

Krzysztof KOCHANEK, Wojciech FELUCH

Politechnika Warszawska – filia w Płocku

ZASTOSOWANIE METODY UOGÓLNIONYCH MOMENTÓW DO ESTYMACJI KWANTYLI PRZEPŁYWÓW MAKSYMALNYCH WYBRANYCH ROZKŁADÓW O GRUBYCH OGONACH

THE ESTIMATION OF FLOOD QUANTILES OF THE SELECTED HEAVY-TAILED DISTRIBUTIONS BY MEANS OF THE METHOD OF GENERALISED MOMENTS

Wprowadzenie

W inżynierii wodnej, strategii przeciwpowodziowej i w gospodarce wodnej istnieje potrzeba określenia przepływu maksymalnego fali powodziowej dla określonego prawdopodobieństwa przewyższenia P , tj. obliczenia kwantyli wysokich rzędów ($x(P)$) na podstawie serii pomiarowej rocznych maksymalnych przepływów w rzece. Wnioskowanie statystyczne dotyczące górnych kwantyli generuje problemy zarówno natury statystycznej, jak i hydrologicznej, których aspekty rzadko traktowane są łącznie. Statystyczna niepewność oszacowań górnych kwantyli wynika m.in. z błędu próbkowania, krótkich szeregów czasowych, ciągów pomiarowych obarczonych błędami, nieznaną funkcji rozkładu prawdopodobieństwa przepływów rocznych maksymalnych (WQ lub $Q_{\max}$), założeń upraszczających, poczynając od niezależnych od siebie i równomiernie rozłożonych elementów w próbie i, w szczególności, bardzo problematycznych założeń co do stacjonarności długiej serii danych i przyszłego zachowania się procesu formowania fali powodziowej.

W analizie częstości występowania powodzi funkcja gęstości prawdopodobieństwa (PDF) jest wybierana mniej lub bardziej subiektywnie spośród asymetrycznych

dwu- lub trzyparametrowych rozkładów zmiennej losowej typu ciągłego. Niektóre z tych rozkładów wprowadzano do hydrologii inżynierskiej ze względu na ich przydatność do modelowania różnorodnych kształtów hydrogramu (Cunnane, 1989). Ponieważ argumenty teoretyczne optujące za wyborem danego typu funkcji mogą być łatwo podważone, znacznie większą rolę praktyczną odgrywa empiryczna przydatność danego modelu, niż argumentacja *a priori*. Jest oczywiste, że konsekwencje wyboru modelu i jego zastosowania w praktyce do oceny częstości występowania powodzi ujawniają się wyraźnie w górnej części ogona rozkładu prawdopodobieństwa. Dlatego też współcześnie wśród badaczy ugruntował się pogląd, że hydrologiczne zjawiska ekstremalne podlegają rozkładom prawdopodobieństwa o grubych ogonach (np. Gumbel, 1958; Kendall, Stuart, 1969; Coles, 2001; Katz i in., 2002). Warto także wspomnieć, że w praktyce nieznany jest rzeczywisty rozkład populacji przepływów maksymalnych rocznych, a co za tym idzie, każdy rozkład dobrany do serii czasowej historycznych pomiarów podczas estymacji parametrów i kwantyli generuje mniejszy lub większy błąd (tzw. błąd modelu).

Równie ważnym etapem analizy częstości powodzi jak dobór modelu jest zastosowanie odpowiedniej metody estymacji parametrów rozkładu prawdopodobieństwa. Dobiera się metodę minimalizującą błędy estymacji kwantyli o niskich prawdopodobieństwach przewyższenia, $P \leq 0,1$. Obecnie do najczęściej stosowanych w praktyce inżynierskiej metod estymacji parametrów rozkładów statystycznych należą metoda największej wiarygodności (MLM) (np. Kaczmarek, 1970; Rao, Hamed, 2000), metoda momentów (MOM) (Rao, Hamed, 2000) oraz stosunkowo nowa metoda momentów liniowych (LMM) (Greenwood i in., 1979; Hosking, 1986; Hosking, Wallis, 1997; Rao, Hamed, 2000). Ciągłe jednak wysiłki naukowców skupiają się na poszukiwaniu innych metod estymacji parametrów modeli lub doskonaleniu metod już istniejących, w tym dobrze znanej, łatwej w użyciu i akceptowanej przez praktyków metody momentów (np. Ma, 1984 i Liang i in., 2014). Jednakże modele o grubych ogonach mają ograniczenia co do istnienia momentów statystycznych, przy czym prawdopodobieństwo istnienia teoretycznego momentu statystycznego dla takiego rozkładu maleje wraz ze wzrostem rzędu momentu. Ponadto metoda momentów charakteryzuje się dużymi błędami estymacji oraz małą odpornością na elementy odstające w próbie, przede wszystkim ze względu na rosnące wraz ze rzędem momentu błędy systematyczne (ang. *bias*, np. Wallis i in., 1974; Strupczewski i in., 2002 a, b, 2005b, 2007; Kochanek i in., 2005).

Aby ominąć problemy wynikające z właściwości klasycznej metody momentów, w niniejszej pracy zaproponowaliśmy zastosowanie uogólnionej (generalizowanej) metody momentów (np. Ashkar, Bobee, 1987; Ashkar, Mahdi, 2003, 2006), polegającej na odejściu od założeń, że momenty powinny być funkcjami parametrów podniesionych do potęgi naturalnej.

Rozkłady prawdopodobieństwa

W analizie częstości występowania powodzi najczęściej stosuje się modele o grubych ogonach, do których zalicza się np. trzyparametrowe rozkłady: logarytmiczno-normalny, generalizowany wartości ekstremalnych (ang. *Generalised Extreme Value* – GEV), generalizowany logarytmiczno-logistyczny (ang. *Generalised Log-Logistic*), Pareto, Weibulla, Wakeby (opisywany aż 5 parametrami). Jednakże tradycyjnie w hydrologii statystycznej rozkłady opisujące zjawiska ekstremalne najczęściej są definiowane jedynie przez 3 parametry: parametr lokacji, skali i kształtu. Dodatkowo charakter wielu hydrologicznych zjawisk ekstremalnych (np. brak ujemnych wartości w próbie) oraz fakt, że tematem zainteresowania w teorii zjawisk ekstremalnych i częstości występowania powodzi są kwantyle o małych prawdopodobieństwach przewyższenia (lokujące się daleko od początku układu współrzędnych), umożliwia rezygnację z parametru lokacji poprzez ustalenie początku lewego ogona funkcji rozkładu prawdopodobieństwa w początku układu współrzędnych. Tym samym upraszcza się analiza i uwiarygadnia się estymacja parametrów i kwantyli dokonywana dla krótkich serii czasowych. Ponadto takie działanie uściśla wywód teoretyczny i obliczenia, przy jednoczesnym zachowaniu ogólności opisu zjawiska. Dwuparametryczne rozkłady o grubych ogonach, w ogólności charakteryzowane parametrami kształtu i skali, stanowią obecnie podstawę warsztatu hydrologów zajmujących się badaniem zjawisk ekstremalnych (Rowiński i in., 2002; Węglarczyk i in., 2002; Strupczewski i in., 2005a).

W niniejszej pracy wykorzystano 3 dwuparametrowe rozkłady prawdopodobieństwa o grubych ogonach:

1. Log-Gumbela (LG) – dwuparametrowa realizacja rozkładu generalizowanego wartości ekstremalnych (ang. *generalised extreme value*, GEV – Johnson i in., 1995);
2. Log-Logistyczny (LL) – dwuparametrowa realizacja rozkładu generalizowanego logarytmiczno-logistycznego (ang. *generalised log-logistic*, GLL – Johnson i in., 1995);
3. Weibulla (We) – dwuparametrowa realizacja rozkładu Weibulla (Johnson i in., 1994).

Wszystkie zastosowane rozkłady należą do rodziny rozkładów wartości ekstremalnych i charakteryzują się grubym prawostronnym „ogonem”, co oznacza, że przy tym samym rzędzie górnego kwantyla (prawdopodobieństwie przewyższenia), np. $P \leq 0,2$, generują one dużo większe wartości kwantyli niż inne modele.

Rozkład Log-Gumbela

Gęstość rozkładu prawdopodobieństwa Log-Gumbela (LG) wyraża się wzorem:

$$f(x) = \frac{\xi}{\alpha} x^{-1/\alpha-1} \exp\{-\xi \cdot x^{-1/\alpha}\} \quad (1)$$

funkcja kwantyla

$$x(P) = \left(-\frac{1}{\xi} \log(1-P) \right)^{-\alpha} \quad (2)$$

gdzie: $\xi > 0$ – parametr skali; $\alpha > 0$ – parametr kształtu; P – prawdopodobieństwo przewyższenia.

Rozkład Log-logistyczny

Gęstość rozkładu prawdopodobieństwa Log-logistycznego (LL) wyraża się wzorem

$$f(x) = \frac{1}{\kappa x} \left(\frac{x}{\alpha} \right)^{-\frac{1}{\kappa}} \left(1 + \left(\frac{x}{\alpha} \right)^{-\frac{1}{\kappa}} \right)^{-2} \quad (3)$$

natomiast funkcja kwantyla

$$x(P) = \alpha \left(\frac{1}{(1-P)} - 1 \right)^{-\kappa} \quad (4)$$

gdzie: $\alpha > 0$ – parametr skali; $\kappa > 0$ – parametr kształtu.

Rozkład Weibulla

Gęstość rozkładu prawdopodobieństwa Weibulla (We) wyraża się wzorem

$$f(x) = \frac{b}{a} \left(\frac{x}{a} \right)^{b-1} \exp \left[-\left(\frac{x}{a} \right)^b \right] \quad (5)$$

natomiast funkcja kwantyla

$$x(P) = a \cdot \left[-\ln(1-P) \right]^{\frac{1}{b}} \quad (6)$$

gdzie: $a > 0$ – parametr skali; $b > 0$ – parametr kształtu

Dokładny opis właściwości oraz zakresu stosowalności powyższych rozkładów prawdopodobieństwa wybiega poza ramy niniejszego artykułu. Szczegółową analizę każdego z wymienionych modeli statystycznych wraz z formułami, metodami estymacji parametrów, metodami obliczania błędów estymacji czytelnik łatwo znajdzie w literaturze hydrologicznej, np. w Rao, Hamel, 2000.

Najczęściej stosowane metody estymacji parametrów

W praktyce inżynierskiej do estymacji parametrów rozkładów statystycznych obecnie najczęściej stosuje się metodę największej wiarygodności (MLM), metodę momentów (MOM) oraz metodę momentów liniowych (LMM). Wyniki uzyskane za pomocą metody uogólnionych (generalizowanych) momentów (MGM) porównano

do wyników uzyskanych za pomocą MLM, MOM i LMM i jakkolwiek trzy ostatnie metody są znane i powszechne w hydrologii, zasługują one także na krótki komentarz i wskazanie ich zalet i wad.

Metoda największej wiarygodności

W metodzie tej zakłada się, że skoro wartości próby losowej ($x_1, x_2, \dots, x_n$) zostały zaobserwowane, to należy im przypisać wysokie prawdopodobieństwo. Podstawą metody jest funkcja wiarygodności zbudowana na podstawie wartości funkcji gęstości prawdopodobieństwa dla poszczególnych elementów próby losowej. Estymacja parametrów polega na takim dobraniu zestawu parametrów rozkładu, aby funkcja wiarygodności osiągnęła maksimum.

Zalety estymatorów MLM wynikają głównie z teoretycznych własności metody MLM. Estymatory MLM:

- są zgodne, $\forall \lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$, gdzie θ i $\hat{\theta}$ są odpowiednio wektorami parametrów i estymatorów parametrów, a n jest liczebnością próby,
- są co najmniej asymptotycznie nieobciążone, $\lim_{n \rightarrow \infty} P(E(\hat{\theta}) - \theta) = 0$,
- są asymptotycznie najefektywniejsze (mała wariancja),
- można je stosować do prób zależnych (nieprostych),
- metoda jest odporna na elementy odstające (ang. *outliers*).

Wady metody są natomiast głównie praktyczne:

- generuje duży błąd systematyczny, gdy „nie trafimy” we właściwy rozkład, szczególnie gdy próba jest mała,
- koncentruje się na głównej masie prawdopodobieństwa – daje wyniki obciążone dużym błędem w przypadku estymacji kwantyli o małym prawdopodobieństwie przewyższenia,
- dla wieloparametrycznych rozkładów prawdopodobieństwa (nawet gdy liczba parametrów > 2) czasem trudno uzyskać wiarygodne rozwiązanie numeryczne w skończonym czasie.

Metoda momentów

Metoda momentów opiera się na założeniu, że moment teoretyczny jest równy swojemu odpowiednikowi empirycznemu (policzonemu na podstawie próby losowej). Na przykład dla rozkładu o dwóch parametrach:

$$\begin{cases} E(X; g_1, g_2) = \bar{x} \\ \text{var}(X; g_1, g_2) = s_x^2 \end{cases} \quad (7)$$

gdzie: $E(.)$ – wartość oczekiwana; $\text{var}(.)$ – wariancja.

Wyrażenia po lewej stronie dotyczą teoretycznego rozkładu prawdopodobieństwa o parametrach (g_1, g_2) , prawa strona układu wyznaczana jest na podstawie próby losowej. Rozwiązaniem układu równań (7) są funkcje estymujące wartości parametrów rozkładu (tzw. Estymatory Metody Momentów). Pogłębioną analizę stosowalności metody momentów w hydrologii przedstawili w swojej fenomenalnej pracy Wallis i in. (1974).

Zalety estymatorów uzyskanych za pomocą metody MOM:

- charakteryzują się stosunkowo małym błędem systematycznym, który jednak **rośnie z liczbą estymowanych parametrów (rzędem momentu)**,
- mają mały średni błąd kwadratowy,
- metoda jest prosta w stosowaniu.

Wady metody:

- brak skończonych momentów dla niektórych rozkładów, w szczególności rozkładów o grubych ogonach,
- mała odporność na elementy odstające.

Metoda momentów liniowych

Jest to metoda stosunkowo nowa, gdyż została zaadoptowana w hydrologii pod koniec lat 1970. Stanowi modyfikację metody momentów probabilistycznych ważonych (PWM) (Greenwood i in., 1979; Hosking i in., 1986; Rao, Hamed, 2000). Metoda momentów liniowych (Hosking, 1990; Hosking, Wallis, 1997) jest szeroko stosowana w teorii zjawisk ekstremalnych z powodu swojej prostoty i satysfakcjonujących wyników dla krótkich prób losowych. W przypadku rozkładów dwuparametrowych układ równań przyjmuje postać

$$\begin{cases} \lambda_1(X; g_1, g_2) = l_1 \\ \lambda_2(X; g_1, g_2) = l_2 \text{ lub } \lambda(X; g_1, g_2) = t \end{cases} \quad (8)$$

gdzie: λ_1, l_1 – pierwszy moment liniowy i jego empiryczna realizacja z próby, λ_2, l_2 – drugi moment liniowy i jego empiryczna realizacja z próby, τ, t – współczynnik liniowy zmienności i jego empiryczna realizacja z próby.

Rozwiązaniem powyższego układu równań są funkcje estymujące wartości parametrów rozkładu (tzw. Estymatory Metody Momentów Liniowych). Zalety metody momentów liniowych:

- reprezentuje poziom jakości wyników zbliżony do MLM,
- charakteryzuje się prostotą rozwiązań,
- jeśli istnieje skończona wartość średnia, to pozostałe momenty liniowe też są skończone,
- wykazuje dużą niezawodność,
- oraz dużą odporność na elementy odstające.

Wady metody LMM:

- dystrybuenta rozkładu, $F(x)$, lub funkcja kwantyla, $x(F)$, muszą występować w analitycznej formie, co nie zawsze jest możliwe,
- konieczność stosowania próby w postaci szeregu rozdzielczego, co unieumożliwia zastosowanie w estymacji zjawisk niestacjonarnych.

Metoda momentów uogólnionych (generalizowanych)

Moment statystyczny w ogólnej postaci wyraża się wzorem

$$M_{prs} = E \left[x^p \cdot F^r \cdot (1-F)^s \right] = \int_0^1 [x(F)]^p \cdot F^r \cdot (1-F)^s dF \quad (9)$$

Gdy przyjmujemy, że zarówno współczynnik $r = 0$, jak i $s = 0$, a $p \in \mathbb{R}$, uzyskujemy wzór na moment uogólniony (porównaj Ashkar, Mahdi, 2006):

$$M_p = \int_0^1 [x(F)]^p dF, \quad (10)$$

gdzie p może być dowolną wartością z zakresu liczb rzeczywistych. Własność ta daje teoretycznie możliwość zminimalizowania charakterystycznych dla metody momentów klasycznych (gdzie $p = 1, 2, 3, \dots$) błędów estymacji wynikających z wyższych potęg momentów. Dane pomiarowe, dla których wykonywana jest estymacja, obarczone również błędami pomiarowymi, podnoszone są do niższych potęg niż w klasycznej metodzie, przez co błędy nie ulegają multiplikacji.

W przypadku rozkładu **Log-Gumbela** wygodniej jest w równaniu (1) wprowadzić następującą re-parametryzację: $\beta = 1/\xi$, $\alpha = \alpha$. Wówczas uogólniony moment przyjmuje postać:

$$M_p = \beta^{-p\alpha} \cdot \Gamma(1 - p \cdot \alpha) \quad (11)$$

W szczególności gdy $p = 1$, uzyskujemy wartość średnią, $M_1 = \beta^{-\alpha} \times G(1 - \alpha)$, (12) natomiast dla $p = -1$, $M_{-1} = \beta^{\alpha} \times G(1 + \alpha)$. (13)

Gdy formuły dwóch dowolnych, ale różnych uogólnionych momentów porównamy do ich odpowiedników empirycznych, policzonych dla serii pomiarowej, uzyskujemy układ dwóch nieliniowych równań z dwiema niewiadomymi, który można rozwiązać stosując wyłącznie metody numeryczne.

Dla rozkładu **log-logistycznego** uogólniony moment przyjmuje postać:

$$M_p = \pi \cdot p \cdot \alpha^p \cdot \kappa \cdot \csc(p\pi\kappa) \quad (14)$$

gdzie $\csc(p\pi\kappa)$ – cosecans iloczynu $p\pi\kappa$.

W szczególności odpowiednio momenty M_1 i M_{-1} , mają więc postać:

$$M_1 = \pi \cdot \alpha \cdot \kappa \cdot \csc(\pi\kappa), \quad (15)$$

$$M_{-1} = \pi \cdot \frac{\kappa}{\alpha} \cdot \csc(\pi\kappa). \quad (16)$$

Podobnie jak w przypadku rozkładu log-Gumbela, układ równań, jaki powstaje po przyrównaniu dwóch dowolnych uogólnionych momentów do odpowiadających im momentów statystycznych z próby, można rozwiązać wyłącznie numerycznie.

Dla rozkładu Weibulla uogólniony moment przyjmuje postać

$$M_p = a^p \cdot \Gamma\left(1 + \frac{p}{b}\right), b + p > 0 \quad (17)$$

Odpowiednio momenty M_1 i M_{-1} :

$$M_1 = a \cdot \Gamma\left(1 + \frac{1}{b}\right), \quad (18)$$

$$M_{-1} = \frac{1}{a} \cdot \Gamma\left(1 - \frac{1}{b}\right), b \geq 1. \quad (19)$$

Układ równań estymatorów a i b rozwiązuje się numerycznie.

Symulacje komputerowe

Do analizy dokładności estymacji kwantyli wysokich rzędów wykorzystano metodę komputerowych symulacji Monte Carlo, która polega na generowaniu ciągu pomiarowego liczb pseudolosowych z założonego rozkładu prawdopodobieństwa ze znanymi parametrami i wartościami kwantyli ($x(P)$). W tym celu założyliśmy parametry populacji (wejściowe): średnią (m), współczynnik zmienności (CV) oraz liczebność próby (N). Następnie na podstawie wygenerowanej próby liczb pseudolosowych (oznaczonej jako v) dokonuje się estymacji parametrów i kwantyli ($\hat{x}_F$) za pomocą określonego rozkładu prawdopodobieństwa: zgodnego (Hipoteza = True, H = T) lub niezgodnego z generatorem (Hipoteza = False, H = F). Oczywiście uzyskane w ten sposób estymatory parametrów i kwantyli różnią się od teoretycznych wartości dla populacji (zadanych dla generatora), czyli są obciążone błędem. Cała operacja powtarzana jest dużą liczbę razy ($v = 1, 2, 3, \dots, S$; $S > 1000$), aby uzyskać uśrednione wartości błędów estymacji.

Spśród estymatorów kwantyli wysokich rzędów, $\hat{x}_F$, do analizy wybrano kwantyle rzędu $P = 0,1$ i $P = 0,01$, czyli dla powodzi o średnim okresie powtarzalności odpowiednio $T = 10$ i 100 lat. Obliczenia przeprowadzono dla następujących opcji i parametrów:

- rozkłady generujące: log-Gumbela, log-logistyczny, Weibulla;
- rozkłady estymujące: log-Gumbela, log-logistyczny, Weibulla;
- metody estymacji: MGM (metoda uogólnionych momentów) oraz MLM, MOM i LMM;
- założona średnia generatora: $m = 10$;
- założone współczynniki zmienności: $CV = 0,25, 0,5$ i $1,0$;

– liczebności próby: $N = 25, 50$ i 100 .

Jako kryterium dobroci dopasowania kwantyli zaproponowano wartości względnego pierwiastka średniego błędu kwadratowego (RRMSE):

$$RRMSE(\hat{x}_T^{(v)}) = \sqrt{E(\hat{x}_T^{(v)} - x_T)^2} / x_T. \quad (20)$$

Wyniki RRMSE przedstawiane są procentach.

Zaproponowane podejście umożliwia porównanie wyników uzyskanych dla różnych kombinacji generatorów i estymatorów oraz różnych metod estymacji.

Wyniki obliczeń

Przypadek szczególny: $p = 1$ i $p = -1$

Narzucając się naturalnie parą momentów uogólnionych jest wartość średnia oraz moment o potęgze $p = -1$. Jednakże wyniki dla tej pary momentów wskazują na nikłą konkurencyjność metody momentów uogólnionych w stosunku do pozostałych popularnych metod estymacji kwantyli stosowanych w statystyce zjawisk ekstremalnych. Obniżenie potęgi do jedności i -1 nie przyniosło bowiem oczekiwanych korzyści w postaci redukcji błędów systematycznych i średnich kwadratowych kwantyli wysokich rzędów. Metoda uogólnionych momentów rzadko plasuje się na pierwszym miejscu spośród wszystkich używanych w obliczeniach metod niezależnie od tego, czy rozkład prawdopodobieństwa estymatora pokrywa się z rozkładem prawdopodobieństwa generatora (sytuacja $H = T$) czy też nie (przypadek $H = F$). Co najwyżej wyniki dla MGM zbliżone są poziomem błędów do wyników uzyskanych dla innych metod, co przy wadach metod MLM, MOM i LMM może stanowić uzasadnienie jej ograniczonej stosowalności w wyjątkowych przypadkach.

Co ciekawe, słabe wyniki MGM nie zależą ani od wartości współczynnika CV ani od wielkości próby. Teoretycznie należałoby się spodziewać, że wyniki dla przypadku $H = T$, małych wartości współczynnika zmienności i dużych prób powinny charakteryzować się mniejszymi błędami niż dla innych kombinacji parametrów wejściowych. Tak też po części jest, jednakże metoda MGM nie wyróżnia się spośród pozostałych metod szczególnie małymi wartościami błędów usprawiedliwiającymi stosowanie tej metody w praktyce.

Przyczyn słabego wyniku metody MGM upatrujemy, po pierwsze, w dużej asymetrii używanych rozkładów ekstremalnych (ich grubych ogonów), teoretycznie bowiem im wyższy rząd momentu, tym bardziej dotyczy on ogonów (obu – lewego i prawego!) rozkładu prawdopodobieństwa. Ponadto analizowane kwantyle są kwantylami o bardzo małym prawdopodobieństwie przewyższenia $x(P=0,1)$

i $x(P=0,01)$, określają odpowiednio dziesięcio- i stuletni okres powtarzalności ekstremalnej powodzi, co plasuje je daleko w prawym ogonie funkcji, na który teoretycznie momenty niskich rzędów nie mają znaczącego wpływu. Należy zatem stwierdzić, że od strony praktycznego zastosowania metoda MGM dla $p = 1$ i $p = -1$ stanowi „ślepy zaułek” i należy ją stosować w obliczeniach jako metodę uzupełniającą.

Optymalna wartość potęgi p z zakresu $p \in (-2, 2)$

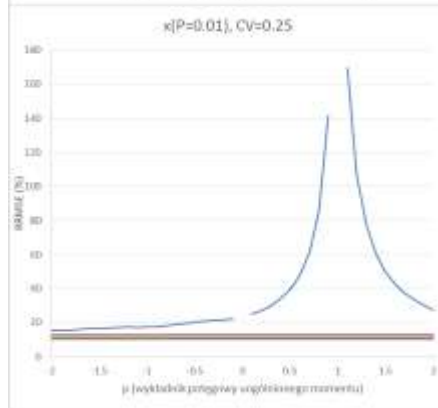
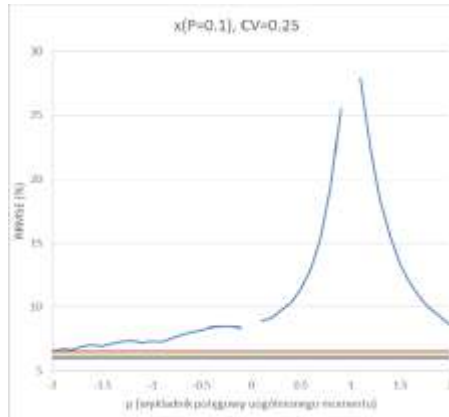
Metodę Monte Carlo wykorzystano ponadto do wyznaczenia optymalnych wartości p z zakresu $p \in (-2, 2)$, które w metodzie MGM wraz z momentem rzędu $p = 1$ (średnia) dadzą niższe wartości błędów RRMSE estymatorów kwantyli powodziowych niż metody konkurencyjne. Podobnie jak w przypadku pary $p = 1$ i $p = -1$ w eksperymencie wykorzystano rozkłady LL, LG i We, zarówno w konfiguracji $H = T$, jak i $H = F$. Założenia co do wartości parametrów poszczególnych generatorów liczb losowych opisano w rozdziale dotyczącym najczęściej stosowanych metod estymacji parametrów. Jakkolwiek dokonano obliczeń dotyczących wszystkich przypadków wymienionych w tym rozdziale, to ze względu na ograniczenia co do objętości tekstu w niniejszym rozdziale pokazano tylko wybrane wyniki, jeśli to możliwe, uogólniając wnioski na pozostałe przypadki. Wybrane wyniki analizy przedstawiono na wykresach (rys. 1-3). Przerwane linie wykresów MGM (kolor niebieski) na rysunkach oznaczają, że niemożliwe jest uzyskanie wyników estymacji, gdy jeden z momentów to średnia, a potęgą drugiego momentu oscyluje wokół wartości $p = 0$ (wówczas drugi moment $M_0 = 1$) lub $p = 1$ (wówczas oba momenty to średnie).

Dla tak rozległych przedziałów założeń nie można uzyskać jednej optymalnej wartości potęgi momentu uogólnionego, aby wraz ze średnią dla wszystkich 3 rozkładów charakteryzowała się najmniejszymi wartościami błędów estymacji kwantyli powodziowych. Mimo to, analiza wyników pozwala wysnuć pewne ogólne wnioski co do zakresu stosowalności metody MGM. Ponieważ metody MLM, MOM i LMM nie stanowią tematu niniejszej pracy i wzajemne porównanie tych metod wybiega poza tematykę artykułu, ich wyniki będą komentowane jedynie w kontekście rezultatów dla metody MGM.

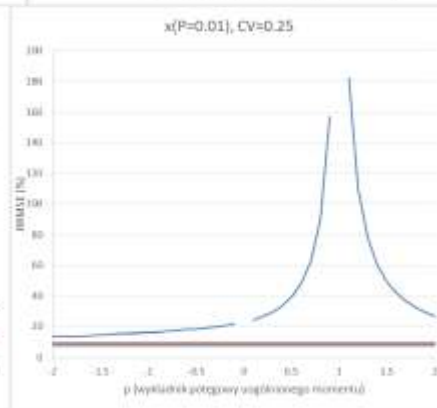
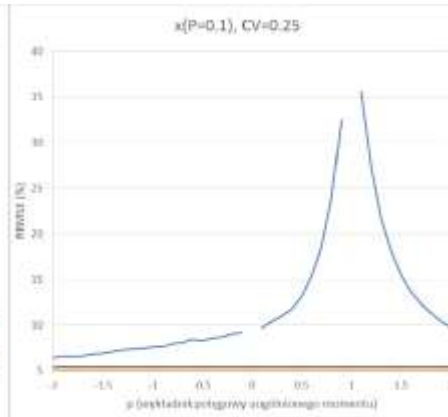
Przebieg zmienności wartości RRMSE wysokich kwantyli estymowanych metodą MGM w zależności od wartości p dla przypadków LG i LL zachowuje się podobnie co do kształtu linii wykresu. Wyniki dla przypadku $H = T$ generalnie charakteryzują się niższymi wartościami RRMSE niż dla przypadków, gdy rozkład populacji jest nieznany. Niezależnie od przypadku $H = T$ czy $H = F$, dla małych wartości współczynnika zmienności ($CV = 0,25$) klasyczne metody estymacji (MLM, MOM i LMM) charakteryzują się niższymi wartościami RRMSE estymatorów

Generator próby losowej z rozkładu Log-Gumbela, N=50

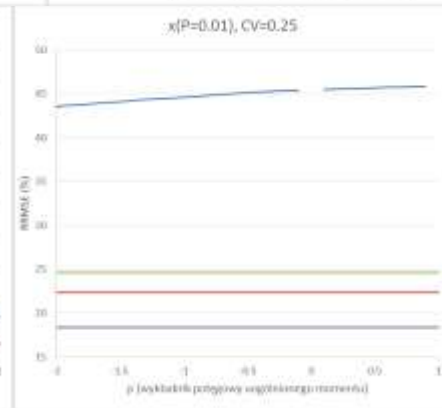
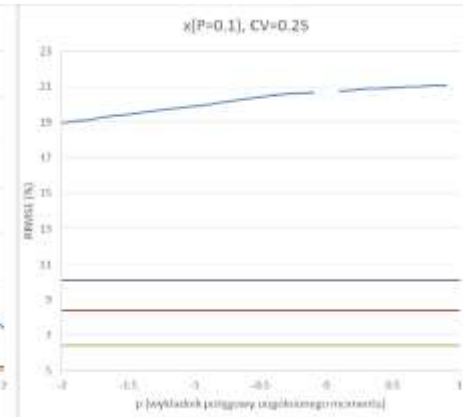
Estymator LG

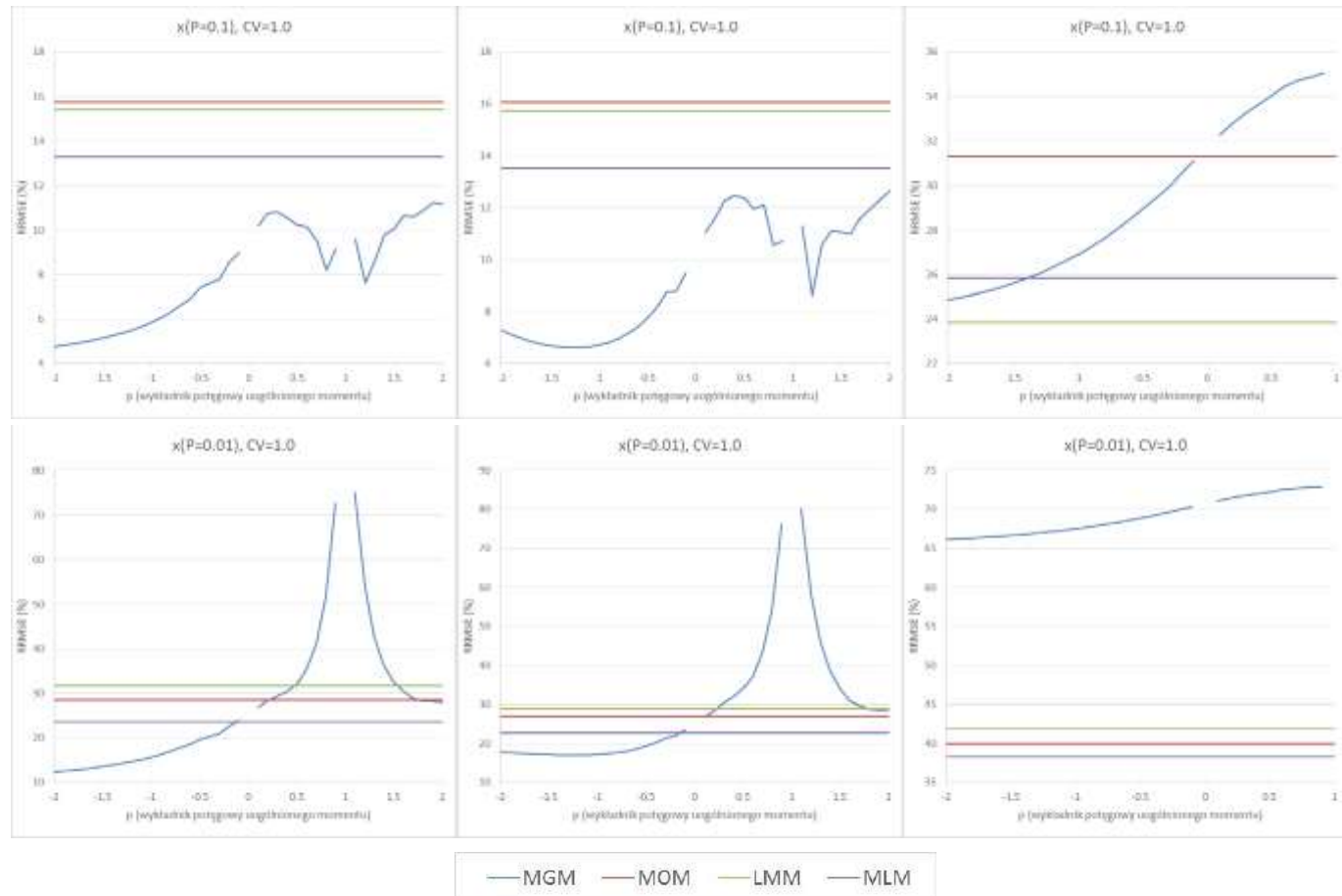


Estymator LL



Estymator We



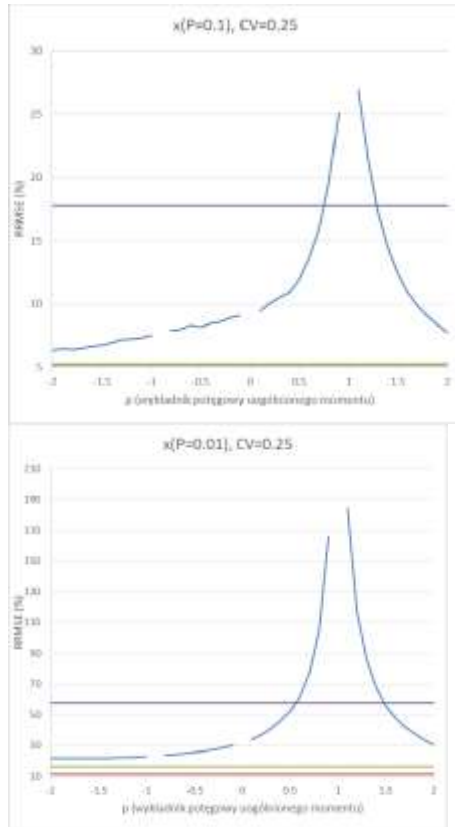


Rys. 1. Wartości względnego pierwiastka średniego błędów kwadratowych dla 4 metod estymacji i 3 modeli. Rozkład populacji Log-Gumbel. Pierwszy moment uogólniony $p = 1$, drugi p zmienny (oś odciętych)

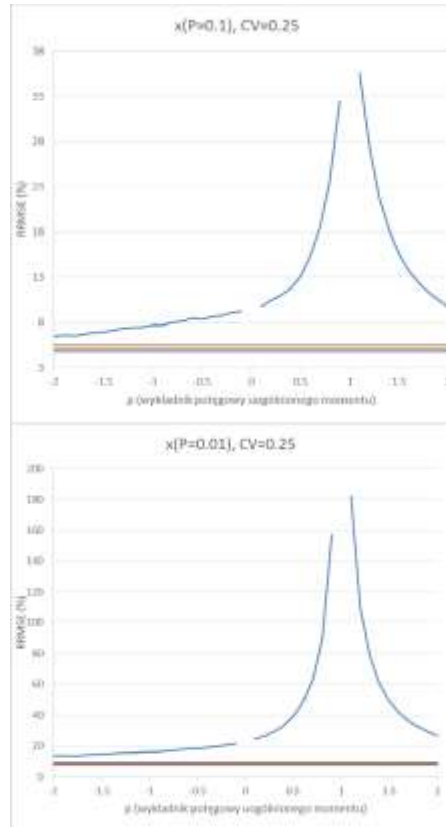
Fig. 1. The absolute mean square error for 4 methods of estimation and 3 models. The population distribution function Log-Gumbel. First generalised moment is mean ($p = 1$), in the second one p varies (horizontal axis)

Generator próby losowej z rozkładu Log-Logistycznego, $N=50$

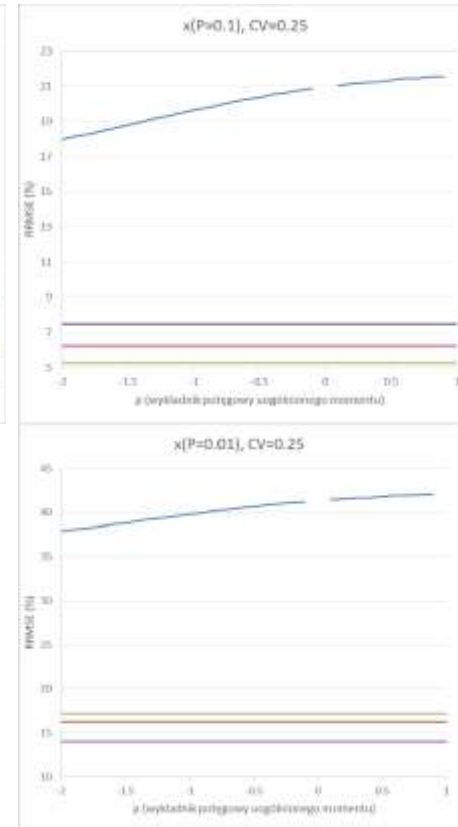
Estymator LG

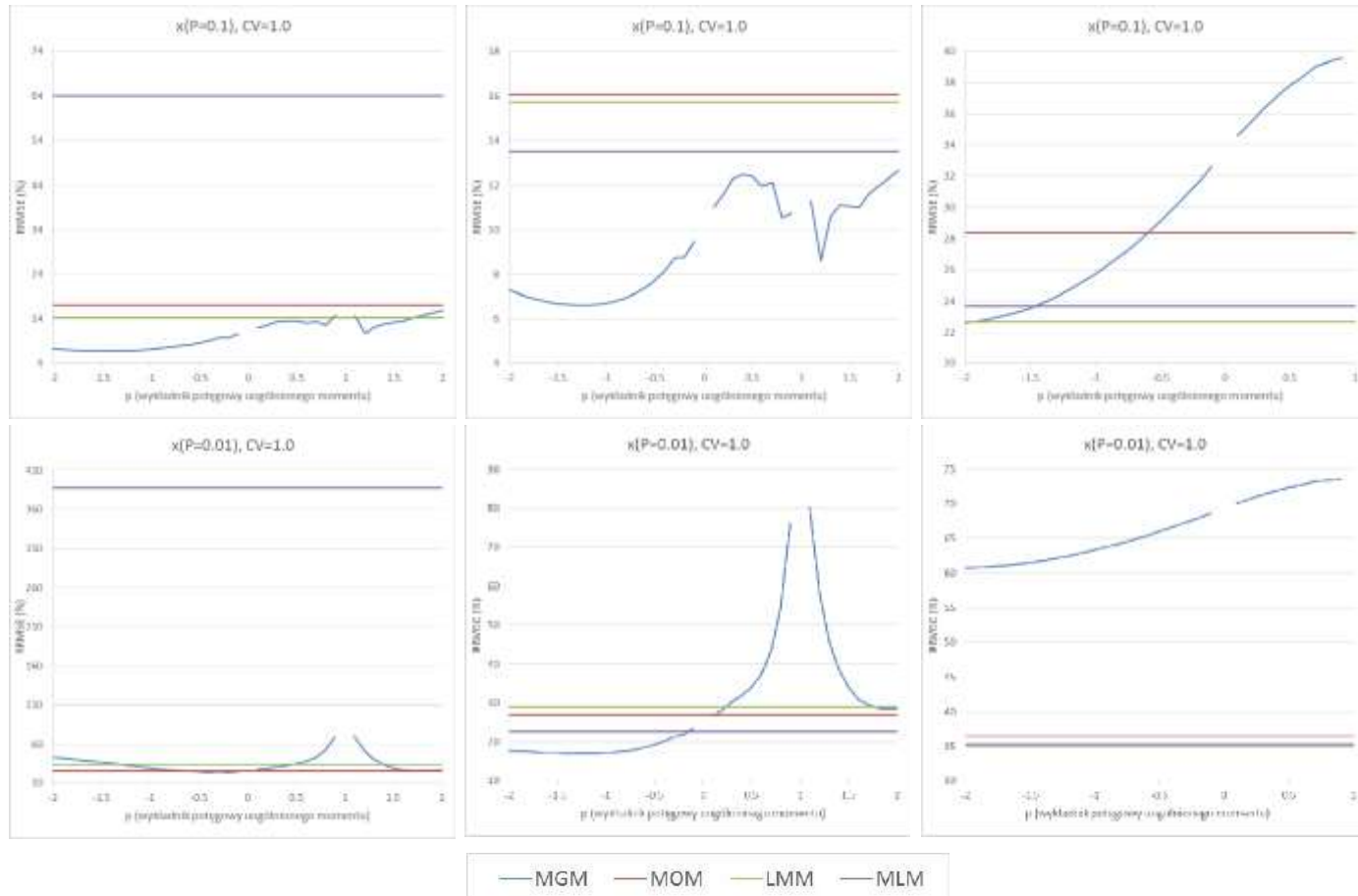


Estymator LL



Estymator We



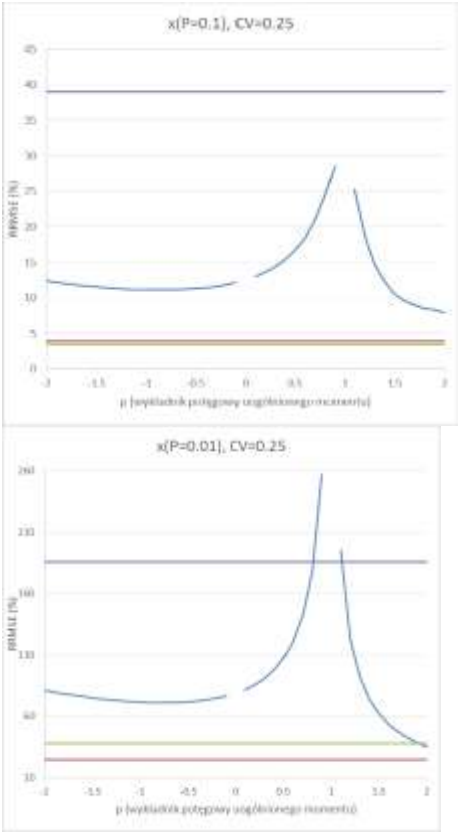


Rys. 2. Wartości względnego pierwiastka średniego błędu kwadratowego dla 4 metod estymacji i 3 modeli. Rozkład populacji Log-Logistyczny. Pierwszy moment uogólniony $p = 1$, drugi p zmienne (oś odciętych)

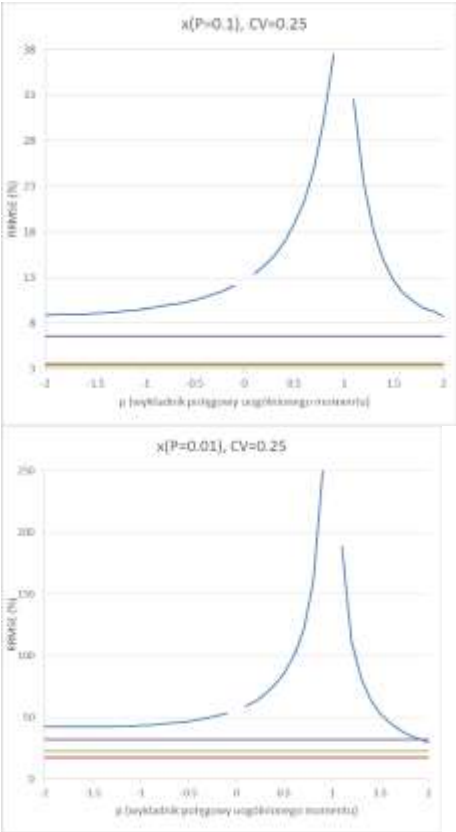
Fig. 2. The absolute mean square error for 4 methods of estimation and 3 models. The population distribution function Log-Logistic. First generalised moment is mean ($p = 1$), in the second one p varies (horizontal axis)

Generator próby losowej z rozkładu Weibulla, N=50

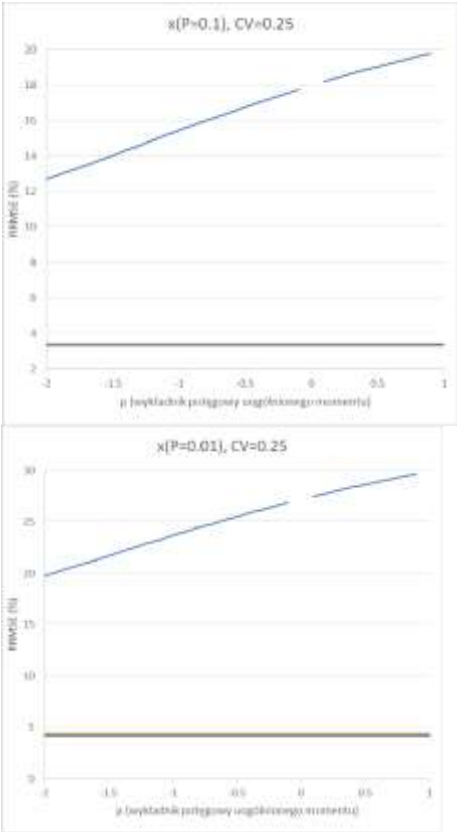
Estymator LG

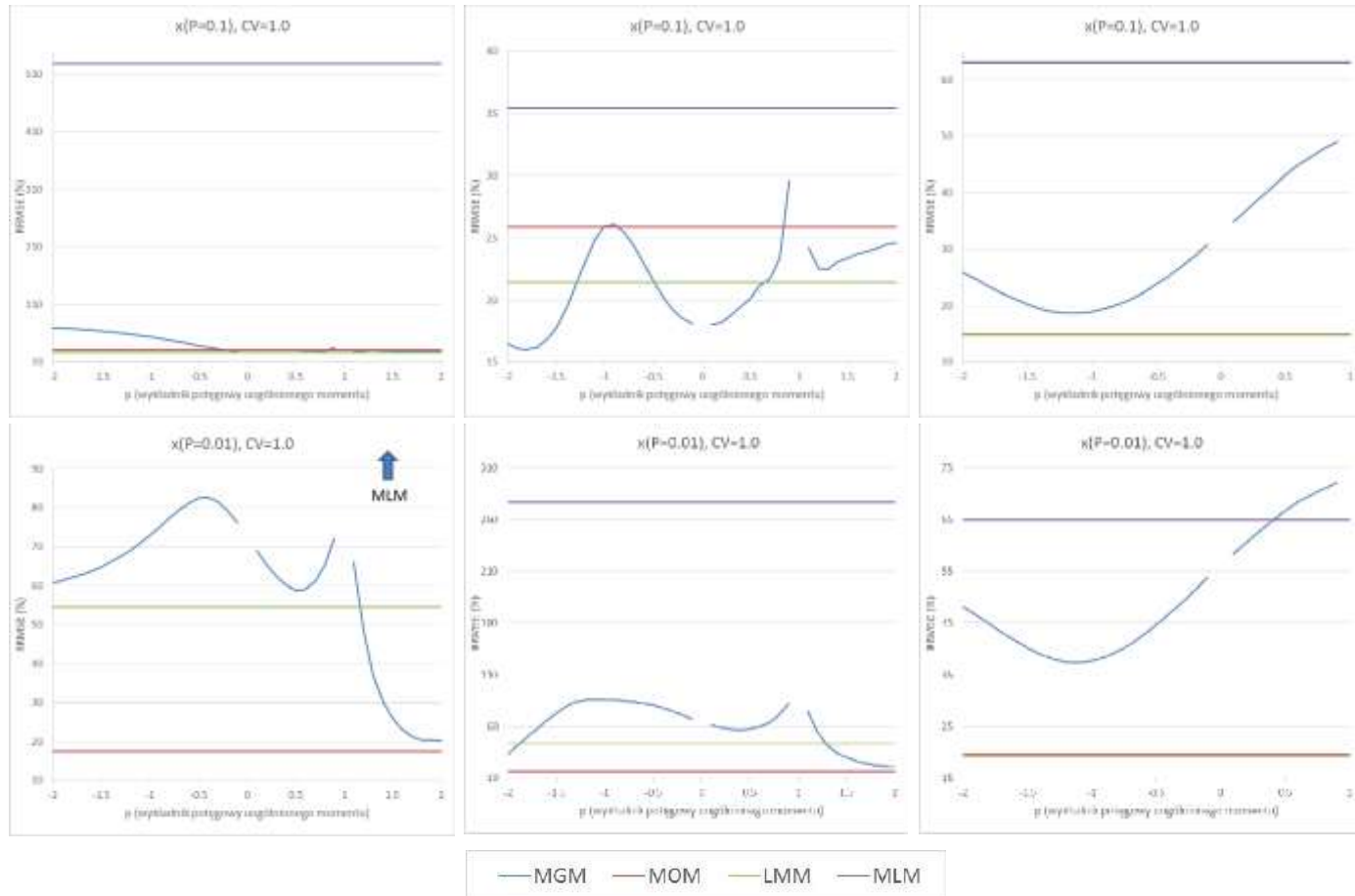


Estymator LL



Estymator We





Rys. 3. Wartości względnego pierwiastka średniego błędów kwadratowych dla 4 metod estymacji i 3 modeli. Rozkład populacji Weibulla. Pierwszy moment uogólniony $p = 1$, drugi p zmienny (oś odciętych)

Fig. 3. The absolute mean square error for 4 methods of estimation and 3 models. The population distribution function Weibull. First generalised moment is mean ($p = 1$), in the second one p varies (horizontal axis)

$\hat{x}(P=0,1)$ i $\hat{x}(P=0,01)$ niż dla metody MGM. Dla dużych wartości współczynnika zmienności ($CV=1$) metoda MGM zyskuje przewagę nad metodami alternatywnymi, zwłaszcza dla przedziału ujemnych wartości współczynnika p . Jednakże dla modeli LG i LL nie jest możliwe wskazanie jednej wartości współczynnika p , rekomendowanego do wykorzystania w metodzie MGM jako gwarantującej mniejsze wartości błędów estymacji kwantyli powodziowych niż metody alternatywne, a jedynie pewnego zakresu zmienności p . W praktycznym zastosowaniu metody MGM dla modeli LL i LG należy stosować ujemne wartości p , przy czym im mniejsze p (bliższe -2), tym mniejszych wartości $RRMSE$ można się spodziewać.

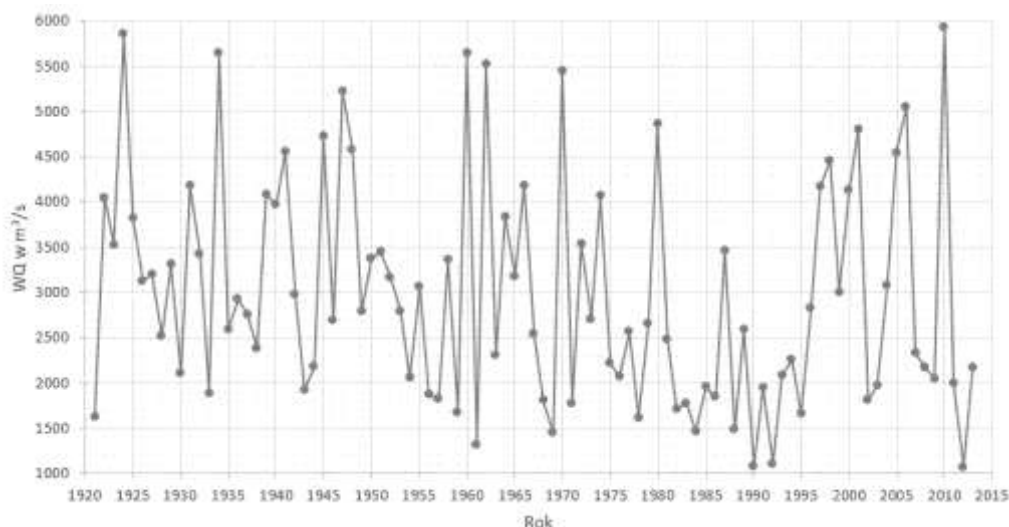
Dla rozkładu We wyniki zachowują się odmiennie od pozostałych dwóch modeli, co wynika wprost ze specyfiki rozkładu Weibulla (Rao, 2000). Przede wszystkim dla $p > 1$ koniunkcja dwóch równań momentów uogólnionych staje się układem sprzecznym. Ponadto dla modelu Weibulla niezależnie od przypadku $H = T$ lub $H = F$, CV i rzędu kwantyla P , rzadko kiedy metoda MGM wykazuje się mniejszymi wartościami błędów $RRMSE$ niż metody alternatywne. W niektórych z rozważanych przypadków (kombinacji: model estymatora kwantyli, CV i N) metoda MGM pod względem dokładności estymacji przewyższa metodę MLM, lecz nie wynika to ze szczególnych właściwości MGM, lecz ograniczonych możliwości MLM. Oznacza to, że metoda MGM nie jest metodą uniwersalną dobrze współpracującą z ze wszystkimi rozkładami o grubych ogonach.

Przykład zastosowania – Wisła, profil Warszawa-Nadwilanówka

Profil wodowskazowy Warszawa-Nadwilanówka Śluza położony jest na lewym brzegu 504,1 km Wisły, w rejonie ujścia rzeki Wilanówki. W miejscu posadowienia wodowskazu Wisła płynie jednym korytem, a brzegi są umocnione; na prawym brzegu znajdują się ostrogi. Brzeg rzeki jest częściowo zagospodarowany (nieopodal znajdują się ogródki działkowe i budynki gospodarcze oraz mieszkalne), a częściowo pokryty gęstymi zaroślami i/lub lasem łęgowym (ols), wpływającymi na reżim powodziowy rzeki.

Dysponujemy 93-letnim ciągiem pomiarów WQ z lat 1921-2013. Hydrogram WQ dla profilu pokazano na rys. 4.

Praktycznie każdy długi ciąg pomiarowy przepływów maksymalnych rzecznych w danym profilu wodowskazowym charakteryzuje się pewną niejednorodnością, co łatwo zauważyć również na załączonym hydrogramie w przypadku pomiarów w profilu Warszawa-Nadwilanówka. Jednocześnie badanie statystyczne niejednorodności ciągu oraz zastosowanie metod statystycznych uwzględniających wybiegałoby znacznie poza zakres tematu zaproponowany w niniejszej pracy. Dla uproszczenia wywodu przyjęto, że przepływy maksymalne roczne w Warszawie-



Rys. 4. Przepływy maksymalne roczne w profilu Wisła – Warszawa-Nadwilanówka

Fig. 4. Annual maximal flows for the Warsaw-Nadwilanówka gauge at the River Vistula

-Nadwilanówce podlegają jednemu rozkładowi prawdopodobieństwa przez cały okres pomiarowy.

Na podstawie próby dokonano estymacji kwantyli rozkładów LG, LL o prawdopodobieństwie przewyższenia $P = 0,001$ do $0,99$, co odpowiada zakresowi danych WQ w profilu Warszawa-Nadwilanówka. Rozkład Weibulla (We), jako dający niezadowolające rezultaty w eksperymencie numerycznym, został wyłączony z obliczeń. Wyniki estymacji kwantyli metodami MLM, MOM, LMM oraz MGM wraz z danymi rzeczywistymi przedstawiono na rys. 5. W metodzie MGM dla wszystkich trzech rozkładów prawdopodobieństwa wykorzystano dwa momenty statystyczne: średnia ($p = 1$) i moment generalizowany z $p = -1,5$, przy czym wybór wartości potęgi drugiego z momentów był podyktowany faktem, że w eksperymencie Monte Carlo właśnie dla tej wartości uzyskiwano stosunkowo niewielkie błędy $RRMSE$ oszacowań wysokich kwantyli oraz stabilnością rozwiązań numerycznych.

Z wykresów wynika, że w przypadku rozkładów LG i LL metoda MGM daje wyniki zbliżone do konkurencyjnych metod. Widoczne jest to szczególnie w przypadku rozkładu log-logistycznego (LL), gdzie wszystkie 4 analizowane metody dają bardzo podobne oszacowania kwantyli w całym zakresie rozważanego przedziału prawdopodobieństwa przewyższenia. Świadczyć to może o dobrej zgodności modelu LL z ciągiem pomiarowym z Warszawy-Nadwilanówki. Warto zwrócić również uwagę, że o ile modele LG i LL wraz z metodą MLM stosunkowo dobrze radzą sobie z odzwierciedleniem wartości kwantyla $x(P=0,1)$, to błędy dla $x(P=0,01)$, sięgają już odpowiednio 4000 i 1500 m³/s! Rozbieżności takie powstają pomimo

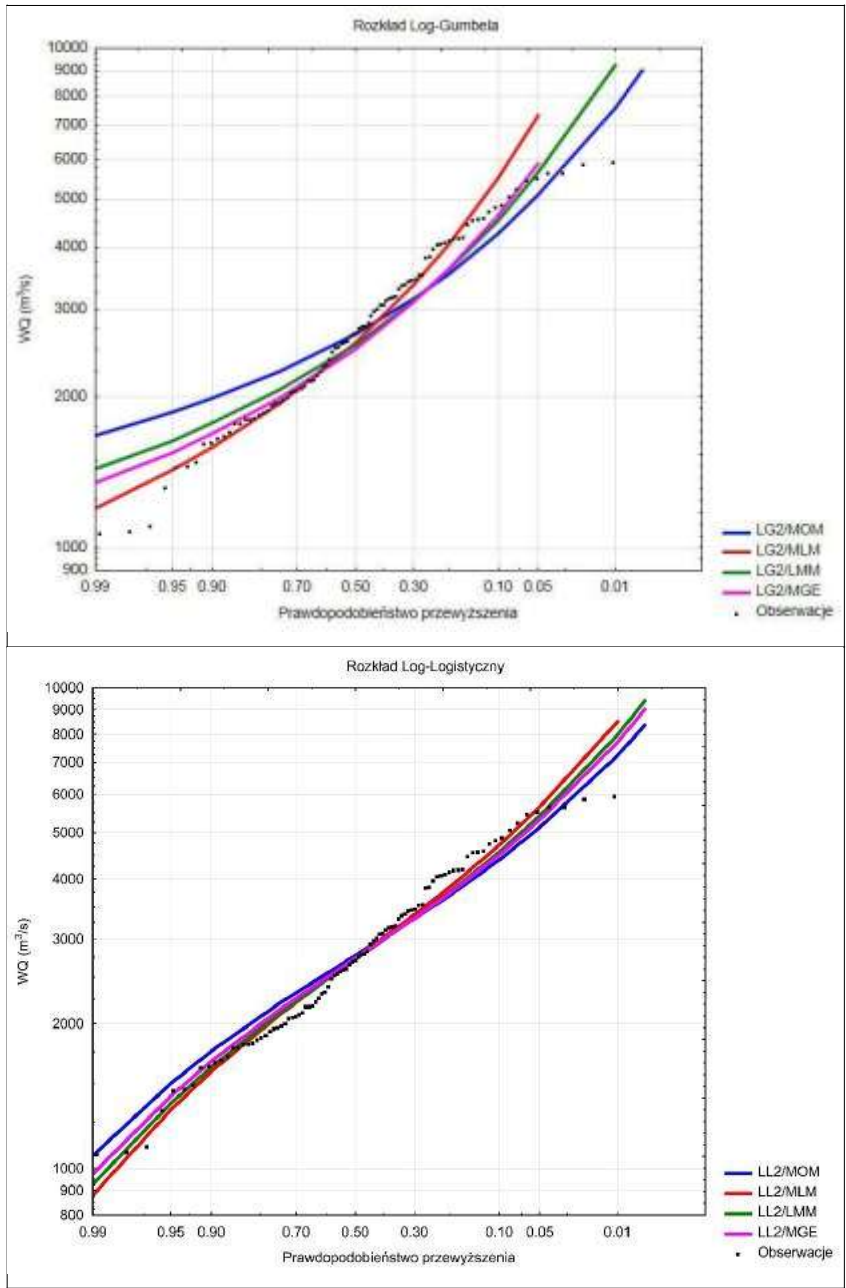
tego, że rozkłady o grubych ogonach teoretycznie powinny lepiej dopasowywać się do ekstremalnych wartości kwantyli. Wynik ten, choć jednostkowy, poddaje w wątpliwość sens stosowania rozkładów o grubych ogonach do analizy częstości występowania powodzi w profilu Warszawa-Nadwilanówka. Co oczywiste, wyniki uzyskane dla danych wodowskazowych w profilu Warszawa-Nadwilanówka nie są w żaden sposób uniwersalne i wnioski wynikające z obliczeń nie powinny być generalizowane na inne wodowskazy lub rozkłady statystyczne.

Wnioski i rekomendacje

W niniejszej pracy zaprezentowano metodę momentów uogólnionych do obliczania kwantyli powodziowych trzech rozkładów gruboogoniastych (log-Gumbela, log-logistycznego i Weibulla). Metoda ta została porównana z trzema innymi stosowanymi często w hydrologii metodami estymacji parametrów i kwantyli. Przeprowadzono eksperyment symulacyjny Monte Carlo, w którym zbadano wartości średnich błędów kwadratowych estymacji kwantyli o małym prawdopodobieństwie przewyższenia ($P = 0,1$ i $P = 0,01$), zachowując przy tym realizm warunków hydrologicznych dla rzek w Polsce: adekwatne zakresy długości serii pomiarowych oraz współczynników zmienności. Następnie pokazano na przykładzie ciągu przepływów maksymalnych rocznych w profilu Warszawa-Nadwilanówka na Wiśle wzajemną zależność danych rzeczywistych, metody MGM i metod konkurencyjnych.

Badania za pomocą metody Monte Carlo wykazały ograniczoną konkurencyjność metody MGM w stosunku do metod alternatywnych. Niestabilność numeryczna oraz stosunkowo wysokie wartości błędów oszacowań kwantyli powodziowych dla dodatnich potęg drugiego z momentów uogólnionych sprawia, że w dodatnim zakresie wartości potęgi momentu uogólnionego MGM nie jest rekomendowana jako alternatywa dla klasycznych metod estymacji. Jednakże w przypadku potęg ujemnych, gdy $p \in (-2,0)$, metoda MGM bardzo często zyskuje przewagę nad metodami alternatywnymi, szczególnie dla wysokich wartości współczynnika zmienności CV, przy czym zysk jest widoczny zarówno przy prawdziwym, jak i fałszywym założeniu modelu populacji. Zatem metoda MGM, będąc pozbawiona wad swoich konkurentek, przy zastosowaniu momentów z ujemnymi potęgami, może stanowić prosta alternatywę do praktycznych zastosowań estymacji kwantyli powodziowych. Przykład dotyczący Warszawy-Nadwilanówki pokazał, że w przypadku rozkładów LG i LL metoda MGM daje porównywalne wyniki do metod konkurencyjnych.

Podsumowując, należy stwierdzić, że jakkolwiek metoda MGM w zastosowanym kształcie (wykorzystując dwa momenty – średnią i moment generalizowany o potęgę p) charakteryzuje się umiarkowanie pozytywnymi właściwościami na



Rys. 5. Zestawienie danych rzeczywistych i wyników modelowania przepływów maksymalnych rocznych na profilu Wisła – Warszawa-Nadwilanówka

Fig. 5. Comparison of the observations with the estimation results for annual maximal flows for Warsaw-Nadwilanówka, the River Vistula

tle metod konkurencyjnych, autorzy polecają jej używanie dla rozkładów LL i LG i ujemnych potęg jednego momentu w zakresie $(-2, 0)$, jako uzupełnienie metod klasycznych, ale nie ich alternatywę.

Materiały wpłynęły do redakcji 9 V 2016.

Literatura

- Ashkar F., Bobee B., 1987, *The generalized method of moments as applied to problems of flood frequency analysis: some practical results for the log-Pearson type 3 distribution*. J. Hydrol. 90, 199-217.
- Ashkar F., Mahdi S., 2003, *Comparison of two fitting methods for the log-logistic distribution*. Water Resour., Res. 39, 8, 1217.
- Ashkar F., Mahdi S., 2006, *Fitting the log-logistic distribution by generalized moments*. Journal of Hydrology, 328, 3-4, 694-703, ISSN 0022-1694, <http://dx.doi.org/10.1016/j.jhydrol.2006.01.014>.
- Coles S., 2001, *An introduction to statistical modeling of extreme values*. Springer, London.
- Cunnane C., 1989, *Statistical distributions for flood frequency analysis*. Operational Hydrol. Rep. No.33, WMO-No. 718, Geneva.
- Greenwood J.A., Landwehr J.M., Matalas N.C., Wallis J.R., 1979, *Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form*. Water Res. Research, 155, 1049-1054.
- Gumbel E.J., 1958, *Statistics of extremes*. Columbia University Press, New York.
- Hosking J.R.M., 1986, *The theory of probability weighted moments*. Research Rep. RC 12210, ss. 160, IBM Research Division, Yorktown Heights, NY.
- Hosking J.R.M., 1990, *L-moments: Analysis and estimation of distributions using linear combinations of order statistics*. J. R. Stat. Soc. Ser. B., 52, 105-124.
- Hosking J.R.M., Wallis J.R., 1997, *Regional Frequency Analysis: An Approach based on L-Moments*. Cambridge University Press, Cambridge, UK.
- Johnson N.L., Kotz S., Balakrishnan N., 1994, *Continuous Univariate Distributions*. T. 1, 2 Ed., Wiley, ss. 784.
- Johnson N.L., Kotz S., Balakrishnan N., 1995, *Continuous Univariate Distributions*. T. 2, 2 Ed., Wiley, ss. 752.
- Kaczmarek Z., 1970, *Metody statystyczne w hydrologii i meteorologii*. WLiK, Warszawa.
- Katz R.W., Parlange M.B., Naveau P., 2002, *Statistics of extremes in hydrology*. Adv Water Resour 25:1287-1304.
- Kendall M., Stuart A., 1969, *The Advanced Theory of Statistics*. McGraw-Hill Press, New York.
- Kochanek K., Strupczewski W.G., Singh V.P., Weglarczyk S., 2005, *Are Parsimonious Flood Frequency Models More Reliable than the True Ones? II. Comparative assessment of the performance of simple models versus the parent distribution*. Acta Geoph. Polonica, 53, 4, 437-457.
- Liang Z., Hu Y., Li B., Yu Z., 2014, *A modified weighted function method for parameter estimation of Pearson type three distribution*. Water Resour. Res., 50, 3216-3228, doi:10.1002/2013WR013653
- Ma X., 1984, *The weighted function method for calculating hydrologic frequency parameters*. J. China Hydrol., 21, 1-8.
- Rao A.R., Hamed K.H., 2000, *Flood Frequency Analysis*. CRC Press.
- Rowiński P.M., Strupczewski W.G., Singh V.P., 2002, *A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: I. Known pdf*, Hydrol. Sc. J. 47, 1, 107-122.

- Strupczewski W.G., Singh V.P., Węglarczyk S., 2002a, *Asymptotic bias of estimation methods caused by the assumption of false probability distribution*. J. Hydrol. 258, 1-4, 122-148.
- Strupczewski W.G., Węglarczyk S., Singh, V.P., 2002b, *Model error in flood frequency estimation*. Acta Geoph. Pol. 50, 2, 279-319.
- Strupczewski W.G., Kochanek K., Węglarczyk S., Singh V.P., 2005a, *On robustness of large quantile estimates of log-Gumbel and log-logistic distributions to largest element of the observation series: Monte Carlo results vs. first order approximation*. Stoch Environ Res Risk Assess, 19, 4, 280-291, doi:10.1007/s00477-005-0232-x
- Strupczewski W.G., Kochanek K., Singh V.P., Węglarczyk S., 2005b, *Are Parsimonious Flood Frequency Models More Reliable than the True Ones? I. Accuracy of Quantiles and Moments Estimation AQME – Method of Assessment*. Acta Geoph. Polonica 53, 4, 419-436.
- Strupczewski W.G., Kochanek K., Węglarczyk S., Singh V.P., 2007, *On robustness of large quantile estimates to largest elements of the observation series*. Hydrol. Process. 20, 1328-1344, doi:10.1002/hyp.6342
- Wallis J. R., Matalas N.C., Slack J.R., 1974, *Just a moment!* Water Resour. Res., 102, 211-219, doi:10.1029/WR010i002p00211
- Węglarczyk S., Strupczewski W.G. and Singh, V.P., 2002, *A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses: II. Assumed pdf*, Hydrol. Sc. J. 471, 123-137.

Streszczenie

W pracy przedstawiamy metodę estymacji kwantyli powodziowych za pomocą uogólnionych momentów jako alternatywę dla klasycznych metod estymacji, czyli metody momentów, momentów liniowych i największej wiarygodności. W swoich teoretycznych założeniach metoda uogólnionych momentów, uwalniając się od założeń naturalnych potęg momentów, umożliwia utrzymanie tych potęg na możliwie niskim poziomie, przez co minimalizuje błędy pomiarowe oraz niejednorodności próby losowej. Wyprowadzono wzory na momenty generalizowane dla trzech często stosowanych w analizie częstości powodzi modeli charakteryzujących się grubym ogonem: rozkłady Log-Gumbela, Log-Logistyczny i Weibulla. Za pomocą metody symulacyjnej Monte Carlo obliczono błędy systematyczne i średnie błędy kwadratowe kwantyli o okresie powtarzalności 10 i 100 lat. Wyliczenia pokazały, że metoda dla wybranych wartości rzędów momentów może stanowić alternatywę dla metod klasycznych.

Dodatkowo przeprowadzono obliczenia na podstawie ciągu przepływów maksymalnych z profilu Warszawa-Nadwilanówka, które wykazały, że metoda uogólnionych momentów nie powinna być stosowana dla rozkładu Weibulla.

Słowa kluczowe: częstość występowania powodzi, metody estymacji, rozkłady statystyczne o grubych ogonach, momenty generalizowane

Summary

This paper considers the estimation of the flood quantiles by means of the method of generalised moments which is the alternative for classical methods of estimation, namely methods of moments, linear moments and maximum likelihood. Theoretically the method of generalised moments is released from the assumption of the natural powers of moments and enables to keep these powers as low as possible diminishing the measurement errors and errors stemming from the heterogeneity of the random sample. The equations for the generalised moments for the three heavy-tailed models

commonly used in the flood frequency analysis: log-Gumbel, Log-Logistic and Weibull. We used the Monte Carlo simulations to calculate the mean square errors of quantiles of the 10- and 100-year floods. The calculations revealed, that the method of generalised moments for certain range of powers can be an alternative for classical methods. Additionally, we carried out the calculations for the annual maximum flows for the Warsaw-Nadwilanówka at the River Vistula, which revealed that method of generalised moments should not be used for the Weibull model.

K e y w o r d s : flood frequency analysis, methods of estimation, heavy-tailed distribution functions, generalised moments

Krzysztof Kochanek
kochanek@igf.edu.pl

Wojciech Feluch
wfeluch@gmail.com

Politechnika Warszawska, filia w Płocku. Wydział Budownictwa, Mechaniki i Petrochemii