

Analiza 2

Analiza zdarzeń ekstremalnych

Antonina Brzeska 246824

1. Krótki opis wybranych danych. Zestaw danych opisuje zanotowane wysokości fal na wybrzeżu Australijskim w miejscowości Mooloolaba. Wybrane pomiary zawierają 29108 rekordów, z czego mierzono je w okresie 20 miesięcy począwszy od stycznia 2017 roku. W każdym miesiącu pobrano podobną ilość danych (47 rekordów dziennie), dlatego można założyć, że analizowane dalej bloki odpowiadają kolejnym miesiącom.

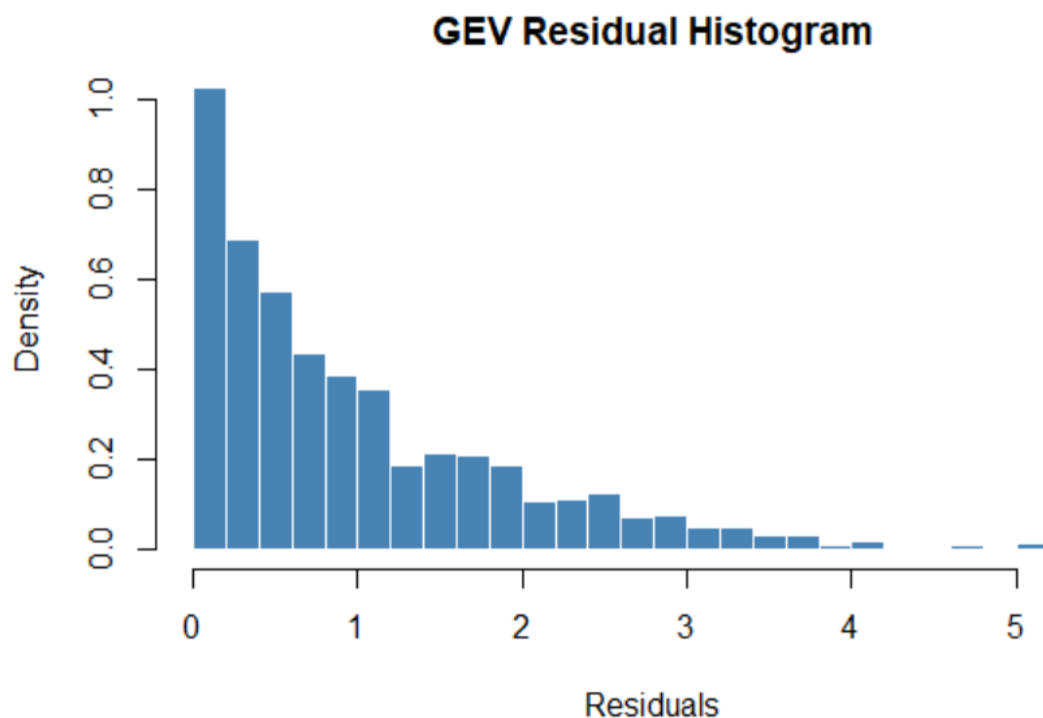
2. Analiza metodą maksimów blokowych (BMM). Wyestymowano parametry rozkładu GEV w oparciu o maksima z ustalonych bloków miesięcznych:

	xi	sigma	mu
Parametry rozkładu GEV	-0.3372591	1.0874319	4.5626598

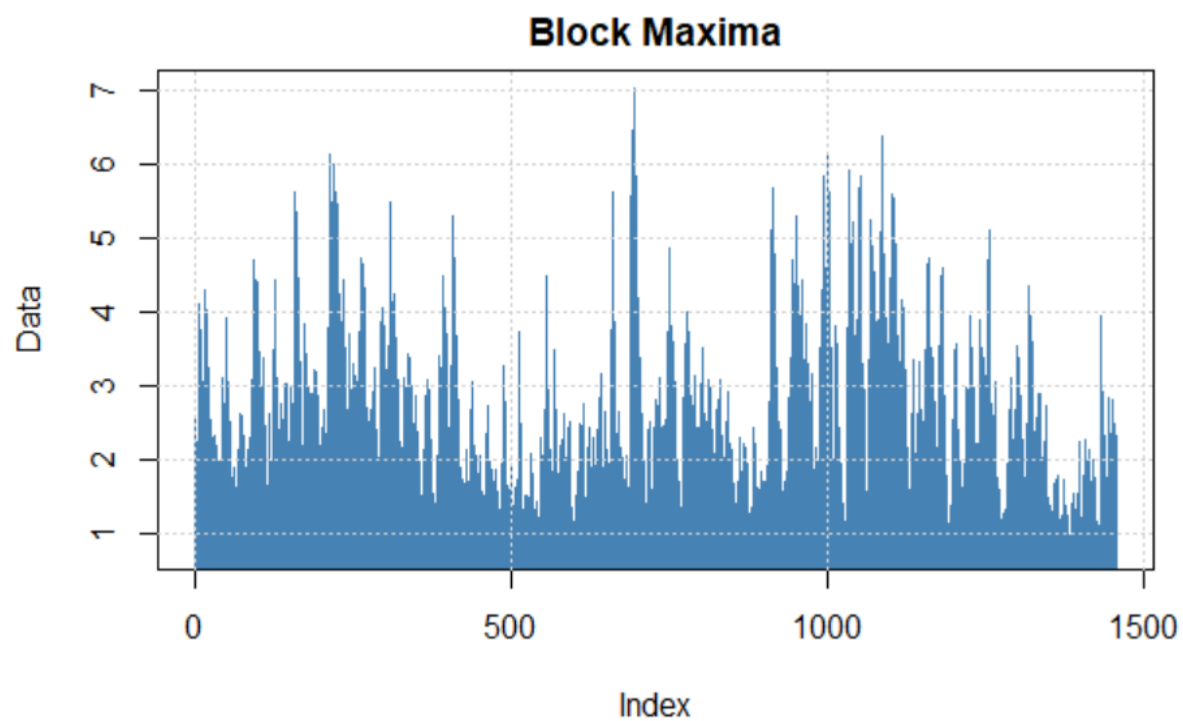
$\xi < 0$, więc dystrybuanta F zbiegająca do rozkładu GEV należy do maksymalnej dziedziny przyciągania Weibulla.

Przeprowadzono analizę oceniającą dobroć dopasowania za pomocą następujących wykresów diagnostycznych:

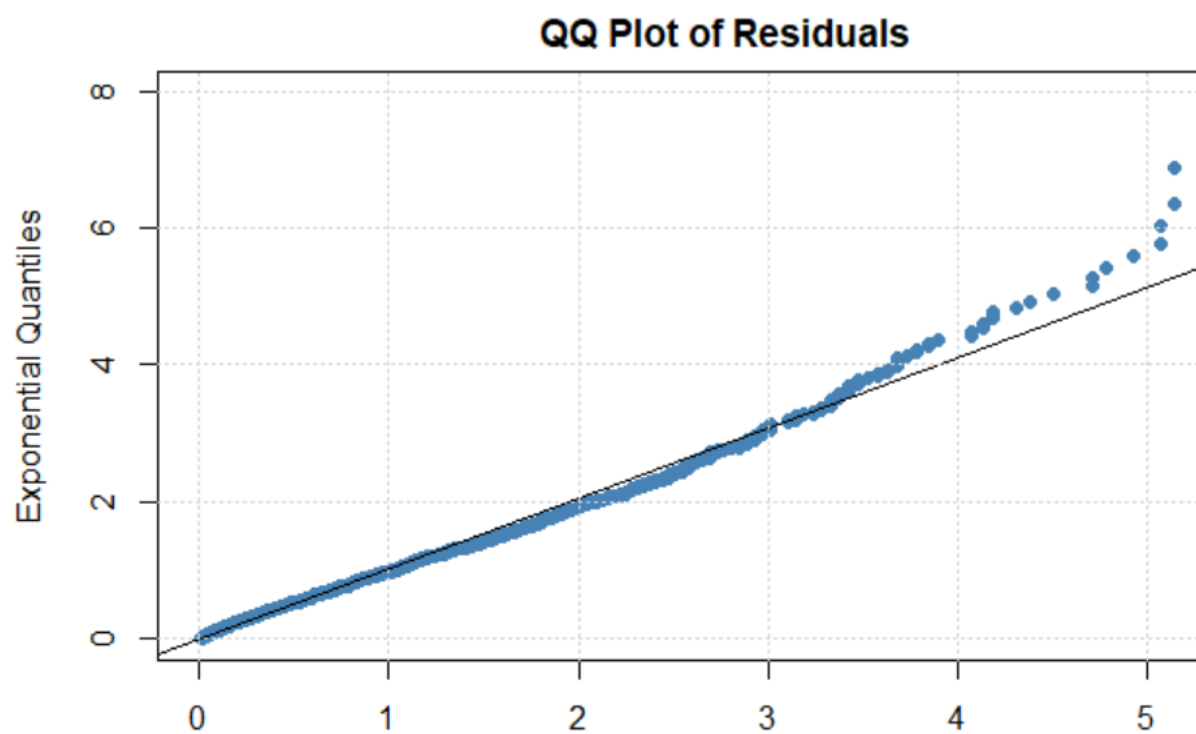
- Dobroć dopasowania na histogramie:



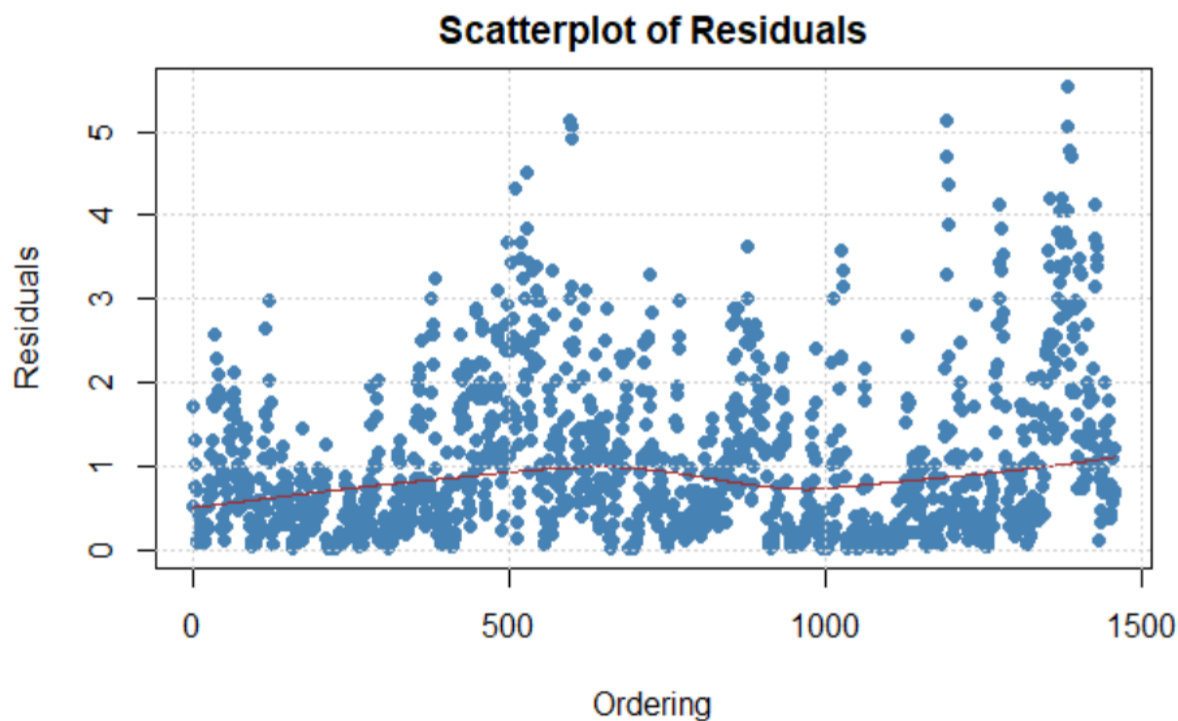
- Wykres maksimów blokowych:



- Wykres kwantyli dopasowanego rozkładu:



- Wykres punktowy:



Wyznaczono poziomy zwrotu x_{20} :

Poziomu zwrotu x_{20}
6.602865

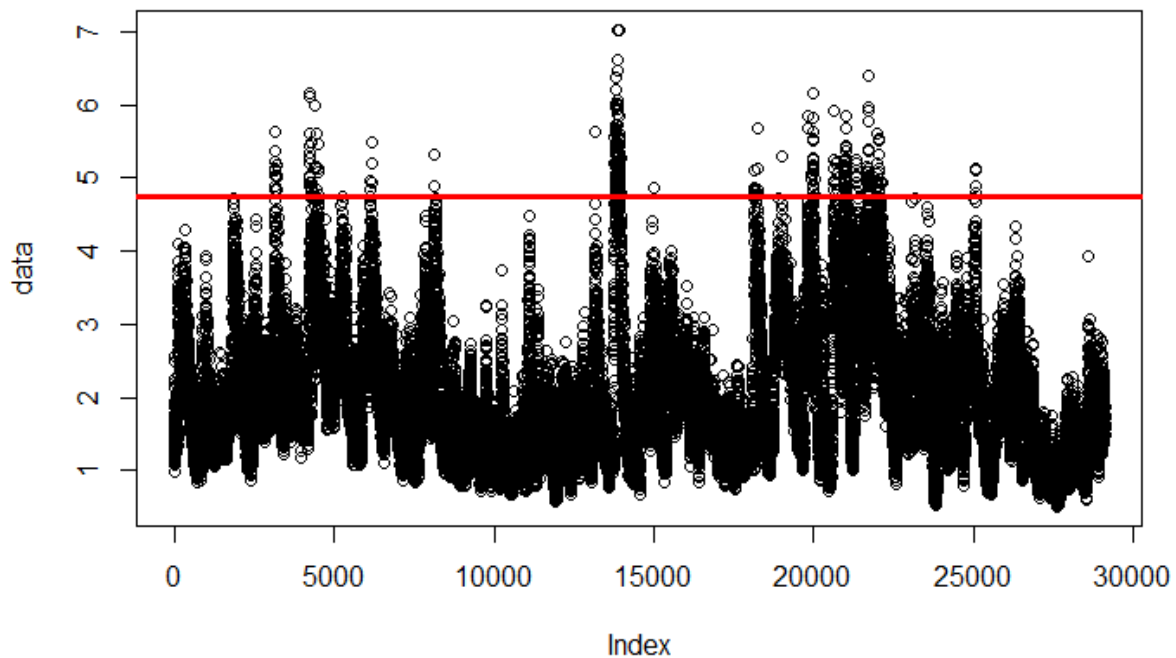
Poziom zwrotu obliczony metodą maksimów blokowych dla 20-stu bloków, z czego każdy blok zawiera średnio $30 \cdot 47$ rekordów (47 spisów dziennie przez 30 dni) wynosi 6.602865. Oznacza to, że średnio raz na miesiąc można spodziewać się, że wysokość fal przekroczy 6,6 metra.

3. Analiza metodą przekroczeń progu (POT). Kwantyl 95% $u = 3.8$ co oznacza, że 5% mierzonych fal przekraczało wysokość 3.8 m. Ustalono próg na tym poziomie i wyestymowano parametry rozkładu GPD dla nadwyżek nad ten próg:

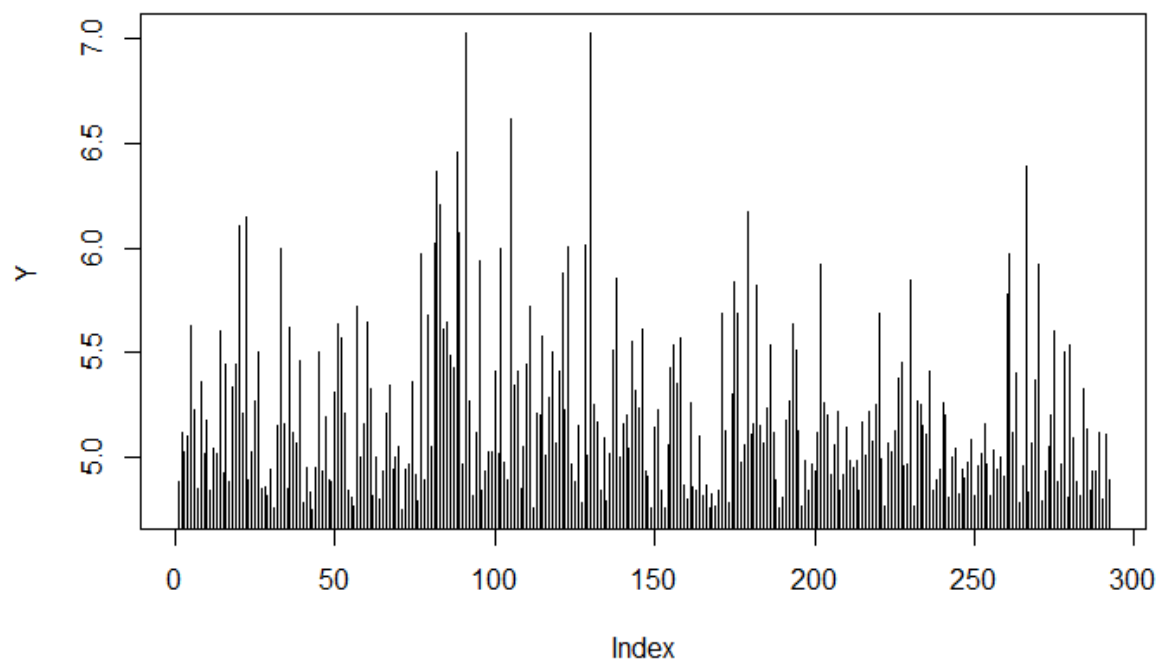
	xi	beta
Parametry rozkładu GPD	-0.1663637	0.7715406

Przeprowadzono analizę oceniającą dobroć dopasowania za pomocą następujących wykresów diagnostycznych:

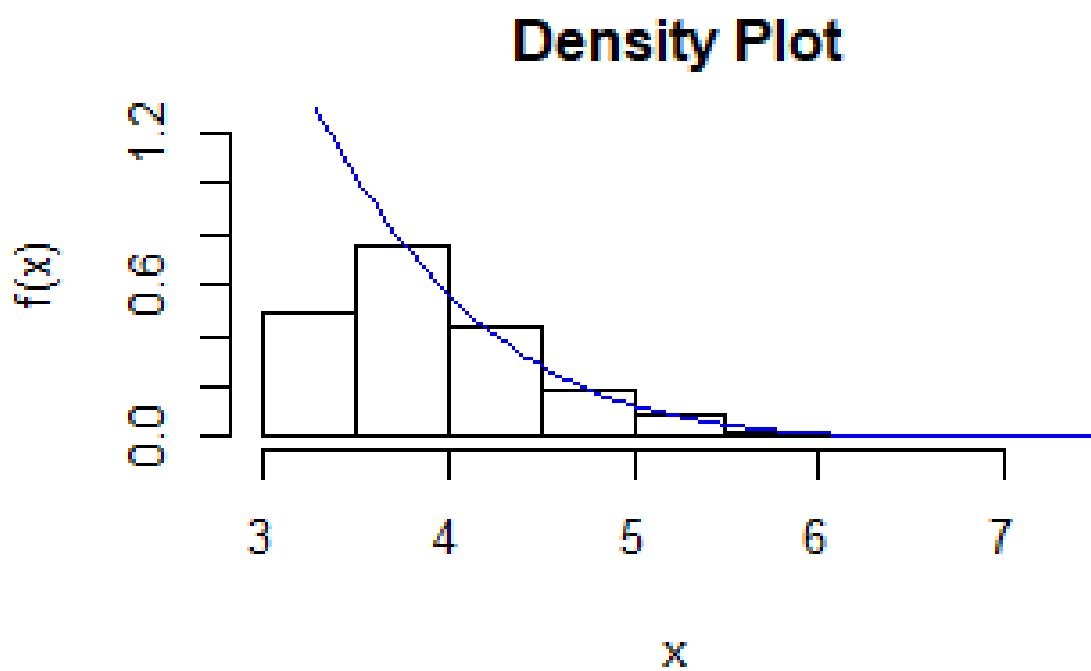
- Wykres rozrzutu z zaznaczonym progiem:



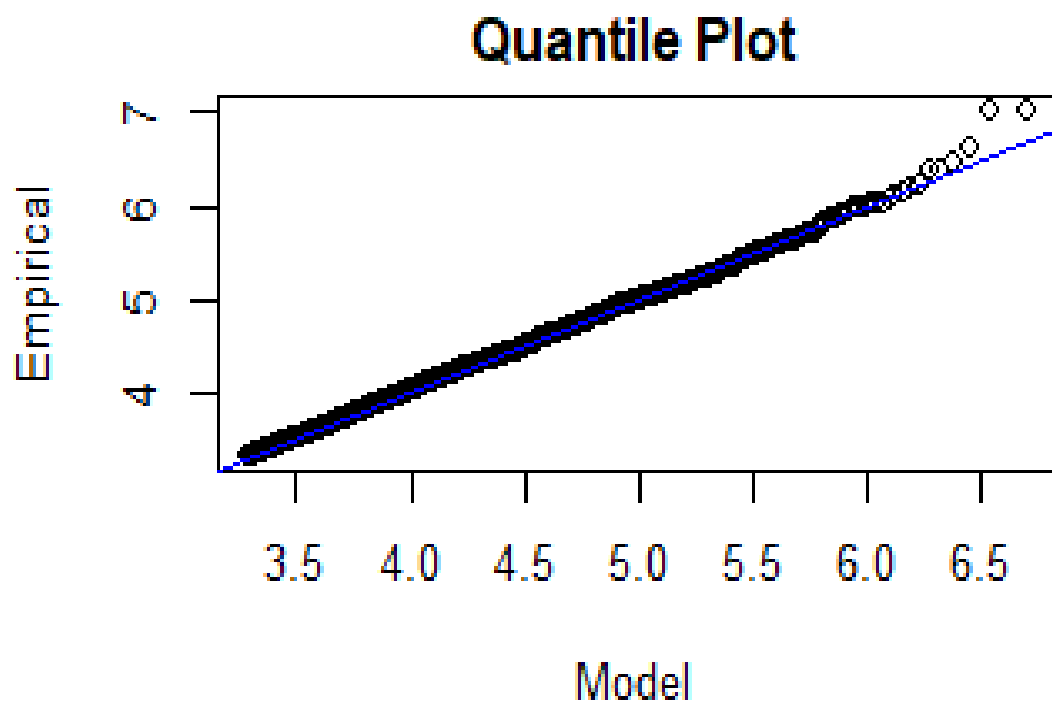
- Wykres nadwyżek nad próg 99%:



- Dobroć dopasowania na histogramie:



- Wykres kwantyli dopasowanego rozkładu:



Wyznaczono poziom zwrotu x_{20} :

Poziomu zwrotu x_{20}
6.689767

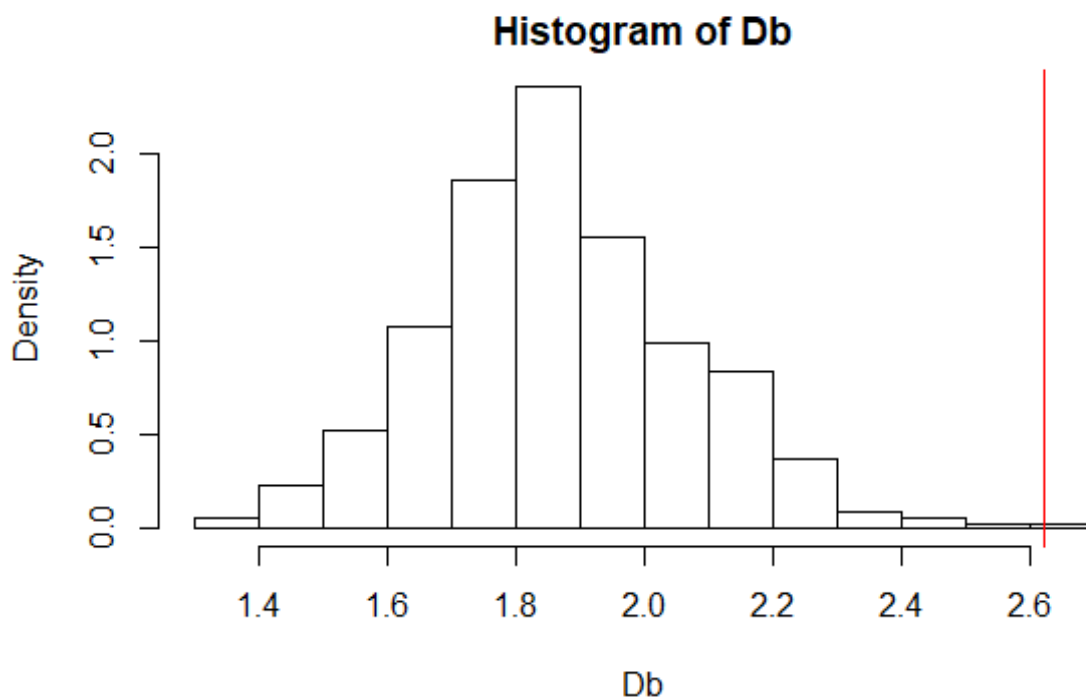
Poziom zwrotu obliczony metodą przekroczeń progu wynosi 6.689767. Oznacza to, że średnio raz na miesiąc wysokość fal będzie miała taką wartość lub ją przekroczy.

4. Analiza metodą Bootstrap. Wygenerowano 10000 prób bootstrapowych. Każda próba jest średnią, medianą lub wartością maksymalną puli 28 danych.

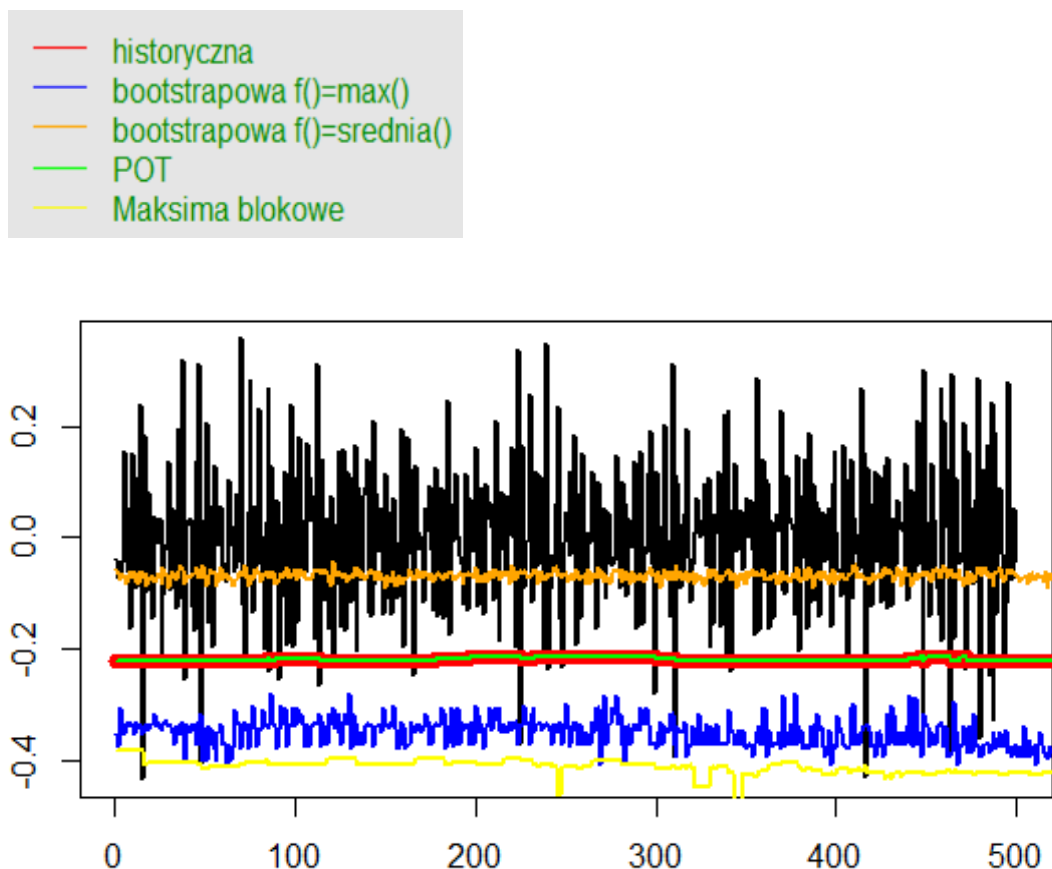
	Kwantyl oryginalnych danych	Bootsrap: funkcja $f()=\max()$	$f()=\text{średnia}()$	$f()=\text{mediana}()$
x_{20}	3.8[m]	5.615[m]	2.20525[m]	2.322766[m]

Wniosek: Przy wyznaczaniu wartości ekstremalnych metodą bootstrapową wybranie średniej lub mediany jako funkcję konwertującą dane jest wyjątkowo nieoptymalne- wyniki są gorsze niż przed generacją próbek. Dobrze sprawdza się natomiast funkcja $\max()$, której wynik przekracza kwantyl oryginalnych danych.

- Histogram próbki bootstrapowej (z funkcją wyliczającą średnie) z zaznaczonym kwantylem:



5. Backtesting. Backtesting przeprowadzono w pętli 500 iteracji testując kolejne przedziały 1000 danych. Przetestowano dane metodą historyczną, bootstrapową z funkcjami max i średnich, metodą POT przekroczeń progu oraz maksimów blokowych.



6. Podsumowanie. Analizując powyższe wykresy oraz obliczenia można zauważyć, że analiza metodą przekroczeń progów osiąga większy poziom zwrotu niż metoda przekroczeń progu (tabela poniżej). Metoda bootstrapowa zwracająca średnie wartości przedziałów danych w tym przypadku jest nieefektywna- uśredniając traci się dane ekstremalne. Stąd wniosek, że w przypadku analizy wysokości fal metoda przekroczeń progu (POT) jest najefektywniejsza.

	Metoda maksimów blokowych	Przekroczeń progu	Bootstrap f=średnia()
Poziom zwrotu x_{20}	6.602865[m]	6.689767[m]	2.20525[m]
Różnica	8,69[cm]		

Źródła danych:

- <https://www.kaggle.com/jolasa/waves-measuring-buoys-data-mooloolaba>