

Doktori értekezés tézisei

Rácz Anita

Budapest
2016



Kemometria és FT-NIR spektroszkópia alkalmazása az élelmiszeranalitikában

Rácz Anita

Budapest
2016

A doktori iskola

megnevezése: Élelmiszertudományi Doktori Iskola

tudományága: Élelmiszertudományok

vezetője: **Dr. Vatai Gyula**
egyetemi tanár, DSc
SZIE, Élelmiszertudományi Kar,
Élelmiszeripari Műveletek és Gépek Tanszék

Témavezetők: **Dr. Fodor Marietta**
habilitált egyetemi docens, PhD
SZIE, Élelmiszertudományi Kar,
Alkalmazott Kémia Tanszék

Dr. Héberger Károly
tudományos tanácsadó, DSc
MTA Természettudományi Kutatóközpont,
Anyag- és Környezetkémiai Intézet

A jelölt a Szent István Egyetem Doktori Szabályzatában előírt valamennyi feltételnek eleget tett, az értekezés műhelyvitájában elhangzott észrevételeket és javaslatokat az értekezés átdolgozásakor figyelembe vette, ezért az értekezés nyilvános vitára bocsátható.

.....
Az iskolavezető jóváhagyása

.....
A témavezető jóváhagyása

.....
A témavezető jóváhagyása

I. A MUNKA ELŐZMÉNYEI

Az élelmiszertudományi és élelmiszeripari vizsgálatokban előszeretettel használnak olyan analitikai eszközöket, amelyek gyors, költséghatékony és lehetőség szerint roncsolásmentes vizsgálatokat biztosítanak. A doktori munkám során használt közeli infravörös spektroszkópia is ezt az irányelvet követi, amely hatékonyan alkalmazható az élelmiszerek beltartalmi paramétereinek ellenőrzésére, mennyiségi és minőségi vizsgálatára. Az élelmiszerek minőségellenőrzése nem csupán a fogyasztók, de a gyártók érdeke is, hiszen hatalmas a verseny az élelmiszerek piacán, ahol a megfelelő minőség garanciája már elengedhetetlen feltételnek számít. Bár a magyar, élelmiszervizsgálatra vonatkozó szabványok sok esetben “elavultnak”, hosszadalmasnak és drágának tekinthetők, a feltörekvően lévő gyors analitikai eljárások – köztük a közeli infravörös spektroszkópia – alkalmazásával egy új, gazdaságilag sokkal kedvezőbb és hatékonyabb irányt mutathatunk meg.

Az ezredfordulót megelőző évtized óta a modern analitikai kémia és ezzel együtt az élelmiszeralitika jelentősége is rohamosan nő. Az egyre modernebb és gyorsabb számítógépek megjelenése hatalmas szerepet játszott az analitikai módszerek fejlődésében, főként a spektroszkópia területén felhalmozódó nagymennyiségű adat feldolgozásában. Az utóbbi évtizedek “adatrobbanásával” járó hosszadalmas számolásokat a nagyteljesítményű számítógépek (akár szuperszámítógépek) használata nem csak lehetővé tette, de meg is könnyítette. A kemometria, mint a statisztika többváltozós kémiai adatokra történő alkalmazási területe már az előző évszázad közepétől egyre népszerűbbé vált, de az igazi áttörést a 80-as és 90-es évek számítógépes fejlődése hozta meg. Olyan cikkek születtek akkor, amelyek a mai napig a kemometriai tudomány alapillérei.

Az élelmiszertudományban is a kemometria számos alkalmazására lelhetünk. Segítségével felderíthetjük az adatainkban rejlő mintázatokat, hasonlóságokat és különbségeket, olyan rejtett összefüggéseket, amelyek mennyiségi és minőségi kiértékelések alapjául szolgálhatnak. A kemometriát előszeretettel társítják a spektroszkópiai kiértékelésekhez, hiszen a spektroszkópiai módszerek közül jó néhány az általuk generált hatalmas adatkészletek miatt elképzelhetetlen lenne az adatok kemometriai értékelése nélkül. A doktori munkám során alkalmazott legfontosabb analitikai eszköz, a közeli infravörös spektroszkópia elengedhetetlen feltétele a sokváltozós adatelemzés, a kemometria. A komplex élelmiszermatrixok spektrumainak kiértékelése csakis a kemometria segítségével történhet meg.

II. A KITŰZÖTT CÉLOK

A doktori dolgozatom a vizsgált mintamátrixok alapján három nagyobb részre tagolható: i) a Q10 tartalmú étrendkiegésztők ii) az energaitalok komplex vizsgálata, valamint iii) az antioxidánsok vizsgálata. A kemometriai módszerek fejlesztéseit a dolgozatomban az előbbieken felsoroltaktól külön tárgyalom. A doktori munkám során a következő célokat fogalmaztam meg az egyes fejezetekhez tartozóan:

Q10 koenzim tartalmú étrendkiegésztők:

- A Q10 koenzim tartalmú étrendkiegésztők hatóanyagtartalmának vizsgálata FT-NIR spektroszkópiával, kalibrációs modell készítése és validálása keresztellenőrzéssel és külső teszt mintákkal. Az eddig elterjedt kromatográfias módszer kiváltása FT-NIR spektroszkópia segítségével.
- A különböző változókiválasztási módszerrel létrehozott modellek összehasonlítása a rangszámkülönbségek összegének módszerével (SRD)

Energaitalok vizsgálata:

- A koffein tartalom meghatározása HPLC-UV eljárással, mint referencia módszerrel. A kromatográfias módszer energaital mintákra történő optimalizálása. Ezt követően FT-NIR módszerrel a regressziós modell megalkotása, valamint belső és külső validálása.
- Cukortartalom meghatározása Schoorl referencia módszerrel, majd FT-NIR regressziós modellépítés a Schoorl referencia módszerrel kapott, illetve a nominális (dobozon feltüntetett) értékek használatával. A létrehozott modellek belső és külső validálása új teszt mintákkal is.
- Az energaitalok osztályozása az FT-NIR spektrumaik alapján. Elsőként cukortartalom szerint, majd a taurinos, arginines és normál (taurint és arginint nem tartalmazó) minták elkülönítése. Az osztályozás elvégzése többféle mintázatfelismerő módszerrel is.

Antioxidánsok vizsgálata:

- Már meglévő antioxidáns kapacitás adatok alapján, két adatkészlet (esettanulmány) felhasználásával, összehasonlító elemzés készítése a meghatározási módszerekről: Az antioxidáns kapacitás módszerek osztályozása és rangsorolása.

Kemometriai módszerfejlesztések:

- Új változóselektációs módszer létrehozása és tesztelése a Q10 koenzim tartalmú étrendkiegészítők FT-NIR spektrumának felhasználásával. A modellek jóságának növelése a közel 2300 változószám csökkentésével.

- Olyan, „*n*-class” vevő-működtető görbék (ROC görbék) megalkotása, melyekkel kettőnél több osztályú elemzések is kiértékelhetővé válnak. A módszer megfelelő validálása és tesztelése az energiaitalok osztályozása során. A mintázatfelismerési módszerek összehasonlítása az „*n*-class” ROC görbék segítségével.

- A véletlen fák (random forest) használata során történő optimalizálási lépés kidolgozása a minél jobb osztályozási modellek létrehozására.

- Regressziós modellek teljesítmény paramétereinek összehasonlítása és rangsorolása. Mennyiségi szerkezet-hatás összefüggések segítségével kapott regressziós modellek adatkészletei alapján, hogy szélesebb körű és általánosabban megfogalmazható eredményt kapjunk.

III. ANYAG ÉS MÓDSZER

A doktori munkám során a kemometriai adatfeldolgozások egy részét FT-NIR spektrumok és klasszikus valamint nagyműszeres analitikai módszerek által kapott adatokon végeztem el. A kemometriai módszerfejlesztések is részét képezték a dolgozatomnak, ezeket az új módszereket az Eredmények fejezetben tárgyalom.

3.1 Az FT-NIR készülék

Az általam használt Bruker MPA NIR/NIT készüléket (Bruker Optik GmbH, Ettlingen, Németország) diffúz-reflexiós és transzmissziós mérési módban is alkalmaztam. Míg a Q10 koenzim tartalmú étrendkiegészítők vizsgálatánál a szilárd mintákhoz megfelelően a forgó mintatartóval ellátott reflexiós egységet, az energiatalokhoz az átfolyó küvetta (1 mm rétegvastagság) termosztálható transzmissziós egységet használtam. Diffúz-reflexiós üzemmódban a spektrumokat 12500 cm^{-1} és 3800 cm^{-1} közötti tartományban, az energiatalok transzmissziós spektrumát 12500 cm^{-1} és 4000 cm^{-1} hullámszám között rögzítettem. A készülék mindkét mérési módjában a spektrális felbontás 8 cm^{-1} volt, a szkennelési sebesség pedig 10 kHz. A készülék ezekben az esetekben 32 „alspektrumot” rögzít, és ezek átlagát kapjuk egy-egy felvétel eredményeként.

3.2 Nagyhatékonyságú folyadékkromatográfia

A Q10 koenzim FT-NIR vizsgálatait megelőzően a hatóanyag tartalom meghatározása HPLC-UV referenciaméréssel történtek. A meghatározás Agilent 1200 HPLC készülék használatával (Agilent Technologies, Santa Clara, CA, USA) történt izokratikus módban Agilent Zorbax XDB C18 HPLC kolonna segítségével (2,1 mm \times 50 mm \times 3,5 μm), UV detektorral 275 nm-en nyomon követve. A kolonnatér hőmérséklete 30 $^{\circ}\text{C}$ volt. Eluensként acetonitril (ACN) : tetrahidrofurán (THF) : víz 65:30:5 térfogat százalékos (v/v %) elegyét használtam. Az áramlási sebesség 0,35 ml/perc volt, az injektálási térfogat pedig 10 μl . A fejlesztett módszer detektálási limitje 0,05 mg Q10 koenzim / g volt.

Az energiatalok HPLC-UV méréseinél, kávék koffein tartalmának meghatározását leíró nemzetközi standard eljárás (ISO 20481:2008) továbbfejlesztett változatát alkalmaztam. A készülék hasonlóan az előző esethez az Agilent 1200 HPLC volt UV detektorral felszerelve. A szabvány módszer módosításán átesett részletei a következők voltak: kollonként egy Agilent

Zorbax XDB C18 kolonnát (4,6 mm × 150 mm × 5,0 μm) választottam. A kromatogramok felvétele izokratikus módban 40 °C-on történt. Áramlási sebességnek 1 ml / perct választottam, az injektálási térfogat pedig 20 μl volt. A futtatási idő 18 perc volt. Az UV detektálás 273 nm-en történt, de párhuzamos mérésenként 260 nm-en csúcstiszaság vizsgálatot is végeztem. Erre azért volt szükség, hogy kimutatható legyen, ha az adott mintában bármilyen egyéb komponens esetleg zavaró tényezőként jelentkezne hasonló retenciós időnél.

3.3 Mintaelőkészítés

A Q10 koenzim minták HPLC méréseinek mintaelőkészítése Vass és munkatársai által közölt publikációban (2014) került részletes ismertetésre.

Az energiailatok HPLC méréseihez ultrahangos fürdő (T2MODX; VWR, Radnor, PA, USA) segítségével szén-dioxid mentesítettem a mintákat 20 percen keresztül. Ezt követően 50 μl-t hígítottam 1600 μl-re desztillált vízzel. A külső kalibrációs mintasort használtam a koffein koncentráció meghatározására, amelyben a pontok rendre a következők voltak: 2,5; 5,0; 10,0 és 20,0 ppm.

Az FT-NIR mérések esetén a Q10 koenzim tablettákat dörzsmozsárban megfelelően elporítottam és homogenizáltam. Az energiailatok mérése során az ultrahangos fürdő segítségével gázmentesített mintákból 10-10 ml-t használtam fel.

3.4 Klasszikus mérési módszer – Schoorl módszer

A Schoorl módszer az élelmiszerek cukortartalmának egyik legelterjedtebb meghatározási módszere, mely során a redukáló cukrokat, illetve hidrolízis segítségével redukáló cukrokká lebomló cukrokat lehet vizsgálni. A meghatározást az OÉTI módszergyűjteményének ÉLK 4.009 szabványa alapján végeztem el.

Az energiailatok cukortartalmának meghatározása során invert cukorra számoltam ki az eredményeket, tekintve, hogy savas hidrolízissel invertáltam a mintákat. Az energiailatokon csak nagyon kevés esetben van feltüntetve a konkrét vegyület, gyakran változnak a cukor, invertcukor, glükóz stb. megfogalmazások. Így az értékek invertcukorra történő megadása egy átlagos, jó közelítésnek mondható.

3.5 Szoftverek, kemometriai módszerek

Az FT-NIR spektrumok kiértékeléséhez az OPUS 7.2 (Bruker Corp., Ettlingen, Németország) és az Unscrambler 9.7 szoftvereket (CAMO Software, Oslo, Norway) alkalmaztam. A mintázatfelismerésre használt kemometriai módszereket (PCA, LDA, RF, BT, PLS-DA) a STATISTICA 12 (Statsoft Inc., Tulsa, OK, USA) program segítségével alkalmaztam. Az SRD és GPCM módszereket MS Excel segítségével használtam. Az „*n*-class” többosztályos ROC görbék esetében egy saját Linux programkódot hoztam létre, amely a doktori dolgozatom mellékletét is képezi.

IV. EREDMÉNYEK

4.1 Kemometriai módszerfejlesztések

4.1.1 Az intervallum szelektivitási arány (iSR)

A módszer alapötletét, a szelektivitási arány módszert, Rajalahti és munkatársai (Rajalahti et al., 2009) dolgozták ki. Az eredeti egyenletet némiképp meg kellett változtatni az „intervallumosítás” miatt: a nevezőbe egy négyzetgyök függvényt (RMSEC) még beillesztettem. A módosított egyenlet a szelektivitási arány alapján a következő:

$$(1) \quad SR_i = R_i^2 / RMSEC_i, i = 1, 2, 3 \dots m.$$

Ahol i az intervallumok számát jelöli. Az egyenletben az R^2 esetén a szabadsági fokokat kihagytam, amely viszont egyenlő mértékben befolyásolja az összes intervallumot (ekvidisztáns felosztás esetén, így a végső ábrán nem okoz aránytalan eltéréseket. A szelektivitási arányhoz hasonlóan, az intervallum SR is, minél nagyobb, annál fontosabb az adott intervallum a modellépítés szempontjából.

4.1.2 „n-class”, avagy több osztályos ROC görbék

A többsztályos ROC görbék esetében, a ROC görbe alatti terület (AUC) úgy számolható ki, hogy az egyik csoportot pozitívként kezeljük az összes többit pedig negatívként és ezt ismétljük annyiszor, ahány csoportunk van.

Az AUC értékek súlyozott átlaga adja meg a teljes osztályozásra adható AUC értéket:

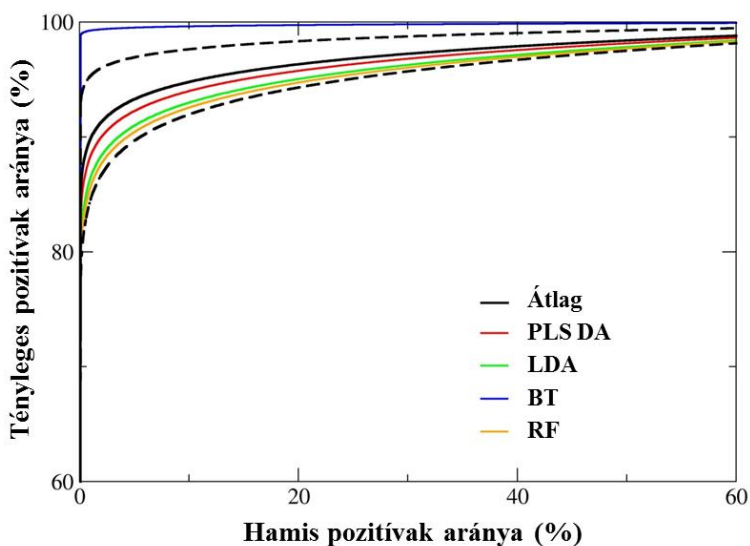
$$(2) \quad \overline{AUC} = \frac{\sum_{i=1}^n N_i AUC_i}{\sum_{i=1}^n N_i},$$

Ahol N_i a csoportokban (osztályokban) lévő minták száma. A Hanley formula segítségével vizualizálni tudunk egy „ROC-szerű” görbét az AUC értékek súlyozott átlagának felhasználásával (összesített AUC érték). A hibaterjedés törvényének alkalmazásával az előbbieken említett átlagos AUC érték variáciája is kiszámolható (Ku, 1966):

$$(3) \quad Var(\overline{AUC}) = \frac{\sum_{i=1}^n N_i^2 Var(AUC_i)}{\sum_{i=1}^n N_i^2}$$

Több módszer ROC görbékkel történő összehasonlításakor, ahhoz hogy el tudjuk dönteni, hogy az adott módszerek szignifikánsan különböznek-e egymástól, az átlag \pm az összesített AUC érték egyszeres standard deviációját alkalmaztam.

A többosztályos ROC görbék segítségével négyféle osztályozási módszer, a véletlen erdők (RF), fejlesztett fák (BT), LDA és PLS DA módszerek teljesítményét hasonlítottam össze. A vizsgálatokhoz energiatalok FT-NIR spektrumait és azok cukortartalom szerinti csoportosíthatóságát osztályozási változóként használtam fel. Két adatkészletet használtam fel az osztályozási módszerek összehasonlítására. Az első esetben 90 energiatal minta FT-NIR spektrumát, majd ugyanezen minták spektrumainak főkomponens-elemzés révén kapott főkomponenseit. A főkomponensek alapján elvégzett osztályozások eredményét az 1. ábrán szemléltetem. A ROC görbék alapján a legjobb osztályozást a fejlesztett fák (BT) módszer segítségével tudtam elérni.



4. ábra: A vizsgált négy osztályozási módszer átlagos ROC görbéi.

A tényleges pozitívák arányát ábrázoltam a hamis pozitívák függvényében. Az átlag AUC értékhez tartozó ROC görbe feketével van jelölve, a hozzá tartozó ± 1 SD pedig szaggatott vonallal. Az ábra az eredeti fókuszált verziója a jobb láthatóság érdekében.

4.2 Q10 tartalmú étrendkiegészítők hatóanyagtartalmának vizsgálata

Összesen 52 különböző Q10 koenzim tartalmú étrendkiegészítő FT-NIR spektrumát vettem fel. A spektrum 12500 cm^{-1} és 9000 cm^{-1} (800–1111 nm) közötti részét kihagytam a kiértékelésből, mivel a spektrum ezen szegmense nem hordozott szisztematikus információt számomra. Az első lépésként a lehetséges spektrális kieső mintákat szűrtem ki az adatkészletből főkomponens-elemzés segítségével. Az így megtalált két spektrális kiesőt kihagytam a további modellépítésekben. Ezt követően az adatelőkezelési lépést és háromféle változó szelektálási módszert alkalmazva (iPLS, iSR és genetikus algoritmus) megalkottam a PLS regressziós modelleket. Többféle adatelőkezelést is kipróbálva a deriválás adta a legjobb eredményt. A modellek belső validálását öt részre osztott kereszt-ellenőrzéssel hajtottam végre és ezt kiegészítettem minden esetben egy külső teszt validálással is. Az 1. táblázatban összefoglaltam a három legjobb regressziós modellem teljesítmény paramétereit. A táblázat első sorában a kiindulási, skálázás és változószelektálás nélküli modell található.

1. táblázat: A Q10 koenzim koncentráció meghatározására szolgáló modellek. Vastagon kiemelve találhatóak a későbbi alkalmazásra javasolható (legjobb) modellek.

R^2	Q^2	RMSEC	RMSECV	RMSEP	Skál.	Komp. szám	Vált. szel.
0,85	0,71	11,26	15,74	-	-	9	-
0,92	0,87	8,16	10,58	8,82	D	7	iPLS
0,90	0,87	8,90	10,85	13,74	D	6	iSR
0,91	0,88	8,48	10,22	11,12	D	6	GA

R^2 : a determinációs együttható négyzete, Q^2 : A kereszt-ellenőrzésre vonatkozó determinációs együttható négyzete, RMSE: átlagos négyzetes hiba (C=kalibrációra vonatkozó, CV=kereszt-ellenőrzésre vonatkozó, P=teszt validálásra vonatkozó) értéke. A D a deriválás rövidítésére szolgál.

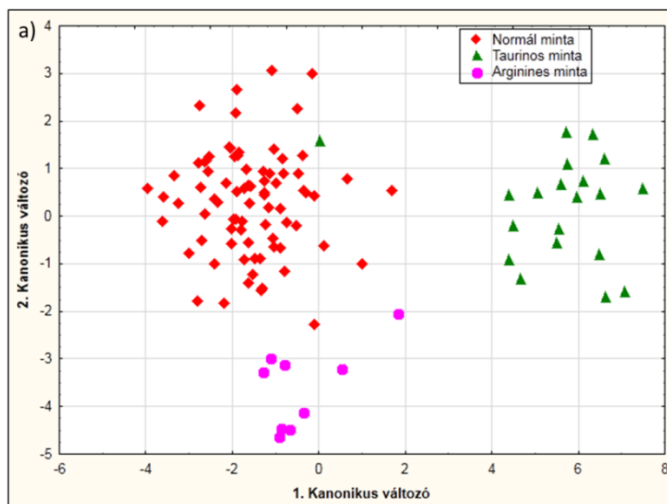
A modelleket SRD módszer segítségével összehasonlítva a legkonzisztensebb és így legjobb modellnek is az iPLS modell tekinthető az előrebecslési értékeket az átlaghoz (mint referenciához) hasonlítva.

4.3 Az energiaitalok FT-NIR spektrumának kemometriai elemzése

Az eredmények bemutatását két csoportra tagolva teszem meg: i) mintázatfelismerési/osztályozási modellek létrehozása valamint ii) regressziós modellépítések az energiaitalok koffein és cukortartalmának becslésére, mivel ez a két szegmens kemometriai szempontból teljesen más módszert és szemléletet igényel.

4.3.1 Energiaitalok osztályozási lehetőségei

Az energiaitalok cukortartalma mellett (lásd. ROC görbék) osztályozni akartam a különböző magyar, szlovák és görög mintákat arginint, taurint és az előbbieket nem tartalmazó csoport szerint. Ehhez összesen 108 energiaital FT-NIR spektrumát vettem fel majd PCA és LDA módszerekkel végeztem el a kiértékelést. A PCA szerepe a kiértékelésben az adatkészlet „redukciója” volt. A spektrumfelvételt követően az adatmátrixot standardizáltam. Az így kapott adatkészletet PCA módszerrel értékeltem ki és a kapott 20 főkomponenst használtam a további LDA elemzésekhez. A kapott osztályozási modell eredménye a 2. ábrán látható. Az osztályozási modell validálását három részre osztott kereszt-ellenőrzéssel és X véletlenszerűségi (randomizációs) teszttel végeztem el. A validálások megfelelően alátámasztották a modell helyességét.

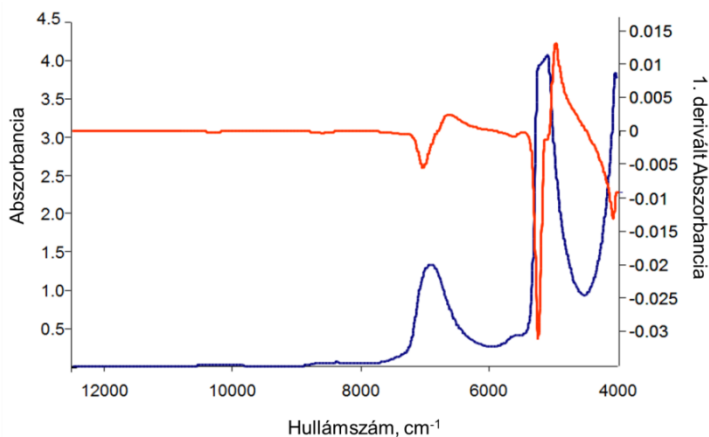


2. ábra: Az energiaitalok osztályozási modellje: elkülönítés az arginines, taurinos és normál (taurint és arginint nem tartalmazó) minták szerint.

4.3.2 Energiaitalok koffein és cukortartalmának meghatározása

Koffein tartalom meghatározása

Első lépésben HPLC-UV technika segítségével 42 eredeti energiital minta koffein tartalmát határoztam meg. Ezen kívül 33 keverék mintát is létrehoztam az eredeti minták felhasználásával. A jellegzetes koffein csúcs a kromatogrammon megközelítőleg 9,5 percnél volt látható. Az FT-NIR mérésekhez 10 ml mintát használtam fel minden mintából. Egy példa-spektrum és annak derivált verziója a 3. ábrán látható.



3. ábra: A vizsgált minták egy szemléletes példa spektruma, ill. annak derivált formája.

Az eredeti spektrumot kékkkel, míg a derivált verzióját pirossal jelöltem.

A koffein koncentrációja a HPLC-UV mérések alapján 118 és 338 mg / 100 ml között változott. A modellépítés folyamatában nem találtam spektrális kiesőnek tekinthető mintákat, így az optimalást és adatelőkezelést követően megalkottam a végső PLS regressziós modellt 8 PLS komponens felhasználásával. A modellt hét részre osztott kereszt-ellenőrzéssel és külső validálással is ellenőriztem 13 új, kereskedelemben kapható mintára. A modell teljesítmény paraméterei a 2. táblázatban összefoglalva találhatóak meg.

Cukortartalom meghatározása

A cukortartalom meghatározásához 71 eredeti és 20 keverék minta (összesen 91 minta) FT-NIR spektrumát használtam fel. A referencia Y értékeként a Schoorl módszerrel meghatározott cukortartalom értékeket alkalmaztam. A teljes mintasorozatból ezzel a technikával 75 mintát tudtam megmérni. Sajnos a Schoorl módszer bár rendkívül népszerű, elég nagy torzítású és szórású (12,4 %), különösen az kisebb koncentrációk tartományában (1-2 g / 100 ml). Ezért úgy döntöttem, hogy az ezzel a módszerrel kapott értékek mellett a nominális (dobozon feltüntetett) értékekkel is végrehajtom a modellépítést.

A spektrális kiesőket ismét PCA elemzés segítségével határoztam meg, amelynek eredményeként két mintát kellett kihagynom a további modellépítési folyamatból. A nominális koncentrációk esetében fennmaradó 89 mintára adatelőkezelésként első deriváltat és standardizálást használtam. A validálási protokoll megegyezett a koffein esetében használtakkal. A Schoorl módszer alapján meghatározott Y értékekkel 73 mintára végeztem el a modellépítést. A cukortartalom koncentráció tartománya 0,1 és 15,3 g / 100 ml között volt. Mindkét modellhez hat PLS komponensre volt szükség a végső modell megalkotásához, amelyek megfelelően pontosnak és robusztusnak bizonyultak. A Schoorl módszert használva referenciaként a modell nagyobb hibával rendelkezett, hiszen maga a referencia módszer is nagyobb hibájú volt. A cukortartalom meghatározásra létrehozott modellek teljesítmény paramétereit a 2. táblázat tartalmazza.

2. táblázat: Az energiai italok koffein és cukortartalmára vonatkozó regressziós modellek teljesítmény paramétereit.

	N	C	R ²	Q ² _{ext}	Q ²	RMSECV	RMSEP
<i>Koffein modell</i>	75	8	0,966	0,898	0,928	16,8	36,3
<i>Cukor modell (Schoorl)</i>	73	6	0,943	0,935	0,919	1,13	1,23
<i>Cukor modell (nominális)</i>	89	6	0,998	0,996	0,995	0,29	0,26

N a mintaszámot jelöli, C a PLS komponensek számát, valamint az RMSECV és RMSEP (lásd. 1. táblázat) értékek mértékegysége a cukor modellek esetén g / 100 ml volt, a koffein modell esetében pedig ppm. Az ext jelölés a külső teszt validálás Q² értékére vonatkozik.

A három modell teljesítmény paramétereit figyelembe véve elmondható, hogy mind a koffein mind a cukortartalom meghatározása sikeresen megtörtént az FT-NIR spektrumok alapján.

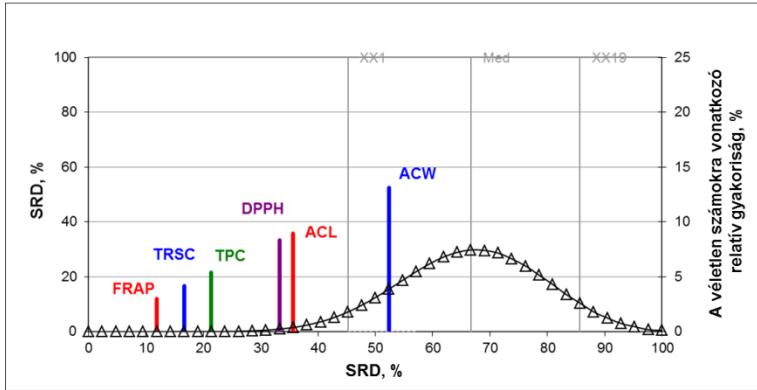
4.4 Antioxidáns kapacitás meghatározási módszerek csoportosítása, rangsorolása

Az antioxidáns kapacitás meghatározási technikák összehasonlításához két adatkészletet használtam fel: az első 13 bogyós gyümölcs mintát (szamóca, málna, piros és fekete ribiszke) tartalmazott, a második pedig 12féle meggy mintát. Összességében hétféle meghatározási módszerrel vizsgálták meg a mintákat, adatkészletekre lebontva a bogyós gyümölcsök antioxidáns kapacitását a FRAP, TPC, TRSC, DPPH, ACL és ACW módszerekkel, míg a meggy mintákét FRAP, TPC, TEAC, ACL és ACW módszerekkel mérték meg. (A rövidítések feloldása a dolgozat rövidítés jegyzékében található.)

A csoportosíthatóságokat és kapcsolatokat a különböző módszerek között a PCA és hierarchikus fürtelemzés (HCA) segítségével tártam fel. Az antioxidáns kapacitás meghatározási technikák rangsorolásához az SRD és GPCM technikákat használtam.

A fürtelemzés során Euklideszi távolságot használtam távolságmértékként és Ward módszert kapcsolási szabályként mindkét adatkészlet esetén. A bogyós és a meggy fajták adatkészlete esetén is az ACW és ACL módszerek elkülönültek a többi módszertől. A fürtelemzés során kapott mintázatot PCA segítségével ellenőriztem. A bogyós gyümölcsök adatkészlete esetén az ACW, ACL és TPC módszerek az első két főkomponens-együttható egymás függvényében ábrázolva nem alkottak csoportokat. A meggy fajták esetén az ACW és ACL módszerek elkülönülő mintázata ismét megfigyelhető volt.

Az SRD elemzés már kimondottan a módszerek közötti rangsorolásra irányult. Az adatkészletek kiegészültek egy referencia, sorátlag oszloppal is. Ez a megoldás az olyan esetekben kedvező, ha nincs egzakt referencia értékünk az adatmátrixban. A 4. ábrán a bogyós gyümölcs adatkészletből számolt skálázott SRD értékek láthatóak.



4. ábra: A skálázott SRD értékek (százalékos) ábrázolása a módszerek függvényében a bogyós gyümölcs adatkészlet esetében. A baloldali Y és X tengelyeken az SRD értékek, míg a jobboldali Y tengelyen a véletlen számokra vonatkozó relatív gyakoriság van feltüntetve. Az ábrán látható Gauss-szerű eloszlás az előbbi relatív gyakoriságokhoz tartozó görbe.

Minél kisebb a kapott SRD érték, annál közelebb van a referencia értékhez (jelen esetben az átlaghoz). Így az eredmény alapján elmondható, hogy a FRAP módszer esett a legközelebb a referenciához, vagyis ez a módszer képes a legkisebb hibával kiváltani a többi módszert. A véletlen számokra vonatkozó relatív gyakoriságok Gauss-szerű görbéjéből pedig láthatjuk. Ebben az esetben majdnem mindegyik antioxidáns kapacitás meghatározási technika jobbnak mutatkozott a véletlen számoknál, kivéve az ACW technikát. Az eredmény validálásához véletlenszerűségi tesztet és hétszeres kereszt-ellenőrzést használtam. A kereszt-ellenőrzés eredményeinek felhasználásával kimutatható volt, hogy a DPPH és ACL módszerek nem különböztek egymástól szignifikánsan.

A meggy adatkészletet vizsgálva a TPC eredményezte a legkisebb SRD értéket, így ebben az esetben ez a módszer képes a többi vizsgált módszer kiváltására. Az ACW módszer viszont kívül helyezkedtek az elfogadható tartományon, míg az ACL határeset. A FRAP és TEAC módszerek ugyanazon SRD értékeket adták, és a validálás során kapott értékek mediánjai Sign és Wilcoxon próbával vizsgálva sem térnek el szignifikánsan a két módszer esetén.

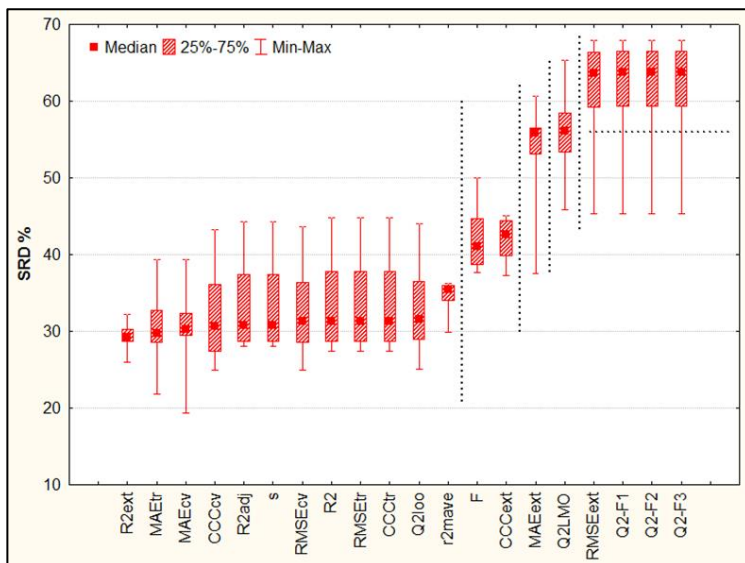
Az általánosított pár-korrelációs módszer segítségével (GPCM) az eredmények rendkívül hasonlóak lettek az SRD során kapottakhoz, viszont így a GPCM technikával a DPPH és ACL módszerek között az első adatkészlet

esetén, míg a FRAP valamint TEAC módszerek között a második adatkészlet esetén különbséget tudtam tenni.

4.5 Regressziós modellek előrejelző képességét meghatározó paraméterek összehasonlítása

A munkám során egyfajta kitekintésképpen, két olyan adatkészletet használtam fel a becslés jóságára vonatkozó paraméterek összehasonlítására, amelyek bár a gyógyszerkémia terültéről származnak, de sokkal szélesebb körű, nagyobb mennyiségű paraméter számolását tették lehetővé. A két vizsgált adatkészlet egyike különböző benzol származékok toxicitás értékeit tartalmazta független (Y) változóként, míg a másik N-szubsztituált maleimidek inhibitor aktivitás értékeit (IC_{50}). A kiértékelésekhez többváltozós lineáris regressziót (MLR) használtam. A kapott modellek teljesítményparamétereit SRD módszerrel rangsoroltam, és hasonlítottam össze. Az első adatkészlet esetén 20 paraméter szerepelt az oszlopokban és összesen hatvan modellt készítettem el ekkor. A sorátlagokat használtam referenciaként, a kapott értékeket pedig később hét részre osztott kereszt-ellenőrzéssel validáltam.

Az 5. ábra a validálási eredményeket szemlélteti egy doboz-bajusz ábra segítségével.



5. ábra: A toxicitási modellek SRD elemzése során kapott eredmények doboz-bajusz ábrája.

Az SRD értékeket az Y tengelyen ábrázoltam a teljesítményparaméterek függvényében. A vízszintes szaggatott vonal a véletlen számok SRD értékeinek eloszlásához tartozó 5 %-os hibahatárt szemlélteti. A függőleges vonalak a páros Wilcoxon próba alapján történő szignifikáns elkülönüléseket jelöli.

Az ábra alapján néhány külső validálási paramétert leszámítva (RMSE_{ext}, Q^2 -F1, Q^2 -F2, Q^2 -F3) a többi paraméter, a véletlen számok SRD eloszlásához tartozó 5 %-os hiba határ alatt van (H_0 hipotézis: az adott paraméter SRD értéke nem különbözik szignifikánsan a véletlen számok SRD értékeitől).

Ugyanezt a kiértékelést elvégeztem a második adatkészletre is, amely eredményeképp nagyon hasonló rangsorhoz jutottam. A vizsgált teljesítmény paraméterek száma ekkor 20 volt, és 70 modellt használtam fel. Bár kissé eltérő sorrendben, de ugyanazon paraméterek kapták a legkisebb SRD értékeket.

Összességében elmondható, hogy a betanító készletre vonatkozó, vagy a kereszt-ellenőrzéssel (vagy a kereszt-ellenőrzéssel egy elem kihagyásos változatával) kapott paraméterek alkották a leginkább reprezentatív és leginkább konzisztens csoportot. A két adatkészletet figyelembe véve a leginkább az RMSE_{cv}, CCC_{cv} és Q^2 _{LOO} és MAE paraméterek ajánlhatók.

V. ÚJ TUDOMÁNYOS EREDMÉNYEK

1. A doktori munkám során a vevő-működtető jelleggörbék (ROC görbék) alapelvét továbbgondolva, többosztályos („ n -class”) ROC görbéket hoztam létre, melyek a különböző módszerek összehasonlítására, értékelésére – beleértve a hibabecslést is – valamint az osztályozó képességük ábrázolására, megjelenítésére is kiválóan alkalmasak. Ezen görbék segítségével energiaitalok FT-NIR spektrumai alapján többféle mintázatfelismerési (osztályozási) eljárást hasonlítottam össze. A véletlen erdő módszerével történő osztályozás modellépítésére kétféle paraméteroptimalási eljárást is kidolgoztam és sikeresen alkalmaztam. Megállapítottam, hogy a vizsgált adatkészleten a fejlesztett fák módszere képes a legjobb modellépítésre. A módszerhez saját fejlesztésű programot használtam, amelynek programkódja a dolgozat mellékletében megtalálható.

2. A Q10 koenzim tartalmú étrendkiegészítők vizsgálata során három regressziós modellt is építettem, melyek segítségével az étrendkiegészítő minták FT-NIR spektrumaiból gyorsan, és kellő pontossággal, környezetkímélő módon meghatározható a Q10 koenzim koncentrációja. A modelleket belső és külső validálásnak is alávettem. A három PLS regressziós modell segítségével kiválthatók az idő- és költségigényes, kevésbé környezetbarát HPLC eljárások is.

3. A Q10 koenzim étrendkiegészítők vizsgálata során, a modellépítésekhez továbbfejlesztettem a szelektivitási arány változókiválasztási módszert, amely az intervallum szelektivitás arány kiszámításán alapul (iSR). Az így kifejlesztett változószelekciós eljárást sikeresen alkalmaztam a modellépítés során.

4. Az energiaitalok vizsgálata során a cukor és koffein tartalom meghatározására FT-NIR spektrumok felhasználásával PLS regressziós modelleket dolgoztam ki. A fejlesztett modellek mindkét komponensre nézve megfelelő pontossággal és robusztussággal rendelkeznek. A modellek mindkét esetben belső és külső validáláson estek át, így a továbbiakban alkalmazhatók az eddigi időigényes, és kevésbé gazdaságos módszerek helyett az energiaitalok mennyiségi meghatározására. Ezen kívül megfelelő osztályozási képességgel rendelkező osztályozási modellt hoztam létre a taurinos, arginines és normál (taurint és arginint nem tartalmazó) minták elkülönítésére. Ez utóbbi modell a minőségellenőrzés, eredetazonosítás vagy hamisítványok kiszűrése során használható.

5. Az antioxidáns kapacitási módszerek vizsgálata és összehasonlítása során megállapítottam, hogy az azonos kémiai háttérrel rendelkező módszerek a PCA és HCA eredmények alapján sokkal hasonlóbb (egységesebb) eredményt mutatnak. Az ACW és ACL módszerek, mindkét adatkészlet vizsgálata során, jelentős eltérést mutatnak a többi módszerhez képest. Statisztikai szempontból leginkább konzisztens módszernek a FRAP és TPC technikákat tartom, amelyek

kiválthatják a többi módszer alkalmazását, amellyel így szűkös időkeret esetén is a legkisebb hibával tudjuk megadni az antioxidáns kapacitás értékeket.

6. A regressziós modellek teljesítményparamétereinek széleskörű összehasonlítása során, jelentős eltérést tapasztaltam a legtöbb külső és belső validálásból származó paraméter között. A belső validáláshoz tartozó paraméterek általában sokkal jobbnak (konzisztensnek) bizonyulnak. Ezek közül is kiemelten jónak tekinthetők az RMSECV, CCC_{cv} és Q²_{LOO} paraméterek. A külső validálási paramétereknek a modellépítések során tapasztalt nagy eltérései viszont új információt hordozhatnak.

VI. KÖVETKEZTETÉSEK ÉS JAVASLATOK

A doktori munkám során számos példán keresztül bemutattam az FT-NIR kutatásokban és a kemometriában rejlő, még máig is gyakran kiaknázatlan lehetőségeket. Az elvégzett kísérleteimmel és elemzéseimmel törekedtem a kemometriai tudományág és az élelmiszertudomány területén minél több fejlesztést és újítást létrehozni.

A Q10 koenzim tartalmú étrendkiegészítők piaca még a mai napig is növekvőben van, hiszen sokáig egyfajta „csodaszerként” kezelték ezt a koenzimet. A manapság kiélezett piaci versenyhelyzetben a különböző formában forgalomba hozott termékek minőségellenőrzése kihangsúlyozandóan, nem csak a fogyasztók, de a gyártók érdeke is. Ebben a szegmensben az általam megalkotott kalibrációs modellek tökéletesen ki tudják váltani a korábban alkalmazott, idő és vegyszerigényes HPLC-s technikát. A modellépítések során kifejlesztett változókiválasztási technikával pedig a kemometria világában új lehetőséget nyújtok a korábban használt változóselektálási lépésekhez képest. A modellépítések e része gyakran kulcsfontosságúnak bizonyul, ahogy ez a Q10 koenzim tartalmú étrendkiegészítők vizsgálatokor is bebizonyosodott számomra. Mivel az iSR változóselektálás megfelelő hatékonyságot mutatott – akárcsak a másik két alkalmazott módszer – így bátran javasolható a későbbi regressziós modellek során történő alkalmazása is.

Az energiatalok területén végzett kutatásaim során szembesültem azzal, hogy az energiatal kereskedelemben betöltött szerepe és fogyasztók nagy száma miatt mekkora jelentőségre tett szert Magyarországon. Ahogy egy évtizeddel ezelőtt hazánkban még éppen csak néhány márka volt forgalomban, úgy az utóbbi években már ez a szám több, mint százra volt tehető. Sajnálatos módon a legnagyobb fogyasztói köre az energiataloknak mindmáig a fiatalok, beleértve a 18 és sokszor 14 év alatti korosztályt is. A rohamos fejlődésnek indult energiatal piacon így szintén fontos – már csak a fogyasztói kört figyelembe véve is – a termékek minőségellenőrzése, különös tekintettel a két legnagyobb egészségügyi kockázattal járó összetevőre: a koffeinre és a cukorra. A kutatásaim során mindkét komponensre nézve sikerült olyan regressziós modelleket építenem, amelyek gyors és környezetbarát módon percek alatt képesek kellően pontos mennyiségi meghatározásra. Ezek a modellek így ismételen kiválthatják a területen gyakorta alkalmazott nagyműszeres analitikai eljárásokat. A vizsgálatok során közel száz energiatalt használtam fel, így a magyarországi piacon található energiatalokra mindkét modell kellőképpen optimálnak tekinthető. Az osztályozás területén pedig sikerült megalkotni egy olyan modellt, amely megfelelő elkülönítést biztosít a taurinos, arginines és normál minták között. Ez a modell az eredetazonosítás és hamisítványok

kiszűrése során juthat nagy szerephez. Az energiatalok spektrumainak felhasználásával pedig kifejlesztettem és teszteltem a többosztályos ROC görbék egy új formációját. Az energiatalok adatait felhasználva rendkívül jó vizuális megjelenítést biztosítva tudtam összehasonlítani többféle osztályozási módszert is. Az ebből származó eredmények következtetéseként érdemes sokszor a még nem túlzottan elterjedt, új kemometriai módszereket is tesztelni, hiszen gyakran a jól bevált és több évtizedes múlttal rendelkező módszerek is csődöt mondhatnak. Esetemben a fejlesztett fák módszere (BT) tűnt ki a többi alkalmazott módszer közül, így ennek a technikának az alkalmazását feltétlenül javaslom osztályozási feladatokra.

Az antioxidáns kapacitás meghatározási technikák világában is számos lehetőséggel találkozhatunk, amelyek közül nehéz eldönteni melyiket érdemes választani a vizsgálatainkhoz. Gyakran ajánlott lehetőség a minél többféle technika használata. A kutatásom során én azt a kérdést tettem fel, hogy melyik módszer tudná visszaadni leginkább a legkisebb hibájú (konszenzusos) értékeket, melyik az, amelyik kiválthatja a többit olyan esetekben, ha nincs elég időnk, vagy túl költségesnek bizonyul több mérés elvégzése. A vizsgálatok során erre a célra legmegfelelőbbnek a FRAP és TPC módszerek bizonyultak. Így e két módszer használata az előbbieken felvázolt esetekben erősen ajánlott, hiszen ezek a módszerek tudták a két vizsgált adatkészlet esetén a konszenzusos eredményt legjobban visszaadni.

A regressziós modellépítések során állandó problémát okoz, a helyes teljesítményparaméterek megfelelő figyelembe vétele, legyen szó FT-NIR spektrumokról vagy gyógyszerkémiail adatokról. A modellnek a különböző teljesítmény paraméterek alapján történő eltérő megítélése miatt nagyon fontos kérdés, hogy melyiket érdemes használni. Az általam elvégzett számítások során bebizonyosodott, hogy bár a különböző külső validálási paraméterek fontos plusz információt adhatnak a modellekről (és bizonyos esetekben elengedhetetlen a külső validálás megléte), a belső validáláshoz tartozó RMSECV vagy éppen CCC_{CV} értékek sokkal inkább konzisztens modelleket válogatnak ki. Az általam elvégzett vizsgálatok így azt tükrözik, hogy a belső validálásra, kereszt-ellenőrzésre és annak eredményeire érdemes nagyobb hangsúlyt fektetni.

AZ ÉRTEKEZÉS TÉMAKÖRÉHEZ KAPCSOLÓDÓ PUBLIKÁCIÓK

Hatástényezővel (IF) rendelkező angol nyelvű folyóiratcikkek

1. **A. RÁCZ**, K. HÉBERGER, M. FODOR (2016): Quantitative determination and classification of energy drinks using near-infrared spectroscopy.
Analytical and Bioanalytical Chemistry, Vol. 408. pp. 6403-6411. (IF=3.125)
2. H. TIMA, **A. RÁCZ**, ZS. GULD, CS. MOHÁCSI-FARKAS, G. KISKÓ (2016): Deoxynivalenol, zearalenone and T-2 in grain based swine feed in Hungary.
Food Additives & Contaminants: Part B. Vol. 7. pp. 275-280. (IF=1.467)
3. F. ANDRIC, D. BAJUSZ, **A. RÁCZ**, S. SEGAN and K. HÉBERGER (2016): Multivariate assessment of lipophilicity scales – computational and reversed phase thin-layer chromatographic indices.
Journal of Pharmaceutical and Biomedical Analysis, Vol. 127. pp. 81-93. (IF=2.867)
4. **A. RÁCZ**, D. BAJUSZ, M. FODOR and K. HÉBERGER (2016): Comparison of classification methods with "n-class" receiver operating characteristic curves: A case study of energy drinks.
Chemometrics and Intelligent Laboratory Systems, Vol. 151. pp. 34-43. (IF=2.321)
5. D. BAJUSZ, **A. RÁCZ** and K. HÉBERGER (2015): Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?
Journal of Cheminformatics, Vol. 7:(1) paper 20. 13 p. (IF=4.547)
Hivatkozások összesen: 36 (MTMT)
6. **A. RÁCZ**, N. PAPP, E. BALOGH, M. FODOR and K. HÉBERGER (2015): Comparison of antioxidant capacity assays with chemometric methods. *Analytical Methods: Advancing Methods and Applications*, Vol. 7. pp. 4216-4224. (IF=1.821)
Hivatkozások összesen: 3 (MTMT)
7. **A. RÁCZ**, A. VASS, K. HÉBERGER and M. FODOR (2015): Quantitative determination of coenzyme Q10 from dietary supplements by FT-NIR spectroscopy and statistical analysis.
Analytical and Bioanalytical Chemistry, Vol. 407:(10). pp. 2887-2898. (IF=3.436)
Hivatkozások összesen: 5 (MTMT)

8. A. RÁCZ, D. BAJUSZ, K. HÉBERGER (2015): Consistency of QSAR models: Correct split of training and test sets, ranking of models and performance parameters.
SAR and QSAR in Environmental Research, Vol. 26:(7-9). pp. 683-700. (IF=1.596)
Hivatkozások összesen: 5 (MTMT)
9. O.H.J. CHRISTIE, A. RÁCZ, J. ELEK, K. HÉBERGER (2014): Classification and unscrambling a class-inside-class situation by object target rotation: Hungarian silver coins of the Árpád Dynasty, ad 997–1301.
Journal of Chemometrics, Vol. 28:(4). pp. 287-292. (IF=1.803)
Hivatkozások összesen: 2 (MTMT)
10. A. RÁCZ, K. HÉBERGER, R. RAJKÓ, J. ELEK (2013): Classification of Hungarian medieval silver coins using X-ray fluorescent spectroscopy and multivariate data analysis.
Heritage Science, Vol. 1:(2) paper 1/1/2. 9 p.
(IF(Chemistry Central Journal)=1.663)
Hivatkozások összesen: 3 (MTMT)
- Nem IF-os folyóirat cikkek:**
11. RÁCZ, A. (2015): Mire jó a kemometria?
Élet és Tudomány, Vol. 70:(4). pp. 118-120.
12. RÁCZ, A., VASS, A., HÉBERGER, K., FODOR, M. (2015): Q10 tartalmú étrendkiegészítők hatóanyagtartalmának mennyiségi meghatározása FT-NIR spektroszkópiával.
Élelmiszer - Tudomány Technológia, Vol. 69:(2). pp. 25-32.

Angol nyelvű könyvfejezetek

1. A. RÁCZ, D. BAJUSZ, K. HÉBERGER:
Chapter title: Cheminformatics/chemometrics in analytical chemistry
In **Cheminformatics – a textbook**
(Editors: Johann Gasteiger, Thomas Engel)
WILEY, 2016 (in press)
2. D. BAJUSZ, A. RÁCZ, K. HÉBERGER:
Chapter 30010, title: Chemical data formats, fingerprints and other molecular descriptions for database analysis and searching
In **Comprehensive Medicinal Chemistry III**.
(Editors: Andy Davis, Colin Edge)
ELSEVIER, 2016 (in press)

3. D. BAJUSZ, A. RÁCZ, K. HÉBERGER:
Chapter title: Which performance parameters are best suited to assess the predictive ability of models?
In **Advances in QSAR modeling with applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences**
(Editor: Kunal Roy)
SPRINGER, 2016 (in press)

Magyar nyelvű konferencia (összefoglaló)

1. RÁCZ, A., FILIP, A., BAJUSZ, D., HÉBERGER, K. (2016): Számításos és vékonyréteg-kromatográfias lipofilcítási indexek összehasonlítása kemometriai módszerekkel.
KeMoMo–QSAR 2016 szimpózium, Szeged, 2016. május
<http://www.chemicro.hu/QSAR/kovetkezo.html>
2. RÁCZ, A., HÉBERGER, K., FODOR, M. (2016): Energiaitalok minőségi és mennyiségi elemzése FT-NIR spektroszkópiával.
Aktualitások a táplálkozástudományi kutatásokban – VI. PhD Konferencia, Budapest, 2016. február
(<http://www.mttt.hu/index.php?content=57>)
3. RÁCZ, A., FODOR, M., HÉBERGER, K. (2015): Energiaitalok egy kemometrikus szemével.
KeMoMo–QSAR 2016 szimpózium, Szeged, 2015. május
<http://www.chemicro.hu/QSAR/20150514.html>
4. RÁCZ, A., VASS, A., HÉBERGER, K., FODOR, M. (2015): Q10 tartalmú étrendkiegészítők minőségellenőrzése FT-NIR módszerrel.
Aktualitások a táplálkozástudományi kutatásokban – V. PhD Konferencia, Budapest, 2015. január. p. 32 (ISBN 978-963-88108-8-5)
5. RÁCZ, A., FODOR, M., HÉBERGER, K. (2014): Változó kiválasztás – avagy legális út regressziós modellek javítására.
KeMoMo–QSAR 2014 szimpózium, Szeged, 2014. május
<http://www.chemicro.hu/QSAR/kivonatok19/kivonat1902.html>
6. RÁCZ, A., PAPP, N., FODOR, M., HÉBERGER, K. (2014): Antioxidáns kapacitás meghatározási technikák összehasonlítása kemometriai módszerek segítségével. Aktualitások a táplálkozástudományi kutatásokban – IV. PhD Konferencia, Budapest, 2014. január. p. 18 (ISBN 978-963-88108-7-8)

7. **RÁCZ, A., ELEK, J., PAPP, G.** (2013): Égetett szeszesitalok tömegspektrometriás és infravörös spektroszkópiás vizsgálatának többváltozós elemzése
KeMoMo–QSAR 2014 szimpózium, Szeged, 2013. április
8. **RÁCZ A., ELEK J., NEMES Z.** (2012): Hogyan segíthet a modern műszeres analitika történelmi kérdések tisztázásában, avagy interdiszciplináris kutatások egy római kori sírkő körül.
XXXV. KEN, Szeged, 2012. október. p. 246 (ISBN: 978-963-315-099-3)
9. **RÁCZ, A., ELEK, J., NEMES, Z.** (2013): Feltárhat-e rejtett összefüggéseket egy római kori sírkőről röntgenfluoreszcenciás elemzés adatok főkomponens-elemzése?
KeMoMo–QSAR 2014 szimpózium, Szeged, 2012. szeptember
10. **RÁCZ, A., ELEK, J., HÉBERGER K., RAJKÓ, R., LENGYEL, A.** (2011): Régészeti leletek XRF vizsgálata és értékelése, avagy mennyi adatra van szükségünk főkomponens elemzéshez
MKE 1. Nemzeti konferencia, Sopron, 2011. május
(<http://www.mkenk2011.mke.org.hu/hu/tudomanynos-program.html>)

Nemzetközi konferencia (összefoglaló)

1. **D. BAJUSZ, A. RÁCZ, K. HÉBERGER** (2015): Revival of an old debate: Cross- vs. External validation in QSAR modeling.
Conferentia Chemometrica 2015, Budapest (Hungary), 2015, September. p. L23 (ISBN: 978-963-7067-31-0)
2. **A. RÁCZ, D. BAJUSZ, K. HÉBERGER** (2015): Large scale statistical comparison of similarity metrics for fingerprint-based calculations.
Conferentia Chemometrica 2015, Budapest (Hungary), 2015, September. p. P04 (ISBN: 978-963-7067-31-0)
3. **A. RÁCZ, D. BAJUSZ, K. HÉBERGER** (2015): *n*-class ROC curves as novel, intuitive tools for method comparison.
Conferentia Chemometrica 2015, Budapest (Hungary), 2015, September. p. P24 (ISBN: 978-963-7067-31-0)
4. **K. HÉBERGER, D. BAJUSZ, A. RÁCZ** (2015): Consistency of QSAR models: correct split of training and test sets, ranking of models and performance parameters. 8th International symposium on computational methods in toxicology and pharmacology integrating internet resources, Chios (Greece), 2015, June. p. 26

5. J. ELEK, A. MEZŐSI, **A. RÁCZ** (2013): Estimation of active ingredient content by multivariate calibration: NIR examination of tablets used in ED treatment. Conferentia Chemometrica 2013, Sopron (Hungary), 2013, September. p. L26 (ISBN: 978-963-9970-38-0)
6. **A. RÁCZ**, J. ELEK, G. PAPP (2013): Multivariate data analysis of Hungarian spirit drinks' IR spectroscopic data. Conferentia Chemometrica 2013, Sopron (Hungary), 2013, September. p. P22 (ISBN: 978-963-9970-38-0)
7. O.H.J. CHRISTIE, **A. RÁCZ**, J. ELEK, K. HÉBERGER (2012): Object target rotation for principal components analysis: metal content data of hungarian silver coins from the Árpád dynasty. XIII. Chemometrics in Analytical Chemistry, Budapest (Hungary), 2012, June. p. 72 (ISBN: 978-963-9970-24-3)
8. **A. RÁCZ**, J. ELEK, K. HÉBERGER, R. RAJKÓ, A. LENGYEL (2012): Principal component analysis of an XI-XV century silver coins' XRF dataset and verification by additional multilinear methods. WSC 8, Drakino, Russia, 2012, February. (<http://wsc.chemometrics.ru/wsc8/presentations/?page=2>)
9. **A. RÁCZ**, J. ELEK, K. HÉBERGER, R. RAJKÓ, A. LENGYEL (2011): Classification of medieval silver coins using X-ray fluorescent spectroscopy and multivariate data analysis. Conferentia Chemometrica 2011, Sümeg (Hungary), 2011, September. p. P33 (ISBN: 978-963-9970-15-1)
10. **A. RÁCZ**, J. ELEK, K. HÉBERGER, R. RAJKÓ, A. LENGYEL (2011): Principal Component Analysis of an XI-XV century silver coins' XRF dataset and verification by additional multilinear methods. Conferentia Chemometrica 2011, Sümeg (Hungary), 2011, September. p. L26 (ISBN: 978-963-9970-15-1)