Primljen / Received: 9.5.2017.
Ispravljen / Corrected: 1.3.2018.
Prihvaćen / Accepted: 17.8.2018.
Dostupno online / Available online: 10.9.2018.

Usporedba metoda nadziranog učenja u svrhu predviđanja srednjeg mjesečnog protoka

Autori:

Dr.sc. Jadran Berbić, mag.ing.aedif.
Državni hidrometeorološki zavod
jberbic@hotmail.com

Izv.prof.dr.sc. Eva Ocvirk, dipl.ing.građ.; dipl.ing.mat.
Sveučilište u Zagrebu
Građevinski fakultet
ocvirk@grad.hr

Doc.dr.sc. Gordon Gilja, dipl.ing.građ.
Sveučilište u Zagrebu
Građevinski fakultet
ggilja@grad.hr

Izvorni znanstveni rad

Jadran Berbić, Eva Ocvirk, Gordon Gilja

Usporedba metoda nadziranog učenja u svrhu predviđanja srednjeg mjesečnog protoka

Dugoročno planiranje hidrotehničkih sustava zahtijeva poznavanje dugoročne dostupnosti vode, najčešće u obliku srednjeg mjesečnog protoka. Uglavnom se koriste znanja iz stohastičke hidrologije, a mogući scenariji dobivaju se generiranjem sintetičkog protoka. Raspolaganje klimatskim modelima nameće mogućnost modeliranja iz budućih scenarija, a pretpostavka u radu je da se za tu svrhu može primjenjivati nadzirano učenje. U radu je analizirana preciznost tri modela nadziranog učenja u tri pristupa i autoregresivnog modela u prvom pristupu, za predviđanje srednjeg mjesečnog protoka, a u ovisnosti o duljini povijesnog niza.

Ključne riječi:
dugoročno planiranje, srednji mjesečni protok, autoregresivni model, nadzirano učenje

Original scientific paper

Jadran Berbić, Eva Ocvirk, Gordon Gilja

Comparison of supervised learning methods for prediction of monthly average flow

Long-term planning of water resources systems requires knowledge of long-term availability of water, most often in the form of monthly average flow information. Knowledge from stochastic hydrology is most often applied, and possible scenarios also involve generation of synthetic flow. The use of climatic models imposes the possibility of modelling based on future scenarios, and it is assumed in the paper that supervised learning can be applied for this purpose. The paper analyses accuracy of three supervised learning models in three approaches and the autoregressive model in the first approach, for predicting monthly average flow as related to the length of a historic dataset.

Key words:
long-term planning, monthly average flow, autoregressive model, supervised learning

Wissenschaftlicher Originalbeitrag

Jadran Berbić, Eva Ocvirk, Gordon Gilja

Vergleich der Methoden des überwachten Lernens zum Zweck der Vorhersage des mittleren monatlichen Durchflusses

Die langfristige Planung von hydrotechnischen Systemen erfordert Kenntnisse über die langfristige Verfügbarkeit von Wasser, meist in Form des mittleren monatlichen Durchflusses. Hauptsächlich werden Kenntnisse aus der stochastischen Hydrologie angewendet, und mögliche Szenarien erhält man durch Erzeugung des synthetischen Durchflusses. Die Verfügung über Klimamodelle drängt die Möglichkeit der Modellierung anhand zukünftiger Szenarien auf, und die Voraussetzung in der Abhandlung ist die, dass zu diesem Zweck das überwachte Lernen angewendet werden kann. In der Abhandlung wurde die Präzision von drei Modellen des überwachten Lernens in drei Ansätzen und des autoregressiven Modells im ersten Ansatz zur Vorhersage des mittleren monatlichen Durchflusses analysiert, abhängig von der Länge der historischen Reihe.

Schlüsselwörter:
langfristige Planung, mittlerer monatlicher Durchfluss, autoregressives Modell, überwachtes Lernen

1. Uvod

Nadolazeći pritisci na vodna dobra u smislu povećanja broja stanovništva, potrebe za energijom i hranom zahtijevaju povećanje učinkovitosti i djelotvornosti proizvodnje [1, 2]. Izraženije klimatske varijacije i promjene uzrokuju češće pojave ekstremno vlažnih i sušnih razdoblja te mijenjaju statističku raspodjelu hidroloških događaja [3-5]. Praktičan alat za simulaciju protoka hidrološki izučenih slivova predstavljaju stohastičke metode i metode nadziranog učenja. Pretpostavka je da se primjenom odgovarajućih simulacijskih modela u gospodarenju vodama mogu analizirati sadašnje i buduće potrebe vezane uz hidrotehničke sustave ako je prisutan dovoljno dug povijesni niz mjerenja. Primjerice, izgradnja kvalitetnog simulacijskog modela nužna je kako bi se proveo simulacijsko-optimizacijski postupak za analizu dostupnosti vode za potrebe ovisne o akumulaciji [6]. Stoga su predviđanja srednjeg mjesečnog protoka mjesec-za-mjesec i dugoročno planiranje od velike važnosti za planiranje i odabir režima rada akumulacija.

U radu je prikazana analiza prihvatljivosti korištenja povijesnog niza protoka ovisno o duljini niza i podacima na raspolaganju. Ispitana je mogućnost primjene autoregresivnog modela (eng. *autoregressive* - AR) i tri metode nadziranog učenja (eng. *supervised learning* - SL) u predviđanju na temelju protoka, te iste tri metode za predviđanje na temelju količine oborine i temperature zraka. Ciljevi analize su: odgovoriti na pitanje kolika je minimalna duljina niza pri kojoj je prihvatljivo primjeniti navedene metode te ispitati mogućnost izgradnje kvalitetnog modela kojim bi se predviđao protok iz rezultata klimatskih modela.

1.1. Pregled i zaključci iz dosadašnjih istraživanja

Strojnim učenjem pronalaze se zakonitosti u podacima te generaliziraju indukcijom. Nadzirano učenje je dio strojnog učenja i umjetne inteligencije kojim se na temelju zadanih podataka (ulaza i izlaza) i pretpostavljene hipoteze (funkcije) pretražuju njeni parametri koji rezultiraju najboljim predviđanjima na neviđenim primjerima, a za rješavanje problema klasifikacije i regresije. Iz pregleda literature može se zaključiti da je u hidrologiji često primjenjivano nadzirano učenje za potrebe predviđanja u stvarnom vremenu (s vremenskim korakom do nekoliko sati) te za kratkoročne i srednjoročne prognoze (1-7 dana), a rjeđe za dugoročne prognoze (mjesec dana) te još rjeđe za dugoročno planiranje. Primjena manjeg vremenskog koraka je zanimljiva zbog prisutnosti većeg broja podataka za gradnju modela te je relativno jednostavno izgraditi kvalitetan model bez korištenja vanjskih varijabli (dakle, protok se predviđa iz samog protoka). S druge strane, tako izgrađen model nije u stanju pouzdano predviđati nekoliko vremenskih koraka unaprijed od trenutnog (jer se s povećanjem broja koraka generira greška), osim uz možebitno uvođenje vremenski usrednjenih varijabli ili primjena vanjskih prediktora (temperature zraka, količine oborine, itd.) kao ulaznih varijabli.

Najpopularniji model SL-a jest umjetna neuronska mreža (*eng. artificial neural network* - ANN) te je prisutna u velikoj većini radova iz predmetnog područja. Gradnja modela ANN-a sastoji se u odabiru težina u sinapsama s ciljem da se učenjem iz primjera minimizira razlika između željenog izlaza i stvarnog izlaza neuronske mreže na temelju odabranog statističkog kriterija [7]. S obzirom na razloge stalnog poboljšanja razumijevanja hidrološkog ciklusa, hidrolozi se tijekom dugogodišnjeg modeliranja više usmjeravaju na fizikalno zasnovane modele, što s vremenom dovodi do gradnje složenijih modela [8]. Glavne prednosti ANN-a, primjerice zaobilaženje problema potpunog razumijevanja procesa otjecanja za hidrološko modeliranje, što je na stvarnoj prostornoj skali složeno, primijećene su već tijekom prethodna dva desetljeća. Nije potrebno uvoditi pretpostavke linearnosti, detaljno opisati složene veze različitih procesa, uporaba podataka je fleksibilnija, modele je moguće relativno brzo izgraditi [9]. Slične prednosti pojavljuju se kod primjene ostalih modela nadziranog učenja. SVM se cijeni zbog sposobnosti generalizacije, strogih teoretskih osnova, relativno jednostavne primjene te robusnosti na problemima regresije i prepoznavanja uzoraka [7]. Jedan od ciljeva ovog rada jest usporediti tri različita modela nadziranog učenja: ANN-a kao popularnog modela, SVM-a kao robusnog, ali u hidrološkoj literaturi manje prisutnog i NNM-a (metoda najbližih susjeda, eng. *Nearest Neighbours Method* - NNM) kao vrlo jednostavnog modela u odnosu na prethodna dva.

Cigizoglu i dr. (2005.) uspoređuju ANN sa stohastičkim autoregresivnim modelom s pomičnim prosjekom (eng. *autoregressive moving average* - ARMA) i modelom višestruke linearne regresije (MLR, *eng. multilinear regression model*) [10]. ANN upotpunjen generaliziranom regresijom (GR) dao je točnije rezultate nego uobičajeni ANN. Dodatno su ANN gradili na seriji generiranih dotoka, što je omogućilo znatno više podataka na raspolaganju i poboljšalo modele, a najtočniji je bio GR ANN. Nilsson i dr. (2006.) koristeći ANN sa 6-12 vanjskih prediktora predviđaju srednje mjesečno otjecanje na slivu. Nakon pokušaja s temperaturom i oborinom, rezultati su poboljšani dodavanjem količine snijega i sezonskih karakteristika, dok vlažnost tla i sjevernoatlantski indeks oscilacija nisu poboljšali točnost predviđanja [11]. Wu i Chau (2010.) su primjenom ANN-a, NNM-a te ARMA modela sa 6-12 ulaza (protoka) predviđali srednji mjesečni protok jedan mjesec unaprijed. Gradnji modela je prethodila metoda rekonstrukcije faznog prostora (eng. *phase space reconstruction* - PSR). NNM i ARMA dali su bolje rezultate nego ANN i kombinacija PSR-ANN, dok je ANN poboljšana pomičnim prosjecima dala najbolje rezultate [12]. Guo i dr. (2011.) uveli su poboljšanja kod ANN-a i SVM-a (metoda potpornih vektora, *eng. support vector machine*) primjenom valne metode i PSR-a, optimizacije rojem čestica kod SVM-a i Levenberg-Marquardt-ovog algoritma kod ANN-a gradeći modele s osam ulaza (protoka). Postignuti su precizniji rezultati, ali sa složenijom procedurom za predviđanje protoka jedan mjesec unaprijed, u odnosu na ANN i SVM [13]. Akiner i Akkoyunlu (2012.) koriste ANN za nadopunu nedostajućih podataka i

predviđanje oborine, a otjecanje u sljedećem desetljeću određuju koristeći predviđenu oborinu u SWAT modelu [14]. Farajzadeh i dr. (2014.) uspoređuju preciznost ARIMA (autoregresivni model s integriranim pomičnim prosjekom) i ANN-a za predviđanje srednjeg mjesečnog otjecanja. Nakon predviđanja oborine modelima, otjecanje je predviđano iz oborine - modelima te koeficijentom otjecanja. ARIMA je dala nešto točnije rezultate, pri čemu je pristup s koeficijentom otjecanja bio precizniji [15]. Terzi (2014.) koristi GP (genetsko programiranje, *eng. genetic programming*) za predviđanje srednjeg mjesečnog dotoka iz oborine s tri postaje i dotoka s dvije postaje te uspoređuje s MLR-om [16].

Na predmetnom području (Vinalić, Cetina) primijenjena je ANN u svrhu kratkoročnog predviđanja dotoka, u radu Matić (2014.). Kroz različite pristupe korištenja ulaznih varijabli (dotoka, pale oborine i temperature zraka) i primjene ANN-a riješen je problem odziva modela u odnosu na stvarni događaj za predviđanje od jednog do deset dana unaprijed. Uspoređeni su modeli vremenske serije (predviđanje dotoka iz dotoka), oborine i otjecanja (predviđanje dotoka iz oborine ili oborine i dotoka) te viševeličinski modeli (predviđanje dotoka iz oborine, temperature itd.). Za predviđanje dotoka primijenjene su direktna i indirektna metoda. Dok se direktnom metodom za svaki vremenski korak gradi zaseban model, indirektnom metodom se koristi jedan model za sve korake te je zbog generiranja greške nepreciznija od prethodne. Modeli vremenske serije su se pokazali najtočnijima, ali je trebalo riješiti problem odziva. Nakon provedenih koraka: uvođenja učestalosti oborine i akumulirane oborine, primjene adaptivnog neuronskog modela s podmodelima za različito doba godine i optimiziranog neuronskog modela, te uvođenjem usrednjavanih varijabli, znatno je povećana preciznost. Gradnja i kalibracija izvedene su na podacima od 2007. do 2011. godine (1826 podataka), a verifikacija na podacima iz 2012. godine (365 podataka) [17]. U pravilu se u literaturi koriste dulji povijesni nizovi, 20 do 40 godina [10, 16], 40 do 60 [11, 12, 15] pa čak i stotinjak godina [13]. Kako kvaliteta modela SL izravno ovisi o količini podataka korištenih za izgradnju modela (vjerojatnost izgradnje kvalitetnog modela povećava se s količinom korištenih podataka zbog veće mogućnosti generalizacije zakonitosti), zanimljivo je ispitati kolika je duljina niza potrebna za gradnju modela sposobnog predviđati izvan domene povijesnog niza, uz zadovoljavajuću preciznost. Prema vremenskom okviru planiranja razlikuju se modeli za dugoročno predviđanje (mjesec-za-mjesec) i za dugoročno planiranje, a svrha rada obuhvaća razvoj modela za obe potrebe. Korišteni su modeli vremenske serije (za 1 vremenski korak unaprijed) te viševeličinski modeli (direktna metoda, s tim da jedan model uči opću zakonitost između ulaznih i izlaznih veličina). Količina podataka iznosila je oko 110-750 (povijesni niz od 10 do redom 65, 62, 60 godina, godine bez mjerenja protoka nisu uračunate), od čega je 60 % korišteno za gradnju, 20 % za kalibraciju, 20 % za verifikaciju, a ostatak (od približno 640 do nula podataka za godine od 10 do redom 65, 62 i 60) za dodatnu verifikaciju modela.

Predviđanja srednjeg mjesečnog protoka pomoću SL-a rjeđe su zastupljena nego predviđanja na kraćoj vremenskoj osnovi, pogotovo za dugoročno planiranje (od navedenih radova [10, 14]). Prema spoznajama autora, nema radova koji analiziraju utjecaj duljine niza na točnost SL-a, a količina podataka izravno utječe na točnost modela.

2. Metodologija istraživanja

2.1. Autoregresivni model: Thomas-Fiering AR(1)

Stohastički procesi u gospodarenju vodama često se opisuju Markovljevim procesima, a za potrebu primjene uvodi se pretpostavka stacionarnosti povijesnog niza. Markovljevi procesi diskretiziraju se diskretnim procesima, Markovljevim lancima [18]. Opći oblik autoregresijskih modela AR(p) reda ρ je [19]:

$$z_t = \sum_{i=1}^{p} \varphi_i z_{t-i} + \varepsilon_t$$

gdje su: z_t vremenski neovisna, normalizirana i standardizirana serija, φ_i autoregresivni koeficijenti, ε_t vremenski neovisne varijable. Najjednostavniji je autoregresivni proces prvog reda AR(1). Za mjesečne normalno raspodijeljene dotoke sa srednjom vrijednošću μ, varijancom σ^2, korelacijom mjesec-za-mjesecom ρ može se primjenjivati Thomas-Fieringov model AR(1) [6, 18]:

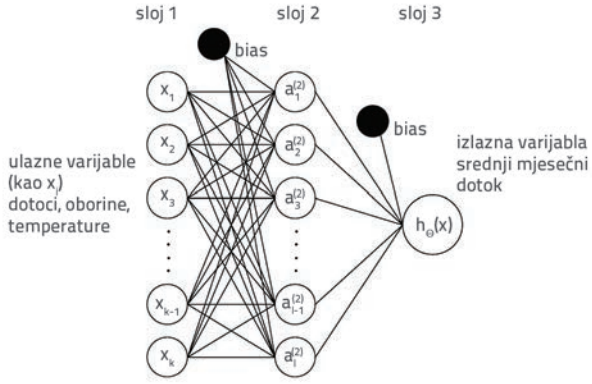
$$Q_{i+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j}(Q_i - \mu_j) + V_i \sigma_{j+1}(1 - \rho_j^2)^{0.5} \quad (1)$$

gdje su: Q_i, Q_{i+1} srednji mjesečni dotok za i+1-i i i-ti mjesec, μ_j, μ_{j+1} po godinama srednji protoci za j-i i j+1-i mjesec, σ_j, σ_{j+1} standardna devijacija j-tog i j+1-og mjeseca (po godinama), ρ_j koeficijent korelacije j-tog i j+1-og mjeseca, V_i nasumično odabrana varijabla iz normalne raspodjele sa srednjom vrijednošću $E[V_i] = 0$ i jediničnom varijancom $E[V_i] = 1$. Ovakav model često je korišten za generiranje sintetičkih dotoka i u stanju je očuvati statističku sličnost s povijesnim nizom (npr. [6]). Model je primijenjen u prvom pristupu (poglavlje 3.), a procedura je programirana u programskom okruženju Python (www.python.org, [20]), koje je korišteno i za sve ostale modele.

2.2. Umjetne neuronske mreže

ANN oponaša princip učenja kakav je prisutan u mozgu, uz pretpostavku da se proces učenja odvija kroz elektrokemijsku aktivnost u mrežama sastavljenih od neurona [21]. ANN se najčešće sastoji od tri sloja: prvog sloja karakteriziranog čvorovima koji su u biti ulazne varijable, skrivenog sloja s čvorovima s aktivacijskom funkcijom te sloja s izlaznim čvorom – predviđenom vrijednošću (npr. protok). Mogućnost variranja broja skrivenih slojeva i čvorova upućuje na činjenicu da je proces iznalaženja odgovarajuće arhitekture ANN-a složen zadatak [21, 22]. U radu je korišten tip višeslojni perceptron (*eng. multilayer perceptron*), s tri sloja, za rješavanje problema regresije. Razlike stvarnih i modeliranih vrijednosti minimizirane su stohastičkim

optimizacijskim algoritmom zasnovanim na gradijentima prvog reda (eng. *Adaptive Moment Esimation* - ADAM). Algoritam je računalno učinkovit, ne zahtijeva mnogo memorije i pogodan je u slučaju veće količine podataka [20, 23] Parametri koji najznačajnije utječu na kvalitetu izgradnje modela su broj unutarnjih slojeva i broj čvorova u sloju, aktivacijska funkcija, intenzitet i moment učenja, maksimalni broj iteracija u optimizaciji greške te tolerancija greške. Postoji još parametara, ali je u pravilu kod primjene SL-a odabir ulaznih varijabli ključan korak.



Slika 1. Struktura troslojne neuronske mreže

Kod troslojne ANN aktivacija $a_j^{(2)}$ u čvoru j = 1, 2, ..., l skrivenog sloja (oznaka 2) računa se na sljedeći način [24, 25]:

$$a_1^{(2)} = g(\theta_{1,0}^{(1)}x_0 + \theta_{1,1}^{(1)}x_1 + \theta_{1,2}^{(1)}x_2 + \ldots + \theta_{1,k-1}^{(1)}x_{k-1} + \theta_{1,k}^{(1)}x_k) = g(z_1^{(1)})$$

$$a_2^{(2)} = g(\theta_{2,0}^{(1)}x_0 + \theta_{2,1}^{(1)}x_1 + \theta_{2,2}^{(1)}x_2 + \ldots + \theta_{2,k-1}^{(1)}x_{k-1} + \theta_{2,k}^{(1)}x_k) = g(z_2^{(1)})$$

$$\vdots \qquad (2)$$

$$a_l^{(2)} = g(\theta_{l,0}^{(1)}x_0 + \theta_{l,1}^{(1)}x_1 + \theta_{l,2}^{(1)}x_2 + \ldots + \theta_{l,k-1}^{(1)}x_{k-1} + \theta_{l,k}^{(1)}x_k) = g(z_l^{(1)})$$

gdje su: g aktivacijska funkcija, $\theta_{j,i}^{(1)}$ težinski utjecaj ulazne varijable x_i na aktivaciju $a_j^{(2)}$, i = 1, 2, ..., k. Indeks k označava broj čvora prvog sloja, indeks i broj čvora skrivenog sloja, a indeks 0 odnosi se na "bias" varijablu. Predviđena veličina računa se prema izrazu:

$$h_\theta(x) = a_1^{(3)} = g(\theta_{1,0}^{(2)}a_0^{(2)} + \theta_{1,1}^{(2)}a_1^{(2)} + \theta_{1,2}^{(2)}a_2^{(2)} + \ldots + \theta_{1,l-1}^{(2)}x_{l-1} + \theta_{1,l}^{(2)}x_l) \qquad (3)$$

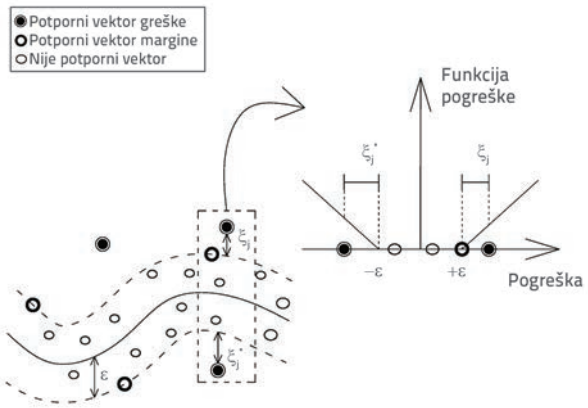
2.3. Metoda potpornih vektora

U klasifikacijskom problemu SVM za odabranu funkciju pronalazi parametre s kojima je funkcija optimalno udaljena od različitih klasa, a u regresijskom postupak se svodi na pronalazak funkcije koja na optimalan način opisuje podatke. Problem je često višedimenzionalan (u poglavlju 3 vidi se da je srednji mjesečni protok opisan kao funkcija barem 6 različitih prediktora) te složen za grafičko prikazivanje. SVM podatke shvaća kao potporne vektore koje aproksimira zadanom hipotezom minimizirajući grešku aproksimacije predviđane varijable. Pritom se unutar definirane margine, odnosno greške ε, mora nalaziti što više točaka. Pristranost i tolerancija količine devijacija većih od greške određuje se parametrom C (eng. *trade-off*), pozitivnom konstantom koja određuje stupanj penalizacije greške. Pristranost i varijanca određuju se kroz minimizaciju zbroja regularizacijskog dijela i greške gradnje modela u izrazu (4) [26, 27]:

$$\min\left(\frac{1}{2}\| w \|^2 + C\sum_{j=1}^{l}(\xi_j + \xi_j^*)\right)^2 \qquad (4)$$

uz uvjete:
$$y_j - \langle w, x_j \rangle - b \le \varepsilon + \xi_j$$
$$\langle w, x_j \rangle + b - y_j \le \varepsilon + \xi_j^*$$
$$\xi_j, \xi_j^* \ge 0$$

gdje su: x_j ulazne varijable iz skupa podataka, y_j varijabla koja se predviđa, w vektor iz prostora ulaznih varijabli, b "bias" varijabla, ξ_j, ξ_j^* "slack" varijable korištene za procjenu odstupanja ulaznih varijabli od margine.



Slika 2. Prikaz podataka (potpornih vektora), hipoteze i margine kod SVM-a (prilagođeno prema [28])

Hipoteza kojom se aproksimira predviđena varijabla je [28]:

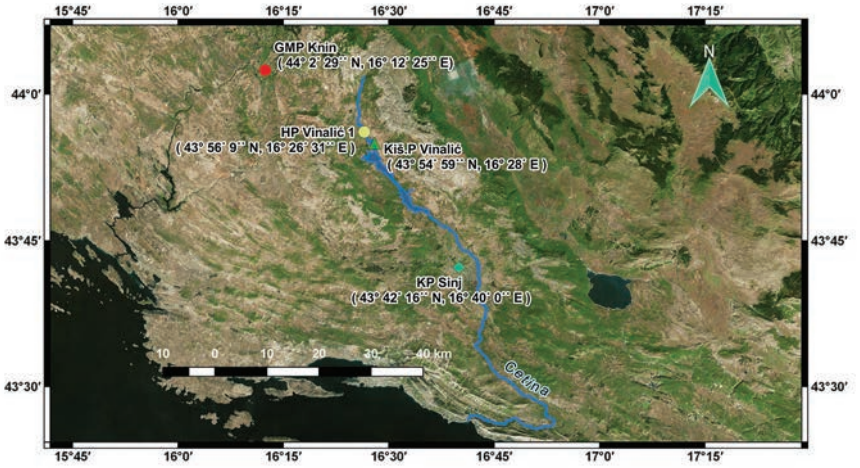
$$g(x) = \sum_{i,j=1}^{i}\left(\alpha_i - \overline{\alpha}_i\right)K\langle u_i u_j \rangle + b$$

gdje su α_i varijable proizašle iz prijelaza na dualni optimizacijski problem, a K oznaka za kernel. Programsko okruženje omogućuje odabir funkcije i parametara kernela (linearna, polinom i stupanj, radijalna osnovna funkcija) te parametra C koji utječu na preciznost predviđanja.

2.4. Metoda najbližih susjeda

Princip NNM-a svodi se na traženje skupa vrijednosti (u dijelu podataka za gradnju modela) najsličnijih zadanima (na dijelu podataka za predviđanje modela). Za to je potrebno pronaći udaljenosti između zadanih i najsličnijih postojećih točaka (k najbližih susjeda). Premalena količina susjeda podrazumijeva veću osjetljivost modela, a prevelika količina smanjenu točnost

zbog utjecaja udaljenijih susjeda. Nakon što nađe najbliže susjede, NNM izračuna srednju vrijednost predviđenih vrijednosti svakog pojedinog susjeda [25, 27]. Definiranje mjere udaljenosti (euklidska, Minkowski itd.) u radu nije značajno utjecalo na točnost rezultata. Uz broj susjeda, težine utjecaja susjeda (jednolike ili ovisne o udaljenosti) imaju značajan utjecaj na točnost modela. Okruženje omogućuje odabir četiri algoritma za pretraživanje najbližih susjeda: *ball tree, kd tree, brute* algoritma te *auto* odabira najboljeg od ova tri. Važni su jer je izračun udaljenosti među susjedima računalno zahtjevan. Brute pretražuje sve moguće opcije, što može biti dugotrajno za velik broj susjeda, a ostala dva koriste logiku stabala za pretraživanje. Kd tree je binarno stablo koje koristi logiku izbjegavanja računanja udaljenosti do točaka za koje se zna da su udaljenije (ako je točka A daleko od B, a C blizu B, onda je C daleko od A). Ovaj algoritam je neučinkovit pri korištenju D-dimenzionalnih mjera udaljenosti za D > 20 (broj varijabli prediktora > 20). Problem je riješen ball tree algoritmom koji, umjesto korištenja Kartezijevog koordinatnog sustava, udaljenosti računa u sfernom koordinatnom sustavu [27, 29].



Slika 3. Situacija i položaj postaja (Izvor kartografske podloge: QGIS, © 2007-2018 RDC ScanEx, http://kosmosnimki.ru/)

3. Korištene podloge i formiranje modela

3.1. Područje istraživanja

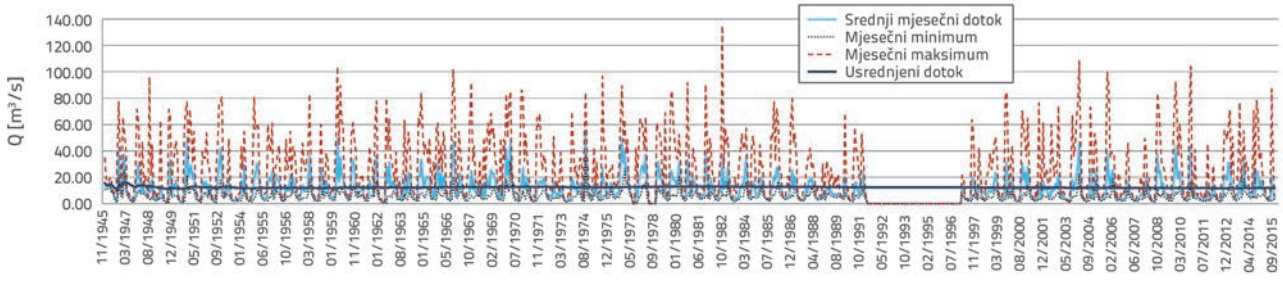
Metodologija i modeli su primijenjeni na mjerenjima protoka rijeke Cetine s postaje Vinalić 1. Na raspolaganju je bio povijesni niz dnevnih protoka od 1946. do 2015. godine, s prekidom u mjerenjima od 1991. do 1997. godine [30]. U načelu se ti protoci mogu shvatiti kao dotoci u akumulaciju Peruća, no s mjerom opreza jer je riječ o krškom području. Na raspolaganju su: akumulirana dnevna oborina (količina pale kiše) i srednja dnevna temperatura zraka (nadalje temperatura) s glavne meteorološke postaje Knin (250 m n.m.) u razdoblju od 1949. do 2015. godine, akumulirana dnevna oborina s kišomjerne postaje Vinalić (350 m n.m.) u razdoblju od 1951. do 2015. godine (prekid 1991. - 1997.) i srednja dnevna temperatura s klimatološke postaje Sinj (308 m n.m.) u razdoblju od 1949. do 2015. godine [31]. Situacija i položaj postaja može se vidjeti na slici 3., a srednji, minimalni, maksimalni te usrednjeni mjesečni protok na slici 4. Na protoke postaje Vinalić 1 formiranje akumulacije nije imalo znatan utjecaj. To je bitno razmotriti u svakoj statističkoj analizi pa i kod SL-a jer koristi princip učenja iz uzorka u podacima. SL može pokriti i promjene prirodno prisutnog protoka nastale izgradnjom uz prisutnost dovoljno instanci za izgradnju modela. Može se pretpostaviti da se dotoci s postaje Vinalić 1 mogu iskoristiti za dugoročnu analizu dostupnosti vode.

3.2. Formiranje modela

Prvi korak čini odabir ulaznih varijabli, tj. prediktora. To su varijable iz kojih se predviđa srednji mjesečni protok, a o tom pitanju u radu su primijenjena tri različita pristupa - predviđanje dotoka:

- iz samog dotoka
- iz oborine i temperature s jedne postaje (postaja Knin)
- iz oborine i temperature s po dvije postaje (oborina s postaja Knin i Vinalić te temperatura s postaja Knin i Sinj).

Prvi pristup može biti odgovarajući za kreiranje modela koji sintetički generira protok ili za predviđanje jedan mjesec unaprijed (eventualno dva do tri mjeseca uz određene preinake).



Slika 4. Dotoci na hidrološkoj postaji Vinalić 1

Tablica 1. Ulazni podaci korišteni u analizi

	Prvi pristup	Drugi i treći pristup	
Veličina	Q...protok [m³/s]	T...temperatura zraka [°C]	P...oborina [mm]
Oznaka	avm, min, max, yavm, avmmin, avmmax	avm, min, max, yavm, avmin, avmax	avm, acc, max, yavm, avacc, avmax
avm, min, max...srednja, minimalna i maksimalna mjesečna vrijednost; yavm, avmin, avmax...srednja, minimalna i maksimalna mjesečna vrijednost usrednjena po svim godinama; avmmin, avmmax...minimalna i maksimalna srednja mjesečna vrijednost po svim godinama acc...akumulirana mjesečna vrijednost; avacc...akumulirana mjesečna vrijednost usrednjena po svim godinama			

Drugi i treći pristup primjenjuju isključivo vanjske varijable te su pogodni za dugoročno planiranje. Kao ulazni podaci definirane su veličine u tablici 1.

Karakteristične veličine prikazane u tablici 1. predstavljaju varijable koje su bile na raspolaganju pri izboru konfiguracija modela. Primjerice, kod prve konfiguracije je jedna od potencijalnih ulaznih varijabli Q_{avmmin} – minimalna srednja mjesečna vrijednost protoka usrednjena po svim godinama. Od svih mjesečnih vrijednosti, njen minimum za razdoblje 1946. - 2015. iznosi 0,56 m³/s, srednja vrijednost 2,70 m³/s, a maksimalna vrijednost 5,51 m³/s, što je prikazano u tablici 2. Analogno vrijedi za ostale fizikalne veličine kod 2. i 3. konfiguracije.

U programskom okruženju napravljena je procedura na osnovi koje se obrađuju i pripremaju podaci za gradnju modela. Za svaki pristup analizirana je korelacija potencijalnih ulaznih varijabli (tablica 1.) sa srednjim mjesečnim protokom. U preliminarnom izboru ulaznih varijabli korištene su one s korelacijom barem 0,55-0,60. Za usrednjene varijable po godinama razmatrala se korelacija sa srednjim dotocima za svaku godinu posebno, a za ulazak u preliminarni izbor prag korelacije trebao je biti zadovoljen u barem 30-40 % vremenskog niza. Rezultat procedure je vremenska serija za postupak modeliranja.

Drugi korak je preliminarna izgradnja modela. S dobivenom vremenskom serijom ispituje se mogućnost modela AR, ANN, SVM i NNM da aproksimiraju protok. Preliminarno se odrede parametri modela u svojstvu minimizacije statističkih mjera greške.

Treći korak je provođenje analize osjetljivosti preciznosti modela za različite konfiguracije ulaznih podataka. Iz prethodno dobivene vremenske serije uklone se ili dodaju

Tablica 2. Statistika korištenih karakterističnih veličina po konfiguracijama

Protoci (Vinalić 1) (1946.-2015.)						
	Q_{min}	Q_{avm}	Q_{max}	Q_{avmmin}	Q_{yavm}	Q_{avmmax}
Min.	0,13	0,56	1,01	0,56	3,38	10,5
Sr. vr.	5,52	11,9	28,4	2,70	11,9	35,9
Maks.	36,9	55,9	135,0	5,51	19,8	55,9
St. dev.	4,04	9,48	24,2	1,52	5,78	15,3
N	765	765	765	12	12	12

Temperatura i oborina (Knin) (1949.-2015.)												
	T_{min}	T_{avm}	T_{max}	T_{avmin}	T_{yavm}	T_{avmax}	P_{avm}	P_{acc}	P_{max}	P_{yavm}	P_{avacc}	P_{avmax}
Min.	-12,4	-3,79	4,00	-5,20	3,49	9,9	0,00	0,00	0,00	0,00	0,00	0,00
Sr. vr.	7,04	13,1	18,9	7,09	13,2	18,9	2,91	88,3	29,1	2,89	87,8	28,1
Maks.	23,2	26,9	31,9	19,3	24,7	28,5	11,5	354	155	8,03	241	63,7
St. dev.	7,48	6,89	6,17	7,06	6,69	5,85	1,94	58,9	18,8	0,94	28,7	6,33
N	731	731	731	732	732	732	732	732	732	732	732	732

Temperatura (Sinj); oborina (Vinalić) (1951.-2015.)												
	Tw_{min}	Tw_{avm}	Tw_{max}	Tw_{avmin}	Tw_{yavm}	Tw_{avmax}	H_{avm}	H_{acc}	H_{max}	H_{yavm}	H_{avacc}	H_{avmax}
Min.	-16,7	-3,13	3,4	-3,82	2,89	8,03	0,00	0,00	0,00	0,43	8,15	6,1
Sr. vr.	7,02	12,7	18,1	6,87	12,6	17,9	2,89	87,9	27,9	2,95	86,9	28,4
Maks.	22,0	26,0	30,4	19,2	23,8	27,6	11,9	356	140	6,53	196	51,2
St. dev.	7,51	6,85	6,22	7,07	6,62	5,92	2,07	62,8	17,3	0,87	27,7	5,33
N	719	719	719	720	720	720	701	701	701	713	713	713

Tablica 3. Odabrani parametri modela SL-a

Pristup	ANN				SVM				NNM		
	Akt. funkcija	Broj čv. u skrivenom sloju	Početni int. učenja	Tolerancija	Kernel	St.	C	γ	Br. susjeda	Težine	Algoritam
1	tang. hip.	25	$5{,}0*10^{-5}$	$2*10^{-9}$	polinom	1	100,0	3,0	7	Jednol.	*brute*
2	tang. hip.	45	$2{,}1*10^{-3}$	$2*10^{-9}$	rbf	/	56,73	0,009	10	Udalj.	auto
3	rektif.	30	$2{,}2*10^{-3}$	$2*10^{-9}$	rbf	/	56,73	0,009	10	Udalj.	auto

pojedine varijable s ciljem povećanja preciznosti modela. Pokazalo se da uglavnom varijable visoko korelirane s protokom imaju većinski doprinos preciznosti modela. No, previše varijabli moglo bi umanjiti preciznost, a pojedine visoko korelirane varijable nisu značajno pridonosile preciznosti te su uklonjene. Ponekad su nešto slabije korelirane varijable značajnije pridonosile povećanju preciznosti modela (npr. $T_{avmin\text{-}2}$). Rezultat trećeg koraka su odabrane konfiguracije modela:

Q_{avm} = f ($Q_{avm\text{-}1}$, $Q_{min\text{-}11}$, $Q_{min\text{-}1}$, $Q_{max\text{-}11}$, Q_{yavm}, $Q_{avmmin\text{-}11}$)
Q_{avm} = f ($T_{avm\text{-}11}$, $T_{avmin\text{-}2}$, $P_{avm\text{-}1}$, P_{avm}, $P_{acc\text{-}11}$, $P_{acc\text{-}2}$, $P_{acc\text{-}1}$, P_{max}, $P_{avacc\text{-}2}$)
Q_{avm} = f ($H_{avm\text{-}11}$, H_{avm}, $H_{acc\text{-}11}$, $H_{acc\text{-}1}$, H_{acc}, H_{max}, H_{yavm}, $H_{avacc\text{-}11}$, H_{avacc}, $T_{avm\text{-}11}$, $T_{avmin\text{-}2}$, $P_{avm\text{-}1}$, P_{avm}, $P_{acc\text{-}11}$, $P_{acc\text{-}2}$, $P_{acc\text{-}1}$, $P_{avacc\text{-}2}$, P_{avacc}, $Tw_{avm\text{-}11}$, $Tw_{avm\text{-}2}$, Tw_{avm})

Kod drugog i trećeg pristupa *T* i *P* se odnose na temperaturu i oborinu s postaje Knin, *Tw* na temperaturu s postaje Sinj, a *H* na oborinu s postaje Vinalić.
Sljedeći korak bio je optimizacija parametara modela. Kod AR-a određene su vrijednosti parametra *t* koje daju najbolje rezultate modeliranih protoka. Po definiciji *t* ima normalnu raspodjelu, a opisuje varijaciju mjesečnog protoka u odnosu na srednju vrijednost. Kod ANN-a ispitan je utjecaj različitih aktivacijskih funkcija (hiperbolički tangens, logistička, identitetska i rektifikacijska funkcija), broja čvorova skrivenog sloja, početnog intenziteta učenja i tolerancije na preciznost. Kod SVM-a ispitan je utjecaj funkcije kernela (linearna, polinomijalna, radijalna osnovna funkcija) i njenog stupnja te parametara *C* i γ. Kod NNM-a variran je broj susjeda, raspodjele težine po susjedima te algoritam računanja udaljenosti. S parametrima određenima u ovom koraku (tablica 3.) provedena je analiza na svim nizovima različitih duljina.
Vremenske serije su pri izgradnji modela uvijek dijeljene kronološki: prvih 60 % godina za izgradnju modela, 20 % sljedećih za kalibraciju te 20 % za verifikaciju modela. U prvom pokusu se za sve pristupe koristila maksimalna količina podataka na raspolaganju: redom 65, 62 i 60 godina. U drugom su uklonjene zadnje godine iz podataka tako da je korišteno redom 60, 60 i 55 godina. Nadalje je uklanjano po 5 godina sve dok nije ostalo 10 godina podataka. Godine koje nisu korištene za postupak izgradnja-kalibracija-verifikacija iskorištene su za dodatnu verifikaciju modela. Tako je u drugom pokusu ostalo 5, 2 i 5 godina za dodatnu verifikaciju, a u zadnjem 55, 52 i 50 godina (tablice 4. do 6.).

3.3. Statističke mjere greške

Pri optimiziranju modela najviše se vodilo računa o postizanju što veće korelacije, manjeg korijena srednje kvadratne greške i većeg koeficijenta determinacije. Koeficijent korelacije *R* predstavlja međupovezanost izmjerene i predviđene varijable. Područje 0-0,25 označava slabu, 0,25-0,6 srednje jaku, a 0,6-1,0 čvrstu korelaciju [32]. Visoke vrijednosti koeficijenta korelacije ne znače nužno da je izgrađeni model sposoban dobro generalizirati. Stoga su korištene i ostale mjere greške: korijen srednje kvadratne greške (eng. *root mean squared error* - RMSE), srednja apsolutna greška (eng. *mean absolute error* - MAE), relativna apsolutna greška (*eng. relative absolute error* - RAE), korijen relativne kvadratne greške (eng. *root relative squared error - RRSE*), koeficijent determinacije ili učinkovitosti (eng. *coefficient of determination* - R^2). Zbog ograničenosti prostora, tablično su prikazani samo R^2 i *RMSE*. Korišteni R^2 je mjera izglednosti predviđanja neviđenih vrijednosti modelom i nije nužna kvadratna vrijednost od *R* (postoji više definicija) te može biti negativan ako model proizvoljno loše predviđa. Vrijednost 1,0 označava apsolutno precizno predviđanje [24]. Izrazi za navedene mjere mogu se naći u istraživanjima područja (primjerice [10, 13, 24, 33]).

4. Rezultati i rasprava

4.1. Prvi pristup

U prvom pristupu pokazano je da AR(1), uz optimizaciju parametra *t*, može dohvatiti visoki raspon vrijednosti protoka. Ti rezultati su svrstani u područje gradnje i kalibracije modela. Parametar *t* po definiciji ima normalnu raspodjelu, no aproksimacijom vrijednosti *t* normalnom raspodjelom točnost rezultata je smanjena pri verifikaciji (tablica 4.). Naime, varijabilnost protoka nije uvijek normalne raspodjele (npr. [18]), a među optimiziranim vrijednostima *t* postoje diskontinuiteti. Modeli SL-a polučili su čvrstu korelaciju, ali s niskom količinom točno opisanih vrijednosti. Globalno prate protoke, ali značajno podcjenjuju vršne vrijednosti (visoki *RMSE*, *MAE*, *RRSE* i *RAE*). Postižu dobru korelaciju pri verifikaciji i verifikaciji izvan duljine niza, dok AR(1) postiže slabiju korelaciju izvan duljine niza. Koeficijenti determinacije iznosa 0,3-0,4 i niže nisu zadovoljavajući. Za sve pristupe prikazani su modelirani i izmjereni dotoci za duljinu povijesnog niza od 45 godina (slike 5., 7. i 9.), zbog zadovoljavajuće preciznosti

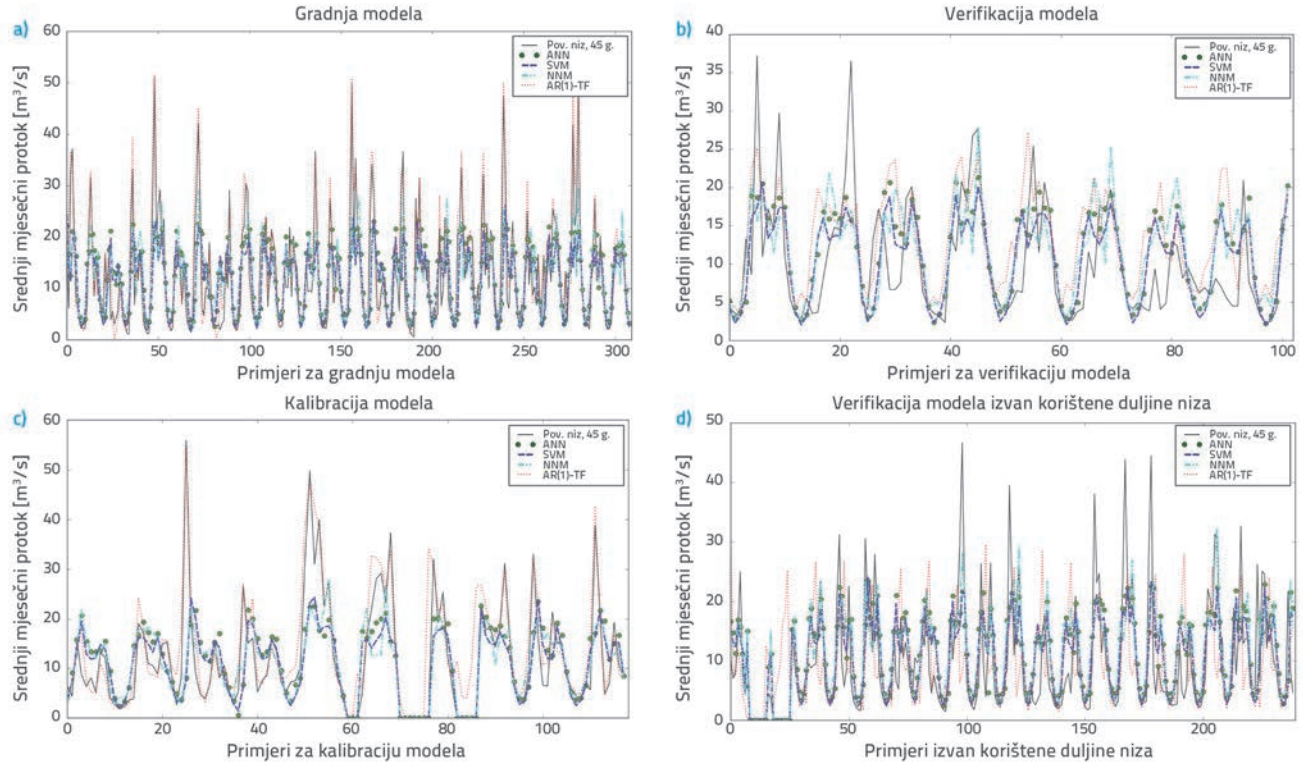
Tablica 4. Koeficijent determinacije i korijen srednje kvadratne greške modela AR i SL-a, prvi pristup

Dulj. niza [god]	Gradnja modela								Kalibracija modela								Verifikacija modela								Verifikacija izv. korištene dulj. n.							
	R^2 (*RMSE*)								R^2 (*RMSE*)								R^2 (*RMSE*)								R^2 (*RMSE*)							
	AR		ANN		SVM		NNM		AR		ANN		SVM		NNM		AR		ANN		SVM		NNM		AR		ANN		SVM		NNM	
65	0,85	(3,96)	0,39	(7,77)	0,38	(7,89)	0,51	(6,97)	0,75	(3,6)	0,35	(5,54)	0,44	(5,14)	0,22	(6,07)	0,49	(7,03)	0,41	(7,51)	0,43	(7,43)	0,33	(8,02)	/	(/)	/	(/)	/	(/)	/	(/)
60	0,84	(4,09)	0,39	(7,88)	0,38	(7,92)	0,52	(6,97)	0,75	(3,71)	0,32	(5,83)	0,39	(5,55)	0,16	(6,5)	0,45	(7,31)	0,45	(7,32)	0,48	(7,11)	0,4	(7,63)	-0,11	(8,48)	0,15	(7,41)	0,19	(7,26)	0,09	(7,67)
55	0,86	(3,72)	0,4	(7,77)	0,37	(7,93)	0,51	(6,99)	0,69	(5,08)	0,43	(6,7)	0,37	(7,04)	0,36	(7,12)	0,22	(7,45)	0,38	(6,51)	0,42	(6,28)	0,4	(6,37)	0	(9,28)	0,4	(7,2)	0,41	(7,13)	0,28	(7,88)
50	0,86	(3,66)	0,41	(7,64)	0,39	(7,76)	0,52	(6,85)	0,72	(5,52)	0,33	(8,2)	0,33	(8,2)	0,35	(8,09)	0,1	(6,89)	0,41	(5,32)	0,47	(5,04)	0,34	(5,63)	0,02	(9,24)	0,39	(7,27)	0,42	(7,11)	0,33	(7,64)
45	0,87	(3,6)	0,45	(7,33)	0,42	(7,47)	0,55	(6,58)	0,75	(5,5)	0,25	(9,15)	0,27	(9,04)	0,27	(9,02)	0,01	(7,41)	0,32	(6,13)	0,37	(5,88)	0,2	(6,64)	-0,01	(9,15)	0,4	(6,98)	0,42	(6,85)	0,37	(7,16)
40	0,86	(3,66)	0,44	(7,32)	0,43	(7,38)	0,52	(6,77)	0,89	(3,66)	0,26	(9,31)	0,27	(9,24)	0,33	(8,28)	0,36	(7,89)	0,36	(7,47)	0,34	(7,6)	0,24	(8,15)	-0,21	(9,53)	0,38	(6,75)	0,41	(6,6)	0,39	(6,7)
35	0,84	(3,95)	0,44	(7,32)	0,43	(7,41)	0,51	(6,88)	0,83	(3,99)	0,37	(7,61)	0,34	(7,8)	0,41	(7,42)	0,35	(9,3)	0,25	(9,6)	0,2	(9,92)	0,23	(9,78)	-0,3	(9,88)	0,36	(6,83)	0,41	(6,6)	0,36	(6,87)
30	0,87	(3,56)	0,4	(7,68)	0,41	(7,63)	0,49	(7,07)	0,83	(3,89)	0,31	(7,88)	0,31	(7,88)	0,31	(7,9)	0,1	(8,97)	0,22	(8,31)	0,21	(8,37)	0,21	(8,38)	-0,09	(9,73)	0,4	(7,14)	0,42	(7,01)	0,39	(7,19)
25	0,86	(3,75)	0,38	(7,93)	0,43	(7,56)	0,5	(7,13)	0,83	(3,83)	0,37	(7,34)	0,39	(7,22)	0,38	(7,26)	0,54	(7,15)	0,27	(8,96)	0,33	(8,59)	0,28	(8,89)	-0,13	(9,83)	0,36	(7,26)	0,39	(7,12)	0,33	(7,47)
20	0,86	(3,7)	0,29	(8,26)	0,43	(7,4)	0,5	(6,93)	0,85	(4,06)	0,24	(9,16)	0,4	(8,11)	0,36	(8,36)	0,58	(6,26)	0,24	(8,37)	0,41	(7,38)	0,37	(7,67)	-0,15	(10,11)	0,26	(8,01)	0,37	(7,37)	0,32	(7,68)
15	0,84	(4,2)	0,39	(8,14)	0,46	(7,65)	0,51	(7,32)	0,65	(3,77)	0,34	(5,19)	0,12	(5,98)	0,16	(5,86)	0,64	(7,23)	0,29	(10,19)	0,45	(8,99)	0,32	(9,96)	-0,22	(10,39)	0,33	(7,57)	0,36	(7,4)	0,31	(7,73)
10	0,75	(5,17)	0,33	(8,45)	0,48	(7,44)	0,42	(7,83)	0,66	(6,27)	0,23	(9,42)	0,23	(9,4)	0,24	(9,34)	0,5	(6,6)	0,36	(7,48)	0,53	(6,41)	0,55	(6,22)	-0,63	(12,09)	0,25	(8,12)	0,35	(7,55)	0,33	(7,65)

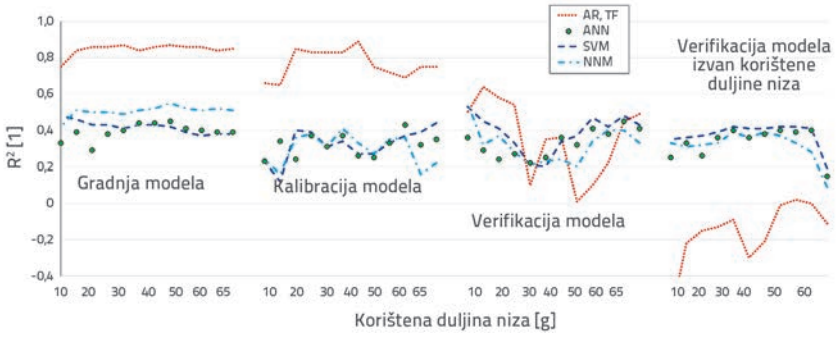
(u drugom i trećem pristupu) i mogućnosti dugoročnog planiranja do 15 godina.

Područje u kojem nedostaju podaci prikazani su vrijednošću 0. Analiza rezultata pokazuje da je R^2 prilikom gradnje i kalibracije modela zadovoljavajući samo za AR(1) model ($R^2 > 0,65$), dok je za ostale u nižem području srednje jačine ($R^2 < 0,45$), s iznimkom nešto većih vrijednosti kod modela NNM pri gradnji ($R^2 < 0,55$). Verifikacija je za većinu modela u nižem području srednje jačine koeficijenta determinacije ($R^2 < 0,50$), kao i verifikacija izvan korištene duljine niza za sve modele ($R^2 < 0,43$). Na temelju navedenoga može se zaključiti da se taj pristup ne preporučuje, osim uz eventualno uvođenje poboljšanja gradnjom hibridnih modela, primjerice analizirajući singularni spektar [33]. Kako je u radu naglasak i na dugoročnom planiranju, potrebno je poslužiti se ostalim pristupima. Koeficijent determinacije i korijen srednje kvadratne greške modela AR i SL-a za prvi pristup dani su u tablici 4.

Na slici 6. dan je grafički prikaz mjera R^2 ovisno o duljini niza za prvi pristup. Kod AR(1) pažnju treba obratiti na parametar *t*, što se



Slika 5. Primjeri uz prvi prostup: a) Izgradnja; b) kalibracija; c) verifikacija; d) verifikacija izvan korištene duljine niza modela za povijesni niz duljine 45 godina



Slika 6. R^2 na svim dijelovima podataka u odnosu na korištenu duljinu niza, prvi pristup

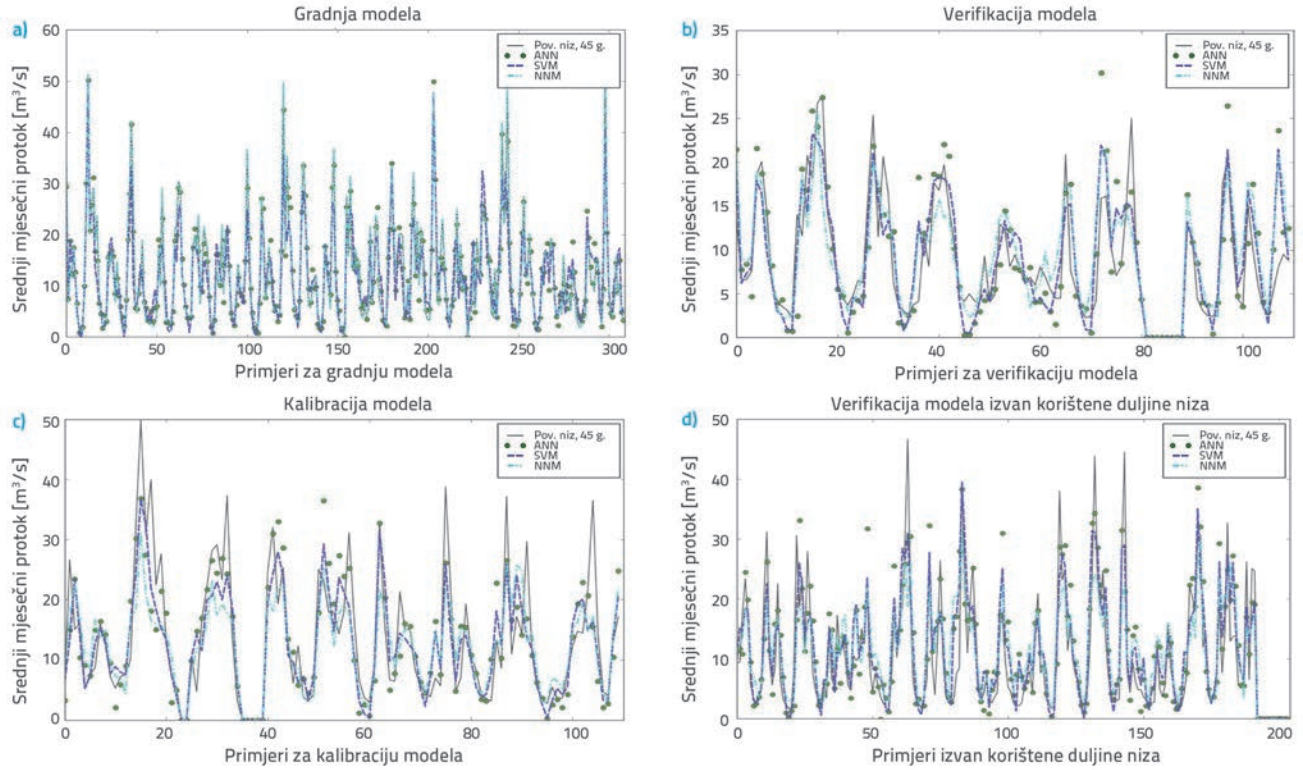
vidi na verifikacijskim dijelovima. Prema zamisli ovog istraživanja, modeli bi se trebali primjenjivati i za dugoročno planiranje, pa je najbolje pratiti mjere greške na verifikacijskim dijelovima. AR(1) ne koristi vanjske prediktore te nije primijenjen u ostala dva pristupa. Kod SL-a, NNM preciznije opisuje protoke pri izgradnji modela, ali preciznost nije očuvana pri kalibraciji i verifikaciji. U drugom i trećem pristupu, a mjestimično i u prvom, ANN i NNM daju veću točnost pri gradnji modela nego SVM. No, kod SVM-a je točnost očuvana pri kalibraciji i verifikaciji. Najpovoljnije kombinacije mjera greške (najveće vrijednosti R^2 i najmanje vrijednosti *RMSE*) dobivene su modelom SVM, za svaku duljinu niza. SVM u drugom i trećem pristupu pokazuje i najmanju varijabilnost mjera greške ovisno o duljini niza. ANN ima vrlo velik izbor parametara i oblikovanja mreže te je moguće da bi se iscrpnim pretraživanjem postigla veća točnost, što može biti vremenski zahtjevno.

4.2. Drugi pristup

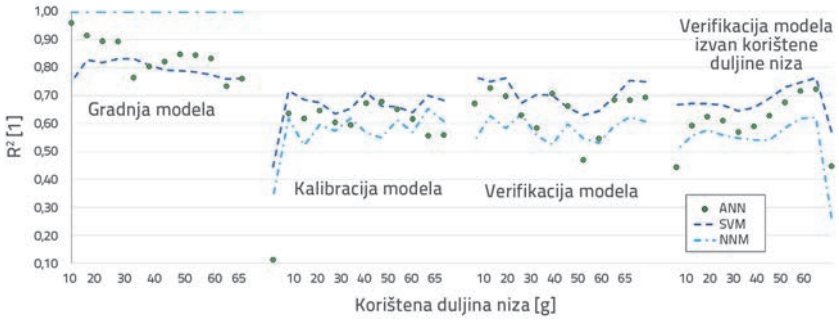
Promijenjene su ulazne varijable, a krajnje korišteni modeli promijenjenih su parametara u odnosu na prvi pristup. Kod NNM-a su, primjenom raspodjele težina "po susjedima" ovisno o udaljenosti, potpuno točno opisani dotoci pri gradnji modela. Naravno, pri kalibraciji i verifikaciji točnost je smanjena. To je važno jer model koji jako dobro aproksimira podatke za gradnju, nema nužno i dobru sposobnost generalizacije na ostalim podacima. No, u ovom slučaju nije postignuta preprilagođenost (eng. *overfitting*) jer se podjednaka točnost postiže pri kalibraciji i verifikaciji (ali ne i gradnji) s jednolikom raspodjelom težina. Kernel s radijalnom osnovnom funkcijom davao je najveću točnost SVM-a. Korelacija svih modela se na verifikaciji i verifikaciji izvan duljine niza nalazi u području čvrste korelacije ($R^2 > 0,44$), izuzev jednog modela NNM ($R^2 = 0,26$) koji se odnosi na najkraći niz predviđanja od dvije godine. Također je potrebno naglasiti da je R^2 za duljinu niza od 45 do 55 godina veći u području verifikacije izvan duljine niza nego u slučaju kalibracije i verifikacije za sve modele. Ekstremne vrijednosti su precijenjene, odnosno podcijenjene, na većem dijelu verifikacije (visoki *RMSE*, *MAE*, *RRSE*, *RAE*). S druge strane, vrijednosti mjere *RMSE* u rasponu su vrijednosti 4,9-7,05 m³/s, što upućuje na značajno povećanje točnosti u odnosu na prvi pristup s rasponima 6,6-12,09 m³/s. Bitna je značajka svih modela da globalno opišu narav protoka, pa i na verifikaciji izvan duljine niza. Kod modela s manjim duljinama niza vidi se da nije trebalo mnogo podataka za prepoznavanje naravi protoka (65 primjera za gradnju modela kod duljine 10 godina). Najpovoljnije mjere greške postignute su ponovno modelom SVM za sve duljine niza. Pretpostavljeno je da se dodavanjem podataka s dvije preostale postaje (u trećem pristupu) mogu izgraditi kvalitetniji modeli. Koeficijent determinacije i korijen srednje kvadratne greške modela SL za drugi pristup dani su u tablici 5.

Tablica 5. Koeficijent determinacije i korijen srednje kvadratne greške SL modela, drugi pristup

Duljina niza [god]	Gradnja modela						Kalibracija modela						Verifikacija modela						Verifikacija izv. korištene dulj. n.					
	R^2 (*RMSE*)						R^2 (*RMSE*)						R^2 (*RMSE*)						R^2 (*RMSE*)					
	ANN		SVM		NNM		ANN		SVM		NNM		ANN		SVM		NNM		ANN		SVM		NNM	
62	0,76	(4,88)	0,76	(4,88)	1,00	(0)	0,56	(4,58)	0,68	(3,89)	0,61	(4,32)	0,69	(5,53)	0,75	(5)	0,61	(6,25)	/	(/)	/	(/)	/	(/)
60	0,73	(5,17)	0,76	(4,91)	1,00	(0)	0,56	(4,66)	0,70	(3,83)	0,65	(4,12)	0,68	(5,46)	0,75	(4,83)	0,62	(5,96)	0,45	(6,74)	0,57	(5,94)	0,26	(7,78)
55	0,83	(4,14)	0,77	(4,81)	1,00	(0)	0,62	(4,46)	0,64	(4,32)	0,57	(4,72)	0,68	(4,96)	0,70	(4,88)	0,59	(5,65)	0,72	(5,31)	0,76	(4,91)	0,62	(6,21)
50	0,84	(4,02)	0,78	(4,74)	1,00	(0)	0,65	(4,97)	0,66	(4,91)	0,61	(5,24)	0,55	(4,48)	0,64	(3,96)	0,53	(4,55)	0,72	(5,41)	0,75	(5,11)	0,62	(6,28)
45	0,85	(3,86)	0,79	(4,54)	1,00	(0)	0,68	(5,89)	0,66	(6,03)	0,55	(6,97)	0,47	(4,55)	0,63	(3,8)	0,55	(4,21)	0,68	(5,38)	0,73	(4,91)	0,58	(6,09)
40	0,82	(4,17)	0,79	(4,5)	1,00	(0)	0,67	(6,16)	0,71	(5,78)	0,57	(7,09)	0,66	(4,74)	0,66	(4,78)	0,60	(5,17)	0,63	(5,36)	0,69	(4,9)	0,54	(5,96)
35	0,80	(4,4)	0,81	(4,34)	1,00	(0)	0,59	(5,79)	0,65	(5,36)	0,62	(5,62)	0,71	(5,95)	0,70	(6,01)	0,52	(7,58)	0,59	(5,48)	0,66	(5,01)	0,54	(5,8)
30	0,76	(4,96)	0,83	(4,2)	1,00	(0)	0,60	(5,61)	0,63	(5,39)	0,57	(5,82)	0,58	(7,2)	0,70	(6,08)	0,56	(7,43)	0,57	(5,75)	0,64	(5,22)	0,55	(5,89)
25	0,89	(3,28)	0,83	(4,13)	1,00	(0)	0,65	(5,53)	0,67	(5,31)	0,60	(5,92)	0,63	(5,65)	0,67	(5,31)	0,63	(5,62)	0,61	(5,81)	0,67	(5,38)	0,56	(6,19)
20	0,89	(3,16)	0,82	(4,15)	1,00	(0)	0,62	(6,06)	0,69	(5,51)	0,52	(6,79)	0,70	(5,68)	0,76	(5,04)	0,58	(6,68)	0,62	(5,71)	0,67	(5,35)	0,58	(6,06)
15	0,92	(2,87)	0,83	(4,1)	1,00	(0)	0,64	(5,54)	0,72	(4,88)	0,62	(5,68)	0,73	(6,01)	0,75	(5,75)	0,63	(7,02)	0,59	(5,95)	0,67	(5,35)	0,56	(6,22)
10	0,96	(1,82)	0,75	(4,55)	1,00	(0)	0,11	(7,95)	0,44	(6,29)	0,34	(6,87)	0,67	(7,13)	0,77	(6,01)	0,54	(8,44)	0,44	(7,05)	0,67	(5,46)	0,50	(6,67)



Slika 7. Primjeri uz drugi pristup: a) Izgradnja; b) kalibracija; c) verifikacija; d) verifikacija izvan korištene duljine niza modela za povijesni niz duljine 45 godina



Slika 8. R^2 na svim dijelovima podataka u odnosu na korištenu duljinu niza, drugi pristup

4.3. Treći pristup

Preciznost modela je povećana u pogledu svih statističkih mjera. Kod ANN-a se najboljom pokazala rektifikacijska funkcija kao aktivacijska. S modelom NNM dogodila se identična stvar kao i u drugom pristupu. Također je i ANN za 10-20 godina dao savršenu točnost pri izgradnji, a znatno smanjenu na ostalim dijelovima. U slučaju ANN-a riječ je o preprilagođenosti te bi sa smanjenjem primjera za gradnju modela trebalo uložiti dodatnu energiju za pretraživanje odgovarajuće arhitekture mreže. SVM je najpreciznji te pokazuje sposobnost očuvanja mjera greške na svim dijelovima podataka. Korelacija na verifikaciji izvan duljine niza za SVM je kod svih slučajeva jednaka ili veća od 0,82. *RMSE* i *MAE*, u rasponu 3,82-5,36 m³/s te 2,95-4,02 m³/s, *RRSE* i *RAE*, u rasponu 0,45-0,57 i 0,42-0,55, najmanji su, a R^2 iznosa 0,67-0,83 je najveći. Također je potrebno naglasiti da je R^2 za sve duljine niza veći u području verifikacije izvan duljine niza nego u slučaju kalibracije i verifikacije za sve modele (jedina iznimka su modeli SVM i NNM za najkraći niz predviđanja od 5 godina). Koeficijent determinacije i korijen srednje kvadratne greške modela SL za treći pristup dani su u tablici 6. Na slici 9. može se vidjeti da svi modeli globalno prate izmjereni protok, NNM najslabije dohvaća

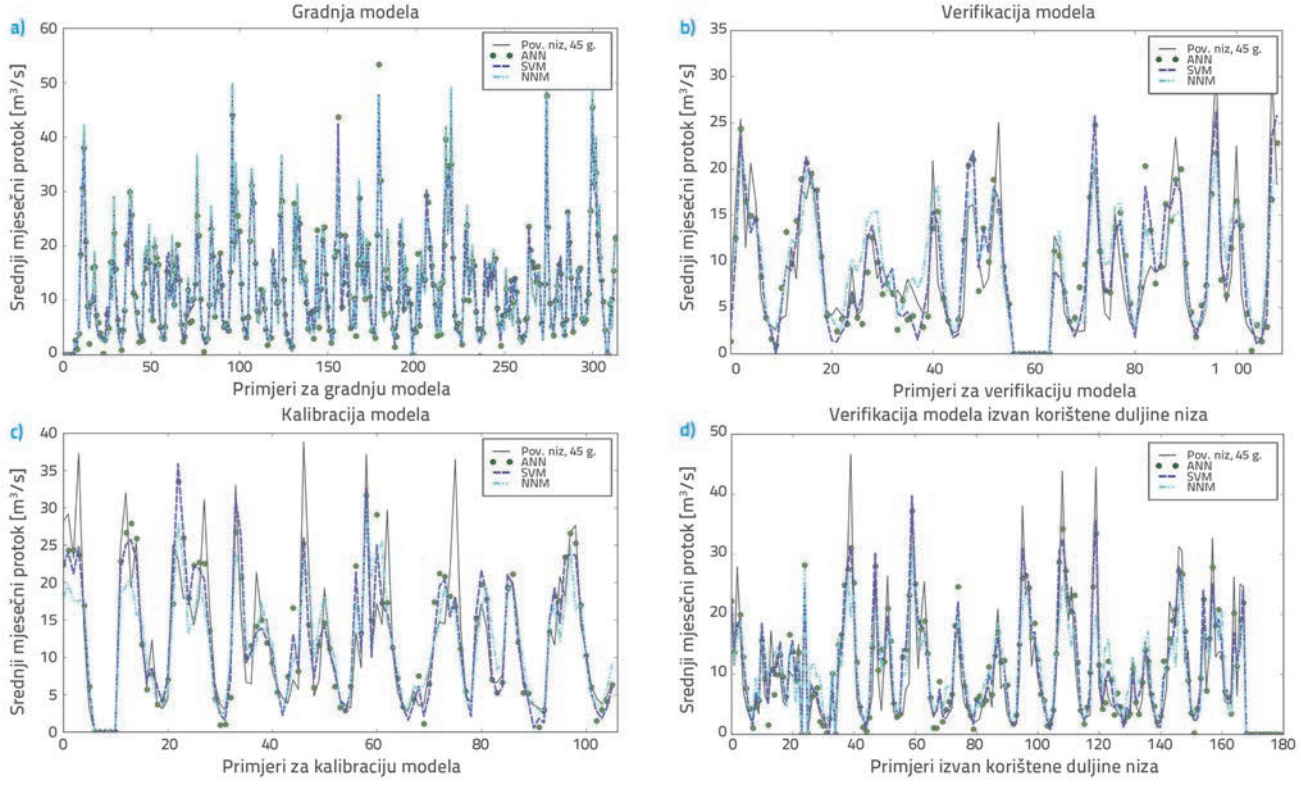
Promatrajući R^2 u drugom pristupu, nizovi 40-60 godina daju veću točnost ($R^2 > 0,7$), dok kraći nizovi ne rezultiraju značajno manjim vrijednostima ($0,65 < R^2 < 0,7$). Na temelju rezultata se može utvrditi sljedeće: odabir prediktora ključan je dio u korištenju SL-a za predviđanje protoka, moguće je odrediti dosad neviđene protoke na temelju predviđanja iz oborine (većinom) i temperature s takvom točnošću da se uz pomoć prediktora može utvrditi načelna dostupnost vode u vremenu. Kod trećeg pristupa, na verifikacijskom dijelu izvan duljine niza, R^2 kod SVM-a za > 40 godina prelazi 0,8, a za 20 do 40 godina je u rasponu 0,7-0,8. Modelirani i izmjereni dotoci za duljinu povijesnog niza od 45 godina za drugi pristup prikazan je na slici 7. Na slici 8. dan je grafički prikaz mjera R^2 ovisno o duljini niza za drugi pristup.

Tablica 6. Koeficijent determinacije i korijen srednje kvadratne greške SL modela, treći pristup

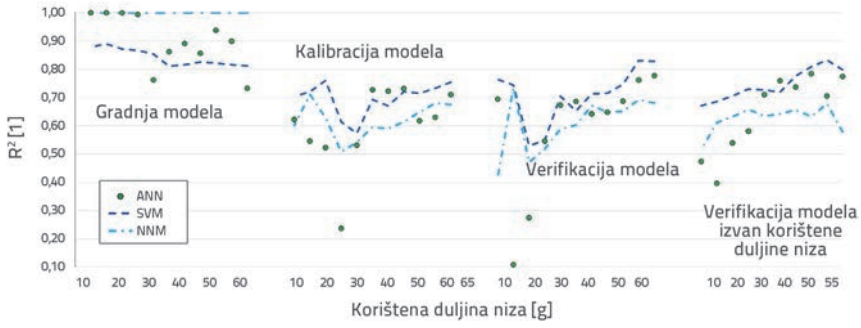
Duljina niza [god]	Gradnja modela						Kalibracija modela						Verifikacija modela						Verifikacija izv. korištene dulj. n.					
	R^2 (*RMSE*)						R^2 (*RMSE*)						R^2 (*RMSE*)						R^2 (*RMSE*)					
	ANN		SVM		NNM		ANN		SVM		NNM		ANN		SVM		NNM		ANN		SVM		NNM	
60	0,73	(5,08)	0,81	(4,27)	1,00	(0)	0,71	(3,61)	0,75	(3,32)	0,67	(3,82)	0,78	(4,76)	0,83	(4,19)	0,68	(5,7)	/	(/)	/	(/)	/	(/)
55	0,90	(3,16)	0,82	(4,27)	1,00	(0)	0,63	(4,25)	0,73	(3,61)	0,68	(3,96)	0,76	(4,89)	0,83	(4,12)	0,69	(5,56)	0,77	(4,04)	0,80	(3,82)	0,58	(5,52)
50	0,94	(2,5)	0,82	(4,23)	1,00	(0)	0,62	(4,79)	0,71	(4,13)	0,65	(4,59)	0,69	(4,81)	0,75	(4,33)	0,65	(5,08)	0,71	(5,21)	0,83	(3,93)	0,68	(5,45)
45	0,86	(3,77)	0,82	(4,16)	1,00	(0)	0,73	(4,87)	0,72	(4,97)	0,61	(5,84)	0,65	(4)	0,71	(3,6)	0,65	(3,99)	0,78	(4,45)	0,81	(4,19)	0,63	(5,81)
40	0,89	(3,23)	0,81	(4,2)	1,00	(0)	0,72	(5,52)	0,67	(6,02)	0,59	(6,71)	0,64	(4,14)	0,71	(3,7)	0,67	(3,96)	0,74	(4,69)	0,78	(4,33)	0,66	(5,36)
35	0,86	(3,64)	0,81	(4,26)	1,00	(0)	0,73	(5,71)	0,69	(6,06)	0,60	(6,95)	0,69	(4,92)	0,65	(5,18)	0,60	(5,54)	0,76	(4,25)	0,72	(4,59)	0,64	(5,19)
30	0,76	(4,72)	0,85	(3,7)	1,00	(0)	0,53	(7)	0,57	(6,68)	0,54	(6,92)	0,67	(6,09)	0,70	(5,79)	0,59	(6,83)	0,71	(4,65)	0,73	(4,52)	0,63	(5,22)
25	0,99	(0,76)	0,87	(3,53)	1,00	(0)	0,24	(9,25)	0,61	(6,58)	0,51	(7,43)	0,55	(5,69)	0,55	(5,67)	0,52	(5,86)	0,58	(5,99)	0,73	(4,81)	0,66	(5,41)
20	1,00	(0,24)	0,87	(3,5)	1,00	(0)	0,52	(6,84)	0,76	(4,85)	0,63	(6,04)	0,27	(8,54)	0,53	(6,88)	0,47	(7,3)	0,54	(6,21)	0,71	(4,96)	0,63	(5,55)
15	1,00	(0,02)	0,89	(3,35)	1,00	(0)	0,55	(6,06)	0,72	(4,76)	0,71	(4,84)	0,11	(8,53)	0,74	(4,58)	0,74	(4,65)	0,40	(7,29)	0,69	(5,26)	0,61	(5,85)
10	1,00	(0)	0,88	(3,07)	1,00	(0)	0,62	(5,44)	0,70	(4,82)	0,60	(5,59)	0,69	(7,36)	0,76	(6,47)	0,42	(10,15)	0,47	(6,78)	0,67	(5,36)	0,52	(6,5)

varijabilnost vrijednosti protoka. ANN podcjenjuje i minimume i maksimume. SVM pokazuje najveću sklonost generaliziranju, ali svi modeli ne uspijevaju doseći lokalne maksimume.

SVM u trećem pristupu najpogodniji je za dugoročnu analizu dostupnosti vode, U svakom slučaju, uputno je napraviti analizu osjetljivosti modela na duljinu niza. SVM je u tome stabilniji od ANN-a i od NNM-a. Dakle, kad je riječ o točnosti i stabilnosti, SVM se može preporučiti za daljnju upotrebu. Modelirani i izmjereni dotoci za duljinu povijesnog niza od 45 godina za treći pristup prikazan je na slici 9. Na slici 10. dan je grafički prikaz mjera R^2 ovisno o duljini niza za treći pristup. Dodatnim uključivanjem podataka s obližnjih postaja zasigurno bi se povećala preciznost modela. No, cilj strojnog učenja je da se sa što manje ulaznih varijabli izgradi dobar model, a također i realne situacije nisu



Slika 9. Primjeri uz treći pristup: a) Izgradnja; b) kalibracija; c) verifikacija; d) verifikacija izvan korištene duljine niza modela za povijesni niz duljine 45 godina



Slika 10. R^2 na svim dijelovima podataka u odnosu na korištenu duljinu niza, treći pristup

uvijek takve da je na raspolaganju znatan broj obližnjih postaja. Uključivanje broja dana u mjesecu s određenom količinom pale oborine također bi moglo pridonijeti preciznosti, Nakon utvrđivanja najbolje konfiguracije modela, poboljšanja se mogu postići spektralnom analizom, valićnom metodom, analizom kaosa, faznom rekonstrukcijom prostora itd, (vidjeti [12, 13, 29]).

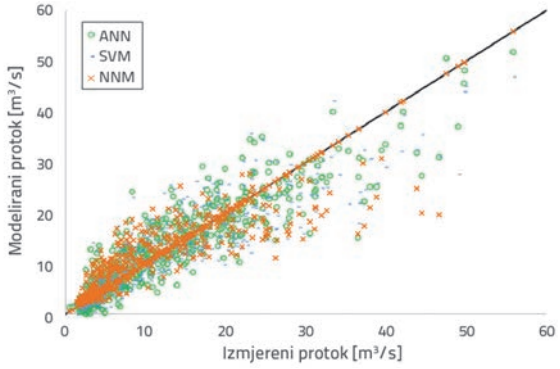
4.4. Statistička analiza rezultata

U radu su izračunane osnovne statističke značajke modela, a prikazane su u tablici 7. U prvom pristupu može se primijetiti značajno podcjenjivanje maksimalnih vrijednosti modela ANN, SVM i NNM, dok AR(1) pokazuje ukupno manja odstupanja a veća u vezi sa srednjom vrijednosti protoka. Kod drugog i trećeg pristupa odstupanja su značajno smanjena. U drugom pristupu ANN pokazuje najmanja odstupanja, a u trećem SVM. Pažnju treba obratiti i na moguću pojavu negativnih vrijednosti protoka kod ANN-a i SVM-a, mada su kod najpreciznijeg modela (SVM, 3. pristup) zanemarive. Za buduće istraživanje predlaže se rješavanje ovih problema optimizacijom modelskih parametara.

Tablica 7. Statističke karakteristike rezultata modela

	Pristup 1					Pristup 2				Pristup 3			
Gradnja	Izmj.	ANN	SVM	NNM	AR1	Izmj.	ANN	SVM	NNM	Izmj.	ANN	SVM	NNM
Srednja vrijednost	12,66	12,61	11,56	12,11	13,89	12,56	12,57	12,00	12,56	12,74	12,79	12,37	12,74
Minimum	0,56	2,20	1,86	2,20	-0,15	0,56	-0,14	-0,94	0,56	0,56	0,23	0,79	0,56
Maksimum	51,02	23,38	26,20	29,53	51,53	55,94	51,27	51,12	55,94	55,94	51,81	47,65	55,94
Koef. asimetrije	1,24	-0,18	-0,07	0,32	1,20	1,44	1,18	1,21	1,44	1,44	1,29	1,32	1,44
Koef. spljoštenosti	1,62	-1,36	-1,10	-0,67	1,25	2,60	1,90	2,56	2,60	2,48	2,17	2,30	2,48
Kalibracija													
Srednja vrijednost	13,87	13,33	12,35	12,77	14,99	14,52	14,21	13,47	13,15	11,31	10,91	11,33	11,39
Minimum	1,84	0,69	1,42	2,65	1,83	2,73	-0,41	0,35	2,02	1,59	-2,61	-0,04	2,38
Maksimum	55,94	23,42	24,29	27,78	54,16	49,84	39,92	36,72	31,06	38,81	34,61	35,86	29,89
Koef. asimetrije	1,52	-0,39	-0,19	0,08	1,30	1,04	0,47	0,51	0,21	1,16	0,54	0,65	0,34
Koef. spljoštenosti	2,58	-1,01	-0,75	-0,48	1,43	0,55	-0,60	-0,24	-0,62	0,81	-0,47	-0,33	-0,58
Verifikacija													
Srednja vrijednost	10,30	11,66	10,72	11,64	13,37	9,36	9,98	9,76	10,21	9,35	9,17	9,80	10,33
Minimum	2,22	2,01	2,00	2,63	3,67	2,22	-1,03	0,44	2,19	1,59	-2,61	-0,04	2,52
Maksimum	37,14	21,26	20,64	27,78	27,46	27,63	30,53	23,23	25,55	31,20	23,25	26,24	21,45
Koef. asimetrije	1,38	-0,22	-0,21	0,24	0,26	1,03	0,61	0,31	0,44	1,20	0,34	0,64	0,18
Koef. spljoštenosti	2,00	-1,46	-1,33	-0,66	-0,93	0,28	-0,51	-0,93	-0,50	0,90	-0,86	-0,38	-0,97
Verifikacija izvan duljine niza													
Srednja vrijednost	10,70	12,17	11,24	12,17	12,74	10,99	12,51	12,06	11,97	10,94	11,67	11,78	11,43
Minimum	1,43	2,51	2,36	2,25	1,65	1,43	-1,52	-1,22	1,90	1,43	-1,18	0,09	2,25
Maksimum	46,65	23,61	24,00	32,31	33,21	46,65	42,31	39,38	31,45	46,65	33,98	39,72	30,84
Koef. asimetrije	1,51	-0,08	0,00	0,24	0,52	1,44	0,67	0,73	0,46	1,51	0,60	0,88	0,47
Koef. spljoštenosti	2,24	-1,39	-1,15	-0,48	-0,22	1,84	-0,26	0,09	-0,43	2,07	-0,47	0,24	-0,43

Na slici 11. prikazan je dijagram rasipanja modeliranih vrijednosti u odnosu na izmjerene vrijednosti protoka za treći pristup. Dijagram obuhvaća vrijednosti iz svih dijelova podataka, zasebno za modele ANN, SVM i NNM. Najveći rasap primjećuje se na modelu NNM, osim na dijelu za gradnju modela, čije se vrijednosti u potpunosti preklapaju s pravcem koji ima idealne vrijednosti. S povećanjem vrijednosti izmjerenog protoka, NNM sve značajnije podcjenjuje vrijednosti. Manji rasap primjećuje se kod modela SVM i ANN, premda oni također imaju tendenciju podcjenjivanja viših vrijednosti protoka. Stoga se za buduća istraživanja preporučuje izračun intervala pouzdanosti modela te integracija tih vrijednosti u rezultate modela, primjerice primjenom kvantil-kvantil regresije, npr. [34].



Slika 11. Dijagram rasipanja modeliranih vrijednosti, treći pristup

5. Zaključak

U radu je analizirana mogućnost predviđanja srednjeg mjesečnog protoka radi dugoročnog predviđanja i planiranja u rješavanju problema vezanih uz dostupnost vode. Primijenjena su tri različita pristupa u kojima su uspoređene tri metode SL-a te stohastička metoda u prvom pristupu. U prvom pristupu SL je bio u stanju opisati prirodu protoka globalno, ali sa značajnim odstupanjima u opisu ekstremnih vrijednosti. AR je u stanju dohvatiti varijabilnost protoka, pri čemu se mora paziti na kvantificiranje varijabilnosti protoka. Kod SL-a treba primjenjivati složenije modele ili dodatno poraditi na odabiru ulaznih podataka. U drugom i trećem pristupu SL je bolje povezao uzročnost i posljedičnost ulaznih podataka i predviđene varijable. Primjena oborine i temperature za predviđanje protoka omogućuje uporabu projekcija iz klimatskih modela, što kod prvog pristupa nije moguće. U pravilu vrijedi da je s većom količinom podataka za gradnju modela SL-a veća i preciznost. No, duljina povijesnog niza ne podrazumijeva nužno i kvalitetno ili nekvalitetno izgrađen model. Međutim, preciznost SVM-a s koeficijentom determinacije 0,7-0,8 za korištenu duljinu niza 20-40 godina je zadovoljavajuća dok za 10 godina (koeficijent determinacije 0,67) preciznost nije znatno slabija. Za daljnje istraživanje preporučuje se dodatno razraditi metodologiju odabira ulaznih varijabli tako da se što učinkovitije koriste raspoloživi podaci.

LITERATURA

[1] Parry, M.: Food and Energy Security: Exploring The Challenges of Attaining Secure and Sustainable Supplies of Food and Energy, FOOD AND ENERGY SECURITY, (2012) 1, pp. 1-2

[2] UN (United Nations): World Population Prospects, 2015 Revision Population Database, 2015

[3] Marton, D., Menšik, P.P., Stary, M.: Using Predictive Model for Strategic Control of Multi-reservoir System Storage Capacity, 13th Computer Control for Water Industry Conference, PROCEDIA ENGINEERING, (2015) 112, pp. 994-1002

[4] IPCC (Intergovernmental Panel on Climate Change): Climate Change 2014, Synthesis Report, Geneva, Švicarska, 2015,

[5] Simonović, S.P.P.: Floods in a Changing Climate, International Hydrology Series, UNESCO, Cambridge, SAD, 2012.

[6] Rubinić, J., Margeta, J.: Dimenzioniranje akumulacija primjenom generiranih protoka, GRAĐEVINAR, 53 (2001) 1, pp. 17-23

[7] Haykin, S.: Neural Networks and Learning Machines, 3, izdanje, Upper Saddle River, New Jersey, SAD, 2013,

[8] Govindaraju, R.S., Rao, A.R.: Artificial Neural Networks in Hydrology, Springer Science & Business Media, 2000.

[9] Abrahart, R., Kneale, P.P.E., See, L.M.: Neural Networks for Hydrological Modelling, CRC Press, 2004

[10] Cigizoglu, H.K.: Generalized regression neural network in monthly flow forecasting, CIV. ENG. ENVIRON, SYST, 22 (2005) 2, pp. 71-84

[11] Nillson, .P.P., Uvo, C.B., Berndtsson, R.: Monthly Runoff Simulation: Comparing and combining conceptual and neural network models, J HYDROL, 321 (2006), pp. 344-363

[12] Wu, C.L., Chau, K.W., Li, Y.S.; River stage prediction based on a distributed support vector regression, J HYDROL, 358 (2008), pp. 96-111

[13] Guo, J., Zhou, J., Qin, H., Zou, Q., Li, Q.: Monthly streamflow forecasting based on improved support vector machine model,EXPERT SYS APPL, 38 (2011), pp.13073-13081

[14] Akiner, M.E., Akkoyunlu, A.: Modeling and forecasting river flow rate from the Melen Watershed, Turkey, J HYDROL, 456-457 (2012), pp. 121-129

[15] Farajzadeh, J., Fard, A.F., Lotfi, S.: Modeling of monthly rainfall and runoff of Urmia lake basin using "feed-forward neural network" and "time series analysis" model, WATER RESOURCES AND INDUSTRY, 7-8 (2014), pp. 38-48

[16] Terzi, O,: A genetic programming approach to river flow modeling, J INTELL FUZZY SYST, 27 (2014), pp. 2211-2219

[17] Matić, P.P.: Kratkoročno predviđanje hidrološkog dotoka pomoću umjetne neuronske mreže, Fakultet elektrotehnike, strojarstva i brodogradnje, Sveučilište u Splitu, Hrvatska, 2014

[18] Loucks, D.P.P., Van Beek, E.: Water Resources Systems Planning and Management, UNESCO, Paris, Francuska, 2005.

[19] Karamouz, M., Szidarovszky, F., Zahraie, B.: Water Resources Systems Analysis with emphasis on Conflict Resolution, Lewis Publishers, Boca Raton, SAD, 2003.

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.P., Weiss, R., Duborg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: Scikit-learn: Machine Learning in Python, J MACH LEARN RES, 12 (2011), pp. 2825-2830

[21] Russel, R., Norvig, P.P.: Artificial Intelligence: A Modern Approach, 3rd Edition, Prentice Hall, SAD, 2010.

[22] Mitchell, T.M.: Machine learning, McGraw Hill Inc,, New York, SAD, 1997.

[23] Kingma, D.P.P., Ba, J.L.: Adam: A Method For Stochastic Optimization, 3rd International Conference for Learning Representations, San Diego, 2015.

[24] Ng, A,: Lectures: Machine Learning, Stanford University, https://www,coursera,org/learn/machine-learning

[25] MacKay, D.J.C.: Information Theory, Inference and Learning Algorithms, Cambridge University Press, Cambridge.

[26] Raghavendra, N.S., Deka, P.P.C.: Support vector machine application in the field of hydrology: A review, APPL SOFT COMPUT, 19 (2014), pp. 372-386.

[27] Marsland, S.: Machine Learning, An Algorithmic Perspective, Chapman & Hall, Boca Raton, SAD, 2015,

[28] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression, STAT COMPUT, 14 (2004), pp. 199-222, https://alex,smola,org/papers/2004/SmoSch04,pdf [dostupno 07.04.2017,]

[29] Python: Scikit-learn user guide, release 0,17, 2015.

[30] Državni hidrometeorološki zavod: Hidrološka baza podataka, HIS 2000, 2017.

[31] Državni hidrometeorološki zavod: Relacijska meteorološka baza podataka, 2017.

[32] Šošić, I., Serdar, V.: Uvod u statistiku, Školska knjiga, Zagreb, 1992,

[33] Latifoğlu, L., Kişi, Ö., Latifoğlu, F.: Importance of hybrid models for forecasting of hydrological variable, NEURAL COMPUT APPLIC, 16 (2015), pp. 1669-1680.

[34] Dogulu, N., López López, P., Solomatine, D.P., Weerts, A.H., Shrestha, D.L.: Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments, HYDROL. EARTH SYST. SCI, 19 (2015), pp. 3181-3201.