



Univerza v Ljubljani  
Fakulteta *za računalništvo  
in informatiko*

# UPORABA MEHKE LOGIKE ZA MODELIRANJE BIOLOŠKIH SISTEMOV NA PRIMERU SIGNALNE POTI MAPK

Lidija Magdevska

Delo je pripravljeno v skladu s Pravilnikom o podeljevanju  
Prešernovih nagrad študentom, pod mentorstvom

doc. dr. *Mihe Moškona*

Ljubljana, 2017



© 2017, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko

Rezultati dela so intelektualna lastnina Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov dela je potrebno pisno soglasje Fakultete za računalništvo in informatiko ter mentorja.



Univerza v Ljubljani  
Fakulteta za računalništvo in informatiko

Lidija Magdevska

## Uporaba mehke logike za modeliranje bioloških sistemov na primeru signalne poti MAPK

### POVZETEK

Signalne poti so ena od osnovnih tem raziskav sistemske biologije, saj le-te regulirajo številne celične procese in so vključene v zunajcelično komunikacijo. Njihova analiza je zahtevna zaradi velikih razlik med časovnimi odzivi različnih proteinov in prepletenosti signalnih poti tako preko skupnih dražljajev kot tudi v primerih, kjer prenos informacij sprožijo različni dražljaji. Pri raziskovanju povezav med gradniki signalnih poti imajo zato pomembno vlogo tudi računski modeli.

Na voljo je širok spekter metod za analizo in razumevanje signalnih poti, ki so primerne za različne vrste problemov. Nekaterne metode zahtevajo natančno poznavanje biokemijskega ozadja problema, za tako zgrajene modele pa je značilna velika natančnost, medtem ko je lahko pri drugih metodah naše poznavanje opazovanega sistema bolj okrnjeno, a so rezultati (napovedi) modela manj natančni.

V zadnjem času stopajo v ospredje metode, ki predstavljajo kompromis med obema ekstremoma. Mednje sodijo tudi metode mehke logike, na katere se osredotočimo v pričujočem delu. Za vsako izmed obstoječih metod opišemo njene prednosti in slabosti ter jo bolj ali manj uspešno apliciramo na signalno pot MAP-kinaze.

**Ključne besede:** mehka logika, računska biologija, modeliranje in simulacija, signalna omrežja, MAP-kinaza



University of Ljubljana  
Faculty of Computer and Information Science

Lidija Magdevska

## **Fuzzy modelling of biological systems and its application to the analysis of MAPK signalling pathway**

### **ABSTRACT**

Signalling pathways are one of the fundamental topics of research in systems biology as they are involved in extracellular communication and regulate various cellular processes. Their analysis is difficult due to two factors: the first presents significant differences in temporal responses of different proteins, and the second presents the fact that they are interleaved through both common stimuli and cases when different stimuli trigger the transmission of the same information. Therefore, computational models play an important role in exploring the connections between different components of the signalling pathways.

A wide spectrum of methods with various applications is available for analysing and understanding of signalling pathways. Some methods require a complex understanding of the problem's biochemical background and result in highly accurate models, while with others, our knowledge may be more limited, which then results in less accurate predictions of the model.

Methods that represent a compromise between the two approaches are recently becoming prevalent. These include fuzzy logic methods, on which we focus in this work. For each such method we describe its advantages and disadvantages and then apply it to the MAP kinase signalling pathway.

**Key words:** fuzzy logic, computational biology, modelling and simulation, signalling networks, MAP kinase





## ZAHVALA

*Iskrena hvala mentorju doc. dr. Mihi Moškoni za vso pomoč pri izdelavi tega dela. Zahvaljujem se tudi ostalim zaposlenim v Laboratoriju za računalniške strukture in sisteme ter članom Skupine za računske metode v sistemski in sintezni biologiji za prijetno delovno vzdušje. Velika zahvala gre seveda prof. dr. Mihi Mrazu in prof. dr. Nikolaju Zimicu za koristne pripombe in strokovno pomoč pri nastajanju pričujočega dela ter Tanji Cvitanović za usmerjanje pri izbiri biološko relevantnega problema.*

*Na koncu bi se rada zahvalila prijateljem in širši družini za spodbude v času izobraževanja. Hvala staršem za materialno podporo in sestri Veri za potrpežljivost. Posebna zahvala pripada tudi Sašu, ki mi je med študijem večkrat pomagal in mi polepšal marsikateri trenutek študentskega življenja.*

— Lidija Magdevska, Ljubljana, september 2017.



## KAZALO

<b>Povzetek</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Zahvala</b>	<b>v</b>
<b>1 Uvod</b>	<b>1</b>
1.1 Cilji . . . . .	2
1.2 Pregled naloge . . . . .	3
<b>2 Osnovni pojmi računske biologije in signalnih omrežij</b>	<b>5</b>
2.1 Sistemska in računska biologija . . . . .	5
2.2 Kemijske zvrsti in njihove koncentracije . . . . .	5
2.3 Signalna omrežja . . . . .	6
2.4 Kemijski procesi v signalnih kaskadah . . . . .	7
2.5 Opis signalne poti MAPK . . . . .	8
2.6 Klasična kaskada mitogen kinaze . . . . .	8
<b>3 Metode za modeliranje signalnih omrežij</b>	<b>11</b>
3.1 Pregled obstoječih metod . . . . .	12
3.2 Analiza ravnotežnega stanja . . . . .	12
3.3 Analiza dinamike sistema . . . . .	13
3.4 Verodostojnost modelov . . . . .	14
<b>4 Osnove mehke logike</b>	<b>15</b>
4.1 Mehka množica . . . . .	15
4.2 Mehka spremenljivka . . . . .	16

4.3	Mehka pravila . . . . .	16
4.4	Mehke relacije . . . . .	17
4.5	Mehko sklepanje . . . . .	18
4.6	Ostrenje . . . . .	19
4.6.1	Težiščna metoda . . . . .	19
4.6.2	Metoda središčnih vsot . . . . .	20
4.6.3	Metode maksimumov . . . . .	20
4.7	Mehki model . . . . .	21
<b>5</b>	<b>Metode za gradnjo mehkih podatkovno vodenih modelov signalne poti</b>	<b>23</b>
5.1	Mehka metoda voditeljev . . . . .	24
5.2	Mehke časovne vrste . . . . .	25
5.3	Metoda večatributne mehke časovne vrste . . . . .	26
<b>6</b>	<b>Rezultati</b>	<b>27</b>
6.1	Predpostavke in poenostavitve mehkih modelov . . . . .	27
6.1.1	Učni podatki . . . . .	28
6.1.2	Pogoji testiranja . . . . .	28
6.2	Referenčni model signalne poti MAPK . . . . .	29
6.3	Uporaba metode FCM za učenje ravnotežnega stanja signalne poti MAPK	35
6.3.1	Polni mehki model ravnotežnega stanja . . . . .	35
6.3.2	Reducirani mehki model ravnotežnega stanja . . . . .	36
6.4	Uporaba metode FCM za učenje dinamičnega modela signalne poti MAPK	38
6.4.1	Polni dinamični mehki model . . . . .	39
6.4.2	Reducirani dinamični mehki model . . . . .	41
6.5	Uporaba metode MAFTS za učenje dinamičnega modela signalne poti MAPK . . . . .	44
6.5.1	Polni dinamični mehki model . . . . .	45
6.5.2	Reducirani dinamični mehki model . . . . .	47
6.6	Primerjava pristopov za gradnjo mehkega modela . . . . .	48
<b>7</b>	<b>Zaključek</b>	<b>53</b>
7.1	Prispevki k znanosti . . . . .	54
7.2	Nadaljnje delo . . . . .	54

**A Priloga**

**61**



# 1 Uvod

Pravilno delovanje signalnih poti je za človeka življenjskega pomena. Napake v njihovem delovanju se namreč pogosto odražajo v obliki bolezni. Številnih tovrstnih bolezni, med katere sodijo avtoimunske bolezni in rak, ne znamo zdraviti vzročno, temveč lahko blažimo le njihove simptome. Vzrok, da teh bolezni ne znamo odpraviti, leži v celičnem sporočanju. Slednje je zelo kompleksno, zato še ni v celoti raziskano, do sedaj pridobljeno znanje pa za zdravljenje marsikatere bolezni ne zadošča.

Obolenja, katerih vzrok je nepravilno delovanje signalnih poti, se pri ljudeh pojavljajo pogosto, zato bi njihovo uspešno zdravljenje predstavljalo velik napredek za človeštvo. Da bi naredili prvi korak k uspehu, je potrebno izboljšati metode za preučevanje signalnih procesov ter pospešiti raziskave na področju celičnega sporočanja. Pomembno orodje pri tem predstavljajo matematični oziroma računski modeli [1].

Z uporabo modelov želimo preveriti smiselnost načrtovanih poskusov že pred njihovo izvedbo ter tako zmanjšati čas in denar namenjen izvedbi eksperimentov, ki hipotetično ne bodo prinesli obetavnih rezultatov. Matematični modeli namreč omogočajo napovedovanje rezultatov bioloških eksperimentov na podlagi predhodno opravljenih meritev.

Prav tako lahko z računskimi modeli ocenimo obnašanje biološkega sistema v primerih, ko eksperimentov z obstoječimi biokemijskimi metodami ni mogoče izvesti.

Za vzpostavitev relevantnega modela je potrebno v proces njegove gradnje vključiti eksperimentalne podatke. To lahko storimo ročno na podlagi predhodnje analize rezultatov meritev ali z uporabo avtomatiziranih orodij. Ker danes uporabniki raje posegamo po slednjih, so se v računski biologiji razvili podatkovno vodeni modeli. Ti omogočajo gradnjo matematičnih modelov neposredno iz zbranih eksperimentalnih podatkov. Pri njihovi uporabi se moramo zavedati, da so meritve dovzetne za napake, zato so za gradnjo modelov primernejši pristopi, ki znajo obravnavati določen nivo negotovosti. Mednje sodijo tudi metode, ki temeljijo na mehki logiki.

Mehki podatkovno vodeni modeli signalnih poti so bili v preteklosti že predstavljeni [2-4], a je analiza njihove koristnosti za namene načrtovanja bioloških poskusov pogosto pomanjkljiva. Avtorji v svojih delih praviloma objavijo le odstopanja učnih podatkov od napovedi mehkih modelov, testna množica pa je ali zelo majhna, ali pa odstopanja modela na njej sploh niso objavljena. Ta pomanjkljivost odpira velike možnosti za pretirano prilagajanje modela učnim podatkom, s čimer zmanjša njegovo verodostojnost in uporabnost. Zavedati se moramo namreč, da so mehki modeli dovzetni za pretirano prilagajanje že zaradi velike količine parametrov, ki jih vsebujejo.

## 1.1 Cilji

V pričujočem delu želimo ovrednotiti kakovost metod za gradnjo podatkovno vodenih mehkih modelov signalnih poti. Pri tem upoštevamo številne kriterije, med katere sodijo minimizacija števila parametrov, ki jih poskušamo oceniti, verodostojnost zgrajenega modela ter možnost gradnje tako dinamičnih kot tudi statičnih modelov. Zanimariti ne smemo niti količine podatkov, ki je potrebna za vzpostavitev modela, saj pomanjkanje zadostne količine eksperimentalnih podatkov predstavlja veliko oviro za sistemsko biologijo.

Metode uporabimo za gradnjo mehkega modela signalne poti kinaz z mitogenom aktiviranih proteinov – MAPK. Ta ima pomembno vlogo pri razvoju rakavih celic, zato je bilo v preteklosti zgrajenih že veliko modelov, ki jo opisujejo s konvencionalnimi metodami računske biologije [5-10]. Zaradi velikega števila obstoječih modelov, ki so nam na voljo za primerjavo, je analiza pravilnosti delovanja naučenega modela lažja in lahko



privede do kakovostnejših zaključkov o uporabljenih računskih metodah.

## 1.2 Pregled naloge

V pričujočem delu predstavimo metode za učenje podatkovno vodenih mehkih modelov signalnih poti in analizo njihove uporabe na primeru signalne poti MAPK. V poglavju 2 predstavimo osnovne koncepte sistemske in računske biologije, ki predstavljajo izhodišče tematike dela. V poglavju 3 opišemo osnovne pristope k modeliranju signalnih omrežij ter razložimo razliko med analizo ravnotežnega stanja in analizo dinamike biološkega sistema. Izpostavimo želene lastnosti matematičnih modelov, s katerimi lahko zagotovimo njihovo verodostojnost in posledično smiselnost njihove uporabe pri načrtovanju poskusov. V poglavju 4 uvedemo osnovne pojme mehke logike, ki jih potrebujemo za razumevanje metod za učenje podatkovno vodenih mehkih modelov. Obstoječe pristope k učenju mehkih modelov podrobneje opišemo v poglavju 5, na signalno pot MAPK pa jih apliciramo v poglavju 6. Simulacijske rezultate mehkih modelov naučenih z vsako izmed metod nato primerjamo s pričakovanimi rezultati ter ovrednotimo verodostojnost metode. Na podlagi rezultatov tudi primerjamo uporabnost metod za različna področja modeliranja signalnih poti. V poglavju 7 povzamemo zaključke in iztočnice za nadaljnje raziskovalno delo.



## 2 Osnovni pojmi računske biologije in signalnih omrežij

V pričujočem poglavju razložimo osnovne pojme računske biologije, ki so predmet obravnave dela, a niso neposredno povezani s področjem računalništva ali matematike.

### 2.1 Sistemska in računska biologija

*Sistemska biologija* (angl. *systems biology*) je veda, katere predmet je celostno razumevanje bioloških procesov, celic in organizmov [11]. *Računska biologija* (angl. *computational biology*) je raziskovalno področje podatkovnega rudarjenja, računalniškega modeliranja in simulacij za potrebe razumevanja dinamike sistemov na področju ved o življenju [12].

### 2.2 Kemijske zvrsti in njihove koncentracije

*Kemijska zvrst* (angl. *chemical species*) predstavlja osnovno entiteto, ki jo opazujemo znotraj biološkega sistema. Primeri opazovanih kemijskih zvrsti so npr. *proteini*, *rastni hormoni*, *molekule DNA* itd. Njihovo prisotnost ponavadi merimo s *koncentracijo* molekul v nekem opazovanem okolju (npr. v celici) [13].

Koncentracija določa prisotnost kemijske zvrsti znotraj nekega sistema. V okviru našega dela predpostavljamo, da je prisotnost določene zvrsti homogena, torej je enaka v vseh segmentih opazovanega sistema. Prav tako predpostavljamo, da je prostornina, znotraj katere potekajo kemijske reakcije in znotraj katere se nahajajo kemijske zvrsti, fiksna. Opazovano okolje je homogeno, sistem pa je v termičnem ravnovesju. Prisotnost posamezne kemijske zvrsti opazujemo v številu elementarnih entitet, t.j. molekul, na enoto prostornine oziroma v molih na liter ( $\frac{mol}{L} = M$ ), kar označujemo tudi z izrazom *molaren*. Pri tem je v enem molu  $N_A = 6,0221415 \times 10^{23}$  elementarnih entitet, ki določajo opazovano kemijsko zvrst. Čeprav se to enoto najpogosteje uporablja pri modeliranju bioloških sistemov, se v praksi uporabljajo tudi druge. V pričujočem delu bomo koncentracijo kemijskih zvrsti, ki se nahajajo znotraj celice, merili v številu molekul na celico, tiste, ki se nahajajo zunaj celice pa v  $\frac{pg}{mL}$  (pikogrami na mililiter).

## 2.3 Signalna omrežja

*Signalna omrežja* (angl. *signaling networks*) skrbijo za prenos informacij iz okolja, v katerem se celice nahajajo, v notranjost celice [14]. Pravilnost njihovega delovanja je ključna tako za adaptacijo in preživetje celice v spremenljivem okolju kot tudi za diferenciacijo in celično smrt. Signalna omrežja sestavljajo *signalne poti* (angl. *signal transduction pathways*), ki skrbijo za aktivacijo in deaktivacijo izbranih proteinov kot odziv na *zunajcelične signale* (angl. *extracellular signals*) [15].

Zunajcelični signali lahko nastopajo v različnih oblikah, kot so vrednost pH, osmotski pritisk, oksidativni stres ali preko signalne molekule, liganda [14, 16]. V nekaterih primerih so ligandi zelo majhne molekule, ki lahko prosto prehajajo skozi celično membrano, v večini primerov pa se vežejo na zunajcelični del membranskega receptorja, s čimer povzročijo konformacijsko spremembo receptorja [16].

Po sprejemu zunajceličnega signala se v celici izvrši zaporedje dogodkov, zato ga imenujemo tudi *dražljaj* (angl. *stimulus*). Običajno prenos signala vključuje vezavo signalne molekule na zunajcelični receptor, fosforilacijo znotrajceličnega encima, ojačanje in prenos signala ter spremembo celične funkcije [14].

## 2.4 Kemijski procesi v signalnih kaskadah

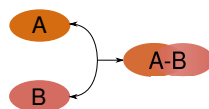
Signalna kaskada je zaporedje biokemijskih reakcij, ki ga sproži dražljaj. Procese, ki se najpogosteje pojavljajo v tem zaporedju, lahko v grobem razdelimo v dve skupini. V prvo umestimo procese, ki jih sestavljata *aktivacija* in *deaktivacija*, v drugo pa procese, ki jih tvorita *asociacija* in *disociacija*.

Najpogostejši reakciji aktivacije in deaktivacije v signalni kaskadi sta fosforilacija in defosforilacija. Sestavljajo ju protein-kinaza in protein, na katerega ta protein-kinaza deluje. Protein-kinaza ob delovanju na protein doda fosfatno skupino (*fosforilacija*), sčasoma pa se ta spontano odstrani s proteina (*defosforilacija*). Shema tega procesa je prikazana na sliki 2.1. Protein-kinaza je označena s K, protein, ki ga fosforilira, oziroma substrat s S in fosforiliran protein s pS.



**Slika 2.1:** Shema procesa fosforilacije in defosforilacije. Protein-kinaza ob delovanju na protein doda fosfatno skupino (fosforilacija), sčasoma pa se ta spontano odstrani s proteina (defosforilacija). Protein-kinaza je označena s K, protein, ki ga fosforilira, oziroma substrat s S in fosforiliran protein s pS. Povezave s puščicami predstavljajo kemijske reakcije, pri čemer je povezava usmerjena od substrata proti produktu, medtem ko povezava s kvadratom predstavlja delovanje zvrsti na reakcijo.

Procese asociacije in disociacije tvorita dve kemijski zvrsti, ki se združita v kompleksnejšo vrst (asociacija), ta pa lahko razpade na prvotni enostavnejši vrsti (disociacija). Shema procesa je prikazana na sliki 2.2. Enostavnejši vrsti sta označeni z A in B, kompleks pa z A-B.



**Slika 2.2:** Shema procesa asociacije in disociacije. Enostavnejši vrsti A in B se združita v kompleks A-B s procesom imenovanim asociacija, med disociacijo pa vrst A-B razpade na A in B.

## 2.5 Opis signalne poti MAPK

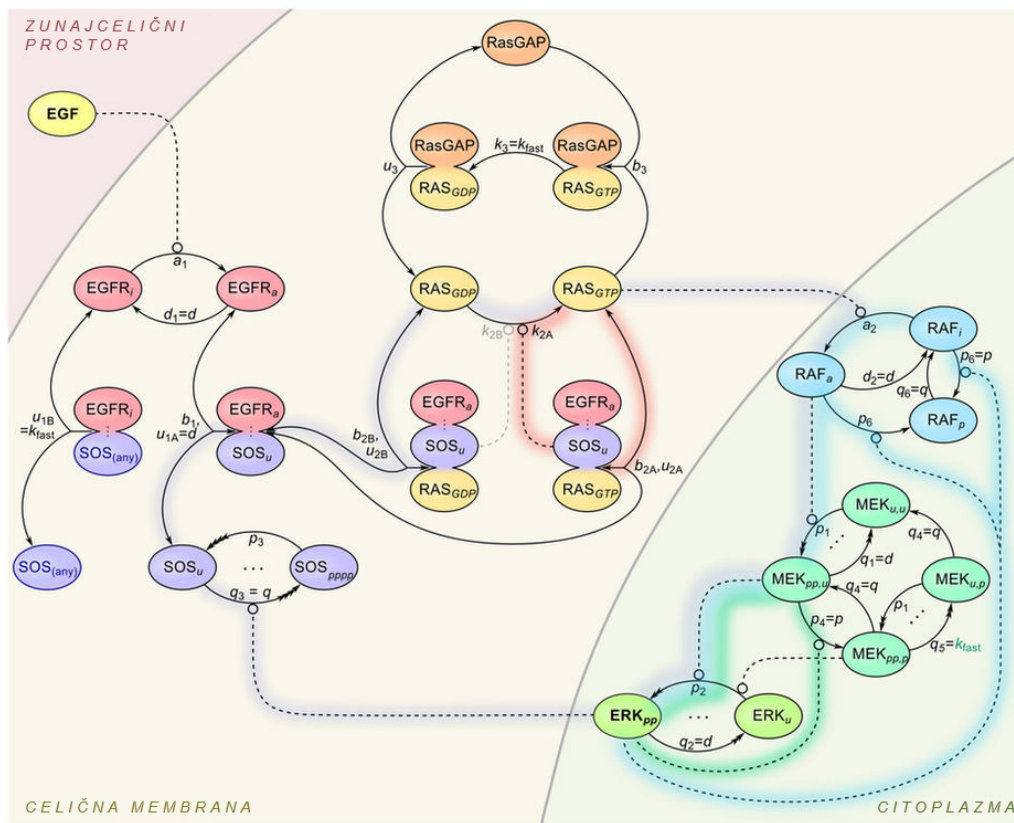
Celice prepoznavaajo in se odzivajo na zunajcelične dražljaje preko specifičnih znotraj-celičnih programov, med katere sodijo tudi signalne kaskade, ki vodijo do aktivacije *kinaz z mitogenom aktiviranih proteinov* (angl. *mitogen-activated protein kinases*, MAPKs) oziroma *MAP-kinaz* [17]. V evkariontskih celicah najdemo številne oblike poti MAPK, ki koordinirajo regulacijo raznolikih celičnih aktivnosti, od genske ekspresije, mitoze in metabolizma do gibanja, preživetja, apoptoze in diferenciacije [17]. Nepravilnosti v njihovem delovanju so prisotne pri raznolikem naboru bolezni, kot so Alzheimerjeva bolezen, Parkinsonova bolezen, amiotrofična lateralna skleroza in različne vrste raka [18]. Vsaka izmed družin MAPK je sestavljena iz množice treh kinaz, ki delujejo zaporedno. Le-te so MAP-kinaze (MAPK), kinaze MAPK (MAPKK) in kinaze kinaze MAPK (MAPKKK).

V pričujočem delu se osredotočimo na le eno družino MAPK in sicer na sesalski modul ERK1/2, znan tudi kot klasična kaskada mitogen kinaze. Klasična kaskada mitogen kinaze je daleč najbolj preučena pot iz družine MAPK, njena deregulacija pa je prisotna v približno tretjini vseh rakavih obolenj pri človeku [19].

## 2.6 Klasična kaskada mitogen kinaze

Modul ERK1/2 (imenovan tudi Raf/MEK/ERK kaskada) je dobil ime po dveh MAPK, ki nastopata v njem. To sta *z zunajceličnim signalom regulirani kinazi* (angl. *extracellular signal-regulated kinase* - ERK) 1 in 2 z oznakama ERK1 in ERK2. Močno ju aktivirajo rastni faktorji, serum in forbolni estri, manj pa ligandi heterotrimernih receptorjev, ki se vežejo na G protein, citokini, osmotski stres in disorganizacija mikrotubulov. Omenjena kaskada je tako sestavljena iz MAPKKK A-Raf, B-Raf in Raf-1, MAPKK MEK1 in MEK2 ter MAPK ERK1 in ERK2 [17]. Poenostavljena shema kaskade je prikazana na sliki 2.3 in je povzeta po viru [5]. Predpostavlja, da signalno pot aktivira *epidermalni rastni hormon* (angl. *epidermal growth factor* - EGF).

Tipično membranski receptorji kot so RTK in na G-protein sklopljeni receptorji, med katere sodi tudi *receptor za epidermalni rastni faktor* (angl. *epidermal growth factor receptor* - EGFR), prenesejo aktivacijske signale Raf/MEK/ERK kaskadi preko različnih izooblik majhnih GTP-vezavnih proteinov Ras. Aktivacija z membrano povezanega Rasa je dosežena preko SOS (angl. *son of sevenless*). SOS stimulira Ras, da spremeni GDP v GTP, ki lahko nato reagira s številnimi efektorskimi proteini, vključujoč izooblike



**Slika 2.3:** Shema klasične kaskade mitogen kinaze, povzeta po viru [5]. EGF aktivira membranski receptor EGFR, ki preko SOS povzroči fosforilacijo Rase. Ras nato aktivira Raf, ki se nadalje veže na MEK in ga fosforilira. Aktivirani MEK lahko nato fosforilira še ERK.

serin/treonin kinaze Raf. Poleg vezave Rase je za aktivacijo Rafe potrebnih tudi nekaj fosforilacij na membrani [17].

Aktivirani Raf se veže na dualni specifični kinazi MEK1 in MEK2, ki ju fosforilira, ti dve pa nato fosforilirata ERK1/2. Po aktivaciji se večina ERK1/2 akumulira v jedru celice. Pri akumulaciji imajo pomembno vlogo zadrževanje v jedru (angl. *nucleus retention*), dimerizacija, fosforilacija in sproščanje iz citoplazemskih sider (angl. *release from cytoplasmic anchors*). Aktivirana ERK1/2 fosforilirata številne substrate v vseh celičnih razdelkih – *kompartmentih* (angl. *compartments*), vključujoč številne membranske proteine, jedrne substrate in citoskeletne proteine [17].

Večina poškodb povezanih z rakom, ki vodijo do konstitutivne aktivacije signalne poti ERK, se pojavi na začetnih korakih poti. Primeri tovrstnih napak so prekomerno

izražanje receptorskih tirozin kinaz in aktivacija njihovih mutacij, produkcija ligandov, ki se vežejo na receptorje na isti ali sosednjih celicah, ter mutacije na Ras in B-Raf [19]. Nadalje, mutacije treh različnih Ras kodonov povzročijo, da je Ras konstitutivno aktiven protein [20], kar privede do nezmožnosti pretvorbe GTP nazaj v GDP. Mutacija treonina 292 v alanin eliminira 90 % fosforilacije MEK1 katalizirane z MAPK, a ne vpliva na aktivnost MEK1 [21]. Mutacije so možne tudi pri akumulaciji ERK1/2 v jedru [19].



# 3 Metode za modeliranje signalnih omrežij

Računski modeli formalizirajo kompleksne biološke in eksperimentalne procese v matematičnem jeziku, kar je uporabno tako za zbiranje kot tudi za analizo kvantitativnih podatkov. Modeliranje je zato postalo nujno potrebno orodje za področje sistemske biologije [22].

Potrebno se je zavedati, da zaenkrat ne obstaja nobena metoda, ki bi zadostila vsem potrebam na področju modeliranja bioloških sistemov. Namesto tega je na voljo širok spekter računskih tehnik, ki so primerne za različne probleme [22]. Izbira metode je odvisna od zastavljenega vprašanja, kvalitete in tipa eksperimentalnih podatkov ter stopnje predhodnega poznavanja omrežja. Prva lastnost opisuje, ali nas zanima mehanizem delovanja ali izraženi fenotip, druga predvsem ali imamo na voljo kvalitativne ali kvantitativne podatke, pri tretji pa običajno upoštevamo seznam interakcij (angl. *interaction map*) ali podrobno biokemijsko pot [23].

### 3.1 Pregled obstoječih metod

Obstoječe metode za modeliranje signalnih poti lahko v grobem razdelimo v dve skupini. V prvo sodijo *podatkovno vodene metode* (angl. *data-driven methods*), ki model zgradijo na podlagi podatkov, v drugo pa *metode vodene z znanjem* (angl. *knowledge-driven methods*), ki model zgradijo na podlagi poznavanja biološkega sistema [22, 24].

Podatkovno vodene metode predstavljajo priljubljeno orodje v sistemski biologiji. Njihov cilj je odkriti korelacije med signali oziroma med signali in celičnimi fenotipi [22, 23]. Njihova slabost je, da zahtevajo velike količine podatkov za gradnjo verodostojnega modela, poleg tega pa je težko interpretirati biološki pomen naučenih interakcij med kemijskimi zvrstmi [24].

Pri metodah vodenih z znanjem zberemo informacije o biokemijskih reakcijah iz literature in jih zapišemo v modele, ki so običajno predstavljene v obliki Boolovih mrež ali navadnih diferencialnih enačb [24]. Vzpostavljeni modeli so praviloma zelo specifični ter dobro opisujejo časovno in prostorsko dinamiko na nivoju posameznih reakcij [23]. Oviro tovrstnih metod predstavlja pomanjkljiva literatura, saj le-ta zaenkrat pokriva samo dele signalnih omrežij z omejenim številom podrobnosti o mehanizmih interakcij, zato je še vedno neprimerna za gradnjo modelov velikih signalnih omrežij, ki bi bili zmožni dobro razložiti in napovedati dinamiko omrežja [24].

Nekatere študije zato priporočajo uporabo hibridnih metod, ki združujejo prednosti obeh pristopov. Z njimi lahko ročno popravimo signalne poti iz literature in podatkovnih baz ali na podlagi eksperimentalnih podatkov identificiramo alternativne poti, ki postanejo aktivne po stimulaciji z drugačnim dražljajem [24]. Med tovrstne metode sodijo metode kot so Bayesova statistika, skriti markovski modeli (angl. hidden Markov model) in nekateri pristopi mehke logike [23, 24]. Prednost teh metod je, da jih lahko apliciramo tudi v situacijah, ko je poznavanje mehanizmov reakcij nepopolno, poznamo pa seznam interakcij med biokemijskimi zvrstmi [23].

### 3.2 Analiza ravnotežnega stanja

Pogosto opazujemo le asimptotsko obnašanje dinamičnega sistema, s čimer okrnemo analizo modelov na opazovanje njegovih bistvenih značilnosti. Čeprav je lahko asimptotsko obnašanje oscilatorno ali celo kaotično, v večini primerov sistem s časom preide v *ravnotežno stanje* (angl. *steady state*) [25].

Koncept ravnotežnega stanja je matematična idealizacija, ki ima pomembno vlogo pri *kinetičnem modeliranju*. Z izrazom kinetično modeliranje označimo matematični opis sprememb lastnosti opazovanega sistema, na primer koncentracij kemijskih zvrsti [26]. V opisih praviloma nastopajo kinetični parametri, med katere sodijo konstante hitrosti reakcij, disociacijske konstante in hitrosti degradacije različnih proteinov. Pravimo, da se sistem nahaja v ravnotežnem stanju, če se makroskopske spremenljivke v opazovanem časovnem intervalu ne spreminjajo znotraj sprejemljive natančnosti [25]. V primeru signalnih omrežij so opazovane makroskopske spremenljivke koncentracije opazovanih proteinov, torej pričakujemo, da bo v ravnotežnem stanju njihova koncentracija konstantna.

### 3.3 Analiza dinamike sistema

Dinamične lastnosti obnašanja nekaterih proteinov, med katere sodijo na primer transkripcijski faktorji, vplivajo na odziv celice na dražljaj, zato analiza ravnotežnega stanja ne zadošča za karakterizacijo odziva. Za dobro razumevanje vpliva stimulusa na celični odziv je zato potrebna tudi karakterizacija časovne dinamike biološkega sistema [2].

Za razliko od statičnih modelov oziroma modelov ravnotežnega stanja dinamični modeli opisujejo spreminjanje koncentracije kemijskih zvrsti v času, včasih pa tudi v prostoru. Čeprav so na voljo številne metode modeliranja, se najpogosteje uporabljajo modeli, ki temeljijo na diferencialnih enačbah.

*Sistemi navadnih diferencialnih enačb* (angl. *ordinary differential equations* - ODE) opisujejo le spreminjanje koncentracij v času, medtem ko *sistemi parcialnih diferencialnih enačb* (angl. *partial differential equations* - PDE) opisujejo tako časovno kot tudi prostorsko dinamiko sistema. ODE-modeli so primernejši za sisteme, kjer opazujemo veliko kemijskih zvrsti, PDE-modeli pa za sisteme, kjer opazujemo malo kemijskih zvrsti, saj se pri večjem številu zvrsti kompleksnost njihovega reševanja zelo poveča. Čeprav modeli, ki temeljijo na diferencialnih enačbah, dobro opisujejo biokemijske hitrostne zakone, je njihova slaba lastnost ta, da mora biti topologija modela znana vnaprej. Prav tako je rezultat modela močno odvisen od izbire vrednosti prostih parametrov, t.j. začetnih koncentracij proteinov in *konstant hitrosti reakcij* (angl. *reaction rate constant*). Ocenjevanje teh parametrov je časovno potratno in zahteva veliko količino učnih podatkov [27].

### 3.4 Verodostojnost modelov

Eden temeljnih ciljev modeliranja bioloških sistemov je gradnja verodostojnih modelov. To pomeni, da se rezultati simulacij modela skladajo z eksperimentalnimi podatki. Pri tem nismo omejeni le na eksperimente, ki so bili že izvedeni, temveč želimo, da bi model pravilno napovedal tudi rezultate eksperimentov, ki jih šele načrtujemo. Obstoječi eksperimentalni podatki omejujejo gradnjo našega modela, saj definirajo pričakovane rezultate pri določenih pogojih.

Modele, ki vsebujejo zelo malo podrobnosti, lahko zato dobro omejimo, a so praviloma pretirano poenostavljeni, zaradi česar je manj verjetno, da bo napoved rezultatov neizvedenih eksperimentov pravilna. Za sofisticirane modele nam po drugi strani primanjkuje eksperimentalnih podatkov, ki bi specificirali podrobne mehanizme, zato je njihova gradnja zahtevnejša. Verodostojnost modelov je torej največja, kadar model vsebuje le tiste podrobnosti, ki so potrebne za prikazovanje in napovedovanje rezultatov eksperimentov, ki nas zanimajo pri zastavljenem biološkem vprašanju [22].

Vsaka podrobnost, ki jo dodamo modelu, vnese vanj tudi parametre. Slednji lahko definirajo začetne koncentracije proteinov, konstanto hitrosti reakcije ali hitrost difuzije kemijske zvrsti skozi citoplazmo. Pogosto so neznani, zato jih je potrebno oceniti iz podatkov, kar zmanjša verodostojnost modela. Zavedati se je potrebno, da vsaka podrobnost, ki je dodana modelu, poveča število zahtevanih parametrov [22].

# 4 Osnove mehke logike

Teorija mehkih množic je bila v preteklosti uspešno aplicirana na številna področja, med katera sodijo krmiljenje, identifikacija sistemov, prepoznavanje vzorcev in procesiranje signalov. Še posebej je zanimiva njena zmožnost obravnave nenatančnosti, ki je prisotna v naravnem jeziku. Zaradi te lastnosti je mehka logika postala močno orodje za reševanje problemov, pri katerih imamo na razpolago strokovno znanje, še bolj pa je prikladna za gradnjo podatkovno vodenih modelov v obliki sistema pravil, ki so preprosta za razumevanje [28].

## 4.1 Mehka množica

**Definicija 1.** Mehka množica (*angl.* fuzzy set)  $\tilde{A}$  elementov  $x$  iz prostora  $U$  je množica urejenih parov

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in U\}, \quad (4.1)$$

kjer je  $\mu_{\tilde{A}}$  pripadnostna funkcija (*angl.* membership function) mehke množice  $\tilde{A}$  z zalogo vrednosti  $[0, 1]$ , na podlagi katere izračunamo stopnjo pripadnosti posameznega elementa  $x$  k mehki množici  $\tilde{A}$  [29].

## 4.2 Mehka spremenljivka

*Mehka spremenljivka* je mehka množica oziroma množica vrednosti pripadnostnih funkcij posameznim lingvističnim ali opisnim množicam [30]. Formalno je določena z naslednjo definicijo.

**Definicija 2.** *Mehka spremenljivka je določena s trojico  $(S, U, T)$ , kjer  $S$  predstavlja ime spremenljivke,  $U$  zalogo ostrih vrednosti, nad katerimi je opazovana mehka spremenljivka  $S$  definirana, in  $T$  nabor mehkih množic  $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ , ki so definirane nad  $U$  in s pomočjo katerih predstavimo vrednosti mehke spremenljivke  $S$  [29].*

Postopek kreiranja mehkih vhodnih spremenljivk imenujemo *mehčanje* ali *fuzifikacija* (angl. *fuzzification*). V sklop mehčanja sodi tudi določitev števila mehkih množic, ki pokrivajo vhodne spremenljivke, ter izbira oblike ustreznih funkcij pripadnosti [29]. Funkcija pripadnosti mehki množici je lahko poljubne oblike, vendar se najraje držimo konveksnih oblik. Če pripadnost ni konveksna, se lahko zgodi, da za ustrezno mehko množico ne veljajo določeni postopki, ki jih sicer teorija ponuja načrtovalcu mehkega sistema [30]. Oblike pripadnostnih funkcij nimajo večjega vpliva na uspešnost mehkega sistema. V našem delu se bomo omejili na gaussove pripadnostne funkcije, ki jih opredelimo s srednjo vrednostjo  $c$  in standardnim odklonom  $\sigma$ , izračunamo pa jih na podlagi izraza

$$\mu_{\tilde{A}}(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}. \quad (4.2)$$

## 4.3 Mehka pravila

*Mehka pravila* opisujejo vzroke in posledice v sistemu, ki ga modeliramo z mehko logiko. Pri tem so vzroki trenutne vhodne mehke spremenljivke oziroma njihove funkcijske pripadnosti, posledice pa trenutne izhodne mehke spremenljivke oziroma ustrezne funkcijske pripadnosti. Obstaja več vrst mehkih pravil, vendar se največ uporabljata dva tipa. To sta pravilo tipa Zadeh-Mamdani in pravilo tipa Takagi-Sugeno. Obe pravili sta oblike IF-THEN-(ELSE), kar pomeni, da v obravnavanem sistemu povezujejo vzroke s posledicami s pomočjo mehkih množic. Pogoj ELSE je zajet implicitno skozi druga pravila [29].

Mehko pravilo tipa Zadeh-Mamdani [31] je oblike

$$\text{IF } x \text{ is } \tilde{A} \text{ THEN } y \text{ is } \tilde{B}, \quad (4.3)$$

kjer sta ( $x$  is  $\tilde{A}$ ) in ( $y$  is  $\tilde{B}$ ) dve mehki trditvi oziroma prepoziciji, ki pomenita, da je vhodna spremenljivka  $x$  v območju mehke množice  $\tilde{A}$  z vrednostjo pripadnosti  $\mu_{\tilde{A}}(x)$ , izhodna spremenljivka  $y$  pa v območju mehke množice  $\tilde{B}$  z vrednostjo pripadnostne funkcije  $\mu_{\tilde{B}}(y)$ . Splošna oblika tega pravila, ki upošteva poljubno število vhodnih spremenljivk, je oblike

$$\text{IF } x_1 \text{ is } \tilde{A}_1 \text{ AND } x_2 \text{ is } \tilde{A}_2 \text{ AND } \dots \text{ AND } x_k \text{ is } \tilde{A}_k \text{ THEN } y \text{ is } \tilde{B} \quad (4.4)$$

Mehko pravilo tipa Takagi-Sugeno [32] se od pravila tipa Zadeh-Mamdani razlikuje v posledičnem delu, saj je podan s funkcijo vhodnih spremenljivk in ne z izhodno mehko množico

$$\text{IF } x \text{ is } \tilde{A} \text{ AND } y \text{ is } \tilde{B} \text{ THEN } z = f(x, y). \quad (4.5)$$

V splošnem je vhodnih pogojev več, funkcija pa je pogosto linearna

$$\text{IF } x_1 \text{ is } \tilde{A}_1 \text{ AND } x_2 \text{ is } \tilde{A}_2 \text{ AND } \dots \text{ AND } x_m \text{ is } \tilde{A}_m \text{ THEN } z = C_0 + C_1 x_1 + \dots + C_m x_m. \quad (4.6)$$

Pravila tipa Takagi-Sugeno so bolj priljubljena pri pristopih, kjer opazujemo kakovost aproksimacije. Čeprav lahko pod določenimi pogoji prevedemo sistem iz ene oblike v drugo, pa je znanje izraženo v pravilih tipa Zadeh-Mamdani transparentnejše (razumljivejše) od znanja izraženega v pravilih tipa Takagi-Sugeno [28].

## 4.4 Mehke relacije

*Mehke relacije* so mehke podmnožice  $X \times Y$ , torej preslikave  $X \rightarrow Y$ . Matematično so definirane z definicijo v nadaljevanju.

**Definicija 3.** *Naj bosta  $X, Y \subseteq \mathbb{R}$  univerzalni množici. Potem je*

$$\tilde{R} = \{(x, y), \mu_{\tilde{R}}(x, y) \mid (x, y) \in X \times Y\} \quad (4.7)$$

*mehka relacija na  $X \times Y$  [29].*

Nadalje lahko definiramo mehke relacije tudi na poljubnih mehkih množicah [29].

**Definicija 4.** Naj bosta  $X, Y \subseteq \mathbb{R}$  univerzalni množici ter  $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X\}$  in  $\tilde{B} = \{(y, \mu_{\tilde{B}}(y)) \mid y \in Y\}$  mehki množici. Potem je

$$\tilde{R} = \{((x, y), \mu_{\tilde{R}}(x, y)) \mid (x, y) \in X \times Y\} \quad (4.8)$$

mehka relacija na  $\tilde{A} \times \tilde{B}$ , če velja:

$$\mu_{\tilde{R}}(x, y) \leq \mu_{\tilde{A}}(x), \forall (x, y) \in X \times Y \quad (4.9)$$

in

$$\mu_{\tilde{R}}(x, y) \leq \mu_{\tilde{B}}(y), \forall (x, y) \in X \times Y. \quad (4.10)$$

Relacija kompozicije ali kratko *kompozicija* dveh mehkih relacij  $R_1(A, B)$  in  $R_2(B, C)$  je relacija  $R(A, C)$ , ki jo dobimo po zaporedni izvedbi relacij  $R_1$  in  $R_2$ . Tipična kompozicija je *max-min*.

**Definicija 5.** Naj bosta  $\tilde{R}_1(x, y)$ ,  $(x, y) \in X \times Y$  in  $\tilde{R}_2(y, z)$ ,  $(y, z) \in Y \times Z$  mehki relaciji. Kompozicija *max-min* relacij  $\tilde{R}_1(x, y) \circ \tilde{R}_2(y, z)$  je mehka množica

$$\tilde{R}_1(x, y) \circ \tilde{R}_2(y, z) = \{((x, z), \max_y \{\min(\mu_{\tilde{R}_1}(x, y), \mu_{\tilde{R}_2}(y, z)) \mid x \in X, y \in Y, z \in Z\})\} \quad (4.11)$$

s pripadnostno funkcijo  $\mu_{\tilde{R}_1 \circ \tilde{R}_2}$  [29].

## 4.5 Mehko sklepanje

Mehko sklepanje ali mehka inferenca je sestavljena iz treh korakov, ki jih imenujemo *agregacija*, *mehka implikacija* in *akumulacija*. Agregacija je potrebna, če imamo v pogojnem delu pravila več kot eno mehko množico. Na podlagi vrednosti vhodnih mehkih spremenljivk za vsako pravilo izračunamo *stopnjo izpolnjenosti pogojnega dela pravila* [30]. Stopnjo izpolnjenosti mehkega pravila  $r$ , pri katerem so spremenljivke  $S_i$  povezane z operacijo AND, dobimo po izrazu

$$\alpha_r = \min_{i=1, \dots, n} \mu_j^i(S_i^{\text{input}}). \quad (4.12)$$

V primeru, da so spremenljivke  $S_i$  povezane z operacijo OR, stopnjo izpolnjenosti pravila izračunamo po izrazu

$$\alpha_r = \max_{i=1, \dots, n} \mu_j^i(S_i^{\text{input}}). \quad (4.13)$$



Avtor	$\mu_r^{B'}(y)$
Zadeh	$\max\{1 - \alpha_r, \min\{\alpha_r, \mu_{B'}(y)\}\}$
Lukasiewicz	$\min\{1, 1 - \alpha_r + \mu_{B'}(y)\}$
Mamdani	$\min\{\alpha_r, \mu_{B'}(y)\}$
Kleene	$\max\{1 - \alpha_r, \mu_{B'}(y)\}$

**Tabela 4.1:** Najbolj uporabne implikacije v mehki logiki.

Mehka implikacija omogoča sklepanje, to je prehod iz IF-dela v THEN-del mehkega pravila. To ni enolično določena logična relacija, saj lahko pri različnih primerih uporabimo različne implikacije [30]. *Sklepe* posameznega pravila izračunamo po enem izmed izrazov iz tabele 4.1.

Akumulacija poskrbi za izračun končne vrednosti izhodne mehke spremenljivke, ki je predstavljena s pripadnostmi k posameznim mehkim množicam izhodne spremenljivke [30]. Dobimo jo z združitvijo sklepov  $\mu_r^B(y)$  vseh pravil po izrazu

$$\mu(y) = \max_r \mu_r^B(y). \quad (4.14)$$

## 4.6 Ostrenje

*Ostrenje* ali *defuzifikacija* (angl. *defuzzification*) je postopek pretvorbe izhodne mehke spremenljivke v končno realno (ostro, angl. *crisp*) numerično obliko. V rabi so različne metode ostrenja.

### 4.6.1 Težiščna metoda

*Težiščna metoda* (angl. *center of gravity* - COG ali *center of area* - COA) sodi med najbolj priljubljene metode ostrenja. Predpostavimo, da lahko rezultat procesiranja  $\tilde{R}$  opišemo z zvezno funkcijo pripadnosti  $\mu_{\tilde{R}}(y)$ . Ostra vrednost  $R'$  je tedaj

$$R' = \frac{\int_0^\infty y \mu_{\tilde{R}}(y) dy}{\int_0^\infty \mu_{\tilde{R}}(y) dy}. \quad (4.15)$$

$R'$  je tista ostra vrednost, ki najbolj ustreza rezultativni mehki množici  $\tilde{R}$ . V primeru, da je pripadnostna funkcija  $\mu_{\tilde{R}}(y)$  le delno zvezna ali diskretna, izračunamo ostro rezultatno vrednost po naslednjem izrazu

$$R' = \frac{\sum_{i=1}^n y_i \mu_{\tilde{R}}(y_i)}{\sum_{i=1}^n \mu_{\tilde{R}}(y_i)}. \quad (4.16)$$

#### 4.6.2 Metoda središčnih vsot

*Metoda središčnih vsot* (angl. *center of sums* - COS) je procesorsko manj zahtevna od težiščne metode. Temelji na uniji  $\tilde{R} = \cup_{i=1}^q \tilde{R}_i$ , pri  $q$  aktivnih pravilih. Prekrivanje delnih rezultatov  $\tilde{R}_i \cap \tilde{R}_j \neq \emptyset$  pride le enkrat do izraza. Ostro vrednost rezultata izračunamo po izrazu

$$R' = \frac{\int_0^\infty y \sum_{i=1}^q \mu_{\tilde{R}_i}(y) dy}{\int_0^\infty \sum_{i=1}^q \mu_{\tilde{R}_i}(y) dy}, \quad (4.17)$$

oziroma v diskretnem primeru po izrazu

$$R' = \frac{\sum_{j=1}^n y_j \sum_{i=1}^q \mu_{\tilde{R}_i}(y_j)}{\sum_{j=1}^n \sum_{i=1}^q \mu_{\tilde{R}_i}(y_j)}, \quad (4.18)$$

kjer je  $n$  število diskretnih vrednosti, ki sodelujejo pri opredelitvi ostrega rezultata  $R'$ .

#### 4.6.3 Metode maksimumov

*Metoda najnižjega maksimuma* (angl. *first of maxima* - FOM) ugotovi, katero mehko pravilo najbolj učinkuje na rezultat. Med vsemi elementi z najvišjo pripadnostjo izberemo najmanjšega. Ostro vrednost rezultata zato izračuna kot

$$R' = \inf_{y \in Y} \{y \in Y \mid y_R(y) = \sup_{y \in Y} \mu_R(y)\}. \quad (4.19)$$

Podobno *metoda najvišjega maksimuma* (angl. *last of maxima* - LOM) med vsemi elementi z najvišjo pripadnostjo izbere največjega (4.20).

$$R' = \sup_{y \in Y} \{y \in Y \mid y_R(y) = \sup_{y \in Y} \mu_R(y)\}. \quad (4.20)$$

*Metoda srednjega maksimuma* (angl. *middle of maxima* - MOM) izračuna srednjo vrednost metod FOM in LOM (4.21).

$$R' = \frac{\sup_{y \in Y} \{y \in Y \mid y_R(y) = \sup_{y \in Y} \mu_R(y)\} + \inf_{y \in Y} \{y \in Y \mid y_R(y) = \sup_{y \in Y} \mu_R(y)\}}{2}. \quad (4.21)$$

## 4.7 Mehki model

Mehki model sestavimo iz treh delov [30]:

- iz mehkih vhodnih in izhodnih spremenljivk ter njihovih mehkih vrednosti,
- mehkih pravil in
- mehkih metod sklepanja ali inferenčnih metod, ki vsebujejo še postopke mehčanja in ostrenja.

Ogrodje mehke logike podpira številne mehanizme prilagodljivosti, med katere sodijo naklon in oblika pripadnostnih funkcij, teža pravil, metode mehčanja in ostrenja ter oblika pravil [23].

Predstavljeni koncepti mehke logike omogočajo vzpostavitev mehkih modelov signalnih poti. V poglavju 5 predstavimo metode, ki omogočajo avtomatsko gradnjo mehkih modelov iz eksperimentalno pridobljenih podatkov, v poglavju 6 pa jih apliciramo na modul ERK1/2 iz družine signalnih poti MAPK. Primerjamo kvaliteto izgrajenih mehkih modelov ter na primeru predstavimo in utemeljimo prednosti, ki jih ima uporaba mehke logike pri izgradnji modelov signalnih poti pred drugimi pristopi modeliranja, ki so v uporabi na področju sistemske biologije.



# 5 Metode za gradnjo mehkih podatkovno vodenih modelov signalne poti

Naš cilj je vzpostavitev metodologije za učenje mehkih modelov na pomanjkljivih kvantitativnih bioloških podatkih. To pomeni, da meritev nismo mogli opraviti pod vsemi možnimi pogoji in da podatkov, ki so na voljo, ni dovolj, da bi omejili tako topologijo kot tudi kvantitativne parametre omrežij. Zato stremimo k razvoju mehkega logičnega sistema, ki minimizira število parametrov, s čimer se izognemo pretiranemu prilagajanju učnim podatkom, hkrati pa poenostavimo logično strukturo za lažjo interpretabilnost modela. Ker želimo predstavljati razmerja med proteini in encimskimi kaskadami, morajo biti tudi matematične relacije biološko relevantne [3]. Zaželeno je tudi, da metoda omogoča uporabo kvalitativnih podatkov. Slednje je namreč preprosteje izmeriti v sklopu eksperimentalnega dela [2].

V pričujočem poglavju predstavimo nekaj metod, ki omogočajo učenje mehkih modelov na podlagi eksperimentalnih podatkov.

## 5.1 Mehka metoda voditeljev

Mehka metoda voditeljev (angl. *fuzzy c-means clustering algorithm* - FCM) [33] je osnovni mehki algoritem za gručenje, ki poskuša najti mehko particijo  $U = [u_{ik}]$  nabora podatkov preko minimizacije posplošenega funkcionala najmanjših kvadratov

$$J_m(X, U, v) = \sum_{k=1}^N \sum_{i=1}^c u_{ik}^m d^2(x_k, v_i), \quad (5.1)$$

kjer so  $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^n$  podatki,  $c$  število skupin v množici  $X$  ( $2 \leq c < N$ ),  $m \geq 1$  stopnja mehčanja za odstranjevanje šuma iz podatkov,  $U$  mehka particija množice  $X$  in  $v = [v_i]$  vektor centrov skupin. Minimizacija poteka iterativno pod naslednjimi pogoji:

$$0 \leq u_{ik} \leq 1; \quad 1 \leq i \leq c, 1 \leq k \leq N, \quad (5.2)$$

$$0 < \sum_{k=1}^N u_{ik} \leq n; \quad 1 \leq i \leq c, \quad (5.3)$$

$$\sum_{i=1}^c u_{ik} = 1; \quad 1 \leq k \leq N. \quad (5.4)$$

Po vsaki iteraciji centre  $v_i$  in pripadnostne stopnje  $u_{ik}$  posodobimo po naslednjem postopku:

$$v_i = \frac{\sum_{k=1}^N u_{ik}^m x_k}{\sum_{k=1}^N u_{ik}^m}; \quad 1 \leq i \leq c, \quad (5.5)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{d(x_k, v_i)}{d(x_k, v_j)} \right)^{\frac{2}{m-1}}}; \quad 1 \leq k \leq N, 1 \leq i \leq c. \quad (5.6)$$

Naj ima mehki model  $n$  vhodnih in  $m$  izhodnih spremenljivk. Tedaj učenje mehkega modela s FCM poteka na  $n+m$  dimenzionalnih podatkih. Vsak podatek je pri tem vektor, ki vsebuje znane vrednosti vhodnih in pričakovane vrednosti izhodnih spremenljivk pri danih vhodih. Tako sestavljene podatke nato gručimo v  $c$  skupin, vsaka skupina pa predstavlja po eno mehko pravilo. Iz centrov skupin določimo tudi pripadnostne funkcije za posamezno spremenljivko.

Metodo lahko uporabimo tako za učenje modela ravnotežnega stanja kot tudi za gradnjo dinamičnega modela. Pri gradnji modela ravnotežnega stanja signalne poti, je edina vhodna spremenljivka koncentracija dražljaja. Izhodni podatki so koncentracije kemijskih zvrsti v ravnotežnem stanju.

Kot vhodne podatke dinamičnega modela signalne poti podamo zaporedje koncentracij kemijskih zvrsti, kot izhodne pa spremembe koncentracij kemijskih zvrsti med dvema zaporednima meritvama. Pri izhodnih podatkih zanemarimo spremembo koncentracije dražljaja, za katerega predpostavimo, da je konstanten skozi celoten čas simulacije.

## 5.2 Mehke časovne vrste

*Mehka časovna vrsta* (angl. *fuzzy time series* - FTS) je napovedni model, ki omogoča modeliranje dinamičnih procesov, katerih opažanja so lingvistične vrednosti. Pri tem predpostavljamo, da so opažanja v nekem trenutku rezultat opažanj iz preteklosti [34]. Strožje je določena z naslednjo definicijo.

**Definicija 6.** Naj bo  $Y(t) \subseteq \mathbb{R}$ ,  $t \in \mathbb{Z}$ , definicijsko območje mehkih množic  $f_i(t)$ ,  $i \in \mathbb{N}$ , in  $F(t) = \{f_i(t) \mid i \in \mathbb{N}\}$ . Potem  $F(t)$  pravimo mehka časovna vrsta na  $Y(t)$ ,  $t \in \mathbb{Z}$ .

Nadalje naj bo  $I$  množica indeksov za  $F(t-1)$  in  $J$  množica indeksov za  $F(t)$ .

**Definicija 7.** Naj za vsak  $f_j(t) \in F(t)$ ,  $j \in J$ , obstajata  $f_i(t-1) \in F(t-1)$ ,  $i \in I$ , in mehka relacija  $R_{ij}(t, t-1)$ , tako da velja  $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$ , kjer  $\circ$  označuje kompozicijo *max-min*. Pravimo, da le  $F(t-1)$  vpliva na  $F(t)$ , kar krajše zapišemo kot

$$F(t-1) \rightarrow F(t). \quad (5.7)$$

**Definicija 8.** Naj za vsak  $f_j(t) \in F(t)$ ,  $j \in J$ , obstajata  $f_i(t-1) \in F(t-1)$ ,  $i \in I$ , in mehka relacija  $R_{ij}(t, t-1)$ , tako da velja  $f_j(t) = f_i(t-1) \circ R_{ij}(t, t-1)$ , kjer  $\circ$  označuje kompozicijo *max-min*. Teda relaciji  $R(t, t-1) = \bigcup_{i,j} R_{ij}(t, t-1)$  pravimo mehka relacija med  $F(t)$  in  $F(t-1)$ , kar lahko zapišemo tudi z naslednjim izrazom:

$$F(t) = F(t-1) \circ R(t, t-1). \quad (5.8)$$

Običajni postopek za gradnjo modela s pomočjo mehkih časovnih vrst je naslednja:

1. Določimo definicijska območja, na katerih bomo definirali mehke množice.
2. Zberemo zgodovinske podatke.
3. Na podlagi zgodovinskih podatkov definiramo mehke množice.
4. Na podlagi zgodovinskih podatkov določimo mehke relacije.
5. Združimo relacije definirane v koraku 4.
6. Model apliciramo na nove podatke in jih po potrebi ostrimo.

### 5.3 Metoda večatributne mehke časovne vrste

Posebna oblika postopka za gradnjo modela s pomočjo mehkih časovnih vrst je metoda večatributne mehke časovne vrste (angl. *multi-attribute fuzzy time series method*), v nadaljevanju tudi MAFTS [35], ki tretji korak algoritma opisanega v razdelku 5.2 opravi s pomočjo mehkega algoritma za gručenje. Sestavljen je iz naslednjih štirih korakov:

1. Gručenje časovne vrste  $S(t)$  v  $c$  skupin z namenom identificiranja vzorcev.
2. Rangiranje posamezne skupine in mehčanje časovne vrste  $S(t)$  v mehko časovno vrsto  $F(t)$ .
3. Določanje mehkih relacij.
4. Napovedovanje novih podatkov in ostrenje rezultatov metode.

Zaradi lažjega učenja predpostavimo, da velja  $F(t-1) \rightarrow F(t)$  oziroma da je naslednje stanje sistema odvisno le od predhodnega in od vrednosti dražljaja. Tak pristop predpostavlja tudi model signalne poti predstavljen v [2].

Učni podatki, ki jih uporabimo, so koncentracije kemijskih zvrsti v zaporednih časovnih trenutkih. Podatke gručimo za vsako kemijsko zvrst posebej, s čimer določimo pripadnostne funkcije za posamezno spremenljivko. Srednje vrednosti gaussovih pripadnostnih funkcij določimo iz rezultatov metode FCM tako, da jim pripišemo vrednosti centrov skupin, standardne odklone pa moramo določiti naknadno.



# 6 Rezultati

Zaradi vloge, ki jo igra družina signalnih poti MAPK pri razvoju bolezni, je bilo v preteklosti zgrajenih kar nekaj modelov te družine signalnih poti [5–10], a se bomo v nadaljevanju osredotočili na modul ERK1/2. Večina obstoječih modelov temelji na ODE, njihovi kinetični parametri pa so bili ocenjeni iz eksperimentalnih podatkov. Ti lahko med drugim predstavljajo dobro izhodišče za gradnjo mehkih modelov in testiranje njihove učinkovitosti. V pričujočem poglavju apliciramo metode za gradnjo mehkih podatkovno vodenih modelov na izbrano signalno pot MAPK. Rezultate zgrajenega modela primerjamo z referenčnim ODE-modelom.

## 6.1 Predpostavke in poenostavitve mehkih modelov

Vse modele bomo gradili s pravili tipa Zadeh-Mamdani. Kot smo že omenili, se bomo omejili na gaussove pripadnostne funkcije tako vhodnih kot tudi izhodnih mehkih spremenljivk. Zaradi enostavnosti bodo teže vseh mehkih pravil enake 1, spremenljivke pa bodo vedno povezane z operacijo AND, saj nas hkrati zanimajo koncentracije vseh prisotnih kemijskih zvrsti. Sklepali bomo po Mamdanijevi metodi, za ostrenje pa bomo

uporabljali težiščno metodo. Kljub naštetim predpostavkam vsi pristopi opisani v nadaljevanju ohranjajo splošnost.

### 6.1.1 Učni podatki

Ker eksperimentalnih podatkov nimamo na voljo, učne podatke generiramo s pomočjo referenčnega ODE-modela. Tako generirani učni podatki simulirajo podatke, ki bi jih sicer pridobili eksperimentalno. Potrebujemo dva različna tipa učnih podatkov, eno množico podatkov za gradnjo modela ravnotežnega stanja in drugo množico za gradnjo dinamičnega modela.

Podatke za učenje statičnega modela pridobimo iz referenčnega modela tako, da ga poženemo za dovolj dolg časovni interval, da se koncentracije proteinov več ne spreminjajo. Empirično smo ugotovili, da za opazovani referenčni model zadošča 30 minut, da sistem vzpostavi ravnotežno stanje. Postopek ponovimo pri šestnajstih različnih vrednostih koncentracije dražljaja EGF in sicer 0, 2, 5, 10, 20, 50, 60, 70, 80, 90, 100, 150 in  $200 \frac{pg}{mL}$ .

Podatke za učenje dinamičnih modelov pridobimo iz referenčnega modela tako, da shranimo vrednosti vseh koncentracij enkrat na minuto čez časovno obdobje 30 minut. Postopek ponovimo pri štirih različnih vrednostih koncentracije dražljaja EGF in sicer 2, 5, 10 in  $20 \frac{pg}{mL}$ . Te vrednosti smo izbrali tako, da so znotraj intervala, na katerem so v viru [5] izvajali poskuse.

### 6.1.2 Pogoji testiranja

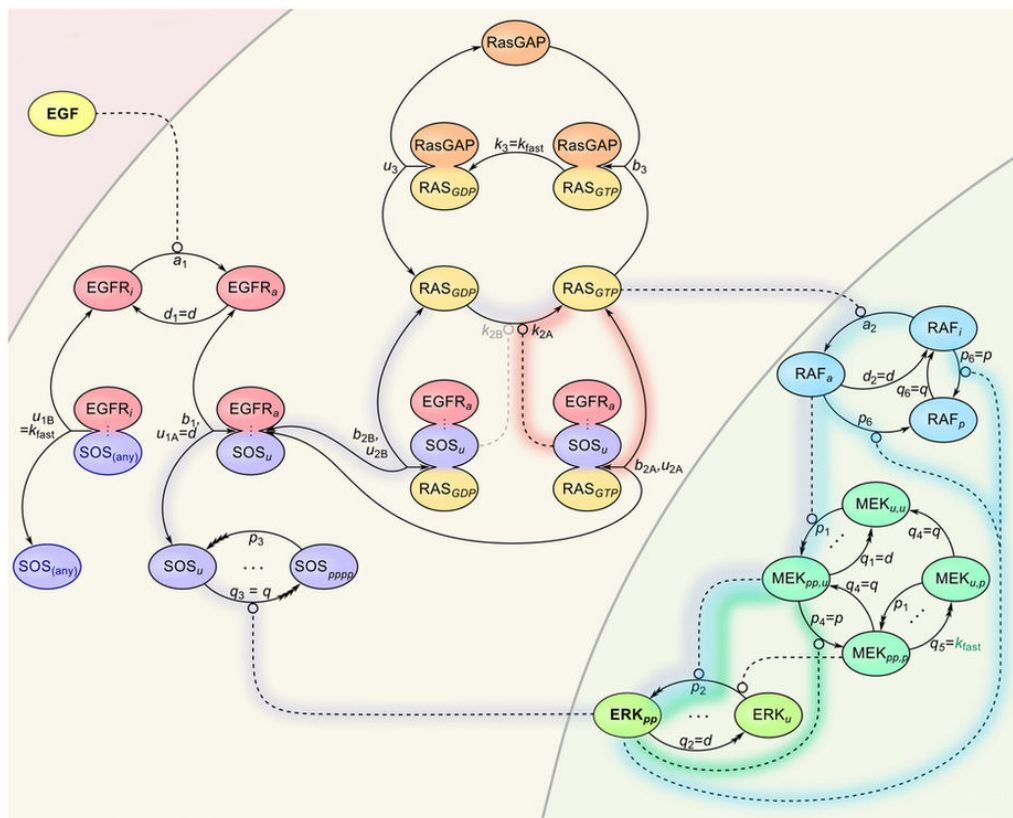
Za testiranje učinkovitosti oziroma pravilnosti metode poskusimo nekoliko spremeniti začetne pogoje in opazujemo obnašanje naučenega modela. Na ta način želimo preveriti, ali se metoda pretirano prilagaja učnim podatkom. Najprej poskusimo spremeniti le vrednost koncentracije EGF, nato pa tudi začetne koncentracije nekaterih kemijskih zvrsti v celici. Rezultate primerjamo z rezultati referenčnega modela.

Za testiranje občutljivosti dinamičnih modelov na spremembe vrednosti koncentracij EGF uporabimo štiri različne koncentracije in sicer 3, 8, 13 in  $18 \frac{pg}{mL}$ , ki so vse znotraj definicijskega območja mehke vhodne spremenljivke EGF, za testiranje občutljivosti statičnih modelov pa uporabimo večji nabor koncentracij. Za testiranje sprememb začetnih vrednosti, kar je smiselno opazovati le v primeru dinamičnih modelov, generiramo začetne pogoje naključno, vendar pri tem pazimo, da so znotraj definicijskih območij opazovanih

koncentracij kemijskih zvrsti.

## 6.2 Referenčni model signalne poti MAPK

Kot referenčni model, na podlagi katerega se učimo ter testiramo mehke modele, smo izbrali ODE-model objavljen v [5]. Njegova shema je prikazana na sliki 6.1. Ker sta ERK1 in ERK2 v različnih tkivih različno izražena [17], v našem poenostavljenem modelu predpostavimo prisotnost le ene ERK. Prav tako upoštevamo le eno MEK, pri kateri prva fosforilacija označuje mesto, ki ga fosforilira Raf [21], druga pa fosforilacijo treonin 292 (dodatek k članku [5]), torej mesta, ki ga fosforilira ERK [21]. Model predpostavlja 27



**Slika 6.1:** Shema ODE-modela signalne poti, ki je povzeta po viru [5]. EGF aktivira membranski receptor EGFR, ki preko SOS povzroči fosforilacijo Rasa. Ras nato aktivira Raf, ki se nadalje veže na MEK in ga fosforilira. Aktivirani MEK lahko nato fosforilira še ERK.

kemijskih reakcij med 31 kemijskimi zvrstmi. Parametri modela so prikazani v tabeli 6.1.

parameter	vrednost parametra	opis
$egfr_{tot}$	$3 \times 10^5 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul EGFR
$ras_{tot}$	$6 \times 10^4 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul Ras
$sos_{tot}$	$10^5 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul SOS
$rasgap_{tot}$	$6 \times 10^3 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul RasGAP
$raf_{tot}$	$5 \times 10^5 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul Raf
$mek_{tot}$	$2 \times 10^5 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul MEK
$erk_{tot}$	$3 \times 10^6 \frac{\text{kopij}}{\text{celico}}$	skupno število molekul ERK
$\alpha_1$	$5 * 10^{-5} (\frac{pg}{mL})^{-1} s^{-1}$	hitrost aktivacije EGFR prek EGF
$\alpha_2$	$10^{-7} (\frac{\text{molekul}}{\text{celico}}) s^{-1}$	hitrost aktivacije Raf prek Ras <sub>GTP</sub>
$\beta_1$	$10^{-5} (\frac{\text{molekul}}{\text{celico}}) s^{-1}$	hitrost vezave aktivnega EGFR in defosforiliranega SOS
$\rho_1$	$10^{-7} (\frac{\text{molekul}}{\text{celico}}) s^{-1}$	hitrost fosforilacije MEK prek Raf
$\rho_2$	$3 * 10^{-6} (\frac{\text{molekul}}{\text{celico}}) s^{-1}$	hitrost fosforilacije ERK prek MEK
$\delta_1$	$0,01 s^{-1}$	hitrost deaktivacije EGFR
$\delta_2$	$0,01 s^{-1}$	hitrost deaktivacije Raf
$\nu_{1A}$	$0,01 s^{-1}$	hitrost disociacije aktivnega kompleksa EGFR-SOS
$\sigma_1$	$0,01 s^{-1}$	hitrost defosforilacije MEK
$\sigma_2$	$0,01 s^{-1}$	hitrost defosforilacije ERK
$\beta_{2A}$	$10^{-6} s^{-1}$	hitrost asociacije SOS in Ras <sub>GTP</sub>
$\beta_{2B}$	$0,1 * \beta_{2A}$	hitrost asociacije SOS in Ras <sub>GDP</sub>
$\nu_{1B}$	$100 s^{-1}$	hitrost disociacije neaktivnega kompleksa EGFR-SOS
$\nu_{2A}$	$1 s^{-1}$	hitrost disociacije kompleksa SOS-Ras <sub>GTP</sub>
$\nu_{2B}$	$1 s^{-1}$	hitrost disociacije kompleksa SOS-Ras <sub>GDP</sub>
$\kappa_{2A}$	$10^{-4} s^{-1}$	stopnja aktivnosti SOS, ko je nanj vezan Ras <sub>GTP</sub>
$\kappa_{2B}$	$0,1 * \kappa_{2A}$	stopnja aktivnosti SOS, ko je nanj vezan Ras <sub>GDP</sub>
$\kappa_{2C}$	$0 s^{-1}$	stopnja aktivnosti SOS, ko nanj ni vezan Ras
$\beta_3$	$10^{-5} s^{-1}$	hitrost asociacije RasGAP in Ras <sub>GTP</sub>
$\kappa_3$	$100 s^{-1}$	hitrost hidrolize Ras <sub>GTP</sub> v Ras <sub>GDP</sub> zaradi RasGAP

$v_3$	$0,01 \text{ s}^{-1}$	hitrost disociacije RasGAP iz RasGDP po hidrolizi
$\rho_3$	$3 * 10^{-9} \text{ s}^{-1}$	hitrost fosforilacije nevezanega SOS prek ERK
$\rho_4$	$6 * 10^{-10} \text{ s}^{-1}$	hitrost fosforilacije MEK prek ERK
$\rho_6$	$6 * 10^{-10} \text{ s}^{-1}$	hitrost fosforilacije Raf prek ERK
$\sigma_3$	$3 * 10^{-4} \text{ s}^{-1}$	hitrost defosforilacije SOS
$\sigma_4$	$3 * 10^{-4} \text{ s}^{-1}$	hitrost defosforilacije MEK
$\sigma_6$	$3 * 10^{-4} \text{ s}^{-1}$	hitrost defosforilacije Raf
$\sigma_5$	$100 \text{ s}^{-1}$	hitrost defosforilacije MEK <sub>T292p</sub>
$f_{\text{ERK}_{pp} \mapsto \text{SOS}_1}$	1	jakost vpliva povratne zanke od ERK <sub>pp</sub> do SOS
$f_{\text{ERK}_{pp} \mapsto \text{MEK}_1}$	1	jakost vpliva povratne zanke od ERK <sub>pp</sub> do MEK
$f_{\text{ERK}_{pp} \mapsto \text{RAF}}$	1	jakost vpliva povratne zanke od ERK <sub>pp</sub> do RAF

**Tabela 6.1:** Oznaka, vrednosti in pomen parametrov, ki opisujejo skupne koncentracije proteinov in hitrosti procesov v obravnavani signalni poti.

Sistem navadnih diferencialnih enačb, ki opisuje model, je zapisan v izrazih (6.1)–(6.23). Simbol  $\times$  v njih označuje množenje, simbol  $\cdot$  pa kompleks dveh zvrsti. Izrazi so povzeti po prilogi k [5], a smo v izrazu (6.12)  $\text{egfr}_a \cdot \text{sos}_u \times \text{ras}_{\text{GTP}}$  nadomestili z  $\text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}}$ , da smo dosegli dinamiko sistema, ki je predstavljena v članku.

$$\begin{aligned}
\frac{d}{dt} \text{egfr}_a &= \alpha_1 \times \text{EGF} \times (\text{egfr}_{\text{tot}} - (\text{egfr}_a + \text{egfr}_i \cdot \text{sos}_u + \text{egfr}_a \cdot \text{sos}_u + \\
&\quad \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\
&\quad \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}})) - \delta_1 \times \text{egfr}_a - \\
&\quad \beta_1 \times \text{egfr}_a \times (\text{sos}_{\text{tot}} - (\text{sos}_p + \text{sos}_{pp} + \text{sos}_{ppp} + \text{sos}_{pppp} + \\
&\quad \text{egfr}_i \cdot \text{sos}_u + \text{egfr}_a \cdot \text{sos}_u + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\
&\quad \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) + \nu_{1A} \times \text{egfr}_a \cdot \text{sos}_u, \tag{6.1}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{sos}_p &= -3 \times \rho_3 \times f_{\text{ERK}_{pp} \mapsto \text{SOS}_1} \times \text{erk}_{pp} \times \text{sos}_p + \\
&\quad 4 \times \rho_3 \times f_{\text{ERK}_{pp} \mapsto \text{SOS}_1} \times \text{erk}_{pp} \times (\text{sos}_{\text{tot}} - (\text{sos}_p + \text{sos}_{pp} + \text{sos}_{ppp} + \\
&\quad \text{sos}_{pppp} + \text{egfr}_i \cdot \text{sos}_u + \text{egfr}_a \cdot \text{sos}_u + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} +
\end{aligned}$$

$$\begin{aligned} & \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) - \\ & \text{sos}_p + 2 \times \text{sos}_{\text{pp}}, \end{aligned} \quad (6.2)$$

$$\begin{aligned} \frac{d}{dt} \text{sos}_{\text{pp}} &= 3 \times \rho_3 \times f_{\text{ERK}_{\text{pp}} \mapsto \text{SOS}_1} \times \text{erk}_{\text{pp}} \times \text{sos}_p + 3 \times \text{sos}_{\text{ppp}} - \\ & 2 \times \text{sos}_{\text{pp}} - 2 \times \rho_3 \times f_{\text{ERK}_{\text{pp}} \mapsto \text{SOS}_1} \times \text{erk}_{\text{pp}} \times \text{sos}_{\text{pp}}, \end{aligned} \quad (6.3)$$

$$\begin{aligned} \frac{d}{dt} \text{sos}_{\text{ppp}} &= 2 \times \rho_3 \times f_{\text{ERK}_{\text{pp}} \mapsto \text{SOS}_1} \times \text{erk}_{\text{pp}} \times \text{sos}_{\text{pp}} + 4 \times \text{sos}_{\text{pppp}} - \\ & 3 \times \text{sos}_{\text{ppp}} - \rho_3 \times f_{\text{ERK}_{\text{pp}} \mapsto \text{SOS}_1} \times \text{erk}_{\text{pp}} \times \text{sos}_{\text{ppp}}, \end{aligned} \quad (6.4)$$

$$\frac{d}{dt} \text{sos}_{\text{pppp}} = \rho_3 \times f_{\text{ERK}_{\text{pp}} \mapsto \text{SOS}_1} \times \text{erk}_{\text{pp}} \times \text{sos}_{\text{ppp}} - 4 \times \text{sos}_{\text{pppp}}, \quad (6.5)$$

$$\begin{aligned} \frac{d}{dt} \text{egfr}_i \cdot \text{sos}_u &= v_{2B} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} - \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u - \\ & v_{1B} \times \text{egfr}_i \cdot \text{sos}_u + \delta_1 \times \text{egfr}_a \cdot \text{sos}_u - \beta_{2B} \times \text{egfr}_i \cdot \text{sos}_u \times (\text{ras}_{\text{tot}} - \\ & (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \\ & \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\ & \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) + v_{2A} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \\ & \beta_{2A} \times \text{egfr}_i \cdot \text{sos}_u \times \text{ras}_{\text{GTP}}, \end{aligned} \quad (6.6)$$

$$\begin{aligned} \frac{d}{dt} \text{egfr}_a \cdot \text{sos}_u &= \beta_1 \times \text{egfr}_a \times (\text{sos}_{\text{tot}} - (\text{sos}_p + \text{sos}_{\text{pp}} + \text{sos}_{\text{ppp}} + \text{sos}_{\text{pppp}} + \\ & \text{egfr}_i \cdot \text{sos}_u + \text{egfr}_a \cdot \text{sos}_u + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\ & \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) - \\ & \beta_{2B} \times \text{egfr}_a \cdot \text{sos}_u \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \\ & \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \\ & \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) - v_{1A} \times \text{egfr}_a \cdot \text{sos}_u + \\ & v_{2A} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + v_{2B} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\ & \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u - \beta_{2A} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \\ & \delta_1 \times \text{egfr}_a \cdot \text{sos}_u, \end{aligned} \quad (6.7)$$

$$\begin{aligned} \frac{d}{dt} \text{ras}_{\text{GTP}} &= \kappa_{2B} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \\ & \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \\ & \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) + \\ & v_{2A} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \beta_{2A} \times \text{egfr}_a \cdot \text{sos}_u \times \text{ras}_{\text{GTP}} + \\ & \kappa_{2B} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} \end{aligned}$$

$$\begin{aligned}
& + \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \\
& \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) + \\
& \kappa_{2A} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \\
& \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \\
& \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) + \\
& v_{2A} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \beta_{2A} \times \text{egfr}_i \cdot \text{sos}_u \times \text{ras}_{\text{GTP}} + \\
& \kappa_{2A} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \\
& \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \\
& \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) - \beta_3 \times (\text{ras}_{\text{gap}_{\text{tot}}} - \\
& (\text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}})) \times \text{ras}_{\text{GTP}}, \quad (6.8)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} &= -v_{2B} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \delta_1 \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} - \\
& \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \beta_{2B} \times \text{egfr}_i \cdot \text{sos}_u \times (\text{ras}_{\text{tot}} - \\
& (\text{ras}_{\text{GTP}} + \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\
& \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\
& \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})), \quad (6.9)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} &= \beta_{2A} \times \text{egfr}_i \cdot \text{sos}_u \times \text{ras}_{\text{GTP}} - \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \\
& v_{2A} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \delta_1 \times \text{egfr}_a \cdot \text{sos}_u \times \text{ras}_{\text{GTP}}, \quad (6.10)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} &= \beta_{2B} \times \text{egfr}_a \cdot \text{sos}_u \times (\text{ras}_{\text{tot}} - (\text{ras}_{\text{GTP}} + \\
& \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \\
& \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}})) - \\
& \delta_1 \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} + \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}} - \\
& v_{2B} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GDP}}, \quad (6.11)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt} \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} &= \alpha_1 \times \text{EGF} \times \text{egfr}_i \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} + \beta_{2A} \times \text{egfr}_a \cdot \text{sos}_u \times \text{ras}_{\text{GTP}} - \\
& v_{2A} \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}} - \delta_1 \times \text{egfr}_a \cdot \text{sos}_u \cdot \text{ras}_{\text{GTP}}, \quad (6.12)
\end{aligned}$$

$$\frac{d}{dt} \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} = -v_3 \times \text{ras}_{\text{GDP}} \cdot \text{ras}_{\text{gap}} + \kappa_3 \times \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}}, \quad (6.13)$$

$$\begin{aligned}
\frac{d}{dt} \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} &= -\kappa_3 \times \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \beta_3 \times (\text{ras}_{\text{gap}_{\text{tot}}} - (\text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}} + \\
& \text{ras}_{\text{GTP}} \cdot \text{ras}_{\text{gap}})) \times \text{ras}_{\text{GTP}}, \quad (6.14)
\end{aligned}$$

$$\begin{aligned} \frac{d}{dt} \text{raf}_a &= -\rho_6 \times f_{\text{ERK}_{pp} \rightarrow \text{RAF}} \times \text{raf}_a - \delta_2 \times \text{raf}_a + \alpha_2 \times \text{ras}_{\text{GTP}} \times (\text{raf}_{\text{tot}} - \\ &\quad (\text{raf}_a + \text{raf}_p)), \end{aligned} \quad (6.15)$$

$$\begin{aligned} \frac{d}{dt} \text{raf}_p &= \rho_6 \times f_{\text{ERK}_{pp} \rightarrow \text{RAF}} \times \text{raf}_a + \rho_6 \times f_{\text{ERK}_{pp} \rightarrow \text{RAF}} \times \text{erk}_{pp} \times (\text{raf}_{\text{tot}} - \\ &\quad (\text{raf}_a + \text{raf}_p)) - \text{raf}_p, \end{aligned} \quad (6.16)$$

$$\begin{aligned} \frac{d}{dt} \text{mek}_{u, \text{T292p}} &= -\text{mek}_{u, \text{T292p}} + \rho_4 \times f_{\text{ERK}_{pp} \rightarrow \text{MEK1}} \times \text{erk}_{pp} \times (\text{mek}_{\text{tot}} - \\ &\quad (\text{mek}_{u, \text{T292p}} + \text{mek}_{p, \text{T292u}} + \text{mek}_{p, \text{T292p}} + \text{mek}_{pp, \text{T292u}} + \\ &\quad \text{mek}_{pp, \text{T292p}})) + \sigma_1 \times \text{mek}_{p, \text{T292p}} + \sigma_5 \times \text{mek}_{p, \text{T292p}} - \\ &\quad 2 \times \rho_1 \times \text{mek}_{p, \text{T292p}} \times \text{raf}_a, \end{aligned} \quad (6.17)$$

$$\begin{aligned} \frac{d}{dt} \text{mek}_{p, \text{T292u}} &= \text{mek}_{p, \text{T292p}} - \sigma_1 \times \text{mek}_{p, \text{T292u}} + 2 \times \sigma_1 \times \text{mek}_{pp, \text{T292u}} - \\ &\quad \sigma_1 \times \text{mek}_{p, \text{T292u}} \times \text{raf}_a + 2 \times \rho_1 \times (\text{mek}_{\text{tot}} - (\text{mek}_{u, \text{T292p}} + \\ &\quad \text{mek}_{p, \text{T292u}} + \text{mek}_{p, \text{T292p}} + \text{mek}_{pp, \text{T292u}} + \text{mek}_{pp, \text{T292p}})) \times \text{raf}_a - \\ &\quad \rho_4 \times f_{\text{ERK}_{pp} \rightarrow \text{MEK1}} \times \text{erk}_{pp} \times \text{mek}_{p, \text{T292u}}, \end{aligned} \quad (6.18)$$

$$\begin{aligned} \frac{d}{dt} \text{mek}_{p, \text{T292p}} &= \rho_4 \times f_{\text{ERK}_{pp} \rightarrow \text{MEK1}} \times \text{erk}_{pp} \times \text{mek}_{p, \text{T292u}} - \sigma_1 \times \text{mek}_{p, \text{T292p}} - \\ &\quad \rho_1 \times \text{mek}_{p, \text{T292p}} \times \text{raf}_a - \text{mek}_{p, \text{T292p}} - \sigma_5 \times \text{mek}_{p, \text{T292p}} + \\ &\quad 2 \times \sigma_5 \times \text{mek}_{pp, \text{T292p}} + 2 \times \sigma_1 \times \text{mek}_{pp, \text{T292p}} + \\ &\quad 2 \times \rho_1 \times \text{mek}_{u, \text{T292p}} \times \text{raf}_a, \end{aligned} \quad (6.19)$$

$$\begin{aligned} \frac{d}{dt} \text{mek}_{pp, \text{T292u}} &= -\rho_4 \times f_{\text{ERK}_{pp} \rightarrow \text{MEK1}} \times \text{erk}_{pp} \times \text{mek}_{pp, \text{T292u}} - \\ &\quad 2 \times \sigma_1 \times \text{mek}_{pp, \text{T292u}} + \rho_1 \times \text{mek}_{pp, \text{T292u}} \times \text{raf}_a + \\ &\quad \text{mek}_{pp, \text{T292p}}, \end{aligned} \quad (6.20)$$

$$\begin{aligned} \frac{d}{dt} \text{mek}_{pp, \text{T292p}} &= \rho_4 \times f_{\text{ERK}_{pp} \rightarrow \text{MEK1}} \times \text{erk}_{pp} \times \text{mek}_{p, \text{T292u}} + \\ &\quad \rho_1 \times \text{mek}_{pp, \text{T292p}} \times \text{raf}_a - \text{mek}_{pp, \text{T292p}} - 2 \times \sigma_1 \times \text{mek}_{pp, \text{T292p}} - \\ &\quad 2 \times \sigma_5 \times \text{mek}_{pp, \text{T292p}}, \end{aligned} \quad (6.21)$$

$$\begin{aligned} \frac{d}{dt} \text{erk}_p &= 2 \times \sigma_2 \times \text{erk}_{pp} - \sigma_2 \times \text{erk}_p + \\ &\quad 2 \times \rho_2 \times \text{mek}_{pp, \text{T292u}} \times (\text{erk}_{\text{tot}} - (\text{erk}_p + \text{erk}_{pp})) - \\ &\quad \rho_2 \times \text{mek}_{pp, \text{T292u}} \times \text{erk}_p - \rho_2 \times \text{mek}_{pp, \text{T292p}} \times \text{erk}_p + \\ &\quad 2 \times \rho_2 \times \text{mek}_{pp, \text{T292p}} \times (\text{erk}_{\text{tot}} - (\text{erk}_p + \text{erk}_{pp})), \end{aligned} \quad (6.22)$$



$$\begin{aligned} \frac{d}{dt} \text{erk}_{pp} = & \rho_2 \times \text{mek}_{pp, T292p} \times \text{erk}_p - 2 \times \sigma_2 \times \text{erk}_{pp} + \\ & \rho_2 \times \text{mek}_{pp, T292u} \times \text{erk}_p. \end{aligned} \quad (6.23)$$

## 6.3 Uporaba metode FCM za učenje ravnotežnega stanja signalne poti MAPK

### 6.3.1 Polni mehki model ravnotežnega stanja

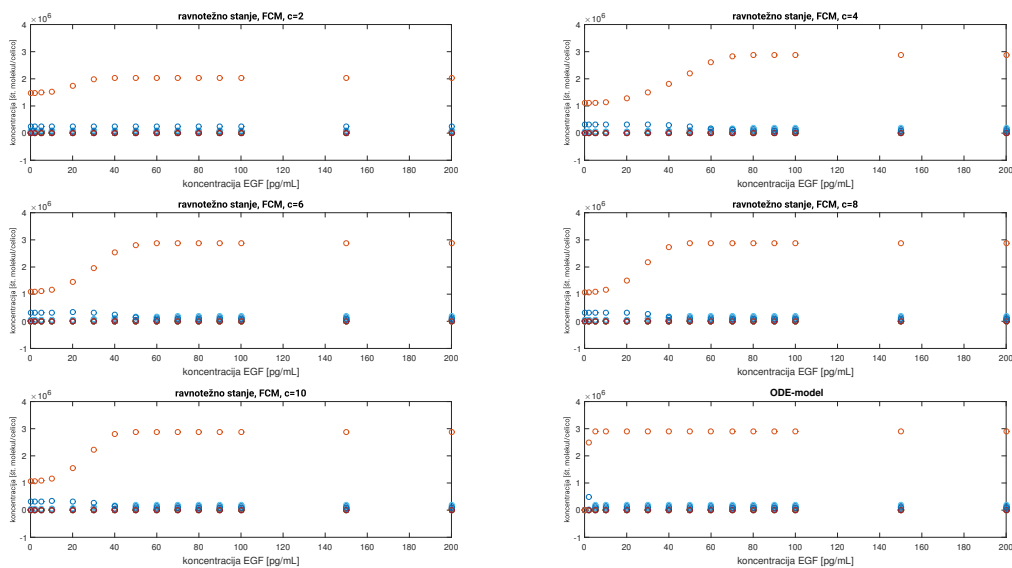
Ravnotežno stanje sistema je odvisno le od vrednosti koncentracije dražljaja EGF, ne pa tudi od njegovega začetnega stanja. Tako lahko zgradimo mehki model, ki ima le eno vhodno spremenljivko, koncentracijo dražljaja EGF, in 23 izhodnih spremenljivk, ki ustrezajo koncentracijam vseh ostalih proteinov v ravnotežnem stanju.

Za učenje mehkih pravil uporabimo algoritem FCM oziroma implementacijo učenja z njim, ki je v programskem paketu MATLAB dostopna pod imenom `genfis3`. Njena slabost je, da rezultati učenja niso vsakič enaki, zato dobimo dobro aproksimacijo naših podatkov le, če metodo poženemo večkrat in izberemo najboljšo izmed generiranih rešitev. Pri našem delu smo metodo pognali 10-krat in kot najprimernejšo rešitev izbrali tisto, ki se najboljše prilega izmerjenim koncentracijam.

Vnaprej moramo določiti tudi število skupin  $c$ , ki določa število mehkih pravil, pa tudi število mehkih vrednosti za vsako izmed spremenljivk. Ker ne obstaja univerzalen postopek, kako to število najboljše izbrati, smo učenje pognali na večjem naboru različnih vrednosti za  $c$ . Ker se želimo izogniti pretiranemu prilagajanju modela na učne podatke, izberemo najmanjšo vrednost za  $c$ , ki vrne zadovoljive rezultate.

Učenje smo izvedli pri vrednostih parametra  $c$  2, 4, 6, 8 in 10. Izkazalo se je, da metoda dobro aproksimira učne podatke pri uporabi štirih skupin. Rezultati simulacije so prikazani na sliki 6.2. Za primerjavo je prikazan tudi rezultat simulacije z referenčnim ODE-modelom.

Da lahko utemeljimo pravilnost delovanja zgrajenih modelov, jih testiramo na vrednostih koncentracije EGF 3, 8, 13, 18, 32, 64, 89, 120 in 156  $\frac{pg}{mL}$ . Rezultati simulacije in primerjava z rezultati simulacije z referenčnim ODE-modelom so prikazani na sliki 6.3. Vidimo, da metoda uspešno aproksimira ravnotežno stanje le za vrednosti koncentracije EGF, ki so večje od 20  $\frac{pg}{mL}$ . Najmanjša vrednost parametra  $c$ , pri kateri je metoda uspešna je enaka 6.



**Slika 6.2:** Rezultati simulacij z enim izmed naborov mehkih modelov ravnotežnega stanja naučenih z metodo FCM na učnih podatkih pri različnih vrednostih parametra  $c$ . Rdeča barva predstavlja koncentracijo  $ERK_{PP}$ , modra pa koncentracijo  $ERK_P$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

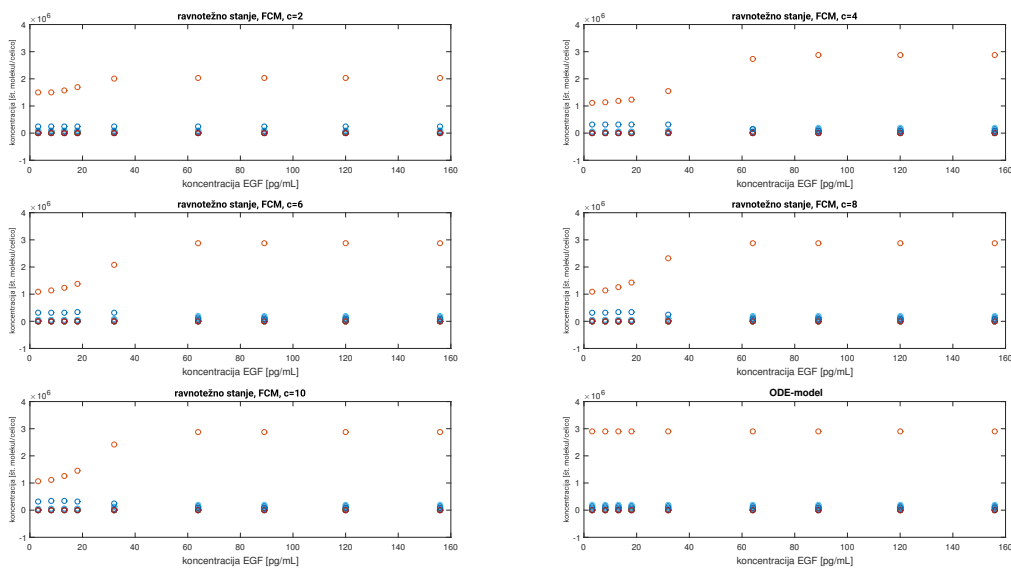
Mehki model porabi za izračun precej manj časa kot referenčni ODE-model, kar je vsekakor velika prednost. Največja ovira pristopa je, da smo morali za njegovo gradnjo aproksimirati 576 parametrov, medtem ko je imel ODE-model le 31 kinetičnih parametrov. Težava mehkega modela naučenega z metodo FCM je tudi v tem, da naučenim mehkim množicam težko pripišemo smiselne lingvistične vrednosti. Tak primer prikazujejo mehke množice (pripadnostne funkcije) vhodne spremenljivke EGF, ki so prikazane na sliki 6.4.

### 6.3.2 Reducirani mehki model ravnotežnega stanja

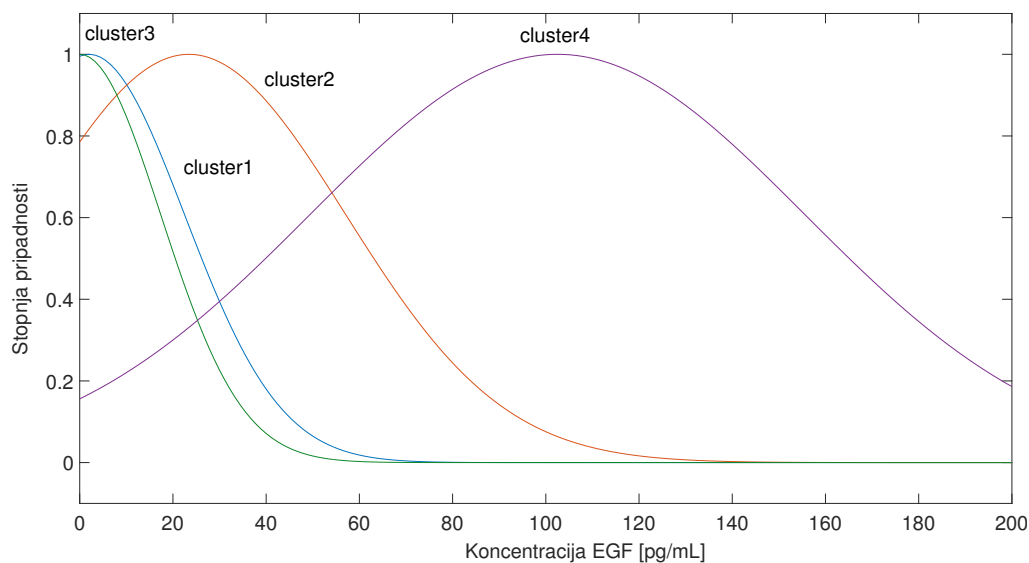
Običajno nas ne zanimajo koncentracije vseh kemijskih zvrsti, ki jih opisuje ODE-model, temveč le nekatere med njimi. Opisi signalne poti MAPK pogosto vsebujejo le tri proteine, to so Raf, MEK in ERK. Pri reduciranem mehkem modelu se bomo zato poskusili naučiti koncentracij  $Raf_a$ ,  $ERK_{PP}$  in vsote  $MEK_{PP,T292u}$  in  $MEK_{PP,T292p}$ . Učne podatke pridobimo na enak način, kot je opisano v razdelku 6.1.1.

Učenje tudi v tem primeru izvedemo za vrednosti parametra  $c$  2, 4, 6, 8 in 10. Izkáže se, da metoda dobro aproksimira učne podatke pri uporabi štirih skupin. Rezultati

### 6.3 UPORABA METODE FCM ZA UČENJE RAVNOTEŽNEGA STANJA SIGNALNE POTI MAPK 37

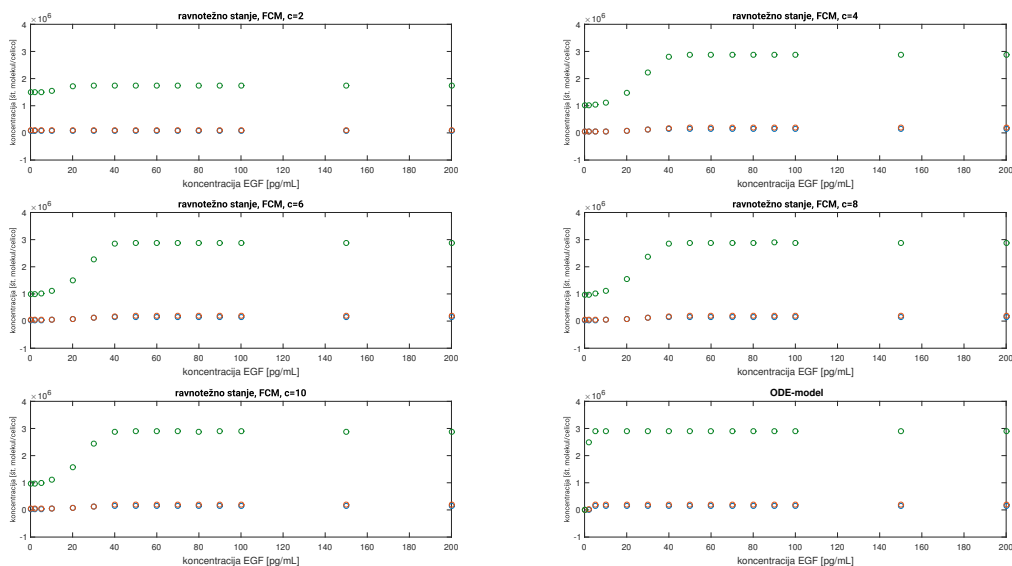


**Slika 6.3:** Rezultati simulacij z enim izmed naborov mehkih modelov naučenih z metodo FCM na testnih podatkih pri različnih vrednostih parametra  $c$ . Rdeča barva predstavlja koncentracijo ERK<sub>pp</sub>, modra pa koncentracijo ERK<sub>p</sub>. Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.



**Slika 6.4:** Pripadnostne funkcije spremenljivke EGF, ki smo se jih naučili z metodo FCM pri vrednosti parametra  $c = 4$ .

simulacije so prikazani na sliki 6.5. Za primerjavo je prikazan tudi rezultat simulacije z referenčnim ODE-modelom.

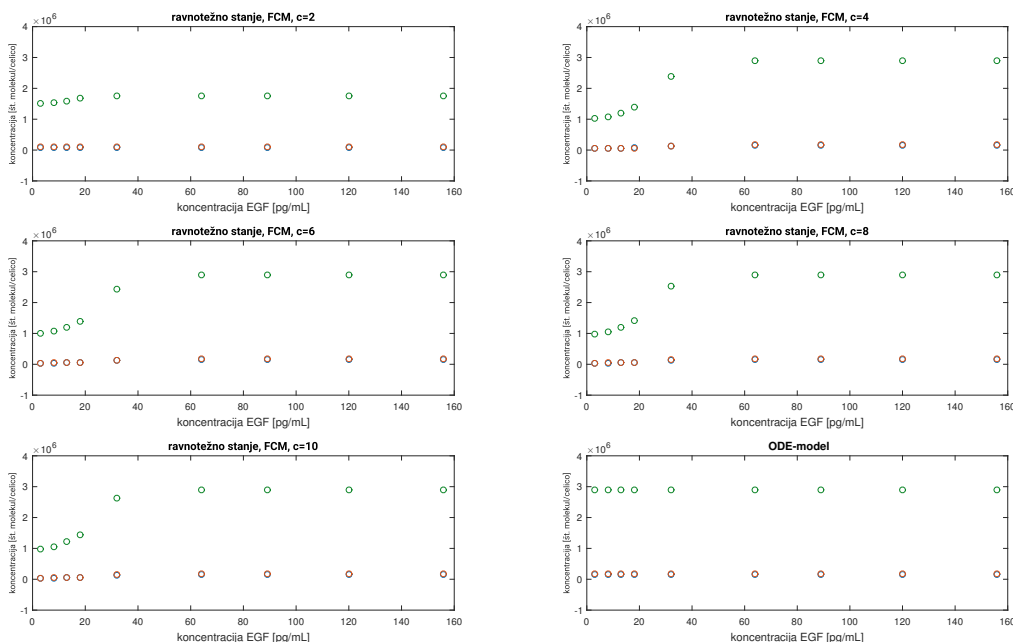


**Slika 6.5:** Rezultati simulacij z enim izmed naborov reduciranih mehkih modelov ravnotežnega stanja naučenih z metodo FCM na učnih podatkih pri različnih vrednostih parametra  $c$ . Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo  $MEK_{pp}$  in zelena koncentracijo  $ERK_{pp}$ .

Da lahko utemeljimo pravilnost delovanja zgrajenih modelov, jih testiramo na vrednostih koncentracije EGF 3, 8, 13, 18, 32, 64, 89, 120 in  $156 \frac{pg}{mL}$ . Rezultati simulacije in primerjava z rezultati simulacije z referenčnim ODE-modelom so prikazani na sliki 6.6. Vidimo, da metoda uspešno aproksimira ravnotežno stanje le za vrednosti koncentracije EGF, ki so večje od  $20 \frac{pg}{mL}$ . Najmanjša vrednost parametra  $c$ , pri kateri je metoda uspešna, je enaka 4. Prav tako kot pri osnovnem polnem modelu tudi tu ohranimo veliko število parametrov v primerjavi z ODE-modelom, saj jih ocenjujemo kar 64.

## 6.4 Uporaba metode FCM za učenje dinamičnega modela signalne poti MAPK

Pri gradnji dinamičnega modela smo ponovno uporabili funkcijo `genfis3`. Metodo smo poglani 10-krat in kot najprimernejšo rešitev izbrali tisto, ki se najbolje prilega izmerjenim koncentracijam. Na ta način bolj kaznujemo napake pri izračunih sprememb koncentracije na začetnih korakih simulacije, s čimer se krivulja, ki je rezultat simulacije, lepše



**Slika 6.6:** Rezultati simulacij z enim izmed naborov reduciranih mehkih modelov naučenih z metodo FCM na testnih podatkih pri različnih vrednostih parametra  $c$ . Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo  $MEK_{pp}$  in zelena koncentracijo  $ERK_{pp}$ .

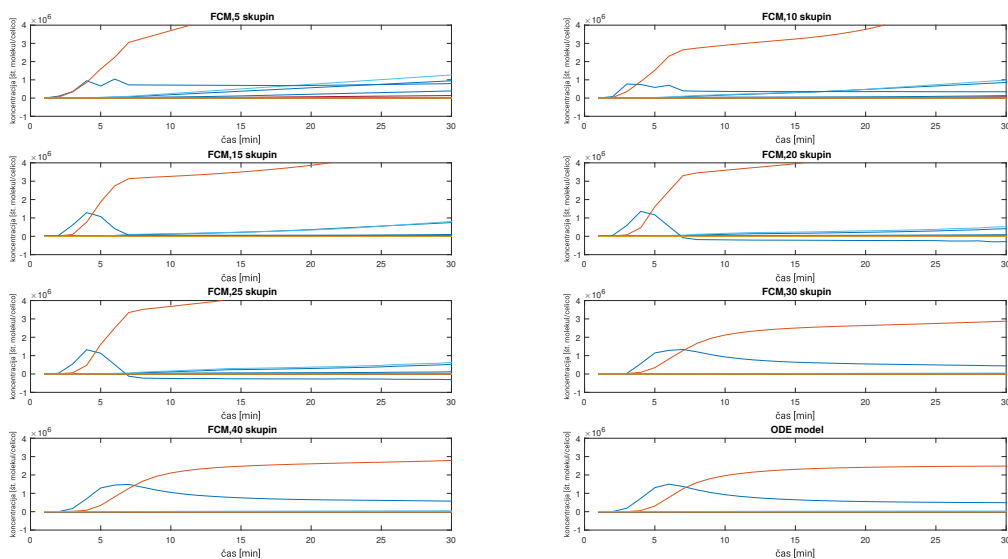
približa krivulji iz učnih podatkov.

#### 6.4.1 Polni dinamični mehki model

Učenje smo izvedli za naslednjih sedem vrednosti parametra  $c$ : 5, 10, 15, 20, 25, 30 in 40. Izkazalo se je, da metoda dobro aproksimira učne podatke šele pri uporabi 30 skupin. Na sliki 6.7 so prikazani rezultati simulacije pri konstantni vrednosti  $EGF = 5 \frac{pg}{mL}$  in začetni vrednosti vseh ostalih koncentracij 0 molekul na celico. Koncentracije enakih kemijskih vrst so na vseh grafih prikazane z enako barvo. Za primerjavo je prikazan tudi rezultat simulacije z referenčnim ODE-modelom.

Za testiranje učinkovitosti oziroma pravilnosti metode, poskusimo nekoliko spremeniti začetne pogoje in opazujemo obnašanje naučenega modela. Najprej poskusimo spremeniti le vrednost koncentracije EGF, nato pa tudi začetne koncentracije nekaterih kemijskih vrst v celici. Rezultate primerjamo z rezultati referenčnega modela.

Uporabili smo štiri različne koncentracije in sicer 3, 8, 13 in 18 pg/ml, ki so vse znotraj definicijskega območja mehke vhodne spremenljivke EGF. Na sliki 6.8 vidimo, da model



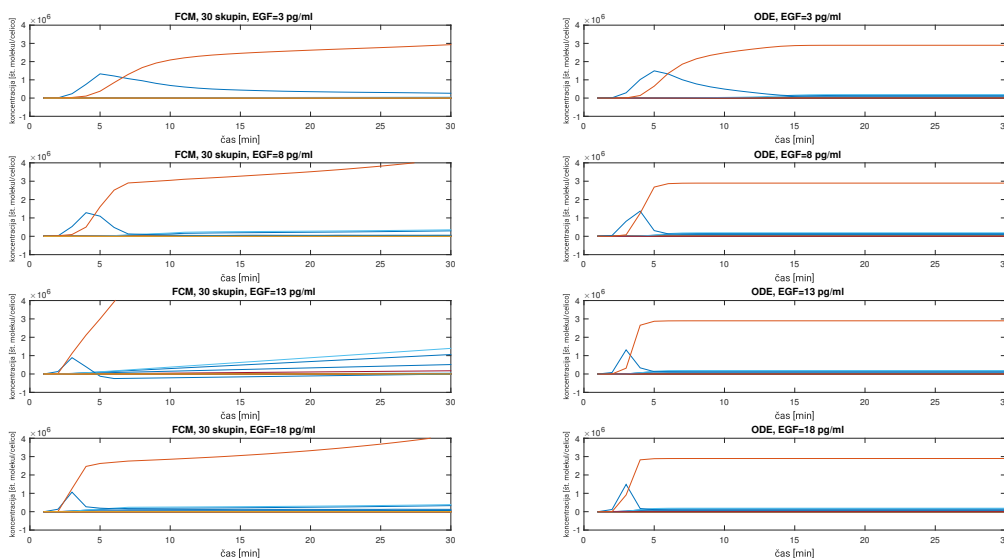
**Slika 6.7:** Rezultati simulacij z enim izmed naborov mehkih modelov naučenih z metodo FCM pri različnih vrednostih parametra  $c$  in koncentracijo  $EGF = 5 \frac{pg}{mL}$ . Rdeča barva predstavlja koncentracijo  $ERK_{pp}$ , modra pa koncentracijo  $ERK_p$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

ne opisuje več dobro opazovanega sistema, če vrednosti koncentracije EGF spremenimo, saj koncentracija  $erk_{pp}$  pretirano naraste pri koncentracijah EGF, ki so večje ali enake  $8 \frac{pg}{mL}$ .

Začetne pogoje smo nato nastavili na naključne vrednosti in pri istih začetnih pogojih pognali simulacije za vse štiri vrednosti koncentracije EGF, pri katerih smo se učili modela. Rezultati simulacije so prikazani na sliki 6.9. Vidimo, da tudi pri tovrstnih spremembah model ne opisuje dobro opazovanega sistema, saj koncentracija  $erk_{pp}$  pretirano naraste pri vseh koncentracijah EGF.

Prav tako kot pri 6.3 tudi v tem primeru naučenim mehkim vrednostim za spremenljivke težko pripišemo smiselna lingvistična imena. Kot primer so na sliki 6.10 prikazane pripadnostne funkcije spremenljivke EGF, ki smo se jih naučili pri parametru  $c = 5$ . Zaradi te lastnosti tudi mehka pravila nimajo biološko relevantne razlage.

Tretji problem je, da z uporabo tega mehkega modela ne zmanjšamo števila parametrov. Za 23 spremenljivk in 30 pravil namreč uvedemo 1380 parametrov, ki določajo gaussove pripadnostne funkcije, medtem ko je ODE-model zahteval le 31 kinetičnih parametrov.



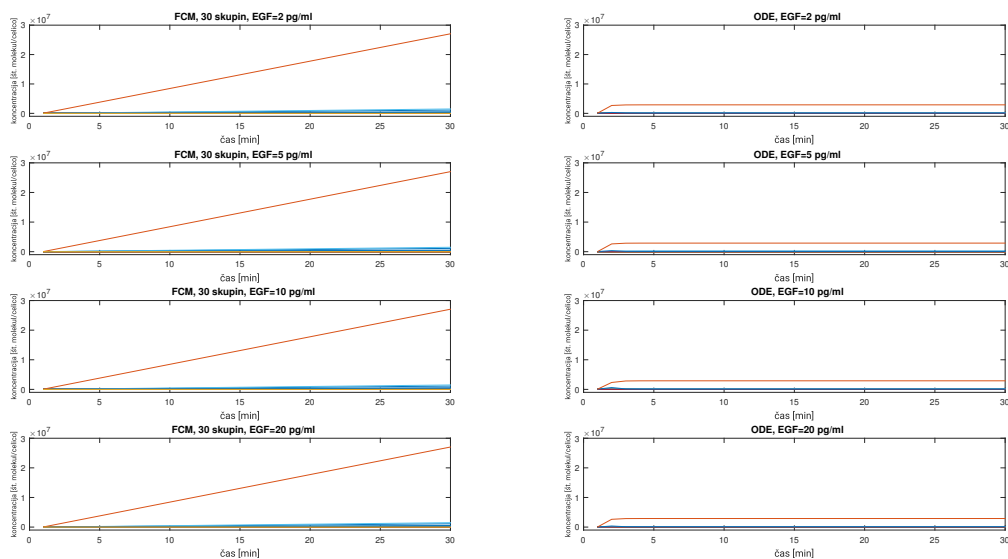
**Slika 6.8:** Rezultati simulacij mehkega modela naučenega z metodo FCM pri vrednosti parametra  $c = 30$  pri različnih vrednostih koncentracije EGF. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Rdeča barva predstavlja koncentracijo  $ERK_{pp}$ , modra pa koncentracijo  $ERK_p$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

#### 6.4.2 Reducirani dinamični mehki model

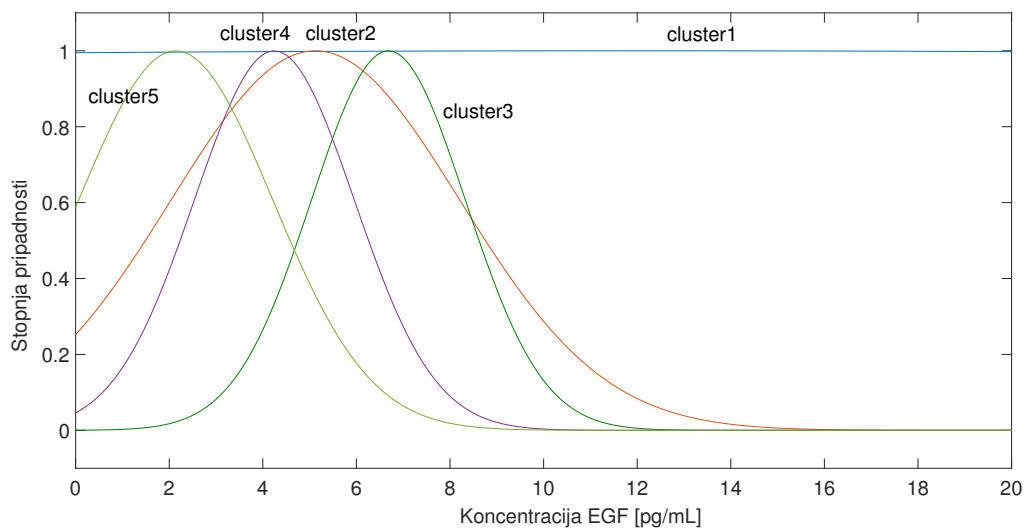
Če želimo, da ima učenje mehkega modela uporabno vrednost za eksperimentaliste, ne moremo pričakovati, da bodo lahko v laboratoriju izmerili vrednosti koncentracij za vseh 23 kemijskih zvrsti, ki nastopajo v ODE-modelu. Zato želimo v tem razdelku preveriti, ali lahko dobro aproksimiramo odziv celice s poznavanjem zgolj dela teh podatkov.

Opisi signalne poti MAPK pogosto vsebujejo le tri proteine, to so Raf, MEK in ERK. Pri reduciranem mehkem modelu se bomo zato poskusili naučiti koncentracij  $Raf_a$ ,  $ERK_{pp}$  in vsote  $MEK_{pp,T292u}$  in  $MEK_{pp,T292p}$ . Učne podatke pridobimo na enak način, kot je opisano v razdelku 6.1.1.

Mehki model zgradimo s pomočjo metode FCM pri različnih vrednostih parametra  $c$ . Ponovno želimo minimizirati njegovo vrednost, pri čemer želimo imeti še vseeno dovolj dober matematični model. Rezultati učenja so prikazani na sliki 6.11. Vidimo, da tudi v tem primeru ne moremo zmanjšati števila parametrov pod 31, kolikor bi jih morali oceniti za gradnjo ODE-modela, saj je najmanjše število skupin, ki približno dobro opišejo obnašanje opazovanega sistema 20, posledično pa opazujemo vsaj 200 različnih



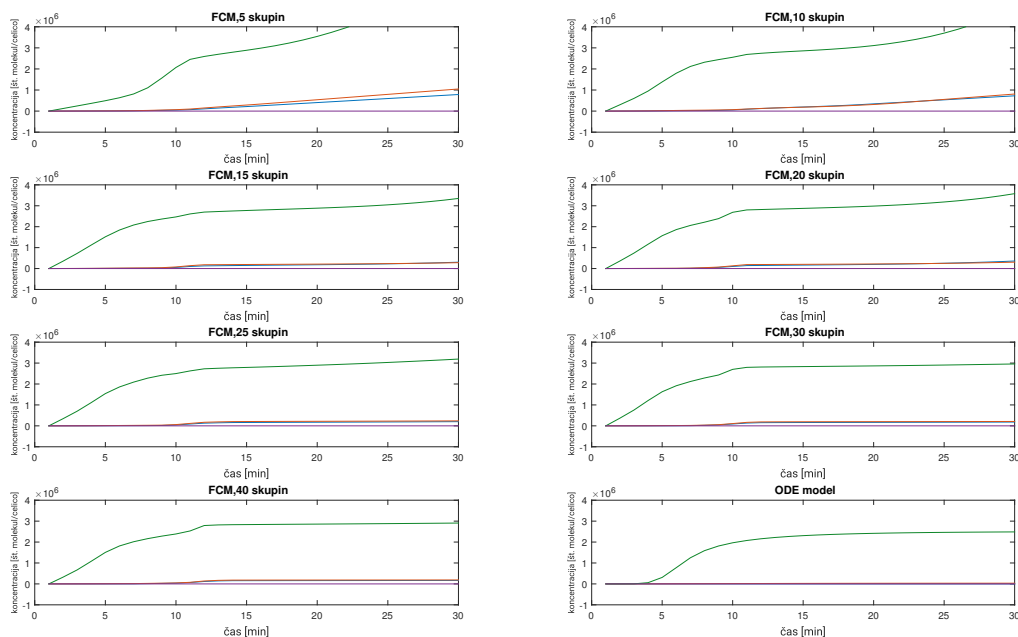
**Slika 6.9:** Rezultati simulacij mehkega modela naučenega z metodo FCM pri vrednosti parametra  $c = 30$  pri naključnih začetnih pogojih in začetnih koncentracijah EGF, ki smo jih uporabili za učenje modela. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Rdeča barva predstavlja koncentracijo  $ERK_{pp}$ , modra pa koncentracijo  $ERK_p$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.



**Slika 6.10:** Pripadnostne funkcije spremenljivke EGF, ki smo se jih naučili z metodo FCM pri vrednosti parametra  $c = 5$ .



parametrov.

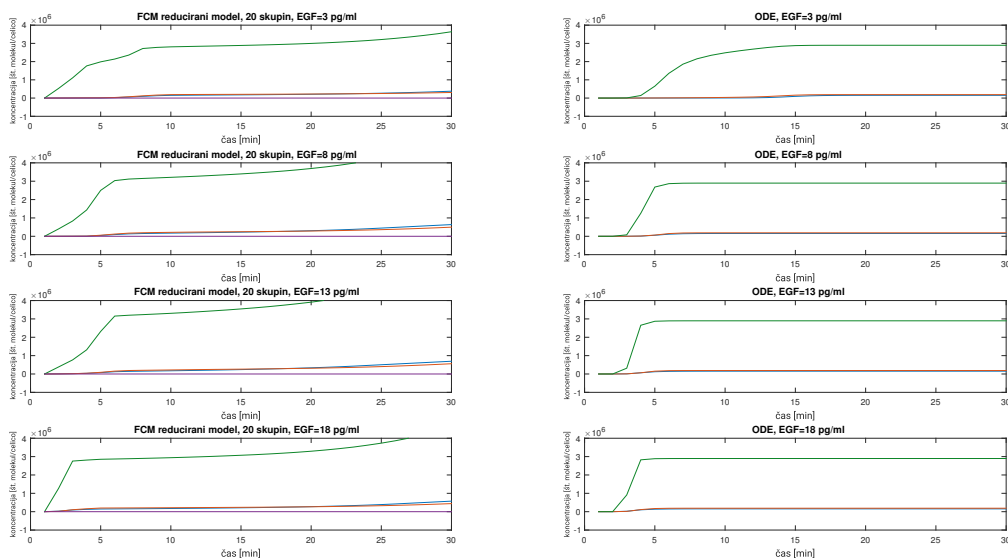


**Slika 6.11:** Rezultati simulacij z enim izmed naborov mehkih reduciranih modelov naučenih z metodo FCM pri različnih vrednostih parametra  $c$ . Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo  $MEK_{pp}$  in zelena koncentracijo  $ERK_{pp}$ .

Vseeno poskusimo narediti še občutljivostno analizo, ki smo jo izvedli tudi pri prejšnjem modelu. Najprej poskusimo spremeniti le vrednost koncentracije EGF, nato pa tudi začetne koncentracije nekaterih kemijskih zvrsti v celici. Rezultate primerjamo z rezultati referenčnega modela.

Uporabili smo štiri različne koncentracije in sicer 3, 8, 13 in 18  $\text{pg/ml}$ , ki so vse znotraj definicijskega območja mehke vhodne spremenljivke EGF. Na sliki 6.12 vidimo, da model zelo slabo opisuje dinamiko opazovanega sistema, če vrednosti koncentracije EGF spremenimo, saj koncentracija  $erk_{pp}$  pretirano naraste pri koncentracijah EGF, ki so višje od 8  $\frac{\text{pg}}{\text{mL}}$ .

Začetne pogoje smo nato nastavili na naključne vrednosti in pri istih začetnih pogojih pognali simulacije za vse štiri vrednosti koncentracije EGF, pri katerih smo se učili modela. Rezultati simulacije so prikazani na sliki 6.13. Vidimo, da tudi pri tovrstnih spremembah model opazovanega sistema ne opisuje dobro, kar je najbolj izrazito pri koncentraciji  $EGF = 10 \frac{\text{pg}}{\text{mL}}$ . Poleg tega se tudi v tem primeru pojavijo vse že prej omenjene



**Slika 6.12:** Rezultati simulacij z enim izmed naborov mehkih reduciranih modelov naučenih z metodo FCM pri vrednosti parametra  $c = 20$  pri različnih vrednostih koncentracije EGF. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo MEK<sub>pp</sub> in zelena koncentracijo ERK<sub>pp</sub>.

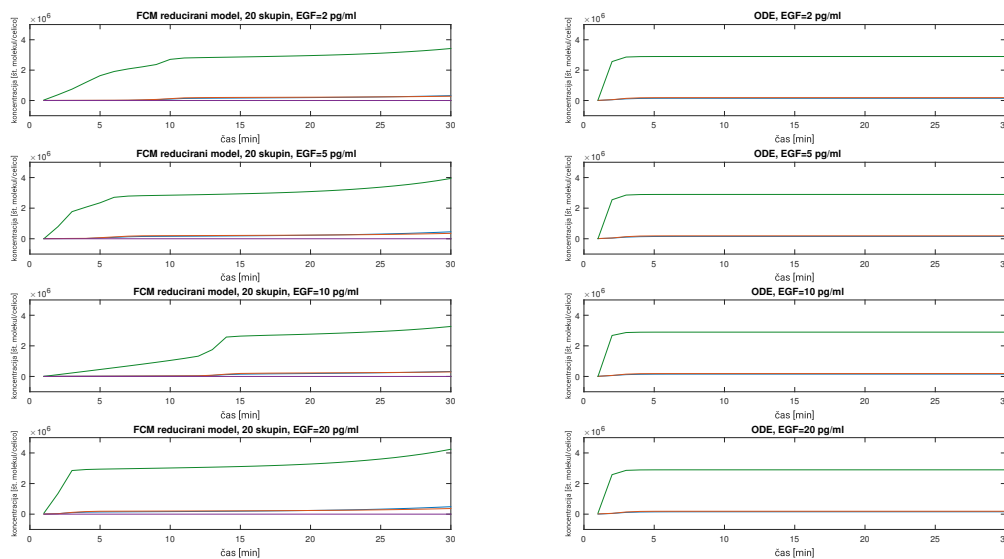
slabosti.

## 6.5 Uporaba metode MAFTS za učenje dinamičnega modela signalne poti MAPK

Metoda za gručenje FCM je v programskem paketu MATLAB implementirana pod imenom `fcm`. Uporabimo jo za gručenje koncentracij posameznih proteinov. Ker gručimo učne podatke za vsak protein posebej, lahko vrednostim pripišemo tudi lingvistična imena. Vnaprej določimo, da bomo vsaki spremenljivki določili po 5 mehkih vrednosti, zato jim dodelimo pripadajoča lingvistična imena *zelo nizka*, *nizka*, *srednja*, *visoka* in *zelo visoka*.

Podobno kot funkcija `genfis3` tudi `fcm` ne vrača enoličnih rezultatov. Ovira, ki se pojavi, je, da so rezultati včasih numeričnega tipa NaN (ni število, angl. *not a number*), posledično pa je zgrajeni model neuporaben, saj z njim ne moremo računati. Težavo smo razrešili tako, da smo učenje ponavljali dokler nismo dobili veljavnega numeričnega rezultata za ocenjevanje centrov skupin.

## 6.5 UPORABA METODE MAFTS ZA UČENJE DINAMIČNEGA MODELA SIGNALNE POTI MAPK45



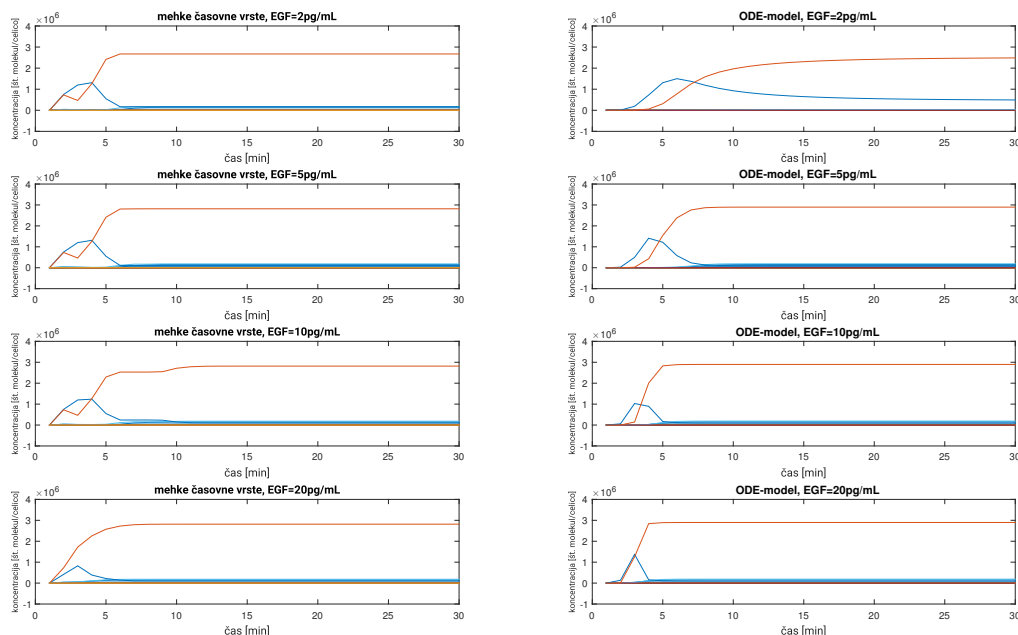
**Slika 6.13:** Rezultati simulacij reduciranega mehkega modela naučenega z metodo FCM pri vrednosti parametra  $c = 20$  pri naključnih začetnih pogojih in začetnih koncentracijah EGF, ki smo jih uporabili za učenje modela. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo  $MEK_{pp}$  in zelena koncentracijo  $ERK_{pp}$ .

### 6.5.1 Polni dinamični mehki model

Po učenju srednjih vrednosti mehkih pripadnostnih funkcij z metodo FCM vsaki dodelimo standardni odklon, ki je enak 3,5 % definicijskega območja mehke spremenljivke (glej tabelo 6.2). S tem zmanjšamo število parametrov, ki se jih moramo naučiti. Za celoten model tako ocenjujemo le še 230 parametrov, saj za vsako mehko spremenljivko določimo, da je spodnja meja njenega definicijskega območja enaka 0, torej ocenjujemo le še srednjo vrednost in zgornjo mejo definicijskega območja. Zgornjo mejo določimo kot maksimalno vrednost, ki jo doseže med učnimi podatki.

Iz prehodov med sosednjimi stanji določimo mehka pravila in preverimo učinkovitost učenja. Prilagajanje učnim podatkom prikazuje slika 6.14. Čeprav prilagajanje ni najboljše, model še vseeno opiše glavne lastnosti sistema. Koncentracija  $ERK_{pp}$  (označena z rdečo) sčasoma narašča, koncentracija  $ERK_{pp}$  (označena z modro) pa na začetku narašče in nato pade, hitrost padanja pa narašča s povečevanjem količine EGF. Ostale koncentracije so ves čas blizu nič.

Poskusimo narediti še občutljivostno analizo, kot smo jo predstavili v razdelku 6.1.2.



**Slika 6.14:** Rezultati simulacij z enim izmed mehkih modelov naučenih z metodo večatributne mehke časovne vrste pri koncentracijah EGF, ki smo jih uporabili za učenje modela. Rdeča barva predstavlja koncentracijo  $ERK_{pp}$ , modra pa koncentracijo  $ERK_p$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

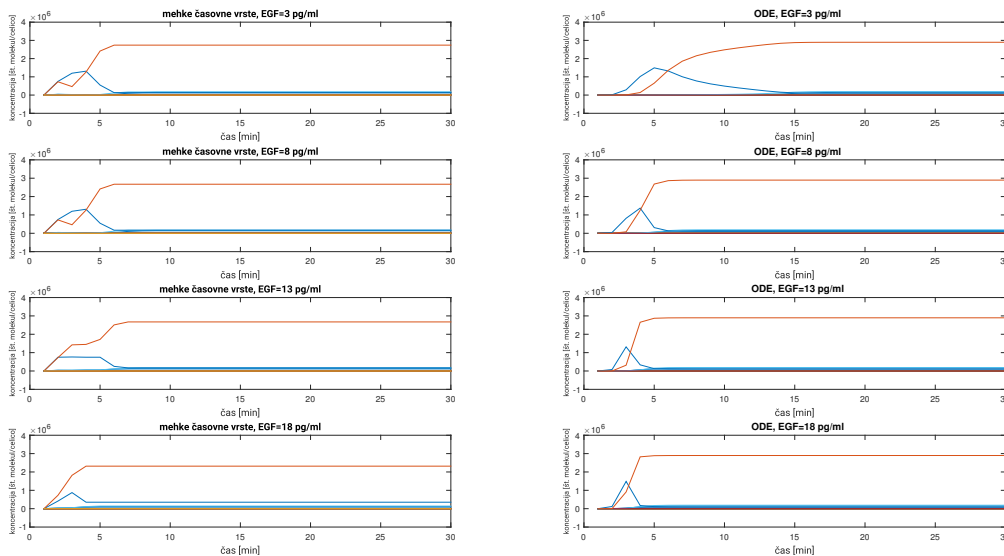
Najprej poskusimo spremeniti le vrednost koncentracije EGF, nato pa tudi začetne koncentracije nekaterih kemijskih zvrsti v celici. Rezultate primerjamo z rezultati referenčnega modela.

Uporabimo štiri različne koncentracije in sicer 3, 8, 13 in 18  $\frac{pg}{mL}$ , ki so vse znotraj definicijskega območja mehke vhodne spremenljivke EGF. Na sliki 6.15 vidimo, da model razmeroma dobro opisuje dinamiko opazovanega sistema, če vrednosti koncentracije EGF spremenimo, saj je odstopanje od modela podobno kot pri učnih podatkih.

Začetne pogoje nato nastavimo na naključne vrednosti in pri istih začetnih pogojih poženemo simulacije za vse štiri vrednosti koncentracije EGF, pri katerih smo se učili modela. Rezultati simulacije so prikazani na sliki 6.16. Vidimo, da tudi pri tovrstnih spremembah model opiše ključne lastnosti opazovanega sistema. Odstopanje od modela je podobno kot pri učnih podatkih.

Za konec utemeljimo še, da je bila izbira standardnega odklona pripadnostnih funkcij primerna. V tabeli 6.2 vidimo, da se model najboljše nauči podatkov pri standardnem

## 6.5 UPORABA METODE MAFTS ZA UČENJE DINAMIČNEGA MODELA SIGNALNE POTI MAPK47



**Slika 6.15:** Rezultati simulacij z enim izmed mehkih modelov naučenih z metodo MAFTS pri različnih vrednostih koncentracije EGF. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Rdeča barva predstavlja koncentracijo ERK<sub>pp</sub>, modra pa koncentracijo ERK<sub>p</sub>. Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

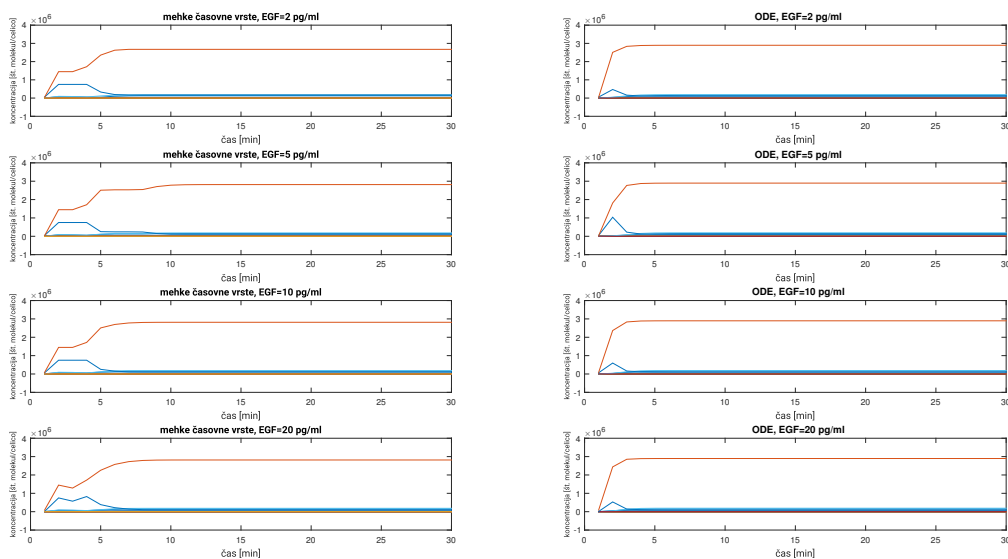
odlonu okrog 3 % definicijskega območja. Izboljšanje bi morda lahko dosegli, če bi dovolili, da imajo različne pripadnostne funkcije različne vrednosti standardnih odklonov.

### 6.5.2 Reducirani dinamični mehki model

Metodo MAFTS, ki se je izkazala za uspešno, poskusimo sedaj uporabiti na manjšem naboru kemijskih zvrsti na enak način, kot smo to naredili z metodo FCM. Razen zmanjšanja količine spremenljivk ohranimo vse lastnosti, ki smo jih opisali v prejšnjem razdelku. Rezultati simulacije so prikazani na sliki 6.17. Vidimo, da pride do velikega odstopanja pri visoki koncentraciji dražljaja EGF.

Vseeno poskusimo narediti še analizo na testni množici. Najprej poskusimo spremeniti le vrednost koncentracije EGF, nato pa tudi začetne koncentracije nekaterih kemijskih zvrsti v celici. Rezultate primerjamo z rezultati referenčnega modela.

Uporabimo štiri različne koncentracije in sicer 3, 8, 13 in 18 pg/ml, ki so vse znotraj definicijskega območja mehke vhodne spremenljivke EGF. Na sliki 6.18 vidimo, da model zelo slabo opisuje dinamiko opazovanega sistema, če vrednosti koncentracije EGF



**Slika 6.16:** Rezultati simulacij z enim izmed mehkih modelov naučenih z metodo MAFTS pri naključnih začetnih pogojih in koncentracijah EGF, ki smo jih uporabili za učenje modela. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Rdeča barva predstavlja koncentracijo  $ERK_{pp}$ , modra pa koncentracijo  $ERK_p$ . Koncentracije vseh ostalih 21 kemijskih zvrsti so zgoščene okoli vrednosti 0.

spremenimo.

Začetne pogoje nato nastavimo na naključne vrednosti in pri istih začetnih pogojih poženemo simulacije za vse štiri vrednosti koncentracije EGF, pri katerih smo se učili modela. Rezultati simulacije so prikazani na sliki 6.19. Vidimo, da pri tovrstnih spremembah model opazovanega sistema ne opisuje dobro.

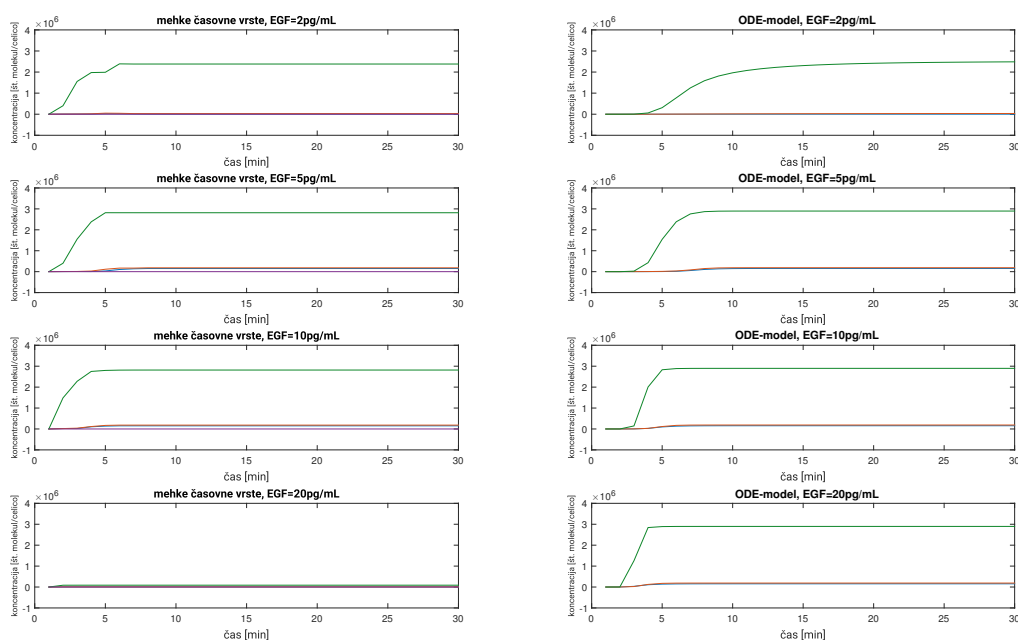
## 6.6 Primerjava pristopov za gradnjo mehkega modela

Uporabili smo dva različna pristopa za gradnjo mehkega modela signalne poti MAPK. Lastnosti obeh pristopov so povzete v tabeli 6.3.

Ugotovimo lahko, da je metoda FCM primerna za gradnjo modela ravnotežnega stanja, saj je natančna in za svojo gradnjo zahteva majhno količino eksperimentalnih podatkov. Prav tako omogoča gradnjo modela z relativno malo kinetičnimi parametri, kar je eden izmed naših ciljev. Med metodama za gradnjo dinamičnega modela je primernejša metoda MAFTS, saj je natančnejša in manj občutljiva na spremembo začetnih pogojev sistema. Zahteva manj podatkov za natančno gradnjo od metode FCM, posledično pa

standardni odklon (% definicijskega območja)	napaka ( $\times 10^7$ )
1	3,24
2	1,12
3	1,08
4	1,09
5	1,10
10	1,16
15	1,86
20	2,09

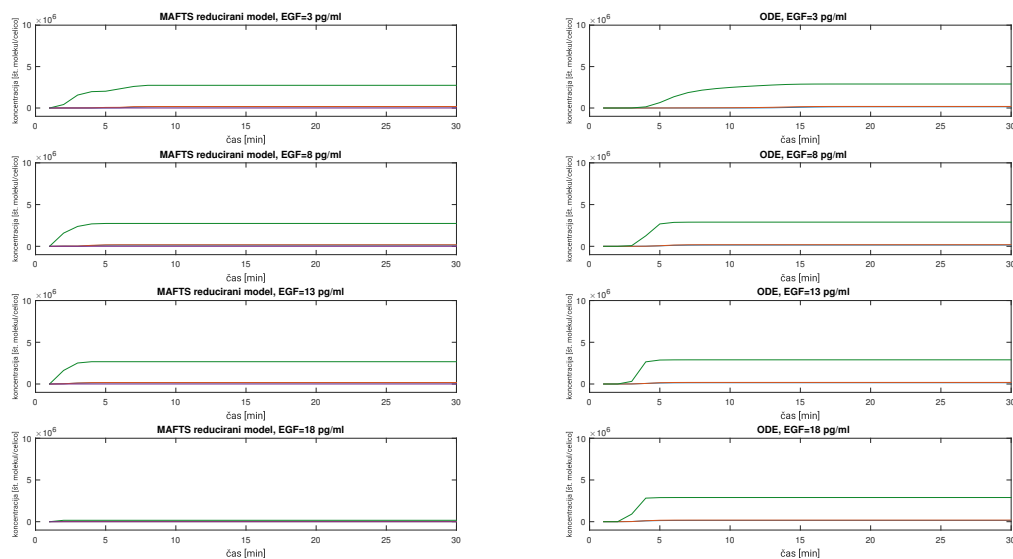
**Tabela 6.2:** Vrednost napake napovedi mehkega modela zgrajenega z MAFTS na testni množici pri uporabi različnih vrednosti standardnih odklonov.



**Slika 6.17:** Rezultati simulacij z enim izmed reduciranih mehkih modelov naučenih z metodo večatributne mehke časovne vrste pri koncentracijah EGF, ki smo jih uporabili za učenje modela. Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo  $MEK_{pp}$  in zelena koncentracijo  $ERK_{pp}$ .

ima zgrajeni model tudi manjšo količino kinetičnih parametrov.

Dodatna prednost metode MAFTS je, da znamo mehkim vrednostim avtomatsko

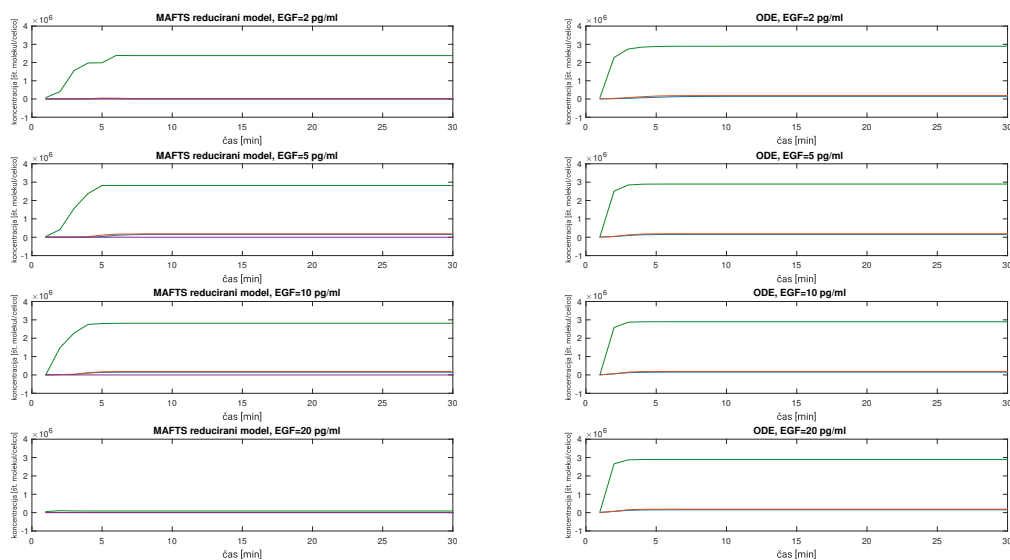


**Slika 6.18:** Rezultati simulacij z enim izmed reduciranih mehkih modelov naučenih z metodo MAFTS pri različnih vrednostih koncentracije EGF. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo MEK<sub>pp</sub> in zelena koncentracijo ERK<sub>pp</sub>.

pripisati lingvistične opise. To odpira možnost za ročno popravljanje zgrajenega modela na podlagi poznavanja lastnosti signalne poti. Ločeno generiranje mehkih množic za vsako izmed mehkih spremenljivk omogoča tudi popravljanje modela po tem, ko imamo na voljo dodatne podatke, saj lahko vključimo v ponovno učenje le nove podatke, medtem ko moramo pri FCM celoten postopek učenja pognati znova.

Kljub vsem naštetim lastnostim zaenkrat nobena od metod še ne omogoča opisa dinamike, ki bi bil enako kvaliteten kot ODE-modeli. Prednost mehkih metod napram ODE-modelom je v njihovi zmožnosti učenja topologije. ODE-model lahko namreč vzpostavimo zgolj ob poznavanju topologije biološkega sistema, medtem ko slednja za gradnjo mehkih modelov ni potrebna. Prav tako ODE-modela ne moremo uspešno reducirati, model pa je praviloma dovolj občutljiv na vrednosti kinetičnih parametrov, da ga je skoraj nemogoče uspešno ročno popravljati po tem, ko smo avtomatsko ocenili vrednosti parametrov. Nobena izmed metod ne omogoča avtomatske vzpostavitve reduciranega modela.





**Slika 6.19:** Rezultati simulacij z enim izmed reduciranih mehkih modelov naučenih z metodo MAFTS pri naključnih začetnih pogojih in koncentracijah EGF, ki smo jih uporabili za učenje modela. V desnem stolpcu so prikazani rezultati simulacije z referenčnim ODE-modelom pri enakih začetnih pogojih. Modra barva predstavlja koncentracijo Raf, rdeča koncentracijo MEK<sub>PP</sub> in zelena koncentracijo ERK<sub>PP</sub>.

lastnost	FCM - ravnotežje	FCM - dinamika	MAFTS	ODE
natančnost	velika	majhna	velika	velika
občutljivost na začetne pogoje	/	velika	majhna	majhna
količina potrebnih podatkov	majhna	velika	srednja	velika
količina parametrov	majhna	velika	srednja	majhna
možnost ročnega spreminjanja	ne	ne	da	ne
inkorporacija novih podatkov	zahtevna	zahtevna	enostavna	enostavna
poznavanje topologije omrežja	ne	ne	ne	da

**Tabela 6.3:** Primerjava pristopov za gradnjo mehkih modelov, ki jih obravnavamo v delu, in ODE-modelov.



# 7 Zaključek

V delu smo obravnavali metode za učenje mehkih modelov signalnih poti iz eksperimentalnih podatkov. Metode smo opisali, ovrednotili ovire njihovih implementacij in jih aplicirali na signalno pot MAPK.

Učenje z vsakim izmed opisanih pristopov smo pognali pri večjem spektru pogojev, izbrali najprimernejši zgrajeni model in preverili njegovo verodostojnost. S tem smo želeli utemeljiti, ali je učna metoda primerna za uporabo v eksperimentalnih pogojih, kjer je namen modeliranja zmanjšanje stroškov biokemijskih poskusov.

Izpostavili smo prednosti mehkih modelov pred konvencionalnimi metodami modeliranja. Ugotovili smo, da za gradnjo mehkih modelov poznavanje topologije signalnega omrežja ni potrebno, medtem ko gradnja ODE-modelov zahteva podrobno poznavanje gradnikov biološkega sistema in interakcij med njimi. ODE-modele praviloma gradimo ročno na osnovi poznavanja biokemijskih relacij med kemijskimi zvrstmi, medtem ko lahko mehke modele gradimo tako ročno kot neposredno iz zbranih biokemijskih podatkov. Pri ročni gradnji mehkih modelov za razliko od gradnje ODE-modelov poznavanje mehanizmov reakcij ni zahtevano, potrebno je le poznavanje interakcij med kemijskimi

zvrstmi. To lastnost si mehki modeli delijo z drugimi pristopi logičnega modeliranja, med katere sodijo tudi Boolove mreže. Druga značilnost logičnega modeliranja je preprostost in razumljivost opisov biološkega sistema, saj so pretežno lingvistični, medtem ko so modeli na osnovi ODE zelo kompleksni, saj vsebujejo komplicirane matematične izraze. Boolove mreže predpostavljajo, da lahko spremenljivke zavzamejo le diskretne vrednosti, zato vzpostavljeni modeli niso realistični. Mehki modeli v tem kontekstu predstavljajo izboljšavo Boolovih mrež, saj lahko opišejo tudi vmesna stanja spremenljivk. Čeprav modeli na osnovi Boolovih mrež ne vsebujejo parametrov, ODE-modeli pa jih vsebujejo precej manj od mehkih modelov, ki enako dobro opisujejo časovno dinamiko biološkega sistema, pa so mehki modeli med vsemi tremi pristopi edini, ki lahko za svojo gradnjo neposredno uporabijo tako kvalitativne kot tudi kvantitativne eksperimentalne podatke. Za gradnjo ODE-modelov oziroma ocenjevanje njihovih parametrov lahko namreč uporabimo le kvantitativne podatke, Boolove mreže pa lahko vgradijo le kvalitativne podatke.

## 7.1 Prispevki k znanosti

Utemeljili smo, da izgradnja mehkega modela kompleksnih signalnih poti še zdaleč ni trivialna, saj so pristopi zelo dovzetni za pretirano prilagajanje modelov učnim podatkom. Dodatna ovira je potreba po veliki količini eksperimentalnih podatkov, ki nam praviloma niso na voljo v taki meri.

Ugotovili smo, da je učenje krivulj z metodo večatributne mehke časovne vrste bolj uspešno od učenja sprememb koncentracij z metodo FCM. Pri obeh metodah je težavno učenje reduciranih modelov, kar zmanjšuje uporabnost metodologije za načrtovanje laboratorijskih poskusov.

## 7.2 Nadaljnje delo

Pričujoče delo odpira številne možnosti za nadaljnje raziskovalno delo. Omejili smo se na specifične lastnosti mehkih modelov, analizirati pa bi bilo možno tudi vpliv izbire metod mehčanja, implikacije, oblike pripadnostnih funkcij in drugih parametrov na uspešnost učenja z opisanimi metodami. S tem pristopom bi lahko omejili in opredelili pogoje, pod katerimi metode obetajo zadovoljive rezultate učenja.

Smiselno bi se bilo osredotočiti tudi na redukcijo modelov, saj imajo reducirani modeli uporabno vrednost za eksperimentaliste. Ugotoviti bi se dalo, kako močno se da model

reducirati in pod kakšnimi pogoji ter katere metode so za to najprimernejše. Možna sta seveda dva pristopa, pri čemer prvi temelji na odstranjevanju spremenljivk iz večjega modela, drugi pa je podoben, kot smo ga uporabili v našem delu, in predpostavlja gradnjo modela iz manjše količine podatkov.

Nadalje bi bilo mogoče razviti robustnejšo mehko metodo od obstoječih. Pri tem bi bilo smiselno upoštevati topološke predpostavke oziroma poznavanje biološkega sistema, s čimer bi omejili prostor rešitev in bolje usmerjali učenje mehkega modela.

Poleg kombinacije ročne in avtomatske vzpostavitve modela bi bil pomemben prispevek tudi ročno popravljanje avtomatsko zgrajenih modelov. Da je to mogoče, morajo imeti naučeni modeli določene lastnosti, med katere sodijo biološko relevantne vrednosti mehkih spremenljivk ter enostavna mehka pravila. Preden bo ročno popravljanje mehkih modelov možno, moramo tudi izboljšati metodologijo, ki nam zagotavlja gradnjo mehkega modela z želenimi lastnostmi. Čeprav že metoda MAFTS omogoča ročno popravljanje modela, ki ga je zgradila, so namreč mehka pravila, ki jih izgradi precej kompleksna, saj vsako pravilo vsebuje vse vhodne spremenljivke, čeprav to morda ni nujno potrebno.

Zanimiva bi bila tudi kombinacija mehkih modelov s konvencionalnimi metodami modeliranja. Možna bi bila uporaba mehke inference za segmente signalne poti, katerih topologije ne poznamo, preostale segmente sistema pa bi lahko opisovali z natančnejšimi metodami, na primer s sistemom diferencialnih enačb.



## LITERATURA

- [1] M. Calder, A. Duguid, S. Gilmore, J. Hillston, Stronger computational modelling of signalling pathways using both continuous and discrete-state methods, in: International Conference on Computational Methods in Systems Biology, Springer, 2006, pp. 63–77.
- [2] Z. Huang, J. Hahn, Fuzzy modeling of signal transduction networks, *Chemical Engineering Science* 64 (9) (2009) 2044–2056.
- [3] M. K. Morris, J. Saez-Rodriguez, D. C. Clarke, P. K. Sorger, D. A. Lauffenburger, Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli, *PLoS computational biology* 7 (3) (2011) e1001099.
- [4] M. Bernardo-Faura, S. Massen, C. S. Falk, N. R. Brady, R. Eils, Data-derived modeling characterizes plasticity of MAPK signaling in melanoma, *PLoS computational biology* 10 (9) (2014) e1003795.
- [5] M. Kočańczyk, P. Kocieniewski, E. Kozłowska, J. Jaruszewicz-Błońska, B. Sparta, M. Pargett, J. G. Albeck, W. S. Hlavacek, T. Lipniacki, Relaxation oscillations and hierarchy of feedbacks in MAPK signaling, *Scientific reports* 7 (2017) 38244.
- [6] A. Levchenko, J. Bruck, P. W. Sternberg, Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties, *Proceedings of the National Academy of Sciences* 97 (11) (2000) 5818–5823.
- [7] Y. Kamioka, S. Yasuda, Y. Fujita, K. Aoki, M. Matsuda, Multiple decisive phosphorylation sites for the negative feedback regulation of SOS1 via ERK, *The Journal of biological chemistry* 285 (2010) 33540–33548.

- [8] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, G. Müller, Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, *Nature biotechnology* 20 (4) (2002) 370–375.
- [9] U. S. Bhalla, Signaling in small subcellular volumes. I. Stochastic and diffusion effects on individual pathways, *Biophysical journal* 87 (2) (2004) 733–744.
- [10] S. Yamada, T. Taketomi, A. Yoshimura, Model analysis of difference between EGF pathway and FGF pathway, *Biochemical and biophysical research communications* 314 (4) (2004) 1113–1120.
- [11] M. Bizzarri, A. Cucina, A. Palombo, The conceptual foundations of systems biology: An introduction, *Systems Biology-Theory, Techniques and Applications*, Nova Science Publishers, Inc., New York (2014) 1–18.
- [12] I. Mellman, T. Misteli, Computational cell biology, *The Journal of cell biology* 161 (3) (2003) 463–464.
- [13] M. Moškon, Modeli in metrike dinamike preklopa v enostavnih bioloških sistemih za potrebe računalniških struktur prihodnosti, *Doktorska disertacija*, Univerza v Ljubljani (2012).
- [14] B. Palsson, *Systems Biology: Properties of Reconstructed Networks*, Cambridge Univ Pr, 2006.
- [15] E. Klipp, W. Liebermeister, Mathematical modeling of intracellular signaling pathways, *BMC Neuroscience* 7 (1).
- [16] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, *Systems Biology. A textbook Second Edition* (2016).
- [17] P. P. Roux, J. Blenis, ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions, *Microbiology and molecular biology reviews : MMBR* 68 (2004) 320–344.
- [18] E. K. Kim, E.-J. Choi, Pathological roles of MAPK signaling pathways in human diseases, *Biochimica et biophysica acta* 1802 (2010) 396–405.
- [19] A. S. Dhillon, S. Hagan, O. Rath, W. Kolch, MAP kinase signalling pathways in cancer, *Oncogene* 26 (2007) 3279–3290.



- [20] F. Chang, L. S. Steelman, J. T. Lee, J. G. Shelton, P. M. Navolanic, W. L. Blalock, R. A. Franklin, J. A. McCubrey, Signal transduction mediated by the Ras/Raf/MEK/ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention, *Leukemia* 17 (2003) 1263–1293.
- [21] A. M. Gardner, R. R. Vaillancourt, C. A. Lange-Carter, G. L. Johnson, MEK-1 phosphorylation by MEK kinase, Raf, and mitogen-activated protein kinase: analysis of phosphopeptides and regulation of activity, *Molecular biology of the cell* 5 (1994) 193–201.
- [22] K. A. Janes, D. A. Lauffenburger, A biological approach to computational models of proteomic networks, *Current opinion in chemical biology* 10 (1) (2006) 73–80.
- [23] B. B. Aldridge, J. Saez-Rodriguez, J. L. Muhlich, P. K. Sorger, D. A. Lauffenburger, Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling, *PLoS computational biology* 5 (4) (2009) e1000340.
- [24] H. Liu, F. Zhang, S. K. Mishra, S. Zhou, J. Zheng, Knowledge-guided fuzzy logic modeling to infer cellular signaling networks from proteomic data, *Scientific reports* 6 (35652).
- [25] R. Heinrich, S. Schuster, *The regulation of cellular systems*, Springer Science & Business Media, 2012.
- [26] W. Dubitzky, O. Wolkenhauer, H. Yokota, K.-H. Cho, *Encyclopedia of Systems Biology*, Springer Publishing Company, Incorporated, 2013.
- [27] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, D. A. Lauffenburger, Logic-based models for the analysis of cell signaling networks, *Biochemistry* 49 (15) (2010) 3216–3224.
- [28] A. E. Gaweda, J. M. Zurada, Data-driven linguistic modeling using relational fuzzy rules, *IEEE Transactions on fuzzy systems* 11 (1) (2003) 121–134.
- [29] H.-J. Zimmermann, *Fuzzy set theory and its applications*, Springer Science & Business Media, 2001.
- [30] J. Virant, *Čas v mehkih sistemih*, Didakta, Slovenija, 1998.

- [31] E. H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies* 7 (1) (1975) 1–13.
- [32] M. Sugeno, An introductory survey of fuzzy control, *Information Sciences* 36 (1-2) (1985) 59–83.
- [33] J. C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences* 10 (2-3) (1984) 191–203.
- [34] Q. Song, B. S. Chissom, Fuzzy time series and its models, *Fuzzy sets and systems* 54 (3) (1993) 269–277.
- [35] C.-H. Cheng, G.-W. Cheng, J.-W. Wang, Multi-attribute fuzzy time series method based on fuzzy clustering, *Expert Systems with Applications* 34 (2) (2008) 1235–1242.

# A Priloga

Pri delu so bile uporabljene naslednje programske datoteke, ki so dostopne na naslovu [http://lrss.fri.uni-lj.si/bio/material/magdevska\\_bsc.zip](http://lrss.fri.uni-lj.si/bio/material/magdevska_bsc.zip).

- `constants.m`: kinetični parametri referenčnega ODE-modela
- `differential.m`: funkcija odvodov referenčnega modela pri podani vrednosti koncentracije EGF
- `steady.m`: implementacija učenja mehkega modela ravnotežnega stanja z algoritmom FCM
- `steady_test.m`: implementacija testiranja mehkega modela ravnotežnega stanja naučenega z algoritmom FCM pri spreminjanju koncentracije EGF
- `steady_reduced.m`: implementacija učenja reduciranega mehkega modela ravnotežnega stanja z algoritmom FCM
- `steady_reduced_test.m`: implementacija testiranja reduciranega mehkega modela ravnotežnega stanja naučenega z algoritmom FCM pri spreminjanju koncentracije

## EGF

- `fcm_learning.m`: implementacija učenja mehkega dinamičnega modela z algoritmom FCM
- `fcm_inittest.m`: implementacija testiranja pravilnosti delovanja mehkega dinamičnega modela naučenega z algoritmom FCM pri spremenjenih začetnih pogojih
- `fcm_test.m`: implementacija testiranja pravilnosti delovanja mehkega dinamičnega modela naučenega z algoritmom FCM pri spreminjanju koncentracije EGF
- `fcm_reduced_learn`: implementacija učenja reduciranega mehkega modela z algoritmom FCM
- `fcm_reduced_inittest.m`: implementacija testiranja pravilnosti delovanja reduciranega mehkega modela naučenega z algoritmom FCM pri spremenjenih začetnih pogojih
- `fcm_reduced_test.m`: implementacija testiranja pravilnosti delovanja reduciranega mehkega modela naučenega z algoritmom FCM pri spreminjanju koncentracije EGF
- `mafts_learn.m`: implementacija učenja mehkega modela z metodo večatributne mehke časovne vrste
- `mafts_inittest.m`: implementacija testiranja pravilnosti delovanja mehkega modela naučenega z metodo večatributne mehke časovne vrste pri spremenjenih začetnih pogojih
- `mafts_test.m`: implementacija testiranja pravilnosti delovanja mehkega modela naučenega z metodo večatributne mehke časovne vrste pri spreminjanju koncentracije EGF
- `mafts_reduced_learn.m`: implementacija učenja reduciranega mehkega modela z metodo večatributne mehke časovne vrste
- `mafts_reduced_test.m`: implementacija testiranja pravilnosti delovanja reduciranega mehkega modela naučenega z metodo večatributne mehke časovne vrste pri spreminjanju koncentracije EGF

- `mafts_reduced_inittest.m`: implementacija testiranja pravilnosti delovanja reduciranega mehkega modela naučenega z metodo večatributne mehke časovne vrste pri spremenjenih začetnih pogojih
- `mafts_sigmatest.m`: implementacija računanja napake mehkega modela naučenega z MAFTS pri različnih vrednostih standardnega odklona pripadnostnih funkcij