

Napovedovanje vremena s konvolucijskimi nevronskimi mrežami

Uroš Perkan¹, Gregor Skok¹, Žiga Zaplotnik^{2,1}

Povzetek

Moderno napovedovanje vremena temelji na numeričnih vremenskih modelih, ki časovno evolucijo atmosferskih polj izračunajo na podlagi osnovnih fizikalnih zakonov. V pričujočem članku opišemo model strojnega učenja za globalno srednjeročno napovedovanje vremena ConvCastNet, ki smo ga razvili v okviru magistrske naloge. ConvCastNet temelji na konvolucijskih nevronskih mrežah in napoved izvaja na 3-stopinjski ekviangularni prostorski mreži za 10 atmosferskih, morskih in kopenskih spremenljivk.

ConvCastNet se napovedovanja nauči tako, da pri učenju upošteva cenilko (t. i. funkcijo izgube), ki meri razliko med kratkoročno napovedjo in istočasnim približkom resničnega stanja ozračja, ki ga določimo iz rekonstrukcij preteklega vremena, t. i. reanaliz, ERA5. V model implementiramo nov način obravnave robov atmosferskih polj in polj latentnih značilnk, tako da je upoštevana Zemljina sferična geometrija. Začetni pogoj za modelsko napoved so polja reanaliz, napoved pa izvedemo z avto regresivnim korakanjem s časovnim korakom 1 dan. ConvCastNet napoveduje dnevno povprečje atmosferskih spremenljivk.

Vremenska napoved z modelom ConvCastNet omogoča natančno napovedovanje razvoja sinoptičnih vremenskih valov. Pri napovedovanju polja geopotenciala na 500 hPa ploskvi ConvCastNet tako doseže 7,8 dni z vrednostjo koeficienta korelacij anomalij (ACC) večjo od 0,6, s čimer se pri tej metriki približa rezultatom najnovjših modelov strojnega učenja in fizikalnemu determinističnemu visokoresolucijskemu modelu (HRES) Evropskega centra za srednjeročne vremenske napovedi (ECMWF).

Model ConvCastNet lahko več dni vnaprej uspešno napove položaje sinoptičnih valov in baričnih tvorb ter uspešno sledi trajektorijam tropskih ciklonov. Zaradi nizke ločljivosti in dnevnega povprečenja atmosferskih polj pa ni sposoben izvedbe koristne napovedi intenzitete tropskih ciklonov, lokalnih ekstremnih padavinskih dogodkov in ostalih lokalnih vremenskih ekstremov. Analiza porazdelitve napak modela v letu 2022 pokaže, da je absolutna napaka vremenskih napovedi največja na baroklinih območjih v visokih geografskih širinah, normiranje napake z naravno variabilnostjo vremena pa pokaže, da te glede na tipično hitrost spreminjanja vremena najhitreje naraščajo v tropih.

Ključne besede: strojno učenje, napovedovanje vremena, konvolucijska nevronska mreža, konvolucijskih kodirnik, povprečna kvadratna napaka, koeficient korelacij anomalij, tropski ciklon, absolutna napaka

Keywords: machine learnig, weather forecasting, convolutional neural network, convolutional encoder-decoder, root mean squared error, anomaly correlation coefficient, tropical cyclones, absolute error

Uvod

Vremenske razmere vplivajo na kopico različnih človekovih dejavnosti, ki neposredno ali posredno vplivajo na blagostanje človeške civilizacije. Neugodne vremenske razmere lahko otežijo kmetovanje, onemogočijo uspešno pridelavo hrane, povzročijo težave v

¹ Univerza v Ljubljani, Fakulteta za matematiko in fiziko, Jadranska ulica 19, Ljubljana

² Evropski center za srednjeročne vremenske napovedi, Robert-Schuman-Platz 3, Bonn, Nemčija

prometu in pridobivanju električne energije, v najslabšem primeru pa povzročijo tudi izgubo premoženja in človeške žrtve. Zaradi njihovega izrazitega vpliva na blagostanje prebivalstva napovedovanje vremena predstavlja enega izmed ključnih virov za priprave na neugodne vremenske dogodke in preprečitev njihovih negativnih posledic.

Zasnove modernega napovedovanja vremena segajo v začetek 20. stoletja, ko je L. F. Richardson pokazal, da lahko sistem parcialnih diferencialnih enačb, ki opisujejo atmosfero (gibalne enačbe, kontinuitetna enačba, termodinamska enačba in plinska enačba) diskretiziramo in rešujemo na mreži točk (Richardson & Lynch, 1922). Prvo numerično vremensko napoved je približno tri desetletja kasneje izvedel J. G. Charney (Charney et al., 1950), od takrat pa je s hitrim napredkom v računski zmogljivosti superračunalnikov, izpopolnjevanjem numeričnih shem, izboljšavo algoritmov za pridobivanje kakovostnega približka začetnega stanja ozračja (t. i. začetni pogoj) in večanjem števila in kakovosti meritev vremenska napoved numeričnih vremenskih modelov dosegla velik napredek v natančnosti na kratkoročnih, srednjeročnih, sezonskih in klimatoloških časovnih skalah (Bauer, 2015).

Dinamika atmosfere je zaradi nelinearnih členov v prognostičnih parcialnih diferencialnih enačbah kaotična, zato je kakovost vremenske napovedi omejena z velikostjo napake v začetem pogoj. Poleg tega napako dodatno večajo izbira numerične metode, parametrizacija podmrežnih procesov in netočnost robnih pogojev. Z nadomestitvijo fizikalnih modelov s statističnimi metodami se pri napovedovanju vremena znebimo potrebe po obravnavi raznoraznih parametrizacij, ki so v fizikalnih modelih dodane, kjer neposredna simulacija procesov ni mogoča (Rasp et al., 2018). Modeli strojnega učenja se lahko povezav med atmosferskimi spremenljivkami in napoved nadaljnjega razvoja vremena učijo na podlagi podatkov o stanju vremena v preteklosti. Tako se znebimo potrebe po izvajanju aproksimacij v fizikalnih enačbah in model napovedovanja učimo tako, da minimizira napako med napovedanim stanjem atmosfere in dejanskim stanjem atmosfere, pridobljenim v procesu asimilacije meritev (npr. Courtier et al., 1994).

Preizkušenih je bilo že veliko različnih arhitektur nevronske mreže. Dueben in Bauer, 2018 sta uporabila preprosto gosto povezano nevronske mreže, Weyn et al., 2019 in Weyn et al., 2021 pa so uporabili konvolucijske nevronske mreže (CNN) in z njimi zgradili konvolucijski kodirnik, imenovan DLWP (angl. *Deep Learning Weather Prediction*). Pokazali so, da je DLWP stabilen in sposoben izvedbe daljših vremenskih napovedih. Izvedli so tako deterministično vremensko napoved, kjer se za vsako lokacijo napove le po eno vrednost vsake izmed modelskih spremenljivk, kot ansambelsko vremensko napoved, kjer se za vsako lokacijo napove verjetnostno porazdelitev vrednosti vsake izmed spremenljivk. Pri slednji so uporabili podatke 10 perturbiranih začetnih pogojev iz reanalize ERA5 (Hersbach et al., 2020) in 32 različno naučenih modelov ter tako dosegli 320 članov ansambla. Ansambelska napoved DLWP je hitra in energijsko učinkovita, vendar po kakovosti ne premaga ansambelske napovedi Evropskega centra za srednjeročne vremenske napovedi (ECMWF).

Dobre rezultate so pokazali modeli osnovani na grafovskih nevronske mrežah (GNN) (npr. Keisler, 2022; Lam et al., 2023; Rabier, 2023). Te delujejo na podatkih, ki so predstavljeni v obliki grafov, torej vsebujejo vozlišča in vmesne povezave. Vsako vozlišče vsebuje svoje podatke, npr. temperaturo, tlak, geopotencial itd., na različnih vertikalnih nivojih. Podatke najprej iz ekviangularne mreže preslikajo v t. i. R-krat zgoščeno ikozaedrično mrežo, v kateri so točke enakomerno porazdeljene po sferi. Nato izvedejo iterativno posodabljanje skritih stanj v vozliščih, na koncu pa skrita stanja dekodirajo nazaj na ekviangularno mrežo in s tem izvedejo en korak napovedi.

Poleg omenjenih CNN in GNN, so bile za napovedovanje preizkušene tudi različne oblike t. i. vision transformer nevronske mreže (npr. Pathak et al., 2022; Bi et al. 2023; Chen et al.

2023), nevronske operator SFNO (angl. *Spherical Fourier Neural Operators*) (Bonev, 2023) in kombinacija reševanja fizikalnih enačb z matematično diskretizacijo dinamičnega jedra in aproksimacijo ostalih fizikalnih procesov z nevronske mreže (Kochkov et al., 2023). Vision transformerji napoved izvajajo s t. i. mehanizmom pozornosti (angl. *attention-mechanism*), SFNO pa z nevronske operatorjem, ki vrednosti uteži določa z upoštevanjem sferične geometrije v Fourierovem prostoru. Vsi omenjeni pristopi na podlagi površinsko uteženih metrik korena povprečne kvadratne napake (angl. *root mean squared error*) (RMSE) in koeficienta korelacij anomalij (angl. *anomaly correlation coefficient*) (ACC) dosegajo podobno kakovost vremenske napovedi kot ECMWF-jev fizikalni visokoresolucijski deterministični model (HRES).

Dosedanje raziskave so pokazale, da so napovedi nekaterih nevronske mrež lahko boljše korelirane z dejanskim razvojem vremena in imajo manjšo povprečno napako kot napoved HRES (npr. Lam et al. 2023; Rabier, 2023; Kochkov et al., 2023). Po drugi strani imajo za razliko od HRES vsi omenjeni pristopi težave s spektralnim glajenjem napovednih polj (Rasp et al. 2023; Bonavita 2023). Posledično modeli strojnega učenja trenutno niso sposobni napovedovanja lokalnih ekstremnih padavinskih, vetrovnih in ostalih dogodkov na majhnih prostorskih skalah, hkrati pa množice njihovih napovedi ne moremo združevati v zanesljiv ansambel, ki bi zagotovil ustrezno oceno verjetnosti ekstremnih dogodkov.

V tem članku povzamemo glavne ugotovitve iz raziskovalne magistrske naloge, ki jo je pripravil Uroš Perkan (Perkan et al., 2023) in rezultate primerjamo z rezultati drugih raziskav na področju globalnega srednjeročnega napovedovanja vremena.

Model ConvCastNet

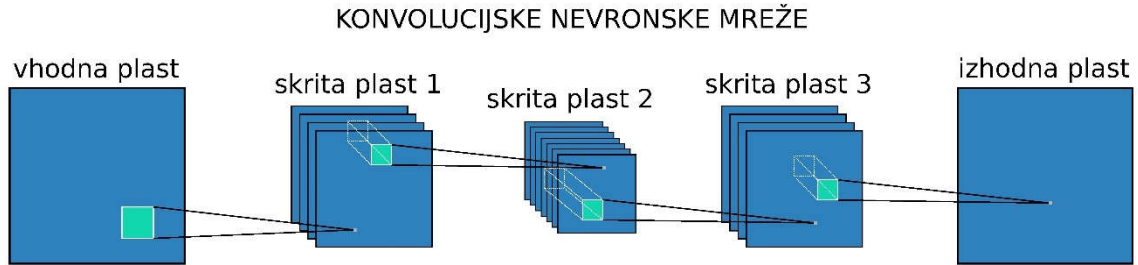
Model ConvCastNet (angl. *Convolutional Forecasting Network*) temelji na konvolucijskih nevronske mrežah. Za razumevanje njegovega delovanja si v tem poglavju najprej ogledamo teorijo konvolucijskih nevronske mrež, nato pa opišemo proces učenja in arhitekturo modela ConvCastNet.

Konvolucijske nevronske mreže

Izrek o neobstoju brezplačnega kosila (angl. *no free lunch theorem*) pravi, da nobena statistična metoda ni superiorna ostalim na vseh podatkovnih množicah (Wolpert & Macready, 1997). Posledično je izbira metode odvisna od podatkovne množice S oz. bolj natančno od lastnosti njenih vhodnih in ciljnih spremenljivk. Izbira tipa nevronske mreže sloni na pričakovani kakovosti in fleksibilnosti arhitekture ter na računski zmogljivosti strojne opreme, ki jo imamo na voljo. Sami smo konvolucijske nevronske mreže izbrali zaradi njihovih sposobnosti prepoznavanja vzorcev v atmosferskih poljih in sicer tako lokalnih značilnosti vremena na majhnih prostorskih skalah, kot globalnih povezav (t. i. telekonekcij), ki jih lahko prepozna v globljih plasteh nevronske mreže. Konvolucijske nevronske mreže imajo tudi sposobnost aproksimacije diferencialnih operatorjev, npr. 2D-Laplaceovega operatorja. Poleg tega smo imeli za učenje modela na voljo eno NVIDIA GeForce RTX 3090 grafično kartico, zato so bili iz izbire izvzeti vision transformerji, saj ti zahtevajo več velikostnih redov večjo računsko zmogljivost.

Konvolucijske nevronske mreže delujejo na podatkih v obliki tenzorjev z dimenzijami (C, H, W) , kjer je (H, W) matrika z dimenzijama višine H in širne W , C pa je število kanalov (tj. število slojev matrik v tenzorju). V modelu ConvCastNet uporabljamo 3-stopinjsko ločljivost, zato je matrika atmosferskih polj oblike $(60, 120)$, na vhodu pa uporabljamo

skupno $C=27$ različnih polj, saj so podatki predstavljeni na več višinskih nivojih. Delovanje konvolucijske nevronske mreže za napovedovanje vremena razložimo s pomočjo slike 1. Prikazan je konvolucijski kodirnik (angl. *convolutional encoder-decoder*), v katerem vhodne plasti najprej kodiramo v nižje dimenzionalne skrite plasti, nato pa dekodiramo nazaj na velikost vhodne plasti.



Slika 1: Prikazana je shema konvolucijske nevronske mreže v obliki konvolucijskega kodirnika. Velikost izhodne plasti je v tem primeru enaka vhodni, vmesne, skrite plasti pa so običajno manjše. Povezave med plastmi so lahko operatorji konvolucije, združevanja maksimalnih vrednosti ali operatorji, ki povečujejo ločljivost plasti.

Stanje nevrona v l -ti plasti $v^{(l)}$ iz izhodov plasti $l-1$ $y^{(l-1)}$ izračunamo kot (Goodfellow et al., 2016):

$$v_{i,j,k}^{(l)} = w_{0i}^{(l)} + \sum_{c=1}^C \sum_{m=1}^{k_1} \sum_{n=1}^{k_2} w_{i,c,m,n}^{(l)} y_{c,j+m-1,k+n-1}^{(l-1)}, \quad (1)$$

kjer je $w_{0i}^{(l)}$ predsodek i -tega izhodnega kanala, $y_{c,j+m-1,k+n-1}^{(l-1)}$ vrednost vhodnega tenzorja na položaju c -tega kanala, na višini $j+m-1$ in širini $k+n-1$, $w_{i,c,m,n}^{(l)}$ pa vrednost 4D tenzorja uteži na (m,n) -tem položaju konvolucijskega jedra velikosti $k_1 \times k_2$, ki povezuje i -ti kanal izhodne plasti s c -tim kanalom vhodne plasti. Izhod nevrona y iz njegovega stanja v izračunamo kot $y_{i,j,k}^{(l)} = \alpha(v_{i,j,k}^{(l)})$, kjer smo z α označili nelinearno aktivacijsko funkcijo. V našem primeru je $\alpha(x) = \text{LeakyReLU}(x) = \begin{cases} x; & x > 0 \\ kx; & x \leq 0 \end{cases}$, kjer je $k = 0.01$. Ta v nevronske mreže vpelje nelinearnost, zaradi katere se močno poveča njena sposobnost reprezentacije zapletenih nelinearnih funkcij. Stanja nevronov l -te plasti izračunamo tako, da po tenzorju izhodov nevronov $l-1$ plasti v horizontalnih smereh s poljubnim korakom premikamo konvolucijsko jedro velikosti (C, k_1, k_2) (na sliki 1 shematsko prikazano s svetlo modrim kvadrom) in na vsakem položaju pomnožimo komponente uteži in istoležne izhode nevronov (t. i. operacija navzkrižne korelacije oz. v literaturi globokega učenja zaradi velike podobnosti pogosto imenovana konvolucija) ter tako dobimo vrednosti značilnik v enem kanalu l -te plasti. Uteži so med premikanjem po tenzorju konstantne. Za vsako izhodno plast uporabimo svoj nabor uteži, ki lahko v vhodnih poljih prepoznavajo druge vzorce.

Poleg konvolucijskega lahko uporabljamo tudi operator združevanja maksimalnih vrednosti (angl. *max pooling*). Ta matriko nevronov agregira tako, da podmatrike oz. jedra sosednjih nevronov velikosti $n \times n$ združi v eno število, tako da izbere maksimalno vrednost med stanji nevronov znotraj jedra. Na ta način zmanjša število nevronov v naslednji plasti in nevronska mreža se je primorana učiti posploševanja vzorcev, ki jih dobi na vhodu nevronske mreže. Na sliki 1 vidimo, da je možno dimenzije plasti nevronov tudi povečevati. To storimo z operatorjem transponirane konvolucije, ki deluje podobno kot običajna konvolucija, saj se spremeni samo pomen parametra korakanja. Ta ne predstavlja več velikosti koraka horizontalnega premika jedra, ampak število dodanih ničel med stanja posameznih nevronov v vhodni plasti. Transponirana konvolucija poskrbi za ustrezno

velikost obleganja vhodnega tenzorja značilnik (tj. dodajanja novih stolpcev in vrstic na robovih tenzorja), tako da postane izhodna plast večja od vhodne.

Gradientni spust

Vrednosti uteži modela želimo nastaviti tako, da minimizirajo napako napovedi. Napaka modela je funkcija uteži nevronske mreže $E(w)$, kjer je w vektor vseh uteži. Izračunamo jo s pomočjo poljubne odvedljive funkcije L , za katero velja $L = L(y, \hat{y}) \geq 0 \forall y, \hat{y}$ in $L = 0$, ko $y = \hat{y}$. Pri tem smo z y označili ciljno oz. resnično vrednost, ki jo želimo dobiti kot izhod nevronske mreže, z \hat{y} pa napoved nevronske mreže. Takšni funkciji rečemo funkcija izgube (angl. *loss function*). Napaka je povprečna vrednost funkcije izgube na podatkovni množici S :

$$E = \frac{1}{|S|} \sum_{p \in S} L_p,$$

pri čemer $|S|$ označuje število primerov v podatkovni množici, $p \in S$ označuje posamezne primere iz S in L_p označuje vrednost funkcije izgube na primeru p . Pogosto izbrana funkcija izgube pri regresijskih modelih, tj. modelih, katerih zaloga vrednosti je kontinuum realnih števil, je vsota kvadratov razlik med modelsko napovedjo \hat{y} in resnično vrednostjo y (MSE). V nekoliko modificirani obliki jo za lepši zapis njenega odvoda zapišemo kot

$$L = \frac{1}{2} \sum_j (\hat{y}_j - y_j)^2.$$

V splošnem velja, da v primeru $\frac{\partial E}{\partial w} > 0$ z večanjem uteži w večamo tudi napako E , v primeru $\frac{\partial E}{\partial w} < 0$ pa z večanjem w napako manjšamo. Uteži zato popravljamo po enačbi

$$w_{t+1} = w_t - \eta \nabla_w E(w_t),$$

kjer je w_{t+1} vektor posodobljenih uteži, w_t vektor uteži pred posodobitvijo, $\eta > 0$ parameter stopnje učenja (angl. *learning rate*), $\nabla_w E(w_t)$ pa vektor odvodov funkcije napake po vseh utežeh. Takšno posodabljanje uteži zagotavlja, da se bo napaka na učni množici manjšala.

Če se uteži posodablajo tako, da sledijo smeri največjega gradienta funkcije napake, izračunanega na celotni podatkovni množici, temu algoritmu pravimo gradientni spust. Za učenje modela ConvCastNet uporabimo optimizirano različico algoritma gradientnega spusta, imenovano Adam (Kingma & Ba, 2017). Ta poleg gradientov pri posodabljanju uteži upošteva tudi njihov prvi in drugi moment. Poleg tega med učenjem spreminjamo tudi velikost parametra stopnje učenja. V ta namen uporabimo načrtovanje stopnje učenja ReduceLROnPlateau (ReduceLROnPlateau, 2023) ki deluje tako, da spremlja spremembe funkcije izgube in se za spremembe stopnje učenja odloča na podlagi števila epoh, tj. števila iteracij skozi vso učno množico, v katerih se napoved modela ni izboljšala. Določiti mu moramo mero potrpežljivosti (angl. *patience*) in faktor zmanjšanja stopnje učenja (angl. *factor*). Faktor določa vrednost, s katero pomnožimo stopnjo učenja ob njenem zmanjšanju, tj. $\eta \rightarrow \eta \cdot factor$, potrpežljivost pa določa število epoh brez izboljšav, po katerih stopnjo učenja pomnožimo z izbranim faktorjem. Poleg tega ima na voljo še nekaj dodatnih parametrov, s katerimi lahko dodatno optimiziramo njegovo delovanje. Za učenje modela

smo izbrali mero potrpežljivosti 6, faktor 0,01 in parameter praga zaznavanja sprememb 10^{-4} .

Modificirali smo tudi funkcijo izgube. Eksperimenti so pokazali, da so rezultati učenja boljši pri uporabi funkcije $L = 0,1 \cdot MSE^{\frac{1}{256}}$. Pri učenju so v vhodnem tenzorju uporabljena polja pri eni sami časovni instanci, z vključitvijo polj ob večih preteklih časih pa ne pridemo do velike izboljšave napovedi, zato je spomin na grafičnem procesorju boljše uporabljen pri vključitvi večih spremenljivk, kot z dodajanjem časovnih instanc obstoječim spremenljivkam. Učenje izvajamo v treh delih. Najprej 20 epoch model učimo napovedovati naslednji časovni korak, nato 20 epoch model učimo z dvema avtoregresivnima korakoma, za tem pa še 20 epoch s štirimi avtoregresivnimi koraki. Na ta način ga postopno učimo kakovostnega večdnevnega napovedovanja vremena. Pri tem vsakič znova ponastavimo načrtovanje stopnje učenja, zato se ta ciklično spreminja in omogoča modelu, da se izogne lokalnim minimumom funkcije napake. Posamezen korak napovedi opišemo kot

$$X(t+1) = \mathcal{S}^{-1}[\mathcal{M}(\mathcal{S}[X(t)])],$$

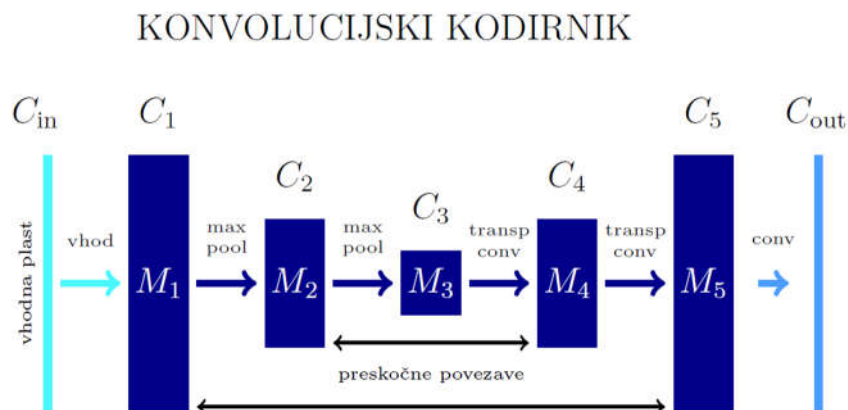
kjer je $X(t)$ tenzor nestandardiziranih spremenljivk ob času t , \mathcal{S} standardizacija podatkov, \mathcal{S}^{-1} inverzna transformacija standardizacije in \mathcal{M} model ConvCastNet. Standardizacijo \mathcal{S} izračunamo prek enačbe

$$X \rightarrow \frac{X - X_{povp}}{X_{std} + \epsilon},$$

kjer je X_{povp} časovno povprečje, X_{std} časovna standardna deviacija, izračunana v vsaki točki mreže, $\epsilon = 10^{-7}$ pa parameter za izboljšavo numerične stabilnosti.

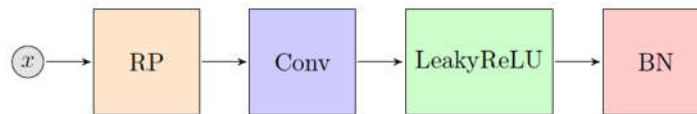
Arhitektura modela ConvCastNet

Model ConvCastNet je konvolucijski kodirnik, shematsko prikazan na sliki 2.



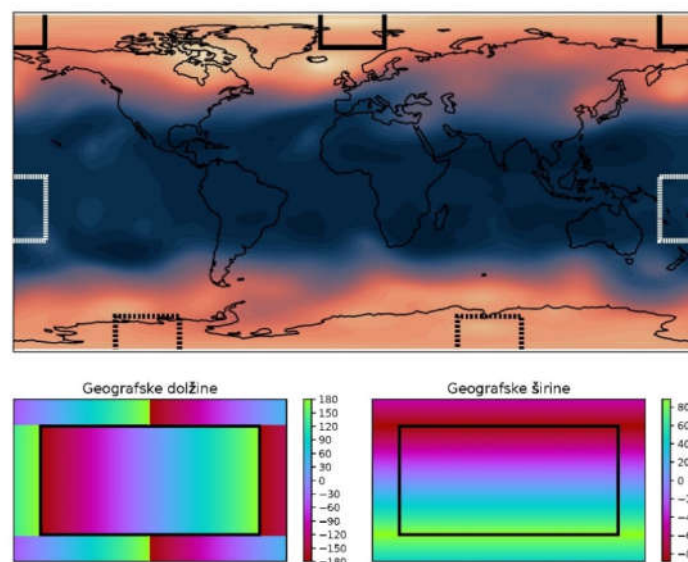
Slika 2: Skica osnovne arhitekture konvolucijskega kodirnika ConvCastNet. Sestavljen je iz vhodnega in izhodnega tenzorja nevronov ter vmesnih modulov M_i . V splošnem je lahko modulov poljubno mnogo. Vhodno polje vsebuje C_{in} , izhodno polje C_{out} , modul M_i pa C_i kanalov. Pri manjšanju ločljivosti med moduli uporabimo operator združevanja maksimalnih vrednosti (max pool), pri večanju pa transponirano konvolucijo (transp conv). Zadnjo plast dobimo iz zadnjega modula s konvolucijo z jedrom velikosti 1×1 . Med moduli so tudi preskočne povezave (angl. skip connections).

Sestavljen je iz petih modulov v kodirniku in štirih modulov v dekodirniku. Posamezen modul je sestavljen iz štirih zaporednih blokov, katerih zgradba je prikazana na sliki 3. V prvih treh modulih kodirnika in vseh modulih dekodirnika je velikost konvolucijskih jeder $k_1 = k_2 \equiv k = 7$, v ostalih pa $k = 3$.



Slika 3: Skica osnovnega bloka v modulih. Vhodno polje x najprej oblegamo (RP-robni pogoji). Nato uporabimo operator konvolucije (Conv) in LeakyReLU aktivacijsko funkcijo, na koncu pa izvedemo še normalizacijo paketov (BN).

Zmanjšanje velikosti konvolucijskega jedra je v modulih M_4 in M_5 uporabljeno zaradi manjše ločljivosti skritih plasti (npr. v modulu M_5 so te velike 3×7), zaradi česar bi konvolucijsko jedro velikosti 7×7 pokrilo velik del ali vso plast nevronov. Posamezna jedra se po vhodnem tenzorju premikajo s korakom velikosti ena. Znotraj modulov so med bloki uporabljene preskočne povezave s skupnim virom (tj. tenzor izhodov prvega bloka je dodan vhodom vseh ostalih blokov znotraj modula), preskočne povezave pa so uporabljene tudi med moduli kodirnika in dekodirnika (tj. tenzorji izhodov modula v kodirnem delu nevronske mreže so dodani vhodom v dekodirnem delu). Število kanalov C je v vseh modulih enako, saj smo opazili, da imajo uteži kanalov manjši prispevek h končni napovedni zmogljivosti kot tiste, ki jih dobimo s povečanjem velikosti konvolucijskih jeder. Pri tem smo izbrali $C = 350$. Med posameznimi moduli kodirnika je uporabljen operator združevanja največjih vrednosti, med moduli dekodirnika pa so uporabljeni operatorji transponirane konvolucije. Oboji uporabljajo jedra velikosti 2×2 in korak 2.



Slika 4: Na zgornjem grafu je prikazano, katere podatke bi morala vključevati konvolucijska jedra, da bi pravilno upoštevala sferične robne pogoje. Na spodnjih grafih so označene geografske širine (levo) in dolžine (desno) podatkov, s katerimi moramo obložiti vhodne podatke, da bo nevronska mreža pravilno upoštevala robne pogoje v Zemljini sferični geometriji. Matrika neobleganih podatkov je na njih prikazana s črnim pravokotnikom, obleganje pa je izvedeno na njenih robovih.0

Znotraj vsakega bloka je vhodno polje x najprej oblegano, tj. na robovih vhodnih polj so dodane vrstice in stolpci, ki poskrbijo, da konvolucijska jedra na robovih prekrivajo ustrezne geografske širine in dolžine, tako da je upoštevana Zemljina sferična geometrija, kot je prikazano na sliki 4. Mejne ploskve matrik se pred obleganjem nahajajo pri geografskih širinah $\varphi = \pm 88,5^\circ$ in geografskih dolžinah $\lambda = \pm 180^\circ$. Posamezna konvolucijska jedra na robovih vhodnih matrik zato nimajo dostopa do ustreznih značilk na drugi strani meja, zaradi česar na teh območjih niso sposobna realistično napovedovati razvoja vremena. Pri prehodu poldnevnik $\lambda = 180^\circ$ se geografska širina φ ohranja, spremeni pa se geografska dolžina $\lambda \rightarrow \lambda \pm 360^\circ$, kjer je predznak odvisen od smeri, iz katere se približujemo robu. Posledično mora ustrezno konvolucijsko jedro na mejnem območju prekrivati tiste elemente matrike, ki jih prikazuje beli črtkani kvadrater. Nekoliko drugače je na območju polov, kjer upoštevamo, da se robne vrstice matrik ne nahajajo na polu, temveč tik pod njim (na $\varphi = \pm 88,5^\circ$). Ob prehodu čez pol se geografska širina ohranja, geografska dolžina pa se spremeni za $\pm 180^\circ$. Ustrezna pokritja konvolucijskih jeder so za ta primer na sliki prikazani s črnimi črtkanimi pravokotniki. Kombinacijo obeh pa je treba upoštevati v ogliščih, kjer mora del konvolucijskega jedra prečkati pol, del pa poldnevnik 180° . Konvolucijsko jedro mora v tem primeru prekrivati območje, predstavljeno s črnim pravokotnikom. Ustrezne geografske širine in dolžine podatkov, ki jih moramo oblegati matrikam so prikazane na spodnjih grafih. V enačbi 1 vidimo, da konvolucijsko jedro na (i, j) tem položaju vhodne matrike s konvolucijo izračuna stanje (i, j) -tega nevrona izhodne plasti. Sosednje značilke vhodnih polj se torej po konvoluciji preslikajo v sosednje značilke skritih plasti, zato pričakujemo, da bodo tudi v teh bližnji nevroni opisovali bližnje vremenske pojave. Posledično ne oblegamo le tenzorjev vhodnih, ampak tudi skritih plasti.

Podatki

Osnovo vseh statističnih analiz predstavlja podatkovna množica. Neodvisno od kakovosti izbrane metode bo končen rezultat odvisen od izbire in kakovosti podatkov. Pri učenju nevronske mreže za globalno napovedovanje vremena potrebujemo podatke o stanju celotne atmosfere v daljšem časovnem intervalu, ki bo vseboval kar največji možni nabor različnih stanj ozračja. V ta namen uporabimo rekonstrukcijo stanja atmosfere – t. i. reanalizo. Ta poleg meritev uporabi tudi fizikalni model atmosfere, s pomočjo katerega v končnem rezultatu dobimo časovno rekonstrukcijo meteoroloških polj po celotni atmosferi. To storimo v procesu asimilacije meritev, kjer definiramo cenilko, ki meri odstopanje stanja atmosfere od kratkoročne modelske napovedi in meritev, pri tem pa ovrednoti tudi sistematsko napako satelitskih meritev in upošteva sistematske napake napovednega modela v stratosferi in mezosferi. Nato s pomočjo znanih statistik napak vseh virov cenilko minimiziramo z algoritmom 4D-variacijske asimilacije (Courtier et al., 1994) in tako dobimo najbolj verjetno stanje atmosfere v danem trenutku.

ERA5 reanaliza je peta in trenutno najnovejša reanaliza ECMWF. Izračunana je za vsako polno uro dneva v obdobju med leti 1940 in 2023, pri tem pa je med drugim upoštevano tudi časovno spreminjanje termodinamskega ozadja zaradi naraščajočih koncentracij toplogrednih plinov in zmanjšanih koncentracij ozona, ki vpliva na model sevalnega prenosa toplote in posledično na diabatno gretje v termodinamski enačbi. Spreminja se tudi spodnji robni pogoj, tj. temperatura površine morja in koncentracija morskega ledu. Poleg tega se spreminja tudi število in kakovost upoštevanih meritev, ki so v zadnjih desetletjih veliko bolj številčne in natančne, kot sredi 20. stoletja. Posledično je tudi reanaliza skladno s kvaliteto meritev bolj kakovostna proti koncu prej omenjenega časovnega intervala. Poleg meritev k

napaki reanalize prispeva tudi napaka izračunov modela, ki izvira iz aproksimacije parcialnih diferencialnih enačb z njihovimi diskretnimi različicami, parametrizacija konvekcije, turbulenca v planetarni mejni plasti in še mnogo drugih faktorjev.

Vse uporabljene spremenljivke so navedene v preglednici 1.

Tabela 1: Uporabljene spremenljivke in njihovi višinski nivoji. Temperatura Zemeljskega površja opisuje temperature površja kopnega, morja in ledu.

Spremenljivke	Višinski nivoji
Geopotencial ϕ [$m^2 s^{-2}$]	850, 700, 500, 250 [hPa]
Zonalen veter u [ms^{-1}]	10 [m], 900, 700, 500, 200 [hPa]
Meridionalen veter v [ms^{-1}]	10 [m], 900, 700, 500, 200 [hPa]
Temperatura zraka T [K]	2 [m], 850, 500 [hPa]
Potencialna vrtinčnost PV [$Km^2 kg^{-1} s^{-1}$]	/
Zračni tlak preračunan na nivo morja p [Pa]	/
Vsota padavin [m]	/
Temperatura Zemeljskega površja T_s [K]	/
Sončno obsevanje na vrhu atmosfere [Wm^{-2}]	/
Masa vodne pare v stolpcu atmosfere [$kg m^{-2}$]	/
Geografska širina [°]	/
Nadmorska višina [m]	/
Masa kopno-morje []	/

Uporabljenih je 5 enonivojskih in 5 večnivojskih ter 3 statične spremenljivke. Slednje so bile dodane z namenom, da bi modelu olajšali implicitno učenje raznoraznih masnih in energijskih tokov (npr. izhlapevanje vode nad morjem in kopnim) in vremenskih pojavov, povezanih z orografijo in geografsko širino. Teh ne napovedujemo, ampak jih nevronska mreža dobiva samo na vhodu.

Podatki so razdeljena na učno, validacijsko in testno množico (glej tabelo 2).

Tabela 2: Leto začetka in konca podatkov v učni, validacijski in testni množici. Končno leto je vključeno v podatkovno množico.

Množica	Začetek	Konec
Učna	1970	2014
Validacijska	2015	2019
Testna	2020	2022

Na učni množici izvajamo posodabljanje uteži, na validacijski poteka testiranje modela med učenjem, na testni pa testiranje naučenega modela. Končni model izberemo na podlagi najmanjše napake napovedi na validacijski množici, izračunane s seštevkom ACC vseh spremenljivk na nestandardiziranih podatkih. V vhodnih poljih standardiziramo vsako spremenljivko posebej in pri tem ločeno obravnavamo vsako lokacijo. Pri tem uporabimo povprečno vrednost in standardni odklon iz obdobja 1950 - 2014. Podatkov iz validacijske in testne množice med standardizacijo ne upoštevamo, saj bi v tem primeru model nekaj

informacij o novih podatkih imel tudi med testiranjem, zaradi česar bi bili rezultati preveč optimistični.

Rezultati

Napovedna zmogljivost

Kakovost vremenskih napovedi se s časom spreminja in je med drugim odvisna od trenutnega dinamičnega stanja ozračja, letnega časa in počasi spreminjajočih se atmosferskih in oceanskih oscilacij, zaradi katerih obstajajo obdobja povečane in zmanjšanje napovedljivosti vremena (Mariotti et al, 2020; Cohen et al., 2019). Posledično se natančnejšo analizo napovedne zmogljivosti izvaja s povprečenjem napak vremenskih napovedi, izvedenih v daljšem časovnem obdobju. Model zato testiramo v testni množici, tako da napoved izvajamo vsak 5. dan in jo povprečimo za vsak dan napovedi posebej. Pri tem uporabljamo metrike površinsko uteženih vrednosti RMSE in ACC. Te izračunamo po enačbah:

$$RMSE = \sqrt{E[c_w(\hat{y} - y)^2]}$$

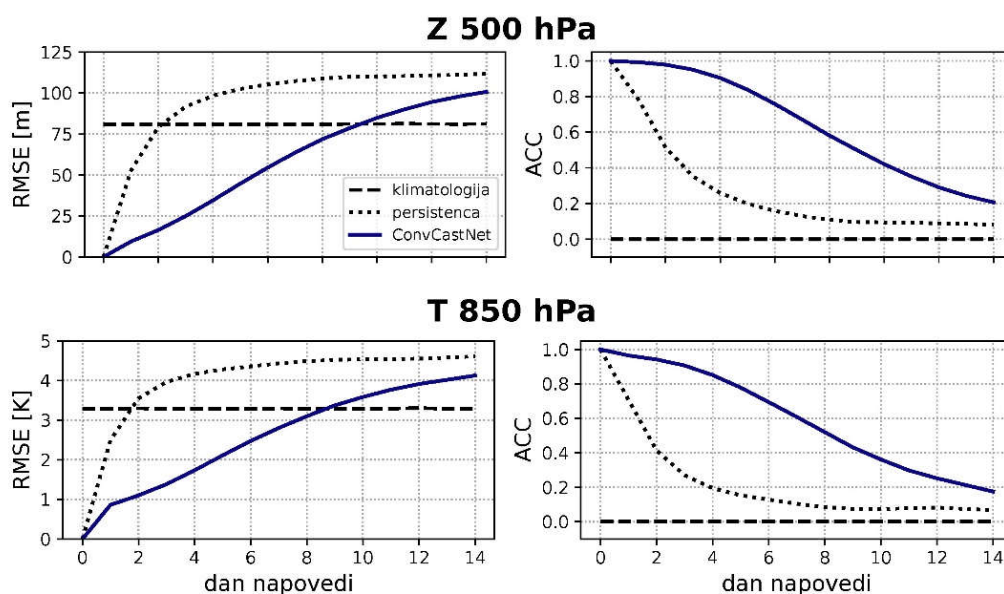
in

$$ACC = \frac{cov(\hat{y} - y_c, y - y_c)}{\sqrt{var(\hat{y} - y_c) var(y - y_c)}}$$

kjer je E pričakovana vrednost, c_w vektor površinskih uteži, ki jih izračunamo po enačbi

$$c_{wi} = \frac{cos(\varphi_i)}{\frac{1}{N_\varphi} \sum_{j=1}^{N_\varphi} cos(\varphi_j)}$$

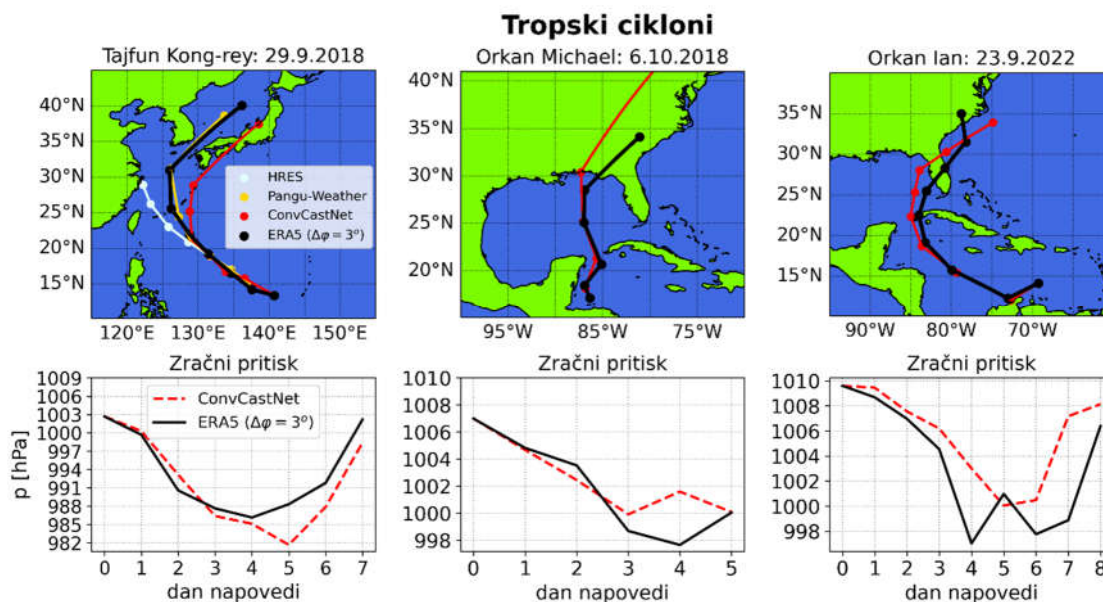
pri čemer N_φ predstavlja število upoštevanih točk v meridionalni smeri (smer sever-jug), \hat{y} napoved modela, y resnično vrednost iz ERA5 reanalize, y_c pa klimatološko povprečje na dan vremenske napovedi, izračunano v obdobju 1950-2014.



Slika 5: Globalna povprečna napaka na površinsko uteženih metrikah RMSE in ACC, izračunana v testni množici.

Površinske uteži upoštevamo tudi v variancah in kovariancah ACC, zato jih izračunamo po enačbah $var(x) = E[c_w(x - \mu_x)^2]$ in $cov(x,y) = E[c_w(x - \mu_x)(y - \mu_y)]$. Rezultati povprečne napovedne zmogljivosti ConvCastNet so za višino geopotencialne ploskve na 500 hPa (z_{500}) in temperaturo zraka na 850 hPa (T_{850}) prikazani na sliki 5. Napako meteoroloških modelov pogosto primerjamo s klimatologijo, tj. napovedjo klimatološkega povprečja, in persistenco, ki privzame, da je stanje ozračja v prihodnosti enako stanju ozračja ob začetku napovedi. Napaka ConvCastNet napovedi je vseh 14 dni precej manjša od napake persistence in pri RMSE metriki več kot 9 dni manjša od napake klimatologije. Rezultati ConvCastNet modela so izrazito boljši od persistence tudi za ostale spremenljivke (ni prikazano), nekoliko slabši pa so pri napovedi vsote padavin, kjer ima že enodnevna napoved precejšnjo napako. Napoved z_{500} lahko primerjamo tudi z modelom DLWP, ki je analizo izvajal na površinsko neobteženih metrikah, kar njihove vrednosti nekoliko zmanjša. Deterministična napoved DLWP nivo RMSE klimatologije doseže po približno 7,5 dneh, kar je približno 1,5 dni prej, kot ConvCastNet. Ansambelska napoved DLWP pa v prvih 14 dneh ne doseže klimatologije, zato je njegova dolgoročna napoved boljša.

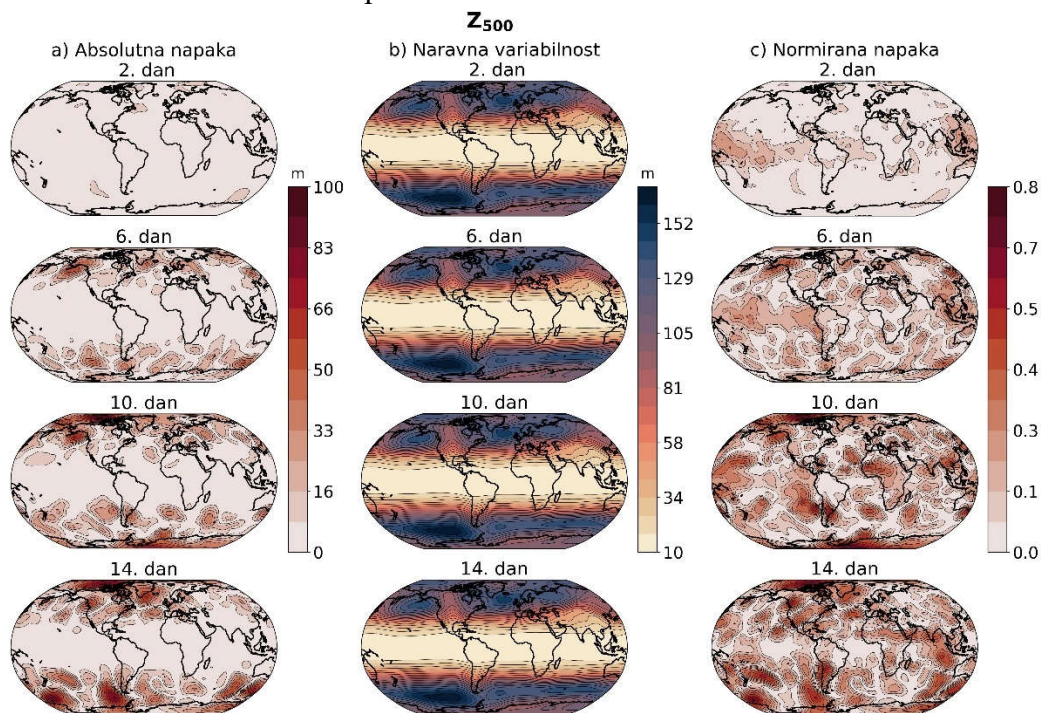
ConvCastNet smo testirali tudi pri napovedovanju poti in intenzitete tropskih ciklonov. Pri tem moramo upoštevati dejstvo, da tako modelska napoved kot ERA5 reanaliza prikazujeta dnevno povprečena in na 3° ločljivost interpolirana polja, zaradi česar tudi vrednosti zračnih tlakov v obeh primerih močno odstopajo od dejanskih, tj. globine ciklonov so močno podcenjene. Razvili smo algoritem za sledenje tropskim ciklonom, katerih središče definiramo v minimumu MSLP. Analizo napovedi smo izvedli na treh različnih tropskih ciklonih, ki so prikazani na sliki 6.



Slika 6: Primerjava napovedi tropskih ciklonov z ERA5 reanalizo. V zgornjih grafih so prikazane napovedane poti središč tropskih ciklonov, pri čemer so v prvem stolpcu prikazane tudi napovedane poti HRES in Pangu-Weather modela, pridobljene iz Bi et al., 2023. Točke so narisane s časovnim razmakom enega dne, pri čemer so za HRES in Pangu-Weather model narisane ob časih 12 UTC. V spodnji vrstici so prikazane vrednosti MSLP v središčih tropskih ciklonov.

Najprej si oglejmo napoved poti tajfuna Kong-rey. Ta je bila izračunana s pričetkom 29. 9. 2018, zato je primerljiva z napovedma modelov Pangu-Weather in HRES [13], ki sta bili inicializirani 30. 9. 2018 ob 00.00 UTC. Ker ConvCastNet napoveduje dnevno povprečna polja, sta napovedi obeh ostalih modelov prikazani ob časih 12 UTC. Najboljšo napoved je izvedel model Pangu-Weather, ki je pravo trajektorijo napovedal vseh 6 dni vnaprej. Pri tem je treba upoštevati, da je primer pridobljen iz njihovega članka (Bi et al., 2023), kjer so ga lahko samostojno izbrali, tako da je njihova napoved dosegla najboljše rezultate. Natančna primerjava množice tropskih ciklonov v letu 2018 je pokazala, da ni signifikantne razlike v kakovosti napovedovanja poti ciklonov med HRES in Pangu-Weather, medtem ko Pangu-Weather bistveno bolj podcenjuje globino ciklonov kot HRES (Ben-Bouallegue et al., 2023). Kakovostni sta bili tudi napovedi HRES in ConvCastNet modelov, ki dejanski trajektoriji pravilno sledita približno 3,5 dni. Za tem HRES tajfun premakne preveč na zahod, ConvCastNet pa na vzhod.

Prikazana je tudi napoved orkana Michael s pričetkom 6. 10. 2018. To lahko primerjamo z napovedjo 100-članskega ansambla modela FourCastNet (Pathak et al., 2022), ki je bila izvedena s pričetkom 7. 10. 2018 ob 00 UTC. V tem primeru je ConvCastNet prve 4 dni boljše sledil dejanski poti orkana kot FourCastNet. Ker sta obe opisani napovedi v našem modelu izvedeni znotraj validacijske množice, smo izvedli še napoved orkana Ian, ki se nahaja znotraj testne množice. Za začetni pogoj smo izbrali 23. 9. 2022, ko se je ta v začetni fazi razvoja nahajal v Karibskem morju. ConvCastNet je tudi v tem primeru pravilno ugotovil premik orkana proti vzhodu in napovedal, da bo ta vplival na zahod Kube in Florido. Na sliki 6 so prikazane tudi napovedane vrednosti MSLP v središčih tropskih ciklonov. Pri tem ponovimo, da se te pri 3° ločljivosti in dnevem povprečenju precej razlikujejo od dejanskih. V vseh primerih je opaziti korelacijo med napovedanimi in resničnimi poglobitvami oz. oslabitvami tropskih ciklonov.



Slika 7: Prikazane so letno povprečne napake in naravna variabilnost z_{500} v letu 2022. V prvem stolpcu (a) je prikazana absolutna vrednost povprečne razlike med napovedjo in reanalizo, v drugem stolpcu (b) naravna variabilnost, v tretjem (c) pa z naravno variabilnostjo normirana absolutna vrednost povprečne napake. Izračunane so za 2., 6., 10., in 14. dan napovedi.

Da bi lažje našli izvore napak našega modela, smo izvedli analizo prostorske porazdelitve napak. Na sliki 7a opazimo, da so absolutne napake največje na območjih barokline nestabilnosti. Ker pa se na teh območjih vreme že naravno hitro spreminja, obravnava absolutne vrednosti napake ni merodajna za analizo izvora napak. Absolutne napake spremenljivk zato normiramo z njihovo naravno variabilnostjo, tj. standardnim odklonom od dolgoletnega povprečja za določeni dan v letu. Tako ugotovimo velikost napak relativno na pričakovano hitrost spreminjanja vremena. Povprečne napake napovedi, zagnane vsak peti dan v letu 2022, so za polje z_{500} prikazane na sliki 7. Polje naravne variabilnosti z_{500} se zaradi dolgoletnega povprečenja podatkov za obdobje napovedi po dnevih skoraj ne spreminja (slika 7b), pričakovano pa je variabilnost največja v visokih geografskih širinah. Tudi absolutna napaka najhitreje raste na območjih z veliko naravno variabilnostjo, zato se za popolnejši vpogled v izvore napak osredotočimo na normirano napako. V drugem dnevu napovedi opazimo, da je ta največja na območju ekvatorialnega zahodnega Pacifika in Oceanije. Normirana napaka začne po daljšem času napovedi enakomerno naraščati tudi po preostalih območjih planeta in postane največja na območjih ponavljajočih se proženj Rossbyevih valov.

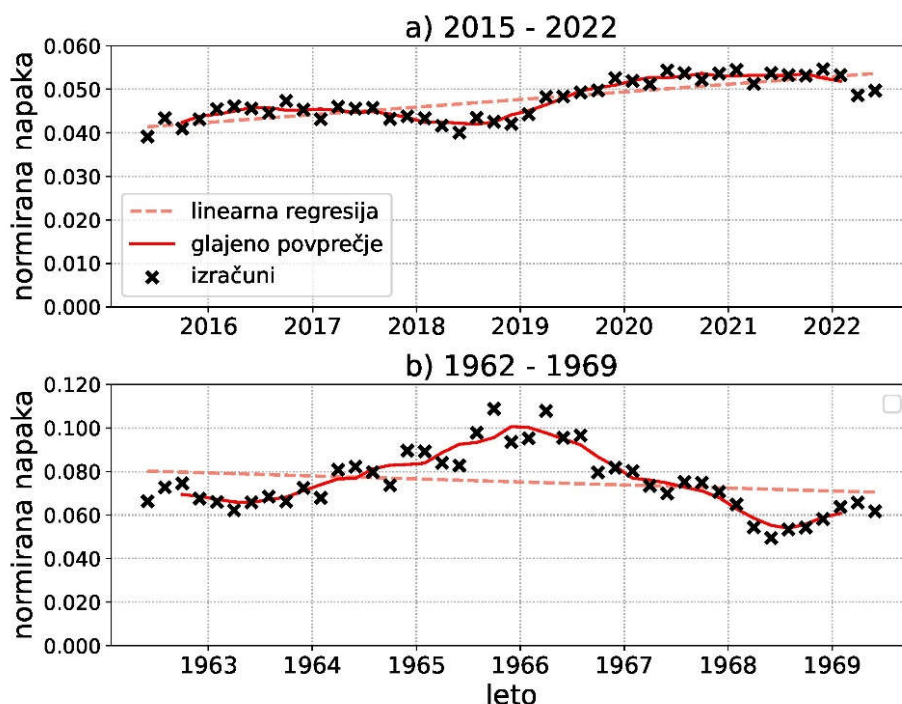
Ogledali smo si tudi spreminjanje globalne normirane napake v obdobju osmih let pred začetkom podatkov učne množice (leto 1970) in obdobju sedmih let po koncu podatkov učne množice (leto 2015) (slika 8). Napaka je izračunana z globalnim povprečenjem letnih normiranih napak 2. dneva napovedi. Pri tem so izračuni ponovljeni vsaka dva meseca, zato da pridobimo boljši vpogled v trend napovedljivosti v izbranih obdobjih. Linearna regresija pri tem pokaže, da se z oddaljevanjem od podatkov učne množice napaka počasi povečuje, kar je skladno s povečevanjem napake napovedi z modelom Pangu-Weather (Ben-Bouallegue et al., 2023). Opazimo pa tudi oscilacije okrog linearnega trenda, ki imajo na izračunanih podatkih podobno periodo kot ENSO, vendar te ne sovpadajo povsem z indeksi ENSO intenzitete (ni prikazano). Napovedljivost je torej odvisna tudi od lokalnega stanja klimatskega sistema. Na podatkih po letu 2015 je ta v povprečju približno 60 % boljša, kot na podatkih starejših ERA5 reanaliz. To je deloma povezano s hiperparametri nevronske mreže (npr. stopnja učenja, velikost konvolucijskega jedra, število kanalov v skritih plasteh,...), ki so bili izbrani za optimizacijo rezultatov v validacijski množici, deloma pa s spremembo lastnosti ERA5 reanalize in klimatskega sistema.

Razprava in zaključek

Ustvarili smo nov model za globalno srednjeročno napovedovanje vremena ConvCastNet, osnovan na arhitekturi konvolucijskega kodirnika. Razvili smo nov način upoštevanja sferičnih robih pogojev pri ekviangularni matrični reprezentaciji sferičnih podatkov v klasičnih konvolucijskih nevronskih mrežah. Obleganje matrik smo izvajali tako na meteoroloških poljih, kot tudi na izhodih skritih plasti. Pri tem smo upoštevali, da se sosednje značilke vhodnih plasti preslikajo v sosednje značilke izhodnih, zaradi česar so v skritih plasteh tudi kodirani vremenski pojavi blizu skupaj. Ugotovili smo, da se tako model nauči napovedovati razvoj vremena na robovih matričnih polj in da vremenske motnje v tem primeru lahko prehajajo prek njih.

Pri učenju modela smo izbrali optimizator Adam in empirično določili funkcijo izgube, ki pri učenju pripelje do najhitrejše konvergence. Uporabili smo načrtovanje stopnje učenja ReduceLROnPlateau, ki se je med preizkušenimi izkazala za najboljšo. Med učenjem smo model testirali na validacijski množici, kjer smo za metriko uporabljali seštevek ACC vseh spremenljivk. Pri analizi napovedljivosti pred obdobjem učne množice in po njem smo pokazali, da je modelska napoved precej boljša v obdobju po letu 2015 kot pred 1970. To

bi lahko nakazovalo na preprileganje modela na novejšje podatke ERA5 reanalize ali pa na večjo napako začetnega pogoja starejših ERA5 reanaliz.



Slika 8: Globalna povprečna normirana napaka 2. dneva napovedi z500 za obdobje 2015 - 2022 (a) in 1962 - 1969 (b). S križci so prikazane izračunane vrednosti napake, s polno rdečo črto glajeno povprečje, s črtkano črto pa linearna regresija. Križci so na časovni osi postavljeni v središča obdobj, v katerih so bili izračunani. Med sosednjimi izračuni je dvomesečen časovni korak. Domena vertikalne skale je na grafu (b) dvakrat večja kot na grafu (a).

Poleg tega se nevronske mreže naučijo napovedovati vreme le v razponu atmosferskih stanj, ki jih opisuje podatkovna množica. Sprememb dinamike atmosfere kot posledice sprememb v klimatskem režimu (npr. zaradi antropogenih podnebnih sprememb ali zaradi naravnih večdekadnih klimatskih oscilacij) zato niso sposobne predvideti, kar pa ne velja za fizikalne modele. Ti imajo na tem področju prednost pred statističnimi, saj fizikalni zakoni opisujejo večjo domeno faznega prostora in so zato aplikativni tudi pri napovedih v spremenjeni klimatologiji.

Model ConvCastNet smo učili na desetih dnevno povprečenih atmosferskih, morskih in površinskih spremenljivkah ter treh statičnih poljih, interpoliranih na 3° ločljivost. Pri napovedovanju poti tropskih ciklonov so napovedane poti več dni vnaprej uspešno sledile dejanski. Pri orkanu Michael je bila napoved ConvCastNet primerljiva napovedi modela FourCastNet (Pathak et al., 2022), pri tajfunu Kong-rey pa nekoliko slabša kot HRES in GraphCast (Lam et al., 2023). V vseh primerih pa so napovedi ConvCastNet zaradi dnevnega povprečenja polj, nizke ločljivosti in modelske napake močno podcenile dejansko intenziteto tropskih ciklonov. ConvCastNet je napram dnevno povprečeni ERA5 reanalizi, interpolirani na 3° ločljivost, napovedoval primerljive vrednosti zračnega tlaka. Iz tega sklepamo, da bi lahko z višjo ločljivostjo modela, izvajanjem direktnih vremenskih napovedi (brez dnevnega povprečenja polj) in z manjšim predsodkom zračnih tlakov v učnih podatkih, izboljšali tudi natančnost napovedi intenzitete tropskih ciklonov. To velja tako za model ConvCastNet kot ostale modele strojnega učenja. Težave pri napovedovanju vremena z modeli strojnega učenja povzročata tudi glajenje napovednih polj s časom. To v kombinaciji z nizko ločljivostjo

podatkov onemogoča izvajanje uporabne vremenske napovedi na mezoskalah ali konvektivnih skalah, kjer se pojavlja veliko ekstremnih vremenskih dogodkov, npr. intenzivni frontalni prehodi, nevihte in povezani ekstremni padavinski dogodki, pobočni viharji, vetrovi, vetrovi v tropskih ciklonih ...

Napovedno zmogljivost modela smo testirali na površinsko uteženih metrikah RMSE in ACC in do rezultatov prišli s povprečenjem napak napovedi, inicializiranimi v testni množici. Pri primerjavi rezultatov z ostalimi modeli moramo upoštevati dejstvo, da ConvCastNet zaradi uporabljene nižje ločljivosti in napovedovanja dnevno povprečenih polj podcenjuje napako napovedi, zato rezultati niso povsem primerljivi z ostalimi modeli. Napovedna zmogljivost ConvCastNet je v prvih dneh napovedi boljša od determinističnega in ansambelskega modela DLWP (Weyn et al., 2021), po daljšem času pa ansambelska napoved DLWP privede do manjše napake. ConvCastNet je na podlagi vrednosti ACC po 7 dneh slabši od GraphCast (Lam et al., 2023), FourCastNet (Pathak et al., 2022), Pangu-Weather (Bi et al., 2023), FengWu (Chen et al., 2023) in HRES (IFS), ki jih razvijajo večje kolaboracije v tehnoloških podjetjih in operativnih prognostičnih centrih in imajo pri tem na voljo sisteme z več velikostnimi redi večjo računsko zmogljivostjo.

Z analizo prostorske porazdelitve velikosti napak napovedi z_{500} smo prišli do zaključka, da je ta največja na območjih, kjer je njihova naravna variabilnost velika. Ta se v polju z_{500} nahaja na območjih zmernih geografskih širin, kjer prihaja do barokline nestabilnosti. Analizo smo izvedli tudi na napakah, ki so normirane z naravno variabilnostjo. Tako smo ugotovili, da se te pri ConvCastNet najprej pojavijo v tropskih območjih, kasneje pa se napaka močno poveča na območjih večje povprečne baroklinosti ozračja. Zakasnitev povečanja napake v zmernih geografskih širinah bi lahko bila posledica relativno bolj natančno določenega začetnega pogoja v zmernih geografskih širinah ali postopne propagacije napake iz tropov. Za zmanjšanje napake modelske napovedi bi lahko uporabili neekvangularno mrežo, kjer bi točke zgostili na območjih z razgibano orografijo, gosto poselitvijo ali na območjih s pričakovano večjo napako vremenskih napovedi. Namesto ekvangularne bi lahko uporabili tudi reducirano Gaussovsko mrežo ali ikozaedrsko mrežo, tj. mrežo v obliki ikozaedrske prizme, ki ohranja razdalje med točkami na sferi. Tako bi se znebili deformacij meteoroloških polj v ekvangularni mreži, ki v visokih geografskih širinah zaradi povečane hitrosti potovanja vremenskih motenj po matriki morebiti prispevajo k napaki modelskih napovedi. Izbira točk bi lahko temeljila tudi na podlagi lokacij izvajanja atmosferskih meritev, zato bi lahko tudi te vključili v proces napovedovanja vremena ali asimilacije meritev. Pri tem bi lahko bila bolj praktična uporaba grafovskih nevronske mreže, kjer je izbira položajev vozlišč poljubna.

V magistrskem delu smo pokazali, da lahko tudi s konvolucijskimi nevronske mrežami naučimo model za globalno napovedovanje vremena, ki se v nekaterih pogledih primerja z novejšimi modeli strojnega učenja. Ena izmed njegovih največjih pomanjkljivosti je nizka ločljivost njegovih napovedi, zaradi česar te niso neposredno uporabne za napovedovanje ekstremnih vremenskih pojavov z velikim vplivom na družbo, ki se tipično dogajajo na mezoskali. Večina novih modelov strojnega učenja iz literature že uporablja $0,25^\circ$ ločljivost, zaradi katere lahko napovedujejo tudi vremenske pojave na manjši prostorski skali. Kljub temu se vsi modeli strojnega učenja spopadajo s težavo glajenja napovedanih meteoroloških polj, zaradi česar po daljšem času niso sposobni natančno napovedovati mezoskalnih in konvektivnih ekstremnih vremenskih dogodkov. ConvCastNet poleg tega napoveduje dnevno povprečna meteorološka polja, kar na področju srednjeročnega napovedovanja vremena ni nujno zelo uporabno. Za razliko od fizikalnih numeričnih modelov pa so njegovi izračuni hitri in energijsko veliko bolj učinkoviti. Povečanje ločljivosti in odstranitev dnevnega povprečenja meteoroloških polj bi rezultate morebiti izboljšalo, vendar to pripelje predvsem do izzivov, povezanih z računsko zmogljivostjo strojne opreme. V prihodnje

načrtujemo izvedbo prve 4-D variacijske asimilacije meritev, izvedene izključno na podlagi modelov strojnega učenja. To želimo doseči z združitvijo modela ConvCastNet in modela strojnega učenja za izvedbo 3-D variacijske asimilacije meritev (Melinc in Zaplotnik, 2023). Uspešnost izvedbe asimilacije meritev z modeli strojnega učenja bi lahko imela še večji doprinos k napovedovanju vremena kot optimizacija napovednega modela samega, saj v operativnih centrih za napovedovanje vremena večino računskih zmogljivosti posvečajo pripravi začetnega pogoja, zato bi z vpeljavo metod strojnega učenja morda lahko sprostili veliko računskih moči. Razvili smo torej model strojnega učenja, katerega kombinacija nizke računske potratnosti in zadovoljive napovedne zmogljivosti vzpodbuja k njegovemu nadaljnjemu izpopolnjevanju in uporabi na področju raziskovanja in analize napovedovanja vremena.

Literatura

- Bauer, P. et al. (2015). The quiet revolution of numerical weather prediction, *Nature* 525, 47-55.
- Ben-Bouallegue, Z. et al. (2023). The rise of data-driven weather forecasting, arXiv:2307.10128.
- Bi, K. et al. (2023). Accurate medium-range global weather forecasting with 3D neural networks, *Nature* 619, 533-538.
- Bonavita, M. (2023). On some limitations of data-driven weather forecasting models, arXiv:2309.08473.
- Bonev, B. et al. (2023). Spherical Fourier Neural Operators: Learning Stable Dynamics on the Sphere, arXiv:2306.03838.
- Charney, J. G. et al. (1950). Numerical Integration of the Barotropic Vorticity Equation, *Tellus* 2, 237-254.
- Chen, K. et al. (2023). FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead, arXiv:2304.02948.
- Cohen, J. et al. (2019). S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts, *WIREs Climate Change* 10, e00567.
- Courtier, P. et al. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society* 120, 1367- 1387.
- Dueben, P. D. & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning, *Geosci. Model Dev.* 11, 3999–4009.
- ECMWF, IFS Documentation CY43R3 - Part III: Dynamics and numerical procedures, <https://www.ecmwf.int/en/elibrary/80319-ifs-documentation-cy43r3-part-iii-dynamics-and-numerical-procedures> (1.12.2023) .
- Goodfellow, I. et al. (2023). *Deep Learning* (MIT Press, 2016) book in preparation for MIT Press.
- Hersbach, H. et al. (2020). The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society* 146, 1999-2049 .
- Keisler, R. (2022). Forecasting Global Weather with Graph Neural Networks, arXiv:2202.07575.
- Kingma, D. P. & Ba, J. (2017). Adam: A Method for Stochastic Optimization, arXiv:1412.6980.
- Kochkov, D. et al. (2023). Neural General Circulation Models, arXiv:2311.07222.
- Lam, R. et al. (2023). Learning skillful medium-range global weather forecasting, *Science* 0.
- Mariotti, A. et al. (2020). Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond, *Bulletin of the American Meteorological Society* 101, E608–E625.
- Melinc & Zaplotnik (2023). Neural-Network Data Assimilation using Variational Autoencoder, arXiv:2308.16073.
- Pathak, J. et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive fourier neural operators, arXiv:2202.11214.
- Perkan, U. (2023): Napovedovanje vremena s konvolucijskimi nevronskimi mrežami (Weather forecasting with convolutional neural networks). Master's thesis, Univerza v Ljubljani, FMF (in Slovenian).

- Rabier, F. (2023). The AIFS is launched. <https://www.ecmwf.int/en/newsletter/177/editorial/aifs-launched> (20.11.2023) .
- Rasp, S. et al. (2018). Deep learning to represent subgrid processes in climate models, PNAS 115, 9684-9689.
- Rasp, S. et al.. (2023). WeatherBench 2: A benchmark for the next generation of data-driven global weather models, arXiv:2308.15560.
- ReduceLROnPlateau (2023). https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html (21.11.2023).
- Richardson, L. F., Lynch, P. (1922). Weather Prediction by Numerical Process, Cambridge University Press.
- Weyn, J. A. et al. (2019). Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data, Journal of Advances in Modeling Earth Systems 11, 2680-2693.
- Weyn, J. A. et al. (2021). Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models, Journal of Advances in Modeling Earth Systems 13 .
- Wolpert, D. H. & Macready, W. G. (1997). No free lunch theorems for optimization, IEEE Transactions on Evolutionary Computation 1, 67-82.