STATISTIČKO MODELIRANJE EKSTREMNIH VREDNOSTI: PRIMENA NA PRORAČUN EKSTREMNIH PROTOKA NA RECI RASINI

STATISTICAL MODELLING OF EXTREME VALUES: APPLICATION TO CALCULATE EXTREME FLOW AT RIVER RASINA

Miljan KOVAČEVIĆ
Ljubo MARKOVIĆ
Lidija BABIĆ

STRUČNI RAD
PROFESSIONAL PAPER
UDK: 556.535:519.2(497.11)

1 UVOD

Građevinski inženjeri često se bave prirodnim rizicima. Ekstremni događaji, kao što su poplave, uragani, zemljotresi, mogu odneti mnogo ljudskih života i izazvati štete od više milijardi dolara.

Paradoksalno, takođe je potrebno znati posledice i neuobičajeno niskih protoka, koji mogu da rezultuju visokim koncentracijama zagađivača. I pored toga, inženjeri moraju da projektuju građevinske konstrukcije tako da mogu da izdrže ekstremne poplave, jake vetrove, a konstrukcije na moru da izdrže neuobičajeno visoke talase. Opstanak datog sistema zavisice od sposobnosti da izdrže ove ekstremne uslove kojima su izloženi, a koji nisu uobičajene vrednosti [7].

Ciljevi ovoga rada jesu

(i) modeliranje godišnjeg maksimalnog protoka na reci Rasini, na jugu centralne Srbije, sa Gumbelovom, Frešetovom i GEV raspodelom,

(ii) upotreba Kolmogorov-Smirnov testa i Anderson-Darling testa prilikom testiranja i upoređivanja empirijskih i teorijskih funkcija raspodele koje predviđaju ove tri raspodele,

(iii) nalaženje potrebnih projektnih vrednosti za dati povratni period.

1 INTRODUCTION

Civil engineers are often concerned with natural hazards. Extreme events, such as floods, hurricanes, and earthquakes, can take many human lives and cause billions of dollars in damages.

Paradoxically, one also needs to address the consequences of unusually low streamflows, which can result in high pollutant concentrations. Additionally, engineers must design construction facilities to withstand extreme floods, high winds and maritime structures to cater for abnormally high sea waves. The survival of a given system depends on its capability to resist those extreme conditions it can be subject to - and not simply the typical values [7].

The objectives of this study are to

(i) model the annual maximum streamflows at river Rasina in south central Serbia with the Gumbel, Frechet and general extreme value distributions,

(ii) use the Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit tests to compare the observed and theoretical cumulative frequencies as predicted by the three distributions,

(iii) obtain required design values for given return periods of exceedance.

Miljan Kovačević, dipl.inž.građ., asistent, Fakultet tehničkih nauka, Univerzitet u Prištini, Kneza Miloša 7, 38220 Kosovska Mitrovica, Srbija
miljankovacevic@yahoo.com
Doc.dr Ljubo Markovic, dipl.inž.građ., Fakultet tehničkih nauka, Univerzitet u Prištini, Kneza Miloša 7, 38220 Kosovska Mitrovica, Srbija
ljubo.markovic@pr.ac.rs
Mr Lidija Babić, dipl.inž.građ., asistent, Fakultet tehničkih nauka, Univerzitet u Prištini, Kneza Miloša 7, 38220 Kosovska Mitrovica, Srbija, lidija.babic@pr.ac.rs

Miljan Kovacevic, BCE., Research Assistant, Faculty of Technical Sciences, University of Pristina, Kneza Miloša 7, 38220 Kosovska Mitrovica, Serbia
miljankovacevic@yahoo.com
Assistant Professor Ljubo Markovic, BCE, Faculty of Technical Sciences, University of Pristina, Kneza Miloša 7, 38220 Kosovska Mitrovica, Serbia
ljubo.markovic@pr.ac.rs
Mr Lidija Babic, CE., Assistant professor, Faculty of Technical Sciences, University of Prishtina, Kneza Miloša 7, 38220 Kosovska Mitrovica, Serbia, lidija.babic@pr.ac.rs

Tok je dugačak 92 km kroz Rasinski region i uliva se u Zapadnu Moravu blizu grada Kruševca. Podaci za dvadeset sedam maksimalnih godišnjih protoka sumirani su u Tabeli 1.

The 92 km long river flows through the Rasina region, and flows into the Zapadna Morava near the city of Kruševac. The data of 27 annual maximum streamflows are summarized in Table 1.

Tabela 1. Maksimalni godišnji protok na reci Rasini Table 1. Annual maximum streamflow at river Rasina					
Year	m ³ /s	Year	m ³ /s	Year	m ³ /s
1966	57.30	1991	67.10	2000	30.10
1968	15.80	1992	82.00	2001	69.40
1969	108.00	1993	32.10	2002	50.30
1970	44.10	1994	55.50	2003	19.40
1985	40.00	1995	162.20	2004	22.50
1987	69.00	1996	52.40	2005	166.90
1988	53.00	1997	131.00	2006	82.60
1989	66.10	1998	99.10	2007	36.80
1990	23.80	1999	94.50	2008	34.80

2 FUNKCIJE RASPODELE EKSTREMNIH VREDNOSTI

Suština teorije ekstremnih vrednosti jesu proučavanja statističkog ponašanja slučajne veličine $X = \max\{X_1, \dots, X_n\}$, gde $\{X_1, \dots, X_n\}$ predstavlja niz slučajnih veličina koje imaju zajedničku funkciju raspodele F . Promenljive X obično predstavljaju vrednosti koje su izmerene u jednakim vremenskim intervalima. X je maksimum posmatranog procesa sastavljenog od n vremenskih jedinica [9]. Maksimalni godišnji protok predstavlja se obično Gumbelovom raspedelom [1]. Alternativa Gumbelovoj raspodeli jeste Frechet-ova raspodela [2] ili GEV raspodela (generalizovana raspodela ekstremnih vrednosti) [3], [4].

2.1 Gumbelova raspodela

Postoje tri tipa raspodele ekstremnih vrednosti. Tip I raspodele je intenzivno razvijan i u slučajevima poplavnih proticaja primenjivao ga je Gumbel; zbog toga je ova raspodela i poznata kao Gumbelova raspodela [3]. Funkcija raspodele za Gumbelovu raspodelu slučajne varijable X data je kao

$$F(x) = \exp[-e^{-(x-\beta)/\alpha}] , -\infty < x < \infty, \quad (1)$$

gde je β parametar lokacije i α je parametar razmere. Srednja vrednost i disperzija X date su respektivno izrazima

$$\mu = E[X] = \beta + \gamma_c \alpha , -\infty < x < \infty, \quad (2)$$

$$\sigma^2 = Var[X] = \frac{\pi^2 \alpha^2}{6} \quad (3)$$

gde γ_c označava Ojlerovu konstantu, približno jednaku 0.5772. Koeficijent asimetrije je 1.1629. Ako su prva dva

2 EXTREME VALUE DISTRIBUTIONS

The core of the extreme value theory is the study of the statistical behavior of $X = \max\{X_1, \dots, X_n\}$, where $\{X_1, \dots, X_n\}$ is a sequence of independent random variables having a common distribution function F . In applications, variables X usually represent values of a process measured on a regular time-scale. Then X is the maximum of the observed process over n time units [9]. The annual maximum streamflows is often represented by the Gumbel distribution [1]. An alternative to the Gumbel distribution is the Frechet distribution [2] or the general (generalized) extreme value (GEV) distribution [3], [4].

2.1 Gumbel Distribution

There are three types of extreme value distributions. The Type I distribution was extensively developed and applied to the flood flows by Gumbel; therefore, it is often referred to as the Gumbel distribution [3]. The cumulative distribution function of the Gumbel random variable X is given as

where β is the location parameter and α is the scale parameter. The mean and the variance of X are given respectively as

where γ_c denotes the Euler constant, approximately equal to 0.5772. The skewness coefficient is 1.1629. If

momenta slučajne promenljive X poznata, vrednost α i β mogu se odrediti metodom momenata na osnovu srednje vrednosti i standardne devijacije raspodele X. Iz gornjih dveju jednačina (2) i (3), možemo dobiti [7],

$$\alpha = \sqrt{\frac{6}{\pi\sigma}} \quad (4)$$

$$\beta = \mu - n_c \alpha \quad (5)$$

Ove dve jednačine mogu se koristiti da se ocene α i β ako imamo dostupan konačan uzorak varijable X, kao što je maksimalni protok za period od n godina.

the first two moments of X are known, the values of α and β can be determined by the method of moments from the mean and standard deviation σ of the distribution X. From the above two equations (2) and (3), one obtains [7],

The two equations can be used to estimate α and β if a finite sample of the values taken by X is available, such as the annual maximum streamflows for a period of n years.

2.2 Frechet-ova raspodela

Frechet-ova raspodela je posebna forma Tip II raspodele ekstremnih vrednosti. EV2 raspodelu je razvio i primenio Frechet (1927). Prvenstveno je razvijena i primenjena na poplavne protoke M.Frecheta [4]. Funkcija raspodele verovatnoće u M.Frechet raspodeli i data je kao

$$F(x) = \exp\left[-\left(\frac{x}{\eta}\right)^{\theta}\right], x > 0 \quad (6)$$

gde $\eta > 0$ predstavlja faktor razmere a $\theta > 0$ je parametar oblika. Moment r-tog reda dat je sledećim izrazom [7],

$$E[X]^r = \eta^r \Gamma\left(1 - \frac{r}{\theta}\right), r < \theta \quad (7)$$

Shodno tome iz (7) imamo

Consequently, from (7) we have

$$F(x) = \exp[-e^{\frac{1}{\theta}}(-(x - \beta)/\alpha)], -\infty < x < \infty, \quad (8)$$

$$\mu = E[X] = \eta \Gamma\left(1 - \frac{1}{\theta}\right), \theta > 1 \quad (9)$$

pošto

since

$$V^2 = \frac{\sigma^2}{\mu^2} = \frac{\Gamma\left(1 - \frac{2}{\theta}\right)}{\Gamma^2\left(1 - \frac{1}{\theta}\right)} - 1, \quad \theta > 2 \quad (10)$$

parametar oblika zavisi samo od koeficijenta varijacije.

Ako prva dva momenta od X postoje i ako su poznata, vrednost parametra α i θ mogu se odrediti iz srednje vrednosti, μ , i disperzije σ^2 iz populacije ekstremnih vrednosti. Međutim, koristeći srednju vrednost i koeficijent varijacije vrlo je jednostavno naći, zato što jednačina (10) ukazuje na to da parametar oblika θ zavisi samo od V. Pošto je V određeno (kao količnik od uzoračke standardne devijacije i srednje vrednosti), jednačina (10) mora se rešiti numeričkim postupkom da bi se našlo θ . Onda, koristeći jednačinu (8),

the shape parameter only depends on the coefficient of variation.

If the first two moments of X exist and are known, the values of the parameters α and θ can be determined from the mean, μ , and the variance, σ^2 , of the extreme value population. However, using the mean and the coefficient of variation V is relatively straightforward, because Eq. (10) indicates that the shape parameter θ depends only on V . After V is estimated (as the ratio of the sample standard deviation to the sample mean), Eq. (10) must be solved via numerical iteration to find θ . Then, using Eq. (8),

$$x_0 = \eta = \frac{\mu}{\Gamma\left(1 - \frac{1}{\theta}\right)}, \theta > 1 \quad (11)$$

možemo oceniti parametar razmere.

one can estimate the value of the scale parameter

2.3 GEV funkcija raspodele

Generalizovanu raspodelu ekstremnih vrednosti (GEV) uveo je od Jenkinson [1], [2] da bi se odredila frekvencija raspodele najvećih vrednosti meteoroloških podataka kada je granična forma ekstremnih vrednosti nepoznata. Funkcija raspodele GEV raspodele data je sa sledećim izrazom [5], [6] [9],

$$F(x) = \exp\left\{-\left[1 + k\left(x - \frac{\varepsilon}{\alpha}\right)\right]^{-\frac{1}{k}}\right\}, \quad (12)$$

gde α predstavlja parametar razmere, ε je lokacijski parametar, a k je parametar oblika. Treba imati na umu da za $k < 0$, GEV raspodela predstavlja EV2 raspodelu i ona je definisana samo za $x > (\varepsilon + \alpha/k)$; za $k > 0$, ovaj model postaje EV3 raspodela, i ona je definisana samo za $x < (\varepsilon + \alpha/k)$; slučaj da je $k = 0$ odgovara Gumbelovoj raspodeli jednačina (1) s parametrom razmere α i parametrom lokacije β [7].

Srednja vrednost i disperzija dati su za GEV raspodelu sa

$$\mu = E[X] = \varepsilon + \left(\frac{\alpha}{k}\right) [\Gamma(1 - k) - 1], k < 1 \quad (13)$$

i

and

$$\sigma^2 = Var[X] = \left(\frac{\alpha}{k}\right)^2 [\Gamma(1 - 2k) - \Gamma^2(1 - k)], k < 0.5 \quad (14)$$

respektivno. Koeficijent asimetrije je dat sa

respectively. The coefficient of skewness is given by

$$\gamma = \text{sign}(k) \frac{-[\Gamma(1 - 3k) + 3\Gamma(1 - k)\Gamma(1 - 2k) + 2\Gamma^3(1 - k)]}{[\Gamma(1 - 2k) - \Gamma^2(1 - k)]^{\frac{3}{2}}}, k > \frac{1}{3}. \quad (15)$$

gde je znak $(k) = +1$ za $k > 0$ i znak $(k) = -1$ za $k < 0$, dok nije definisano za $k < -1/3$; dakle možemo primetiti da parametar oblika zavisi samo od koeficijenta asimetrije ako treći moment postoji. Ako prva tri momenta od X_{max} postoje i ako su poznati, ε , α , i k mogu se oceniti na osnovu srednje vrednosti, disperzije i koeficijenta zakrivljenosti podataka. Pošto jednačina (15) ukazuje na to da k samo zavisi od koeficijenta asimetrije za $k > -1/3$, možemo rešiti ovu jednačinu po k , zamenjujući uzorački koeficijent asimetrije [7]. Onda, na osnovu jednačine (14) možemo naći parametar razmere kao

where $\text{sign}(k) = +1$ for $k > 0$ and $\text{sign}(k) = -1$ for $k < 0$, while it is not defined for $k < -1/3$; therefore, one notes that the shape parameter only depends on the coefficient of skewness if the third moment exists. If the first three moments of X_{max} exist and are known, the values of the three parameters ε , α , and k can be determined from the mean, the variance, and the skewness coefficient of the data. Since Eq. (15) indicates that k only depends on the coefficient of skewness for $k > -1/3$, one can solve this equation in k by substituting the sampling skewness coefficient [7]. Then, from Eq. (14) the scale parameter is found as

$$\alpha = \sqrt{\frac{k^2 \sigma^2}{[\Gamma(1 - 2k) - \Gamma^2(1 - k)]}} \quad (16)$$

gde je uzoračka disperzija zamenjena $\sigma^2 = Var[X]$

Na kraju, lokacijski parametar može se izračunati iz

where the sample variance is substituted for $\sigma^2 = Var[X]$

Finally, the location parameter is computed from

$$\varepsilon = \mu - \left(\frac{n}{k}\right) [\Gamma(1-k) - 1] \quad (17)$$

gde je uzoračka srednja vrednost zamena za μ .

where the sample mean is substituted for μ .

3 TEST STATISTIČKIH HIPOTEZA KOLMOGOROV-SMIRNOVA

Kolmogorov-Smirnovljev test statističkih hipoteza jeste neparametarski test koji se odnosi češće na funkciju raspodele nego na funkciju gustine kontinualne varijable. Nije primenljiv na diskretne varijable. Test statistika, kod dvostranog testa, jeste maksimalna apsolutna vrednost razlike (što je obično vertikalno rastojanje) između empirijske i hipotetičke funkcije raspodele [8], [9].

Za kontinualnu varijablu X neka $x(1), x(2), \dots, x(n)$ predstavlja statistiku poretka veličine n , to su vrednosti poređane u rastućem poretku. Empirijska ili uzoračka funkcija $F_n(x)$ jeste step funkcija. Ovo će dati proporciju vrednosti koje ne prelaze x i definisana je kao

$$F_n(x) = \frac{k}{n}, x_{(k)} \leq x < x_{(k+1)} \quad (18)$$

$$F_n(x) = 0, x < x_{(1)} \quad (19)$$

$$F_n(x) = 1, x \geq x_{(n)} \quad (20)$$

Neka $F_0(x)$ predstavlja kompletno definisanu teorijsku funkciju raspodele CDF. Nulta hipoteza H_0 jeste da je prava funkcija raspodele CD od F_X ista kao i $F_0(x)$.

Kriterijumski test je maksimalna apsolutna vrednost razlike između $F_n(x)$ i $F_0(x)$, formalno definisana kao

$$D_n = \sup_x |F_n(x) - F_0(x)| \quad (21)$$

Gore navedena mera devijacije je za dvostrani test koji se uobičajeno koristi. Za velike vrednosti n , Smirnov (1948) daje granične vrednosti distribucije za $\sqrt{n} D_n$ [7] kao

$$\lim_{n \rightarrow \infty} [Pr[\sqrt{n} D_n \leq z]] = \left(\frac{\sqrt{2\pi}}{z}\right) \sum_{k=1}^{\infty} \exp\left[-(2k-1)^2 \frac{\pi^2}{(8z)^2}\right] \quad (22)$$

Dakle, može se izračunati da su kritične vrednosti za

$D_{n,\alpha}$ za velike uzorke, recimo $n > 35$, su $1.3581\sqrt{\frac{1}{n}}$ i $1.6276\sqrt{\frac{1}{n}}$ za $\alpha = 0.05$ i 0.01 , a to je za verovatnoće od 0.95 i 0.99 u jednačini (22) respektivno. Za manje veličine uzoraka, kritične vrednosti za $D_{n,\alpha}$ date su u tabeli C.7, u apendiksu C. [7].

4 ANDERSON-DARLINGOV TEST STATISTIČKIH HIPOTEZA

Anderson-Darlingov test osmišljen je da dodeli veću težinu repovima raspodele, gde su neočekivano velike ili male vrednosti, zvane izuzeci, ponekad locirane. Ovo je moguće ako se podeli razlika između empirijske funkcije

3 KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST

The Kolmogorov-Smirnov goodness-of-fit test is a nonparametric test that relates to the cdf rather than the pdf of a continuous variables. It is not applicable to discrete variables. The test statistic, in a two-sided test, is the maximum absolute difference (that is, usually the vertical distance) between the empirical and hypothetical cdfs. [8], [9].

For a continuous variate X let $x(1), x(2), \dots, x(n)$ represent the order statistics of a sample of size n , that is, the values arranged in increasing order. The empirical or sample distribution function $F_n(x)$ is a step function. This gives the proportion of values not exceeding x and is defined as

Let $F_0(x)$ denote a completely specified theoretical continuous cdf. The null hypothesis H_0 is that the true cdf of X is the same as $F_0(x)$.

The test criterion is the maximum absolute difference between $F_n(x)$ and $F_0(x)$, formally defined as

The foregoing measure of deviation is for a two-sided test which is commonly used. For large values of n , Smirnov (1948) gives the limiting distribution of $\sqrt{n} D_n$, as [7],

Thus, one can compute that the critical values $D_{n,\alpha}$

for large samples, say, $n > 35$, are $1.3581\sqrt{\frac{1}{n}}$ and $1.6276\sqrt{\frac{1}{n}}$ for $\alpha = 0.05$ and 0.01 , that is, for probabilities of 0.95 and 0.99 in Eq. (22), respectively. For smaller sample sizes, critical values $D_{n,\alpha}$ are given in Table C.7, of Appendix C. [7].

4 ANDERSON-DARLING GOODNESS OF FIT TEST

The Anderson-Darling test is devised to give heavier weighting to the tails of a distribution where unexpectedly high or low values, called outliers are sometimes located. This is made possible if one divides

raspodele $F_n(x)$ i teorijske funkcije raspodele $F_0(x)$ [9], [10],

the difference between the empirical cdf $F_n(x)$ and theoretical cdf $F_0(x)$ by [9], [10],

$$\sqrt{(F_0(x)[1 - F_0(x)])} \quad (23)$$

Posle dizanja na kvadrat test statistika postaje

After squaring the test statistic becomes

$$\frac{\ln [F_0 [x_{(i)} + \ln \{1 - F_0 [x_{((n-i+1))}]\}]]}{n} \quad (24)$$

gde $x(1), x(2), \dots, x(n)$ predstavljaju osmatranja u rastućem poretku. Pošto su funkcije raspodele u rasponu 0-1, njihovi logaritmi su negativni, pa zbog toga imamo da je desna strana jednačine (24) negativna.

Apsolutna vrednost sume je takođe veća od n, pa sve to rezultuje pozitivnom vrednošću za A^2 . Za velike vrednosti test statistika A^2 , odbija se nulta hipoteza da $F_n(x)$ i $F_0(x)$ imaju istu funkciju raspodele [7]. Pretpostavimo da je distribucija u potpunosti poznata.

Shodno tome, kritične vrednosti A^2 za pet vrednosti značajnosti α , za gornji rep, jesu $A_{0.10}^2 = 1.933$, $A_{0.05}^2 = 2.492$, $A_{0.025}^2 = 3.070$, i $A_{0.01}^2 = 3.857$, respektivno [7].

Ove asimptotske vrednosti važe za aproksimativno $n > 10$. Mogu se koristiti za test bilo koje raspodele.

Koristeći (1), (6), i (12), vrednosti povratnog perioda od T godina za Gumbelovu, Frechet, i GEV raspodelu date su respektivno kao [7],

where $x(1), x(2), \dots, x(n)$ are the observations ordered in ascending order. Because the cdfs are in the range 0-1, their logarithms are negative and hence the summation on the righthand side of Eq. (24) is negative.

The absolute value of the summation is also greater than n, thus resulting in a positive value of A^2 . For large values of the test statistic A^2 , the null hypothesis that $F_n(x)$ and $F_0(x)$ have the same distribution is rejected [7]. Let us assume that the distribution is completely specified.

Accordingly, critical values A^2 at five α levels of significance, at the upper tail, are

$A_{0.10}^2 = 1.933$, $A_{0.05}^2 = 2.492$, $A_{0.025}^2 = 3.070$, and $A_{0.01}^2 = 3.857$, respectively [7].

These asymptotic values hold approximately for $n > 10$. They apply to tests on any type of distribution.

Using (1), (6), and (12), the return level with return period of T years for the Gumbel, Frechet, and GEV models are respectively given as [7],

$$x(T) = b - d \ln \left[\frac{T}{T-1} \right] \quad (25)$$

$$x(T) = x_0 \left(\frac{\ln T}{T-1} \right)^{\frac{1}{\beta}} \quad (26)$$

$$x(T) = \varepsilon + \frac{1}{k \left[1 - \left(-\frac{\ln(T-1)}{T} \right)^k \right]} \quad (27)$$

5 NUMERIČKI REZULTATI

Srednja vrednost, standardna devijacija, koeficijent varijacije i koeficijent asimetrije za maksimalnu godišnju vrednost protoka na reci Rasini, na mernoj stanici Ravni jesu 65.40 m³/s, 39.65 m³/s, 0.368, i 1.1629.

Gumbelov model: Iz izraza (4), vrednost

$$\hat{\alpha} = \frac{\sqrt{6}}{\pi} \hat{\sigma} = 0.780 \times 39.65 = 30.93 \text{ m}^3/\text{s}$$

ocenjen je parametar razmere Gumbelove distribucije metodom momenata. Iz izraza (5), parametar lokacije ocenjen je kao

5 NUMERICAL RESULTS

The mean, standard deviation, coefficient of variation, and coefficient of skewness for the annual maximum streamflow record at river Rasina, station Ravni are 65.40 m³/s, 39.65 m³/s, 0.368, and 1.1629.

Gumbel Model: From (4), the value

is the estimated scale parameter of the Gumbel distribution by the method of moments. From (5), the location parameter is estimated as

$$\hat{\beta} = \hat{\mu} - n_c \hat{\alpha} = 65.40 - 0.5772 \times 30.927 = \frac{47.55 \text{ m}^3}{s}$$

Nulta hipoteza H_0 jeste da maksimalni godišnji protok ima Gumbelovu raspodelu. Vrednost $\alpha = 0.05$ predstavlja nivo značajnosti koji je primenjen. Kritična oblast za K-S test je [7],

$$D_n > 0.254$$

Zapažena vrednost za D_n iz (21) je $d_n = 0.0932$ i ona je manja od kritične vrednosti 0.254, i zbog toga se nulta hipoteza **ne odbija**.

Izračunata vrednost A^2 iz izraza (24) za A-D test je 0.216, što je manje od kritične vrednosti 2.492 i nulta hipoteza takođe se **ne odbija**. Drugim rečima, Gumbelova raspodela obezbeđuje dobro slaganje s maksimalnim godišnjim protokom.

Frechet-ov model: Koristeći metod momenata, može se dobiti ocena parametra oblika Frechet raspodele. Iz izraza (10), imamo

$$\frac{\Gamma\left(1 - \frac{2}{\theta}\right)}{\Gamma^2\left(1 - \frac{1}{\theta}\right)} = 1 + C_v^2 = 1 + \left(\frac{39.65}{65.40}\right)^2 = 1.368$$

jednačina je rešena numerički da bismo dobili

$$\hat{\theta} = 3.19$$

Onda, iz jednačine (11) dobijamo ocenu za η kao

$$\hat{\eta} = \frac{\hat{\mu}}{\Gamma\left(1 - \frac{1}{\hat{\theta}}\right)} = \frac{65.40}{\Gamma\left(1 - \frac{1}{3.19}\right)} = 49.55$$

Da bismo testirali nultu hipotezu H_0 da maksimalni godišnji protok ima Frechet funkciju raspodele kao što je određeno ranije sa $\eta = 49.55$ i $\hat{\theta} = 3.19$, računamo vrednost D_n i dobili smo $d_n = 0.2208$ za K-S test. Vrednost d_n je manja od kritične vrednosti 0.254, i zbog toga se nulta hipoteza ne odbija. S druge strane, izračunata vrednost A^2 za A-D test je 9.504, što je više nego što je kritična vrednost 2.492, te se zbog toga nulta hipoteza **odbija**.

GEV model: Da bismo ocenili parametar oblika za GEV funkciju raspodele metodom momenata, koeficijent asimetrije 1.1629 zamenjen je u izraz (10) ili (11), koji je posle rešen po k, numeričkim putem. Dakle, ocena za k dobijena je kao

$$k = -0.00385$$

$$\alpha = \sqrt{\frac{k^2 \sigma^2}{[\Gamma(1 + 2k) - \Gamma^2(1 + k)]}} = \sqrt{\frac{(-0.00385)^2 39.65^2}{[\Gamma(1 - 2 \cdot 0.00385) - \Gamma^2(1 - 0.00385)]}} = 30.76$$

$$\varepsilon = \mu - \left(\frac{\alpha}{k}\right) [1 - \Gamma(1 + k)] = 65.40 - \frac{30.76}{-0.00385} [1 - \Gamma(1 - 0.00385)] = 47.53$$

The null hypothesis H_0 is that the annual maximum has a Gumbel distribution. The value of $\alpha = 0.05$ as the level of significance is adopted. The critical region for the K-S test is [7],

The observed value of D_n in (21) is $d_n = 0.0932$ which is less than the critical value 0.254, and therefore the null hypothesis is **not rejected**.

The calculated value of A^2 in (24) for A-D test is 0.216 which is less than the critical value of 2.492 and the null hypothesis is also **not rejected**. In other words, the Gumbel distribution provides a reasonable good fit for the annual maximum streamflow.

Frechet Model: Using the method of moments, the estimate of shape parameter of the Frechet distribution can be obtained. From (10), we have

and the equation is solved by numerical iterations to obtain

Then, from (11) we obtain the estimate of η as

To test the null hypothesis H_0 that the annual maximum streamflow has a Frechet distribution as specified earlier with $\eta = 49.55$ and $\hat{\theta} = 3.19$, we compute the observed value of D_n and obtain $d_n = 0.2208$ for K-S test. The value of d_n is less than the critical value 0.254, and therefore the null hypothesis is not rejected. On the other hand, the calculated value of A^2 for A-D test is 9.504 which is more than the critical value of 2.492 and the null hypothesis is **rejected**.

GEV Model: To estimate the shape parameter of the GEV distribution by the method of moments, the sample skewness of 1.1629 is substituted into (10) or (11), which is then solved for k by numerical iterations. Thus, the estimate of k is obtained as

Da bismo testirali nultu hipotezu H_0 da maksimalni godišnji protok ima GEV raspodelu kao što je specificirano ranije sa $\xi = 47.53$, $\alpha = 30.76$, i $k = -0.00385$, računamo posmatranu vrednost D_n i dobijeno je $d_n = 0.0921$ za K-S test. Vrednos $d_n = 0.0921$ manja je od kritične vrednosti 0.254 i zbog toga se nulta hipoteza **ne odbija**. S druge strane, uračunata vrednost za A^2 za A-D test je 0.2115, što je manje od kritične vrednosti 2.492 i zbog toga se nulta hipoteza **ne odbija**.

To test the null hypothesis H_0 that the annual maximum has a GEV distribution as specified earlier with $\xi = 47.53$, $\alpha = 30.76$, and $k = -0.00385$, we compute the observed value of D_n and obtain $d_n = 0.0921$ for K-S test. The value of $d_n = 0.0921$ is less than the critical value 0.254, and therefore the null hypothesis is **not rejected**. On the other hand, the calculated value of A^2 for A-D test is 0.2115 which is less than the critical value of 2.492 and the null hypothesis is **not rejected**.

Tabela 2. Povratni period za Gumbelovu i GEV funkciju raspodele
Table 2. Return period for Gumbel and GEV distribution

Povratni period Return period	Kvantil Quantile	Gumbelova funkcija raspodele Gumbel distribution	GEV funkcija raspodele GEV distribution
T	q	x(T)	x(T)
10,00	0,9000	117,1538614	117,0520306
20,00	0,9500	139,4181391	139,4175834
50,00	0,9800	168,2369627	168,4596899
100,00	0,9900	189,8326156	190,2910487
200,00	0,9950	211,3494696	212,1011913
1000,00	0,9990	261,1913993	262,847438

6 ZAKLJUČAK

Razmotrili smo tri modela maksimalnog godišnjeg protoka na reci Rasini: Gumbelov Model, Frechet model i GEV model.

Vrednosti za K-S test i A-D test u slučaju testiranja statističkih hipoteza za Frechet funkciju raspodele iznad su gornjih graničnih vrednosti i Frechet model je **odbijen**.

Gumbelov model i GEV model predstavljaju adekvatan model raspodele verovatnoće u ovoj praktičnoj situaciji.

Ovaj model korišćen je da se nađu projektne vrednosti koje bi predvidele povratne periode odgovarajućeg perioda. Na osnovu vrednosti za oba testa (K-S test i A-D test), GEV funkcija raspodele ima neznatno bolje slaganje sa empirijskim podacima za nivo značajnosti $\alpha=0,5$.

Analiza ekstremnih hidroloških događaja igra važnu ulogu u planiranju i analizi vodoprivrednih sistema, jer slučajne varijable koje opisuju poplave, olujne padavine, kao i niski protoci, predstavljaju merodavne projektne vrednosti u ovakvim inženjerskim projektima. Opšteprihvaćen je koncept da se životi i imovina moraju zaštititi od ekstremnih događaja. Primenjeni metod u ovom radu može se koristiti za analizu ekstremnih vrednosti različitih varijabli, kao što su, na primer, hidrološke, seizmičke, vulkanske, te vetar, nivo mora, saobraćajni protok na putevima, čvrstoće materijala i slično.

6 CONCLUSION

We consider three distributions to model the annual maximum streamflow at river Rasina: Gumbel Model, Frechet Model and GEV Model.

Values of K-S test and A-D goodness-of-fit test for the Frechet distribution are above the threshold values and Frechet model is **rejected**.

Gumbel model and GEV model are appropriate probability model in this practical situation.

We use this model to find required design values to predict the return level with certain return period. Based on the values of both tests (K-S and A-D test) GEV distribution has a slightly better fit with empirical data for level of significance $\alpha=0,5$.

Extreme value analysis of hydrological processes plays a relevant role in water resources planning and analysis, because random variables describing flood, storm rainfall and low flows are essential to predict design values in engineering projects. It is widely accepted that life and property need to be protected against the effects of extreme events. The applied method in this paper can be used for analyzing extreme values of different variables, such as for example hydrological, seismic, volcanic, wind, sea-level, traffic flow on highway, strength of material and other.

7 LITERATURA REFERENCES

- [1] A.F. Jenkinson, "Estimation of maximum floods," World Meteorological Organization, Technical Note, no. 98, ch. 5, pp. 183-257, 1969.
- [2] A. F. Jenkinson, "The frequency distribution of the annual maximum (or minimum) value of meteorological elements," O. J. Royal Meteorol. Soc., vol. 81, pp. 158-171, 1955.
- [3] E. J. Gumbel, Statistics of Extreme, New York: Columbia University Press, 1958.
- [4] M. Frechet, "Sur la loi de probabilit de l'ecart maximum," Annales de la Socit Polonaise de Mathmatique, vol. 6, pp. 93-117, 1927.
- [5] M. R. Rajabi and R. Modarres, "Extreme value frequency analysis of wind data from Isfahan, Iran," J. WindEng. Ind. Aerodyn., vol. 96, pp. 78-82, 2008.
- [6] N. I. Ramesh and A. C. Davison, "Local models for exploratory analysis of hydrological extremes," J. Hydrol., vol. 256, pp. 106-119, 2002.
- [7] N. T. Kottegoda and R. Rosso, Applied Statistics for Civil and Environmental Engineers, Blackwell Publishing Ltd. 2008, ch 7, pp 405-478.
- [8] N. V. Smirnov, "Tables for estimating the goodness of fit of empirical distributions," Annals of Mathematical Statistics, vol. 19, 279, 1948.
- [9] P. Chiou, W. Miao and T. C. Ho, The annual maximum wind speed at Pisa Airport in Italy, International Journal of Environmental Science and Development, Vol. 3, No. 3, June 2012.
- [10] T. W. Anderson and D. A. Darling, "A test of goodness of fit," J. Am. Stat. Assoc., vol. 49, pp. 765-769, 1954.

REZIME

STATISTIČKO MODELIRANJE EKSTREMNIH VREDNOSTI: PRIMENA NA PRORAČUN EKSTREMNIH PROTOKA NA RECI RASINI

Miljan KOVAČEVIĆ
Ljubo MARKOVIĆ
Lidija BABIĆ

Inženjeri su upoznati s neizvesnošću koja je povezana s funkcijama raspodele prilikom prirodnih fenomena ili onih u koji su uključeni ljudi. Čak i ako je poznata matična distribucija, u hipotetičkom slučaju, ostaje problem njene funkcionalne prezentacije. Moguće je izabrati adekvatan probabilistički model u praktičnoj situaciji, razmatrajući podatke koji su dostupni, da bi se opisao fenomen koji nas interesuje, te potom procenili parametri i odgovarajući rizik. Podaci za dvadeset sedam vrednosti godišnjih maksimalnih protoka na reci Rasini, na jugu centralne Srbije, od 1966. do 2008, analizirani su i modelirani. Modeli ekstremnih vrednosti za podatke razmatrani su i upoređivani. Nakon toga, dobijena je projektna vrednost sa odgovarajućim povratnim periodom.

Ključne reči: Anderson-Darling test, projektna vrednost, Frešet-ova raspodela, GEV distribucija, testiranje statističkih hipoteza, Gumbelova raspodela, Test Kolmogorov-Smirnova, povratni period

SUMMARY

STATISTICAL MODELLING OF EXTREME VALUES: APPLICATION TO CALCULATE EXTREME FLOW AT RIVER RASINA

Miljan KOVACEVIC
Ljubo MARKOVIC
Lidija BABIC

Engineers are aware of the uncertainties associated with the probability distributions of natural phenomena or those involving human beings. Even if one knows the parent distribution, in a hypothetical case, its functional representation remains a problem. One can select an appropriate probability model in a practical situation, considering the data available, to describe the phenomenon of interest, and then estimate the parameters and assess the risks involved. The data of 27 annual maximum streamflows at river Rasina in south central Serbia from 1966 to 2008 were analyzed and modeled. Three extreme value models for the data were considered and compared. Subsequently, the required design value with a given return period of exceedance was obtained.

Key words: Anderson-Darling test, design value, Frechet distribution, general extreme value distribution, goodness-of-fit, Gumbel distribution, Kolmogorov-Smirnov test, return period