisnak, amelyiknél ez a különbségérték a legnagyobb [2]. A pusztán átlagokból készült dendrogramra alapozott döntéssel szemben, legyen az szubjektív [3] vagy valamilyen indexre [4] alapozott, a CCDA az összes megfigyelés felhasználásával képes meghatározni az optimális csoportszámot.

Az optimális csoportosításban szereplő különböző csoportok tagjai általában hasonlók, de nem szükségszerűen homogének. Mindez akkor fordul elő, ha a különbségérték pozitív. Ekkor a csoportosítás jobb, mint a vizsgált véletlenszerű beosztások 95%-a, azaz szignifikánsan jobb, mint a véletlen, így a csoportosítás tagjai nem tekinthetők homogénnek. Ilyen esetekben a csoportokon belüli legkisebb különbségek megtalálása érdekében a CCDA az első lépésben talált optimális csoportosításnak egynél több tagból álló (al)csoportjait vizsgálja tovább iteratív módon mindaddig, amíg a legnagyobb különbségérték negatív nem lesz. Ekkor egyik csoportosítás sem jobb szignifikánsan, mint egy véletlenszerű csoportosítás, emiatt a vizsgált (al)csoportban levő származási helyeket homogénnek tekintjük. Abban az esetben, ha egy (al)csoport csupán egy tagból/származási helyből áll, azt nem lehet tovább bontani, így ilyenkor az a tag/származási hely önmagában alkot egy homogén egységet.

3. Eredmények a kombinált klaszter- és diszkriminancia analízissel (CCDA)

Az optimális csoportosításra jó példával szolgának a budapesti termálvizek, melyekből 1960–2009 között 27 kútból/forrásból származó kémiai (Na^++K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-) és hőmérsékleti adatait vizsgáltuk a CCDA-módszerrel. A számítások eredményeként a mért értékek standardizált átlagaira kapott dendrogramhoz tartozó d1,…,d27 különbségértékek között a hetedik csoportosításnak volt a legnagyobb különbségértéke (d7* = 70,2%). Az így kapott hét csoport (SG1,…,SG7) tekinthető az optimális csoportosításnak. A számítási eredmények nyomán megállapítható, hogy mindegyik csoport a gravitáció által vezérelt áramlási rendszerek különböző részeit reprezentálja. A csoportok geokémiai, illetve hőmérsékleti adatai az áramlási rendszerekben lévő különbözőségeket tükrözik. Külön csoportot alkotnak az északi megcsapolódási terület, Csillaghegy, Pünkösd– és Római fürdő kútjai (SG2), melyeknek utánpótlódási területei viszonylag közel, a Pilisben találhatók [5], lokális és intermedier áramlási rendszerekhez kapcsolódnak. Ezek jellemzői geokémiai és hőmérsékleti adataikban tükröződnek, ennél a csoportnál láthatjuk a legalacsonyabb értékeket. A rózsadombi megcsapolódási terület langyos (Lukács fürdő forrásai, SG1) és termálvizei (Lukács fürdő 4-es kút és Antal–forrás, SG6) külön-külön csoportokat alkotnak. A langyos források utánpótlódási területe a Budai–hegység [5], vizüket lokális és intermedier áramlási rendszerekből nyerik. Valamennyi termálvíz hozzákeveredés is megfigyelhető esetükben, ami az északi megcsapolódási terület kútjaihoz képest magasabb Na^++K^+, Cl- és SO_4^- tartalomban nyilvánul meg. A termálvizek, így az SG6 alcsoport tagjai is, regionális áramlási rendszerekből származnak, melyet magasabb hőmérsékletük és oldott anyag-tartalmuk jelez. A déli, Gellért-hegyi megcsapolódási területhez a Gellért és Rudas fürdő kútjai és forrásai (SG4) tartoznak. Az itt levő termálvizek magasabb oldottanyag-tartalommal (jellemzően Ca_2^+, Mg_2^+, HCO_3^- és SO_4^{2-}), de alacsonyabb hőmérséklettel jellemezhetők összehasonlítva az északabbra – a Rózsadomb előterében vagy a Városligetben – található termálvizekkel. A karbonátos medence fedett részén, a pesti oldalon elhelyezkedő kutak szintén külön csoportokat alkotnak. A Margitsziget–II, Széchenyi–I, –II (SG5) csoport regionális áramlási rendszer része, a rózsadombi terület folytatásában elhelyezkedő kutakat tartalmazza, ahol a karbonátos kőzetek egyre nagyobb vastagságban üledékkel fedettek, megnövekedett Na^+ és Cl^- tartalommal és hőmérséklettel. A Csepeli fürdő termálkút (SG3) egyetlen tagból álló csoport, regionális áramlási rendszer része. Különbözőségét a Gellért és Rudas fürdő kútjaitól és forrásaitól (SG4) magasabb hőmérséklet mellett, magasabb Na^++K^+, Ca^{2+}, Mg^{2+}, Cl^- és HCO_3^- tartalma jelzi. A Dagály Béke–kút, Margitsziget-III (SG7) csoport tagjai egy természetes megcsapolódáshoz kapcsolhatók, az egykori Fürdő-sziget környezetében található kutakat tartalmazza. A csoport tagjai alacsonyabb hőmérséklettel és kisebb koncentrációban előforduló medence eredetű komponensekkel (Na^+, Cl^-) jellemezhetők az SG6 és SG5 csoportokkal összehasonlítva [2]. A termálvíz

hasznosítása során lényeges ismernünk a tározó paraméterei, a hidrogeológiai feltételek és a geokémiai jellemzők mellett, a kutak és források csoportjait, ugyanis a csoportok tagjai azonos hidrogeológiai háttérrel rendelkeznek, így megteremthető a vízhasznosítás biztonsága.

A CCDA alapötletének további fontos alkalmazási lehetősége az egy rendszerben meglévő legnagyobb különbségek kimutatása. Egy „vonal" mentén elhelyezkedő monitoring hálózat tagjai esetében ez páronkénti összehasonlítások felhasználásával elvégezhető. Ekkor azt vizsgáljuk, hogy a származási helyek egy adott párjának milyen különbségértéke van, azaz az éppen vizsgált két származási hely mennyire különül el egymástól a véletlenhez képest. Ha azonos a mintaszám a pároknál, akkor a kapott páronkénti különbségértékek egymással is összehasonlíthatók. Mindez akkor a leginkább informatív, ha a párok pontfelhője a paramétertérben nem diszjunkt. Ellenkező esetben a különbségértékek szétválásról igen, de annak pontos mértékéről nem adnak információt. Egy „lineáris" rendszer – például egy folyó mintavételi pontjai [6], [7] vagy egymást követő időintervallumok [8, 2] – esetében jól interpretálhatók az egymást követő párok különbségértékei, mert rávilágítanak a rendszerben bekövetkező legnagyobb változások helyeire, illetve idejére.

A Duna jó példa térben egy lineáris rendszerre, ebből következően értelmezhetők a folyásiránnyal megegyezően, az egymást követő mintavételi pontok páronkénti összehasonlításának eredményeiként kapott különbségértékek. A számítások az 1994–2004 évek kétheti–havi gyakoriságú mintavételezésből származó adataira készültek [7], melyek a legfontosabb kationok és anionok mellett a vízhozam, kémiai és biológiai oxigénigény, összes foszfor és klorofill–a paramétereket tartalmazták, a Duna magyarországi szakaszának 12 mintavételi pontján. A legnagyobb különbséget Rajka és Győrzámoly mintavételi pontok között detektáltuk (2,65%). További hasonló mértékű különbségek voltak Komárom – Almásneszmély (1,44%), Almásneszmély – Szob (1,95%) és Nagytétény – Dunaföldvár (1,59%) között, míg ennél kisebb, de detektálható különbségek jellemzik Győrzámoly – Komárom (0,13%), Szob – Budapest (0,13%), Dunaföldvár – Fajsz (0,16%) és Fajsz – Baja (0,01%) mintavételi pontpárokat. Negatív különbségértékeket kaptunk Budapesttől északra és délre (-0,67%), továbbá Baja – Mohács (-0,42%) és Mohács – Hercegszántó (-0,29%) pontpárok esetén, így ezek szignifikánsan nem különböznek egymástól. Összességében a kapott eredmények alapján a Duna 12 mintavételi pontjából 9 homogén csoport volt elkülöníthető. Közülük hét önálló, – Rajka, Győrzámoly, Komárom, Almásneszmély, Szob, Dunaföldvár, Fajsz – míg egy csoport két, – Budapest, Nagytétény – egy pedig három – Baja, Mohács, Hercegszántó – mintavételi pontból állt. Így a 12 mintavételi pontból a jövőben legalább 9 megtartása javasolt a vizsgálatba bevont paraméterkör és időszak alapján [7].

CCDA–val elvégeztük a Kis–Balaton Vízvédelmi Rendszer (KBVR) monitoringhálózatának optimalizációját. A vizsgálat az 1993-2009 közötti időszak heti–kétheti mintavételezéséből származó adataira történt, alapvetően szervetlen és szerves vízminőségi paraméterekre. Az eredmények szerint a KBVR estében a mintavételi pontok optimális csoportszáma három. Az első csoportot elsősorban az 1985-ben átadott eutróf tó mintavételi pontjai alkották, a másodikat az 1992-ben elárasztott makrofita vegetációval borított vizes élőhely mintavételi pontjai, míg a harmadik csoportot a 205-ös mintavételi pont önállóan alkotta, ami a környezetétől izolált kazettában helyezkedik el. A 12 mintavételi pontból 10 egyedülálló és egy kettő tagból álló homogén csoportra vált szét. A 12 mintavételi pontból ahhoz, hogy megfelelően figyelhessük a KBVR állapotát és folyamatait (legalább) 11 megtartása szükséges [7].

Az egymást követő időintervallumok speciális lineáris rendszernek foghatók fel. Ilyen a Budapest területén levő 27 kútban/forrásban 1960–2009 között mért kémiai (Na^++K^+, Ca_2^+, Mg_2^+, Cl^-, SO_4^{2-}, HCO_3^-) és hőmérsékleti adat, melyekre lehetőség nyílt évtizedes felbontásban megnézni, mikor következtek be a legnagyobb változások. Az eredmények alapján megállapítható, hogy a vizsgált paraméterek mért értékeiben szignifikáns változások következtek be az egyes évtizedek között (pozitív különbségértékek). Megállapítást nyert, az utolsó vizsgált évtized (2000–2009) adatai különböznek leginkább a többi évtized adataitól, de ezen kívül az időbeli változásoknak nincs egyértelmű szerkezete [2].

A Balaton helyes mintavételezéséhez jelentős gazdasági érdekek fűződnek, ugyanis csak így lehet megőrizni a tó jó állapotát és vízminőségét. A Víz Keretirányelv (VKI; [10]) a Balatont egyetlen víztestként határozza meg és egy víztest jellemzéséhez egy mintavételi pontot rendel. A VKI életbe lépése előtt szükségessé vált az „egy tó: egy víztest" koncepció Balatonra vonatkoztatott helyességének vizsgálata és esetleges felülbírálata. Választ kellett keresni, hogy a víztesten belül az eltérő vízminőség alapján hány víztájat lehet kijelölni, illetve ezek alapján hány reprezentatív megfigyelési pont megtartására van szükség? A vizsgált adathalmaz 10 mintavételi pont, 1985–2004 között történt évenkénti négy mintavételezésének, a tápanyagháztartás, és az általános vízkémia paramétereinek mérési eredményeit

tartalmazta. A hierarchikus klaszteranalízis többlépcsős alkalmazásának eredményei alapján a Balaton viselkedésében az időpontoknak három jól elkülöníthető csoportja van. A mintavételi pontoknak – a mintavételi időpontok csoportjaihoz tartozó felosztások közös részei alapján – öt csoportját lehet elkülöníteni [9]. Erre az adathalmazra a CCDA objektív módon szintén öt víztájnak a létezését tárta fel, melyeken belül levő mintavételi pontok homogéneknek tekinthetők [7]. A csoportosítást (víztáj felosztást) leginkább befolyásoló paramétereket két csoportra lehetett osztani. Egyikbe tartoznak az eutro- és oligotrofizációhoz kapcsolódó, tápanyagháztartáshoz elsődlegesen köthető paraméterek, míg a másik, a szervetlen paraméterek csoportja [11]).

A számításokra alkalmazott szoftver a Combined Cluster and Discriminant Analysis (CCDA), melynek fejlesztői: KOVÁCS Solt, KOVÁCS József és TANOS Péter [11]. A program 2013–2014-ben készült, felhasználható minden operációs rendszer alatt, amely támogatja a szabadon használható R programcsomagot (http://cran.r-project.org/). A CCDA fejlesztését az R statisztikai szoftvercsomag tette lehetővé, különösképpen annak base és stats csomagjai (R Core Team, 2013). A CCDA program és dokumentációja elérhető a http://cran.r-project.org/web/packages/ccda/ címen. Programozási nyelv: R, a program mérete: 8,69 kB.

A módszer implementációja néhány R függvényből áll. Ezek közül a legfontosabb a ccda.main, ami két lépést hajt végre. Az alapcsoportosítást (I. lépés) hierarchikus klaszterezés segítségével kapjuk meg (hclust, stats csomagok), Ward módszerét használva a mért paraméterek átlagaira. A magciklus (II. lépés) a ccda.main függvényben az lda függvényt használja a lineáris diszkriminancia analízishez a MASS csomagból [11]. A percentage nevű segédfüggvény az lda kimeneti adataiból a helyesen klasszifikált esetek arányát számolja ki. Ez utóbbi függvény használható egyrészt a dendrogramból kialakuló csoportosítások, illetve a véletlenszerű beosztásoknál a helyesen klasszifikált megfigyelések arányának kiszámításához. Az eredmények értékelése (III. lépés) a felhasználóra van bízva. A döntéshozatalt a ccda.main outputja/eredménye segíti. A ccda.main eredményei: a helyesen klasszifikált esetek (ratio), a véletlenszerű beosztások 95%-os kvantilise (q95) és az ezek közötti különbségérték (difference). Ezeket a számított eredményeket az alapcsoportosítás (dendrogram) minden beosztására megkapjuk. Ezen eredmények, azaz a ratio, q95, illetve ezek különbségének megjelenítésével a plot.ccda.result egy vizuális segédletet nyújt a felhasználónak, hogy dönthessen a további csoportokra bontásról. A plot.ccda.cluster az alapcsoportosítás dendrogramját rajzolja ki.

4. Következtetések

Összességében a CCDA célja nem csupán a hasonló [12, 3, 13, 14, 15, 16, 17, 9, 18], hanem homogén csoportok keresése a csoportok közötti legkisebb különbségek megtalálására [19, 1, 7]. A módszer alkalmas az optimális csoportosítás meghatározására [2] és páronkénti összehasonlítások alkalmazásával képes egy rendszerben meglévő legnagyobb különbségek kimutatására [6, 8, 7]. A CCDA nem csak a föld- és környezettudományok, hanem más szakterületeken is felhasználásra került [20, 21] és az alkalmazási területek várhatóan folyamatosan bővülnek.

Irodalomjegyzék

1. KOVÁCS, J., KOVÁCS, S., MAGYAR, N., TANOS, P., HATVANI, I. G., ANDA, A., 2014: Classification into homogeneous groups using combined cluster and discriminant analysis. *Environmental Modelling and Software*, **57**, 52–59
2. KOVÁCS, J., ERŐSS, A., 2017: Statistically optimal grouping using combined cluster and discriminant analysis (CCDA) on a geochemical database of thermal karst waters in Budapest. *Applied Geochemistry* **84**, 76–86.
3. DÉRI-TAKÁCS, J., ERŐSS, A., KOVÁCS, J., 2015: The chemical characterization of the thermal waters in Budapest, Hungary by using multivariate exploratory techniques. *Environmental Earth Sciences* **74(12)**, 7475–7486.
4. DAVIES, D.L., BOULDIN, D.W., 1979: A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, 224–227.
5. ALFÖLDI, L., BÉLTEKY, L., BÖCKER, T., HORVÁTH, J., KORIM, K., RÉMI, R., 1968: *Budapest Hévizei*. Hungarian Institute for Water Resources Research Budapest, 365 pp, Budapest
6. CHAPMAN, D,V., BRADLEY, C., GETTEL, G,M., HATVANI, I.G., HEIN, T., KOVÁCS, J., LISKA, I., OLIVER, D,M., TANOS, P. TRÁSY, B., VÁRBÍRÓ, G., 2016: Developments in water quality monitoring and management in large river catchments using the Danube River as an example. *Environmental Science & Policy* **64**, 141–154.

7. KOVÁCS, J., KOVÁCS, S., HATVANI, I., G., MAGYAR, N., TANOS, P., KORPONAI, J., BLASCHKE, A.P., 2015: Spatial optimization of monitoring networks on the examples of a river, a lake–wetland system and a sub–surface water system. *Water Resources Management* **29(14)**, 5275–5294
8. TANOS, P., KOVÁCS, J., KOVÁCS, S., ANDA, A., HATVANI, I.G., 2015: Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environmental Monitoring and Assessment* **187(9)**, 575.
9. KOVÁCS, J., NAGY, M., CZAUNER, B., KOVÁCS, I.SZ., BORSODI, A.K., HATVANI. I.G., 2012:, Delimiting sub-areas in water bodies using multivariate data analysis on the example of Lake Balaton (W Hungary). *Journal of Environmental Management* **110**, 151–158.
10. WFD, 2000: *Directive of the European Parliament and of the Council 2000/60/EC Establishing a Framework for Community Action in the Field of Water Policy*. European Union, Luxembourg. PE-CONS 3639/1/00 REV 1
11. KOVÁCS, S., KOVÁCS, J., TANOS, P., 2014: Combined Cluster and Discriminant Analysis, https://cran.r-project.org/web/packages/ccda/ccda.pdf, 1–6.
12. BARICZA, Á., BAJNÓCZI, B., KOVÁCS, J., MAY, Z., SZABÓ, M., SZABÓ, C., TÓTH M., 2018: Chemical durability of lead – bearing glazes in sulphuric acid solutions – Laboratory experiments performed on Zsolnay architectural ceramics from Budapest (Hungary). *International Journal of Architectural Heritage* **12(2)**, 216-236.
13. HATVANI, I.G., KOVÁCS, J., KOVÁCSNÉ SZÉKELY, I., JAKUSCH, P., KORPONAI, J., 2011: Analysis of long term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. *Ecological Engineering* **37(4)**, 629–635.
14. HATVANI, I.G., CLEMENT, A., KOVÁCS, J., KOVÁCS, I.S., KORPONAI, J., 2014. Assessing water–quality data: The relationship between the water quality amelioration of Lake Balaton and the construction of its mitigation wetland. *Journal of Great Lakes Research*, **40(1)**, 115–125.
15. KOVÁCS, J., MÁRKUS, L., CSEPREGI, A., 1997: Grouping of Wells by Groundwater Levels and Chemical Data, International Conference on Applied Mathematic, 287, Hong Kong
16. KOVÁCS, J., VID, G., MAUCHA, L., BERÉNYI ÜVEGES, J., 2005: Az Aggteleki–karszt nagy forrásainak és a Baradla illetve a Béke–barlangban a járattalp alatt észlelt vizek kémiai összetevőinek vizsgálata többváltozós adatelemző módszerekkel. In: Veress, M. (Ed.), *Karsztfejlődés* X. Berzsenyi Dániel Főiskola Természetföldrajzi Tanszék, Szombathely, 107–120.
17. KOVÁCS, J. TANOS, P., KORPONAI, J., KOVÁCSNÉ SZÉKELY, I., GONDÁR, K., GONDÁR–SŐREGI, K., HATVANI, I., G., 2012: Analysis of Water Quality Data for Scientists. In: Kostas, V., Dimitra, V. (Eds.), *Water Quality Monitoring and Assessment*. InTech Open Access Publisher, 65–94, Rijeka
18. KOVÁCS, J., BODNÁR, N., TÖRÖK, Á., 2016: The application of multivariate data analysis in the interpretation of engineering geological parameters, *Open Geosciences* **8(5)**, 52–61
19. FARICS, É., FARICS, D., KOVÁCS, J., HAAS, J., 2017: Interpretation of sedimentological processes of coarse-grained deposits applying a novel combined cluster and discriminant analysis. *Open Geosciences* **9(1)**, 525–538.
20. BÁNFI, R., POHNER, ZS., KOVÁCS, J., LUZICS, SZ., NAGY, A., DUDÁS, M., TANOS, P., MÁRIALIGETI, K., VAJNA, B,. 2015: Characterisation of the large-scale production process of oyster mushroom (Pleurotus ostreatus) with the analysis of succession and spatial heterogeneity of lignocellulolytic enzyme activities. *Fungal Biology*, **119(12)**, 1354–1363.
21. NOVÁK, M., PALYA, D., BODAI, ZS., NYIRI, Z., MAGYAR, N., KOVÁCS, J., EKE, ZS., 2017: Combined Cluster and Discriminant Analysis an efficient Chemometric Approach in Diesel Fuel Characterization. *Forensic Science International* **270**, 61–69.