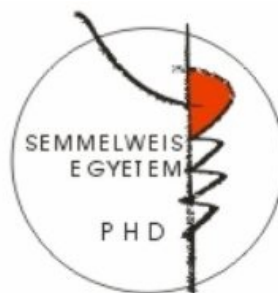


Kinetikus áramlási citometriás mérések értékelése

Doktori értekezés

Dr. Kaposi Ambrus

Semmelweis Egyetem
Klinikai Orvostudományok Doktori Iskola



Témavezető: Dr. Vásárhelyi Barna egyetemi tanár, az MTA doktora

Hivatalos bírálók: Dr. Lacza Zsombor tud. főmunkatárs, Ph.D.
Dr. Nagy György egyetemi adjunktus, Ph.D.

Szigorlati bizottság elnöke: Dr. Smeller László egyetemi tanár, az MTA doktora

Szigorlati bizottság tagjai: Dr. Matkó János egyetemi tanár, az MTA doktora
Dr. Ferenczy György tud. főmunkatárs, Ph.D.

Budapest
2013

Tartalomjegyzék

1. Rövidítések jegyzéke	3
2. Bevezetés	4
2.1. Áramlási citometria	5
2.1.1. Kompenzálás	7
2.1.2. Kinetikus mérések	7
2.2. Adatelemzés	8
2.2.1. Elemző szoftverek	8
2.2.2. Szekvenciális kézi kapuzás	10
2.2.3. Exploratív adatelemzés	11
2.2.4. Hipotézisek vizsgálata	12
2.2.4.1. Jellemző: abszolút vagy relatív sejtszám	12
2.2.4.2. Eloszlás jellemzői	14
2.2.4.3. Jellemzők számítása kinetikus mérésekből	19
3. Célkitűzések	25
4. Módszerek	26
4.1. Dekvantálás	26
4.1.1. Idő paraméter dekvantálása	26
4.1.2. Dekvantálás általános esetben	27
4.2. Helyettesítés	27
4.3. Függvény-illesztés	29
4.3.1. Függvények	29
4.3.2. Illesztési módszerek	34
4.3.3. Illesztés jóságának becslése	36
4.4. Standardizálás	37
4.5. Paraméterek eloszlásának generálása	37
4.5.1. Medián módszer	37
4.5.2. Medián eloszlása	38

4.5.3.	Kvantilis módszer	38
4.5.4.	Hálós illesztés	39
4.6.	Összehasonlítás	41
4.7.	Mérések	41
4.7.1.	1-5-10-25 sorozat	41
4.7.2.	5-20-50 sorozat	44
5.	Eredmények	45
5.1.	Helyettesítés	45
5.2.	Függvény-illesztés	46
5.2.1.	Függvények	46
5.2.2.	Illesztési módszerek	49
5.2.3.	Illesztés jóságának becslése	49
5.3.	Standardizálás	49
5.4.	Paraméterek eloszlásának generálása	50
5.4.1.	Medián módszer	50
5.4.2.	Medián eloszlása	54
5.4.3.	Kvantilis módszer	55
5.4.4.	Hálós illesztés	58
5.5.	A FacsKin által implementált módszer	59
6.	Megbeszélés	63
6.1.	Továbbfejlesztési lehetőségek	68
7.	Következtetések	71
8.	Összefoglalás	73
9.	Summary	74
10.	Irodalomjegyzék	75
11.	Saját publikációk jegyzéke	86
11.1.	Az értekezéshez kapcsolódó közlemények	86
11.2.	Az értekezéshez nem kapcsolódó közlemények	87
12.	Köszönetnyilvánítás	89

1. fejezet

Rövidítések jegyzéke

APC	Allophycocyanin (fluoreszcens festék)
AUC	Area Under Curve (görbe alatti terület)
PB	Probability Binning (valószínűségi „rekeszelés”)
CA	Cluster Analysis (klaszterelemzés)
CD4	Cluster of Differentiation 4 (sejtfelszíni glikoprotein bizonyos T sejteken)
CD8	Cluster of Differentiation 8 (transzmembrán glikoprotein bizonyos T sejteken)
CV	Cross Validation (kereszt-validáció)
FACS	Fluorescent Activated Cell Separator (áramlási citométer szortoló képességekkel)
FCM	Flow Cytometer (áramlási citométer)
FCS	Flow Cytometry Standard (fájlformátum FACS mérési adatok tárolására)
FITC	Fluorescein Isothiocyanate (fluoreszcens festék)
FSC	Forward Scatter (előrefelé irányuló szórás)
KS	Kolmogorov-Smirnov (statisztikai teszt)
NM	Nelder-Mead (illesztő algoritmus)
PBMC	Peripheral Blood Mononuclear Cells (perifériás vér mononukleáris sejtei)
PCA	Principal Component Analysis (főkomponens-analízis)
PBS	Phosphate Buffer Saline (foszfátot tartalmazó pufferoldat)
PE	Phycoerythrin (fluoreszcens festék)
PHA	Phytohemagglutinin
ROS	Reactive Oxygen Species (reaktív oxigén-gyökök)
SAD	Sum of Absolute Deviation (abszolút eltérésösszeg)
SANN	Simulated ANNealing („szimulált hűtés” illesztő algoritmus)
SSC	Side Scatter (oldalra irányuló szórás)
SSD	Sum of Squared Deviation (négyzetes eltérésösszeg)
Th1	T-helper 1-es típusú limfocita
Th2	T-helper 2-es típusú limfocita

2. fejezet

Bevezetés

Az áramlási citometria egy vizsgálóeljárás, mely rövid idő alatt nagyszámú sejt fényelnyelési, -szórási és fluoreszcencia tulajdonságait képes detektálni. Fő erőssége, hogy a sejtekről egyenként ad információt, így a mérés során nem csak az átlagos sejt tulajdonságairól nyerünk adatokat, hanem a vizsgált tulajdonság szórásáról, sejtek között való eloszlásáról is, így a sejtek egyes alcsoportjai is elkülöníthetővé válnak. Sejt-alcsoportok megoszlásának megállapítására évtizedek óta használják az áramlási citometriát az orvosi diagnosztikában ([1], [2]) és sejtélettani kutatásokban [3]. A sejtben időben gyorsan változó mennyiségű anyagok detektálására kifejlesztett fluoreszcens festékek lehetővé tették az intracelluláris folyamatok valós idejű nyomon követését áramlási citométerrel is [4]. Az ilyen típusú, ún. kinetikus mérések értékelése ennek a dolgozatnak a témája.

Ebben a fejezetben összefoglaljuk az áramlási citometria technikájának lényegét (2.1. alfejezet), majd irodalmi áttekintést adunk az áramlási citometriás mérések értékeléséről (2.2. alfejezet), külön kitérve a kinetikus mérésekre jelenleg alkalmazott módszerekre (2.2.4.3. alfejezet). Munkánk célkitűzése egy kinetikus méréseket értékelő eljárás készítése (3. fejezet), mely objektív módon jellemzi a méréseket és statisztikai összehasonlításokra ad lehetőséget. A 4., Módszerek fejezetben az értékelés különböző fázisai szerint csoportosítva ismertetjük azokat az eljárásokat, amelyeket megvizsgáltunk abból a szempontból, hogy mennyire felelnek meg céljainknak. Az 5., Eredmények fejezetben összehasonlítjuk az ismertetett eljárásokat, és megadjuk az általunk legmegfelelőbbnek talált algoritmust. Megbeszélés (6. fejezet) és a továbbfejlesztési lehetőségek ismertetése (6.1. alfejezet) után a tézisek tételes megfogalmazásával és összefoglalással zárjuk a dolgozatot (7., 8. fejezetek).

2.1. Áramlási citometria

Az áramlási citométer (FCM) egyik legfontosabb alkatrésze az ún. áramlási cella, melyen egyesével haladnak keresztül szuszpenzióban levő részecskék (elsősorban sejtek) a hidrodinamikai fókuszálás elve alapján. A készülékben monokromatikus fény világítja meg a keresztülhaladó sejtet, és a sejt által emittált fényt egy optikai rendszer segítségével fotoelektron-sokszorozóval vagy fotodiódával detektálja. Minden sejtről külön rögzítésre kerül az áthaladás időpontja¹, az előre² irányuló fényszórás (FSC, Forward Scatter), a 90°-ban oldalra szórt fény intenzitása (SSC, Side Scatter), és optikai szűrők segítségével az oldalra szórt fény adott hullámhossztartományokban mért intenzitása (fluoreszcens paraméterek). Az FSC paraméter a sejt méretével, az SSC a granuláltságával arányos, és már ezen két paraméter alapján is elkülöníthetők pl. a periferiás vér mononukleáris sejtjei (PBMK) közül a limfociták (1. ábra). Speciálisabb sejt-alcsoportok elkülönítésre használhatók olyan fluoreszcens festékek, melyek egyrészt a gerjesztő fény hatására fluoreszkálnak, másrészt konjugálva vannak valamilyen sejt felszíni vagy intracelluláris molekulához specifikusan kötődő anyaggal (pl. antitesttel), így a fluoreszcens paraméter értéke a jelölt molekula sejtben levő mennyiségével arányos. Ilyen módon pl. anti-CD4 APC festékkel konjugált antitesttel elkülöníthetők a CD4+ limfociták a CD4- limfocitáktól (2. ábra).

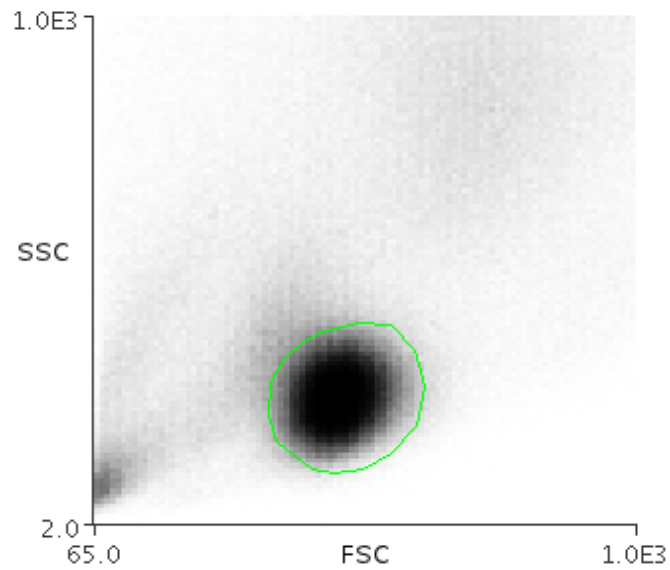
Egy FCM mérés eredményeit egy olyan táblázatban lehet összegezni, melynek egy sora egy sejtnek, egy oszlopa pedig egy paraméternek felel meg (pl. 1. táblázat). A táblázatot, mint sejtek paraméter-vektorainak halmazát formalizáljuk: $data \subset \mathbb{R}^{npar}$, ahol $data$ a mérés eredménye, $npar$ a paraméterek száma. A mérési eredmény általában szabványos FCS formátumban kerül rögzítésre ([5], [6], [7], [8]), mely a táblázat mellett metaadatokat tartalmaz.

1. táblázat. Egy FCM mérés eredményének eleje. Egy sor egy sejtnek, egy oszlop egy paraméternek felel meg. A Time paraméter másodpercben adja meg a sejt mérésének idejét, az FSC és SSC paraméterek felelnek meg az előre- és oldalra szórásnak, a többi paraméterérték (fluoreszcens paraméterek) egy relatív skálán adja meg a paraméternévnek megfelelő fluoreszcens festékhez beállított hullámhossztartományban az adott sejt által emittált fény intenzitását.

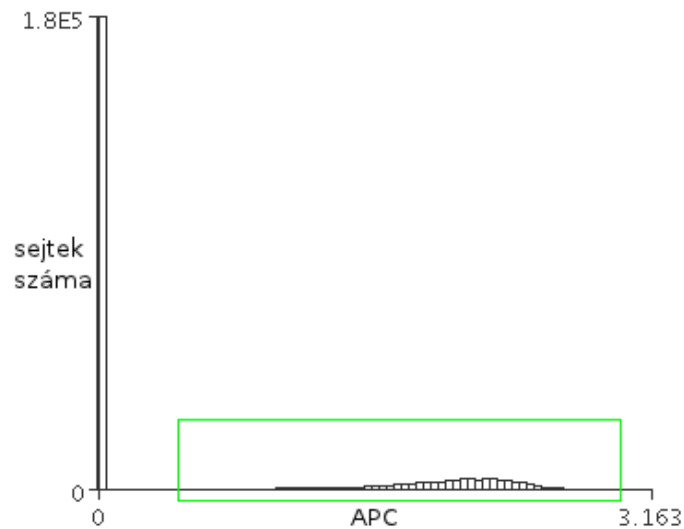
FSC	SSC	FITC	PE	PE-Cy7	APC	APC-Cy7	Time
63353,3	125472,2	8034,5	781,2	282,8	1453,7	627,3	0,001
54528,8	64740,1	6496,4	1727,5	681,4	429,5	717,5	0,002
81142,5	44227,6	5153,4	488,8	2795,1	8,6	522,7	0,003
86443,5	45844,4	19496,7	305,3	977,3	68,4	93,2	0,004
35268,1	122954,8	5138,4	608,0	436,5	890,5	537,8	0,005

¹a mérés kezdetéhez viszonyítva

²a megvilágító fénynyaláb tengelyének irányába



1. ábra. FSC-SSC alcsoport szerinti kapuzás (gate-elés). Az ábrán PBMC sejtek áramlási citometriás mérésének FSC és SSC paraméter szerinti kétdimenziós hisztogramja (density plot) látható. Minél sötétebb egy tartomány, annál nagyobb a sejtek száma. A limfociták (körberajzolt terület) ezen paraméterek alapján elkülöníthetők a többi mononukleáris sejtől.



2. ábra. APC paraméter szerinti alcsoport kapuzás (gate-elés). Az ábrán PBMC sejtek áramlási citometriás mérésének APC paraméter szerinti hisztogramja látható. Az ábrázolás a limfociták kiválasztása után történt. Az APC festék CD4-antitesttel volt kombinálva, a CD4+ sejtek az ábrán körbe vannak rajzolva.

Az FCM másodpercenként több tízezer sejtet képes detektálni [3], egy mérés időtartama perces nagyságrendű, tehát pl. egy 10 perces mérés során akár 10 millió sejtről kaphatunk információt; egy sejtről egy mai FCM akár 20 paramétert is képes rögzíteni [9], a keletkezett adathalmaz tehát több, mint százmillió értéket tartalmazhat.

2.1.1. Kompenzálás

A mérés során fontos, hogy olyan festékeket használjunk különböző anyagok jelölésére, melyeknek emissziós spektruma nincs átfedésben, vagy csak kicsit fed át³. A lehetséges átfedések miatt az egyes paraméterek értékeit ún. kompenzálással módosítjuk, hogy értékeik a tényleges, adott festék miatti fluoreszcenciát jobban közelítsék [10]: a tényleges mérések előtt kontroll mérésekre van szükség, minden egyes fluoreszcens festékkel egy mérést kell végezni, melynek során csak azzal az egy festékkel vannak megfestve a sejtek. Az áramlási citométer minden ilyen mérésre rögzíti, hogy az aktuálisan használt festék mért intenzitásának átlagosan hány százaléka jelenik meg a többi fluoreszcens paraméter értékében (ezeknek a százalékoknak, ha a mért fluoreszcens festék emissziós spektruma 0 értékű mindegyik másik paraméter hullámhossz-tartományában, 0-nak kellene lennie). Ezeket az értékeket az FCM a $SPILL \in [0,1]^{npar \times npar}$ mátrixban rögzíti [8], ahol $SPILL_{ij}$ megadja, hogy az i . paraméter fluoreszcenciájának várhatóan hány százaléka jelenik meg a j . paraméter értékében, ha a sejt csak az i . paraméteren mért festékkel van megfestve. A $SPILL$ mátrix felhasználásával egy tényleges mérés kompenzálása a következőképp történik⁴:

$$data_{compensated} \leftarrow \bigcup_{d \in data} \{SPILL^{-1} * d\}$$

2.1.2. Kinetikus mérések

A legtöbb FCM mérés során a sejtek állapota a mérés teljes időtartama alatt megegyezik, tehát a fluoreszcens paraméterek eloszlása nem függ az idő paraméter értékétől. Kinetikus mérések esetében létezik olyan paraméter, melynek eloszlása függ az időtől (3. ábra). Ilyen mérések azok, melyek során egy, a mérés időtartamával összemérhető hosszúságú folyamat játszódik le a sejtekben, mely valamilyen intracelluláris anyag fluoreszcenciájának változásával nyomon követhető. Ilyen anyag pl. a NADPH, mely intrinsic fluoreszcenciával bír, tehát megfelelő hullámhosszú excitáció hatására detektálható jelet ad

³Lényeges még, hogy az FCM fényforrása képes legyen mindegyik festéket excitálni, ideális esetben az excitációs spektrumuk teljesen átfed.

⁴ d oszlopvektor ($npar \times 1$ méretű mátrix), $*$ a mátrixszorzás, $SPILL^{-1} = COMP$ a $SPILL$ mátrix inverze (compensation matrix, lásd [8]), \leftarrow értékadás

[11], mennyiségének változása az emittált fény intenzitásának regisztrálásával nyomon követhető. Specifikus festékanyagok használatával autofluoreszcenciát nem mutató intracelluláris anyagok szintje is követhető. Ilyen anyagok ill. jellemzők és a kimutatásukra szolgáló festékek a következők [11]:

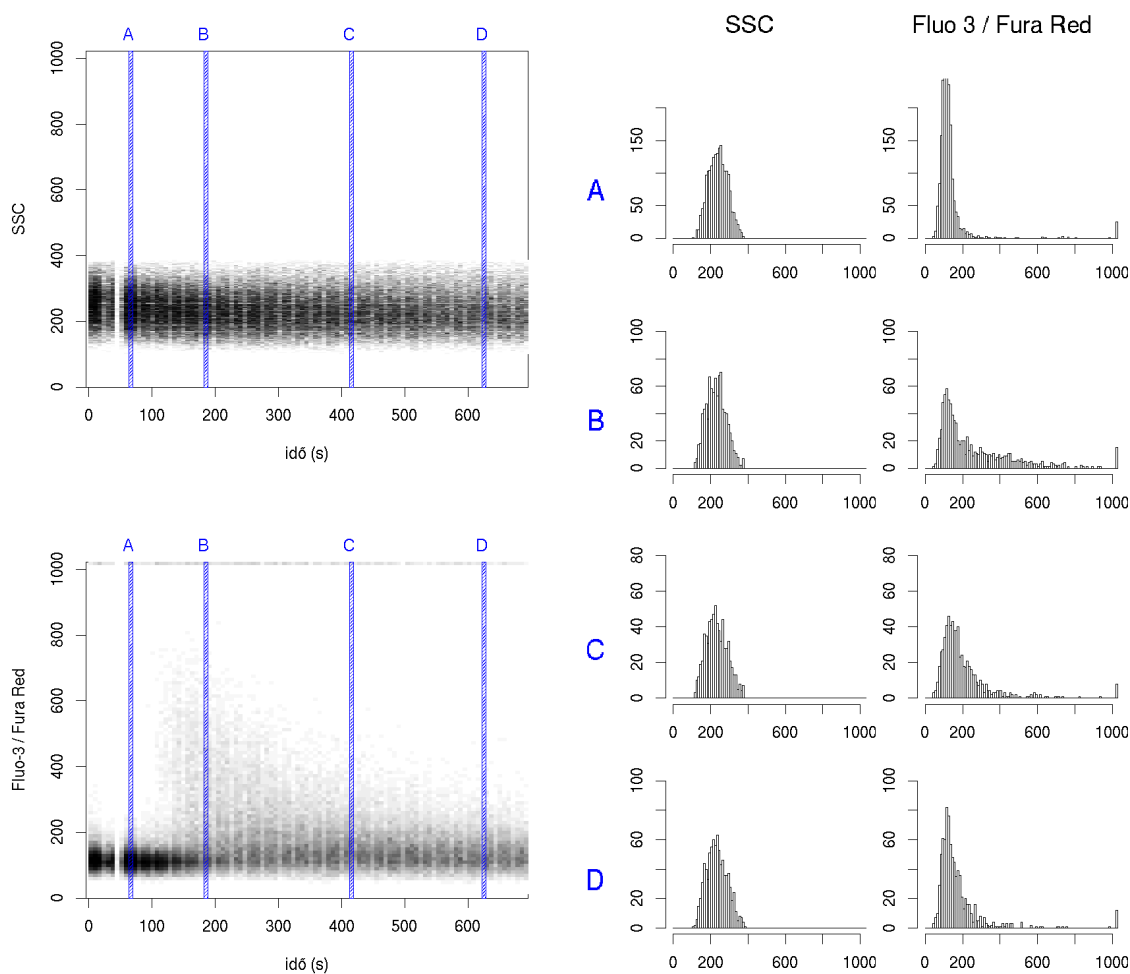
- **citoplazmatikus Ca^{2+}** : Indo-1, Fluo-3, Fluo-3 / Fura Red arány, Fluo-4, Calcium Green, Calcium Orange ([12], [13], [14], [15], [16])
- **mitokondriális Ca^{2+}** : Rhod-2, AM ([17])
- **K^+** : PBFI
- **Na^+** : SBFI, Sodium Green
- **intracelluláris pH**: Karboxi SNARF AM, BCECF AM, 5-CFDA AM
- **plazmamembrán-potenciál**: DiBAC4(3), DiSBAC2(3), DiOC2(3) ([18], [19])
- **mitokondrium-potenciál**: JC-1, TMRM
- **ROS**: DAF-FM diacetát (NO-specifikus), DHE, DCFHDA ([19])

Megfelelő emissziós spektrumú festékek kiválasztásával ezek közül több folyamat akár egyszerre is mérhető ugyanazokban a sejtekben [20]. Az FCM kinetikus mérések a single-cell technikákhoz (fluoreszcens mikroszkópia, patch-clamp) hasonló kinetikus adatokat szolgáltatnak, azzal a lényeges különbséggel, hogy FCM esetében egy adott sejtről csak egyetlen időpillanatban készül mérés, és arról, hogy a sejt a későbbiekben hogyan viselkedik, csak a többi sejt paraméter-értékeiből következtethetünk. Az FCM-módszer előnye azonban, hogy sokkal több adatot szolgáltat (a mintavétel akár kevert sejtpopulációkra is reprezentatív lehet), és a mérés során megmaradnak a sejt-sejt interakciók. Munkacsoportunk a kinetikus mérések közül legtöbbit a limfociták aktivációja során történő kalcium-szint változás detektálásával, a kalcium-flux-szal [21] foglalkozott. Az aktiváció közben az immunsejtek jeltovábbító molekulák (citokinek, kemokinek) termelésével befolyásolják egymás működését, ezért ezen interakciók megőrzése a biológiai vizsgálódás szempontjából előnyös.

2.2. Adatelemzés

2.2.1. Elemző szoftverek

Az FCM mérések értékelése specifikus elemző szoftverek segítségével történik, melyek az FCS formátumú mérési adatfájlokat ([5], [6], [7], [8]) beolvassák, lehetővé teszik a sej-



3. ábra. Bal oldal: kétdimenziós idő - SSC ill. idő - Fluo-3 / Fura Red hisztogramok egy kalcium-flux mérés során. A sötétebb területek több sejtnak felelnek meg. A Fluo-3 / Fura Red arány az intracelluláris Ca^{2+} -koncentrációval arányos [15]. Az FCM mérés előtt a PBMC sejteket $25 \mu\text{g/ml}$ PHA-val stimuláltuk (4.7.2. alfejezet). A mért sejtek közül ki lettek választva a CD4+ limfociták. Jobb oldal: 4 reprezentatív időintervallumban (A, B, C, D) ábrázoltuk az SSC ill. Fluo-3/Fura Red paraméterek eloszlását. Látható, hogy az SSC paraméter eloszlása időben állandó (bal oldali egymás alatti négy hisztogram), míg a Fluo-3 / Fura Red paraméter eloszlásában kb. 110 s-ig egy magasabb értékek felé tolódás játszódik le, ezután pedig egy alacsonyabb értékek felé tolódás (jobb oldali egymás alatti négy hisztogram).

tek kiválasztását, a kiválasztott sejtpopulációk statisztikai jellemzését és különböző mérések összehasonlítását. Léteznek kereskedelmi termékek, ilyenek a CellQuest Pro (BD, Franklin Lakes, NJ, USA), CytoPaint (Leukocyte, Inc., Pleasanton, CA, USA), FCAP Array (Soft Flow Inc., Pécs, Magyarország), FCSExpress (De Novo Software, Los Angeles, CA, USA), FlowJo (Tree Star, Inc., Ashland, OR, USA), Infinicyt (Cytognos S.L., Salamanca, Spain), VenturiOne (Applied Cytometry, Dinnington, Sheffield, UK) és ingy-

nes szoftverek: Cytospec (Purdue University, West Lafayette, IN, USA), WinMDI (FACS Core Facility, The Scripps Research Institute, La Jolla, CA, USA). A flowCore [22] egy nyílt forráskódú Bioconductor [23] csomag, mely R nyelven [24] valósít meg egy standard könyvtárat FCM adatok tárolására, elemzésére. Sok más FCM adatelemző csomag épít rá.

2.2.2. Szekvenciális kézi kapuzás

Az értékelés első lépése a vizsgálni kívánt sejtek kiválasztása (kapuzása, gate-elése), mely hagyományosan kézzel, egydimenziós hisztogramban egy tartomány kijelölésével (lásd 2. ábra) vagy kétdimenziós hisztogramban poligon⁵ vagy téglalap alakú tartományok kijelölésével (lásd 1. ábra) történik [25]. A kétdimenziós hisztogram változatai a scatter-plot (ahol egy pont egy sejtnak felel meg, és a pontok sűrűsége mutatja a sejtek számát) és a contour-plot (szintvonalak jelölik az azonos sűrűségű területeket) vagy ezek kombinációi. A kézi módszer előnye, hogy egyszerű és könnyen értelmezhető. Hátránya, hogy egyszerre csak két (esetleg három-dimenziós hisztogramokkal három) paraméter szerint választhatunk ki alcsoportokat és szubjektív. Használatával önálló limfocita alcsoportokat fedeztek fel, melyeknek az immunfolyamatok szabályozása szempontjából külön szerepük van [26].

A kézi kapuzást finomíthatják félig-automatizált módszerek, pl. egy kétcsúcsú hisztogram egyik felének kapuzásakor a két csúcs közötti legalacsonyabb értéket a szoftver automatikusan megkeresheti. Ez az ötlet kiterjeszhető több-dimenzióra, ilyen pl. a FlowJo (Tree Star, Inc., Ashland, OR, USA) „magnetic gate” funkciója, amely automatikusan egy szintvonalnál (ahol a sejtek sűrűsége azonos) húzza meg a kapu határát.

A mai FCM készülékek egyetlen sejtről akár 20 paramétert is képesek rögzíteni [9], ezeket kézi kapuzással és kettesével-hármasával megjelenítéssel végigelemezni nagyon időigényes feladat. Nem beszélve az ún. klaszter-citometriáról [27], mely technikailag megvalósítja sok ezer FCM mérés automatikus elvégzését, ezzel óriási adathalmazokat generálva.

A kapuzás folyamatának rögzítése elengedhetetlen az értékelés reprodukálhatóságának biztosításához. A kapuk mások számára is elérhetővé tehetők pl. az XML formátumot használó gatingML szabvánnyal [28], melynek használatát az új FCS szabvány [8] is javasolja. Az MIFlowCyt szabvány [29] az egy citometriás mérés megértéséhez és reprodukálhatóságához szükséges minimális információt (MI) határozza meg.

⁵Ebben az esetben ún. ray-casting algoritmus használatos annak eldöntésére, hogy egy adott sejt benne van-e a poligonban (PIP, Point In Polygon probléma).

2.2.3. Exploratív adatelemzés

Ha pontosan tudjuk, hogy milyen alcsoportot vizsgálunk, akkor a kézi módszer jól használható, ha azonban nagyszámú, sokparaméteres FCM adathalmazból szeretnénk új információt nyerni, exploratív adatelemző, adatbányász módszerekre van szükség. Bayes hálók alkalmazásával FCM adatok alapján olyan új jelátviteli folyamatokat fedeztek fel, amelyeket korábban klasszikus biokémiai módszerekkel nem tudtak azonosítani [30]. A később leírt Probability Binning (PB) eljárás is használható exploratív adatelemzésre azért, hogy képes automatikusan kiválasztani sejteket a Frequency Difference Gating módszerével [31]. A PB eljárás n paraméter használata esetén n -dimenziós rekeszekre osztja a mérést, és meghatározza, hogy melyek azok a rekeszek, melyek két mérés között a bennük levő sejtek mennyisége alapján leginkább különböznek. A kapukat ez alapján állapítja meg. Hátránya, hogy a mérési beállításokra nagyon érzékeny. Exploratív elemzésre használatos még a klaszteranalízis (CA), mely sokdimenziós távolságok alapján hoz létre klasztereket a mérési adatokból. Alkalmas trendek megállapítására [32], a mérések automatikus csoportosítására (akár átfedő csoportokkal is, [33], [34]) és kapuk megállapítására [35]. Régóta használatos az FCM mérések értékelésére ([36], [37], [38], [39]), de nem alakult ki standard klaszterező algoritmus [40], és az eljárások biológiai validálása még hátra van [9].

Ha sok paraméter generálja a százalékos megoszlást (az alcsoportokat), ezeknek megjelenítése ill. értelmezése a sok alcsoport miatt nehéz: n paraméter esetén pl. maximálisan 2^n alcsoport lehetséges, ami 18 mért paraméter esetén is már százezres nagyságrendű szám. Ennek a problémának a megoldására fejlesztették ki a SPICE (Simplified Presentation of Incredibly Complex Evaluations, NIAID, NIH, USA) nevű szoftvert. Ez a program a szokásos kézi kapuzási módszer után egyszerű módon, pl. kördiagramok segítségével képes megjeleníteni a kapott alcsoportok megoszlását. Kvantitatív adatokat is szolgáltat, melyek statisztikai összehasonlításokra alkalmasak. Szép példák a szoftver használatára [41] és [42].

A sokdimenziós adathalmaz megjelenítését más módon segíti elő a főkomponens analízis (PCA): a meglévő paraméterek kombinálásával kevesebb számú új, olyan paramétereket hoz létre, melyek nem korrelálnak egymással, így csökkenti az adathalmaz dimenzióját [43]. Szintén használatos FCM adatok exploratív elemzésére [44], [33].

Az exploratív elemző módszerek általában nem adnak kvantitatív eredményeket, hanem a mérési adatokban jelen levő olyan trendekre hívják föl a figyelmet, melyeket azután külön, ezekre a trendekre kidolgozott biológiai eljárásokkal lehet validálni (a trendek könnyen lehet, hogy csak véletlenek, és ezek az eljárások nem adnak szignifikancia értéket).

2.2.4. Hipotézisek vizsgálata

Konkrét hipotézisek FCM-mel való ellenőrzésére a keletkezett nagy adathalmaz miatt Bashashati és mtsai az értékelés szabványosítását javasolják [45]. Eszerint egy mérés értékelése a következő lépésekből áll:

1. **Minőségellenőrzés:** (esetleg már a mérés közben) a mérési hibák kiküszöbölésére.
2. **Standardizálás**⁶: a nem-biológiai eltérések kiküszöbölése, pl. különböző gyártóktól származó antitestek által okozott eltérések, különböző erősítési beállítások a mérőeszközön stb.
3. **Outlierek eltávolítása:** pl. halott sejtek, dupla események (amikor két sejtet egyszerre mér az FCM).
4. **Automatizált kapuzás:** objektív és szisztematikus eljárás szükséges az ismert és még nem ismert sejt-alcsoportok meghatározására.
5. **Alcsoportok elnevezése:** hogy a különböző mérések között csak ugyanazokat a sejt-alcsoportokat hasonlítsuk össze, szükséges ezeket elnevezni, hogy pl. az egyik mérés limfocitáit ne akarjuk összehasonlítani egy másik mérés granulocitáival. Ez a lépés lehet az előzőnek is része.
6. **Jellemzők számítása:** az egyes alcsoportok számszerű jellemzőinek (melyekkel majd összehasonlításokat végzünk) kiszámítása, pl. százalékos megoszlása, sejtek száma, medián, szórás.
7. **Értelmezés:** a vizsgálat tárgyától függően statisztikai összehasonlítások végzése, a mérések csoportokba sorolása stb.

A 2.2.2. pontban szó volt a 4-5. lépésre használható módszerekről. A 6. lépés attól függően, hogy milyen adatokra van szükségünk, többféle lehet. Ezeket foglalja össze a 2. táblázat, az alábbiakban pedig részletesen írunk róluk.

2.2.4.1. Jellemző: abszolút vagy relatív sejtszám

Abszolút sejtszámot állapítanak meg pl. a HIV fertőzés nyomon követésére (CD4+ sejtek μ l-enkénti száma). A relatív sejtszámok meghatározása (pl. hematológiában) az áramlási citometria egyik leggyakoribb alkalmazása. A sejtszámok megállapítására használható a kapuzás, ilyenkor a kapuzás végén megmaradt sejtek számát használjuk fel. Ha ismerjük

⁶Normalizálásnak is nevezik, de mi standardizálásnak hívjuk, hogy a normál eloszlásra hozás transzformációval ne keveredjen.

2. táblázat. FCM mérésekből számítható jellemzők.

Alfejezet	Numerikus jellemző	Hány paraméterből számítjuk
2.2.4.1	abszolút/relatív sejtszám kapuk alapján	egy paraméter
2.2.4.1	abszolút/relatív sejtszám becsléssel (Overton [46], Lampariello [49], Watson [51], PB [47])	egy paraméter
2.2.4.2	középértékek: átlag, geometriai átlag, medián	egy paraméter
2.2.4.2	modell-paraméterek	több-paraméteres is lehet
2.2.4.2	KS-teszttel számított $ x $ ill. \bar{r}_t érték (kontrollhoz képest)	egy paraméter
2.2.4.2	PB-vel számított T érték (kontrollhoz képest)	több-paraméteres is lehet
2.2.4.2	PB-vel számított ujjlenyomat (kontrollhoz képest)	több paraméter
2.2.4.3	kinetikus mérésekből számított görbék paraméterei (pl. maximum, maximum elérési ideje, meredekség)	két paraméter: idő, kinetikus paraméter

az adott paraméter(ek) eloszlásait, akkor pontosabb eredményt kapunk, ha eloszlásfüggvényeket illesztünk a tapasztalati eloszlásra, és pl. egy kétcsúcsú hisztogram kettévágása helyett két átfedésben levő sűrűségfüggvényt illesztünk, s ezek közül az egyik görbe alatti területe lesz a kívánt sejtszám. A leggyakoribb eset az, amikor van egy kontroll mérésünk, melyben a vizsgált paraméternek egy-csúcsú hisztogramja van, ill. egy teszt mérésünk, melyben a kontroll-eloszlást pozitív sejtek tarkítják. A kontroll mérést kivonva a teszt mérésből megkapjuk a csak pozitív sejteket, ebből a százalékos megoszlásuk kiszámítható. A kivonás egy fejlettebb változatát Overton dolgozta ki [46], mely figyelembe veszi azokat az eseteket, amikor a kontroll hisztogram értéke valamelyik osztályban nagyobb, mint a teszt hisztogramé. Ilyenkor a negatív különbséget hagyományosan 0-nak veszik, míg Overton módszere kumulatív módon hozzáadja ezeket a negatív különbségeket más osztályokban levő pozitív különbségekhez. A később leírt PB módszer szintén használható a pozitív sejtek arányának meghatározására, és ebben ugyanolyan jól működik, mint Overton módszere [47]. Lampariello és mtsai kidolgoztak egy matematikai modellt negatív és gyengén pozitív sejtek közös hisztogramjának alakjára ([48], [49]). A modell ún. Johnson S_U négy-paraméteres eloszlásokat használ [50], és robusztus az FCM méréseknél jellemző zajosságra. Az eloszlás paramétereinek meghatározása maximum-likelihood módszerrel, Newton-típusú algoritmussal történik. A módszert Watson fejlesztette tovább [51], modellje konkrét eloszlást nem feltételez, de az eloszlások szimmetriájára és ferdeségére érzékeny. Előnye, hogy konfidencia limitet is ad a százalékos megoszlások megbízható-

ságára. Ezek a paraméteres módszerek sajnos nem használhatók akkor, ha nagymértékű kompenzálásra van szükség (lásd 2.1.1. alfejezet), mert ilyenkor az eloszlás kiszélesedik [10].

Az abszolút sejtszámokat (koncentrációkat) különböző csoportokban össze lehet hasonlítani klasszikus statisztikai tesztekkel (pl. t-próba, ANOVA, nem-normál eloszlás esetén Kruskal-Wallis teszt [52]). Az arányok összehasonlítására specifikus statisztikai tesztek léteznek ([53], [54], [55]).

2.2.4.2. Eloszlás jellemzői

A sejtszámok megállapításához előzetes ismeretekkel kell rendelkezniünk egy (vagy több) paraméter eloszlásáról ill. a sejtek alcsoportjairól. Ha a nyers adatokat, tehát magukat a paraméter-eloszlásokat használjuk, ilyen feltételezésekkel nem élünk, és meghagyjuk annak lehetőségét, hogy olyan kisebb különbségeket (pl. egy eloszlás valamilyen finom irányú eltolódása) is kimutassunk, melyek esetleg egy kézzel nagyjából behúzott kapuval eltűnnének. Másik előny a több-paraméteres eloszlások jellemzésének lehetősége.

Egyparáméteres eloszlások megjelenítésére egyszerű és jól használható eljárás a box and whisker's plot [56], mely megjeleníti a mediánt, egy dobozzal a kvartiliseket (25. és 75. percentilis vagy 0,25 ill. 0,75 kvartilis) és bajszokkal (whiskers) a tartományt (feltéve, hogy a dobozon nem nyúlnak jobban túl, mint az interkvartilis terjedelem másfélszerese; ha igen, akkor az ezen kívül levő adatokat outliereknek veszi, és külön pontokkal jeleníti meg). Példák box plotra a 27. és 29. ábra.

Eloszlások statisztikai összehasonlításához valamilyen konkrét numerikus értéket kell egy-egy eloszlásból számítanunk. Ekkor minden eloszlást (vagyis minden mérést) egy darab számmal jellemzünk. Ez lehet az **átlag** (mean), amely az eloszlás „fizikai” súlypontja (a távolabbi értékeket nagyobb súllyal vesszük figyelembe). Az átlag akkor jó jellemzője az eloszlásnak, ha az normális. Ha log-normális, használható helyette a **geometriai átlag**, mely megegyezik a logaritmizált adatok átlagával. Ha nem teszünk fel semmit előzetesen az eloszlásról, akkor a **mediánt** célszerű használni, mely az eloszlás felezőpontja, nem érzékeny az outlierekre, és megfelelő mérőszáma a „centrális tendenciának” [50]. Illeszthetünk modellt is az eloszlásokra. FCM mérések eloszlásait leírták már Weibull-eloszlással [57] és Johnson-féle S_U eloszlásokkal [48], [49]. Ilyenkor a **modell paramétereivel** jellemezzük az eloszlást. A modell-illesztés módszer használható a mérés értékelésének 2. lépésére (standardizálás) is, hiszen egy szorzó (nagyító) vagy eltolás paraméter figyelmen kívül hagyható, ha feltételezzük, hogy a mérés során ilyen típusú, számunkra nem érdekes melléktermékek adódtak. Ha nincs a modellnek a melléktermékeknek megfelelő paramétere, átparaméterezéssel kaphatunk ilyen.

Felmerülhet, hogy hisztogramokat közvetlenül, statisztikai tesztekkel hasonlítsunk

össze. Pl. két mérés átlaga összehasonlítható t-tesztel, ha normál eloszlásúak. Ha nem normál eloszlásúak, Mann-Whitney U teszt használható, ez utóbbi esetben a mediánokat hasonlítjuk össze. Ekkor megkapjuk az átlagot ill. mediánt, mint az eloszlás jellemzőit, és egy p-értéket is. Egy fejlettebb hisztogram-összehasonlító eljárás a **Kolmogorov-Smirnov** (KS) teszt [58], mely nemcsak a közép (átlag, medián) eltolódására érzékeny, hanem az eloszlás alakjának változásait is figyelembe veszi. Young már 1977-ben javasolta FCM adatok összehasonlítására [59]. A KS teszt két minta (kumulatív) eloszlásfüggvény közti maximális különbség (D_{n_1, n_2} -vel jelölik, ahol n_1 és n_2 a két minta elemszáma) eloszlását adja meg. Ha a két minta ugyanabból a populációból származik (nullhipotézis), akkor $\sqrt{\frac{n_1 * n_2}{n_1 + n_2}} * D_{n_1, n_2}$ -nek ún. Kolmogorov-eloszlása van, melyre léteznek kritikus érték táblázatok [60] ill. közelítő számítások [61]. FCM mérések során jellemzők a műszer beállításából vagy a fluoreszcens festékek mennyiségének különbségeiből következő szisztematikus különbségek, melyekből adódóan (standardizálás nélkül) kétszer pontosan ugyanazt a mintát lemérve is szignifikáns különbségeket adnak a fenti tesztek [62]. Habár az FCM mérések elemszáma rendkívül nagy, a mintavételezés módszere szisztematikus torzításokat rejt magában, emiatt szükséges sok mérés elvégzése, és a mérésekből egy konkrét numerikus érték számítása. Lampariello a KS teszt egy módosított változatát dolgozta ki [63], mely kontroll-mérések alapján megállapítja azt a D_{krit} értéket, mely biológiailag szignifikáns. Ebben az esetben a statisztikailag szignifikáns különbség lehet biológiailag inszignifikáns.

Az eljárás a következő: n kontroll hisztogramunk van, és egy teszt hisztogramunk. Először egy standardizálással kapcsolatos lépés következik, amikor megállapítjuk, hogy mekkora az átlagos eltolás (shift) a kontroll hisztogramok között, és ehhez képest mennyire van eltolva a teszt hisztogram. Ha az eltolás nagy, akkor kimondjuk, hogy biológiailag szignifikáns a különbség. Ha nem nagy, akkor jön egy többszörös, módosított KS teszt.

1. lépés: a kontroll hisztogramok tapasztalati sűrűségfüggvényét $c_i(k)$ -val jelöljük, $i \in \{1, 2, \dots, n\}$, a teszt hisztogramét $t(k)$ -val, $k \in \{1, 2, \dots, nclass\}$ (feltételezzük, hogy mindegyik hisztogram ugyanúgy $nclass$ osztályra van bontva). Az i . hisztogram átlagos osztálya:

$$\bar{c}_i = \frac{\sum_{j=1}^{nclass} j * c_i(j)}{size_i} \quad \text{ahol } size_i = \sum_{j=1}^{nclass} c_i(j) \quad (\text{az } i. \text{ minta elemszáma})$$

Az összes hisztogram átlagos osztálya:

$$\bar{c} = \frac{\sum_{i=1}^n \bar{c}_i}{n}$$

Egy adott kontroll hisztogram eltolása:

$$s_i = \text{round}(\bar{c} - \bar{c}_i)$$

A maximum eltolás, ami a kontroll mintákban szerepelhet:

$$S_M = \max_{i=1}^n |s_i|$$

A teszt hisztogram átlagos osztályát (\bar{t}) ugyanúgy számítjuk, mint a kontroll hisztogramokét. A teszt hisztogram eltolása a kontroll hisztogramokhoz képest:

$$x = \text{round}(\bar{c} - \bar{t})$$

Ha $|x| > S_M$, akkor megállapítjuk, hogy a teszt minta biológiailag szignifikánsan különbözik a kontrolloktól. Ha nem, akkor jön a 2. lépés.

2. lépés: egy többszörös, módosított KS teszt, melynek során a kumulatív hisztogramok különbségének számításánál figyelembe vesszük az 1. lépés során meghatározott eltolásokat, tehát a megfelelően eltolat kumulatív eloszlásfüggvényeket vonjuk ki egymásból, és ezt megteesszük minden kontroll mintára. A $c_i(k)$ és $t(k)$ értékekből számított kumulatív eloszlásfüggvényeket $C_i(k)$ -val ill. $T(k)$ -val jelöljük (pl. $T(k) = \sum_{j=1}^k t(j)$).

Minden kontroll hisztogramra ($i \in \{1, 2, \dots, n\}$) kiszámítjuk az alábbi értéket:

$$r_{it} = \frac{1}{D_{crit,it}} * \max_{j=1}^{n_{class}} |C_i(j - s_i) - T(j - x)|$$

$D_{crit,it}$ a KS teszt kritikus értéke, $C_i(j - s_i)$ az i . kumulatív eloszlásfüggvény s_i -vel eltolva (standardizálva), $T(j - x)$ hasonlóképpen. \bar{r}_t az r_{it} -k átlaga, ez lesz az új statisztika, amit egy L kritikus értékkel hasonlítunk össze:

$$L = \mu + 3 * \sigma$$

ahol \bar{r} az átlaga, σ_r pedig a szórása az alábbi r_{il} arányoknak:

$$r_{il} = \frac{1}{D_{crit,il}} * \max_{j=1}^{n_{class}} |C_i(j - s_i) - C_l(j - s_l)| \quad i \in \{1, 2, \dots, n - 1\}, l \in \{i + 1, \dots, n\}$$

Tehát elvégezzük az eltolással módosított KS tesztet az összes kontroll-párra, majd ezekből számítunk kritikus értéket \bar{r}_t -re⁷. Ha $\bar{r}_t > L$, akkor a teszt mérés biológiailag

⁷Az L kritikus érték kiszámításakor a $3 * \sigma$ szórás-értéket a Csebisev-egyenlőtlenség azon következménye alapján állapította meg Lampariello, hogy tetszőleges eloszlás 88,88889%-a ebben a tartományban van, normál eloszlásnál ez 99,7%.

szignifikánsan különbözik a kontrolloktól, egyébként csak statisztikailag lehet szignifikáns.

Brescia és mtsai ROS mérésekre fejlesztették tovább Lampariello eljárást, azt még robusztusabbá téve (átlag helyett pl. mediánt használnak) [62].

A KS-teszt ezen módosított változata esetében az eloszlásból számított jellemző az $|x|$ ill. az \bar{r}_t érték, melyeknek kiszámításához kontroll eloszlásokat veszünk igénybe. A kontroll eloszlások használata volt az alapja a **Probability Binning** (PB) eljárásnak is, melyet Roederer és mtsai dolgoztak ki ([47], [64]) Cox χ^2 statisztikára épülő eljárásának [65] továbbfejlesztéseként. A PB eljárás által számított jellemző az ún. T érték, mely egy adott kontroll eloszlástól való eltéréssel arányos. Előnye, hogy a Lampariello (és a SED⁸ [66], Overton [46]) módszerrel ellentétben pozitív és negatív irányú eltéréseket is jelez, és nagysága arányos az eloszlások közötti különbséggel, tehát valódi metrikaként alkalmazható: a különböző mérések aszerint csoportosíthatók vele, hogy mennyire térnek el a kontroll méréstől⁹. A biológiailag szignifikáns T érték empirikusan meghatározható. A PB eljárás fő előnye azonban, és ez volt kidolgozásának fő motivációja, hogy nem csak egy, hanem tetszőleges számú paraméter együttes összehasonlítására alkalmas.

A PB eljárás a következő: a kontroll mintát rekeszekre osztjuk, de a klasszikus hisztogram-számítási módszerrel ellentétben nem egyenlő nagyságú rekeszekre, hanem egyenlő számú sejtet tartalmazó rekeszekre. Így az eloszlás azon részén, ahol több érték van, kisebbek lesznek a rekeszek, azon a részén, ahol kevesebb van, nagyobbak. Ez a módszer általánosítható több dimenzióra: kezdetben egy rekeszünk van. Az eljárás minden lépésben az összes rekeszt kettéosztja (elfelezi) aszerint a paraméter szerint, amelynek legnagyobb a varianciája. Így k lépés után 2^k db rekesz keletkezik, és azon paraméterek mentén lesz több rekesz, melyeknek értékei jobban különböznek.

A teszt mintára megszámloljuk, hogy (a kontroll minta rekeszeit használva) melyik rekeszben hány sejt van. Jelöljük c_i -vel az i . rekeszben levő kontroll sejtek relatív gyakoriságát¹⁰, s_i -vel az i . rekeszben levő teszt sejtek számát, ekkor, ha B rekeszünk van, az alábbi χ^2 értéket számítjuk:

$$\chi^2 = \sum_{i=1}^B \frac{(c_i - s_i)^2}{c_i + s_i}$$

Ennek a statisztikának, feltéve, hogy nincs eltérés a kontroll és a teszt eloszlások között, $\overline{\chi^2} = \frac{B}{E}$ átlaggal és $\sigma_{\chi^2} = \frac{\sqrt{B}}{E}$ szórással normál eloszlása van. Ez a tulajdonság

⁸lásd lejjebb az „Egyéb módszerek” részben

⁹Használható sejtcsoportok százalékos arányának meghatározására [47], sőt, kapuzásra is [31].

¹⁰Rekeszben levő kontroll sejtek száma osztva az összes kontroll sejt számával; ez jó közelítéssel meg-
egyezik az összes kontroll sejt száma osztva a rekeszek számával, kis eltérés akkor lehet, ha előbbi nem többszöröse utóbbinak.

használható a T -metrika számítására, mely tetszőleges χ_m^2 mérésből megállapított értékre kiszámítható:

$$T = \max\left(0, \frac{\chi_m^2 - \bar{\chi}^2}{\chi^2}\right)$$

Ha a teszt eloszlás nem tér el a kontroll eloszlásoktól, ennek értéke 0, minél inkább eltér, annál nagyobb az értéke. A normál eloszlás eloszlásfüggvénye alapján kiszámíthatók a T -értékhez tartozó p -értékek (T annak felel meg, hogy hány szórásnyival tér el az átlagtól az adott teszt eloszlás), vagyis az eltérésekre szignifikancia szint is számolható. Ezeknek azonban hasonlóan korlátozott a jelentősége, mint a KS-tesztből számolt p -értékeknek, melyek lehetnek biológiailag inszignifikánsak. Maguk az eloszlásból származtatott T -értékek viszont különböző mérések között összehasonlíthatók klasszikus statisztikai módszerekkel, csakúgy, mint az eloszlások átlagai, mediánjai stb. Az értelmezésben annyi a különbség, hogy ezek nem az eloszlás közepét jellemzik, hanem az eloszlásnak egy kontrolltól való eltérésének mértékét, tehát függnek a kontroll eloszlástól. A kontroll eloszlást lehet több mérésből összeönteni (pl. ha nincs kontroll, az összes mérés összeöntve lehet az), vagy ha ismert egy valamilyen biológiai és mérés technikai indokok alapján megfelelő kontroll, azt lehet használni.

Ahelyett, hogy egy numerikus metrikát használnánk egy mérés jellemzésére, Rogers és mtsai magát a rekeszek szerinti eloszlást használja az eloszlás jellemzésére, melyet citometriás ujjlenyomatnak (CF, Cytometric Fingerprint) nevez [67]. A kontroll minta rekeszeinek meghatározásának algoritmusát is továbbfejlesztette, miszerint nem csak sokdimenziós téglatesteket, hanem politópokat határoz meg. Az eljárás szabadon hozzáférhető implementációja a `flowFP` csomag [68].

Egyéb módszerek: FCM egy-dimenziós hisztogramok összehasonlítására több, nem széles körben elterjedt javaslat is született. Ilyenek Bagwell Bayes-tételre alapozott eljárása [69] ill. D-max (SED) algoritmus [66]. Zeng és mtsai hisztogramok összehasonlításának egy információelméletre épülő módszerét dolgozták ki [70]. A hisztogramokra különböző erősségű simítási módszereket alkalmaztak, majd a kapott, simított görbék első deriváltjait karakterláncokká kódolták, és a karakterláncok információelméleti távolságát határozták meg. Parikh és mtsai konfidencia intervallumok helyett predikciós intervallumok használatát javasolják [71], melyek jobban megközelíthetik az FCM méréseknél használatos gondolkodásmódot, ahol érdekesebb az, hogy egy paraméternek milyen értéke lesz a jövőre vonatkozóan adott valószínűséggel, mint hogy egy hipotetikus végtelenül nagy mérésorozat esetén mennyi lenne az értéke. Algoritmusuk paraméteres, alkalmazásához további vizsgálatok szükségesek.

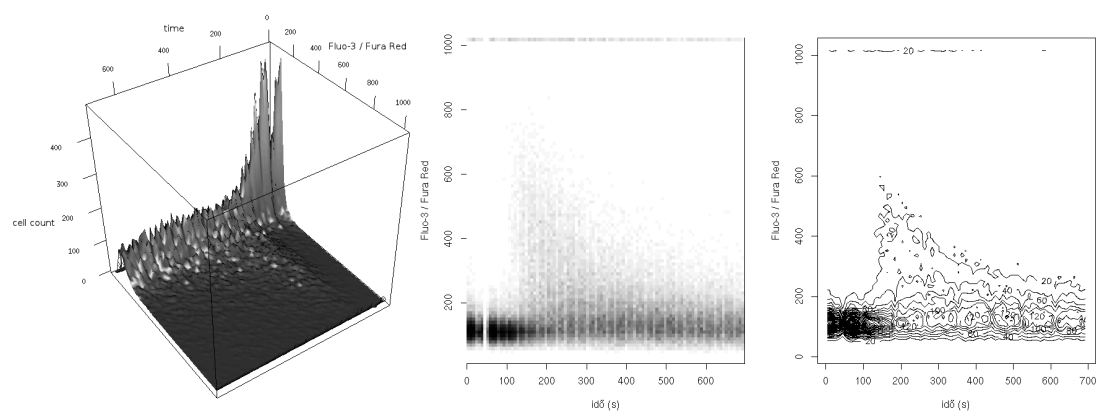
Több-dimenziós, nem exploratív összehasonlításokra példa Davey és mtsainak közle-

ménye [72], akik 3 módszert, a PCA felügyelt változatát, a PCR-t (Principal Component Regression), a PLSR-t (Partial Least Square Regression) és a mesterséges neuronhálókat (ANN) hasonlították össze mikrobiológiai minták felismerésére. A méréseket tanító halmazra (training set, ez alapján állították be a módszerek paramétereit) és teszt halmazra (test set, a módszereket ezen próbálták ki) osztották, legjobbnak az ANN-t találták.

2.2.4.3. Jellemzők számítása kinetikus mérésekből

Kinetikus mérések azok, melyekben az idő paraméter és valamelyik másik paraméter egymástól nem független (3. ábra). Az ilyen mérések értékelésére, hasonlóan ahhoz, ahogy az idősor-analízisre vagy túlélési görbék vizsgálatára speciális statisztikai eljárásokat fejlesztettek ki [73], az előbbiektől eltérő értékelési módszereket alkalmaznak az irodalomban.¹¹

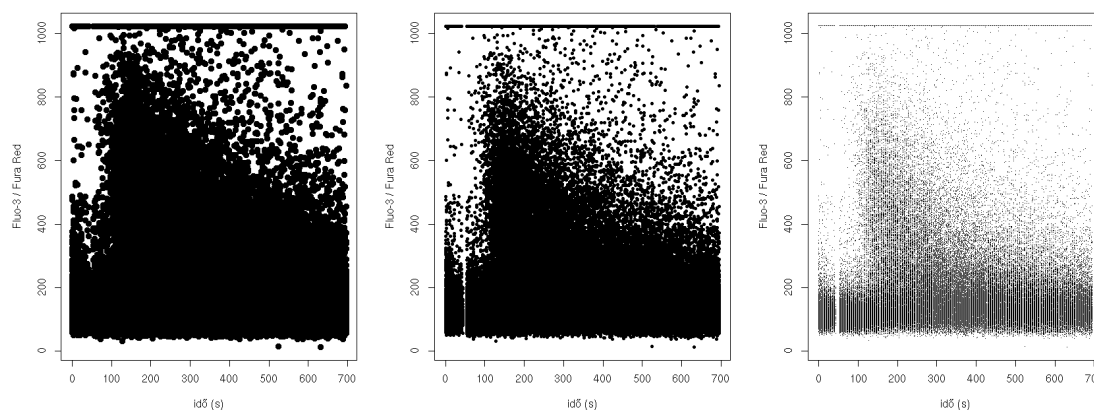
Ábrázolás: a kinetikus méréseket általában kétdimenziós hisztogrammal ábrázolják, az x tengelyen az időt, az y tengelyen a kinetikus paramétert feltüntetve. A z tengely¹² ábrázolhat sejtsűrűséget, ez megjeleníthető 3D-s módon (pl. [4]), színárnyalatokkal vagy szintvonalakkal (lásd 4. ábra). A z tengely ábrázolhatja a sejteket magukat pontonként, ilyenkor, ha elég kicsik a pontok, a megjelenés hasonló egy sűrűség-ábrához (scatter-plot vagy dot-plot, pl. [75], [76], [77], [78]). A kapott kép információtartalma erősen függ attól, hogy mekkora pontméretet használunk (5. ábra) és hogyan választjuk meg az y tengely széleit.



4. ábra. Kinetikus mérések ábrázolása sejtsűrűséggel különböző módokon. Balról jobbra: 3D-s ábrázolás, színárnyalatok (density plot), szintvonalak (contour plot). Az itt ábrázolt kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél 25 $\mu\text{g/ml}$ PHA-t használtunk.

¹¹Az irodalomban használt módszereket táblázatban foglaltuk össze a FacsKin programról szóló könyvfejezetben [74].

¹²vagyis maga a felület, ha 2-dimenziós ábráról van szó



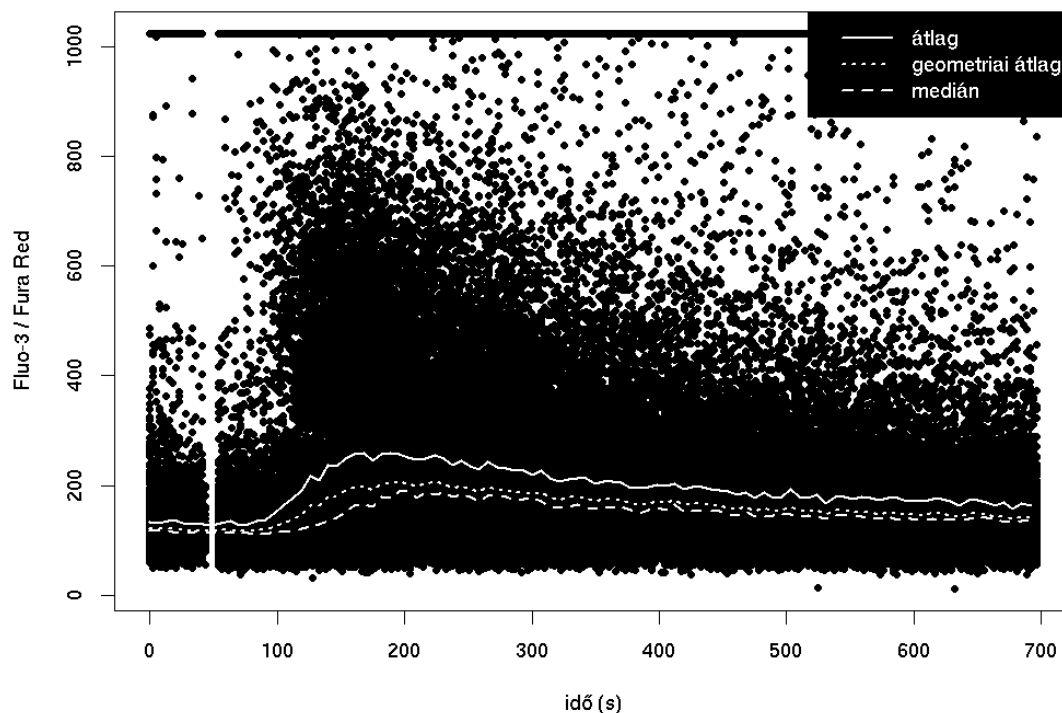
5. ábra. Kinetikus mérések ábrázolása pontokkal (scatter plot) különböző pont-méretekkel. A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél $25 \mu\text{g/ml}$ PHA-t használtunk.

Görbék számítása: az előbb említett ábrák alapján a kinetikus mérések csak erősen szubjektív módon jellemezhetőek, emiatt szokás az adathalmazokból görbéket számítani. Ez megtehető úgy, hogy a mérés időtartamát adott mennyiségű, pl. 100 egyenlő hosszúságú intervallumra osztjuk, majd mindegyik intervallumban kiszámítjuk az ott levő sejtek kinetikus paraméter-értékének átlagát (pl. [4], [79], [80], [76], [77]), geometriai átlagát (pl. [16]) vagy mediánját (pl. [81]). Az így kapott görbék magassága függ az alkalmazott középérték-számító eljárástól, lásd 6. ábra. Az átlagot a kiugró (outlier) értékek erősen elhúzzák, sőt, gyakori, hogy az FCM az eloszlást maximalizálja egy adott értéknél (a bemutatott ábrákon a maximális érték 1023), és az összes ennél nagyobb intenzitású sejt értéke ennyi lesz. Ez az átlag értékét (ebben az esetben lefele) befolyásolja, míg a mediánra nincs hatással. A geometriai átlag akkor használható, ha az eloszlás log-normális, ekkor logaritmikus transzformáció után megegyezik az átlaggal.

A görbe értékei kiszámíthatók adott konkrét időközönként (pl. 10 s-onként, 8 s-os szünetekkel [79], 10 s-onként [80], 25 s-onként [77], 30 s-onként [78], 60 s-onként [82]) vagy egy intervallum folyamatos csúsztatásával, ún. mozgóátlag segítségével (pl. [76], [16]). A 7. ábra bemutatja, hogyan függ a görbe zajossága az időintervallum hosszától. A kapott görbe simítható különböző módszerekkel, melyek közül egy az ún. lowess-eljárás [83], használatára egy példa [81]. Különböző erősségű simításokat mutat be ugyanarról a mérésről a 8. ábra.

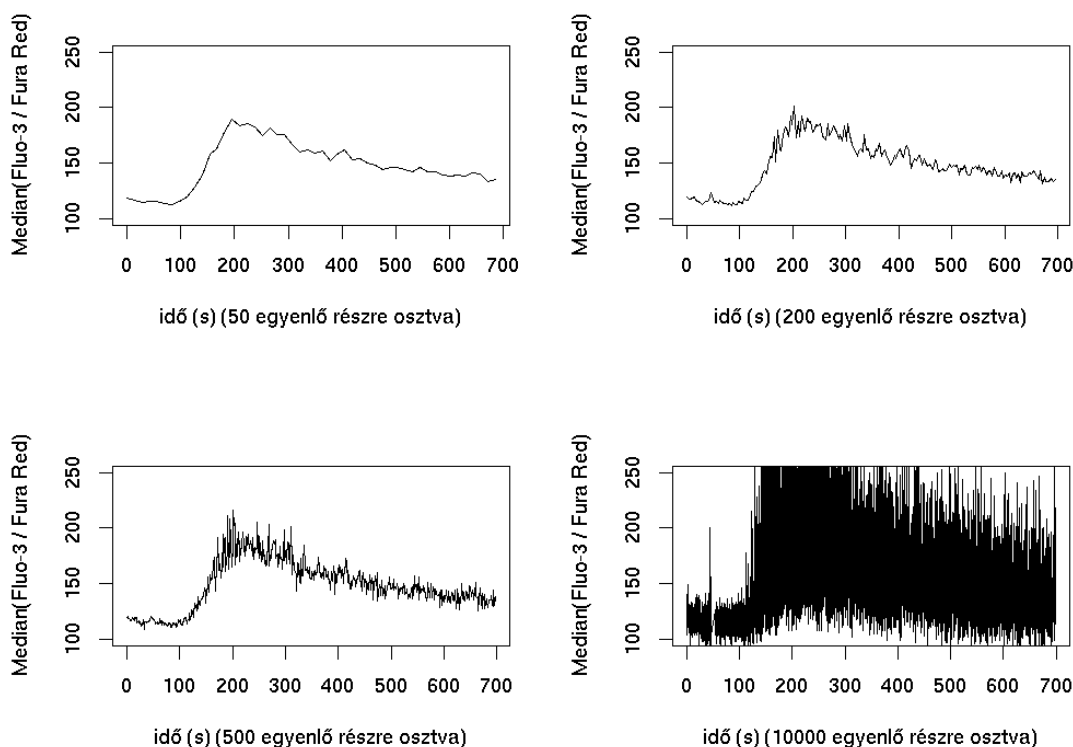
Hogy különböző mérésekből származó görbék összehasonlíthatók legyenek, standardizálni¹³ szokás őket, vagyis az y értékeket egy számmal szorozni, hogy 1-el kezdődjenek, vagy pontos koncentráció-értéket mutassanak. Ez utóbbihoz kalibrálni kell a készüléket,

¹³a standardizálást normalizálásnak is nevezik



6. ábra. Különböző középérték-számítási módszerek bemutatása. A mérés időtartama 100 azonos hosszúságú intervallumra lett osztva, és mindegyik időtartamban meghatároztuk az átlagot (egybefüggő görbe), geometriai átlagot (pontozott görbe) és mediánt (vonalkázott görbe). A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél $25 \mu\text{g/ml}$ PHA-t használtunk.

ezt Omann és mtsai [79] és Norgauer és mtsai [80] megtették. Az egyszerűbb, 1-re való standardizáláshoz általában a mérés első részének átlagával/mediánjával osztják a görbét. Ez lehet az első 5 s [81], 20 s [77], 30 s [78] vagy 40 s [75], [82], [84]. Ha készül külön alapvonal-mérés, akkor annak középértékét lehet erre a célra felhasználni.

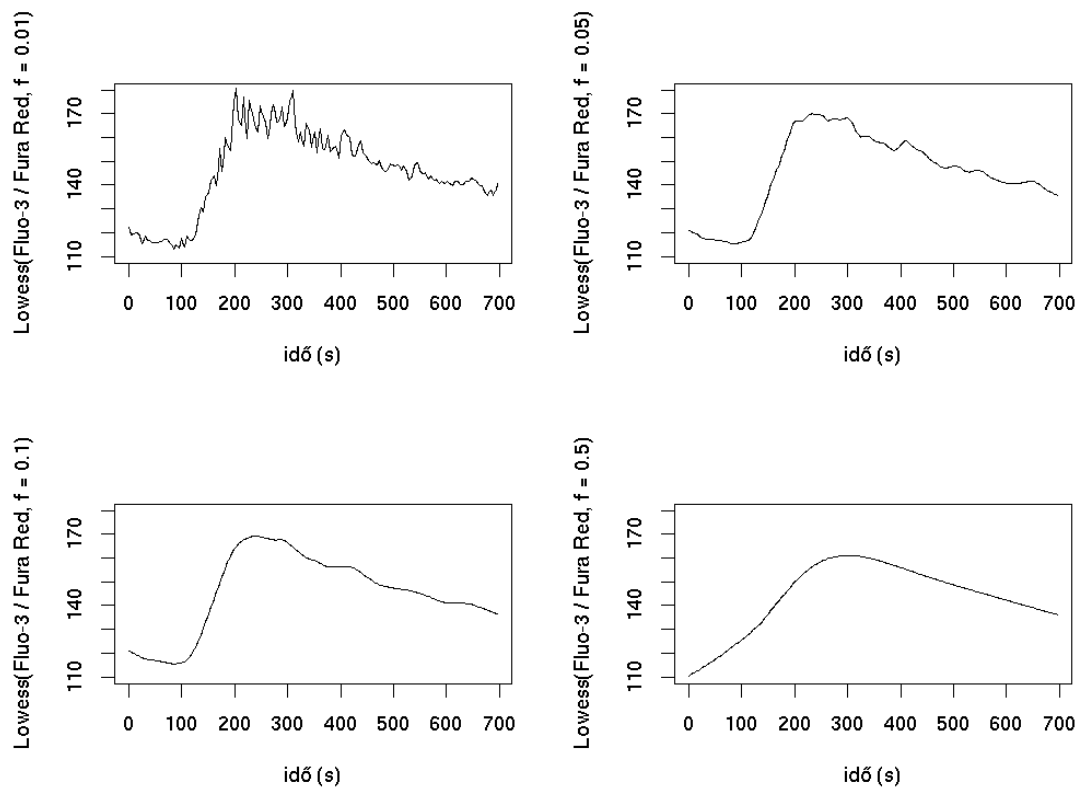


7. ábra. Különböző finomságú medián simítás (lásd x tengelyek felirata). A mérés időtartamát 50, 200, 500 ill. 10000 egyenlő hosszúságú részre osztottuk, és mindegyik intervallumban ábráztuk a mediánt. A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél 25 $\mu\text{g/ml}$ PHA-t használtunk.

Görbék paramétereinek számítása: a görbékből az irodalomban a következő paramétereket olvasták le (9. ábra):

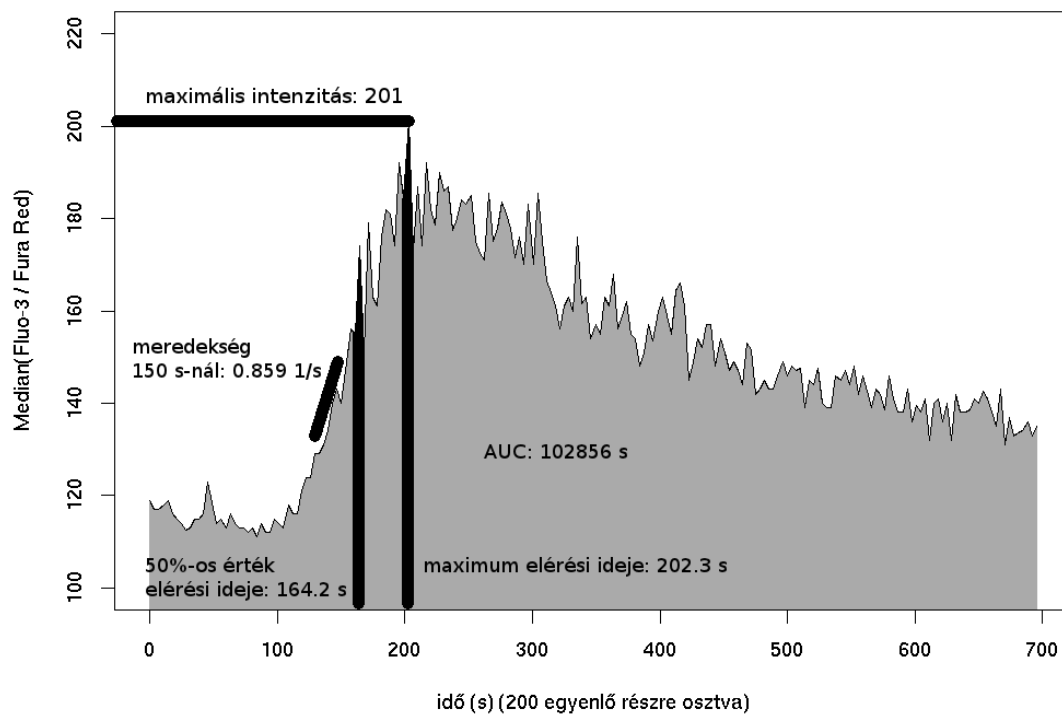
- intenzitás adott idővel a mérés kezdete után [79], [75], [80], [82], [84], [78]
- maximális intenzitás [80], [76], [77], [16], [81]
- maximális intenzitás elérési ideje [81]
- maximum intenzitás érték 50%-ának elérési ideje [4]
- meredekség adott idővel a mérés kezdete után [82]
- AUC (görbe alatti terület) [81]

A paraméterekhez biológiai jelentés társítható, pl. a maximális intenzitás a maximális sejten belüli kalcium-koncentrációval arányos, a maximum elérésének ideje a kalciumválasz gyorsaságával, míg a meredekség azzal, hogy milyen hirtelen növekszik meg ez a



8. ábra. Különböző finomságú lowess-típusú [83] simítás: az f simítási paraméter értékei 0,01, 0,05, 0,1, 0,5. A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél $25 \mu\text{g/ml}$ PHA-t használtunk.

kalcium-koncentráció. A görbe alatti terület a kalcium-atomok mérés során citoplazmában töltött idejének összegével arányos. Ezekkel a számított paraméterekkel jellemezve a méréseket statisztikai összehasonlítások végezhetők, pl. t-teszt (pl. [75]), Mann-Whitney U teszt (pl. [82], [84], [81]), párosított Wilcoxon-teszt (pl. [78], [81]) stb.



9. ábra. Példa paraméterekre, melyek egy kinetikus görbéből leolvashatók. A függőleges tengelyen levő adatokat mértékegység nélkülnek vettem.

3. fejezet

Célkitűzések

Célul tűztük ki egy, a kinetikus áramlási citometriás mérések leírására és összehasonlítására szolgáló algoritmus kifejlesztését, mely

1. objektív numerikus értékeket ad, ezáltal statisztikai összehasonlítást tesz lehetővé,
2. nem függ a felhasználó által ad-hoc módon megadott beállításoktól, így reprodukálható eredményeket ad,
3. a kinetika különböző aspektusait (pl. maximum, válasz gyorsasága) képes számszerűsíteni,
4. megfelelő a kalcium-flux mérések kinetikájának leírására, de kellően robusztus ahhoz, hogy más, hasonlóan bonyolult kinetikájú méréseket leírjon,
5. hatékonyan implementálható és a kutatásban felhasználható.

4. fejezet

Módszerek

Egy kinetikus FCM mérés során sejtekről állapítunk meg paramétereket. Innentől ezek közül kettővel foglalkozunk, az idő paraméterrel és az ún. kinetikus paraméterrel. Mérésen innentől kinetikus FCM mérést értünk, pontosabban $t, r : [1, ncell] \rightarrow \mathbb{R}$ vektorokat¹. t_i az i . sejthez tartozó időpont, r_i az i . sejthez tartozó kinetikus paraméter érték, $ncell$ a sejtek száma.

Hogy kifejlesszük a célkitűzéseknek megfelelő eljárást, mely tkp. numerikus jellemzőket (lásd 2.2.4.3. alfejezet) számít ki a kinetikus mérésekből, az alább leírt algoritmusokat implementáltuk R [24] nyelven. Az algoritmusokat aszerint csoportosítottuk, hogy a jellemzők számításának mely lépésében vesznek részt. Először a paraméterek lehetséges transzformációiról írunk (4.1.), majd a mérési adathalmaz redukálásáról (4.2.), ezután függvények illesztéséről a (redukált) adathalmazra (4.3.). Standardizálásra (4.4) és eloszlások meghatározására (4.5) van szükség ahhoz, hogy a függvényeket össze tudjuk hasonlítani (4.6). Az egyes alfejezetekben leírjuk az általunk arra a lépésre kipróbált különböző megoldásokat. Az algoritmusokat elsődlegesen a 4.7. alfejezetben leírt méréseken teszteltük.

4.1. Dekvantálás

4.1.1. Idő paraméter dekvantálása

Bizonyos FCM-ek a mérési adatokat olyan formátumban adják meg, melyben az idő paraméter felbontása pl. 2,56 s, ami azt jelenti, hogy tetszőleges sejtek közötti időkülönbség ennek többszöröse. Mivel tudjuk, hogy az FCM szekvenciálisan mérte meg a sejteket, és ebben a sorrendben kapjuk meg a táblázatban az eredményeket, az alábbi algoritmussal újraszámoljuk az időpontokat, hogy minden sejthez különböző időpont legyen rendelve,

¹Jelölés: $[a, b] \subset \mathbb{N}$, $[a, b] = \{a, a + 1, a + 2, \dots, b\}$

amely ráadásul jobban megközelíti a tényleges mérési időpontot.

Legyen Δ a legkisebb pozitív különbség, ami két sejt idő paramétere között van:

$$\Delta = \min \left\{ x - y \mid x, y \in \bigcup_{i \in [1, n_{\text{cell}}]} \{t_i\} \right\}.$$

Azokra az egymás után következő időértékekre, melyek egyenlőek, vagyis minden i, j -re, melyekre $t_{i-1} < t_i$ és $t_j < t_{j+1}$ és $t_i = t_j$ (úgy definiálva, hogy $t_0 < t_1$ és $t_{n_{\text{cell}}} < t_{n_{\text{cell}}+1}$) a következőt definiáljuk:

$$(t'_i, t'_{i+1}, t'_{i+2}, \dots, t'_j) = \left(t_i, t_i + \frac{1}{j-i+1} * \Delta, t_i + \frac{2}{j-i+1} * \Delta, \dots, t_i + \frac{j-i}{j-i+1} * \Delta \right)$$

A dekvantált idővektor az így kapott t' vektor lesz.

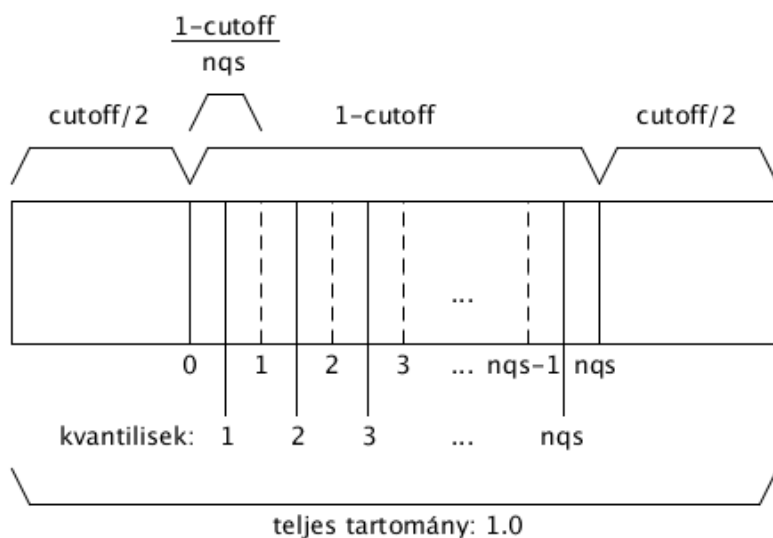
4.1.2. Dekvantálás általános esetben

Ha nem használhatunk fel olyan többlet-tudást, mint az idő paraméternél az előbb, mégis szeretnénk, hogy minden érték egy $v : [1, n] \rightarrow \mathbb{R}$ vektorban különböző legyen, akkor, ha a számok lebegőpontos értékeként vannak tárolva, a mantissza m biten van ábrázolva ($m > 6$), és v -nek a K halmazbeli indexű elemei egyenlőek (tehát $|\bigcup_{i \in K} \{v_i\}| = 1$ és $\forall i \in [1, n] \setminus K : v_i \neq v_j$, ahol $j \in K$), a $v_j * 2^{-m+3} * [0, 1, 2, \dots, |K| - 1]$ halmaz elemeit (ahol $j \in K$) adjuk hozzá a K indexű vektorelemekhez tetszőleges sorrendben, így mindegyik szám különbözni fog, és az értékük csak a nem szignifikáns biteken módosul kissé tetszőleges nagysegrendű számok esetén. Kis esély ugyan van arra, hogy a módszerrel eddig nem egyenlő számokat egyenlővé teszünk, de FCM mérési adatok esetén ez elenyésző, mert a készülékek mérési pontatlansága jóval nagyobb, mint a lebegőpontos számábrázolás pontatlansága, emiatt a mért értékek a legkevésbé szignifikáns biteken jellemzően megegyeznek. Ez a dekvantálási algoritmus a `flowFP` [68] saját dekvantálási algoritmusának (`dequantize` függvény) továbbfejlesztett változata.

4.2. Helyettesítés

A mérést felosztjuk nts egyenlő hosszúságú egymást követő időintervallumra², s minden időintervallumban meghatározunk nqs kvantilist *cutoff* levágással, mégpedig a követke-

²Annak érdekében, hogy ezekben az intervallumokban egyenletesen oszoljanak meg a sejtek akkor is, ha az intervallumok hossza összemérhető a készülék felbontásával, szükséges az idő paraméter dekvantálása, lásd 4.1.1. alfejezet.



10. ábra. n db egyenletesen elosztott kvantilis kétoldali $cutoff$ nagyságú levágással.

zöket (10. ábra):

$$q_s = \frac{cutoff}{2} + \frac{1 - cutoff}{2nqs} + \frac{1 - cutoff}{nqs} [0, nqs - 1]$$

Ha megfelelően nagy számokat választunk nts és nqs értékeként, és nem áll fönn $ncell \ll nts * nqs$, a mérés helyettesíthető $nts * nqs$ db számmal, vagyis egy $qdata \in \mathbb{R}^{nqs \times nts}$ mátrix-szal, ahol $qdata_{ij}$ a j . időintervallumban az i . kvantilis értéke, ill. az időintervallumok kezdetének $times \in \mathbb{R}^{nts}$ vektorával. Ha vannak olyan időintervallumok, ahol kevesebb, mint nqs mérési adat van, ott a kvantilis-értékek pontatlanok lesznek, szélsőséges esetben nem lesznek kvantilis értékek, ilyenkor azt az időintervallumot kihagyjuk, és nts kisebb lesz.

A helyettesítés során létrejövő adathalmazzal gyorsabban végezhetőek számítások, mint az eredetivel. A helyettesítéssel ugyan elveszítjük az eredeti, $ncell$ -nek megfelelő, akár milliós szabadsági fokot, de, mivel a statisztikai tesztek szempontjából érdekes elemszám a mérések száma, nem pedig a méréseken belüli sejttség, ezt az eredeti szabadsági fokot nem is használnánk fel. Szintén elveszítjük az idő paraméter és a sejttség összefüggését: az értékelést nem fogják jobban befolyásolni azok az időintervallumok, ahol több sejt van; tehát egyenletesítjük az idő szerinti sejt-eloszlást.

4.3. Függvény-illesztés

Kinetikus FCM méréseket jellemző eljárásunk kifejlesztésének alapötlete volt, hogy az idő és a kinetikus paraméter összefüggését egy függvénnyel jellemezzük, vagyis függvényt illesszünk a mérésre. Ez azt jelenti, hogy egy megfelelő algoritmussal úgy állítjuk be a függvény paramétereit, hogy a mérési pontoktól való eltérés lehetőleg minimális legyen. Ilyen eljárás pl. egyenes függvény ($y = a * x + b$) illesztésére a lineáris regresszió [85], amely az eltérés-négyzetösszeget (Sum of Squared Deviation, SSD) minimalizálja, a függvény paramétereit a (meredekség) és b (tengelymetszet).

4.3.1. Függvények

A matematikában jól ismert S-alakú **logisztikus** függvény általános képlete:

$$\text{logist}(t; b, c, d, e) = c + \frac{d - c}{1 + \left(\frac{t}{e}\right)^b}$$

Értelmezési tartománya \mathbb{R}^+ , értékészlete $[\min(c, d), \max(c, d)]$.

Paramétereit:

- b : meredekséggel kapcsolatos paraméter, ha pozitív, akkor $\text{logist}(0) = d$ és $\lim_{\infty}(\text{logist}) = c$; ha negatív, akkor $\text{logist}(0) = c$ és $\lim_{\infty}(\text{logist}) = d$; ha 0, akkor a görbe átmegy $\frac{c+d}{2}$ értékű konstans függvénybe.
- c : b -től függően a kezdő- vagy végső érték
- d : b -től függően a végső- vagy kezdő érték
- e : a $\frac{c+d}{2}$ érték elérési ideje

A logisztikus függvénnyel jól leírhatók olyan időbeli folyamatok, melyek egy konstans értékről indulnak, majd fokozatosan egy másik konstans értéket érnek el. A biológiában gyakran használják pl. növekedési vagy dózis-hatás görbék leírására. Egy módosított változatát a **hormesis** nevű biológiai jelenség [86] leírására dolgozták ki Cedergreen és mtsai [87]. A hormesis bizonyos dózis-hatás görbék jellegzetessége, miszerint kis dózisokra ellentétes reakció alakul ki, mint nagyobbakra: egy adott dózistartományon belül a hatás dózsisfüggő módon emelkedik, egy adott dózisszint elérése után pedig dózsisfüggő módon csökken. A jelenséget először toxikológiában figyelték meg [88]. A Cedergreen és mtsai által kidolgozott függvény képlete:

$$\text{hormesis}(t; b, c, d, e, f) = c + \frac{d - c + f * \exp\left(\frac{-1}{t^{0,25}}\right)}{1 + \left(\frac{t}{e}\right)^b}$$

A függvényt a következőképp módosítottuk (az e paraméter bekerült a számlálóba az $f*$... levágó-függvény alakját módosítandó):

$$hormesis_1(t; b, c, d, e, f) = c + \frac{d - c + f * \exp\left(\frac{-1}{(t*e)^{0.25}}\right)}{1 + \left(\frac{t}{e}\right)^b}$$

A $hormesis_1$ függvény értelmezési tartománya \mathbb{R}^+ , b, c, d, e paramétereinek értelme hasonló a logisztikus függvény paramétereihöz, a függvény határértéke $b < 0$ esetén $+\infty$ -ben $d + f$. $f = 0$ esetén megegyezik a logisztikus függvénnyel ($f = b = 0$ esetén pedig konstans függvény). Alakja a következőképp függ a paramétereiktől:

- $c > d$: a függvény alapvetően negatív meredekségű
 - $f < 0$: a függvény végig negatív meredekségű
 - $0 < f < c - d$: a függvény először negatív, majd pozitív meredekségű, értéke $+\infty$ -ben kisebb a kezdőértéknél
 - $f > c - d$: a függvény először negatív, majd pozitív meredekségű, értéke $+\infty$ -ben nagyobb a kezdőértéknél
- $c < d$: a függvény alapvetően pozitív meredekségű
 - $f > 0$: a függvény végig pozitív meredekségű
 - $c - d < f < 0$: a függvény először pozitív, majd negatív meredekségű, értéke $+\infty$ -ben nagyobb a kezdőértéknél
 - $f < c - d$: a függvény először pozitív, majd negatív meredekségű, értéke $+\infty$ -ben kisebb a kezdőértéknél

Az előbbi függvények paramétereinek értelmezése nem triviális, hiszen pl. a logisztikus függvény egyik paraméterének előjelétől függ, hogy másik két paramétere mit is jelent (a függvény kezdő értékét vagy végső értékét). Emiatt kidolgoztunk egy 5 olyan függvényből álló függvény-készletet, melyeknek az összes paramétere könnyen értelmezhető, és a „matematikai” és „biológiai” paraméterek egybeesnek³. Ezt újabb paraméterek bevezetésével, régiak kiküszöbölésével és paraméter-megszorításokkal értük el. Az 5 függvény a következő:

1. Konstans függvény.

$$constant(t; y) = y$$

³A logisztikus ill. a hormesis függvény paramétereiből numerikus módszerekkel ugyan kiszámíthatók ugyanezek a biológiailag használható paraméterek, de az, hogy a paramétereknek nincs egyszerű értelmezések, azzal is jár, hogy illesztésnél az induló paramétereket nehéz megbecsülni (lásd 4.3.2. alfejezet).

Paraméter megszorítás: $y \geq 0$

Paraméter:

- y (**konstans érték**, constant value): a függvény értéke tetszőleges időpontban

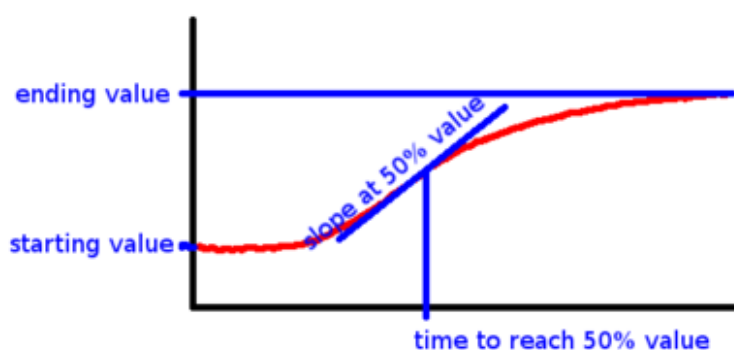
2. **Pozitív logisztikus** függvény (11. ábra): ez a függvény egy adott y_0 értékről indul és értéke folyamatosan növekedve végül végtelenben egy adott y_2 értéket ér el. A logisztikus függvény átparaméterezése.

$$\text{logist}_+(t; y_0, y_2, x_1, m_1) = \frac{y_0 + (y_2 - y_0)}{1 + \exp\left(\frac{4 \cdot m_1 \cdot (-t + x_1)}{y_2 - y_0}\right)}$$

Paraméter megszorítások: $y_0, y_2, x_1, m_1 \geq 0$ and $y_0 < y_2$

Paraméterek:

- y_0 (**kezdő érték**, starting value): a függvény határértéke $-\infty$ -ben
- y_2 (**végső érték**, ending value): a függvény határértéke $+\infty$ -ben (nagyobb, mint a kezdő érték)
- x_1 (**50%-os érték elérési ideje**, time to reach 50% value): az az időpont, amikor a függvény értéke a kezdőérték és végső érték átlaga
- m_1 (**meredekség az 50%-os értéknél**, slope at 50% value): az 50%-os értéknél a függvény meredeksége (pozitív)



11. ábra. A pozitív logisztikus függvény (logist_+) paraméterei.

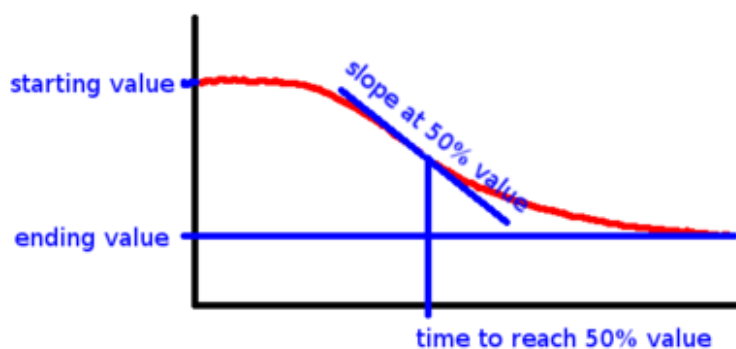
3. **Negatív logisztikus** függvény (12. ábra): ez a függvény egy adott y_0 értékről indul és értéke folyamatosan csökkenve végül végtelenben egy adott y_2 értéket ér el. A logisztikus függvény átparaméterezése.

$$\text{logist}_-(t; y_0, y_2, x_1, m_1) = \frac{y_0 + (y_2 - y_0)}{1 + \exp\left(\frac{4 \cdot m_1 \cdot (-t + x_1)}{y_2 - y_0}\right)}$$

Paraméter megszorítások: $y_0, y_2, x_1 \geq 0, m_1 \leq 0$ and $y_2 < y_0$

Paraméterek:

- y_0 (**kezdő érték**, starting value): a függvény határértéke $-\infty$ -ben
- y_2 (**végső érték**, ending value): a függvény határértéke $+\infty$ -ben (kisebb, mint a kezdő érték)
- x_1 (**50%-os érték elérési ideje**, time to reach 50% value): az az időpont, amikor a függvény értéke a kezdőérték és végső érték átlaga
- m_1 (**meredekség az 50%-os értéknél**, slope at 50% value): az 50%-os értéknél a függvény meredeksége (negatív)



12. ábra. A negatív logisztikus függvény (*logist₋*) paraméterei.

4. **Pozitív dupla logisztikus** függvény (13. ábra): ez a függvény egy adott y_0 értékről indul, értéke folyamatosan növekedve elér egy y_1 maximumot, majd innen folyamatosan csökkenve végül végtelenben egy adott y_2 értéket ér el. Ez a függvény két logisztikus függvény egymáshoz illesztése úgy, hogy az illesztésnél az összes deriváltjuk megegyezik.

$$dlogist_+(t; y_0, y_1, y_2, x_1, x_{d0}, x_{d2}, m_0, m_2) =$$

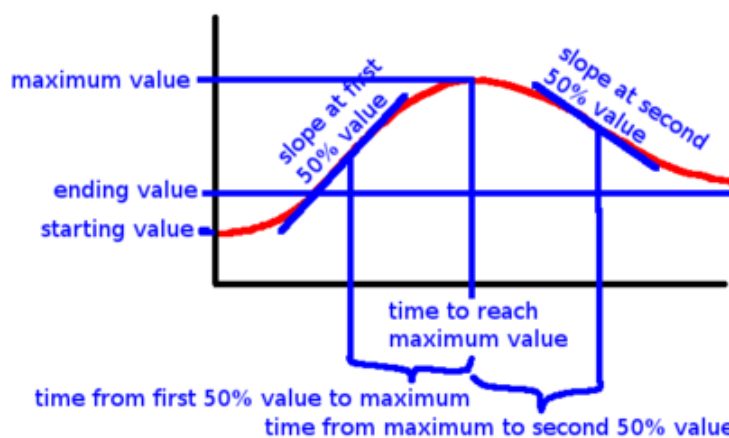
$$= \begin{cases} y_0 + \frac{y_1 - y_0}{1 + \left(\frac{x_1 - t}{x_{d0}}\right)^{\frac{4 * x_{d0} * m_0}{y_1 - y_0}}} & \text{if } t < x_1 \\ y_2 + \frac{y_1 - y_2}{1 + \left(\frac{t - x_1}{x_{d2}}\right)^{\frac{4 * x_{d2} * m_2}{y_2 - y_1}}} & \text{if } t \geq x_1 \end{cases}$$

Paraméter megszorítások: $y_0, y_1, y_2, m_0, x_1, x_{d0}, x_{d2} \geq 0, m_2 \leq 0,$
 $x_{d0} \leq x_1, y_1 > y_0, y_1 > y_2$

Paraméterek:

- y_0 (**kezdő érték**, starting value): a függvény határértéke $-\infty$ -ben

- y_1 (**maximum érték**, maximum value): a függvény maximumának értéke
- y_2 (**végző érték**, ending value): a függvény határértéke $+\infty$ -ben
- x_1 (**maximum érték elérési ideje**, time to reach maximum value): az az időpont, amikor a függvény értéke megegyezik a maximum értékkel
- x_{d0} (**az első 50%-os érték és a maximum közötti idő**, time from the first 50% value to maximum): a távolság az első 50%-os érték és a maximum elérési ideje között (előbbi a kezdő érték és a maximum érték átlaga)
- x_{d2} (**a maximum és a második 50%-os érték közötti idő**, time from maximum to the second 50% value): a távolság a maximum érték és a második 50%-os érték elérési ideje között (ez utóbbi a maximum érték és a végző érték átlaga)
- m_0 (**meredekség az első 50%-os értéknél**, slope at first 50% value): az első 50%-os értéknél a függvény meredeksége (pozitív)
- m_2 (**meredekség a második 50%-os értéknél**, slope at second 50% value): a második 50%-os értéknél a függvény meredeksége (negatív)



13. ábra. A pozitív dupla logisztikus függvény ($dlogist_+$) paraméterei.

5. **Negatív dupla logisztikus** függvény (14. ábra): ez a függvény egy adott y_0 értékről indul, értéke folyamatosan csökkenve elér egy y_1 minimumot, majd innen folyamatosan növekedve végül végtelenben egy adott y_2 értéket ér el. Ez a függvény két logisztikus függvény egymáshoz illesztése úgy, hogy az illesztésnél az összes

deriváltjuk megegyezik.

$$dlogist_{-}(t; y_0, y_1, y_2, x_1, xd_0, xd_2, m_0, m_2) =$$

$$= \begin{cases} y_0 + \frac{y_1 - y_0}{1 + \left(\frac{x_1 - t}{xd_0}\right)^{\frac{4 * xd_0 * m_0}{y_1 - y_0}}} & \text{if } t < x_1 \\ y_2 + \frac{y_1 - y_2}{1 + \left(\frac{t - x_1}{xd_2}\right)^{\frac{4 * xd_2 * m_2}{y_2 - y_1}}} & \text{if } t \geq x_1 \end{cases}$$

Paraméter megszorítások: $y_0, y_1, y_2, m_2, x_1, xd_0, xd_2 \geq 0, m_0 \leq 0, xd_0 \leq x_1, y_1 < y_0, y_1 < y_2$

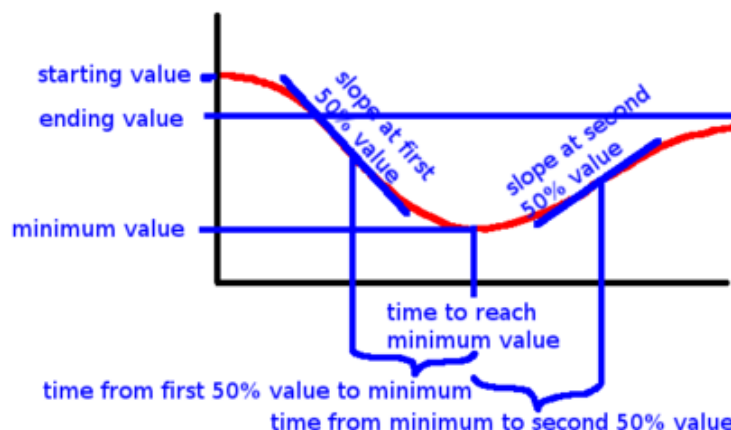
Paraméterek:

- y_0 (**kezdő érték**, starting value): a függvény határértéke $-\infty$ -ben
- y_1 (**minimum érték**, minimum value): a függvény minimumának értéke
- y_2 (**végző érték**, ending value): a függvény határértéke $+\infty$ -ben
- x_1 (**minimum érték elérési ideje**, time to reach minimum value): az az időpont, amikor a függvény értéke megegyezik a minimum értékkel
- xd_0 (**az első 50%-os érték és a minimum közötti idő**, time from the first 50% value to maximum): a távolság az első 50%-os érték és a minimum elérési ideje között (előbbi a kezdő érték és a maximum érték átlaga)
- xd_2 (**a minimum és a második 50%-os érték közötti idő**, time from maximum to the second 50% value): a távolság a maximum érték és a minimum 50%-os érték elérési ideje között (ez utóbbi a maximum érték és a végző érték átlaga)
- m_0 (**meredekség az első 50%-os értéknél**, slope at first 50% value): az első 50%-os értéknél a függvény meredeksége (negatív)
- m_2 (**meredekség a második 50%-os értéknél**, slope at second 50% value): a második 50%-os értéknél a függvény meredeksége (pozitív)

A fenti függvényeknek számítható egy további, AUC paramétere, mely a függvény integráltja 0-tól egy adott t időpontig.

4.3.2. Illesztési módszerek

Egy minta átlaga a minta elemeitől való négyzetes eltérést (SSD-t) minimalizálja. Ha a $x, y : [1, n] \rightarrow \mathbb{R}$ vektorokra, ahol x egyenletes eloszlású és y szórása független x -től, egy SSD-t minimalizáló illesztő-módszerrel függvényt illesztünk, a függvény a mozgó-átlaghoz közel fog elhelyezkedni (a feltételes átlagra illesztünk). Egy minta mediánja a



14. ábra. A negatív dupla logisztikus függvény (*dlogist_*) paraméterei.

minta elemeitől való abszolút eltérést (SAD-ot) minimalizálja. És hasonlóképpen, ha egy SAD-ot minimalizáló módszerrel illesztünk, a függvény a mozgómediánhoz közel fog elhelyezkedni (a feltételes mediánra illesztünk). Utóbbira lásd 19. ábra.

Emiatt, ha a mérést egy $qdata \in \mathbb{R}^{nqs \times nts}$ mátrix-szal ill. $times \in \mathbb{R}^{nts}$ vektorral helyettesítettük (4.2. alfejezet), a $times$, $qdata_{floor(nqs/2)}$ adatokra (időintervallumonkénti mediánokra) illesztett függvény ugyanoda fog illeszkedni, mint az eredeti mérésre SAD minimalizáló algoritmussal illesztett függvény. Ez akkor igaz, ha a sejtek idő szerinti eloszlása egyenletes, ill. a kinetikus paraméter szórása nem függ az idő paramétertől. Ha ez nincs így, akkor a mediánokra illesztés olyan eredményt ad, mintha egyenletesítettük⁴ volna az eloszlást ezekből a szempontokból: az illesztést ugyanannyira fogják befolyásolnia különböző időpontok.

Mivel az FCM eloszlások nem normálisak, a mozgóátlagnak megfelelő SSD-minimalizáló illesztő megoldások nem megfelelőek, a centrális tendencia mérésére a mozgómediánnak megfelelő SAD-t minimalizáló illesztési módszert használtunk.

Nem lineáris függvények illesztésére nem létezik mindig legjobb megoldást adó eljárás, ehelyett iterációs megoldásokat használhatunk, melyek a paraméterek értelmezési tartománya által meghatározott problématerben keresik a minimumot pl. a minimalizálandó függvény különböző paraméterek szerinti parciális deriváltjának értékei alapján. Sok különböző ilyen módszer létezik [89], mi a következőket használtuk:

- Newton-módszer [90]: tetszőleges függvényt képes minimalizálni, szükségesek a függvény második deriváltjai, R-ben megvalósítja a `stats` csomag `nlm` függvénye.
- Gauss-Newton módszer: SSD-t minimalizál, R-ben megvalósítja a `stats` csomag

⁴Az egyenletesítés elméletileg pl. történhet úgy, hogy úgy súlyozzuk az egyes értékeket, hogy az össz-súlyok idő szerint egyenletesen oszoljanak meg.

`nls` függvénye.

- Gauss Newton módszer robusztus változata (Iterated Reweighted Least Squares, IWLS): SSD-t minimalizál, R-ben megvalósítja a `robustbase` csomag `nls` függvénye.
- BFGS módszer (Broyden, Fletcher, Goldfarb és Shanno) [91]: egy kvázi-Newton módszer, függvényértékeket és gradienseket használ, kvantilis-regresszió (tehát pl. medián regresszió, vagyis SAD-minimalizálás) céljára implementálja a `quantreg` csomag `nls` függvénye.
- NM módszer (Nelder-Mead) [92]: szimplex algoritmus tetszőleges függvény minimalizálására, R-ben megvalósítja a `stats` csomag `optim` függvénye.
- SANN módszer (Simulated ANNealing) [93]: szimulált hűtés algoritmus, R-ben megvalósítja a `stats` csomag `optim` függvénye.

A fenti illesztő algoritmusok elindításához becsléseket kell adni a paraméterekre. Hogy az algoritmus megtalálja -e a tényleges minimumot, erősen függ az induló paraméterektől [94]. Az induló paraméterbecslést az adathalmaz `lowess`-módszerrel [83] történő simításával és a paramétereknek a simított görbéről való közvetlen leolvasásával végezzük. A logisztikus és hormesis függvény b ill. g paramétere nem olvasható le közvetlenül a simított görbéről, ezeket egy-paraméteres optimalizálással határozzuk meg, a `golden section search` algoritmus módosított változatával (az R `stats` csomagjának `optimize` függvénye). Egy illesztés kimenete az eltérésösszeg (SSD vagy SAD) érték és az illesztett paraméterek.

4.3.3. Illesztés jóságának becslése

Az illesztés jóságát az eltérésösszeg (SAD/SSD) értéke jellemzi, minél kisebb ez az érték, annál kevésbé szórnak az adatok a függvény körül, tehát annál jobb az illesztés. Általánosságban, ha egy függvény minél több paraméteres, annál kisebb lesz ez az érték, de annál valószínűbb az is, hogy az adatokban tendenciózan meglevő minták helyett a véletlen szóródás miatti mintákat fogja követni a függvény. Ezt a jelenséget túlillesztésnek (overfit) hívják [95].

A túlillesztés elkerülésére F-tesztet lehet használni: ha a szórások nem különböznek szignifikánsan, akkor a kevesebb paraméteres modellt használjuk.

A túlillesztés elkerülésére dolgozták ki a kereszt-validáció (cross validation, CV) módszerét [96] is. Egy gyakorlatias bemutatása a témának Moore prezentációja [97].

Az n -szeres kereszt validáció (n -fold cross validation) a következőképp működik:

1. Az adathalmazt, amire illesztünk, n db random egyenlő csoportra osztjuk.
2. Minden lépésben az n csoport közül egy másik lesz a teszt halmaz (test set), a maradék $n - 1$ a tanító halmaz (training set). A tanító halmazra illesztjük az adott modellt, és a teszt halmazon számítjuk az eltérésösszeget (SSD-t vagy SAD-ot, az illesztési módszernek megfelelően). Így, ha a modell sok paraméteres, igaz, hogy követni fogja a véletlen szóródás miatti mintákat a tanuló halmazban, de ezek a minták a teszt halmazban nem lesznek meg, így a rugalmas modell (sok paraméter) ellenére növekedni fog az eltérésösszeg.
3. Az előbbi lépést megismételjük n -szer, és a kapott eltéréseket összegezzük. Minél kisebb lesz ez az összeg, annál jobban illeszkedik a modell az adatokra.

4.4. Standardizálás

Hogy különböző körülmények között felvett mérések összehasonlíthatók legyenek, standardizálás szükséges (lásd még 2.2.4.2. alfejezet). A vízszintes irányú (idő szerinti) standardizálás a paraméterek megfelelő választásával elérhető, emiatt használjuk a „maximum és az 50%-os értékek közötti különbség”-típusú paramétereket a dupla logisztikus függvényeknél ahelyett, hogy abszolút idő-paramétereket használnánk, pl. „50%-os érték elérési ideje”. Így ezek a paraméterek olyan mérések esetén is összehasonlíthatók lesznek, melyek időben eltérően kezdődtek, az eltérő kezdést pedig a „maximum elérési ideje” paraméter fejezi ki. A függőleges irányú standardizáláshoz az összes olyan paramétert, melyeknek van ilyen komponense (tehát az y , y_0 , y_1 , y_2 , m_0 és m_2 paraméterek a különböző függvényeknél) elosztjuk egy adott értékkel, ami a következő lehet:

- a függvény kezdő paramétere. Ekkor a függvény 1-ről fog kezdődni, pl. a pozitív logisztikus függvény kezdő érték paramétere 1 lesz.
- a mérés kezdeti szakaszának medián értéke. Ha készült alapvonal-mérés, akkor ennek mediánja. Ekkor a kezdő érték lehet 1-től különböző is, a mérés kezdetén (a 0 időpont környékén) fog 1-et felvenni a függvény.

4.5. Paraméterek eloszlásának generálása

4.5.1. Medián módszer

Medián módszernek nevezzük azt, amikor egy mérésre SAD módszerrel (4.3.2. alfejezet) 1 db függvényt illesztünk, mely így a mérés mediánját jellemzi. A medián módszerrel egy

mérésből annyi paramétert kapunk, ahány paraméteres a függvény. Ilyenkor egy értékből áll az eloszlás, tehát tkp. nem generálunk eloszlást.

4.5.2. Medián eloszlása

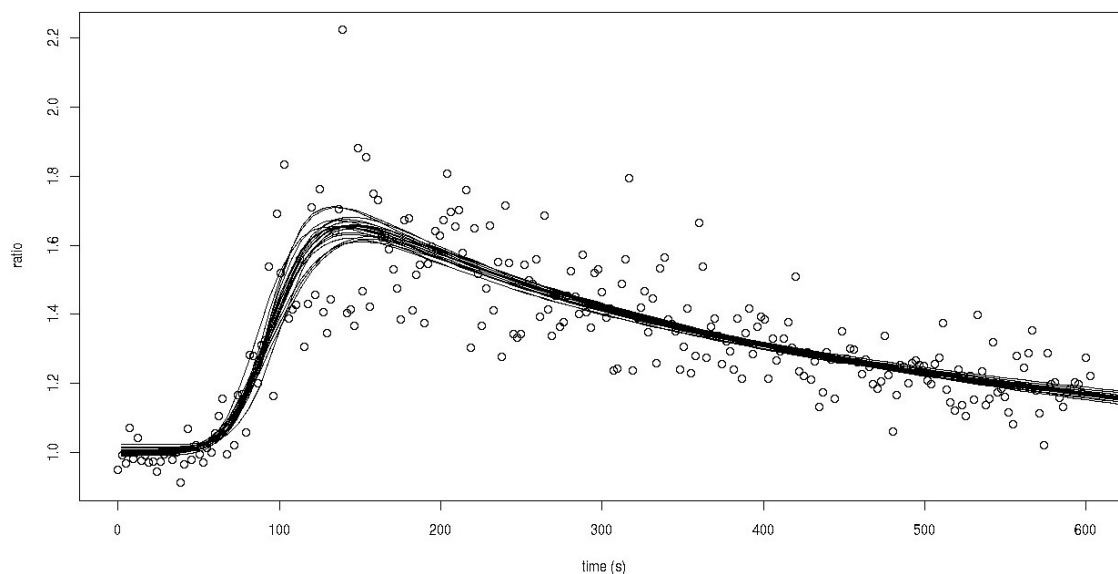
A medián módszert használva a paraméterek azt jellemzik, hogy a mérés közepe hogyan viselkedik, de nem adnak semmilyen adatot a mérés többi tartományáról, ill. hogy hogyan szórhatnak ezek a paraméterek. A paraméterek eloszlása (és így konfidencia intervalluma) bizonyos illesztő módszerek esetén megbecsülhető, pl. Gauss-Newton módszer használatakor profile likelihood-ok használatával [89]. Számított paraméterek (pl. hormesis függvény maximum paramétere) esetén az illesztett paraméterek becsült eloszlásaiból delta módszerrel (delta method, [98]) kiszámíthatók a számított paraméterek becsült eloszlásai. Ha az illesztő módszerünk nem ad eloszlásokat, vagy a delta módszer nehezen alkalmazható, mert nem lehet kifejezni képlettel valamelyik paramétert, újra-mintavételezési eljárással, bootstrap módszerrel [99] is megkaphatjuk azokat: a mérési adatokkal megegyező méretű visszatéves mintákat veszünk a mérési adatokból, és ezekre végzünk el az illesztéseket (lásd 15. ábra). Az így kapott függvények az eredeti mérésre illesztett függvény körül fognak szórni, és az egyes függvények megfelelő paraméterei megadják annak a paraméternek az eloszlását, konfidencia intervallumát.

Ezek az eloszlások tkp. a medián görbe paramétereinek eloszlásai (sampling distribution-jei), egy adott paraméter eloszlásából számított szórás megfelel az adott paraméter standard hibájának, vagyis minél nagyobb az elemszám, annál szűkebb eloszlásokat kapunk. Ez a módszer tehát nem végezhető el helyettesítés (4.2. alfejezet) után, mert az eredeti szabadsági fokot lecsökkenti, és így a konfidencia intervallum szélessége nem a mérestől, hanem a helyettesítéshez használt nts és nqs értékektől fog függeni.

4.5.3. Kvantilis módszer

Annak érdekében, hogy a mérési adatoknak ne csak a közepét jellemezzük, egy másik lehetőség a kvantilis⁵ szerinti rétegzett mintavétel a mérési adatokból: a 4.2. alfejezetben leírt $qdata \in \mathbb{R}^{nqs \times nts}$ egyes soraira, vagyis minden $i \in [1, nqs]$ -re a $times$ és $qdata_i$ vektorokra illesztünk, így összesen nqs db függvényt kapunk. Az nqs db függvény megfelelő paraméterei megadják annak a paraméternek az eloszlását (pl. a maximum paraméterre lásd 16. ábra). Ez az eloszlás a paraméter tényleges eloszlása az összes sejt között, feltételezve, hogy a kinetikus folyamat párhuzamosan halad a sejtekben, vagyis ha egy sejtnek egy adott időpontban kicsi a kinetikus paraméter értéke, akkor később is kicsi lesz. Egy adott kvantilishez tartozó függvény az ahhoz a kvantilishez tartozó sejteket helyettesíti.

⁵Ha $nqs = 100$, akkor percentilisekről van szó.



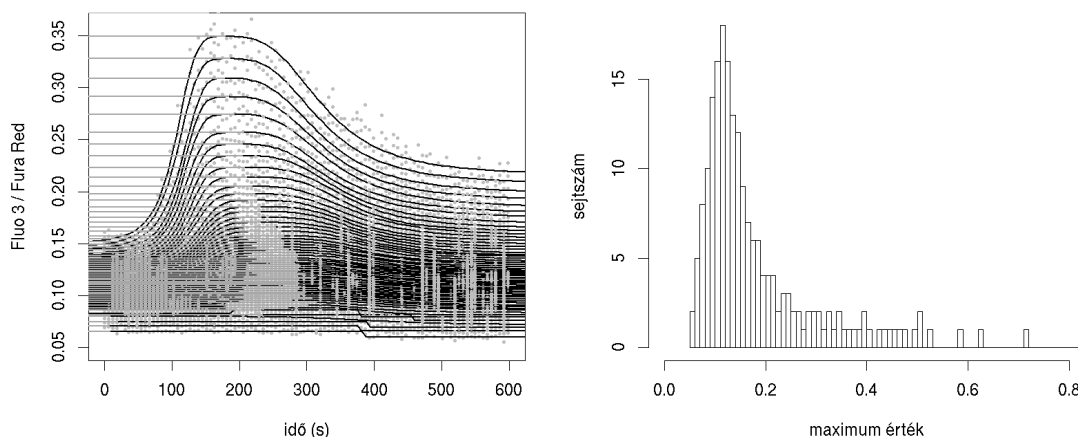
15. ábra. Dupla logisztikus függvény illetve 20 db bootstrap mintára. Az üres körök a 100 időintervallum mediánjait, a fekete görbék a 20 függvényt ábrázolják. A mintákaz visszatevéssel vettük a mérésből, méretük megegyezik az eredeti adathalmaz méretével. A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél $25 \mu\text{g/ml}$ PHA-t használtunk.

Ahol több sejt van, ott sűrűbben vannak a kvantilisok, így több függvény tartozik ahhoz a területhez, ennek megfelelően egy adott paraméter eloszlásában ahhoz a területhez több érték tartozik, így az eloszlás alakját az a terület jobban befolyásolja. Hogy a függvények alakja minden kvantiliséknél konzekvensen hasonló legyen, elkülönítettük a pozitív és negatív függvényeket a logisztikus és a dupla logisztikus függvényeken belül, lásd az 5 függvényből álló készletet a 4.3.1. alfejezetben.

4.5.4. Hálós illesztés

A kvantilis módszer a dupla logisztikus függvény illesztésekor $8 * nqs$ paramétert illeszt, vagyis kvantilisenként külön-külön 8-at. Ha ismert a paraméterek kvantilis-függése, ez használható arra, hogy az illesztett függvény paramétereit feltételes paraméterekként, a kvantilis függvényében adjuk meg. Az illesztés pedig egy lépésben, az nqs számú függvény hozzájuk tartozó kvantiliseitől való eltérések összegét minimalizálva történik.

Ha pl. mindegyik paramétert egy kvantilistól függő harmadfokú polinommal adunk meg ($param = a * q^3 + b * q^2 + c * q + d$, ahol $q \in [0, 1]$ a kvantilis és a, b, c, d a kvantilis-függést meghatározó paraméterek), egy reprezentatív mérésre a 17. ábrán látható illesztést kapjuk. A hálós módszerrel egyszerre az összes kvantilisre illesztett függvények a 18. ábrán láthatók. Látszik, hogy a függvények jól követik a kvantilis értékeket, kivéve az 1. kvantilis esetén, ahol a függvény kicsit feljebb került.



16. ábra. Dupla logisztikus függvény illesztése egy calcium flux mérés különböző kvantilisereire (bal oldal, szürke pontok felelnek meg az egyes kvantilis értékeknek, a szürke vonalak a maximum értékek) és a maximum paraméter eloszlásának meghatározása (jobb oldal). A mérés időtartama $nts = 100$ egyenlő hosszúságú időintervallumra lett osztva, és minden intervallumban meghatároztunk $nqs = 201$ kvantilist a 4.2. alfejeztben megadott módszer szerint. A Fluo-3 / Fura Red arány az intracelluláris Ca^{2+} -koncentrációval arányos [15]. Az FCM mérés előtt a PBMC sejteket $10 \mu\text{g/ml}$ PHA-val stimuláltuk. A mért sejtek közül ki lettek választva a CD4+ limfociták.

Első-, másod- és harmadfokú polinomokkal és normál eloszlás kvantilis-függvényével (ezt úgy módosítva, hogy ha a szórás negatív, a $\lambda q \rightarrow 1 - q$ függvénnyel kompozitáljuk, hogy kvantilistól függően csökkenni is tudjanak az értékek) próbáltuk leírni a kvantilis-függést. Utóbbinál a kinetikus paraméter értékeinek illesztés előtt való logaritmálását is kipróbáltuk. A kvantilis-függést leíró függvényeknek paramétereit (meta-paramétereket, mert ezek határozzák meg a végleges paramétereket) külön-külön illesztettük minden mérésnél, ill. használtunk olyan modellt is, ahol pl. az $xd0$ és $xd2$ és $x1$ paramétereknél meg egyezett valamelyik metaparaméter (normál eloszlás kvantilisfüggvénye metafüggvény esetén ilyen lehet a szórás, mely egy dimenzió mentén nagyjából állandó). Különböző helyettesítési beállításokat használtunk: nqs értékének 100-at és 20-at is használtunk, és 0 és 0,5 között különböző *cutoff* értékekkel próbáltuk ki az illesztéseket. A dupla logisztikus függvény általános változatát (mely egyszerre írja le a pozitív és negatív függvényt) és egy olyan változatát is használtuk, melyben nem szükséges megadni az $m2$ paramétert, ennek értékét a többi paraméterből számítjuk.

A hálós illesztés folyamata általánosságban így nézett ki:

1. Helyettesítés.
2. Induló becslés számítása az egyes kvantilisekre külön-külön.

3. Az induló becslést finomítjuk kvantilisenkénti függvény-illesztéssel (NM algoritmus).
4. A hálós illesztéshez induló metaparaméterek meghatározása pl. lineáris regresszióval.
5. Hálós illesztés, minden paramétert valahány újabb metaparaméterrel becsülünk, melyek között lehetnek átfedések is.
6. A hálós illesztéssel kapott paraméterek kvantilisfüggéseinek és a kvantilisenkénti illesztéssel kapott paramétereknek az összehasonlítása.

4.6. Összehasonlítás

Különböző méréscsoportok összehasonlítására a medián módszert (4.5.1. alfejezet) használva a méréscsoportokat különböző szempontok szerint (az illesztett függvény paramétereinek megfelelően) hasonlíthatjuk össze klasszikus statisztikai módszerekkel (t-teszt, Mann-Whitney U-teszt, Kruskal-Wallis teszt stb). Ha pl. pozitív logisztikus függvényt illesztettünk 4.3.1. a szempontok, melyek szerint a méréscsoportok összehasonlíthatók: kezdő érték, végső érték, 50%-os érték elérési ideje, meredekség az 50%-os értéknél.

Ha kvantilis módszert használunk (4.5.3. alfejezet), egy mérésből n_{qs} db értéket kapunk minden paraméterre. Ilyenkor egy ilyen eloszlásból általános dekvantálás (4.1.2. alfejezet) után a PB módszerrel (2.2.4.2. alfejezet) egy T-értéket számolunk valamilyen kontroll csoport segítségével, ezek a T értékek pedig már nemcsak az eloszlás közepét jellemzik, mint a kvantilis módszernél, hanem az egész eloszlást. A T értékeket az előbb leírt statisztikai tesztekkel különböző méréscsoportok között összehasonlíthatjuk.

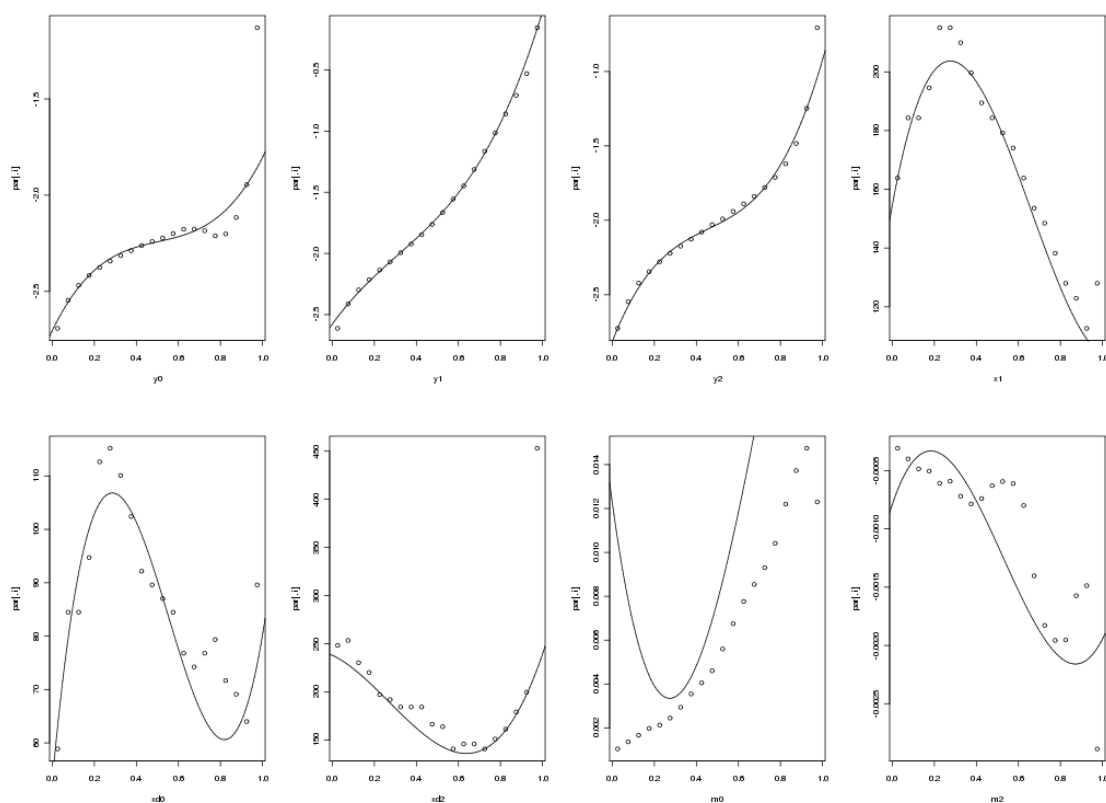
A bootstrap illesztésre nem dolgoztunk ki összehasonlító módszert, míg hálós illesztés esetén a metaparaméterek hasonlíthatók össze a fent leírt módokon.

4.7. Mérések

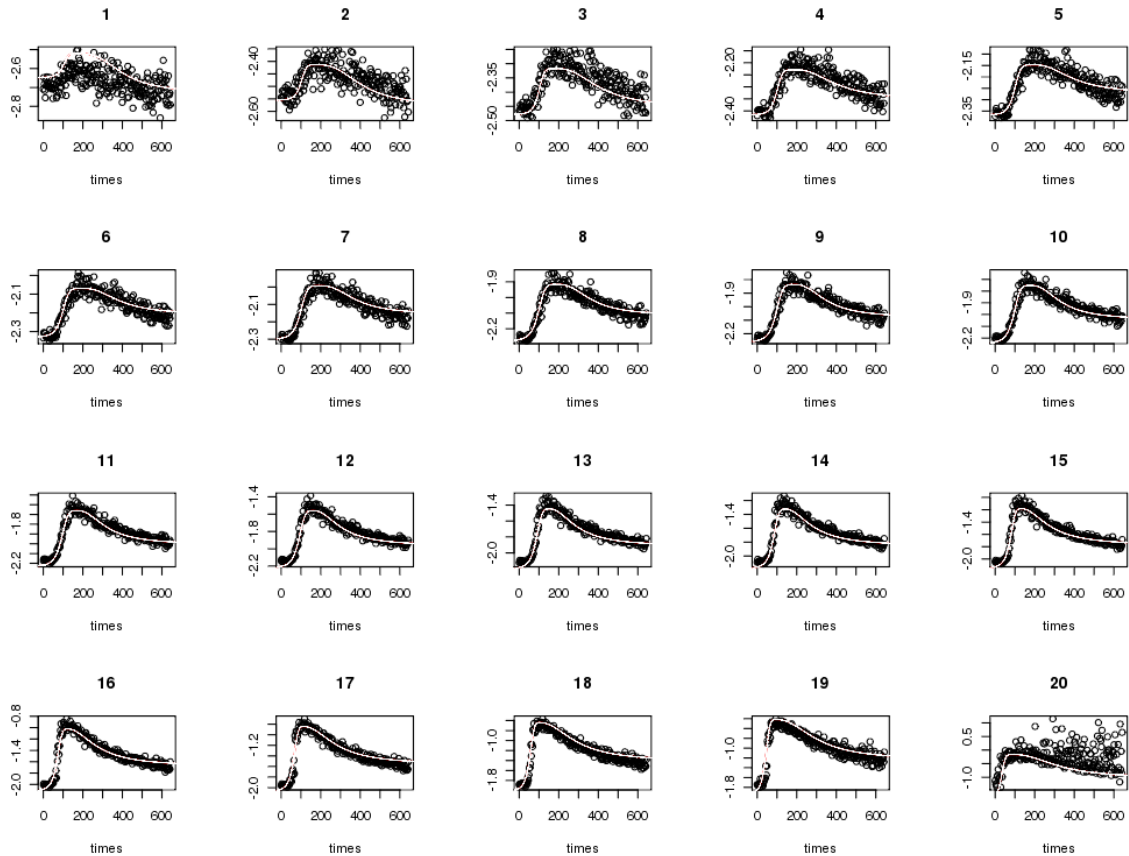
A fenti algoritmusokat elsősorban az alábbi, kifejezetten az értékelési eljárás kifejlesztésének elősegítésére végzett méréseken próbáltuk ki.

4.7.1. 1-5-10-25 sorozat

A PBMC sejteket standard sűrűség gradiens centrifugálással (Ficoll Paque, Amersham Biosciences AB, Uppsala, Sweden, 25 perc, 400 g, 22°C) választottuk el 10 ml frissen,



17. ábra. Hálós illesztés. A mérés időtartama $nts = 100$ egyenlő hosszúságú időintervallumra lett osztva, és minden intervallumban meghatároztunk $nqs = 20$ db kvantilist a 4.2. alfejezetben megadott módszer szerint. A dupla logisztikus függvényt illesztettük az egyes kvantilisokra, a paraméterek értékeit a kvantilistól függően üres körökkel ábráztuk. Az x tengelyeken a kvantilis, az y tengelyeken a paraméter értéke szerepel. Ezenkívül az egyes paraméterek kvantilisfüggéseit harmadfokú polinomokkal (4 paraméter) leírva, mivel a függvény 8-paraméteres, egy összesen $4 * 8 = 32$ -paraméteres modellt is illesztettünk a mérésre Nelder-Mead (NM, [92]) algoritmussal. Az illesztett polinomokat fekete görbékkel ábrázoluk. A kalcium-flux mérés a 4.7.2. alfejezetben bemutatott mérések közül az, melynél $25 \mu\text{g/ml}$ PHA-t használtunk.



18. ábra. Hálós illesztés, a 17. ábra paramétereinek megfelelő függvények. Az ábrák az egyes időintervallumokon belüli kvantilis adatokat üres körökkel mutatják. Fehér görbével vannak megjelenítve a 32-paraméteres illesztés adott kvantilisnek megfelelő görbéi. A kvantiliseket 1-től 20-ig sorszámoztuk az ábrák tetején.

Lítium heparinos csőbe (BD Pharmingen, San Jose, CA, USA) vett perifériás vénás vérből. A vért egészséges önkéntesektől vettük. A sejtszuspenziót kétszer mostuk PBS-ben (Phosphate Buffer Saline), majd a sejteket 30 percig 30°C-on 4 µg/ml Fluo-3 AM és 10 µg/ml Fura Red AM (Invitrogen, Carlsbad, CA, USA) keverékével inkubáltuk sötétben. Mosás után APC-anti-CD4 konjugált monoklonális antitesttel (BD Pharmingen, San Jose, CA) jelöltük a sejteket a gyártó javaslatának megfelelően. Mosás után mérésig a sejteket szobahőmérsékleten (21°C), sötétben tartottuk. A mérést egy 488 és egy 633 nm-es lézerrel felszerelt BD FACSAria áramlási citométerrel (BD Biosciences, San Jose, CA, USA) végeztük. A Fluo-3 és a Fura Red jeleket egy 610 nm-es alul áteresztő dikroikus tükörrel választottuk el, a Fluo-3 jelét 530/30 nm-es, a Fura Red jelét 695/40 nm-es sávszűrővel detektáltuk.

Először egy alapvonalat vettünk fel 30 s-ig, majd 1, 5, 10 ill. 25 µg/ml PHA-t (Sigma, Bonnem, Belgium) adtunk az egyes mintákhoz.

Ezután 10 percen keresztül mértük a kinetikus folyamatot. A sejtek közül SSC-FSC paraméterek alapján kézi kapuzással kiválasztottuk a limfocitákat, és az APC paraméter alapján a CD4+ sejteket.

4.7.2. 5-20-50 sorozat

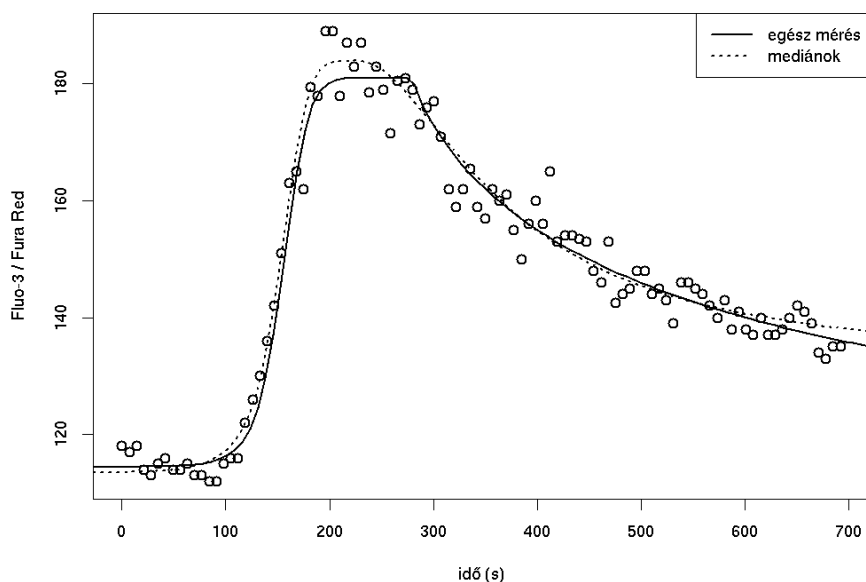
A mérési körülmények és az előkészítés megegyezett az 1-5-10-25 sorozatával, csak 120 s-os alapvonalat vettünk fel, és a PHA stimuláció is más volt: 4 db mérés 5 µg/ml PHA stimulációval történt, 4 db mérés 20 µg/ml és további 4 db mérés 50 µg/ml-el. A vért hasonlóképpen egy egészséges önkéntes adta a méréshez.

5. fejezet

Eredmények

5.1. Helyettesítés

A 4.2. alfejezetben leírt helyettesítés elvégzését hatékonysági szempontok miatt mindenképp szükségesnek láttuk, hiszen pl. a 4.7.2. alfejezetben leírt $25 \mu\text{g/ml}$ PHA-val stimulált mérésre a $dlogist_+$ függvényt egy az egyben illeszteni 153 percig tartott, míg a medián értékekre mindez 30 másodpercet vett igénybe ugyanazon a számítógépen. A két függvény nagyon hasonlóan illeszkedik (19. ábra), a különbség az egyenletesítő hatás következménye lehet.



19. ábra. Dupla logisztikus függvény illesztése ugyanarra a mérésre (4.7.2. alfejezet, $25 \mu\text{g/ml}$ PHA) egyszerre az egész mérésre illetve (folytonos görbe), és csak a 100 egyenlő hosszúságú időintervallum mediánjaira (szaggatott görbe).

FCM mérések során előfordul, hogy a sejtek száma (az idő paraméter eloszlása) nem egyenletes, ill. hogy szünetek vannak a mérésben (pl. mert épp a PHA hozzáadása történik a sejtekhez, vagy valamilyen mérési hiba történt). Ezen, kisebb sejtszámmal rendelkező mérési szakaszok kevésbé tudnák alakítani az illesztett függvények alakját, ha nem történne egyenletesítés.

5.2. Függvény-illesztés

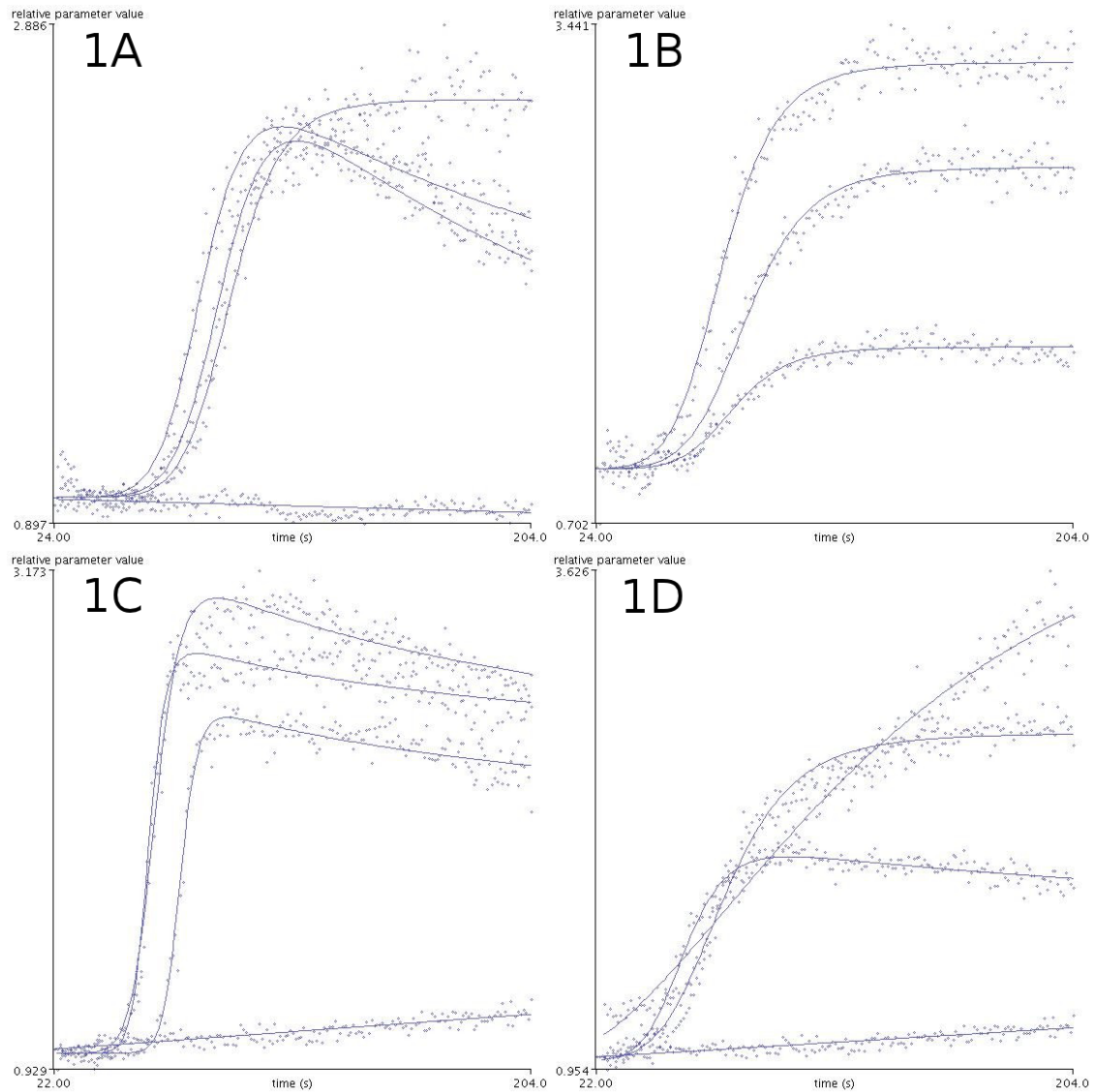
5.2.1. Függvények

Minden függvény megfeleltethető egy kinetikai modellnek, melyet megpróbálhatunk a mérés leírására használni. Ezen kinetikai modellek megfelelőségét azzal tudjuk igazolni, ha sok mérésre illeszkednek. Pl. abban reménykedtünk, hogy a pozitív logisztikus függvény (4.3.1. alfejezet) megfelelően leírja azokat a méréseket, melyek medián értékei egy konstans értékről kezdődnek, majd folyamatosan növekednek, és nincs leszálló águk.

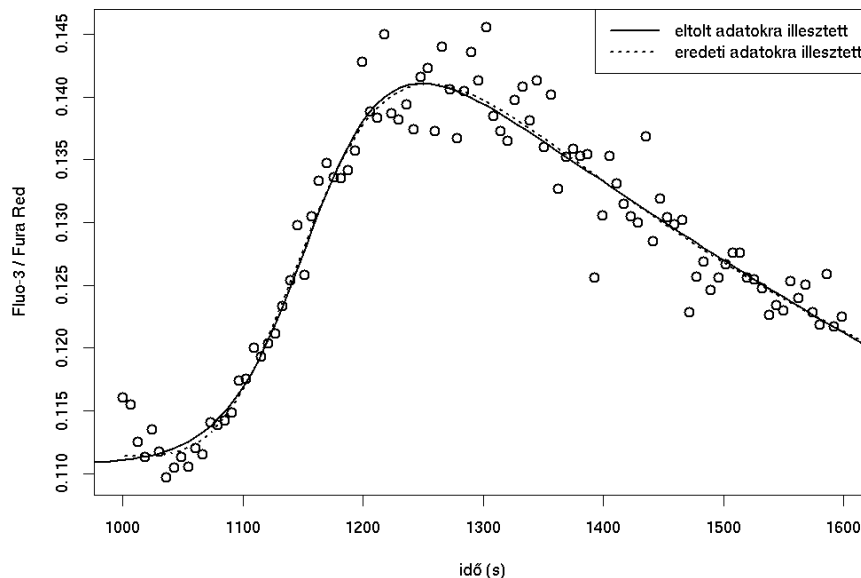
A lineáris ($y = a * x + b$), logisztikus vagy hormesis függvény jól illeszkedett a 4.7.2. alfejezetben megadott mérések mediánjaira, valamint Bailey és Macardle [16] méréseire [100]. Utóbbiakra néhány példa a 20. ábra.

A hormesis függvény hátránya, hogy vízszintesen nem tolható el, vagyis nem ugyanúgy illeszkedik két mérésre, melyek csak abban különböznek, hogy az egyik összes időpontjához hozzá lett adva egy szám (21. ábra), ez a jelenség viszont gyakori FCM méréseknél. Ennek lehet az is a következménye, hogy a függvény leszálló ága tapasztalataink szerint nem tud megfelelő meredekségű lenni, emiatt bizonyos mérésekre nem illeszthető. Másik hátránya, hogy paraméterei nehezen becsülhetők. Ugyanez elmondható a hormesis₁ függvényről is. Előnye, hogy van olyan paramétere (f), melyet 0-nak választva átmegy egyszerűbb függvénybe. Ezt az előnyt azonban nem használjuk ki, lásd 5.2.3. alfejezet.

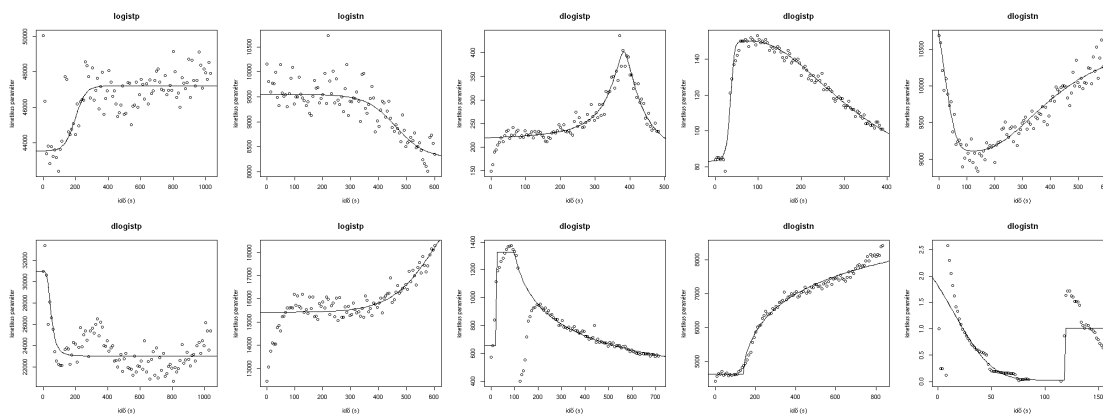
Az 5 függvényből álló készlet előnye, hogy paraméterei nagyon könnyen becsülhetők, az idő irányú eltolás az x_1 paraméter módosításával, az y irányú eltolás pedig az y_0 , y_1 , y_2 paraméterek módosításával érhető el. Nagyszámú kinetikus FCM mérés mediánjaira illesztés alapján úgy találtuk, hogy a (pozitív/negatív) logisztikus függvény tetszőleges olyan mérést le tud írni, melyben egy olyan emelkedő/ereszkedő folyamat van, melynek maximum egy inflexiós pontja van, míg a dupla logisztikus függvény tetszőleges egy „huplit” tartalmazó mérést le tud írni. Néhány reprezentatív mérésre illesztett függvényt mutat a 22. ábra, melyen az is látszik, milyen adatokra nincs megfelelő modellünk (függvényünk).



20. ábra. Bailey és Macardle cikkében [16] publikált néhány reprezentatív mérésre medián értékeire illesztett egyenes, logisztikus ill. hormesis függvények képe (a mediánok pontokkal vannak ábrázolva). A mérésekben CD31+ Jurkat sejteket stimuláltak anti-CD3-mal (1A, 1C) vagy thapsigarginnal (1B, 1D), a kalcium mennyiséget Fluo-3 / Fura Red aránnyal (1A, 1B) ill. Indo-1 festékekkel (1C, 1D) követték. Az ábra megfelel a cikkben szereplő 1. ábrának.



21. ábra. Hormesis függvény illesztése ugyanarra a mérésre (4.7.2. alfejezet, $10 \mu\text{g/ml}$ PHA) kétszer, de egyik esetben 1000 s-ot hozzáadtunk minden idő-értékhez. Az illesztés NM módszerrel SAD minimalizálással történt. Az eltolt adatokra való illesztés SAD értéke 0,1752, az eredeti adatokéra illesztésé 0,1693. Az eltolt adatokat, az ezekre illesztett függvényt (folytonos görbe) és az eredeti adatokra illesztett görbét 1000 s-al eltolva (szaggatott görbe) ábrázoltuk.



22. ábra. 5 „jó” (első sor) és 5 „rossz” illesztés (második sor) az 5 függvényből álló készlettel. Az illesztett függvény neve az egyes ábrák fölött szerepel. A „rossz” illesztések esetén látszik, hogy olyan kinetikát követnek a mérések, melyeket az általunk használt függvények nem tudnak leírni. Ilyen esetekben érdemes a mérésnek egy olyan részére illeszteni, melyet le tudnak írni (pl. maximum egy „hupli” van benne). A mérések mind kinetikus áramlási citometriás mérések, medián helyettesítést ($nqs = 1$ és $nts = 100$, lásd 4.2. alfejezet) végeztünk, és a kapott adatokra illesztettük a függvényeket NM ill. SANN módszerrel (4.3.2. alfejezet). A CV érték alapján választottuk ki a legjobb függvényt (4.3.3. alfejezet).

5.2.2. Illesztési módszerek

Az illesztő algoritmusokat medián helyettesítéssel ($nqs = 1$ és $nts = 100$, lásd 4.2. alfejezet) kapott adatokra hasonlítottuk össze. A mediánok szórása már jóval kisebb, mint az eredeti mérési adatoké, emiatt az SSD-t minimalizáló algoritmusok is szóba jöhetnek, de, mivel kiugró értékek még a mediánok között is előfordulnak (lásd 22. ábra), a robosztusabb SAD-minimalizáló eljárásokat választottuk. Úgy találtuk, hogy a Newton-módszer és a BFGS módszer ritkábban konvergál a tesztadatainkra, mint az R alapértelmezett függvényei között megtalálható *optim* eljárás által implementált SANN és NM algoritmusok. Az illesztés jósága nagyon függött az induló paramétereiktől, emiatt, hogy az illesztés felügyelet nélkül, automatikusan végrehajtható legyen, legjobbnak azt találtuk, ha több helyről (többféle paraméter-értékről) indítjuk el az illesztést. A következő módszert találtuk legjobbnak az induló becslésekre:

- A lowess módszer különböző simítási fokozataival végezzük az induló becslést ($f = 0,05, 0,1, 0,15, 0,2, 0,35$ simítási fokozatokkal).
- Az induló meredekség paramétereiket különböző értékekkel szorozzuk (0,01, 0,1, 1, 10, 100).
- A kétféle algoritmust felváltva is használjuk.
- Minden újabb illesztési lépés esetén, ha az előző lépés kimeneti SAD-ja kisebb volt, mint az aktuális paraméterbecslés, az előző lépés kimeneti paramétereit használjuk az újabb lépés induló paramétereiként.

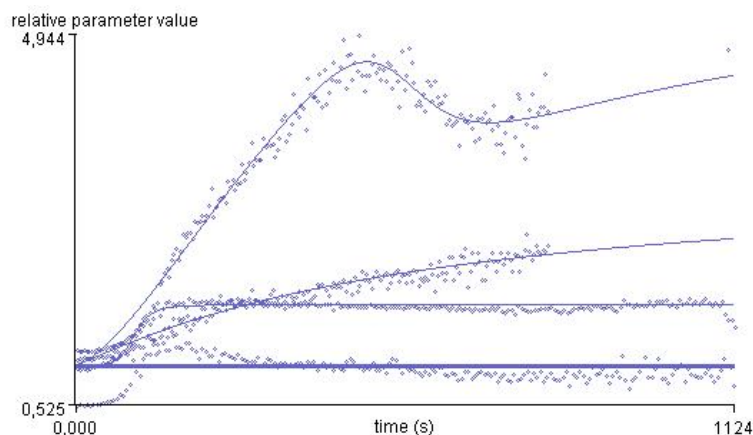
5.2.3. Illesztés jóságának becslése

A 10-szeres CV-t találtuk megfelelőnek erre a célra. Az F-teszt azért sem alkalmazható, mert nem négyzetes eltérésekkel dolgozunk, így nem szórást számolunk (lásd előző pont), és a reziduálisok gyakran nem normál eloszlásúak.

Mivel CV-vel állapítjuk meg az illesztés jóságát, nem okoz problémát, hogy az 5 függvényből álló modellben nincs olyan paraméter, melyet 0-nak választva átmenne a függvény egy egyszerűbb modellbe, mert paraméterenkénti tesztek helyett egyszerre, egy CV értékkel jellemezzük az illesztés jóságát.

5.3. Standardizálás

Ha az illesztett függvény kezdő paraméterével standardizálunk, az illesztés jóságától függ, hogy a mérés kezdetén mekkora értékeket találunk. A 23. ábra mutatja, hogy ilyenkor



23. ábra. 4 kinetikus mérésre illesztett függvény a függvények kezdő paramétereit alapján standardizálva. A legalsó, vastag egyenes függvény nem illeszkedett rá az adatokra, emiatt az adatok a standardizálás következtében lejjebb csúsztak.

nem a mért adatok, hanem a függvény alapján történik a standardizálás, így rossz illesztés esetén a mért adatok másik értékről kezdődnek.

Emiatt a mérés első szakasza alapján standardizálást látjuk helyesnek.

5.4. Paraméterek eloszlásának generálása

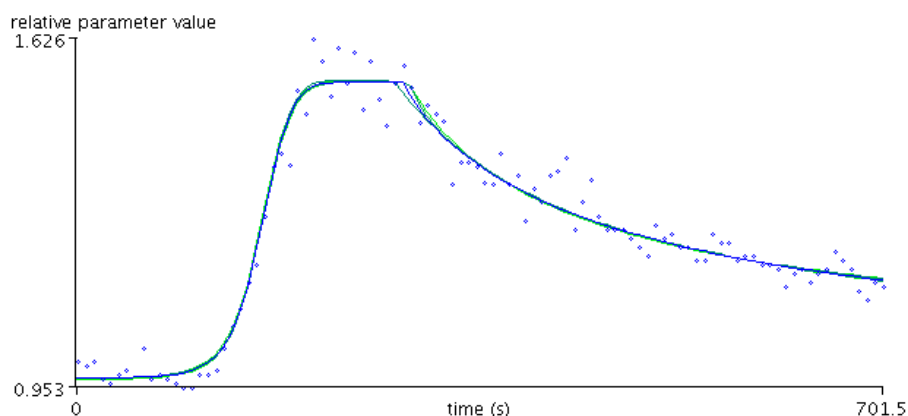
5.4.1. Medián módszer

A medián módszer megfelel a 2.2.4.3. alfejezetben leírt, sokak által használt simító módszerek egy továbbfejlesztett változatának, mely (a célkitűzésekben megfogalmazottaknak megfelelően):

1. objektív numerikus értékeket számít a mérésekből, mégpedig az illesztett függvény paramétereit, melyeket aztán statisztikai összehasonlításokra használni lehet. Különböző mérések összehasonlításakor ugyanazt a modellt (függvényt) kell használni az összehasonlításhoz.
2. a mérési adatokon kívül más bemenő adattól nem függ, és reprodukálható eredményeket ad: 40-szer ugyanazt a mérést (4.7. alfejezet, 25 $\mu\text{g}/\text{ml}$) kiértékelve a pozitív dupla logisztikus függvény paramétereit 36 esetben 3 tizedesjegyig mind megegyeztek, 4 esetben pedig a mérés alapján megjósolható paraméterekben alig különböztek, míg néhány olyan paraméterben, melyet a mérés idejének rövidsége miatt nem lehet pontosan megállapítani (a második 50%-os értékre vonatkozó paraméterek), nagyobb, akár 70%-os eltérés is volt tapasztalható (24. ábra és 3. táblázat). Az, hogy

az algoritmus nem pontosan ugyanoda illeszt minden esetben, valószínűleg az NM vagy SANN algoritmusok R-beli implementációjának köszönhető.

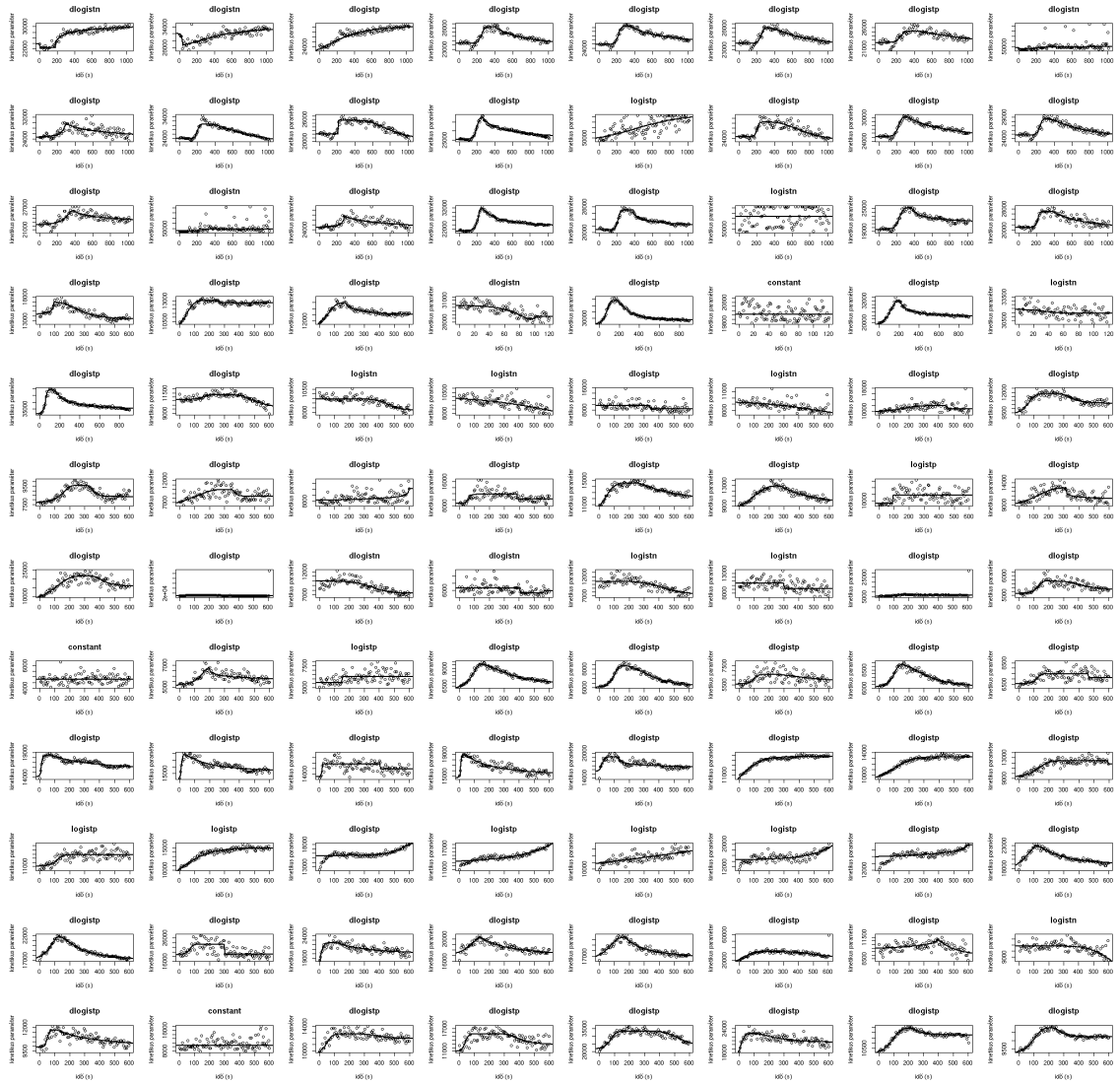
3. megadja a CV érték alapján, hogy melyik a legjobb modell a mérésre, és a modellek megfelelő függvény paramétereivel a kinetika különböző aspektusait képes jellemezni.
4. megfelelő a kalcium-flux mérések és egyéb kinetikus mérések leírására: 1846 kinetikus mérésre (lásd következő pont), melyek nagy része kalcium-flux, illesztettük az 5 függvényből álló sorozatot, és CV módszerrel meghatároztuk a legjobban illeszkedő függvényt. Szemrevételezéssel megállapítva ezek közül 2 esetben (0,07 %) a függvény nem illeszkedett jó helyre, tehát az illesztő algoritmus nem találta meg a minimumot, 68 esetben (3,68 %) pedig a mérések olyan kinetikát mutattak, melyet az 5 függvényünk közül egyik sem képes leírni (22. ábra, második sor). Az összes illesztés közül 96 taláalomra kiválasztottat mutat a 25. ábra. Mitochondriális kalciumszint, ROS termelés és plazmamembrán-potenciál mérésekre is alkalmazható volt az eljárásunk, lásd [20].
5. A medián módszert (a kvantilis módszerrel együtt, lásd lejjebb) egy felhasználóbarát számítógépes programban, a FacsKin-ben implementáltuk, működését az 5.5. alfejezetben foglaltuk össze. A programot a kutatók számára az interneten keresztül elérhetővé tettük a <http://www.facskin.com> címen. A honlapon 76 kutató regisztrált, és 2011. január 27. és 2012. szeptember 26. között 1846 mérést értékelték ki a módszerrel. Toldi és mtsai a medián módszer használatával vizsgálták limfocitákban a kalcium-flux változásait újszülöttekben [101], 1-es típusú cukorbetegségben [102], preeclampszában [103], sclerosis multiplexben [104] és rheumatoid arthritisben [105]. Egészségesekben a Th1 és Th2 limfocita alcsoportok kalciumkinetikájának különbségeire és azok okaira világítottak rá Toldi és mtsai [106].



24. ábra. A pozitív dupla logisztikus függvény 40-szer illesztve ugyanarra a mérésre (4.7. alfejezet, 25 $\mu\text{g/ml}$) a medián módszerrel. Az illesztett függvényeket ábrázoltuk (36 esetben teljesen egybeestek, 4 esetben kicsit eltértek).

3. táblázat. A pozitív dupla logisztikus függvény 40-szer illesztve ugyanarra a mérésre (4.7. alfejezet, 25 $\mu\text{g/ml}$) a medián módszerrel. A táblázat az illesztések eredményeit tartalmazza.

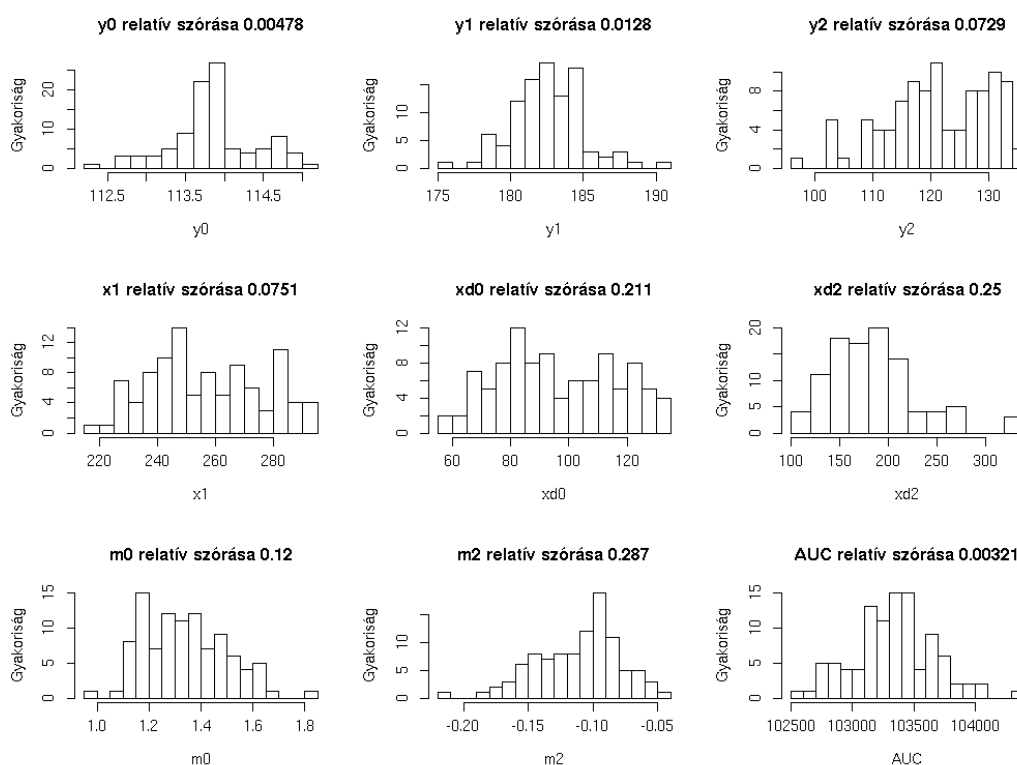
	36 illesztés	1 illesztés	1 illesztés	1 illesztés	1 illesztés
kezdő érték	113,924	113,939	113,667	113,9	113,922
maximum érték	181,014	180,995	181,078	181,03	181,51
végő érték	112,587	106,422	119,942	104,949	119,978
maximum érték elérésének ideje	284,489	289,477	288,274	289,546	274,688
első 50%-os érték és a maximum közötti idő	124,337	129,317	129,042	129,543	114,303
maximum és második 50%-os érték közötti idő	198,615	241,239	147,747	253,108	163,042
meredekség az első 50%-os értéknél	1,238	1,243	1,251	1,182	1,256
meredekség a második 50%-os értéknél	-0,075	-0,061	-0,098	-0,058	-0,099
AUC	102985	103022	103079	103038	103067



25. ábra. 96 függvény-illesztés laboratóriumunkban végzett calcium flux mérésekre medián módszerrel. Minden kis ábrán a legkisebb CV értékű függvényt ábrázoltuk, nevét az ábra fölé írtuk.

5.4.2. Medián eloszlása

A medián függvény paramétereinek eloszlásait FCM mérések esetén olyan keskenyek találtuk, ami összehasonlítások végzésére nem alkalmas. Arra azonban alkalmas lehet, hogy megjósolja, melyek azok a paraméterek, melyeknek nagy a szórása, tehát értékük akár egészen más is lehetne anélkül, hogy az illesztés jósága (SAD) nagyon megnőne. Ez azt jelenti, hogy azoknak a paramétereknek az értékét csak pontatlanul tudjuk a mérési adatokból megjósolni. A 24. ábrán látható illesztés paramétereinek eloszlását meghatároztuk bootstrap módszerrel. Ahogy a 26. ábráról leolvasható, a mérés időtartama alapján nem egészen megjósolható paraméterek ($xd2$, $m2$) szórása nagyobb, mert azok többféle értékkel is illeszkedhetnek ugyanazokra a mérési adatokra úgy, hogy közben a SAD érték alacsony marad. Az $xd0$ és $m1$ paraméterek szórása is nagy, ez azt is jelzi, hogy a paraméterek nem teljesen függetlenek, könnyen illeszkedik ugyanoda a függvény kicsit különböző paraméterekkel, és az ebből a szempontból leginkább összefüggő paraméterek ezek.



26. ábra. 100 visszatevéses mintavétel a 4.7.2. alfejezetben leírt $25 \mu\text{g/ml}$ mérésből, majd $nts = 100$ és $nqs = 1$ helyettesítés után pozitív dupla logisztikus függvény illesztése. Az illesztett paraméterek hisztogramjai fölött az adott paraméter relatív szórását (variációs tényezőjét) is feltüntettük. Jól látszik, hogy legjobban a második 50%-os értékkel kapcsolatos paraméterek ($xd2$, $m2$) szórnak, aztán az első 50%-os értékkel kapcsolatos paraméterek ($xd0$, $m0$), legstabilabb pedig az AUC .

5.4.3. Kvantilis módszer

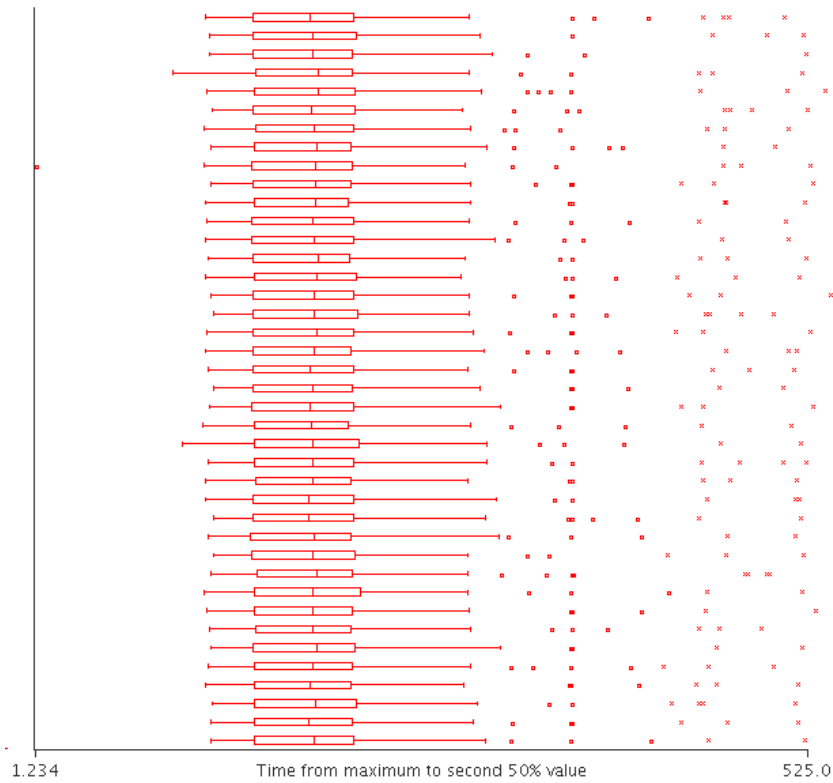
A kvantilis módszer a mérés teljes terjedelméről szolgáltat információt, és a paramétereknek jóval szélesebb eloszlásokat kapjuk, mint az előző módszernél. Úgy találtuk, hogy $cutoff = 0$, $nts = 100$ és $nqs = 201$ beállításokkal, vagyis összesen 20100 adattal történő helyettesítés esetén a kvantilis módszer már annyi adatot szolgáltat, melyből egy eloszlás könnyen kirajzolható, de még nem tart túl hosszú ideig az illesztés¹. A levágást nem láttuk szükségesnek, mert a mérések szélei is beletartoznak a teljes képbe, és az itt levő kiugró (outlier) paraméter-értékek csak kismértékben módosították a kimenetet. A pontos 201 értékre azért esett a választásunk, mert így a mediánnak és a kvartiliseknek is pontosan megfelel egy kvantilis (a 101. ill. az 51. és 151.).

A kvantilis módszert a célkitűzések szempontjából megvizsgálva:

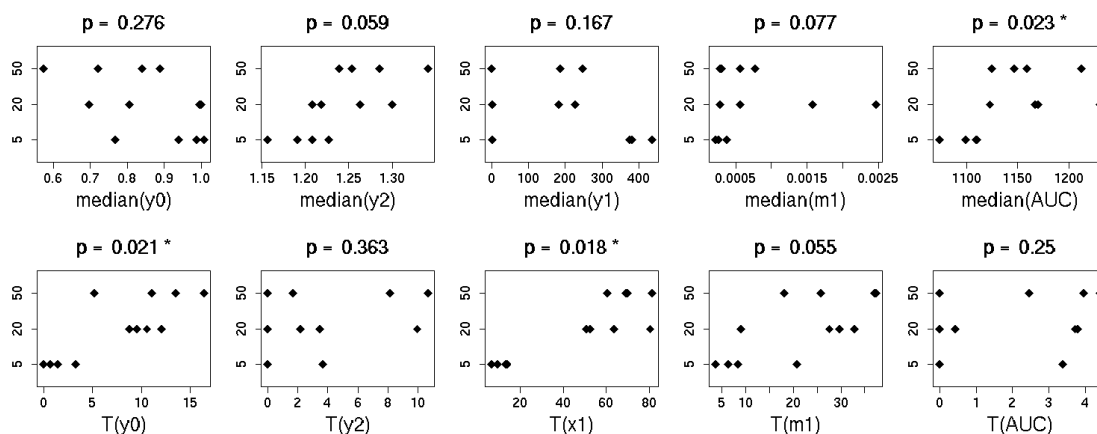
1. Az eloszlásokból számított T értékekkel olyan numerikus eredményeket ad, melyekkel statisztikai összehasonlítások végezhetők.
2. reprodukálhatóság: 40-szer ugyanarra a mérésre illesztve kvantilis módszerrel, és az összes eloszlásból az egyik randomhoz képest T értéket számolva a T-értékre minden értékelés mindegyik paraméterére 0-t kaptunk, a legdiverzebb paraméter eloszlását mutatja a 27. ábra.
3. A kinetika különböző aspektusait a medián módszer kiterjesztéseként 201 értékből álló eloszlásokkal adja meg.
4. A függvények illeszkedését a medián módszernél bemutattuk.
5. A FacsKin program tkp. a kvantilis módszert implementálja, és ezzel automatikusan a medián módszert is, mert a 101. kvantilis a medián. A program felkínálja a választási lehetőséget, hogy csak a medián függvényt használjuk, vagy az összes kvantilist. Egy mérés-sorozat értékelése mindkét módszerrel és a paraméterek láthatók a 28. ábrán. A kétféle módszerrel mások a szignifikáns paraméterek, az AUC paraméterre vonatkozóan ennek magyarázata látható a 29. ábrán. A kvantilis módszer némely paraméter (példánkban az *AUC*) esetén kitágítja az eloszlásokat a teljes mért tartomány figyelembe vételével, és esetleg megszünteti a medián módszernél szignifikáns eltéréseket (persze nagyobb mérésszámmal lehet, hogy kimutatható a különbség). Más paraméterek (példánkban az y_0 és x_1) esetén az eloszlásból származó plusz információval szignifikánssá tehet a medián módszerrel nem szignifikáns eltéréseket. A módszer hormesis függvénnyel működő változatát [100] és az

¹Az 5.5. alfejezetben leírt R implementációval az $5 * 201$ illesztés az esetek 50%-ában 4 óra alatt végrehajtható egy Intel(R) Core(TM) i7 950 3,07 GHz-es processzorral (1055 mérés kiértékelésének ideje alapján; az $5 * 201$ -ben az 5 az 5-féle függvénynek felel meg).

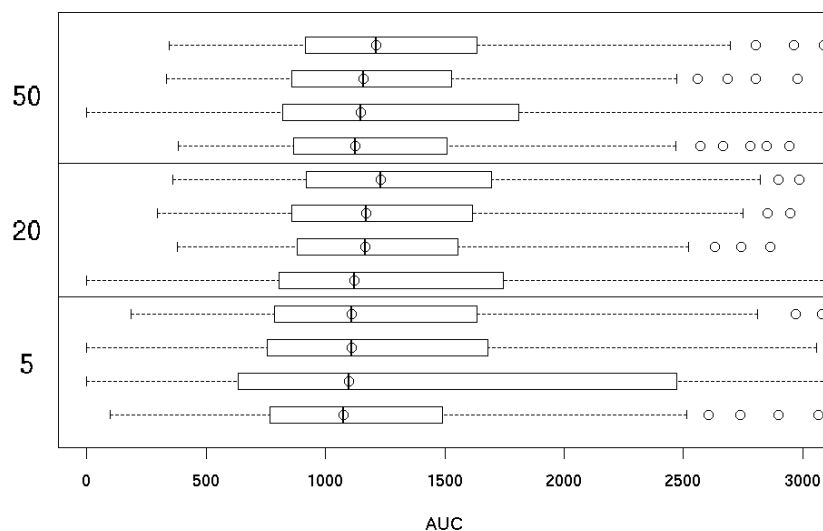
5 függvényből álló készletet használó, FacsKin programban megvalósított változatát [74] publikáltuk. Kutatásban új biológiai megállapításokra a kvantilis módszert egyelőre nem használták, a kvantilisokra illesztett függvényekből számolt eloszlások biológiai értelmezése még várat magára.



27. ábra. A pozitív dupla logisztikus függvény 40-szer illesztve ugyanarra a mérésre (4.7. alfejezet, 25 $\mu\text{g/ml}$) a kvantilis módszerrel. A maximum és a második 50%-os érték paraméter eloszlásait ábrázoltuk box plottal (mindegyik 201 értékből áll), ez volt a legdiverzebb paraméter az illesztések során.



28. ábra. A 4.7.2. alfejezetben leírt mérések értékelése medián és kvantilis módszerrel, pozitív logisztikus függvényt illesztve. Az első sorban a medián függvény paraméterei, a második sorban a kvantilisenkénti illesztésekből kapott paramétereloszlásokból a 05-ös csoportot kontrollul véve PB módszerrel számított T értékek vannak megjelenítve. Minden paraméterre végeztünk egy Kruskal-Wallis tesztet, ennek p-értékei vannak megadva az egyes ábrák fölött, a $p < 0,05$ értékeket *-al jelöltük



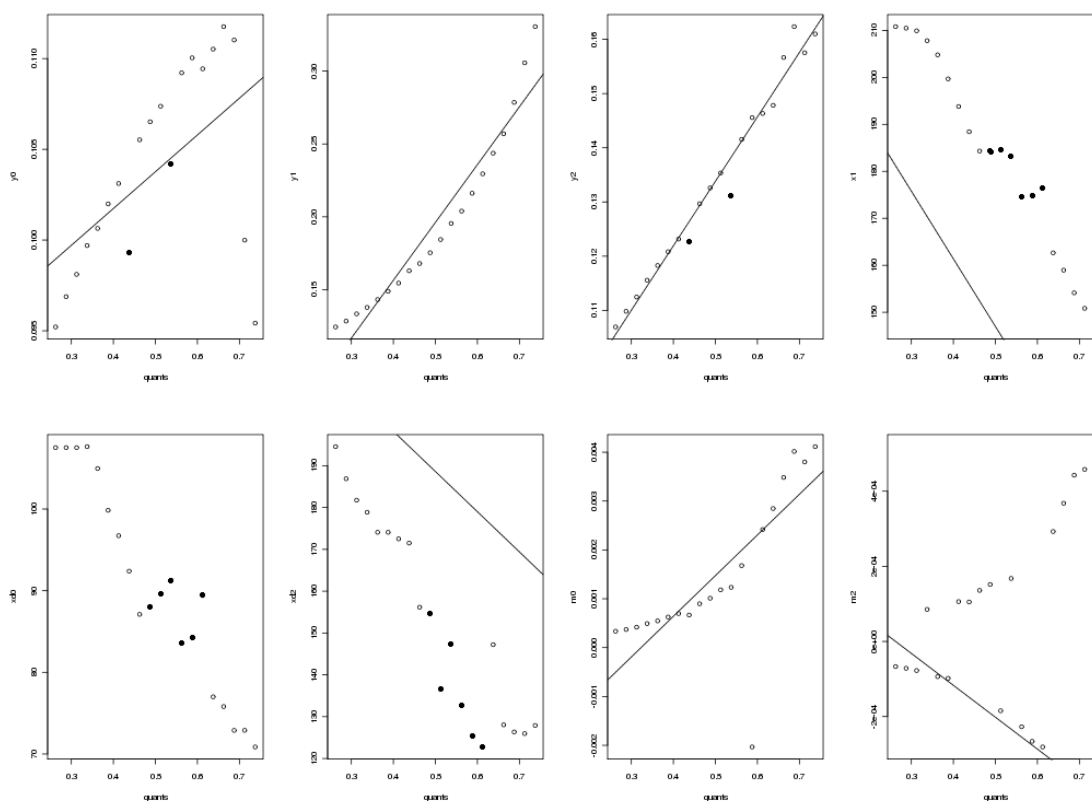
29. ábra. A 4.7.2. alfejezetben leírt mérések értékelése kvantilis módszerrel, az AUC paraméterek eloszlásainak ábrázolása box plottal. A medián függvény AUC paramétere körrel van jelölve, ez általában nagyon közel volt az eloszlás mediánjához. Látszik, hogy, bár az 5-ös csoportban a mediánok mind kisebbek a másik két csoport legkisebb értékénél is, ha az egész eloszlásokat figyelembe vesszük, ez a különbség eltűnik. Ez lehet a magyarázata, hogy a 28. ábrán látható módon medián módszerrel az AUC paraméterek szignifikánsan különböznek, míg a kvantilis módszerrel nem.

5.4.4. Hálós illesztés

A 4.5.4. alfejezetben leírt hálós illesztés azzal az előnnyel kecsegtet, hogy a paraméterek elméleti eloszlásait a kvantilis-függésen keresztül megadja, így a kvantilis-függések metaparamétereit² közvetlenül össze tudjuk hasonlítani anélkül, hogy T értékeket vagy hasonló, az eloszlások különbözőségére jellemző számokat kellene számítanunk. Néhány reprezentatív mérésre (4.7.2. alfejezet) a köbös metafüggvény jól illeszkedett, lásd pl. 18. ábra. Mivel egy hálós illesztés során sok paramétert illesztünk egyszerre, az NM eljárás a sokdimenziós térben nem tud messzire menni a kiinduló paraméterektől, ezért a kiinduló becslés jósága nagyon fontos. Ezeket egy előzetes, egyesével a kvantilisre illesztésből tudjuk megbecsülni, mely sokszor zajos, emiatt nehezen becsülhetők meg az induló paraméterek. Bár valószínűleg az általunk használt metafüggvények nem eléggé robusztusak tetszőleges mérések leírására, mégis, végül más miatt vetettük el ezt a módszert: a hálós illesztés mesterségesen csökkenti a paraméterek szórását. A probléma oka, hogy a dupla logisztikus függvény paraméterei között összefüggések vannak, melyek miatt egy függvény két vagy többféle paraméter-kombinációval is tud ugyanarra a mérésre illeszkedni. A kiinduló értéktől (ill. az illesztő algoritmus működésében levő véletlenszerűségtől, lásd 3. táblázat) függ, hogy melyik lokális eltérésösszeg-minimumot, és így melyik paraméter-kombinációt választja az algoritmus. Ha kvantilisenként külön illesztünk, néha az egyik, néha a másik típusú paraméterezés nyer, mint az a 30. ábrán látszik: az 1. (y_0) és 3. (y_2) paraméterek esetében egy szinte kvantilisenként folytonos görbét írnak le a paraméterek (üres körök), de a 8. és 12. kvantilisnél (fekete körök) egy másik paraméter-kombinációt választott az illesztés. Hasonlóképpen a 4. (x_1), 5. (xd_0) és 6. (xd_2) paraméter esetén a 10-15. kvantilisek esetén másik paraméter-kombinációt választott az algoritmus, mint az összes többi kvantilisnél (a 13. kvantilis esetén ez talán nem igaz, mert az közelebb esik a többi kvantilis-érték paraméterei által kirajzolt görbéhez). A 8. paraméter esetén látszik legjobban a két különböző típusú paraméter-kombináció, ahol szinte felváltva követnek két különböző lefutású görbét a paraméterek, melyek közül a hálós illesztés metafüggvénye, az egyenes látványosan az egyiket preferálja. A hálós illesztés több különböző paraméter-kombináció közül egy típusút választ ki, de könnyen előfordulhat, hogy különböző mérések esetén különbözőt, amivel különböző mérések összehasonlítása esetén mesterségesen sok különbséget találnánk. Ez a probléma a kvantilis módszernél kevésbé jelentkezik, hiszen a 201 kvantilis esetén néha az egyik, néha a másik típusú paraméter-kombinációt választja az illesztés, ezzel a szórást nagyobb értéken tartva. Medián módszer használatakor ez a probléma szintén jelentkezhethet, emiatt érdemes a paraméterek relatív szórását (variációs tényezőjét, lásd 26. ábra) megtekinteni, és ha ez az érték nagy, a

²vagy az ezekből számított, az eloszlás vizsgálni kívánt tulajdonságait (pl. szélesség, szórás, közép) jellemző értékeket

paramétert nem szabad összehasonlításokra felhasználni.



30. ábra. Hálós illesztés lineáris metafüggvénnyel a 4.7.2. fejezetben leírt $25 \mu\text{g/ml}$ PHA mérésre. A helyettesítés $nqs = 20$ és $cutoff = 0,5$ beállításokkal történt. A metafüggvény a dupla logisztikus függvény kvantilisfüggését adja meg. Az x tengelyen a kvantilis, az y-on az egyes paraméterek értékei láthatók. Körrel jelöltük az egyesével (kvantilis módszer) illesztett paramétereket, egyenes vonallal az egész mérésre egyszerre illesztett lineáris metafüggvényt. Az 1. (y_0) és 3. (y_2) paraméterek esetén feketével vannak megjelölve az összefüggő, e többletől eltérően illeszkedett paraméterek. A 4. (x_1), 5. (xd_0) és 6. (xd_2) paramétereknél hasonlóképpen.

5.5. A FacsKin által implementált módszer

A következő módszert implementáltuk a **FacsKin** programban Java, R és C programozási nyelveken:

1. FCS fájlok beolvasása, kézi kapuzása (egy- és kétdimenziós³), skálák állíthatók logaritmikusra, poligon-, téglalap- és inverz kapuzás).
2. Idő paraméter szerinti dekvantálás (4.1.1. alfejezet).

³Kétdimenziós sűrűség-ábrázolás (density plot), lásd 2.2.4.3. alfejezet.

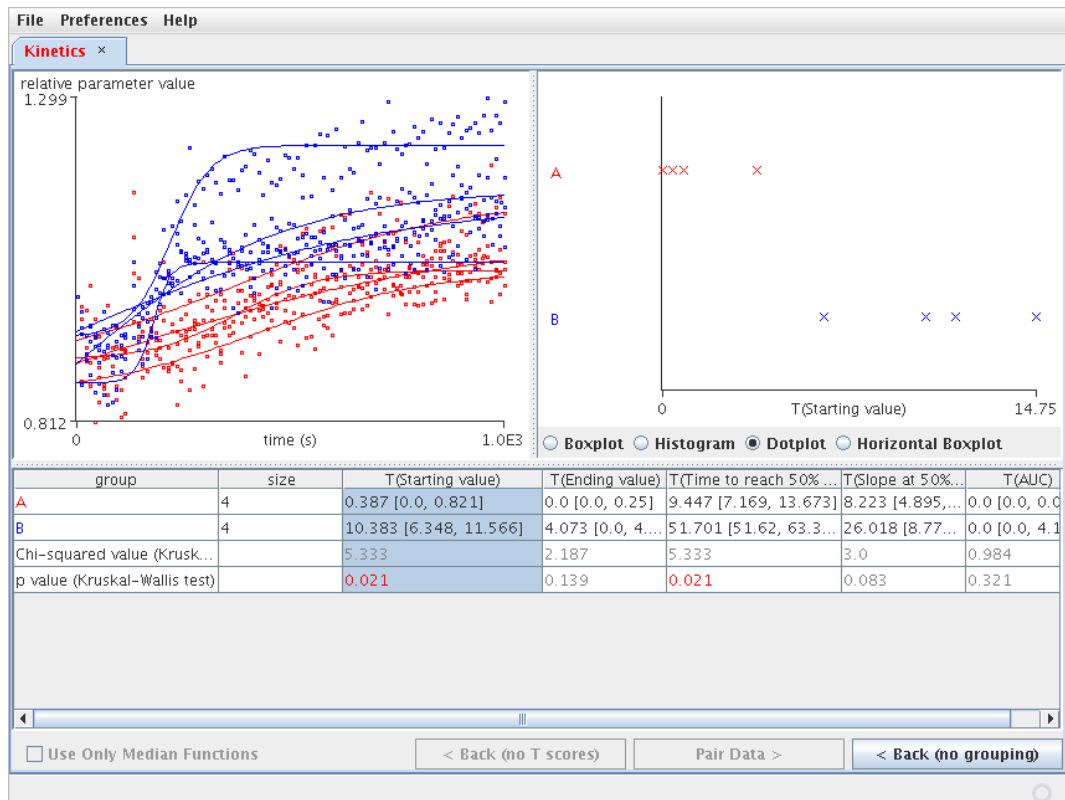
3. A mérés helyettesítése $nts = 100$ egyenlő hosszúságú időintervallumban felvett $nqs = 201$ kvantilissal $cutoff = 0$ levágással (4.2. alfejezet).
4. A kvantilis módszer (4.5.3. alfejezet) alkalmazásával a 201 db kvantilisre a konstans, pozitív és negatív logisztikus, pozitív és negatív dupla logisztikus függvények (4.3.1. alfejezet) illesztése a 4.3.2. alfejezetben megadott illesztési módszerrel: SAD minimalizálás, NM és SANN minimalizáló módszerek használata többféle induló paraméterértékről. A középső (101.) függvény a medián módszernek (4.5.1. alfejezet) megfelelő függvény lesz.
5. Az illesztések jóságának becslése 10-fold CV módszerrel (4.3.3. alfejezet). A CV értékeket a medián és a kvartilis értékekre számoljuk ki, majd összegezzük.
6. Az illesztett függvények standardizálása a mérések tetszőleges hosszú első szakaszának mediánja alapján (4.4. alfejezet).
7. A mérések tetszőleges csoportokba sorolása ill. méréspárok kijelölése.
8. Mérések jellemzőinek számítása a következő módokon:
 - Medián módszer (4.5.1. alfejezet):
 - Ha párosítottak a mérések, a paramétereket páronként kivonjuk egymásból, majd a különbségeket paraméterenként az előző pontban meghatározott csoportokba soroljuk.
 - Ha nem párosítottak a mérések, paraméterenként az előző pontban meghatározott csoportokba soroljuk őket.
 - Kvantilis módszer (4.5.3. alfejezet):
 - Ha párosítottak a mérések, a párok közül az egyiket kontrollnak véve a PB módszerrel (2.2.4.2. alfejezet) meghatározzuk a párok közötti eltéréssel arányos T-értéket, majd ezeket soroljuk csoportokba.
 - Ha nem párosítottak a mérések, paraméterenként valamelyik csoportot vagy az összes mérést együtt kontrollnak véve mindegyik mérésre meghatározzuk PB módszerrel a T értékeket, majd ezeket soroljuk csoportokba.
9. Összehasonlítás: Kruskal-Wallis teszttel [52] a csoportokat összehasonlítjuk. Két csoport esetén ez a teszt megegyezik a Mann-Whitney U-teszttel.

Az 1. és a 6-9. lépéseket felhasználóbarát Java felületen [107] implementáltuk, míg a 2-5. lépéseket R és C programozási nyelven. Utóbbiak egy dedikált nagy teljesítményű

számítógépen hajtódnak végre. A két program közötti kommunikáció hálózaton keresztül FTP protokollal történik. A 31. és a 32. ábrákon látható a FacsKin program felhasználói felülete.



31. ábra. FacsKin 0.6.4 felhasználói felülete. 8 mérés van megnyitva, ezek közül a felhasználó négyet az A, négyet a B csoportba sorolt. A $logist_+$ modell van kijelölve, a modell paraméter-eloszlásai medián [kvartilisek] formában a táblázatban vannak megjelenítve. Egy sor egy mérésnek, egy oszlop (a SAD fejlécű oszloptól jobbra) egy paraméternek felel meg. A kezdő érték (starting value) paraméter ki van jelölve és a különböző mérésekben ennek a paraméternek az eloszlása box and whisker's plottal [56] megjelenítve a jobb felső ábrán. A bal felső ábrán az egyes mérések medián értékei és az ezekre illesztett függvény látható. A sejtek humán PBMC-ből kikapuzott CD4+ limfociták Fluo-3 és Fura Red kalcium-szenzitív festékekkel megfestve [15] és 5 ill. 50 μM PHA-val stimulálva. A kinetikus paraméter érték a Fluo-3 / Fura Red arány. A mérések az alapvonalhoz mediánjára (2 perc) lettek standardizálva.



32. ábra. FacsKin 0.6.4 felhasználói felülete T értékek számítása és csoportosítása után. A mérések megegyeznek a 31. ábra méréseivel. Kontroll csoportként az A csoport lett kijelölve, az ettől való individuális eltéréseket számolta ki a program PB módszerrel, majd ezeket csoportonként összegezte paraméterenként medián [kvartilisek] formájában. A kezdő érték paraméter T értékei pont-ábrázolással meg vannak jelenítve a két csoportban. A program kiírja a Kruskal-Wallis teszt χ^2 értékét és p-értékét.

6. fejezet

Megbeszélés

Watson így ír az FCM mérések értékeléséről [108]: „We collect very large, so-called, statistically valid data sets. This is a frequently voiced attribute of the various cytometric technologies. However, we seldom use this to its full advantage.”¹ A nagyszámú mérési adat által nyújtott előnyöket gyakran nem használjuk ki. Ennek egyik példája, hogy a kézi kapuzással megállapított sejtszámlálás még mindig általánosan elterjedt értékelési módszer [45]. Erre magyarázat lehet pl. az áramlási citometriát használó biológusok idegenkedése a matematikai eljárásoktól. A trend azonban változni látszik: egyre több új fajta, akár bonyolult adatbányász módszereket használó publikáció jelenik meg, amilyen a klaszter analízis ([37], [38]) vagy a mesterséges neuronháló [72] felhasználása FCM mérések értékelésére. Az FCM mérések tárolására vonatkozó szabványok [5] fejlődése az értékelés reprodukálhatóságát segíti elő azzal, hogy a folyamat ad-hoc jellegű lépéseit legalább rögzítik olyan formában, hogy az újra ugyanúgy lejátszható legyen. Ezt néhány folyóirat, pl. a Cytometry Part A kötelezővé is teszi annak, aki FCM méréseket publikál. Az értékelés teljes folyamatának szabványosítására is történnek kísérletek [45], és a mérések kevés adattal történő prezentációjára is, mely mégis minden fontos információt tartalmaz (citometriás ujjlenyomat [67], lásd 2.2.4.2. alfejezet). A citometriás ujjlenyomat alapját képező Probability Binning (PB, [47]) eljárás FCM-alkalmazásairól az elmúlt tíz évben számos publikáció született.

A kinetikus FCM mérések (amilyen a kalcium-flux) értékelésében azonban hasonló folyamatok egyelőre nem látszanak. A fentebb leírt módszerek nem használhatók az idő paraméterrel kapcsolatban, mely egészen más jellegű, mint a többi: sejtek csoportosítására nem alkalmas, csak a kinetikus paraméterrel együtt hordoz információt. A citometriás ujjlenyomatok számolására használatos `flowFP` R csomag pl. nem is veszi figyelem-

¹„Igen nagyszámú, ún. statisztikailag érvényes adathalmazokat gyűjtünk. Ezt a különböző citometriás technikák előnyeként gyakran szokták emlegetni. Mégis, csak ritkán használjuk ki teljesen az ezáltal nyújtott előnyöket.”

be az idő paramétert az alcsoportok számításánál. Az ilyen mérések leírására a simító módszerek a legelterjedtebbek, mint amilyen az intervallumonkénti középérték-számítás, mozgóátlag, lokális regresszió, lowess [83], spline-ok stb. Ezek kimenetele számos beállítási tényezőtől függ, melyekről nincs megegyezés: simítás finomsága, intervallumok szélessége, különböző középértékek, vízszintes és függőleges standardizálás mikéntje. A paraméterek értékei, melyek egy ilyen simított görbéről leolvashatók, erősen függenek ezektől a beállításoktól. Ez megnehezíti különböző, mások által publikált mérések egymással való összevetését (2.2.4.3. alfejezet).

A simítási módszerek alkalmasak az adatok megjelenítésére és minták exploratív analízisére, hipotézisek felállítására, de nem használhatók a hipotézisek igazolására, vagyis adatelemzésre [109]. Matematikai modell használatával a kinetikus folyamatról feltett hipotéziseink explicitté és így verifikálhatóvá válnak [110]. Egy modell általában egy függvény, melynek van bemenő és kimenő értéke (ezekből több is lehet, ilyenkor pl. vektor formában adjuk meg), és különböző paraméterei. A modellt a mérési adatokra illesztve, tehát a paramétereit úgy változtatva, hogy a kimenő érték és a mért adatok közötti eltérés minimális legyen, megvizsgálható, hogy a mérés megfelel-e a modellnek ill. a modell paraméterei és azok megbízhatósága számszerűsíthetők. A megbízhatóságot felhasználva meghatározhatunk szignifikancia értékeket, melyekből statisztikai értelemben következtéseket tudunk levonni.

Egy modell lehet empirikus vagy mechanisztikus. Előbbi az adatok általános összefüggéseit tükrözi, és paraméterei nem feleltethetők meg közvetlen biológiai folyamatoknak. Utóbbi közvetlenül a biológiai működést írja le fizikai vagy kémiai egyenletekre építve, és olyan paraméterei vannak, mint disszociációs konstansok, enzimaktivitások, koncentrációk stb. A mechanisztikus modellek hátránya, hogy sokszor túl bonyolultak (nagyon sok paraméterük van), ha pedig leegyszerűsítjük őket, akkor csak nagyon szűk tartományokban működnek, kevés mérésre illeszkednek.

Az immunsejtek aktivációjakor is lezajló kalcium-szint változás leírására sokféle mechanisztikus modellt dolgoztak ki ([111], [112], [113], [114]), melyek jól illeszkednek a fluoreszcens mikroszkópiával vagy patch-clamp mérési technikával kapott adatokra: ilyenkor összesen egy sejt viselkedésének leírására van szükség. A kalcium nyugalmi koncentrációja 50-200 nM, míg aktiváció után 1 μ M-os szintig növekedhet, majd elkezd csökkenni vagy oszcillálni a nyugalmi érték és e magas koncentráció között [115]. Az oszcilláció megléte vagy nem-léte, amplitúdója és frekvenciája adja a folyamat információ-tartalmát, kódolja a sejt specifikus reakcióját [116].

Munkánk motivációja az volt, hogy a laboratóriumunkban áramlási citométerrel mért kalcium-flux mérések elemzésére nem ismertünk megfelelő eszközt. Ezekben a mérésekben milliós nagyságrendű sejt relatív kalcium-szintjét állapítjuk meg, de mindegyik sejt-

ről csak egyetlen adott idő-pillanatban kapunk információt, tehát a kalcium-görbéjének egyetlen pontját ismerjük csak. Arra, hogy előtte meg utána milyen volt abban a sejtben a görbe lefutása, csak a többi sejtben mért értékek alapján tudunk következtetni. Az oszcillációt az adatok közvetlenül nem ábrázolják, helyette azt látjuk, hogy a sejtek kalcium-szintje átlagosan emelkedik, és a szórás is megnövekszik (4. ábra). Utóbbira magyarázat lehet az oszcilláció is: ugyan megnő a maximális kalcium-szint, de sok sejt épp akkor, amikor megmérjük, egy oszcillációs periódus közepén vagy alján van, így alacsony értékeket is mérünk. Ezekből a mérési adatokból a sejtek önálló kalcium-görbéit visszakapni hasonló feladat egy Fourier-transzformációhoz, amikor az összesített időbeli lefolyás görbéjéből az azt előállító szinusz-görbék frekvencia-spektrumát határozzuk meg azzal a gondolattal, hogy az tkp. a sok különálló görbe összegzésekképpen adódik. Kalcium-flux esetében a háttérben lezajló oszcillációs görbék frekvencia- és amplitúdó-eloszlásait kellene megbecsülnünk a mérési adatok alapján. A képet tarkítják a reagáló, de nem oszcilláló-, ill. a nem-reagáló sejtek és a kalcium-szenzitív festékek sejtekben való mennyiségének különbségei.

Mivel (1) arról, hogy milyenek lehetnek a sejtek individuális görbéi, ennyi sejt esetén nagyon keveset tudunk, és (2) célunk volt, hogy módszerünk ne csak limfocitákra ill. ne csak kalcium-szint változásokra működjön, úgy döntöttünk, hogy mechanisztikus helyett empirikus modellt használunk, amely képes leírni az olyan kinetikákat, melyben maximum egy „hupli” van, tehát a kinetikus paraméter egy adott értékről indulva növekszik, elér egy maximumot, majd csökken egy újabb meghatározott értékig (vagy épp fordítva, csökkenés, majd egy minimum elérése után növekszik). Tapasztalataink alapján az FCM kalcium-flux mérések túlnyomó többségének mediánjai ilyen, vagy egyszerűbb kinetikát követnek (25. ábra). Olyan függvényeket, vagyis paraméteres modelleket kerestünk, melyek ilyen típusú kinetikát írnak le. A biológiában gyakran használt 4-paraméteres logisztikus függvény módosított változatát, az 5-paraméteres hormesis függvényt [87] vizsgáltuk meg először, majd, mivel vízszintesen nem eltolható és paraméterei nehezen becsülhetők, egy ún. dupla logisztikus függvényt dolgoztunk ki, mely tkp. két logisztikus függvény összeillesztve úgy, hogy egyik 180° -al tükrözve van. A dupla logisztikus függvénynek így 8 paramétere van, bemenő értéke az idő, kimenő értéke pedig a kinetikus paraméter becsült értéke.

A függvényt a mérési adatok mozgómedián görbéjének egzakt leírására használtuk. Átlag helyett azért használtuk a mediánt, mert a kinetikus paraméter eloszlása tapasztalataink alapján nem normális, még csak nem is log-normális a sok kiugró érték miatt, a medián pedig ezekre nem érzékeny. Ez ellentétben áll azzal, hogy az irodalomban a mozgóátlagot használják leginkább kinetikus FCM mérésekből görbe számítására ([4], [79], [80], [76], [77]). Az átlag hátránya amellet, hogy értékét a kiugró értékek nagyon

megnövelik, az, hogy nem stabil: a mért intenzitás készüléktől függő felső korlátja miatt különböző festék-koncentrációknál, különböző mérési beállítások mellett ugyanannak a mérésnek a mozgóátlaga egészen máshol lehet.

Az illesztett függvényekből olyan paraméterek olvashatók le, melyekhez biológiai jelentés társítható: maximum értéke, maximum elérési ideje (t_{max}), kezdő érték, végső érték, 50%-os értékek elérési idejének maximumtól való távolsága, meredekség ezeknél az értékeknél, AUC. Ezek a paraméterek megfeleltethetők az irodalomban a simított görbék-ből számított paramétereknek, melyeket a 2.2.4.3. alfejezetben listáztunk:

- intenzitás, meredekség adott idővel a mérés kezdete után: az ilyen típusú, szubjektív időpont-választástól függő paraméterek helyett adjuk meg az ennél több információt hordozó kezdő, végső és maximális intenzitás paramétereket. Ha mégis, biológiai szempontból érdekes lenne konkrét idővel a mérés kezdete után az (átlagos) intenzitás, ez kiszámítható az illesztett függvényből is, ráadásul fix kontroll-pontokhoz (mint pl. a maximum) képest is.
- maximális intenzitás, maximum intenzitás elérési ideje, AUC: ezek a paraméterek szerepelnek a modellünkben.
- maximum intenzitás érték 50%-ának elérési ideje: ezt közvetlenül nem adjuk meg, hanem ehelyett a maximumtól való távolságát, de egy kivonással kiszámítható.

A módszerünkkel illesztett dupla logisztikus függvény paramétereit felhasználtuk arra, hogy a Th1 és Th2 limfocita alcsoport kalcium-kinetikájának különbségeit karakterizáljuk, és azokra magyarázatokat találjunk [106]. Az AUC, meredekség az első 50%-os értéknél (slope), maximum és maximum elérési ideje (t_{max}) paramétereket használtuk erre a célra. Ruthenium Red-del, az MCU (mitokondriális kalcium uniporter) gátlószerével kezelt sejteket vizsgálva az AUC, slope és maximum értékek alapján arra lehetett következtetni, hogy a mitokondriális kalcium felvétel Th1 sejtekben a limfocita aktiváció kezdeti és csúcs szakaszában is jelen van, míg Th2 sejtekben csupán a kalcium beáramlás csúcán van szerepe. Thapsigarginnal, az endoplazmás retikulum (ER) kalcium-pumpájának (SERCA) gátlószerével kezelt mérések esetén a slope érték nagyobb mértékű növekedésből arra lehetett következtetni, hogy az ER kalcium felvétele hamarabb indul meg Th2 sejtekben, mint Th1 sejtekben.

A T-sejtek kalcium-szintjének fenntartásában fontos kálium-csatornák specifikus gátlásával az AUC, maximum és t_{max} paraméterek alapján Toldi és mtsai módszerünkkel kimutatták, hogy nagyobb mennyiségű kálium hagyja el kompenzatorikusan a sejtet a limfocita aktiváció során a Th1 sejt típusban, mint a Th2 sejt típusban [101]. Ezen csatornák gátlása akár potenciális sejt típus-szelektív terápiára adhat reményt. Ugyancsak az

AUC, maximum és t_{max} paraméterek használatával Toldi és mtsai kimutatták, hogy a limfociták kalcium kinetikája és különböző kálium csatorna-gátlószerekre való érzékenysége egészséges várandós nőkben specifikus mintázattal bír, amely hiányzik preeclampsziában [103].

A CD4+, Th1, Th2 és CD8+ limfocita alcsoportokban a Kv1.3 és IKCa1 kalciumcsatornák specifikus gátlószereinek hatását egészségesekben és IFN- β tával kezelt ill. nem kezelt sclerosis multiplex (SM) betegekben karakterisztikus kinetikus paraméterváltozásokkal lehetett leírni [104], melyekből a jövőbeli terápiás lehetőségekre tudunk következtetni. 1-es típusú diabeteses betegek esetén a limfociták a Kv1.3 csatornák gátlására érzékenyebben reagáltak, mint egészségesekben az AUC és maximum paraméterek tekintetében, míg a t_{max} értékben nem találtunk különbséget [102].

Az AUC, maximum, maximum elérési ideje és az 50%-os érték meredeksége paraméterek kutatásban való felhasználására tehát vannak példáink. Ahhoz, hogy a többi, általunk megadott paramétert is fel lehessen hasonló célokra használni, szükséges, hogy értékük jól becsülhető legyen a mérési adatokból. Ha például a második 50%-os érték a mérési időtartamon kívül esik, csak nagy bizonytalansággal állapítható meg a meredekség a görbe azon pontján, bár következtetni lehet a görbe még mérési időn belül lefelé induló szakaszának alakjából. A paraméterek bizonytalansága azok eloszlásával, és így szórásával megállapítható bootstrap módszerrel (lásd 26. ábra). A nagy relatív szórású paraméterek értéke az adott mérésből nehezen becsülhető, így azok következtetésekre nem használhatók fel. A nagy szórás a paraméterek közötti összefüggésekből is eredhet (lásd 30. ábra).

A mozgómediánt függvénnyel helyettesíteni, majd a függvény paramétereit felhasználni a mérés leírására ahhoz hasonló, mint amikor két méréscsoportot pl. FSC paraméter szerint szeretnénk összehasonlítani, és ezt úgy tesszük, hogy minden mérés medián FSC értékét kiszámoljuk, majd ezeket az értékeket használjuk a statisztikai teszt bemeneteként. A medián értékeknél azonban több információt hordoznak a teljes FSC-eloszlások, melyeket a KS-teszt módosított változatával [63] vagy PB-módszerrel [47] össze is tudunk hasonlítani. Ezt a módszert szívesen kiterjesztenénk a kinetikus paraméterekre is: ne csak a kinetikus paraméter mediánját, vagyis a medián függvény paraméterét használjuk fel összehasonlításra, hanem annak teljes eloszlását!

Egy kinetikus paraméter (amilyen a maximum, maximum elérési ideje stb.) abban különbözik az FSC paramétertől, hogy nem tudjuk értékét minden sejtre meghatározni, és ezzel az előbbi módszert alkalmazni. Megállapíthatjuk a medián értékét vagy megállapíthatjuk egy lehetséges eloszlását azzal a feltételezéssel élve, hogy a kinetikus folyamat a sejtekben párhuzamosan halad. Ezt neveztük kvantilis módszernek. Ilyenkor különböző kvantilisekre illesztjük a függvényeket, majd az illesztett függvények megfelelő paramé-

tereit összeöntve kapjuk meg a paraméter-eloszlásokat. Az egyes sejtek feltételezett kinetikus paraméter-értéke tehát annak a függvénynek a paramétere lesz, melyet a sejtnak megfelelő kvantilisre illesztettünk. A feltételezés, hogy a különböző sejtekben párhuzamosan zajlanak a folyamatok, valószínűleg nem igaz (hiszen különböző sejtcsoportokat mérünk egyszerre, némelyikben oszcillál a kalcium-szint stb.), de egy empirikus (és nem mechanisztikus) modellben szerepelhet az adatok összefüggéseinek feltárására. A kvantilis módszerrel kapott paraméter-eloszlások előnye, hogy a mérés teljes tartományát leírják, és a paraméterek között meglévő összefüggések miatti bizonytalanságot a sokszori illesztéssel a kvantilis szerinti eloszlásba viszik át, de biológiai értelmezésük még hátra van.

Az illesztések során egy öt függvényből álló modell-készletet használunk, melyben elkülönítettük a fölfelé és lefele haladó logisztikus ill. dupla logisztikus függvényeket. Ennek előnye, hogy a kvantilis módszer során minden kvantilisre ugyanolyan irányú függvény illeszkedik, ezzel csökkentve a szélső kvantiliseknél az esetleges kiugró értékek miatti más irányú illesztésekből adódó szórást. Az adott mérésre legjobban illeszkedő modellt 10-szeres kereszt-validáció (CV) módszerével választjuk ki a túlillesztést elkerülendő.

A hálós illesztés a kvantilis módszer szigorítása azzal, hogy a paraméterek kvantilis-függését egy metafüggvénnyel írjuk le, nem hagyjuk szabadon. A módszer hátránya, hogy mesterségesen csökkenti a paraméterek szórását, ezáltal mesterséges különbségeket állapíthat meg egyébként kevésbé eltérő mérések között. A kalcium-görbék kinetikája a valóságban kevésbé függ össze a kvantilissal, mint azt a hálós illesztés javasolja, de jobban, mintha random mintákat vennénk az adathalmazból, és ezekre illesztenénk függvényeket (melyeknek, mivel a mérési adatok szórása hatalmas, mind nagyon különböző alakjuk lenne). A kettő között van a kvantilis módszer, amely valószínűleg még mindig túl szigorú az összefüggés megállapítása szempontjából.

A medián és kvantilis módszert felhasználóbarát módon implementáltuk, és a tudományos közösség rendelkezésére bocsátottuk a <http://www.facsKin.com> honlapon. A honlapon a program felhasználói dokumentációja is elérhető.

6.1. Továbbfejlesztési lehetőségek

A függőleges irányú standardizálás helyett a függvények y_0 , y_1 , y_2 paraméterei közül egy kivételével a többi lehetne relatív paraméter, így egyszerűbbé válna a FacsKin program felhasználói felülete, és külön standardizáló számítások nélkül összehasonlíthatóvá válnának bizonyos paraméterek. Ezt a műveletet a vízszintes, idő irányú paramétereknél megtettük, de ott relatív helyett különbség-paraméterekre van szükség, mert az idő

intervallum-skála, míg a fluoreszcens intenzitás arány-skála.

A paraméterek összefüggései (non-linearitásuk) megnöveli a szórásukat és megnehezíti a non-lineáris módszerekkel történő illesztésüket. Egy lineáris modell egy lépésben, iterációk nélkül illeszkedik, míg non-lineáris modellekre nem ismerünk általánosan működő illesztő eljárást. A non-linearitás két összetevőből áll ([117], [118]): (1) az intrinsic non-linearitás a fizikai jelenségből, a mért adatokból következik, ezen a modell egyszerűsítésével változtathatunk csak, de akkor nem biztos, hogy leírja a jelenséget, (2) paraméter-hatás non-linearitás, melyen átparaméterezéssel lehet segíteni. A (dupla) logisztikus függvény átparaméterezésével a paraméterek közötti összefüggések esetleg csökkenthetők lennének, igaz, hogy ennek az lehet az ára, hogy a biológiai értelmezhetőségük veszik el. Előny lenne viszont a jobb (gyorsabb, gyakrabban konvergáló) illesztés.

A bootstrap módszer által adott szórásokat figyelembe lehetne venni az összehasonlításoknál, ehhez azonban több mérésre kellene egyszerre illeszteni. Ekkor szükség lenne egy egyszerűbb modellekké átparaméterezhető függvényre, és összehasonlításnál meg kellene határozni a nullhipotézist a következőképpen:

1. Először megadjuk, hogy mely paraméterek mely méréseknél vagy méréscsoportoknál egyeznek meg, míg mely paraméterek lehetnek különbözőek (utóbbiaknak esetleges egyezőségére nem vagyunk kíváncsiak: ilyen lehet pl. a logisztikus függvénynél az 50%-os érték elérési ideje paraméter, ha tudjuk, hogy a mérések egymáshoz képest időben eltolva kezdődhetnek).
2. Az illesztést elvégezzük így, majd elvégezzük úgy is, hogy nem választjuk közösen a megadott paramétereket.
3. A két különböző illesztés eltérésösszegének összehasonlításával megkapjuk, hogy a paraméterek valójában különböztek-e a méréseknél: ha az első illesztés eltérésösszege szignifikánsan nagyobb, mint a másodiké, a paraméterek különböznek.
4. Hasonlóan, ha egy paraméter bootstrap konfidencia intervalluma tartalmazza a 0-t, érdemes illeszteni az egyszerűbb modellt (így lenne megadható a modell-választás). Az egyszerűbb modellek közé beletartozhatna ugyanaz a függvény úgy is, hogy két paramétere megegyezik, pl. a kezdő- és a végső érték, ezt is lehetne tesztelni bootstrap konfidencia intervallumokkal.

Az általunk fejlesztett szoftver a mérések értékelését (az illesztést) egyesével, mérésenként végzi, és nagy adatmennyiség mellett a több mérésre illesztés igen sok számítást igényelne. A helyettesítés (4.2. alfejezet) mindenképp szükséges lenne ennek felgyorsításához.

Az FCM kinetikus méréseket érdemes lenne nagy mennyiségű sejten elvégzett sejtenkénti, egyesével mérésekkel összevetni, ugyanolyan mintákon, ekkor több adatokat kaphatnánk a tényleges sejtenkénti kalcium-görbék eloszlásáról, és a kvantilis módszer biológiai értelméről. Ezen adatok ismeretében esetleg kifejleszthetők lennének a kalcium-flux folyamat sejtelettani alapjaira építő mechanisztikus modellek is, melyekben közvetlen biológiai értelmezéssel rendelkező paramétereket illeszthetnénk, mint amilyen a kezdeti kalcium-koncentráció, különböző transzmembrán fehérjék, csatornák, pumpák mennyisége, sejtípus stb.

További fejlesztési irány lehetne különböző, a sejtben egyszerre lezajló folyamatokra (melyek mérhetők is egyszerre, lásd [20]) egyszerre történő függvény-illesztés. Ehhez szintén mechanisztikus modell lenne a legmegfelelőbb. Ilyen méréseknél további információt ad, hogy egy sejtről két különböző értéket is figyelembe tudunk venni egyszerre.

7. fejezet

Következtetések

1. Kifejlesztettünk egy függvény-készletet, mely a kinetikus FCM mérések nagy részét képes leírni. Mindegyik függvény különböző kinetikus modellnek felel meg: a konstans függvény a kinetikus folyamat hiányának, a pozitív logisztikus függvény az adott szintről induló, emelkedő, majd egy adott szintet elérő folyamatnak, a negatív logisztikus függvény a hasonló, csak csökkenő folyamatnak, a pozitív dupla logisztikus függvény az adott szintről induló, egy maximumot elérő, majd egy alacsonyabb, adott szintre csökkenő folyamatnak, a negatív dupla logisztikus függvény a hasonló, csak átmeneti minimummal rendelkező folyamatnak.
2. Algoritmusunk képes a kinetikus FCM mérésre a legjobban illeszkedő függvényt meghatározni (10-szeres kereszt validáció módszerével), és ezáltal a méréseket kvalitatív módon osztályozni. Az egyes paraméterek mérésből való megjósolhatóságának számszerűsítésére relatív szórásukat használjuk.
3. A függvények paraméterei a kinetikus folyamat különböző aspektusait írják le, melyekhez különböző biológiai jelentés társulhat: a maximum a kinetikus reakció nagyságát, a maximum elérési ideje annak gyorsaságát, az 50%-os értékek maximumtól való távolsága az indulás ill. a lecsengés gyorsaságát, míg meredekségei azt írják le, hogy ez a lecsengés mennyire hirtelen történik. A kezdő- és végső érték a mérés előtti ill. utáni kinetikus paraméter szintre enged következtetni. Az AUC paraméter a kinetikus folyamat összesített méretét adja meg, kalcium-flux esetén pl. a kalcium-ionok citoplazmában eltöltött idejének összegével arányos.
4. A különböző paraméterek alapján a méréseket a paramétereknek megfelelő aspektusok mentén tudjuk összehasonlítani, ezzel finomítva az összességében, pl. az AUC paraméter által adott különbségek háttérét, ill. ki tudunk mutatni olyan különbségeket, melyek összesítve nem látszanak.

5. Módszerünket számos esetben sikerrel alkalmazták kalcium-flux adatok elemzésére és biológiai következtetések levonására. Az utóbbi három évben 6 ezzel kapcsolatos publikáció jelent meg.
6. A módszert implementáló algoritmus a tudományos közvélemény számára elérhető a <http://www.facskin.com> honlapon egy felhasználóbarát program formájában. A program képes mérések kézi kapuzására, függvények illesztésére, az illesztett függvények paramétereinek megadására, mérések csoportosítására és paraméterek alapján történő összehasonlítására.
7. Az egyelőre biológiailag nem validált, ún. kvantilis-módszerrel a kinetikus FCM méréseknek nem csak a közepét (mediánját), hanem teljes terjedelmét le tudjuk írni.

8. fejezet

Összefoglalás

Háttér: kinetikus áramlási citometriás mérések értékelésére az elterjedt módszer az idő szerinti simítás, majd a kapott görbe különböző paramétereinek leolvasása. Ezen értékeket nagymértékben befolyásolják az alkalmazott simítási módszer beállításai.

Célkitűzések: célul tűztük ki egy olyan algoritmus kifejlesztését, mely alkalmas kinetikus áramlási citometriás mérésekből numerikus jellemzők kiszámítására, nem függ a felhasználó által megadott ad-hoc beállításoktól, képes paraméterek formájában számszerűsíteni a kinetika különböző aspektusait, használható kalcium-flux és más kinetikus mérések értékelésére a kutatásban.

Módszerek: kifejlesztettünk egy függvénycsaládot a különböző kinetikájú mérések leírására. A függvényeket nem-lineáris regressziós módszerekkel illesztettük a mérésekre. Hogy ne csak a mérés közepét írjuk le, bootstrap, kvantilis ill. hálós módszerekkel többször illesztettük ugyanazt a függvényt a mérésre.

Eredmények: a kinetikus mérések túlnyomó része leírható az általunk kifejlesztett függvénycsaláddal. Legtöbbször konvergáló illesztési módszernek a Nelder-Mead és SANN algoritmus kombinációját találtuk. A T-limfocita aktiváció különböző betegségek (1-es típusú diabetes, SM, preeclampsia, RA) és körülmények (újszülött kor, bizonyos kalcium-pumpák és kálium-csatornák gátlása) hatására bekövetkező specifikus változásainak leírására a medián függvény maximum, maximum elérési idő, 50%-os érték mereksége és AUC paraméterei megfelelőnek bizonyultak.

Konklúzió: kifejlesztettünk egy algoritmust kinetikus mérések leírására és összehasonlítására, és elérhetővé tettük a tudományos közösség számára a <http://www.facskin.com> honlapon. A honlapon 76 kutató regisztrált és az utóbbi másfél évben 1846 mérést értékelt ki. Az elmúlt három évben 6 tudományos közlemény jelent meg, melyekben a program által adott eredményeket biológiai következtetésre használták. A program kvalitatív módon csoportosítja a kinetikus méréseket, és numerikus paraméterekkel jellemzi a mérést, elkülönítve a kinetikus folyamat különböző aspektusait.

9. fejezet

Summary

Background: the common method for evaluating kinetic flow cytometric measurements is smoothing along time and calculating different parameters from the resulting curve. The exact values of these parameters depend on the properties of the smoothing method used.

Aims: our aim was to develop an algorithm suitable for calculating numerical attributes from kinetic measurements which do not depend on ad-hoc user-defined settings, are capable of numerically characterizing different aspects of the kinetic process and are suitable for describing Calcium flux and other kinetic measurements in medical research.

Methods: we developed a family of functions describing measurements with different kinetics. We fitted these functions to the measurements using non-linear regression methods. In order to describe not only the middle of the measurement but the whole range of values, we fitted the same function multiple times to the same measurement using bootstrap, quantile and metafunction methods.

Results: the majority of kinetic measurements can be described by the family of functions we developed. We found the combination of the Nelder-Mead and SANN algorithms to have the best convergence properties. We were able to use the maximum, time to reach maximum, slope at 50% value and AUC parameters to describe the change of T-lymphocyte activation in different diseases (type 1 diabetes, SM, preeclampsia, RA) and conditions (newborns, specific inhibition of certain Potassium and Calcium channels).

Conclusion: we developed an algorithm for the description and comparison of kinetic measurements and made it available to the scientific community on the <http://www.facskin.com> website. 76 researchers have registered on the website and during the past 1.5 years 1846 measurements were analyzed using the program. During the last 3 years, 6 papers were published using the results given by the program to establish biological conclusions. The program classifies measurements qualitatively by determining the best fitting function and describes the measurement by objective numerical parameters separating different aspects of the kinetic process.

10. fejezet

Irodalomjegyzék

- [1] Drouet M, Lees O. (1993) Clinical applications of flow cytometry in hematology and immunology. *Biol Cell*, 78: 73-8.
- [2] Jaye DL, Bray RA, Gebel HM, Harris WA, Waller EK. (2012) Translational applications of flow cytometry in clinical practice. *J Immunol*, 188: 4715-9.
- [3] Rieseberg M, Kasper C, Reardon KF, Scheper T. (2001) Flow cytometry in biotechnology. *Appl Microbiol Biotechnol*, 56: 350-60.
- [4] June CH, Ledbetter JA, Rabinovitch PS, Martin PJ, Beatty PG, Hansen JA. (1986) Distinct patterns of transmembrane calcium flux and intracellular calcium mobilization after differentiation antigen cluster 2 (E rosette receptor) or 3 (T3) stimulation of human lymphocytes. *J Clin Invest*, 77: 1224-1232.
- [5] Spidlen J, Shooshtari P, Kollmann TR and Brinkman RR. (2011) Flow cytometry data standards. *BMC Res Notes*, 4: 50.
- [6] Murphy RF, Chused TM. (1984) A proposal for a flow cytometric data file standard. *Cytometry*, 5: 553-5.
- [7] Seamer LC, Bagwell CB, Barden L, Redelman D, Salzman GC, Wood JC, Murphy RF. (1997) Proposed new data file standard for flow cytometry, version FCS 3.0. *Cytometry*, 28: 118-22.
- [8] Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombey P, Hyun B, Hubbard B, Lange S, Lefebvre R, Leif R, Novo D, Ostruszka L, Treister A, Wood J, Murphy RF, Roederer M, Sudar D, Zigon R, Brinkman RR. (2010) Data File Standard for Flow Cytometry, version FCS 3.1. *Cytometry A*, 77A: 97-100.

- [9] Lugli E, Roederer M, Cossarizza A. (2010) Data analysis in flow cytometry: the future just started. *Cytometry A*, 77: 705-13.
- [10] Roederer M. (2001) Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. *Cytometry A*, 45: 194-205.
- [11] Mészáros G, Rónai K, Toldi G, Kaposi A, Vásárhelyi B, Treszl A. (2008) Sejt-élettani folyamatok jellemzése „real-time” áramlási citometriás módszerrel. *Magyar Immunológia*, 7: 22-29.
- [12] Grynkiewicz G, Poenie M, Tsien RY. (1985) A new generation of Ca²⁺ indicators with greatly improved fluorescence properties. *J Biol Chem*, 260: 3440–3450.
- [13] Minta A, Kao JP, Tsien RY. (1989) Fluorescent indicators for cytosolic calcium based on rhodamine and fluorescein chromophores. *J Biol Chem*, 264: 8171-8.
- [14] June CH, Rabinovitch PS. (1990) Flow cytometric measurement of intracellular ionized calcium in single cells with indo-1 and fluo-3. *Methods Cell Biol*, 33: 37–58.
- [15] Novak EJ, Rabinovitch PS. (1994) Improved sensitivity in flow cytometric intracellular ionized calcium measurement using fluo3/Fura Red fluorescence ratios. *Cytometry*, 17: 135-41.
- [16] Bailey S, Macardle PJ. (2006) A flow cytometric comparison of Indo-1 to fluo-3 and Fura Red excited with low power lasers for detecting Ca²⁺ flux. *J Immunol Methods*, 311: 220–225.
- [17] Hajnoczky G, Robb-Gaspers LD, Seitz MB, Thomas AP. (1995) Decoding of cytosolic calcium oscillations in the mitochondria. *Cell*, 82: 415–424.
- [18] Mello de Queiroz F, Ponte CG, Bonomo A, Vianna-Jorge R, Suarez-Kurtz G. (2008) Study of membrane potential in T lymphocytes subpopulations using flow cytometry. *BMC Immunol*, 9: 63.
- [19] Nagy G, Koncz A, Perl A. (2003) T cell activation-induced mitochondrial hyperpolarization is mediated by Ca²⁺ and redox-dependent production of nitric oxide. *J Immunol*, 171: 5188–5197.
- [20] Mészáros G, Szalay B, Toldi G, Kaposi A, Vásárhelyi B, Treszl A. (2012) Kinetic Measurements Using Flow Cytometry: New Methods for Monitoring Intracellular Processes. *Assay Drug Dev Technol*, 10: 97-104.

- [21] Vig M, Kinet JP. (2009) Calcium signaling in immune cells. *Nat Immunol*, 10: 21-27.
- [22] Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N. *flowCore: flowCore: Basic structures for flow cytometry data*. R package version 1.22.3.
- [23] Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5: R80.
- [24] R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.r-project.org>
- [25] Herzenberg LA, Tung J, Moore WA, Parks DR. (2006) Interpreting flow cytometry data: A guide for the perplexed. *Nat Immunol*, 7: 681–685.
- [26] Lugli E, Troiano L, Cossarizza A. (2009) Investigating T cells by polychromatic flow cytometry. *Methods Mol Biol*, 514: 47–63.
- [27] Edwards BS, Zhu J, Chen J, Carter MB, Thal DM, Tesmer JJ, Graves SW, Sklar LA. (2012) Cluster cytometry for high-capacity bioanalysis. *Cytometry A*, 81: 419-29.
- [28] Spidlen J, Leif RC, Moore W, Roederer M, Brinkman RR. (2008) Gating-ML: XML-based Gating Descriptions in Flow Cytometry. *Cytometry A*, 73A: 1151–1157.
- [29] Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, Hyun B, Jansen K, Kollmann T, Kong M, Leif R, McWeeney S, Moloshok TD, Moore W, Nolan G, Nolan J, Nikolich-Zugich J, Parrish D, Purcell B, Qian Y, Selvaraj B, Smith C, Tchuvatkina O, Wertheimer A, Wilkinson P, Wilson C, Wood J, Zigon R, International Society for Advancement of Cytometry Data Standards Task Force, Scheuermann RH, Brinkman RR. (2008) MIFlowCyt: The minimum information about a flow cytometry experiment. *Cytometry A*, 73A: 926–930.
- [30] Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308: 523–529.

- [31] Roederer M, Hardy RR. (2001) Frequency difference gating: A multivariate method for identifying subsets that differ between samples. *Cytometry*, 45: 56–64.
- [32] Larsen M, Sauce D, Arnaud L, Fastenackels S, Appay V, Gorochov G. (2012) Evaluating cellular polyfunctionality with a novel polyfunctionality index. *PLoS One*, 7: e42403.
- [33] Lugli E, Pinti M, Nasi M, Troiano L, Ferraresi R, Mussi C, Salvioli G, Patsekina V, Robinson JP, Durante C, Cocchi M, Cossarizza A. (2007) Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data. *Cytometry A*, 71A: 334–344.
- [34] Kalina T, Stuchly J, Janda A, Hrusak O, Ruzickova S, Sediva A, Litzman J, Vlkova M. (2009) Profiling of polychromatic flow cytometry data on B-cells reveals patients' clusters in common variable immunodeficiency. *Cytometry A*, 75A: 902–909.
- [35] Lo K, Brinkman RR, Gottardo R. (2008) Automated gating of flow cytometry data via robust model-based clustering. *Cytometry A*, 73A: 321–332.
- [36] Murphy RF. (1985) Automated identification of subpopulations in flow cytometric list mode data using cluster analysis. *Cytometry*, 6: 302–309.
- [37] Maynadié M, Picard F, Husson B, Chatelain B, Cornet Y, Le Roux G, Campos L, Dromelet A, Lepelley P, Jouault H, Imbert M, Rosenwadj M, Vergé V, Bissières P, Raphaël M, Béné MC, Feuillard J. (2002) Immunophenotypic clustering of myelodysplastic syndromes. *Blood*, 100: 2349–2356.
- [38] Kitsos CM, Bhamidipati P, Melnikova I, Cash EP, McNulty C, Furman J, Cima MJ, Levinson D. (2007) Combination of automated high throughput platforms, flow cytometry, and hierarchical clustering to detect cell state. *Cytometry A*, 71A: 16–27.
- [39] Duque RE. (2012) Minimal disease detection of B-cell lymphoproliferative disorders by flow cytometry: multidimensional cluster analysis. *Hematology*, 17 Suppl 1: S63-5.
- [40] Wilkins MF, Hardy SA, Boddy L, Morris CW. (2001) Comparison of five clustering algorithms to classify phytoplankton from flow cytometry data. *Cytometry*, 44: 210–217.
- [41] De Rosa SC, Lu FX, Yu J, Perfetto SP, Falloon J, Moser S, Evans TG, Koup R, Miller CJ, Roederer M. (2004) Vaccination in humans generates broad T cell cytokine responses. *J Immunol*, 173: 5372–5380.

- [42] Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, Lederman MM, Benito JM, Goepfert PA, Connors M, Roederer M, Koup RA. (2006) HIV nonprogressors preferentially maintain highly functional HIV-specific CD81 T cells. *Blood*, 107: 4781–4789.
- [43] Wold S, Esbensen K, Geladi P. (1987) Principal component analysis. *Chemom Intell Lab Syst*, 2: 37–52.
- [44] Mann RC, Popp DM, Hand RE Jr. (1984) The use of projections for dimensionality reduction of flow cytometric data. *Cytometry*, 5: 304–307.
- [45] Bashashaty A, Brinkman R. (2009) A survey of flow cytometry data analysis methods. *Adv Bioinform*, 2009:584603.
- [46] Overton WR. (1988) Modified histogram subtraction technique for analysis of flow cytometry data [published erratum appears in *Cytometry*. 1989;10(4):492–494]. *Cytometry*, 9: 619–626.
- [47] Roederer M, Treister A, Moore W, Herzenberg LA. (2001) Probability binning comparison: a metric for quantitating univariate distribution differences. *Cytometry*, 45: 37-46.
- [48] Lampariello F. (1994) Evaluation of the number of positive cells from flow cytometric immunoassays by mathematical modelling of cellular autofluorescence. *Cytometry*, 15: 294–301.
- [49] Lampariello F, Aiello A. (1998) Complete mathematical modelling method for the analysis of immunofluorescence distributions composed of negative and weakly positive cells. *Cytometry*, 32:241–254.
- [50] Hahn GJ, Shapiro SS: *Statistical Models in Engineering*. John Wiley & Sons, Inc., New York, 1967.
- [51] Watson JV. (2001) Proof without prejudice revisited: immunofluorescence histogram analysis using cumulative frequency subtraction plus ratio analysis of means. *Cytometry*, 43: 55-68.
- [52] Hollander M and Wolfe DA. *Nonparametric Statistical Methods*. New York: John Wiley & Sons, 1973:115-120.
- [53] Sokal RR, Rohlf FJ. *Biometry*. W. H. Freeman and Company, San Francisco, 1969:607.

- [54] Newcombe, Robert G. (1998) Interval estimation for the difference between independent proportions. Comparison of eleven methods. *Stat Med.* 17: 873-90.
- [55] Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*, 3rd Edition, 2003.
- [56] Chambers JM, Cleveland WS, Kleiner B, Tukey PA. *Graphical Methods for Data Analysis*. Wadsworth & Brooks/Cole, 1983.
- [57] Sladek TL, Jacobberger JW. (1993) Flow cytometric titration of retroviral expression vectors: Comparison of methods for analysis of immunofluorescence histograms derived from cells expressing low antigen levels. *Cytometry*, 14: 23–31.
- [58] Conover WJ. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, 1971:309-314.
- [59] Young IT. (1977) Proof without prejudice: Use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. *J Histochem Cytochem*, 25: 935–941.
- [60] Pearson ES, Hartley HO ed. *Biometrika Tables for Statisticians*. 2. Cambridge University Press, 1972:117–123.
- [61] Marsaglia G, Tsang WW, Wang J. (2003) Evaluating Kolmogorov's distribution. *Journal of Statistical Software*, 8: 1-4.
- [62] Brescia F, Sarti M. (2008) Modification to the Lampariello approach to evaluate reactive oxygen species production by flow cytometry. *Cytometry A*, 73: 175-9.
- [63] Lampariello F. (2000) On the use of the Kolmogorov-Smirnov statistical test for immunofluorescence histogram comparison. *Cytometry*, 39: 179-88.
- [64] Roederer M, Moore W, Treister A, Hardy RR, Herzenberg LA. (2001) Probability binning comparison: a metric for quantitating multivariate distribution differences. *Cytometry*, 45: 47-55.
- [65] Cox C, Reeder JE, Robinson RD, Suppes SB, Wheelless LL. (1988) Comparison of frequency distributions in flow cytometry. *Cytometry*, 9: 291–298.
- [66] Bagwell C. (1996) A journey through flow cytometric immunofluorescence analyses. *Clin Immunol Newsletter*, 16: 33–37.

- [67] Rogers WT, Moser AR, Holyst HA, Bantly A, Mohler ER, Scangas G, and Moore JS. (2008) Cytometric fingerprinting: quantitative characterization of multivariate distributions. *Cytometry A*, 73: 430–41.
- [68] Holyst H, Rogers W (2009). flowFP: Fingerprinting for Flow Cytometry. R package version 1.14.0.
- [69] Bagwell CB, Hudson JL, Irvin GL. (1979) Non-parametric flow cytometry analysis. *J Histochem Cytochem*, 27: 293-6.
- [70] Zeng Q, Wand M, Young AJ, Rawn J, Milford EL, Mentzer SJ, Greenes RA. (2002) Matching of flow-cytometry histograms using information theory in feature space. *Proc AMIA Symp*, 2002: 929-33.
- [71] Parikh HH, Li WC, Ramanathan M. (1999) Evaluation of an alternative to the Kolmogorov-Smirnov test for flow cytometric histogram comparisons. *J Immunol Methods*, 229: 97-105.
- [72] Davey HM, Jones A, Shaw AD, Kell DB. (1999) Variable selection and multivariate methods for the identification of microorganisms by flow cytometry. *Cytometry*, 35: 162–168.
- [73] Cox DR. (1972) Regression models and life-tables. *J R Statist Soc B*, 34: 187-220.
- [74] Kaposi A, Toldi G, Mészáros G, Szalay B, Veress G, Vásárhelyi B. Experimental conditions and mathematical analysis of kinetic measurements using flow cytometry - the FacsKin method. In: *Flow Cytometry/Book 1*. Schmid I (ed.). Intech, 2012.
- [75] Lund-Johansen F, Olweus J. (1992) Signal transduction in monocytes and granulocytes measured by multiparameter flow cytometry. *Cytometry*, 13: 693-702.
- [76] Rijkers GT, Griffioen AW. (1993) Changes in free cytoplasmic magnesium following activation of human lymphocytes. *Biochem J*, 289: 373-377.
- [77] do Céu Monteiro M, Sansonetty F, Gonçalves MJ, O'Connor JE. (1999) Flow cytometric kinetic assay of calcium mobilization in whole blood platelets using Fluo-3 and CD41. *Cytometry*, 35: 302-310.
- [78] Schepers E, Glorieux G, Dhondt A, Leybaert L, Vanholder R. (2009) Flow cytometric calcium flux assay: evaluation of cytoplasmic calcium kinetics in whole blood leukocytes. *J Immunol Methods*, 348: 74-82.

- [79] Omann GM, Harter JM. (1991) Pertussis toxin effects on chemoattractant-induced response heterogeneity in human PMNs utilizing Fluo-3 and flow cytometry. *Cytometry*, 12: 252-259.
- [80] Norgauer J, Dobos G, Kownatzki E, Dahinden C, Burger R, Kupper R, Gierschik P. (1993) Complement fragment C3a stimulates Ca²⁺ influx in neutrophils via a pertussis-toxin-sensitive G protein. *Eur J Biochem*, 217: 289-94.
- [81] Szalay B, Mészáros G, Cseh A, Acs L, Deák M, Kovács L, Vásárhelyi B, Balog A. (2012) Adaptive Immunity in Ankylosing Spondylitis: Phenotype and Functional Alterations of T-Cells before and during Infliximab Therapy. *Clin Dev Immunol*, 2012: 808724
- [82] Jakubczak B, Wasik M, Popko K, Demkow U. (2006) Kinetics of calcium ion concentration accompanying signal transduction in neutrophils from children with increased susceptibility to infections. *J Physiol Pharmacol*, 57: 131-137.
- [83] Cleveland WS, LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35: 54.
- [84] Demkow U, Winklewski P, Potapinska O, Popko K, Lipinska A, Wasik M. (2009) Kinetics of calcium ion concentration accompanying transduction of signals into neutrophils from diabetic patients and its modification by insulin. *J Physiol Pharmacol*, 60: 37-40.
- [85] Dinya E. *Biometria az orvosi gyakorlatban*. 3. átdolgozott, bővített kiadás. Medicina, Budapest, 2011: 373-380.
- [86] Calabrese EJ, Baldwin LA. (2002) Defining hormesis. *Hum Exp Toxicol*, 21: 91-97.
- [87] Cedergreen N, Ritz C, Streibig JC. (2005) Improved empirical models describing hormesis. *Environmental Toxicology and Chemistry*, 24: 3166-72.
- [88] Rozman KK, Doull J. (2003) Scientific foundations of hormesis, part 2: maturation, strengths, limitations, and possible applications in toxicology, pharmacology, and epidemiology. *Crit Rev Toxicol*. 33: 451-462.
- [89] Bates DM, Watts, DG *Nonlinear Regression Analysis and Its Applications*, Wiley, 1988.
- [90] Schnabel RB, Koontz JE, Weiss BE. (1985) A modular system of algorithms for unconstrained minimization. *ACM Trans. Math. Software*, 1985: 419-440.

- [91] Nash JC. Compact Numerical Methods for Computers. Linear Algebra and Function Minimisation. Adam Hilger, 1990.
- [92] Nelder JA, Mead RA. (1965) Simplex algorithm for function minimization. *Computer Journal*, 7: 308-313.
- [93] Belisle CJP. (1992) Convergence theorems for a class of simulated annealing algorithms on \mathbb{R}^d . *J Applied Probability*, 29: 885-895.
- [94] Motulsky, H, Christopoulos, A. Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Oxford University Press, USA, 2004: 134-135.
- [95] Hawkins DM. (2004) The Problem of Overfitting. *J Chem Inf Comput Sci*, 44: 1-12.
- [96] Picard RR, Cook RD. (1984) Cross-Validation of Regression Models. *Journal of the American Statistical Association*. 79: 575-583.
- [97] Moore AW. Cross-validation for detecting and preventing overfitting (presentation). School of Computer Science, Carnegie Mellon University. <http://www.autonlab.org/tutorials/overfit10.pdf> (2012. március 31-én elérhető)
- [98] Oehlert GW. (1992) A Note on the Delta Method. *The American Statistician*, 46: 27-29.
- [99] Moore DS, McCabe GP, Craig B. Introduction to the Practice of Statistics, Seventh Edition. Freeman and Company Book. 2010. Chapter 16: Bootstrap Methods and Permutation Tests.
- [100] Kaposi AS, Veress G, Vásárhelyi B, Macardle P, Bailey S, Tulassay T, Treszl A. (2008) Cytometry-acquired calcium-flux data analysis in activated lymphocytes. *Cytometry A*, 73: 246-53.
- [101] Toldi G, Treszl A, Pongor V, Gyarmati B, Tulassay T, Vásárhelyi B. (2010) T-lymphocyte calcium influx characteristics and their modulation by Kv1.3 and IKCa1 channel inhibitors in the neonate. *Int Immunol*, 22: 769-74.
- [102] Toldi G, Vásárhelyi B, Kaposi A, Mészáros G, Pánczél P, Hosszufalusi N, Tulassay T, Treszl A. (2010) Lymphocyte activation in type 1 diabetes mellitus: the increased significance of Kv1.3 potassium channels. *Immunol Lett*, 133: 35-41.

- [103] Toldi G, Stenczer B, Treszl A, Kollár S, Molvarec A, Tulassay T, Rigó J, Vásárhelyi B. (2011) Lymphocyte calcium influx characteristics and their modulation by Kv1.3 and IKCa1 channel inhibitors in healthy pregnancy and preeclampsia. *Am J Reprod Immunol*, 65: 154-63.
- [104] Toldi G, Folyovich A, Simon Z, Zsiga K, Kaposi A, Mészáros G, Tulassay T, Vásárhelyi B. (2011) Lymphocyte calcium influx kinetics in multiple sclerosis treated without or with interferon beta. *J Neuroimmunol*, 237: 80-6.
- [105] Toldi G, Bajnok A, Dobi D, Kaposi A, Kovács L, Vásárhelyi B, Balog A. (2013) The effects of Kv1.3 and IKCa1 potassium channel inhibition on calcium influx of human peripheral T lymphocytes in rheumatoid arthritis. *Immunobiology*, 218: 311-6.
- [106] Toldi G, Kaposi A, Zsembery A, Treszl A, Tulassay T, Vásárhelyi B. (2012) Human Th1 and Th2 lymphocytes are distinguished by calcium flux regulation during the first 10 min of lymphocyte activation. *Immunobiology*, 217: 37-43.
- [107] Java programozási nyelv, Oracle Inc., Redwood Shores, CA, USA. <http://java.oracle.com> (2012. szeptember 24-én elérhető)
- [108] Watson JV. (2001) A brief history of numbers and statistics with cytometric applications. *Cytometry*, 46: 1-22.
- [109] Motulsky HJ, Ransnas LA. (1987) Fitting curves to data using nonlinear regression: a practical and nonmathematical review. *FASEB J*, 1: 365-374.
- [110] Motulsky H, Christopoulos A. (2004) Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Oxford University Press, USA, 2004: 58-59.
- [111] Meyer T, Stryer L. (1988) Molecular model for receptor-stimulated calcium spiking. *Proc Natl Acad Sci USA*, 85: 5051-5055.
- [112] Politi A, Gaspers LD, Thomas AP, Höfer T. (2006) Models of IP₃ and Ca²⁺ oscillations: frequency encoding and identification of underlying feedbacks. *Biophys J*, 90: 3120-33.
- [113] Tang Y, Stephenson JL, Othmer HG. (1996) Simplification and analysis of models of calcium dynamics based on IP₃-sensitive calcium channel kinetics. *Biophys J*, 70: 246-63.

- [114] Szabó A. Intracelluláris Ca^{2+} -dinamika vizsgálata. Diplomamunka, BME Matematika Intézet, 2010.
- [115] Hess SD, Oortgiesen M, Cahalan MD. (1993) Calcium Oscillations in Human T and Natural Killer Cells Depend upon Membrane Potential and Calcium Influx. *J Immunol*, 150: 2620-2633.
- [116] Launay P, Cheng H, Srivatsan S, Penner R, Fleig A, Kinet JP. (2004) TRPM4 regulates calcium oscillations after T cell activation. *Science*, 306: 1374-7.
- [117] Karolczak M, Mickiewicz A. (1995) Why to Calculate, When to Use, and How to Understand Curvature Measures of Nonlinearity. *Current Separations*, 14: 1.
- [118] Ratkowsky DA. *Nonlinear Regression Modeling*. Marcel Dekker, Inc., New York, 1983.

11. fejezet

Saját publikációk jegyzéke

11.1. Az értekezéshez kapcsolódó közlemények

Összesített impakt faktor: 17,351, első szerzőként: 3,729.

- Nemzetközi közlemények:

1. **Kaposi A**, Veress G, Vasarhelyi B, Macardle P, Bailey S, Tulassay T, Treszl A. (2008) Cytometry-acquired calcium-flux data analysis in activated lymphocytes. *Cytometry Part A*, 73A: 246-53. IF: 3,729
2. Toldi G, Vásárhelyi B, **Kaposi A**, Mészáros G, Pánczél P, Hosszúfalusi N, Tulassay T, Treszl A. (2010) Lymphocyte activation in type 1 diabetes mellitus: The increased significance of Kv1.3 potassium channels. *Immunol Lett*, 133: 35-41. IF: 2,526
3. Toldi G, Folyovich A, Simon Z, Zsiga K, **Kaposi A**, Mészáros G, Tulassay T, Vásárhelyi B. (2011) Lymphocyte calcium influx kinetics in multiple sclerosis treated without or with interferon beta. *J Neuroimmunol*, 237: 80-86. IF: 2,959
4. Mészáros G, Szalay B, Toldi G, **Kaposi A**, Vásárhelyi B, Treszl A. (2012) Kinetic Kinetic Measurements Using Flow Cytometry: New Methods for Monitoring Intracellular Processes. *ASSAY and Drug Development Technologies*, 10: 97-104. IF: 1,727
5. Toldi G, **Kaposi A**, Zsembery Á, Treszl A, Tulassay T, Vásárhelyi B. (2012) Human Th1 and Th2 Lymphocytes are Distinguished by Calcium Flux Regulation During the First Ten Minutes of Lymphocyte Activation. *Immunobiology*, 217: 37-43. IF: 3,205
6. Toldi G, Bajnok A, Dobi D, **Kaposi A**, Kovács L, Vásárhelyi B, Balog A. (2012) The effects of Kv1.3 and IKCa1 potassium channel inhibition on cal-

cium influx of human peripheral T lymphocytes in rheumatoid arthritis. Immunobiology, 218: 311-6. IF: 3,205

- Könyvfejezet:

7. **Kaposi A**, Toldi G, Mészáros G, Szalay B, Veress G, Vásárhelyi B. Experimental conditions and mathematical analysis of kinetic measurements using flow cytometry - the FacsKin method. In: Flow Cytometry/Book 1. Schmid I (ed.). Intech, 2012. ISBN 979-953-307-355-1.

- Magyar közlemény:

8. Mészáros G, Rónai K, Toldi G, **Kaposi A**, Vásárhelyi B, Treszl A. (2008) Sejtéletani folyamatok jellemzése „real-time” áramlási citometriás módszerrel. Magyar Immunológia, 7: 22-29.

11.2. Az értekezéshez nem kapcsolódó közlemények

Összesített impakt faktor: 3,705, első szerzőként: 0.

- Nemzetközi közlemények:

9. Treszl A, **Kaposi A**, Hajdú J, Szabó M, Tulassay T, Vásárhelyi B. (2007) The Extent to Which Genotype Information May Add to the Prediction of Disturbed Perinatal Adaptation: None, Minor, or Major? *Pediatr Res*, 62: 610-4. IF: 2,7
10. Gombos P, Langer RM, Korbely R, Varga M, **Kaposi A**, Dinya E, and Müller V. (2010) Smoking Following Renal Transplantation in Hungary and Its Possible Deleterious Effect on Renal Graft Function. *Transplantation Proceedings*, 42: 2357–2359. IF: 1,005

- Könyvfejezet:

11. Mihalik A, **Kaposi A**, Kovacs IA, Nanasi T, Palotai R, Rak A, Szalay MS, Csermely P. How creative elements help the recovery of networks after crisis: lessons from biology. *Networks in Social Policy Problems*. B Vedres and M Scotti (ed.). Cambridge University Press, Cambridge, 2011.

- Magyar közlemény:

12. Lantos B, Riesz P, Gulácsi A, Keszthelyi A, Szűcs M, **Kaposi A**, Romics I. (2009) Hólyagrák-diagnosztikai program klinikai felhasználása. Uroonkológia, 6: 51-56.

12. fejezet

Köszönetnyilvánítás

2006 őszén, harmadéves orvostanhallgató koromban kapcsolódtam be TDK hallgatóként az I. sz. Gyermekklinikán működő áramlási citometriás laboratórium munkájába. Az itt tapasztalt szakmailag színvonalas és emberileg befogadó, támogató közeg nélkül az orvosi egyetemet sem tudtam volna elvégezni. Köszönettel tartozom témavezetőmnek, Vászárhegyi Barna professzornak, aki megteremtette ezt a közeget, biztatott és bármilyen probléma esetén mindig elérhető volt. Szeretnék köszönetet mondani Tulassay Tivadar professzor úrnak személyes támogatásáért és a laboratórium munkájához szükséges feltételek megteremtéséért. A laboratóriumban dolgozó Dr. Treszl Andrásról sokat tanultam az orvosi statisztikáról, Dr. Toldi Gergely, Dr. Mészáros Gergő és Dr. Szalay Balázs hallgatótársaim pedig a FacsKin program fő felhasználóiként hasznos észrevételeikkel, kérdéseikkel, igényeikkel segítettek elő a program fejlődését. Ők adták a motivációt a program fejlesztésére. A laboratórium kiváló atmoszférájának kialakításáért ők is felelősek voltak.

Köszönettel tartozom Veress Gábor matematikusnak, akinek szakmai irányítása alatt fejlesztettem az algoritmust. Minden közös megbeszélésünkön volt valamilyen izgalmas, továbblépést jelentő ötlete.

Köszönöm Dr. Dinya Eleknek, hogy megtanított az orvosi statisztika alapjaira, és soha véget nem érő, ezzel kapcsolatos feladatokkal látott el. Nagyon sokat tanultam tőle.

A FacsKin programmal kapcsolatos értékes megjegyzéseikért köszönettel tartozom Éliás Gergelynek, Diviánszky Péternek és Egri Ádámnak, valamint a program összes felhasználójának.

Köszönöm azon önkéntesek és betegek felajánlását és áldozatkészségét, akik vérmin-tákkal segítettek a kísérletek elvégzését.