

STATISTIČKA ANALIZA VELIKIH VODA U PRISUSTVU IZUZETAKA

Jasna PLAVŠIĆ*, Dragutin PAVLOVIĆ
Univerzitet u Beogradu – Građevinski fakultet

Vladislava MIHAILOVIĆ
Univerzitet u Beogradu – Šumarski fakultet

Borislava BLAGOJEVIĆ
Univerzitet u Nišu – Građevinsko-arhitektonski fakultet

REZIME

Statistička analiza velikih voda se radi na nizovima hidroloških osmatranja, koja u rezultate ove analize unose neizvesnosti u pogledu ograničene dužine osmatranja, teškoća u merenju vrlo velikih protoka i njihove procene iz zone ekstrapolovanih krivih protoka, kao i zbog različitih mehanizama generisanja oticaja. Definisane probabilističkog modela velikih voda je otežano i zbog postojanja izuzetaka u nizovima, tj. značajno velikih ili malih vrednosti u odnosu na većinu. U radu se daje pregled postupaka i statističkih testova za otkrivanje izuzetaka od standardnih do najnovijih, kao i načini postupanja sa izuzetnim vrednostima. Najveća pažnja posvećena je otkrivanju i tretmanu donjih izuzetaka, koji u velikoj meri utiču na raspodelu a time i na usvojene računске velike vode. Poseban osvrt je dat na najnovije preporuke iz američkog Biltena 17C. Komentarišu se pretpostavke o raspodeli populacije, procedure testiranja, postupanje sa transformisanim nizovima i ukazuje na moguće poteškoće u primeni. Sama primena testova i procedura prikazana je u praktičnom delu, gde se statistička analiza velikih voda u prisustvu izuzetaka sprovodi za karakteristične primere sa hidroloških stanica u Srbiji na nizovima formiranim metodama godišnjih maksimuma i prekoračenja preko praga. Primeri obuhvataju i gornje i donje izuzetke. Komentarišu se razlike u rezultatima prilikom primene različitih procedura u prisustvu izuzetaka i predlaže niz iskustvenih mera kako bi ocena velikih voda bila što pouzdanija.

Ključne reči: velike vode, statistička analiza, maksimalni godišnji protoci, metod prekoračenja preko praga, izuzeci, statistički testovi, raspodele verovatnoće velikih voda

1. UVOD

Najveće neizvesnosti u vezi sa rezultatima statističke analize velikih voda potiču od ograničenih nizova hidroloških osmatranja, kao i od neizvesnosti u merenjima ekstremnih velikih voda i njihove procene iz ekstrapolovanih krivih protoka. U praktičnim primenama statističke analize velikih voda ove neizvesnosti dovode do poteškoća u identifikaciji raspodele verovatnoće velikih voda. Jedan od tipičnih problema koji se javlja je taj da su maksimalne godišnje velike vode u različitim godinama nastale različitim mehanizmima generisanja oticaja, npr. usled topljenja snega ili usled obilnih padavina. To je slučaj kada vrednosti u uzorku maksimalnih godišnjih protoka potiču iz mešavine raspodela, što zahteva razdvajanje uzorka na dve ili više grupa podataka prema mehanizmu generisanja velikih voda, a potom kombinovanje odvojeno ocenjenih raspodela u složenu raspodelu velikih voda. Međutim, broj poplava nastalih usled jednog od mehanizama oticaja je obično mali za pouzdanu ocenu raspodele verovatnoće.

Problem postojanja izuzetaka u nizovima velikih voda usko je povezan sa problemom nepoznate raspodele populacije velikih voda i neizvesnostima u merenju i oceni protoka velikih voda. Prisustvo izuzetaka u uzorku otežava formulisanje probabilističkog modela i prilagođavanje teorijskih raspodela osmatranjima. S obzirom da je pronalaženje korektnog modela koji će omogućiti ekstrapolaciju velikih voda izvan raspona osmotrenih vrednosti veoma važno za ocenu računskih velikih voda, uticaj gornjih izuzetaka na izbor teorijske raspodele je veliki. Takođe je poznato da donji izuzeci mogu značajno da utiču ne samo na izbor najbolje raspodele, već i

* Dobitnik nagrade iz fondacije "Dr Vujica Jevđević" za 2005. godinu.

na ocene parametara raspodela a time i na ocene računskih velikih voda.

Najpopularniji metod predložen u literaturi za otkrivanje izuzetaka i postupanje sa njima zasniva se na pristupu datom u poznatom Biltenu 17B tj. preporukama za proračun velikih voda u SAD (IACWD, 1982). Predlog novih preporuka za proračun velikih voda pod nazivom Bilten 17C objavljen je ove godine (England Jr. et al, 2016) i nalazi se u fazi stručne revizije. Ova verzija američkih preporuka se fokusira samo na donje izuzetke i predlaže novu metodologiju u kojoj se male vrednosti sa donjeg kraja raspodele odbacuju iz proračuna ukoliko značajno utiču na oblik raspodele. Ova metodologija se na neki način približila metodi prekoračenja iznad praga (tj. POT, prema *peaks over threshold*), čiji je takođe cilj da se raspodela velikih voda oceni na osnovu informacija o ekstremnim velikim vodama, a ne i na osnovu uobičajenih talasa velikih voda koje ne izazivaju nikakve probleme na slivu (Plavšić, 2005).

U okviru istraživanja sprovedenih u okviru izrade „Preporuka za ocenu velikih voda na profilima hidroloških stanica u Srbiji“ (Blagojević i sar., 2013, 2014a, 2014b) sprovedene su analize velikih voda sa ciljem da se utvrde razlike koje proizilaze iz različitih pristupa otkrivanja izuzetaka i postupanja sa njima kao i da se identifikuju problemi koji mogu nastati u primeni predložene postupaka. Preliminarna analiza raspoloživih podataka (Blagojević i sar., 2014a) je pokazala da glavni problemi u prepoznavanju izuzetaka proizilaze iz primene tradicionalnog Grubbs-Beck testa za detekciju izuzetaka, koji važi za normalno raspoređene podatke (ili podatke transformisane u normalno raspoređene), pri čemu se pretpostavlja da osmotrene velike vode prate druge tipove raspodela (npr. Gumbelova ili log-Pirson III tipa). Pored toga, nizovi godišnjih maksimuma protoka u kojima se javljaju izuzeci obično potiču iz mešavine raspodela, pa ostaje pitanje da li su ti izuzeci zaista izuzetne vrednosti male verovatnoće pojave u razmatranoj populaciji ili su samo „obične“ vrednosti iz populacije sa drugačijom raspodelom.

U ovom radu ispituju se postupci detekcije izuzetaka i njihovog tretmana u statističkoj analizi velikih voda. Analiziraju se rezultati različitih metoda za detekciju izuzetaka na raspoloživim nizovima maksimalnih godišnjih protoka na 70 hidroloških stanica u Srbiji. Rad nastavlja istraživanja započeta u radu Plavšić i sar. (2014). Razmatra se uticaj otkrivenih izuzetaka na ocene kvantila velikih voda u svetlu pretpostavke o raspodeli populacije velikih voda i daju se preporuke za statističku analizu velikih voda u praksi.

2. DETEKCIJA IZUZETAKA I POSTUPANJE SA NJIMA

2.1 Test Grubsa i Beka

Test Grubsa i Beka, koji je prvobitno predložen u radovima Grubbsa (1969) i Grubbsa i Becka (1972), koristi se za ispitivanje prisustva bilo gornjih bilo donjih izuzetaka u nizovima. Primenjuje se pod pretpostavkom da su podaci (u originalnoj formi ili nakon neke transformacije) normalno raspoređeni. Ako su x_i ($i = 1, 2, \dots, n$) podaci iz uzorka obima n uređeni u rastućem redosledu, test statistika Grubbs-Bek testa se definiše kao standardizovano odstupanje najmanje ili najveće vrednosti u uzorku x_1 tj. x_n :

$$K = \frac{\bar{x} - x_1}{s} \quad \text{ili} \quad K = \frac{x_n - \bar{x}}{s} \quad (1)$$

gde su \bar{x} i s uzoračka srednja vrednost i standardna devijacija, dok je K test-statistika koja zavisi od praga značajnosti α i obima uzorka n . Za prag značajnosti od 10% i dužinu niza n , za K važi sledeći približan izraz (Stedinger et al., 1993):

$$K_{10\%} = -0.9043 + 3.345\sqrt{\log n} - 0.4046 \log n \quad (2)$$

2.2 Postupanje sa izuzecima prema Biltenu 17B

Preovlađujući metod za detekciju izuzetaka i njihovo tretiranje u hidrološkim nizovima zasniva se na postupku datom u Biltenu 17B (IACWD, 1982). Ova publikacija daje smernice za proračun velikih voda u kojima se preporučuje primena jednostranog Grubbs-Bek testa za detekciju izuzetaka. U opisanom postupku logaritmovani protoci iz uzorka sa oba kraja raspodele se porede sa pragom koji odgovara kritičnoj vrednosti kontrolne statistike Grubbs-Bek testa za prag značajnosti od 10% (jednačina 2).

Iako je originalni Grubbs-Bek test namenjen testiranju jednog izuzetka (samo najmanje ili najveće vrednosti u nizu), Bilten 17B preporučuje da se donjim izuzecima smatraju svi logaritmovani protoci ispod donjeg praga definisanog kao:

$$y_d = \bar{y} - s_y \cdot K_{10\%} \quad (3)$$

gde su \bar{y} i s_y srednja vrednost i standardna devijacija logaritmovanog niza protoka. Analogno, gornjim izuzecima se smatraju sve vrednosti iznad gornje granice:

$$y_g = \bar{y} + s_y \cdot K_{10\%} \quad (4)$$

Prema Biltenu 17B, redosled testiranja donjih i gornjih izuzetaka zavisi od koeficijenta asimetrije logaritmovanog niza protoka c_{sy} . Za uzorke sa koeficijentom asimetrije $c_{sy} < -0.4$ prvo se testiraju donji izuzeci, dok se gornji izuzeci testiraju na nizu iz kojeg su uklonjeni detektovani donji izuzeci. Za uzorke sa koeficijentom asimetrije $c_{sy} > -0.4$ prvo se testiraju gornji izuzeci, a zatim donji. U ovom slučaju kada je fokus na gornjim izuzecima, u Biltenu 17B se pravi razlika između nizova velikih voda iz perioda sistematskog osmatranja i onih koji sadrže podatke izvan perioda sistematskog osmatranja (istorijske velike vode). Za gornje izuzetke koji se detektuju u nizovima sa istorijskim velikim vodama, Bilten 17B daje uputstva za korekciju statistika tih nizova. U takvim nizovima donji izuzeci se testiraju pre korekcije statistika ako je $-0.4 < c_{sy} < 0.4$, odnosno posle korekcije statistika ako je $c_{sy} > 0.4$. U nizovima bez istorijskih velikih voda nema korekcije statistika niti uklanjanja gornjih izuzetaka. Međutim, Bilten 17B preporučuje da se ocena velikih voda u slučaju postojanja velike neizvesnosti vezane za ocenu koeficijenta asimetrije nizova (kao što je slučaj kratkih nizova, ali i slučaj prisustva gornjih izuzetaka) bazira na regionalno korigovanoj vrednosti koeficijenta asimetrije. Time se zapravo „reguliše“ gornji kraj raspodele i ne dozvoljava se uticaj gornjeg izuzetka na raspodelu.

Ako se detektuju, donji izuzeci se uklanjaju iz niza i određuje se uslovna raspodela F_1 za podatke iznad praga y_d (za tzv. zasečen skup podataka). Raspodela velikih voda F tada se preračunava uzimajući u obzir uslovnu raspodelu F_1 i verovatnoću da se godišnji maksimum nađe ispod praga y_d :

$$F(x) = \frac{n_d}{n} - \frac{n - n_d}{n} F_1(x) \quad (5)$$

gde je n ukupan broj podataka, a n_d broj donjih izuzetaka, odnosno broj logaritmovanih protoka u nizu manjih od praga y_d definisanog jednačinom (3).

2.3 Bilten 17C

U Biltenu 17C (England Jr. i sar, 2016) opisuje se postupak sa tzv. potencijalno uticajnim malim vrednostima u nizovima velikih voda (engl. *potentially-influential low floods, PILF*). Ova procedura se zasniva na testu koji autori nazivaju višestrukim Grubs-Bek testom (engl. *multiple Grubbs-Beck test*) koji je prikazan u radu Cohna i sar. (2013). Test je sličan metodi Rosnera (1975), ali bez apriori određenog broja izuzetaka. Cilj

ove procedure je da se prepoznaju sve osmotrene vrednosti velikih voda koje ne predstavljaju nikakve ekstreme, a značajno utiču na oblik raspodele godišnjih maksimuma. Predloženi pristup predstavlja veliki korak napred za statističku analizu velikih voda i hidrološku statistiku uopšte.

Nulta hipoteza višestrukog Grubs-Bek testa je da svi osmotreni podaci u uzorku potiču iz populacije nezavisnih i normalno raspoređenih promenljivih. Test se sastoji od identifikacije broja k najmanjih vrednosti u nizu koje su neubičajeno male pod pretpostavkom važnija nulte hipoteze. Ako je $y_{(1:n)}, y_{(2:n)}, \dots, y_{(n:n)}$ niz logaritmovanih osmotrenih protoka uređen u rastući redosled, testira se da li je on konsistentan sa normalnom raspodelom kroz statistiku:

$$W_k = \frac{y_{(k:n)} - m_k}{s_k} \quad (6)$$

gde je $y_{(k:n)}$ k -to najmanje osmatranje u uzorku, dok su

$$m_k = \frac{1}{n-k} \sum_{i=k+1}^n y_{(i:n)} \quad (7)$$

$$s_k = \sqrt{\frac{1}{n-k-1} \sum_{i=k+1}^n (y_{(i:n)} - m_k)^2} \quad (8)$$

aritmetička sredina i standardna devijacija niza $y_{(k+1:n)}, y_{(k+2:n)}, \dots, y_{(n:n)}$, tj. svih vrednosti većih od $y_{(k:n)}$ za koje se smatra da nisu donji izuzeci, pa m_k i s_k odražavaju osobine populacije ekstremnih velikih voda. Testiranje se sprovodi u dva pravca: počevši od medijane niza ($k = n/2$) ka najmanjoj vrednosti ($k = 1$), i obrnuto. U oba prolaza traži se granična vrednost k^* pri kojoj je nulta hipoteza testa odbačena za svako $k < k^*$. Kao broj donjih izuzetaka onda se usvaja veće k^* iz dva prolaza.

Statistika W_k je manja što je k -to najmanje osmatranje udaljenije po veličini od ostalih podataka. Cohn i sar. (2013) su prikazali raspodelu statistike W_k u opštem obliku koji je prilično komplikovana, što nije pogodno za određivanje p -vrednosti ovog testa u praksi. Radna verzija Biltenu 17C ne nudi tablice sa kritičnim vrednostima statistike W_k , već je se predlaže primena softvera kao što su HEC-SSP (U.S. Army Corps of Engineers, 2016) ili PeakFQ (Flynn et al., 2006) u koje je ovaj test ugrađen.

U proceduri u Biltenu 17C identifikovani izuzeci se koriste kao donji prag ispod koga se vrednosti iz niza ne

uzimaju u proračun parametara log-Pirson III raspodele. Zbog gornjih izuzetaka i istorijskih velikih voda, može se zadati i gornji prag da bi se i te vrednosti ili isključile iz ocene parametara ili uključile na validan način kroz korekciju asimetrije iz regionalnih zavisnosti i promenom vrednosti empirijske raspodele za istorijske vrednosti (slično kao u Biltenu 17B). Ovaj postupak u celini je nazvan algoritam očekivanih momenata (engl. *expected moments algorithm, EMA*).

2.3 Drugi testovi za izuzetke

Na primenu Grubs-Bek testa kod velikih voda prema postupku u Biltenu 17B mogu se uputiti dve kritike. Prva je ta da se ovim testom može testirati samo jedan izuzetak a ne više njih odjednom, dok se druga kritika odnosi na kršenje pretpostavke o normalnosti podataka (Spencer and McCuen, 1996).

Ako se Grubs-Bek test primeni više puta uzastopno tj. uklanjanjem npr. donjeg izuzetka pa ponovnim testiranjem, moć testa se smanjuje u ponovljenim primenama (Tietjen i Moore, 1972). Uzastopna primena testa takođe može dovesti do efekta maskiranja izuzetka ako postoji grupa podataka odvojena od ostalih podataka; drugim rečima, uzastopnim testiranjem se može desiti da se jedan izuzetak ne detektuje u grupi od više potencijalnih izuzetaka. S druge strane, ukoliko se primene testovi kojima se testira više izuzetaka odjednom, može doći do efekta utapanja (engl. *swamping effect*) tj. do proglašavanja svih podataka u izdvojenoj grupi izuzecima, dok je zapravo samo jedan podatak izuzetak.

Grubbs (1950) je razvio i test za testiranje više izuzetaka odjednom kao grupu od k vrednosti na jednom kraju raspodele, sa nulom hipotezom da svi podaci u uzorku potiču iz iste normalne raspodele. Ovaj test su kasnije takođe preporučili (Tietjen i Moore, 1972). Kontrolna statistika ovog testa za testiranje k najvećih vrednosti u uzorku glasi:

$$L_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (9)$$

gde je

$$\bar{x}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} x_i \quad (10)$$

Kada se ispituje k najmanjih vrednosti, brojač sume u brojiocu u jednačini (9) i u jednačini (10) ide od $i = k + 1$ do $i = n$. Tablice sa kritičnim vrednostima za L_k za dužine uzorka do $n = 50$ i sa grupama do $k = 10$ izuzetaka dali su Tietjen i Moore (1972), a za dužine uzorka do $n = 100$ i grupama do $k = 4$ dali su Verma i Quiroz-Ruiz (2006).

Za testiranje više izuzetaka, Tietjen i Moore (1972) preporučuju postupak unazad da bi se izbegao efekat maskiranja. Na početku testa se pretpostavi velika vrednost k ; ako nulta hipoteza nije odbačena, k se smanji za jedan i testiranje ponovi. Ovakav postupak ipak ne sprečava efekat utapanja. Tietjen i Moore (1972) su razmatrali nekoliko pristupa da se pretpostavi broj izuzetaka k ako se njihov broj ne može naslutiti na osnovu vizuelne inspekcije, pa su preporučili da se za k uzme broj podataka iznad ili ispod najveće razlike između susednih vrednosti u uređenom nizu.

McCuen (2002) je predložio test Rosnera (1975) kojim može da se identifikuje više od jednog izuzetka, ali se broj k potencijalnih izuzetaka mora zadati unapred. Test se primenjuje na donje i gornje izuzetke istovremeno tako što se posmatra k vrednosti najudaljenijih od srednje vrednosti. Svih k podataka smatraju se izuzecima ako se odbaci nulta hipoteza da nema izuzetaka, ali nijedan od pojedinačnih podataka se ne može smatrati izuzetkom ako se nulta hipoteza ne odbaci. Međutim, ovaj test sa istovremenim ispitivanjem oba kraja raspodele se ne može smatrati pogodnim za ispitivanje izuzetaka u nizovima velikih voda zato što je uticaj donjih i gornjih izuzetaka na raspodele velikih voda potpuno drugačiji i oni ne mogu biti tretirani na isti način. Ako se i gornji i donji izuzeci uklone iz uzorka, on će se svakako približiti normalnoj raspodeli, ali to nije cilj u statističkoj analizi velikih voda, gde se ne želi izostavljanje podataka o ekstremno velikim protocima i time gubitak vrednih informacija o gornjem kraju raspodele.

Svi pomenuti testovi i kritične vrednosti test-statistika se zasnivaju na pretpostavci o normalnosti podataka. To znači da odbacivanje nulte hipoteze za neki niz može biti posledica odstupanja od normalne raspodele, a ne posledica prisustva izuzetaka. Uobičajena preporuka jeste da se podaci transformišu kako bi se približili normalnoj raspodeli. Uobičajene transformacije su logaritamska, Vilson-Hilferti, Box-Cox i druge. Logaritamska transformacija je najčešća u hidrološkim praksi. Iako je opšte mišljenje da se logaritmovanjem niz dovodi do dovoljnog stepena slaganja sa normalnom raspodelom, ova transformacija istovremeno može izazvati pojavu

donjih izuzetaka, jer male vrednosti u nizu dobijaju veću težinu u logaritmovanom uzorku.

2.4 Postupanje se izuzecima

U analizi velikih voda, kada se ona sprovodi na nizovima formiranim iz godišnjih maksimuma, gornji i donji izuzeci imaju različit uticaj na raspodelu verovatnoće, pa i njihov tretman treba da bude drugačiji. Gornji izuzeci zaista mogu da budu retki i izuzetni događaji, ali mogu biti i rezultat greške u merenjima, neizvesnosti pri ekstrapolaciji krivih protoka i slično. Izuzetne vrednosti takođe mogu biti rezultat nekog antropogenog uticaja a ne prirodnih procesa. To govori da se svaka vrednost u nizu mora pažljivo proveriti pre nego što se pristupi statističkoj analizi. S druge strane, ne očekuje se da su donji izuzeci rezultat mernih grešaka niti neizvesnosti u krivama protoka, ali oni značajno utiču na oblik raspodele verovatnoće.

Postupak uklanjanja donjih ili gornjih izuzetaka iz uzorka naziva se zasecanje, pa se dobija zasečen uzorak. Ovo je postupak koji se u analizi velikih voda generalno sprovodi na donjim izuzecima. Drugi pristup je korekcija izuzetaka tako da se njihov uticaj na uzorak smanji (u statističkoj literaturi ova korekcija se naziva *winsorising*). Vukmirović i Pavlović (2000) su dali tri moguće korekcije za gornje izuzetke: 1) zamena drugom najvećom vrednošću iz iste godine, 2) zamena drugom najvećom vrednošću u uzorku, ili 3) zamena ocenom 50-godišnje ili 100-godišnje velike vode sa susjednih stanica. Iako autori to nisu naglasili, u trećoj opciji se misli na regionalnu ocenu velikih voda za razmatrani sliv u odnosu na susjedne slivove.

Korekcija gornjeg izuzetka se generalno smatra boljim postupkom od njegovog izbacivanja iz niza osim u slučaju osnovane sumnje da se radi o mernoj greški (BLFUW, 2011). Zadržavanje gornjih izuzetaka u nizu bez korigovanja izgleda kao opšte prihvaćen pristup jer se smatra da oni nose značajne informacije o ekstremnim velikim vodama.

U hidrološkoj statističkoj literaturi (npr. Kottegoda and Rosso, 2008) takođe se preporučuje da se parametri raspodela u prisustvu izuzetaka ocenjuju na osnovu robusnih statistika kao što su srednja vrednost zasečenog uzorka, međukvartilno rastojanje itd. umesto onih statistika koje su osjetljive na prisustvo izuzetaka. Međutim, za primenu takvih statistika uz parametarske raspodele verovatnoće potrebno je još istraživanja u hidrološkoj statistici.



Slika 1. Hidrološke stanice sa raspoloživim nizovima maksimalnih godišnjih protoka.

3. METODOLOGIJA I PODACI

U ovom radu se razmatra više pristupa za detekciju izuzetaka:

- 1) procedura iz Biltena 17B (oznaka 17B) za donje i gornje izuzetke,
- 2) originalni Grubs-Bek test za detekciju jednog izuzetka, primenjen sekvencijalno na uzorak uz uklanjanje detektovanih donjih izuzetaka (oznaka GB),
- 3) testiranje prisustva višestrukih izuzetaka testom Grubbsa (1950) prema postupku koji su prikazali Tietjen i Moor (1972) (oznaka TM),
- 4) testiranje prisustva višestrukih donjih izuzetaka generalizovanim Grubs-Bek testom prema Biltenu 17C (oznaka B17C).

Ovi pristupi su primenjeni na logaritmovane protoke, tako da je pretpostavljeno da podaci prate log-normalnu raspodelu.

Na raspolaganju su bili podaci o maksimalnim godišnjim protocima na 70 hidroloških stanica u Srbiji (slika 1). Površine slivova domicilnih reka do razmatranih sta-

nica kreću se od 84 km² do 37 200 km², dok su površine slivova Dunava, Tise i Save do stanica u Srbiji znatno veće (Sava do Sremske Mitrovice 88 000 km², Tisa do Sente 142 000 km² i Dunav do Bezdana 210 000 km² i do Smedereva 526 000 km²). Dužine nizova variraju od oko 40 do 90 godina. Manje od polovine nizova ima podatke za sve godine osmatranja, što izaziva zabrinutost u vezi sa pouzdanošću analiza velikih voda. Od ukupnog broja, za dalja razmatranja odbačeno je 19 stanica koje su prestale sa radom, imaju duže prekide ili su na lokacijama pod uticajem rada akumulacija.

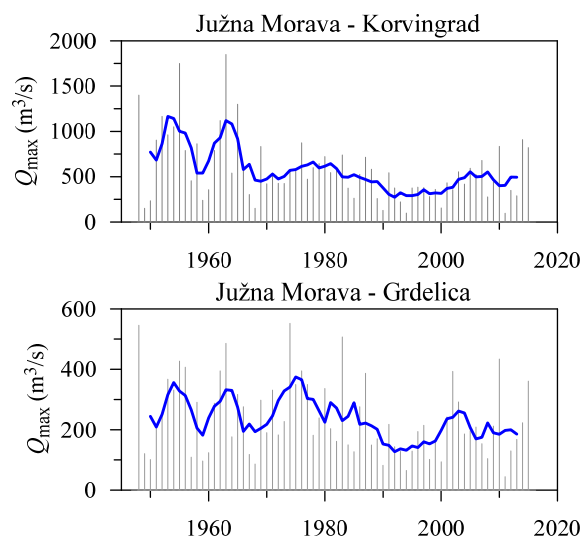
Preduslov za statističku analizu je nezavisnost i homogenost podataka u uzorku. Na svim nizovima sprovedeno je testiranje slučajnosti neparametarskim testom koraka, a nulta hipoteza o slučajnosti na pragu značajnosti od 5% je odbačena u slučaju 9 stanica. Ovakvi rezultati se odnose na stanice na kojima vremenska serija godišnjih maksimuma pokazuje persistenciju tj. autokorelaciju, nasuprot nizovima slučajnih vrednosti (na slici 2 dati su primeri niza sa persistencijom i slučajnog niza, Mojsinje i Grdelica na Južnoj Moravi). Testiranje homogenosti je sprovedeno neparametarskim Men-Vitni (Mann-Whitney) testom (Kottogoda and Rosso, 2008) na svim nizovima tako što su posmatrane dve polovine niza. Nulta hipoteza o homogenosti je odbačena na 28 stanica od 51. Nehomogenost u razmatranim nizovima se uglavnom uočava kroz grupisanje većih vrednosti u jednoj polovini niza i manjih vrednosti u drugoj polovini. U većem broju slučajeva ekstremnije velike vode su se događale u prvoj polovini niza, što možda ukazuje na preduzimanje mera na slivu zbog kojih su u kasnijem periodu izostale ekstremne vrednosti. Primer jednog takvog niza je stanica Korvingrad na Južnoj Moravi (gornji dijagram na slici 2). Ovakve promene u nizovima treba dodatno istražiti pre nego što se pristupi statističkoj analizi kako bi se ustanovili izvori neizvesnosti i eventualne nestacionarnosti.

Slaganje empirijske raspodele sa log-normalnom raspodelom testirano je za sve nizove maksimalnih godišnjih protoka testovima značajnosti koeficijenta korelacije dijagrama verovatnoće (engl. *probability plot correlation coefficient*, PPCC) i Anderson-Darlinga (AD) na pragu značajnosti od 5%. Ova dva testa su pokazala konsistentne rezultate na većini stanica. Prema PPCC testu, za 9 nizova od 51 se ne može reći da prate log-normalnu raspodelu, a prema AD testu za 8 nizova (rezultati dva testa nisu isti na ukupno 3 stanice, ali je u sva tri slučaja *p*-vrednost testa bliska usvojenoj kritičnoj vrednosti praga značajnosti od 5%). Nizovi logaritmovanih protoka koji ne ispunjavaju uslov normalnosti su uglavnom sa

stanica u čijim se nizovima javljaju izuzeci, ali to nije pravilo.

Za nizove maksimalnih godišnjih protoka sa detektovanim izuzecima sprovedena je statistička analiza u kojoj je primenjeno više teorijskih raspodela, među kojima su tri raspodele generalno pokazale najbolje slaganje sa empirijskim raspodelama: log-normalna, log-Pirson tipa III i opšta raspodela ekstremnih vrednosti (GEV).

Pored analize godišnjih maksimuma, za nekoliko stanica sprovedena je i analiza prekoračenja preko praga tj. POT metoda. U ovoj metodi, čije su osnove postavili Todorović i Zelenhasić (1970), formiraju se nizovi prekoračenja preko izabranog praga u koje može ući više maksimuma talasa iz jedne godine, dok iz neke druge godine bez značajnih velikih voda ne mora ući nijedna vrednost. Smatra se da ovakvi nizovi nose više informacija o velikim vodama i da daju pouzdanije ocene kvantila velikih voda (Plavšić, 2005, 2006). Postupak analize POT metodom detaljno su objasnili Vukmirović (1990), Vukmirović i Petrović (1997) i Plavšić (2005).



Slika 2. Primer niza koji ne zadovoljava test slučajnosti (gore) i primer niza za koji se prihvata hipoteza o slučajnosti (dole).

4. REZULTATI I DISKUSIJA

Rezultati testiranja prisustva izuzetaka u razmatranim nizovima je prikazano u tabelama 1, 2 i 3. Da bi rezultati bili konsistentni, hipoteze o prisustvu izuzetaka su razmatrane na pragu značajnosti od 10% u svim testovima.

4.1 Donji izuzeci

Donji izuzeci su detektovani u devet nizova (tabela 1). Samo u četiri od devet nizova donji izuzeci su detektovani svim metodama. U tri slučaja donji izuzeci su detektovani samo testom Tietjen-Moora (TM) i višestrukim GB testom u proceduri iz Biltena 17C, a samo TM testom u još dva slučaja. Detektuje se različit broj izuzetaka: testovi koji testiraju više izuzetaka odjednom kao rezultat imaju više značajnih izuzetaka, dok postupak prema Biltenu 17B i uzastopna primena GB testa detektuju samo po jedan izuzetak u ovim nizovima. Postupak iz Biltena 17C po pravilu daje najveći broj donjih izuzetaka, ali se tu zapravo ne radi o izuzecima već o k

najmanjih vrednosti u nizu koje značajno utiču na oblik raspodele.

U tabeli 1 nizovi sa donjim izuzecima su poređani po rastućem koeficijentu asimetrije logaritmovanog niza c_{sy} da bi se ukazalo na redosled testiranja prema postupku iz Biltena 17B. Prvih pet nizova u ovoj tabeli ima $c_{sy} < -0.4$, a među njima dva niza imaju po jedan donji izuzetak prema ovoj metodi. Posle uklanjanja donjeg izuzetaka, u ovim nizovima nije bilo gornjih izuzetaka. Pored ovih nizova, prema metodi iz Biltena 17B donji izuzeci su detektovani i u nizovima na stanicama Beli Brod i Valjevo na Kolubari, koji su prvo testirani na gornje izuzetke, ali oni nisu detektovani.

Tabela 1. Donji izuzeci u nizovima godišnjih maksimuma protoka prema različitim metodama (n je obim uzorka).

Reka/Stаница	Površina sliva (km ²)	n	Osm. min. (m ³ /s)	Osm. maks. (m ³ /s)	Koef. asim. c_{sy}	Broj izuzetaka k			
						B17B	GB1	TM	B17C
V. Morava/Lj. Most	37320	68	353	2354	-0.67	1	1	3	8
V. Morava/Varvarin	31548	68	336	2550	-0.55	0	0	3	3
V. Morava/Bagrđan	33446	67	354	2930	-0.51	0	0	3	3
Bjelica/Guča	239	53	10.2	232	-0.51	0	0	2	0
J. Morava/Mojsinje	15390	65	115	1830	-0.48	1	1	2	2
Kolubara/Beli Brod	1896	57	27.3	954	-0.35	1	1	1	1
Nišava/Niš	3870	90	40.0	1200	-0.31	0	0	4	6
Lugomir/Majur	427	53	3.35	431	-0.27	0	0	4	0
Kolubara/Valjevo	340	60	7.2	396	-0.12	1	1	1	1

Tabela 2. Nizovi sa donjim (DO) i gornjim (GO) izuzetkom (n označava obim uzorka; GO* označava broj gornjih izuzetaka posle uklanjanja donjih izuzetaka).

Reka/Stаница	Površina sliva (km ²)	n	Osm. min. (m ³ /s)	Osm. maks. (m ³ /s)	Koef. asim. c_{sy}	Tip	Broj izuzetaka k			
							B17B	GB1	TM	B17C
Kolubara/Slovac	995	62	28.7	776	-0.41	DO	1	1	1	2
						GO	/	1	1	
						GO*	1	1	1	
Studenica/Ušće	540	62	9.84	276	0.24	DO	1	1	1	1
						GO	1	1	1	
						GO*	1	1	1	

Tabela 3. Gornji izuzeci u nizovima godišnjih maksimuma protoka prema različitim metodama (n je obim uzorka).

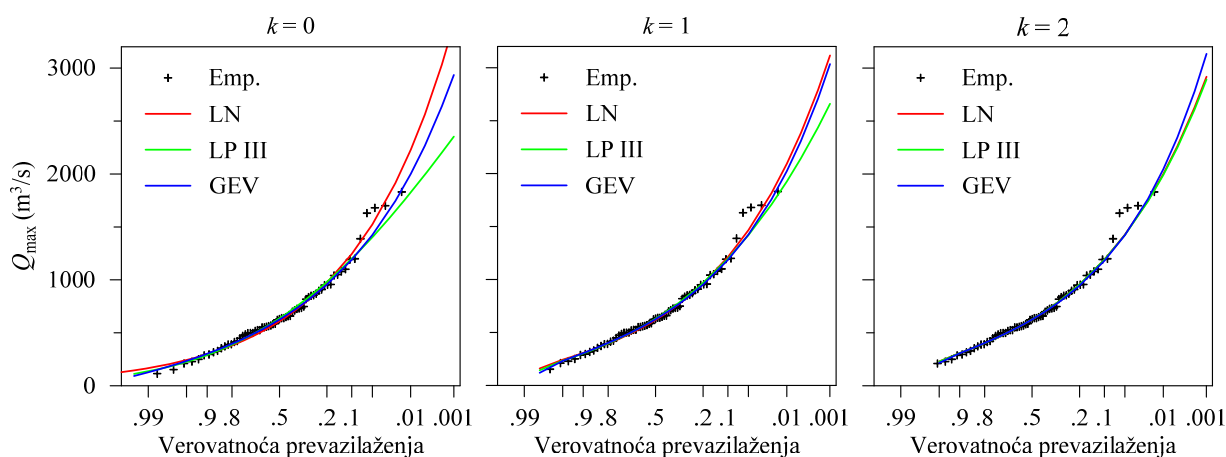
Reka/Stаница	Površina sliva (km ²)	n	Osm. min. (m ³ /s)	Osm. maks. (m ³ /s)	Koef. asim. c_{sy}	Broj izuzetaka k		
						B17B	GB1	TM
Lužnica/Svođe	319	53	6.6	488	-0.07	1	1	1
Z. Morava/K. Stena	3077	87	99.4	1250	0.30	1	1	3
Lopatnica/Bogutovac	116	55	7.39	200	0.39	1	1	1
Ljig/Bogovađa	679	60	21.9	651	0.42	1	1	1
Vlasina/Svođe	350	60	4.62	578	0.48	1	1	1
Ibar/Lopatnica Lakat	7818	68	115	1520	0.55	1	2	3
Z. Morava/Jasika	14721	65	258	2150	0.56	1	1	2

U tabeli 2 je prikazano da je u nizu sa stanice Slovac na Kolubari ($c_{sy} < -0.4$) prema Biltenu 17B prvo testiran i detektovan donji izuzetak, a zatim je taj izuzetak uklonjen iz niza da bi bio testiran i detektovan gornji izuzetak. Na stanici Ušće na Studenici prvo je testiran i detektovan gornji izuzetak, a zatim i donji izuzetak.

Efekat uklanjanja različitog broja donjih izuzetaka prikazan je na slici 3 za stanicu Mojsinje na Južnoj Moravi. Na ovom primeru se može videti da izostavljanje izuzetaka više utiče na oblik log-normalne i log-Pirson III raspodele nego na oblik opšte raspodele ekstremnih vrednosti. Gornji kraj log-normalne raspodele se spušta sa izbacivanjem donjih izuzetaka zbog smanjene standardne devijacije logaritmovanih protoka iako se njihova srednja vrednost povećava, pa se kvantili velikih voda prema ovoj raspodeli smanjuju (tabela 4). Raspodela log-Pirson III se prilagođava većem koeficijentu asimetrije pa daje i veće kvantile. Opšta raspodela ekstremnih vrednosti (GEV) je najotpornija na uklanjanje izuzetaka

jer se njen gornji kraj u zoni razumne ekstrapolacije krive raspodele (npr. do povratnog perioda 100 godina) veoma malo pomera. Međutim, u zoni većih povratnih perioda i ova raspodela se pomera sa uklanjanjem donjih izuzetaka, što je u skladu sa većim neizvesnostima u ovoj zoni. Slični rezultati se dobijaju i za druge nizove sa donjim izuzecima.

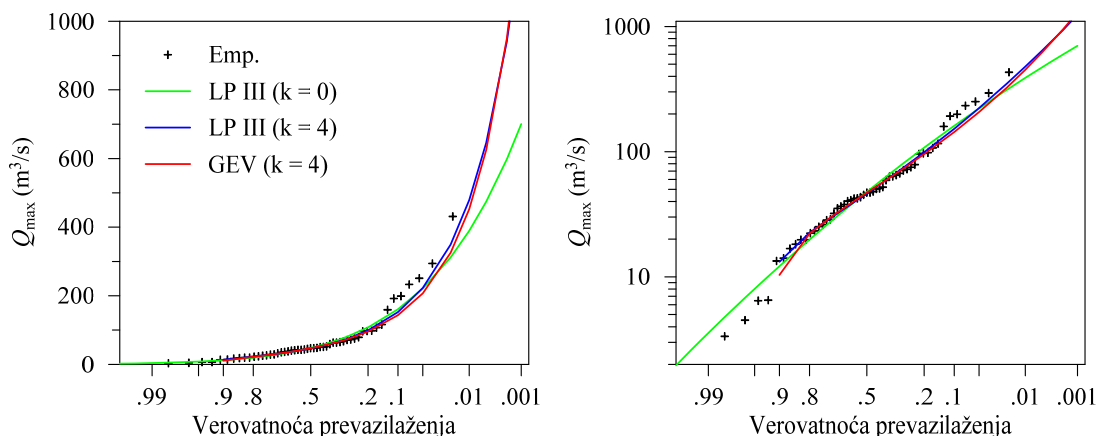
Dva niza u kojima su detektovani višestruki donji izuzeci jedino prema testu Tietjen-Moora (stanice Guča na Belici i Majur na Lugomiru) su specifična po tome što ovi izuzeci ne odstupaju značajno po veličini u odnosu na naredne vrednosti u nizu, ali čine uočljivu grupu. U slučaju stanice Majur na Lugomiru, prema TM testu detektovana je grupa od četiri donja izuzetka koja se uočava tek kada se empirijska raspodela nacrtala sa protocima u logaritamskoj razmeri (slika 4). I u ovim primerima GEV raspodela je najotpornija na otklanjanje cele grupe izuzetaka, dok je log-Pirson III raspodela znatno osetljivija.



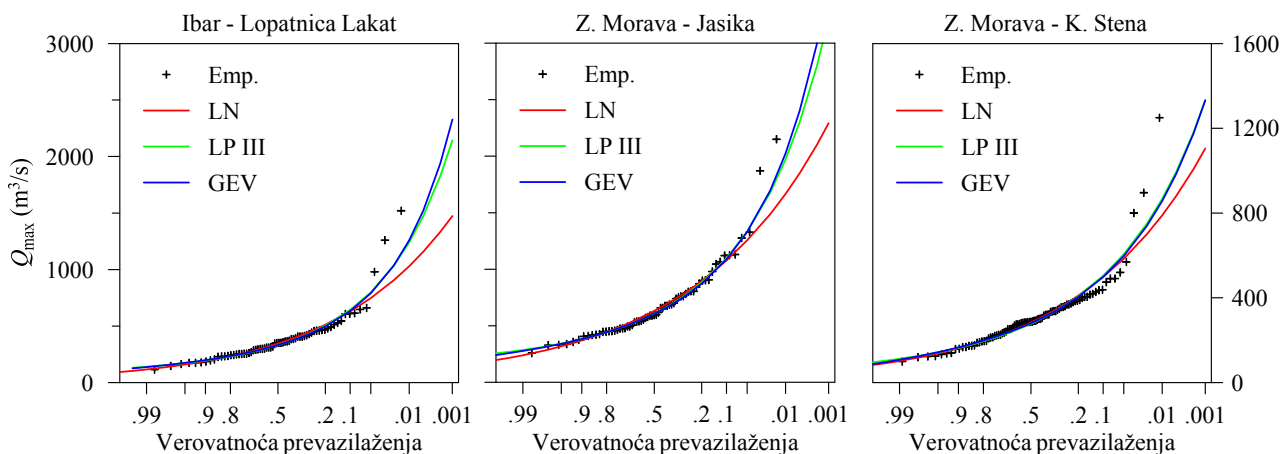
Slika 3. Uticaj donjih izuzetaka na raspodelu maksimalnih godišnjih protoka na stanici Mojsinje na Velikoj Moravi (k je broj donjih izuzetaka uklonjenih iz niza).

Tabela 4. Statistike logaritmovanih nizova i računске 100-godišnje i 500-godišnje velike vode posle uklanjanja izuzetaka na stanici Mojsinje na Južnoj Moravi (u zagradama su procentualne razlike u odnosu na ceo niz).

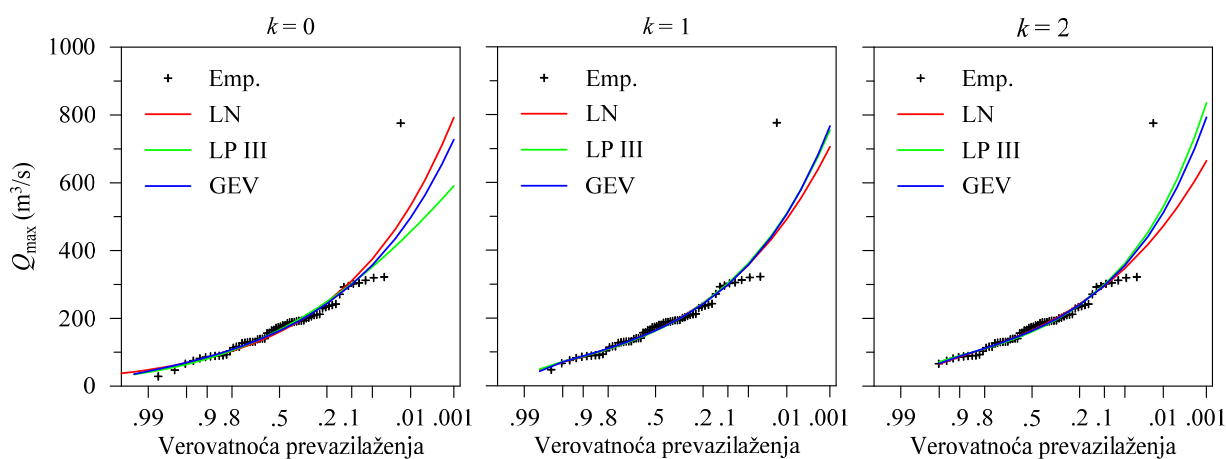
Broj izuzetaka	m_y	s_y	c_{sy}	X_{100} (m^3/s)			X_{500} (m^3/s)		
				LN	LP3	GEV	LN	LP3	GEV
$k=0$	2.784	0.242	-0.481	2228	1825	2001	3032	2202	2638
$k=1$	2.795	0.226	-0.217	2029	1925	2023	2790	2436	2708
				(-6.1%)	(+5.5%)	(+1.1%)	(-8.0%)	(+10.6%)	(+2.6%)
$k=2$	2.805	0.214	-0.013	2000	1991	2044	2628	2608	2778
				(-10.3%)	(+9.0%)	(+2.2%)	(-13.3%)	(+18.4%)	(+5.3%)



Slika 4. Uticaj 4 donjih izuzetaka prema testu Tietjen-Moora na raspodelu maksimalnih godišnjih protoka na stanici Majur na Lugomiru (k je broj donjih izuzetaka uklonjenih iz niza), u aritmetičkoj (levo) i logaritamskoj (desno) razmeri.



Slika 5. Nizovi sa višestrukim gornjim izuzecima prema testu Tietjen-Moora.



Slika 6. Uticaj donjih izuzetaka na raspodelu maksimalnih godišnjih protoka na stanici Slovaca na Kolubari (k je broj donjih izuzetaka uklonjenih iz niza) u prisustvu jednog gornjeg izuzetaka.

4.2 Gornji izuzeci

Gornji izuzeci su detektovani u sedam nizova (tabela 3), pri čemu tri primenjena testa identifikuju izuzetke u istim nizovima, ali je broj detektovanih izuzetaka različit. U ovoj grupi svi nizovi osim jednog imaju pozitivnu asimetriju logaritmovanih protoka. Rezultati testa Grubs-Beka prema Biltenu 17B i u uzastopnoj primeni daju iste rezultate osim u slučaju stanice Lopatnica Lakat na Ibru čija jedna najveća vrednost predstavlja gornji izuzetak u celom nizu, dok se sledeća najveća vrednost iskazuje kao gornji izuzetak tek posle uklanjanja najveće vrednosti. U slučaju ove stanice su se očigledno javile tri izuzetno velike vode (slika 5), koje je kao izuzetke detektovao samo TM test. Prema ovom testu, još dve stanice imaju više gornjih izuzetaka (Jasika i Kratovska Stena na Zapadnoj Moravi, takođe prikazane na slici 5), što može biti i efekat utapanja, ali istovremeno može biti i indikacija da veći broj izuzetaka ukazuje da se u ovim slučajevima radi o mešavini raspodela (što zahteva dodatne analize; primer primene mešovitenih raspodela prikazan je u radu Stojkovića i sar., 2016).

4.3 Izuzeci na oba kraja

Rezultati testova na dva niza u kojima su pronađeni i donji i gornji izuzeci su dosledni (tabela 2), tj. svi testovi detektuju po jedan donji i jedan gornji izuzetak, osim postupka iz Biltenu 17C prema kome u nizu na stanici Slovac postoje dve potencijalno uticajne male vrednosti. Kada se iz oba niza ukloni donji izuzetak, gornji izuzetak je i dalje značajan. Efekat uklanjanja donjeg izuzetka iz ova dva niza je isti kao za ostale nizove sa donjim izuzecima (rezultati za stanicu Slovac su prikazani na slici 6 i u tabeli 5).

Tabela 5. Računske 100- i 500-godišnje velike vode posle uklanjanja k izuzetaka na stanici Slovac na Kolubari.

k	X_{100} (m^3/s)			X_{500} (m^3/s)		
	LN	LP3	GEV	LN	LP3	GEV
0	534	456	498	710	552	656
1	492	509	507	638	676	682
	(-8%)	(12%)	(2%)	(-10%)	(22%)	(4%)
2	472	529	512	605	734	700
	(-12%)	(16%)	(3%)	(-15%)	(33%)	(7%)

I u ovom slučaju se opšta raspodela ekstremnih vrednosti (GEV) veoma malo pomera sa uklanjanjem izuzetaka. Razlika u kvantilima velikih povratnih perioda pre i posle uklanjanja izuzetaka je svega nekoliko procenata, što se može objasniti time da gornji ekstremi, koji su stalno prisutni u nizu bez obzira na prisustvo donjih vrednosti, zapravo „fiksiraju“ položaj gornjeg kraja

GEV raspodele, dok su ostala dva tipa raspodele znatno osetljivija na donje nego na gornje izuzetke.

4.4 Poređenje zasecanja niza s donje strane sa metodom pikova

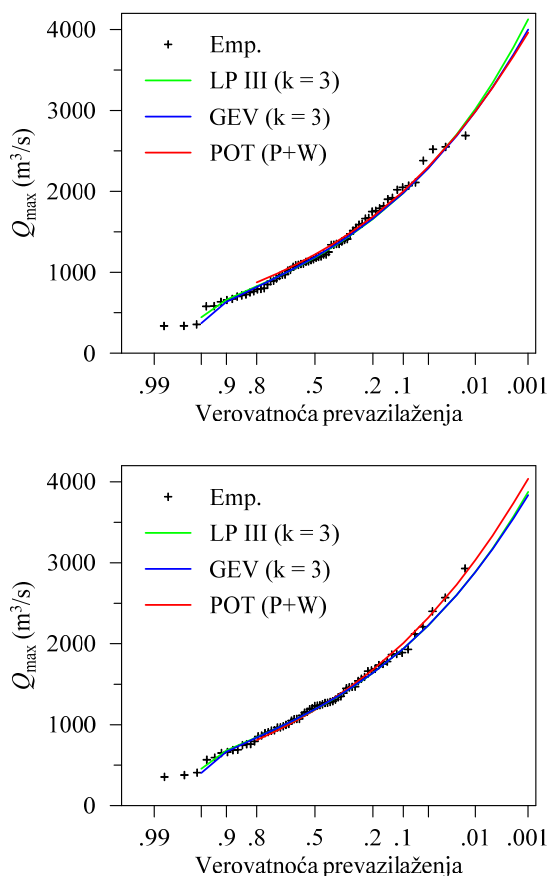
Broj potencijalno uticajnih donjih vrednosti prema Biltenu 17C može da bude neočekivano veliki. Na primer, čak 8 takvih vrednosti je detektovano u nizu na stanici Ljubičevski Most na Velikoj Moravi, a raspon tih vrednosti je od 353 do 685 m^3/s . Odstranjivanje većeg broja vrednosti sa donjeg kraja niza, ili čak i odstranjivanje manjeg broja vrednosti ali sa relativno velikim rasponom, približava postupak iz Biltenu 17C ka postupku koji se primenjuje u metodi prekoračenja preko praga (tj. POT metodi).

Na slici 7 su na primerima stanica Varvarin i Bagrdan prikazane raspodele dobijene prema Biltenu 17C (gde je primenjena log-Pirson III raspodela na niz iz koga su uklonjeni donji izuzeci uz uzimanje u obzir uslovne raspodele zasečenog skupa podataka prema jednačini 5) i prema POT metodi (u oba slučaja primenjen je Poason-Vejbulov model za prekoračenja preko praga od 800 m^3/s). U ovim primerima se pokazalo da POT metoda i zasecanje niza godišnjih maksimuma sa donje strane daju skoro identične rezultate. Ovu hipotezu bi trebalo ispitati i na drugim nizovima.

4. ZAKLJUČAK

Primena različitih testova za detekciju izuzetaka u nizovima maksimalnih godišnjih protoka, pod pretpostavkom važenja log-normalne raspodele, može da da različite rezultate. Najveće razlike u rezultatima testiranja se javljaju kada postoji grupa vrednosti koja se izdvaja od ostatka niza, pa testovi koji su namenjeni testiranju samo jedne izuzetne vrednosti ne detektuju izuzetke. Testovi u kojima se odjednom testira grupa izuzetnih vrednosti najčešće i identifikuju celu grupu kao izuzetke, a ređe samo jednu vrednost.

Rezultati testiranja prisustva izuzetaka se najviše odražavaju na tretman donjih izuzetaka u nizu, koji se po pravilu odbacuju i statistička analiza se sprovodi na zasečenom skupu podataka s donje strane. Uobičajene raspodele koje se koriste u statističkoj analizi velikih voda su generalno osetljive na izostavljanje donjih izuzetaka, što je u ovom radu pokazano za log-normalnu i log-Pirson III raspodelu na primerima razmatranih nizova. Opšta raspodela ekstremnih vrednosti se pokazala kao najotpornija na prisustvo donjih izuzetaka, što je čini pogodnom za preliminarne analize.



Slika 7. Poređenje raspodela maksimalnih godišnjih protoka za zasečene nizove godišnjih maksimuma i prema POT metodi za stanice Varvarin (gore) i Bagrdan (dole) na Velikoj Moravi.

S druge strane, objavljivanje predloga novih preporuka za statistički proračun velikih voda u SAD kroz Bilten 17C u fokus je stavio potrebu da se iz raspoloživih nizova godišnjih maksimuma izvuku suštinske informacije o zaista ekstremnim velikim vodama i iz analize odbace male vrednosti. Ovakav postupak je ne samo praktičan već približava metodu godišnjih maksimuma metodi prekoračenja preko praga, za koju je već poznato da daje pouzdanije ocene velikih voda.

U ovom radu je na raspoloživim primerima za stanice na našoj teritoriji pokazano da se, uz pažljivu analizu, metodom godišnjih maksimuma uz odbacivanje donjih izuzetaka i metodom prekoračenja preko praga dolazi do praktično iste raspodele velikih voda. Ovakve zaključke bi sigurno trebalo proveriti na više primera kako bi se mogle dati konkretne preporuke za praksu.

Donji izuzeci i njihov tretman su nesumnjivo veoma važni za statističku analizu velikih voda, ali su i gornji izuzeci ti koji često otežavaju izbor najboljih raspodela i pouzdanu ocenu velikih voda malih verovatnoća pojave izvan raspona osmotrenih podataka. Postojanje gornjih izuzetaka često navodi na razmišljanja o tome da podaci u nizu potiču iz mešavine raspodela. Primenom metode prekoračenja preko praga ili uz identifikaciju potencijalno uticajnih donjih vrednosti prema Biltenu 17C povećavaju se mogućnosti za identifikaciju mešavine raspodela.

Problem izuzetaka u nizovima velikih voda je očigledno povezan i s mnogim drugim problemima kao što je nepoznavanje raspodele populacije velikih voda, nehomogenost podataka, relativno kratki nizovi iz kojih se potencijalno mešovite raspodele teško identifikuju. Zbog toga i dalje ne postoji jednostavna i „jednosmerna“ procedura za statističku analizu velikih voda. U praksi je i dalje neophodno da se svaki niz pažljivo razmotri, a naročito ocene najvećih zabeleženih podataka u svetlu pouzdanosti krivih protoka i merenja u ekstremnim uslovima. Uz sav oprez, preporučuje se primena robustnih metoda i raspodela (kao što je rad sa zasečenim skupovima podataka, metoda prekoračenja ili primena GEV raspodele). Raspoložive nizove godišnjih maksimuma i njihovu statističku analizu bi svakako trebalo sprovesti i u regionalnom kontekstu kako bi se izbegli nekonzistentni rezultati.

ZAHVALNOST

Ovaj rad je nastao u okviru istraživanja u projektu tehnološkog razvoja TR 37005 kod Ministarstva prosvete, nauke i tehnološkog razvoja pod nazivom „Ocena uticaja klimatskih promena na vodne resurse Srbije“. Autori se takođe zahvaljuju Republičkom hidrometeorološkom zavodu Republike Srbije za stavljanje podataka za istraživanje na raspolaganje.

LITERATURA

- [1] Blagojević B., Mihailović V., Plavšić J. (2013) New Guidelines for Flood Flow Assessment at Hydrologic Stations in Serbia. Electronic Proceedings of the International Conference on Flood Resilience: Experiences in Asia and Europe 5-7 September 2013, Exeter, United Kingdom. Djordjević, S., Butler, D., Chen, A. (Eds.). ISBN: 978-0-9926529-0-6. E1_298_Blagojevic, 10 p.
- [2] Blagojević B., Mihailović V., Plavšić J. (2014a) Outlier treatment in the flood flow statistical analysis. Proc. Int. Conf. on Contemporary

- Achievements in Civil Eng., Faculty of Civil Engineering Subotica, University of Novi Sad, Subotica, Serbia, 24-25 April 2014, pp. 603-609, DOI: 10.14415/konferencijaGFS2014.081.
- [3] Blagojević B., Mihailović V., Plavšić J. (2014b) Statistička analiza velikih voda na profilima hidroloških stanica: potreba za promenom pristupa, *Vodoprivreda* 46: 267-272.
- [4] BLFUW (2011). Leitfaden Verfahren zur Abschätzung von Hochwasserkennwerten (Guidelines for Flood Flow Assessment), P. Lorenz (ed.), Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien, Austria.
- [5] Cohn, T. A., J. F. England, C. E. Berenbrock, R. R. Mason, J. R. Stedinger, and J. R. Lamontagne (2013) A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series, *Water Resour. Res.*, 49, 5047–5058, doi:10.1002/wrcr.20392
- [6] England Jr., J.F., Cohn T.A., Faber B.A., Stedinger J.R., Thomas Jr., W.O., Veilleux A.G., Kiang J.E., and Mason R.R. (2016) Guidelines for Determining Flood Flow Frequency – Bulletin 17C: U.S. Geological Survey Techniques and Methods 4–BXX, draft report, doi: 10.3113/tm4–BXX/.
- [7] Flynn K.M., Kirby W.H., Hummel P.R. (2006) User's Manual for Program PeakFQ Annual Flood-Frequency Analysis Using Bulletin 17B Guidelines: U.S. Geological Survey, Techniques and Methods Book 4, Chapter B4, 42 pgs.
- [8] Grubbs, F.E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Stat.*, 21(1), 27-58.
- [9] Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples, *Technometrics*, 11(1), 1–21.
- [10] Grubbs, F. E., and G. Beck (1972). Extension of sample sizes and percentage points for significance tests of outlying observations, *Technometrics*, 14(4), 847–854.
- [11] IACWD (1982). Guidelines for determining flood flow frequency: Bulletin 17B, Interagency Advisory Committee on Water Data, Hydrol. Subcomm., Washington, D.C.
- [12] Kottegoda, N., Rosso R. (2008). Applied Statistics for Civil and Environmental Engineers, Blackwell.
- [13] McCuen, R.H. (2002). Outlier Detection. Chapter 4 in *Modelling Hydrologic Change: Statistical Methods*, Lewis Publishers, CRC Press.
- [14] Plavšić, J. (2005) *Analiza rizika od poplava pomoću prekidnih slučajnih procesa*, doktorska disertacija, Građevinski fakultet Univerziteta u Beogradu, 318 str.
- [15] Plavšić, J. (2006) Neizvesnosti u analizi velikih voda metodom parcijalnih serija, *Vodoprivreda*, br. 219-221, 38 (1-3), str. 41-50.
- [16] Plavšić, J., Mihailović, V. and Blagojević, B. (2014) Assessment of methods for outlier detection and treatment in flood frequency analysis. Proc. Mediterranean Meeting on Monitoring, modelling, early warning of extreme events triggered by heavy rainfall, University of Calabria, Cosenza, E. Ferrari and P. Versace (eds.), ISBN 978-88-6822-268-0, pp. 181-192.
- [17] Rosner, B. (1975). On the detection of many outliers, *Technometrics*, 17(2), 221–227.
- [18] Spencer, C.S., and McCuen, R.H. (1996) Detection of Outliers in Pearson type III data. *J. Hydrol. Eng.*, 1(1), 2-10.
- [19] Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E. (1993) Frequency Analysis of Extreme Events. In: *Handbook of Applied Hydrology*, D. Maidment (ed.), Mc-Graw Hill.
- [20] Stojković M., Prohaska S., Zlatanović N. (2016) Estimation of flood frequencies from data sets with outliers using mixed distribution functions, *J. Appl. Statistics*, doi: 10.1080/02664763.2016. 1238055.
- [21] Tietjen, G., and R. Moore (1972). Some Grubbs-type statistics for the detection of several outliers, *Technometrics*, 14(3), 583–597.
- [22] Verma, S.P., and Quiroz-Ruiz, A. (2006). Critical values for 22 discordancy test variants for outliers in normal samples up to sizes 100, and applications in science and engineering. *Revista Mexicana de Ciencias Geológicas*, 23(3), 302-319.
- [23] Todorović, P. i Zelenhasić, E. (1970) A stochastic model for flood analysis, *Water Resour. Res.*, 6(6): 1641-1648.
- [24] U.S. Army Corps of Engineers (2016) Statistical Software Package HEC-SSP User's Manual, Version 2.1, US Army Corps of Engineers – Hydrologic Engineering Center.
- [25] Vukmirović V. (1990) *Analiza verovatnoće pojave hidroloških veličina*, Naučna knjiga.
- [26] Vukmirović, V., Pavlović, D. (2000). Utvrđivanje kriterijuma za izbor merodavnih velikih voda. Izveštaj za naučno istraživački projekat TSI 114 „Savremene metode u hidrotehnici“, Građevinski fakultet Univerziteta u Beogradu.
- [27] Vukmirović, V. i Petrović, J. (1997) Flood flow analysis using renewal processes, UNESCO-IHP V Tech. Documents in Hydrology No. 11 (Annual FRIEND-AMHY meeting, Thessaloniki, 1995), 159-169.

FLOOD FLOW FREQUENCY ANALYSIS IN THE PRESENCE OF OUTLIERS

by

Jasna PLAVŠIĆ

University of Belgrade – Faculty of Civil Engineering, jplavsic@grf.bg.ac.rs

Dragutin PAVLOVIĆ

University of Belgrade – Faculty of Civil Engineering

Vladislava MIHAILOVIĆ

University of Belgrade – Faculty of Forestry

Borislava BLAGOJEVIĆ

University of Niš – Faculty of Civil Engineering and Architecture

Summary

Flood flow frequency analysis is performed on the series of hydrological observations. The observations introduce uncertainties in the analysis results, inherited from limited data sets, operational difficulties in high flow measurements, flood flow estimates in the rating curve extrapolation zone, as well as different mechanisms of runoff generation. Difficulties in defining flood flow probabilistic model arise in the presence of outliers (significantly larger or smaller value compared to the remaining data set values). The paper provides an overview of methods for the detection of outliers from traditional practice to the latest research results, as well as procedures for outlier treatment in the further statistical analysis, focusing especially on the recommendations from the recently published Bulletin 17C in the USA. The assumptions about the population, procedures of outlier detection, the treatment of

transformed data sets, and potential difficulties in the application are highlighted. Statistical tests and procedures are applied to flood data at 70 hydrological stations in Serbia. Analysis of flood flow frequency in the presence of outliers is performed by the methods of the annual maxima and peaks over threshold (POT). The examples include both upper and lower outliers. Comments are given on the differences in the results from the applied outlier detection and treatment procedures. Recommendations to the practitioners are given in the conclusions in order to reduce uncertainty in the flood flow frequency analysis in the presence of outliers.

Key words: flood flows, statistical analysis, annual maximum flow, peaks over threshold, outliers, statistical tests, flood flow frequency

Redigovano 14.11.2016.