

**Kvantitatív szerkezet – hatás
összefüggések keresése új
kemometriai módszerekkel**

Témavezető: **Héberger Károly**,
tudományos tanácsadó
OTKA témaszám: T037684

MTA Kémiai Kutatóközpont, 2006

Tartalomjegyzék*

A résztvevők tervezettől való eltérésének okai	3
Az elvégzett munka megfelel az eredeti célkitűzéseknek	3
Zárójelentés témák szerinti bontásban	3
HPLC oszlopok, kromatográfiai rendszerek osztályozása	3
Gázkromatográfiai mennyiségi szerkezet-retenció összefüggések (QSRR) keresése	5
Új változóselektálási módszerek kidolgozása	7
Egyéb, környezetvédelmi, gyakorlati jellegű szerkezet – hatás vizsgálatok	8
Az eredeti szerződésbeli programhoz fűzött megjegyzések	9
Az eredmények gazdasági hasznosíthatósága	12
A kutatási tevékenység rövid összefoglalása (magyarul)	12
A kutatási tevékenység rövid összefoglalása (angolul)	13
Melléklet:	
Különlevonatlista az adott időszakra (2002-2004)	13

* Mivel az egyes mezőkbe többszöri próbálkozás után sem sikerült szöveget beírni vagy bemásolni, kénytelen voltam az egyes kérdésekre adott válaszokat egyetlen szöveges dokumentumba integrálni.

A résztvevők tervezettől való eltéréseinek okai:

Az eredeti szerződésben két doktoráns (Farkas Orsolya és Stadler Krisztián) szerepelt. Már az első éves részjelentésben megírtam, hogy Stadler Krisztián helyett Vanyúr Rozáliát vontuk be a munkába. Sajnos azonban ő sem maradt véglegesnek. Ma már egyikük sem dolgozik az intézetben. Stadler Krisztián Amerikába távozott, Vanyúr Rozália 2003 augusztusától nem dolgozik a csoportunkban, később az intézettől is megvált. Ezért Forlay-Frick Pétert, aki akkor a Brüsszeli Egyetem munkatársa volt, kértem fel a számítások elvégzésére, aminek nagy örömmel tett eleget.

Megjegyzem meg, hogy a 4 éves periódus alatt 7 PhD-s fordult meg a csoportunkban (és 3-an készítették diplomamunkát). Az első évben munkát kezdő négy közül ma már csak egy dolgozik nálunk, egy szülési szabadságon van.

Az elvégzett munka megfelel az eredeti célkitűzéseknek.

Megjegyzendő, hogy a négy év folyamán jelentős előrehaladást értünk el és ezért jelentősen kibővítettük az elérendő célokat, többek között azzal a háttérfeltevással, hogy ezért majd könnyebben kapunk a pályázat folytatására pénzt. Talán nem szükségtelen megemlíteni, hogy várakozásainkban keserűen csalódtunk.

Hangsúlyozni szeretném azonban, hogy nem arról van szó, hogy minden lehető és lehetetlen problémával elkezdtünk foglalkozni, és minden cikkre, amire megtehettem ráírtam a nevemet. Ez könnyen ellenőrizhető a különlevonat listám és az OTKA jelentés benyújtott cikkei listájának összehasonlításával (ld. melléklet). Mindig szem előtt tartottuk az eredeti OTKA célkitűzést: a mennyiségi szerkezet - hatás összefüggések keresését új kemometriai módszerekkel.

Megjegyzendő még, hogy az eredeti szerződés 3 évre szólt, melyet egy évvel külföldi tartózkodás miatt meg kellett hosszabbítani. A hosszabbításra engedélyt kaptunk.

Az összes munkatervben szereplő elemzést elvégeztük, sőt cikket is írtunk a tervezett témákból, (néhányikból többet is). Kétségtelen, hogy nem mindegyik jelent meg.

Zárójelentés témák szerinti bontásban:

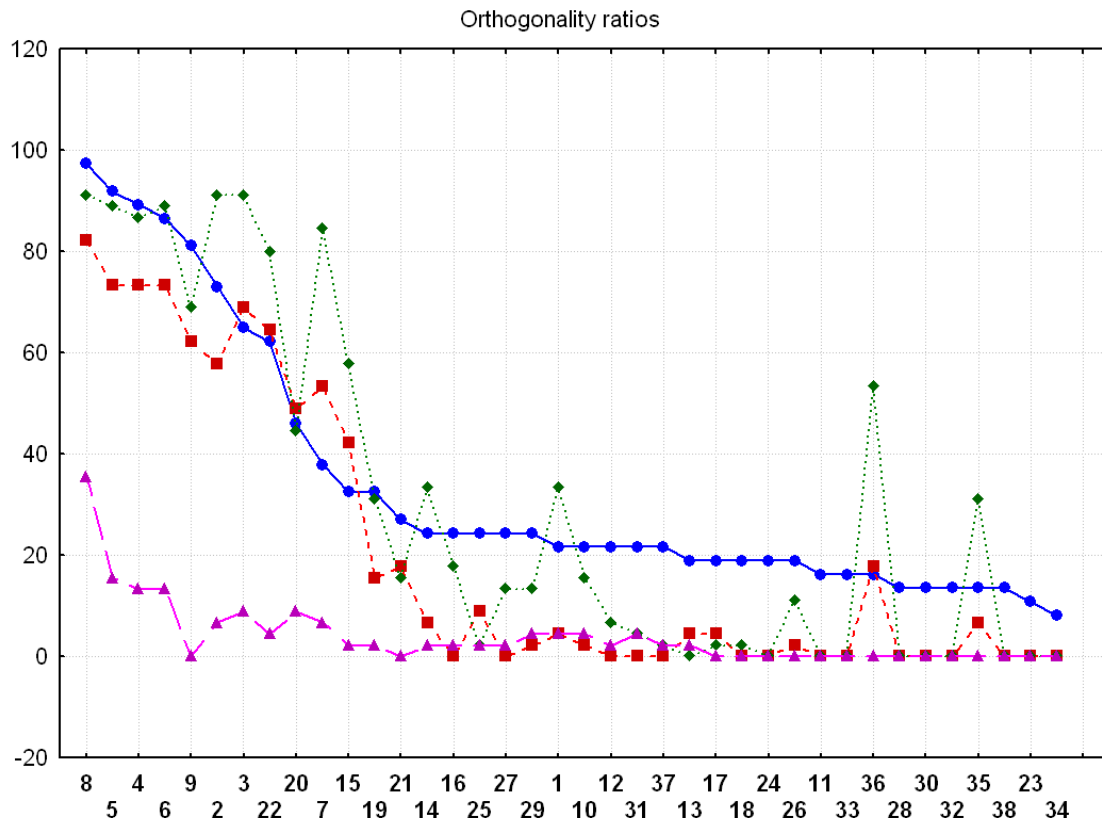
A mondandót témák szerint, nem az elért eredmények tudományos jelentősége vagy pedig időrend szerinti bontásban csoportosítottuk.

HPLC oszlopok, kromatográfiás rendszerek osztályozása:

- 1) Kemometriai módszerek alkalmazásával folyadék kromatográfiás oszlopok is tesztelhetők. Az adott célnak megfelelő legjobb oszlop, mobil fázis, hőmérséklet, pH és más körülmények kiválaszthatók. Három modell vegyületet választottunk a kromatográfiás oszlopok és rendszerek jellemzésére: benzooesavat, N,N-dimetil-anilint és vancomycint. Az elméleti tényezőszám és a szimmetria faktor segítségével végeztük a rendszerek közötti különbségtételt. Főkomponens elemzés mutatta meg, mely oszlopok szolgáltatnak hasonló, közel hasonló, illetve nagyon eltérő elválasztást. A PCA eredményeit láthatóvá tevő ábrák segítenek a kromatográfiás rendszerek helyettesíthetőségének eldöntésében.

Melléklet: 96./ Péter Forlay-Frick, Jenő Fekete and Károly Héberger*, Classification and replacement test of HPLC systems using principal component analysis, *Anal. Chimica Acta*, **536** /1-2/ 71-81 (2005) if(2004)=2.588

2) A HPLC-s rendszerek osztályozása során, az elért eredményeknek köszönhetően, kapcsolódtunk egy nagyobb kutatócsoport tevékenységéhez (vezető professzor Yvan Vander Heyden, Brüsszel). Javasoltuk, hogy az általuk használt, korrelációs koefficiens kiszámításán alapuló módszert más módszerekkel is egészítsük ki és hasonlítsuk össze a módszereket. A többi módszer között a párkorrelációs módszer nagyon jó eredményeket szolgáltatott. Az összevetést és a levonható következtetéseket összefoglaló (review) cikkben publikáltuk. Összehasonlítottuk a korrelációs koefficiensen a Spearman ρ -n, Kendall τ -n alapuló nem-paraméteres módszereket és a párkorrelációs módszert háromféle nem-paraméteres és egy paraméteres szelekciós kritériummal. A feladat minél eltérőbb tulajdonságú kromatográfiás (HPLC) rendszerek kiválogatása volt, azzal a céllal, hogy a gyógyszervegyületeknek, metabolizmusaiknak, és szennyezőinek elválasztását minél kevesebb munkával és minél kevesebb kromatográfiás rendszer használatával tudjuk megoldani. A módszerek összehasonlítására új mértékegységet, az ún. ortogonalitási arányt definiáltuk; az x tengelyen a kromatográfiás rendszerek sorszámát, az y tengelyen az ortogonalitási arányokat ábrázoltuk, különböző színnel, eltérő szimbólummal jelölve az egyes módszereket (ábra)



A módszereket méltányos módon hasonlítottuk össze és a „legortogonálisabb” rendszereket kiválogattuk, megadtuk.

Melléklet: 99./ Péter Forlay-Frick, Elke Van Gyseghem, Károly Héberger* and Yvan Vander Heyden*, Selection of orthogonal chromatographic systems based on parametric and non-parametric statistical tests (review), *Anal. Chimica Acta*, **539** /1-2/ 1-10 (2005) if(2004)=2.588

Gázkromatográfiai mennyiségi szerkezet-retenció összefüggések (QSRR) keresése

- 3) E témakörben is végeztünk módszerek, modellek összehasonlítására vonatkozó vizsgálatokat. Egyik esetben modell vegyületekként az alifás alkoholokat választottuk, mely vegyületcsoport retenciós indexét mások is előre jelezték már. Az irodalomban lévő vita, bár sok új meglátással árnyalható, az utóbbi évek fejleményeinek következtében, máig sem jutott nyugvópontra. Leegyszerűsítve, a vitát úgy lehet összefoglalni, hogy a matematikusok szerint a peremregresszió nyújtja a legjobb megoldást előrejelzésre, extrapolálásra; a gyakorlati szakemberek pedig gyakran demonstrálják, hogy a parciális legkisebb négyzetek módszere (PLS) jobb előrejelzéseket szolgáltat. A módszerek összehasonlításának azonban komoly akadályai vannak. Ha mindegyik változóselektálási módszert a lehető legjobb előrejelzési képessége alapján hasonlítanánk össze, akkor különböző szabadsági fokú, különböző számú deskriptort tartalmazó modelleket kellene összevetni, ami nem méltányos és nem is tisztességes. De ez fordítva is igaz. Ha az azonos szabadsági fokú modelleket hasonlítjuk össze, akkor azok nem a legoptimálisabb előrejelzési képességüket mutatják. Innen származott az ötlet, hogy azonos modellépítési módszert használjunk (MLR) azonos (közel optimális) szabadsági fokkal, de teszteljük az egyes módszereket a változóselektálási képesség szempontjából. Ez annál is inkább igazolható, mert az egyes modellezési módszereket önállóan is használják változóselektálásra. Az így szelektált változókat másféle modellezési módszer bemeneteként is használják.

Vizsgálatunkban a következő módszereket hasonlítottuk össze a változóselektálási képesség szempontjából. Peremregresszió (RR), PLS, párkorrelációs módszer (PCM), előrele történő változóbevonás (FS) és legjobb alrendszer szelekció (BSS). Sem a peremregresszió, sem a PLS nem képes a Kováts indexek előrejelzését megfelelően megoldani. A PCM reális, értelmezhető eredményeket ad, de nem alkalmas predikcióra, míg jó előrejelzési képességgel az FS és a BSS módszer rendelkezik. Kiválogattuk a legfontosabb változókat is: ezek az elektrotopológiai állapot index, a molekulatömeg és a molekula méretét és alakját jellemző WHIM indexek. Az előrejelzésre alkalmas modelleket ezúttal is kereszt-ellenőrzéssel igazoltuk.

Melléklet: 95./ Orsolya Farkas*, Károly Héberger, Comparison of Ridge Regression, Partial Least Squares, Pair-wise Correlation, Forward- and Best Subset Selection Methods for Prediction of Retention Indices for Aliphatic Alcohols, *J. Chem. Inf. Modeling*, **45** /2/ 339 -346, (2005) if(2004)=2.810

- 4) A mennyiségi szerkezet - retenció összefüggések témájában telített oxigén-, nitrogén- és kénatomokat tartalmazó heterociklusos vegyületek retenciós indexeit jeleztük előre, a vegyületek molekulászerkezetéből számított deskriptorokkal. A retenciós

indexeket Igor G. Zenkevich, a Szentpétervári Egyetem professzora szolgáltatta. Jó néhány változószelektálásra szolgáló módszert is összehasonlítottunk a predikcióra alkalmas modellek építése során (pl. MLR, PLS), és megállapítottuk, hogy az általános hiedelmekkel ellentétben a PLS nem képes a legjobb, legszignifikánsabb változókat kiválogatni. Az elméletből származó változók mellett a forráspontot is használtuk a modellezésben. E változó szerepe nem hanyagolható el. A forráspont lényegesen jobb deskriptor a retenciós indexek előrejelzésére, mint bármelyik molekulaszervezetből számított; sőt, ha azok közül a legjobbakat együttesen vizsgáljuk, még akkor is. Az előrejelzésre alkalmas modelleket kereszt-ellenőrzéssel igazoltuk.

Melléklet: 88./ Orsolya Farkas*, Károly Héberger, and Igor G. Zenkevich, Quantitative Structure – Retention Relationships XIV. Prediction of Gas Chromatographic Retention Indices for Saturated O-, N- and S-heterocyclic Compounds, *Chemometrics Intell. Lab. Syst.* **72** /2/ 173-184 (2004) if(2004)=1.899

- 5) A mennyiségi szerkezet-retenciós összefüggések keresése során Kováts indexek elemzésekor figyeltünk fel arra a tényre, hogy az indexek hőmérsékletfüggése nem lineáris, hanem görbült, rendszerint minimumos összefüggéssel adható meg, de találtunk maximumra is példát. Tulajdonképpen érthetetlen, hogy a tudományos közönség miért ragaszkodott az Antoine egyenlet szerint (hiperbola típusú) leíráshoz, mely függvénynek szakadása van jól definiált helyen. A maximummal és minimummal ellentétben a szakadást az empirikus megfigyelések nem igazolták vissza. Egy paraméterekben lineáris összefüggést javasoltunk a Kováts indexek hőmérsékletfüggések leírására:

$$I = A + B / T_c + C \ln T_c$$

ahol T_c az izoterm oszlophőmérséklet és A, B, C empirikus konstansok. Bár az egyenlet fölénye a korábbi Antoine egyenlettel szemben laikus számára is nyilvánvaló, a konstansok elméleti magyarázata nem volt megoldott. Erre tettünk kísérletet az *Acta Chromatographica*-ban megjelent cikkünkkel (a folyóirat szerkesztő bizottságának tagjaként nem volt lehetőség az egyébként megtisztelő felkérés elhárítására). Ha a cikkben foglalt magyarázatot elfogadjuk, további termodinamikai mennyiségek számolhatók: az oldás aktiválási energiája és az egy metilén egységre jutó kémiai potenciál.

Melléklet: 84./ Teresa Kowalska, Károly Héberger and Miklós Görgényi: Temperature Dependence of Kováts Indices in Gas Chromatography. Explanation for Empirical Constants using Transition State Theory, *Acta Chromatographica*, No. **13**, 60-68 (2003) if(2004)=0.698

- 6) Mennyiségi szerkezet retenció összefüggéseket kerestünk poláros vegyületek ciano és nitro alkánok példáján is. A fent említett logaritmikus egyenlet, a konstansok elméleti magyarázata lehetővé teszi oldási aktiválási energia függések meghatározását a szénatomszám függvényében. Három különböző filmvastagságú oszlopon végzett mérések adatai csak csekély eltérést mutattak és bizonyítják az egyenlet általános érvényét.

Melléklet: 90./ Károly Héberger*, Miklós Görgényi, and Teresa Kowalska: Effect of film Thickness on the retention index of nitro and cyano compounds. Interpretation of the logarithmic equation, *J. Chromatogr. Sci.* **42** /6/ 288-292 (2004) if(2004)=1.166

7) Folytattuk a gázkromatográfiás adatokból a termodinamikai mennyiségek számítását is. Normál alkánokat és a McReynolds tesztvegyületek retenciós adatait dolgoztuk fel és oldási hőkapacitásokat, azok hőmérsékletfüggését számoltuk ki. A modell egy egyszerű elméleti modellel, kvantumkémiailag számítások segítségével (harmonikus oszcillátor, rigid rotor feltételezésével) értelmezhetővé vált. Az oldási hőkapacitás a vizsgált hőmérséklet tartományban csaknem folytonos csökkenő görbével adható meg.

Melléklet: 101./ Zoltán A. Fekete, Károly Héberger*, Zoltán Király, Miklós Görgényi, Temperature dependence of Solvation Heat Capacities by Gas Chromatography, *Anal. Chim Acta*, **549** /1-2/ 134-139 (2005) if(2004)=2.588

8-9) Inverz gázkromatográfiában a polimerek és oldószerek illetve tömőanyagok és oldószerek közötti kölcsönhatásokat jellemeztük főkomponens elemzés segítségével. A tesztvegyületek száma – hasonlóságuk alapján – csökkenthető volt, s a különböző szerkezetű polimerek és töltelékanyagok, módosítók csoportosíthatóvá váltak. Ezen vizsgálatok során figyeltünk fel arra a tényre, hogy a Flory-Huggins paraméterek, amelyek az összetevők közötti kölcsönhatásokat jellemzik, függenek a tesztvegyületek minőségétől, bár definícióik alapján nem kellene, hogy fűgjenek. A Zhao-Choi eljárással a paraméterek oldószer függése eliminálható. A főkomponens elemzés segítségével a Flory-Huggins paraméterek megfelelő meghatározására szolgáló tesztvegyületek száma jelentősen csökkenthető.

Melléklet: 98./ Károly Héberger, Kasylda Milczewska and Adam Voelkel*: Principal Component Analysis of Polymer-filler Interactions in Inverse Gas chromatography, *Colloids and Surfaces, A: Physiochem.* **290** /-/ 29-37 (2005) if(2004)=1.513

Melléklet: 104./ Adam Voelkel, Kasylda Milczewska, Károly Héberger*, Inverse Gas Chromatographic and Chemometric Tools for Determination of Interactions between the Components of Polymeric Composition, *Analytica Chimica Acta*, (accepted)

Új változószelektálási módszerek kidolgozása

10) Régóta foglalkozunk új változószelektálási módszerek kidolgozásával. A Szegedi Egyetem Főiskolai Karával, Rajkó Róberttel való együttműködésben sikeresen általánosítottuk a pár-korrelációs módszert több változóra. Gyakorlati példák (1) a gyökaddíciós reakciók sebességi állandóját befolyásoló faktorok kiválasztása, (2) a gázkromatográfiás oszlopok polaritását megszabó tényezők közötti választás) sikerült bebizonyítani az általánosított pár-korrelációs módszer (generalized pair-correlation method – GPCM) előnyeit a többi módszerrel szemben. Nyolc különböző elven alapuló (és eltérő szelekciós kritériumot használó) módszert hasonlítottunk össze. Az összehasonlítás során kiderült, hogy a polaritást legjobban magyarázó faktort a népszerű változószelektálási módszerek (PLS, genetikus algoritmus, mesterséges

neuronhálók), nem képesek kiválasztani a legjobb változók közé, míg a GPCM (és más népszerű (korrelációs koefficiensen alapuló) és kevésbé népszerű módszerek (Höskuldsson lényegi változós módszere, principal variable method: PVM)) igen. A felfedezésnek messzemenő következményei lettek. E módszer előnyeit felismerve javasoltunk együttműködést az NSF-fel, együttműködő partner John H. Kalivas, Pocatello, IH, USA. Az együttműködést támogatták az MTA-OTKA-NSF projekt keretében. Jelenleg is sikeresen folytatódik.

Melléklet: 79./ Károly Héberger* and Róbert Rajkó, Generalization of Pair-Correlation Method (PCM) for Nonparametric Variable Selection, *J. Chemometrics*, **16** /8-10/ 436-443 (2002) if(2002)=1.376

11) A párkorrelációs módszer továbbfejlesztésén is dolgoztunk. Bár ismert, hogy nagyon jó, magyarázó változókat választ ki, tulajdonságait osztályozási feladatokban még senki nem vizsgálta. Ha a párkorrelációs módszert szukcesszíven, bizonyos szabályok alapján végezzük (az egyenértékűnek tűnő változók közül a kevésbé valószínűt tartjuk meg modellben), akkor a PCM is alkalmas osztályozási feladatokra, sőt, összehasonlítása a Prokrusztész rotációval arra utal, hogy a csoportokat és kiugró értékeket jobban képes elkülöníteni, ill. kiemelni.

Melléklet: 87./ Károly Héberger*, José M. Andrade, Procrustes Rotation and Pair-Wise Correlation: a Parametric and a Non-parametric Method for Variable Selection, *Croatica Chemica Acta* **77** /1-2/ 117-125 (2004) if(2004)=0.924

Egyéb, környezetvédelmi, gyakorlati jellegű szerkezet – hatás vizsgálatok

12) Az utóbbi időben egyre erősebb az igény, hogy környezetvédelmi adatok elemzésével foglalkozzunk. Az óriási tömegben mért és (elvben) rendelkezésre álló adatok többváltozós elemzése nem vagy alig történt meg hazánkban; óriási potenciál kihasználatlan. Ezért fogadtuk el a Miskolci Egyetemen dolgozó kollégák ajánlatát a Sajó vízének tisztaságára jellemző adatok elemzésére és a miskolci levegőben az ózon koncentráció előrejelzésére vonatkozóan. Bár mindkét témát sikeresen befejeztük, sőt cikkírássra is sor került, a Sajó-víz adatok elemzését összefoglaló cikk ezidáig még nem jelent meg, számomra nem érthető okokból. Az ózon koncentráció előrejelzésével foglalkozó cikk azonban megjelent, a főkomponens elemzés segítségével a napszakok jól elkülöníthetők, az ózon koncentrációt befolyásoló faktorok (szélsebesség és irány, napsugárzás, nedvességtartalom, NO, NO₂, CO, SO₂ koncentrációk és hőmérséklet) csoportosíthatók. Az ózon koncentráció előrejelzését végül is főkomponens regresszióval (PCR) oldottuk meg, de összehasonlítottuk három módszerrel (PCR, PLS, MLR), melyek a szignifikáns változók.

91./ Attila Lengyel*, Károly Héberger, László Paksy, Olivér Bánhidi, Róbert Rajkó Prediction of Ozone Concentration in Ambient Air using Multivariate Methods, *Chemosphere* **57** /8/ 889-896 (2004) if(2004)=2.359

13) Nem akartuk elhárítani a megtisztelő felkérést a *Molecules* című folyóirat különszámába való cikk írására. Mennyiségi szerkezet - antioxidáns hatás

összefüggést állítottunk fel flavon alapvázú vegyületek példáján. Bár a flavonok irodalma óriási (a források száma jóval 10 000 felett van) és gyökfogó-antioxidáns hatásuk is ismert, szerkezetükből az antioxidáns hatás megadására szinte még nem is történt próbálkozás. Az irodalomból származó adatokra a parciális legkisebb négyzetek módszerével (PLS) illesztettünk előrebecslésre alkalmas modellt. A flavonok csoportosítása (X score-ok) ábrázolásával történt. Farkas Orsolya a vegyület 2D szerkezetének reprezentációját és a leíró deszkriptorok számítását is önállóan végezte.

Melléklet: 93./ Orsolya Farkas, Judit Jakus, Károly Héberger, Quantitative structure – antioxidant activity relationships of flavonoid compounds, *Molecules*, **9** /12/ 1079-1088 (2004) if(2004)=0.676

14) Szabad aminosavak és biogén aminok elemzését végeztük el főkomponens-elemzéssel és lineáris diszkriminancia analízissel borok eredetvizsgálata (technológia, bortermelő hely, szőlőfajta és termelési év) szempontjából. Meghatároztuk, mely változóknak van a legnagyobb megkülönböztető ereje, hatása. Az egyes fehér, illetve vörös borok megkülönböztethetősége, termőhelye, szőlőfajta és évjárat tekintetében 62-68 % között mozog, ami kereskedelmi (nem autentikus) minták esetén nem lebecsülhető eredmény.

Melléklet: 86./ Károly Héberger, Elemér Csomós and Livia Simon-Sarkadi, Principal Component and Linear Discriminant Analysis of Free Amino Acids and Biogenic Amines in Hungarian Wines *J. Agric. Food Chem.* **51** /27/ 8055-8060 (2003) if(2003)=2.102

15) A nagyszerű ACE & CC 2003 (Advances in Chromatography and Electrophoresis and Conferentia Chemometrica) szervezésének folyamánként a konferenciakötetben (mely a *Chemometrics Intelligent Laboratory Systems* folyóirat 72/2 számaként jelent meg, vendégszerkesztő Héberger Károly, if (2003) = 2,229) összefoglaltuk a magyarországi kemometriai kutatások helyzetét az utolsó 10 év angol nyelven megjelent és idézett folyóiratközleményei alapján. Az összefoglalás a legjelentősebb felfedezések felsorolásával kezdődik, majd történelmi perspektívából elemzi az országban folyó kemometriai kutatásokat. Kitér a kemometriával foglalkozó csoportok, tudományos iskolák szerinti csoportosításra, végül a sikeres alkalmazásokat (szoftvereket is) ismerteti. Az összefoglaló közlemény írása során konzultáltam a kutató csoportok vezetőivel, egyes tagjaival, hogy lehetőleg ne szubjektív, inkább általános, „testületi” lényegű véleményt tükrözzön.

Melléklet: 89./ Károly Héberger, Chemometrics in Hungary (the last 10 years) [review], *Chemometrics Intell. Lab. Syst.* **72** /2/ 115-122 (2004) if(2004)=1.899

Az eredeti szerződésbeli programhoz fűzött megjegyzések:

2002. Környezetvédelmi adatok többváltozós elemzése a Miskolci Egyetemmel való együttműködésben. Cél új összefüggések feltárása, az antropogén szennyezési profil

megkülönböztetése a természetes folyamatoktól. (Sajó folyó víztisztasági adatainak elemzése).

Az együttműködés megvalósult, több adatkészleten jónéhány (>5) főkomponens-elemzést végeztünk. Az antropogén szennyezési profil csakugyan megkülönböztethető a természetes folyamatoktól. A cikket megírtuk, jelenleg a társszerzőknél pihen.

2002. Környezetvédelmi adatok többváltozós elemzése a Miskolci Egyetemmel való együttműködésben. Cél: az ózonkoncentrációt befolyásoló faktorok felderítése és főkomponens-regresszió segítségével az ózonkoncentráció előrejelzése, modellezése. (Miskolc levegőminőségi adatainak elemzése).

A kitűzött célt maradéktalanul sikerült megvalósítani, a részletek a 12) pontban a 9. oldalon találhatóak.

2003 Kvantitatív szerkezet – hatás összefüggések keresése gyökreakciókban. A metil gyök vinil típusú alkénekre történő reakciója sebességi állandóinak előrejelzése parciális legkisebb négyzetek módszerének segítségével, a fontos faktorok megállapítása, a modell validálása.

Farkas Orsolya optimalta a reakciópartnerek molekulaszervezetét, kiszámította a leíró változókat, és PLS modelleket építettünk a sebességi állandók előrejelzésére. Megállapítottuk, hogy a poláris tulajdonságok és a reakcióhő befolyásolja leginkább a sebességi állandót. Véletlen számok segítségével validáltuk a modelleket és az eredményeket bemutattuk a „Free Radical Reactions” Gordon konferencián. Mivel az érdeklődés súlypontja átkerült a biológiai gyökök tulajdonságainak felderítésére, a publikálásról nem erőltettünk. A modelleket és a reakciókat befolyásoló faktorokat példaként beletettük a változószelektálási módszerünk továbbfejlesztéséről szóló cikkbe ld. még 10) pont a 8. oldalon.

2003 A belga és litván együttműködésben mért levegőszennyezők (BTEX) főkomponens-elemzése.

Az adatok főkomponens-elemzését többféle adatelőkezelési eljárást használva végeztük el. Az adatok interpretálása azonban annyira ellentmondásos volt, hogy nem volt lehetőség a társszerzők nézeteinek összehangolására, bár egy kézirat elkészült.

2003 Az utolsó tíz évben publikált magyar kemometriai tárgyú munkák összeszedése és összefoglaló (review) cikkben való ismertetése.

Megtörtént, lásd 15) pont a 10. oldalon.

2004 Az új típusú nemparaméteres változószelektálási eljárásunk, a Pár-korrelációs Módszer (PCM) több változóra való általánosítását teszteljük és fejlesztjük tovább környezetvédelmi és kvantitatív szerkezet – hatás (toxicitás) adatokon.

Megtörtént, részletek találhatóak a 3. és 4. pontban (a 6. ill. a 7. oldalon), De toxicitási adatok előrejelzésével is foglalkoztunk, Dubrovnikban posztert mutattunk be a témában, és jelenleg úgy néz ki, hogy cikkírára is sor fog kerülni.

2004 A parciális sorbarendezési (Partial order ranking) algoritmust és a PCM-et egyesíteni szeretnénk előrejelzési és változószelektálási céllal.

Az egyesítést megcsináltuk, Hasse diagrammokat elemeztünk, de előrejelzésre annyira használhatatlan modelleket kaptunk, hogy a további vizsgálatoktól eltekintettünk, noha az elv sokat ígérő lenne.

Egyéb elemzések: A Poznani Egyetemmel való együttműködésben polaritási jellemzők és gázkromatográfiás retenciós adatok elemzését, a Szegedi Egyetemmel való együttműködésben Tisza vízminőségi és egyéb környezetvédelmi adatok elemzését, a Veszprémi Egyetemmel való együttműködésben Balaton és csapadékvizek nehézfém tartalmának elemzését vettük tervbe.

A Poznani Egyetemmel való együttműködésből két cikk is született lásd 8-9) pontok a 8. oldalon. A Szegedi Egyetemmel nem sikerült az együttműködést tető alá hozni; többszöri ígéret ellenére sem kaptam adatokat. A Balaton vizének elemzése Hlavay professzor úr sajnálatos halála miatt hiúsult meg.

De dolgozom termodinamikai adatok gázkromatográfiás adatokból való meghatározásán, a retenciós adatok nemlineáris illesztésén is. Több, régóta elkezdett, de nem lezárt témával is szeretnék foglalkozni, mint pl. bronztárgyak, kőeszközök neutronaktivációs analízissel kapott adatainak feldolgozása, borelemzések, termodinamikai adatok előrejelzése a molekulaszervezetből, stb.

A termodinamikai adatok számításával több cikkben is foglalkoztunk, lásd 5-7) pontok a 6. és 7. oldalon. A bronztárgyak elemzése is elkészült.

Mindegyik bekezdésben említett témából tudományos cikk várható.

Bár az eredeti dokumentumban 8 bekezdés szerepelt a 15 elkészült és megjelent cikkkel (plusz meg néhány előkészületben van) nem vallottunk szegyet, sőt!

A Pár korrelációs módszer megvalósításából szoftver csak kedvező esetben készülhet. Bár ismételt kísérletet tettem a párkorrelációs módszer felhasználóbarát szoftverének elkészíttetésére, (Rajkó Róbert, Büki András) a felajánlható összeg csekély volta miatt nem lehetett megvalósítani.

Az eredmények gazdasági hasznosíthatósága

Az ortogonális kromatográfiás rendszerek kiválogatására, összehasonlítására szolgáló módszereket minden gyógyszercégnél – ahol gyógyszervegyületek, szennyezők, metabolitok elválasztásával foglalkoznak – használni lehetne és nagymértékben csökkenthető lenne az analízisek kidolgozására fordított idő (az oszlopok, körülmények száma, stb.). Nemcsak költségcsökkentést lehetne elérni azonban, hanem növelni, lehetne az azonosítás biztonságát; valószínűsíthetjük, hogy nem maradnának átfedő csúcsok a kromatogramban. Bár a módszert csak gyógyszervegyületek elválasztásában teszteltük, nyilvánvaló, hogy bármilyen más elválasztási feladat optimálására is alkalmas. Sajnos azonban kevés lehetőséget látok a módszer alkalmazására. Ha jobb eredményeket szolgáltat, akkor az alkalmazók azt kritikaként élik meg saját korábbi munkájuk, nem optimált tevékenységük kritikájaként és mindent elkövetnének azért, hogy a módszer jó tulajdonságai ne derülhessenek ki, ne manifesztálódjanak. Ha esetleg egy különleges

elválasztási feladat esetén nem adna jobb eredményt, akkor meg miért használnák? (A módszer alkalmazhatóságáról alkotott ítélet meglévő adatokon alapszik, hasonló, de más feladatokra valószínűleg beválik, ám mint minden valószínűségi megállapítás, ez sem abszolút érvényű.)

Borok eredetvizsgálatát autentikus minták segítségével, rutin kemometriai módszerekkel meg tudnánk oldani. Ez segítene a borhamisítás elleni küzdelemben, ami állami, esetleg fogyasztóvédelmi feladat. Az államnak, fogyasztóvédelemnek ilyesmire se pénze, se energiája nincsen. Némi érdeklődést és fizetőképes keresletet is tapasztaltunk a feltehető borhamisítók részéről, nekik azonban nem kívánjuk tudásunkat értékesíteni.

Az inverz gázkromatográfias adatok elemzésével polimereket, fogászati tömőanyagokat értékelni, bevizsgálni, osztályozni, minőségüket ellenőrizni tudnánk, viszonylag egyszerű eszközökkel (gázkromatográf, számítógép, adatelemzési szoftverek). Egyelőre a polimergyártók nem ismerik, nem értik, mire jó, hogyan használható az az osztályozási, csoportosítási eljárás, ami lengyel kollégákkal közösen dolgoztunk ki.

Több QSRR modellünk jól használható alkoholok, szerves O, N, és S atomokat tartalmazó, heterociklusos vegyületek azonosítására, retenciós index könyvtárak kiegészítésére.

Megjegyzendőnek tartom, hogy bár az alap kutatási eredmények alkalmazását nagyon is kívánatosnak tartom, az alkalmazáshoz más (pl. menedzseri) tulajdonságokra van szükség, mint a felfedezések megtételére, a módszerek kidolgozására. Magyarországon hiányzik az összekötő kapocs, senki nincs, aki tőkét kockáztatna olyasmire, amit meg sem ért. Tevékenységünk reklámozására pedig mi nem vagyunk felkészülve.

A kutatási tevékenység rövid összefoglalása (magyarul)

Kemometriai módszerek alkalmazásával folyadékkromatográfias oszlopokat teszteltünk minél eltérőbb tulajdonságú kromatográfias (HPLC) rendszerek kiválogatása céljából, összehasonlítottuk az eddigi módszereket, és új mértékegységet definiáltunk, az ún. ortogonalitási arányt. QSRR modelleket építettünk változószelektálási és előrejelzési céllal, alkoholok és heterociklusos vegyületek példáján. Sem a peremregresszió, sem a PLS nem képes a Kováts indexek előrejelzését megfelelően megoldani. Predikciós modelljeink identifikálásra használhatók. Gázkromatográfias retenciós adatokból termodinamikai mennyiségeket számítottunk. Inverz gázkromatográfias adatok elemzésével polimerek, tömőanyagok osztályozását sikerült megoldani. Továbbfejlesztettük, általánosítottuk a pár-korrelációs módszert (PCM), és új PCM-en alapuló módszert dolgoztunk ki osztályozásra is. Az ózon koncentráció előrejelzését is megoldottuk főkomponens regresszióval, meghatároztuk az ózon koncentrációt befolyásoló faktorok közül melyek szignifikánsak. Borok eredetvizsgálatával (technológia, bortermelő hely, szőlőfajta és termelési év szempontjából), és flavonok hatásának előrejelzésével is foglalkoztunk. Összefoglaltuk az utóbbi tíz év magyarországi szerzőhöz köthető

kemometriai munkáit tudományos jelentőség, történelmi perspektíva, tudományos iskolák és alkalmazások (szofteverek) szerint.

A kutatási tevékenység rövid összefoglalása (angolul)

Using advanced chemometric methods HPLC systems were tested to select diverse (so called orthogonal) chromatographic systems; compared the available methods and defined a new measure for comparison, the orthogonality ratio. QSRR models were built to recognize the features of variable selection methods and to predict gas chromatographic retention data. On the contrary to the general belief, neither the ridge regression nor PLS is able to select proper features for prediction. Our validated models are suitable for identification purposes. Thermodynamic quantities were also calculated from gas chromatographic retention data. Classifications of polymers and fillers were elaborated by analyzing inverse gas chromatographic data. The pair-correlation method (PCM) has been generalized and a novel method based on successive application of PCM was developed for classification tasks. The concentration of ozone in air was predicted using principal component regression; and the significant factors influencing O₃ concentration was determined. Authenticity of wines was examined according to technology, geographic region, grape variety and year of vintage. The antioxidant effect of flavones was predicted using descriptors calculated from their molecular structure. The activity of the Hungarian chemometric community was reviewed according to scientific significance, historical perspective scientific schools (groups) and applications (softwares).

Melléklet:

Különlevonatlista az adott időszakra (2002-2004); a zárójelentésbe benyújtott 15 publikációt félkövérrel jeleztük.

74./ Károly Héberger and Róbert Rajkó: Variable Selection using Pair-Correlation Method. Environmental Applications.

SAR and QSAR in Environmental Research **13** /5/ 541-554 (2002) if(2002)=1.082

75./ Rozália Vanyúr, Károly Héberger, István Kövesdi and Judit Jakus: Prediction of Photosensitizer Properties Affecting Tumoricidal Effect in Photodynamic Therapy Using Artificial Neural Network,

Photochemistry and Photobiology **75** /5/ 471-478 (2002) if(2002)=2.241

76./ András Németh, Tamás Vidóczy, Károly Héberger, Zsolt Kúti and János Wágner: MECHGEN: Computer Aided Generation and Reduction of Reaction Mechanisms.

J. Chem. Inf. Comput. Sci. **42** /2/ 208-214 (2002) if(2002)=2.902

77./ Elemér Csomós, Károly Héberger and Livia Simon-Sarkadi, Principal Component Analysis of Biogenic Amines and Polyphenols in Hungarian Wines

J. Agric. Food Chem. **50** /13/ 3768-3774 (2002) if(2002)=1.915

78./ Károly Héberger, Miklós Görgényi and Teresa Kowalska, Temperature Dependence of Kováts Indices in Gas Chromatography Revisited.

- J. Chromatogr. A*, **973** /1-2/ 135-142 (2002) if(2002)=3.098
- 79./ Károly Héberger and Róbert Rajkó, Generalization of Pair-Correlation Method (PCM) for Nonparametric Variable Selection**
J. Chemometrics **16** /8-10/ 436-443 (2002) if(2002)=1.376
- 80./ Károly Héberger, Chemometrics, Present Developments, Prognosable Trends
Magy. Kém. Lapja, **57** /9/ 345-348 (2002) [in Hungarian] if=0.207
- 81./ Annamaria Jakab, Károly Héberger, Esther Forgács: Comparative Analysis of Different Plant Oils by High-performance Liquid Chromatography-atmospheric Pressure Chemical Ionization Mass Spectrometry,
J. Chromatogr. A, **976** /1-2/ 255-263 (2002) if(2002)=3.098
- 82./ Annamaria Jakab, Kornél Nagy, Károly Héberger, Károly Vékey, Esther Forgács: Differentiation of Vegetable Oils by Mass Spectrometry Combined with Statistical Analysis,
Rapid Commun. in Mass Spectr. **16** /24/ 2291-2297 (2002) if(2002)=2.372
- 83./ Miklós Görgényi, Károly Héberger: Minimum in the Temperature Dependence of the Kováts Retention Indices of Nitroalkanes and Alkanenitriles on an Apolar Phase
J. Chromatogr. A, **985** /1-2/ 11-19 (2003) if(2003)=2.922
- 84./ Teresa Kowalska, Károly Héberger and Miklós Görgényi: Temperature Dependence of Kováts Indices in Gas Chromatography. Explanation for Empirical Constants using Transition State Theory**
Acta Chromatographica, No. **13**, 60-68 (2003) if(2004)=0.698
- 85./ Rozália Vanyúr, Károly Héberger, and Judit Jakus: Prediction of Anti-HIV-1 Activity of a Series of Tetrapyrrole Molecules
J. Chem. Inf. Comput. Sci. **43** /6/ 1829-1836 (2003) if(2003)=3.078
- 86./ Károly Héberger Elemér Csomós and Livia Simon-Sarkadi, Principal Component and Linear Discriminant Analysis of Free Amino Acids and Biogenic Amines in Hungarian Wines**
J. Agric. Food Chem. **51** /27/ 8055-8060 (2003) if(2003)=2.102
- 87./ Károly Héberger, José M. Andrade, Procrustes Rotation and Pair-Wise Correlation: a Parametric and a Non-parametric Method for Variable Selection,**
Croatia Chemica Acta **77** /1-2/ 117-125 (2004) if(2004)=0.924
- 88./ Orsolya Farkas, Károly Héberger, and Igor G. Zenkevich, Quantitative Structure–Retention Relationships XIV. Prediction of Gas Chromatographic Retention Indices for Saturated O-, N- and S-heterocyclic Compounds.**
Chemometrics Intell. Lab. Syst. **72** /2/ 173-184 (2004) if(2004)=1.899
- 89./ Károly Héberger Chemometrics in Hungary (the last 10 years) [review]**
Chemometrics Intell. Lab. Syst. **72** /2/ 115-122 (2004) if(2004)=1.899
- 90./ Károly Héberger, Miklós Görgényi, and Teresa Kowalska: Effect of film Thickness on the retention index of nitro and cyano compounds. Interpretation of the logarithmic equation**
J. Chromatogr. Sci. **42** /6/ 288-292 (2004) if(2004)=1.166
- 91./ Attila Lengyel, Károly Héberger, László Paksy, Olivér Bánhidi, Róbert Rajkó Prediction of Ozone Concentration in Ambient Air using Multivariate Methods,**
Chemosphere **57** /8/ 889-896 (2004) if(2004)=2.359
- 92./ Vanyúr Rozália, Héberger Károly és Jakus Judit, QSAR Prediction of Anti-HIV-1 Activity of Porphirin Derivatives

- Magy. Kém. Folyóirat*, **109-110** /2/ 71-75, (2004). [in Hungarian] if=0.207
- 93./ Orsolya Farkas, Judit Jakus, Károly Héberger: Quantitative structure – antioxidant activity relationships of flavonoid compounds**
Molecules, **9** /12/ 1079-1088 (2004) if(2004)=0.676
- 94./ Milan Hájek, Claude Guillou Serge Rezzi, Fabiano Reniero and Károly Héberger, Biochemical Profiling by ¹NMR: Introductory review, European Commission, Joint Research Centre, Ispra, Italy, EUR 21483 EN (2004)
- 95./ Orsolya Farkas, Károly Héberger, Comparison of Ridge Regression, Partial Least Squares, Pair-wise Correlation, Forward- and Best Subset Selection Methods for Prediction of Retention Indices for Aliphatic Alcohols**
J. Chem. Inf. Modeling, **45** /2/ 339 -346, (2005) if(2004)=2.810
- 96./ Péter Forlay-Frick, Jenő Fekete and Károly Héberger, Classification and replacement test of HPLC systems using principal component analysis**
Anal. Chimica Acta, **536** /1-2/ 71-81 (2005) if(2004)=2.588
- 97./ Miklós Görgényi, Károly Héberger: Solvation Thermodynamics of n-Alkanes in Polymer Phases by Gas Chromatography
J. Sep. Sci. **28** /6/ 506-512 (2005). if(2004)=1.927
- 98./ Károly Héberger, Kasylda Milczewska and Adam Voelkel: Principal Component Analysis of Polymer-filler Interactions in Invers Gas chromatography**
Colloids and Surfaces, A: Physiochem. **290** /-/ 29-37 (2005) if(2004)=1.513
- 99./ Péter Forlay-Frick, Elke Van Gyseghem, Károly Héberger and Yvan Vander Heyden, Selection of orthogonal chromatographic systems based on parametric and non-parametric statistical tests (review)**
Anal. Chimica Acta, **539** /1-2/ 1-10 (2005) if(2004)=2.588
- 100./ Francesca Serra, Claude G. Guillou, Fabiano Reniero, Luciano Ballarin, Maria I. Cantagallo, Michael Wieser, Sundaram S. Iyer, Károly Héberger and Frank Vanhaecke: Determination of the geographical origin of green coffee by principal component analysis of carbon, nitrogen and boron stable isotope ratios
Rapid Commun. in Mass Spectr. **19** /15/ 2111-2186 (2005) if(2004)=2.750
- 101./ Zoltán A. Fekete, Károly Héberger, Zoltán Király, Miklós Görgényi, Temperature dependence of Solvation Heat Capacities by Gas Chromatography**
Anal. Chim Acta, **549** /1-2/ 134-139 (2005) if(2004)=2.588
- 102./ Serge Rezzi, David. E. Axelson, Károly Héberger, Fabiano Reniero, Carlo Mariani and Claude Guillou, Classification of olive oils using high throughput flow ¹H NMR fingerprinting with Principal Component Analysis, Linear Discriminant Analysis and Probabilistic Neural Networks
Analytica Chimica Acta, **552** /1-2/ 13–24 (2005) if(2004)=2.588
- 103./ E. Van Gyseghem, B. Dejaegher, R. Put, P. Forlay-Frick, A. Elkihel, M. Daszykowski, K. Héberger, D.L. Massart, Y. Vander Heyden, Evaluation of chemometric techniques to select orthogonal chromatographic systems
Journal of Pharmaceutical and Biomedical Analysis (accepted) if(2004)=1.509
- 104./ Adam Voelkel, Kasylda Milczewska, Károly Héberger, Inverse Gas Chromatographic and Chemometric Tools for Determination of Interactions between the Components of Polymeric Composition**
Analytica Chimica Acta, **559** /1-2/ 221-226 (2005) if(2004)=2.588