

FREQUENCY ANALYSIS OF EXTREME RIVER FLOWS: SELECTED METHODS AND THEIR APPLICATION

Analiza częstości ekstremalnych przepływów rzecznych: wybrane metody i ich zastosowanie

Sisay E. Debele

Streszczenie

Celem pracy jest rozwiązanie praktycznych problemów związanych z analizą wartości ekstremalnych przepływu w warunkach stacjonarnych i niestacjonarnych. Problemy te dotyczą:

- Metod wykrywania zależności i trendów w ciągach chronologicznych przepływów ekstremalnych oraz oceny ich istotności.
- Oceny wpływu zależności ekstremów sezonowych na oszacowanie kwantyli projektowych rozkładu przepływów rocznych w modelu alternatywy zdarzeń w przypadku stacjonarnego reżimu hydrologicznego.
- Porównania i oceny metod niestacjonarnej analizy częstości wartości ekstremalnych.
- Zalet i ograniczeń stosowania pakietu Generalized Additive Model for Location, Scale and Shape parameters (GAMLSS) w zagadnieniach hydrologicznych.
- Wykorzystania niejednorodnego cechowanego procesu Poissona w niestacjonarnej analizie POT (peak over threshold).

Praca składa się z dwóch rozdziałów wstępnych, gdzie przedstawiono wyżej wymienione problemy w kontekście potrzeby analizy wielowymiarowej i niestacjonarnej, przegląd rozkładów używanych w analizie statystycznej wartości ekstremalnych wraz z metodami ich estymacji, oraz dalszych czterech rozdziałów podzielonych na dwie części dotyczące metod analizy stacjonarnej i niestacjonarnej.

Opis metod omawianych w pracy jest przedstawiony i ilustrowany przykładami dotyczącymi przepływów maksymalnych. Opisanie metod mogą być również wykorzystane do przepływów minimalnych Q_{min} np. przez transformację tych ostatnich do postaci $1/Q_{min}$ (dla cieków stałych, gdy $Q_{min} > 0$) lub $-Q_{min}$. Wówczas minimalne wartości przepływu stają się maksymalnymi i mogą być analizowane opisanymi metodami, a wyznaczone rozkłady, przez transformację odwrotną do zastosowanej, wykorzystane do wyznaczenia kwantyli projektowych.

Należy jednak podkreślić, że zgodnie z założeniami projektu CHIHE (Climate Change Impact on Hydrological Extremes), w którego ramach wykonana była ta praca, wszystkie obliczenia zostały przeprowadzone na podstawie maksymalnych przepływów dobowych.

W części poświęconej analizie stacjonarnej rozwiązano problem oceny wpływu zależności ekstremów sezonowych na oszacowanie kwantyli projektowych rozkładu ekstremów przepływów rocznych w modelu alternatywy zdarzeń w przypadku stacjonarnego reżimu hydrologicznego. W podejściu sezonowym wykorzystującym maksima zimowe i letnie do modelowania ekstremów rocznych zwykle zakłada się, że ekstrema sezonowe są niezależne, przez co możliwe jest zastosowanie modelu produktowego w postaci:

$$F(x) = P(X \leq x) = P(W \leq x) \cdot P(S \leq x) = F_W(x) \cdot F_S(x), \quad (1)$$

gdzie $F(x)$ jest dystrybuantą rozkładu maksimów rocznych, $F_W(x)$ and $F_S(x)$ są dystrybuantami rozkładów maksymalnych przepływów sezonowych zimowych (W) i letnich (S).

Niezależność ekstremów sezonowych jest kwestionowana przez praktyków wskazujących, że częstą sytuacją w analizie hydrogramów rocznych jest występowanie maksimum letniego na opadającej gałęzi wezbrania zimowego, co stanowi przesłankę do istnienia dodatkowej zależności maksimów zimowych i letnich. Zależność ta została przeanalizowana na podstawie danych ze stacji hydrologicznych w dorzeczu Sanu, gdzie w 4 na 20 analizowanych par serii obserwacyjnych stwierdzono słabą dodatnią, ale istotną zależność między maksimami sezonowymi zimy i lata. Podobny wynik obserwujemy w innych regionach Polski. W takim przypadku model produktowy (1) powinien być zastąpiony rozkładem dwuwymiarowym z zależnymi rozkładami brzegowymi np. w postaci funkcji łączącej – kopula (C), które zastosowano w tej pracy. Wówczas równanie (1) przybiera postać:

$$F(x) = C(F_W(x), F_S(x)) \quad (2)$$

Na podstawie nierówności Fréchet'a-Hoeffding'a udowodniono, że górne kwantyle rozkładu maksimów rocznych otrzymane przy założeniu niezależności (1) są większe niż w przypadku uwzględnienia dodatkowej zależności maksimów sezonowych (2). Zatem stosowanie modelu produktowego prowadzi do przeszacowania kwantyli projektowych, czyli, z punktu widzenia projektowania budowli jest bezpieczne. Idea przeszacowania przepływów projektowych przez klasyczny model produktowy (1) może być atrakcyjna dla projektantów, ponieważ pozwala na zredukowanie do pewnego stopnia kosztów planowanych inwestycji, jeśli zastosować model z zależnością. Jednak słaba zależność maksimów sezonowych i małe liczebności serii obserwacyjnych oraz niejednoznaczność związana z wyborem odpowiedniej funkcji kopula sprawiają, że wyniki estymacji są dużym stopniu niepewne, a zatem tradycyjne podejście w postaci modelu produktowego, nawet w przypadku silnej zależności maksimów sezonowych, jest rekomendowane. W przypadku dominacji jednego typu wezbrań, gdy np. największe wezbrania w serii maksimów rocznych są pochodzenia roztopowego, a wezbrania deszczowe są z reguły mniejsze, nawet bardzo silna dodatnia zależność między przepływami nie ma wpływu na ocenę górnych kwantyli rocznych, które są bardzo bliskie kwantylom sezonu dominującego, w tym przypadku kwantylom przepływów maksymalnych pochodzenia roztopowego.

Studium przypadku dotyczyło 6 posterunków spośród analizowanych 20 w dorzeczu Sanu, gdzie zależność między maksimami sezonowymi była istotna lub prawie istotna na poziomie 5%. Przeprowadzona analiza potwierdziła wnioski przedstawione powyżej.

Wyniki badań zostały opublikowane w:

The impact of seasonal flood peak dependence on annual maxima design quantiles

S. E. Debele, E. Bogdanowicz and W. G. Strupczewski

Hydrological Sciences Journal – Journal Des Sciences Hydrologiques, 2017

<https://doi.org/10.1080/02626667.2017.1328558>

W części poświęconej analizie niestacjonarnej porównano trzy metody estymacji w niestacjonarnej analizie częstości powodzi. Były to

- Metoda największej wiarygodności (MNW)
- Metoda dwuetapowa (TS)
- Metody Generalized Additive Model for Location, Scale and Shape parameters (GAMLSS)

Metoda MNW jest powszechnie znaną i stosowaną metodą w praktyce hydrologicznej. **Dwuetapowa metoda estymacji TS**, jak sama nazwa wskazuje, odbywa w dwóch etapach. Pierwszy etap polega na zastosowaniu jednoczesnej estymacji zależnych od czasu wartości średniej i odchylenia standardowego metodą ważonych najmniejszych kwadratów (ang. WLS) przez rozwiązanie układu równań, co w przypadku założenia trendów liniowych prowadzi do układu 4 równań:

$$\begin{cases} \sum_{t=1}^T \frac{1}{s_t^2} (y_t - m_t) \cdot [t,1] = 0 \\ \sum_{t=1}^T \frac{1}{s_t^3} \left\{ (y_t - m_t)^2 - s_t^2 \right\} \cdot [t,1] = 0 \end{cases} \quad (3)$$

gdzie y_t , m_t i s_t oznaczają odpowiednio: obserwowane wartości przepływów maksymalnych w kolejnych latach okresu obserwacji $t=1, \dots, T$; wartość średnią oraz odchylenie standardowe w roku t , zaś $[t,1]$ jest macierzą Jacobi'ego funkcji trendu względem jej parametrów. W drugim etapie metody dokonuje się transformacji obserwowanej zmiennej losowej do postaci standaryzowanej:

$$z_t = \frac{y_t - m_t}{s_t} \quad (4)$$

gdzie z_t są wartościami stacjonarnej zmiennej losowej o średniej 0 i odchyleniu standardowym 1. Na podstawie standaryzowanych wartości z_t estymowany jest parametr kształtu rozkładu i wyznaczone kwantyle rozkładu. Re-impozycja trendu na wyznaczone wartości kwantyli $\hat{z}_t(p)$ jest dokonywana przez:

$$\hat{y}_t(p) = m_t + s_t \cdot \hat{z}_t(p) \quad (5)$$

W **metodzie GAMLSS** zakłada się, że niezależne obserwacje y_i , dla $i = 1, \dots, n$, mają rozkład prawdopodobieństwa o gęstości $f_Y(y_i | \theta^i)$ z $\theta^i = (\theta_1^i, \dots, \theta_p^i)$ wektorem parametrów o długości p ($p \leq 4$) odpowiadających parametrom położenia, skali i kształtu zmiennej Y . W metodzie GAMLSS parametry rozkładu są powiązane ze zmiennymi objaśniającymi przez monotoniczne funkcje łączące $g_k(\cdot)$. GAMLSS zawiera szereg możliwych funkcji łączących; w tej pracy użyto parametrycznej wersji w postaci:

$$g_k(\theta_k) = \mathbf{X}_k \boldsymbol{\beta}_k \quad (6)$$

gdzie θ_k są wektorami o długości n , \mathbf{X}_k jest macierzą zmiennych objaśniających rzędu $n \times m$, $\boldsymbol{\beta}_k$ jest wektorem parametrów o długości m .

Powszechną praktyką w badaniach hydrologicznych jest traktowanie serii obserwacyjnych przepływów ekstremalnych, jako niestacjonarnych i stosowanie narzędzi statystycznych do wykrycia tej niestacjonarności. Jednak stacjonarność i niestacjonarność są własnościami procesu, a nie serii obserwacyjnej. Zgodnie z Koutsoyiannis'em deterministyczna funkcja reprezentująca trend może być określona wyłącznie na drodze dedukcji, niezależnie od danych (tzn. *a priori*, na podstawie modelu, który mógłby je przewidzieć). Zwykle, w przeciwieństwie do powyższego stwierdzenia, wnioskowanie o "tendencjach" i "przesunięciach" w średnich jest prowadzone przez indukcję na podstawie danych (*a posteriori*), stąd wyznaczone funkcje trendu nie są deterministyczne i nie reprezentują niestacjonarności procesu leżącego u ich podstaw.

Prowadząc grę z Naturą o możliwość przewidywania przyszłości, możemy jedynie monitorować dynamikę naszych serii czasowych (co może być użyteczne w odniesieniu do najbliższej przyszłości) oraz doskonalić modele klimatyczne i hydrologiczne, mając nadzieję, że dadzą nam najlepszą ocenę przyszłych ekstremów, jaką jesteśmy w stanie osiągnąć w dzisiejszych czasach.

Pozostając w przeświadczeniu, że analiza trendów w kwantylach projektowych maksymalnych przepływów sezonowych i rocznych może być przydatną przesłanką do podejmowania odpowiedzialnych decyzji w projektowaniu urządzeń narażonych na wpływ wielkich wód lub zabezpieczających przed tym wpływem można stosować dwa sposoby zakładając:

1. Trendy w parametrach rozkładu prawdopodobieństwa.
2. Trendy w momentach rozkładu (średniej i odchyleniu standardowym).

Z uwagi na ograniczoną długość hydrologicznych serii pomiarowych abstrahuje się tutaj od zmian typu rozkładu.

To drugie z wymienionych podejść, rekomendowane przez Strupczewskiego i Kaczmarka, Kochanka oraz Strupczewskiego i in., jest zgodne z klasyczną analizą klimatologiczną polegającą na analizach czasowej zmienności średniej i odchylenia standardowego, oraz pozwala na porównanie wyników uzyskanych z różnych modeli. To podejście zostało zastosowane w pracy.

Parametr kształtu, w związku z wysoką niepewnością estymacji (nawet w przypadku stacjonarnym), jest przyjmowany jako stały we wszystkich wykorzystywanych modelach.

Zgodnie z zasadą brzytwy Ockhama przyjęte postaci trendu w średniej i odchyleniu standardowym powinny być jak najprostsze, co oznacza zastosowanie liniowej funkcji trendu. Przeprowadzona analiza innych prostych funkcji trendu (trend logarytmiczny lub wykładniczy) potwierdziła zasadność użycia trendu liniowego.

W pracy zastosowano podejście sezonowe, które jest zalecane z uwagi na większą podatność sezonowych procesów formujących wezbrania, a co za tym idzie przepływów maksymalnych, na zmienność i zmiany klimatu. Ważne cechy dynamiki zmian sezonowych mogą być maskowane, gdy analizuje się tylko maksima roczne. Do opisu częstości maksimów sezonowych wykorzystano 5 rozkładów, powszechnie stosowanych w hydrologii w analizach wartości maksymalnych (do estymacji metodą MNW i TS) oraz 5 rozkładów z rodziny rozkładów GAMLSS, które tylko w 2 przypadkach pokrywają się z tymi powszechnie używanym w hydrologii. Jeden z nich (Pearson 3) został dołączony w ramach tej pracy do algorytmów GAMLSS przy wykorzystaniu narzędzi oprogramowania pakietu R. Drugi rozkład

to rozkład Weibulla figurujący w rodzinie rozkładów GAMLSS w zreparametryzowanej postaci pod nazwą RGE (Reverse Generalized Extreme). Dla każdego rozważanego posterunku wodowskazowego przeanalizowano cztery opcje trendu w ocenie dopasowania do danych sezonowych:

- Opcja 0 – (stacjonarność) oznaczająca brak trendu w średniej i odchyleniu standardowym
- Opcja 1 – trend tylko w wartości średniej
- Opcja 2 – trend tylko w odchyleniu standardowym
- Opcja 3 – trend w średniej i odchyleniu standardowym

Następnie zakładając niezależność maksimów w obu sezonach obliczono kwantyle rozkładu maksimów rocznych w każdym roku analizowanego okresu według wzoru (1). Ponadto, w celu złagodzenia wpływu wyboru rozkładu zastosowano podejście wielomodelowe (MMA). Przebieg zmienności kwantyla rozkładu maksimów rocznych zależy od przebiegu kwantyli sezonowych. Gdy linie trendów kwantyli sezonowych nie przecinają się, wówczas trend w kwantylu rocznym prawie dokładnie pokrywa się z trendem większego kwantyla sezonowego. Gdy linie trendów przecinają się, trend w kwantylu rocznym jest krzywoliniowy wpisany w kąt utworzony przez przecinające się linie trendów sezonowych i położony powyżej obu linii trendów sezonowych.

Przeprowadzone obliczenia dla 10 zlewni reperowych uwzględnionych w projekcie CHIHE oraz badania symulacyjne Monte Carlo wykazały, że:

- Wyniki estymacji przeprowadzonej metodą MNW i GAMLSS są silnie zależne od przyjętych rozkładów, w związku z czym wybór rozkładu ma zasadnicze znaczenie w estymacji trendów w kwantylach projektowych maksimów sezonowych i rocznych.
- Metoda TS jest najmniej wrażliwa na wybór rozkładu, a wyniki estymacji kwantyla 99% są najbardziej ze sobą zgodne. Metody MNW i GAMLSS dają z reguły większe wartości kwantyli sezonowych i rocznych.
- Ocena dewiacji (deviance statistics) potwierdza, że metoda TS częściej wykrywa trend niż MNW i GAMLSS.
- Estymowane momenty uzyskane porównywanymi metodami są bardziej zgodne w sezonie zimowym niż letnim.
- Zależność wyników estymacji kwantyli od wyboru typu rozkładu (związane z tzw. błędem wyboru modelu) można złagodzić stosując podejście wielomodelowe polegające na uśrednieniu kwantyli tego samego rzędu przy użyciu średniej ważonej z wagami wyznaczonymi na podstawie wartości kryteriów Akaike (AIC).
- Symulacje Monte Carlo wykazały, że zarówno przy prawdziwej jak i fałszywej hipotezie dotyczącej przyjętego rozkładu metoda TS daje mniejszy względny średni błąd standardowy (RRMSE) kwantyla 99% niż MNW, natomiast względne obciążenie (RB) kwantyla jest mniejsze w przypadku MNW. Metoda GAMLSS dała najlepsze rezultaty w wychwyceniu trendu w odchyleniu standardowym, estymacji parametru kształtu oraz estymacji kwantyla 99% (zwłaszcza dla długich serii czasowych) zarówno pod względem RRMSE jak i RB. Metoda TS jest lepsza w wychwytywaniu trendu w

wartości średniej oraz w estymacji kwantyla 99% w krótkich seriach czasowych i dlatego rekomendujemy ją w tym przypadku.

Ogólnie należy stwierdzić, że uwzględnienie zmian klimatu w niestacjonarnej analizie częstości powodzi stwarza wiele problemów natury teoretycznej i praktycznej. Ponieważ celem nie jest tutaj jedynie opis zmian reżimu wielkich wód w okresie objętym obserwacjami, ale ekstrapolacja wyników na długoterminowe horyzonty planowania, pojawiają się problemy oceny wiarygodności i niepewności stosowanych metod a wraz z nimi również pytanie o zasadność uwzględniania trendów statystycznie istotnych, ale nieistotnych z fizycznego (hydrologicznego) punktu widzenia.

W pracy porównano kwantyle uzyskane na podstawie projekcji przepływów maksymalnych otrzymanych z połączenia 7 modeli klimatycznych i modelu odpływu w okresie 1971-2100.

Część wyników została opublikowana w:

1. **A comparison of three approaches to non-stationary flood frequency analysis**
S. E. Debele, W. G. Strupczewski, E. Bogdanowicz
Acta Geophys. (2017) 65(4): 863–883, Special Issue: Impacts of climate and land use change on hydrological processes
<https://doi.org/10.1007/s11600-017-0071-4>
2. **Around and about an application of the GAMLSS package to non-stationary flood frequency analysis**
S. E. Debele, E. Bogdanowicz, W. G. Strupczewski
Acta Geophys. (2017) 65(4): 885–892, Special Issue: Impacts of climate and land use change on hydrological processes
<https://doi.org/10.1007/s11600-017-0072-3>

Trzecią częścią analizy niestacjonarnej jest wykorzystanie niejednorodnego cechowanego procesu Poissona w niestacjonarnej analizie POT (peak over threshold). W tej części zostały przedstawione teoretyczne podstawy stacjonarnej i niestacjonarnej analizy POT, metody wyboru progu, estymacja intensywności procesu Poissona w wersji stacjonarnej i niestacjonarnej oraz niestacjonarna estymacja cech – przepływów maksymalnych. Podejście POT do analizy wartości maksymalnych jest często preferowane w stosunku do analizy maksymalnych przepływów rocznych i sezonowych z uwagi na wykorzystanie pełniejszej informacji o wielkich wodach. Dodatkową zaletą jest możliwość oddzielnej analizy trendów w częstości występowania i w wielkości przepływów maksymalnych, czego nie da się rozdzielić w analizie niestacjonarnych rozkładów maksymalnych przepływów rocznych/sezonowych. Przy wyborze największych przepływów kulminacyjnych istotną rolę pełni odpowiedni dobór progu, od którego zależy wynik analizy.

Publikacja w przygotowaniu.