

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA FIZIKO

PRAKTIKUM STROJNEGA UČENJA V FIZIKI

**4. naloga: Integracija dinamičnih sistemov v času - meteorološka
napoved**

Žiga Šinigoj, 28222025

Ljubljana, november 2023

1 Uvod

Naloga je sestavljena iz dveh delov. V prvem delu bomo s pomočjo rekurentnih nevronske mreže (RNN) poskusili čim bolj napovedati dinamiko Lorentzovega '63 modela. V drugem naloge pa bomo s pomočjo konvolucijskih nevronske mreže (CNN) poskusili čim bolj napovedati vreme oz. potencial Z500. V obeh primerih uporabimo nevronske mreže za integracijo kaotičnih dinamičnih sistemov.

2 Lorenzov '63 sistem

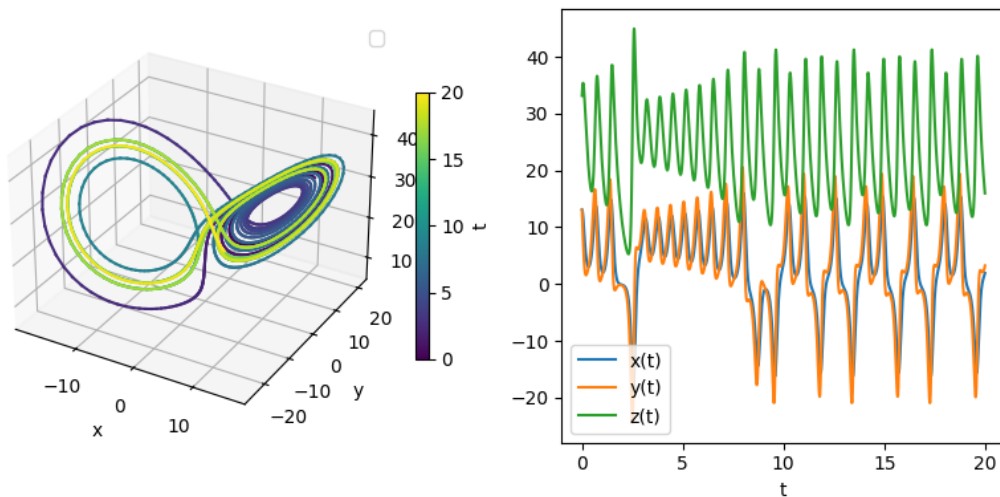
Lorenzov '63 reduciran sistem enačb opisuje gibanje tekočine med med neskončnima plosčama z različno temperaturo. Za podane vrednosti parametrov modela postane dinamičen sistem nelinearen, neperiodičen in disipativen. Velja za enega preprostejših modelov atmosferske dinamike, ki pa še vedno vključuje nelinearnosti in kaos (tako kot atmosferski sistem).

Cilj naloge je najbolje napovedati dinamiko takega sistema s pomočjo RNN. Nabor podatkov za nevronske mreže dobimo z numerično integracijo sistema Lorentzovih enačb (metoda Runge-Kutta). Čaovni razvoj integriranega sistema je prikazan na sliki 1. Podatkovni set vsebuje 10^6 točk do časa $t = 100$ ($dt = 0.01$). Ker je sistem neperiodičen lahko vzamemo različne odseke integriranega sistema in jih uporabimo kot učno, validacijsko in testno množico za nevronske mreže. Učna množica vsebuje prvih $2 \cdot 10^5$ točk, validacijska množica pa $5 \cdot 10^4$ točk. Tako množico sem uporabil pri iskanju najboljšega modela. Ko sem našel optimalen model sem ga za napoved učil na prvih $7 \cdot 10^5$ točkah in validiral na naslednjih 10^5 točkah. Pri iskanju optimalnih parametrov sem šel po predlaganem vrstnem redu. Ko sem našel nek optimalni hiperparameter, sem ga uporabil pri iskanju naslednjega optimalnega hiperparametra. Optimalne parametre sem poiskal za modele: RNN many-to-many, RNN many-to-one, LSTM many-to-many, LSTM many-to-one.

Za napoved vzamemo 1000 začetnih pogojev iz konca podatkovnega seta in z integriranjem do poljubnega časa dobimo časovni razvoj sistema ter našo testno množico. Za merjenje kvalitete napovedi bomo uporabili funkcijo izgube *val_mean_squared_error* in metriko

$$E(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left((X_i^f - X_i^t)^2 + (Y_i^f - Y_i^t)^2 + (Z_i^f - Z_i^t)^2 \right)} \quad (1)$$

kjer s superskriptom f označimo napoved iz nevronske mreže, s superskriptom t označimo *resnico*. Metrika pove povprečno odstopanje napovedi nevronske mreže, kjer povprečimo po 1000 različnih trajektorijah.



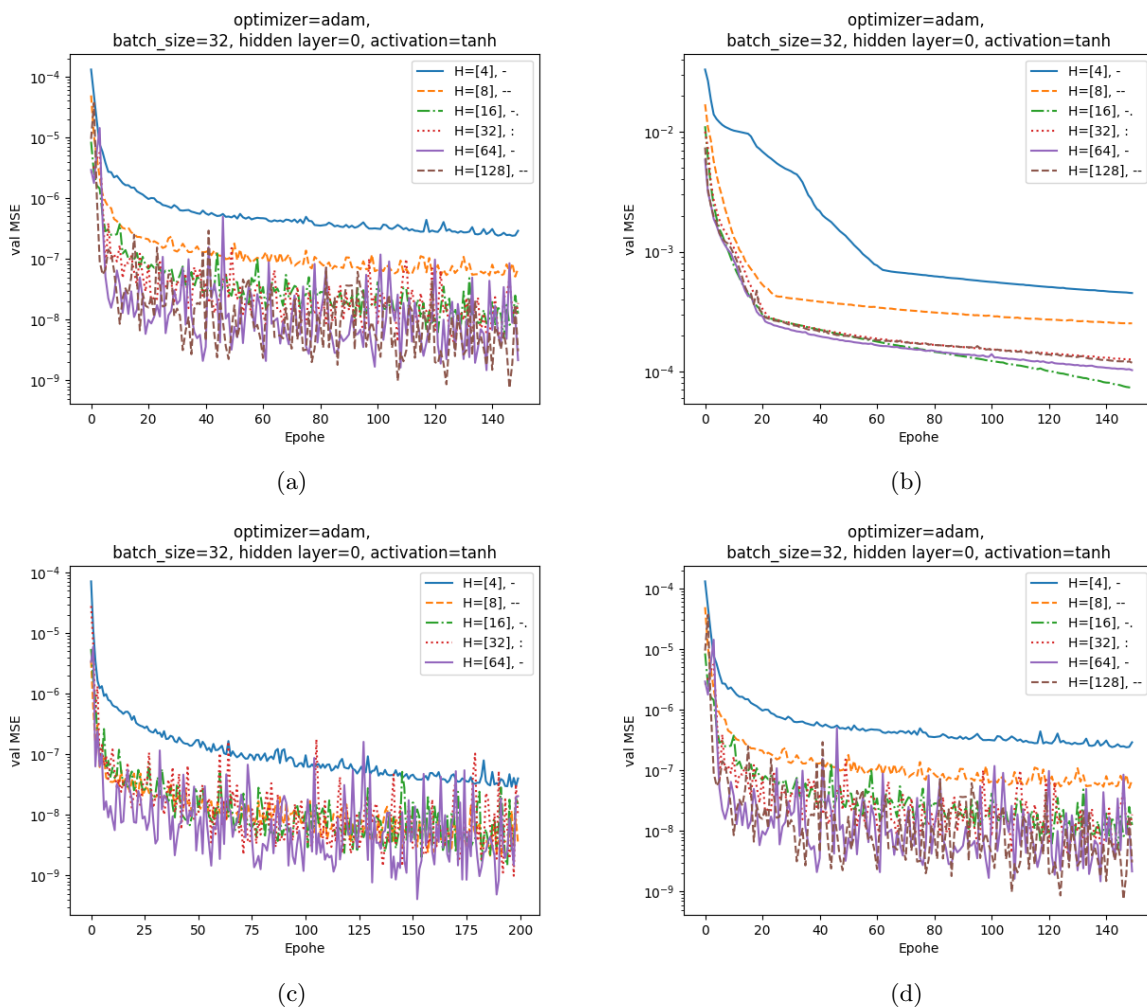
(a)

Slika 1: a) Časovni razvoj Lorentzovega sistema '63. b) Projekcija časovnega razvoja Lorentzovega sistema '63 na ravnino.

2.1 Velikost skritega stanja

Za začetno mrežo sem vzel RNN mrežo iz navodil. Najprej je bilo potrebno poiskati optimalno velikost skritega stanja. Rezultati učenja so prikazani na sliki 2. Na splošno je bilo časovno veliko bolj potratno poiskati optimalne parametre za many-to-one metodo, saj ob vsakem časovnem koraku napovemo iz k vrednosti samo naslednjo vrednost. Pri many-to-many metodi napovemo po k vrednosti in zato naredimo manj korakov za napoved enake testne množice. Načeloma nam večje skrito stanje da boljšo napoved. Vseeno pa ni preveč smiselno pretiravati, saj nad neko velikostjo skritega stanja se model bistveno ne izboljša. Z večanjem skritega stanja večamo tudi časovno zahtevnost. Optimalna velikost skritih stanj:

- RNN mto (many-to-one): $H=16$
- RNN mtm (many-to-many): $H=16$
- LSTM mto (many-to-one): $H=16$
- LSTM mtm (many-to-many): $H=32$



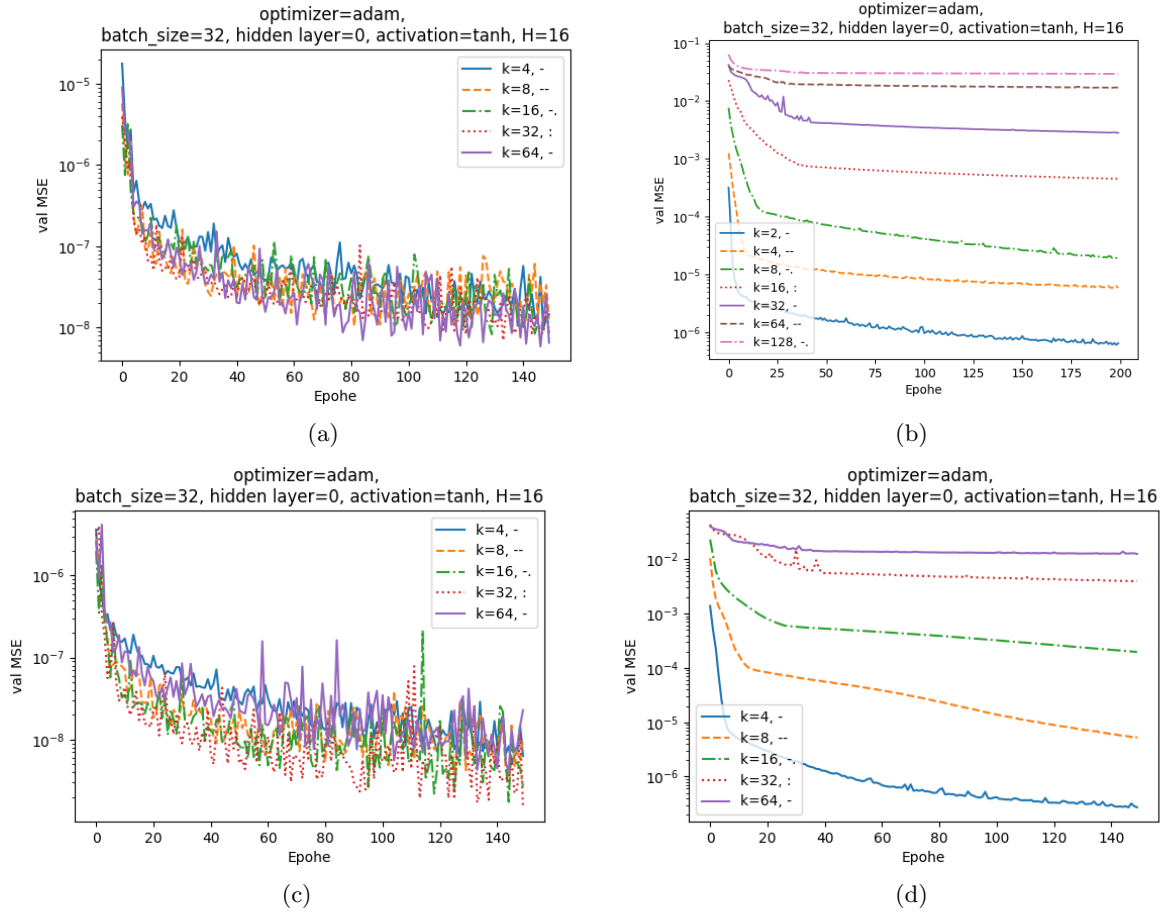
Slika 2: Validacijska funkcija izgube pri različnih velikostih skritega stanja- H . a) RNN many-to-one model. b) RNN many-to-many model. c) LSTM many-to-one model d) LSTM many-to-many model.

2.2 Dolžina vhoda pri učenju

Z optimalnimi velikostmi skritih stanj sedaj poiščemo optimalno dolžino vhoda k . Pri many-to-many metodi napovemo tudi k novih stanj. Kot je bilo omenjeno v navodilu, funkcija izgube ni dobra metrika za dolžino vhoda, kar je tudi razvidno iz slike 3. Funkcija izgube se manjša z manjšanjem k , saj je najmanjša napaka,

če vedno napovemo samo naslednje stanje. V ta namen bom za manj-to-many modele vzel $k = 10$ in bom po določanju optimalnih parametrov prek metrike $E(t)$ (enačba 1) poiskal optimalne k . Pričakovali bi, da večji k da boljše rezultate, ampak pri many-to-many modelih so vsi k primerljivega velikostnega reda. Izbrani optimalni k :

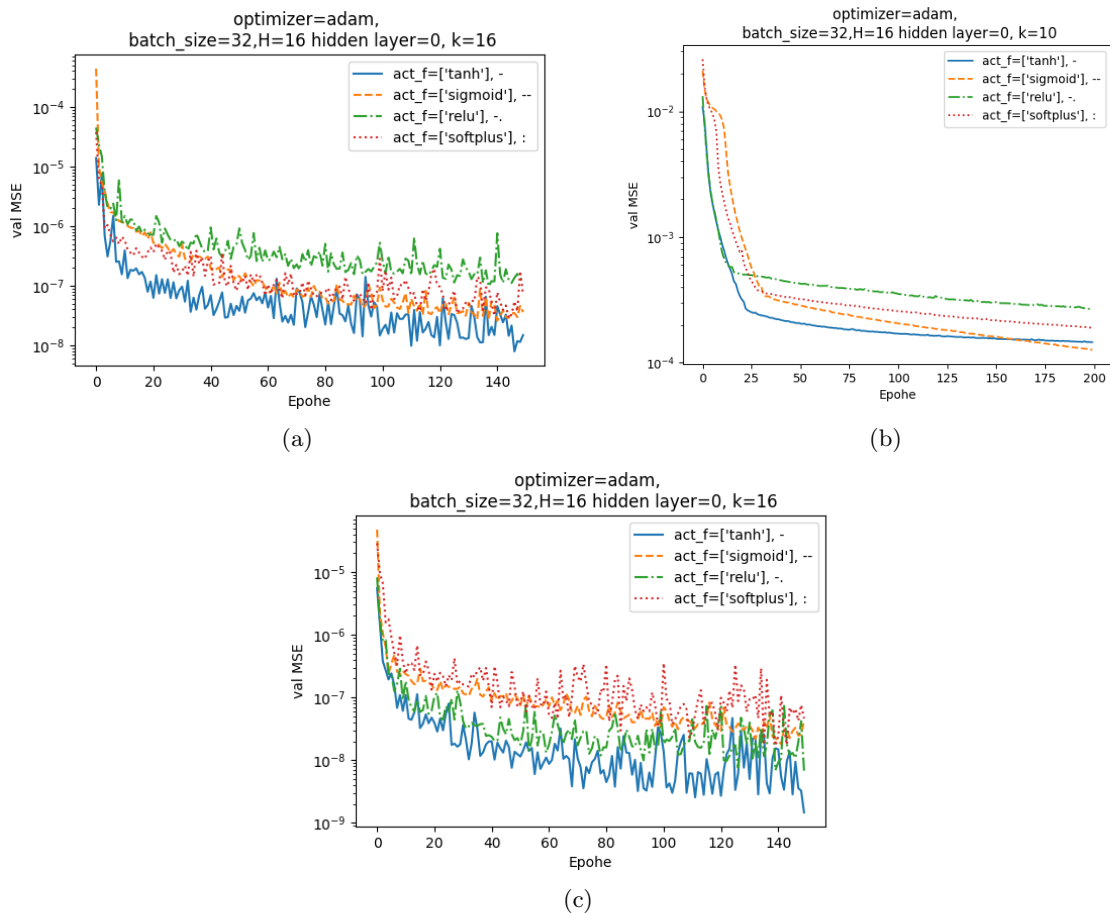
- RNN mto: $k = 32$
- RNN mtm: $k = 10$
- LSTM mto: $k = 32$
- LSTM mtm: $k = 10$



Slika 3: Validacijska funkcija izgube dolžinah vhoda pri učenju- k . a) RNN many-to-one model. b) RNN many-to-many model. c) LSTM many-to-one model d) LSTM many-to-many model.

2.3 Aktivacijska funkcija

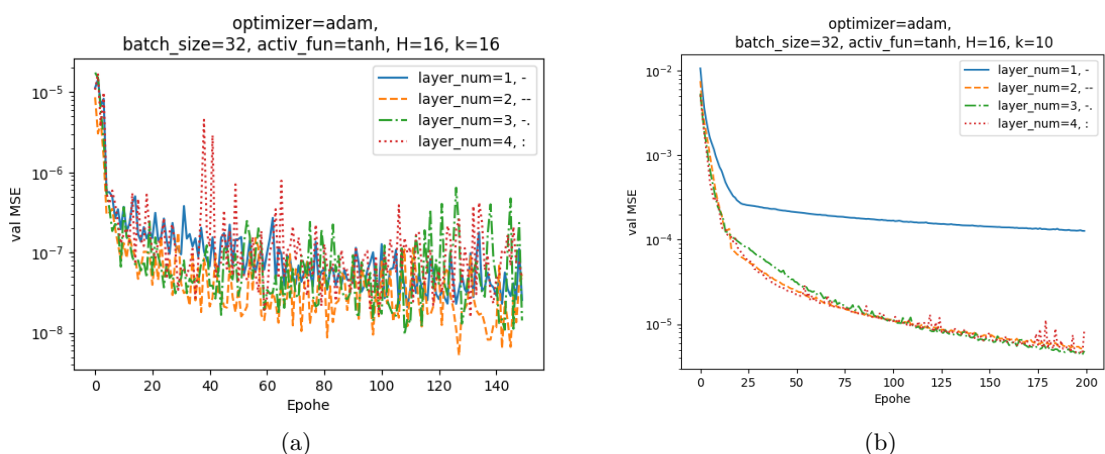
Funkcijo izgube pri različnih aktivacijskih funkcijah prikazuje slika 4. Predpostavka, da lahko uporabimo enako aktivacijsko funkcijo pri *mto* in *mtm* je smiselna, do določenega časa učenja. Za daljše čase pa se izkaže, da to ni nujno res (slika 4b). Pri RNN many-to-one je *sigmoid* aktivacijska funkcija boljša od *tanh* za dovolj dolge čase učenja. V našem primeru se modeli ne učijo toliko časa oz. bi prišlo ob takih časih do prepreleganja in bi dobili slabše rezultate. Za naše potrebe je optimalna aktivacijska funkcija *tanh*.



Slika 4: Validacijska funkcija izgube pri različnih aktivacijskih funkcijah-*act_f*. a) RNN many-to-one model. b) RNN many-to-many model. c) LSTM many-to-one model

2.4 Število skritih plasti

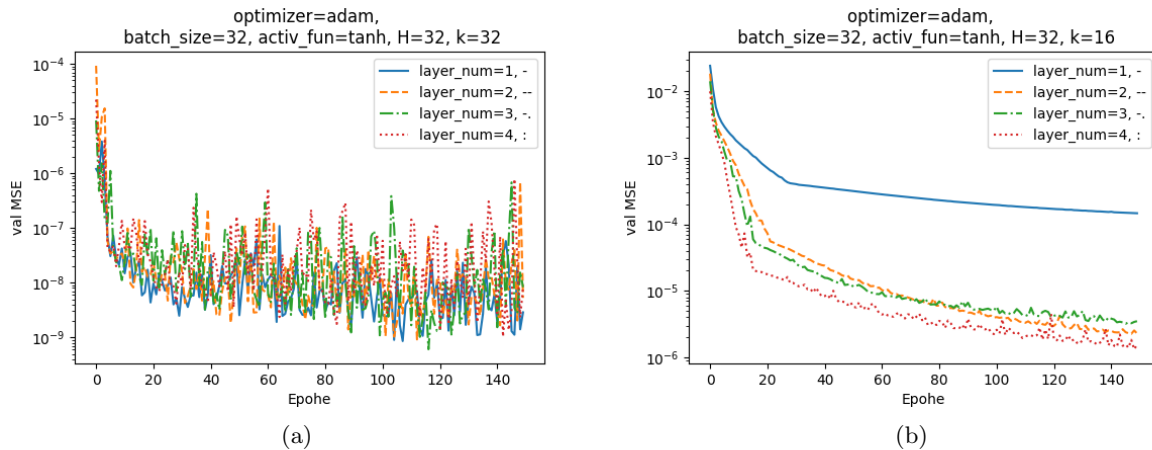
Kot zadnji hiperparameter je potrebno določiti optimalno število plasti mreže (slika 5). Parameter *num_layer* označuje število plasti modela brez zadnje plasti. Zadnja plast je odvisna od modela *mtm* ali *mtm*.



Slika 5: Validacijska funkcija izgube pri različnem številu plasti-*layer_num* (brez upoštevanja zadnje plasti). a) RNN many-to-one model. b) RNN many-to-many model.

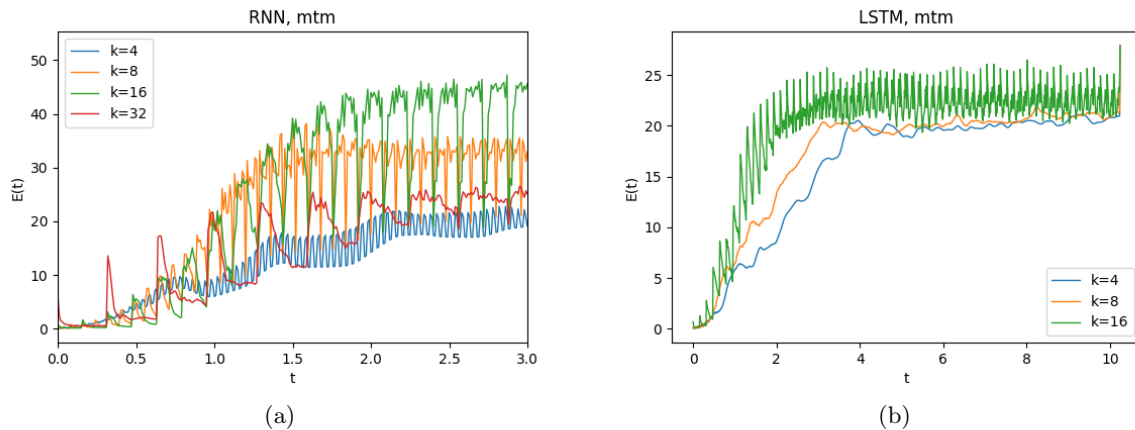
Pri RNN *mtm* sta optimalni 2 plasti + zadnja plast, pri RNN *mtm* modelu pa tri plasti + zadnja plast. Pri LSTM *mtm* sem vzel za optimalno dve plasti + zadnja plast in za LSTM *mtm* štiri plasti + zadnja plast. Pri določanju optimalnega števila plasti in aktivacijskih funkcij lahko opazimo, da doseže nižjo vrednost funkcije izgube LSTM mreža, kar mogoče implicira na to, da bo LSTM mreža boljše napovedovala dinamiko

Lorentzovega '63 sistema od RNN mreže. Sedaj imamo določene skoraj vse optimalne hiperparametre. Potrebno je še bolje določiti k pri many-to-many modelu. V ta namen sem pogledal metriko $E(t)$ teh modelov pri različnih k (slika 7).



Slika 6: Validacijska funkcija izgube pri različnem številu plasti- $layer_num$ (brez upoštevanja zadnje plasti). a) LSTM many-to-one model. b) LSTM many-to-many model.

Za LSTM dobimo boljšo napoved pri majhnih k . Mogoča razlaga je ta, da je model hidrodinamski in upošteva samo lokalne spremembe polj. Časovne korelacije očitno pojemajo hitro in prevelik k pokvari napoved. Zanimivo je tudi to, da se pri večjih k pojavljajo večje oscilacije (slika 7a). Pri RNN modelu, $k = 32$ se zdi, da je ob tako velikem k , prva napoved najbolj zgrešena in nato se napoved izboljša (valovi na grafu). Pri RNN mtm nisem bil ravno prepričan kateri je optimalni k . Odločil sem se za $k = 32$, saj je v povprečju najmanjši (tudi glede na poznejše čase). Mogoče bi bilo bolje če bi vzel $k = 4$, saj ima manjši povprečen naklon.



Slika 7: Povprečno odstopanje napovedi od pravih vrednosti $E(t)$ pri časovnem razvoju sistema pri različnih dolžinah vhoda- k . a) RNN many-to-many model. b) LSTM many-to-many model.

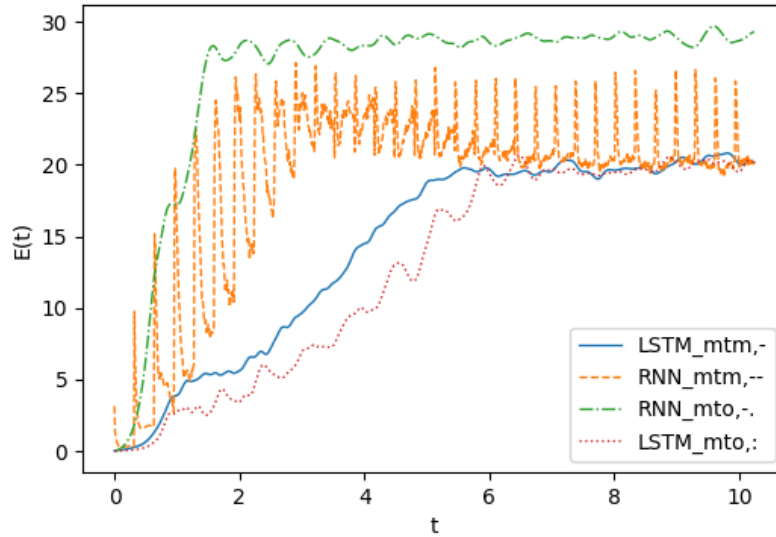
2.5 Najboljši model

Sedaj, ko imamo optimalne parametre, lahko pogledamo metriko $E(t)$ (slika 8). Pri RNN mto se za $k = 32$ pojavijo velike oscilacije in boljšo napoved dobimo če je k manjši, recimo $k = 4$. Optimalni set parametrov je tako:

- RNN mto: $H=16$, $k=4$, tanh, 2 plati + zadnja plast
- RNN mtm: $H=16$, $k=32$, tanh, 3 plati + zadnja plast
- LSTM mto: $H=16$, $k=32$, tanh, 2 plati + zadnja plast

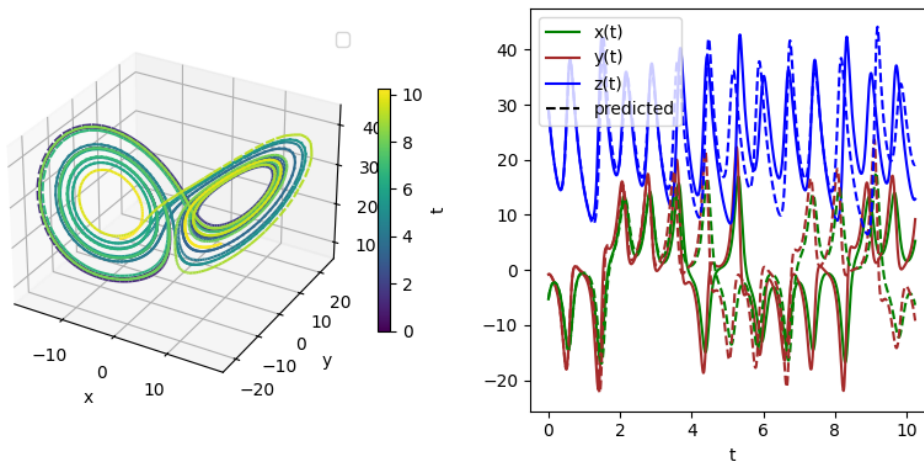
- LSTM mtm: $H=32$, $k=4$, tanh, 4 plasti + zadnja plast

Iz slike 8 je razvidno, da je LSTM mreža veliko boljše za napoved Lorentzovega '63 sistema. Napoved bi mogoče lahko izboljšal pri RNN mtm če bi vzel $k = 4$. RNN mreže precej slabo napovejo dinamiko in sicer jo napovejo samo za čas $t \approx 1$. Nasprotno pa LSTM precej dobro napovejo dinamiko, mto model celo nekoliko boljše od mtm modela.



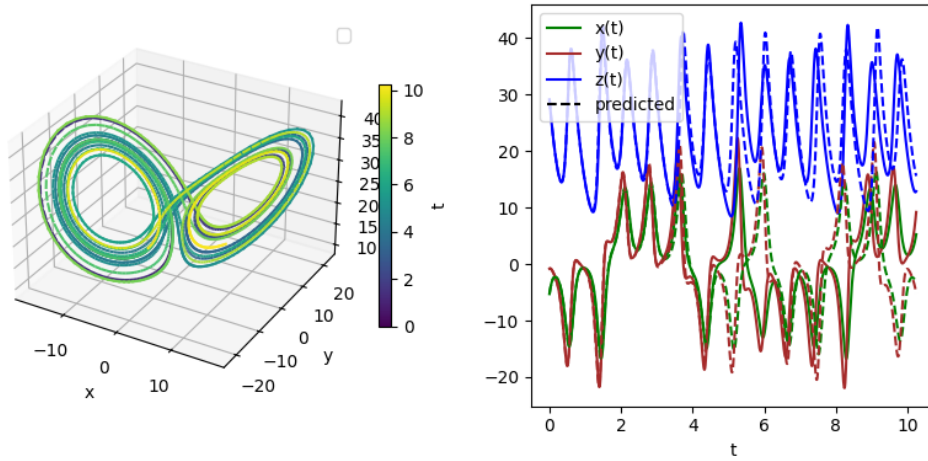
Slika 8: Povprečno odstopanje napovedi od pravih vrednosti $E(t)$ pri časovnem razvoju sistema za različne modele pri optimalnih parametrih.

Napoved v realnem prostoru najboljših modelov (LSTM mtm in mto) prikazujeta sliki 9 in 10. Napovedi sta si precej podobni, okrog časa $t \approx 4$ (mtm) oz $t \approx 4.5$ (mto) se napoved zelo dobro prilega pravih vrednostim, potem pa mreža še vedno po obliki dobro napove dinamiko, a zgreši na katerem krilu atraktorja je sistem v danem času.



(a)

Slika 9: a) Napovedan časovni razvoj (črtkana črta) LSTM many-to-many modela in pravi časovni razvoj (polna črta) Lorentzovega '63 sistema.

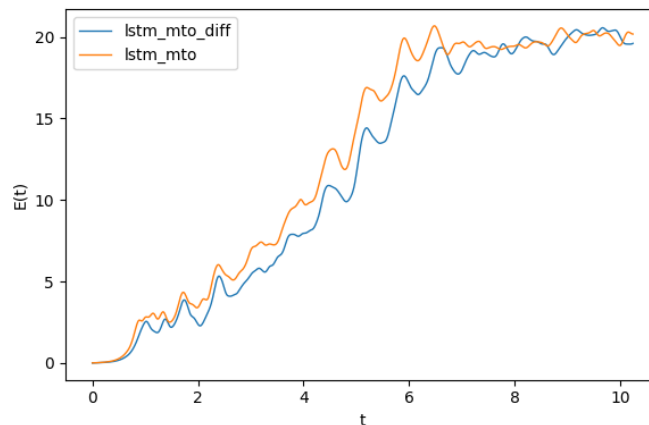


(a)

Slika 10: a) Napovedan časovni razvoj (črtkana črta) LSTM many-to-one modela in pravi časovni razvoj (polna črta) Lorentzovega '63 sistema.

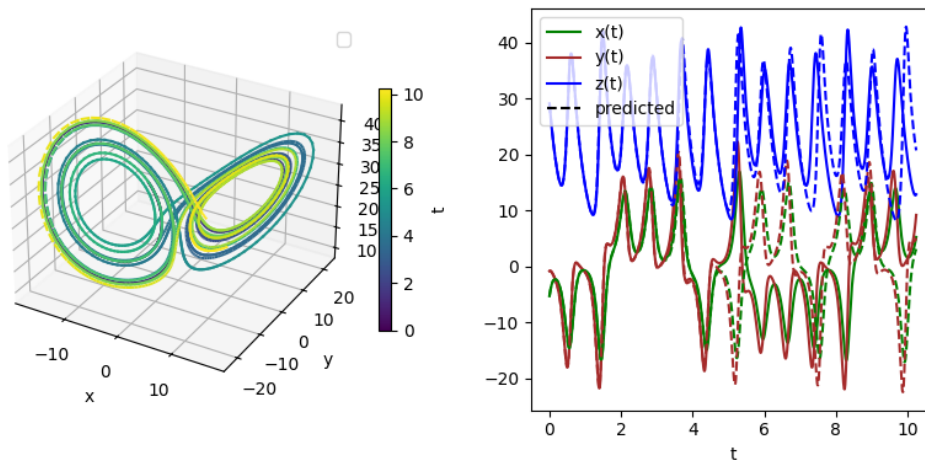
2.6 Napovedovanje z diferencami

Dinamiko z nevronske mreže je mogoče napovedovati tudi preko napovedovanja spremembe položaja glede na prejšnji korak (z diferencami). Metoda še nekoliko izboljša rezultat napovedi. Vzel sem najboljšo mrežo - LSTM mto. Pri napovedi je bilo potrebno popraviti normalizacijo. Mreža dobiva na vhod k predhodnih vektorjev in napove razliko med naslednjim vektorjem in vektorjem v katerem je trenutno sistem. Vhodni podatki so normirani in zato so normirane tudi izhodne difference. Ker je koda za normalizacijo izhodnih vektorjev že znana, sem trenutnemu normiranemu vektorju prištel izračunano izhodno diferenco in šele nato inverzno normiral vsoto. Na ta način sem dobil izhodne vektorje. Povprečno odstopanje od pravih vrednosti prikazuje slika 11. Napoved v realnem prostoru prikazuje slika 12. Mreža natančno napove dinamiko do časa $t \approx 5$, nato zgreši krilo atraktorja, vendar še vedno precej dobro napove dinamiko.



(a)

Slika 11: a) Povprečno odstopanje napovedi optimalnega modela LSTM mto od pravega časovnega razvoja sistema z in brez uporabe diferenc.



(a)

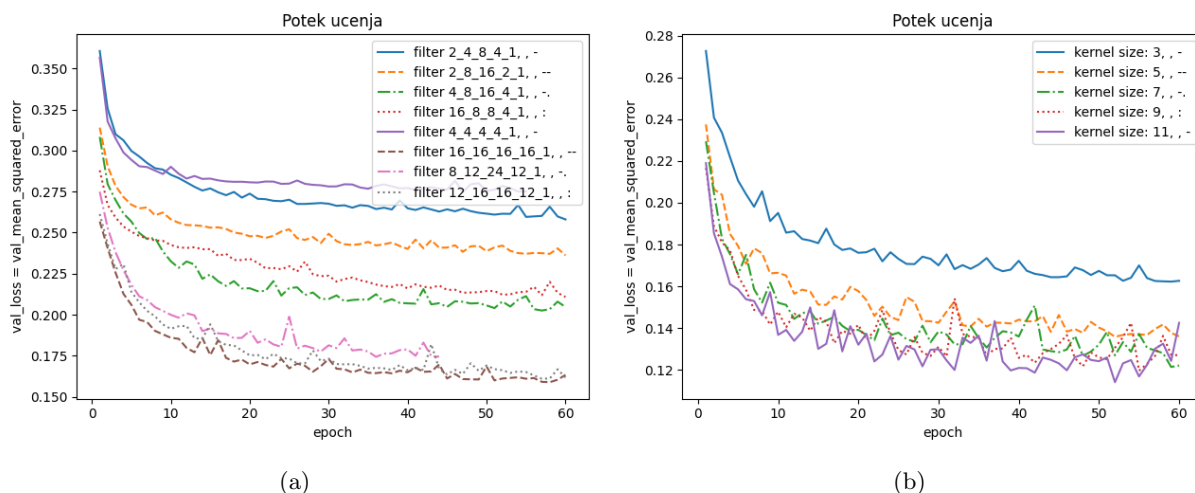
Slika 12: a) Napovedan časovni razvoj (črtkana črta) LSTM many-to-one modela z diferencami in pravi časovni razvoj (polna črta) Lorentzovega '63 sistema.

3 Meteorološka napoved

V drugem delu naloge je bilo potrebno poiskati najboljšo konvolucijsko nevronske mrežo (CNN) za opis potenciala Z500 pri prostorski resoluciji 4° in časovnim korakom 1 dan. Za začetno konvolucijsko mrežo sem vzel mrežo iz navodil naloge. Uspešnost mrež bomo ocenjevali na podlagi metrik RMSE in ACC. Napoved mreže je uspešna, če je večja od persistence (boljša od trenutnega stanja) in klimatologije (boljša od dolgoletnega povprečja tistega dne). Najprej poiščemo nekaj optimalnih hiperparametrov.

3.1 Število filtrov, dimenzija konvolucijskih jeder, dodajanje plasti,...

Najprej sem poiskal optimalno število filtrov (slika 13). V našem primeru velja, da večje število filtrov bolj minimizira funkcijo izgube. Hkrati je tudi pomembna razporeditev velikosti filtrov, neugodna je razporeditev filtrov, kjer velikost pada s plastmi. Veliko boljša je simetrična razporeditev filtrov po plasteh z maksimumom na sredini plasti, kot pa enakomerna razporeditev po plasteh (slika 13a).



(a)

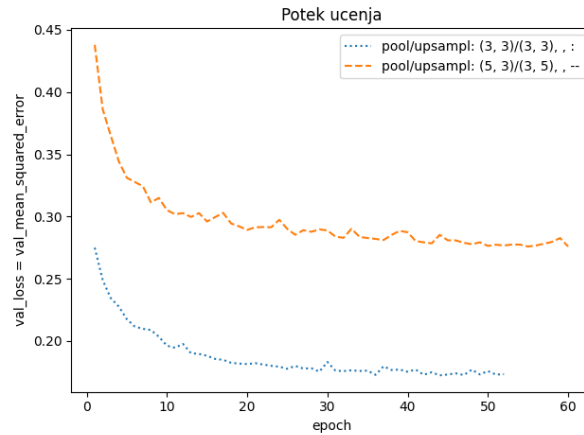
(b)

Slika 13: a) Funkcija izgube pri različnem številu filtrov v plasteh. Oznaka *filter x_y_z* označuje x filtrov v prvi plast, y filtrov v drugi plasti in z filtrov v tretji plasti. b) Funkcija izgube pri različni velikosti jedra (kernel size), kjer ima prva konvolucijska plast večje jedro za $kernel\ size + 2$.

Pri odvisnosti od velikost jedra (slika 13b), večje jedro bolj minimizira funkcijo izgube. Razlike so bolj opazne pri majhnih jedrih, pri velikih jedrih pa se z večanjem števila jeder, ne bistveno izboljša model.

Kljub zgornji odvisnosti je potrebno biti pri velikosti jedra previden, saj večje jedro bolj povpreči lokalno polje in izgubimo lahko za nas morebiti pomembne podrobnosti. V tem primeru se izkaže, da večje jedro deluje boljše.

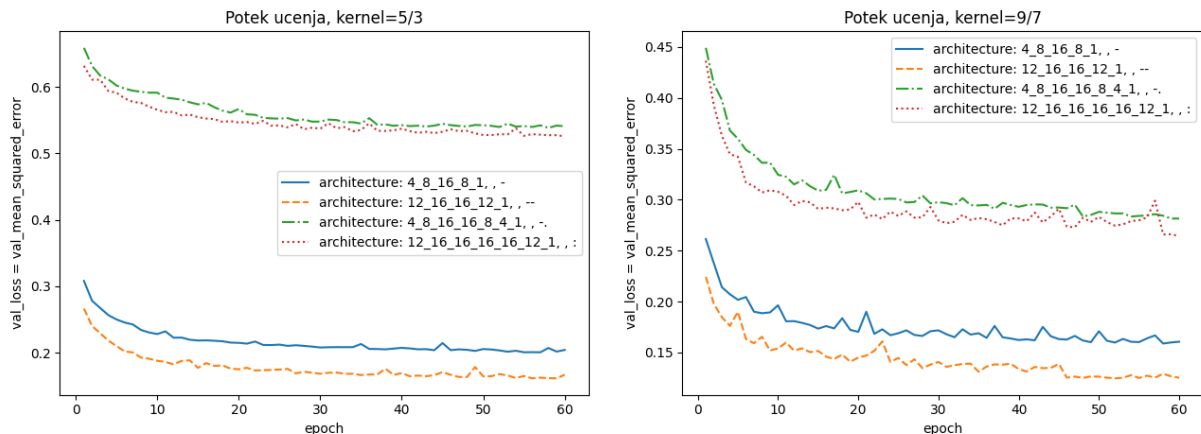
Za osnovno mrežo sem pogledal kako na funkcijo izgube vpliva sprememba dimenzije pooling in upsampling plasti (slika 14). Večji pooling je slabši za dani primer. Pri dodajanju plasti ali poolingu je potrebno paziti na ohranjanje dimenzij po rekonstrukciji, zato niso možne poljubne vrednosti poolinga in upsamplinga. Za 7 plasti obstaja samo en način za tvorbo mreže (do identitete in permutacij natančno) in sicer (3,3,5/5,3,3), kjer številka pred / označuje dimenzijo poolinga v dani plasti (štetje plasti od leve proti desni) in številka / označuje dimenzijo upsamplinga.



(a)

Slika 14: Funkcija izgube pri različnih poolingih in upsamplingih.

Odvisnost od dodajanja plasti prikazuje slika 15, kjer sem pogledal odvisnost za 2 različni velikosti jeter (5/3 in 9/7). Z večanjem plasti nismo izboljšali nevronske mreže. Dobro bi bilo pogledati rezultate za manjšo mrežo vendar sem pozabil na to, da je lahko osnovna mreža še manjša.



(a)

(b)

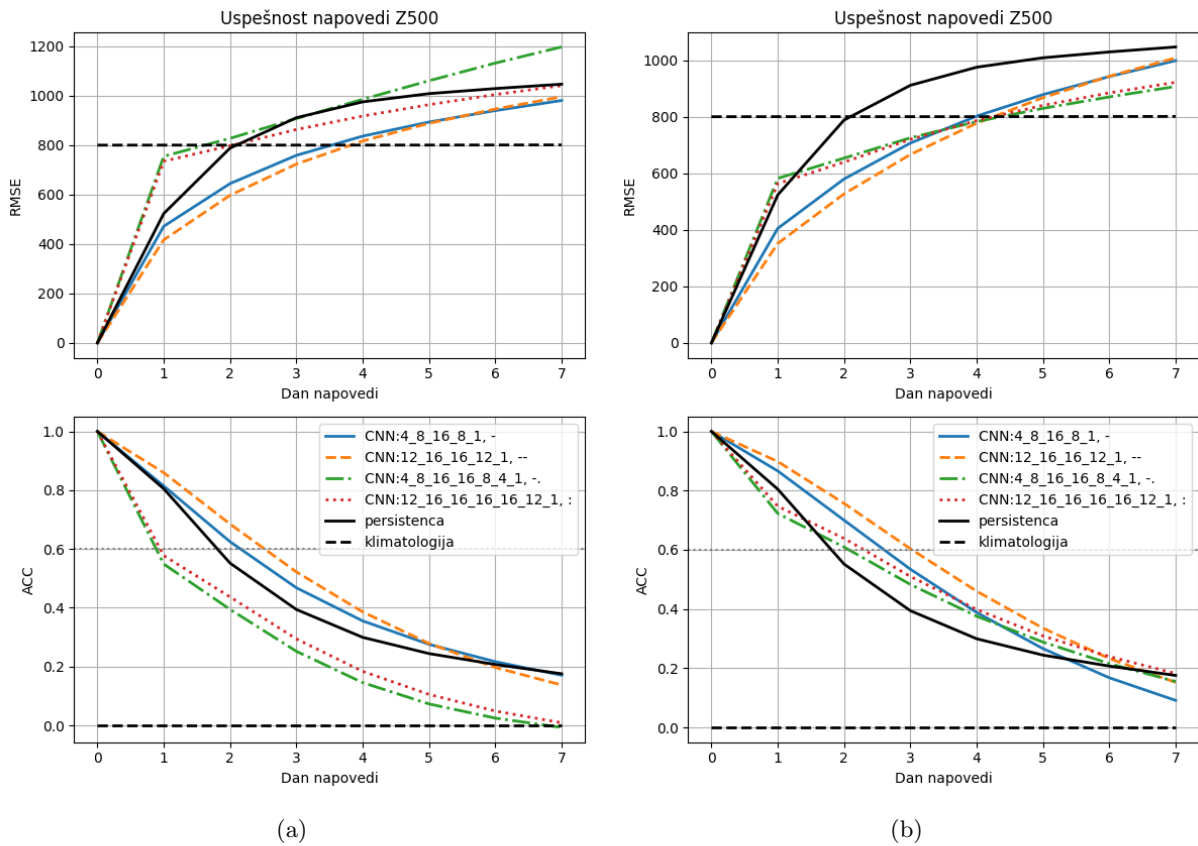
Slika 15: Funkcija izgube pri različnih arhitekturah CNN. a) Pri velikosti jeter 5/3, kjer 5 označuje velikost prvega jedra in 3 velikost vseh ostalih. b) Pri velikosti jeter 9/7, kjer 9 označuje velikost prvega jedra in 7 velikost vseh ostalih.

3.2 Optimalni model

Optimalni parametri so:

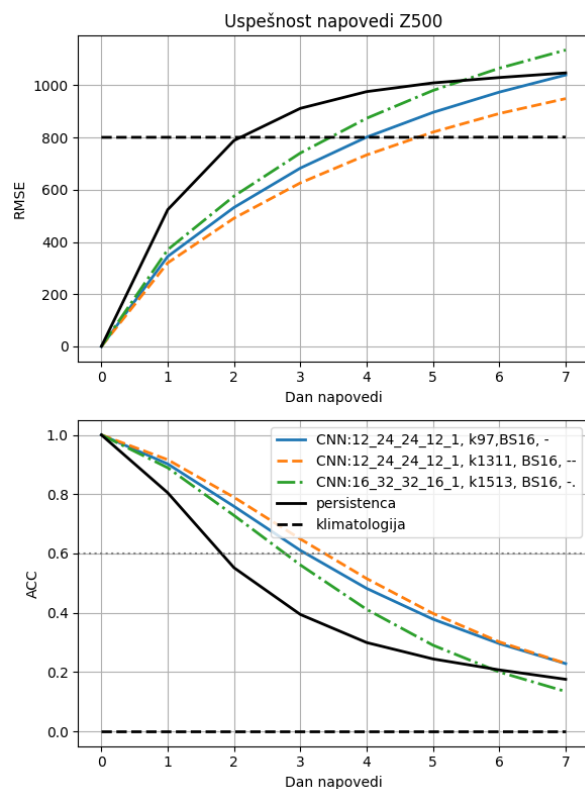
- pooling/upsampling: 3
- velikost jeder: 9/7 ali 5/3

Za modele iz predhodnih grafov sem pogledal uspešnost napovedi po metrikah RMSE in ACC (slika 16). Modeli s sedmimi plastmi so slabi. Klimatologijo in persistenco premagata le modela s petimi plastmi. Model z več filtri je bolj uspešen, kar je pričakovano. Pomembna opazka je ta, da velikost jedra precej vpliva na uspešnost napovedi. Modeli z večjo velikostjo jeder nam dajo precej boljšo napoved (slika 16b). Iz tega lahko sklepam, da sta najpomembnejša parametra mreže za dani primer število filtrov in velikost jeder. Mogoče ima velik vpliv tudi dimenzija poolinga oz. upsamplinga, a je zaradi ohranitve dimenzij možnih manj kombinacij.



Slika 16: a) Uspešnost napovedi modelov pri velikosti jeder 5/3. b) Uspešnost napovedi modelov pri velikosti jeder 9/7.

Da bi dobil še boljši model sem povečal število filtrov in ob tem spreminjal velikost jeder (slika 17). Izkaže se, da modeli z več filtri potrebujejo za boljše napovedovanje tudi večje jedro. Za optimalni model sem dobil model CNN: 12_24_24_24_12_1, 13/11, kjer dana števila, ločena s podčrtajem, označujejo število filtrov v plasteh. Štetje plasti se začne na levi. Optimalni model boljše napove potencial Z500 od klimatologije in persistence za 3 dni v prihodnost, nato pa vrednost ACC modela pade pod 0.6. Ko se to zgodi, napoved ni več zanesljiva.



(a)

Slika 17: Uspešnost napovedi modelov pri različnih številih jeter v plasti (ločena s podčrtaji) in različnimi velikostmi jeter, kjer oznaka x/y, označuje velikost x prvega jedra in velikost y vseh ostalih. BS označuje batch velikost.

3.3 Napoved vremena

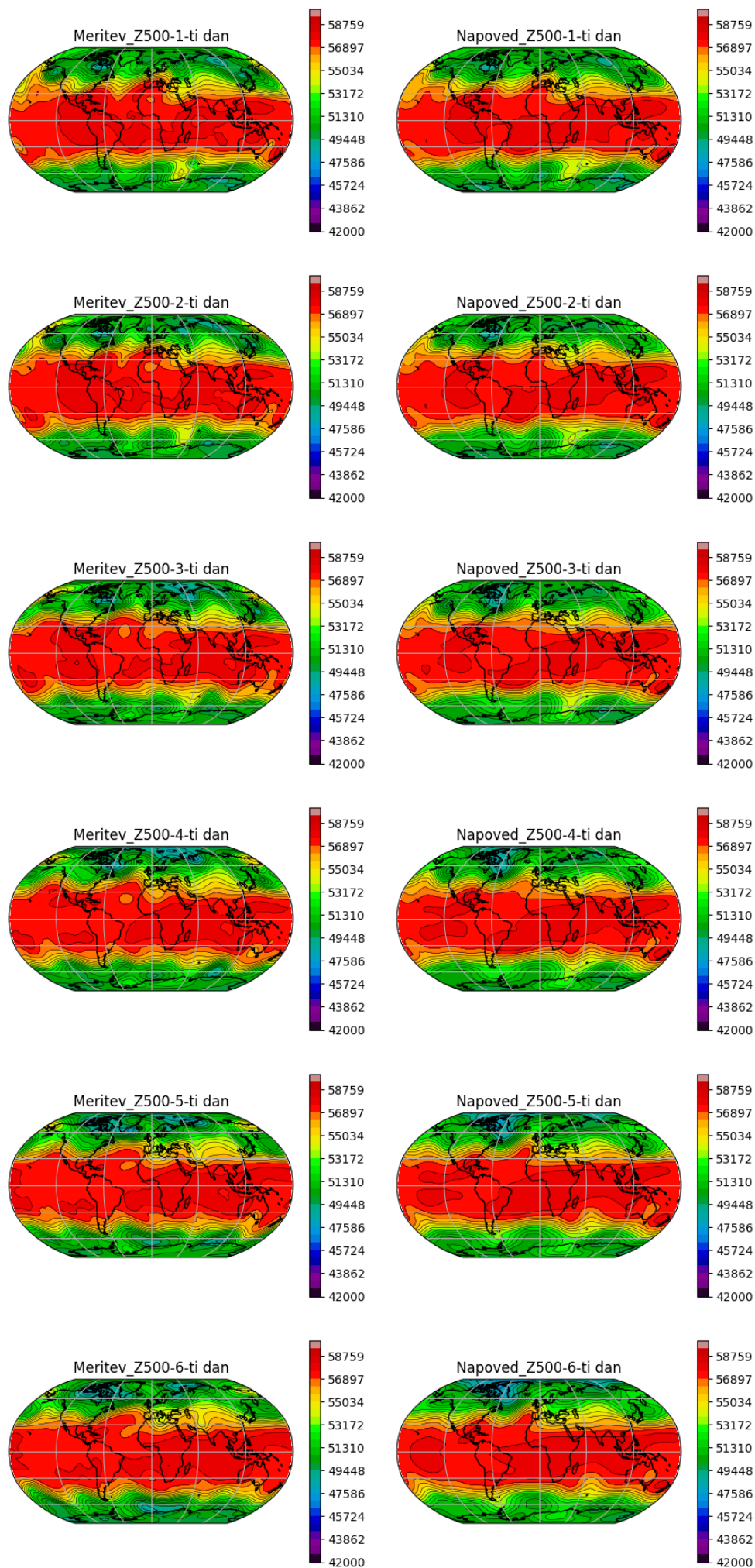
Z optimalnim modelom sedaj napovemo vreme oz. Z500. Kot testno množico vzamemo potencial Z500 iz leta 2022. Meritve in napovedi so prikazane na sliki 18. Odstopanja napovedi od meritev so prikazana na sliki 19. Napoved optimalnega modela relativno dobro napove vrednost potenciala Z500 za nekaj dni (2-3). Odstopanja se večajo s časom (slika 19). Največja odstopanja so v območju Rossbyjevih valov, saj je tam tudi dinamika najbolj spremenljiva.

3.4 Vpliv robnih pogojev, kopnega in morja ter dodajanja plasti

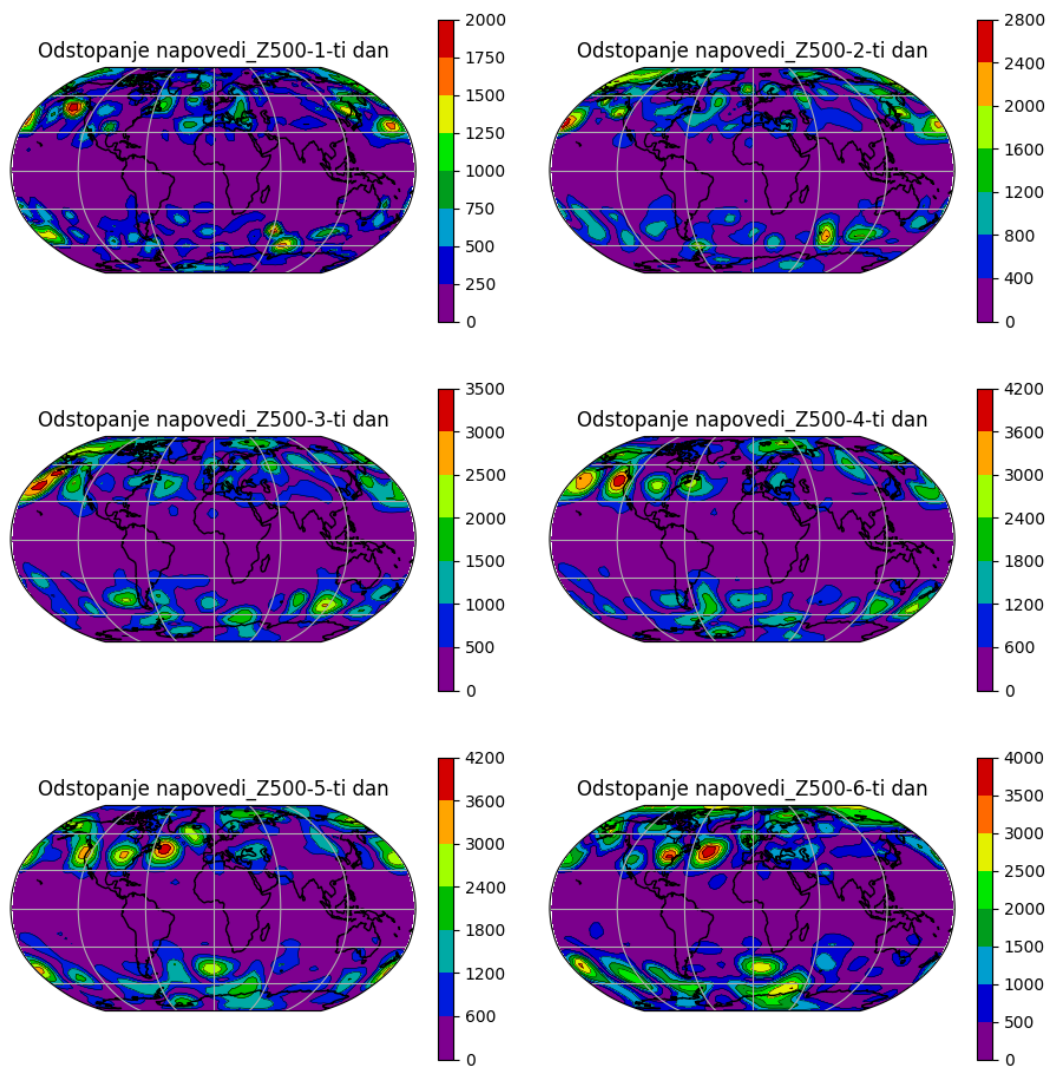
Pogledal sem tudi vpliv morja in kopnega, robnih pogojev, dodajanja plasti brez dodanega poolinga ali unsamplinga in dodajanja meritev izpred leta 1970 v učno množico (slika 20). Te spremembe niso izboljšale napovedi optimalnega modela (za nekatere je očitno). Je pa zanimivo pogledati odstopanja napovedi od meritev v primeru ko imamo na datumski meji periodične robne pogoje in na polih dodane ničle (zero padding). Odstopanja prikazuje slika 21. Pričakovano se večja odstopanja pojavijo ob polih.

4 Zaključek

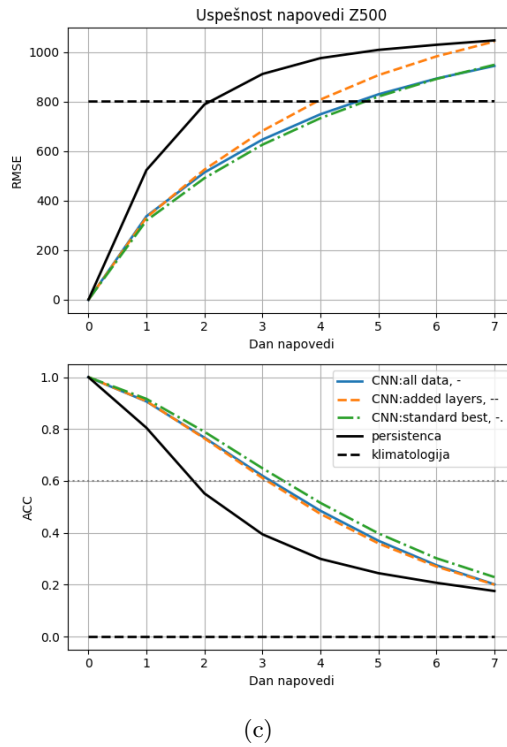
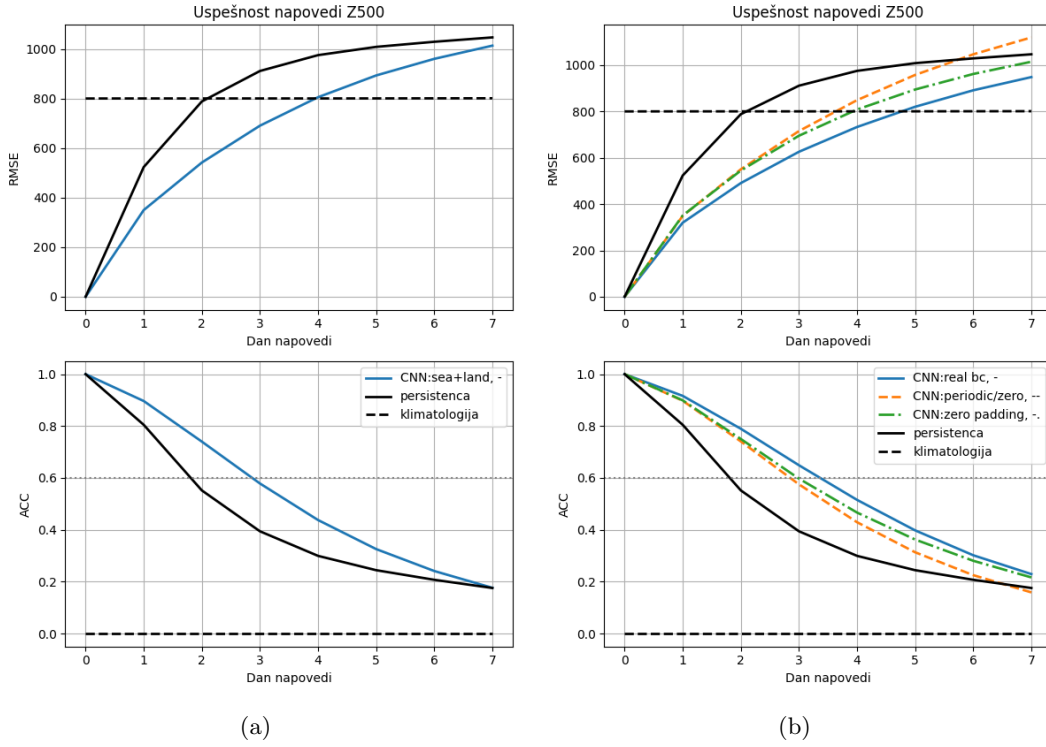
Naloga je obravnavala integracijo dinamičnih sistemov z rekurentnimi in konvolucijskimi nevronske mrežami. Obe nevronske mreži sta precej uporabni za napovedovanje dinamike dinamičnih sistemov. Največji izziv predstavlja določanje optimalnih hiperparametrov za dano mrežo.



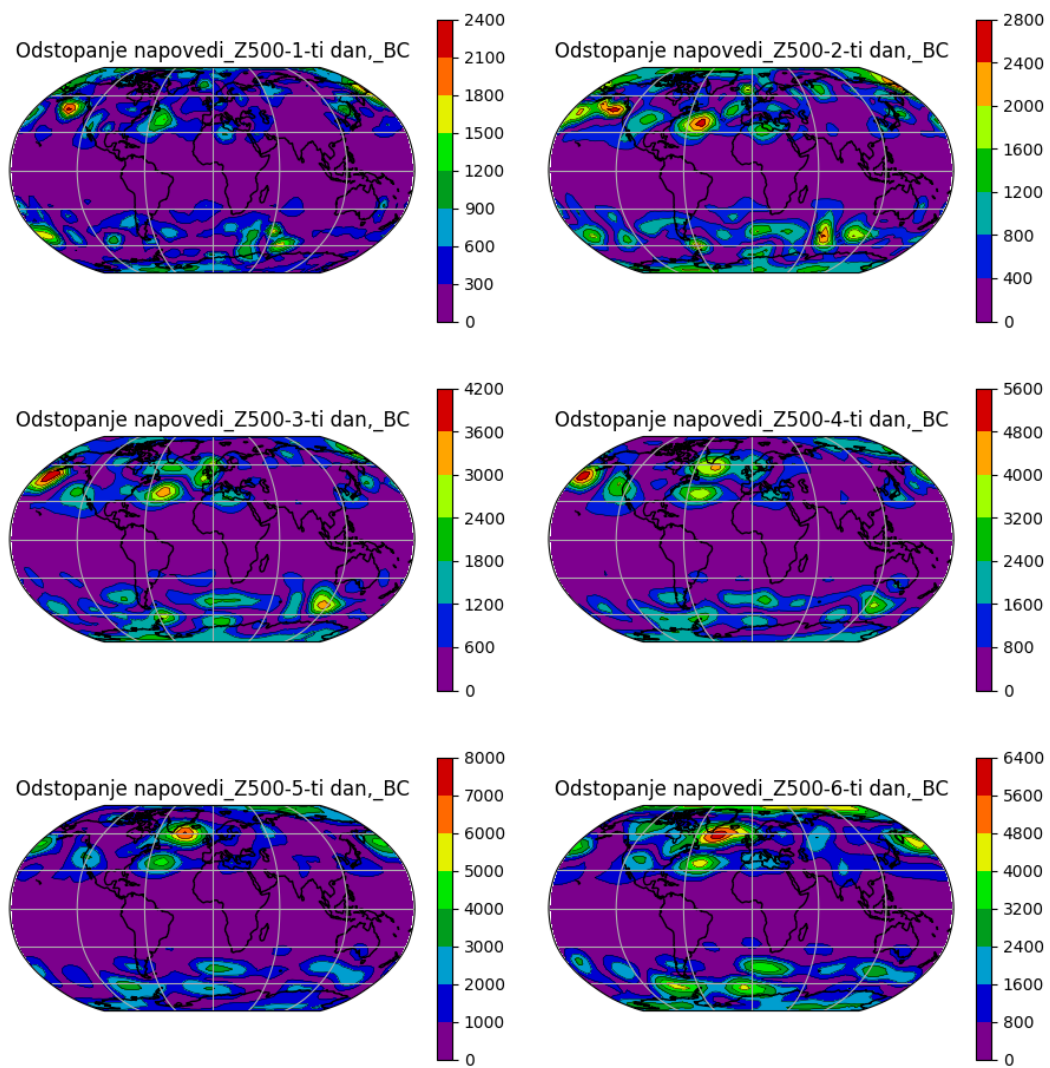
Slika 18: Meritev (desno) in napoved (levo) potenciala Z500 za prvih 6 dni leta 2022. Barvna skala predstavlja vrednost potenciala Z500.



Slika 19: Odstopanja napovedi od meritev potenciala Z500 za prvih 6 dni leta 2022.



Slika 20: a) Uspešnost napovedi optimalnega modela z dodanim poljem morja in kopnega. b) Uspešnost optimalnega modela s pravimi robnimi pogoji (real bc), periodičnimi na datumski meji in zero paddingom na polih (periodic/zero) in zero paddingom povsod (zero padding). c) Uspešnost napovedi optimalnega modela z dodanimi podatki izpred leta 1970 (all data), samo optimalni model (standard best) in optimalni model z dodanimi konvolucijskima plastema brez poolinga/upsamplinga velikosti 24 na pooling in unsampling strani.



Slika 21: Odstopanja napovedi od meritev potenciala Z500 za prvih 6 dni leta 2022 pri napačnih robnih pogojih.