

Görbeillesztés nélküli komponensprofil-kinyerés az analitikai kémiában (MTA doktori értekezés összefoglaló)

RAJKÓ Róbert^{a,*}

^aNeumann János Egyetem GAMF Műszaki és Informatikai Kar, Izsáki út 10., 6000 Kecskemét, Magyarország

Bevezetés és célkitűzések

Az analitikai kémiai mérések során jelentős esetben kapunk mátrixba vagy általánosan többdimenziós adattömbbe rendezhető mérési adatokat. Ezek kiértékelésére, rendszerezésére, elméleti vizsgálatára alakult ki az 1970-es évektől kezdve a kemometria mint kémiai diszciplína, melynek létrejötté *Wold*, *Kowalski* és *Massart* kutatók nevéhez köthetik. Említésre méltó, hogy Magyarországon az 1980-as évek elejétől szerveznek Kemometria elnevezésű szemináriumokat, szimpóziumokat, majd 1997-ben *Conferentia Chemometrica '97* néven az első hazánkban rendezett nemzetközi konferencia is sikeresen megvalósult.

A kemometria a többváltozós/többdimenziós kémiai vagy hasonló jellegű mérési adatok kiértékelésére szolgáló módszerek összességét jelenti. Egy kicsit részletesebben, a kemometria a kémiai és analitikai kémiai információtartalom minél teljesebb (közel optimális) kinyeréséhez szükséges matematikai statisztikai, lineáris algebrai, konvex geometriai, információtechnológiai és számítástechnikai, ill. formál logikai módszerek és eljárások célirányos alkalmazásait foglalja magába. A kemometria tehát a kémia, a matematika és az informatika mérési adat-központú határtudománya, mégis vannak olyan elméleti jellegű kemometriai levezetések, amikor az adatokat egyes tulajdonságaik alapján általánosítjuk és így konkrét mérési adattól függetlenül végezhetjük elméleti vizsgálódásainkat. Nagyon gyakran számítógéppel szimulált „mérési” adatokkal (is) dolgozunk. Az elméleti kemometria keretében, a molekulamodellezés elnevezés analógiájára, adatmodellezésnek nevezhetjük az előbb vázolt tevékenységet (az informatikában használt adatmodell fogalom eltér az előzőekben leírtaktól, hiszen ott egy logikai adatbázishoz rendelik és programozási technikákkal hozzák kapcsolatba).

Wallace már 1960-ban megfogalmazta azt az adatmátrixokra vonatkozó alapvetést, ami a kemometria alapkövét jelentette: “Ha az elegyenben lévő minden komponensnek eltérő a spektruma és a kifejlődő (elúciós) koncentráció profilja, akkor az adatmátrix *algebrai rangja* (*pseudorangja* zaj esetén) és a rendszerben lévő *komponensek száma* között a kapcsolat egyértelmű. Az egyetlen feltétel a *Bouguer-Lambert-Beer-törvény* érvényesülése.” ($A = C \cdot S^T + R$,

ahol C a koncentráció mátrix, S az egységnyi koncentrációhoz tartozó tiszta komponensek spektrumait tartalmazó mátrix, A az adatmátrix, pl. egy spektroszkópiai mérés során az adott mintákra kapott, több hullámhosszon mért abszorbanancia értékek, R a reziduális (maradék) mátrix, ami gyakran a mérési hibához (zajhoz) rendelhető, az értékelhető jelinformációtól elkülönült rész). Kezdetben tehát a lineáris algebra, de főként a mátrixaritmetika vizsgálata jelentette a fő irányvonalat, mennyire lehet ezen elveket analitikai kémiai mérések adatainak kiértékelésére használni. A legfontosabb alapmódszer a *szinguláris érték felbontás* (Singular Value Decomposition, SVD). Általában a nagyon kicsi szinguláris értékhez tartozó szinguláris vektorokat elhagyjuk, így dimenzióredukciót érhetünk el. A következő lépésben, a *főkomponens elemzés* (Principal Component Analysis, PCA) eredménye alakítható ki az SVD felhasználásával. A főkomponens elemzés alap gondolata az, hogy az eredeti adatmátrixot két mátrix szorzatával adjuk meg. Végtelen sok ilyen felbontás létezik, viszont olyan megkötésekkel, mint ortogonalitás és normálás, a merev forrástól eltekintve egyértelmű felbontást kapunk. Minden adatredukciós eljárás egyik fontos lépése az adatok előkészítése (data pre-processing). Itt csak utalunk a centrálásra, a skálázásra, az egységnyi hosszra történő normálásra. A fentiekben vázolt feladat megoldására (C és S együttes becslése az A adatmátrix dekompozíciójával, felbontásával) két markáns irányvonal alakult ki:

- az összes C és S megengedett megoldás feltérképezése (Self-modeling Curve Resolution – SMCR: görbeillesztés nélküli komponensprofil-kinyerés),
- egyetlen C és S megengedett megoldás megtalálása (Multivariate Curve Resolution – MCR: többváltozós görbefulbontás).

Mindkét irányvonal kényszerfeltételeket (constraints) használ a megengedett megoldás-halmaz (set of feasible solutions, SFS), vagy más kifejezéssel megengedett megoldás-tartomány (feasible region, FR) esetleges szűkítésére. A legkézenfekvőbb (természetes) kényszerfeltétel a nemnegativitás: a koncentráció sohasem lehet negatív és a legtöbb spektrum (jól beállított alapvonallal mérve!) szintén nem tartalmazhat negatív értéket (de kivétel pl. a cirkuláris dikroizmus (CD) spektroszkópia).

* E-mail: Prof.Rajko@gmail.com

Az értekezésem címének megfelelően, a görbeillesztés nélküli komponensprofil-kinyerés (mint a kemometria egyik meghatározó területe) témakörében megjelent publikációimat kívántam összefoglalni önálló, jól meghatározott tudományos kérdés (görbeillesztés nélküli komponensprofil-kinyerés) megoldásával foglalkozó 2005-2020 között megjelentetett eredményekkel.

A mérési adatok modellezése kapcsán első körben elméleti fejlesztéseket tűztem ki magamnak, amely *Lawton és Sylvestre* 1971-ben, ill. *Borgen* 1985-86-ban publikált eredményeinek tovább gondolásával indult a megengedett megoldás-tartományok számítógépes geometriai módszerekkel történő analitikus meghatározásával. Azután *Henry* cikkének eredményét vizsgáltam meg és a természetes dualitás igazolását szándékoztam elvégezni minimális kényszer-feltételeket alkalmazva. Tauler cikkére reagálva a megengedett megoldás-tartományok határainak számítását és a forgatási bizonytalanság meghatározását készítettem tüzetesen elemezni. Felismerve, hogy az LS módszer L_1 normán, a *Borgen* módszer viszont az első-sajátvektor-értékei-1 normán alapul, a Borgen normák általános megfogalmazását és vizsgálatát gondoltam megvalósítani. *Abdollahi és mtsai* kétkomponensű rendszereket vizsgált numerikus módszerekkel (*mcrbands* szkript és rácsmódszer), viszont én felismertem, hogy a feladat elvégezhető analitikus levezetésekkel is. Ismert-profil kényszerfeltétel alkalmazását határoztuk el fajtaalapú részecske-raj-optimalizálás alkalmazásával. Tovább lépve, nemcsak az ismert-profil, hanem az unimodalitás kényszerfeltételek alkalmazása és a Borgen módszer dualitáson alapuló kidolgozása mellett döntöttünk. A Borgen módszer alkalmazása néha megmagyarázhatatlan hibaüzenettel végződött, melynek kivizsgálását kívántuk elvégezni és ennek kapcsán az adat-alapú egyértelműen meghatározható profilbecslés definícióját és kimutatását kidolgozni. Az MCR és SMCR területen dolgozó néhány nemzetközi kiválósággal karöltve egy áttekintő (review) cikk megírását határoztuk el. Két- és háromkomponensű három-utas adattömbök nemnegatív egyértelmű felbonthatóságának boncolgatását választottuk következő feladatként. Tovább lépve, kétkomponensű három-utas adattömbökre analitikus megoldástartományt meghatározó módszer kifejlesztését céloztuk meg. Az elméleti blokk lezárásaként, ritkás profil-mátrixokra történő nemnegatív többváltozós görbefeletbontás során az L_0 , L_1 és L_2 normák szerepét kívántuk tisztázni. A mérési adatok kiértékeléséhez kapcsolódó gyakorlati fejlesztéseket és alkalmazásokat a PLS regresszióhoz, a túlillesztés elkerülésére, alternatív érvényesítési eljárás kidolgozásával indítottuk. A következőkben az MCR-ALS algoritmus néhány meglepő gyakorlati tulajdonságainak összegyűjtését és elemzését határoztam el. *Lopes et al.* cikkével kapcsolatban kritikai megjegyzéseimnek publikálását döntöttem el. HPLC/FT-IR mérések során az eluens-hatás kemometriai eliminációját céloztuk meg. A továbbiakban a meloxikám (ME) és a mannit (MA) mint hordozóanyag fizikai és olvadék keverékének kioldódási viselkedésének jellemzését választottuk. Folytatásként a ME-MA binér rendszerek kioldódási vizsgálata során

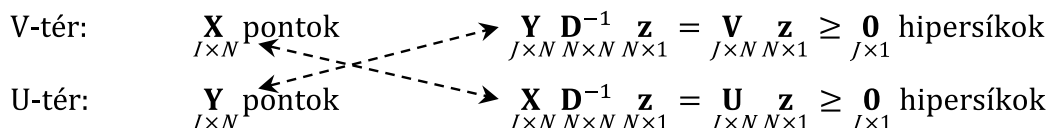
a görbeillesztés nélküli komponensprofil-kinyerő módszer alkalmazhatóságának vizsgálata mellett döntöttünk. Érdekes újdonságnak ígérkezett a lokális rang kényszerfeltétel vizsgálata és a lokális rangvesztés fogalmának bevezetése és értelmezése. Befejezésként, kalibrációs mintákból hiányzó zavaró komponensek jelenléte esetén koncentráció meghatározáshoz dualitás alapú általánosított rang-eltüntető algoritmus kifejlesztése vetődött fel, realisztikus hibabecsléssel, mivel pl. ez utóbbival az akkor rendelkezésre álló GRAM módszerek nem rendelkeztek.

Tudományos megállapítások tézisei (T1-T20)

T1. Számítógépes geometriai módszerekkel a megengedett megoldás-tartományokat analitikusan határoztam meg háromkomponensű rendszerekben [P1]. 1971-ben *Lawton és Sylvestre* (LS) vezette be a módszert és az elnevezést (SMCR) két-komponensű rendszerek felbontására. Háromkomponensű rendszerekre *Borgen* és munkatársai általánosították *Lawton és Sylvestre* módszerét, sajnos azonban a cikkek alapján sokáig nem tudták rekonstruálni, ill. maradéktalanul értelmezni az algoritmust. Ún. számítógépes geometriai módszereket alkalmaztam a kemometriai irodalomban elsőként, így egy igen gyors Matlab implementációt sikerült készítenem. Az általam *Borgen grafikon*-nak (*Borgen plot*) elnevezett ábra elkészítéséhez először az adatmátrixunk normalizálását végeztem el az egyes sorok, ill. oszlopok összegeit felhasználva. A normalizálás okozta dimenzió-csökkenés miatt az absztrakt 2-dimenziós síkon lévő adatpontok konvex burkának meghatározását a *conv-hull* Matlab függvénnyel végeztem (első számítógépes geometriai módszer). Megkaptuk a belső poligont (inner polygon), amit egy megengedett megoldás szimplexe – vagyis három komponens esetén háromszög –, egyetlen oldala sem metszhet, legfeljebb érintheti. A *Fourier-Motzkin eliminációs módszer* (FMEM) egy lineáris egyenlőtlenségrendszerből a redundáns egyenlőtlenségeket szűri ki. A redundáns egyenlőtlenségek észlelése a csúcsok és szélső irányok listájának kezelésével a dupla-deszkripció (double description (DD)) módszerrel történt, amit a Matlabban a *cddmex* harmadik fél által fejlesztett C/C++ alacsony szintű programnyelven készített és előre lefordított külső függvénnyel valósítottam meg (második számítógépes geometriai módszer). A nem redundáns felsíkok a külső poligont (outer polygon) határozzák meg. A megengedett megoldás-tartományok a belső és a külső poligonok között keresendők, tehát a nemnegatív profilok absztrakt pontjai a külső poligonon belül helyezkednek el, és egy háromszög csúcsaként a belső poligont körbeveszik, azaz lineáris konvex kompozícióként az adatpontok előállíthatók. Ebben a tanulmányban vezettük be az Average Orthogonal Distance from the Linear Subspace (AODLS) elnevezésű mértéket, ami egy pontthalmaz átlagos ortogonális távolságát adja meg egy lineáris altértől. Az AODSL segítségével kimutattuk, hogy a *Lawton és Sylvestre* által használt adathalmaz akár négy komponens is tartalmazhat kettő helyett.

T2. Igazoltam a természetes dualitást a minimális kényszerfeltételű görbeillesztés nélküli komponensprofil-kinyerés során [P2]. Ebben a közleményben sikerült általánosan igazolnom, hogy ha a koncentráció és spektrális profilokra csak a nemnegativitás feltételezéssel élünk, akkor a V-térben lévő pontok számával fog megegyezni az U-térben lévő hipersíkok száma és fordítva. Sőt a konvex csúcsok definícióját értelmezve azt a szigorúbb megállapítást tehetjük, hogy a V-térben lévő külső konvex csúcsok által meghatározott

lapok (facets) megfelelnek az U-térben lévő megfelelő belső konvex csúcsoknak és fordítva. Mivel a belső konvex csúcsokhoz a megfelelő térbe transzformált profilok, mint pontok rendelhetők, így az SMCR feladat redukálható a belső politópok (három komponens esetén 2-dimenzióban poligonok) meghatározására és így az esetlegesen időigényes és bonyolult számítógépes geometriai módszerek mellőzhetők és/vagy egyszerűbbre cserélhetők. A természetes dualitást a következő összefüggés-hálózattal illusztráltam:



T3. Számítottam a megengedett megoldás-tartományok határait és meghatároztam a forgatási bizonytalanság mértékét [P3]. Gemperline vezette be először a megengedett megoldás-tartományok terjedelmének (range of feasible solutions) meghatározását az egyes komponensek profiljainak görbe alatti területe (azaz az L1 normájuk) maximalizálásával, ill. minimalizálásával. Később Tauler egy másik vektornormát, az L2 euklideszi normát alkalmazta és bevezette a jel hozzájárulás függvényt (signal contribution function, SCF). Az elképzelést, hogy két szélsőértékkel elég lesz jellemezni a forgatási bizonytalanságot, nyilvánvalóan Lawton és Sylvestre kétkomponensű megoldása sugallta. Azonban kimutattam, hogy két szélsőérték általában nem elegendő. Ehhez az általam fejlesztetett és korábban bemutatott Borgen-plot rajzoló Matlab szkriptemet használtam fel. A rotációs bizonytalanság számszerűsítésére bevezettem a komponensek koncentráció és spektrum szerinti Borgen grafikonbeli megengedett megoldástartományainak területe és a külső poligon területe között számolt arányt, valamint a Gemperline-Tauler módszerrel kapott maximum és minimum SCF értékek különbségét. A Gemperline-Tauler módszer komponensenként, a koncentráció és spektrum hatás megkülönböztetése nélkül nyújt információt, míg a terület-arány külön a koncentrációra és külön a spektrumokra is meghatározható. A probléma megvilágításához képzeljük el, hogy az egyik komponens spektruma ismert. Így az ehhez tartozó rotációs bizonytalanság nulla lesz. Azonban a koncentrációhoz tartozó rotációs bizonytalanság nem feltétlenül lesz nulla. A Gemperline-Tauler módszerrel kapott maximum és minimum SCF értékek különbsége viszont ebben az esetben hibás eredményt ad: ha az érték nulla, akkor nem kezelte megfelelően a koncentrációban maradt rotációs bizonytalanságot; ha értéke nem nulla, akkor nem megfelelően kezelte az ismert spektrumot.

T4. Defináltam a Borgen-normát és tisztáztam alkalmazhatóságát [P4]. A Borgen-normák definícióját és részletes leírását adtam meg, amely téma először jelent meg a kémi-

ai szakirodalomban. A Borgen-normák általános alakját a következőképpen határoztam meg: . Bemutattam még a Borgen normák esetére az általános levezetéseket az SMCR-rel kapcsolatban. Az egyik legfontosabb és legmeglepőbb következtetésem, hogy a rotációs bizonytalanság mértéke függ az alkalmazott Borgen normától a minimális kényszerfeltételű (nemnegatív összetétel- és jel-profilok vételezése) SMCR esetén. Tisztáztam, hogy a Borgen grafikonok segítségével visszatranszformált profilsávokat utólag normalizálhatjuk, vagy változatlanul hagyhatjuk. A nem normalizált sávterületek alkalmasak a komponensek rotációs bizonytalanságának összehasonlítására egy kiválasztott Borgen normára vonatkozóan; míg a normalizált sávterületek a különböző Borgen-normákból adódó rotációs bizonytalanság összehasonlítására használhatók egy kiválasztott komponensre. Ezenkívül megállapítottam, hogy a különböző Borgen-normákkal elvégzett SMCR vizuális validációs eljárásaként használható a valódi profilok unimodalitásának eldöntésére a profilsávok alapján, előzetes fizikai vagy kémiai ismeretek vagy az adatokra vonatkozó feltételezések nélkül.

T5. Kétkomponensű rendszerek megengedett megoldás-tartomány határait értelmeztem és analitikus számítási módszert adtam meg [P5]. Korábban Abdollahi és mtsai egy kétkomponensű rendszer SMCR-éhez a transzformációs mátrix elemeinek kiszámításához és sávhatárok meghatározásához a Tauler-féle *mcrbands* Matlab szkriptet és a Maeder-féle rácsmódszert használták. Egyik módszer sem analitikus, hanem iteratív, meglehetősen nagy számításigénnyel. Régóta köztudott, hogy Lawton és Sylvestre megközelítése analitikusan, tehát nem iteratív módon adhatja a megengedett megoldás sávhatárait. Elsőként a szakirodalomban, egyértelmű kapcsolatot mutattam be Lawton és Sylvestre, Tauler, valamint Maeder megközelítései között. A Tauler-Gemperline-féle SCF maximumának és minimumának kétkomponensű rendszerekre vonatkozó számításához szükséges transzformációs mátrix elemeit a következőképpen adtam

meg az 1. komponensre: $\bar{t}_{12,MAX,1} = \frac{\bar{x}_{2,xBl}}{\bar{x}_{1,xBl}}, \bar{t}_{21,MAX,1} = \frac{\bar{x}_{1,xAu}}{\bar{x}_{2,xAu}}; \bar{t}_{12,MIN,1} = \frac{\bar{x}_{2,xBu}}{\bar{x}_{1,xBu}}, \bar{t}_{21,MIN,1} = \frac{\bar{x}_{1,xAl}}{\bar{x}_{2,xAl}}$
 illetve a 2. komponensre: $\bar{t}_{12,MAX,2} = \frac{\bar{x}_{2,xBu}}{\bar{x}_{1,xBu}}, \bar{t}_{21,MAX,2} = \frac{\bar{x}_{1,xAl}}{\bar{x}_{2,xAl}}; \bar{t}_{12,MIN,2} = \frac{\bar{x}_{2,xBl}}{\bar{x}_{1,xBl}}, \bar{t}_{21,MIN,2} = \frac{\bar{x}_{1,xAu}}{\bar{x}_{2,xAu}}$

A könnyebb ellenőrizhetőség végett tanulmányomhoz mellékeltem a részletes számításokat elvégző Matlab szkriptet, ill. néhány mesterségesen előállított hibaszint mellett a minimum és maximum SCF-ekhez tartozó sávhatárokat is, és megfigyeltem a hibaszint okozta jellegzetes torzulásokat. Megemlítem, hogy amíg a *Tauler-féle mcrbands* és a *Maeder-féle rácsmódszer* az eredeti kezdő profilbecslésekhez ad határokat, addig az LS-alapú számításom csak normalizált határokat ad, azonban ez utóbbi eljárás analitikus és nem-iteratív.

T6. *Speciális rácskereső módszerrel az ismert-profil kényszerfeltétel alkalmazását dolgoztam ki a forgatási bizonytalanság csökkentésére* [P6]. Háromkomponensű rendszerek esetén az összes összetartozó megengedett-megoldás halmazt egy speciális rácskeresés módszerrel határoztuk meg, ami a fajtaalapú részecske-raj-optimalizálás (Species-based Particle Swarm Optimization, SPSO) algoritmuson alapul. A módszert zajmentes és zajos adatsor segítségével ellenőriztük. Az eredmények azt mutatták, hogy a módszer alkalmas az adatok elemzésére, azonban meglehetősen időigényes. A kifejlesztett módszerrel az egyenlőség kényszerfeltételt (ismert koncentráció vagy spektrális profil figyelembevétel: ismert-profil) is alkalmaztuk és várakozásunknak megfelelően a forgatási bizonytalanság és a megengedett megoldás-tartományok drasztikusan csökkentek. Részletekbe menően megmutattuk még, hogy az ismert-profil kényszerfeltétel a komplementer komponensekre vonatkozóan egyenesek megfelelő szakaszaira egyszerűsödő megengedett-tartományokat eredményez.

T7. *Az ismert-profil és unimodalitás kényszerfeltételek alkalmazását dolgoztam ki dualitás alapú görbeillesztés nélküli komponensprofil-kinyeréshez* [P7]. A görbeillesztés nélküli komponensprofil-kinyerés (SMCR) módszerei a mátrix alapú adatkészletekhez a teljes megengedett megoldás-tartományokat adják meg csak nemnegatív kényszerfeltétel alkalmazásával. Úgy tűnik, hogy konvex geometriai szemlélet is szükséges a görbefeblontó módszerek megértéséhez és fejlesztéséhez, mivel a gyakran komplikált (lineáris) algebrai koncepciók kizárólagos alkalmazása megakasztotta a Borgen módszer alaposabb megértését 20 éven keresztül. Az analitikai SMCR módszereket vizsgáltuk felül és írtuk le a dualitás alapelveihez kapcsolódó egyszerű fogalmak segítségével. Ezenkívül, először a szakirodalomban, az egyenlőség (ismert-profil) és unimodalitás kényszerfeltételeket sikeresen implementáltuk a Lawton–Sylvestre módszerhez. A Borgen grafikonok megrajzolásához is egy korszerűbb eljárást fejlesztettünk az ismert-profil kényszerfeltétel alkalmazásához. Két- és háromkomponensű HPLC-DAD adatsorokat szimuláltunk és

elemeztük őket az újabb kényszerfeltételek felhasználásával és azok nélkül. Az *LSandBP* Matlab szkriptünk a dualitás elvén alapuló egyszerűsítéseket tartalmazza így a Borgen grafikonok néhány másodperc alatt elkészültek. A dualitás elvén alapuló egyszerűsítés pl., hogy az egyik absztrakt térben (mondjuk az oszlop térben) konvex burok algoritmus-sal meghatározott belső poligon a duális absztrakt térben (mondjuk a sor térben) a külső poligont határozza meg, ami igaz fordítva is, azaz a másik absztrakt térben (mondjuk a sor térben) meghatározott belső poligon az első absztrakt tér (mondjuk az oszlop tér) külső poligonját rögzíti. Így nem kell a számításgényes dupla-deszkripció (DD) (*cddmex*) számítógépes geometriai algoritmust igénybe vennünk. A másik egyszerűsítés, hogy elegendő csak az egyik absztrakt térben meghatározni a Borgen szimplexekeket (Borgen háromszögek három komponens esetén), mert a dualitás elvének alkalmazásával a másik absztrakt térben iteráció nélkül megkapjuk azokat.

T8. *Bevezettem az adat-alapú egyértelműen meghatározható profilbecslés fogalmát és tételyszerűen megadtam kimutatási lehetőségeit* [P8]. A Manne tételek kritikáját és új elvek megfogalmazását végeztük el. Sikertelt bemutatnunk egy ellenpéldát, amikor *Manne 1. és 2. tétele* teljesül, de a felbontás nem lesz egyértelmű. A megtalált ellenpélda ösztönzött bennünket, hogy áttekintsük újra az egyértelmű felbontás lehetőségeit, ill. a Manne tételek helyett megfelelő megfogalmazást találjunk. A Manne tételek a profil-alapú egyértelműségekre vonatkoznak, míg az általunk bevezetett eljárás az adat-alapú egyértelműségekre vonatkozik. Két formában fogalmazhatjuk meg az adat-alapú egyértelműen meghatározható profilokra vonatkozó tételünket:

1. forma – Ha a Borgen grafikonon (vagy annak általánosításán háromnál több komponens esetén) találunk olyan pontot, amely mind a belső és mind a külső poligonok valamely csúcspontjaihoz tartozik és ennek következtében a Borgen háromszögek (általánosan Borgen szimplexekek) közös csúcspontja, ez a pont egyértelmű megoldás lesz az adott komponens részére.
2. forma – Ha egy komponensnek van szelektív ablaka az egyik irányban/módban/útban (pl. spektrum) és a másik irányt/módot/utat (pl. koncentráció) tekintve ennek a komponensnek van zéró ablaka, ahol az összes zavaró komponens megjelenik, akkor ennek a komponensnek a második irányban/módban/útban található (koncentráció) profilja egyértelműen meghatározható.

Elsőként a kemometriai szakirodalomban az alábbi egyértelműségi kategóriákat vezettük be: Nem-egyértelmű felbonthatóság (non-uniqueness), azaz egyetlen profil sem

adható meg egyértelműen. Töredékes-egyértelműség (fractional uniqueness): egy komponens egyetlen profilja adható meg egyértelműen, a többi nem. Részleges-egyértelműség (partial uniqueness): egy komponens összes profilja egyértelműen megadható. Teljes-egyértelműség (full uniqueness): az összes komponens összes profilja egyértelműen megadható.

T9. *Részletekbe menően áttekintettem és összehasonlítottam az újabb görbeillesztés nélküli komponensprofil-kinyerő módszereket [P9]. Ez az áttekintő mű az addig megjelent görbeillesztés nélküli komponensprofil-kinyerő módszerek áttekintését adta meg. Kiemelendő, hogy a társszerzők között Prof. Maeder, Prof. Tauler, Prof. Abdollahi és Prof. Neymeyr is ott voltak, az alkotás igazi nemzetközi kooperációban készült. A forgatási bizonytalanság feltárásával kapcsolatban szintetikus és valós adatokat elemezve megállapítottuk, hogy bár a végeredmény nagyon hasonló a bemutatott eljárások algoritmusai lényegi különbséget mutatnak. A Borgen-Rajkó grafikonok számítása explicit módon történik, így nagyon gyorsan megkapjuk az eredményt. Ez az elméleti megfontolásokon alapuló módszer érvényesítésre (validation) használható. Az egyetlen hátránya, hogy valós, zajjal terhelt adatokra nem mindig alkalmazható. A szimplex és poligon tágitás (simplex and polygon inflation) módszerek eredménye nagyon hasonlít az előző elméleti alapú módszerrel kapottra. Az előnye ennek a közelítő módszernek, hogy valós zajjal terhelt adatokkal is működik. Az MCR-BANDS algoritmus teljesen különböző. Ez talán a legkönnyebben elérhető program és bármilyen komponens szám esetén használható. Az algoritmus könnyen implementálható, főképp Matlab-ban, mivel ott készen elérhető nemlineáris optimalizációs eljárások használhatók nemlineáris kényszerfeltételekkel. Itt a zajhatást újraminintázó, pl. bootstrapping eljárásokkal térképezhetjük fel. A hátránya, hogy csak közelíteni tudja a valós sávmegoldásokat, mivel csak két extrém, norma-alapú megoldás alapján dolgozik.*

T10. *Igazoltam, hogy Kruskal elégséges és szükséges feltétele két-, ill. háromkomponensű háromutas adat-tömbökre sérülhet és legfeljebb csak elégséges maradhat [P10]. A méréseink során az egyértelmű felbontás az egyik legvonzóbb tulajdonság, hogy trilineáris adattömböt hozunk létre kapcsolt technikákkal. Sajnos a triadikus dekompozíció identifikálása még nem teljesen feltárt, annak ellenére, hogy elméleti és gyakorlati fontossága ismert. A rang-vesztéssel (rank deficiency) sújtott lódingú (loading) adattömbökre egyelőre Kruskal által bevezetett egyenlőtlenség alkalmazható. Vizsgálódásainkból levontuk a következtetést, hogy Kruskal elégséges és szükséges feltétele két-, ill. háromkomponensű háromutas adattömbökre sérülhet és legfeljebb csak elégséges maradhat. Ten Berge és Sidiropoulos igazolta, hogy esetekben a nulla értékek speciális mintázata esetén lehetséges egyértelmű felbontás annak ellenére, hogy a Kruskal egyenlőtlenség nem teljesül, ezt most mi igazoltuk és esetekre is. Az adat-alapú egyértelműség elvének alkalmazásával a következő feltételeknek kell teljesülniük az egyértelmű felbontáshoz a Kruskal egyenlőtlenség nem*

teljesülése esetén: 1) trilinearitás, 2) szelektív ablakok jelenléte, és végül 3) a speciálisan elhelyezkedő zérusok mintázata. Fontos konklúzió, hogy a Kruskal egyenlőtlenség teljesülése nem feltétlenül szükséges a kényszerfeltételes trilineáris dekompozícióhoz két-, ill. háromkomponensű háromutas adattömbök esetében.

T11. *Analitikusan megadtam és értelmeztem a megengedett megoldás-tartományokat kétkomponensű három-utas adattömbök esetén [P11]. A mátrix alapú adatok helyett kétkomponensű három-utas (azaz 3D) adattömböket vizsgáltunk. Így módon az LS módszer egy újabb általánosítását kaptuk. Az elemzett adattömbök alapján a következő megállapításokat tettük. $k\{2,2,2\}$ adattömbök esetén a bemutatott algoritmusunk mindig gyorsan és megfelelően működött, ugyanazt az egyértelmű megoldást szolgáltatva, mint pl. az iteratív eljárás alapján PARAFAC. $k\{2,2,1\}$ adattömbök esetén a Kruskal egyenlőtlenség sérül: , tehát az egyértelmű megoldás nem biztosított. A koncentráció mátrix k -rangja 1 (rangátfedés (rank overlap)), ami a profilok lineáris arányosságának, esetleg megegyezésének következtében alakul ki. Ebben az esetben meglepő módon az összes megengedett megoldás-tartományba eső megoldás trilineáris lesz, azaz nem jelent előnyt a trilineáris kényszerfeltétel alkalmazása, nem kapunk egyértelmű megoldásokat. Fontos következtetésünk, hogy ez a fajta három-utas adattömb bilineáris adatként kezelendő: minden egyes mátrixszelet vizsgálata ugyanarra az eredményre vezet. Adattömbként kezelés csak a szabotosságot és pontosságot csökkentheti a mérési zaj hatásától függően. Kimondtuk, hogy ezeket a megállapításokat általánosíthatjuk $k\{N,N,1\}$ típusú adattömbökre, amikor az utolsó út szerinti profilok megegyeznek, azaz $k_3=1$ (N helyett).*

T12. *Ritkás profil-mátrixok esetén tisztáztam az L_0 , L_1 és L_2 normák szerepét nemnegatív többváltozós görbefeletbontás során [P12]. Az SMCR módszer segítségével határoztunk meg ritkás (sparse) profilokat és kiemeltük, hogy normális adatsor esetén ezek a külső poligonon helyezkednek el, ami az elválasztó határ a csak pozitív és a negatív értékeket is tartalmazó profilok absztrakt pontjai között. A legritkásabb megoldások a külső poligon csúcspontjai lesznek. Másrészt, ha az adatmátrixunk csak nullákat tartalmazó sorokkal vagy oszlopokkal is rendelkezik (abnormális adatsor), akkor a megengedett megoldás-tartomány belső pontjainak megfelelő profilok ugyanannyi nullát fognak tartalmazni (a nullák elhelyezkedése ezekben a profilokban függ a csak nullákat tartalmazó sorok (oszlopok) mátrixon belüli elhelyezkedésétől: egy teljesen nulla sor az adatmátrixban egy nulla elemet eredményez a koncentráció profilokban, és egy teljesen nulla oszlop az adatmátrixban egy nulla elemet eredményez a spektrális profilokban), de még ekkor is a külső poligon pontjainak megfelelő profilok több nullával fognak rendelkezni. A hasznos tartomány (region of interest, ROI) módszerrel pl. nagyon könnyen kaphatunk abnormális adatsorokat. A szakirodalomban a ritkás megoldás meghatározására az L_1 norma vagy az L_1 és L_2 normák kombinációjának minimalizálását javasolják (LASSO*

és elasztikus háló (elastic net)), azonban megállapítottuk, hogy ez csak egyetlen komponens egyetlen profiljának ritkás meghatározásánál használható. Amennyiben 2, 3 vagy több komponens ritkás profiljait kell egyszerre meghatározni, akkor az L_0 norma minimalizálását vagy az L_2 norma maximalizálását és az L_1 norma alkalmazásának mellőzését javasoljuk.

T13. *Alternatív érvényesítési eljárást fejlesztettem a túlillesztés elkerülésére* [P13]. Ebben a tanulmányunkban kritikusan áttekintettük a többváltozós kalibrációnál jelentkező túlillesztés problémáját és a hagyományos érvényesítés (validation) alapú eljárást, ami a túlillesztés elkerülését célozza. Kifejlesztettünk egy alternatív eljárást, ami egy randomizációs-permutációs teszten alapul; ez lehetővé teszi, hogy minden egyes komponenshez, amit a modellbe illesztünk statisztikus szignifikancia szintet rendeljünk. A következő megállapításokat tettük: (1) alternatív eljárásunk enyhé feltételezések mellett alkalmazható, (2) felhasználóbarát, csak a permutációk számát és a szignifikancia küszöbszintet kell megadni, (3) az eredmények gyakran azonosak az érvényesítés alapúakkal (pl. Unscrambler vagy SIMCA), de most azokat teljesen objektív alapon kaptuk, (4) helyettesítheti az érvényesítést a komponensek kiválasztásához, de kiegészítheti a szokásos diagramot (RMSEP becslések vs. komponensek) is, (5) a randomizációs tesztet nem csak PLS regresszióra alkalmazhatjuk, hanem többutas kalibrációhoz is, (6) tömörítésre lehet szükség nagy adathalmazok (big data) esetén, de arra figyelni kell, hogy a tömörítés ne okozzon függést a minták között, (7) a bemutatott randomizációs teszt a kalibrációs adatkészleten működött, hogy a kereszt-érvényesítés objektivitását biztosítsuk, azonban semmi akadálya, hogy ezt külső adatkészlettel érvényesítésre is alkalmazzuk.

T14. *Az MCR-ALS algoritmus meglepő gyakorlati tulajdonságát tártam fel és az orvoslásához egy új kényszerfeltételt vezettem be* [P14]. Az MCR-ALS algoritmus egy meglepő problémáját tártam fel: az iterációk során kapott rész megoldások és még a végső megoldás is az $\mathbf{R}$ adatmátrix nullterében lesznek. Megmutattam, hogy a Borgen grafikonon jól illusztrálható az iterációs nyomvonal és az, hogy a nullterben lévő nemnegatív (rész)megoldás az $\mathbf{R}$ mátrix sor-, ill. oszlopterébe vetítve már negatív elemeket eredményez a profilokban. A probléma orvoslásához egy új kényszerfeltételt vezettem be: becsült profilok zéró ortogonális részzel (zero orthogonal part for the estimated profiles).

T15. *NIR hiperspektrális adatok kiértékelése során alkalmazott minimális térfogat kritériummal kapcsolatban tettem kritikai megjegyzéseket* [P15]. Egy megjegyzés közleményt jelentettem meg egy tanulmánnyal kapcsolatban, amit Lopes et al. közölt az *Analytical Chemistry*-ben. Elsőként azt jegyeztem meg, hogy a tiszta pixel nélküli hiperspektrumok esetére a földtudományok területén megjelent munkára hivatkoztak, pedig a kémiai szakirodalomban a CCA (convex constraint analysis) algoritmust már korábban leközölték. Másodszorban arra hívtam fel a

figyelmet, hogy a forgatási bizonytalanság nem az MCR-ALS vagy más algoritmus tulajdonsága, hanem a bilineáris adatmátrix belső sajátja! Lopes et al. a nemnegativitást alkalmazták kényszerfeltételként, valamint egy szükséges feltételt: „a szimplex minden $(p-1)$ dimenziós oldala, azaz lapja (facet) tartalmazza $p-1$ spektrális adatvektort”. Ez kemometriailag azt jelenti, hogy minden komponensre vonatkozóan léteznie kell legalább $p-1$ spektrális sávnak (hullámhossz, hullámszám stb.), ahol csak ez egyik komponensnek lesz nulla az elnyelése. Rávilágítottam, hogy ez olyan erős megkötés, ami csak ritkán fordul elő a gyakorlatban. Továbbiakban illusztráltam, hogy a koncepció, amit használtak nem minden esetben ad elfogadható eredményt: nemnegatív kényszerfeltétel alkalmazása mellett a minimális térfogatú szimplex (három komponensű rendszerek esetén minimális területű háromszög) a duális térben negatív értékeket is tartalmazó profilok absztrakt pontjait eredményezheti. Egy további kritikus problémaként az alkalmazott normalizálást jártam körül.

T16. *Kifejlesztettem egy kemometriai módszert, amellyel a HPLC/FT-IR mérések során az eluens-hatást küszöbölhetjük ki fizikai oldószerelimináció nélkül* [P16]. A HPLC/FT-IR mérések során az átfolyós folyadékcélával megvalósított IR detektálás előnyeit (gyorsaság, aránylag egyszerű kivitelezés, sokféle eluens alkalmazhatósága) kihasználva a fizikai oldószereliminációt és annak hátrányos tulajdonságait (csak illékony eluens alkalmazható, az elpárologtatás során meghatározandó komponens is távozhat, bonyolult kivitelezés) kemometriai módszerre cseréltük és egy megfelelő algoritmus segítségével vontuk ki az eluens hatását. Kifejlesztettünk egy új iterációs eljárást, az OSSS-IU-PARAFAC-ot (Objective Subtraction of Solvent Spectrum with Iterative Use of PARAFAC). Zajmentes, de a valóságot jobban közelítő elúciós profilok alkalmazása esetén a PARAFAC2 módszert kellett iteratívan alkalmazni a kielégítő kvalitatív és kvantitatív eredmények eléréséhez. Sort kerítettünk a mért adattömb értelemszerű megcsonkítására, vagyis kihagytuk azokat az időszakaszokat a hozzájuk tartozó spektrummal együtt, amelyekhez csak eluens spektrum rendelhető (konkrétan: az első kromatográfiás csúcs előtti és az utolsó kromatográfiás csúcs utáni spektrumsorozatok). A redukált adattömbbel végrehajtott kétlépéses OSSS-IU-PARAFAC2 eljárás az elúciós és koncentrációprofilokat már jól visszaadta, de a kívánt spektrumokat csak egy újabb szubjektív elimináció végrehajtása után kaptuk meg.

T17. *MCR-ALS módszert alkalmaztam a meloxikám-mannit binér rendszer fiziko-kémiai jellemzésére* [P17]. Vizsgáltuk a meloxikám (ME) és a mannit (MA) mint hordozó anyag fizikai és olvadék keverékének (PM - physical mixture, ill. MP - melted product) kioldódási viselkedését. A kemometriai feladat ennél a problémánál az volt, hogy megválaszoljam azt a kérdést miszerint a röntgendiffrakciós mérések alapján értelmezni tudjuk-e az olvadékforma legelőnyösebb kioldódási viselkedését. MCR-ALS módszert alkalmaztam és az eredményekből arra következtettem, hogy az olvadékban egy új forma alakult ki (blend), ami 1:10 arány mel-

lett nagyobb mértékben van jelen, mint 3:7 arány esetén. Megfigyeltük, hogy a fizikai keverékben elhanyagolható, míg a cseppentő módszerrel kapott mintában a legnagyobb mértékben található az új alakulat. A legjobb kioldódást a cseppentő módszerrel kapott termék mutatta.

T18. *Görbeillesztés nélküli komponensprofil-kinyerő módszert alkalmaztam meloxikám-mannit binér rendszerek kioldódási adatainak kiértékelésére* [P18]. Kioldódási vizsgálatok többváltozós kemometriai módszerekkel történő kiértékelése jelent meg az Analytical Chemistry-ben 2006-ban. Úgy találtuk, hogy Wiberg és Hultin által használt kemometriai eljárások nem alkalmazhatók az általunk vizsgált meloxikám-mannit fizikai keverék rendszer kioldódási eredményeinek értékelésére. Többek között azért nem, mert az eltérő részecskeméretű meloxikámnak eltérő kioldódási görbét mértünk (itt arra kell figyelni, hogy az eltérő méretű meloxikám UV-Vis spektruma azonos ugyan, de a kioldódási görbék eltérőek), ezért pl. a PARAFAC nem alkalmazható (a profil nem azonos az összes mintára). Sajnos a PARAFAC2 alkalmazása sem adott kielégítő eredményeket. Ezért az adatsort kettébontottam a kétféle részecskeméretnek megfelelően és az általam kifejlesztett (és korábban részletezett) SMCR módszert megvalósító Matlab programot használtam. Megállapítottuk, hogy az ME2 jelzéssel el látott részecskeméretű meloxikám kioldódása kétszer jobb, mint a többi mintáé. A spektrális sávmegoldások az ME1 és az ME2 adatkészletekre nagyon hasonlóak lettek, igazolva eljárásunk következetességét. Legjobb tudomásunk szerint ez az első beszámoló az SMCR módszerek alkalmazásáról háromkomponensű kioldódási adatokra. A sávmegoldások természetesen csak kvalitatív képet tudnak adni, de a legtöbb esetben a sávok elég szűkek ahhoz, hogy a mérési hibán belül kvantitatív következtetéseket is levonhassunk.

T19. *Definiáltam és értelmeztem a lokális rang-vesztés fogalmát* [P19]. A lokális rang információ kulcsszerepet játszik a többkomponensű kémiai rendszerek kapcsán felmerülő görbefelbontásnál. Lokális rang kényszerfeltételt alkalmazva a forgatási bizonytalanság mértéke csökkenthető és bizonyos esetekben meg is szüntethető minek következtében egyértelmű felbontást kapunk. A szakirodalomban szokásosan azt feltételezik, hogy a lokális rang egyenlő, vagy nagyobb (zaj hatása miatt) mint a kémiai komponensek száma a vizsgált (soronkénti vagy oszloponkénti) mátrixrészletekben. Megállapítottuk, hogy a lokális rang olyan matematikai fogalom, amely esetleg nincs teljes összhangban a kémiai komponensek számával. Így a forgatási bizonytalanság csökkentése érdekében a lokális rang kényszerfeltétel alkalmazása az MCR módszerekben helytelen megoldásokhoz vezethet! Ez a probléma az általunk bevezetett szak kifejezés szerint a „lokális rang-vesztés” (local rank deficiency) következménye. Felhívtuk a figyelmet arra, hogy ha a rang-vesztés csak a választott ablakban fordul elő és nem a teljes mátrixban, akkor pusztán a lokális rang alapján csökkenteni a megengedett megoldás-tartományt kockázatosnak tűnik, hacsak nincs megbízható kémiai információnk az adott ablakra vonatkozóan. Tehát ebben a

munkánkban igazoltuk, hogy az adatmátrixok mikrostruktúrája, amit a Borgen grafikon tár fel, alapvető a lokális rang értelmezése terén. Természetes folytatása ez korábbi eredményünknek, miszerint Manne görbefelbontási tételei bár szükségesek, de általában nem elégségesek, így az általunk előzőleg bevezetett és részletesen bemutatott adat-alapú egyértelműséget (data-based uniqueness) kell használnunk.

T20. *Dualitás alapú általánosított rang-eltüntető algoritmust fejlesztettem, kalibrációs mintában nem szereplő zavaró komponensek esetében, koncentráció meghatározáshoz, realisztikus hibabecsléssel* [P20]. Az analitikai kémia egyik leggyakrabban felmerülő problémája a komplex kémiai minták elemzése úgy, hogy a (zavaró) komponensek többségét nem veszik figyelembe a kalibrációs modellalkotás során. Ezekben az esetekben kívánatos, hogy mennyiségi információt tudjunk szerezni egy kiemelt komponensről (analit) anélkül, hogy a minta többi összetevőjével foglalkoznánk. Azt a tulajdonságot, hogy ismeretlen összetevők jelenlétében is meg tudjuk határozni az analit mennyiségét, másodrendű előnynek (second-order advantage) nevezzük. A háromutas adatok elemzésére többféle módszer létezik, amelyek többsége a másodrendű előnyök felhasználásával oldja meg a kalibrációs feladatot, ezen módszerek közül az általánosított rang-eltüntető módszer (Generalized Rank Annihilation Method, GRAM) nem iteratív és két bilineáris adatmátrixszal dolgozik: . Egy újszerű algoritmust mutattunk be a másodrendű előny elérésére, amely a dualitáson alapul. A javasolt módszerhez a matematikai képletek mellett informatív geometriai szemléltetést is adtunk. A szimulációs és kísérleti adatok felhasználásával a vizsgálat eredményei azt mutatták, hogy a dualitás elve alapján megbízható számítható, amelyből könnyen és gyorsan meghatározható a tényleges koncentráció. A javasolt módszer további előnye, hogy adatvezérelt hibaintervallumokat tudunk számolni. Ezzel szemben a meglévő GRAM algoritmusok önmagukban nem tudnak ilyen jellegű információt szolgáltatni, hibaszámítás csak számítógép-intenzív módszerek (pl. Jackknife, Bootstrap stb.) segítségével végezhető.

Hivatkozások

- P1. Rajkó, R.; István, K. Analytical solution for determining feasible regions of Self-Modeling Curve Resolution (SMCR) method based on computational geometry. *J. Chemom.* **2005**, *19*, 448–463. <https://doi.org/10.1002/cem.947>
- P2. Rajkó, R. Natural duality in minimal constrained self modeling curve resolution. *J. Chemom.* **2006**, *20*, 164–169. <https://doi.org/10.1002/cem.999>
- P3. Rajkó, R. Computation of the range (band boundaries) of feasible solutions and measure of the rotational ambiguity in self-modeling/multivariate curve resolution. *Anal. Chim. Acta* **2009**, *645*, 18–24. <https://doi.org/10.1016/j.aca.2009.04.033>
- P4. Rajkó, R. Studies on the adaptability of different Borgen norms applied in self-modeling curve resolution (SMCR) method. *J. Chemom.* **2009**, *23*, 265–274. <https://doi.org/10.1002/cem.1221>

- P5. Rajkó, R. Additional knowledge for determining and interpreting feasible band boundaries in self-modeling/multivariate curve resolution of two-component systems. *Anal. Chim. Acta* **2010**, 661, 129–132.
<https://doi.org/10.1016/j.aca.2009.12.025>
- P6. Beyramysoltan, S.; Rajkó, R.; Abdollahi H. Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method. *Anal. Chim. Acta* **2013**, 791, 25–35.
<https://doi.org/10.1016/j.aca.2013.06.043>
- P7. Beyramysoltan, S.; Abdollahi H.; Rajkó, R. Newer developments on self-modeling curve resolution implementing equality and unimodality constraints. *Anal. Chim. Acta* **2014**, 827, 1–14.
<https://doi.org/10.1016/j.aca.2014.03.019>
- P8. Rajkó, R.; Abdollahi H.; Beyramysoltan, S.; Omidikia, N. Definition and detection of data-based uniqueness in evaluating bilinear (two-way) chemical measurements. *Anal. Chim. Acta* **2015**, 855, 21–33.
<https://doi.org/10.1016/j.aca.2014.12.017>
- P9. Golshan, A.; Abdollahi, H.; Beyramysoltan, S.; Maeder, M.; Neymeyr, K.; Rajkó, R.; Sawall, M.; Tauler, R. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modeling analyses of multivariate data. *Anal. Chim. Acta* **2016**, 911, 1–13.
<https://doi.org/10.1016/j.aca.2016.01.011>
- P10. Rajkó, R.; Omidikia, N.; Abdollahi, H.; Kompany-Zareh, M. On uniqueness of the non-negative decomposition of two- and three-component three-way data arrays. *Chemom. Intell. Lab. Syst.* **2017**, 160, 91–98.
<https://doi.org/10.1016/j.chemolab.2016.12.001>
- P11. Omidikia, N.; Abdollahi, H.; Kompany-Zareh, M.; Rajkó, R. Analytical solution and meaning of feasible regions in two-component three-way arrays. *Anal. Chim. Acta* **2016**, 939, 42–53.
<https://doi.org/10.1016/j.aca.2016.08.026>
- P12. Omidikia, N.; Ghaffari, M.; Rajkó, R. Sparse non-negative multivariate curve resolution: L0, L1, or L2 norms? *Chemom. Intell. Lab. Syst.* **2020**, 199, 103969.
<https://doi.org/10.1016/j.chemolab.2020.103969>
- P13. Faber, N.M.; Rajkó, R. How to avoid over-fitting in multivariate calibration – the conventional validation approach and an alternative. *Anal. Chim. Acta* **2007**, 595, 98–106.
<https://doi.org/10.1016/j.aca.2007.05.030>
- P14. Rajkó, R. Some surprising properties of Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) algorithms. *J. Chemom.* **2009**, 23, 172–178.
<https://doi.org/10.1002/cem.1228>
- P15. Rajkó, R. Comments on “Near-Infrared Hyperspectral Unmixing Based on a Minimum Volume Criterion for Fast and Accurate Chemometric Characterization of Counterfeit Tablets”. *Anal. Chem.* **2010**, 82, 8750–8752.
<https://doi.org/10.1021/ac101962x>
- P16. István, K.; Rajkó, R.; Keresztury, G. Towards the solution of the eluent elimination problem in HPLC/IR measurements by chemometric methods. *J. Chromatogr. A* **2006**, 1104, 154–163.
<https://doi.org/10.1016/j.chroma.2005.11.131>
- P17. Nassab, P.R.; Rajkó, R.; Szabó-Révész, P. Physicochemical characterization of meloxicam-mannitol binary systems. *J. Pharm. Biomed. Anal.* **2006**, 41, 1191–1197.
<https://doi.org/10.1016/j.jpba.2006.02.055>
- P18. Rajkó, R.; Nassab, P.R.; Szabó-Révész, P. Self-modeling curve resolution method applied for the evaluation of dissolution testing data: A case study of meloxicam-mannitol binary systems. *Talanta* **2009**, 79, 268–274.
<https://doi.org/10.1016/j.talanta.2009.03.068>
- P19. Akbari Lakeh, M.; Rajkó, R.; Abdollahi, H. Local rank deficiency caused problems in analyzing chemical data. *Anal. Chem.* **2017**, 89, 2259–2266.
<https://doi.org/10.1021/acs.analchem.6b03134>
- P20. Tavakkoli, E.; Abdollahi, H.; Rajkó, R. New duality based generalized rank annihilation algorithm for determining analyte concentration with realistically estimated error level for practical data sets. *Chemom. Intell. Lab. Syst.* **2020**, 203, 104058.
<https://doi.org/10.1016/j.chemolab.2020.104058>

Self-modeling curve resolution in analytical chemistry

Róbert Rajkó's scientific endeavors have evolved significantly over the course of his career, encompassing a wide range of analytical and statistical methodologies. His research trajectory can be divided into several distinct phases.

Early Career Focus: Initially, Rajkó concentrated on the evaluation of univariate and multivariate measurement data, as well as the mathematical statistical analysis of related calibration techniques. During this period, he employed both robust and traditional parameter estimation methods to enhance the accuracy and reliability of analytical results.

Mid-Career Developments: As his research progressed, Rajkó shifted his focus towards:

1. Variable selection techniques
2. Quantitative Structure-Activity Relationship (QSAR) studies

These areas of study allowed him to delve deeper into the relationships between chemical structures and their biological activities, contributing to the field of chemometrics.

Recent Research Interests: In recent years, Rajkó has turned his attention to the theoretical examination and practical applicability of hidden-variable data evaluation and modeling procedures. His work primarily involves:

- Two-dimensional (matrix) data arrays
- Higher-dimensional (tensor) data structures

This research has significant implications for the analysis of complex, multi-dimensional datasets in various scientific disciplines.

Doctoral Dissertation and Expertise: Rajkó's doctoral dissertation for the Hungarian Academy of Sciences showcases his expertise in the field of self-modeling curve resolution (SMCR), a pivotal area within chemometrics. The dissertation presents a coherent and concise compilation of 20 selected publications from 2005 to 2020, focusing on curve-fitting-free component profile extraction. His contributions to this field have earned him international recognition as an expert. Rajkó has published numerous scientific papers in leading journals, both as a sole author and in collaboration with prominent chemometricians, doctoral students, and young postdoctoral researchers from around the world.

The Borgen-Rajkó Plots: A significant milestone in Rajkó's career was the (re)creation of the Borgen-Rajkó plots in 2005. This innovation laid the foundation for numerous subsequent achievements in the realm of analytical measurement problems that can only be solved through band solutions. The term "Borgen-Rajkó plots" has gained traction in the scientific community, with an increasing number of independent authors adopting and referencing this concept in their work. In conclusion, Róbert Rajkó's scientific contributions have significantly advanced the field of chemometrics, particularly in the areas of data analysis, curve resolution, and component profile extraction. His work continues to influence and shape the landscape of analytical chemistry and related disciplines.