



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

**NEUMANN JÁNOS
INFORMATIKAI KAR**



DIPLOMAMUNKA

**OE-NIK
2022**

Hallgató neve:
Hallgató törzskönyvi száma:

**Puskás Melánia
T/008831/FI12904/N**

Óbudai Egyetem
Neumann János Informatikai Kar
Biomatika és Alkalmazott Mesterséges Intelligencia Intézet

DIPLOMAMUNKA FELADATLAP

Hallgató neve: **Puskás Melánia**
Törzskönyvi száma: **T/008831/FI12904/N**
Neptun kódja: **W7WUGQ**

A diplomamunka címe:

**Élettani folyamatok paraméterbecslése mesterséges intelligencia
használatával**
Estimation of physiological processes using artificial intelligence

Intézményi konzulens: **Dr. Drexler Dániel András**
Külső konzulens:

Beadási határidő: **2022. december 15.**



A feladat

A jövő orvoslásában az egyik ígéretes irányvonal a terápiák matematikai és mérnöki módszereken alapuló optimalizálásában rejlik, amelyekkel a betegségek kezelése személyre szabottan történik, ellentétben a ma használt, ad hoc, nagy populációs átlagra kifejlesztett, ritkán optimalizált terápiákkal. A terápioptimalizálás számtalan előnnyel jár, ezek közül a legfontosabb az alacsonyabb dózis, ami kevesebb mellékhatással, alacsonyabb költségekkel jár, és bizonyos esetekben előnyös hatással lehet a gyógyszerrel szemben kialakuló rezisztencia leküzdésében is. A terápioptimalizáláshoz azonban szükség van a gyógyszer hatását leíró modellre, ami alapján az optimalizálás elvégezhető. Ez a modell függ a betegség típusától, a páciensről és az alkalmazott gyógyszertől. A személyre szabás alapja, hogy meg tudjuk adni az adott páciens modelljében szereplő paramétereket, amik alapján a terápioptimalizálás elvégezhető. Ehhez parametrikus identifikációra van szükség, amihez sok mérés kell, ami a betegállásban nehezen kivitelezhető, azonban kísérleti körülmények között van rá lehetőség.

A feladat egy tumornövekedési modell alapján egy paraméterbecslő rendszer létrehozása neurális hálózat segítségével. A neurális hálózat tanításához rendelkezésre állnak valós mérési adatokból identifikált modellparaméterek, amik kemoterápiás szerrel kezelt emlőrákos egerekből származnak. Az egyes egereket leíró modellparaméterek alapján tanító adatokat lehet generálni, a cél, hogy a neurális hálózat a bemeneti és kimeneti mérési adatok alapján megmondja az egyed modellparamétereit. A végső cél, hogy a rendszer olyan páciensek modellparamétereit is meg tudja mondani a mérések alapján, amelyek adatait nem használtuk fel a tanítás során, így egy olyan, a gyakorlatban használható eszközt kapunk, ami kísérletek alapján való tanítás után viszonylag kevés mérésből meg tudja adni a páciensek modellparamétereit.

A diplomamunkának tartalmaznia kell:

- A mesterséges neurális hálózatok és identifikációs algoritmusok témakörében végzett irodalomkutatást.
- A mesterséges neurális hálózatok és identifikációs algoritmusok kombinálásával létrehozott paraméterbecslő eljárás rendszertervét.
- A paraméterbecslő eljárás implementálását.
- A paraméterbecslő eljárás tesztelését és más algoritmusokkal való összehasonlítását.
- Az eredmények értékelését.



Dr. Eigner György
intézetigazgató

A diplomamunka elévülésének határideje: **2024. december 15.**
(OE TVSz 55.§ szerint)

A diplomamunkát beadásra alkalmasnak tartom:

.....
külső konzulens

.....
intézményi konzulens



HALLGATÓI NYILATKOZAT

Alulírott hallgató kijelentem, hogy a szakdolgozat / diplomamunka saját munkám eredménye, a felhasznált szakirodalmat és eszközöket azonosíthatóan közöltem. Az elkészült szakdolgozatomban / diplomamunkámban található eredményeket az egyetem és a feladatot kiíró intézmény saját céljára térítés nélkül felhasználhatja.

Budapest, 2022. december 12.



.....
hallgató aláírása



KONZULTÁCIÓS NAPLÓ

Hallgató neve: Neptun kód: Tagozat

Puskás Melánia W7WUGQ Nappali

Telefon: Levelezési cím (pl: lakcím):

06/70-575-1031 2322 Makád, Rákóczi Ferenc utca 35.

Szakdolgozat / Diplomamunka¹ címe magyarul:

ÉLETTANI FOLYAMATOK PARAMÉTERBECSLÉSE MESTERSÉGES INTELLIGENCIA HASZNÁLATÁVAL

Szakdolgozat / Diplomamunka² címe angolul:

ESTIMATION OF PHYSIOLOGICAL PROCESSES USING ARTIFICIAL INTELLIGENCE

Intézményi konzulens:

Külső konzulens:

Dr. Drexler Dániel András

Kérjük, hogy az adatokat nyomtatott nagybetűkkel írja!

Alk.	Dátum	Tartalom	Aláírás
1.	2021.09.21.	Az eddigi munka áttekintése	
2.	2021.10.16.	Az elvégzendő feladatok átbeszélése	
3.	2021.11.10.	Az előrehaladás ellenőrzése	
4.	2021.12.03.	Beszámoló és prezentáció előkészítése	

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4³ tantárgy követelményét teljesítette, beszámolóra / védésre⁴ bocsátható.

Budapest, 2021.12.06.



intézményi konzulens

¹ Megfelelő aláhúzendő!

² Megfelelő aláhúzendő!

³ Megfelelő aláhúzendő!

⁴ Megfelelő aláhúzendő!



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Neumann János Informatikai Kar

KONZULTÁCIÓS NAPLÓ

Hallgató neve: Neptun kód: Tagozat

Puskás Melánia W7WUGQ Nappali

Telefon: Levelezési cím (pl: lakcím):
06/70-575-1031 2322 Makád, Rákóczi Ferenc utca 35.

Szakdolgozat / Diplomamunka⁵ címe magyarul:

ÉLETTANI FOLYAMATOK PARAMÉTERBECSLÉSE MESTERSÉGES INTELLIGENCIA
HASZNÁLATÁVAL

Szakdolgozat / Diplomamunka⁶ címe angolul:

ESTIMATION OF PHYSIOLOGICAL PROCESSES USING ARTIFICIAL INTELLIGENCE

Intézményi konzulens:

Külső konzulens:

Dr. Drexler Dániel András

Kérjük, hogy az adatokat nyomtatott nagybetűkkel írja!

Alk.	Dátum	Tartalom	Aláírás
1.	2022.02.23.	Az eddigi munka áttekintése, továbbfejlesztési lehetőségek megvitatása.	
2.	2022.03.16.	Előrehaladás ellenőrzése.	
3.	2022.04.06.	Eredmények ellenőrzése, megvitatása.	
4.	2022.05.04.	Prezentáció előkészítése.	

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4⁷ tantárgy követelményét teljesítette, beszámolóra / védésre⁸ bocsátható.

Budapest, 2022.05.10.


.....
intézményi konzulens

⁵ Megfelelő aláhúzendő!

⁶ Megfelelő aláhúzendő!

⁷ Megfelelő aláhúzendő!

⁸ Megfelelő aláhúzendő!



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Neumann János Informatikai Kar

KONZULTÁCIÓS NAPLÓ

Hallgató neve: Neptun kód: Tagozat

Puskás Melánia W7WUGQ Nappali

Telefon: Levelezési cím (pl: lakcím):

06/70-575-1031 2322 Makád, Rákóczi Ferenc utca 35.

Szakdolgozat / Diplomamunka⁹ címe magyarul:

ÉLETTANI FOLYAMATOK PARAMÉTERBECSLÉSE MESTERSÉGES INTELLIGENCIA HASZNÁLATÁVAL

Szakdolgozat / Diplomamunka¹⁰ címe angolul:

ESTIMATION OF PHYSIOLOGICAL PROCESSES USING ARTIFICIAL INTELLIGENCE

Intézményi konzulens:

Külső konzulens:

Dr. Drexler Dániel András

Kérjük, hogy az adatokat nyomtatott nagybetűkkel írja!

Alk.	Dátum	Tartalom	Aláírás
1.	2022.09.09.	Az eddigi munka áttekintése, a féléves tervek egyeztetése.	
2.	2022.10.12.	Az eredmények ellenőrzése.	
3.	2022.11.09.	Dolgozat felépítésének megbeszélése.	
4.	2022.12.05.	Dolgozat és prezentáció ellenőrzése.	

A Konzultációs naplót összesen 4 alkalommal, az egyes konzultációk alkalmával kell láttamoztatni bármelyik konzulenssel.

A hallgató a Szakdolgozat I. / Szakdolgozat II. (BSc) vagy Diplomamunka 1 / Diplomamunka 2 / Diplomamunka 3 / Diplomamunka 4¹¹ tantárgy követelményét teljesítette, beszámolóra / védésre¹² bocsátható.

Budapest, 2022.12.12.

.....
intézményi konzulens

⁹ Megfelelő aláhúzendő!

¹⁰ Megfelelő aláhúzendő!

¹¹ Megfelelő aláhúzendő!

¹² Megfelelő aláhúzendő!

KIVONAT

A jövő orvoslásában egy ígéretes irányvonal lehet a terápiák matematikai és mérnöki módszereken alapuló optimalizálása, mellyel a daganatos betegségek kezelése személyre szabható, ellentétben a ma használt, átlagra kifejlesztett, magas dózisokat alkalmazó kezelésekkel. A terápia optimalizálásához szükség van egy jól működő matematikai modellre, mely képes leírni a gyógyszer hatását és a tumor növekedését, illetve a személyre szabáshoz szükségünk van a páciens egyedi modellparamétereire.

Feladatom egy olyan rendszer létrehozása, mely zajos mérési adatok alapján meghatározza a kezelendő beteg egyedi paramétereit, melyek által megtörténhet a terápia optimalizálása. A paraméterek értékeire általam létrehozott és betanított neurális hálózatok segítségével adok becslést. A hálózatokat nagy tartományban generált paraméterekkel leírt virtuális páciensekből származó zajos mérésekkel tanítottam be. A virtuális páciensek paramétereit valódi mérési eredmények alapján illesztett modellparaméterek felhasználásával hoztam létre. A valódi mérések rendkívül zajosak és a rajtuk lévő zaj függ a mért tumor térfogatától, így feladatom egy olyan transzformáció meghatározása, mellyel a zaj függetleníthető a térfogattól. A szimuláció során minél realisabb zajos mérési adatokat generálok egy általam felépített zajmodell segítségével. A zajmodellt a valódi mérések illesztési eredményei és a paraméterváltozások alapján hoztam létre úgy, hogy a mérések eloszlását és különböző eloszlási paramétereket vizsgáltam. Ehhez rendelkezésemre álltak digitális tolómérővel és MRI-vel mért, állatkísérletekből származó tumortérfogat adatok. A zajmodellt beépítettem a tanító adatokat generáló algoritmusomba, így valósághűbb virtuális mérési adatok generálhatók, ezáltal a neurális hálózatok is pontosabb eredményeket produkálhatnak.

A hálózatok által meghatározott paraméterek bemenetül szolgálhatnak későbbi identifikációs algoritmusoknak. A kutatás végére szeretném elérni, hogy a tumornövekedési modell paramétereit valódi mérések alapján is meghatározhatók legyenek. Kutatásom az Innovációs és Technológiai Minisztérium ÚNKP-22-2 kódszámú Új Nemzeti Kiválósági Programjának támogatásával készült.

ABSTRACT

In future medicine, a promising direction may be the optimization of therapies based on mathematical and engineering methods, with which the treatment of cancer diseases can be customized, in contrast to the treatments used today, developed for the average, using high doses. In order to optimize therapy, you need a well-functioning mathematical model that can describe the effect of the drug and tumor growth, and for personalization, we need the unique model parameters of patients.

My task is to create a system that, based on noisy measurement data, determines the unique parameters of the patient to be treated, which can be used to optimize the therapy. I estimate the values of the parameters with the help of neural networks that I have created and trained. I trained the networks with noisy measurements from virtual patients described with parameters generated in a large range. I created the parameters of the virtual patients using model parameters fitted based on real measurement results. Real measurements are extremely noisy and the noise on them depends on the volume of the measured tumor, so my task is to define a transformation that can be used to make the noise independent of the volume. During the simulation, I generate noise measurement data as realistic as possible using a noise model I have built. I created the noise model based on the matching results of the real measurements and the parameter changes by examining the distribution of the measurements and different distribution parameters. For this, I had tumor volume data from animal experiments measured with a digital caliper and MRI. I incorporated the noise model into my training data generation algorithm, so more realistic virtual measurement data can be generated, and thus the neural networks can produce more accurate results.

The parameters defined by the networks can serve as input for later identification algorithms. By the end of the research, I would like to achieve that the parameters of the tumor growth model can be determined based on real measurements. My research was supported by the ÚNKP-21-2 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

Tartalomjegyzék

1. Bevezetés	11
2. Elméleti áttekintés	14
3. Megvalósítás	18
3.1. A zajos mérések simítása és a mérési zaj becslése MRI adatok nélkül	18
3.2. Valószínűségi eloszlások illesztése és zajmodell felépítése MRI adatok nélkül	25
3.3. Digitális tolómérő mérési zajának vizsgálata MRI mérések alapján	36
3.4. Zaj függetlenítéséhez felhasználható transzformáció keresése valószínűségi eloszlások illesztésével és a zajmodell felépítése MRI alapján	40
3.5. Tanító adatok generálása a felépített zajmodell beépítésével	49
4. Eredmények és kiértékelésük	55
4.1. Virtuális páciensek szimulált zajos mérései	55
4.2. Neurális hálózatok teljesítményének kiértékelése az MRI adatok nélkül illesztett zajmodellel generált adatok alapján	57
4.3. Neurális hálózatok teljesítményének kiértékelése az MRI adatokkal illesztett zajmodellel generált adatok alapján	63
5. Konklúzió	67
Irodalomjegyzék	69
Ábrák jegyzéke	72

1. fejezet

Bevezetés

A daganatok okozta megbetegedések jelentik a második leggyakoribb halálokot világszerte. A rák visszaszorítása és korszerű kezelése súlyos népegészségügyi problémát jelent. A diagnosztizált rákbetegek száma folyamatosan emelkedik, globálisan sok rákos beteg nem kaphat időben megfelelő diagnózist és kezelést. A GCO (Global Cancer Observatory), az Egészségügyi Világszervezet interaktív webes platformja alapján 2020-ban majdnem 10 millió ember halt bele rákos megbetegedésébe és több mint 19 millió új beteget diagnosztizáltak. Az előrejelzések alapján ezek a számok a jövőben tovább fognak emelkedni. A rákos betegek kezelése több ismert módszerrel is történhet, például kemoterápiával vagy sugárkezeléssel. Ezek átlagra kifejlesztett, empirikus úton megtervezett terápiák, melyek káros hatásairól is ismertek. Az ilyen terápiáknak a legnagyobb hiányossága az, hogy nem veszik figyelembe a páciensek egyedi sajátosságait. A kutatás, melybe lehetőségem adódott becsatlakozni, egy egyedi, személyre szabott és optimalizált terápiát ígér, melyben kisebb dózisokat alkalmazunk, így a kezelések súlyos mellékhatásai is csökkenthetők. Az alacsony dózisok alkalmazásával csökkenhet az esélye a rezisztencia kialakulásának, illetve a kis dózisok mellett szól a kezelési költségek csökkenése is, mely által szélesebb körben elérhetőbbé válhat a terápia.

Kutatásunk alapja a számítógépes terápiaoptimalizálás, mely a jövő orvoslásának az egyik ígéretes eszköze lehet. Az optimalizált terápiához szükség van egy jól működő matematikai modellre, mely leírja a gyógyszer hatását és a tumor növekedésének dinamikáját. Az optimalizált terápia akkor lesz személyre szabott, ha a matematikai modell paramétereit a kezelendő páciens egyedi paraméterei határozzák meg, tehát a terápia személyre szabáshoz szükségesek a páciensek egyedei paraméterei. A kutatás lényege, hogy rákos egereket kezelünk kemoterápiás szerrel, majd a mérési eredmények alapján modellt illesztünk. A modellillesztés alapján generálunk egy egyedi terápiát és ezt teszteljük. A modell illesztése egy kritikus lépés, hiszen ez határozza meg a modell alapú terápia hatékonyságát. A valódi kísérletekből kevés mérés áll rendelkezésünkre és az illesztési problémának sok megoldása lehet, emiatt fontos a megfelelő kezdeti becslés megadása. A terápia hatékonyságát tehát a páciensek egyedi modellparamétereinek a becslése határozza meg. Ha konstans paraméterértékeink vannak és egy specifikus célcsoportra koncentrálunk, tehát ennek a célcsoportnak a paramétereit szeretnénk megbecsülni, akkor egy igen hatékony és gyakran használatos módszer a kevert hatású modellek alkalmazása [1, 2]. A kutatásunk során a páciensek kísérleti egerek, melyek ikrek, így genetikai kódjuk biztosan azonos, tehát specifikus célcsoportnak tekinthetők. Paraméterértékeik között elvárásaink szerint kicsik a különbségek. A kevert hatású modell a paraméterek értékének várható értékét becsüli és az egyedek kö-

zötti különbségeket zajként modellezi. A kevert hatású modell (és a többi ehhez hasonló statisztika módszer) előnye az, hogy figyelembe veszik, hogy a különböző egyedek paraméterértékeinek hasonlóknak kellene lennie. Egy nagy hátrányuk viszont, hogy szükségük van a paraméterek értékének egy kezdeti becslésére. Ha ez a becslés nem megfelelő, akkor a kutatás későbbi fázisaiban használatos kereső algoritmusok nem működnek. Ilyenkor vagy nem konvergálnak, vagy egy lokális minimumba beragadnak. Munkám során ezt a problémát neurális hálózatokkal oldom meg, melyeket arra használok, hogy megbecsüljék a modellparaméterek kezdeti értékét. Ezeket a kezdeti értékeket a további szakaszokban alkalmazott különböző statisztikai alapú identifikációs algoritmusok fel tudják használni, mint kezdeti paraméterértékek. Ez egy hiánypótló megoldás, mert az identifikációs algoritmusok mind lokális kereső algoritmusok, amelyeknél a kezdeti becslést vagy manuálisan adják meg, vagy egyéb automatizált módszerekkel (például genetikus algoritmussal) általában véletlenszerűen keresik. Kutatásunk célja egy olyan rendszer megalkotása, mely mérési adatok alapján meghatározza a kezelendő beteg egyedi paramétereit, amelyek alapján a terápia optimalizálása megtörténhet.

A diplomamunkámban a zaj modellezésére koncentrálok. A valódi mérések rendkívül zajosak és az előzetes munkámból már kiderült, hogy a rajtuk lévő zaj nagyban függ a mért tumor térfogattól, így az egyik feladatom egy olyan transzformáció meghatározása, mellyel a zaj függetleníthető a térfogattól. A szimuláció során minél realisabb zajos mérési adatokat generálok az általam felépített zajmodell egy továbbfejlesztett verziójának segítségével. Ez a zajmodellt szintén a valódi mérések illesztési eredményei és a paraméterváltozások alapján hoztam létre úgy, hogy a mérések valószínűségi eloszlását és különböző eloszlási paramétereket vizsgáltam.

A diplomamunkámban alapigazságnak először nem a simított értékeket vettem, hanem rendelkezésemre álltak MRI-vel mért, állatkísérletekből származó tumor térfogatok és a hozzájuk tartozó digitális tolómérővel mért térfogatok is, így az MRI mérések szolgáltak alapigazsággént. Az MRI mérések és a hozzájuk tartozó digitális tolómérővel mért mérések egy 2013-as adathalmazból származnak, amely a Richter Gedeon Nyrt. közreműködésével valósult meg. A jelenlegi kísérleteinkhez nem áll rendelkezésre kisállat MRI, így a digitális tolómérővel mért tumor térfogatokon lévő zaj modellezése a jelenlegi kutatásainkhoz elengedhetetlen.

Mivel a méréseink száma igen kevés, így a hozzájuk tartozó zaj is rendkívül alul van mintavételezve ahhoz, hogy a különböző zajszűrési technikák alkalmazhatók legyenek a feladat megoldásához. Emiatt a valószínűségi eloszlások illesztésével, majd felhasználásával egy olyan megfelelő transzformációt kerestem, mellyel függetleníjük a zajt a tumor térfogattól, illetve egy olyan simítási paramétert is keresünk, mellyel a digitális tolómérővel mért térfogatok simított verziója felhasználható alapigazságnak, tehát a legjobban hasonlít az MRI-vel mért térfogatok jellegére. A dolgozatom fő eredménye tehát a 2013-as adathalmazt felhasználva egy összetettebb zajmodell felépítése, amelyet utána a jelenleg is futó kísérleti eredményeken alkalmazva is megvizsgállok. Ezután a kísérleti jelenlegi eredményeink alapján létrehozott virtuális páciensek tumor térfogat méréseire a zajmodell alapján zajt generálok. Tehát, a zajmodellt beépítem a tanító adatokat generáló algoritmusomba, így még valóságosabb virtuális mérési adatok generálhatók, ezáltal a neurális hálózatok is sokkal pontosabb eredményeket produkálnak. A dolgozat végén a neurális hálózatok teljesítményét teszt adatokon és valódi kísérleti adatokon is kiértékelem.

A fentiek részletezése a következőképpen kerül bemutatásra. Az 2. fejezetben a mun-

kám során felhasznált matematikai modellt mutatom be, mely leírja a tumor és az alkalmazott kemoterápiás szer dinamikáját egy differenciálegyenlet-rendszer segítségével. A modellben lévő lineáris és nemlineáris tagok felelnek a gyógyszer szervezetre kifejtett hatásáért és a gyógyszer kimosódásáért. Ezt a dinamikai modellt használok fel a virtuális páciensek generálásához és az ezekhez szimulált tumortérfogat mérésekhez, melyekre majd az általam generált zaj kerül.

A 3. fejezetben részletezem a valódi kísérletekből származó méréseket. Megkeresem a megfelelő transzformációt MRI adatok nélkül és MRI adatok alapján is, mely által függetleníthető a zaj a tumor méretétől, majd a valószínűségi eloszlások illesztését részletezem. Felépítem a zajmodellt és bemutatom az alkalmazását egy választott valódi kísérleti egér mérésein. A mérési zaj modellezése után a tanító adatok generálását mutatom be.

Az egyedi paraméterek becslését neurális hálózatokkal végzem, amiket virtuális páciensekből nyert adatok alapján tanítok be. A hálózatok bemenetei a mérések, tehát a teljes tumor térfogatok, illetve a mérésekhez tartozó dózisok, kimenetei pedig a páciens egyedi paramétereinek a becsült értékei. A virtuális pácienseket paraméterhalmazok írják le, amik valódi mérési adatok alapján lettek létrehozva paraméter-identifikáció segítségével. A virtuális páciensek felhasználásával tanító adatokat generáltam egy általam készített algoritmus-sal, majd a teljes tumor térfogatokra a felépített zajmodell segítségével zajt generáltam. A páciensek paraméterértékeit bizonyos időközönként megváltoztattam, ezután a tanító adatokat lenormálva a neurális hálózatokat betanítottam. Szintén a 3. fejezetben részletezem a használt algoritmust és a tanító adatgenerálás folyamatát, továbbá azt, hogy hogyan generálok mérési zajt a tanító adatokra.

A 4. fejezetben bemutatom az általam szimulált zajos méréseket, ezt néhány virtuális páciens generált mérési hibáján fogom részletezni. Ezután az elkészült hálózatok teljesítményét fogom vizsgálni, melyek feladata, hogy a generált zajos adatok alapján becslést adjanak az adott páciens időben változó paramétereinek az értékére. A paraméterbecslést és az általánosítóképességet megvizsgálom a tanult virtuális páciensek, illetve a tanítás alatt nem látott páciensek adatainak felhasználásával. Továbbá tesztelem a hálózatokat valódi kísérleti méréseken is. Összehasonlítom továbbá a teljesítményt az előző dolgozatomban szereplő egyenletes eloszlású zajos adatokon tanult neurális hálózat teljesítményével.

2. fejezet

Elméleti áttekintés

A szakirodalmakban többféle modell is megtalálható, amelyek leírják a tumor növekedését különböző feltételek mellett, illetve különböző folyamatok figyelembevételével. A tumor kezelés nélküli természetes növekedését általában lineáris, exponenciális és logisztikus modellekkel [3, 4] írják le, illetve elterjedt az exponenciális és lineáris modellek kombinációja is [5]. A lineáris tumornövekedés állandó növekedési sebességet feltételez. Alkalmazták már például áttétes vesesejt karcinóma természetes tumornövekedésének leírására [6]. Az exponenciális növekedés feltételezi, hogy a daganat növekedési sebessége arányos a daganat súlyával. Ezt széles körben használják arra, hogy leírják a természetes tumornövekedést a tumornövekedésgátló modellekben [7, 8]. A lineáris és exponenciális növekedési modellek kombinációját kibővítették egy zsugorodási kifejezéssel, amely a természetes daganat-elhalást írja le. Például egy ilyen zsugorodási taggal rendelkező modellt alkalmaztak az előrehaladott stádiumú daganatos betegek természetes tumornövekedésének leírására [9]. Gyakori a Gompertz-modell használata is, mely széles körben használatos a biológia számos területén. Gyakran használják az állatok és növények növekedésének, valamint a baktériumok és rákos sejtek számának vagy térfogatának leírására. A Gompertz-modell alapvető jellemzője, hogy modellezi a tumornövekedési ráták exponenciális csökkenését, tehát azt hogy a tumorsejtek növekedési üteme az idő függvényében csökken [10, 11].

A leggyakrabban használt tumormodellek azonban a Hahnfeldt modellt [12] veszik alapul, illetve gyakran annak valamilyen módosított változatát [13, 14, 15]. A Hahnfeldt modell nemlineáris egyenleteket használ fel arra, hogy meghatározza a tumor és az érrendszer dinamikáját, de ezt anélkül, hogy figyelembe venné a gyógyszerek szervezeten belüli mozgását (farmakokinetika). Ellentétben a Hahnfeldt modellel és annak különböző változataival, a következőkben bemutatott modell egy viszonylag egyszerű felépítésű modell, melynek differenciálegyenletei lineáris kifejezéseket és egy nemlineáris kifejezést tartalmaz és az egyenletek tartalmazzák a farmakokinetikát, tehát a gyógyszer kimosódását, illetve a farmakodinamikát, amely a gyógyszer szaturált hatását írja le [16].

Ahhoz, hogy a terápiát optimalizálni tudjuk egy olyan matematikai modellre van szükségünk, amely leírja a tumor dinamikáját és a gyógyszer-szervezeten belüli mozgását is. A valódi kísérleti mérésekből kevés áll rendelkezésünkre a neurális hálózatok betanításához, így a valódi páciensek alapján virtuális pacienseket generálunk. Ezután szimuláció segítségével leszimuláljuk a hozzájuk tartozó tumor térfogat méréseket (a páciensek generálását a 3.5. alfejezetben részletezzük). Ezzel az a célunk, hogy kellőképpen megsokszorozzuk a tanító adatokat. A neurális hálózatoknak az a feladata, hogy meghatározzák a modell paramétereinek becsült értékét, adott bemeneti és kimeneti párok alapján, tehát a hálózat be-

tanítását felügyelt tanítási forma szerint végezzük. A hálózat által becsült paraméterértékek határozzák meg a személyre szabott terápia hatékonyságát.

A matematikai modell, melyet felhasználunk, képes leírni a tumor dinamikáját és a gyógyszer mozgását. Ez egy négy állapotváltozós differenciálegyenlet-rendszerrel leírt modell, mely a következő [2]:

$$\dot{x}_1 = (a - n)x_1 - b \frac{x_1 x_3}{ED_{50} + x_3}, \quad (2.1)$$

$$\dot{x}_2 = nx_1 + b \frac{x_1 x_3}{ED_{50} + x_3} - wx_2, \quad (2.2)$$

$$\dot{x}_3 = -(c + k_1)x_3 + k_2 x_4 + u, \quad (2.3)$$

$$\dot{x}_4 = k_1 x_3 - k_2 x_4. \quad (2.4)$$

Az első egyenlet (2.1) az élő tumor dinamikáját írja le, ahol x_1 [mm³] állapotváltozó az élő tumorsejt térfogatának időfüggvénye. Az második egyenlet (2.2) a halott tumor dinamikáját határozza meg, x_2 [mm³] pedig a halott tumorsejt térfogatának időfüggvényét jelöli. A (2.3) egyenletben x_3 [ng/ml] a központi rekesz (kompartment) gyógyszer szintjét jelöli, mely esetünkben a vérben lévő gyógyszer szérumszintjének időfüggvénye és ez az egyenlet írja le a gyógyszer dinamikáját a vérben. A (2.4) egyenlet pedig a gyógyszer dinamikáját határozza meg a szövetekben, ahol x_4 [ng/ml] állapotváltozó a szövetekben lévő gyógyszer szérumszintjének időfüggvényét jelöli. A rendszer bemenete u , mely az injekciós ráta [mg/kg/nap], ezt fecskendezzük a központi rekeszbe. A rendszer kimenetét az élő tumorsejt térfogat és a halott tumorsejt térfogat összege határozza meg és y jelöli:

$$y = x_1 + x_2. \quad (2.5)$$

A valódi kísérletek során y a mért változó, tehát ez a változó írja le a kemoterápiás szerrel kezelt egerek teljes tumor térfogatát. A valóságban az állatházban dolgozó kollégák digitális tolómérővel mérik bőrön keresztül a tumor szélességét és hosszúságát, és egy közelítő képlettel számítjuk ezekből a teljes tumor térfogatokat. A mért értékek mérési zajjal rendelkeznek és a dolgozatomban ezt a mérési zajt fogom modellezni a valósághűbb virtuális mérések szimulálásához. A mérési hiba becslésére MRI méréseket veszek alapigazsággként (3. fejezetben részletezzük). A modell további paramétereinek leírásait, illetve mértékegységeit a 2.1 táblázat foglalja össze.

A modell (2.1)-(2.4) feltételezi, hogy az u bemenet folytonos az időben. A kísérletek alatt azonban a bemenetek az injekciók, melyek adott napon és adott dózisban kerülnek befecskendezésre. Ezt úgy modellezzük, hogy a virtuális páciensek és méréseik szimulálása során az u bemenetet elhagyjuk az egyenletből és x_3 gyógyszer szint hirtelen megváltozását tekintjük bemenetnek. A megvalósítás során minden egyes t_κ időponthoz hozzárendelünk egy d_κ dózist, ahol $\kappa = 1, 2, \dots$, és K az injekciók száma. A bemenet tehát impulzív, mivel a rendszer időállandójához képest nagyon rövid idő alatt megy végbe a gyógyszer szint megváltozása az injekció hatására. Ezért tekinthetjük úgy, mintha x_3 -ban egy szakadás lenne t_κ időpontban, melynek mértéke d_κ [2], azaz

$$x_3(t_\kappa^+) = x_3(t_\kappa^-) + d_\kappa. \quad (2.6)$$

A differenciálegyenletekben szereplő kifejezések reakciókinetikai analógiával szemléltethetők [1]. Legyen X_1 egy fiktív anyagfajta, ami az élő tumorsejtek térfogatát reprezentál-

Paraméter	Leírás	Mértékegység
a	Tumor növekedési ráta	1/nap
b	Inhibíciós ráta	1/nap
c	Módosított klírensz	1/nap
ED_{50}	Medián effektív dózis	ng/ml
k_1	A gyógyszer áramlási sebességi együtthatója a centrális kompartmentből a perifériás kompartmentbe	1/nap
k_2	A gyógyszer áramlási sebességi együtthatója a perifériás kompartmentből a centrális kompartmentbe	1/nap
n	Nekrotikus ráta	1/nap
w	Kimosódási ráta	1/nap

2.1. táblázat. A modell paramétereinek leírása és mértékegysége

ja, X_2 a halott tumorsejtek térfogatát, X_3 a centrális kompartmentben (vérben) lévő gyógyszer szérumszintjét és X_4 a perifériás kompartmentben (szövetekben) lévő gyógyszer szérumszintjét reprezentálja. Ezeket felhasználva a modellben szereplő élettani folyamatokat a következő reakciókinetikai egyenletek írják le [2, 17]:

$$X_1 \xrightarrow{a} 2 X_1, \quad (2.7)$$

$$X_1 \xrightarrow{n} X_2, \quad (2.8)$$

$$X_2 \xrightarrow{w} O, \quad (2.9)$$

$$X_1 + X_3 \xrightarrow{b, ED_{50}} X_2, \quad (2.10)$$

$$X_3 \xrightarrow{c} O, \quad (2.11)$$

$$X_3 \xrightleftharpoons[k_2]{k_1} X_4. \quad (2.12)$$

A tumorsejtek osztódását (2.7) határozza meg (proliferáció), ahol a a tumor növekedési ráta, mely egy sebességi együttható. A reakciólépést leíró kifejezés az $\dot{x}_1 = ax_1$, amit tömeg-hatás kinetikát [18] alkalmazva kaphatunk meg.

A (2.8) írja le a tumorsejtek elhalását (nekrózist), sebességi együtthatója az n nekrotikus ráta. Szintén a tömeg-hatás kinetikát felhasználva ez az egyenlet módosítja az osztódó tumorsejtek térfogatát: $\dot{x}_1 = -nx_1$; és az elhalt tumorsejtek térfogatát: $\dot{x}_2 = nx_1$. Ez a tag azt a nekrozist írja le, ami nem függ a kezeléstől. A halott tumorsejtek egy idő után kimosódnak a szervezetből. Ezt a lépést modellezi a (2.9) reakció, w sebességi együtthatóval, mely a kimosódási ráta. A kimosódás sebessége tehát a $-wx_2$ kifejezéssel írható le.

Egy bizonyos idő után a gyógyszer hatására az élő tumorsejtek halott tumorsejteké válnak. Ezt a (2.10) reakciólépés modellezi. A gyógyszer hatásának leírásakor a medián effektív dózist (ED_{50}) használjuk, meghatározva a $x_1 x_3 / (ED_{50} + x_3)$ kifejezést, ami a tumor térfogatban lineáris, a gyógyszer szintben azonban nem (erről még szó lesz a következőkben). b sebességi együttható a gyógyszer maximális hatását jelöli. Ezek segítségével megkaphatjuk a gyógyszer tumor térfogatra gyakorolt hatását. Az egyenlet ezek által az élő tumorsejtek és az elhalt tumorsejtek térfogatára is hatást gyakorol. Az élő tumorsejtek ezek alapján leírhatók a következő kifejezéssel:

$$\dot{x}_1 = -b \frac{x_1 x_3}{(ED_{50} + x_3)}, \quad (2.13)$$

illetve az elhalt tumorsejtek is leírhatók a következőképpen:

$$\dot{x}_2 = b \frac{x_1 x_3}{(ED_{50} + x_3)}. \quad (2.14)$$

Korábban már említésre került, hogy a gyógyszerkonzentráció függvényében az $x_1 x_3 / (ED_{50} + x_3)$ kifejezés a modellben egy nemlineáris tag. Ez írja le, hogy a gyógyszer hatása szaturált, mivel egy adott határ után a gyógyszer szérumszintjének növelése csak kis mértékben növeli a gyógyszer hatékonyságát. Ezt nevezik farmakodinamikának (helyenként farmakodinámiának). A medián effektív dózis, melyet az ED_{50} paraméter jelöl, azt a gyógyszerkoncentrációs szintet adja meg, ahol a gyógyszer hatása eléri az 50%-ot. A gyógyszer kiürülését a (2.11) egyenlet határozza meg. Szintén egy sebességi együttható a c -vel jelölt paraméter (módosított klírens), amely szintén egy sebességi együttható. Ez a reakció határozza meg azt, hogy a teljes szervezetből $c x_3$ sebességgel ürül ki a gyógyszer. A (2.12) reakció lépés egy kétkompartment modellt ír le, mely meghatározza, hogy milyen gyorsan áramlik egyik kompartmentből a másikba a gyógyszer. Az egyik kompartment a centrális kompartment, a másik pedig a perifériás kompartment. Esetünkben centrális kompartment a vért, a perifériás pedig a szöveteket jelenti. A vérből a szövetekbe áramló gyógyszer sebességét k_1 sebességi együttható határozza meg, tehát az áramlás sebessége $k_1 x_3$. Visszafelé az áramlás sebességi együtthatója k_2 , így a szövetekből a vérbe $k_2 x_4$ sebességgel áramlik át a gyógyszer. A (2.11) és (2.12) egyenletek írják le együtt a gyógyszer farmakokinetikáját.

A felhasznált eloszlásokat és a neurális hálózatok architektúrája megtalálható az előző TDK munkámban [19].

3. fejezet

Megvalósítás

3.1. A zajos mérések simítása és a mérési zaj becslése MRI adatok nélkül

A valódi kísérletekből összesen 49 egér (az egerekre a dolgozatban innentől a sorszámukkal hivatkozunk) zajos tumor térfogat mérései állnak rendelkezésünkre. Az első lépésünk, hogy becslést adunk a zaj értékére az összes mérési pontban. Ehhez szükségünk van a zajmentes mérésekre, viszont ezek nem állnak rendelkezésünkre, így a zajos tumortérfogat méréseket simítjuk, annak érdekében, hogy kapjunk egy becslést a zajmentes mérési értékekre. A simítás után a zaj értékét az egyes pontokban könnyedén megbecsülhetjük a simított és az eredeti görbe különbségével. Ebben a fejezetben a 16-os, 13-as és a 8-as egér méréseinek feldolgozását mutatjuk be részletesebben.

A simítást MatlabR2021a környezetben végezzük, egy általunk készített egyszerű algoritmussal, mely a Matlab fit függvényét (beépített görbe és felület illesztő függvény) használja fel. A fit függvény a bemenetén megkapja a mérési időpontokat és a mért tumor-térfogatokat, majd ráilleszt egy, az adatokhoz jól illeszkedő görbét.

Többféle típusú görbét lehet illeszteni a függvénnyel. Esetünkben a simító spline görbe (smoothing spline curve) típus a legmegfelelőbb, így azt választottuk. Mivel egyszeri simítás után nem minden esetben érhető el a kívánt illeszkedés, így a simító algoritmust többször lefuttatuk. Azt, hogy hányszor futott le az egyes egerek esetében az algoritmus, egy statisztikai paraméterrel határoztuk meg. A fit függvény egy visszatérési értéke a simított görbe értékei mellett egy úgynevezett statisztikai illeszkedési jóságot mérő mutatók értékeit (melyek a goodness-of-fit paraméterek) tároló vektor. Ezen értékek közül, mi az illeszkedés jóságát az R^2 (R-Square) paraméterrel mértük, mely esetünkben a simítás jóságát mutatja. Ennek a paraméternek a maximális értéke egy.

Ha R^2 elérte az egyet, akkor a fit függvény nem tudja tovább simítani a görbét. Az R^2 a statisztikában az eredeti értékek és a becsült értékek közötti korreláció négyzetét adja meg. Ha n a zajos mérések számát jelöli, y_i az i -edik zajos mérést és $\hat{y}_i$ jelöli az i -edik zajos mérés zaj nélküli becsült értékét, továbbá $\bar{y}$ a zajos mérések átlaga, akkor az R^2 a következőképpen definiálható az i -edik mérésre:

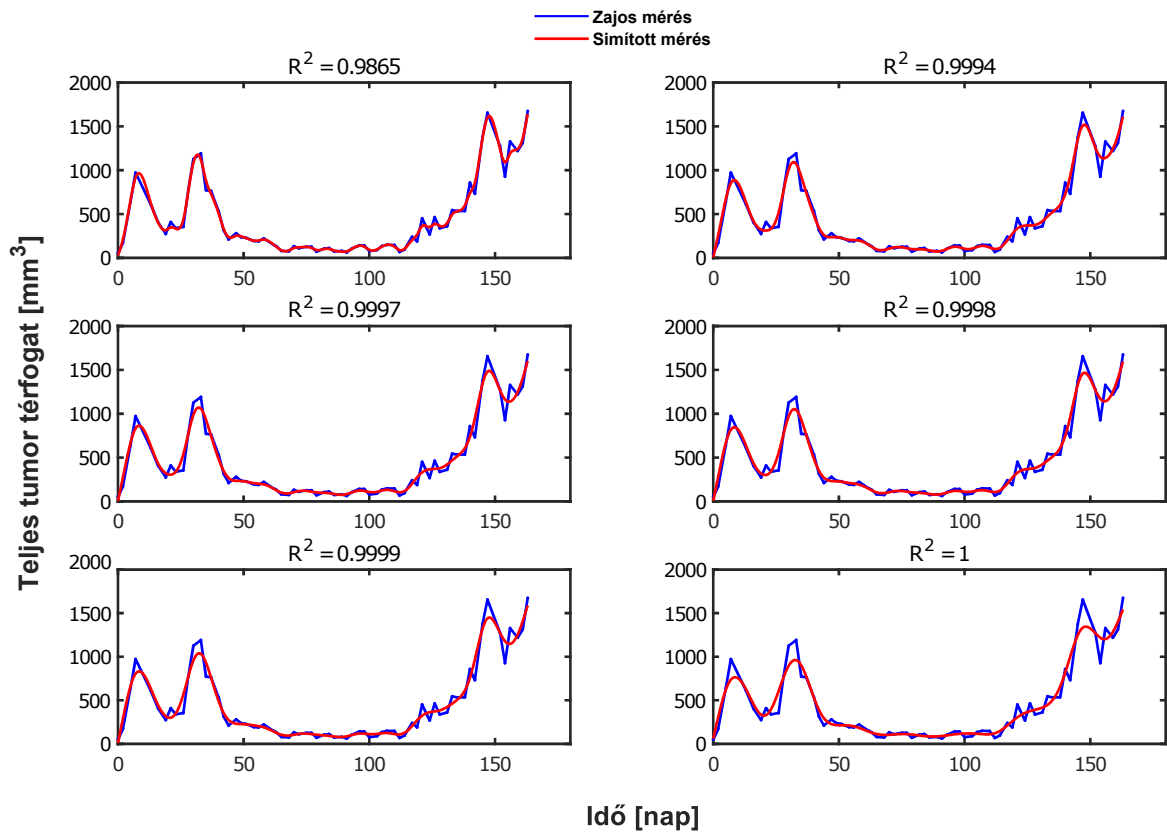
$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{SSR}}{\text{SST}}, \quad (3.1)$$

ahol SSR a regresszió négyzetösszege, SST pedig teljes négyzetösszeg. Az R^2 értéke min-

dig nulla és egy között van. Az egyhez közelebbi érték azt jelzi, hogy az illesztés az átlagra vonatkozó adatok teljes eltérésének közel 100%-át írja le, tehát az illesztett görbe már nem változhat. A simítás során az algoritmus minden iterációban ellenőrzi az R^2 értékét, és amíg nem éri el az egyet, addig tovább simítja a görbét.

A fit függvényt először a zajos mérésre hívjuk meg, majd ha nem értük el a megfelelő R^2 értéket, akkor az adott iterációban simított görbét simítjuk tovább, hiszen minden görbét egyszer tudunk simítani a fit függvény segítségével, így ugyanannak a simítása több iteráción át sem hozna változást.

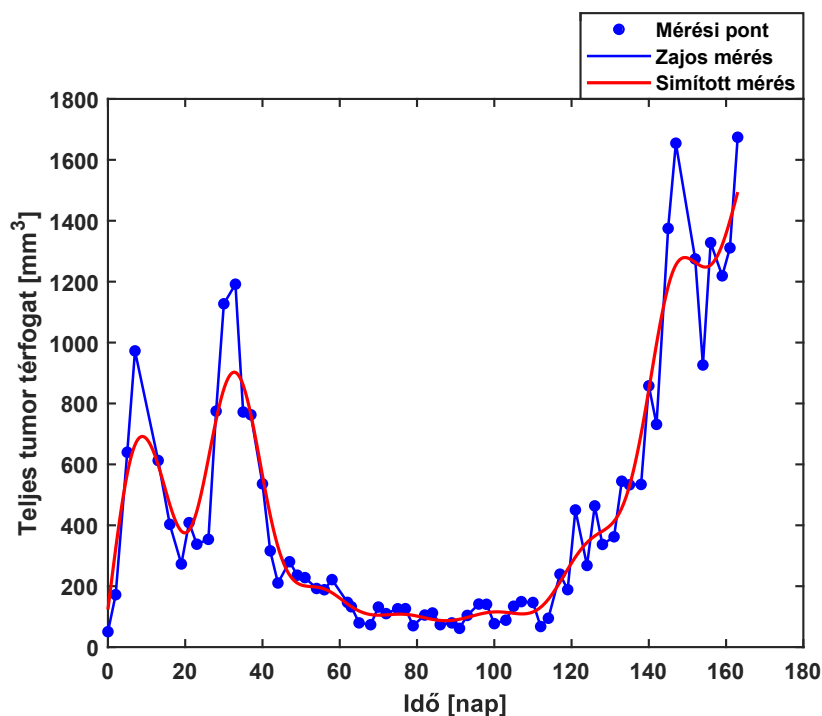
Az algoritmus működését jól szemlélteti a 3.1. ábra, melyen a 16-os egér zajos mérésének simítása látható hat iteráción keresztül. A hat ábra hat iteráció eredményét mutatja, minden diagram tetején megjelenítve az aktuális iterációban az R^2 értékét. A függőleges tengely minden ábrán a teljes tumor térfogata mm^3 -ben, a vízszintes tengely pedig a mérési időpontokat adja meg napokban mérve.



3.1. ábra. A 16-os egér zajos méréseinek (kék görbe) simítási eredményei (piros görbe) hat iteráción keresztül, a simítás jósága a diagramok tetején az R^2 értékével van megadva.

A bal felső ábrán az első simítás látható, ahol a kék görbe az eredeti zajos mérést, a piros görbe pedig az első simított mérést ábrázolja. Az R^2 értéke 0.9865 és a simított görbe szinte teljesen illeszkedik a kék görbére. Mivel a simított görbénk még nagyon követi a zajos görbét, így tovább simítva azt, megkaptuk a jobb felső sarokban lévő mérési eredményt. Itt már az 50. nap előtti csúcsnál és a 150. napnál lévő csúcsnál látható, hogy piros görbe laposodik és az R^2 értéke is nőtt. Az eljárást tovább futtatva további négy iteráción keresztül simított az algoritmus, míg elérte a végleges formáját a piros, tehát a zajmentes mérési értékeket becsülő görbe.

Az utolsó iterációban, melynek eredménye a jobb alsó sarokban látható, az R^2 értéke elérte az egyet, tehát az algoritmus leállt, mivel a további simítások már nem hoznának ennél jobb eredményt. Összehasonlítva az első és utolsó diagramot, a legszembetűnőbb különbség a csúcsoknál van, de az apróbb zajok is jóval simábbak lettek az 50. és 100. nap között. A 16-os egér zajos mérésének simítási végeredményét külön kiemelve a 3.2. ábra mutatja, ahol a kék teli körök a valódi mérési pontokat jelölik.



3.2. ábra. A 16-os egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).

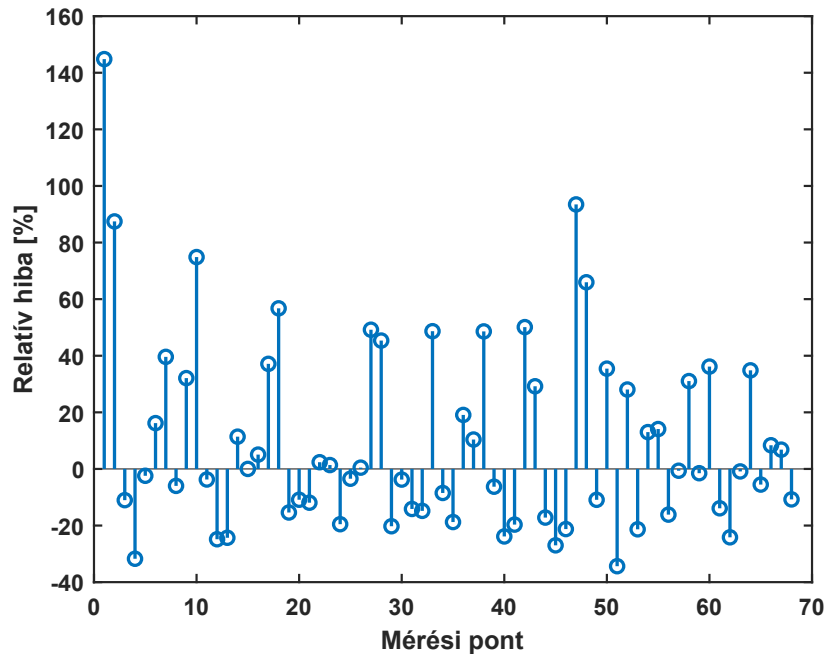
Ezután a következő lépés a simított és valódi mérés közötti hiba számítása. Relatív hibát, tehát százalékos hibát számolunk a két görbe között az adott mérési pontokban. A százalékos relatív hiba a következő összefüggés alapján számítható:

$$\text{hiba} = \frac{\text{becsült zaj nélküli mérési pontok} - \text{eredeti zajos mérési pontok}}{\text{eredeti zajos mérési pontok}} \cdot 100. \quad (3.2)$$

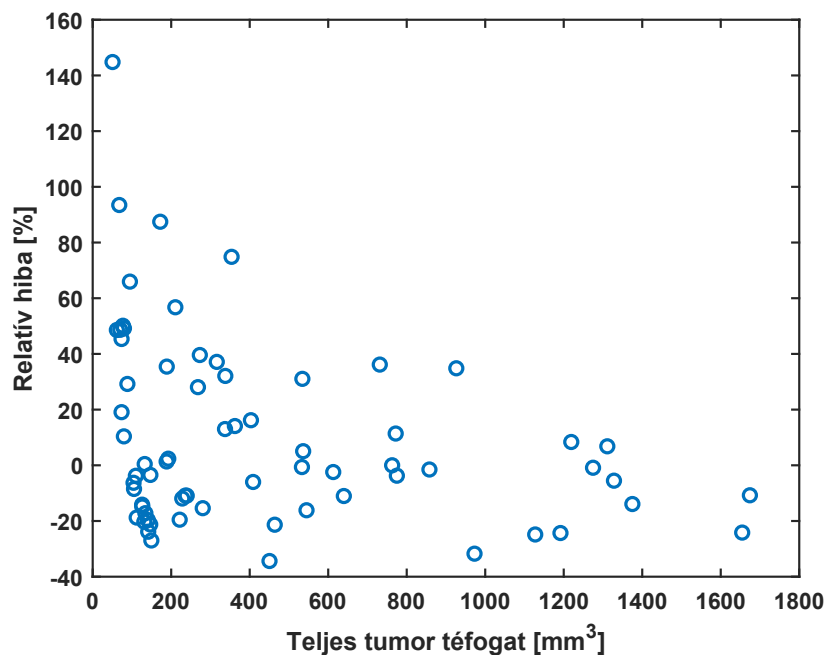
A 3.3. ábra a 16-os egér mérései közötti relatív hibáit mutatja az egyes mérési pontokban. A függőleges tengely a hiba értéke százalékban (üres kék körök jelölik), a vízszintes tengelyen pedig a mérési pont látható. Ezen az ábrán feltűnő az, hogy a pozitív értékű hibák nagysága eltér a negatív értékű hibák nagyságától.

Mivel nekünk fontos a zajmodell felépítése szempontjából, hogy tudjuk azt, hogy az aktuális mérésre mekkora zajt kell majd generálni, így ezeket a relatív hibaértékeket megvizsgáltuk egy szóródás diagramon is. A 3.4. ábra mutatja a függőleges tengelyen a relatív hiba értékét, a vízszintes tengelyen pedig a tumor térfogatok értékét. Összefoglalva tehát, a szóródási diagram tudja nekünk megmondani azt, hogy az egyes tumorméretekhez mekkora nagyságrendű hiba tartozik.

A diagram egyértelműen mutatja, hogy a kisebb tumortérfogatokhoz bizonyos esetekben nagyobb hiba tartozhat. A szóródási diagram egy még hasznosabb tulajdonsága az,



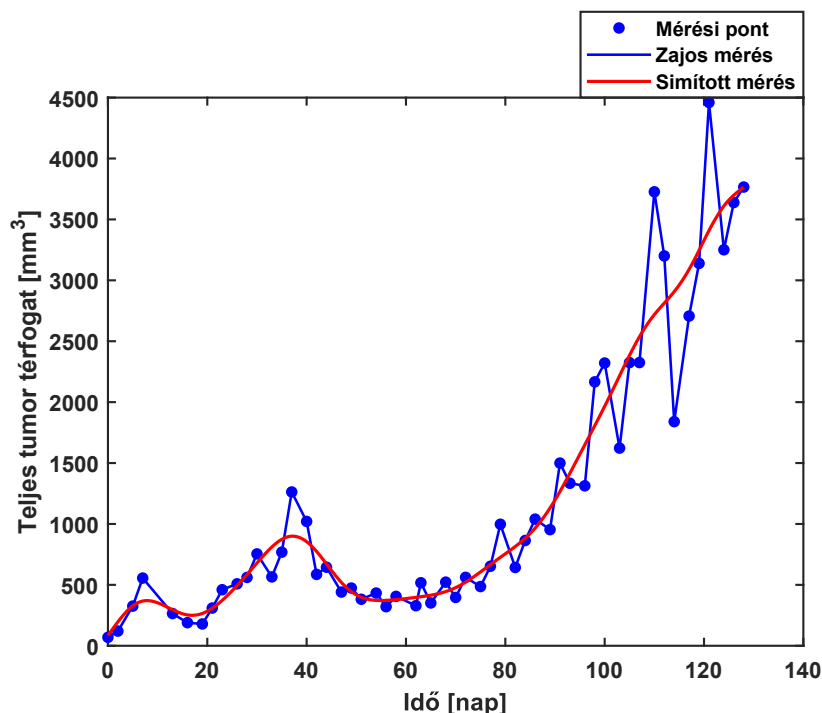
3.3. ábra. A 16-os egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök) minden egyes mérési pontban.



3.4. ábra. A 16-os egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.

hogy nem csak a kék körök értéke számít, hanem a helyzetük is. Tehát ez alapján elmondható, hogy a kisebb tumortérfogatokat nehezebb mérni, hiszen arányaiban több pont van a diagram bal oldalán a kis tumortérfogatoknál. Ez a valódi kísérletek során is igaz, hiszen egy kisebb tumort megmérni nehezebb az általunk használt tolómérővel. A hibák eloszlása ez alapján biztosan nem egyenletes. A hibákról elmondható továbbá az ábra alapján,

hogy szinte exponenciálisan csökkennek a tumor méretének növekedésével a 16-os egér méréseinél.



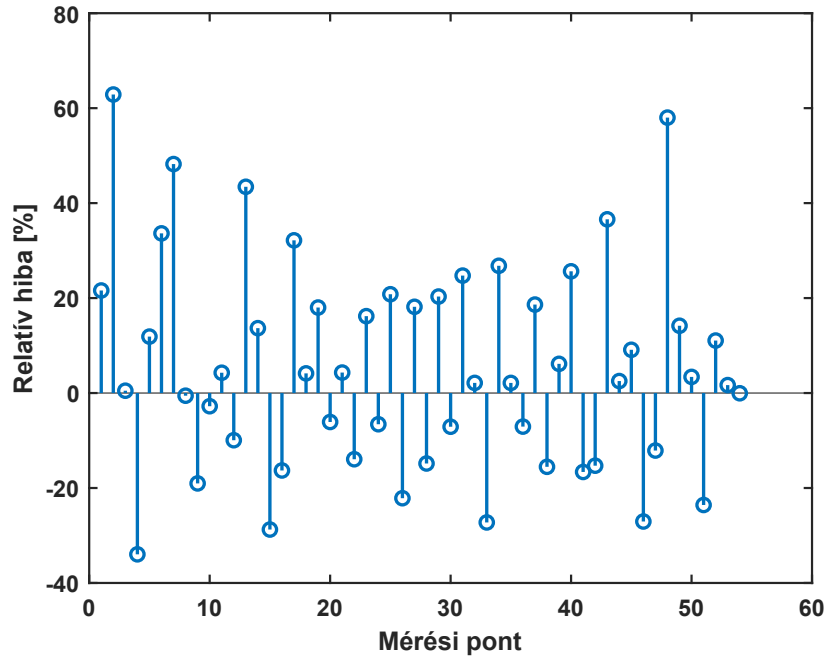
3.5. ábra. A 8-as egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).

A további két egér mérései csak szemléltetésképpen kerülnek bemutatásra, azzal a céllal, hogy kapjunk egy képet arról, hogy mennyire különbözőek az egyes páciensek egymástól. A 3.5. ábra a 8-as egér méréseinek végleges simítását mutatja. Ezen az ábrán egy olyan mérési sorozatot látni, ahol a 100. és a 140. nap közötti mérési értékek, melyeket a kék teli körök mutatnak, egyértelműen zajosak. A tumor térfogata túlságosan ingadozik a valódi zajos mérések alapján. A simított görbe (piros) pedig egy folyamatos tumornövekedést mutat, mely sokkal inkább közelíti a valóságot.

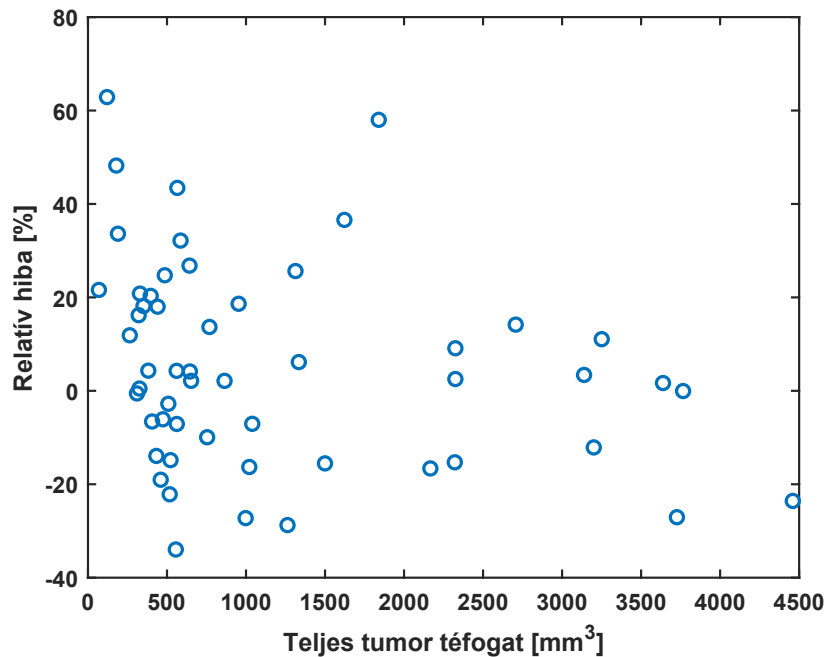
Szintén megnéztük a relatív hibát a két görbe között. Ennek eredményét mutatja a 3.9. ábra. Ezen az ábrán a negatív és a pozitív hibák között nincs akkora különbség, mint a 16-os egérnél. Ennek a hibának is megnéztük a szóródási diagramját, amit a 3.7 ábra mutat. Kicsit egyenletesebb, mint a 16-os egér esetében, illetve egyértelműen nem exponenciálisan csökkennek a hibák értékei a tumor térfogat növekedések függvényében, már inkább egy logaritmikus csökkést mutatnak.

A harmadik bemutatásra kerülő egér a 13-as, melynek a simított mérési eredménye a 3.8. ábrán látható. Ez egy viszonylag kevés ideig kezelt egér, csak 80 napnyi mérési eredmény tartozik hozzá. A mérési görbe és a simított görbe közötti különbség jól látható, illetve a szintén nagy ingadozásokat a tumor térfogatában az algoritmus ennél az egyedenél is szépen kisimította.

A simítás után a relatív hibák a 3.9. ábra szerint alakultak az egyes mérési pontokban. Ezen a hibadiagramon látni a legjobban azt, hogy mennyire különbözik a pozitív és a negatív hibák eloszlása, míg a pozitívok esetében egy exponenciális csökkenést látunk a negatív oldal szinte egyenletesnek tekinthető. Ha megvizsgáljuk a 3.10. ábrát, mely a 13-as egér

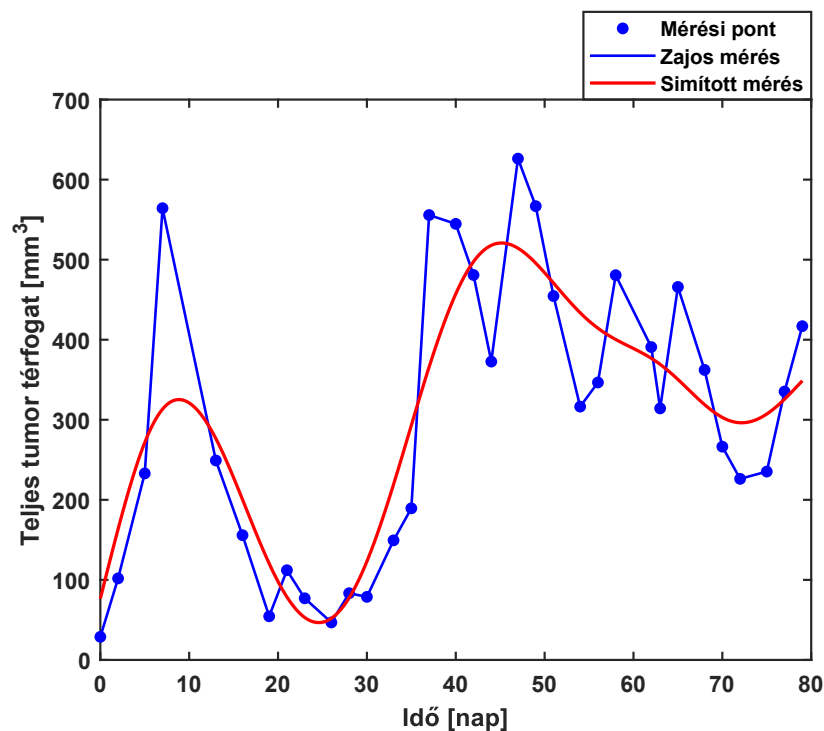


3.6. ábra. A 8-as egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök), minden egyes mérési pontban.

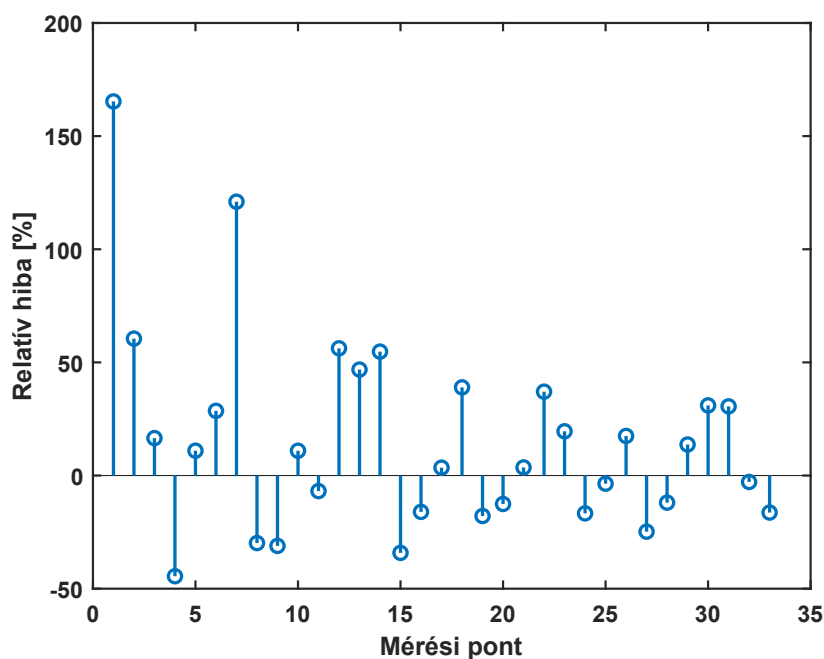


3.7. ábra. A 8-as egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.

mérési hibáinak szóródását mutatja, akkor kijelethető, hogy ennél az egérnél ténylegesen exponenciálisan csökkent a mérési zaj a tumor növekedésével párhuzamosan, illetve ennél a mérési zajnál nem tömörülnek annyira a diagram bal oldalára a kék körök, hasonlóan vannak elszórva minden tumorméreten. A kisebb tumoroknál még mindig nagyobb a zaj, mint a nagyobb tumorok esetén.

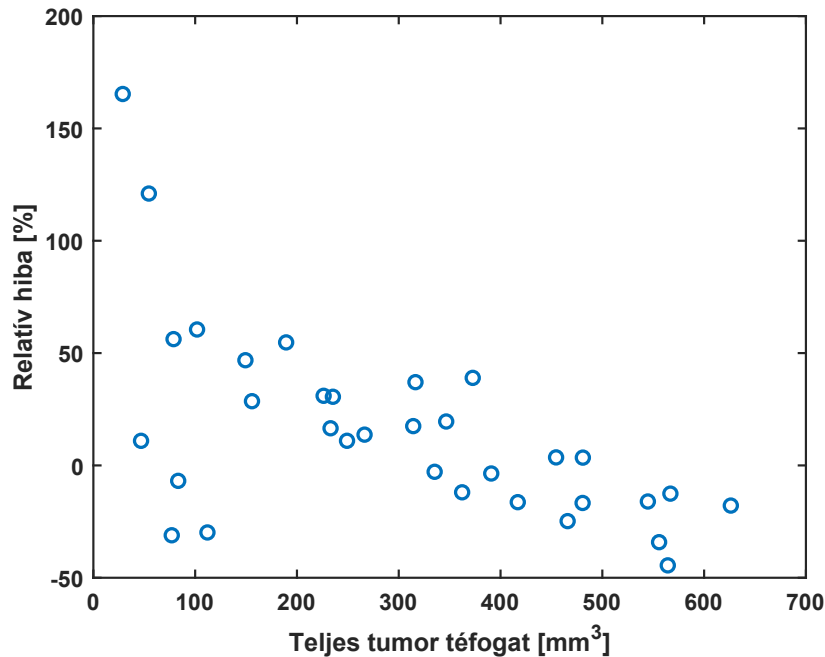


3.8. ábra. A 13-as egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).



3.9. ábra. A 13-as egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök), minden egyes mérési pontban.

A következő lépés a valószínűségi eloszlások mérési zajra való illesztése. A 3.3., 3.6. és 3.9. hibadiagramok, illetve 3.4., 3.7. és 3.10. szóródási diagramok alapján, továbbá a többi, be nem mutatott egérhez tartozó mérési zaj viselkedése alapján arra jutottunk, hogy



3.10. ábra. A 13-as egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.

a hibákat pozitív és negatív részekre kettébontva végezzük el a valószínűségi eloszlások illesztését. Külön illesztünk eloszlást a pozitív zajra és a negatívra, hiszen egyértelműen különböző eloszlást mutattak. Az is megfigyelhető volt az ábrákon, hogy nagyon eltérőek a nagyságrendek egy-egy egérhez tartozó mérési zajnál, illetve a kisebb tumortérfogatokhoz általánosságban nagyobb mérési zaj tartozott. Ezt kiegyenlítve az eloszlások illesztése előtt zajfehérítést hajtottunk végre.

3.2. Valószínűségi eloszlások illesztése és zajmodell felépítése MRI adatok nélkül

Az előző 3.1. alfejezetben bemutatott egerekhez tartozó mérési zajokat ebben a fejezetben kettébontjuk pozitív és negatív mérési zajra, majd fehérítjük mind a két részt külön-külön egy általunk meghatározott transzformációval. Ezután a [19] TDK dolgozatban bemutatott valószínűségi eloszlásokat illesztjük szintén mind a két részre. A valószínűségi eloszlások illesztési eredményét a 16-os egeren részletesen bemutatjuk, a 13-as és 8-as egérnél már csak a végeredmény kerül bemutatásra. A legjobban illeszkedő valószínűségi eloszlással generálunk egy mérési zajt az adott pácienshez, majd azt visszatranszformáljuk és összehasonlítjuk az eredeti mérési zajjal.

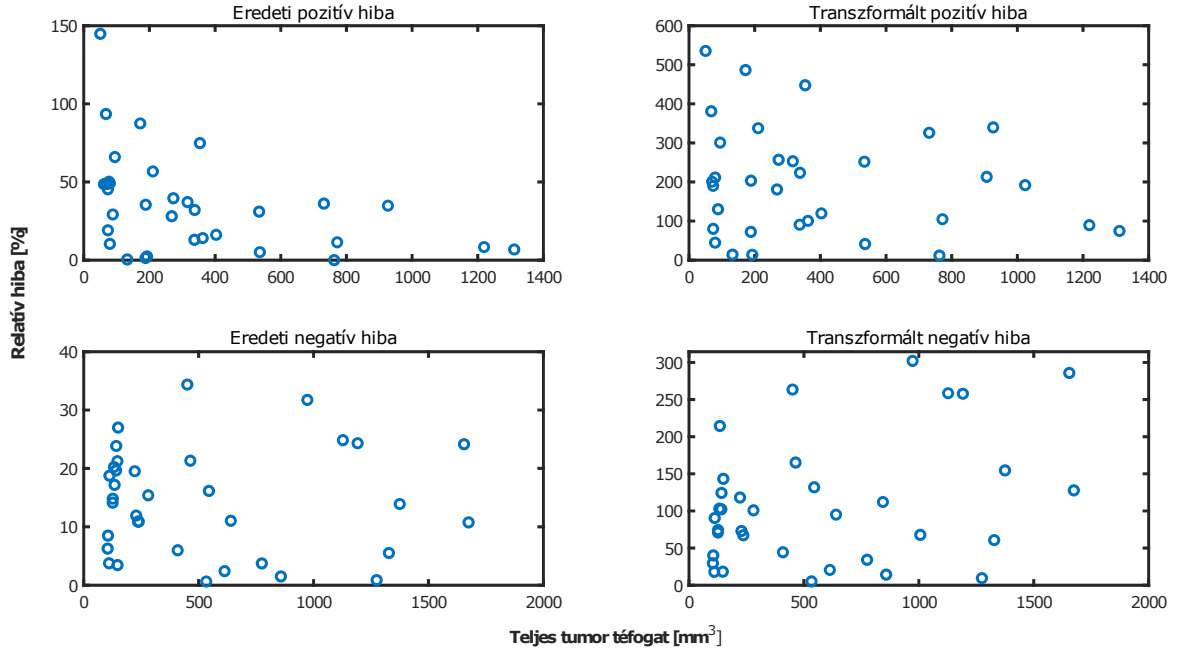
Első lépésben zajt fehérítünk, tehát kiegyenlítjük a nagyságrendeket a zaj értékei között, illetve függetlenítjük a zajt a tumor térfogatól. A heurisztika alapja az volt, hogy a mérési hiba esetünkben az egyik dimenzió méréséhez (szélesség/hosszúság) adódik hozzá, de mi térfogatokkal dolgozunk a valóságban, ami ezek köbétől függ. Ezért olyan transzformációt alkalmaztunk a zajfehérítés során, amiben megszoroztuk a hibákat a térfogatok köbgyökével, ezáltal elérve a zaj függetlenítését a tumor méretétől. A felhasznált transzformáció a

következő:

$$\tilde{e} = \Phi(y)e \quad (3.3)$$

$$\Phi(y) = \sqrt[3]{y}, \quad (3.4)$$

ahol $\tilde{e}$ a transzformált hiba, y a tumortérfogat, és $\Phi(y)$ fehérítésre használt transzformáció. A transzformáció jelenlegi formája a pozitív hibákra értendő, a negatív hibáknál a hibavektornak a mínusz egyszeresét vesszük és azt szorozzuk meg a mérések köbgyökével, amit figyelembe veszünk a visszatranszformálás során is.

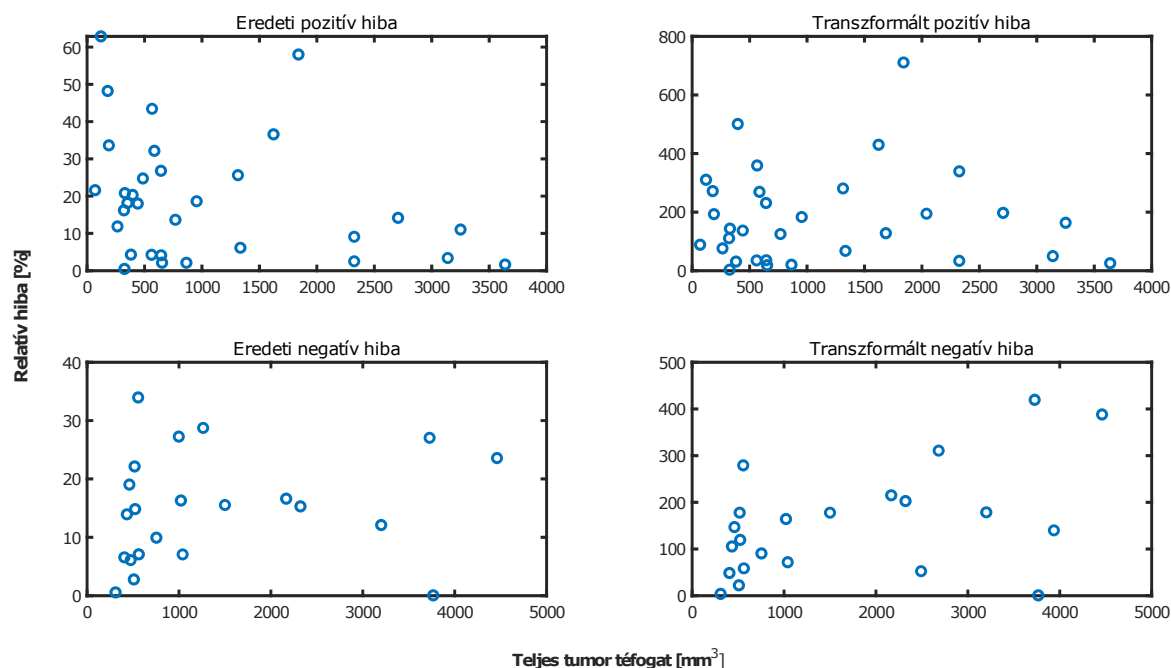


3.11. ábra. A 16-os egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).

A zaj kettébontását és a kettébontott részek fehérítésének eredményét a szóródási diagramok összehasonlításával mutatjuk be. A 3.11. ábra a 16-os egérhez tartozó zaj szóródási diagramjait foglalja össze. A vízszintes tengelyek minden diagramnál a teljes tumor térfogatát mutatják mm^3 -ben, a függőleges tengelyek pedig a relatív hibát százalékban.

A bal felső diagramon a 16-os egér eredeti pozitív hibáinak szóródását látjuk. Mellette, jobbra fent jelenítettük meg ennek a transzformált verzióját. A pozitív hibákat nagyjából sikerült függetleníteni a zajtól, több kék kör jelent meg a nagyobb tumortérfogatoknál a fent lévő jobb oldali ábrán. Az eredeti pozitív zaj inkább exponenciális csökkenést mutatott, a transzformáció után viszont egy kisebb meredekségű lineárisabb csökkenés figyelhető meg. A negatív részen, melyet az ábra lenti két diagramja szemléltet, nem olyan látványos a változás, mint a pozitív részekben. Itt is sikerült a hibákat a térfogatoktól függetlenebbé tenni.

A 3.12. ábra a 8-as egérhez tartozó zaj szóródási diagramjait mutatja. Az elrendezés megegyezik a 3.11. ábra elrendezésével. A pozitív hibák esetében 1000 mm^3 és 2500 mm^3 között feltűnő a változás a jobb oldali transzformált hibákat mutató ábrán, a bal oldali eredetihez képest. A zaj egyenletesebben oszlik el a tumortérfogatok függvényében. A negatív



3.12. ábra. A 8-as egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).

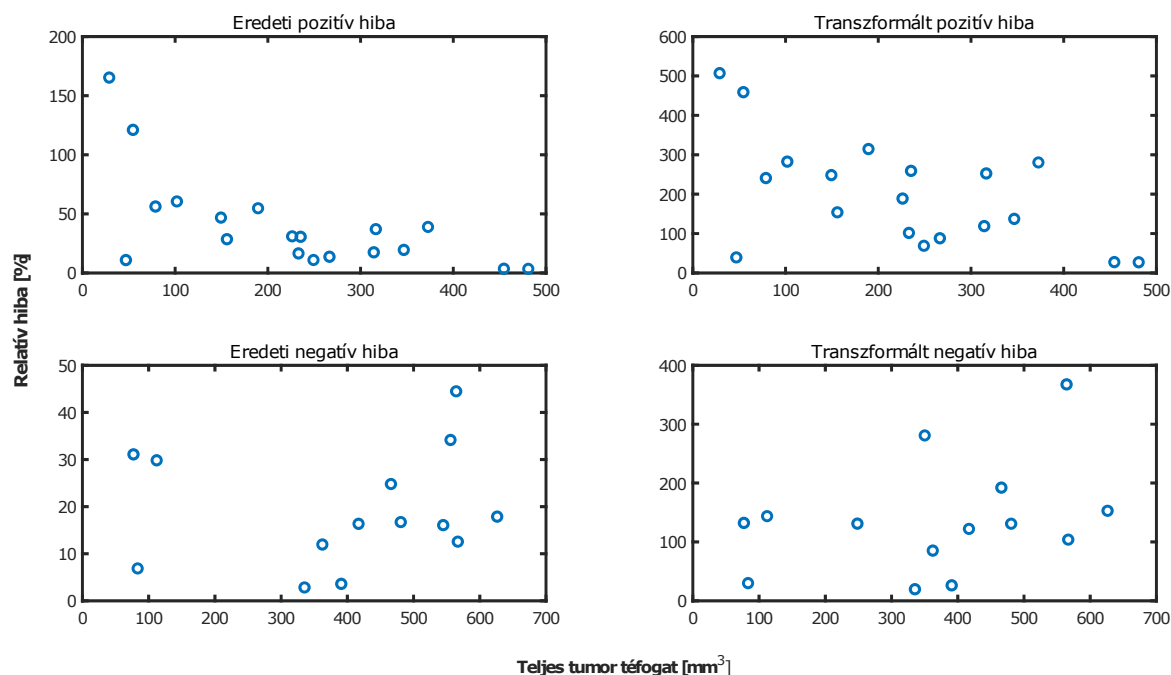
hibáknál is sikerült többnyire egyenletesebbé tenni a hibák szóródását, de a 8-as és 16-os egérnél is igaz, hogy a negatív hibák már eredetileg is egyenletesebb eloszlást mutatottak a pozitív hibáknál.

A 13-as egérhez is elkészítettük a zajferítés eredményét bemutató szóródási diagramokat, ez tekinthető meg a 3.13. ábrán. Elrendezése megegyezik az eddig bemutatottakkal. A kiválasztott egerek közül az ehhez tartozó hibákon sikerült a legjobban a fehérítés, ez betudható annak is, hogy ennél az egérnél kevesebb mérési állt a rendelkezésünkre.

A fent megjelenített pozitív zajokat sikerült egyenletesebbé transzformálni (3.7) segítségével, ez szintén igaz a negatív zajokra is. A pozitív esetben 100 mm^3 és 400 mm^3 között szinte teljesen egyenletesen oszlanak el a transzformált mérési hibák, míg az eredeti hibák eloszlása egyértelműen exponenciális csökkenést mutatott. A negatív rész itt is egyenletesebb volt a pozitívnál, de sikerült szépíteni rajta a fehérítés során.

Ezt a fehérítést a (3.7) segítségével mind a 49 egérhez tartozó pozitív és negatív hibáján elvégeztük, majd a transzformált értékekre illesztettük rá a [19] TDK dolgozatban bemutatott valószínűségi eloszlásokat. A 16-os egér eredményeit részletezi a következő két táblázat. A 3.1. táblázat a pozitív hibákra, míg a 3.2. táblázat a negatív hibákra való illesztési eredményeket mutatja.

A táblázatok első oszlopa az illesztett valószínűségi eloszlás nevét, a második oszlop az adott eloszlás paramétereinek értékét, melyek a 16-os egérhez tartozó mérési zajra való illesztés alapján adódnak. A harmadik oszlop az eloszlások paramétereinek konfidencia intervallumát adja meg, mely a statisztikában az intervallumbecslés egy központi eszköze. Mivel egy adott mintát vizsgálunk, így ezek a paraméterértékek erre a minta adathalmazra vonatkoznak. A konfidencia intervallum megadja az adott minta alapján, hogy "véltetően" hol lehet a valódi sokaságra igaz érték. A táblázatok negyedik oszlopában az Anderson-



3.13. ábra. A 13-as egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).

Darling teszt eredményét jelenítettük meg minden eloszláshoz, mely egy statisztikai érték, ami azt mutatja meg, hogy egy adott minta egy adott valószínűségi eloszlásból származik-e [20, 21]. Az Anderson-Darling tesztre egy beépített függvényt használtunk, mely visszaad egy adstat változót. Ennek a változónak az értékét adtuk meg a táblázatban, mely minél kisebb, annál nagyobb valószínűséggel tartozik a minta a vizsgált eloszlásba.

A két táblázatban kiemeltük bordó színnel a 16-os egérhez tartozó hibákra legjobban illeszkedő eloszlásokat. A 3.1. táblázat adatai alapján az általánosított szélsőérték-eloszlás illeszkedett a legjobban a hibákra, ennek az AD statisztika értéke a legkisebb, mindössze 0,3120. Ezt követi kicsit magasabb értékkel a logisztikus, fél-normális, és a normális eloszlás, ami esetleg még jó illeszkedést mutat. A pozitív hibákon több olyan valószínűségi eloszlás is van, mely tipikusan rossz illeszkedést mutat ezen egér adatain, ilyen például az inverz Gauss-eloszlás, mely a legmagasabb AD statisztika értékkel (35,6203) bír vagy a Birnbaum-Saunders eloszlás a 14,7212-es AD statisztika értékével.

A 3.2. táblázatban szintén a 16-os egérhez tartozó hibák illesztési eredményei vannak, viszont ez a táblázat a negatív hibára vonatkozó eredményeket foglalja össze. A kiemelt eloszlás a Weibull-eloszlás, melynek AD statisztika értéke 0,3219, ami hasonlóan kicsi, mint ami a pozitív hibák esetében az általánosított szélsőérték-eloszlásnál adódott. A negatív oldalon viszont további kiemelendő valószínűségi eloszlások vannak, mivel több AD statisztika érték is közelít a Weibull-eloszlásához. Ilyen például a Gamma-, a stabil vagy szintén az általánosított szélsőérték-eloszlás. A negatív hibákra szinte minden eloszlás elfogadhatóan illeszkedik, nincsenek olyan egyértelműen kizárható eloszlások, mint a pozitív hibák esetében például az inverz Gauss-eloszlás, ami egy 35 feletti AD statisztika értéket mutatott. Ez betudható annak is, hogy a negatív hibák minden egér esetében egyenletesebb eloszlást mutattak, nem voltak annyira változatosak, mint a pozitívak.

Valószínűségi eloszlás	Paraméterek	Konfidencia intervallum	AD statisztika
Exponenciális eloszlás	$\mu_{exp} = 197,35$	[144,17; 286,71]	1,0169
Általánosított Pareto-eloszlás	$\xi_{gp} = -0,49$ $\sigma_{gp} = 299,56$ $\theta_{gp} = 0$	[-0,86; -0,11] [186,13; 482,11] -	0,5688
Birnbaum-Saunders eloszlás	$\beta_b = 8,60$ $\gamma_b = 5,01$	[2,06; 15,15] [3,15; 6,88]	14,7212
Gamma-eloszlás	$a_g = 0,93$ $b_g = 211,79$	[0,61; 1,42] [122,12; 367,32]	1,1135
Nakagami-eloszlás	$\mu_{na} = 0,39$ $\omega = 60009,60$	[0,26; 0,57] [34810,30; 103451,00]	0,6620
Szélsőérték-eloszlás	$\mu_{ext} = 274,39$ $\sigma_{ext} = 159,48$	[216,48; 332,30] [125,34; 202,87]	1,4353
Általánosított szélsőérték-eloszlás	$\xi_{gev} = 0,01$ $\sigma_{gev} = 113,81$ $\mu_{gev} = 128,90$	[-0,32; 0,36] [83,67; 154,80] [82,79; 175,01]	0,3120
Weibull-eloszlás	$A_w = 201,08$ $B_w = 1,06$	[144,35; 280,08] [0,79; 1,42]	0,8949
Rayleigh-eloszlás	$B_r = 173,219$	[148,04; 208,78]	3,9605
Rician-eloszlás	$s_{ri} = 14,20$ $\sigma_{ri} = 172,97$	[0; 991,43] [129,75; 230,59]	3,9628
Inverz Gauss-eloszlás	$\mu_{ig} = 197,35$ $\lambda_{ig} = 2,03$	[-465,48; 860,20] [1,05; 3,01]	35,6203
Normális eloszlás	$\mu_n = 197,35$ $\sigma_n = 147,36$	[145,10; 249,61] [118,51; 194,92]	0,5776
Fél-normális eloszlás	$\mu_{hn} = 0$ $\sigma_{hn} = 244,96$	- [197,58; 322,44]	0,5355
Logisztikus eloszlás	$\mu_l = 185,77$ $\sigma_l = 82,62$	[136,47; 235,07] [62,24; 109,67]	0,4703
Log-normális eloszlás	$\mu_{ln} = 4,66$ $\sigma_{ln} = 1,78$	[4,02; 5,29] [1,43; 2,36]	2,8975
Log-logisztikus eloszlás	$\mu_{ll} = 4,98$ $\sigma_{ll} = 0,76$	[4,55; 5,41] [0,56; 1,03]	1,5500
Student féle t-eloszlás	$\mu_s = 197,35$ $\sigma_s = 145,11$ $\nu_s = 2,30e + 06$	[148,25; 246,45] [114,48; 183,94] [398760; 1,33e + 07]	0,5877
Stabil eloszlás	$\alpha = 2$ $\beta_s = 0,99$ $\gamma_s = 102,57$ $\delta = 197,24$	[0; 2] [-1; 1] [0, ∞] [- ∞ , ∞]	0,5865

3.1. táblázat. A 16-os egérhez tartozó pozitív hibák illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve bordóval a legjobban illeszkedő eloszlást.

Ezt a vizsgálatot elvégezve minden egérre, megkaptuk az egerek mérésein lévő becsült mérési zaj eloszlását mind a 49 darab egérre. Szemléltetésképpen a már ismert három egérre

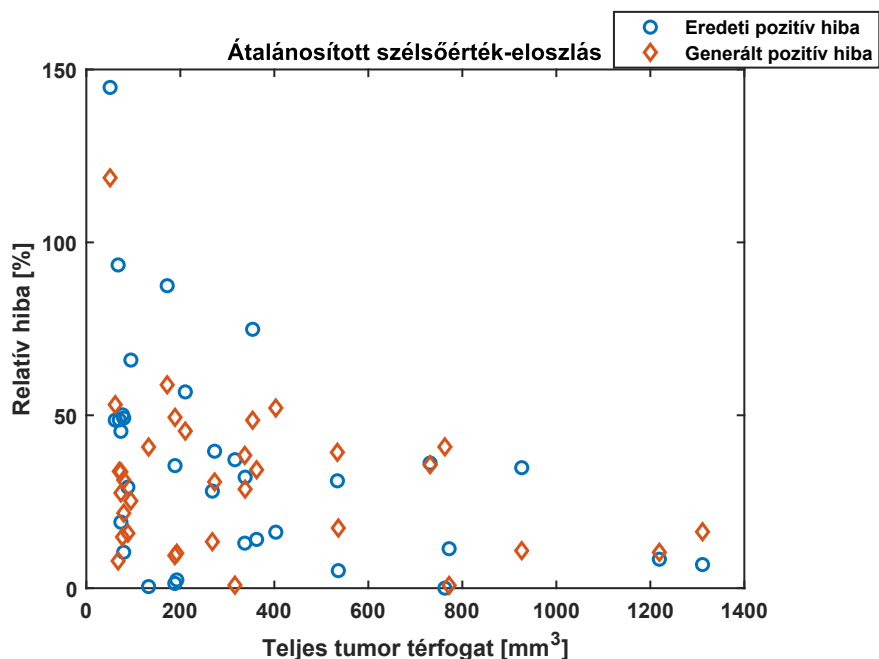
Valószínűségi eloszlás	Paraméterek	Konfidencia intervallum	AD statisztika
Exponenciális eloszlás	$\mu_{exp} = 105,26$	[77,54; 151,13]	0.7852
Átalánosított Pareto-eloszlás	$\xi_{gp} = -0,37$ $\sigma_{gp} = 146,68$ $\theta_{gp} = 0$	[-0,74; -0.01] [90,99; 236.44] -	0,4581
Birnbaum-Saunders eloszlás	$\beta_b = 63,36$ $\gamma_b = 1,11$	[43,37; 83,36] [0,84; 1,37]	1,2602
Gamma-eloszlás	$a_g = 1,47$ $b_g = 71,43$	[0,61; 1,42] [43,07; 118,46]	0,3306
Nakagami-eloszlás	$\mu_{na} = 0,51$ $\omega = 17785,20$	[0,34; 0,76] [11213,20; 28208,90]	0,4155
Szélsőérték-eloszlás	$\mu_{ev} = 149,93$ $\sigma_{ev} = 94,78$	[116,46; 183,41] [75,20; 119,44]	2,5697
Általánosított szélsőérték-eloszlás	$\xi_{gev} = 0,20$ $\sigma_{gev} = 53,05$ $\mu_{gev} = 63,03$	[-0,16; 0.56] [38,43; 73,22] [41,90; 84,16]	0,3576
Weibull-eloszlás	$A_w = 113,63$ $B_w = 1,28$	[86,53; 149,22] [0,98; 1,66]	0,3219
Rayleigh-eloszlás	$B_r = 94,30$	[80,93; 112,99]	4,0902
Rician-eloszlás	$s_{ri} = 5,19$ $\sigma_{ri} = 94,24$	[0; 626,50] [73,71; 120,49]	4,0921
Inverz Gauss-eloszlás	$\mu_{ig} = 105,26$ $\lambda_{ig} = 65,36$	[60,00; 149,52] [34,73; 95,98]	1.7167
Normalis eloszlás	$\mu_n = 105,26$ $\sigma_n = 83,07$	[76,73; 133,80] [67,19; 108,84]	1,3782
Fél-normális eloszlás	$\mu_{hn} = 0$ $\sigma_{hn} = 133,36$	- [108,16; 173,96]	0,3950
Logisztikus eloszlás	$\mu_l = 94,47$ $\sigma_l = 44,25$	[69,21; 119,74] [33,42; 58,88]	0,9567
Log-normális eloszlás	$\mu_{ln} = 4,28$ $\sigma_{ln} = 0,99$	[3,93; 4,62] [0,80; 1,30]	0,7378
Log-logisztikus eloszlás	$\mu_{ll} = 4,37$ $\sigma_{ll} = 0,54$	[4,05; 4,68] [0,41; 0,72]	0,5920
Student féle t-eloszlás	$\mu_s = 87,68$ $\sigma_s = 57,07$ $\nu_s = 3,13$	[62,43; 112,92] [34,65; 94,02] [0,81; 12,14]	0,8079
Stabil eloszlás	$\alpha = 1.34$ $\beta_{st} = 0,99$ $\gamma_{st} = 39,25$ $\delta_{st} = 70,27$	[0; 2] [-1; 1] [0, ∞] [- ∞ , ∞]	0,3438

3.2. táblázat. A 16-os egérhez tartozó negatív hibák illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve bordóval a legjobban illeszkedő eloszlást.

bemutatjuk, hogy milyen eredmények jönnek ki, ha a számukra megfelelő eloszlásokkal generálunk rájuk mérési zajt.

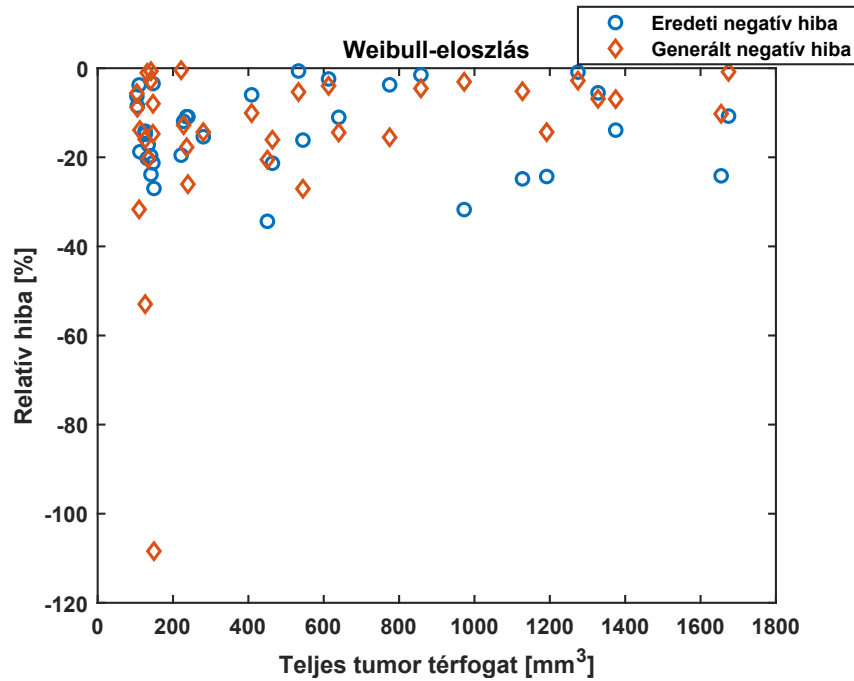
A 16-os egér lesz az első, mivel nála már ismerjük, hogy a pozitív hibák leírhatók az általánosított szélsőérték-eloszlással, a negatívak pedig a Weibull-eloszlással. A hiba újragenerálásához megadjuk a 3.1. és 3.2. táblázatokban szereplő paraméterekkel a megfelelő eloszlásokat, majd az eloszlással egy véletlen vektort generálunk, melynek elemszáma megegyezik a kívánt hibák elemszámával. Ez a 16-os egérre levezetve azt jelenti, hogy lesz egy általánosított szélsőérték-eloszlású véletlen elemekből álló vektorunk a pozitív hibákra, és lesz egy Weibull-eloszlású véletlen elemekből álló vektorunk a negatív hibákra. Ezeket a vektorokat a (3.7) transzformáció inverzével visszatranszformáljuk (nem elfelejtve a negatív értékeknek szánt, visszatranszformált hibák mínusz egyszeresét venni), annak érdekében, hogy ismét függjenek a hibák a tumor térfogatának értékétől.

Egy ilyen generált zaj eredménye látható a 3.14. ábrán, ami a 16-os egérhez tartozó eredeti pozitív hibákat mutatja a kék körökkel és az új általunk generált pozitív hibákat a narancs színű rombuszokkal. Az ábrán egyértelműen látszik, hogy sikerült a zajt reprodukálni az általános szélsőérték-eloszlással. A kék köröket jól követik a narancs rombuszok. A kisebb tumor térfogatokon több mérés van. A pozitív zaj exponenciális csökkenését a tumortérfogatok növekedésének függvényében visszakaptuk. A negatív mérési zaj újrage-

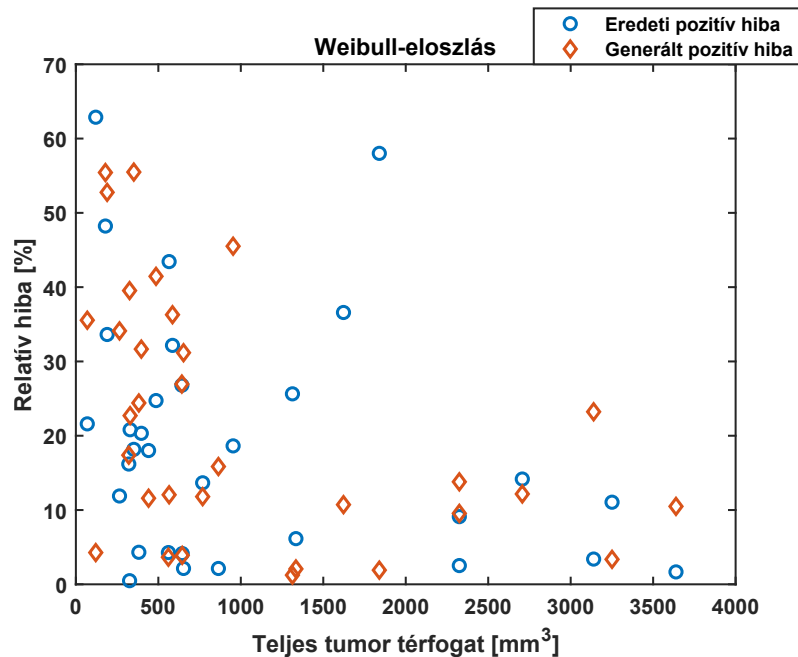


3.14. ábra. A 16-os egér eredeti pozitív hibáinak szóródása (kék körök) és az általánosított szélsőérték-eloszlással generált pozitív hibák szóródása (narancs rombuszok).

nerálásának eredményét a 3.15. ábra mutatja szintén a 16-os egérhez. Itt Weibull-eloszlást alkalmaztunk a mérési zaj reprodukálásánál, ami elfogadható eredményt hozott. Vannak azonban kilógó értékek, ahol a generált hiba (narancs rombuszok) nagyságrendje nem illeszkedik az eredeti hibához (kék körök). A kisebb tumortérfogatokhoz tartozó nagyobb és gyakoribb hibákat sikerült újragenerálni a negatív oldalon is. A következő egér a 8-as egerünk. A hozzá tartozó hibákhoz szintén ez a két eloszlás illeszkedett a legjobban, annyi különbséggel, hogy megcserélődtek az eredmények. A pozitív hibákat írta le jobban a Weibull eloszlás, a negatívakra pedig az általánosított szélsőérték illeszkedett a legjobban az AD statisztika alapján.

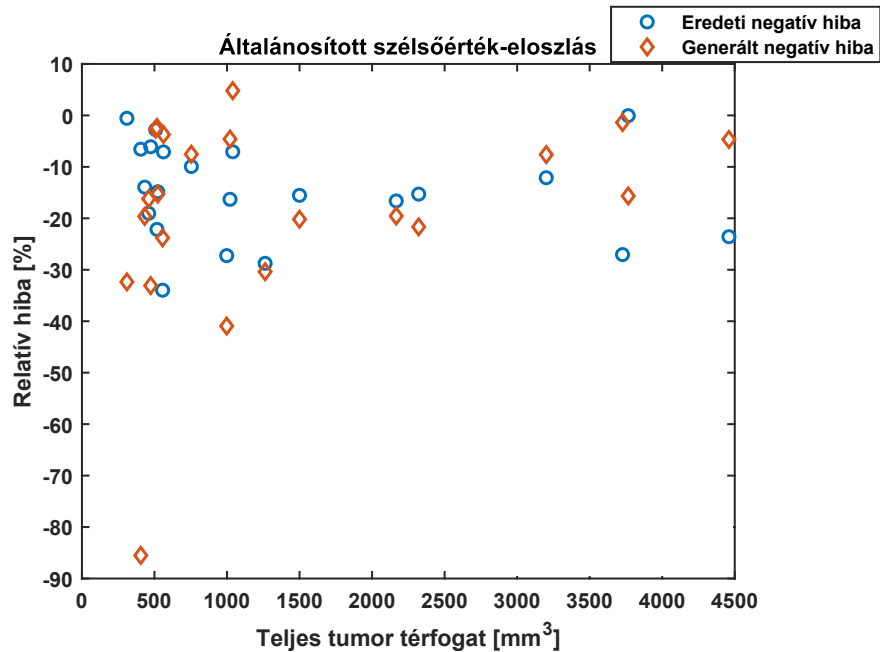


3.15. ábra. A 16-os egér eredeti negatív hibáinak szóródása (kék körök) és a Weibull-eloszlással generált negatív hibák szóródása (narancs rombuszok).



3.16. ábra. A 8-as egér eredeti pozitív hibáinak szóródása (kék körök) és a Weibull-eloszlással generált pozitív hibák szóródása (narancs rombuszok).

A 3.16. ábra a 8-as egérhez tartozó pozitív hibákat mutatja, szintén a kék körök mutatják az eredeti szóródást és a narancs rombuszok az általunk generált pozitív mérési hibák szóródását. Összességében sikerült reprodukálni a zajt, nagyjából követik a narancs rombuszok a kék körök elhelyezkedését. Ellentétben a 16-os egér eredményeivel, itt a Weibull-eloszlással generáltuk a pozitív hibákat. Nagyságrendileg megegyezik az új hibák értéke az



3.17. ábra. A 8-as egér eredeti negatív hibáinak szóródása (kék körök) és az általánosított szélsőérték-eloszlással generált negatív hibák szóródása (narancs rombuszok).

eredetiek értékével.

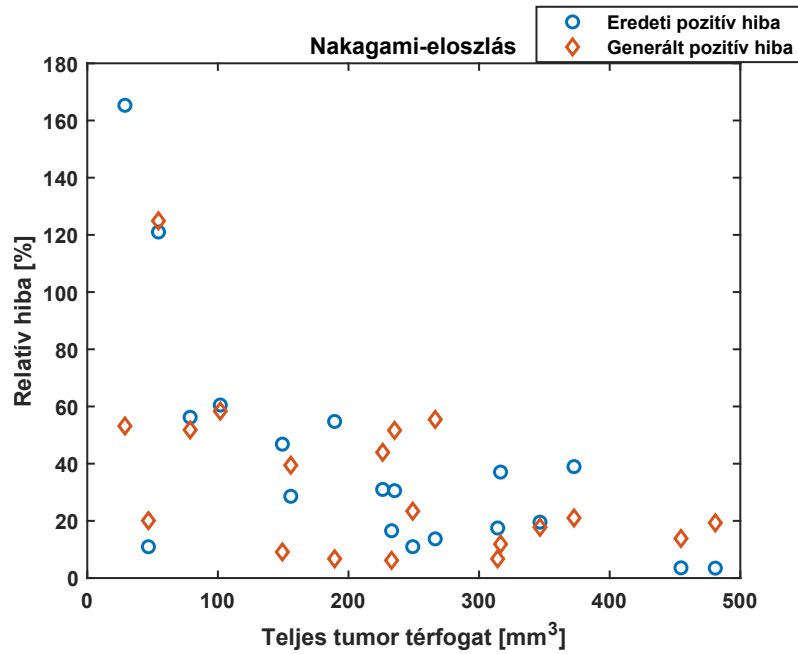
A 8-as egér negatív hibáit az általánosított szélsőérték-eloszlás segítségével modelleztük le. Ennek eredményét jelenítettük meg a 3.17. ábrán. Egy kilógó hibaértékünk van az 500 mm³-es tumor térfogat előtt a -90%-os hiba környékén. Ettől eltekintve sikerült a hibát reprodukálni, a narancs rombuszok (generált hiba) szinte tökéletesen lekövettek a kék körök helyzetét (eredeti hiba).

Az esetek többségében a 49 egér alapján a hibák ezt a két eloszlást, tehát a Weibull-eloszlást és az általánosított szélsőérték-eloszlást követik (ez részletezésre kerül majd a dolgozat későbbi szakaszában). Akadtak kivételes esetek is, ilyen eset a 13-as egér is, melynél már említettük, hogy kevesebb mérés állt a rendelkezésünkre. Ez is lehet az oka annak, hogy a hozzá tartozó hibák teljesen más eloszlásokat követnek, hiszen nem állt rendelkezésre elegendő minta a kevés számú mérés miatt. A 13-as egérhez tartozó pozitív hibákra a Nakagami-eloszlás illeszkedett a legjobban, míg a negatív hibákra a Student féle t-eloszlás (megjegyzendő, hogy ez az eloszlás csak ennél az egy egérnél jött ki a legjobbnak).

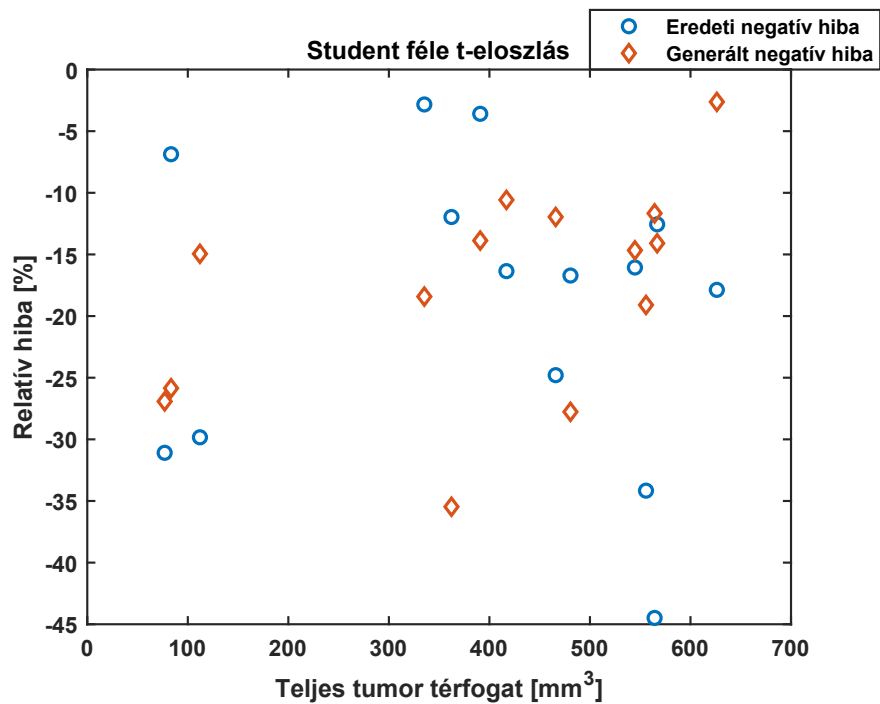
A pozitív hibákra tehát a Nakagami-eloszlás illeszkedett a legjobban a 13-as egér esetében. A pozitív hiba újragenerálásának az eredményét mutatja a 3.18. ábra, ahol igazából az látható, hogy a generált hiba nagyon szépen illeszkedik az eredeti hibákra. Az exponenciális csökkenést megtartotta a tumor térfogatok növekedésének a függvényében és nincsenek kilógó hibák sem, amik nem illeszkednének az eredeti zajra.

A 13-as egér negatív hibáit a Student féle t-eloszlás írta le a legjobban, ennek eredménye a 3.19. ábrán jelenik meg. Itt is elmondható, akárcsak a pozitív esetben, hogy az eloszlással a hiba szépen újragenerálható, nincsenek kilógó értékek. A generált hiba jól követi az eredeti mérési hibákat, a Student féle-t eloszlás illeszkedik a 13-as egérhez tartozó mérési zajra.

Összességében tehát elmondható, hogy a mérési zaj modellezése sikeres volt. A modellezés lépései a következők: A mérési zaj becslése után a $\Phi(\cdot)$ transzformációt alkalmazva,



3.18. ábra. A 13-as egér eredeti pozitív hibáinak szóródása (kék körök) és a Nakagami-eloszlással generált pozitív hibák szóródása (narancs rombuszok).



3.19. ábra. A 13-as egér eredeti negatív hibáinak szóródása (kék körök) és a Student féle t-eloszlással generált negatív hibák szóródása (narancs rombuszok).

a zajt fehérítjük, majd a megfelelő eloszlást meghatározva újrageneráljuk. A generált zajt a $\Phi(\cdot)$ transzformáció inverzével visszatranszformáljuk. Ezt alkalmazva minden egyes egér méréseiből származó hiba legenerálható.

Mivel a célunk, hogy ezt a mérési zajt a virtuális páciensek szimulált méréseire generál-

juk rá, így szükség van egy, vagy kettő valószínűségi eloszlásra, mely képes leírni az összes egérre a pozitív és a negatív hibákat. A valószínűségi eloszlás kiválasztása után a tanító adatok legenerálhatók és a szimulált méréseikre a globálisan megválasztott eloszlásokkal egy reális zaj generálható.

A 3.3. táblázat összefoglalja, hogy mely eloszlások illeszkedtek legtöbbször a legjobban az egerek összességéhez tartozó pozitív, illetve negatív hibákhoz. A táblázat első oszlopa az eloszlás megnevezését tartalmazza, a második oszlop az adott eloszlás legjobb illeszkedésének darabszámát a pozitív hibákon, a harmadik oszlop pedig a negatív hibákon. A táblázat alapján az adott mintaszám mellett (49 egér) a pozitív hibákra a legtöbbször a Weibull-eloszlás illeszkedett. Ez 12 egérhez tartozó pozitív hibára való illeszkedést jelent. A második helyen a stabil eloszlás áll 10 darab sikeres illeszkedéssel, majd az általánosított szélsőérték-eloszlás 8 darab illeszkedéssel.

A negatív hibákra legtöbbször az általánosított szélsőérték-eloszlás illeszkedett, ez 12 alkalmat foglal magában. Utána következik a stabil eloszlás 8 alkalommal, majd a Weibull-eloszlás összesen 7 alkalommal. A három lehetséges valószínűségi eloszlás tehát mind a két oldalon a Weibull-, a stabil és az általánosított szélsőérték-eloszlás.

Valószínűségi eloszlás	Pozitív hibákra [db]	Negatív hibákra [db]
Exponenciális eloszlás	3	0
Általánosított Pareto-eloszlás	4	7
Birnbaum-Saunders eloszlás	0	0
Gamma-eloszlás	4	5
Nakagami-eloszlás	4	6
Szélsőérték-eloszlás	0	0
Általánosított szélsőérték-eloszlás	8	12
Weibull-eloszlás	12	7
Rayleigh-eloszlás	0	0
Rician-eloszlás	0	0
Inverz Gauss-eloszlás	1	0
Normális eloszlás	0	1
Fél-normális eloszlás	1	0
Logisztikus eloszlás	0	0
Log-normális eloszlás	1	1
Log-logisztikus eloszlás	2	0
Student féle t-eloszlás	0	1
Stabil eloszlás	10	8

3.3. táblázat. Felhasznált valószínűségi eloszlások legjobb illeszkedésének darabszáma a pozitív és negatív hibákra az egerek összességén.

Az eloszlásokat külön illesztjük a 49 egér pozitív és negatív értékű mérési hibáira, és az eloszlások paramétereinek vizsgáljuk várható értékét és szórását. A három legjobban illeszkedő eloszlás paramétereinek várható értékét és szórását mutatja a 3.4. táblázat, kettébontva a negatív és pozitív hibákra való illeszkedés eredményeképp kapott paraméterértékeket. A várható értéket és a szórást a 49 egér együttes illesztési eredményeiből számoltuk.

A várható érték alapján tudjuk majd újragenerálni az eloszlással a hibákat, viszont számunkra a globális eloszlás melletti döntést jobban befolyásolhatja a paraméter értékeinek a

Valószínűségi eloszlás	Paraméter	Várható érték		Szórás	
		Pozitív	Negatív	Pozitív	Negatív
Weibull-eloszlás	A_w	294,992	173,026	539,609	232,529
	B_w	0,950	1,091	0,220	0,252
Stabil eloszlás	α_{st}	1,130	1,305	0,329	0,407
	β_{st}	0,999	0,948	0,003	0,280
	γ_{st}	82,335	320,386	105,937	1,90e+03
	δ_{st}	96,777	102,906	69,443	144,370
Általánosított szélsőérték-eloszlás	ξ_{gev}	0,548	0,390	0,372	0,438
	σ_{gev}	97,014	64,425	63,777	30,083
	μ_{gev}	93,024	70,526	45,533	30,994

3.4. táblázat. A hibákra legjobban illeszkedő valószínűségi eloszlások paramétereinek várható értéke és szórása, pozitív és negatív hibákra való illeszkedések szerint kettébontva a 49 egér illesztése alapján.

szórása, hiszen az minél kisebb, annál közelebb állnak egymáshoz a paraméterértékek egy-egy paraméter esetében. A szórás alapján a negatív és pozitív esetben is az általánosított szélsőérték-eloszlást kellene választanunk a tanító adatok generálásához. Ezt viszont azért nem tesszük, mert a 49 egérből összesen 8 darab volt olyan, mely méréseihez tartozó pozitív és negatív hibákra ugyanaz az eloszlás illeszkedett, illetve az általánosított szélsőérték a pozitív hibák esetében csak harmadik helyen szerepelt.

A stabil eloszlás γ_{st} paraméterének szórása a negatív hibák esetén nagyon nagy, így azt már az elején elvetjük a negatív hibák generálásához és megválasztjuk az általános szélsőérték-eloszlást erre a célra. A pozitív esetben pedig a Weibull-eloszlás került kiválasztásra, hiszen az esetek többségében jól illeszkedett a pozitív hibákra, és az általánosított szélsőérték-eloszlást, ami esetlegesen jobb lehetne a Weibull-nál a szórása alapján, már elvetettük. A tanító adatokban szereplő virtuális mérésekre tehát a pozitív hibákat a Weibull-eloszlással generáljuk, a negatív hibákat pedig az általánosított szélsőértékkal.

3.3. Digitális tolómérő mérési zajának vizsgálata MRI mérések alapján

Ha a kísérleti tumorok térfogatának méréséről beszélünk, akkor a legpontosabb mérésre szolgáló eszköz jelenleg a kisállat MRI, melyet napjainkban üzemeltetni irreálisan drága, rezsiköltsége hatalmas. Éppen ezért Magyarországon nagyon kevés helyen elérhető, csak két-három darab van belőle az országban és azok is nagyon be vannak ütemezve, nehezen férnek hozzá a kísérleteket végző kollégák. Sajnos a kisállat MRI-vel való mérés abból a szempontból is drága és bonyolult, hogy az állatokat minden egyes vizsgálathoz el kell altatni. Arról nem is beszélve, hogy az állat egészségének sem kedvez a sűrű MRI mérés, hiszen az altatás előtt az állatoknak koplalnia kell, a vizsgálat előtt 4-5 órával már nem ehetnek és nem ihatnak, emiatt a Richter együttműködésével végzett kísérletben résztvevő egereknél (ahol egy héten háromszor történt mérés, így altatás is) drasztikus súlycsökkenés volt megfigyelhető. Ezek által a gyakorlatban a kísérleti tumorok mérésének legelterjedtebb mérési eszköze a digitális tolómérő, melyet mi is alkalmazunk a kísérleteink során.

A mérési zaj modellezése elengedhetetlen feladat ebben az esetben, hiszen a digitális tolómérővel való méréseknek több hibakomponense is van. Az egyik fő probléma az, hogy az állat bőre is mérésre kerül. Ez a kísérlet elején (amikor még a tumor kicsi) egy jelentősebb mérési zajt okoz, mint a kísérlet előrehaladtával, hiszen a nagyobb tumor térfogatok függvényében a bőr által okozott hiba folyamatosan csökken, de igazából soha nem tűnik el teljesen. Egy másik fontos komponense a hibának a tumor tolómérővel való összenyomása, hiszen egy viszonylag puha állagú daganat mérésénél egyáltalán nem mindegy, hogy a mérést végző állatházás dolgozó mennyire nyomja össze a mérendő tumort a tolómérővel a mérés közben. A harmadik probléma pedig, hogy a tumor az állat testének bármely részén



3.20. ábra. Egereken végzett kísérletek során a tumor az egér testének bármely részén megjelenhet, többféle amorf alakot felvehet, melynek szélességét és hosszúságát digitális tolómérővel mérjük.

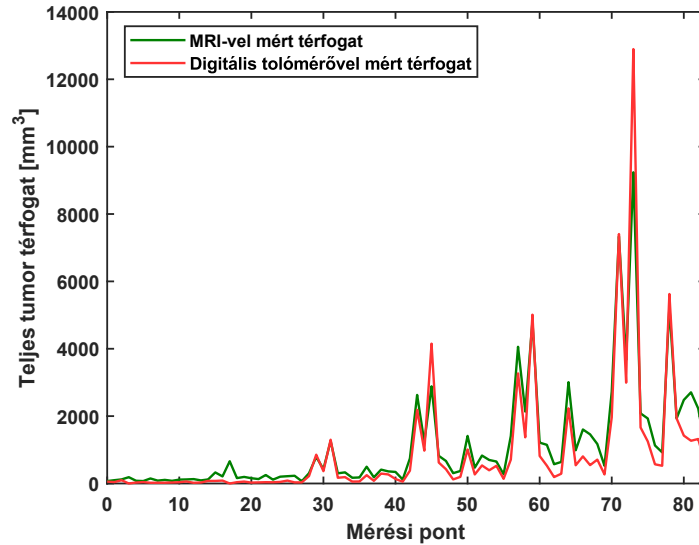
megjelenhet és többféle amorf alakot felvehet, ami miatt mérése még bonyolultabbá válik. A 3.20. ábra szemlélteti, hogy a tumorok nagyjából hogyan helyezkedhetnek el az állat bőre alatt, illetve hogy az alakjuk sem mindig határozható meg pontosan. A digitális tolómérővel a daganat szélességét és hosszúságát mérjük, majd egy közelítőképlettel következtetünk a tumor térfogatára. A jelenleg is alkalmazott közelítőképletünk egy ellipszoid alakot feltételez és a következőképpen számítható:

$$V = \frac{\pi}{2} (\text{szélesség} \cdot \text{hosszúság})^{3/2}. \quad (3.5)$$

A Richter Gedeon Nyrt. együttműködéséből származó kísérleti adathalmaz 14 egér, egyedenként 6 mérését tartalmazza, tehát összesen 84 mérési pont áll rendelkezésünkre tolómérővel és MRI-vel egyaránt. A 3.21. ábra mutatja az MRI mérést, illetve a digitális tolómérő általi mérést a 84 adatpontra. Az ábrán a zöld görbe jelzi az MRI méréseket, a piros görbe pedig a digitális tolómérővel mért megfelelőjét. A kettő közötti relatív, százalékos hiba számításához az MRI méréseket vesszük alapigazsággként, tehát a következőképpen számítjuk a két görbe különbségét:

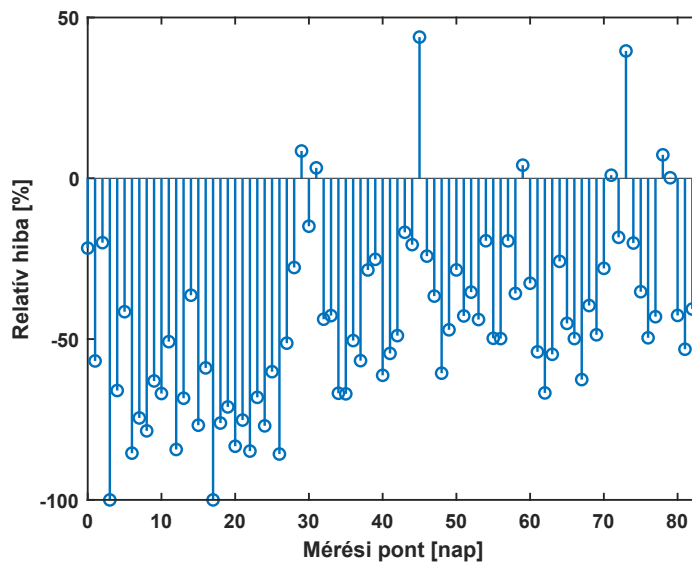
$$\text{relatív hiba} = \left(\frac{\text{digitális tolómérő mérések} - \text{MRI mérések}}{\text{MRI mérések}} \right) \cdot 100 \quad (3.6)$$

A (3.6) számítás eredményét mutatja a 3.22. ábra, melyen egyértelműen látszik a negatív irányú hibák többsége, illetve hogy a hibák negatív és pozitív irányban nem egyforma nagyságrendben mozognak. A függőleges tengely a relatív hibákat mutatja százalékban (a



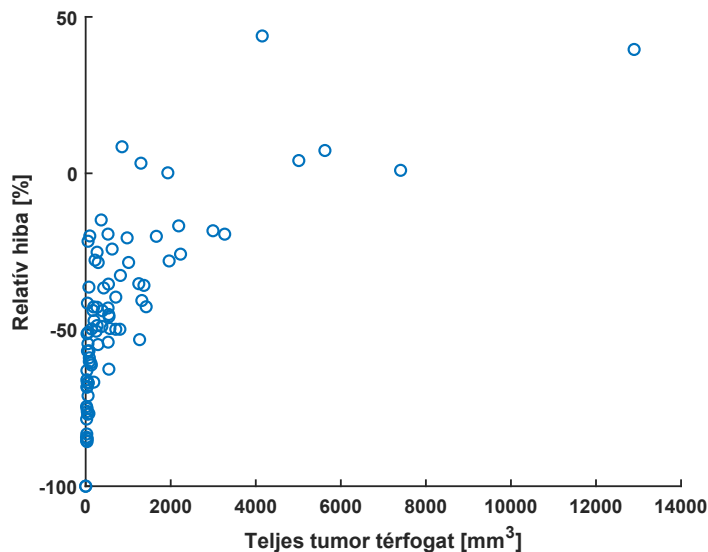
3.21. ábra. MRI-vel mért tumor térfogat mérések (zöld görbe) és digitális tolómérővel mért tumor térfogat mérések (piros görbe).

kék karikák jelölik a tényleges értéket az egyes mérési pontokon). Ez az eredmény már a 3.21. ábra alapján is látható volt, mivel a piros görbe, ami a tolómérővel mért értékeket szemlélteti, a legtöbb esetben a zöld görbe, tehát az MRI mérések alatt van.



3.22. ábra. Digitális tolómérő relatív mérési hibája az MRI mérésekhez képest.

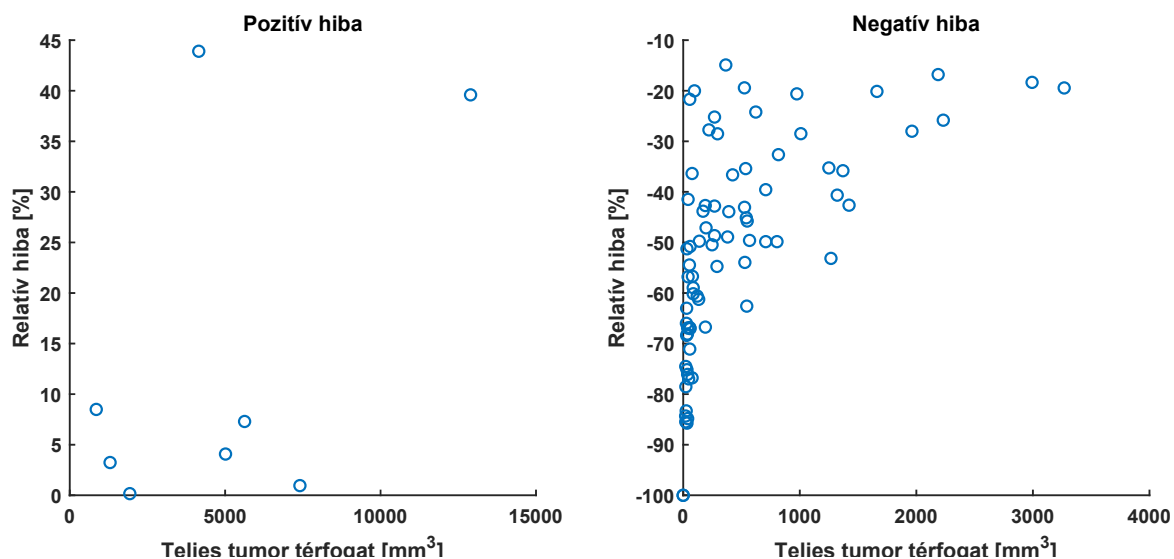
Mivel nekünk fontos a zajmodell felépítése szempontjából, hogy tudjuk azt, hogy az aktuális mérésre mekkora zajt kell majd generálni, így ezeket a relatív hibaértékeket megvizsgáltuk egy szóródás diagramon is. A 3.23. ábra mutatja a függőleges tengelyen a relatív hiba értékét, a vízszintes tengelyen pedig a tumor térfogatok értékét. Összefoglalva tehát, a szóródási diagram tudja nekünk megmondani azt, hogy az egyes tumorméretekhez mekkora nagyságrendű hiba tartozik. Már az előző féléves munkám során is megállapításra került, hogy a zaj erősen függ a tumor térfogatától [19]. A 3.23. ábrán is szépen látszik, hogy a kisebb tumor térfogatokhoz nagyobb mértékű hiba társul, míg a tumor növekedésével egyre



3.23. ábra. Digitális tolómérő és MRI mérések (alapigazság) közötti relatív mérési hibák szóródási diagramja.

kisebb a mérési zaj.

A pozitív és negatív hibákat külön vizsgáljuk, mert nem lehetünk biztosak abban, hogy mind a kettőre ugyanaz a valószínűségi eloszlás illeszkedik a legjobban. A hibák külön lebontott szóródási diagramját a 3.24. ábra mutatja. Szintén látszik a térfogatoktól való erős függés, viszont ezt csak a negatív hibákra jelenthetjük ki. A pozitív hibákból nem áll rendelkezésünkre elegendő számú ahhoz, hogy megvizsgáljuk őket, illetve hogy meghatározzuk a rájuk legjobban illeszkedő eloszlást. Ez a mintanagyság kevés ahhoz, hogy megalapozott döntést hozzunk a legjobb illeszkedéssel kapcsolatban. Pozitív esetben összesen 8 pontunk van, így a negatív hibák összesen 76 mérési ponton jelennek meg. A negatív hibák esetében ez a szám elegendő az illesztéshez.



3.24. ábra. Digitális tolómérő és MRI mérések (alapigazság) közötti pozitív és negatív relatív mérési hibák szóródási diagramja.

A digitális tolómérő és MRI által mért mérések közötti negatív hibákat vesszük alapul tehát a zajmodell felépítéséhez. Célunk egy olyan transzformáció keresése, mellyel transzformálva a zajt, az függetlenebb lesz a tumor térfogatától. A függetlenítésre azért van szükség, hogy illeszteni tudjuk a valószínűségi eloszlásokat. A transzformáció keresését a valószínűségi eloszlások segítségével végezzük. A 3.4. alfejezetben részletezzük a próbált transzformációkat, illetve a valószínűségi eloszlásokra való illeszkedés kiértékelését.

3.4. Zaj függetlenítéséhez felhasználható transzformáció keresése valószínűségi eloszlások illesztésével és a zajmodell felépítése MRI alapján

Első lépésben a zajt transzformáljuk, tehát kiegyenlítjük a nagyságrendeket a zaj értékei között, illetve függetlenítjük a zajt a tumor térfogatától. A heurisztika alapja az volt, hogy a mérési hiba esetünkben az egyik dimenzió méréséhez (szélesség/hosszúság) adódik hozzá, de mi térfogatokkal dolgozunk a valóságban, ami ezek köbétől függ, emiatt a felírt transzformációk gyökös, illetve köbös kifejezések. Az általunk leírt összefüggésben a hibát szorozzuk meg egy transzformációval:

$$\tilde{e} = \Phi(y)e, \quad (3.7)$$

ahol $\tilde{e}$ a transzformált hiba, y a tumor térfogat, és $\Phi(y)$ függetlenítéshez használt transzformáció, amelyből négy változatot is kipróbáltunk:

$$\Phi_1(y) = f\sqrt[3]{y + g}, \quad (3.8)$$

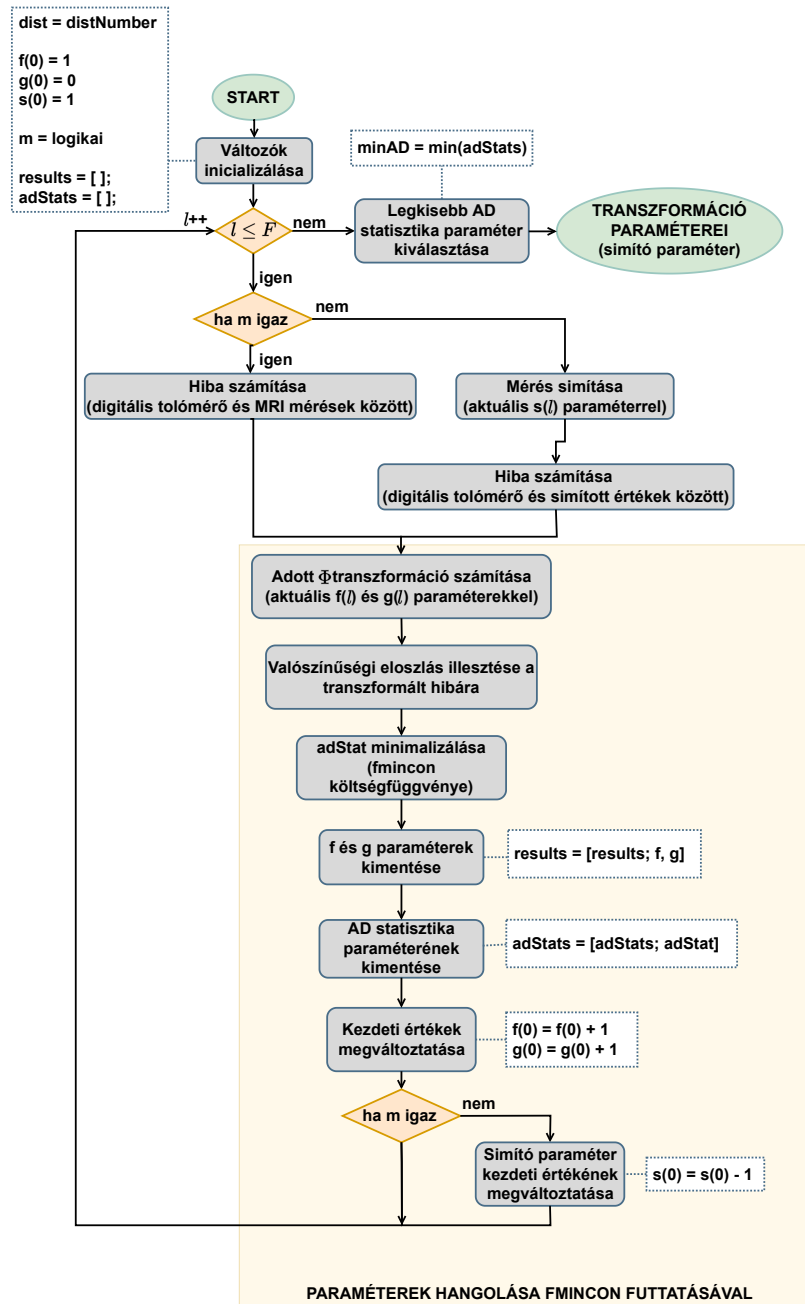
$$\Phi_2(y) = f\sqrt[3]{y} + g, \quad (3.9)$$

$$\Phi_3(y) = \sqrt[4]{y + g}, \quad (3.10)$$

$$\Phi_4(y) = f\left(\frac{1}{3}y + g\right), \quad (3.11)$$

ahol f és g paramétereket a 3.25. ábrán látható algoritmus alapján hangoljuk be. A cél az, hogy a hibát transzformáljuk (3.8)-(3.11) egyikével, majd a transzformált hibára eloszlást illesztünk és futtatunk egy Anderson-Darling tesztet. Az Anderson-Darling teszt (vagy statisztika) egy statisztikai értéket ad vissza, ami azt mutatja meg, hogy egy adott minta egy adott valószínűségi eloszlásból származik-e [20, 21]. A 3.25. ábrán lévő algoritmus az f és g paramétereket úgy hangolja, hogy az Anderson-Darling teszt statisztikai értékét minimalizálja, mivel ennek a változónak az értéke minél kisebb, annál nagyobb valószínűséggel tartozik a minta a vizsgált eloszlásba. Tehát, keressük azt az f és g paramétert, mellyel transzformálva a hibát, az a lehető legjobban illeszkedik az aktuálisan vizsgált eloszlásra. A következőkben az algoritmus részletes működését mutatjuk be.

A folyamat elején létrehozuk a változókat. A dist változóval választhatjuk ki a vizsgálni kívánt eloszlást a 3.5. táblázat alapján. A dolgozat további szakaszaiban az eloszlásokra szintén a táblázatban hozzárendelt sorszámukkal (ID) hivatkozunk. Ezután megadjuk a hangolendő paraméterek kezdeti értékét ($f(0)$ és $g(0)$). Az f paraméter kezdetben soha nem lehet nulla, mivel a (3.8), (3.9) és (3.10) egyenletekben szorzótényező, mely nulla értéket kapva kinullálná a zajt is a függetlenítés közben. A g paramétert kezdetben nullából indítva



3.25. ábra. A függetlenítéshez felhasznált transzformáció paramétereit hangoló algoritmus, mely opcionálisan a simított mérésekhez tartozó simítási paramétert is hangolja az Anderson-Darling teszt paraméterének minimalizálásával.

keressük. Az s paraméter egy simítási paraméter, melyet arra használunk, ha az algoritmus a hibát nem az MRI és tolómérő különbségből, hanem a tolómérővel mért mérések és ennek simítása közötti különbségből veszi (a fejezet későbbi szakaszában részletezzük). Ennek kezdeti értéke egy, mivel ez a paraméter egy $[0, 1]$ intervallumon mozgó paraméter. Mikor értéke egyhez közeli, akkor kevésbé simítja le a görbét, ellenkező esetben nullához közelítve egy nagyobb mértékű simításra számíthatunk. Az m logikai paraméterrel állít-

Valószínűségi eloszlás	ID
Exponenciális eloszlás	1
Birnbaum-Saunders eloszlás	2
Szélsőérték eloszlás	3
Gamma eloszlás	4
Általánosított szélsőérték eloszlás	5
Általánosított Pareto eloszlás	6
Fél-normális eloszlás	7
Inverz Gauss eloszlás	8
Logisztikus eloszlás	9
Log-logisztikus eloszlás	10
Lognormál eloszlás	11
Nakagami eloszlás	12
Normális eloszlás	13
Rayleigh eloszlás	14
Rician eloszlás	15
Stabil eloszlás	16
Student-féle t eloszlás	17
Weibull eloszlás	18

3.5. táblázat. A felhasznált valószínűségi eloszlások és a hozzájuk rendelt sorszámok.

hatjuk, hogy a simítási paraméter hangolását szeretnénk-e vagy sem. A `results` vektorba mentjük ki majd az aktuális hangolt paramétereket, így ennek értéke kezdetben egy üres vektor. A `adStats` vektor pedig az Anderson-Darling statisztikai érték mentésére szolgál majd minden iterációban. Egy iteráció veszi kezdetét, melynek l ciklusváltozója F értékig megy. F egy tetszőleges érték, melyet mi 300-ra állítottunk be. Ez annyit jelent, hogy 300 féle kezdeti értékből keresi az algoritmus a megfelelő f és g paraméterértékeket. A ciklus belsejében megvizsgáljuk, hogy a mérési hibát ténylegesen számoljuk az MRI által mért mérésekből vagy csak a tolmérővel mért értékeink vannak meg (esetünkben a jelenlegi kísérletekre csak ez az eset igaz) és simítást alkalmazunk majd az alapigazság előállítására. Utóbbi esetben a zajra csak becslést tudunk adni a simítás által, melyek simítási paraméterét szintén hangolhatjuk az algoritmussal. A hiba számítása után az algoritmus kiszámítja a transzformált hibát a (3.8)-(3.11) egyikével, majd illeszti a választott valószínűségi eloszlást a negatív zajra. Ezután kiszámítja az Anderson-Darling teszt eredményét, melyet az `adStat` változó tárol el. Ezután mentjük az f és g paraméter aktuális értékét, ezt követően az `adStat` értékét is és megváltoztatjuk a kezdeti feltételeket. A 3.25. ábrán a sárga keret jelöli az algoritmusnak azt a részét, amely a paraméterek hangolásáért felel. Ehhez a Matlab egy beépített függvényét használjuk. Az `fmincon` függvény egy minimalizálóként működik, tulajdonképpen egy probléma minimumát keresi különböző feltételek mellett. Esetünkben az Anderson-Darling teszt eredményét minimalizálja az f és g paraméterek hangolásával.

Az algoritmust felhasználva a [19] TDK dolgozatban bemutatott eloszlások mindegyikére illesztettük a függetlenített zajt mind a négy transzformációt kipróbálva. Ezek eredményét mutatják a 3.6-3.9. táblázatok transzformációkra lebontva. A célunk az, hogy kiválasszuk azt a transzformációt, mellyel a lekisebb Anderson-Darling statisztikai értéket kapjuk. Minden táblázat első oszlopa a valószínűségi eloszlás sorszáma, a második oszlop-

ban az f paraméter hangolásának kezdeti értéke van, ezt $f(0)$ jelöli. A harmadik oszlopban a g paraméter kezdeti értéke található, ezt $g(0)$ jelöli. A negyedik és ötödik oszlop az f és g paraméterek intervallumát foglalja össze az egyes valószínűségi eloszlásokhoz tartozóan. Tehát azt, hogy a hangolás során az $f_{\min}$ -nek mekkora intervallumot adtunk meg, amelyben a paraméter értéke mozoghat a hangolás során. Ez az f paraméternél kezdetben $[0;800]$ közötti, ami csak abban az esetben módosul, ha a behangolt paraméterérték közel van a limithez, vagy maga a limit. Ilyenkor ráengedtük még egyszer az algoritmust, meg-növelve az intervallumot. A g paraméternél pedig $[0;800]$ közötti a kezdeti intervallum. A hatodik és hetedik oszlop a behangolt paraméterek értékét mutatja, az utolsó oszlop pedig az Anderson-Darling statisztikai értéket (AD stat.), mely minél kisebb annál jobban illesz-kedik a vizsgált eloszlás a transzformált zajra.

ID	$f(0)$	$g(0)$	f intervallum	g intervallum	f	g	AD stat.
1	20	10	$[0,1; 800]$	$[0; 800]$	21,2601	$7,18e - 05$	13,5934
2	14	4	$[0,1; 800]$	$[0; 800]$	79,1192	$2,94e - 06$	0,6337
3	193	183	$[0,1; 800]$	$[0; 800]$	192,2945	252,1495	0,3580
4	30	20	$[0,1; 800]$	$[0; 800]$	29,0197	2,6710	0,2831
5	274	264	$[0,1; 800]$	$[0; 800]$	270,5885	266,6929	0,1823
6	295	285	$[0,1; 800]$	$[0; 800]$	293,1813	222,9182	6,3495
7	38	28	$[0,1; 800]$	$[0; 800]$	37,0012	0,0001	11,5357
8	31	21	$[0,1; 800]$	$[0; 800]$	30,0113	3,0617	0,6562
9	278	268	$[0,1; 800]$	$[0; 800]$	279,4325	794,6474	0,2608
10	154	144	$[0,1; 800]$	$[0; 800]$	153,5603	6,1281	0,2314
11	32	22	$[0,1; 800]$	$[0; 800]$	31,0129	3,0144	0,5432
12	32	22	$[0,1; 800]$	$[0; 800]$	31,0355	2,1664	0,3090
13	252	542	$[0,1; 800]$	$[0; 1000]$	252,8823	836,0499	0,1968
14	12	2	$[0,1; 800]$	$[0; 1000]$	11,0037	0,0349	3,5297
15	60	550	$[0,1; 1000]$	$[0; 800]$	11,7354	889,4794	0,2016
16	135	125	$[0,1; 800]$	$[0; 800]$	103,2698	2,6305	0,1978
17	256	246	$[0,1; 800]$	$[0; 800]$	251,1756	797,2271	0,2065
18	260	250	$[0,1; 800]$	$[0; 1000]$	259,7756	842,6568	0,1921

3.6. táblázat. Φ_1 -es transzformációval függetlenített, digitális tolómérőhöz tartozó negatív hibák (MRI mérések alapján számítva) illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve zölddel a legjobban illeszkedő valószínűségi eloszlást (Általánosított szélsőérték eloszlás).

A 3.6. táblázat a Φ_1 -es (3.8) transzformációval függetlenített zaj illeszkedését foglalja össze. Négy eloszláson kívül mindegyik AD statisztika egy alatti értéket vesz fel. A legjobban illeszkedő eloszlások az általánosított szélsőérték, normális, stabil és Weibull eloszlások lettek. Mindegyik 0,1 körüli AD statisztika értéket vesz fel. Ha a legjobbat kell kiválasztani a táblázat alapján, akkor a Φ_1 -es transzformációval függetlenített zaj az általánosított szélsőérték eloszlásra illeszkedik a legjobban, 0,1823-as AD statisztika értékkel. Tehát a táblázat szerint a transzformációnk ebben az esetben a következőképpen néz ki:

$$\Phi_1(y) = 270,5885 \sqrt[3]{y} + 266,6929 \quad (3.12)$$

ID	$f(0)$	$g(0)$	f intervallum	g intervallum	f	g	AD stat.
1	18	8	[0,1; 800]	[0; 800]	20,1955	0,3073	13,6464
2	284	383	[0,01; 800]	[0; 800]	0,0102	408,4142	1,2718
3	17	7	[0,1; 800]	[0; 800]	4,8417	35,7959	0,2141
4	305	2295	[0,1; 800]	[0; 5000]	0,1001	2335,1775	0,6344
5	93	83	[0,1; 800]	[0; 800]	12,5043	108,9257	0,2402
6	58	48	[0,1; 800]	[0; 800]	0,6352	12,0588	3,8343
7	86	76	[0,1; 800]	[0; 800]	149,2962	1,9662	11,6043
8	5	4	[0,01; 800]	[0; 1500]	0,0102	1495,9676	1,3555
9	31	21	[0,1; 800]	[0; 800]	22,5693	32,0587	0,2342
10	295	285	[0,1; 800]	[0; 800]	373,4194	202,0871	0,9198
11	222	3212	[0,1; 800]	[0; 800]	0,1001	8999,0701	1,1628
12	4	3	[0,1; 800]	[0; 800]	2,6714	4,3970	0,6329
13	25	15	[0,1; 800]	[0; 800]	1,1731	55,0660	0,3125
14	174	164	[0,1; 800]	[0; 800]	312,9618	14,4794	3,7900
15	181	230	[0,1; 800]	[0; 800]	0,0101	300,9732	0,2802
16	130	120	[0,1; 800]	[0; 800]	351,4332	1,6616	0,2781
17	258	248	[0,1; 800]	[0; 800]	60,1436	26,1502	0,2525
18	280	779	[0,01; 800]	[0; 1000]	0,0101	877,5183	0,3013

3.7. táblázat. Φ_2 -es transzformációval függetlenített, digitális tolómérőhöz tartozó negatív hibák (MRI mérések alapján számítva) illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve zölddel a legjobban illeszkedő valószínűségi eloszlást (Szélsőérték eloszlás).

A Φ_2 -es transzformációval (3.9) függetlenített negatív zajra is elvégeztük az illesztéseket. Az eredményeket a 3.7. táblázat foglalja össze. Ennél a transzformációnál nem voltak jobb eredmények, mint az előzőnél, de itt sem rosszak az AD statisztika értékek. Ebben a táblázatban a 0,2 körüli értékek lettek a legkisebbek a következő eloszlásokkal: szélsőérték eloszlás, általánosított szélsőérték, logisztikus eloszlás, Rayleigh, Rician és Stabil eloszlás. A legjobban illeszkedő eloszlás a szélsőérték eloszlás volt 0,2141-es AD statisztika értéket kapott. A második transzformációnk tehát a táblázat szerint:

$$\Phi_2(\mathbf{y}) = 4,8417\sqrt[3]{\mathbf{y}} + 35,7959. \quad (3.13)$$

A harmadik transzformáció a Φ_3 -as transzformáció, ahol a legjobban illeszkedő eloszlás a Weibull eloszlás lett. Mellette megjelennek még hasonló AD statisztikával eloszlások, mint például a normális eloszlás, log-logisztikus, logisztikus és lognormál eloszlás, általánosított Pareto és fél-normális eloszlás is. A Weibull-nál már megfigyelhető az is, hogy a g paraméter intervallumának maximumát meg kellett emelni 800-ról 5000-re az optimális megoldás érdekében. Eddig ez az eloszlás és ez a transzformáció adta ki a legkisebb AD statisztika paramétert, amelynek értéke 0,1394. A harmadik transzformáció ezáltal:

$$\Phi_3(\mathbf{y}) = {}^{0,7919}\sqrt{\mathbf{y}} + 4516,2918. \quad (3.14)$$

A negyedik a Φ_4 -es transzformációnk ((3.11)) egy kicsit eltér a többitől, hiszen sem köbös, sem pedig gyökös kifejezés nem található benne. A 3.9. táblázat foglalja össze erre a transzformációra az illeszkedés eredményeit. A Rician eloszlás adta a legkisebb AD

ID	$f(0)$	$g(0)$	f intervallum	g intervallum	f	g	AD stat.
1	5	4	[0,1; 800]	[0; 800]	1,2014	0,5100	0,9916
2	4	3	[0,01; 800]	[0; 800]	2,6726	6,6726	0,2138
3	5	4	[0,1; 800]	[0; 800]	6,4773	0,4017	0,2185
4	7	6	[0,1; 800]	[0; 800]	2,5905	3,7617	0,1983
5	5	4	[0,1; 800]	[0; 800]	2,5999	2,8674	0,1910
6	285	284	[0,1; 800]	[0; 800]	1,3540	0,5351	1,1410
7	5	4	[0,1; 800]	[0; 800]	1,4803	0,2107	1,1446
8	4	3	[0,1; 800]	[0; 800]	2,3584	6,8146	0,2231
9	6	5	[0,1; 800]	[0; 800]	3,8052	4,3105	0,1432
10	5	4	[0,1; 800]	[0; 800]	2,5782	4,4132	0,1736
11	2	1	[0,1; 800]	[0; 800]	2,3778	6,5513	0,1878
12	2	1	[0,1; 800]	[0; 800]	3,1337	2,0264	0,3056
13	301	1300	[0,1; 800]	[0; 5000]	0,8127	4300,37261	0,1455
14	3	2	[0,1; 800]	[0; 800]	2,1112	0,4829	0,6129
15	29	19	[0,1; 800]	[0; 800]	2,9833	789,9030	0,2055
16	14	4	[0,1; 800]	[0; 800]	2,9933	1,5875	0,2135
17	11	1	[0,1; 800]	[0; 800]	3,4291	1,1017	0,2559
18	40	30	[0,1; 800]	[0; 5000]	0,7919	4516,2918	0,1394

3.8. táblázat. Φ_3 -as transzformációval függetlenített, digitális tolómérőhöz tartozó negatív hibák (MRI mérések alapján számítva) illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve zölddel a legjobban illeszkedő valószínűségi eloszlást (Weibull eloszlás).

statisztika értéket (0,1546), amely a második legjobb AD statisztika érték, ha a négy transzformációval való függetlenítést egyben vizsgáljuk. Ennél a transzformációnál mind a két paraméter intervallumát meg kellett változtatni a futtatás során. Magasabbak lettek a nagyságrendek. Az f paraméter esetében 2000 a maximum és a behangolt értéke 1000 körüli, így a g paraméternél 5000-re emeltük a maximumot. A negyedik transzformáció a következőképpen írható fel:

$$\Phi_4(\mathbf{y}) = 1094,6865 \left(\frac{1}{3}\mathbf{y} + 3418,3361 \right). \quad (3.15)$$

A négy transzformációt vizsgálva a legjobban illeszkedő eloszlás a Weibull eloszlás lett abban az esetben, ha a hibákat a Φ_3 -as transzformációval függetlenítjük a zajtól. Ehhez a vizsgálathoz az MRI és digitális tolómérő által mért méréseket vettük alapul, azon belül is a negatív mérési hibákat, mivel a pozitív irányúak száma kevés volt a megalapozott eloszlásvizsgálathoz. Azonban a jelenlegi kísérleteink során nem áll rendelkezésünkre kisállat MRI, így egy olyan megoldást kell keresnünk a amellyel a zaj MRI mérések nélkül is reálisan becsülhető. Mivel az MRI-t vettük eddig alapigazsággként, így kijelenthető, hogy a valódi zaj eloszlása a jelenlegi eszközeink és információink alapján a Weibull eloszláshoz hasonlít leginkább a vizsgált eloszlások közül.

A következő lépésben feltesszük, hogy nem ismerjük az MRI mérést csak a hozzá tartozó digitális tolómérővel mért méréseket. Azt tudjuk, hogy a Weibull eloszláshoz a leghasonlóbb a mérési zaj eloszlása. A feladatunk az, hogy találjunk egy olyan simítási paramétert a

ID	$f(0)$	$g(0)$	f intervallum	g intervallum	f	g	AD stat.
1	11	1	[0,1; 800]	[0; 800]	11,1482	19,6301	1,9080
2	65	55	[0,1; 800]	[0; 800]	64,0649	5,8740	0,6758
3	65	555	[0,1; 800]	[0; 2000]	64,8783	1222,1630	0,4990
4	255	245	[0,1; 800]	[0; 800]	254,0941	575,2470	0,8165
5	231	721	[0,1; 800]	[0; 2000]	260,0002	749,9995	0,2139
6	226	216	[0,1; 800]	[0; 800]	229,9650	1,4968	1,2865
7	261	251	[0,1; 800]	[0; 800]	260,0105	102,5851	3,5052
8	14	4	[0,1; 800]	[0; 800]	13,2120	10,9338	0,9082
9	23	263	[0,1; 800]	[0; 2000]	271,8415	822,0692	0,1918
10	103	93	[0,1; 800]	[0; 800]	105,1184	407,9779	0,5360
11	123	113	[0,1; 800]	[0; 800]	123,5958	354,8224	0,8599
12	168	158	[0,1; 800]	[0; 800]	167,2501	717,1091	0,5626
13	17	507	[0,1; 800]	[0; 2000]	18,4087	829,6786	0,2856
14	12	2	[0,1; 800]	[0; 800]	15,2642	312,4699	2,2878
15	1097	1097	[0,1; 2000]	[0; 5000]	1094,6865	3418,3361	0,1545
16	13	3	[0,1; 800]	[0; 800]	9,5877	797,9955	0,2921
17	269	259	[0,1; 800]	[0; 800]	796,5176	743,0479	0,2233
18	134	624	[0,1; 800]	[0; 2000]	134,6982	1701,6327	0,2233

3.9. táblázat. Φ_4 -es transzformációval függetlenített, digitális tolómérőhöz tartozó negatív hibák (MRI mérések alapján számítva) illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve zölddel a legjobban illeszkedő valószínűségi eloszlást (Rician eloszlás).

Matlab fit függvényének, mely az úgynevezett 'smoothingpline' simítógörbét felhasználva úgy simítja a tolómérővel végzett méréseket, hogy a kettőjük közötti hiba a lehető lejobban hasonlítson a Weibull eloszlásra. Emiatt a 3.25. algoritmust úgy futtatjuk, hogy adja vissza az optimális simító paramétert is, tehát az m logikai változónk ebben az esetben hamis, így elérjük, hogy az algoritmus a hiba számítását a tolómérő által mért mérések és a tolómérő mérések simításának különbségéből számítsa. Megnézzük, hogy a hangolt simítási paraméterek mellett megkapjuk-e a Weibull eloszlást ezekre a becsült hibákra a legjobbnak, illetve újrahangoljuk a Φ_3 transzformáció f és g paraméterét is a simított értékekből számított hibákra. Ennek eredményét mutatja a 3.10. táblázat. A legkisebb AD statisztika értéket a simított alapigazsággal számolva is a Weibull eloszlás kapta, amely számunkra azt jelenti, hogy a simított mérésekkel adhatunk becslést a hiba mértékére, ha a simítási paramétert 0,8805-re állítjuk be, illetve a zajt a következő összefüggéssel függetleníthetjük a tumor térfogatoktól:

$$\Phi_3(\mathbf{y}) = {}^{1,0032}\sqrt{\mathbf{y} + 107,0034}. \quad (3.16)$$

A többi eloszlásra is végigfuttattuk az algoritmust, hogy megnézzük milyen simítási paramétereket kínál különböző valószínűségi eloszlások mellett. Illetve így az is kiemelhető, hogy a Weibull eloszlás bizonyult a legjobbnak a jelenleg vizsgált eloszlások közül, hiszen az AD statisztika paramétere ennek az eloszlásnak a legkisebb. A kezdeti értékeket a 3.8. táblázat alapján megtartottuk, hiszen a Φ_3 -as transzformációt használjuk fel továbbra is. Az f paraméter intervallumát [0.1;5] közé állítottuk át. Ennek oka az volt, hogy gyakran

adott 400 és még annál is nagyobb paraméterértéket, viszont ez egy gyökös kifejezés esetén szinte már lenulázza a hiba mértékét, így ezt beszabályoztuk. A g paraméter esetében az intervallumot változatlanul hagytuk.

Sorrend	smooth	f	g	AD stat.
1	0,8913	1,5185	101,4136	0,1590
2	0,8870	0,7745	69,5365	0,1105
3	0,0003	3,2819	127,4617	0,1868
4	0,8814	1,2871	84,7327	0,0842
5	0,0362	1,3719	86,9656	0,1636
6	0,0038	0,8713	116,4180	0,0967
7	0,1898	1,6766	9,4521	0,1670
8	0,0152	0,2320	623,4633	0,1608
9	0,0001	2,6900	16,6903	0,1118
10	0,9143	0,5887	354,8083	0,0937
11	0,9347	0,6425	242,2067	0,0784
12	0,8811	1,9793	30,0606	0,0762
13	0,0069	0,9559	799,9878	0,1491
14	0,0099	2,6824	27,2238	0,1549
15	0,0098	2,6717	28,7406	0,1719
16	0,4797	2,0586	299,2636	0,2263
17	0,0001	4,9879	98,3524	0,1753
18	0,8805	1,0032	107,0034	0,0725

3.10. táblázat. Φ_3 -as transzformációval függetlenített, digitális tolómérőhöz tartozó negatív hibák (simított mérések alapján számítva) illesztési eredményei a különböző valószínűségi eloszlásokra, kiemelve zölddel a legjobban illeszkedő valószínűségi eloszlást (Weibull eloszlás).

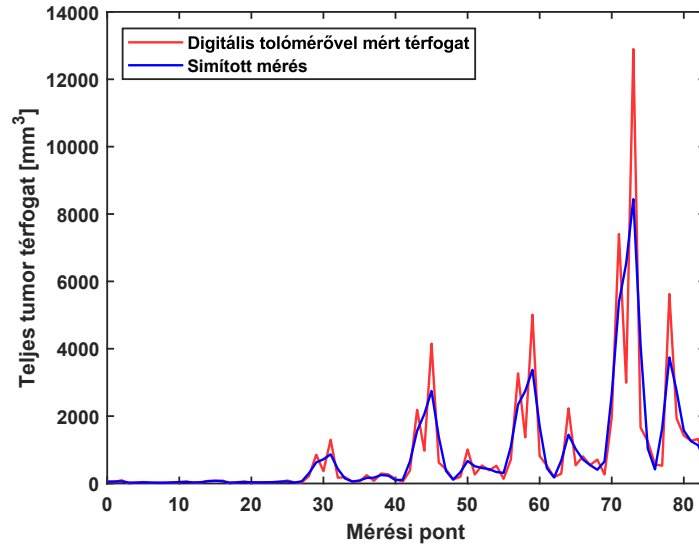
Az általunk meghatározott simítási paraméter ($s = 0,8805$) segítségével a digitális tolómérővel végzett mérést simítottuk. Ennek eredménye látható a 3.26. ábrán. A vízszintes tengelyen a mérési pontokat jelöljük, a függőlegesen pedig a teljes tumor térfogatát. Összehasonlítva a 3.21. ábrával a jelleg nagyon hasonló.

A zajmodell tehát az előző (3.4. alfejezet) részben tárgyaltak szerint a következőképpen építhető fel a Weibull eloszlás alapján:

$$C(x, A_w, B_w) = \begin{cases} \frac{B_w}{A_w} \left(\frac{x}{A_w} \right)^{B_w-1} \exp \left(-\frac{x}{A_w} \right)^{B_w} \cdot \frac{1}{\sqrt[3]{y+g}}, & \text{ha } x \geq 0 \\ 0, & \text{ha } x = 0. \end{cases} \quad (3.17)$$

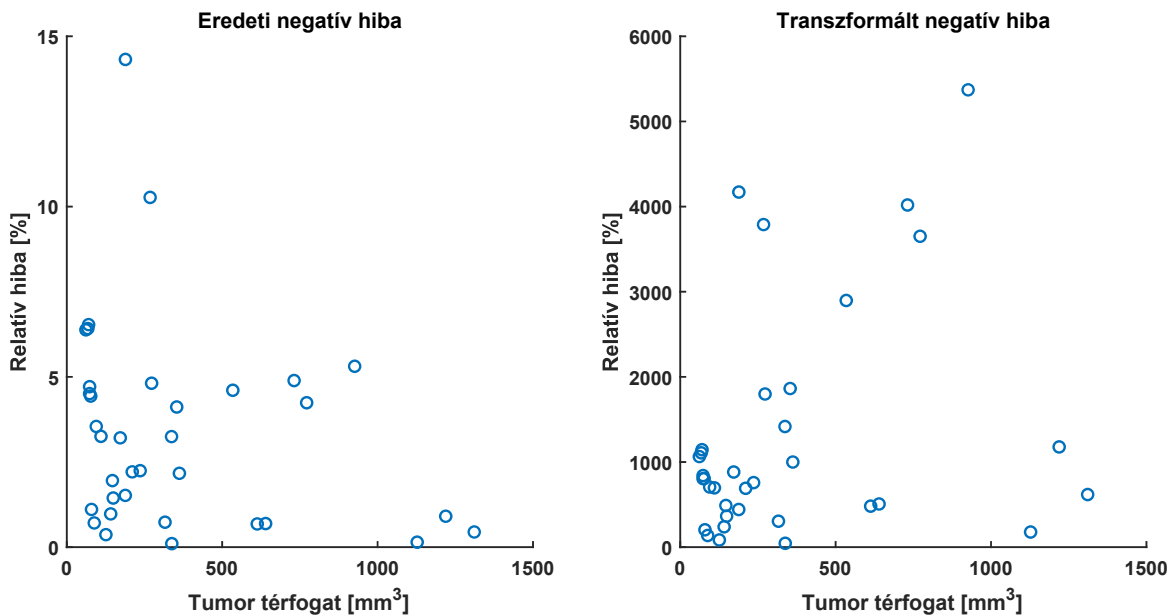
Ez az eloszlás (jelölje ezt most C) írja le a zajt. A jelenlegi, rendelkezésre álló információink alapján ez a legrealisabb, amit le tudunk modellezni. A (3.10) transzformáció inverzével és a Weibull eloszlás segítségével a zaj a mérésekre generálható.

A modell felépítésének első lépése a gyakorlatban, hogy a (3.10) transzformáció és a hozzá tartozó paraméterek segítségével a zajt függetlenítjük a térfogatoktól. A 3.27. ábrán egy kísérleti egér eredeti negatív mérési hibájának szóródási diagramját jelenítettük meg a bal oldalon és ennek a hibának a függetlenített, transzformált verzióját a jobb oldalon.



3.26. ábra. Digitális tolómérővel mért mérési hiba simítása az általunk meghatározott simítási paraméterrel ($s = 0,8805$).

A (3.16). transzformációval elértük, hogy az egyes tumor térfogatokhoz nagyjából hasonló nagyságrendű hibaértékek tartozzanak, így könnyedén illesztettük rájuk az eloszlásokat.



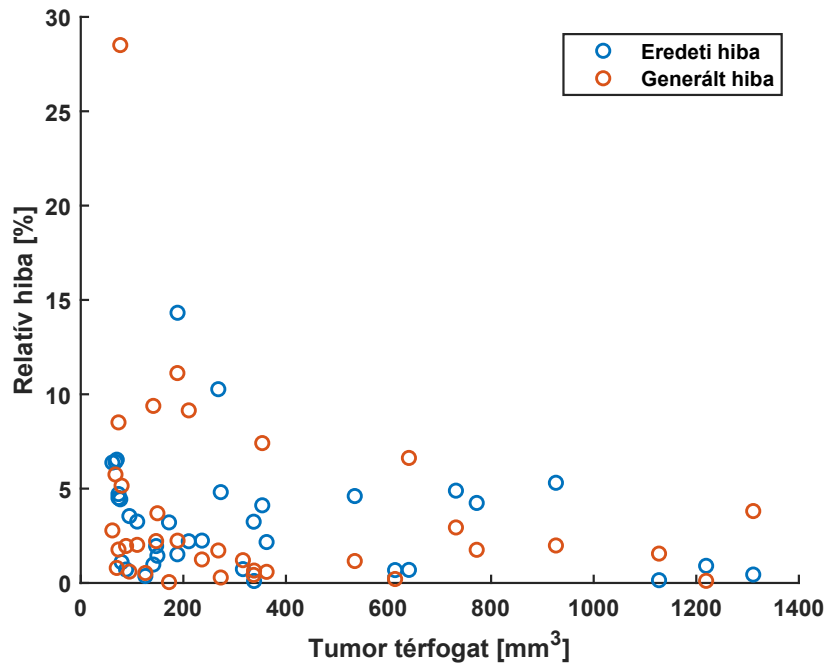
3.27. ábra. 16-os kísérleti egér eredeti negatív mérési hibájának szóródási diagramja (bal oldalon) és a függetlenítet, transzformált hiba (jobb oldalon).

Az eloszlás paramétereit véletlenszerűen generáljuk. A jelenlegi adathalmaz 49 egér mérését tartalmazza. Ezeket a számított simítási paraméterrel lesimítottuk. Számoltunk egy hibát, majd a negatív hibákat transzformáltuk és rájuk illesztettük a Weibull eloszlást. Megkapva ezáltal az eloszlás paramétereit minden egyedre (a pozitív esetre is megvizsgáltuk az AD statisztikákat, illetve a paraméterek szórását, a modell pozitív hibákra is alkalmazható). Ezekből szórást és várható értéket számítottunk, amelyek segítségével generáljuk újra az eloszlás paramétereit, amikor zajt szeretnénk létrehozni. A (3.10) transzformáció inverzével

a hibát a transzformált zajból könnyedén visszakaphatjuk:

$$e = \frac{1}{\sqrt[3]{y+g}} \cdot \tilde{e}. \quad (3.18)$$

Egy ilyen generált zaj eredménye látható a 3.28. ábrán, ami a 16-os egérhez tartozó eredeti negatív hibákat mutatja (szemléltetés kedvéért abszolútértékben) a kék körökkel és az új általunk generált negatív hibákat a narancs színű körökkel. Az ábra jól szemlélteti, hogy reprodukálható a zaj a zajmodell segítségével. A kék köröket jól követik a narancs körök. A kisebb tumor térfogatokon több mérés van. A pozitív zaj exponenciális csökkenését a tumor térfogatok növekedésének függvényében visszakaptuk.

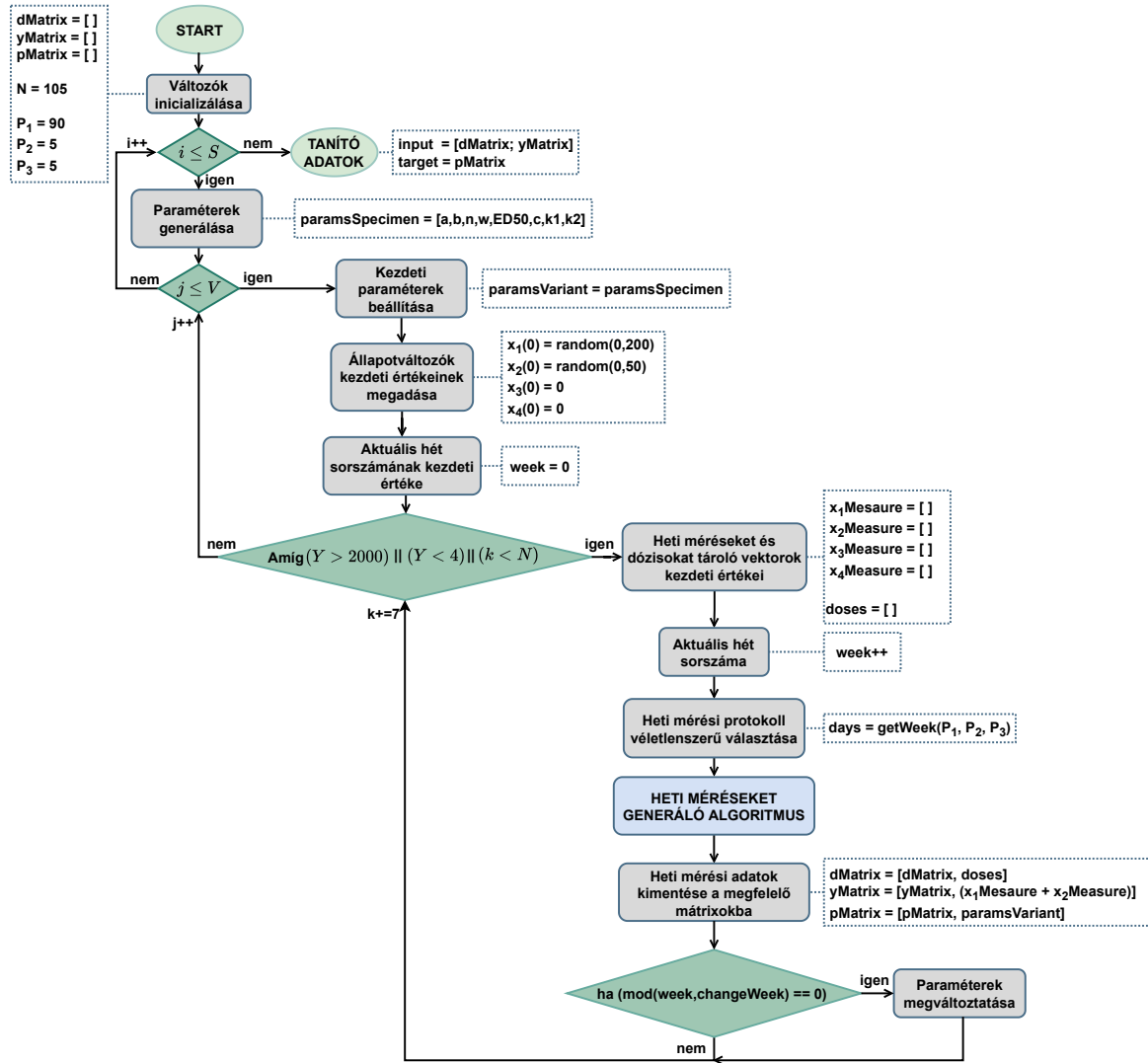


3.28. ábra. A 16-os egér eredeti negatív hibáinak szóródása (kék körök) és az általunk meghatározott eloszlással generált negatív hibák szóródása (narancs körök).

3.5. Tanító adatok generálása a felépített zajmodell beépítésével

Ebben a fejezetben bemutatunk egy (az előző éves munkám alatt készített [22]) tanító adatokat generáló algoritmust, melybe a felépített zajmodellt beépítjük. A tanító adatokat generáló algoritmust a 3.29. és a 3.31. folyamatábrák mutatják.

A folyamat elején minden változót inicializálunk. A *dMatrix* fogja tárolni a heti dózisokat és a *yMatrix* pedig a dózisokhoz tartozó heti méréseket. A *pMatrix* a heti paraméterértékek tárolására szolgál majd. Ezek kezdeti értéke az inicializálás során egy-egy üres vektor. Az *N* változó adja meg a kísérlet hosszát napokban, ez a jelenlegi kísérleti protokoll szerint 105 nap. A *P*₁, *P*₂ és *P*₃ változó a három különböző heti protokoll valószínűségét adja meg. Az első protokoll valószínűsége 90% (*P*₁ = 90), a második és harmadik protokoll



3.29. ábra. Tanító adatokat generáló algoritmus folyamata

generálásának a valószínűsége 5% ($P_2 = 5$ és $P_3 = 5$). Tehát az első eset fog leggyakrabban legenerálódni, mely alap protokollnak számít a jelenlegi kísérleti eljárás során.

Ezután egy ciklus kezdődik, ami a különböző egereken megy végig (S jelöli az egerek darabszámát). Az i -edik egér esetében legenerálódnak az adott egér paraméterei. Ez úgy valósul meg, hogy az egérkísérlet alapján minden paraméterre van egy minimum és egy maximum értékünk melyek között egy random számot generálunk, mint paraméterérték. Ezt minden paraméterre elvégezve állnak elő az i -edik egér egyedi paraméterei. A paraméterek generálására egy külön függvényt készítünk, mely bemenetként a paraméterek minimum és maximum értékét várja. A paraméterek minimum, illetve maximum értékei a 3.11 táblázatban tekinthetők meg [19].

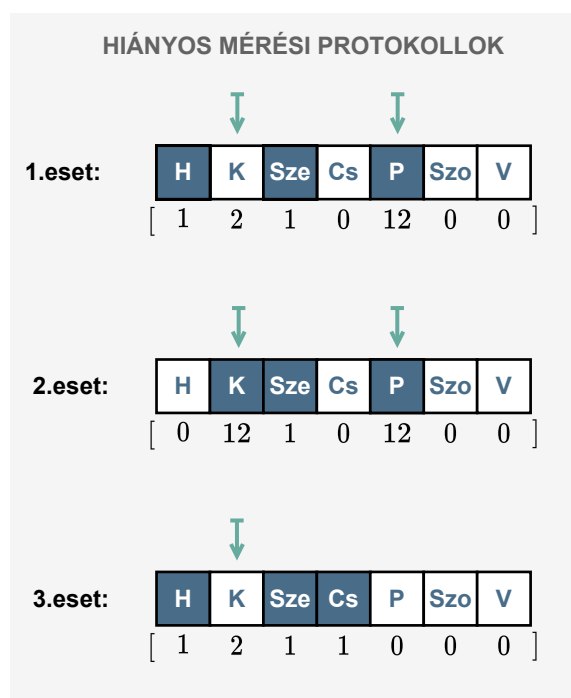
Ezután egy újabb iteráció veszi kezdetét, mely a különböző egerek variánsait generálja le (minden i -edik egérnek V darab variánsát hozzuk létre). A ciklusból akkor lépünk ki, ha j ciklusváltozó eléri a maximális variáns számot (V). Egy egér variánsai mindig az egér paraméterértékeit kapják meg kezdetben, de fontos változás az eddigi munkámhoz képest, hogy ezek most csak kezdeti paraméterértékek, amelyek idővel változni fognak. Tehát egy

Paraméter	Minimum érték	Maximum érték
a	0,5627	1,7245
b	0,0113	5,3656
n	0,4383	1,6300
w	$6,1007e-04$	38,3816
ED_{50}	0,0129	58,3554
c	1,4247	1,4436
k_1	11,0058	11,3104
k_2	16,1376	16,5169

3.11. táblázat. Paraméterek valódi mérésekből származó minimum és maximum értéke

egér variánsain azt értjük, hogy a variáns paramétereinek kezdeti értékei megegyeznek az egér paraméterértékeivel. Ahány variánsa van egy egérnek, annyszor indítjuk a paramétereket ugyanabból a kezdeti értékből. Fontos továbbá, hogy a variánsok minden esetben különböző kezeléseken esnek át, úgy ahogy ez az előzetes munkám során is megtörtént [17, 23], hiszen a dózisok véletlenszerűen generálódnak, viszont itt nem csak a befecskendezett injekció dózisa különbözik, hanem a mérési protokoll is pár hetente megváltozhat a 3.30. ábrán látható esetek alapján. A következő lépés tehát, hogy a **paramsVariant** vektort, mely egy variáns paramétereinek kezdeti értékét fogja tárolni, egyenlővé tesszük a már legenerált **paramsSpecimen** vektorral, mely egy egyed/egér paramétereit tartalmazza. Miután az i -edik egér j -edik variánsának beállítottuk a paraméterek kezdeti értékét, beállítjuk a differenciálegyenlet-rendszerünk állapotváltozóinak kezdeti értékeit is, melyekre majd az ode45 megoldónak szüksége lesz. Mivel időablakos paraméterbecslés a célunk, így ezek kezdeti értékét is véletlenszerűen variáljuk, hogy minél többféle esetet lefedjünk és a hálózat a kísérlet bármely szakaszában alkalmazható legyen. Az élő tumorsejt (x_1) térfogata, melyet $x_1(0)$ jelöl, kezdetben 0 mm^3 és 200 mm^3 között generálódik. A halott tumorsejt térfogatának kezdeti értéke ($x_2(0)$) is egy véletlen érték, azonban ez egy kisebb intervallumban mozog, 0 mm^3 és 50 mm^3 között. A vérben lévő gyógyszerszint kezdeti értéke ($x_3(0)$) nulla és a szövetekben lévő gyógyszerszint kezdeti értéke ($x_4(0)$) is nulla, hiszen még nem történt injekciózás, így a gyógyszer sem áramlott át a vérből szövetekbe.

A paraméterek értékeit minden variáns esetében hetente, illetve pár hetente változtatjuk meg, emiatt szükségünk van arra, hogy a tanító adat generálás során követni tudjuk az aktuálisan generált hét sorszámát. Azt, hogy éppen hányadik heti méréseket generáljuk az adott variánshoz, a *week* változó tárolja, melynek kezdeti értéke minden variáns esetében nulla, mivel még egy hetet sem generáltunk le. Ezután egy *while* ciklus indul, melynek három leállási feltétele van: ha a teljes tumor térfogat a héten meghaladja a 2000 mm^3 -t vagy 4 mm^3 alá csökken, illetve ha az aktuális nap eléri a maximális kísérleti napok számát. A folyamatábrán Y jelöli a heti teljes tumor térfogatot, k ciklusváltozó az aktuális mérési napot, és N a kísérlet idejét, ami 105 nap. A három feltételt logikai vagy-kapcsolat választja el egymástól, tehát bármelyik bekövetkezik a heti mérés generálása során, akkor az algoritmus azt a hetet még befejezi, majd leáll. A teljes tumor térfogat megszorításai a valódi kísérleteket követik. Ha a tumor térfogata meghaladja a 2000 mm^3 -t, akkor az állatot sajnos el kell altatni, mivel az elviselhető tumor túl nagyra nő. A másik véglet pedig az, amikor a tumor térfogat 4 mm^3 -nél kisebb, mivel ebben az esetben olyan kicsi, amit már nem tudunk mérni a kísérletek során alkalmazott digitális tolómérővel. Ebben az esetben az algoritmus



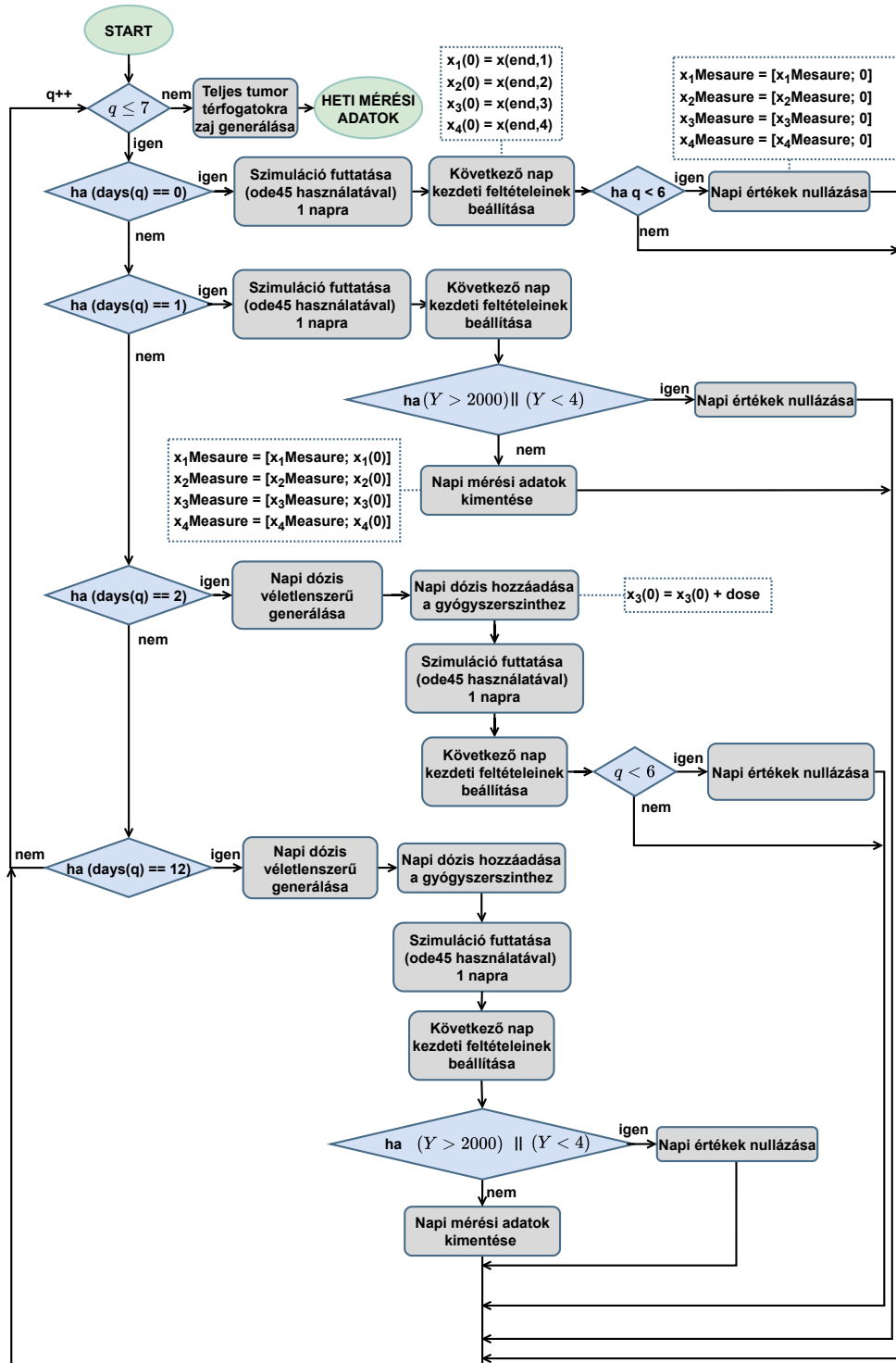
3.30. ábra. Hiányos mérési protokollok a jelenlegi kísérletekben

szintén kilép a ciklusból és egy új variánssal folytatja a generálást. A valóságban ilyenkor abbahagyjuk a kezelést és amikor a tumor újra nőni kezd, akkor folytatjuk, tehát ez a valódi eset a szimuláció során ténylegesen felfogható egy új variánsként.

A ciklusban létrehozunk négy vektort, melyek a négy állapotváltozó heti értékeit fogják tárolni, napi bontásban. Ezek a Measure-végű vektorok, amelyek kezdetben üresek. A heti dózisoknak is létrehozunk egy üres vektort `doses` névvel. Ezt követően megnöveljük eggyel az aktuális hét sorszámát, majd legeneráljuk a jelenlegi hétre vonatkozó mérési protokollt, melyet `days` változó fog tárolni. Ezt egy függvény fogja nekünk megválasztani (`getWeek`), melynek három bemeneti paramétere az esetek valószínűségét adja meg sorrendben, ezek a P_1 , P_2 és P_3 valószínűségek. A heti protokoll választása után legenerálódnak a mérések az adott hétre (ezt a későbbiekben részletezzük, az al folyamat megtekinthető a 3.31. ábrán). A heti mérési adatokat kimentjük a megfelelő mátrixokba, majd ha az aktuális hét sorszáma megfelelő, akkor megváltoztatjuk a paraméterértékeket. A paraméterek pontosan akkor változnak, amikor a hét sorszáma osztható maradék nélkül egy bizonyos `changeWeek` értékkel. Ez az érték egy és négy közötti egész szám, tehát a paraméterek 4, 3, 2 vagy 1 hétig egyformák egy variáns esetében majd bizonyos százalékkal önmagukhoz képest megváltoznak pozitív vagy negatív irányba véletlenszerűen. Az algoritmus leáll, ha az összes egyed összes variánsa elkészült.

A 3.31. ábra a heti mérések generálásának algoritmusát mutatja. Ez az algoritmus az előző folyamat (3.29. ábra) kékkel kiemelt pontját részletezi. Amikor ehhez a részhez érünk, akkor már a heti mérési protokoll kiválasztásra került, és legenerálódott az adott héthez tartozó vektor is, melyet a 3.30. ábrán is megjelentítettünk. Ez az algoritmus tulajdonképpen ezen a vektoron megy végig és elvégzi az egyes napi teendők szimulációját.

Az elején egy ciklus veszi kezdetét, mely egy hetet szimulál, tehát addig fut, amíg a q ciklusváltozó el nem éri a 7-et. Ezután ha a vektorban 0 érték szerepel az adott napon,



3.31. ábra. Heti méréseket generáló algoritmus

tehát nincs sem mérés sem dózis aznap, akkor az ode45 függvénymegoldó segítségével leszimuláljuk az állapotváltozók aznapi értékét (tehát az adott napi tumor térfogat változásokat) és ezek végértékét beállítjuk a következő nap kezdeti feltételeként. Azért a vektorok végértékét használjuk fel, mert azok tartoznak az 1-es időpillanathoz, mivel az ode45 függvénymegoldó mindig egy napot szimulál, így idővektora egy nullától egyig tartó vektor. Az

algoritmus minden olyan pontján, ahol az ode45 lefuttatásra kerül, megvalósul a szakadás, tehát az impulzív bemenet, melyet (2.6) összefüggéssel írtunk le. Ezután meg kell vizsgálnunk, hogy ez a teendőmentes nap hétköznapra vagy hétvégére esik. Ha hétköznapra esett, akkor q ciklusváltozónk kisebb, mint 6, tehát ez egy hiányos mérési nap, így a heti méréseket tároló Measure-végű vektorokba a napi mérési érték helyett egy nulla konstans értéket teszünk, majd megyünk tovább a következő napra.

Ha a heti protokoll következő napját leíró szám értéke 1, akkor az számunkra egy mérési napot jelöl. Az ode45 segítségével újra futtatunk egy napi szimulációt, kimentjük a következő nap kezdeti feltételeit, majd megvizsgáljuk a mérési értékeket. Ha a tumor térfogata 4 mm^3 és 2000 mm^3 között van, akkor kimentjük az állapotváltozók értékeit, mint napi mérési eredmény a Measure-végű vektorokba. Ha azonban a tumor aktuális térfogata nem megfelelő, akkor kinullázzuk ezt a napi mérési értéket is, majd ugrunk a következő napra.

Ha 2-es értékű nap következik, akkor befecskendezés történik az adott napon. Ilyenkor generálunk egy véletlenszerű napi dózist 0.1 mg/kg és 8 mg/kg között, mivel a valódi mérésekben is ezen a skálán mozognak a befecskendezett mennyiségek. Ezt követően leszimuláljuk az injekció beadását, a vérben lévő gyógyszerszinthez ($x_3(0)$) hozzáadjuk a dózis értékét, majd ezzel a módosított kezdeti feltétellel szimuláljuk le a méréseket az ode45 segítségével. Beállítjuk az új kezdeti feltételeket a következő napra és nullázzuk a méréseket a Measure vektorokban, mivel ezen a napon csak befecskendezés történt, mérés nem.

Az utolsó eset, amikor egy napon mérés és befecskendezés is történik. Ezt a napot 12-vel jelöltük a heti protokollt leíró vektorban. Ilyenkor szintén generálódik egy dózis, melyet hozzáadunk a vérben lévő gyógyszerszinthez, majd így szimuláljuk le a méréseket az adott napra. Kimentjük a következő nap kezdeti feltételeit, majd ellenőrizzük a teljes tumor térfogatának méretét. Ha megfelel, akkor kimentjük a mérési adatokat, ha nem felel meg, akkor nullázzuk a napi mérési értékeket.

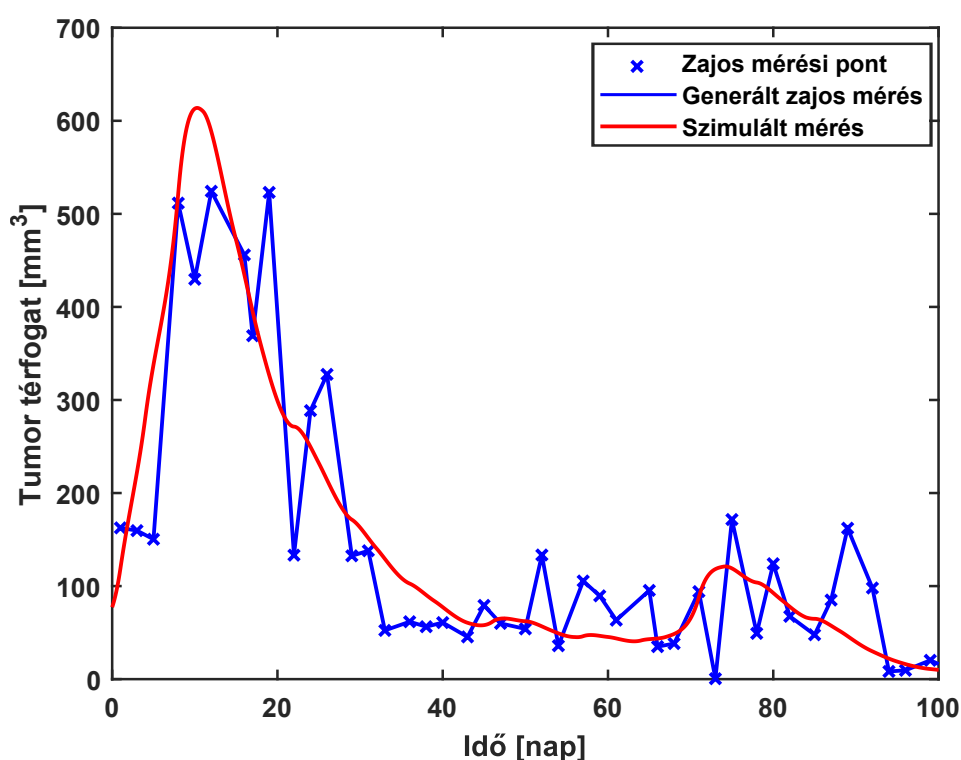
Az algoritmus minden heti protokoll esetén lefut, tehát a legjobb esetben 15-ször, hiszen 105 napos a kísérlet, mely 15 hetet jelent. Persze előfordul, hogy a teljes tumor térfogata egy variáns esetében nem megfelelő, ekkor kevesebb hétig él az adott variáns. Ha esetleg a hét közepén ér el a tumor egy kritikus értéket, akkor azon a héten minden további mérés nullázásra kerül, majd új variáns következik. A tanult hálózatok számára 100 egér 80 variánsát generáltuk le, ez összesen 8000 egeret jelent. A tanító adatok a felügyelt tanítási forma miatt be- és kimeneti mintapárookra oszlottak. A hálózatok tanító adataiból körülbelül 120000 hét mérési adataiból és a hozzájuk tartozó időben változó paraméterértékekből álltak. Az input mátrix $7 \times (\text{kb.}) 120000$ -es (két dózis, öt mérés), a target mátrix pedig $8 \times (\text{kb.}) 120000$ -es (8 paraméter). A körülbelüli érték amiatt nem pontos, mert nem látni előre, hogy egy-egy variáns megéli-e a 15 hetet.

4. fejezet

Eredmények és kiértékelésük

4.1. Virtuális páciensek szimulált zajos mérései

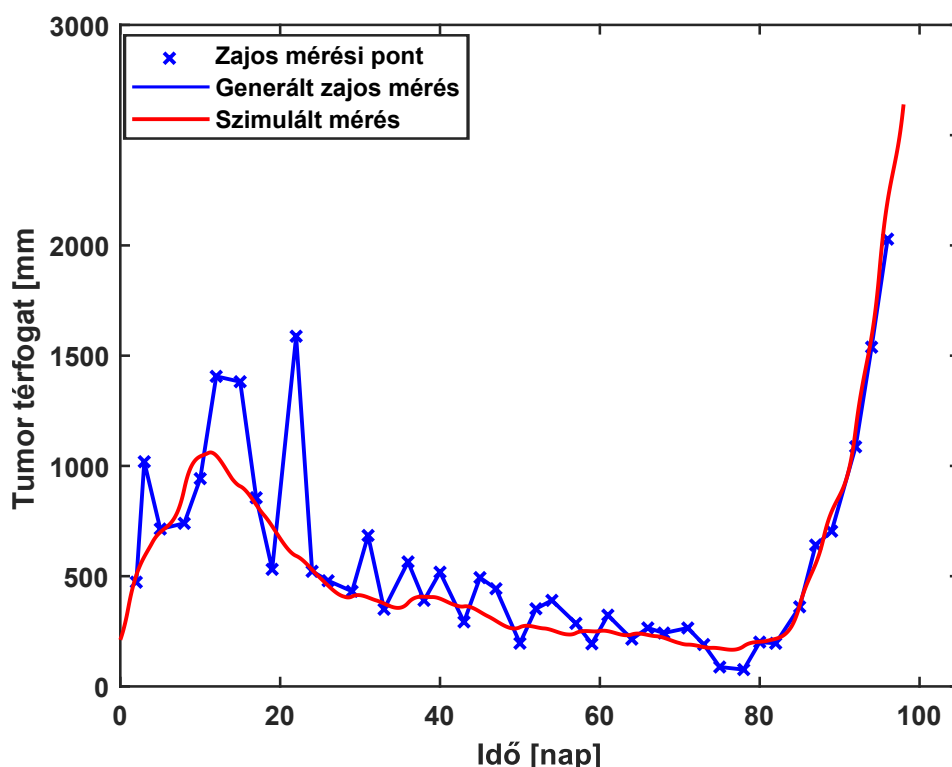
A tanító adatok szemléltetésére kiválasztottunk három virtuális páciens a különböző tanító adathalmazokból (a tanító adathalmazokra a rajtuk tanult hálózatok sorszámaival fogunk hivatkozni). Ezeken a példákön bemutatjuk, hogy a modellezett zaj hogyan jelenik meg a virtuális páciensek szimulált mérésein, illetve hogy a tanító adatokat generáló algoritmus hogyan működik. Az egyszerűség kedvéért a beadott dózisokat nem ábrázoljuk.



4.1. ábra. A #27-es tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).

Az 4.1. ábrán a #27-es tanító adathalmazban lévő virtuális páciens szimulált tumor térfogat mérését a piros görbe mutatja, míg a kék görbe azt a zajos mérést szemlélteti, amit

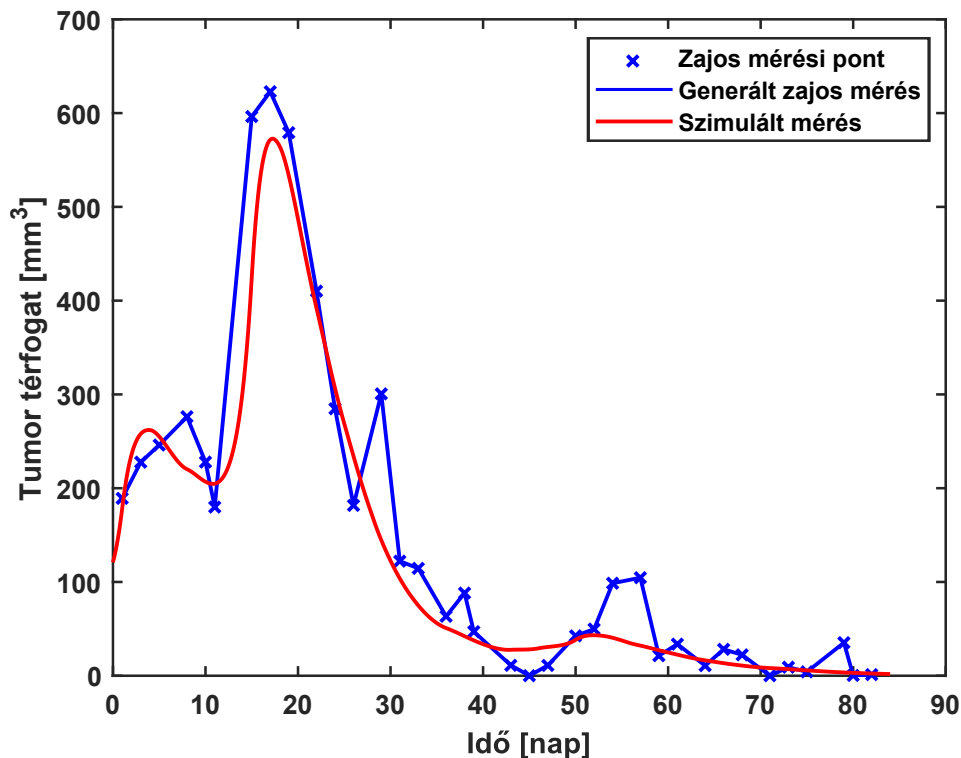
az általunk felépített zajmodellt használva generáltunk a szimulált (piros) értékekre. A kék keresztek a zajos generált mérési pontokat mutatják. A vízszintes tengelyen a kísérlet ideje látható napokban, a függőleges tengely pedig a szimulált tumor térfogatokat jeleníti meg mm^3 -ben. A #27-es tanító adathalmazban a paraméterek 3 hetente változtak meg önmagukhoz képest $\pm 20\%$ -os intervallumon belül generált véletlen százalékos értékkel. A piros görbén észrevehetőek a jellegváltások például a 20. nap környékén. Ezek a jellegbeli váltások is jelzik a paraméterek megváltozását. A generált zajos mérés mutatja a 3.2. alfejezetben tárgyaltakat, mi szerint a kisebb tumor térfogatokhoz nagyobb zaj társul. Ez a 40. és 80. nap között a legszembetűnőbb. A kísérlet ideje majdnem elérte a 105 napot, a tumor térfogata valószínűleg kisebb volt, mint 4 mm^3 , és az algoritmus leállt.



4.2. ábra. A #13-as tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).

A #13-as tanító adathalmaz egy virtuális páciensének méréseit mutatja a 4.2. ábra. Viszonylag sűrű paraméterváltozáson ment keresztül ez a páciens, 2 hetente megváltoztak a paraméterei önmagukhoz képest egy $\pm 10\%$ -os intervallumon belül generált véletlen százalékos értékkel. Ezt a szimulált piros görbe sűrűbb jellegváltozása is mutatja. A zajos mérés itt is nagyon hasonlít a valódi páciensek zajos méréseire. Egyértelműen mutatja azt, hogy a pozitív hibák előfordulásának valószínűsége nagyobb, mint a negatív hiba előfordulásának aránya. Ennél a páciensnél egyértelműen azért állt le az algoritmus, mert a tumor térfogata meghaladta a 2000 mm^3 -t.

A 4.3. ábra az #1-es tanító adathalmazból származó pácienset mutatja. A valódi mérésekhez nagyon hasonlóan a tumor térfogata az első napokban csökken, majd hirtelen növekedésnek indul és újra lecsökken (piros görbe). A mérési zajra ismételten igaz, hogy a kisebb tumor térfogatokra nagyobb zaj generálódott, ahogyan azt vártuk. Az #1-es adathalmazba



4.3. ábra. A #1-es tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).

a páciensek paraméterei 4 hetente változtak meg $\pm 5\%$ -os intervallumon generált random értékkel önmagukhoz képest. Ez látszik a piros görbén, hiszen alig van jellegváltozás. Az algoritmus leállt a 80. nap környékén, hiszen a teljes tumor térfogata 4 mm^3 alá csökkent.

4.2. Neurális hálózatok teljesítményének kiértékelése az MRI adatok nélkül illesztett zajmodellel generált adatok alapján

Négy darab neurális hálózatot hoztunk létre, melyek egy már az előzetes munkáimból származó tanító adathalmazokon tanultak, amihez a jelenlegi dolgozatban felépített zajmodell segítségével hozzáadtunk két különböző valószínűségi eloszlás szerinti mérési zajt. A pozitív hibákat a Weibull-eloszlás segítségével generáltuk, míg a negatívakat az általánosított szélsőérték eloszlással.

A hálózatok mindegyike egy háromrétegű, előrecsatolt architektúrát kapott. A rejtett rétegek sorrendben 50 és 20 neuront tartalmaznak, a kimeneti réteg pedig a paraméterek darabszámával megegyezően 8 neuront alkalmaz. A neuronszámok, illetve a rétegek megválasztása tapasztalati úton történt. A neuronszámok megválasztása 100 felett soknak bizonyult, a futási idő megnövekedett, a teljesítmény egyre romlott, így az első rétegben 50 neuront alkalmaztunk. A tanításokból származó tapasztalatok alapján a második rejtett rétegben elegendő az első rejtett réteg neuronjainak közel felét alkalmazni a megfelelő tel-

jesítmény eléréséhez, így a második rejtett rétegben 20 neuron került alkalmazásra.

A tanítás alatt a neurális hálózatok csak a tanító (70%) és a validációs (15%) adatokat használták fel és a tesztelési (15%) adatokat nem. A Matlab beépített statisztika összesítő grafikonját [24] használjuk a neurális hálózatok paraméter becslő képességének a kiértékelésére a tanult, illetve a teszt adatok összességén. Megnézzük, hogy a hálózatok mekkora hibával képesek visszaadni a virtuális páciensek paramétereit ($a, n, b, ED_{50}, w, c, k_1, k_2$). A diagramokat úgy rendeztük el, hogy a függőleges tengelyen a százalékos relatív hiba jelenlen meg, mely a következő összefüggés alapján számolható:

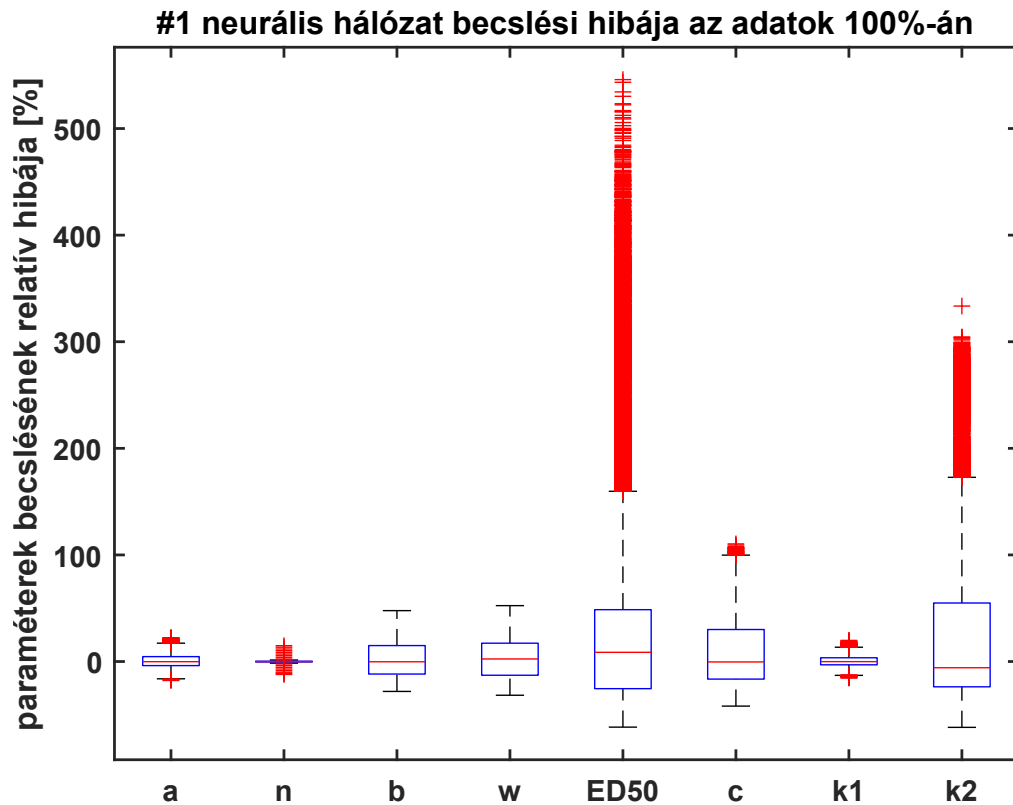
$$\text{hiba} = \frac{\text{kimeneti adatok} - \text{cél adatok}}{\text{cél adatok}} \cdot 100, \quad (4.1)$$

illetve a vízszintes tengelyen megjelenítjük, hogy a hibák mely paraméterhez tartoznak. A diagramot készítő algoritmus növekvő sorrendbe rakja az adatokat és készít egy statisztikai ábrát. Az ábrákon (4.4 - 4.11. ábra) a kék keret területe jelzi az adatok 25-75%-át, míg a benne lévő piros, vízszintes vonal az adatok mediánja. Minél kisebb a kék doboz területe, és minél jobban közelít a medián a nulla %-hoz, a becslési értékek annál jobbabbak. Az eljárás a piros keresztekkel jelölt értékeket kilógó adatnak ítélte meg. Az értékek, melyek nem tartoznak sem a kilógó adatok, sem az adatok 25-75%-a közé, azokat az algoritmus a talpas vonallal zárt szaggatott vonalakkal jelzi [24].

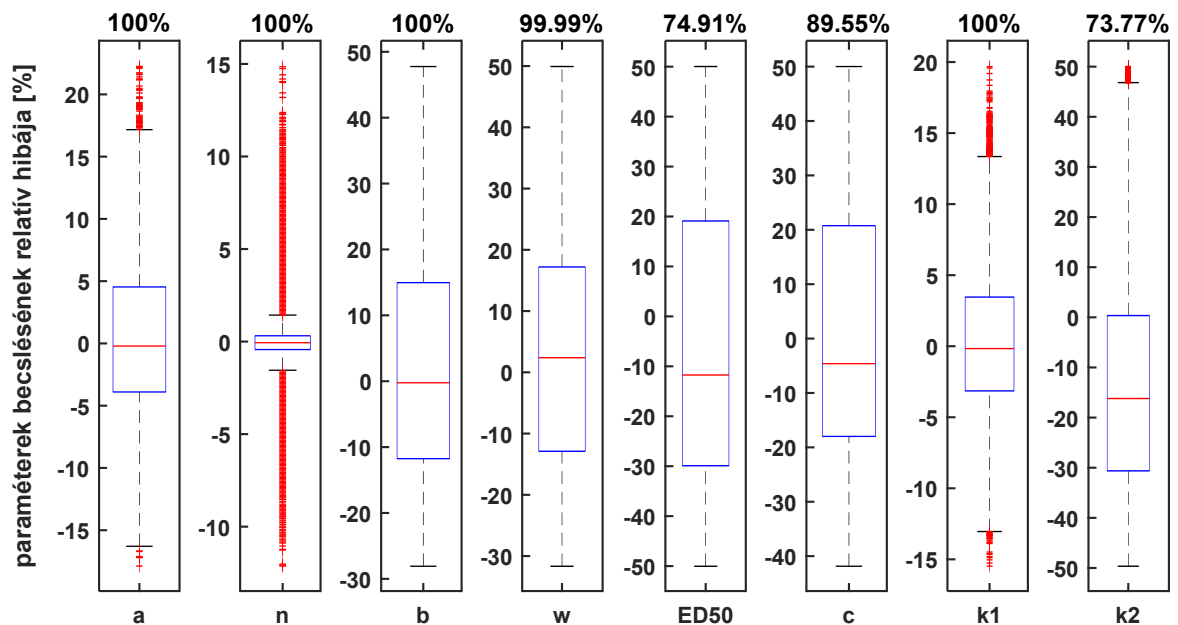
A 4.4. ábrán az #1 hálózat becslésének relatív százalékos hibáit jelenítettük meg paraméterekre lebontva. Ez a hálózat egy olyan tanító adathalmazon tanult, melyben minden virtuális páciens minden paraméterértéke 4 hetente megváltozott önmagához képest $\pm 5\%$ -os intervallumon belül generált véletlen százalékos értékkel. A képen az összes adat ábrázolva van, tehát azok is, amiken tanult a hálózat és a maradék 15% is, melyek a teszt adatok, így a tanítás során nem lettek felhasználva. Látható, hogy a hibák mediánja minden paraméter becslésénél a nulla százaléknál van. Három olyan paraméter is van, melyeknél nagyon kicsik a dobozok és kevés a kilógó minta, tehát a hibák a 0%-körül vannak az adatok nagy hányadán, ilyenek az a, b és k_1 paraméterek. A b és w paraméterek esetén nincsenek kilógó adatok. Van azonban két paraméter, az ED_{50} (medián-effektív dózis) és a k_2 (sebességi együttható), melyek széthúzzák a skálát a nagy mennyiségű kilógó adat miatt. A rendszert leíró modell ((2.1)-(2.4)) egy korábbi verziójának az analízisekor kiderült, hogy az ED_{50} paraméter megváltozására a modell kimenete csak nagyon kis mértékben érzékeny [25]. Ez lehet az egyik oka annak, hogy ez a paraméter viszonylag nehezen becsülhető.

Mivel az algoritmus által kilógó adatnak ítélt hibák (piros kereszttel jelölt minták) széthúzzák a skálát, így készítettünk egy másik ábrát, a könnyebb vizsgálhatóság érdekében, melyen az $\pm 50\%$ -os pontosságba eső hibákat tudjuk vizsgálni. Tulajdonképpen levágtuk azokat a kilógó adatokat minden paraméter esetében, melyek százalékos hibája ezen a $\pm 50\%$ intervallumon kívülre esett. Egy ilyen eredményt mutat az #1-es hálózat esetében a 4.5. ábra, szintén az adatokat paraméterekre lebontva. Az ábra tetjén megjelenítjük a százalékos értékeket, melyek azt mutatják, hogy az összes adatra adott becslés hány százaléka esik bele ebbe a $\pm 50\%$ -os pontossági sávba.

Az a, n, b és k_1 paraméterek esetében a legjobb a becslése a neurális hálózatnak, hiszen az adatok 100%-a ezen a pontosságon belül van. Ha jobban megvizsgáljuk a függőleges tengelyek skálázását, akkor külön kiemelendő az n paraméter, mely adatainak összessége a $\pm 15\%$ -os pontosságon belül van, de az a és k_1 paraméterek összessége is egy szűk pontosságban, a $\pm 20\%$ -os pontossági sávban helyezkedik el. A w paraméter majdnem 100%-a bele esik a vizsgált $\pm 50\%$ pontosságba. Az ED_{50}, k_2 paraméterek értékei sem rosszak ahhoz



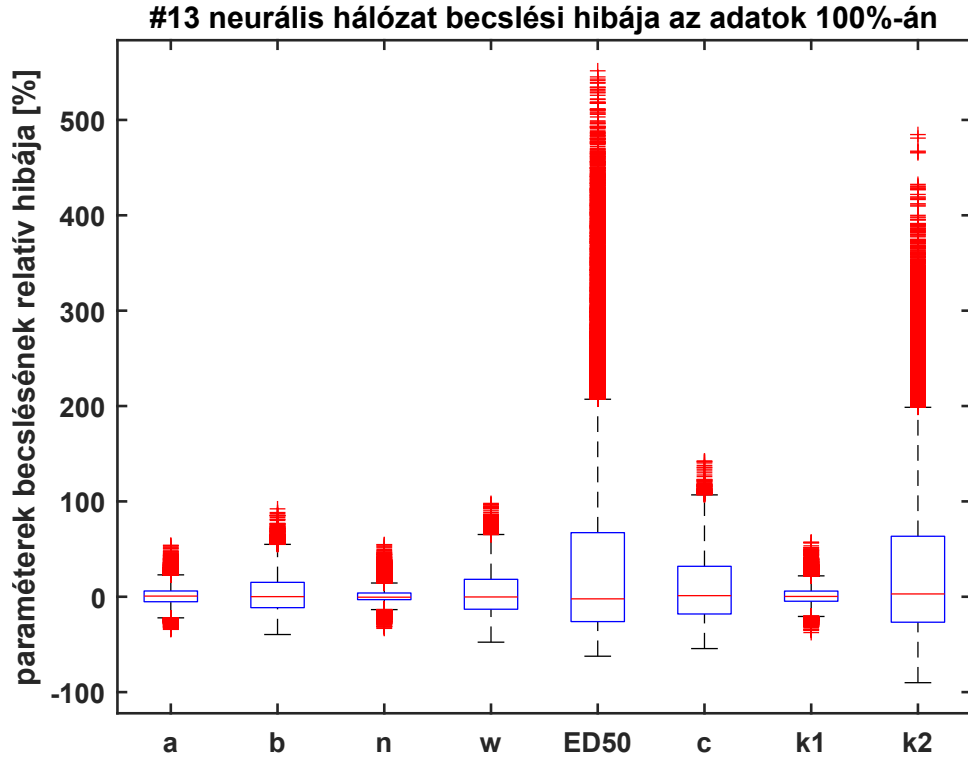
4.4. ábra. #1 neurális hálózat becslésének pontossága az adatok összességén



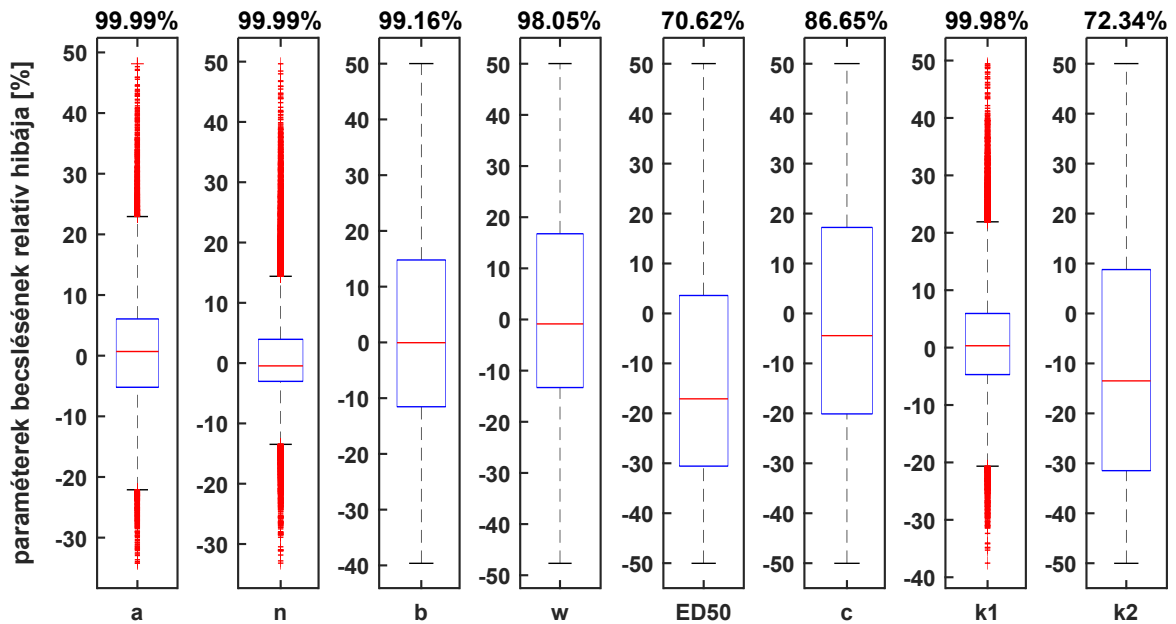
4.5. ábra. #1 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül

képest, hogy a 4.4. ábrán több kilógó mintát is megfigyelhattunk ezeknél a paramétereknél. Megállapítható ezek alapján, hogy az #1-es hálózat képes becslést adni az időben változó paraméterértékekre a tanító adathalmazban lévő valósághibb zajos mérések mellett.

A 4.6. ábrán a #13-as neurális hálózat becslésének relatív százalékos hibáit láthatjuk



4.6. ábra. #13 neurális hálózat becslésének pontossága az adatok összességén

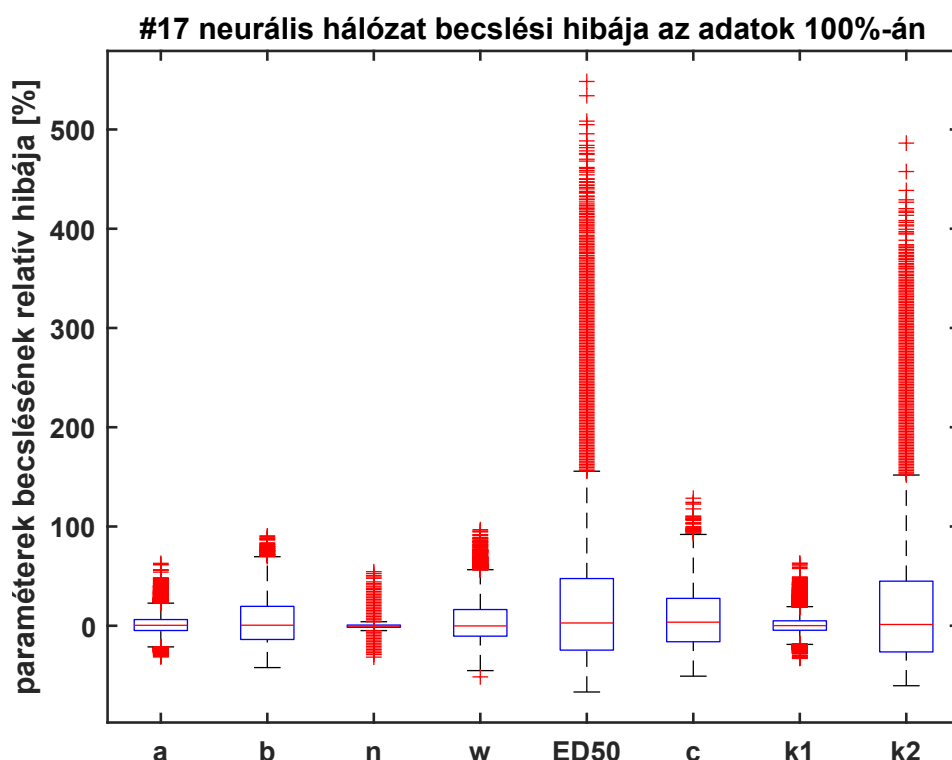


4.7. ábra. #13 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül

szintén paraméterekre lebontva. A hálózat tanító adathalmazában generált virtuális páciensek paraméterei minden második hét végén egy $\pm 10\%$ -os intervallumon belüli véletlen értékkel megváltoztak önmagukhoz képest. Az ábrán is látható, hogy mindegyik paraméternek van kilógó adata és mindegyik nagyobb skálán mozog, mint az előző hálózat esetében. Ez betudható a nagyobb mértékű és gyakoribb paraméterváltozásnak. Az ED_{50} és k_2 pa-

paramétereknek még mindig széthúzzák a skálát, nekik van a legtöbb kilógó adatuk. Illetve ennél a hálózaton már, nincsen olyan paraméter, aminek nem lenne kilógó adata. A kék dobozok is nagyobbak, tehát a hibák is jobban szóródnak, nem olyan összetartóak a nulla körül, mint az első hálózat esetében.

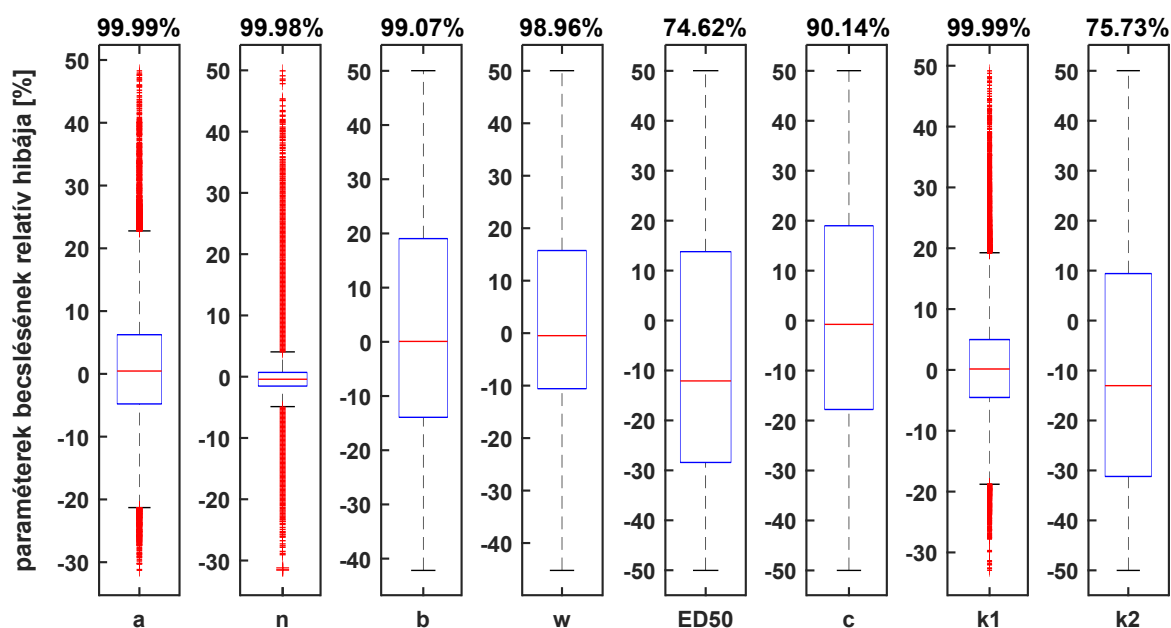
A könnyebb vizsgálhatóság érdekében itt is megvizsgáltuk, hogy az adatok összességének hány százaléka esett bele a $\pm 50\%$ -os pontossági sávba. Ennek eredménye tekinthető meg a 4.7. ábrán. Az ED_{50} , c és k_2 paramétereken kívül minden paraméterre igaz, hogy a hibák több, mint 99%-a ezen az intervallumon belül mozog, kivéve a w paramétert, ahol közel 90% a becslési pontosság ezen a hibasávon. Összességében erről a hálózatról is kijelenthetjük, hogy képes megbecsülni a modell időben változó paramétereit a zajos mérési adatokból.



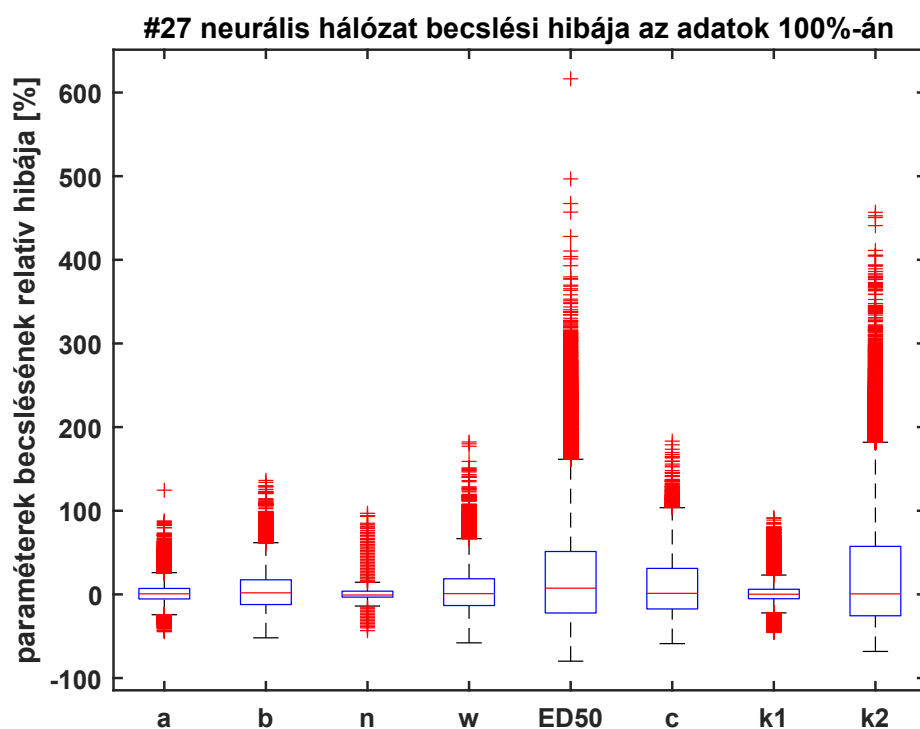
4.8. ábra. #17 neurális hálózat becslésének pontossága az adatok összességén

A #17-es hálózat becslési eredményét mutatja a 4.8. ábra. Ez a hálózat minden negyedik hét végén változtatott paraméterhalmazzal tanult, ahol a paraméterek önmagukhoz képest egy $\pm 15\%$ -os intervallumon generált véletlen értékkel változtak meg. Hasonlóan az eddigiekhez az ED_{50} , c és k_2 paramétereken rosszabbul teljesít a hálózat, mint a többi paraméter esetén és a legtöbb kilógó adatot is ezek a paraméterek hozzák magukkal. A 4.9. ábrán a $\pm 50\%$ -os pontossági sávot vizsgálva kijelenthetjük, hogy az ED_{50} paraméter (mely alaptól is rosszul identifikálható) és a k_2 paraméter a legnehezebben becsülhető paraméterek, de fontos megjegyezni, hogy ezek értéke sem rossz. Az ED_{50} paraméter hibáinak majdnem 75%-a ezen az intervallumon mozog, a k_2 paraméter esetében pedig már majdnem 65%-ról beszélünk. A többi paraméterre az előzőekhez hasonlóan egy biztosabb, pontosabb becslést lehet adni.

A #27-es hálózat az utolsó bemutatott hálózat, a becslési pontosságát a 4.10. ábrán vizsgáljuk. A legnagyobb paraméterváltoztatásos adatokon tanult., ahol a paraméterek a tanító

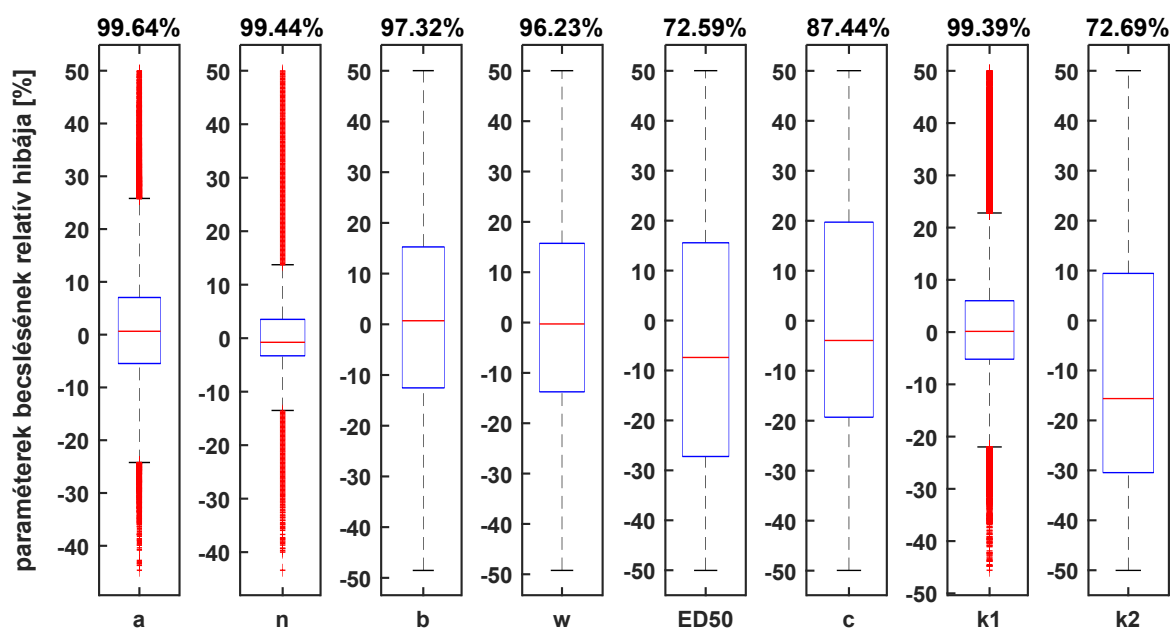


4.9. ábra. #17 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül



4.10. ábra. #27 neurális hálózat becslésének pontossága az adatok összességén

adathalmazban egy $\pm 20\%$ -os intervallumon belüli véletlen értékkel változtak meg önmagukhoz képest minden negyedik hét végén. A kilógó minták száma megnőtt mindegyik paraméternél, és ha megvizsgáljuk a relatív százalékos hibákat mutató fűhőleges tengelyt, akkor az is észrevehető, hogy egy nagyságrenddel nagyobb, mint az előző hálózatok esetén. Mivel a paraméterek akár egy maximum 20%-os változások is áteshettek, így a hibák növekedése is betudható a nagy paraméterváltásoknak. Ha megvizsgáljuk a hálózat becslé-



4.11. ábra. #27 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül

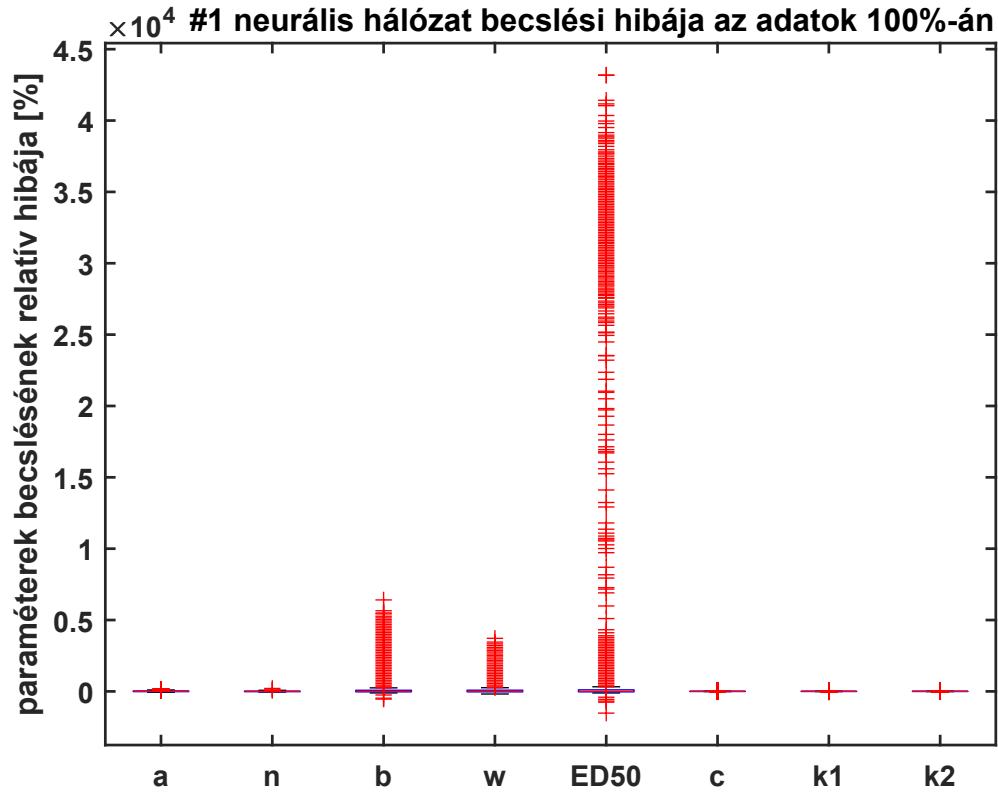
sének relatív százalékos hibáit a $\pm 50\%$ -os pontosságon belül a 4.9. ábrán, akkor láthatjuk, hogy a százalékos értékek majdnem mindegyike még a nagy paraméterváltások ellenére is még mindig 90% feletti. Arról nem is beszélve, hogy a hibák eloszlása sem egyenletes. Az ED_{50} , c és k_2 paraméterek szintén a legnehezebben becsülhetők ezen hálózat számára is.

4.3. Neurális hálózatok teljesítményének kiértékelése az MRI adatokkal illesztett zajmodellel generált adatok alapján

Kettő neurális hálózatot mutatunk be, melynek egyike az előzetes munkáim során tanult be egy olyan adathalmazon, amelyen egyenletes uniform zaj volt a teljes tumor térfogat méréseken (ez lesz a továbbiakban a #2-es neurális hálózat). A másik hálózat tanító adathalmazát hasonlóan építettük fel azzal a különbséggel, hogy a zajmodellünk segítségével generáltuk rá a mérési zajokat a tumor térfogat mérésekre (a továbbiakban ez a #1-es hálózat). A hálózatok teljesítményét teszteljük a tanító és teszt adathalmazon, valamint 10 heti valódi kísérleti mérésen. Teljesítményüket ebben az alfejezetben összehasonlítjuk és kiértékeljük. A két hálózat olyan tanító adathalmazon tanult, melyben minden virtuális páciens minden paraméterértéke hetente megváltozott önmagához képest $\pm 5\%$ -os intervallumon belül generált véletlen százalékos értékkel. A hálózatok architektúrája megegyezik az előző fejezetben tárgyalt hálózati architektúrával.

A zajmodelles. ábrán az #1 hálózat becslésének relatív százalékos hibáit jelenítettük meg paraméterekre lebontva. Ez a hálózat a zajmodellt beépítve kapta meg a tanító adathalmazt. A képen az összes adat ábrázolva van, tehát azok is, amiken tanult a hálózat és a maradék 15% is, melyek a teszt adatok, így a tanítás során nem lettek felhasználva. Látható, hogy a hibák mediánja minden paraméter becslésénél a nulla százaléknál van. Van azonban

egy paraméter az ED_{50} (medián-effektív dózis) mely rendkívül széthúzza a skálát a nagy mennyiségű kilógó adat miatt. A rendszert leíró modell ((2.1)-(2.4)) egy korábbi verziójának az analizisekor kiderült, hogy az ED_{50} paraméter megváltozására a modell kimenete csak nagyon kis mértékben érzékeny [25]. Ez lehet az egyik oka annak, hogy ez a paraméter viszonylag nehezen becsülhető.

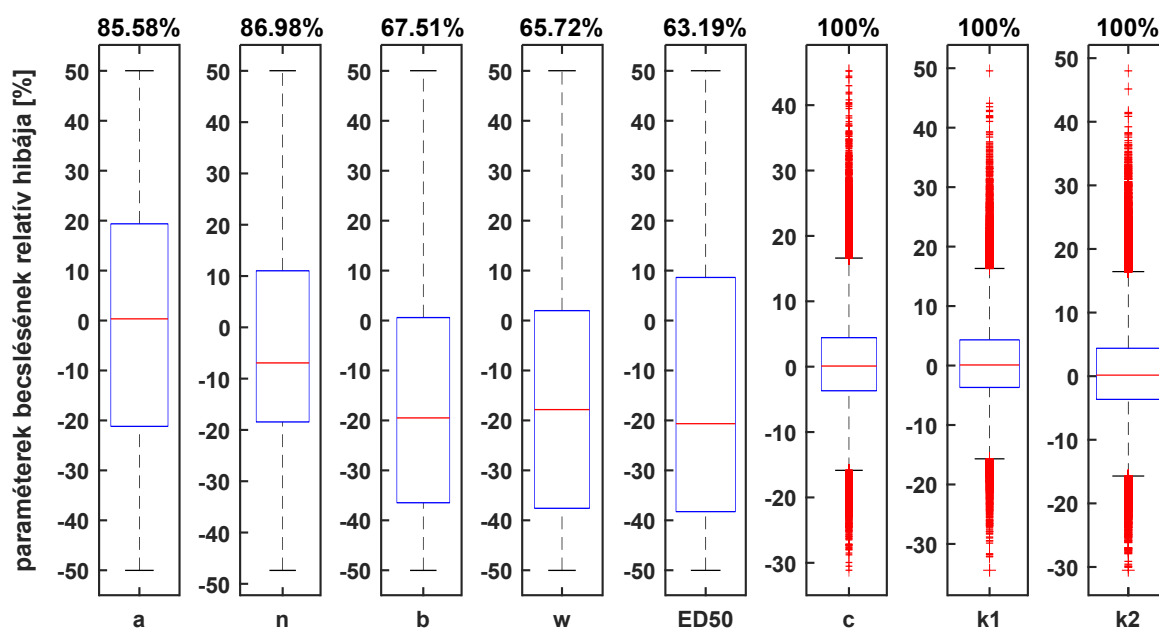


4.12. ábra. #1-es neurális hálózat becslésének pontossága az adatok összességén (zajmodellt beépítve kapott tanító adathalmaz).

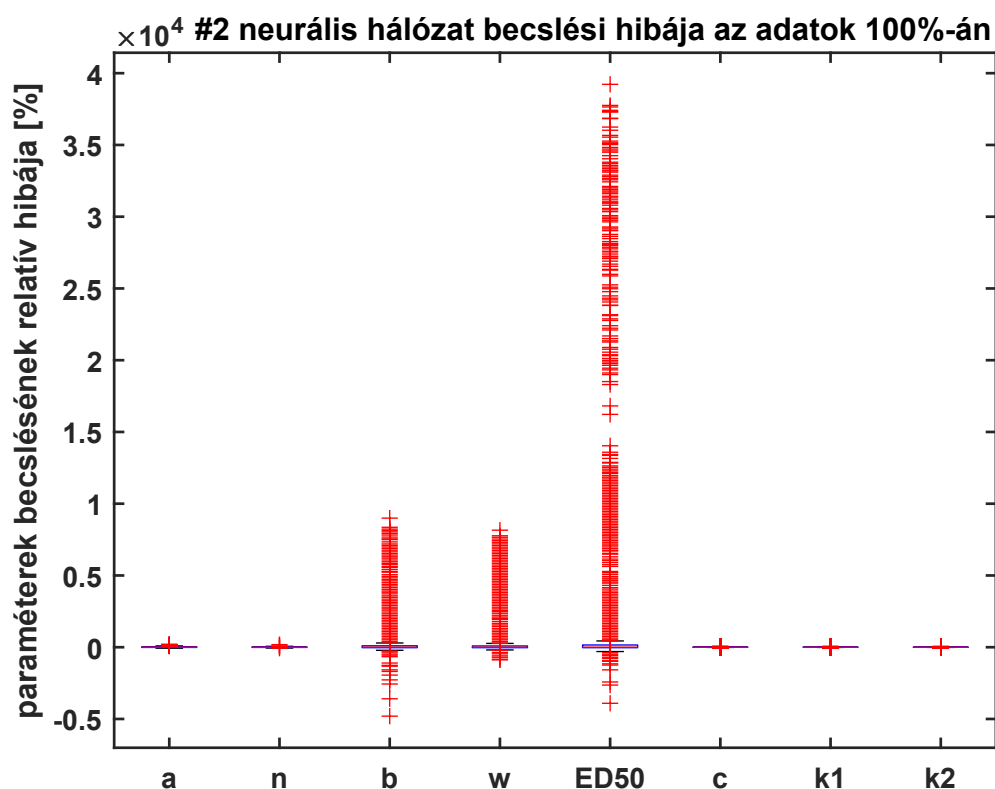
Mivel az algoritmus által kilógó adatnak ítélt hibák (piros kereszttel jelölt minták) széthúzzák a skálát, így készítettünk egy másik ábrát, a könnyebb vizsgálhatóság érdekében, amelyen az $\pm 50\%$ -os pontosságba eső hibákat tudjuk vizsgálni. Tulajdonképpen levágtuk azokat a kilógó adatokat minden paraméter esetében, melyek százalékos hibája ezen a $\pm 50\%$ intervallumon kívülre esett. Egy ilyen eredményt mutat az #1-es hálózat esetében a 4.13. ábra, szintén az adatokat paraméterekre lebontva. Az ábra tetején megjelenítjük a százalékos értékeket, melyek azt mutatják, hogy az összes adatra adott becslés hány százaléka esik bele ebbe a $\pm 50\%$ -os pontossági sávba.

Három paraméterről elmondható, hogy a hibák több mint 100%-a ezen az intervallumon belül van, ezek a c , k_1 és k_2 paraméterek. Az a , n paraméterek esetében sem rossz a teljesítmény, itt a százalékos értékek 85% körül mozognak, tehát a hibák összességének közel 90%-a ezen a pontosságon belül van. A b , w illetve ED_{50} paraméterek értékei sem rosszak ahhoz képest, hogy a 4.12. ábrán sok kilógó mintát figyelhettünk meg, a hibák több, mint 60%-a ezen a $\pm 50\%$ -os hibaintervallumon belül mozog. A hálózat regressziós értékéből is következtethettünk arra, hogy képes egy viszonylag jó becslést adni az időben változó paraméterértékekre. Ennek a hálózatnak a regressziós értéke 0,67 volt.

A 4.14. ábrán a #2-es neurális hálózat becslésének relatív százalékos hibáit láthatjuk



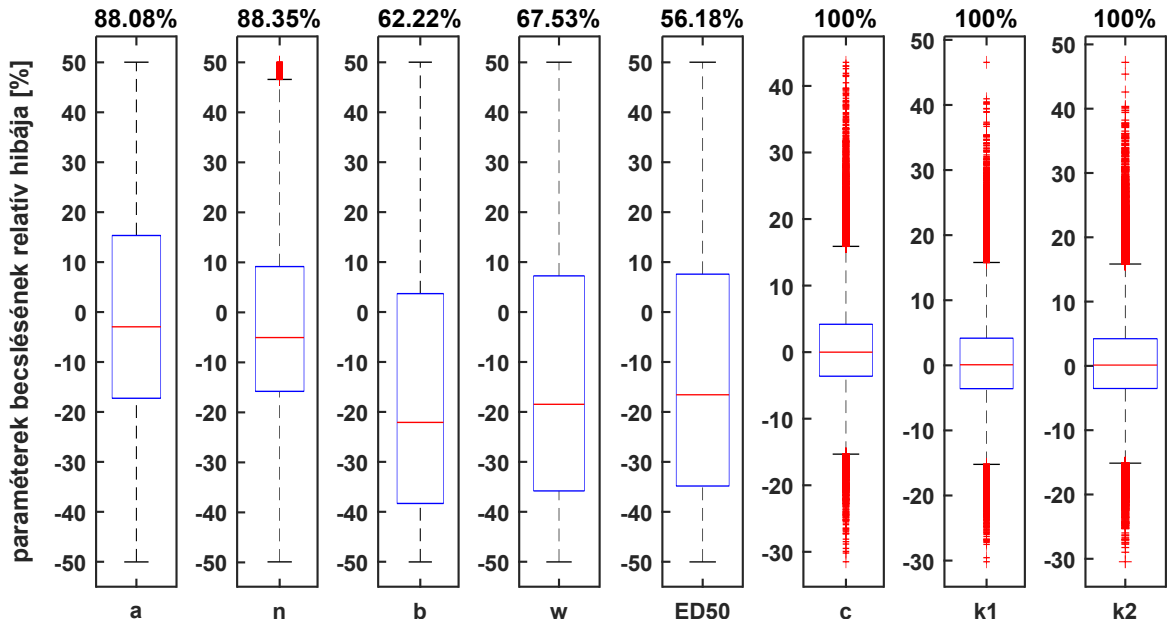
4.13. ábra. #1 neurális hálózat becslésének pontossága az adatokon a $\pm 50\%$ -os pontosságon belül



4.14. ábra. #2-es neurális hálózat becslésének pontossága az adatok összességén (zajmodell nincs beépítve, $\pm 10\%$ -os uniform zaj volt a térfogaton).

szintén paraméterekre lebontva szintén körülbelül 120000 adat eloszlását paraméterenként. A hálózat tanító adathalmazában generált virtuális páciensek paraméterei szintén minden héten változtak $\pm 5\%$ -ot önmagukhoz képest. Ez a hálózat olyan tanító adathalmazon ta-

nult, ahol a teljes tumor térfogatokra egy $\pm 10\%$ -os uniform zajt generáltunk. A 4.14. ábrán megint elhúzzák a skálát a b , w és ED_{50} paraméterek, így ismét megvizsgáljuk a $\pm 50\%$ -os hibaintervallumon belüli hibasávot.



4.15. ábra. #2 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül

A 4.15. ábrán szintén láthatjuk, hogy ez a hálózat is a c , k_1 és k_2 paraméterekre adja a legjobb becslést itt a hibák összessége, 100%-a ezen az intervallumon mozog. Az a és n paramétereket a statisztikai ábra szerint jobban becsüli ez a hálózat, viszont a kritikus paramétereket, mint például az ED_{50} rosszabbul. A hálózat regressziós értéke 0,78 volt, melyből arra következtethetnénk, hogy a hálózat jobb becslést ad a paraméterekre.

A hálózatokat teszteltük valódi méréseken. Összesen 10 heti mérés érkezett a bemene-tükre, így minden hétre nyolc paramétert adtak vissza a kimeneten. Az #1-es hálózat, mely a zajmodell alapján tanult be a zajos méréseken minden paraméterre reális értéket adott vissza, míg a #2-es neurális hálózat többször is irreális, negatív paraméterértékeket adott a kimeneten, annak ellenére, hogy a regressziós értéke jobb volt (0,78 szemben 0,67-el).

5. fejezet

Konklúzió

Egy klinikai kísérletekkel validált tumormodellt alkalmazva (2. fejezet), valódi zajos *in silico* méréseket generáltunk (3. fejezet) különböző heti hiányos mérési protokollokat alkalmazva, és ezeket felhasználtuk háromrétegű, előrecsatolt neurális hálózatok tanítására (2. fejezet). A virtuális terápia generálása alapján a teljes tumor térfogata (ami a kísérletekben mért mennyiség) meghatározható az élő és halott tumorsejtek térfogataként, így a valódi kísérletek *in silico* reprodukálhatók, és az eredmények neurális hálózat tanítására alkalmazhatók.

A dolgozat fő eredménye az előző dolgozatokhoz képest az a zajmodell, amivel *in silico* méréseket tudunk generálni a neurális hálózat tanításához. Az *in silico* mérések előállításánál a valódi kísérleti elrendezést kell modellezni, vagyis a mérések tolmérővel történnek, és a tumor térfogat közelítéssel van megbecsülve. A matematikai modell azonban térfogatértéket generál, tehát a mérési elrendezésből adódó eltéréseket egy zajmodellel írjuk le.

A zajmodell létrehozásához felhasználtunk olyan méréseket, amiben szerepelnek MRI és digitális tolmérővel mért mérések is, és az MRI méréseket tekintettük pontos térfogatnak. Megkerestük azt a nemlineáris transzformációt, ami függetleníti a zajt a tumor térfogattól, majd megkerestük azt az eloszlást, ami a legjobban leírja a zajt. Az eredmények alapján általánosítottam a módszert olyan mérések analízisére, amikor nem áll rendelkezésre MRI (vagy egyéb, a tolmérőnél pontosabb) mérés. Ezt a mérések simításával érem el, a simítás paraméterének a behangolását az MRI mérések alapján végzem a kiadódó eloszlás összehasonlításával. A végeredmény egy zajmodell, ami tartalmazza a térfogatfüggetlenítő transzformációt, az eloszlást, illetve a simítási paramétert.

A zajmodell eloszlásának a paramétereit azokon a kísérleti adatokon határoztuk meg, amelyhez hasonló kísérleti elrendezésben (hasonló egér és tumor) a neurális hálózatot a jövőben alkalmazni fogjuk. A neurális hálózatokat a virtuális betegekkel betanítottuk, paraméterbecslő és általánosító képességüket megvizsgáltuk (4. fejezet). Az elkészített neurális hálózatok első rejtett rétegükben 50, a második rejtett rétegükben pedig 20 neuron felhasználásával készültek, és bemenetükre 100 virtuális páciens 80 féle variánsát generáltuk tanítás és tesztelés céljából. A variánsok paraméterei az egér paraméterértékeit kapták kezdeti értéként, tehát ahány variánsa volt egy egérnek, annyiszor indultak azonos kezdeti értékkel a paramétereik értékei. A szimuláció során ezek a paraméterértékek hetente vagy több hetente önmagukhoz képest bizonyos százalékkal megváltoztak pozitív vagy negatív irányba. Az eredmények mutatják, hogy a hálózatok képesek a virtuális páciensek időben változó paramétereinek becslésére. A neurális hálózatok egy valódi méréssorozaton is tesztelve lettek. Ezen a ponton összehasonlításra került egy olyan neurális hálózat,

mely egyenletes zajos méréseken tanult be és egy olyan is, mely a zajmodell felhasználásával generált zajos adathalmazon tanult. Utóbbi reális paramétereket adott a kimeneten, míg az egyenletes zajos adatokon tanult hálózat sokszor irreális, negatív paraméterértéket adott vissza.

A 3.5. alfejezetben ismertetett heti mérési protokollok, melyek a tanítás alapját adták, az aktuális kísérletek által lettek meghatározva. A későbbiekben a hálózatok a kísérlet során alkalmazhatók az első kísérleti hetet követően. Az aktuális heti méréseket, (azokon a napokon amikor nincs mérés nulla mérési eredmény áll) a hálózat bemeneteire kötve, a hálózat képes becslést adni a paraméterek értékeire, melyek alapján a következő kísérleti hétre a terápia meghatározható. Az ezt követő hetekben ezek alapján a hálózatok a kísérlet teljes ideje alatt alkalmazhatók, használatukkal a paraméterek esetleges heti változásai lekövethetők.

A jelenlegi bonyolultságú neurális hálózatokkal azt az eredményt kaptuk, hogy a mérések hiányossága csak kis mértékben befolyásolja a hálózatok teljesítményét. Megvizsgálva a hiányos és nem hiányos méréseken tanult hálózatokat a regressziós értékek csak kis mértékben térnek el egymástól. Az azonban egyértelmű, hogy az időben változó paraméterek miatt az eddigiekhez képest nőtt a feladat bonyolultsága, hiszen a jelenlegi hálózatok mindegyikének teljesítménye (a regressziós értékek alapján) romlott az előző munkákban alkalmazott konstans paraméterértékekkel tanított időablakos hálózatokhoz képest. A hálózatok bonyolultságát a jövőben néhány kiemelt esetben mindenképpen vizsgálnunk kell. A legjobban teljesítő hálózatok további fejlesztéseken mennek majd keresztül, bonyolultságuk növelésre kerül, melynek hatását a kutatás további szakaszaiban megvizsgáljuk. Jelenleg erre azért nem volt lehetőség, mert az egyes hálózatok tanítása időigényes, akár több napig is eltarthat, amíg egy hálózat tanítása lefut. A problémát tovább bonyolította, hogy ezek neurális hálózatok már egy újabb paraméterhalmazból tanultak. A paraméterek skálája az előzőekhez képest megváltozott az új eredmények és illesztések alapján. A hetente vagy több hetente való megváltozás szimulálása pedig minden esetben a jelleggörbe változását vonja maga után, melyet a hálózatoknak mindig kezelnie kell a hiányos adatok mellett.

A zajmodellből megérthető a valódi mérési zaj, ez később az identifikációnál felhasználható. A neurális hálózatok a valódi zajos mérések szimulációs adatain is képesek betanulni, és megfelelő becslést adni a paraméterek értékeire. A rendszerrel megadható, hogy milyen gyakori és milyen mértékű paraméter megváltozás esetén milyen teljesítményt hoznak az egyes neurális hálózatok. A hálózatok a jövőben lehetővé tehetik a paraméterek változásainak lekövetését a kísérleti idő alatt, ezáltal lehetőséget adva az egyedi terápia generálásának. Kutatásunk végeredménye lehet egy piacképes orvostechikai eszköz, amely képes a terápiát személyre szabni daganatos betegek számára azáltal, hogy a páciensen végzett mérések alapján megmondja a páciens egyedi tumormodelljének paramétereit, ami alapján a személyre szabott terápia legenerálható.

Kutatásom az Innovációs és Technológiai Minisztérium ÚNKP-22-2 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs alapról finanszírozott szakmai támogatásával készült.

Irodalomjegyzék

- [1] Dániel András Drexler, Tamás Ferenci, Anna Lovrics, and Levente Kovács. Tumor Dynamics Modeling based on Formal Reaction Kinetics. *Acta Polytechnica Hungarica*, 16:31–44, 2019.
- [2] Dániel András Drexler, Tamás Ferenci, András Füredi, Gergely Szakács, and Levente Kovács. Experimental data-driven tumor modeling for chemotherapy. In *Proceedings of the 21st IFAC World Congress*, pages 16466–16471, 2020.
- [3] Ricard D Kaloshi G Idbaih A Alentorn A Psimaras D Honnorat J Delattre JY Grenier E Ducray F Samson A Ollier E, Mazzocco P. Analysis of temozolomide resistance in low-grade gliomas using a mechanistic mathematical model. *Fundamental and Clinical Pharmacology*, 31:347–358, 2017.
- [4] Peyre M Ricard D Calvez V Tod M Cajavec-Bernard B Idbaih A Psimaras D Dainese L Pallud J Cartalat-Carel S Delattre JY Honnorat J Grenier E Ducray F Ribba B, Kaloshi G. A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clinical Cancer Research*, 18:5071–5080, 2012.
- [5] Kalamarides M Trocóniz IF Ribba B. Ouerdani A, Goutagny S. Mechanism-based modeling of the clinical effects of bevacizumab and everolimus on vestibular schwannomas of patients with neurofibromatosis type 2. cancer chemother pharmacol. *Cancer Chemotherapy and Pharmacology*, 77:1263–1273, 2016.
- [6] Carter AA Chiparus O Hollaender N Kim H Motzer RJ Sarr C. Stein A, Wang W. Dynamic tumor modeling of the dose-response relationship for everolimus in metastatic renal cell carcinoma using data from the phase 3 record-1 trial. *BMC Cancer*, pages 2–10, 2012.
- [7] Friberg LE. Bender BC, Schindler E. Population pharmacokinetic-pharmacodynamic modelling in oncology: a tool for predicting clinical response. *br j clin pharmacol. British Journal of Clinical Pharmacology*, 79:56–71, 2015.
- [8] Hoff PM Van Cutsem E Zuideveld KP Jorga K Fagerberg J Bruno R Claret L, Girard P. Model-based prediction of phase iii overall survival in colorectal cancer on the basis of phase ii tumor dynamics. *Journal of Clinical Oncology*, 27:4103–4108, 2009.
- [9] Boehm M Bootle D Juric D Quadt C Goggin TK De Buck SS, Jakab A. Population pharmacokinetics and pharmacodynamics of byl719, a phosphoinositide 3-kinase antagonist, in adult patients with advanced solid malignancies. *British Journal of Clinical Pharmacology*, 78:543–555, 2014.

- [10] Kathleen M. C. Tjørve and Even Tjørve. The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family. *PLOS ONE*, 12(6):1–17, 06 2017.
- [11] Cristina Vaghi, Anne Rodallec, Raphaëlle Fanciullino, Joseph Ciccolini, Jonathan P. Mochel, Michalis Mastri, Clair Poignard, John M. L. Ebos, and Sébastien Benzekry. Population modeling of tumor growth curves and the reduced gompertz model improve prediction of the age of experimental tumors. *PLOS Computational Biology*, 16(2):1–24, 02 2020.
- [12] P. Hahnfeldt, D. Panigrahy, J. Folkman, and L. Hlatky. Tumor development under angiogenic signaling: A dynamical theory of tumor growth, treatment response, and postvascular dormancy. *Cancer Research*, 59:4770–4775, 1999.
- [13] J. S. Lowengrub, H. B. Frieboes, F. Jin, Y.-L. Chuang, X. Li, P. Macklin, S. M. Wise, and V. Cristini. Nonlinear modelling of cancer: bridging the gap between cells and tumours. *Nonlinearity*, 23(1):R1–R9, 2010.
- [14] Philipp M. Altrock, Lin L. Liu, and Franziska Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews. Cancer*, 15(12):730–745, 2015.
- [15] Angela M. Jarrett, Ernesto A. B. F. Lima, David A. Hormuth, Matthew T. McKenna, Xinzeng Feng, David A. Ekrut, Anna Claudia M. Resende, Amy Brock, and Thomas E. Yankeeelov. Mathematical models of tumor cell proliferation: A review of the literature. *Expert Review of Anticancer Therapy*, 18(12):1271–1286, 2018.
- [16] B Meibohm and H Derendorf. Basic concepts of pharmacokinetic/pharmacodynamic (pk/pd) modelling. *International Journal Of Clinical Pharmacology And Therapeutics*, 35:401–413, 1997.
- [17] Melánia Puskás and Dániel A. Drexler. Tumor model parameter estimation for therapy optimization using artificial neural networks. In *IEEE International Conference on Systems, Man, and Cybernetics - 2021*, pages 1254–1259, 2021.
- [18] János Tóth, Attila László Nagy, and Dávid Papp. *Reaction kinetics: exercises, programs and theorems*. Springer, 2018.
- [19] Puskás Melánia. Preklinikai kísérletek mérési hibájának modellezése és felhasználása neurális hálózatok tanításához I. TDK dolgozat, ÓE-NIK, 2021/22/1.
- [20] George Marsaglia and John Marsaglia. Evaluating the anderson-darling distribution. *Journal of Statistical Software*, 9(2):1–5, 2004.
- [21] The mathworks inc.: Anderson-darling test. <https://www.mathworks.com/help/stats/adtest.html>. Megtekintve: 2022-04-19.
- [22] Puskás Melánia. Élettani folyamatok para méterbecslése mesterséges intelligencia használatával III. TDK dolgozat, ÓE-NIK, 2021/22/1.
- [23] Puskás Melánia. Élettani folyamatok paraméterbecslése mesterséges intelligencia használatával. TDK dolgozat, ÓE-NIK, 2020/21/1.

- [24] The Mathworks Inc.: Visualize summary statistics with box plot documentation. <https://www.mathworks.com/help/stats/boxplot.html>. Megtekintve: 2022-11-10.
- [25] M. Siket, G. Eigner, and L. Kovács. Sensitivity and identifiability analysis of a third-order tumor growth model. In *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*, pages 417–422, 2020.

Ábrák jegyzéke

3.1. A 16-os egér zajos méréseinek (kék görbe) simítási eredményei (piros görbe) hat iteráción keresztül, a simítás jósága a diagramok tetején az R^2 értékével van megadva.	19
3.2. A 16-os egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).	20
3.3. A 16-os egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök) minden egyes mérési pontban.	21
3.4. A 16-os egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.	21
3.5. A 8-as egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).	22
3.6. A 8-as egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök), minden egyes mérési pontban.	23
3.7. A 8-as egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.	23
3.8. A 13-as egér zajos méréseit ábrázoló görbe (kék görbe) a valódi mérési pontokkal (kék teli körök) és a simított, becsült zajmentes méréseket ábrázoló görbe (piros görbe).	24
3.9. A 13-as egér zajos és simított tumor térfogat mérése közötti relatív hiba értéke (kék üres körök), minden egyes mérési pontban.	24
3.10. A 13-as egér zajos és simított tumor térfogat mérése közötti relatív hiba szóródási diagramja.	25
3.11. A 16-os egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).	26

3.12. A 8-as egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).	27
3.13. A 13-as egér eredeti pozitív hibáinak szóródása (balra fent), pozitív hibáinak szóródása a transzformáció után (jobbra fent), illetve eredeti negatív hibáinak szóródása (balra lent) és negatív hibáinak szóródása a transzformáció után (jobbra lent).	28
3.14. A 16-os egér eredeti pozitív hibáinak szóródása (kék körök) és az általánosított szélsőérték-eloszlással generált pozitív hibák szóródása (narancs rombuszok).	31
3.15. A 16-os egér eredeti negatív hibáinak szóródása (kék körök) és a Weibull-eloszlással generált negatív hibák szóródása (narancs rombuszok).	32
3.16. A 8-as egér eredeti pozitív hibáinak szóródása (kék körök) és a Weibull-eloszlással generált pozitív hibák szóródása (narancs rombuszok).	32
3.17. A 8-as egér eredeti negatív hibáinak szóródása (kék körök) és az általánosított szélsőérték-eloszlással generált negatív hibák szóródása (narancs rombuszok).	33
3.18. A 13-as egér eredeti pozitív hibáinak szóródása (kék körök) és a Nakagami-eloszlással generált pozitív hibák szóródása (narancs rombuszok).	34
3.19. A 13-as egér eredeti negatív hibáinak szóródása (kék körök) és a Student féle t-eloszlással generált negatív hibák szóródása (narancs rombuszok).	34
3.20. Egereken végzett kísérletek során a tumor az egér testének bármely részén megjelenhet, többféle amorf alakot felvehet, melynek szélességét és hosszúságát digitális tolómérővel mérjük.	37
3.21. MRI-vel mért tumor térfogat mérések (zöld görbe) és digitális tolómérővel mért tumor térfogat mérések (piros görbe).	38
3.22. Digitális tolómérő relatív mérési hibája az MRI mérésekhez képest.	38
3.23. Digitális tolómérő és MRI mérések (alapigazság) közötti relatív mérési hibák szóródási diagramja.	39
3.24. Digitális tolómérő és MRI mérések (alapigazság) közötti pozitív és negatív relatív mérési hibák szóródási diagramja.	39
3.25. A függetlenítéshez felhasznált transzformáció paramétereit hangoló algoritmus, mely opcionálisan a simított mérésekhez tartozó simítási paramétert is hangolja az Anderson-Darling teszt paraméterének minimalizálásával.	41
3.26. Digitális tolómérővel mért mérési hiba simítása az általunk meghatározott simítási paraméterrel ($s = 0,8805$).	48
3.27. 16-os kísérleti egér eredeti negatív mérési hibájának szóródási diagramja (bal oldalon) és a függetlenített, transzformált hiba (jobb oldalon).	48

3.28. A 16-os egér eredeti negatív hibáinak szóródása (kék körök) és az általunk meghatározott eloszlással generált negatív hibák szóródása (narancs körök).	49
3.29. Tanító adatokat generáló algoritmus folyamata	50
3.30. Hiányos mérési protokollok a jelenlegi kísérletekben	52
3.31. Heti méréseket generáló algoritmus	53
4.1. A #27-es tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).	55
4.2. A #13-as tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).	56
4.3. A #1-es tanító adathalmaz egy virtuális páciensének szimulált teljes tumor térfogat mérése (piros görbe) és a generált zajos mérés (kék görbe), a zajos mérési pontokkal (kék keresztek).	57
4.4. #1 neurális hálózat becslésének pontossága az adatok összességén	59
4.5. #1 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül	59
4.6. #13 neurális hálózat becslésének pontossága az adatok összességén	60
4.7. #13 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül	60
4.8. #17 neurális hálózat becslésének pontossága az adatok összességén	61
4.9. #17 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül	62
4.10. #27 neurális hálózat becslésének pontossága az adatok összességén	62
4.11. #27 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül	63
4.12. #1-es neurális hálózat becslésének pontossága az adatok összességén (zajmodellt beépítve kapott tanító adathalmaz).	64
4.13. #1 neurális hálózat becslésének pontossága az adatokon a $\pm 50\%$ -os pontosságon belül	65
4.14. #2-es neurális hálózat becslésének pontossága az adatok összességén (zajmodell nincs beépítve, $\pm 10\%$ -os uniform zaj volt a térfogaton).	65
4.15. #2 neurális hálózat becslésének pontossága az adatokon az $\pm 50\%$ -os pontosságon belül	66