

Univerza v Ljubljani
Fakulteta za matematiko in fiziko



Določanje elastičnih konstant v nematskem tekočem kristalu

Avtor: **Simon Perovnik**

Predavatelj: doc. dr. Simon Čopar

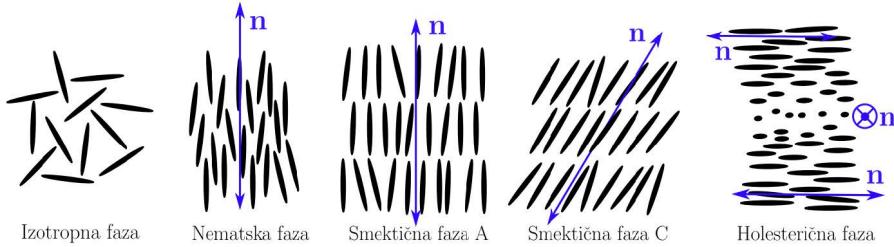
Asistent: asist. Jaka Zaplotnik, mag. fiz.

Tretja naloga pri Praktikumu strojnega učenja v fiziki

Ljubljana, november 2022

1 Uvod

Tekoči kristali (TK) so nadvse zanimivi materiali, saj posedujejo tako nekatere lastnosti tekočine, kot tudi kristala (le te se kažejo predvsem v stopnji urejenosti gradnikov). Ločimo več različnih faz ureditve, pri katerih se za pomemben parameter kategorizacije izkaže orientacija direktorskega vektorja \mathbf{n} . Gre za vektor, ki kaže v smeri preferenčne orientacije molekul. Ko nastopi izotropna faza takšna smer ne obstaja, če pa prečkamo mejo faznega prehoda ($T < T_c$) se lahko kristal uredi na različne načine, ki so prikazani na sliki (1)



Slika 1: Različne faze tekočega kristala.

Podobno kot druge materiale, lahko tudi TK deformiramo na različne načine. Za opis teh procesov vpeljemo Fran-Oseenovo gostoto električne proste energije

$$f_{FO} = \frac{K_{11}}{2}(\nabla \cdot \mathbf{n})^2 + \frac{K_{22}}{2}(\mathbf{n} \cdot \nabla \times \mathbf{n})^2 + \frac{K_{33}}{2}(\mathbf{n} \times \nabla \times \mathbf{n})^2. \quad (1)$$

Pri tem smo vpeljali tri elastične konstante (K_{11} , K_{22} in K_{33}). Prosti energiji pa lahko dodamo še en člen - to bo prispevek k energiji, ki bo posledica vsiljene strukture kristala na robovih (temu rečemo površinsko sidranje). Kristale lahko namreč pripravimo na tak način, da lahko na samem robu določimo, kako se bodo molekule orientirale. Pri tem ločimo več tipov sidranja, v tej nalogi pa se bomo posvetili predvsem pravokotni in horizontalni ureditvi sidranja.

Če celoten TK postavimo v območje električnega ali magnetnega polja, s katerima lahko vplivamo na orientacijo molekul bomo dobili še dva prispevka k prosti energiji - f_E in f_M . Če na celico pošljemo električni signal, bo ta pulz povzročil spremembo ureditve molekul - le te se bodo do neke mere poravnale s poljem, nato pa bodo zanimalo nazaj proti stacionarni legi. Temu procesu pravimo relaksacija. Njeno dinamiko lahko preko proste energije opišemo na način

$$\frac{\partial^2 \theta}{\partial z^2}(K_{11} \cos^2 \theta + K_{33} \sin^2 \theta) + \frac{K_{33} - K_{11}}{2} \left(\frac{\partial^2 \theta}{\partial z^2} \right) \sin 2\theta = \gamma_1 \frac{\partial \theta}{\partial z} \quad (2)$$

kjer θ predstavlja kot odklona vektorja \mathbf{n} od z osi, γ_1 pa je viskozna konstanta. Orientacije direktorskega vektorja ne moremo direktno meriti, lahko pa o njej sklepamo, če TK obsijemo z laserjem in gledamo prepuščeno svetlobe - TK bo namreč deloval kot fazni retarder (zaradi dvolomnosti bo spremenil polarizacijo svetlobe). V tej nalogi se bomo poslužili ravno te povezave med fazno retardacijo prepuščene svetlobe in pa elastičnimi konstantami.

2 Naloga

Naš končen cilj naloge je čim boljše napovedati vrednosti elastičnih konstant iz eksperimentalnih meritev intenzitet prepuščene svetlobe. Ker teoretično tega ne znamo eksplisitno izračunati, bomo za napoved uporabili nevronsko mrežo. Te bomo najprej natrevali na računalniško generiranih podatkih, kjer bomo žrebali naključno začetno neravnovesno orientacijo molekul in iz nje najprej izračunali potek relaksacije nato pa iz te izračunali še kako se med tako relaksacijo spreminja faza (ozioroma ob uporabi prekrižanih polarizatorjev kar intenziteta) prepuščene svetlobe.

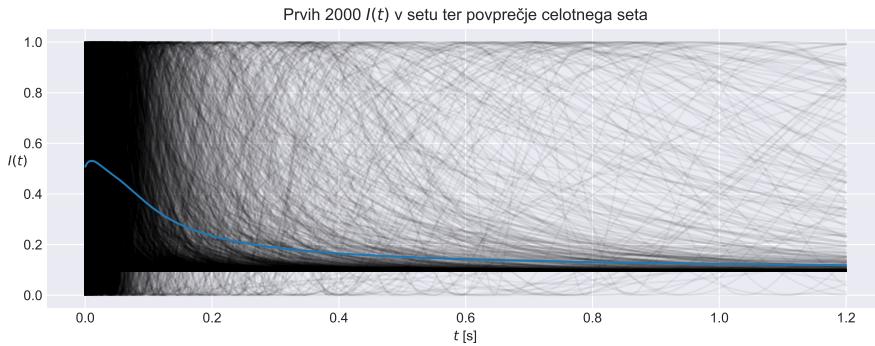
V prvem delu se bomo posvetili primeru, ko velja $K_{11} = K_{33}$. V tem primeru se dinamika relaksacije bistveno poenostavi (drugi člen povsem odpade). Zanimalo nas bo predvsem kako sestaviti optimalno nevronsko mrežo za

naš specifičen problem. Tu ne bomo uporabili še nobenih eksperimentalnih podatkov, pač pa bomo zgolj preverjali uspešnost napovedi na testnem setu generiranih potekov relaksacije.

V drugem delu bomo problem posplošili na primer, ko dopustimo, da se elastični konstanti med seboj razlikujeta. Na koncu pa bomo z natreuirano mrežo napovedali vrednosti elastičnih konstant realnega vzorca na katerem smo opravili meritve $I(t)$ in zanj že poznamo pravi vrednosti.

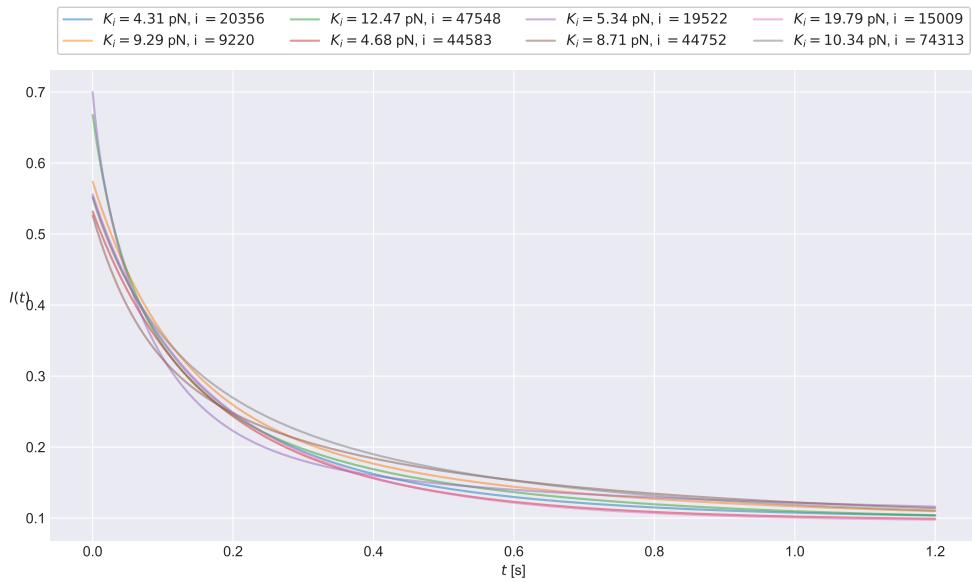
3 Podatkovni set

Set za učenje nevronske mreže je za nas že generiran - tu imamo na voljo poteke relaksacije pri valovnih dolžinah $\lambda = 500\text{nm}$ in $\lambda = 505\text{nm}$, prav tako pa lahko izbiramo med čistim in zašumljenim signalom.

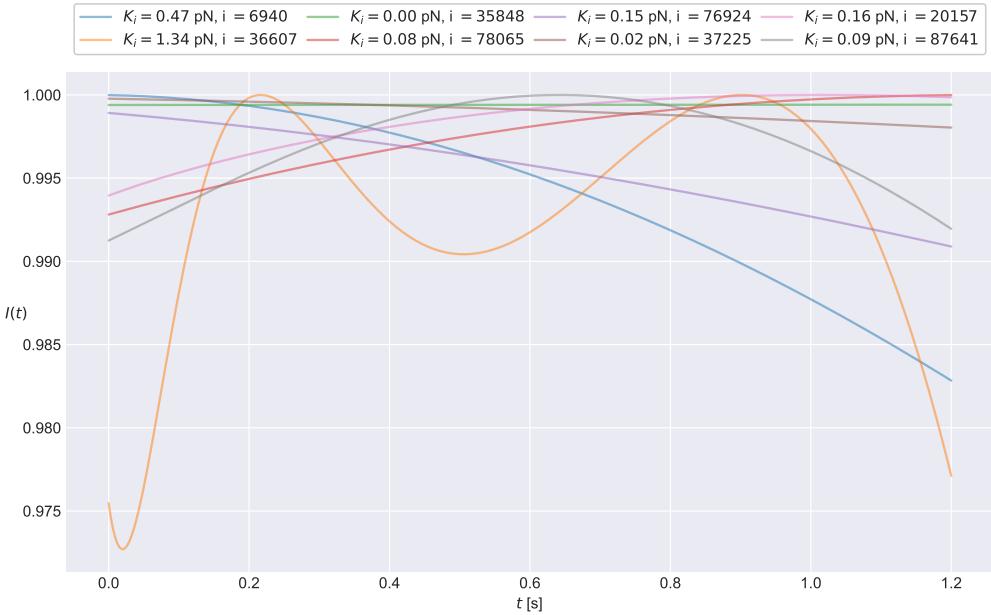


Slika 2: Računalniško generirane $I(t)$.

Set v večji meri sestavlja intenzitete ki se razmeroma hitro ustalijo pri vrednosti okoli 0.1, vidimo pa da je prisotnih tudi nekaj krivulj, ki imajo bistveno drugačen potek in morda sploh še niso dosegle relaksacije.

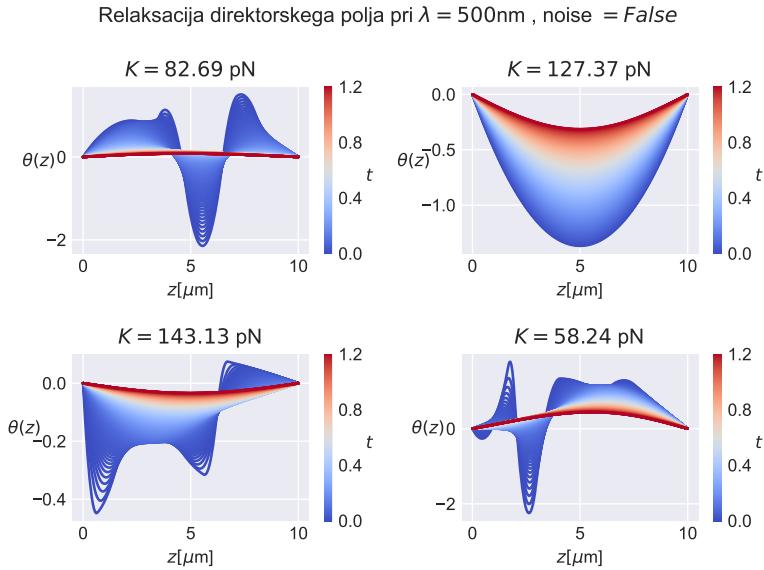


Slika 3: Primeri krivulj, ki najmanj odstopajo od povprečja.



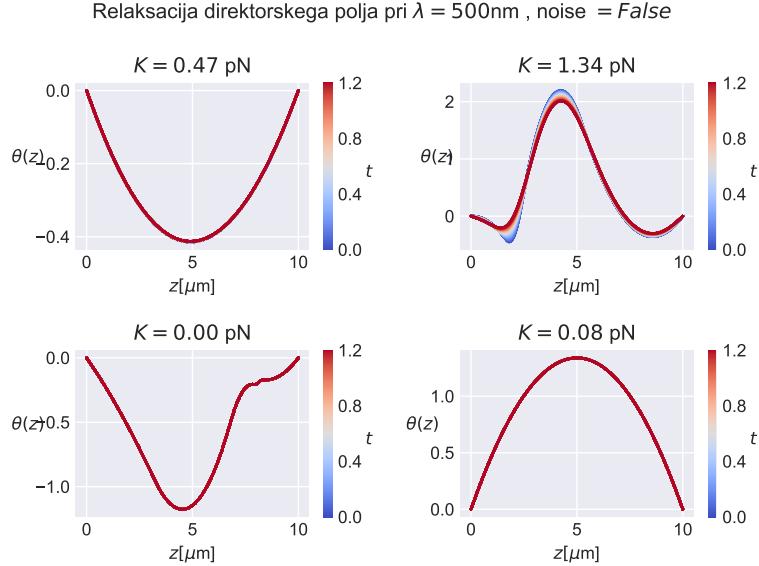
Slika 4: Primeri krivulj, ki najbolj odstopajo od povprečja.

Pogledamo si lahko tudi kako se spreminja direktorski profil TK med procesom relaksacije.



Slika 5: Nekaj primerov relaksacije direktorskega polja.

Vidimo lahko, da so si začetni profili lahko precej različni, vsi pa se relaksirajo v stanje, ko so molekule v plasti na obeh robovih postavljeni pravokotno na rob ($\theta = 0$), vmes pa dobimo profil, ki nekoliko spominja na paraboličnega, le ta pa je enkrat obrnjen v eno smer, drugič v drugo.



Slika 6: Primeri relaksacije direktorskega polja za intenzitete, ki so najbolj odstopale od povprečja.

Kot smo prej pravilno sklepali, krivulje $I(t)$, ki so najbolj odstopale od povprečja zares predstavljajo še nerelaksirane profile, kar je razvidno iz slike (6). Očitno se je pri generaciji časovnega poteka (kar se odraža v grafu intenzitete) tu nekaj zalomilo, saj sami začetni profili ne delujejo nič kaj zelo posebnega, kar samo po sebi ne bi moglo konvergirati v ravnovesno stanje.

4 $K_{11} = K_{33}$

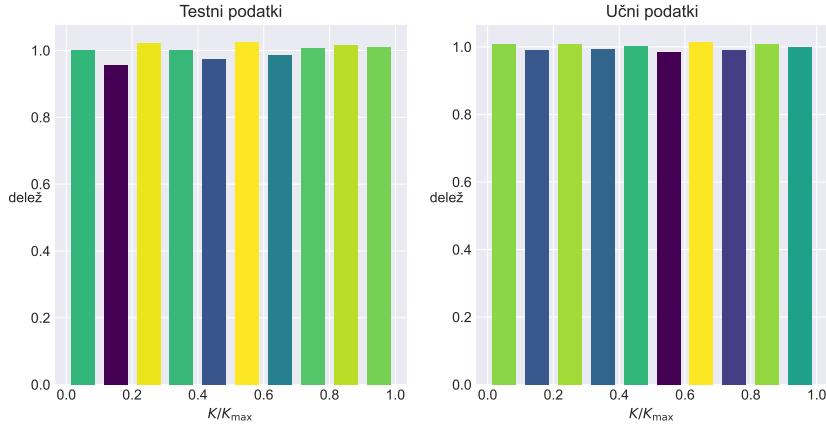
4.1 Sestava nevronske mreže

Podatkovni set moramo najprej pripraviti preden jih bomo lahko uporabili za treniranje nevronskega mrež (NM). V tem primeru bo to zelo enostavno, saj bomo zgolj reskalirali vrednosti elastičnih konstant na območje $[0, 1]$. Na tej točki bi lahko iz seta izločili še vse krivulje, kjer ni prišlo do relaksacije, saj naš set eksperimentalnih podatkov ne vsebuje takšnih primerov, tako da nevronskim mrežam znanje pridobljeno na le-teh ne bo prineslo kakšne dodatne vrednosti. Žal nam je za implementacijo tega zmanjkalo časa.

Ker še ne vemo, kako bo videti optimalna NM začnimo z nečim preprostim. Sedaj bomo poskusili s preskušanjem različnih naborov hiperparametrov dobiti občutek, kakšna mreža bo najbolj primerna za reševanje našega problema. Začeli bomo z dvema skritima plastema s po 100 in 200 nevroni, ker napovedujemo konstanto pa bo zadnja plast vsebovala zgolj en nevron. Glede na pretekle izkušnje bomo za aktivacijske funkcije povsod izbrali **swish** funkcije, za optimizator pa **Adam** s hitrostjo učenja 0.001.

V tem razdelku bomo poskusili minimizirati funkcijo izgube (mean square error) za vsak hiperparameter posebej, kasneje pa se bomo tega iskanja lotili nekoliko bolj sistematično z metodo *Grid search*.

Pred tem bomo set podatkov razdelili še na učni in testni set. To bomo storili na način, da bomo prvih 80% krivulj uvrstili med učni set, preostale pa v testni.

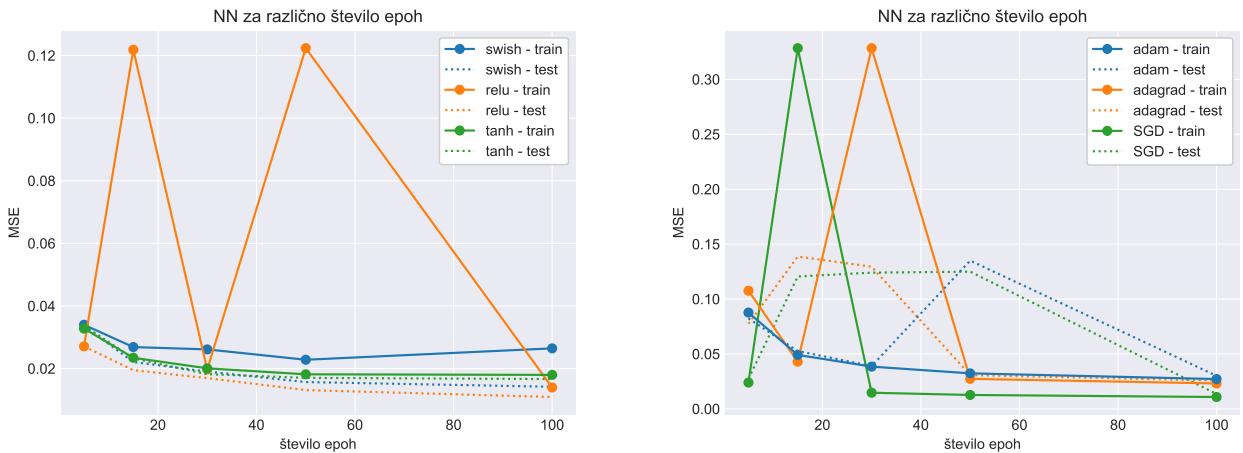


Slika 7: Uravnoteženost podatkov v učnem in testnem setu.

Pri tem je pomembno, da sta oba seta med seboj uravnotežena, kar v našem primeru tudi v dovoljšni meri sta.

4.1.1 Variacija posamičnega hiperparametra

Poglejmo si nekaj ključnih hiperparametrov in njihov vpliv na končno napako testnega in učnega seta. Pri tem bomo vseskozi uporabljali t.i. 'K-fold' validacijo za zmanjšanje zašumljenosti vrednosti funkcije izgube

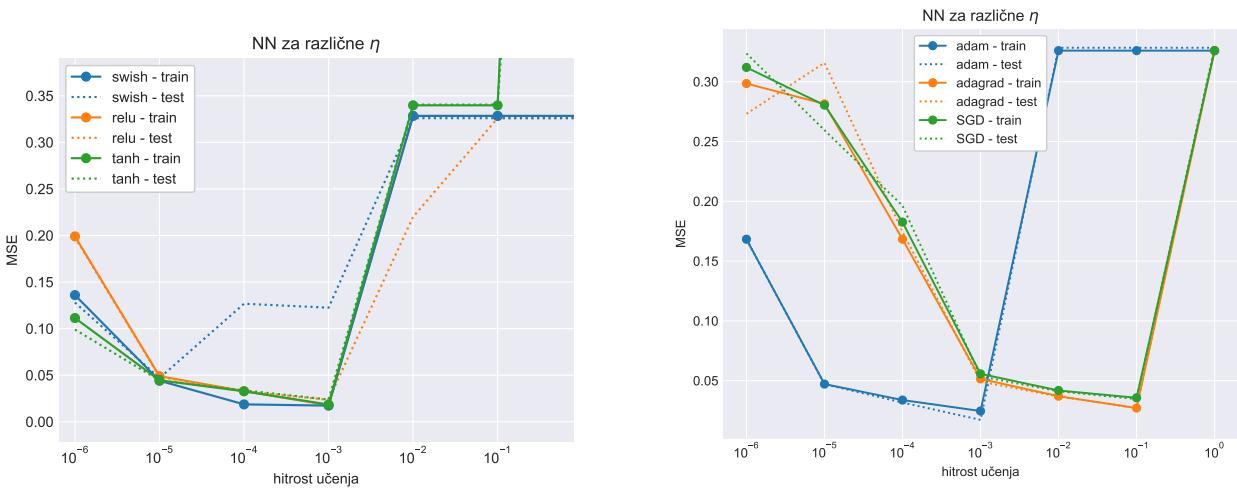


Slika 8: Vpliv števila epoch.

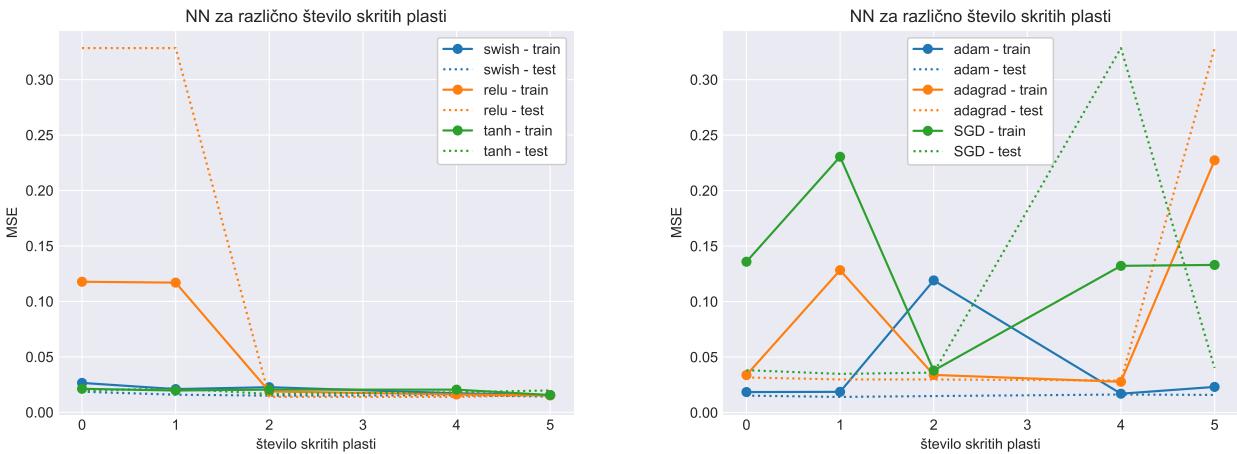
Po pričakovanjih bomo s povečevanjem števila epoch dosegli praviloma boljšo natančnost. Dokler se omejimo na majhno število le-teh, je lahko prisotnega veliko šuma v mrežah, zato (najbolj opazno pri test aktivacijske funkcije relu) lahko napaka pri posamezni simulaciji bistveno naraste, saj je vpliv naključnosti še zelo velik. Zanimivo je, da je kljub temu skoku, funkcija relu praviloma najboljše minimizirala napako na testnem setu.

Pri testu različnih optimizatorjev so uspešnosti za majhne vrednosti epoch zelo variabilne, ko pa poženemo učenje pri sto epochah pa se vrednosti napake nekako ustalijo. Dobro je, da še ne prihaja do over-fittinga, saj se obe napaki (za testni in učni set) še dobro ujemata.

Ker si zaradi časovne potratnosti ne moremo privoščiti da bi vse simulacije pri iskanju optimalnih hiperparametrov izvajali pri tako velikem številu epoch, se bomo zadovoljili s številom 30.



Slika 9: Vpliv hitrosti učenja η .



Slika 10: Vpliv števila skritih plasti. Opomba: skala na x osi je napačna, točke napake so vrisane za število skritih plasti, ki je enako 2, 3, 4, ,

Čeprav so grafi zaradi prekratkega učenja podvrženi velikemu šumu se zdi, da bo optimalno, če hitrost učenja nastavimo na 0.001, pri globini mreže pa se zdi, da dodatne plasti hitro ne dodajo več bistvenega izboljšanja.

4.1.2 Grid search

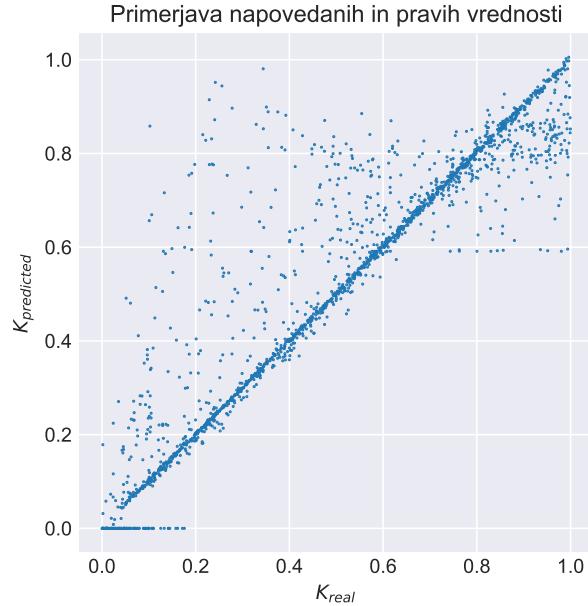
Iskanja optimalnih hiperparametrov se lahko lotimo tudi nekoliko bolj sistematično z uporabo metode 'Grid search', pri kateri preverimo bistveno večje število različnih kombinacij vrednosti hiperparametrov, saj v splošnem ni nujno da minimizacija napake pri posameznem hiperparametru vodi do globalnega minimuma nabora vseh možnih kombinacij in je tako potrebno dejansko preveriti 'vse' različne kombinacije. Tak pristop pa seveda hitro vodi k veliki časovni zahtevnosti saj se število potrebnih simulacij zelo hitro povečuje.

V splošnem nam je na voljo zelo veliko hiperparametrov, ki bi jih lahko poskusili spremenljati, osredotočili pa se bomo le na nekaj najbolj bistvenih - aktivacijske funkcije (zaradi preprostosti bomo na vseh plasteh uporabili enako funkcijo), optimizator, hitrost učenja, globina mreže, število neuronov na zadnji plasti in na vsaki nezadnji plasti, ter velikost 'batch-a'. Da zmanjšamo število ponovitev bomo iskanje razdelili na dva ločena iskanja

minimuma, s čimer sicer izgubimo nekaj splošnosti, a upamo da bo dobijena kombinacija vseeno dober približek optimalnih hiperparametrov. Rezultati iskanja se nahajajo v ločenih .txt datotekah, ki sta priloženi poročilu. Kot optimalen se je izkazala struktura s tremi skritimi plastmi s po 200 in na zadnji plasti s 400 nevtroni, kjer so vse aktivacijske funkcije 'relu'. Kot optimalen optimizator se je izkazal Adam s hitrostjo učenja $\eta = 0.001$, pri-merna velikost 'batch-ov' je 32, število epoh pa 100 (lahko da je optimalno to še povečati, a tega nismo testirali).

4.2 Pregled učinkovitosti nevronske mreže

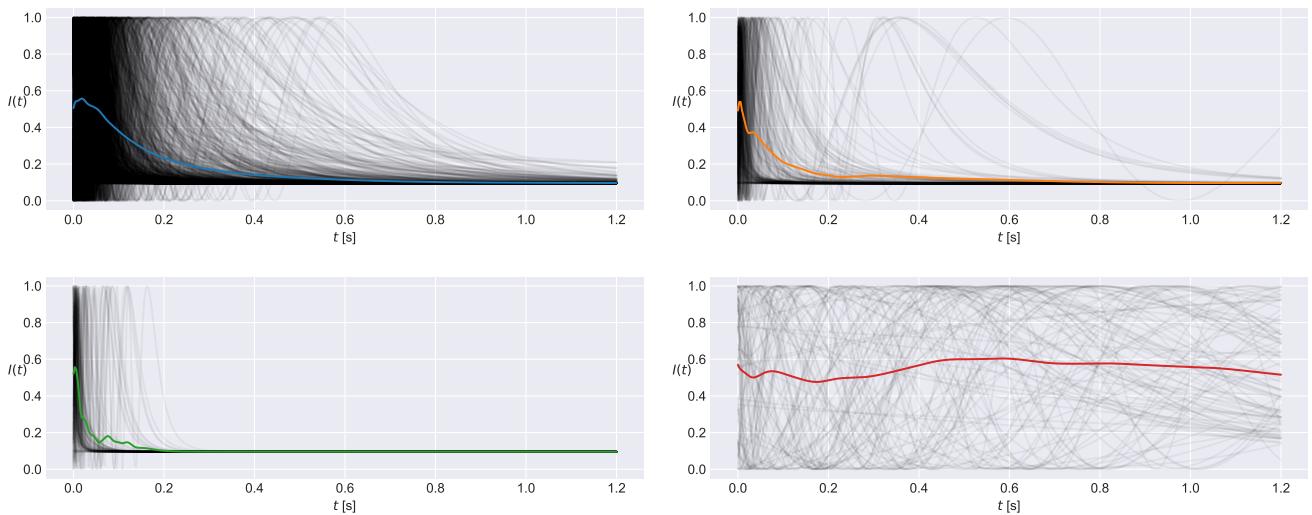
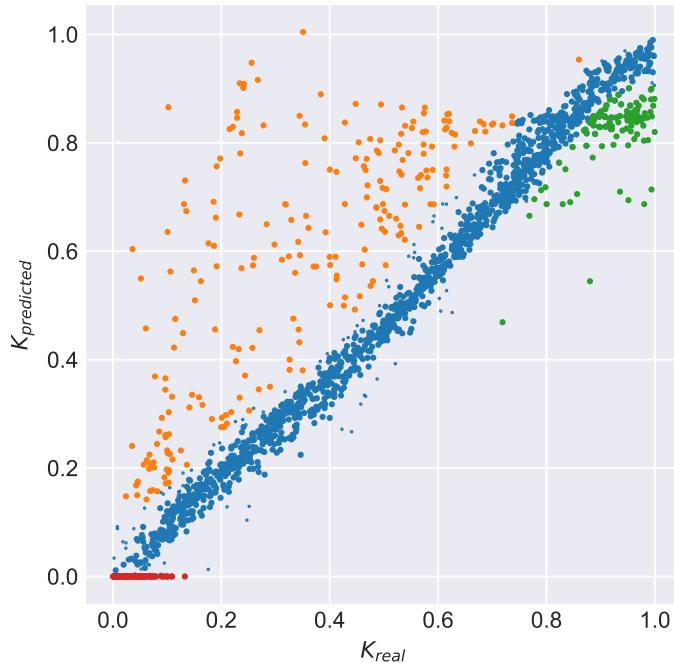
Z naučeno NM lahko sedaj napovemo vrednosti elastične konstante in primerjamo v kakšni meri se napovedi ujemajo z pravimi vrednostmi. Za velikost testnega vzorca bomo vzeli 19% celotnega seta. Zato, da bomo lahko primerjali učinkovitost mreže, ki smo jo poiskali s pomočjo 'grid search-a' si poglejmo najprej rezultat generiran z mrežo katere hiperparametre smo določili z variacijo posamičnega parametra.



Slika 11: Primerjava napovedanih in dejanskih vrednosti elastične konstante.

V splošnem se zdi ujemanje še kar dobro, saj je opaziti močen trend po diagonali, kjer so predstavljene vrednosti, za katere je mreža zelo dobro napovedala K . Opazimo lahko, da se nam pojavi nekaj na videz ločenih območij, kjer je ujemanje slabše. Za podrobnejšo analizo smo implementirali t.i. *lasso tool*, s katerim lahko na grafu označimo poljubno množico točk in jih povežemo s pripadajočimi krivuljami $I(t)$. Za vsak izbor nato izrišemo graf vseh krivulj in pa njihovo 'povprečno' krivuljo.

Primerjava napovedanih in pravih vrednosti - kategorizirano

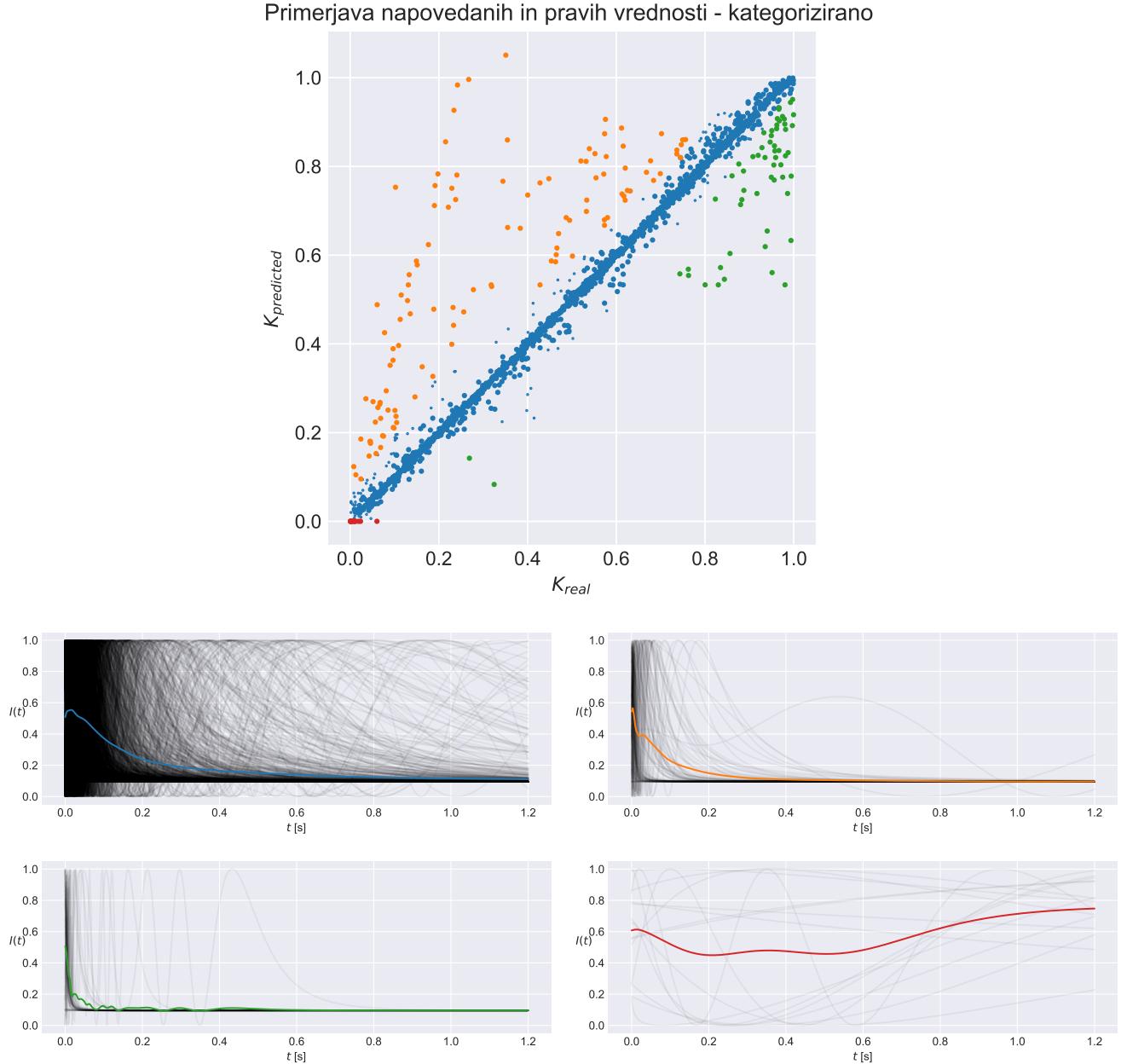


Slika 12: Kategorizacija napovedi nevronske mreže, kjer smo hiperparametre poiskali z minimizacijo napake za vsak hiperparameter posebej.

- Modra - Dobro ujemanje z napovedjo. Tu se nahaja veliko različnih krivulj, največ tistih, ki se relaksirajo nekoliko pozneje je prav tako v tem delu.
- Oranžna - Po večini krivulje, ki se nekoliko hitreje relaksirajo.
- Zelena - Intenziteti profili, ki se še hitreje relaksirajo, imajo zelo oster vrh.
- Rdeča - Krivulje, kjer še ni prišlo do relaksacije.

4.3 Pregled učinkovitosti za hiperparametre najdene z Grid search-em

Analizo rezultatov lahko napravimo tudi na napovedih, ki smo jih pridobili z mrežo naučeno pri hiperparametrih poiskanih z Grid search-em.

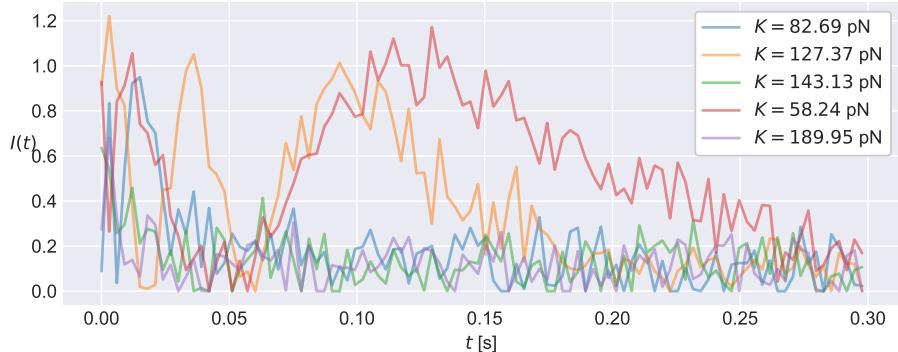


Slika 13: Kategorizacija napovedi nevronske mreže, kjer smo hiperparametre poiskali z Grid search-em.

- Modra - Očitno boljše ujemanje kot v prejšnjem razdelku, za vse več krivulj, tudi takih, ki se kasneje (ali pa sploh še ne) relaksirajo se napoved dobro ujema z resnično vrednostjo.
- Ostali profili so podobni kot v prejšnjem razdelku, še vedno nam dela težave intenzitete, ki zelo hitro pada na nič, in te, ki se sploh ne relaksirajo pravilno.

4.4 Učenje mreže z zašumljenimi podatki

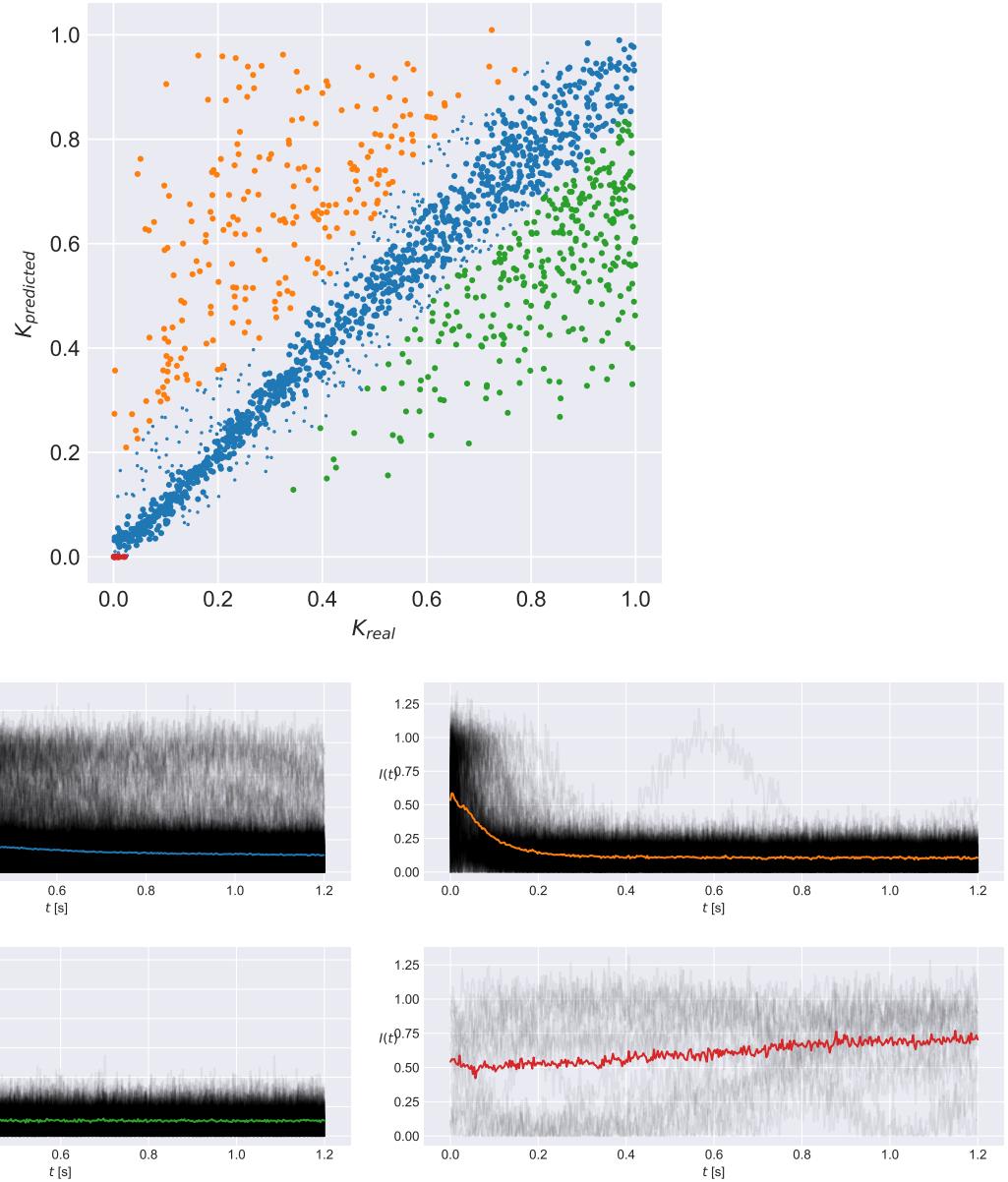
Na voljo so nam tudi podatki, kjer je na vsako mersko točko primešan naključen bel šum. Nekaj takih krivulj je prikazanih na sliki (14). Pogledamo si lahko več kombinacij, kako se naša mreža odziva v kolikor je učni ali pa testni set zašumljen.



Slika 14: Primer nekaterih zašumljenih krivulj $I(t)$.

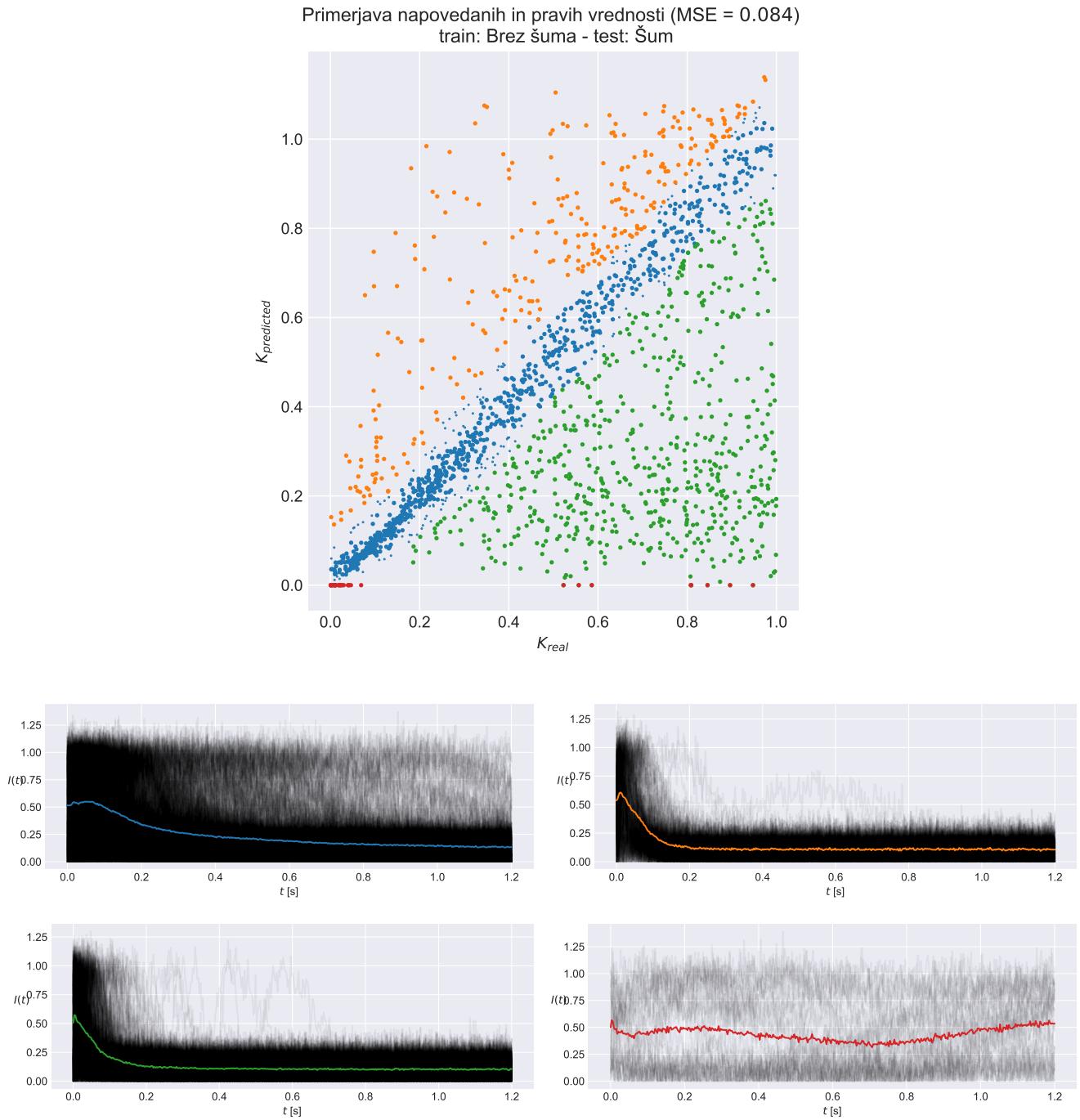
4.4.1 Učenje - s šumom, test - s šumom

Primerjava napovedanih in pravih vrednosti (MSE = 0.033)
train: Šum - test: Šum



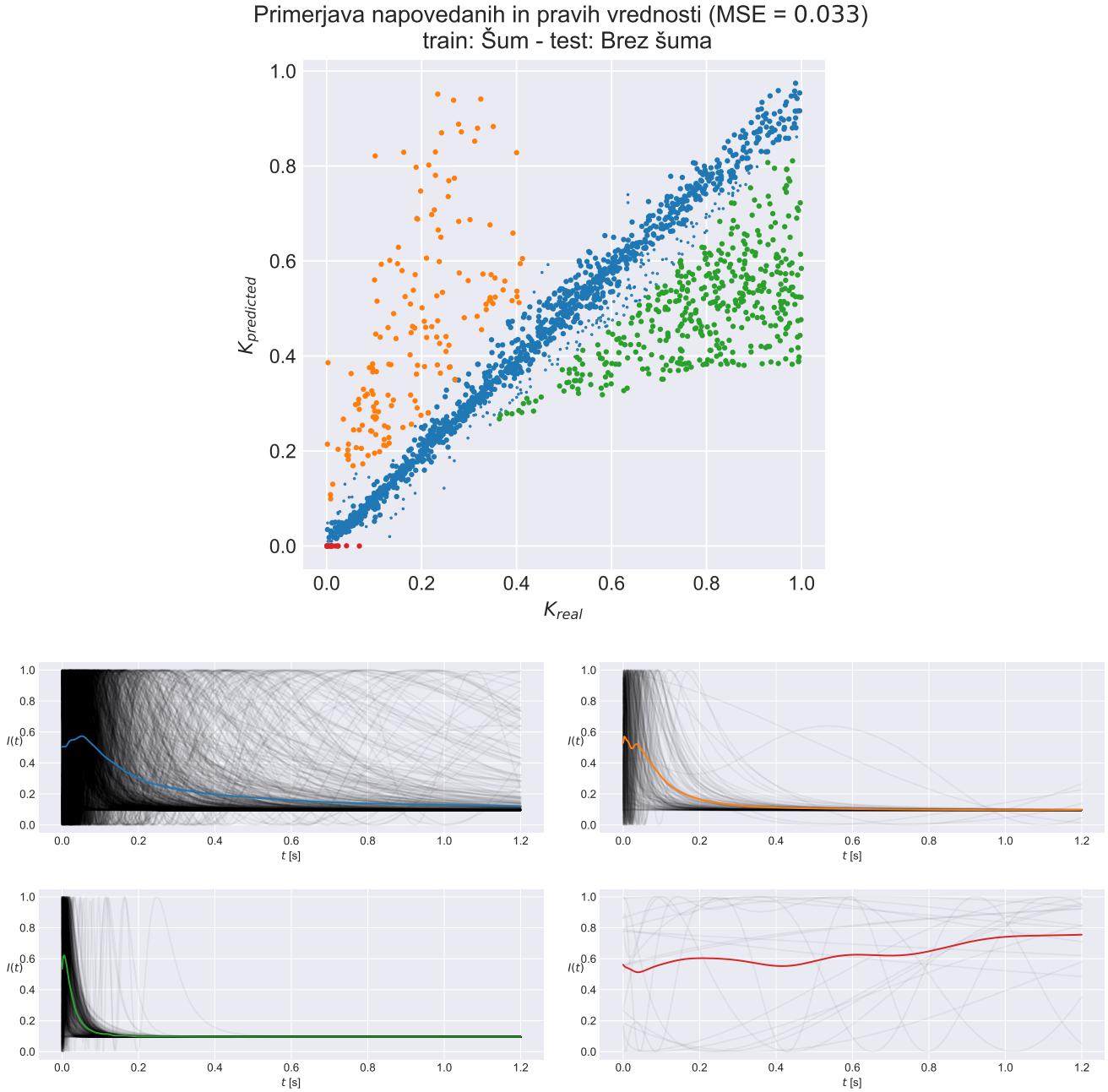
Slika 15: Analiza napovedi za zašumljen testni set podatkov, kjer smo mrežo učili na pravtako zašumljenih podatkih.

4.4.2 Učenje - brez šuma, test - s šumom



Slika 16: Analiza napovedi za testni set podatkov, kjer smo mrežo učili na pravtako zašumljenih podatkih.

4.4.3 Učenje - s šumom, test - brez šuma



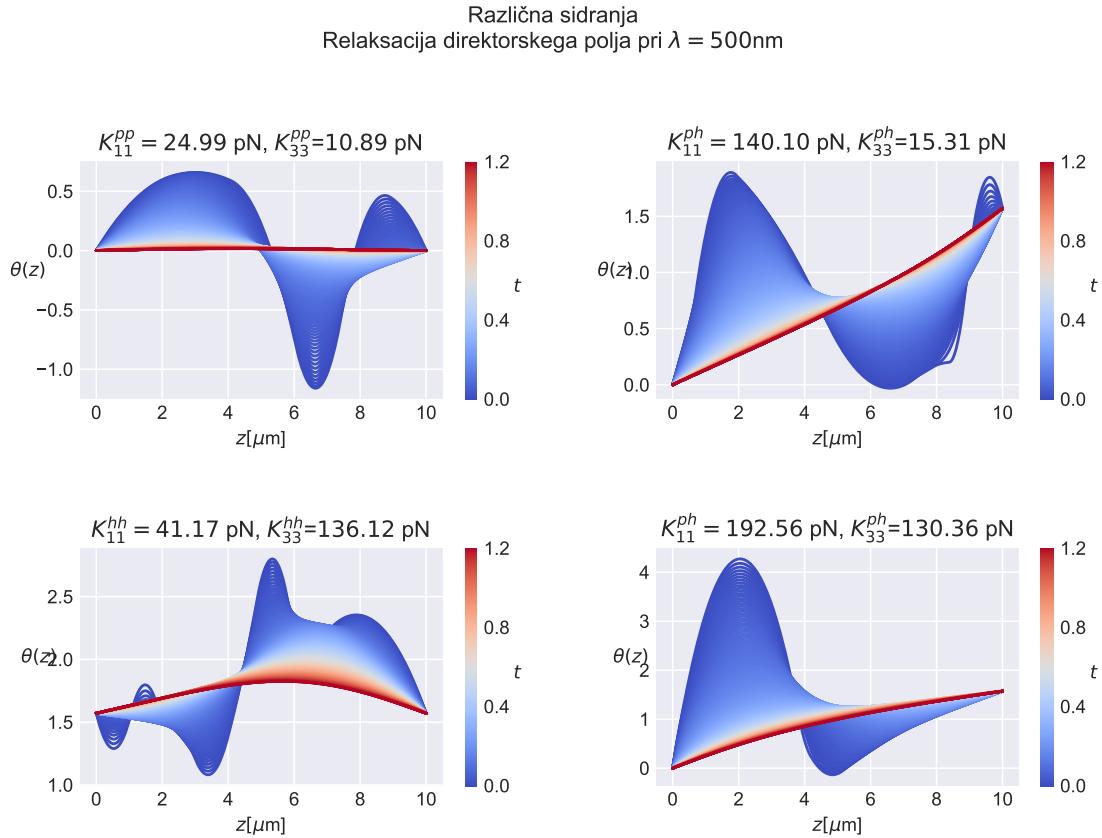
Slika 17: Analiza napovedi za zašumljen testni set podatkov, kjer smo mrežo učili na 'čistih' podatkih.

5 $K_{11} \neq K_{33}$

Posplošitev na določevanje obeh elastičnih konstant v splošnem predstavlja veliko razliko v obravnavanem modelu. Ker pa so za nas učni podatki že pripravljeni je modifikacija nevronske mreže, da bo primerna tudi za posložen primer, zelo enostavna, saj potrebujemo le spremeniti število neutronov na izhodu iz 1 na 2.

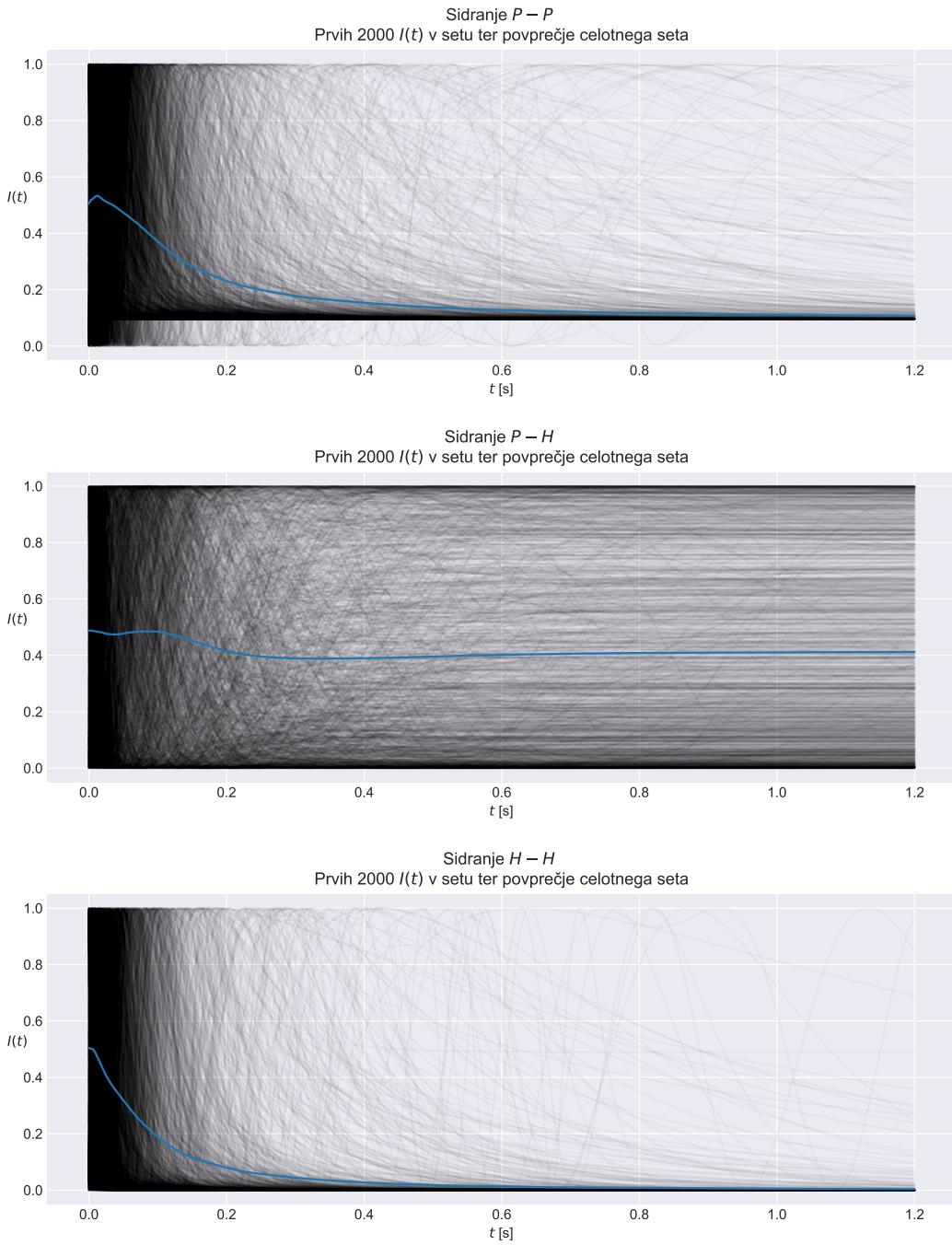
5.1 Različne kombinacije sidranja

Molekulam lahko s skrbno pripravo robnih površin med katere vnesemo tekoči kristal na robu vsilimo orientacijo. V tem razdelku bomo primerjali kako dobro lahko z našo nevronsko mrežo napovedujemo elastični konstanti pri različnih sidranjih. Na sliki (18) je izrisanih nekaj primerov relaksacij pri različnih sidranjih. Tudi iz robnih točk se točno vidi, za kakšno vrsto sidranja gre pri posamezni relaksaciji. Opazimo lahko različnost v končni relaksaciji, kjer ravnovesno stanje ni vedno ravna črta. Morda je zakriviljenost znak, da pravi minimum še ni bil dosežen.



Slika 18: Relaksacije pri različnih načinih sidranja.

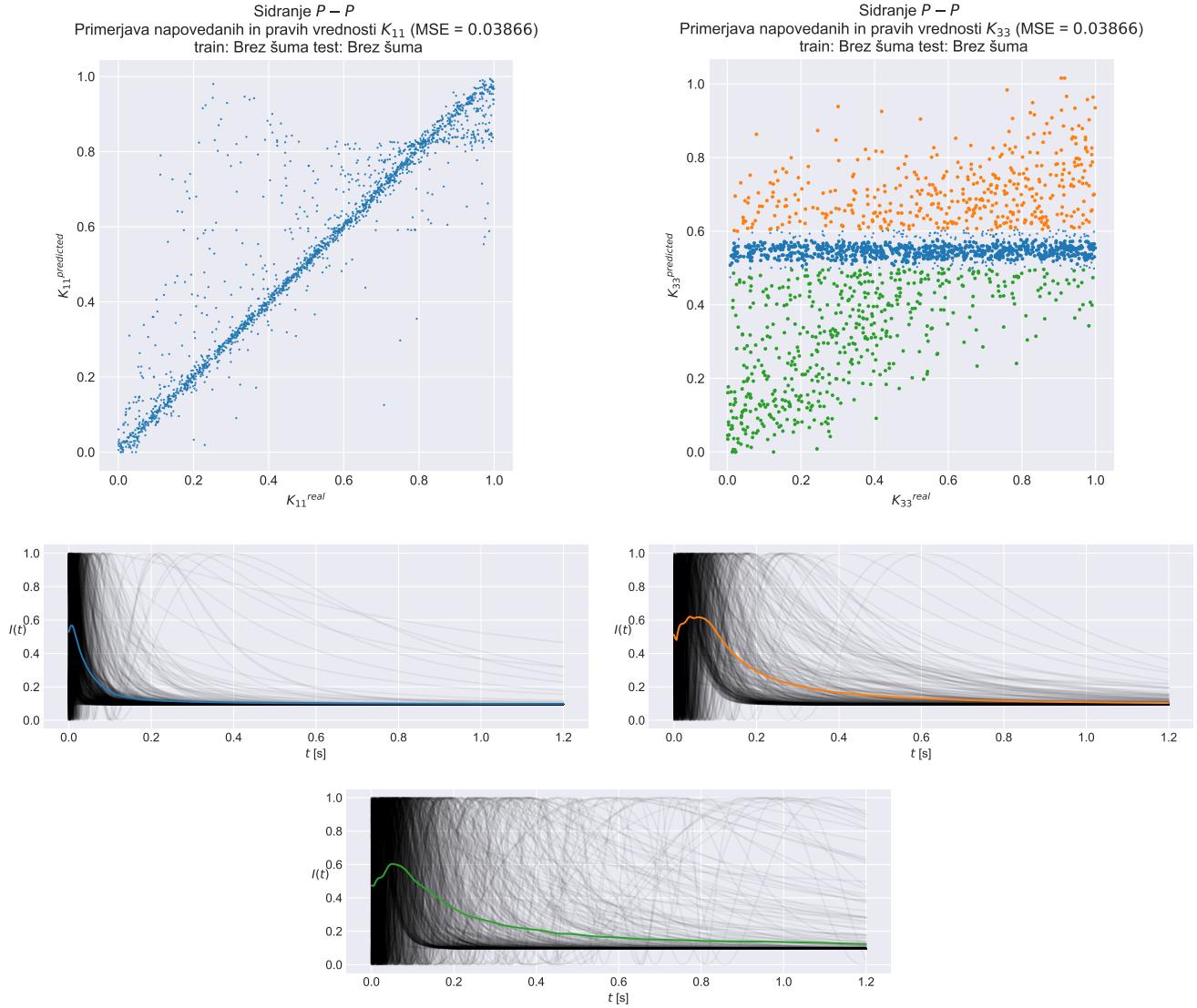
Pogledali si bomo kako dobro napoveduje NM velikosti elastičnih konstant pri različnih sidranjih. Pri tem bomo vseskozi delali le s čistimi podatki.



Slika 19: Pregled intenzitet različnih sidranj.

Sidranje P - P je edino, ki se relaksira na način, da je tudi končna amplituda neničelna. Pravtako lahko pri istem sidranju opazimo nekoliko več počasnih relaksacij, kot pri recimo H - H. Močno izstopa tudi sidranje P - H, katere povprečna intenziteta se giblje okoli sredine območja, kar nakazuje na to, da je v setu prisotnih veliko krvulj, ki se še niso relaksirale oziroma se relaksirajo v različne končne vrednosti. Mislimo si, da bo napoved najslabša ravno pri tej vrsti sidranja.

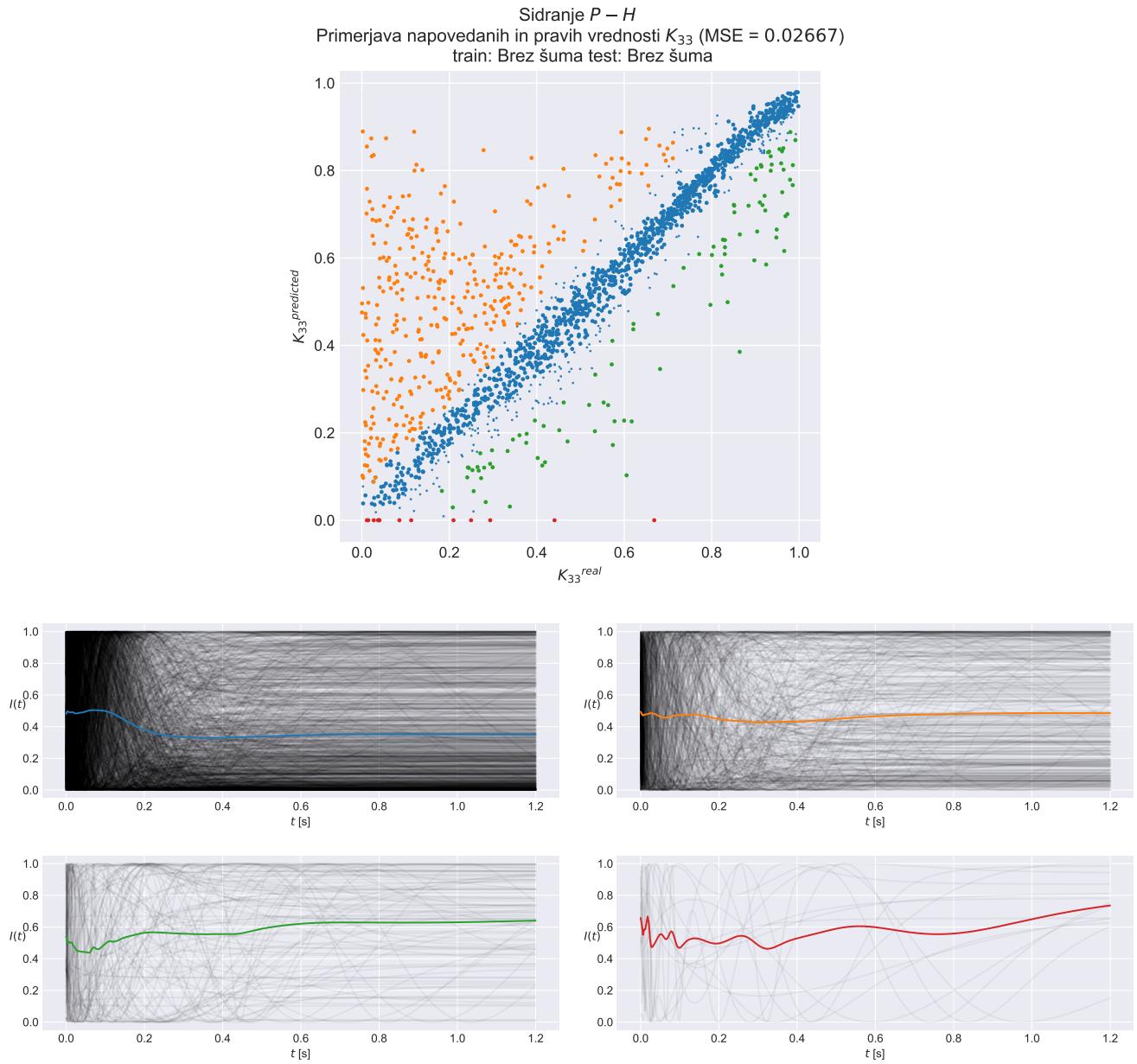
5.1.1 Sidranje P - P



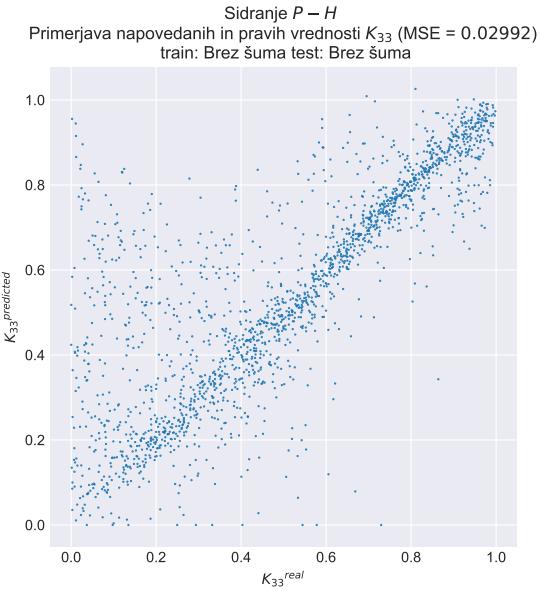
Slika 20: Rezultati napovedi NM pri sidranju P - P na podatkih brez šuma. Levo je prikazana napoved za K_{11} , desno pa za K_{33} .

Ujemanje pri K_{11} je primerljivo dobro, kot pri prejšnjih primerih, medtem ko je napoved za K_{33} bistven slabša. Izoblikuje se tudi nekakšen gostejši pas, ki je prikazan z modro. Raziskali smo vse tri razdelke, na katere se razdeli graf če predpostavimo, da je gostejši pas zaključena celota, a tudi iz podrobne analize ne uspemo priti do kakšnih jasnih zaključkov.

5.1.2 Sidranje P - H



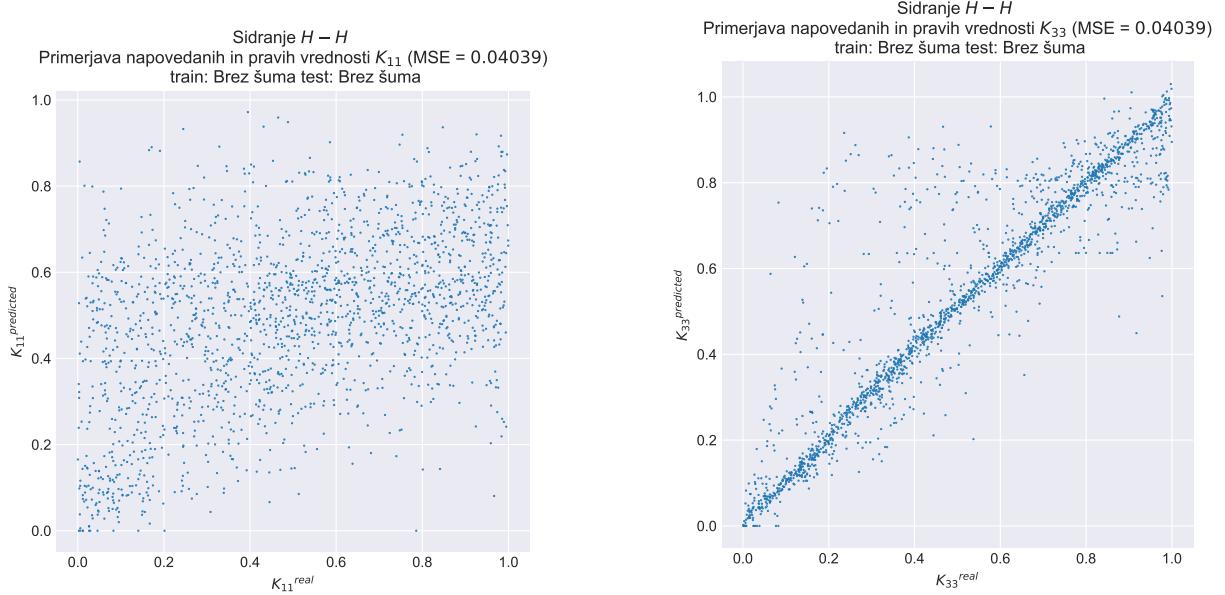
Slika 21: Rezultati napovedi za K_{11} pri sidranju P - H na podatkih brez šuma.



Slika 22: Rezultati napovedi za K_{33} pri sidranju P - H na podatkih brez šuma.

Rezultati pri sidranju P - H so zelo dobri, saj smo dobili v tem primeru najnižji MSE. Opazimo pa lahko, da se povprečna intenziteta pri vseh izoliranih delih giba okoli vrednosti 0.5, kar pomeni, da so povsod približno enako zastopane glede na sredinsko os pri 0.5 simetrične krivulje. Iz tega bi lahko tudi sklepali, da je v NM na nek način zapisana invarianca na takšno zrcaljenje vhodnih podatkov.

5.1.3 Sidranje H - H



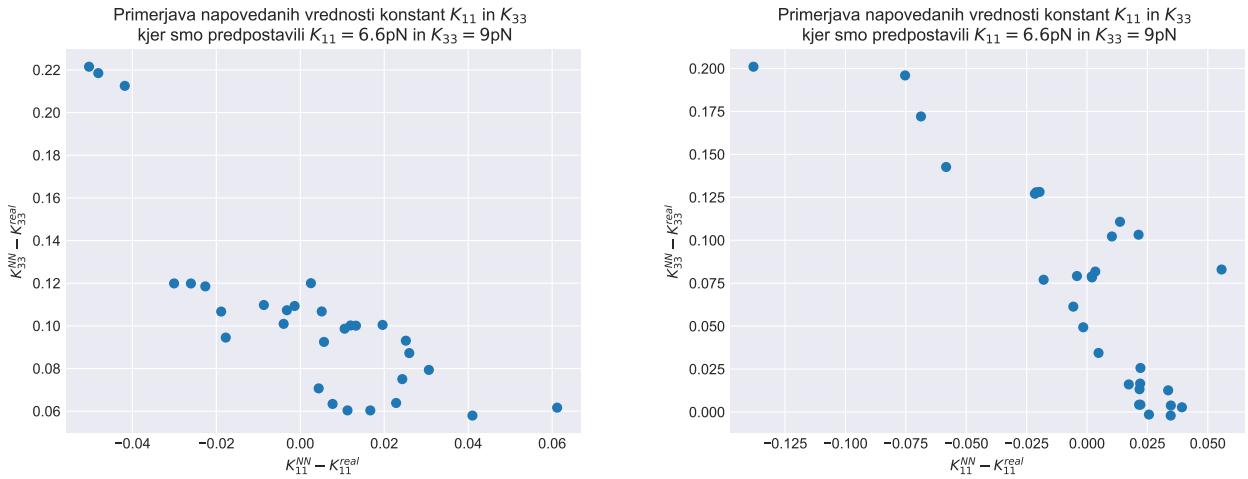
Slika 23: Rezultati napovedi za K_{11} in K_{33} pri sidranju H - H na podatkih brez šuma.

Podobno kot pri sidranju P - P, so napovedi za eno elastično konstanto dobre, pri drugi pa bistveno slabše, le da sta vlogi v tem primeru ravno zamenjani. V tem primeru nismo dobili gostejšega pasu kot pri sidranju H - H.

5.2 Napoved za eksperimentalno izmerjene meritve

Na voljo imamo tudi sicer majhen nabor izmerjenih intenzitet na dejanskem TK za katerega je bilo že ugotovljeno, da sta $K_{11} \approx 6.6\text{pN}$ in $K_{33} \approx 9.0\text{pN}$. Meritve so podane tu na nekoliko drugačen način, kot smo jih generirali, zato jih moramo sprva vzorčiti na enak interval, kot so vzorčene generirane intenzitete. Pravtako moramo biti pozorni, da so bile meritve opravljene pri $\lambda = 505\text{nm}$, kjer je bilo sidranje tipa P - H.

Uporabili bomo enako NM kot prej, čeprav ni zagotovila, da bodo izbrani hiperparametri optimalni tudi v tem primeru.



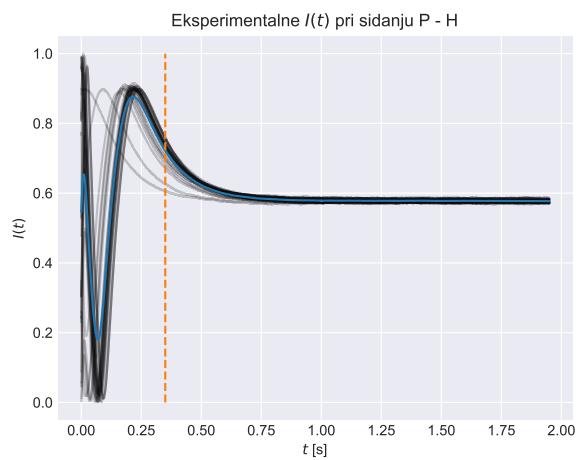
Slika 24: Dve napovedi za elastični konstanti na podlagi eksperimentalnih meritov.

Ker je vzorec tako zelo majhen (30) meritov, se lahko končna napoved za vsako ponovitev precej razlikuje. To se vidi tudi na sliki (24). Vidimo da je ujemanje v splošnem boljše za K_{11} . Omeniti je treba tudi, da so prikazane normirane vrednosti konstant, tako odstopanje za recimo 0.2 predstavlja 20% napako.

5.2.1 Povečanje seta meritov

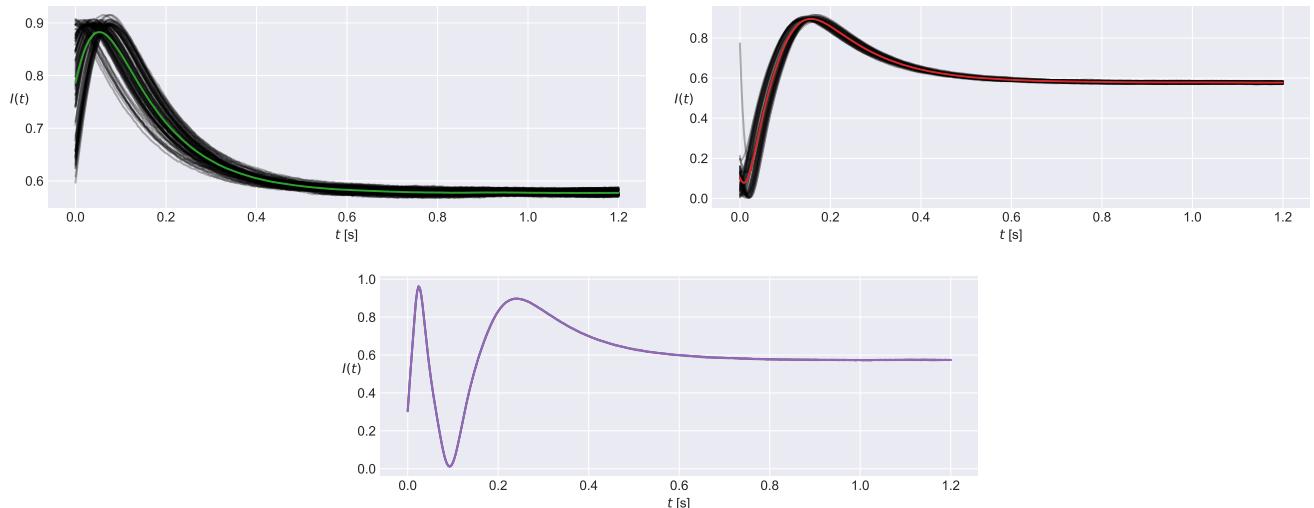
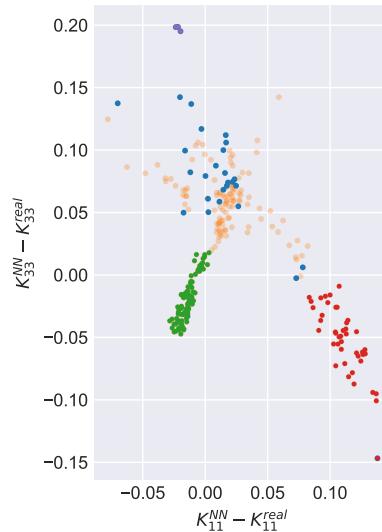
Da bi si nekoliko razširili testni set se lahko poslužimo sicer zelo enostavnega a efektivnega trika. Glede na to, da so eksperimentalne meritve trajale 1.95 sekund, vse naše generirane pa zgolj 1.2, lahko iz merjenih signalov tvorimo nove meritve, tako da vzamemo za nek t_{\max} zakasnjen signal, ki bo skupaj trajal 1.2 sekundi. S tem smo efektivno dodali novo meritev, kjer smo začeli meriti kako se polje relaksira pri nekoliko drugačnim začetnim pogojem. Najprej se moramo vprašati, kolikšen je t_{\max} , da bo intenzitetni profil še smiseln. V ta namen pregledamo vse meritve, ki so prikazane na sliki (25). Glede na izkušnjo videnih generiranih $I(t)$ ocenimo, da je primeren t_{\max} okoli 0.30.

Sedaj lahko naš set razsirimo na poljubno velikost s tem, da dodajamo naključno žrebane izseke že zmerjenih intenzitet. To smo implementirali na način, da smo za vsak nov izsek naključno žrebali meritve, ki jo bomo vzorčili in pa naključen čas, pri katerem bomo začeli.



Slika 25: Eksperimentalno izmerjene krivulje $I(t)$.

Primerjava napovedanih vrednosti konstant K_{11} in K_{33}
kjer smo predpostavili $K_{11} = 6.6\text{pN}$ in $K_{33} = 9\text{pN}$



Slika 26: Rezultati napovedi za eksperimentalne podatke z razširjenim setom.