

ÚJ PARAMÉTERBECSLŐ ELJÁRÁSOK
ALKALMAZÁSA AZ ANALITIKAI KÉMIAI
MÉRÉSKIÉRTÉKELÉS GYAKORLATÁBAN

RAJKÓ RÓBERT
okleveles vegyész

EGYETEMI DOKTORI ÉRTEKEZÉS

SZEGED
1995

Sonnet-to Science

*Science! true daughter of Old Time thou art!
Who alterest all things with thy peering eyes.
Why preyest thou thus upon the poet's heart,
Vulture, whose wings are dull realities?
How should he love thee? or how deem thee wise,
Who wouldst not leave him in his wandering
To seek for treasure in the jewelled skies,
Albeit he soared with an undaunted wing?
Hast thou not dragged Diana from her car?
And driven the Hamadryad from the wood
To seek a shelter in some happier star?
Hast thou not torn the Naiad from her flood,
The Elfin from the green grass, and from me
The summer dream beneath the tamarind tree?*

/Edgar Allan Poe/

Szonett a Tudományhoz

*Tudomány, te, a vén idő szülötte,
Ha mire nézel, más lesz az, sötét.
Szárnyad a zord való, mordképű ölyv, te,
Miért fosztod ki a költő szívét?*

*Hogyan becsüljön téged, hogy szeressen,
Ha nem hagyod, hogy bolygva a merő
Ékköves égen, kincseket keressen?
Hisz oly bátran szállt az egekre ő!*

*Nem szállítottad földre Dianát,
Nem verted fák közül ki a driádot,
Hogy költözzön egy jobb csillagra át?*

*Nem a habok közül a szép najádot,
A pázsitról a tündért s magamat
Az álmaimból nyári fák alatt?*

/fordította Komlós Aladár/

Tartalomjegyzék

Bevezetés	1
1 Az irodalom áttekintése	5
1.1 Paraméterbecslési modellek	6
1.1.1 Paraméterbecslés elsőfajú modellhiba esetén	7
1.1.1.1 Sztochasztikus modell	7
1.1.1.2 Fuzzy halmazokon alapuló modell	19
1.1.2 Paraméterbecslés másodfajú modellhiba esetén	22
1.1.3 Paraméterbecslés harmadfajú modellhiba esetén	25
2 Kalibrációs függvények paramétereinek becslése	26
2.1 Egyváltozós kalibrációs függvények robusztus becslése	26
2.2 Többváltozós kalibrációs függvények robusztus becslése	32
2.3 Lineáris fuzzy regresszió	37
2.3.1 Az eredeti egyváltozós lineáris fuzzy regresszió módosítása	38
2.3.2 Lineáris fuzzy regresszió általánosítása geometriai úton	41
2.4 Fuzzy lineáris regresszió	47
3 Kalibráció során nyert adatok jellemzése	53
3.1 Konfidencia intervallum Gauss-eloszlás teljesülése esetén	53
3.2 Konfidencia intervallumok szerkesztése bootstrap módszerrel	55
4 Karl Fischer titrálás automatikus vezérlése	67
4.1 A Karl Fischer titrálásról általában	67
4.2 Az új vezérlő algoritmus leírása	68
5 Összefoglalás	74

Függelék	77
A disszertációval kapcsolatos közlemények és előadások felsorolása	84
Köszönetnyilvánítás	89
Irodalomjegyzék	89

Ábrák jegyzéke

0.1	Az analitikai kémiai rendszerek felépítése, az analitikai információ és a kémiai információ kapcsolata	4
1.1	Az $L_1(a, d)$, $L_2(b, e)$ és az $L_\infty(c, f)$ vektornormák 2 ill. 3 dimenziós alakjai	11
1.2	Fuzzy paraméterek, mint háromszög alakú fuzzy számok . . .	24
2.1	Otto és Bandemer által javasolt gráf a helyi legjobb fuzzy közelítéshez	40
2.2	Két általunk javasolt gráf, melyekkel elkerülhető a szövegben említett probléma	40
2.3	A fuzzy kimenő adatok e_i terjedelme és az y_i tagsági értékek fordított arányú kapcsolatának illusztrálása	48
2.4	Fuzzy lineáris regresszióval kapott kalibrációs egyenesek ábrái	51
3.1	Séma a bootstrap eljárás alkalmazására	56
3.2	3000 bootstrap mintából meghatározott a^* és b^* paraméterek közötti korreláció	58
3.3	3000 bootstrap mintából számított a^* paraméter eloszlása és a 90%-os megbízhatósági szintű konfidencia intervallum . . .	59
3.4	Az 1. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 5.0$ -nél $y = 6.5$)	62
3.5	A 3. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 5.0$ -nél $y = 6.5$)	63
3.6	Az 1. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 3.0$ -nál $y = 4.5$)	63
3.7	A 2. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 3.0$ -nál $y = 4.5$)	64

3.8	Az 1. mintára vonatkozó jószolt koncentráció eloszlása (a kiugró érték $x = 0.0$ -nál $y = 1.5$)	64
3.9	A 3. mintára vonatkozó jószolt koncentráció eloszlása (a kiugró érték $x = 0.0$ -nál $y = 1.5$)	65
3.10	Talajvíz jószolt Mg koncentrációjának eloszlása különböző paraméterbecslő eljárások alkalmazása mellett	66
4.1	Karl Fischer titrálás során kapott tipikus titrálási görbe bi-amperometriás (dead stop) végpontjelzéssel	69

Táblázatok jegyzéke

2.1	A kalibrációs adatok ICP-AES mérésekhez	31
2.2	A becsült paraméterek értékei a szövegben említett paraméterbecslő módszerekkel meghatározva	31
2.3	A vizsgált paraméterbecslő eljárások tulajdonságai	32
2.4	Szimulált adatokra illesztett egyenes paraméterei és a számított értékek eltérései	41
2.5	A Bűnügyi Technikai Intézetben előállított standardok által meghatározott kalibrációs egyenes paraméterei és a számított értékek százalékos eltérései	42
2.6	Az előbbi táblázatban szereplő standardok és az OMH által előállított standardok felhasználásával kapott kalibrációs egyenes paraméterei és a számított értékek százalékos eltérései	43
2.7	Atomabszorpciós spektrometriásan mért adatsor az irodalomból átvéve	49
2.8	A fuzzy lineáris regresszió eredménye és a módosított számításmenet illusztrálása	50
3.1	Szimulált kalibrációs adatok, valamint a szimulált mérési jelek	62
3.2	Talajvíz Mg koncentrációjának ICP-AES módszerrel történő meghatározásához használt kalibrációs adatsor	65
3.3	Talajvíz Mg koncentrációjának jóslt értékei 90%-os megbízhatósági szintű konfidencia intervallummal, különböző paraméterbecslő eljárások alkalmazása mellett	66

Bevezetés

A gazdasági élet egyre több területén van és lesz szükség az analitikai kémiai mérések által szolgáltatott minőségi és mennyiségi jellemzők felhasználására. A piacokon csak a jó minőségű árut lehet haszonnal eladni és ezek előállításának folyamatában a minőségbiztosítás meghatározó szerepet kapott [85,86,87,88,89,90,91,92]. A minőség betartásához és ellenőrzéséhez tehát nagy számú analitikai meghatározás szükséges, melyeket legtöbbször automatikus sorozatmérések keretében hajtanak végre. A mintavétel és a mintaelőkészítés az analitikai mérés legérzékenyebb pontja, de rögtön ezek után a kalibráció helyes végrehajtása és alkalmazása következik. A nemrégiben Burger Kálmán akadémikus által írt "A mennyiségi analízis alapjai: kémiai és műszeres elemzés" című igazán nagyszerű könyve kellő mélységű bevezetést ad az alkalmazott módszerek kémiai, fizikai-kémiai valamint műszertechnikai kivitelezésük háttéréről, de sajnálatos módon a nyert adatok feldolgozásához ill. kiértékeléséhez szükséges kemometriai módszereket meg sem említi.

Bujtás Piroska és Leisztner László az "Analitikai mérési eredmények minőségbiztosítása" című munkájukban az egzakt tárgyalás mellőzésével, a gyakorló analitikusok általános matematikai és statisztikai felkészültségéhez alkalmazkodva vázolják a szükséges kemometriai módszereket. Jelen disszertáció ezen két mű kiegészítőjeként a kalibráció kérdéskörében próbál gyakran új nézőpontból szemlélt régi és új eljárások bemutatásával az analitikusok segítségére sietni, hogy a jól elvégzett kémiai kalibráció a legmegfelelőbb kemometriai módszerrel kerüljön kiértékelésre. Ezáltal a lehető legtöbb és legmegbízhatóbb információk birtokába juthatunk.

Annak illusztrálására, hogy miért kell a klasszikus, több évtizede elfogadott kiértékelő eljárásokat felülvizsgálni és olykor újakkal helyettesíteni, vizsgáljuk meg a normalitás (igen gyakran minden ellenőrzés nélküli) elfogadásának helyességét.

A normális vagy Gauss-eloszlás (a következőkben inkább a Gauss-elosz-

lás elnevezést használjuk, mivel mint kiderül, a többi eloszlás nagyon is nem abnormális) jogos használatának támaszát gyakran a centrális határeloszlástételekben látják biztosítottak. Persze sokan elfelejtik, hogy bizonyos feltételeknek teljesülniük kell a tételek alkalmazásához. A legegyszerűbb megfogalmazásban nagyon sok azonos eloszlású, véges szórású véletlen változó átlaga lesz Gauss-eloszlású. Napjaink műszereinek pontossága azonban, egyre inkább kis számú, eltérő eloszlású hibatagokból tevődik össze. A tétel Ljapunov-féle alakja még menthetné a helyzetet, hiszen az a nem azonos eloszlásból származó adatok átlagaira határozza meg a normális eloszlást, ha az eloszlások első, második és harmadik momentuma létezik, valamint a harmadik momentumoknak összességükben egyre kisebbeknek kell lenniök a szórásokhoz képest. Nem szabad figyelmen kívül hagyni azonban az olyan eloszlásokat sem, melyeknek nincs véges szórásuk (pld. Cauchy-eloszlás). A fő probléma azonban az, hogy a mérőműszerek által szolgáltatott jelek többsége is csak néhány különböző eloszlás kombinációja, valamint a gyakorlatban mindig véges mintaszámmal dolgozunk, így érvényes Cramér nevezetes tétele, mely szerint a normális eloszlásnak minden faktora normális eloszlású. Ez azt jelenti, hogy véges számú valószínűségi változó összege csak akkor lehet Gauss-eloszlású, ha már mindegyik eleve Gauss-eloszlású volt. Ilyet a gyakorlatban megkövetelni nem lehet, és így a Gauss-eloszlás mellett (a gyakorlati tapasztalatokkal egyetértésben) az elméleti megfontolások sem szolgáltatnak kellő támaszt. Álljon itt Poincaré [108] szarkasztikus megjegyzése ezzel kapcsolatban:

Mindenki hisz a normális eloszlás univerzalitásában:

a FIZIKUSOK azért, mert azt hiszik, hogy a matematikusok igazolták logikai szükségyszerűségét,

a MATEMATIKUSOK pedig azért, mert úgy hiszik, hogy a fizikusok laboratóriumi méréseikkel bizonyították azt.

A gyakorlatban Clancey [30] csaknem 50000 kémiai analízis eredményét megvizsgálva 250 féle eloszlást tudott azonosítani. Ezek 10-15 %-a volt csak egyértelműen Gauss-eloszlású és kb. 50 %-ot tettek ki a szimmetrikus eloszlások. Így a gyakorlatban is cáfolni lehetett az *a priori* meglévő Gauss-eloszlást.

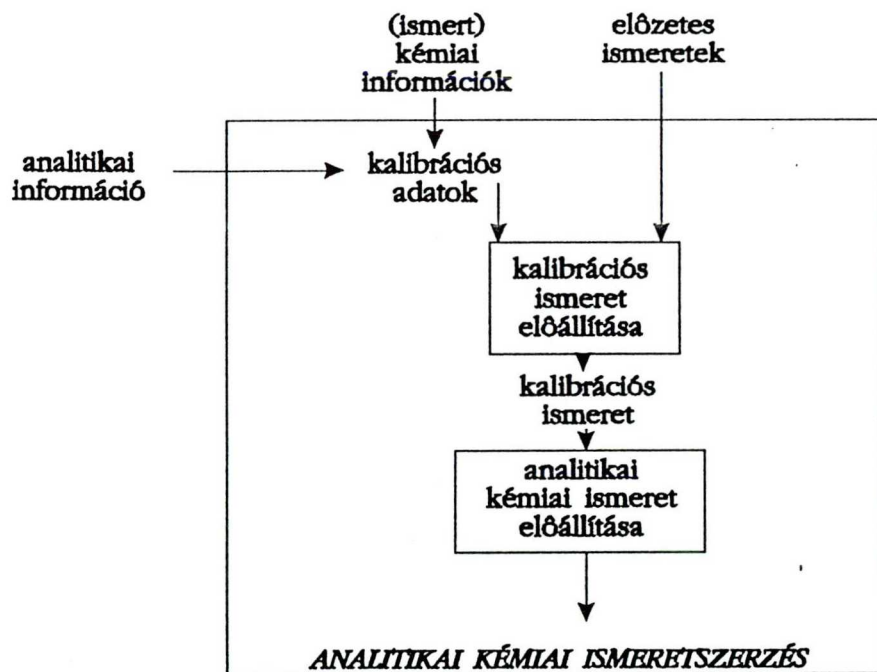
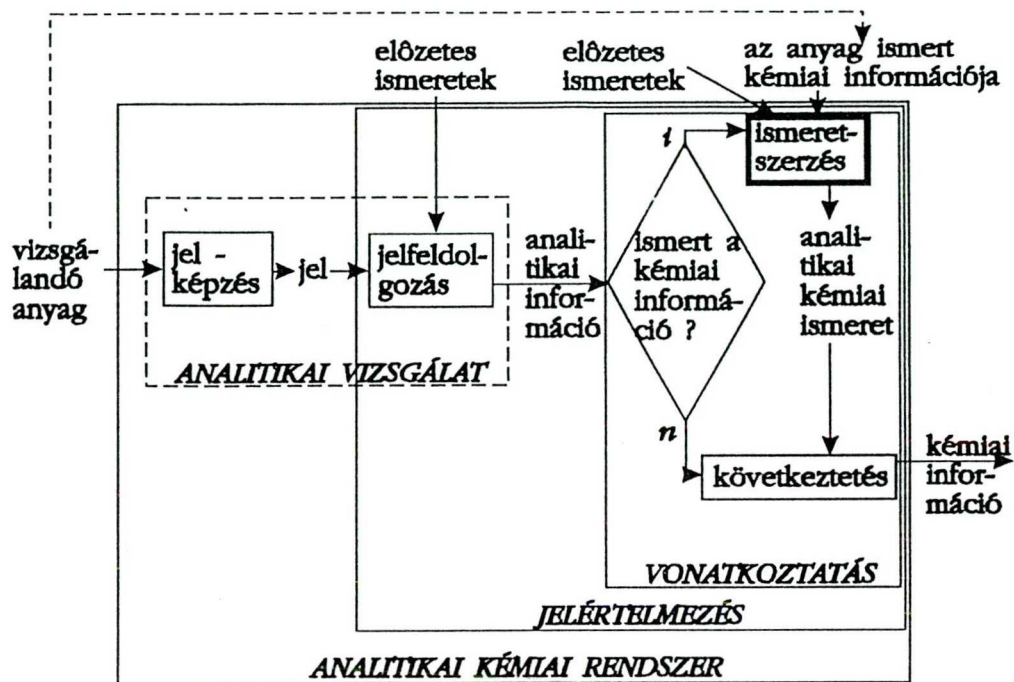
A kémiai mérések kiértékelésével, a kémiai méréselmélettel a kemometria (chemometrics), mint új tudományág foglalkozik. A Chemometrics Society által adott meghatározás szerint a kemometria

- az optimális mérési folyamatok és kísérletek tervezésére, vagy kiválasztására
- kémiai adatok elemzésével maximális kémiai információ előállítására

szolgáló matematikai és statisztikai módszereket alkalmazó kémiai elvek összessége.

Ma már számos kemometriával foglalkozó könyv [20,84,116] és összefoglaló tanulmány [1,17,21,22,46,98,129] áll rendelkezésünkre az alapvető módszerek közötti eligazodás megkönnyítésére.

Veress [133] monográfiájában a 0.1 ábrán látható módon építi fel az információszerzés folyamatát. Dolgozatomban az analitikai kémiai ismeretszerzésben használt kalibráció témakörével foglalkozom. A kalibrációs ismeret legtöbbször az alkalmazott kalibráló függvény paramétereinek meghatározását jelenti, azonban a regressziós módszer robusztussága (mely fogalmat a későbbiekben részletesen elemzünk) általában kívánatos kell legyen, hiszen a matematikai modellt nem dolgozhatjuk ki az egyelőre nem ismeretes általánosabb körülményekhez, azonban az szükséges lehet, hogy a feltételektől való kis eltérés csak kis torzítást okozzon a végeredményekben. Az analitikai kémiai mérések kiértékelése jórészt empirikus modellválasztásokon alapulnak, így nagy szükség van a modell érzéketlen viselkedésére. Robusztus tulajdonsággal bíró eljárásra példaként említhető még a kvantummechanikában alkalmazott perturbációszámítás, amit arra az esetre dolgoztak ki, amikor a vizsgált fizikai rendszer Hamilton-operátora csak közelítőleg számítható, azonban az csak kicsit tér el egy egzaktul megoldható probléma Hamilton-operátorától és várható, hogy a megoldás is csak kicsit tér el a pontosan megoldható feladat megoldásától.



0.1 ábra: Az analitikai kémiai rendszerek felépítése, az analitikai információ és a kémiai információ kapcsolata

1. fejezet

Az irodalom áttekintése

A kalibráció során alkalmazható paraméterbecslő eljárások irodalomi áttekintése egy újszemponútú rendszerezés keretében Bard [9] útmutatásainak továbbgondolása alapján történik. Felhasználtuk még a Beck és Arnold [14] által összeállított és részletesen tárgyalt alapvető statisztikai feltevéseket, melyeket 8 csoportba sorolhatunk és teljesülésük ill. nem teljesülésük szerint kapcsolat teremthető az általunk elképzelt csoportosítással:

1. Additív mérési hiba
 - 0 nem additív
 - 1 igen, additív
2. Nulla várható értékű mérési hiba
 - 0 nem nulla várható értékű
 - 1 igen, nulla várható értékű
3. Állandó varianciájú hibák
 - 0 nem állandó variancia
 - 1 igen, állandó variancia
4. Nem korrelált hibák
 - 0 korrelált hibák
 - 1 igen, nem korreláltak a hibák
5. A hibák normális eloszlása

- 0 nem normális eloszlású hibák
 - 1 igen, normális eloszlású hibák
6. A hibák leírhatók ismert statisztikai paraméterekkel
- 0 a hibák kovariancia mátrixa csak egy szorzó faktortól eltekintve ismert
 - 1 igen, a hibák kovariancia mátrixa teljesen ismert
7. A független változók hiba nélküliek
- 0 nem, a független változók hibákkal terheltek
 - 1 igen, a független változók hiba nélküliek
8. A hiba eloszlások paraméterei nem véletlenszerűek és nincs rájuk vonatkozó előzetes információnk.
- 0 nem, véletlen paraméterek, de nincs előzetes információ
 - 1 igen, nemvéletlen paraméterek és nincs előzetes információ
 - 2 normális eloszlású paraméterek ismert várható értékkel és kovariancia mátrixszal, minden mérés ugyanebből az eloszlásból származik
 - 4 nemvéletlen paraméterek, de ismeretesek bizonyos normális szubjektív előzetes információk

1.1 Paraméterbecslési modellek

A kalibrációs függvény paramétereinek meghatározása egy paraméterbecslési modell fellállítását igényli:

$$g(\mathbf{z}, \mathbf{p}) = 0, \quad (1.1)$$

ahol a $g(\cdot)$ implicit függvény teremt kapcsolatot a $\mathbf{z}$ változók között, a becsülendő $\mathbf{p}$ paraméterekkel. A $\mathbf{z}$ adatok két részre oszthatók: $\mathbf{v}$ melyek hibamentesen mérhetők és $\mathbf{w}$ melyek csak hibával terhelten mérhetők.

Az ilyen modell pontosabb megfogalmazásához esetenként azt kell megvizsgálni, mi mindent tudunk a kiválasztott kalibrációs rendszerről. Sokkal jobban hiszünk a kalibrációs függvény alakjában, mint méréseink pontosságában? Vagy éppen fordítva, méréseink pontosak, azonban a függvény nem írja le tökéletesen a mért folyamatot? Talán mind a mérés, mind a függvény terhelt bizonyos előre nem kalkulálható zavaró tényezőkkel? Az alkalmazott modell típusa függ az előző kérdésekre adott válaszoktól.

- a, Tételezzük fel, hogy az adatok mérési hibával terhelve, de az alkalmazott függvény pontosan leírja a kalibrációs folyamatot:

$$g(\mathbf{v}, \mathbf{w}, \mathbf{p}) = 0, \quad (1.2)$$

A fenti megfogalmazáson alapuló modellhibát elsőfajú modellhibának nevezzük.

- b, Tegyük fel, hogy a kalibrációs függvény a bizonytalan számos hatás figyelmen kívül hagyása miatt, ehhez képest a mérések hiba nélkül kivitelezhetőek:

$$g(\mathbf{v}, \mathbf{p}) = \gamma, \quad (1.3)$$

Az így jelentkező modellhiba a másodfajú modellhiba.

- c, Mind a mérés, mind a függvény felállítása csak bizonytalansággal végezhető el:

$$g(\mathbf{v}, \mathbf{w}, \mathbf{p}) = \gamma, \quad (1.4)$$

Ennek során a harmadfajú modellhibát kell kezelnünk.

1.1.1 Paraméterbecslés elsőfajú modellhiba esetén

1.1.1.1 Sztochasztikus modell

Kendall [72,73] adta meg az (1.2) függvény paramétereinek meghatározásához szükséges feltételrendszert, feltéve hogy (1.2) az $y = f(x, \mathbf{p})$ explicit alakra hozható:

- a. Az $y = f(x, \mathbf{p})$ folytonos függvény határos és zárt (azaz kompakt) a $B \times A \supset \mathbb{R}^2 \times \mathbb{R}^{q+1} (q \geq 0)$ tartomány felett.
- b. Az $y = f(x, \mathbf{p})$ függvény Taylor-sorba fejthető a tartomány minden pontjában, ahol f definiálva van.
- c. A mérési hibák szuperponáltak (additív hibák)
 $y_j(\text{megfigyelt érték}) = y_j^0(\text{valódi érték}) + u_j^y(\text{mérési hiba})$ és
 $x_j(\text{megfigyelt érték}) = x_j^0(\text{valódi érték}) + u_j^x(\text{mérési hiba}).$
- d. A valódi értékek kielégítik az explicit függvénykapcsolatot, azaz
 $y_j^0 = f(x_j^0, \mathbf{p}).$
- e. A mérési hibák az U_j^k folytonos valószínűségi változók realizációi, várható értékük $E[U_j^k] = 0$, k itt és a továbbiakban x -et vagy y -t helyettesíti.

Az a.-e. feltételek még mindig nem elegendőek a probléma megoldásához. További megszorításokat kell tennünk U_j^k -re vonatkozóan:

f. Az U_j^k -k sztochasztikusan függetlenek.

g. Az U_j^k -k ismert, véges varianciával rendelkeznek $\infty > D[U_j^k] > 0$.

Ekkor a $(\hat{\mathbf{p}}, \hat{x}_j, \hat{y}_j)$ megoldást a következő nemlineáris normál egyenletrendszer adja:

$$QF \equiv \sum_{j=1}^n \left((y_j - f(x_j, \mathbf{p}))^2 \cdot D^{-2}[U_j^y] + (x_j - \check{x}_j)^2 \cdot D^{-2}[U_j^x] \right). \quad (1.5)$$

QF minimálissá válásához a következő egyenleteknek kell teljesülniük:

$$-\frac{1}{2} \frac{\partial QF}{\partial x_j} = 0, \quad -\frac{1}{2} \frac{\partial QF}{\partial y_j} = 0, \quad -\frac{1}{2} \frac{\partial QF}{\partial p_i} = 0. \quad (1.6)$$

A fenti normál egyenletrendszer általában nemlineáris, így megoldásához numerikus úton juthatunk el, pld. Newton iteráció segítségével. A nemlineáris paraméterbecslő eljárásokról jó összefoglaló található [131,132] irodalmakban.

Vizsgáljuk meg a legegyszerűbb esetet, amikor a függvénykapcsolat lineáris $y = a_0 + a_1 x$. Milyen további feltételeknek kell teljesülniük?

h. A mérési hibák homoszkedasztikusak, azaz $D^2[U_j^k] = \beta^k$.

$$\hat{x}_j = \frac{\hat{a}_1(y_j - \hat{a}_0) + \frac{\beta^y}{\beta^x} x_j}{\frac{\beta^y}{\beta^x} + \hat{a}_1^2} \quad (1.7)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \quad (1.8)$$

$$\hat{a}_1 = \frac{\sum_{j=1}^n (y_j - \bar{y})^2 - \frac{\beta^y}{\beta^x} \sum_{j=1}^n (x_j - \bar{x})^2}{2 \sum_{j=1}^n (y_j - \bar{y})(x_j - \bar{x})} + \frac{\sqrt{4 \frac{\beta^y}{\beta^x} \sum_{j=1}^n ((y_j - \bar{y})(x_j - \bar{x}))^2 + (\sum_{j=1}^n (y_j - \bar{y})^2 - \frac{\beta^y}{\beta^x} \sum_{j=1}^n (x_j - \bar{x})^2)^2}}{2 \sum_{j=1}^n (y_j - \bar{y})(x_j - \bar{x})}, \quad (1.9)$$

$$\text{ahol } \bar{y} = \sum_{j=1}^n \frac{y_j}{n} \text{ és } \bar{x} = \sum_{j=1}^n \frac{x_j}{n}.$$

$\hat{a}_1$ csak abban az esetben számítható, ha a $\frac{\beta^y}{\beta^x}$ arány ismert. Ha $\beta^y = \beta^x$, akkor ortogonális regressziós problémáról beszélünk. Ha $\beta^x = 0$, vagyis az i . feltétel teljesülése esetén direkt kalibrációról, a $\beta^y = 0$ esetén indirekt kalibrációról beszélhetünk. Az irodalomban az indirekt kalibrációra az *inverz kalibráció* elnevezés is elterjedt [76], jelen dolgozatban az inverz kalibráció terminust a kalibrációs függvény inverzének használatára tartjuk fenn, azaz a mérési jelekhez tartozó koncentrációk becslését végrehajtó eljárásra.

A következő feltétel teljesülése esetén a paraméterek meghatározásához még egyszerűbb formulákat kapunk.

- i. Az x_j értékek nem valószínűségi változók reprezentánsai, hanem ismert, konkrét értékek, azaz $x_j(\text{megfigyelt érték}) = x_j^0(\text{valódi érték})$

$$\hat{a}_0 = \frac{\sum_{j=1}^n x_j^2 \sum_{j=1}^n y_j - \sum_{j=1}^n x_j \sum_{j=1}^n x_j y_j}{n \sum_{j=1}^n x_j^2 - \left(\sum_{j=1}^n x_j \right)^2} \quad (1.10)$$

$$\hat{a}_1 = \frac{n \sum_{j=1}^n x_j y_j - \sum_{j=1}^n x_j \sum_{j=1}^n y_j}{n \sum_{j=1}^n x_j^2 - \left(\sum_{j=1}^n x_j \right)^2} \quad (1.11)$$

Fokozatosan szigorodó feltételek láncolatán keresztül jutottunk el a legkisebb négyzetek paraméterbecslő eljárással kapott formulákhoz, melyek által minimális varianciájú (hatásos) és torzítatlan becslést kapunk a paraméterekben lineáris függvények körében. Fontos megemlíteni, hogy az eddigi eredményeink elérése során az u_j^k mérési hibák eloszlásait csak az f, g, h . (az első és második momentumokra vonatkozó) és i . feltételekkel rögzítettük, így pld. típusáról sem rendelkezünk [74]. Egyértelműen meg kell adnunk a mérési hibák eloszlását leíró összefüggést, ha a becsült paraméterek valamilyen klasszikus módszerrel történő jellemzését is el kell végeznünk. Ez utóbbi probléma egy más úton történő megoldására a későbbiekben még visszatérünk.

Kalibrációs feladatok során az $y = f(x, \mathbf{p})$ függvény következő alakját használhatjuk (átterve a valószínűségszámításban megszokott jelölésrendszerre, az f és g betűk ezután valószínűségi sűrűségfüggvényeket jelölnek):

$$\eta = C(x, \mathbf{p}) + \delta, \quad (1.12)$$

ahol η a mérőkészülék által szolgáltatott jel, mint valószínűségi változó, x a bállított koncentráció (feltételezeten hibamentes), $\mathbf{p}$ a $C(\cdot)$ kalibrációs függvény meghatározandó paramétervektora, δ a mérési hibát reprezentáló valószínűségi változó.

Az ismert paraméterbecslési eljárásokat az alábbi 3 kategória valamelyikébe sorolhatjuk származtatásuk szerint:

- vektornormák minimalizálása alapján előállított eljárások,
- eloszlásfüggvények funkcionáljai alapján előállított eljárások,
- teszt statisztikák alapján előállított eljárások.

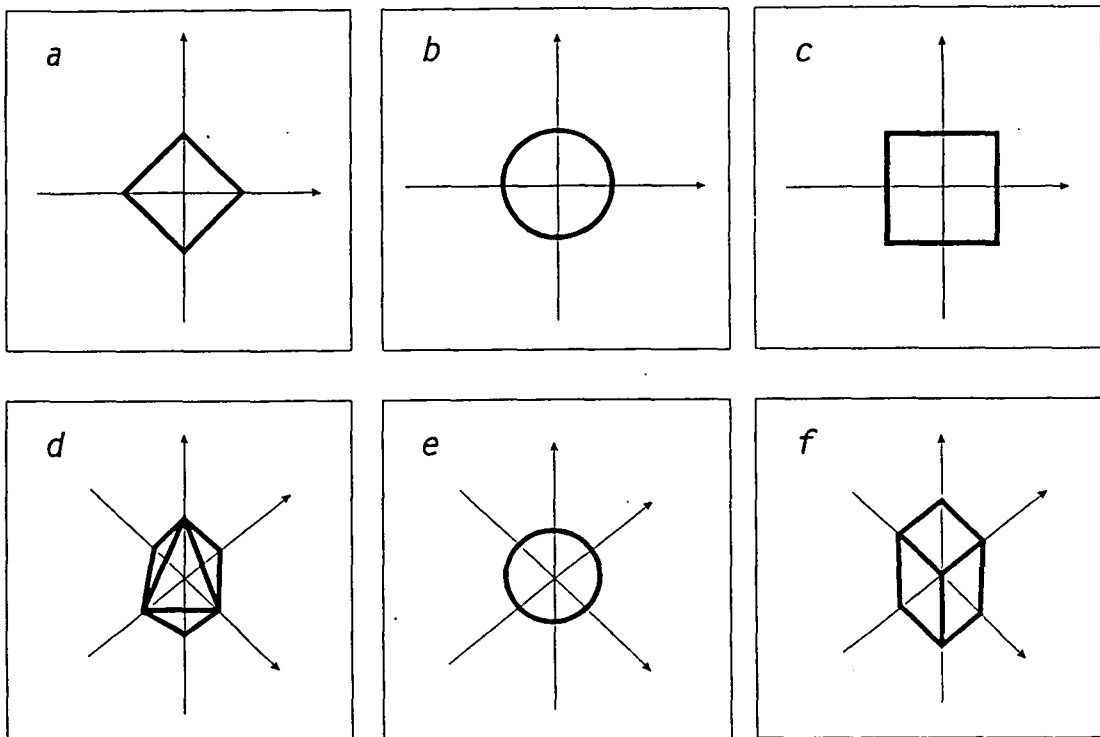
Ezen újszerű csoportosításnak megfelelően fogunk megvizsgálni néhány jellemző esetet az idevonatkozó irodalomra támaszkodva.

Vektornormák minimalizálása alapján előállított eljárások. A vektornormák a távolság fogalom általánosításai. Az előbb levezetett legkisebb négyzetek módszerét például az Euklidészi norma segítségével kaptuk meg. Ez a norma több más normával rokonítható, melyeket összefoglalóan L_p normáknak nevezünk:

$$L_p \text{ norma : } L_p \equiv \| \underline{x} \|_p = \left[\frac{1}{n} \sum |x_i|^p \right]^{\frac{1}{p}} \quad (1.13)$$

Mint említettük a vektor normák a távolság fogalom általánosításai, így vektorok különbségeire is értelmezhetjük őket. Minimalizálva most ezt a normát megkapjuk a legkisebb távolságot a megfigyelési pontok (kalibrációs pontok) által alkotott vektor és azon lineáris altér között, melynek dimenziója a kalibrációs függvény paramétereinek számától függ. Ha a kalibrációs függvény nem lineáris a paraméterekre vonatkozóan, akkor bonyolultabbá válik a helyzet, hiszen az altér sem lesz lineáris. Nézzünk néhány példát az L_p normákra.

$$\begin{array}{ll}
 p & \\
 1 & L_1 \equiv \| \underline{x} - \underline{x}_0 \|_1 = \frac{1}{n} \sum |x_i - x_0| \quad \text{medián : } x_0 \left\{ \begin{array}{ll} x_{(\frac{n-1}{2})} & \text{ha } n \text{ páratlan} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{ha } n \text{ páros} \end{array} \right. \\
 2 & L_2 \equiv \| \underline{x} - \underline{x}_0 \| = \sqrt{\frac{1}{n} \sum |x_i - x_0|^2} \quad \text{átlag : } x_0 = \frac{1}{n} \sum x_i \\
 \vdots & \\
 \infty & L_\infty \equiv \| \underline{x} - \underline{x}_0 \|_\infty = \max_i |x_i - x_0| \quad \text{sávközép : } x_0 = \frac{x_1 + x_n}{2},
 \end{array}$$



1.1 ábra: Az L_1 (a,d), L_2 (b,e) és az L_∞ (c,f) vektornormák 2 ill. 3 dimenziós alakjai

ahol $x_{(j)}$ a j . rendezett megfigyelés és $\underline{x}$ a megfigyelések vektora. Egyváltozós esetben $\underline{x}_0$ a jósolt helyparaméterből alkotott vektor (a vektor összes eleme egyenlő x_0 -val). Különböző becslők neveit és formuláit is feltüntettük, melyeket a normák minimalizálásával nyertük. Az 1.1 ábrán a fenti három norma alakját vizsgálhatjuk meg, mint 2 ill. 3 dimenziós vektorok hosszát.

Magasabb dimenziókban az alakjuk hipergömb, hiperoktaéder és hiperkocka lesz.

Többváltozós esetben x_0 egy olyan altér lesz, melyet a kalibrációs függvény paraméterei feszítenek ki. L_1 minimalizálása egy lineáris programozási feladattal oldható meg [119,44]. Az L_2 probléma a jól ismert legkisebb négyzetek módszerét szolgáltatja [113]. L_∞ az extrém értékek statisztikájában játszik fontos szerepet [49], valamint *minimax* problémák esetén [58]. L_∞ minimalizálása szintén lineáris programozási feladat.

A következő távolság mértéket meglepő módon definiáljuk, úgy hogy az

összegzés helyett produktum képzést használunk: [123]

$$P \equiv \left\{ \prod [\varepsilon^2 + x_i^2] \right\}^{\frac{1}{2n}}, \quad (1.14)$$

ahol

$$\varepsilon^2 = \frac{3 \sum \frac{x_i^2}{(\varepsilon^2 + x_i^2)^2}}{\sum \frac{1}{(\varepsilon^2 + x_i^2)^2}}.$$

ε -t tehát a fenti rekurzív formulával definiálhatjuk és az adatok egyfajta szórtságát fejezi ki, a koncentrálódás mértékét jelzi. A becslő eljárás neve, melyet a P minimalizálásával nyerünk *leggyakoribb érték* (most frequent value) [123].

Végül megemlítünk még egy vektor normát, melyet az L_p -t definiáló képletből úgy nyerhetünk, hogy az összegzést a medián képzés operátorára cseréljük:

$$M_p \text{ norm : } M_p = \left[\frac{1}{n} \text{ median } \{ |x_i|^p \} \right]^{\frac{1}{p}}. \quad (1.15)$$

Az irodalom csak a $p = 2$ esettel foglalkozik és az eljárás a *négyzetek legkisebb mediánja* (least median of squares) nevet viseli [109].

Eloszlásfüggvények funkcionáljai alapján előállított eljárások. A gyakorlatban legtöbbször feltételezhetjük, hogy bármely becslés csak a mérési adatok tapasztalati eloszlásától függ, azaz a becslést az F_n tapasztalati eloszlásfüggvény funkcionáljának tekinthetjük. A T becslőt gyakran valamilyen hipotetikusan feltételezett F valószínűségi eloszlásból származtatjuk. Így $T(F_n)$ egy nemparametrikus becslő lesz, de a tulajdonságait a $T(G)$ viselkedésének tanulmányozásával kell levezetnünk a megvalósuló G számára F környezetében. Ez a probléma az elsőfajú modellhiba példája: vegyünk egy modellt, amely leírja a mérési jel és a koncentráció közötti kapcsolatot a mérési hibák eloszlásával együtt, de ez az eloszlás csak feltételezésen alapul. Használhatjuk azonban az F eloszlásfüggvényt G helyett.

Most már megfogalmazhatjuk a robusztusság kvalitatív definícióját [50]: a becslés robusztus, ha a becslések eloszlása egyenletesen folytonos funkcionálja az anyaeloszlásnak, azaz közel fekvő anyaeloszlásokhoz a becslések közel fekvő eloszlásai tartoznak

(a Prohorov-távolság G és F között kicsi).

A robusztusságot kvantitatívan kétféleképpen mérhetjük. Egy becslő globális érzékenységet az összeomlási ponttal (breakdown point) [58,50,37]

adhatjuk meg (az irodolamban a kvantitatív robusztusság fogalom is használatos), mely egyszerűen fogalmazva a kiugró értékek (melyek akár végtelen nagyok is lehetnek) azon legkisebb hányadát jelenti, amelynél a becslő már minden határon túli (gyakorlatilag végtelen nagy) értékeket is szolgáltat, tehát a jelentős perturbációk hatását képes leírni.

A hatásfüggvény (influence function) [51,52], vagy infinitezimális robusztusság lokálisan jellemzi a robusztusságot. Leírja, hogy a becslő hogyan változik egyetlen pontban infinitezimális perturbáció hatására. Egy becslő robusztus, és eképpen az elsőfajú modellhibát is képes kezelni, ha az összeomlási pontja elegendően nagy valamint a hatásfüggvénye korlátos és folytonos.

Következzen néhány példa az előzőekben részletezett gondolatok alapján levezetett paraméterbecslő eljárásokra.

Momentumok módszere. Legyen a ξ véletlen változó k paraméteres eloszlásfüggvénye a következő:

$$P(\xi < x) = F(x; a_1, a_2, \dots, a_k) = \int_{-\infty}^x f(t; a_1, a_2, \dots, a_k) dt.$$

$m_j(\xi) = E[x^j] = \int_{-\infty}^{\infty} x^j f(x; a_1, a_2, \dots, a_k) dx$ az F eloszlás j . elméleti momentuma. Tegyük fel, hogy az összes elméleti momentum kifejezhető az ismeretlen a_j paraméterekkel, azaz $m_j(\xi) = m_j(a_1, \dots, a_k)$. A tapasztalati momentumokat felhasználva a következő egyenletrendszert nyerjük [134]:

$$m_j(a_1, \dots, a_k) = \frac{1}{n} \sum_{i=1}^n x_i^j, \quad j = 1, 2, \dots, k. \quad (1.16)$$

Néha előfordul, hogy (1.16) megoldása nem egyértelmű. R. A. Fisher kimutatta, hogy a momentumokkal való paraméterbecslés erősen aszimmetrikus eloszlások esetén kevésbé hatékony. Ráadásul a magasabb rendű momentumok egyre érzékenyebbek a mérési hibára a növekvő kitevők miatt, emiatt a momentumok módszerével nem nyerhetünk robusztus becslőt.

Legnagyobb valószínűség (maximum likelihood) elve. Tekintsük az x_i független megfigyeléseket melyek azonos, ismeretlen θ paraméterű eloszlásból származnak. Ebben az esetben az egyesített sűrűségfüggvény az egyedi x_i sűrűségfüggvényeinek szorzata lesz:

$$\mathcal{L}(\underline{x}; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (1.17)$$

Ez a maximum likelihood függvény [134]. Gyakran sokkal egyszerűbb $l = \ln \mathcal{L}$ alakkal dolgozni, mivel a szorzás (1.17)-ben összegzésre módosul. Bármelyik likelihood függvény maximumát meghatározva, a legvalószínűbb $\hat{\theta}$ becslést kapjuk θ -ra vonatkozóan. Bár néha kederül, hogy ez torzított.

A következő példán mutatjuk be a maximum likelihood elvének lényegét. Laplace- és Gauss-eloszlások T helyparamétereit fogjuk becsülni és feltételezzük, hogy az S skálaparaméterek azonosak.

$$\mathcal{L}_L(T, S) = \prod \left[\frac{1}{2S} \exp\left(-\frac{|x_i - T|}{S}\right) \right]$$

$$\min -\ln \mathcal{L}_L(T, S) \sim \min \sum \frac{|x_i - T|}{S} \quad (1.18)$$

$$\mathcal{L}_G(T, S) = \prod \left[\frac{1}{S\sqrt{2\pi}} \exp\left(-\frac{(x_i - T)^2}{2S^2}\right) \right]$$

$$\min -\ln \mathcal{L}_G(T, S) \sim \min \sum \frac{(x_i - T)^2}{2S^2} \quad (1.19)$$

Ha $S > 0$, akkor S nem játszik szerepet a minimumok megtalálásánál, így (1.18) és (1.19) ugyanazt az eredményt szolgáltatja T -re mint az L_1 és L_2 normák minimalizálása. Ez általánosan is igaz. Az f_p supermodell helyparaméterének maximum likelihood becslése azonos lesz az L_p normák minimalizálásával kapott becsléssel:

$$\max \mathcal{L}_p(T) = \max \prod \left[\frac{p^{1-\frac{1}{p}}}{2\Gamma(\frac{1}{p})} \exp\left(-\frac{|x_i - T|^p}{pS^p}\right) \right] \Rightarrow \min L_p \quad (1.20)$$

Ahogy az előzőekben leírtuk a maximum likelihood elvét egy adott $f(x)$ eloszlás esetén a következőkben lehet megfogalmazni: $\max \sum \ln f(x_i, T)$ vagy $\min \sum -\ln f(x_i, T)$. Huber [58] általánosítása a $-\ln f$ függvénynek egy differenciálható ϱ függvényre való kicserélését jelenti:

$$\min \sum \varrho(x_i, T). \quad (1.21)$$

Sokkal praktikusabb megoldani a

$$\sum \psi(x_i, T) = 0 \quad (1.22)$$

egyenletet, ahol $\psi(x, T) = \frac{\partial \varphi(x, T)}{\partial T}$. Mivel T egy helyparaméter becslése, így

$$\sum \psi(x_i - T) = 0. \quad (1.23)$$

$$\varphi(x_i) = \frac{\psi(x_i - T)}{x_i - T} \quad (1.24)$$

felhasználásával $\sum \varphi(x_i) \cdot (x_i - T) = 0$ egyenletet nyerjük, melynek a megoldása

$$T = \frac{\sum \varphi(x_i) \cdot x_i}{\sum \varphi(x_i)}. \quad (1.25)$$

Ez a formula meglehetősen hasonlít egy súlyozott átlagra, azonban a súlyok nem függetlenek az adatoktól, ők maguk is statisztikák. Az (1.25) által definiált becslőket M -becslőknek nevezzük.

Minimális információ veszteség elve. Az előző levezetések során feltételeztük, hogy a hibaeloszlás típusa pontosan ismert. A gyakorlatban nagyon gyakran inkább valamilyen helyettesítő eloszlásfüggvénnyel operálunk, mivel vagy nem ismerjük pontosan az előforduló eloszlást, vagy kényelmesebb egy egyszerűbb formulájú, de hasonló tulajdonságú eloszlással számolnunk.

Ha az adott $f(x)$ sűrűségfüggvényt valamilyen $g(x)$ sűrűségfüggvénnyel helyettesítünk a relatív információt, vagy I -divergenciát definiálhatjuk [32, 65, 77, 106, 117, 123]:

$$I_g(f) = \int_{-\infty}^{\infty} f(x) \log_2 \frac{f(x)}{g(x)}. \quad (1.26)$$

A helyettesítés után az információ veszteség (az I -divergencia) minimalizálásával a g alkalmazásával kapott T helyparaméter és S skálaparaméter becsléseket fogadjuk el az f jellemzésére.

Tegyük fel, hogy $g(x; T)$ (ne felejtjük el, hogy az elsődleges feladat a T meghatározása) szimmetrikus és differenciálható függvény T -re vonatkozóan, valamint az integrálás és a differenciálás művelete felcserélhető. Ekkor a minimumot a jól ismert feltételek teljesülése esetén kapjuk:

$$\begin{aligned} \frac{d I_g(f; T)}{dT} &= 0 \\ \int_{-\infty}^{\infty} \frac{\partial g(x; T)}{\partial T} \cdot \frac{f(x)}{g(x; T)} dx &= 0 \end{aligned} \quad (1.27)$$

és

$$\frac{d^2 I_g(f; T)}{dT^2} > 0$$

$$\int_{-\infty}^{\infty} \left[\frac{\partial g(x; T)}{\partial T} \cdot \frac{1}{g(x; T)} \right]^2 f(x) dx - \int_{-\infty}^{\infty} \frac{\partial^2 g(x; T)}{\partial^2 T} \cdot \frac{f(x)}{g(x; T)} dx > 0. \quad (1.28)$$

(1.28) egyenlet biztosan teljesül, ha

$$\int_{-\infty}^{\infty} \frac{\partial g(x; T)}{\partial T} \cdot \frac{f(x)}{g(x; T)} dx = 0 \quad (1.29)$$

igaz.

Az I -divergencia minimumát az (1.27) és (1.29) egyidejű teljesülésével érhetjük el alkalmasan megválasztott T és S értékekkel.

Két példát tekintsünk, melyekben Gauss- ill. Cauchy-eloszlásokat szerepeltetünk helyettesítő eloszlásként. A Gauss-eloszlás sűrűségfüggvénye:

$$g(x; T) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-T)^2}{2\sigma^2}}.$$

Behelyettesítve ezt (1.27) és (1.29) egyenletekbe kapjuk:

$$T = \int_{-\infty}^{\infty} x f(x) dx.$$

Ez a jólismert E várható érték és

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E)^2 f(x) dx,$$

az $f(x)$ varianciája. Cauchy-eloszlás esetén a sűrűségfüggvény:

$$g(x; T) = \frac{1}{\pi} \frac{S}{S^2 + (x - T)^2}.$$

A helyettesítések után a következő kifejezéseket nyerjük:

$$T = \frac{\int_{-\infty}^{\infty} \frac{x}{S^2 + (x-T)^2} f(x) dx}{\int_{-\infty}^{\infty} \frac{1}{S^2 + (x-T)^2} f(x) dx}.$$

Ez a T az előzőekben már ismertetett leggyakoribb érték becslőt szolgáltatja.

A skálaparaméter becslése:

$$S^2 = \frac{3 \int_{-\infty}^{\infty} \frac{(x-M)^2}{[S^2+(x-M)^2]^2} f(x) dx}{\int_{-\infty}^{\infty} \frac{1}{[S^2+(x-M)^2]^2} f(x) dx}.$$

Erről a formuláról azt kell tudni, hogy ez már robusztus, míg a maximum likelihood elve alapján származtatott becslés érzékenyebb.

Megemlítjük még, hogy más becslőeljárásokat, melyeket az eloszlásfüggvények funkcionáljai alapján nyerhetünk (pld. Bayes módszer, Kálmán-sűrő stb.), nem részletezünk, mivel jelen munkában azzal a feltételezéssel élünk, hogy a paraméterek időben nem változnak.

Teszt statisztikák alapján előállított eljárások. A mérés során kapott $X_1, X_2, \dots, X_n$ minta álljon független, azonos $F(x - \theta)$ eloszlásból származó véletlen változókból. Vizsgálhatjuk a következő tesztet, melyben a nullhipotézist állítjuk szembe egy alternatív hipotézissel: az $F(x - \theta)$ helyparaméterére vonatkozóan:

$$\begin{aligned} H_0 : \theta &= 0 \quad \text{versus} \\ H_A : \theta &\neq 0. \end{aligned} \tag{1.30}$$

A statisztikai tesztet alapvetően két részre osztjuk a mintateret: az egyikben elfogadjuk a H_0 nullhipotézist, a másikban elvetjük azt. A felosztást egy alkalmasan választott V statisztika felhasználásával valósítjuk meg. A nullhipotézisre vonatkozóan V -re egy eloszlást vezethetünk le, melyet felhasználva konkrét valószínűséggel vetjük el illetve fogadjuk el a nullhipotézist.

χ^2 statisztika. Tekintsünk n független, azonos várható értékű, de különböző szórású normális eloszlású véletlen változóból álló mintát. A következő statisztikát vezetjük be:

$$\chi^2 = \sum \frac{(x_i - T)^2}{\sigma_i^2} \tag{1.31}$$

ahol χ^2 chi-négyzet eloszlású véletlen változót jelent $n - 1$ szabadsági fokkal. Minimalizálva χ^2 -et a paraméter súlyozott legkisebb négyzetes becslését nyerjük [134]:

$$T = \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}.$$

Általános pontfüggvények. Legyen a_i az első n pozitív egész pontfüggvénye és legyen értéke nem negatív és nem csökkenő: $0 \leq a_1 \leq \dots \leq a_n$. Definiáljuk a következő statisztikát

$$V = \sum a(R_i) \cdot |x_i - T|, \quad (1.32)$$

ahol R_i a $|x_i - T|$ sorrendjét jelenti a $|x_1 - T|, \dots, |x_n - T|$ sorban. Minimalizálva V -t T -re egy becslő eljárást nyerünk, melyet R -becslőnek hívunk [58,66]. Az $a(i) \equiv 1$ pontfüggvénnyel az L_1 -normával azonos eredményhez jutunk, azaz a minta mediánhoz. Gyakran alkalmazzák a Wilcoxon pontfüggvényt, ahol $a(i) = i$.

Sorrend korreláció. Tételezzük fel, hogy az adataink rendezettek, $x_1 \leq x_2 \leq \dots \leq x_n$.

$$N = \sum_{1 \leq i < j \leq n} |x_j - x_i|$$

a pozitív $x_j - x_i$ különbségek száma. Valamely b valós változóra definiáljuk az $r_i(b) = y_i - bx_i$, $i = 1, \dots, n$ reziduálist. Tekintsük a következő, sorrend korreláción (a Kendall által bevezetett tau [71]) alapuló statisztikát [114]:

$$V(b) = \frac{\sum_{1 \leq i < j \leq n} |x_j - x_i| \cdot |r_j(b) - r_i(b)|}{\sqrt{N \binom{n}{2}}}. \quad (1.33)$$

$V(b)$ a 0 érték becslője lesz és így alkalmas b választásával kell $V(b)$ értékét zérushoz minél közelebb beállítani. Mivel $V(b)$ nem növekszik b szerint, így lesz egy intervallum, ahol $V(b)$ egyenlő lesz nullával. b alkalmas becslésének ezen intervallum közepét választjuk. Tekintsük a következő N különböző pár halmazát:

$$\hat{b}_{ij} = \frac{y_j - y_i}{x_j - x_i}, \quad x_j > x_i. \quad (1.34)$$

$\hat{b}_{ij}$ választjuk mint b -re vonatkozó becslést, mivel ebben az esetben lesz $V(b)$ egyenlő nullával. Pontosan ez az algoritmus jelent meg [45]-ban az előbbi részletes levezetésre vonatkozó hivatkozás nélkül.

U-statisztika. Az U -statisztikát Hoeffding [55] vezette be. Legyen $x_1, \dots, x_n$ egy minta és $\Phi(x_1, \dots, x_m)$ egy függvény m argumentummal. Te-

kintsük a következő statisztikát:

$$U = \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{(\alpha_1, \dots, \alpha_m)} \Phi(x_{\alpha_1}, \dots, x_{\alpha_m}), \quad (1.35)$$

ahol $\sum_{(\alpha_1, \dots, \alpha_m)}$ az $(\alpha_1, \dots, \alpha_m)$ összes permutációjára vonatkozó összegzést jelent úgy, hogy $1 \leq \alpha_i \leq n$, $\alpha_i \neq \alpha_j$ ha $i \neq j$, $i, j = 1, \dots, m$.

Siegel [118] az U -statisztika robusztus változatát vezette le, kicserélve a szummázást az egymásba ágyazott mediánokra:

$$\hat{\theta} = \text{median}^m \{ \theta(x_{\alpha_1}, \dots, x_{\alpha_m}) \}. \quad (1.36)$$

1.1.1.2 Fuzzy halmazokon alapuló modell

A halmazelméletben a közönséges halmazokhoz egy karakterisztikus függvényt rendelhetünk a $\{0, 1\}$ két elemű halmazból vett értékekkel. Zadeh 1965-ben [135] felvetette, hogy a karakterisztikus függvény értékkészlete legyen a $[0, 1]$ zárt intervallum és így a halmazhoz való tartozást ezzel a szubjektív bizonytalansággal fejezhetjük ki. A kiterjesztett karakterisztikus függvényt tagsági függvénynek, a tagsági függvénnyel jellemzett halmazokat fuzzy halmazoknak hívjuk.

A fuzzy halmazok elmélete nagyon gyorsan teret hódított, magyarul Nagy Csaba [93] foglalta össze az elméletre vonatkozó fontosabb ismereteket, majd nem sokkal ezután egy konkrét alkalmazási példával is találkozhatunk [64]. A hagyományos és a fuzzy valószínűségi mezők kapcsolatáról jelenik meg dolgozat [15] az elmélet hazai tovább fejlesztőinek műhelyéből. A legfrissebb összefoglaló [120] korunk tudományában és a mindennapok technikai vívmányaiban éri tetten a fuzzy logikát.

A fuzzy halmazok elméletének nemzetközi szakirodalma igen jelentékeny, így arra vállalkozunk, hogy csak a minket érdeklő regressziós feladatokkal foglalkozókat tekintsük át.

Zadeh 1968-ban közölt tanulmánya [136] a fuzzy események valószínűség-mértékét definiálja, ezzel újabb fejlődési hullámot indítva el. Sorra jelennek meg a cikkek a fuzzy halmazok és a statisztika kapcsolatáról [95, 48, 67, 68, 75, 39, 60]. Sugeno disszertációjában [124] kidolgozza a fuzzy mértékek és fuzzy integrálok elméletét, majd egy alkalmazással is találkozunk [96]. Az érdeklődők témában való részletes elmélyülését két alapmű [69, 40] segíti.

A fuzzy halmazok elméletének alkalmazását a regressziós problémák megoldásában Jajuga [62] dolgozta ki a legegyszerűbben, fuzzy halmazokat alkalmazott a legkisebb négyzetek módszerének súlyozott változatához. Cel-

miñs [25,26] és Diamond [35,34] dolgoztak ki eljárásokat a legkisebb négyzetek módszerének fuzzy változókra való alkalmazására. Elszakadva a legkisebb négyzetek módszerétől más fuzzy regressziós eljárásokat is bemutatnak [27,28,11,12,13,10,121,63,94,47]

Bandemer dolgozata [4] az analitikai kémiában jelentős fejlődést indított el a robusztus fuzzy regressziós eljárások fejlesztése területén, mely a fuzzy halmazok egyéb felhasználását is elősegítette [103,102,8,97,101,104,99,100,56,57,79]. A fuzzy halmazok segítségével végzett adatfeldolgozási módszerekről a Bandemer által szerkesztett kiadványokból [5,6,7] nyerhetünk további ismereteket.

Vizsgáljuk meg részletesen az első fuzzy halmazon alapuló robusztus regressziós eljárás [102] lényegét.

Az analitikai kémiai meghatározások helyessége elsősorban a kalibrálás jóságán múlik. A kalibrációs folyamat során a megfelelő analitikai mérőgörbe kiválasztása, majd alkalmazása a feladat. Általában ismert összetételű mintasorozattal elvégzett mérések után görbeillesztéssel határozzák meg a mérőgörbe paramétereit, majd az egyenlet inverzének segítségével az ismeretlen összetételű minta koncentrációjára lehet következtetni.

Általában feltételezik, hogy a mintasorozat koncentrációja pontosan ismert és csak a mért jelben mutatkozhat bizonytalanság. A fuzzy halmazokon alapuló regressziós eljárás kezelni tudja a mintasorozat koncentrációjában mutató bizonytalanságot is a megfelelő tagsági függvény alkalmazásával. A módszer használata során feltesszük, hogy a kalibrációs függvénykapcsolat pontosan leírja a kalibrációs folyamatot, pld. mert fizikai törvényszerűségek teljesüléséből származtattuk. Ekkor az M_i fuzzy halmaz, mellyel egy fuzzy megfigyelést definiálunk, megadható a tagsági függvényével (membership function): $m_i(x, y), x \in X, y \in Y$, ahol x a koncentráció, y a mérési jel. Ha egy bizonyos pontban több tagsági függvény értéke nem nulla, akkor összevonhatjuk őket:

$$m_M(x, y) = \max_i m_i(x, y). \quad (1.37)$$

A halmazt amelyben egy m_N tagsági függvény nullától különböző értéket vesz fel az N fuzzy halmaz tartójának (support) hívjuk:

$$\text{supp } N \equiv \{(x, y) \in X \times Y : m_N(x, y) > 0\}. \quad (1.38)$$

A tartók lehetnek körök, vagy ellipszisek és a tagsági függvények parabolo-

idok lesznek:

$$m_i(x, y) = \left[1 - \frac{(x - x_i)^2}{u_i^2} - \frac{(y - y_i)^2}{v_i^2} \right]^+, \quad (1.39)$$

ahol u_i^2 és v_i^2 a tartó ellipszis féltengelyeinek hossza, jellemezve az x és y változókban jelentkező bizonytalanságot. $[z] = \max\{z, 0\}$, azaz $m_i(x, y)$ értéke sosem lesz negatív.

A megfigyelések fuzzy jellemzőit a kalibrációs függvényben szereplő paraméterek terébe a várt számosság (expected cardinality) elve alapján vihetjük át. Vegyünk egy lineáris $y = a_0 + a_1x$ függvénykapcsolatot. Egy x_0 kiválasztott pontban az y -nak $m_M(x_0, a_0 + a_1x_0)$ tagsági értéke van. Ez az érték más és más lesz különböző $x_0 \in X$ értékek esetében. Ezért a paraméterek lehetőségének (possibility) megadásához a megfelelő értéket úgy kell számítanunk, hogy a lineáris kapcsolat

$$G(a_0, a_1) = \{(x, a_0 + a_1x) \in X \times Y; x \in X\} \quad (1.40)$$

teljes gráfja megfeleljen a mi egyesített fuzzy megfigyeléseinknek. Integráljuk a tagsági értékeket (1.40) gráf mentén:

$$m_E(a_0, a_1) = \int_{G(a_0, a_1)} m_M(x, y) dx, \quad (1.41)$$

feltéve, hogy az integrál létezik. Az $m_E(a_0, a_1)$ az M megfigyelések *relatív számossága* (relative cardinality) a $G(a_0, a_1)$ gráf mentén. Az eddigiek érvényben maradnak, ha az X környezetet leszűkítjük azon pontokra, melyekben van legalább egy $(x, y) \in X \times Y$ pont amire $m_M(x, y) > 0$, azaz

$$X_0 \equiv \text{supp}(\text{proj}_X M) = \left\{ x \in X : \sup_{y \in Y} m_M(x, y) > 0 \right\}. \quad (1.42)$$

Itt $\text{proj}_X M$ azt jelenti, hogy az M fuzzy megfigyeléseket leképezzük az X -re a maximumokkal (supremum) számolva, az m_M tagsági függvények által meghatározott $y \in Y$ -ra tekintettel. Az előzőleg nyert m_E értéket összehasonlíthatjuk az ún. *fuzzy helyi legjobb közelítéssel* (fuzzy local (best) approximation), melyet gráfjával adunk meg (lineáris függvénykapcsolat esetében X felett: $y = a_0 + a_1x; (x, y) \in X \times Y; (a_0, a_1) \in A$, ahol A egy adott paraméterterület része):

$$G(f) \{(x, f(x)) : x \in X\}. \quad (1.43)$$

A gráf az összes $(x, y) \in X \times Y$ pontból áll, ahol $m_M(x, y)$ maximális feltéve, hogy $y \in Y$, azaz

$$f(x) = \arg \sup_{y \in Y} m_M(x, y), \quad (1.44)$$

az egyenletben az x változó mint paraméter szerepel, ezért az egyenlet jelentése: vedd m_M y argumentumának azon értékeit ahol az $m_M(x, y)$ függvény felveszi legnagyobb értékeit. $G(f)$ -et hívjuk az $m_M(x, y)$ felület *maximális nyomának* X -en.

Most kiszámíthatjuk

$$m_E(f) = \int_{G(f)} m_M(x, y) dx = \int_{X_0} \sup_{y \in Y} m_M(x, y) dx. \quad (1.45)$$

Mivel $m_E(f) > 0$ és $m_E(a_0, a_1) \leq m_E(f)$ bevezetjük

$$m_c(a_0, a_1; f) = \frac{m_E(a_0, a_1)}{m_E(f)} \quad (1.46)$$

értéket, amely azon közelítés fokát fejezi ki, melyet a relatív számosság és a helyi (legjobb) közelítés összevetéséből nyertünk. Az m_c és az m_E tagsági függvények az $(a_0, a_1) \in A$ paraméterek függvényei és ezek határozzák meg az A halmaz felett az $\tilde{A}$ fuzzy halmazt, melyet az ismeretlen paraméterek fuzzy becslésének hívunk.

A szerzők a kalibráció eredményeit felhasználó fuzzy becslést is leírnak, majd az elméleti részben tárgyalta egy részletes algoritmus közlésével foglalják össze.

Az Otto és Bandemer [102] által kidolgozott robusztus fuzzy regressziós eljárás többváltozós általánosításáról a *Kalibrációs függvények paramétereinek becslése* című fejezet *Lineáris fuzzy regresszió* című szakaszában foglalkozunk.

1.1.2 Paraméterbecslés másodfajú modellhiba esetén

A legszélesebb körben elterjedt mérési modell szerint valamilyen zaj szuperponálódik a valódi értékekre és ebből származóan a becsült paraméterek és a jósolt függvényértékek is hibával terheltekké válnak. Ennek a modellnek az alternatívájaként állíthatjuk, hogy a mérési hibák az alkalmazott modell nem adekvát jellegéből származik. Ebben az esetben feltesszük, hogy pontosan tudunk mérni, de abban már nem vagyunk biztosak, hogy az alkalmazott modell jó-e. Mindenképpen választanunk kell valamilyen modellt, a legtöbb

esetben minél egyszerűbbet a kezelhetőség szempontja szerint. Valóban, lehetetlen az összes kapcsolatot matematikailag megformulázni, némelyiket el kell hanyagolni. A mérési hibák ekkor abból származnak, hogy nem tudjuk kellőképpen leírni a rendszerünket. Ezek a hibák tehát a sohasem ismert helyes modell és az alkalmazott modell közötti különbségből adódnak, azaz az eltérések a rendszert meghatározó paraméterek fuzzy jellegével magyarázhatók, semmint megfigyelési hibákként. Ez a probléma így már nem is sztochasztikus, hiszen a rendszernek nincsen véletlen jellegű tulajdonsága, a megoldás a fuzzy elmélet alkalmazásával érhető el. A linearitás feltételezése nagyon gyakori az analitikai kémiában, ezért Tanaka és munkatársai [126] nyomán bemutatjuk annak kezelését, amikor ez a feltételezés sérül.

Az analitikai kémiában gyakran találkozunk lineáris, vagy linearizált kalibrációs függvényekkel, azonban a használatukhoz szükséges feltételek ritkán teljesülnek. Így elfogadjuk a linearitás teljesülését tudva azt, hogy ez nem mindig teljesen korrekt, ezért ez a döntés szubjektív tartalommal ruházható fel.

Ezt a problémát az $\mathcal{F}(y)$ fuzzy függvények használatával oldhatjuk meg. Ezekben a függvényekben a paraméterek fuzzy számok [40]: $\tilde{Y} = f(x, \tilde{A})$.

Most az $\mathcal{F}(y)$ az összes $\tilde{Y}$ -on értelmezett fuzzy részhalmaz halmaza és $\tilde{Y}$ tagsági függvényét a következőképpen definiáljuk:

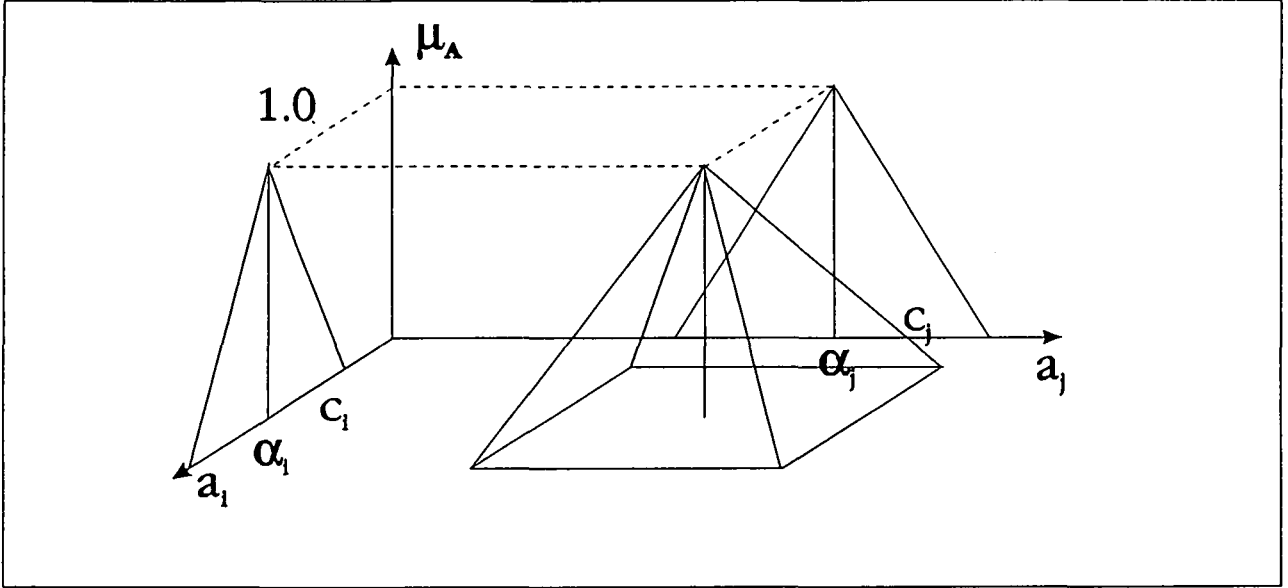
$$\mu_{\tilde{Y}}(y) = \begin{cases} \max_{\{a|y=f(x,a)\}} \mu_{\tilde{A}}(a) & \{a \mid y = f(x, a)\} \neq \emptyset \\ 0, & \text{egyébként.} \end{cases} \quad (1.47)$$

Az 1.2 ábrán láthatjuk a fuzzy paraméterek fogalmának megvilágítását. $\tilde{A}$ tagsági függvénye:

$$\mu_{\tilde{A}}(a) = \min_j [\mu_{\tilde{A}_j}(a_j)] \quad (1.48)$$

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|\alpha_j - a_j|}{c_j} & \text{ha } \frac{|\alpha_j - a_j|}{c_j} < 1 \\ 0, & \text{egyébként.} \end{cases} \quad (1.49)$$

α_j a fuzzy szám centruma és c_j a terjedelem, a centrum körüli szóródás mértéke. A háromszög alakú fuzzy számok értelmezése a következő: a centrum leírja a leginkább lehetséges paraméter értéket, a terjedelem pedig a pontosságát határozza meg, vagyis a rendszer határozatlanságát tükrözi.



1.2 ábra: Fuzzy paraméterek, mint háromszög alakú fuzzy számok

Az $\tilde{Y} = \tilde{A}_1 x_1 + \dots + \tilde{A}_j x_j + \dots + \tilde{A}_p x_p$ fuzzy lineáris függvény (háromszög alakú $\tilde{A}$ fuzzy paraméterekkel) tagsági függvénye:

$$\mu_{\tilde{Y}}(y) = \begin{cases} 1 - \frac{|y - \underline{x}^t \underline{\alpha}|}{\underline{c}^t |\underline{x}|} & \underline{x} \neq 0 \\ 1, & \underline{x} = 0, \underline{y} = 0 \\ 0, & \underline{x} = 0, \underline{y} \neq 0 \end{cases} \quad (1.50)$$

ahol $|\underline{x}| = (|x_1|, \dots, |x_n|)^t$, és $\underline{c}$ és $\underline{\alpha}$ jelenti a paraméterek centrumát és terjedelmét vektor alakban.

A becsült fuzzy lineáris modellnek le kell fednie az összes y_j adatpontot bizonyos H küszöb érték mellett, melyet az eljárás alkalmazója választott korábban, azaz $\mu_{\tilde{Y}}(y_i) \geq H$ ($i = 1, 2, \dots, n$) a nem fuzzy, azaz pontos y_i adatokra vonatkozóan. Ezt az egyenlőtlenséget átírhatjuk felhasználva az $\mu_{\tilde{Y}}(y)$ tagsági függvényt (1.50)-ből:

$$(1 - H)\underline{c}^t |\underline{x}| - |y - \underline{x}^t \underline{\alpha}| \geq 0, \quad \underline{x} \neq 0. \quad (1.51)$$

Természetes követelmény, ha olyan paramétereket választunk, melyek c_j terjedelmeinek összege a legkisebb, más szavakkal a megoldás a legkevesbé

bizonytalanabb fuzzy lineáris függvény kell legyen figyelembe véve a megfigyeléseket és egy jól választott H értéket.

A fuzzy paraméterek kiszámítása így egy egyszerű lineáris programozási feladat lesz:

$$\begin{aligned}
 &\text{Minimalizálandó} && s = \sum_{j=1}^p c_j \\
 &\text{Feltételek} && (1 - H) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}_i^t \underline{\alpha} \geq y_i, \\
 & && (H - 1) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}_i^t \underline{\alpha} \leq y_i, \\
 & && c_i \geq 0, \\
 & && i = 1, 2, \dots, n,
 \end{aligned} \tag{1.52}$$

ahol s a teljes bizonytalanság.

Tanaka és munkatársai [126] által kidolgozott fuzzy lineáris regressziós eljárás robusztus változtát a *Kalibrációs függvények paramétereinek becslése* című fejezet *Fuzzy lineáris regresszió* című szakaszában ismertetjük.

Az eddig elmondottak alapján talán érdemes még egyszer felhívni a figyelmet arra, hogy mi a különbség a *lineáris fuzzy regresszió* és a *fuzzy lineáris regresszió* között.

Lineáris fuzzy regresszióról beszélünk, ha a feltételezett lineáris kalibrációs modell adekvát és a mérési hibák matematikai statisztikai értelmezése helyett a szubjektív, előzetes információkat is magába foglaló fuzzy számokat használjuk a bizonytalanság értelmezésére.

Fuzzy lineáris regresszió esetén a lineáris modell pontos teljesülése kétséges, fuzzy számokat használunk a modellben előforduló paraméterek leírására és ebből fakadóan kapunk eltéréseket a mért és számított értékek között.

1.1.3 Paraméterbecslés harmadfajú modellhiba esetén

Ebben az esetben a mérési adatok és a közöttük fennálló függvénykapcsolat egyaránt csak bizonytalansággal adhatók meg, azonban a feladat sokrétűségét tekintve ez a fajta modellhiba jelen dolgozatban nem kerül részletesebb tárgyalásra.

2. fejezet

Kalibrációs függvények paramétereinek becslése

2.1 Egyváltozós kalibrációs függvények robusztus becslése

Lineáris összefüggés alkalmazására az egyik legjobb példa az ICP-AES (Inductively Coupled Plasma Atomic Emission Spectrometry, induktívan csatolt plazma - atomemissziós spektrometria) mérésekhez használt kalibrációs görbe. Egy ICP spektrométer a koncentráció tartományban hat nagyságrendben képes lineáris jelet produkálni, így kiváló példája az elsőfajú modellhibának, amikor a függvénykapcsolat előzetesen ismert, azonban a feltételezett hibaeloszlás nem teljesül.

Olyan valós kalibrációs adatokon mutatjuk be az általunk vizsgált paraméterbecslő eljárások alkalmazását, melyeket talaj- és ivóvíz környezetvédelmi vizsgálatához alkalmaztak. Meg kell jegyeznünk, hogy ezeket az adatokat számos kalibrációs adatsorból válogattuk, éppen annak illusztrálására, hogy a felhasznált paraméterbecslő eljárások mennyire érzékenyek az esetleg előforduló kiugró értékekre ill. a feltételezett hibaeloszlástól való eltérésre. A kiugró értékek létezését a berendezésben ill. a minta előkészítés során alkalmanként előforduló ismeretlen eredetű hibák eredményezik.

A következőkben felsoroljuk a felhasznált becslő módszereket, közöljük azok angol elnevezését, valamint az azon alapuló rövidítéseket, melyekkel a továbbiakban hivatkozni fogunk rájuk.

- legkisebb négyzetek [134] (Least Sum of Squares, LS),

- abszolút eltérések legkisebb összege [44,119,128] (Least Sum of Absolute Residuals, LSA),
- abszolút eltérések legkisebb maximuma [128] (Least Maximum of Absolute Residuals, LMA),
- iteratíván újraszúlyozott legkisebb négyzetek 6-os ill. 9-es hangoló konstanssal [107] (Iteratively Reweighted Least Sum of Squares with Tuning Constants 6 and 9, IRLS6, IRLS9),
- leggyakoribb érték [123] (Most Frequent Value, MFV),
- egyszerű medián vagy kombinatorikus [114,45] (Single Median, SM),
- ismételt medián [118] (Repeated Median, RM),
- négyzetek legkisebb mediánja [109] (Least Median of Squares, LMS).

Ezek után röviden ismertetjük a paraméterbecslő módszerek algoritmusait. (A feltételezett lineáris modell $y_i = a_0 + a_1 x_i + \varepsilon_i$):

Legkisebb négyzetek módszere (LS)

$$\hat{a}_{0_{LS}} = \frac{\sum x_j^2 \sum y_j - \sum x_j \sum x_j y_j}{n \sum x_j^2 - (\sum x_j)^2} \quad (2.1)$$

$$\hat{a}_{1_{LS}} = \frac{n \sum x_j y_j - \sum x_j \sum y_j}{n \sum x_j^2 - (\sum x_j)^2} \quad (2.2)$$

Abszolút eltérések legkisebb összege módszer (LSA)

$$\begin{array}{ll} \text{Minimalizálandó} & \sum_{j=1}^n s_j \\ \text{Feltételek} & b_1 - b_2 + a_1 x_i - a_2 x_i + s_i \geq y_i, \\ & b_1 - b_2 + a_1 x_i - a_2 x_i - s_i \leq y_i, \\ & a_1, a_2, b_1, b_2, s_i \geq 0, \\ & i = 1, 2, \dots, n. \end{array} \quad (2.3)$$

$$\hat{a}_{1_{LSA}} = a_1 - a_2 \quad (2.4)$$

$$\hat{a}_{0_{LSA}} = b_1 - b_2 \quad (2.5)$$

Abszolút eltérések legkisebb maximuma módszer (LMA)

$$\begin{array}{ll}
 \text{Minimalizálendő} & s \\
 \text{Feltételek} & b_1 - b_2 + a_1 x_i - a_2 x_i + s \geq y_i, \\
 & b_1 - b_2 + a_1 x_i - a_2 x_i - s \leq y_i, \\
 & a_1, a_2, b_1, b_2, s \geq 0, \\
 & i = 1, 2, \dots, n.
 \end{array} \tag{2.6}$$

$$\hat{a}_{1_{LMA}} = a_1 - a_2 \tag{2.7}$$

$$\hat{a}_{0_{LMA}} = b_1 - b_2 \tag{2.8}$$

Iteratívan újrásúlyozott legkisebb négyzetek módszere k=6 és k=9 hangoló konstanssal (IRLS)

$$\hat{a}_0^{(0)} = \hat{a}_{0_{LS}} \text{ és } \hat{a}_1^{(0)} = \hat{a}_{1_{LS}} \tag{2.9}$$

$$\hat{a}_0^{(i)} = \frac{\sum w_j x_j^2 \sum w_j y_j - \sum w_j x_j \sum w_j x_j y_j}{\sum w_j \sum w_j x_j^2 - (\sum w_j x_j)^2} \tag{2.10}$$

$$\hat{a}_1^{(i)} = \frac{\sum w_j \sum w_j x_j y_j - \sum w_j x_j \sum w_j y_j}{\sum w_j \sum w_j x_j^2 - (\sum w_j x_j)^2} \tag{2.11}$$

$$w_j = \begin{cases} (1 - (\frac{r_j}{kS})^2)^2 & \text{if } |r_j| < kS \\ 0 & \text{egyébként} \end{cases} \tag{2.12}$$

$$r_j = y_j - \hat{a}_0^{(i-1)} - \hat{a}_1^{(i-1)} x_j \tag{2.13}$$

$$S = \text{median}_j \{|r_j|\} \quad \text{ha } |r_j| > 0 \tag{2.14}$$

$k = 6$ vagy 9

$$\hat{a}_{0_{IRLS}} = \hat{a}_0^{(k)}, \hat{a}_{1_{IRLS}} = \hat{a}_1^{(k)} \quad \text{ha} \quad \left| \frac{\hat{a}_0^{(k)} - \hat{a}_0^{(k-1)}}{\hat{a}_0^{(k)}} \right| < \delta \text{ és } \left| \frac{\hat{a}_1^{(k)} - \hat{a}_1^{(k-1)}}{\hat{a}_1^{(k)}} \right| < \delta$$

$$\delta = 0.001$$

Leggyakoribb érték módszere (MFV)

$$\hat{a}_0^{(0)} = \hat{a}_{0_{LS}} \quad \text{és} \quad \hat{a}_1^{(0)} = \hat{a}_{1_{LS}} \quad (2.15)$$

$$\hat{a}_0^{(i)} = \frac{\sum w_j x_j^2 \sum w_j y_j - \sum w_j x_j \sum w_j x_j y_j}{\sum w_j \sum w_j x_j^2 - (\sum w_j x_j)^2} \quad (2.16)$$

$$\hat{a}_1^{(i)} = \frac{\sum w_j \sum w_j x_j y_j - \sum w_j x_j \sum w_j y_j}{\sum w_j \sum w_j x_j^2 - (\sum w_j x_j)^2} \quad (2.17)$$

$$w_j = \frac{\varepsilon^2}{\varepsilon^2 + r_j^2} \quad (2.18)$$

$$r_j = y_j - \hat{a}_0^{(i-1)} - \hat{a}_1^{(i-1)} x_j \quad (2.19)$$

$$(\varepsilon^{(l)})^2 = \frac{3 \sum \frac{(\varepsilon^{(l-1)})^4 r_j^2}{((\varepsilon^{(l-1)})^2 + r_j^2)^2}}{\sum \frac{(\varepsilon^{(l-1)})^4}{((\varepsilon^{(l-1)})^2 + r_j^2)^2}} \quad (2.20)$$

$$\varepsilon = \varepsilon^{(m)} \quad \text{ha} \quad \left| \frac{\varepsilon^{(m)} - \varepsilon^{(m-1)}}{\varepsilon^{(m)}} \right| < \delta \quad (2.21)$$

$$\hat{a}_{0_{MFV}} = \hat{a}_0^{(k)}, \hat{a}_{1_{MFV}} = \hat{a}_1^{(k)} \quad \text{ha} \quad \left| \frac{\hat{a}_0^{(k)} - \hat{a}_0^{(k-1)}}{\hat{a}_0^{(k)}} \right| < \delta \quad \text{és} \quad \left| \frac{\hat{a}_1^{(k)} - \hat{a}_1^{(k-1)}}{\hat{a}_1^{(k)}} \right| < \delta$$

$$\delta = 0.001$$

Egyszerű medián vagy kombinatorikus módszer (CM)

$$\hat{a}_{1_{CM}} = \underset{1 \leq i < j \leq n}{\text{median}} \left\{ \frac{y_j - y_i}{x_j - x_i} \right\}_{x_i \neq x_j} \quad (2.22)$$

$$\hat{a}_{0_{CM}} = \underset{j}{\text{median}} \{y_j - \hat{a}_{1_{CM}} x_j\} \quad (2.23)$$

Ismételt medián (RM)

$$\hat{a}_{1_{RM}} = \underset{i}{\text{median}} \left\{ \underset{j}{\text{median}} \left\{ \frac{y_j - y_i}{x_j - x_i} \right\} \right\} \quad \text{ha } x_i \neq x_j \quad (2.24)$$

$$\hat{a}_{0_{RM}} = \underset{j}{\text{median}} \{y_j - \hat{a}_{1_{RM}} x_j\} \quad (2.25)$$

Négyzetek legkisebb mediánja (LSM)

$$\min_{1 \leq j < k \leq n} \underset{i}{\text{median}} \{r_{ijk}^2\} \implies \hat{a}_{0_{LSM}}, \hat{a}_{1_{LSM}} \quad (2.26)$$

$$r_{ijk} = y_i - \hat{a}_{0_{jk}} - \hat{a}_{1_{jk}} x_i \quad (2.27)$$

$$\hat{a}_{0_{jk}} = \frac{x_j y_k - x_k y_j}{x_j - x_k} \quad \text{ha } x_j \neq x_k \quad (2.28)$$

$$\hat{a}_{1_{jk}} = \frac{y_j - y_k}{x_j - x_k} \quad \text{ha } x_j \neq x_k \quad (2.29)$$

A 2.1 táblázat a kalibrációs adatokat tartalmazza, míg a 2.2 táblázat a korábban említett becslő eljárásokkal kapott paraméterek értékeit tünteti fel.

Az adatok kiértékeléséhez egy ún. minőségjellemző faktort (Quality Coefficient, QC) definiálunk:

$$QC = 100 \cdot \sqrt{\frac{\sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2}{n - 1}}, \quad (2.30)$$

ahol y_i a mért, $\hat{y}_i$ az LS módszer által becsült független változó és n az adatok száma, kivéve azokat, melyeknél az y_i mért érték nulla. Az analitikai eljárás pontosságát (ill. pontatlanságát) figyelembe véve előzetesen QC -re egy kritérium értéket lehet megadni. Ha QC ennél a kritérium értéknél nagyobb, akkor azt a hipotézist, hogy nincs kiugró érték elvetjük és robusztus módszert kell alkalmaznunk.

A 2.1 táblázatban közölt mérésekhez 10 %-os kritérium értéket választottunk. A QC értékek 69.2 %, 23.7 %, 497.3 %, 22.9 %, 49.8 % és 13.8 %-nak adódtak a Mo, Cr, Co, Pb és Ni (221,6 nm és 231.6 nm) elemekre

Koncentráció (ppm)	A vizsgált elemekre vonatkozó mérési jelek (tetszőleges egységben)					
	Mo	Cr	Co	Pb	Ni (221.6 nm)	Ni (231.6 nm)
0	8.19	-23.4	-1.83	13.46	6.4	28.33
0	16.05	-19.4	-2.45	6.4	7.47	30.56
0.25	171.9	210.9	261.3	112	223.7	220.8
0.25	180.6	213.7	260.1	119.2	215.6	218.6
0.5	406	420.4	430.8	217	437.9	410.2
0.5	414.5	423.2	431.4	207.7	430.7	407.9
1	810.7	843.2	860.3	419.6	897.3	828.3
1	818.2	840.6	859.6	428.7	886.8	826.1

2.1 táblázat: A kalibrációs adatok ICP-AES mérésekhez

Methods	Mo		Cr		Co		Pb		Ni*		Ni**	
	a_0	a_1	a_0	a_1	a_0	a_1	a_0	a_1	a_0	a_1	a_0	a_1
LS	-3.07	814.5	-11.9	858.5	17.2	846.1	9.84	413.0	0.47	886.3	21.9	798.7
LSA	8.19	802.5	-2.4	845.6	1.3	859.0	13.5	407.1	2.67	884.1	19.0	807.1
LMA	-6.29	802.1	-12.8	864.0	21.7	862.1	8.7	412.7	-3.37	889.8	19.8	797.7
IRLS6	11.1	802.3	-11.7	858.0	-1.45	862.1	10.0	412.9	0.85	885.8	22.2	799.1
IRLS9	-1.88	813.6	-11.8	858.3	15.7	847.2	9.91	412.9	0.63	886.1	22.0	798.8
MFV	9.36	803.9	1.31	840.8	-1.96	862.1	10.3	412.1	6.4	880.4	22.9	801.3
SM	5.25	808.4	-9.15	854.6	0.39	860.2	9.65	412.9	5.1	880.0	26.7	782.7
RM	2.01	813.5	-1.55	844.5	2.25	857.7	10.4	410.8	6.2	870.9	28.9	765.2
LMS	12.1	802.5	2.4	839.6	-1.34	862.0	11.8	407.1	8.98	848.6	29.4	761.1

* 221.6 nm hullámhosszon mérve

** 231.6 nm hullámhosszon mérve

2.2 táblázat: A becsült paraméterek értékei a szövegben említett paraméterbecslő módszerekkel meghatározva

vonatkozóan. A kobalt esetében a QC érték meglehetősen nagyra adódott (497.3 %), melynek magyarázata az, hogy a 0.25 ppm koncentrációnál kapott jel kiugró érték és ennek következtében az LS egyenes meglehetősen távolra került a 0 ppm koncentrációnál mért jeltől.

Az adatok elemzéséből a következőket állapíthatjuk meg. A legjobb becsléseket az LMS adta. Az MFV is jól szerepelt a Mo, Cr, Pb, Co és Ni (221.6 nm) esetében. Az RM által jóslott kalibráló egyenes elfogadható a Cr, Pb és a Ni esetében mindkét hullámhossznál, az IRLS6 által szolgáltatott pedig a Mo és Co mérésekre. Az SM jónak mutatkozott az Pb és a Ni (231.6 nm) esetében, az LSA csak a Mo esetében. Az IRLS9 és az LS közel azonosan szerepeltek, de a mezőny végére kerültek. Egyértelműen a legtorzítottabb eredményt az LMA adta, mutatva a kiugró értékekre vonatkozó kifejezett érzékenységet.

Ezen gyakorlati és más elméleti megfontolásokon nyugodva a 2.3 táblázatban közöljük az általunk vizsgált paraméterbecslő eljárásokra vonatkozó fontosabb tulajdonságokat és egyben gyakorlati felhasználásukhoz segítséget nyújtó ajánlásainkat.

Módszerek	Származás	Összeomlási pont	Robusztusság	Ajánlás
LMA	$\min L_{\infty}$	0 %	egyáltalán nem robusztus	nem ajánlott
LS	$\min L_2$	0 %	egyáltalán nem robusztus	csak nagyszámú adat esetén (30-50 mérési pont)
LSA	$\min L_1$	0 %	modellérzékeny pontok hatása*	csak jól tervezett kísérleteknél
IRLS9	M-becslő	20 %	kissé robusztus	csak nagyszámú adat esetén (15-30 mérési pont)
IRLS6	M-becslő	25 %	robusztus	} kisszámú adat esetén is (8-15 mérési pont)
MFV	M-becslő	25 %	robusztus	
SM	rang-korreláció	30 %	robusztus	
RM	U-statisztika	50 %	nagyon robusztus	
LMS	$\min M_2$	50 %	nagyon robusztus	

* modellérzékeny pontok (leverage points) érzékennyé teszik a becslő eljárást a nem kiegyensúlyozott kísérleti beállítás miatt

2.3 táblázat: A vizsgált paraméterbecslő eljárások tulajdonságai

2.2 Többváltozós kalibrációs függvények robusztus becslése

Ebben a szakaszban a többváltozós kalibrációs függvények paramétereinek robusztus becsléséről lesz szó. Tekintsük át a felhasználásra kerülő alapfogalmakat. Egy helyparaméter becslő, mely n pontból álló mérésre vonatkozik, az $\underline{x}_1, \dots, \underline{x}_n$ vektorfüggvénye lesz:

$$\underline{t}_n = \underline{t}(\underline{x}_1, \dots, \underline{x}_n). \quad (2.31)$$

Hasonlóan egy kovariancia becslő (skálaparaméter becslő):

$$\underline{C}_n = \underline{C}(\underline{x}_1, \dots, \underline{x}_n), \quad (2.32)$$

mely mindig egy szimmetrikus, pozitív definit mátrix.

A gyakorlatban csak olyan becslőket alkalmazunk, melyek a változók permutációjára invariánsak:

$$\underline{t}(\underline{x}_{\pi(1)}, \dots, \underline{x}_{\pi(n)}) = \underline{t}(\underline{x}_1, \dots, \underline{x}_n) \quad (2.33)$$

$$\underline{C}(\underline{x}_{\pi(1)}, \dots, \underline{x}_{\pi(n)}) = \underline{C}(\underline{x}_1, \dots, \underline{x}_n), \quad (2.34)$$

azaz a mérési pontok átszámozása nem változtatja meg a becslést.

A becslők közötti választás fontos kritériuma lehet a különböző ekvivarianciák teljesülése. Egy helyparaméter becslő *eltolás ekvivariáns* (translation equivariant), ha

$$\underline{t}(\underline{x}_1 + \underline{v}, \dots, \underline{x}_n + \underline{v}) = \underline{t}(\underline{x}_1, \dots, \underline{x}_n) + \underline{v} \quad (2.35)$$

minden $\underline{v}$ vektorra. Ez azt jelenti, hogy a mérési adatok eltolása esetén az új becslés az eredeti adatokra vonatkozó becslés azonos mértékű eltolásával nyerhető. Tehát pld. a nullapont megváltozása esetén (hőmérséklet adatok termodinamikai és Celsius-skálán) a becslés is az eltolásnak megfelelően alakul.

Egy helyparaméter becslő *affin ekvivariáns* (affine equivariant), ha

$$\underline{t}(\underline{A}\underline{x}_1 + \underline{v}, \dots, \underline{A}\underline{x}_n + \underline{v}) = \underline{A}\underline{t}(\underline{x}_1, \dots, \underline{x}_n) + \underline{v} \quad (2.36)$$

minden nonsinguláris $\underline{A}$ mátrixra és bármely $\underline{v}$ vektorra. Ez azt jelenti, hogy a helyparaméter becslő az affin transzformáció hatására ugyanúgy változik, mint ahogy a várható érték operátor hat: $E[\underline{A}\underline{x} + \underline{v}] = \underline{A}E[\underline{x}] + \underline{v}$. Az affin ekvivariancia úgy tűnik természetes követelmény a helyparaméter becslők számára, ám számos többváltozós becslő nem elégíti ki ezt a feltételt. Ha $\underline{A}$ csak ortogonális mátrix lehet, akkor ezen gyengébb feltétellel teljesülő ekvivarianciát *ortogonális ekvivarianciának* (orthogonal equivariance) nevezzük. Ekkor a transzformáció a merev mozgásoknál, úgymint transzláció, rotáció, reflexió alkalmazotthoz lesz hasonló. A gyakorlatban pld. a mértékegység váltások miatt lehet hasznos az affin ill. az ortogonális ekvivariancia.

A kovariancia becslő akkor affin ekvivariáns, ha

$$\underline{C}(\underline{A}\underline{x}_1 + \underline{v}, \dots, \underline{A}\underline{x}_n + \underline{v}) = \underline{A}\underline{C}(\underline{x}_1, \dots, \underline{x}_n)\underline{A}' \quad (2.37)$$

minden nemszinguláris $\underline{A}$ mátrixra és bármely $\underline{v}$ vektorra. Tehát a kovariancia becslő invariáns az eltolásra, és ugyanúgy változik, ahogy a variancia operátor hat: $V[\underline{A}\underline{x} + \underline{v}] = \underline{A} V[\underline{x}] \underline{A}'$.

Mivel a robusztus becslő nagy (legalábbis nem nulla) összeomlási ponttal és korlátos hatásfüggvénnyel rendelkezik, vizsgáljuk meg ezen fogalmak gyakorlatban használható matematikai megfogalmazását.

Donoho és Huber [37] fogalmazta meg a véges mintákra definiálható adatromlások fajtáit:

1, ε -szennyezett: az eredeti n elemű mintához (X) m tetszőleges értéket fűzünk ($X^\sharp$). A rossz értékek aránya:

$$\varepsilon = \frac{m}{n + m}$$

2, ε -cserélt: az n elemű minta (X) m méretű tetszőleges részhalmazát tetszőleges értékekre cseréljük ($X^\sharp$). A rossz értékek aránya:

$$\varepsilon = \frac{m}{n}$$

3, ε -módosított: tartozzon az X mintához F_n , az $X^\sharp$ mintához $G_{n^\sharp}$ tapasztalati mérték úgy, hogy

$$\pi(F_n, G_{n^\sharp}) \leq \varepsilon,$$

ahol π tetszőleges távolságfüggvény. n és $n^\sharp$ eltérhet egymástól.

Az előzőek figyelembe vételével véges mintákra a következő összeomlási pont definiálható:

$$\varepsilon^*(X, t) = \inf \left\{ \varepsilon : \sup \left| t(X^\sharp) - t(X) \right| = \infty \right\} \quad (2.38)$$

Az összeomlási pont a becslő globális jellemzője, az adatok nagyobb mértékű szennyezettségének hatását írja le.

Amennyiben a $\underline{t}_n$ helyparaméter becslő funkcionálnak tekinthető az összes valószínűségeloszlást tartalmazó tér felett: $\underline{t}_n = \underline{t}(P_n)$, akkor a $\underline{t}_n$ becslő $IF(\cdot)$ hatásfüggvénye bizonyos P eloszlás esetén (a $t(\cdot)$ funkcionál Gateaux deriváltja a P eloszlásnál, a δ_x Dirac mérték irányában):

$$IF(x; t, P) = \lim_{\varepsilon \rightarrow 0} \frac{t((1 - \varepsilon)P + \varepsilon \delta_x) - t(P)}{\varepsilon}. \quad (2.39)$$

A hatásfüggvény a becslő lokális tulajdonsága kis mértékű szennyezettség esetén. Leírja, hogy a becslő infinitezimális perturbáció hatására hogyan viselkedik egy pontban.

A fenti hatásfüggvényen alapulva definiálható a durva hiba érzékenység is:

$$\gamma^*(t, P) = \sup_x ||\text{IF}(x; t, P)|| \quad (2.40)$$

Az alapfogalmak tárgyalása után vegyük sorra a robusztus többváltozós paraméterbecslőket. Először tekintsük az iteratíván újrásúlyozott legkisebb négyzetek módszerének többváltozós és általánosabban tárgyalt variánsát. Collins [31] tárgyalta általánosan a többváltozós M -becslőket, melyek a következő egyenlet megoldásaként adódnak:

$$\sum_{i=1}^n \frac{\psi(||\underline{x}_i - \underline{t}||)}{||\underline{x}_i - \underline{t}||} (\underline{x}_i - \underline{t}) = 0. \quad (2.41)$$

Azonban, ha a skálaparaméter nem rögzített, a kapott becslő nem lesz affin ekvivariáns.

Maronna [83], majd Huber [58] mutatott be affin ekvivariáns M -becslőt a következő egyenletrendszer megoldásaként, ahol a helyparaméter és a kovariancia becslést szimultán végezzük:

$$\sum_{i=1}^n v_1(d(\underline{x}_i, \underline{t}, \underline{C})) (\underline{x}_i - \underline{t}) = 0 \quad (2.42)$$

$$\sum_{i=1}^n \{v_2(d(\underline{x}_i, \underline{t}, \underline{C})) (\underline{x}_i - \underline{t})(\underline{x}_i - \underline{t})' - v_3(d(\underline{x}_i, \underline{t}, \underline{C})) \underline{C}\} = 0, \quad (2.43)$$

ahol v_1 , v_2 és v_3 alkalmasan választott valós függvények, valamint

$$d(\underline{v}, \underline{m}, \underline{M}) = \sqrt{(\underline{v} - \underline{m})' \underline{M}^{-1} (\underline{v} - \underline{m})} \quad (2.44)$$

az ún. Mahalanobis távolság. Amennyiben $\underline{M} = \underline{I}$, akkor a közönséges euklideszi távolságot kapjuk. Ahogy említettük, így affin ekvivariáns becslőket kapunk, de az összeomlási pont mindig

$$\varepsilon^* \leq \frac{1}{p+1} \quad (2.45)$$

lesz.

Törekvések indultak a nagy összeomlási ponttal és az affin ekvivarianciával egyaránt rendelkező többváltozós becslők megalkotására. Egymástól

függetlenül Stahel és Donoho PhD dolgozatukban közöltek egy lehetséges megoldást. Az $\underline{x}_i$ pont kiugró értéként való figyelembevételének mértékét határozták meg:

$$r_i = \sup_{\|\underline{u}\|=1} \frac{|\underline{u}' \underline{x}_i - \text{median}(\underline{u}' \underline{X})|}{\text{MAD}(\underline{u}' \underline{X})}, \quad (2.46)$$

ahol MAD (Median Absolute Deviation) a mediántól való abszolút eltérés:

$$\text{MAD}(\underline{v}) = \text{median}_{1 \leq i \leq n} |v_i - \text{median}_{1 \leq j \leq n} v_j|. \quad (2.47)$$

Legyen $w_i = w(r_i)$, ahol $w(r) \rightarrow 0$, ha $r \rightarrow \infty$ és $w(r)r$ korlátos, ekkor a becslő

$$\underline{t}_w(\underline{X}) = \frac{\sum_{i=1}^n w_i \underline{x}_i}{\sum_{i=1}^n w_i}. \quad (2.48)$$

Donoho megmutatta, hogy ez a becslő affin ekvivariáns és összeomlási pontja:

$$\varepsilon^*(\underline{t}_w, \underline{X}) = \frac{\lfloor \frac{n+1}{2} \rfloor - p}{n}, \quad (2.49)$$

ahol $\lfloor y \rfloor$ az a legnagyobb egész, mely y -nál kisebb vagy egyenlő vele.

Rousseeuw [110] újabb javaslatokat tett nagy összeomlási pontú affin ekvivariáns többváltozós paraméterbecslőkre. Ezek egyike a minimális térfogatú ellipszoid becslő:

$t_h(\underline{X})$ = azon legkisebb ellipszoid középpontja, amely az $\underline{X} = (x_1, \dots, x_n)$ pontok közül legalább h pontot tartalmaz ($p+1 \leq h \leq n$),

akkor

$$\varepsilon^*(t_h, \underline{X}) = \frac{\lfloor \frac{n}{2} \rfloor - p + 1}{n}, \text{ ha } h = \lfloor \frac{n}{2} \rfloor + 1 \quad (2.50)$$

és

$$\varepsilon^*(t_h, \underline{X}) = \lfloor \frac{n+p+1}{2} \rfloor, \text{ ha } h = \lfloor \frac{n-p+1}{2} \rfloor. \quad (2.51)$$

Egyváltozós esetben ez a becslő annak a legrövidebb intervallumnak a hosszával és középpontjával lesz azonos, melybe legalább h adatpont esik. Sajnos a helyparaméter becslő gyenge konvergenciával tart egy olyan határeloszláshoz, mely nem a normális eloszlás.

Ez a gyenge konvergencia javul a minimum kovariancia determináns becslő esetén:

$t_h(\underline{X})$ = az $\underline{X} = (x_1, \dots, x_n)$ mintából azon h pontok átlaga, melyekre a tapasztalati kovariancia mátrix determinánsa minimális ($p+1 \leq h \leq n$),

ekkor

$$\varepsilon^*(t_h, X) = \frac{\lfloor \frac{n}{2} \rfloor - p + 1}{n}, \text{ ha } h = \lfloor \frac{n}{2} \rfloor + 1 \quad (2.52)$$

és

$$\varepsilon^*(t_h, X) = \lfloor \frac{n + p + 1}{2} \rfloor, \text{ ha } h = \frac{\lfloor \frac{n-p+1}{2} \rfloor}{n}. \quad (2.53)$$

Egyváltozós esetben ez a becslő azon h db adatpont átlagát és varianciáját szolgáltatja, melyekre ez a variancia a legkisebb. Ez a becslő már aszimptotikusan normális $\sqrt{n}$ -es konvergencia sebességgel.

A többváltozós τ -becslő [80] azon t_n helyparaméter, mely minimalizálja a

$$|C| \left\{ \sum_{i=1}^n \rho_2(d(x_i, t, C)) \right\}, \quad (2.54)$$

figyelembe véve

$$\frac{1}{n} \sum_{i=1}^n \rho_1(d(x_i, t, C)) = b_1, \quad (2.55)$$

ahol $0 < b_1 < \sup \rho_1(\cdot)$. Az összeomlási pontja:

$$\varepsilon^*(t_n, X) = \frac{\lceil n r_1 \rceil}{n}, \quad (2.56)$$

ahol $r_1 = \frac{b_1}{\sup \rho_1(\cdot)}$. Ez utóbbi becslő rendelkezik a legjobb tulajdonságokkal, affin ekvivariáns, nagy összeomlási pontú, korlátos hatásfüggvényű, $\sqrt{n}$ -es konvergencia sebességgel normális eloszlású és megfelelő mértékben efficiens.

2.3 Lineáris fuzzy regresszió

A fuzzy elmélet lényegében Zadeh [135] korszkalkotó dolgozatával kelt életre, amelyben a $(c.f.(x) \in \{0, 1\})$ karakterisztikus függvény fogalmát általánosította a $(m.f.(x) \in [0, 1])$ tagsági függvény definiálásával. Így bizonyos szubjektív információk rendelkeznek a halmazhoz való tartozás megfogalmazásához.

Otto és Bandemer [102] fejlesztette ki az első robusztus fuzzy regressziós eljárást, melyet számos kutató vizsgált [56,57]. Otto és Bandemer [102] azonban csak az egyváltozós esetet mutatta be példákon keresztül. Közölték ugyan az általánosítás lehetőségét, melyet többszörös integrálok segítségével lehet megvalósítani, de az integrálási határok meghatározása során a

magasabb fokú (5 vagy a feletti) polinomok gyökeit analitikusan nem tudjuk megoldani, így az egyébként is számítás igényes feladat újabb numerikus feadattal bővülne. Probléma még az optimális paraméterek meghatározása, melyre Hu és munkatársai [56] javasolták a szimplex módszert. Sajnos ezen módszer csak a lokális maximumokat képes megtalálni, így más, globális optimum kereső eljárást kell keresni.

2.3.1 Az eredeti egyváltozós lineáris fuzzy regresszió módosítása

Az általánosítás előtt néhány kritikus megjegyzést kell tennünk az eredeti algoritmus kapcsán. A következő lineáris kalibrációs függvényt tételezzük fel:

$$y = m_1 x + m_2. \quad (2.57)$$

Azon ponthalmazt, melyen a tagsági függvény értéke nem nulla az M fuzzy halmaz tartójának hívjuk:

$$\text{supp}M = \{(x, y) \in X \times Y : m.f.M(x, y) > 0\}. \quad (2.58)$$

Ha a tartó egy ellipszis, a tagsági függvény:

$$m.f.i(x, y) = \left[1 - \frac{(x - x_i)^2}{u_i^2} - \frac{(y - y_i)^2}{v_i^2} \right]^+, \quad (2.59)$$

ahol u_i és v_i jelöli az ellipszis féltengelyeinek hosszát, és $[v]^+ = \max\{v, 0\}$.

Az M egyesített megfigyelés tagsági függvénye:

$$m.f.M(x, y) = \max_i m.f.i(x, y). \quad (2.60)$$

A megfigyelések fuzzy jellegét átvihetjük a paraméter térbe a várt számosság (*expected cardinality*) [40] alkalmas módosításával, amit relatív számosságnak (*relative cardinality*) nevezünk:

$$m.f.E(m_1, m_2) = \int_{G(m_1, m_2)} m.f.M(x, y) dx \quad (2.61)$$

$$G(m_1, m_2) = \{(x, m_1 x + m_2) \in X \times Y; x \in X\}. \quad (2.62)$$

Az $m.f.E(m_1, m_2)$ nyert értéket összehasonlíthatjuk a lineáris függvénykapcsolat legjobb helyi fuzzy közelítésével *fuzzy local (best) approximation*:

$$m.f.E(f) = \int_{G(f)} m.f.M(x, y) dx \quad (2.63)$$

$$G(f) = \{(x, f(x)) : x \in X\} \quad (2.64)$$

$$f(x) = \arg(y) \sup_{y \in Y} m.f.M(x, y). \quad (2.65)$$

Bevezethetjük az (m_1, m_2) paraméterpár tagsági függvényét:

$$m.f.c(m_1, m_2; f) = \frac{m.f.E(m_1, m_2)}{m.f.E(f)}. \quad (2.66)$$

Ábrázolva $m.f.c(m_1, m_2; f)$ értékeit a két paraméter függvényében, megkereshetjük a válaszfüggvény maximumát, melyet meghatározó két paraméter érték lesz a paraméterek fuzzy becslése.

Az említett kritikus észrevételek egyike az előbb vázolt egyváltozós esetre vonatkozik. Ha a tartók körök ($u_i = v_i$ minden i -re), akkor nincs probléma. Azonban nem természetes, hogy a koncentrációnál (x) jelentkező mérési hiba közel azonos a mért jel (y) hibájával. Valós kalibrációs problémák esetében a kalibrációs standardok beállítási hibája sokkal kisebb, mint a mérési jelhez tartozó hiba. Ez garantálja az ismeretlen oldatra vonatkozó pontosabb koncentráció becslést. Így az a természetesebb, ha tartónak ellipszist választunk, x -re vonatkozóan kisebb féltengelytávolsággal.

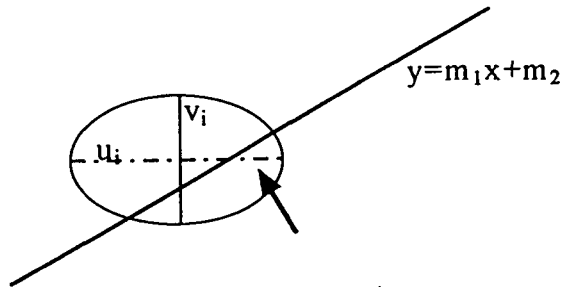
Ha a tartók ellipszisek a következő problémák merülnek fel. A 2.1 ábrán jól látható, hogy ha az Otto és Bandemer által javasolt görbét választjuk a helyi legjobb fuzzy közelítéshez, akkor $m.f.E(f)$ értéke szükségtelenül nagy lesz csökkentve ezzel $m.f.c(m_1, m_2; f)$ értékét, ami bizonytalanabb paraméterbecsléshez vezet.

A 2.2 ábrán két megoldást javasolunk, melyek közül mi az elsőt alkalmazzuk.

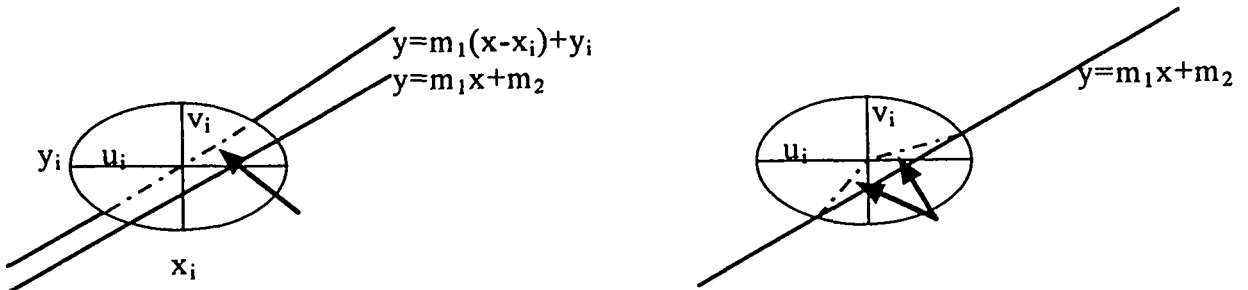
A 2.4 táblázatban szimulált adatokkal hasonlítjuk össze a két eljárást, ahol a FUZZY 1 az Otto és Bandemer által levezetett, a FUZZY 2 pedig az általunk módosított algoritmus.

A továbbiakban vizsgáljuk meg vizes alkoholminták gázkromatográfián meghatározott méréseinek kalibrációs eredményét. A 2.5 táblázatban 9, a 2.6 táblázatban a 9-hez további 3 mérési eredményt és az adott módszer által kapott paraméterekkel számított százalékos eltéréseket tüntettük fel. A táblázatok végén szerepelnek a két módszerrel meghatározott paraméterpár értékei.

Az első 9 minta a laboratóriumban készült, törzsoldat megfelelő hígításával, míg az utolsó 3 minta az OMH által kiadott standard minta volt.



2.1 ábra: Otto és Bandemer által javasolt gráf a helyi legjobb fuzzy közelítéshez



2.2 ábra: Két általunk javasolt gráf, melyekkel elkerülhető a szövegben említett probléma

konc. x	jel y	eltérés		
		LS	FUZZY 1	FUZZY 2
1	1.1	0.32	-0.02	0.00
2	2.0	-0.04	-0.04	0.00
3	3.1	-0.20	0.14	0.20
4	3.8	-0.76	-0.08	0.00
5	6.5	0.68	1.70	1.80
	m_1	1.26	0.92	0.90
	m_2	-0.48	0.20	0.20

2.4 táblázat: Szimulált adatokra illesztett egyenes paraméterei és a számított értékek eltérései

Ebben az esetben úgy vehetjük, hogy a hígított minták sorozatát *szennyeztük* a standardokkal. A számításokból jól látszik, hogy a fuzzy regresszió robusztus tulajdonsága révén kevésbé volt érzékeny a perturbációra.

2.3.2 Lineáris fuzzy regresszió általánosítása geometriai úton

Az egyik legtermészetesebb módja az általánosításnak a többszörös integrálok alkalmazása. A már említett okok miatt azonban a többváltozós geometriai ismeretek használata kivitelezhetőbb algoritmust szolgáltat. Sajnos az n -dimenziós geometriai fogalmakkal kellő mélységben csak két könyvben [122,70] találkozhatunk, így rengeteg saját fejlesztést kellett végeznünk a működő algoritmus összeállításához.

Az általánosításhoz néhány természetes megszorítással kellett élni. A mérési pontok bizonytalanságának jellemzéséhez hiperellipszoid tartókat választottunk, tehát a pontokhoz rendelhető tagsági függvény hiperparaboloid lett. Felírhatjuk a lineáris modellt az y 'függő' és az $x_1, x_2, x_3, \dots, x_d$ (vagy röviden jelölve $\underline{x}$) 'független' változók között. Mivel általában minden változó kisebb-nagyobb mérési hibával terhelt, így $\underline{x}$ -et inkább magyarázó (*explanatory*) változónak, y -t pedig magyarázott (*explained*) változónak hívjuk. Tehát a beállítható $\underline{x}$ változó magyarázza a hatására kialakult y változót. Ezekután a lineáris modell:

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + \dots + m_{d-1} x_{d-1} + m_{d+1}, \quad (2.67)$$



mintaszám	koncentráció x	jel (terület arány) y	százalékos eltérés	
			LS	FUZZY 2
1.	0.244	0.2407	3.88	0.75
2.	0.272	0.2758	5.32	2.61
3.	0.478	0.4781	-0.42	-1.89
4.	0.832	0.8423	-1.68	-2.43
5.	0.928	0.9654	0.72	0.08
6.	0.952	0.9649	-1.98	-2.61
7.	1.976	2.0657	-0.33	-0.52
8.	2.914	3.0819	0.40	0.33
9.	3.698	3.9018	-0.03	-0.05
			m_1	1.063
			m_2	-0.028
				1.061
				-0.020

2.5 táblázat: A Bűnügyi Technikai Intézetben előállított standardok által meghatározott kalibrációs egyenes paraméterei és a számított értékek százalékos eltérései

mintaszám	koncentráció x	jel (terület arány) y	százalékos eltérés	
			LS	FUZZY 2
1.	0.244	0.2407	2.27	0.75
2.	0.272	0.2758	3.68	2.61
3.	0.478	0.4781	-2.40	-1.89
4.	0.832	0.8423	-3.81	-2.43
5.	0.928	0.9654	-1.38	0.08
6.	0.952	0.9649	-4.14	-2.61
7.	1.976	2.0657	-2.53	-0.52
8.	2.914	3.0819	-1.80	0.33
9.	3.698	3.9018	-2.25	-0.05
10.	1.000	1.0932	3.31	4.77
11.	2.000	2.2382	4.21	6.09
12.	3.000	3.3828	4.49	6.50
			m_1	1.061
			m_2	-0.020

2.6 táblázat: Az előbbi táblázatban szereplő standardok és az OMH által előállított standardok felhasználásával kapott kalibrációs egyenes paramétereit és a számított értékek százalékos eltéréseit

ahol m_i az i . együttható vagy paraméter és a feladat az összes m_i becslése a kísérleti adatok felhasználásával.

(2.67) átírható két vektor szorzására egy hipersík egyenletét eredményezve:

$$\begin{pmatrix} x_1, & x_2, & x_3, & \dots, & x_{d-1}, & x_d \equiv y, & 1 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_{d-1} \\ m_d \equiv -1 \\ m_{d+1} \end{pmatrix} = 0. \quad (2.68)$$

A (2.68) kifejezés segítségével felírható a hipersík normál egyenlete is, felhasználva az n_i iránycosinusokat:

$$\begin{pmatrix} x_1, & x_2, & x_3, & \dots, & x_{d-1}, & x_d \equiv y, & 1 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ \vdots \\ n_{d-1} \\ n_d \\ n_{d+1} \equiv p \end{pmatrix} = 0, \quad (2.69)$$

ahol

$$n_i = \cos \alpha_i = \mu m_i \quad \text{ha } i = 1, \dots, d \quad (2.70)$$

és

$$n_{d+1} = p = \mu m_{d+1} = -\frac{|m_{d+1}|}{\sqrt{\sum_{i=1}^d m_i^2}}, \quad (2.71)$$

ahol

$$\mu = -\text{sign}(m_{d+1}) \frac{1}{\sqrt{\sum_{i=1}^d m_i^2}}.$$

$|p|$ a hipersík távolsága a Descartes-féle derékszögű koordináta-rendszer origójától. A (2.69) egyenlet által megadott hipersík vagy metszi, vagy nem

metszi a d -dimenziós hiperellipszoidot, ami a mért pontok bizonytalanságát fejezi ki. A metszet újra hiperellipszoid lesz, de $d - 1$ dimenziós [122]. Ezek a csökkentett dimenziójú hiperellipszoidok lesznek az alapjai annak d dimenziós hiperparaboloidnak, melyet a $d + 1$ dimenziós hiperparaboloid tagsági függvény metszeteként kaptunk. Összegezve ezen d -dimenziós hiperparaboloidok térfogatait kapjuk az $m.f.E(\underline{m}^*)$ relatív számosság értékét [102] egy bizonyos $m_1^*, m_2^*, m_3^*, \dots, m_{d+1}^*$ paraméter halmazt vizsgálva. Az $m.f.E(f)$ helyi legjobb fuzzy közelítést [102] azon d -dimenziós hiperparaboloid térfogatok összegzésével nyerjük, melyeket akkor kapunk, ha a hipersíkot minden egyes mérési ponthoz eltoljuk, ezzel a térfogatösszegek lehetséges legnagyobb értékét szolgáltatva. Mivel $m.f.E(f) > 0$ és $m.f.E(\underline{m}) \leq m.f.E(f)$ minden $\underline{m}$ -re, számolhatjuk

$$m.f.c(\underline{m}; f) \equiv \frac{m.f.E(\underline{m})}{m.f.E(f)} \quad (2.72)$$

értéket, melynek maximális értékét meghatározó paraméter halmaz lesz a fuzzy becslés, így a megfigyelések fuzzy jellegét a paraméter térbe vittük át.

Minden elméleti vonatkozást áttekintettünk ahhoz, hogy formába öntessük a számításokhoz szükséges algoritmust. Egy d -dimenziós hiperparaboloid térfogatát, melynek alapja egy $d - 1$ -dimenziós hiperellipszoid (a levezetések mellőzésével) a következő két eset megkülönböztetésével nyerjük, ha $d - 1$ páros, azaz $d - 1 = 2n$:

$$\frac{h(\pi h)^n}{(n + 1)!} \prod_{j=1}^{2n} u_j \quad (2.73)$$

és ha $d - 1$ páratlan, azaz $d - 1 = 2n + 1$:

$$\frac{h\sqrt{h}(\pi h)^n (n + 1)! 2^{(2n+3)}}{(2n + 3)!} \prod_{j=1}^{2n+1} u_j, \quad (2.74)$$

ahol h a hiperparaboloid magassága és u_i jelenti a hiperellipszoid féltengelyeinek hosszát.

Az i . méréshez tartozó d -dimenziós hiperellipszoid egyenlete:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \\ 1 \end{pmatrix}' \begin{pmatrix} \frac{1}{v_{i1}^2} & 0 & \cdots & 0 & -\frac{x_{i1}}{v_{i1}^2} \\ 0 & \frac{1}{v_{i2}^2} & \cdots & 0 & -\frac{x_{i2}}{v_{i2}^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{v_{id}^2} & -\frac{x_{id}}{v_{id}^2} \\ -\frac{x_{i1}}{v_{i1}^2} & -\frac{x_{i2}}{v_{i2}^2} & \vdots & -\frac{x_{id}}{v_{id}^2} & \left(\sum_{j=1}^d \frac{x_{ij}^2}{v_{ij}^2} - 1 \right) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \\ 1 \end{pmatrix} = 0, \quad (2.75)$$

ahol v_{ij} az i . mérésnél a j . változó bizonytalanságát jelöli (az i . hiperellipszoid j . féltengelyének hossza) és x_{ij} az i . mérésnél a j . változó megfigyelt értéke.

A megfelelő módon választott $\underline{T}$ transzformációs mátrix (lásd pld. [38]-ben 4 dimenziós esetre) elemei csak a (2.69)-ben szereplő iránycosinusok értékeitől függ. Alkalmazva ezt a transzformációs mátrixot (2.75)-re, $\underline{x}'\underline{T}'\underline{V}\underline{T}\underline{x} = 0$ -t nyerjük $\underline{x}'\underline{V}\underline{x} = 0$ -ból.

$\underline{T}'\underline{V}\underline{T}$ jelöljük $\underline{M}(d)$ -vel és általános elemekkel felírva kapjuk

$$\underline{M}(d) = \left(\begin{array}{cccc|ccc} a_{11} & a_{12} & \cdots & a_{1\,d-1} & a_{1d} & a_{1\,d+1} \\ a_{21} & a_{22} & \cdots & a_{2\,d-1} & a_{2d} & a_{2\,d+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{d-1\,1} & a_{d-1\,2} & \cdots & a_{d-1\,d-1} & a_{d-1\,d} & a_{d-1\,d+1} \\ \hline a_{d1} & a_{d2} & \cdots & a_{dd-1} & a_{dd} & a_{dd+1} \\ \hline a_{d+1\,1} & a_{d+1\,2} & \cdots & a_{d+1\,d-1} & a_{d+1\,d} & a_{d+1\,d+1} \end{array} \right). \quad (2.76)$$

A hipersík egy olyan $d - 1$ -dimenziós elforgatott hiperellipszoidot metsz ki melyet $\underline{M}(d - 1)$ -el jellemezhetünk:

$$\underline{M}(d - 1) =$$

$$\left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1\,d-1} & (pa_{1d} + a_{1\,d+1}) \\ a_{21} & \cdots & a_{2\,d-1} & (pa_{2d} + a_{2\,d+1}) \\ \vdots & \ddots & \vdots & \vdots \\ a_{d-1\,1} & \cdots & a_{d-1\,d-1} & (pa_{d-1\,d} + a_{d-1\,d+1}) \\ \hline (pa_{d1} + a_{d+1\,1}) & \cdots & (pa_{dd-1} + a_{d+1\,d-1}) & (p^2a_{dd} + p(a_{dd+1} + a_{d+1\,d}) + a_{d+1\,d+1}) \end{array} \right). \quad (2.77)$$

Főtengely-transzformációval meghatározhatjuk az $\underline{M}(d - 1)$ mátrix λ_{ij} sajátértékeit. Ezen sajátértékek segítségével kiszámolhatjuk a kimetszett

$d-1$ -dimenzós hiperellipszoid féltengelyeinek hosszát és a középpontját meghatározó $x_{ij}^\ominus$ koordináták értékeit (az $\ominus$ jelzést használjuk a $d-1$ -dimenziós hiperellipszoid középpontjának megkülönböztetésére a d -dimenzióstól):

$$u_{ij} = \sqrt{\frac{-(p^2 a_{dd} + p(a_{dd+1} + a_{d+1d}) + a_{d+1d+1})}{\lambda_{ij}}}, \quad (2.78)$$

$$x_{ij}^\ominus = \frac{p a_{jd} + a_{j\ n+1}}{\sqrt{-\lambda_{ij} (p^2 a_{dd} + p(a_{dd+1} + a_{d+1d}) + a_{d+1d+1})}}. \quad (2.79)$$

A hiperparaboloid h magasságát (2.79) és az i . méréshez tartozó tagsági függvényt felhasználva nyerjük:

$$h_i = \left[1 - \sum_{j=1}^d \frac{(x_{ij}^\ominus - x_{ij}^*)^2}{v_{ij}^2} \right]^+, \quad (2.80)$$

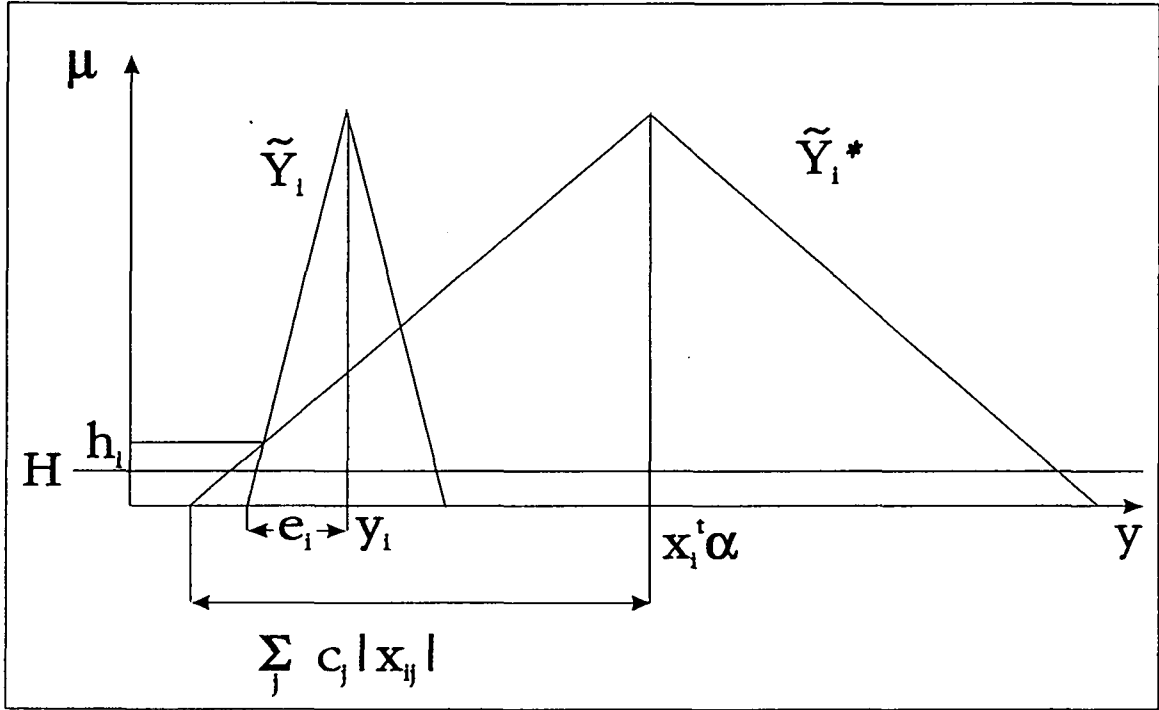
ahol $\underline{x}^*$ az eredeti mérési pont $\underline{T}$ által transzformált koordinátáját jelenti, $x_{id}^\ominus = p$ és $[v]^+ = \max\{v, 0\}$.

A fentebb vázolt módon kiszámolható a $m.f.c(\underline{m}; f)$ értéke bármely $\underline{m}$ paraméter vektorra. Sajnos nincs direkt út $m.f.c(\underline{m}; f)$ maximumának meghatározására. A *genetikus algoritmus* [81] globális optimum kereső eljárással már végeztünk néhány biztató kísérletet, így a közel jövőben szándékozunk egy olyan programcsomagot összeállítani, melyben a többváltozós lineáris fuzzy regresszió által becsült optimális paramétervektort a genetikus algoritmus segítségével keressük meg.

2.4 Fuzzy lineáris regresszió

A Tanaka és munkatársai [126] által kidolgozott fuzzy lineáris regressziós eljárásról kiderült, hogy nem robusztus, így kidolgoztuk egy módosított változatát. A következő iteratív eljárást használjuk. Az y_i adatok számára tagsági függvényeket származtathatunk az előzőleg kiszámolt paraméterek α_j centrumai és c_j terjedelmei felhasználásával. Ezeket az értékeket felhasználva az (y_i, e_i) fuzzy kimenő adatokat határozhatjuk meg, ahol e_i terjedelmeket jelent. Talán meglepő, de ez utóbbi nem fordítottan arányos az y_i tagsági értékével, ami jól látható a 2.3 ábrán.

$$e_i = \left[\sum_{j=1}^p c_j |x_{ij}| - \frac{|y_i - \underline{x}_i^t \underline{\alpha}|}{1 - H} \right]^+ \quad (2.81)$$



2.3 ábra: A fuzzy kimenő adatok e_i terjedelme és az y_i tagsági értékek fordított arányú kapcsolatának illusztrálása

és nagyobb értékéhez y_i nagyobb tagsági értéke tartozik. Az első lépésben nem fuzzy kimenő adatokkal meghatározzuk a fuzzy paramétereket, majd e_i -ket számítjuk mindig az előző lépés eredményéből.

A fuzzy regressziós paramétereket fuzzy kimenő adatokkal a következő lineáris programozási feladat megoldásaként adhatjuk meg [127]:

$$\begin{aligned}
 &\text{Minimalizálandó} && s = \sum_{j=1}^p c_j \\
 &\text{Feltételek} && (1 - H) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}_i^t \underline{\alpha} \geq y_i + (1 - H)e_i, \quad (2.82) \\
 &&& (H - 1) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}_i^t \underline{\alpha} \leq y_i + (H - 1)e_i, \\
 &&& c_i \geq 0, \\
 &&& i = 1, 2, \dots, n.
 \end{aligned}$$

A módosított eljárás robusztusságát az irodalomban nemrégén közölt atomabszorpciós spektrometriásan mért adatsoron [56] szemléltetjük.

Adatsor	Elem	Koncentráció (ppb)	Jel
2.	Al	0	0.000
		20	0.084
		40	0.194
		60	0.245
		80	0.330
5.	Cu	20	0.097
		40	0.182
		60	0.256
		80	0.331
		100	0.417
8.	Mn	0	0.000
		10	0.110
		20	0.226
		40	0.430
		60	0.637
		80	0.765
		100	0.870

2.7 táblázat: Atomabszorpciós spektrometriásan mért adatsor az irodalomból átvéve

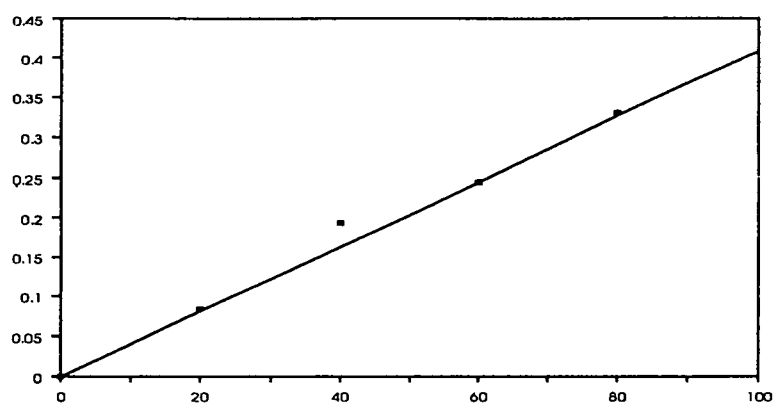
		Adatok (jelzés az ábrán)		
		2. adatsorra (a)	5. adatsorra (b)	8. adatsorra (c)
1. iteráció	$a_0 \pm c_0$	0.0 ± 0.0	0.019 ± 0.0	0.0 ± 0.0
	$a_1 \pm c_1$	0.004466 ± 0.0004259	0.003988 ± 0.0000972	0.01000 ± 0.001444
	$\mu_{Y_1^c}$	1.0000	0.1000	1.0000
	$\mu_{Y_2^c}$	0.3731	0.1000	0.3077
	$\mu_{Y_3^c}$	0.1000	0.6143	0.1000
	$\mu_{Y_4^c}$	0.1000	0.1000	0.4808
	$\mu_{Y_5^c}$	0.1970	0.9228**	0.5731
	$\mu_{Y_6^c}$			0.6971
2. iteráció	$a_0 \pm c_0$	0.0 ± 0.0		0.0 ± 0.0
	$a_1 \pm c_1$	0.004083 ± 0.007758		0.008700 ± 0.01653
	$\mu_{Y_1^c}$	1.0000		1.0000
	$\mu_{Y_2^c}$	0.9850		0.8609
	$\mu_{Y_3^c}$	0.9012		0.8427
	$\mu_{Y_4^c}$	1.0000		0.8760
	$\mu_{Y_5^c}$	0.9946*		0.8840
	$\mu_{Y_6^c}$			0.9470
3. iteráció	$a_0 \pm c_0$			0.0 ± 0.0
	$a_1 \pm c_1$			0.01037 ± 0.01056
	$\mu_{Y_1^c}$			1.0000
	$\mu_{Y_2^c}$			0.9619
	$\mu_{Y_3^c}$			0.9437
	$\mu_{Y_4^c}$			0.9770
	$\mu_{Y_5^c}$			0.9851
	$\mu_{Y_6^c}$			0.9511
	$\mu_{Y_7^c}$			0.8990*

* A leállási feltétel akkor teljesült, mikor a mérések több mint felének a tagsági függvény értéke elérte vagy meghaladta a 0.9 (= H) értéket.

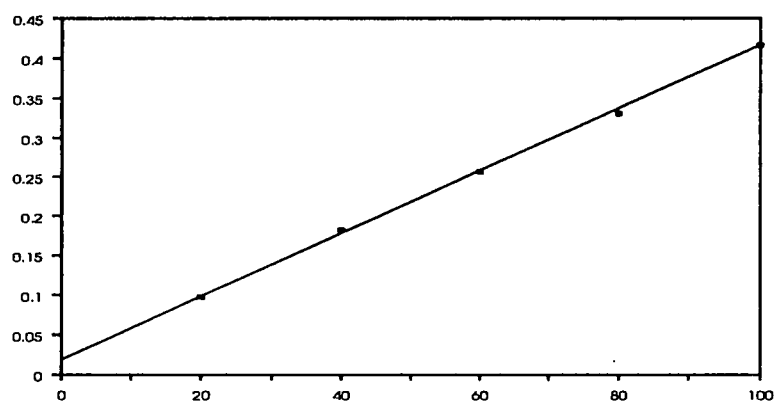
** A leállási feltétel akkor teljesült, mikor a mérések több mint felének a tagsági függvény értéke elérte vagy meghaladta a 0.1 (= H) értéket.

2.8 táblázat: A fuzzy lineáris regresszió eredménye és a módosított számításmenet illusztrálása

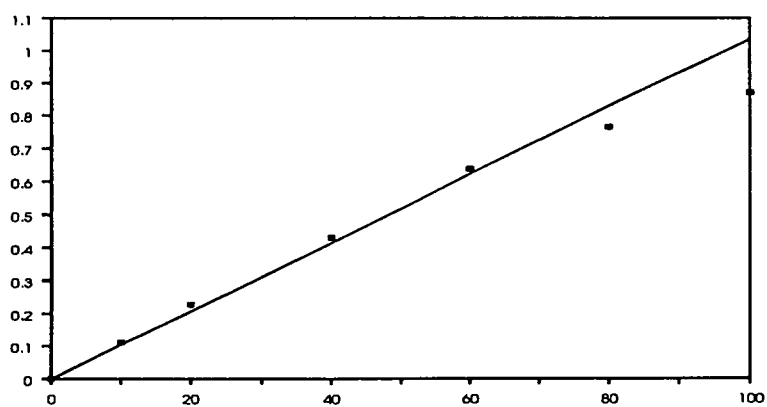
a



b



c



2.4 ábra: Fuzzy lineáris regresszióval kapott kalibrációs egyenesek ábrái

A 2.4 ábrán megfigyelhető a módosított fuzzy lineáris regresszió robusztus tulajdonsága. A irodalomban nem régen jelent meg egy eljárás a linearitás tesztelésére [130], de úgy gondoljuk, hogy a mi eljárásunk könnyen alkalmazható többváltozós esetben is és nincs korlátozás az adatok számára vonatkozóan. Egynél több iterációs ciklus már természeténél fogva jelzi a linearitás sérülését, azonban a robusztus tulajdonság miatt a kalibrációs adatok felhasználhatók maradnak.

3. fejezet

Kalibráció során nyert adatok jellemzése

3.1 Konfidencia intervallum Gauss-eloszlás teljesülése esetén

Az $y = a_0 + a_1 x$ egyváltozós lineáris függvénykapcsolat esetén az egyenes $\hat{a}_1$ meredekségét és $\hat{a}_0$ tengelymetszetét kell becsülnünk az n db mérési pontból:

$$\hat{a}_1 = \frac{n \sum x_j y_j - \sum x_j \sum y_j}{n \sum x_j^2 - (\sum x_j)^2} \quad (3.1)$$

$$\hat{a}_0 = \frac{\sum x_j^2 \sum y_j - \sum x_j \sum x_j y_j}{n \sum x_j^2 - (\sum x_j)^2}, \quad (3.2)$$

melyek a későbbi vizsgálódásokhoz felírhatók másképpen is:

$$\hat{a}_1 = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sum (x_j - \bar{x})^2} \quad (3.3)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}, \quad (3.4)$$

ahol $\bar{x} = \frac{\sum x_j}{n}$ és $\bar{y} = \frac{\sum y_j}{n}$.

Mivel a feltevések miatt $\text{Var}(y_j) = \sigma^2$ minden j -re és $\text{Var}(\bar{y}) = \frac{\sigma^2}{n}$, a fenti paraméterek varianciái:

$$\text{Var}(\hat{a}_1) = \text{Var} \left(\frac{\sum (x_j - \bar{x}) y_j}{\sum (x_j - \bar{x})^2} \right) = \frac{\sum (x_j - \bar{x})^2}{(\sum (x_j - \bar{x})^2)^2} \text{Var}(y_j) = \frac{\sigma^2}{\sum (x_j - \bar{x})^2} \quad (3.5)$$

$$\text{Var}(\hat{a}_0) = \text{Var}(\bar{y} - \hat{a}_1 \bar{x}) = \text{Var}(\bar{y}) + \text{Var}(\hat{a}_1 \bar{x}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_j - \bar{x})^2} \right). \quad (3.6)$$

A kalibrációhoz felhasznált x_i értéknél a becsült $\hat{y}_i$ varianciájának számításához az $\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$ helyett a $\hat{y}_i = \bar{y} + \hat{a}_1(x_i - \bar{x})$ összefüggést használjuk, mivel $\hat{a}_0$ és $\hat{a}_1$ nem független becslések.

$$\text{Var}(\hat{y}_i) = \text{Var}(\bar{y}) + (x_i - \bar{x})^2 \text{Var}(\hat{a}_1) = \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2} \right) \quad (3.7)$$

A kalibrációs függvényt általában a mért jelek alapján történő koncentráció becslésére használják. Tekintsük a következő összefüggést:

$$\tilde{y} = \bar{y} + \hat{a}_1(x - \bar{x}) + \varepsilon, \quad (3.8)$$

ahol $\tilde{y} = \frac{\sum y_i}{m}$, azaz az ismeretlen koncentrációjú oldatra vonatkozó m mérés átlaga. Az ε mérési hiba varianciája:

$$\text{Var}(\varepsilon) = \text{Var}(\tilde{y}) + \text{Var}(\bar{y}) + \text{Var}(\hat{a}_1(x - \bar{x})) = \sigma^2 \left(\frac{1}{m} + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_j - \bar{x})^2} \right). \quad (3.9)$$

A fenti egyenletekben szereplő σ^2 a gyakorlatban legtöbbször nem ismert, így annak becslését alkalmazzuk:

$$\sigma^2 \approx s^{*2} = \sqrt{\frac{\sum (y_j - \hat{a}_0 - \hat{a}_1 x_j)^2}{n - 2}}. \quad (3.10)$$

Amennyiben ε Gauss-eloszlású mérési hibát jelent a következő hányados Student-eloszlású lesz:

$$t = \frac{\tilde{y} - \bar{y} - \hat{a}_1(x - \bar{x})}{s^* \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_j - \bar{x})^2}}}. \quad (3.11)$$

$(1 - \alpha)$ biztonsági szint mellett

$$t \leq |t_\alpha| \quad (3.12)$$

teljesül, ahol t_α az α szignifikancia szinthez és $n-2$ szabadsági fokhoz tartozó Student-eloszlás táblázatbeli értéke.

t értékét (3.11) alapján (3.12)-be helyettesítve, majd négyzetre emelés után a következő másodfokú egyenlőtlenséget nyerjük

$$\left(\hat{a}_1^2 - \frac{s^{*2} t_\alpha^2}{\sum (x_j - \bar{x})^2} \right) (x - \bar{x})^2 - 2\hat{a}_1(\tilde{y} - \bar{y})(x - \bar{x}) + (\tilde{y} - \bar{y})^2 - s^{*2} t_\alpha^2 \left(\frac{1}{m} + \frac{1}{n} \right) \leq 0, \quad (3.13)$$

melynek megoldásaként kapott konfidencia intervallum [18,134]:

$$x < \left| t_\alpha \frac{s^*}{\gamma} \sqrt{\frac{\gamma}{\hat{a}_1} \left(\frac{1}{m} + \frac{1}{n} \right) + \frac{(\tilde{y} - \bar{y})^2}{\hat{a}_1^2 \sum (x_j - \bar{x})^2}} \right| + \bar{x} + \frac{\tilde{y} - \bar{y}}{\gamma}, \quad (3.14)$$

ahol

$$\gamma = \hat{a}_1 - \frac{s^{*2} t_\alpha^2}{\hat{a}_1 \sum (x_j - \bar{x})^2}. \quad (3.15)$$

Amennyiben (3.15)-ben a második tag közel nulla, azaz $\gamma \approx \hat{a}_1$ a következő elterjedtebb és egyszerűbb formula használható

$$x < \hat{x} + \left| t_\alpha \frac{s^*}{\hat{a}_1} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{\sum (x_j - \bar{x})^2}} \right|, \quad (3.16)$$

ahol $\hat{x}$ a kalibrációs egyenes által becsült koncentráció:

$$\hat{x} = \bar{x} + \frac{\tilde{y} - \bar{y}}{\hat{a}_1}. \quad (3.17)$$

3.2 Konfidencia intervallumok szerkesztése bootstrap módszerrel

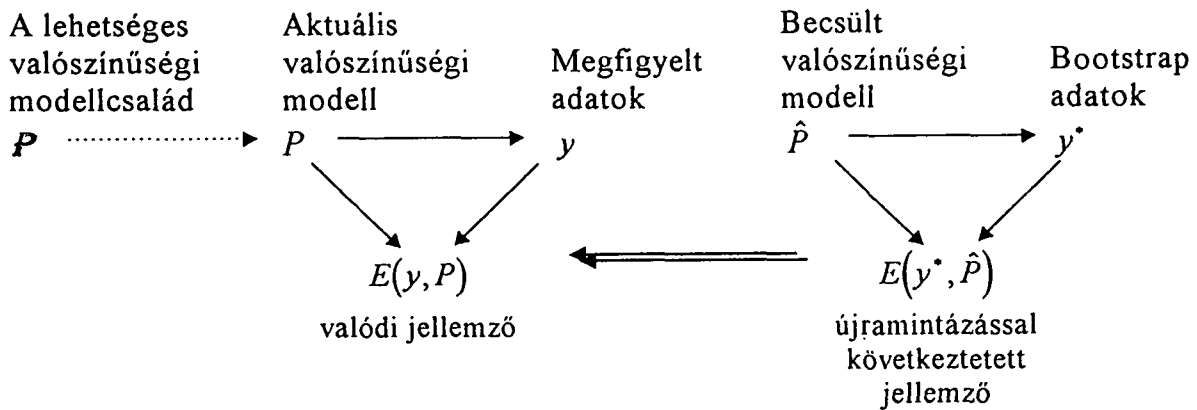
Az előző szakaszban a legkisebb négyzetek módszerével meghatározott paraméterekkel és a mérési hiba eloszlására vonatkozóan a Gauss-eloszlás feltételezésével konfidencia intervallumot vezettünk le az ismeretlen koncentráció becsülésének jellemzésére. A másodfajú hiba kezelésére bemutatott robusztus eljárások alkalmazása esetén nem használhatjuk ezeket az eredményeket, hiszen a feltételezett eloszlástól (amely nem is mindig a Gauss-eloszlás) kis-mértékű eltérés megengedett.

Ha bizonyos statisztikai leírás egzakt megoldása nem létezik, akkor választanunk kell a teoretikus approximáció (pld. sorbafejtés) és a tiszta numerikus approximáció (pld. Monte Carlo módszerek) között. A bootstrap

módszer [42,82] legfontosabb tulajdonsága, hogy releváns statisztikai jellemzők szimulációja, úgymint variancia, torzítás, konfidencia intervallum, stb. minimális feltételek betartása mellett.

A gyakorlatban a kalibráció elvégzéséhez szükséges mérések száma erősen korlátozott a költséges vegyszerek használata ill. az időigényes meghatározások miatt. Így csak csekély számú mérést lehet elvégezni. A kisszámú minta értékeléséhez nem használhatjuk a normalitást megkövetelő statisztikai módszereket. A kisszámú minták jellemzésére alkalmas, sztochasztikus approximáción alapuló bootstrap eljárás alkalmazása kínálkozik megoldásul.

Az 3.1 ábrán a bootstrap eljárás alkalmazásának sémáját mutatjuk be. A valódi P valószínűségi modell $E[y, P]$ várható értékét becsüljük az $E[y^*, \hat{P}]$ bootstrap becsléssel. A kritikus lépést kettős nyílal jeleztük.



3.1 ábra: Séma a bootstrap eljárás alkalmazására

A regresszióanalízisnél a bootstrap eljárást kétféle újramintázási stratégiával alkalmazhatjuk:

1. Újramintázás reziduálisokon alapulva

Tekintük a következő regressziós modellt: $y_i = f(x_i, \beta) + \varepsilon_i$, ahol y_i a függő változó (mérési jel), x_i a független változó (koncentráció), ε_i a mérési hibát reprezentáló véletlen változó. Először a $\hat{\beta}$ paramétert kell meghatározni a legkisebb négyzetek, vagy más pld. robusztus paraméterbecslő eljárással. Ezután az illesztett értékek számolhatók: $\hat{y}_i = f(x_i, \hat{\beta})$. Ha feltételezzük, hogy az ε_i hibák homoszkedasztikusak és az F eloszlás függvényük becsülhető a reziduálisok $\hat{F}$ tapasztalati eloszlásával, akkor $\{(x_i, y_i^*), i = 1, \dots, n\}$ szimulált adatokat állíthatunk elő $y_i^* = \hat{y}_i + \varepsilon_i^*$ alkalmazásával, ahol ε_i^* -t visszatevéses mintavétellel választjuk az eredeti $\{\varepsilon_i, i = 1, \dots, n\}$ reziduáli-

sok közül, ahol természetesen $\hat{\varepsilon}_i = y_i - \hat{y}_i$. Az új adatokkal m alkalommal elvégezve az illesztést a $\hat{\beta}^*$ és $\hat{y}_i^*$ bootstrap becsléseket kapjuk.

2. Újramintázás adatpárokon alapulva

Ezt az eljárást ajánlott, ha a hibák heteroszkedasztikusak. Ebben az esetben az (x_i, y_i) adatpárokat tekintjük, melyet az F együttes eloszlás ír le. Az (x_i^*, y_i^*) bootstrap mintákat az eredeti $\{(x_i, y_i), i = 1, \dots, n\}$ adatpárok közül véletlenszerűen kiválasztva nyerjük. Meg kell említenünk ennek az újramintázási módszernek néhány hátrányos tulajdonságát. A $Var(\varepsilon_i)$ változhat x_i ill. $f(x_i, \beta)$ -val. Néhány statisztikai jellemző függ a $D = (x_1, \dots, x_n)$ tervezett beállítástól, azonban a szimulált adatpárok különbözni fognak D -től. Végül gyakran előfordul, a sikertelen újramintázás (bizonyos adatpárok túl gyakran szerepelnek a mintában), aminek következtében a paraméterek becslése nem vitelezhető ki a szingularitás miatt.

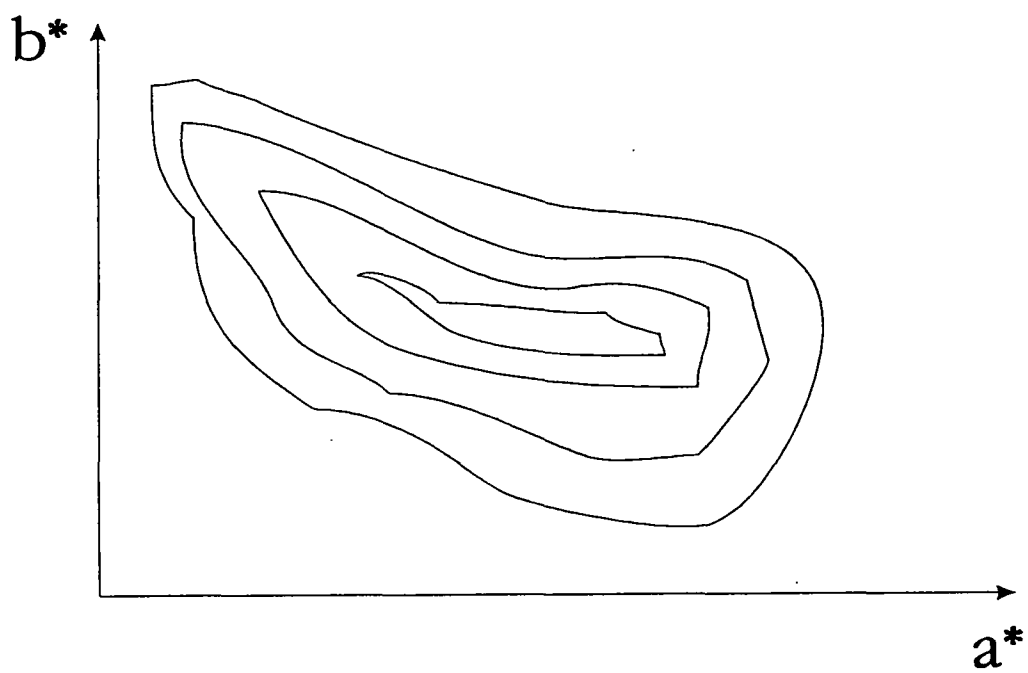
A bootstrap módszer regressziós függvények paramétereinek jellemzésére történő alkalmazásával a matematikai statisztikai irodalomban számtalan példával találkozunk [16,112,41,43,125,115,36,33]. Az analitikai kémiában történő felhasználásra is történtek próbálkozások [24,54,2], főleg az illesztő függvények paramétereinek korrelációját vizsgálták ill. a paraméterek eloszlását és konfidencia intervallumait határoztak meg. Bemutatjuk a 3.2 és a 3.3 ábrákat a bootstrap újramintázási algoritmus által nyerhető információk szemléltetésére.

Bonate [19] írta le először a becsült koncentrációk jellemzését bootstrap módszerrel meghatározott variancia becslések segítségével.

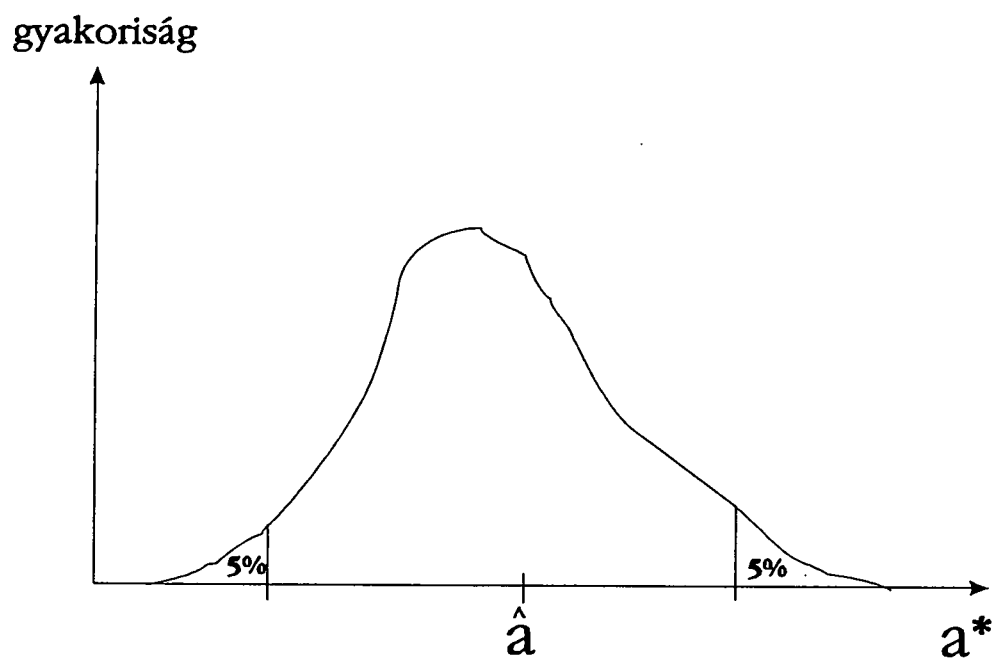
A következőkben olyan eljárást ismertetünk, melynek során konfidencia intervallumok határozhatók meg a becsült koncentrációkra vonatkozóan. A kalibrációs függvény paramétereit különböző robusztus becslőkkel határoztuk meg és a bootstrap módszer segítségével tapasztalati sűrűségfüggvényük felhasználásával nyertük a kívánt konfidencia intervallumokat.

Nézzük meg részletesen ezen *kiterjesztett bootstrap technika* alkalmazását.

Az alkalmazott paraméterbecslő eljárások mindegyikét a 2. fejezetben már korábban ismertettük, így csak felsoroljuk őket: LS, IRLS6, IRLS9, CM, RM, LSM. Ezekkel a becslőkkel a reziduálison alapuló bootstrap újramintázással 3000 paraméterpárt határoztunk meg. Ezeket a paraméter becsléseket használjuk az alkalmas koncentráció becslés megkeresésére a *tapasztalati maximum likelihood* módszer segítségével. Egy bizonyos koncentrációhoz



3.2 ábra: 3000 bootstrap mintából meghatározott a^* és b^* paraméterek közötti korreláció



3.3 ábra: 3000 bootstrap mintából számított a^* paraméter eloszlása és a 90%-os megbízhatósági szintű konfidencia intervallum

tartozó mérési jel sűrűségfüggvénye meghatározható a *nemparametrikus kernel sűrűségfüggvény becselővel* [61]

$$\hat{f}(t) = \frac{1}{3000h} \sum_{B=1}^{3000} K\left(\frac{t - y^B}{h}\right), \quad (3.18)$$

felhasználva, hogy

$$y^B = a_0^B + a_1^B x, \quad (3.19)$$

ahol az $\frac{1}{\pi} \frac{h}{h^2 + (t - y_i)^2}$ Cauchy-eloszlás sűrűségfüggvényét alkalmaztuk $K(\cdot)$ kernel függvénynek, h a skálaparaméter becslése, és a_0^B és a_1^B a B . bootstrap paraméterpár. Azért kell a Cauchy-eloszlást alkalmazni, mert ezen eloszlás széles szárnyaival az esetleg előforduló extrém értékeket is kellően elsimítja. Ráadásul a Cauchy-eloszlás a Student-eloszlások családjának 1 szabadsági fokhoz tartozó tagja, és a Student-eloszlásokat az előző szakasz tanúsága szerint a klasszikus statisztikai módszerekkel meghatározott konfidencia intervallumoknál alkalmazzák.

A h értékét a következő robusztus skálaparaméter becselővel számoljuk [111]:

$$h = 1.4826 \left(1 + \frac{5}{n-1}\right) m_T, \quad (3.20)$$

ahol $2m_T$ azon legkisebb tartomány, amely az adatok legalább felét tartalmazza.

A t_e jóslt mérési jelet úgy választjuk meg, hogy az $\hat{f}(t_j)$ értékek szorzata a lehető legnagyobb legyen:

$$\max \prod_{j=1}^m \hat{f}(t_j) \Rightarrow t_e, \quad (3.21)$$

ahol t_j jelzi az ismeretlen koncentrációjú mintára vonatkozó j . mérési jelet.

Felhasználva a jóslt t_e értékét, a becsült koncentrációra vonatkozóan nyerhetünk tapasztalati sűrűségfüggvényt:

$$\hat{f}(z) = \frac{1}{3000h} \sum_{B=1}^{3000} K\left(\frac{z - x^B}{h}\right), \quad (3.22)$$

felhasználva, hogy

$$x^B = \frac{t_e - a_0^B}{a_1^B}, \quad (3.23)$$

ahol a_0^B és a_1^B ugyanaz mint fentebb, és t_e a mért jelek tapasztalati maximum likelihood becslése.

A konfidencia intervallumot a szokásos módon számíthatjuk:

$$\int_{-\infty}^{x_l} \hat{f}(z)dz = \frac{\alpha}{2}, \quad \int_{x_r}^{-\infty} \hat{f}(z)dz = \frac{\alpha}{2}, \quad (3.24)$$

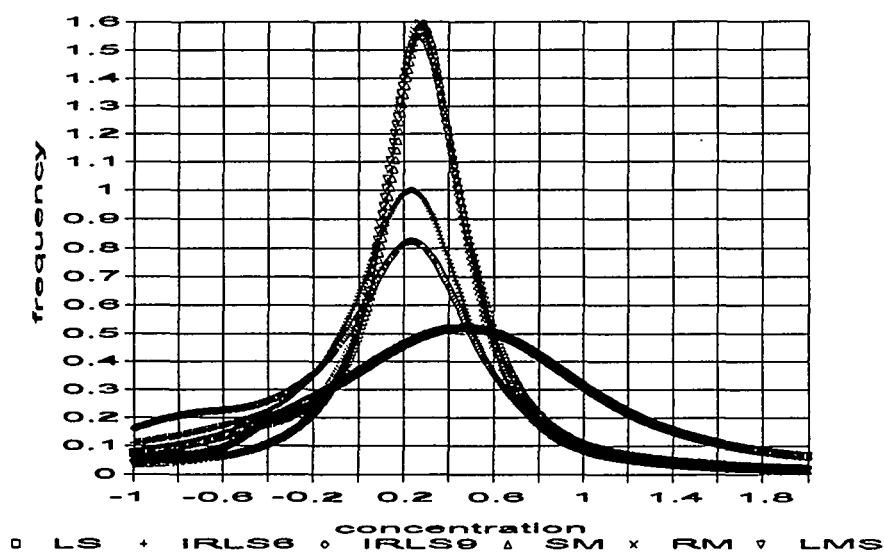
ahol x_l és x_r a konfidencia intervallum bal ill. jobb oldali értéke és α az előre megadott szignifikancia szint. Mivel a jósolt koncentráció becslése robusztus módszerrel történik, így ez az intervallum aszimmetrikus is lehet, jobban megközelítve a valós eseteket.

A fentebb vázolt eljárás bemutatására tekintsük a 3.1 táblázatban lévő 3 db szimulált kalibrációs adatsort, ahol aláhúzással jelöltük az általunk elhelyezett kiugró értékeket. A táblázatban feltüntettünk még három szimulált mérési jelsorozatot (mindegyik 3 párhuzamos mérést tartalmaz), melyek az ismeretlen koncentrációjú mintához tartoznak. A feladat az ismeretlen koncentrációk becslése különböző paraméterbecslő eljárásokkal kiértékelt kalibrációs függvények segítségével. A 3.4-3.9 ábrákon láthatjuk az eredményeket. Általánosságban megállapíthatjuk, hogy a robusztus tulajdonsággal bíró paraméterbecslőkkel kiértékelt kalibrációval torzítatlan és kevésbé bizonytalan (a tapasztalati sűrűségfüggvény a módusz közelében keskenyebb ill. a gyakorisáérték nagyobb) koncentráció jóslást kapunk, mint a legkisebb négyzetek módszerének használata esetében. A robusztus becslők jó tulajdonsága a kiugró pont közelébe eső mérési jel alapján történő koncentráció becslés esetén mutatkozik meg igazán. Jól látható ez a 3.5, 3.7 és 3.8 ábrákon. Érdekes eredmény, hogy ha a nulla pont közelében van a kiugró pont (3.8 és 3.9 ábrák), csak az LMS szolgáltat torzítatlan és kis bizonytalanságú koncentráció becslést, az origótól távolabbi pontokra még az RM és SM robusztus becslők is igen torzított jóslást adnak.

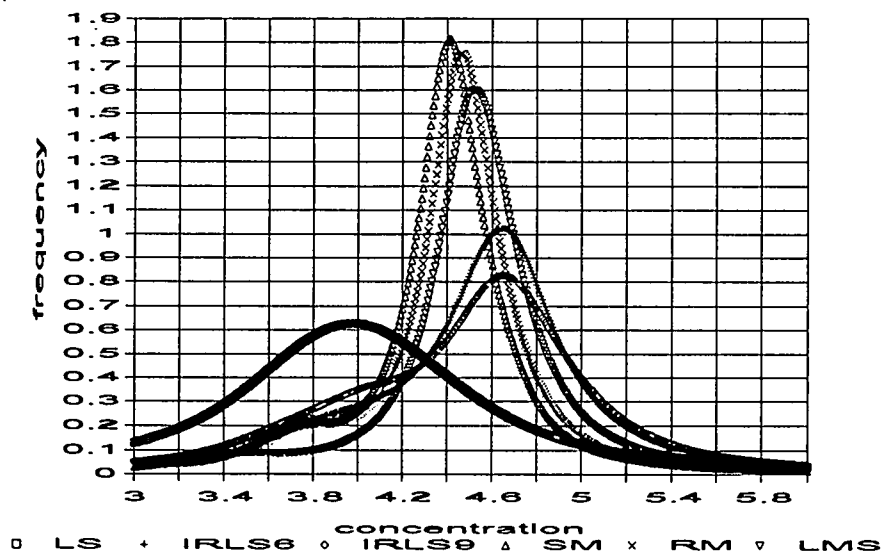
Végül bemutatunk egy gyakorlati példát, ahol talajvizek Mg koncentrációját kellett ICP-AES módszerrel meghatározni. A 3.2 táblázatban tüntettük fel a kalibrációs adatsort, valamint az ismeretlen Mg koncentrációjú oldatra kapott mérési jeleket. A 3.10 ábrán a különböző paraméterbecslőkkel meghatározott kalibrációs egyenes által jósolt koncentráció eloszlásokat vizsgálhatjuk. A 3.3 táblázat a számított konfidencia intervallumokat adja meg 90%-os megbízhatósági szint mellett.

Kalibrációs adatok							
	X	0.0	1.0	2.0	3.0	4.0	5.0
1.	Y	0.0	1.1	2.0	3.1	3.8	<u>6.5</u>
2.	Y	0.0	1.1	2.0	<u>4.5</u>	3.8	5.1
3.	Y	<u>1.5</u>	1.1	2.0	3.1	3.8	5.1
Mérések							
	1. minta	2. minta		3. minta			
	0.24	2.83		4.51			
	0.29	2.88		4.56			
	0.32	2.91		4.59			

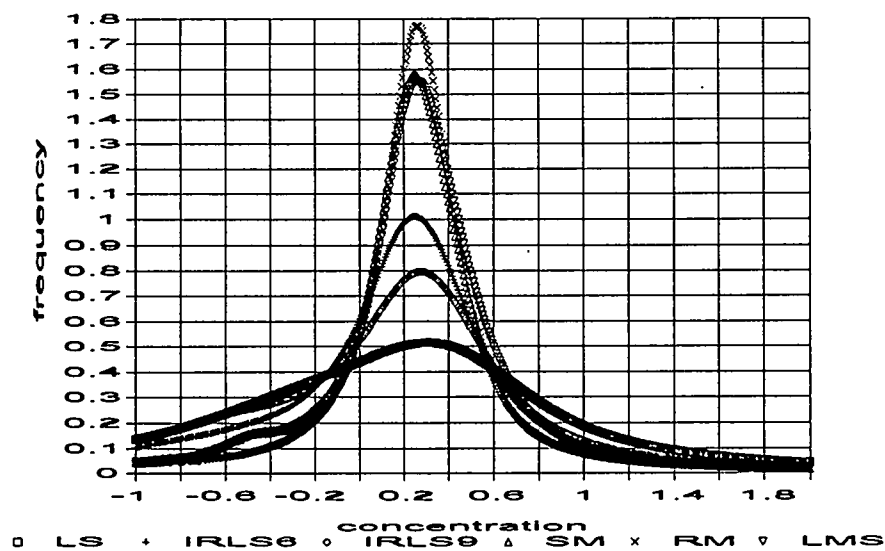
3.1 táblázat: Szimulált kalibrációs adatok, valamint a szimulált mérési jelek



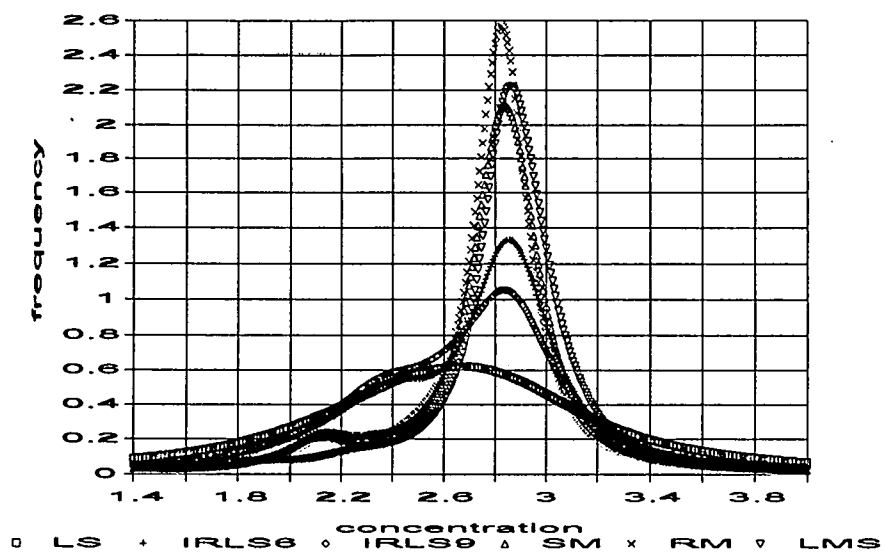
3.4 ábra: Az 1. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 5.0$ -nél $y = 6.5$)



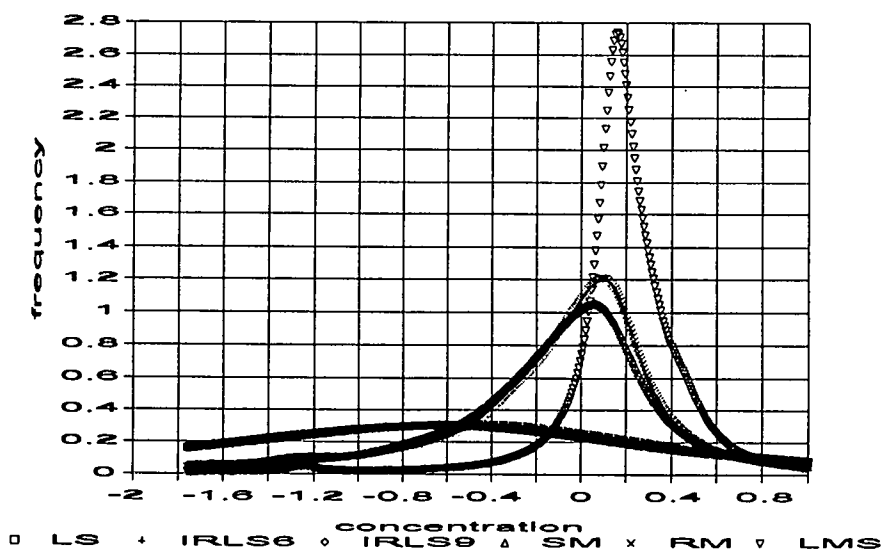
3.5 ábra: A 3. mintára vonatkozó jósolt koncentráció eloszlása (a kiugró érték $x = 5.0$ -nél $y = 6.5$)



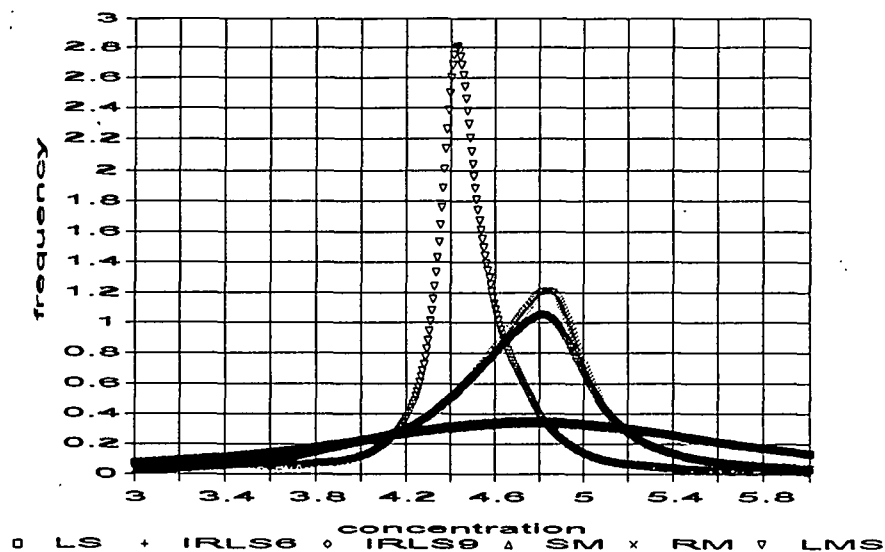
3.6 ábra: Az 1. mintára vonatkozó jósolt koncentráció eloszlása (a kiugró érték $x = 3.0$ -nál $y = 4.5$)



3.7 ábra: A 2. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 3.0$ -nál $y = 4.5$)



3.8 ábra: Az 1. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 0.0$ -nál $y = 1.5$)



3.9 ábra: A 3. mintára vonatkozó jóslt koncentráció eloszlása (a kiugró érték $x = 0.0$ -nál $y = 1.5$)

Kalibrációs adatok			
konc.	1. mérés	2. mérés	3. mérés
ppm	mérési jelek (a.u.)		
0.0	154.6	156.8	153.6
0.5	2418	2413	2401
1.0	4685	4709	4642
1.5	6951	6887	6917
2.0	9727	9738	9788

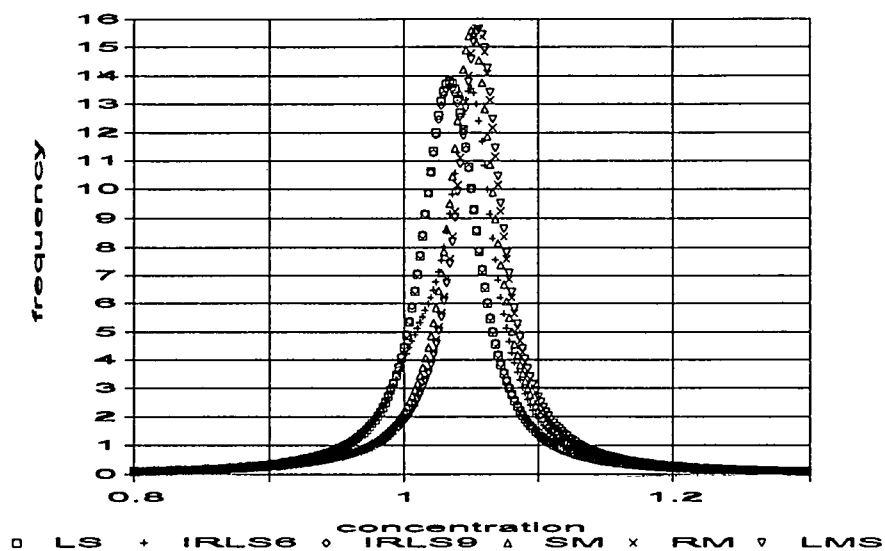
Az ismeretlen koncentrációjú oldatra vonatkozó mérések (a.u.)

4894

4935

5003

3.2 táblázat: Talajvíz Mg koncentrációjának ICP-AES módszerrel történő meghatározásához használt kalibrációs adatsor



3.10 ábra: Talajvíz jósolt Mg koncentrációjának eloszlása különböző paraméterbecslő eljárások alkalmazása mellett

	x_l	x_e	x_r
LS	0.907	1.034	1.162
IRLS6	0.914	1.050	1.170
IRLS9	0.908	1.034	1.162
CM	0.922	1.050	1.176
RM	0.928	1.056	1.182
LSM	0.926	1.056	1.182

3.3 táblázat: Talajvíz Mg koncentrációjának jósolt értékei 90%-os megbízhatósági szintű konfidencia intervallummal, különböző paraméterbecslő eljárások alkalmazása mellett

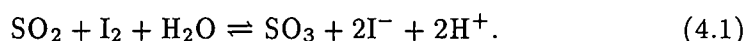
4. fejezet

Karl Fischer titrálás automatikus vezérlése

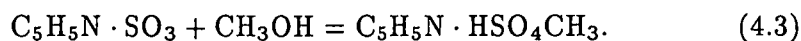
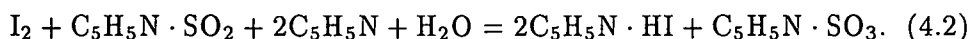
Ebben a fejezetben egy olyan új algoritmust mutatunk be, melyet az automatizált nedvességtartalom meghatározásra, azaz a számítógéppel vezérelt Karl Fischer titrálásra fejlesztettünk ki. Legyen ez példa egy komplex analitikai probléma kemometria szemszélétű megoldására.

4.1 A Karl Fischer titrálásról általában

A Karl Fischer-módszer kémiai vízmeghatározás. Lényegében az alábbi reakción alapul [23]:



A reakció reverzibilis, a keletkező savas komponensek megkötésével az egyensúly a vízmegkötés irányába tolható el. E célra piridint használnak fel, a reakció általában metanolos közegben játszódik le, ahol a metanol nemcsak oldószerként szerepel, de a kén-dioxid oxidációja során keletkező kén-trioxidot is szolvatálja. Így végeredményben a folyamat a következő egyenletekkel írható le:



A gyakorlatban használt oldatokban az egyensúly eltolásának érdekében alkalmazott összetétel miatt a hatóértéket a jódkoncentráció határozza meg. A víztartalom meghatározást megghamisítják a jódtartalmat megváltoztató zavaró kísérő anyagok (szerves peroxidok, oxidáló anionok, oxidáló kationok, redukálószeresek, bórsav, acetón stb.), így azokat a meghatározás előtt ki kell nyerni. Ez metanolos, esetleg etilén-glikolos extrakcióval történhet, pld. olajok, zsírok víztartalmának extrakciója.

A meghatározás során az analizálandó mintát vízmentes metanolban (vagy olyan metanolban, amelynek víztartalmát a minta oldása előtt titráltuk le Karl Fischer-oldattal, és így víztelenítettük) oldjuk, majd a Karl Fischer mérőoldattal titráljuk. A titrálás végpontját általában három módszer szerint lehet jelezni:

- vizuálisan, a sárgából barnába való színváltozás észlelésével,
- potenciometrikus (biamperometriás (dead stop)),
- fotometrikus úton.

Az első esetben a titrálást addig kell végezni, amíg a barna szín megmarad, kb. 20 másodpercig. Túltitrálás esetén más árnyalatú barna keverékszín keletkezik, amely jól felismerhető. A végpont jelzése ezen az úton csak olyan anyagoknál lehetséges, amelyek nem vagy igen kevésbé színezettek.

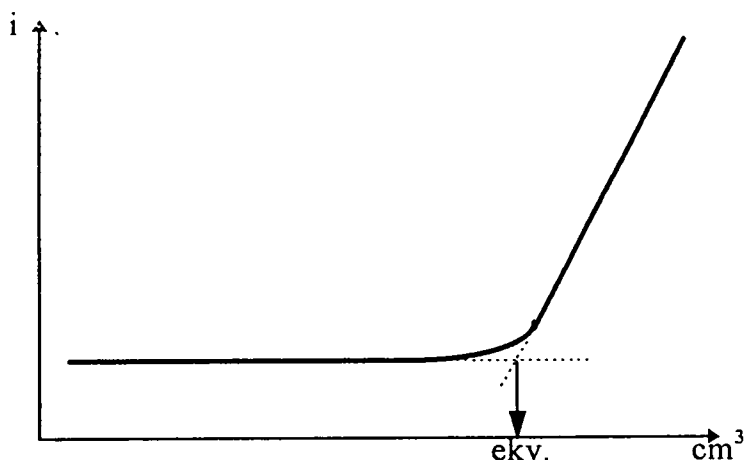
A potenciometrikus titrálás előnye, hogy pontosabb és sötét színű anyagok is titrálhatók. Az összehasonlító elektród alkalmazásának problematikus volta miatt a gyakorlatban a dead stop módszer használatos. Az oldatba merülő két Pt-elektrodot kis feszültséggel (15-20 mV) polarizálják. A cellán minaddig nem megy át áram, amíg a reagens fölöslegbe nem kerül, vagyis a katód depolarizálódik. Az áram megindulását ampermérő jelzi.

A fotometrikus végpontjelzés tiszta és átlátszó, ill. színezett anyagok esetén használható.

A Karl Fischer-mérőoldat hatóértéke állás közben változik, így azt naponta ellenőrizni kell. Ez ismert kristályvíztartalmú sók (pld. bóráx) ill. ismert víztartalmú kalibráló oldatok (pld. HYDRANAL-EICHSTANDARD 5,00) segítségével végezhető el.

4.2 Az új vezérlő algoritmus leírása

Karl Fischer titrálás során kapott titrálási görbét mutatunk be a 4.1 ábrán. A gyakorlatban nem a két egyenes metszésével kapott pontot veszik



4.1 ábra: Karl Fischer titrálás során kapott tipikus titrálási görbe biamperometriás (dead stop) végpontjelzéssel

végpontnak, hanem előzetes mérések alapján előre kijelölnek egy pontot és a továbbiakban eddig az áramerősség értékig titrálnak. Ennél az eljárásnál csak a legelső mérést kell figyelemen kívül hagyni, mivel az esetleges túlادagolás minden további mérésnél kiküszöbölődik. Mivel mindig a titráló edényben lévő metanolos oldat víztelenítésével kell kezdeni a sorozatos elemzést, így hasznos mérés nem megy veszendőbe. Ezzel a módszerrel az értékes titrálószer mennyiségével lehet takarékoskodni.

A megoldandó feladat tehát adott végpontig történő titrálás automatikus vezérlése [3,29,78,105,59]. A fejlesztés kezdeti szakaszában a vezérlő algoritmus lineáris függvény segítségével jósolta meg az adagolandó Karl Fischer-oldat térfogatát. 5 egymást követő mérési pontra egy egyenest illesztettünk a legkisebb négyzetek módszerével. (A továbbiakban minden

összegzés $j=1$ -től 5-ig értendő).

$$\underline{X} = \begin{pmatrix} V_1 & 1 \\ V_2 & 1 \\ V_3 & 1 \\ V_4 & 1 \\ V_5 & 1 \end{pmatrix} \quad \underline{y} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{pmatrix} \quad \underline{p} = \begin{pmatrix} m \\ t \end{pmatrix}, \quad (4.4)$$

ahol V_j a titrálószer térfogata, I_j a j . térfogathoz tartozó áramerősség, m és t pedig az egyenes meredeksége és tengelymetszete. A részletes levezetés mellőzésével a paraméterek legkisebb négyzetes becslése:

$$\underline{p} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{y} \quad (4.5)$$

$$\underline{X}' \underline{X} = \begin{pmatrix} \sum V_j^2 & \sum V_j \\ \sum V_j & 5 \end{pmatrix} \quad \underline{X}' \underline{y} = \begin{pmatrix} \sum V_j I_j \\ \sum I_j \end{pmatrix} \quad (4.6)$$

$$(\underline{X}' \underline{X})^{-1} = \frac{1}{5 \sum V_j^2 - (\sum V_j)^2} \begin{pmatrix} 5 & -\sum V_j \\ -\sum V_j & \sum V_j^2 \end{pmatrix} \quad (4.7)$$

$$\underline{p} = \begin{pmatrix} m \\ t \end{pmatrix} = \begin{pmatrix} \frac{5 \sum V_j I_j - \sum V_j \sum I_j}{5 \sum V_j^2 - (\sum V_j)^2} \\ \frac{\sum V_j^2 \sum I_j - \sum V_j \sum V_j I_j}{5 \sum V_j^2 - (\sum V_j)^2} \end{pmatrix}. \quad (4.8)$$

A t tengelymetszetet felírhatjuk

$$t = \frac{\sum I_j}{5} - m \frac{\sum V_j}{5} = \bar{I} - m \bar{V} \quad (4.9)$$

alakban is.

A regressziós egyenes meredekségét nem befolyásolja a koordináta tengelyek kezdőpontjainak elhelyezkedése, ezért

$$m = \frac{\sum (V_j - \bar{V})(I_j - \bar{I})}{\sum (V_j - \bar{V})^2} \quad (4.10)$$

$$t = \bar{I} - m \bar{V}.$$

(4.10) dinamikus kiszámításához, azaz az adatpontok hozzáadásának ill. elvételének megengedése esetén [53] számítsuk ki a

$$q_n = \sum (V_j - \bar{V})^2$$

$$r_n = \sum (V_j - \bar{V})(I_j - \bar{I}) \quad (4.11)$$

rekurzív összefüggéseket. Egyszerű belátni, hogy tetszőleges V, I hozzáadásával ill. elvételével a következő képletek adódnak:

$$\begin{aligned} q_{n\pm 1} &= q_n \pm \frac{n-1}{n} (V_j - \bar{V}_{n\pm 1})^2 \\ r_{n\pm 1} &= r_n \pm \frac{n-1}{n} (V_j - \bar{V}_{n\pm 1})(I_j - \bar{I}_{n\pm 1}). \end{aligned} \quad (4.12)$$

Az átlagok rekurzív számításához szükséges összefüggések:

$$\bar{V}_{n\pm 1} = \frac{n \bar{V}_n \pm V}{n \pm 1} \quad (4.13)$$

$$\bar{I}_{n\pm 1} = \frac{n \bar{I}_n \pm I}{n \pm 1}. \quad (4.14)$$

Az egyenes paramétereit az előzőek felhasználásával egyszerűen számíthatjuk:

$$m = \frac{r_{n\pm 1}}{q_{n\pm 1}} \quad (4.15)$$

$$t = \bar{I}_{n\pm 1} - m \bar{V}_{n\pm 1}. \quad (4.16)$$

Ezekkel a rekurzívan számítható paraméterekkel kapott egyenes segítségével megjósolható a következő adag Karl Fishcer-oldat, figyelembe véve, hogy mindig azonos áramerősség emelkedést kívánatos elérni. A gyakorlatban történő alkalmazás során kiderült, hogy a vezérlő algoritmus gyakran eredményezett változó mértékű túlادagolást. Ezért át kellett dolgoznunk az eljárást nemlineáris jósló függvény alkalmazásával. A másodfokú polinomot választottuk, mivel aránylag egyszerűen kezelhető, de már kellően összetett ahhoz, hogy a kritikus görbületet kielégítő pontossággal leírja.

A stabilis rekurzív formulák az előzőekkel azonos módon vezethetők le.

$$\underline{X} = \begin{pmatrix} V_1^2 & V_1 & 1 \\ V_2^2 & V_2 & 1 \\ V_3^2 & V_3 & 1 \\ V_4^2 & V_4 & 1 \\ V_5^2 & V_5 & 1 \end{pmatrix} \quad \underline{y} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{pmatrix} \quad \underline{p} = \begin{pmatrix} b_2 \\ b_1 \\ b_0 \end{pmatrix} \quad (4.17)$$

$$\underline{p} = (\underline{X}' \underline{X})^{-1} \underline{X}' \underline{y} \quad (4.18)$$

$$\underline{X}' \underline{X} = \begin{pmatrix} \sum V_j^4 & \sum V_j^3 & \sum V_j^2 \\ \sum V_j^3 & \sum V_j^2 & \sum V_j \\ \sum V_j^2 & \sum V_j & 5 \end{pmatrix} \quad \underline{X}' \underline{y} = \begin{pmatrix} \sum V_j^2 I_j \\ \sum V_j I_j \\ \sum I_j \end{pmatrix} \quad (4.19)$$

$$\begin{aligned}
(X'X)^{-1} &= \\
&= \frac{1}{5 \left(\sum V_j^2 \sum V_j^4 - (\sum V_j^3)^2 \right) - (\sum V_j)^2 \sum V_j^4 + \sum V_j^2 (2 \sum V_j \sum V_j^3 - (\sum V_j^2)^2)} \cdot \\
&\cdot \begin{pmatrix} 5 \sum V_j^2 - (\sum V_j)^2 & \sum V_j \sum V_j^2 - 5 \sum V_j^3 & \sum V_j \sum V_j^3 - (\sum V_j^2)^2 \\ \sum V_j \sum V_j^2 - 5 \sum V_j^3 & 5 \sum V_j^4 - ((\sum V_j)^2)^2 & \sum V_j^2 \sum V_j^3 - \sum V_j \sum V_j^4 \\ \sum V_j \sum V_j^3 - (\sum V_j^2)^2 & \sum V_j^2 \sum V_j^3 - \sum V_j \sum V_j^4 & \sum V_j^2 \sum V_j^4 - (\sum V_j^3)^2 \end{pmatrix} \\
&\quad (4.20)
\end{aligned}$$

A levezetések részletezése nélkül a korábban vázolt gondolatmenetet követve kapjuk:

$$\begin{aligned}
q_1 &= \sum (V_j - \bar{V})^2 \\
q_2 &= \sum (V_j^2 - \bar{V}^2)^2 \\
q_{12} &= \sum (V_j - \bar{V})(V_j^2 - \bar{V}^2) \\
q_{b_1} &= \sum (I_j - \bar{I}) \left((I_j - \bar{I})q_2 - (I_j^2 - \bar{I}^2)q_{12} \right) \\
q_{b_2} &= \sum (I_j - \bar{I}) \left((I_j^2 - \bar{I}^2)q_1 - (I_j - \bar{I})q_{12} \right).
\end{aligned} \tag{4.21}$$

Tetszőleges V, I hozzáadása ill. elvétele során a paraméterek rekurzív becsléséhez szükséges képletek:

$$\begin{aligned}
q_1(n\pm 1) &= q_1(n) \pm \frac{n-1}{n} (V - \bar{V}_{n\pm 1})^2 \\
q_2(n\pm 1) &= q_2(n) \pm \frac{n-1}{n} (V^2 - \bar{V}^2_{n\pm 1})^2 \\
q_{12}(n\pm 1) &= q_{12}(n) \pm \frac{n-1}{n} (V - \bar{V}_{n\pm 1})(V^2 - \bar{V}^2_{n\pm 1}) \\
q_{b_1}(n\pm 1) &= q_{b_1}(n) \pm (I - \bar{I}_{n\pm 1}) \left((V - \bar{V}_{n\pm 1})q_2(n\pm 1) - (V^2 - \bar{V}^2_{n\pm 1})q_{12}(n\pm 1) \right) \\
q_{b_2}(n\pm 1) &= q_{b_2}(n) \pm (I - \bar{I}_{n\pm 1}) \left((V^2 - \bar{V}^2_{n\pm 1})q_1(n\pm 1) - (V - \bar{V}_{n\pm 1})q_{12}(n\pm 1) \right),
\end{aligned} \tag{4.22}$$

ahol

$$\begin{aligned}
\bar{V}_{n\pm 1} &= \frac{n\bar{V}_{n\pm 1} + V}{n\pm 1} \\
\bar{V}^2_{n\pm 1} &= \frac{n\bar{V}^2_{n\pm 1} + V^2}{n\pm 1} \\
\bar{I}_{n\pm 1} &= \frac{n\bar{I}_{n\pm 1} + I}{n\pm 1}.
\end{aligned} \tag{4.23}$$

Ezen összefüggések felhasználásával a paraméterekre adódó rekurzív becslések:

$$b_1 = \frac{q_{b_1}(n\pm 1)}{D} \tag{4.24}$$

$$b_2 = \frac{q_{b_2}(n\pm 1)}{D} \tag{4.25}$$

$$b_0 = \bar{I}_{n\pm 1} - b_1 \bar{V}_{n\pm 1} - b_2 \bar{V}^2_{n\pm 1}, \tag{4.26}$$

ahol $D = q_1(n \pm 1)q_2(n \pm 1) - q_{12}^2(n \pm 1)$.

A fentebb levezetett paraméterekkel kapott jósló függvénnyel a túlada-golás veszélye elhárult, így az automatikusan vezérelt titrálással a nedves-ségmeghatározás az ISO 9004 [91] előírásainak megfelelő minőségbiztosítási követelményeknek megfelelően végrehajtható, ami a kész program függelék-ben közölt kezelési útmutatójából is kitűnik.

5. fejezet

Összefoglalás

Jelen disszertációban közölt tudományos eredményeimet pontokba szedve foglalom össze.

1. A robusztus regressziós eljárásokkal foglalkozó irodalom áttekintését egy újszerű csoportosítás szerint végeztem el, definiálva az első-, második- és harmadfajú modellhibát
2. Egyváltozós kalibrációs függvények paramétereinek becslésére robusztus regressziós eljárásokat vizsgáltam, felhasználva az ún. minőségjellemző faktort (quality coefficient)
3. Az irodalomban megjelent lineáris fuzzy regresszió egyváltozós változatát úgy módosítottam, hogy a kapott paraméterekre vonatkozó szükségtelenül nagymértékű bizonytalanságot csökkenteni lehetett
4. Az irodalomban megjelent lineáris fuzzy regresszió olyan általánosítását végeztem el, mely nem a többszörös integrálok alkalmazásának segítségével történt (melyek analitikus kezelése gyakran kivitelezhetetlen), hanem n -dimenzós geometriai megfontolások alapján
5. Az irodalomban megjelent fuzzy lineáris regressziós eljárást úgy módosítottam, hogy az robusztus becslővé vált, így az a kalibrációs modell linearitásának kismértékű sérülése esetén is biztonságosan használható marad
6. A bootstrap módszer segítségével algoritmust dolgoztam ki a robusztus eljárásokkal kiértékelt kalibrációkkal jósolt koncentrációk konfidencia intervallummal történő jellemzésére

7. Stabil és pontos vezérlő algoritmust fejlesztettem a Karl Fischer-módszerrel történő vízmeghatározás automatikus irányítására, a kezelő program elkészítése során figyelembe vettem az ISO 9004 minőségbiztosítási előírásokat
8. Munkám során az összes felvetődött numerikus problémához saját fejlesztésű, Turbo Pascalban írt programokat használtam:
 - a robusztus regressziós eljárások Monte-Carlo módszerrel történő vizsgálatához
 - regresszióval kapott reziduálisok grafikus szemléltetéssel történő statisztikai vizsgálatához
 - a PREGO nevű LOTUS 1-2-3 táblázatkezelőbe integrált robusztus regressziós eljárásokat és a bootstrap módszert alkalmazó program fejlesztéséhez
 - lineáris programozási feladat megvalósításához ill. az abszolút eltérések legkisebb összege (LSA), az abszolút eltérések legkisebb maximuma (LMA) és a fuzzy lineáris regresszió alkalmazásához szükséges felhasználói felület kialakításához
 - robusztus regressziós eljárások kalibráció során történő felhasználásához
 - a bootstrap eljárás segítségével megvalósított konfidencia intervallumok megadását végző program fejlesztéséhez
 - Karl Fischer-módszerrel megvalósított automatikus vezérlésű nedvességtartalom meghatározásához
 - a többváltozós négyzetek legkisebb mediánja (LMS) és a többváltozós lineáris fuzzy regresszió paramétereinek meghatározása globális optimumkereső (genetikus algoritmus) eljárással (Unix alatt C-ben fejlesztve)

Az eddig elért eredmények továbbfejlesztése további elméleti és gyakorlati vizsgálódásokat igényel. Már most körvonalazódik azonban egy lehetséges haladási irány, mely a kísérlettervezési módszerek, a különböző heurisztikus eljárások (genetikus algoritmus, mesterséges neuronhálózatok), a hipotézisvizsgálatok és a többváltozós regresszió általánosabb alkalmazását (mintafelismerést, osztályozást) jelöli ki. Az elkövetkező időszakban e területeken szeretnék mélyreható tudományos vizsgálatokat folytatni.

Novel Parameter Estimating Procedures Applied for Evaluating Analytical Measurements

Summary

I showed the following scientific results in this dissertation.

1. To sum up the robust regressions in the literature was used a new concept for the systematization defining the model error of first, second and third kind
2. Some robust procedures were investigated to estimate parameters of univariate linear calibration function using the quality coefficient
3. The univariate version of the linear fuzzy regression was modified to reduce the inappropriate uncertainty of the predicted parameters
4. The linear fuzzy regression was generalized not by multiple integration, but by n -dimensional geometric concepts
5. The fuzzy linear regression was modified to become robust, so it can also use for only nearly fulfilled linear calibration model
6. An algorithm was developed using the bootstrap method to characterize the predicted concentrations obtained from calibration evaluated by robust procedures
7. Stable and precise algorithm was developed for automatic controlling the moisture determination by Karl Fischer method, the program was written taking into consideration the specifications of ISO 9004
8. Developing several computer programs in Turbo Pascal for Dos and in C for Unix

Függelék

KARL-FISCHER TITRÁLÁS VEZÉRLÉSE SZEMÉLYI SZÁMÍTÓGÉPPEL című program kezelési útmutatója

A program indítása

A programot a `kf.bat` nevű file futtatásával lehet elindítani. A képernyőn bejelentkezik a főmenü a következő választási lehetőségeket felkínálva:

- FAKTOROZÁS
- MÉRÉS
- EREDMÉNYEK MEGTEKINTÉSE
- JEGYZŐKÖNYV
- PARAMÉTEREK BEÁLLÍTÁSA
- VÉGE

A menü pontokat a "Tab" ill. a "Shift Tab" billentyűk alkalmazásával tudjuk kijelölni, majd az "Enter" billentyű lenyomásával jutunk a kiválasztott almenübe.

A menü pontokat alkalmazásuk logikai sorrendjében tárgyaljuk.

PARAMÉTEREK BEÁLLÍTÁSA

Ebben a menü pontban a titrálás körülményeit befolyásoló legfontosabb paramétereket lehet beállítani.

A képernyőn megjelenik a paraméterek listája, a változtatható paraméterek mellett kék mezőben sárga számmal ill. piros mezőben sárga számmal jelenik meg az aktuális érték.

A BÜRETTA, a DEAD-STOP TITRÁLÓ ADAPTER és a pH-MÉRŐ KÉSZÜLÉK paramétereinek beállítása válik lehetővé. A paraméterek közül megint csak a "Tab" és "Shift Tab" billentyűkkel választhatunk. A további teendőket a megfelelő résznél ismertetjük, de fehér karakterekkel mindig megjelenik egy tömör ismertető a képernyő jobb szélén. Ebből a menü pontból az "Esc" billentyű megnyomásával távozhat.

BÜRETTA

A "Maximális térfogat adag" címszó mellett be lehet állítani a maximálisan megengedett kiadagolható titrálószer mennyiségét μl -ben. Az adat változtatását négyféleképpen valósíthatjuk meg. Csökkenthetjük az értékét $10\mu\text{l}$ -rel a "Kurzor Fel" billentyűvel, csökkenthetjük az értékét $100\mu\text{l}$ -rel a "Page Up" billentyűvel ill. növelhetjük az értékét $10\mu\text{l}$ -rel a "Kurzor Le" billentyűvel és növelhetjük az értékét $100\mu\text{l}$ -rel a "Page Down" billentyűvel.

A "Monitorozásnál várakozási idő" címszó mellett a jelstabilitás elérésének megállapításához a jelek egymás utáni beolvasása között szükséges várakozási időt lehet beállítani ms -okban. Ezt az értéket a keveredés biztosításához szükséges idő is befolyásolhatja. Az adat változtatását ismét négyféleképpen valósíthatjuk meg. Csökkenthetjük az értékét 100 ms -mal a "Kurzor Fel" billentyűvel, csökkenthetjük az értékét 1000 ms -mal a "Page Up" billentyűvel ill. növelhetjük az értékét 100 ms -mal a "Kurzor Le" billentyűvel és növelhetjük az értékét 1000 ms -mal a "Page Down" billentyűvel.

A megjelenő értékek azonnal rögzülnek, így nem kell semmilyen más billentyűvel elfogadtatni azt. A "Tab" vagy a "Shift Tab" billentyűk segítségével újabb paramétert választhatunk ki.

DEAD-STOP TITRÁLÓ ADAPTER

Végpont kapcsolási szint

A "Végpont kapcsolási szint" címszó mellett azon áramrősség értéket kell beállítani skálarész egységben, amelynél a titrálás befejezése szándékunk szerint bekövetkezik. Ehhez állítsa a dead-stop titráló egység mutatós műszer üzemmódváltó gombját "U Pol" polarizációs feszültség méréshez! Ezután a forgatógomb segítségével állítsa be a megfelelő skálarész értéket! Ha sikerült, ezt jelezze az "ENTER" billentyű lenyomásával, ekkor a képernyőn

is megjelenik a kiválasztott érték. Ezután a művelet után a polarizációs feszültséget mindig újra be kell állítani!

Az árammérő végkitérése mindig $50\mu A$ lesz, az áramváltás iránya "INCR.", azaz növekvő állapotban legyen, és a kapcsolás "MAN.", azaz manuális, kézi vezérlés állapotban legyen.

Polarizáló feszültség

A "Polarizáló feszültség" címszó alatt lehet beállítani a polarizációs feszültséget mV -okban. Ehhez állítsa a dead-stop titráló egység mutatós műszer üzemmódváltó gombját "U Pol" polarizációs feszültség méréshez! Ezután a forgatógomb segítségével állítsa be a megfelelő feszültség értéket! Az "ENTER" billentyű lenyomásával a képernyőn is megjelenik a kiválasztott érték. Ha ebben vagy az előző beállításnál a képernyőn megjelenő értékek és a készülék által mutatott értékek jelentősen eltérnek (jobban mint 1–2%), akkor a következő pontban részletezett módon hitelesítse a pH-mérőt.

Ebből a menü pontból csak akkor tud kilépni, ha utoljára a polarizációs feszültséget állította be!

pH-MÉRŐ KÉSZÜLÉK

pH-mérő hitelesítése

A "pH-mérő hitelesítése" címszó alatt kell elvégezni a hitelesítést, ha az előző pontban említett eltérések adódnak ill. a napi méréskezdés, újbóli gép bekapcsolás utáni első tevékenységként. Először mindig ellenőrizzük, hogy a Dead-Stop készülék és a pH-mérő nullpontja azonos legyen! Ehhez állítsa a Dead-Stop titráló egység mutatós műszer üzemmódváltó gombját "U-pol" polarizációs feszültség méréshez, majd a forgatógomb segítségével nulázza ki a műszert. A pH-mérő készülék "BUFFER" feliratú gombjával állítsa be a $0mV$ értéket és ha kész nyomjon "ENTER"-t! A pH-mérő hitelesítésénél állítsa a Dead-Stop titráló egység mutatós műszer üzemmódváltó gombját "U Pol" polarizációs feszültség méréshez! Ezután a forgatógomb segítségével állítson be $250mV$ feszültség értéket! Ha sikerült, nyomjon ENTER-t. Most a forgatógombbal $500mV$ feszültséget állítson be! Ha ezzel is elkészült, ismét nyomjon ENTER-t, így egy kétpontos kalibrációt végzett a pH-mérő hitelesítése végett. Ezekután újból állítsa be az előző pontban említett paramétereket az ott említettek alapján!

FAKTOROZÁS

Naponta ill. bizonyos idő eltelte után (ami lehet néhány óra is) a Karl-Fischer-oldat faktorát újból meg kell határozni. Mindig legyen biztos benne, hogy előzőleg a parmétereket a megfelelő értékekre állította!

A faktorozás a metanol oldat víztelenítésével kezdődik, mely a titráló edényben a tisztítás során esetlegesen bent maradt víznyomok eltávolítását célozza. A metanol-tartalmú bürettaból töltsön a cellába metanolt, majd a cella rázogatóásával öblítse le a cella belső falait! Zárja el a metanolt tartalmazó büretta csapját, majd nyissa ki a Karl-Fischer-oldatot tartalmazó büretta csapját és nyomja meg az "ENTER" billentyűt!

Az "ENTER" billentyű lenyomásával elindítottuk a titrálást, mely teljesen automatikusan megy végbe. A képernyőn folyamatosan nyomon követhetjük az eseményeket.

A "Fogyott mérőoldat térfogata" címszó mellett a már kiadagolt Karl-Fischer-oldat mennyiségét láthatjuk μl -ben.

A "Cellán átfolyó áram" címszó mellett láthatjuk az előzőleg beállított végpontot skr. (azaz skálarész) egységben, valamint az aktuális áramértéket szintén skr. egységben. Láthatunk még két értéket melyek a monitorozás alatt leolvasott értékek átlaga és az átlag szórása. Az újabb adagolás nem következik be addig, amíg a szórás értéke egy adott érték alá nem csökken. Ezzel biztosítjuk az áramerősség állandóságát, azaz a teljes elkeveredést.

A "Végpont átcapások száma" címszó mellett annak maximális értéke és aktuális értéke látható.

Az "Átcapás után eltelt idő" címszó mellett annak minimális értéke és aktuális értéke látható másodpercekben. Minden átcapás után a minimális értéknek megfelelő ideig várakozik a program, majd újra leolvassa az áramerősséget, ha az újra a végpont alá esett vissza a titrálás folytatódik, amíg a maximálisan megengedett végpont átcapások számát el nem értük.

A metanol oldat víztelenítése után a valódi faktorozás következik. A képernyőn megjelenik a faktorozás sorszáma és az, hogy összesen hány faktorozást kell elvégezni, pld. *1. faktorozás a 3-ból*. Ezen a képernyőn kell beadni a bemért bórax tömegét is. Figyelem! Minden titrálás előtt töltsse fel a Karl-Fischer-oldatot tartalmazó bürettát!

A bórax tömegének beírása után, majd az "ENTER" billentyű megnyomásával elkezdődik a titrálás. A képernyőn az előbb jellemzett címszók jelennek meg, melyek jelentése ugyanaz, mint ahogy már elmagyaráztuk. A titrálás befejeződése során három eset fordulhat elő:

1. A titrálás normálisan befejeződik, az áramerősség elérte a végpontot, vagy a végpont átcapások száma elérte maximális értékét.
2. A fogyott mérőoldat mennyisége túl csekély. Ekkor a program nagyobb mennyiségű anyaggal ismételteti meg a mérést.
3. A fogyott mérőoldat mennyisége túl sok. Ekkor a program kisebb mennyiségű anyaggal ismételteti meg a mérést.

A faktorozáshoz előírt összes titrálás sikeres befejeződésével a program kiírja a Karl-Fischer-oldat faktorát és ezen érték szórását. Az eredmény automatikusan egy faktor állományba kerül.

A faktorszámítás a következő képletek alapján történik:

$$f_{\text{bórax}_i} = \frac{m_{\text{bórax}_i} \cdot 0.47238}{V_i} \quad (5.1)$$

$$f_{\text{bórax}} = \sum_{i=1}^n \frac{f_{\text{bórax}_i}}{n}, \quad (5.2)$$

ahol $f_{\text{bórax}}$ a Karl Fischer-oldat bóraxszal meghatározott hatóértéke (1ml oldat hány mg vizet mér), n a faktorozáshoz szükséges ismételt titrálások száma, $m_{\text{bórax}_i}$ az i . titráláshoz bemért bórax tömege mg-ban, $0.47238 \text{ mg}(H_2O)/1 \text{ mg}(\text{bórax})$ a bóraxban lévő víz relatív mennyisége és V_i az i . titrálás során fogyott K-F-oldat térfogata.

A faktorozást az "Esc" billentyűvel lehet abbahagyni valamilyen vész-helyzet esetén, hacsak nem vagyunk adatbeviteli utasításnál. Ekkor írjunk be egy elfogadható adatot és csak ezután nyomjuk meg az "Esc" billentyűt.

MÉRÉS

A mérés során a vizsgált anyag víztartalmát határozzuk meg az előzőleg faktorozott Karl-Fischer-oldattal. Először a mérendő anyag azonosítóját kell megadnunk, melyhez az összes fehér billentyűn lévő karaktert felhasználhatjuk, szerkeszthetjük, törölhetünk karaktereket.

Az azonosító beadása után a mérendő anyag mértékegységét kell kiválasztanunk, annak függvényében, hogy térfogatát (ml), vagy tömegét (mg) mértük. A szóköz ("Space") billentyűvel váltogathatjuk a mértékegységet, amíg a megfelelő meg nem jelenik, ezt az "ENTER" billentyűvel fogadjuk el.

Ezután a program az anyag tömegét/térfogatát (attól függően, hogy mit választottunk előzőleg) kéri be. Már most jelzi, hogy hanyadik mérésnél tartunk a szükségesen elvégzendők közül, pld. *1. mérés a 3-ból*.

Az "ENTER" billentyű lenyomásával elindítottuk a titrálást, mely teljesen automatikusan megy végbe. A képernyőn folyamatosan nyomon követhetjük az eseményeket.

A "Fogyott mérőoldat térfogata" címszó mellett a már kiadagolt Karl-Fischer-oldat mennyiségét láthatjuk μ l-ben.

A "Cellán átfolyó áram" címszó mellett láthatjuk az előzőleg beállított végpontot skr. (azaz skálarész) egységben, valamint az aktuális áramértéket szintén skr. egységben. Láthatunk még két értéket melyek a monitorozás alatt leolvasott értékek átlaga és az átlag szórása. Az újabb adagolás nem következik be addig, amíg a szórás értéke egy adott érték alá nem csökken. Ezzel biztosítjuk az áramerősség állandóságát, azaz a teljes elkeveredést.

A "Végpont átcapások száma" címszó mellett annak maximális értéke és aktuális értéke látható.

Az "Átcapás után eltelt idő" címszó mellett annak minimális értéke és aktuális értéke látható másodpercekben. Minden átcapás után a minimális értéknek megfelelő ideig várakozik a program, majd újra leolvassa az áramerősséget, ha az újra a végpont alá esett vissza a titrálás folytatódik, amíg a maximálisan megengedett végpont átcapások számát el nem értük.

A titrálás befejeződése során három eset fordulhat elő:

1. A titrálás normálisan befejeződik, az áramerősség elérte a végpontot, vagy a végpont átcapások száma elérte maximális értékét.
2. A fogyott mérőoldat mennyisége túl csekély. Ekkor a program nagyobb mennyiségű anyaggal ismételteti meg a mérést.
3. A fogyott mérőoldat mennyisége túl sok. Ekkor a program kisebb mennyiségű anyaggal ismételteti meg a mérést.

A méréshez előírt összes titrálás sikeres befejeződésével a program kiírja a vizsgált anyag víztartalmát és ezen érték szórását.

Az eredmény automatikusan egy eredmény állományba kerül, melyet az "EREDMÉNYEK MEGTEKINTÉSE" menü pontban nézhetünk meg.

EREDMÉNYEK MEGTEKINTÉSE

Ha az adott alkönyvtárban nem létezik eredmény állomány, akkor a program figyelmeztetése után kilépünk a menü pontból.

Ha léteznek eredmény állományok akkor ezeket a képernyő bal oldalán egy oszlopban láthatjuk. A "Kurzor Le" ill. a "Kurzor Fel" billentyűkkel járhatjuk végig az állományokat, elérve a képernyőre nem fértek is. Az állományok havi osztásban tartalmazzák az eredményeket, azaz egy adott hónapban mért értékek egy állományba kerülnek. A "Kurzor Jobb" billentyűvel betekinthetünk a kiválasztott eredmény állományba.

Az adott eredmény állomány tartalmát kapjuk a képernyőre írva. Ha az adott hónapban több mérést is tároltunk, akkor azokat a "Kurzor Le" ill. a "Kurzor Fel" billentyűkkel érhetjük el. Megtudhatjuk, hogy az állomány hány mérést tartalmaz, éppen melyik adatot nézzük. A mérés dátumát, a mért anyag azonosítóját, a mért anyag víztartalmát és a víztartalom szórást. A képernyőn megjelenő információkat bekapcsolt nyomtató esetén a "Print Screen" billentyű megnyomásával nyomtathatjuk ki további feldolgozás végett.

Újabb állományok kijelöléséhez a "Kurzor Bal" billentyű megnyomásával jutunk és ismét az ezen fejezet elején mondottak lépnek érvénybe.

JEGYZŐKÖNYV

Az ISO 9004 által megkövetelt jegyzőkönyv formátumot lehet előállítani a meglévő eredmény állományok tartalmának felhasználásával.

VÉGE

Ennél a menü pontnál a program befejezését lehet kérni, melyet egy megerősítés után érhetünk el.

A disszertációval kapcsolatos közlemények és előadások felsorolása

Közlemények

1. **Rajkó Róbert** (1988–1989): Mérési hibával terhelt koncentráció és jel esetén fuzzy halmazok alkalmazása a kalibráció elvégzéséhez.
Tudományos Diákköri Dolgozat, JATE.
2. **Rajkó Róbert** (1988): Robusztus eljárások alkalmazása a kémiai kalibráció során.
Egyetemi Diplomadolgozat, JATE.
3. **Rajkó Róbert**, Horváth István és Huhn Péter (1989): Paraméterbecslés a fazzi halmazok segítségével.
Magyar Kémiai Folyóirat, **95**, 323–326.
4. Horváth István, **Rajkó Róbert** és Huhn Péter (1989): Robusztus regressziós módszerek alkalmazása a lineáris kalibrációs modellben, I.
Magyar Kémiai Folyóirat, **95**, 327–335.
5. **R. Rajkó** (1994): Treatment of model error in calibration by robust and fuzzy procedures.
Analytical Letters, **27**(1), 215–228.

6. **Rajkó Róbert** (1994): Lineáris és linearizált függvénykapcsolatok kiértékelése.
Élelmiszeripari Főiskola, Tudományos Közlemények, **17**, 44–52.
7. Szabó Gábor, **Rajkó Róbert**, Kovács Erzsébet, Papp Gézáné, Hotya Líviusz (1994): A mikrohullámú termikus kezelés hatása a szójabab minőségére.
Élelmiszeripari Főiskola, Tudományos Közlemények, **17**, 12–22.
8. **R. Rajkó**: Novel chemometrical methods in analytical chemistry. Part I. A theoretical approach.
Journal of Chemometrics, submitted for publication
9. **R. Rajkó**: Parameter estimation of multivariate linear fuzzy regression.
Acta Alimentaria, submitted for publication
10. **R. Rajkó**, C. Lucasius, Jr. and G. Kateman: Parameter estimation of multivariate linear fuzzy regression searching for global optimum with a genetic algorithm.
in preparation
11. **R. Rajkó**, Cs. Judik and L. Sebők: A new algorithm for computer controlled moisture determination by Karl Fischer titration.
in preparation
12. **R. Rajkó**, I. Horváth, A. Csikkel-Szolnoki and S. Veres: Novel chemometrical methods in analytical chemistry. Part II. A practical approach.
in preparation
13. G. Szabó and **R. Rajkó**: Novel experimental methods for reducing enzyme activity by microwave energy.
in preparation
14. **R. Rajkó** and G. Szabó: Advanced experimental design in food science.
in preparation
15. K. Héberger and **R. Rajkó** : Variable selection by pair-correlation method in QSAR.
in preparation

16. **R. Rajkó** and K. Héberger : Selection criteria with a test statistics for pair-correlation method.
in preparation
17. K. Héberger and **R. Rajkó** : Generalization of the pair-correlation method for variable selection and model building.
in preparation

Előadások

1. **Rajkó Róbert**: Véralkohol-koncentráció meghatározása során fuzzy halmazok alkalmazása a kalibráció elvégzéséhez.
MKE Ifjú Kémikusok Köre XI. Tudományos Szimpózium Szeged, 1988. november 8-10.
2. **R. Rajkó**: Calibration with imprecise signals and concentrations based on fuzzy theory.
XVI. Yugoslav Congress of Students of Pure and Applied Chemistry with International Participation, Novi Sad, Yugoslavia, December 14-17, 1988.
3. **Rajkó Róbert**: Mérési hibával terhelt koncentráció és jel esetén fuzzy halmazok alkalmazása a kalibráció elvégzéséhez.
XIX. Országos Tudományos Diákköri Konferencia, Szeged, 1989. április 5-7.
4. **Rajkó Róbert**, Dombi József: A fuzzy elmélet alkalmazása a kémiai mérések értékelése során.
KEMOMETRIA '89, Csopak, 1989. október 16-18.
5. **Rajkó Róbert**: Néhány robusztus módszer összehasonlítása variancia analízissel.
MKE Ifjú Kémikusok Köre XII. Tudományos Szimpózium Szeged, 1989. november 7-9.
6. **R. Rajkó**: Novel statistical methods in analytical chemistry.
in Proceedings of Conference CHEMOMETRICS II, Brno, Czechoslovakia, September 3-6, 1990.
7. **Rajkó Róbert**: Fuzzy halmazok alkalmazása az analitikai célú számítástechnikában.
KEMOMETRIA '90, Csopak, 1990. október 15-17.

8. Horváth István, **Rajkó Róbert**: Analitikai sorozatmérések értékelésére szolgáló új kemometriai módszerek.
A Szerves és Gyógyszeranalitikai Munkabizottság ülése, Szeged, 1991. május 23.
9. Horváth István, **Rajkó Róbert**: Robusztus becslési eljárások értékelése variancia analízissel lineáris kalibráció esetén.
KEMOMETRIA '91, Balatonlelle, 1991. október 14-16.
10. **Rajkó Róbert**: Neuronhálózatok alkalmazása az analitikai kémiában.
KEMOMETRIA '91, Balatonlelle, 1991. október 14-16.
11. **R. Rajkó**: On robust calibration with bootstrapping.
Research discussion, Nijmegen, NL, March 19, 1992.
12. **R. Rajkó**: Multivariate regression with LMS and based on fuzzy theory using a genetic algorithm.
Research discussion, Nijmegen, NL, June 18, 1992.
13. Horváth István, **Rajkó Róbert**, Csúri Péter: Egy eljárás az optimális kalibráció megkeresésére.
Vegyészkonferencia, Szombathely, 1992. július 6-8.
14. **R. Rajkó**: Treatment of model errors in calibration.
COMPANA '92, Jena, Germany, August 24-27, 1992.
15. I. Horváth, **R. Rajkó**, P. Csúri: A suggestion to search for the optimum calibration.
COMPANA '92, JENA, Germany, August 24-27, 1992.
16. **R. Rajkó**, Z. Kupihár and R. Rónyai: Smoothing the signal of HPLC by median based methods.
HPLC '93, Hamburg, Germany, May 9-14, 1993.
17. **Rajkó Róbert**: Új eljárás robusztus becslésekkel kapott kalibrációs eredmények jellemzésére.
KEMOMETRIA '93, Balatonfüred, 1993. október 18-20.
18. **Rajkó Róbert**: Robusztus regressziós eljárások és a Bootstrap módszer alkalmazása az analitikai kémiában.
MKE Ifjú Kémikusok Köre XVI. Tudományos Szimpózium Szeged, 1993. november 1-3.

19. **R. Rajkó:** Characterization of results obtained from robust calibration by bootstrap resampling method.
International Chemometrics Research Meeting, Veldhoven, the Netherlands, July 3-7, 1994, accepted for presentation
20. **R. Rajkó, C. Lucasius, Jr., G. Kateman:** Parameter estimation of multivariate linear fuzzy regression searching for global optimum with a genetic algorithm.
International Chemometrics Research Meeting, Veldhoven, the Netherlands, July 3-7, 1994, accepted for presentation
21. **Rajkó Róbert:** Többváltozós lineáris fuzzy regresszió paramétereinek meghatározása.
X. Élemiszertudományi Konferencia, Budapest, 1994. június 9-10.
22. **Rajkó Róbert:** Robusztus regressziós eljárásokkal kiértékelt mérési eredmények jellemzése Bootstrap módszer segítségével.
Az MTA ÉKB, a MÉTE és a KÉKI 269. Tudományos Kollokviuma, Budapest, 1994. szeptember 30.
23. **Rajkó Róbert:** Többváltozós lineáris fuzzy regresszió és más robusztus többváltozós paraméterbecslő eljárások.
KEMOMETRIA '94, Zamárdi, 1994. október 5-6.
24. **Rajkó Róbert:** Kalibráció a kémiai méréseknél.
Analitikai Napok '95, Budapest, 1995. január 17-18. (felkért előadóként)
25. **R. Rajkó and G. Szabó:** Novel data processing methods in food science.
9th World Congress of Food Science and Technology, Budapest, July 30 - August 4, 1995, accepted for presentation
26. **G. Szabó and R. Rajkó:** Novel experimental methods for reducing enzyme activity by microwave energy.
9th World Congress of Food Science and Technology, Budapest, July 30 - August 4, 1995, accepted for presentation

Köszönetnyilvánítás

Köszönetet mondok Dr. Szabó Gábor tanszékvezető egyetemi docensnek, hogy számomra jelen dolgozat elkészítését lehetővé tette.

Hálásan köszönöm Prof. drs. Gerrit Kateman nyugalmazott egyetemi tanárnak (Catholic University of Nijmegen, the Netherlands) a lehetőséget, hogy még az aktív korában általa vezetett tanszéken 1992-ben fél éves posztgraduális képzésen vehettem részt, melynek keretében a kemometria tudomány időszerű kérdéseivel foglalkozhattam (a képzés költségeit a TEMPUS IMG fedezte). Megköszönöm Dr. Lutgarde Buydens, Dr. Carlos Lucasius, Jr. és Dr. Klaas Faber által kezdeményezett, olykor élesen kritikus hangvételű megbeszéléseket, melyekből nagyon sokat tanultam.

Megköszönöm Csikkel Csabáné dr. és Dr. Veres Sándor egyetemi adjunktusoknak, hogy ICP-AES kalibrációs mérési eredményeiket rendelkezésemre bocsájtották.

Köszönet illeti Dr. Horváth István egyetemi docenst, aki révén megismerkedtem tudományterületemmel, a kemometriával.

Végül megköszönöm feleségemnek és kislányomnak a sok türelmet, megértést és biztatást, melyek nélkül e dolgozat nem jöhetett volna létre.

Irodalomjegyzék

- [1] T. Aishima and S. Nakai (1991). Chemometric in flavor research. *Food Reviews International*, 7(1):33–101.
- [2] Yu.I. Alimov and A.B. Shaevich (1993). Performance of the bootstrap and jack-knife methods in evaluating data. *Zhurnal Analiticheskoi Khimii*, 48(5):847–850. Translated from Russian original.
- [3] J.M. Ariano and W.F. Gutknecht (1976). On-line, computer-controlled potentiometric analysis system. *Analytical Chemistry*, 48(2):281–287.
- [4] H. Bandemer (1985). Evaluating explicit functional relationships from fuzzy observations. *Fuzzy Sets and Systems*, 16:41–52.
- [5] H. Bandemer (1985), editor. *Problems of Evaluations of Functional Relationships from Random-Noise or Fuzzy Data. D 170*, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig.
- [6] H. Bandemer (1988), editor. *Some Applications of Fuzzy Set Theory in Data Analysis. D 187*, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig.
- [7] H. Bandemer (1990), editor. *Some Applications of Fuzzy Set Theory in Data Analysis II. D 197*, VEB Deutscher Verlag für Grundstoffindustrie, Leipzig.
- [8] H. Bandemer and M. Otto (1986). Fuzzy theory in analytical chemistry. *Mikrochimica Acta (Wien)*, II:93–124.
- [9] Y. Bard (1974). *Nonlinear Parameter Estimation*. Academic Press, New York and London.

- [10] A. Bardossy (1990). Note on fuzzy regression. *Fuzzy Sets and Systems*, 37:65–75.
- [11] A. Bardossy, I. Bogardi, and L. Duckstein (1988). *Fuzzy regression in hydrology, Part I: Theory and sample illustration*. Working paper, University of Nebraska-Lincoln.
- [12] A. Bardossy, I. Bogardi, L. Duckstein, and W.E. Kelly (1988). *Fuzzy regression in hydrology, Part II: Actual and potential applications*. Working paper, University of Nebraska-Lincoln.
- [13] A. Bardossy, I. Bogardi, and W.E. Kelly (1989). Geostatistics utilizing imprecise (fuzzy) information. *Fuzzy Sets and Systems*, 31:311–328.
- [14] J.V. Beck and K.J. Arnold (1977). *Parameter Estimation in engineering and science*. Wiley, New York.
- [15] A. Benczúr and S. József (1989). A fuzzy σ -algebrák generáltságáról. *Alkalmazott Matematikai Lapok*, 14:37–43.
- [16] P. Beran (1984). Bootstrap methods in statistics. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 86:14–30.
- [17] J.C. Berridge (1989). Chemometrics in pharmaceutical analysis. *Analytica Chimica Acta*, 223:149–159.
- [18] A. Bezegh and G. Veres (1980). A kalibrációs egyenes statisztikai jellemzése. *Magyar Kémikusok Lapja*, 35(4–5):198–204.
- [19] P.L. Bonate (1993). Approximate confidence intervals in calibration using the bootstrap. *Analytical Chemistry*, 65(10):1367–1372.
- [20] R.G. Brereton (1990). *Chemometrics. Applications of Mathematics and Statistics to laboratory Systems*. Ellis Horwood, New York.
- [21] R.G. Brereton (1987). Chemometrics in analytical chemistry. A review. *Analyst*, 112:1635–1657.
- [22] S.D. Brown and R.S. Bear, Jr. (1993). Chemometric techniques in electrochemistry: A critical review. *Critical Reviews in Analytical Chemistry*, 24(2):99–131.
- [23] K. Burger (1981). *A mennyiségi kémiai analízis alapjai*. Medicina, Budapest.

- [24] M.S. Caceci (1989). Estimating error limits in parametric curve fitting. *Analytical Chemistry*, 61(20):2324–2327.
- [25] A. Celmiņš (1987). Least squares model fitting to fuzzy vector data. *Fuzzy Sets and Systems*, 22:245–269.
- [26] A. Celmiņš (1987). Multidimensional least-squares fitting of fuzzy models. *Mathematical Modelling*, 9:669–689.
- [27] R.L. Chang (1979). Applications of fuzzy sets in curve fitting. *Fuzzy Sets and Systems*, 2:67–74.
- [28] S. Chen (1988). Analysis for multiple fuzzy regression. *Fuzzy Sets and Systems*, 25:59–65.
- [29] T.F. Christiansen, J.E. Busch, and S.C. Krogh (1976). Successive determinations of calcium and magnesium in drinking water by complexometric, potentiometric digital titration to two equivalence points. *Analytical Chemistry*, 48(7):1051–1056.
- [30] V.J. Clancey (1947). Statistical methods in chemical analyses. *Nature*, 159:339–340.
- [31] J.R. Collins (1982). Robust M -estimators of location vectors. *Journal of Multivariate Analysis*, 12:480–492.
- [32] I. Csiszár (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066.
- [33] D. De Angelis, P. Hall, and G.A. Young (1993). A note on coverage error of bootstrap confidence intervals for quantiles. *Math. Proc. Camb. Phil. Soc.*, 114:517–531.
- [34] P. Diamond (1988). Fuzzy least squares. *Information Sciences*, 46:141–157.
- [35] P. Diamond (1987). Least squares fitting of several fuzzy variables. In *Preprints of Second IFSA Congress*, pages 329–331, Tokyo.
- [36] G. Dikta (1993). A note on the rate of convergence of the bootstrap. *Journal of Approximation Theory*, 75:112–114.

- [37] D.L. Donoho and P.J. Huber (1983). The notion of breakdown point. In K.A. Doksum P.J. Bickel and J.L. Hodges Jr. (1983), editors, *A Festschrift for Erich L. Lehmann*, pages 157–184, Wadsworth, Belmont, California.
- [38] D.A. Douglas:1991 (1991). Fourth-dimensional conic sections. *Journal of Undergraduate Mathematics*, 23:21–24.
- [39] D. Dubois and H. Prade (1986). Fuzzy sets and statistical data. *European Journal of Operational Research*, 25:345–356.
- [40] D. Dubois and H. Prade (1980). *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York.
- [41] B. Efron (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, 72(1):45–58.
- [42] B. Efron (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.
- [43] B. Efron and R. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1(1):54–77.
- [44] S.P. Ellis and S. Morgenthaller (1992). Leverage and breakdown in l_1 regression. *Journal of the American Statistical Association*, 87:143–148.
- [45] M. Feinberg (1985). Program for robust calibration using a combinatorial algorithm for regression (CAR). *Trends in Analytical Chemistry*, 4:189–190.
- [46] M. Feinberg and E. Bugner (1989). Chemometrics and food chemistry: Data validation. *Analytica Chimica Acta*, 223:223–235.
- [47] S. Frühwirth-Schnatter (1992). On statistical inference for fuzzy data with applications to descriptive statistics. *Fuzzy Sets and Systems*, 50:143–165.
- [48] B.R. Gaines (1978). Fuzzy and probability uncertainty logics. *Information and Control*, 38:154–169.

- [49] E.J. Gumbel (1982). *Statistics of Extremes*. Volume of , Columbia University Press, edition.
- [50] F.R. Hampel (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42:1887–1896.
- [51] F.R. Hampel (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69:383–393.
- [52] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- [53] P. Henrici (1985). *Numerikus analízis*. Műszaki Könyvkiadó, Budapest.
- [54] R. Henrion, G. Henrion, K. Szukalski, I. Fabian, A. Thiesies, and P. Heininger (1991). Parametric and bootstrap estimations of confidence intervals for representative sample weights. *Fresenius' Journal of Analytical Chemistry*, 340:1–5.
- [55] W. Hoeffding (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19:293–325.
- [56] Y. Hu, J. Smeyers-Verbeke, and D.L. Massart (1990). An algorithm for fuzzy linear calibration. *Chemometrics and Intelligent Laboratory Systems*, 8:143–155.
- [57] Y. Hu, J. Smeyers-Verbeke, and D.L. Massart (1990). Outlier detection in calibration. *Chemometrics and Intelligent Laboratory Systems*, 9:31–44.
- [58] P.J. Huber (1981). *Robust Statistics*. Wiley, New York.
- [59] J. Inczédy (1984). *Folyamatos és automatikus analízis*. Műszaki Könyvkiadó, Budapest.
- [60] H. Inoue (1991). A strong law of large numbers for fuzzy random sets. *Fuzzy Sets and Systems*, 41:285–291.
- [61] A.J. Izenman (1991). Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86:205–224.

- [62] K. Jajuga (1986). Linear fuzzy regression. *Fuzzy Sets and Systems*, 20:343–353.
- [63] S. József (1992). On the effect of linear data transformations in possibilistic fuzzy linear regression. *Fuzzy Sets and Systems*, 45:185–188.
- [64] S. József, M. Kovács, K. M. Dobos, and I. Zsigmond (1986). A termés átlag ingadozása: a törvényszerűségek matematikai elemzése. *SZIGMA*, 19:307–323.
- [65] G. Jumarie (1990). *Relative Information. Theories and Applications*. Springer-Verlag, Berlin Heidelberg.
- [66] J. Jureckova (1971). Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*, 42:1328–1338.
- [67] A. Kandel (1978). Fuzzy statistics and forecast evaluation. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-8(5):396–401.
- [68] A. Kandel and W.J. Byatt (1978). Fuzzy sets, fuzzy algebra, and fuzzy statistics. *Proceedings of the IEEE*, 66(12):1619–1630.
- [69] A. Kaufmann (1975). *Introduction to the Theory of Fuzzy Subsets. Vol. 1: Fundamental Theoretical Elements*. Academic Press, New York.
- [70] M.G. Kendall (1961). *A Course in the Geometry of n Dimensions*. Griffin, London.
- [71] M.G. Kendall (1955). *Rank correlation methods*. Griffin, London.
- [72] M.G. Kendall (1951). Regression, structure and functional relationship, Part I. *Biometrika*, 38:11–25.
- [73] M.G. Kendall (1952). Regression, structure and functional relationship, Part II. *Biometrika*, 39:96–108.
- [74] M.G. Kendall and A. Stuart (1965). *The Advanced Theory of Statistics, Vol. 1*. Griffin, London.
- [75] E.P. Klement, W. Schwyhla, and R. Lowen (1981). Fuzzy probability measures. *Fuzzy Sets and Systems*, 5:21–30.
- [76] R.G. Krutchkoff (1967). Classical and inverse regression methods of calibration. *Technometrics*, 9(3):425–439.

- [77] S. Kullback (1959). *Information Theory and Statistics*. Wiley, New York.
- [78] D.J. Leggett (1978). Microcomputer controlled potentiometric titration system for equilibrium studies. *Analytical Chemistry*, 50(6):718–722.
- [79] V. Liebich and M. Otto (1990). Applications of fuzzy theory to spatially resolved analysis of solids. *Analytica Chimica Acta*, 239:61–72.
- [80] H.P. Lopuhaä (1990). Multivariate τ -estimators for location and scatter. *Manuscript*.
- [81] C.B. Lucasius, Jr. (1993). *Towards Genetic Algorithm Methodology in Chemometrics*. PhD thesis, Katholieke Universiteit Nijmegen, the Netherlands.
- [82] E. Mammen (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*. Springer-Verlag, New York.
- [83] R.A. Maronna (1976). Robust M -estimates of multivariate location and scatter. *The Annals of Statistics*, 4:51–67.
- [84] D.L. Massart, B.G.M. Vandeginste, Y. Michotte, and L. Kaufman (1988). *Chemometrics: a textbook*. Elsevier, Amsterdam.
- [85] MSZ 18969/3–86 (1987). A VÁLLALATI MINŐSÉGSZABÁLYOZÁS FELÜLVIZSGÁLATA. A részletes felülvizsgálat kérdéseinek jegyzéke. 1987. január.
- [86] MSZ 18995–1989 (1990). MINŐSÉGÜGYI FOGALOMMEGHATÁROZÁSOK. 1990. május 1.
- [87] MSZ EN 29000:1991 (1992). Minőségirányítási és minőségbiztosítási szabványok kiválasztásának és alkalmazásának irányelvei. 1992. július 1.
- [88] MSZ EN 29001:1992 (1992). Minőségügyi rendszerek minőségbiztosítási modellje a tervezés, a fejlesztés, a termelés, a felszerelés és a vevőszolgálat területén. 1992. június 1.
- [89] MSZ EN 29002:1992 (1992). Minőségügyi rendszerek minőségbiztosítási modellje a termelés és a felszerelés területén. 1992. június 1.

- [90] MSZ EN 29003:1992 (1992). Minőségügyi rendszerek minőségbiztosítási modellje a végellenőrzés és a vizsgálat területén. 1992. június 1.
- [91] MSZ EN 29004:1992 (1992). A minőségirányításra és a minőségügyi rendszer elemeire vonatkozó irányelvek. 1992. június 1.
- [92] R.A. Nadkarni (1993). ISO 9000. Quality management standards for chemical and process industries. *Analytical Chemistry*, 65(8):387A–395A.
- [93] Cs. Nagy (1981). Bevezetés a fuzzy halmazok elméletébe. *SZIGMA*, 14:53–70.
- [94] S.P. Nicolescu and R. Viertl (1992). A comparison inference for fuzzy data with applications to descriptive statistics. *Fuzzy Sets and Systems*, 48:341–350.
- [95] T. Okuda, H. Tanaka, and K. Asai (1978). A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. *Information and Control*, 38:135–147.
- [96] T. Onisawa, M. Sugeno, Y. Nishiwaki, H. Kawai, and Y. Harima (1986). Fuzzy measure analysis of public attitude towards the use of nuclear energy. *Fuzzy Sets and Systems*, 20:259–289.
- [97] M. Otto (1987). Chemometrics for material analysis. *Mikrochimica Acta (Wien)*, I:445–453.
- [98] M. Otto (1990). Chemometrics in kinetic analysis. Plenary lecture. *Analyst*, 115:685–688.
- [99] M. Otto (1990). Fuzzy expert systems. *Trends in Analytical Chemistry*, 9(2):69–72.
- [100] M. Otto (1990). Fuzzy sets: Applications to analytical chemistry. *Analytical Chemistry*, 62(14):797A–802A.
- [101] M. Otto (1988). Fuzzy theory explained. *Chemometrics and Intelligent Laboratory Systems*, 4:101–120.
- [102] M. Otto and H. Bandemer (1986). Calibration with imprecise signals and concentrations based on fuzzy theory. *Chemometrics and Intelligent Laboratory Systems*, 1:71–78.

- [103] M. Otto and H. Bandemer (1986). A fuzzy method for component identification and mixture evaluation in the ultraviolet spectral range. *Analytica Chimica Acta*, 191:193–204.
- [104] M. Otto, W. Wegscheider, and E.P. Lankmayr (1988). A fuzzy approach to peak tracking in chromatographic separations. *Analytical Chemistry*, 60:517–521.
- [105] T. Pap, M. Molnár, and J. Inczédy (1980). Automatikus titráló rendszer. *Magyar Kémiai Folyóirat*, 86(12):567–571.
- [106] A. Perez (1967). Information-theoretical risk estimates in statistical decisions. *Kybernetika*, 3:1–21.
- [107] G.R. Phillips and E.M. Eyring (1983). Comparison of conventional and robust regression in analysis of chemical data. *Analytical Chemistry*, 55:1134–1138.
- [108] H. Poincaré (1912). *Calcul des Probabilités*. Gauthier-Villars, Paris, 2nd edition.
- [109] P.J. Rousseeuw (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880.
- [110] P.J. Rousseeuw (1983). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz (1983), editors, *Mathematical Statistics and Applications*, pages 283–297, Reidel, Dordrecht.
- [111] P.J. Rousseeuw and A.M. Leroy (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- [112] N. Schenker (1985). Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association*, 80:360–361.
- [113] S.R. Searle (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.
- [114] P.K. Sen (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63:1379–1389.

- [115] Q.-M. Shao and H. Yu (1993). Bootstrapping the sample means for stationary mixing sequences. *Stochastic Processes and their Applications*, 48:175–190.
- [116] M.A. Sharaf, D.L. Illman, and B.R. Kowalski (1986). *Chemometrics*. John Wiley & Sons, New York.
- [117] J.E. Shore and R.W. Johnson (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26(1):26–37. Corrections: *IEEE Transactions on Information Theory*, IT-26(6):942–943, November 1983.
- [118] A.F. Siegel (1982). Robust regression using repeated medians. *Biometrika*, 69(1):242–244.
- [119] R.L. Sileken Jr. and H.O. Hartley (1973). Two linear programming algorithms for unbiased estimation of linear models. *Journal of the American Statistical Association*, 68:639–641.
- [120] D. Singer (1994). Fuzzy halmazok és a lehetőségek logikája. *Magyar Tudomány*, (2):163–173.
- [121] T. Solymosi and J. Dombi (1990). Fitting functions to data with error bounds: fuzzy regression with ERRGO. In *International Conference on Operations Research Vienna*, Austria.
- [122] D.M.Y. Sommerville (1929). *An Introduction to the Geometry of n Dimensions*. Dover Publications, New York.
- [123] F. Steiner (1988). Most frequent value procedures (A short monograph). *Geophysical Transactions*, 34:139–260.
- [124] M. Sugeno (1974). *Theory of Fuzzy Integrals and its Applications*. PhD thesis, Tokyo Institute of Technology.
- [125] C.D. Sutton (1993). Computer-intensive methods for testing about the mean of an asymmetrical distribution. *Journal of the American Statistical Association*, 88:802–810.
- [126] H. Tanaka, S. Uejima, and K. Asai (1980). Fuzzy linear regression model. In *International Congress on Applied Systems Research and Cybernetics*, Acapulco, Mexiko.

- [127] H. Tanaka, S. Uejima, and K. Asai (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-12(6):903–907.
- [128] P. Valkó and S. Vajda (1987). *Műszaki-tudományos feladatok megoldása személyi számítógéppel*. Műszaki Könyvkiadó, Budapest.
- [129] B.G.M. Vandeginste (1990). Chemometrics in the daily laboratory practice. *Fresenius' Journal of Analytical Chemistry*, 337:786–793.
- [130] P. Vankeerberghen, J. Smeyers-Verbeke, A. Thielemans, and D.L. Massart (1992). The slope ranking method, an algorithm for diagnosing linearity of calibration lines. *Analisis*, 20:103–109.
- [131] K. Varga (1980). Reakciókinetikai paraméterek számítógépes meghatározása mérési eredményekből, I. *Magyar Kémikusok Lapja*, 35(6):618–624.
- [132] K. Varga (1981). Reakciókinetikai paraméterek számítógépes meghatározása mérési eredményekből, II. *Magyar Kémikusok Lapja*, 36(1):6–13.
- [133] G.E. Veress (1984). A kémiai mérések matematikai és rendszertechnikai jellemzése. *A Magyar Tudományos Akadémia Veszprémi Akadémiai Bizottsága Monográfiái*, 9(1).
- [134] I. Vincze (1968). *Matematikai statisztika ipari alkalmazásokkal*. Műszaki Könyvkiadó, Budapest.
- [135] L.A. Zadeh (1965). Fuzzy sets. *Information and Control*, 8:338–353.
- [136] L.A. Zadeh (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23:421–427.