

Razvoj metodologije za izdelavo modelov 7-dnevne napovedi nizkih pretokov

Luka Štravs¹, Mitja Brilly¹

Povzetek

Na podlagi uporabe tako teoretično razvitega znanja o dogajanju v obdobju nizkih pretokov kot tudi metode strojnega učenja M5 za generiranje regresijskih dreves smo razvili metodologijo izdelave modelov napovedi nizkih pretokov, ki so zelo primerni za operativno rabo. Prikazana in na dveh primerih slovenskih rek je uspešno implementirana metodologija razvoja empiričnih modelov nizkih pretokov za 7 dni vnaprej. Variabilnosti v obnašanju recesijskih delov hidrogramov kot posledici raznolikih faktorjev smo se izognili tako, da smo namesto konstantne vrednosti upoštevali variabilnost recesijskega koeficienta v različnih hidroloških razmerah, ki smo jih opisali s spremembo vrednosti nizkega pretoka od včeraj na danes (dQ) in vrednostjo srednjega dnevnega pretoka danes oziroma v dnevu izdelave napovedi (Q_t).

Uvod

Določanje nizkih pretokov je zelo pomembno za različne dejavnosti v okviru vodarstva. Suša s pomanjkanjem vode je namreč največji izziv za upravljanje z vodami zaradi pomanjkanja vira, od katerega so odvisni življenje in mnoge gospodarske dejavnosti. Varnost uporabnikov vode je tesno povezana z razpoložljivimi vodnimi viri, ki jih v suši praviloma primanjkuje. Opredelitev količine razpoložljivih virov vode v sušnih obdobjih je osnovno izhodišče za načrtovanje in izkoriščanje vodnih virov.

V naši raziskavi smo na podlagi uporabe hidrološkega znanja o dinamiki pretočnih količin v sušnih oziroma recesijskih obdobjih in uporabe metode strojnega učenja M5 razvili metodologijo razvoja enostavnih modelov napovedi nizkih pretokov. Metoda je bila implementirana na podlagi podatkov dveh vodomernih postaj na pritokih reke Save, in sicer na podatkih vodomerne postaje Podhom na Radovni in podatkih vodomerne postaje Bodešče na Savi Bohinjki.

Opis porečij

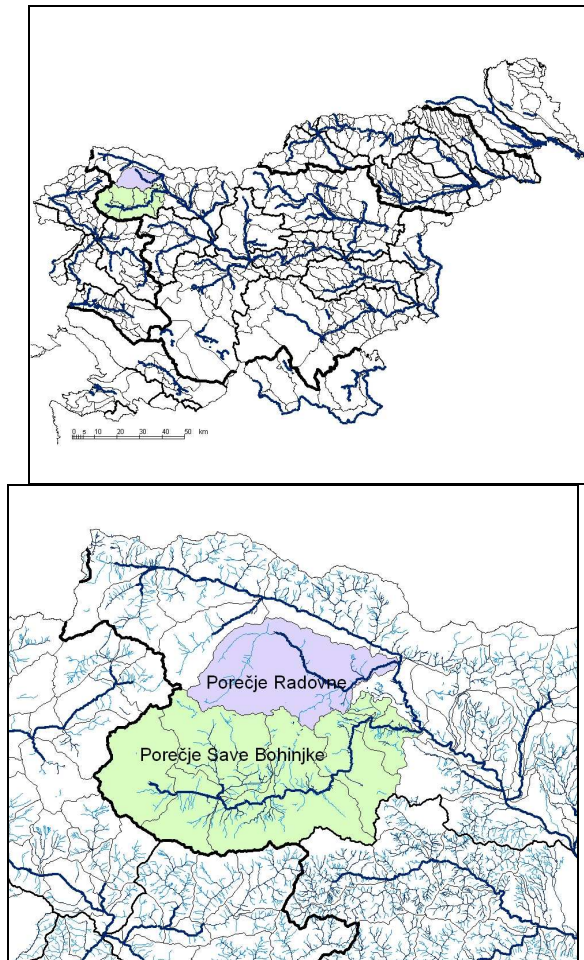
Za gradnjo modelov in za preveritev metodologije smo uporabili podatke o srednjih dnevni pretokih na Savinih pritokih Radovni in Savi Bohinjki. Porečje Save v Sloveniji meri 10746 km², kar znaša več kot 50 % ozemlja Slovenije; dolžina vodotoka Save na ozemlju Slovenije je 220.7 km.

Vodotok	Vodomerna postaja	Prispevno območje [km ²]	Obdobje	Q ₉₅	Q ₈₀	Q ₅₀	Q ₃₀
Radovna	Podhom	165,6	1991-2003	1,9	2,6	4,8	7,4
Sava Bohinjka	Bodešče	354,5	1991-2003	4,1	6,0	11,4	19,3

¹ Luka Štravs, univ.dipl.inž.gradb., prof.dr.Mitja Brilly, Fakulteta za gradbeništvo in geodezijo, Jamova 2, 1000 Ljubljana.

Preglednica 1 - Podatki o prispevnih območjih vodomernih postaj in značilne nizkopretočne karakteristike

Radovna in Sava Bohinjka skupaj s Savo Dolinko, Tržiško Bistrico, Kokro, Soro, Ljubljano, Kamniško Bistrico, Savinjo in Krko so glavni pritoki osrednje slovenske reke Save. Porečji Radovne in Save Bohinjke sta locirani na severozahodnem delu Slovenije (Slika 1), značilne nizkopretočne karakteristike in velikost prispevnih površin za vodomerni postaji Podhom in Bodešče za obdobje 1991-2003 pa so podane v Preglednici 1.



Slika 1 - Prikaz porečij Radovne in Save Bohinjke

Metode **Podatki**

Za določitev krivulj trajanja pretokov in karakterističnih statistik srednjih dnevnih pretokov Q_{95} (vrednost srednjega dnevnega pretoka, ki je presežen v 95 % časa), Q_{80} , Q_{50} in Q_{30} (Tabela 1) na posameznih vodomernih postajah smo uporabili podatke o srednjih dnevnih pretokih v letih od 1991 do 2003 na vodomernih postajah Podhom na Radovni in Bodešče na Savi Bohinjki; podatke je posredovala Agencija Republike Slovenije za okolje.

Za gradnjo oziroma razvoj modelov napovedi nizkih pretokov smo uporabili podatke iz obdobja od vključno leta 1991 do vključno leta 2002, za verifikacijo modelov napovedi pa

smo uporabili podatke o srednjih dnevni pretokih za leto 2003, katerega lahko uvrstimo med bolj sušna obdobja v zadnjem obdobju. Za verifikacijo modelov napovedi nizkih pretokov smo izbrali 5 sušnih obdobj v letu 2003. Modeli napovedi so bili verificirani samo na tistih podatkih, ko je srednji dnevni pretok v času napovedi (Q_t) znašal manj kot 150 % vrednosti Q_{30} , torej pretoka, ki je bil presežen vsaj 30 % časa v obdobju od 1991 do 2003. Tako smo modele preverili tudi na podatkih o srednjem dnevnem pretoku, ki niso bili vključeni v razvoj modelov.

Metoda strojnega učenja M5 za razvoj regresijskih dreves

Za razvoj modelov v obliki regresijskih dreves, ki predstavljajo modele variabilnega recesijskega koeficienta k , smo v naši raziskavi uporabili metodo strojnega učenja M5 iz programskega orodja WEKA. To orodje je bilo razvito na Univerzi Waikato na Novi Zelandiji (Witten & Frank, 2000). S tem, ko metodi strojnega učenja ponudimo dovolj vhodnih (atributov) in izhodnih podatkov, ki opisujejo modelirani proces, se le-ta iz posameznih primerov nauči vzorce obnašanja v okviru modeliranega procesa. Pri tem se celotni podatkovni prostor vhodnih spremenljivk deli na podprostore, katerih meje so v procesu razvoja modela optimizirane, znotraj posameznih podprostorov celotnega podatkovnega n -dimenzijskega prostora pa veljajo določeni vzorci, ki se jih lahko aproksimira z npr. nominalno ali konstantno numerično vrednostjo in linearno funkcijo. Odločitvena drevesa se generirajo na podlagi iterativnega deljenja podatkov v podatkovne podprostore celotnega podatkovnega prostora, pri čemer je cilj, da je razdalja med posameznimi podatkovnimi podprostori čim večja (Breiman et al., 1984; Quinlan 1986, 1992; Kompare, 1995; Mitchell, 1997; Witten & Frank, 2000; Solomatine & Dulal, 2003).

Glavne komponente odločitvenih dreves so odločitvena vozlišča, veje in listi. Odločitveni proces se začne pri glavnem oziroma zgornjem odločitvenem vozlišču, v katerem je specificiran razdelitveni kriterij. Glede na odgovor na ta razdelitveni kriterij se odločitveno drevo razdeli v veje. Vsaka veja lahko vodi do podrejenih vej ali pa zaključnih delov odločitvenih dreves, ki se imenujejo listi in predstavljajo vrednost, ki jo zavzame model. Rezultati modeliranja so torej modeli v obliki odločitvenih dreves, ki predstavljajo zaporedje pravil. Ta pripeljejo do vrednosti razreda, numerične vrednosti ali linearne funkcije uporabljenih atributnih spremenljivk. Glede na to lahko razdelimo vrste odločitvenih dreves v:

- klasifikacijska drevesa z nominalnimi vrednostmi razreda kot listi modela,
- regresijska drevesa s konstantnimi numeričnimi vrednostmi kot listi modela in
- modelna drevesa z linearnimi kombinacijami oziroma funkcijami uporabljenih atributov kot listi modela.

Razvoj modela napovedi s konstantno vrednostjo recesijskega koeficienta k

V prvi fazi smo želeli izdelati modele napovedi nizkih pretokov za n -dni vnaprej na podlagi najbolj poznane enačbe recesijskega dela hidrograma (Tallaksen, 1995):

$$Q_{t+n} = Q_t \cdot e^{-k \cdot n}$$

(1)

kjer predstavljajo Q_t in Q_{t+n} srednji dnevni pretok na posamezni vodomerni postaji v dnevih t in $t+n$, n število dni vnaprej, za katero se izdela napoved (npr. za 1 dan vnaprej $n = 1$), in k recesijski koeficient konstantne vrednosti.

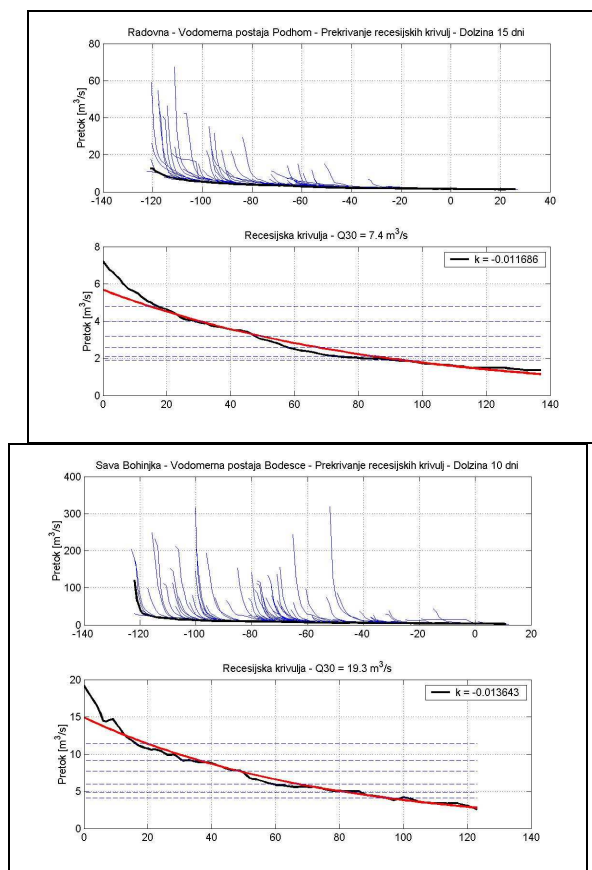
Enačba (1) se v literaturi (Tallaksen, 1995) pojavlja tudi v naslednjih podobnih oblikah:

$$Q_{t+n} = Q_t \cdot e^{-n/C} \quad (1a)$$

$$Q_{t+n} = Q_t \cdot e^{-a1 \cdot n} \quad (1b)$$

$$Q_{t+n} = Q_t \cdot k^n \quad (1c)$$

kjer predstavljajo C , $a1$ in k konstante in je n število dni vnaprej, za katero se izdela napoved.



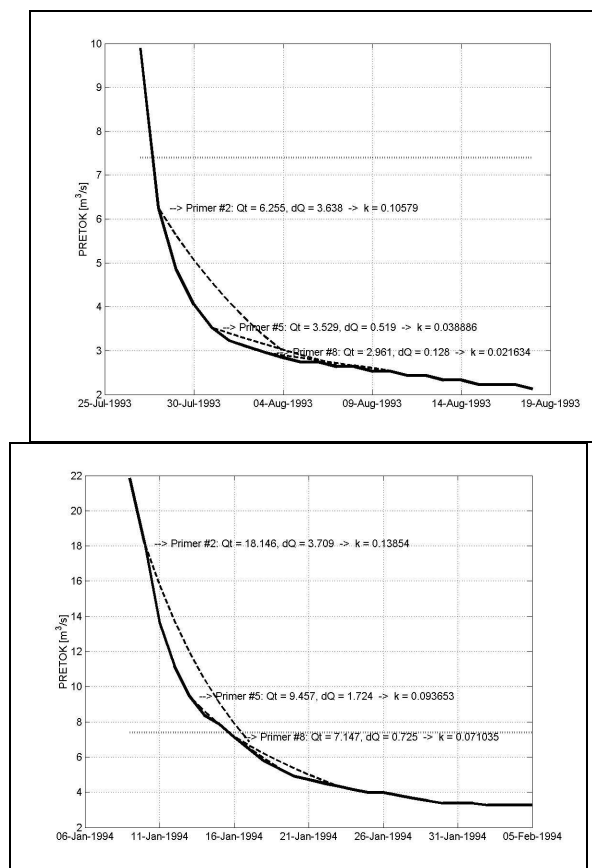
Slika 2 - Določitev konstantne vrednosti koeficienta k na podlagi določitve glavne recesijske krivulje

Za vsako izmed vodomernih postaj smo določili glavno recesijsko krivuljo (angl. master recession curve) in ji določili eksponentno krivuljo z eksponentnim koeficientom, ki se ji najbolj prilega po kriteriju najmanjšega odstopanja kvadratov (Slika 2). Tako smo določili

konstantno vrednost recesijskega koeficienta k , ki se ob uporabi v enačbi (1) uporabi za izdelavo 7-dnevne napovedi nizkih pretokov.

Razvoj modela napovedi s spremenljivo vrednostjo recesijskega koeficienta k

Za razliko od modela napovedi s konstantno vrednostjo recesijskega koeficienta smo na podlagi analize recesijskih obdobj v letih od 1991 do 2002 razvili modele variabilnega recesijskega koeficienta k , ki se ga kasneje uporabi v modelu napovedi nizkih pretokov za 7 dni vnaprej. Razvili smo torej regresijsko drevo variabilnega recesijskega koeficienta k iz enačbe (1), ki ni konstanten, temveč variira glede na vrednost pretoka v dnevu izdelave napovedi (Q_t) in glede na spremembo pretoka v zadnjih 24 urah (dQ).



Slika 3 - Zbiranje atributov (Q_t in dQ) in razredov (koeficient k) posameznih primerov iz podatkov o srednjih dnevni pretokih na Radovni

Osnovna oblika modela torej ostane enaka (enačba 1), le recesijski koeficient ni več konstanten v vseh hidroloških pogojih v trenutku izdelave napovedi, temveč je odvisen od spremembe pretoka v zadnjih 24 urah dQ in vrednosti pretoka v dnevu izdelave napovedi Q_t :

$$k = f(Q_t, dQ)$$

Modele oziroma regresijska drevesa recesijskega koeficienta k smo razvili tako, da smo za obdobje let od 1991 do 2002 identificirali vsa obdobja, ko je bil pretok v upadanju (Slika 3). Za vsako tako obdobje in vrednosti nižje od Q_{30} smo določili koeficient eksponencialne krivulje, ki se najbolje prilega vrednostim pretoka za 7 dni vnaprej. Za

vsak tako dobljeni k smo odčitali še vrednost pretoka v trenutku izdelave napovedi Q_t in vrednost spremembe pretoka glede na prejšnji dan dQ . Tako smo za vsako posamezno postajo pridobili večje število trojic (primerov), kjer sta dQ in Q_t atributa (oziroma vhodni spremenljivki), razred (oziroma modelirana spremenljivka) v obliki konstante vrednosti pa je recesijski koeficient k (Preglednica 2).

Primer #	Q_t [m ³ /s]	dQ [m ³ /s]	k
1	4,76	0,400	0,0617
2	4,19	0,570	0,0538
3	4,01	0,180	0,0518
4	3,66	0,350	0,0443
5	3,49	0,170	0,0419
6	3,33	0,160	0,0365
7	3,17	0,160	0,0291
8	3,02	0,150	0,0211
9	2,88	0,140	0,0178
10	2,74	0,140	0,0148
...
549	2,64	0,127	0,0301
550	2,52	0,122	0,0200

Preglednica 2 - Prikaz posameznih zbranih trojic (primerov) Q_t , dQ in k , na podlagi katerih metoda M5 sama zgradi regresijsko drevo recesijskega koeficienta k .

Tako smo za VP Podhom na Radovni zbrali 550 primerov (Tabela 2) in za VP Bodešče na Savi Bohinjki 340 primerov, iz katerih smo z uporabo metode strojnega učenja M5 razvili regresijski drevesi, ki predstavljata model recesijskega koeficienta k v funkciji dQ in Q_t .

Rezultati

Za vrednost konstantnega recesijskega koeficienta k za vodomerno postajo Podhom je bil določen $k = 0.01168$ in za vodomerno postajo Bodešče določen $k = 0.01364$ (Slika 2). Razvita modela napovedi nizkih pretokov s konstantnim k iz enačbe (1) za 7 dni vnaprej sta torej:

- za vodomerno postajo Podhom na Radovni:

$$Q_{t+n} = Q_t \cdot e^{-0.01168 \cdot n}$$

- in za vodomerno postajo Bodešče na Savi Bohinjki:

$$Q_{t+n} = Q_t \cdot e^{-0.01364 \cdot n}$$

Na podlagi uporabe metode strojnega učenja M5 smo za vsako vodomerno postajo razvili regresijsko drevo (Slika 4), ki predstavlja model variabilnega recesijskega koeficienta k . Ta se ob uporabi v enačbi (1) uporabi za izdelavo 7-dnevne napovedi nizkih pretokov.

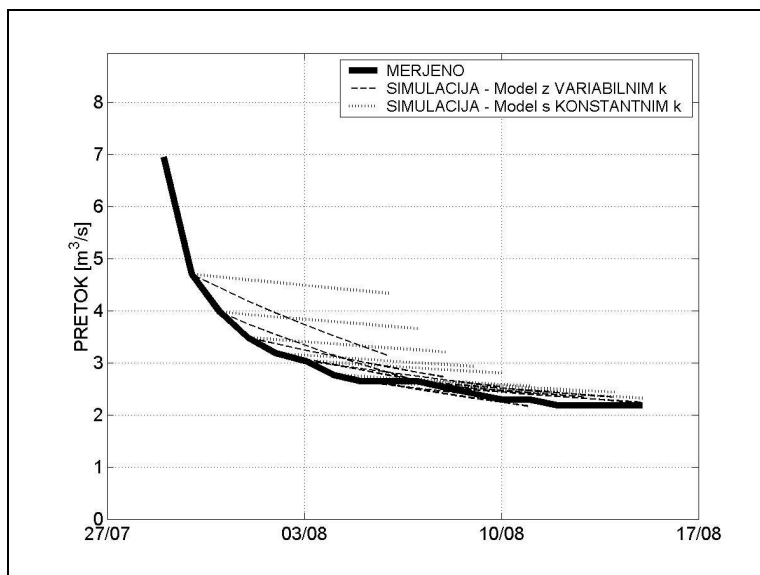
Radovna - VP Podhom	Sava Bohinjka - VP Bodešče
$dQ \leq 0,207$: $dQ \leq 0,123$: $Q_t \leq 2,2$: $k = 0,00813$ $Q_t > 2,2$: $k = 0,01730$ $dQ > 0,123$: $k = 0,02580$ $dQ > 0,207$: $dQ \leq 0,554$: $k = 0,03520$ $dQ > 0,554$: $k = 0,05770$	$dQ \leq 1,52$: $dQ \leq 0,727$: $Q_t \leq 6,52$: $k = 0,0269$ $Q_t > 6,52$: $k = 0,0418$ $dQ > 0,727$: $k = 0,0548$ $dQ > 1,52$: $dQ \leq 3,18$: $k = 0,0752$ $dQ > 3,18$: $k = 0,0990$

Slika 4 - Regresijski drevesi recesijskega koeficienta k za vodomerni postaji Podhom na Radovni (levo) in Bodešče na Savi Bohinjki (desno)

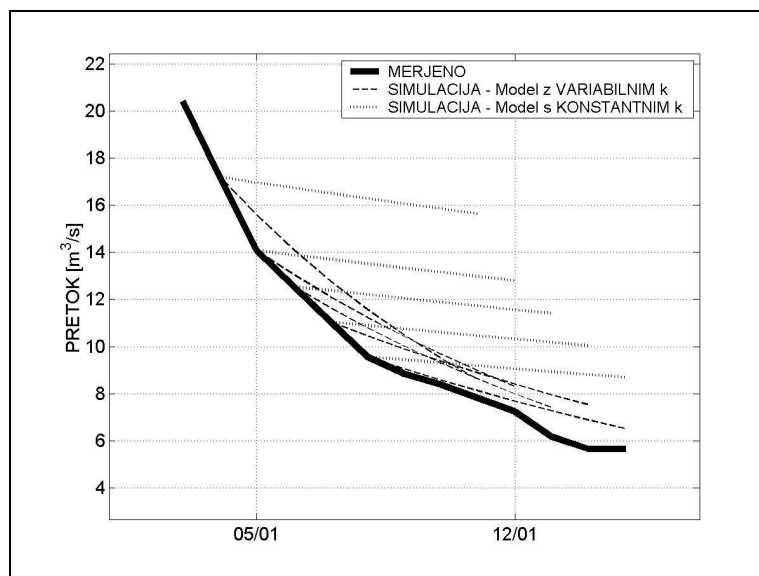
Model recesijskega koeficienta k za VP Bodešče se uporabi na naslednji način:

1. če je dQ manjši od ali enak $0,727 \text{ m}^3/\text{s}$ in Q_t manjši od ali enak $6,52 \text{ m}^3/\text{s}$, potem se v enačbi (1) za izdelavo napovedi uporabi koeficient $k = 0,0269$;
2. če je dQ manjši od ali enak $0,727 \text{ m}^3/\text{s}$ in Q_t večji od $6,52 \text{ m}^3/\text{s}$, potem se v enačbi (1) za izdelavo napovedi uporabi koeficient $k = 0,0418$;
3. če je dQ večji od $0,727 \text{ m}^3/\text{s}$ in manjši od ali enak $1,520 \text{ m}^3/\text{s}$, potem se v enačbi (1) za izdelavo napovedi uporabi koeficient $k = 0,0548$;
4. če je dQ večji od $1,520 \text{ m}^3/\text{s}$ in manjši od ali enak $3,180 \text{ m}^3/\text{s}$, potem se v enačbi (1) za izdelavo napovedi uporabi koeficient $k = 0,0752$;
5. če je dQ večji od $3,180$, potem se v enačbi (1) za izdelavo napovedi uporabi koeficient $k = 0,0990$;

Maksimalna vrednost pretoka, pri kateri se lahko izdela 7-dnevna napoved za vodomerno postajo Podhom je enaka $Q_{30} = 7,4 \text{ m}^3/\text{s}$ in je enaka $Q_{30} = 19,3 \text{ m}^3/\text{s}$ za vodomerno postajo Bodešče.



Slika 5 - Primerjava in preveritev modelov napovedi s konstantnim in modelov napovedi z variabilnim recesijskim koeficientom na podatkih o nizkih pretokih iz leta 2003 na VP Podhom na Radovni



Slika 6 - Primerjava in preveritev modelov napovedi s konstantnim in modelov napovedi z variabilnim recesijskim koeficientom na podatkih o nizkih pretokih iz leta 2003 na VP Bodešče na Savi Bohinjki

Modeli 7-dnevne napovedi srednjega dnevnega pretoka v sušnih obdobjih so bili testirani na podatkih o srednjih dnevni pretokih iz leta 2003. Tako je bil model napovedi nizkih pretokov na VP Podhom na Radovni (Tabela 3) testiran na sedmih sušnih obdobjih, model napovedi nizkih pretokov na VP Bodešče na Savi Bohinjki pa na petih daljših sušnih oziroma recesijskih obdobjih (Tabela 4) v letu 2003. Za vsako izbrano sušno obdobje smo za vsak dan kontinuirano simulirali 7-dnevno napoved (Sliki 5 in 6) in jo primerjali z dejansko dinamiko spremembe pretoka v naslednjih 7 dneh. Za vsako tako obdobje smo izračunali povprečno absolutno in povprečno relativno napako (Preglednici 3 in 4).

# obdobja	Model - spremenljivi k		Model - konstantni k	
	dQ _{AVG} [m ³ /s]	dQ _{AVG} [%]	dQ _{AVG} [m ³ /s]	dQ _{AVG} [%]
1	0.14	3.41	0.43	10.24
2	0.12	3.73	0.24	7.43
3	0.04	2.12	0.06	3.01
4	0.11	5.21	0.20	9.59
5	0.13	4.50	0.09	3.09
6	0.20	7.44	0.44	16.29
7	0.17	7.11	0.46	19.28

Preglednica 3 - Primerjava natančnosti modelov s spremenljivim in modelov s konstantnim recesijskim koeficientom za VP Podhom

# obdobja	Model - spremenljivi k		Model - konstantni k	
	dQ _{AVG} [m ³ /s]	dQ _{AVG} [%]	dQ _{AVG} [m ³ /s]	dQ _{AVG} [%]
1	1.10	13.58	3.63	45.20
2	0.44	10.62	0.69	16.54
3	0.53	9.45	0.66	11.77
4	1.23	12.27	2.51	25.84
5	0.42	8.13	0.97	18.67

Tabela 4 - Primerjava natančnosti modelov s spremenljivim in modelov s konstantnim recesijskim koeficientom za VP Bodešče

Primerjava natančnosti modelov (Tabeli 3 in 4) pokaže, da je model z variabilnim recesijskim koeficientom večinoma precej bolj natančen kot model s konstantnim recesijskim koeficientom. Medtem ko je relativna povprečna napaka pri napovedih z modelom s konstantnim recesijskim koeficientom znašala tudi do 45 % (primer za obdobje #1 na VP Bodešče), pa je bila maksimalna povprečna relativna napaka obeh modelov z variabilnim recesijskim koeficientom nekaj več kot 13.5 % (primer za obdobje #1 na VP Bodešče). Razen za primer obdobja #5 na VP Podhom je bila napaka modela z variabilnim recesijskim koeficientom k občutno manjša.

Iz strukture modelov recesijskega koeficienta k v obliki regresijskih dreves za VP Podhom na Radovni in VP Bodešče na Savi Bohinjki pa se vidi tudi naslednje:

- razvita modela sta povsem v skladu s pričakovanji, saj s povečevanjem razlike v vrednosti srednjega pretoka med včeraj in danes (dQ v modelu) raste tudi absolutna vrednost recesijskega koeficienta k ;
- prav tako se v strukturnih delih modelov, kjer se kot odločitveni kriterij pojavi Q_t , vidi, da je absolutna vrednost recesijskega koeficienta večja pri višjih vrednoti Q_t ;
- v modelu kot odločitveni kriterij prevladuje atribut dQ , kar pomeni, da variiranje recesijskega koeficienta ni v toliki meri odvisno od vrednosti srednjega dnevnega pretoka v dnevu izdelave napovedi (Q_t), temveč da je dQ boljši indikator dogajanja v naslednjih 7 dneh kot pa vrednost srednjega dnevnega pretoka Q_t .

Zaključki

Na podlagi uporabe tako teoretično razvitega znanja o dogajanju v obdobju nizkih pretokov kot tudi metode strojnega učenja M5 za generiranje regresijskih dreves smo razvili zelo uporabne modele napovedi nizkih pretokov, ki so zelo primerni za operativno rabo. Prikazana in na dveh primerih uspešno implementirana je bila metodologija razvoja empiričnih modelov nizkih pretokov. Variabilnosti v obnašanju recesijskih delov hidrogramov kot posledici raznolikih faktorjev (na primer vpliv sezonskosti ali predhodne vlažnosti) smo se izgonili tako, da smo namesto konstantne vrednosti upoštevali variabilnost recesijskega koeficienta v različnih hidroloških razmerah, ki smo jih opisali s spremembo vrednosti nizkega pretoka od včeraj na danes (dQ) in vrednostjo srednjega dnevnega pretoka danes oziroma v dnevu izdelave napovedi (Q_t).

Iz rezultatov v obliki modelov, ki so predstavljeni kot regresijska drevesa, se vidi, da nam o dinamiki vrednosti pretoka v naslednjih nekaj dneh več pove sprememba vrednosti pretoka (dQ) kot pa sama vrednost pretoka v dnevu, ko se izdelava napoved (Q_t). Model napovedi z variabilnim recesijskim koeficientom, ki ga izrazimo kot funkcijo vrednosti srednjega dnevnega pretoka (dQ) in spremembe le-tega v dnevu izdelave napovedi (Q_t), predstavlja z vidika obnašanja modelov napovedi nizkih pretokov izrazito izboljšavo.

Še enkrat je bila prikazana uspešnost in primernost uporabe modernih matematično-modelarskih orodij strojnega učenja v hidrologiji. Vseeno pa menimo, da je poznavanje področja dela, v okviru katerega se razvijajo modeli oziroma regresijska drevesa z metodami strojnega učenja, nepogrešljiva postavka, brez katere sta razvoj in še posebej interpretacija razvitih modelov skoraj nemogoča oziroma lahko interpretacija včasih pripelje do absurdnih situacij.

Literatura

- Breiman L, Friedman JH, Olshen RA, Stone CJ. (1984) Classification and regression trees. Wadsworth, Belmont.
- Kompare B. (1995) The use of artificial intelligence in ecological modelling. Ph. D. Thesis, Royal Danish School of Pharmacy, Copenhagen, Denmark.
- Mitchell T. (1997) Machine Learning. MIT Press and The McGraw-Hill Companies, Inc.
- Quinlan JR. (1986) Induction of Decision Trees. Machine Learning 1: 81-106.
- Quinlan JR. (1992) Learning with continuous classes. In: Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence, pp 343–348.
- Solomatine DP, Dulal KN. (2003) Model trees as an alternative to neural networks in rainfall-runoff modelling. Hydrological Sciences Journal 48: 399–411.
- Tallaksen, L.M. (1995) A review of baseflow recession analysis. Journal of Hydrology 165, 349-370.
- Witten I H, Frank E. (2000) Data mining: Practical machine learning tools and techniques with java implementations. Morgan Kaufmann Publishers, San Francisco, USA.