

# PENENTUAN NILAI AMBANG BATAS SEBARAN PARETO TERAMPAT DENGAN *MEASURE OF SURPRISE*\*

Yumna Karimah<sup>1</sup>, Aji Hamim Wigena<sup>2</sup>, Agus M Soleh<sup>3‡</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia, yumna\_karimah@apps.ipb.ac.id

<sup>2</sup>Department of Statistics, IPB University, Indonesia, ajiwigena@gmail.com

<sup>3</sup>Department of Statistics, IPB University, Indonesia, agusms@apps.ipb.ac.id

‡corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)**

**Vol 3 No 2 (2019), 161 - 168**

Copyright © 2019 Yumna Karimah, Aji Hamim Wigena, Agus M Soleh. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Extreme rainfall can result in natural disasters such as floods and landslides. These natural disasters will cause damage and losses to the surrounding environment. Prevention of damage from natural disasters can be done by extreme rainfall estimation. Estimates of extreme rainfall are based on Generalized Pareto Distribution (GPD) which requires threshold value information. The threshold value can be determined by two methods, namely Measure of Surprise (MOS) and Mean Residual Life Plot (MRLP). The purpose of this study is to determine and compare the threshold values of MRLP and MOS. The data used are 10-day and monthly rainfall data. The results of this study indicate that the procedure of MOS is shorter and easier than that of MRLP. Based on the cross validation result, the log-likelihood value of MOS is larger than that of MRLP, then MOS is better than MRLP.

**Keywords:** ambang batas, GPD, mean residual life plot, measure of surprise, sebaran pareto terampat.

---

\*Received Mei 2019; Accepted Jun 2019; Published online on Jun 2019

## 1 Pendahuluan

Di Indonesia terdapat dua pola musim yaitu musim kemarau dan musim hujan. Pada musim hujan, curah hujan akan cenderung lebih tinggi dibandingkan musim kemarau. Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), musim hujan ditetapkan berdasarkan jumlah curah hujan dalam satu dasarian (10 hari) sama atau lebih dari 50 mm dan diikuti oleh beberapa dasarian berikutnya. Tingginya curah hujan pada musim hujan dapat menghasilkan curah hujan ekstrem. Terjadinya curah hujan ekstrem dapat diketahui dengan menentukan nilai ambang batas curah hujan ekstrem tersebut.

Perbedaan karakteristik setiap daerah akan menyebabkan curah hujan di setiap daerah berbeda. Curah hujan yang berbeda menyebabkan ambang batas curah hujan ekstrem juga berbeda. Terjadinya curah hujan ekstrem akan berdampak bagi masing-masing daerah. Dampak yang ditimbulkan dapat berupa banjir dan longsor. Oleh karena itu, diperlukan informasi awal mengenai ambang batas curah hujan ekstrem agar dapat mewaspadaai bencana alam yang akan terjadi.

Penentuan ambang batas curah hujan ekstrem secara statistika dapat menggunakan Teori Nilai Ekstrem (*Extreme Value Theory*, EVT). EVT adalah sebuah teori untuk memodelkan ekor dari suatu sebaran (Vicari, 2010). Penentuan ambang batas curah hujan ekstrem dibagi menjadi dua yaitu *Block Maxima* dan *Peaks Over Threshold* (POT). *Block Maxima* adalah pengambilan nilai-nilai maksimum dari setiap periode waktu atau blok sebagai nilai ekstrem, sedangkan POT adalah pengambilan nilai-nilai yang melampaui ambang batas (*threshold*) sebagai nilai ekstrem. *Block Maxima* didekati dengan *Generalized Extreme Value* (GEV) dan POT didekati dengan sebaran pareto terampat/*Generalized Pareto Distribution* (GPD).

Metode penentuan ambang batas pada POT menggunakan metode grafik (Coles et al., 2001), yaitu menggunakan *Mean Residual Life Plot* (MRLP) dan *Threshold Choice* (TC). Selain itu, nilai persentil ke- $n$  ( $P_n$ ) dapat digunakan untuk menentukan ambang batas sebagai alternatif. Pada penelitian Widyanto (2013) dan Soleh (2015), penentuan ambang batas curah hujan ekstrem dengan persentil 90, MRLP, dan TC. Penentuan ambang batas pada POT dari ketiga metode tersebut memiliki kelemahan yaitu penentuan dilakukan secara visual sehingga tergantung subjektivitas peneliti. Manurung (2018) menggunakan metode *Measure of Surprise* (MOS) sebagai alternatif untuk mengatasi kelemahan metode sebelumnya. Menurut Weaver (1948) metode MOS digunakan untuk menentukan nilai ambang batas pada data ekstrem karena sifatnya yang lebih objektif.

Penelitian ini menggunakan metode MOS dan MRLP untuk menentukan nilai ambang batas curah hujan ekstrem. Data yang digunakan adalah data dasarian pada stasiun Pondok Betung. Ambang batas dari kedua metode tersebut akan dibandingkan untuk menentukan yang terbaik.

## 2 Metodologi

### 2.1 Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari BMKG. Data BMKG berisi informasi curah hujan pada periode 1977-2009 di stasiun Pondok Betung, Jakarta Selatan. Data curah hujan berupa data dasarian dengan jumlah amatan sebanyak 1206 amatan. Satuan curah hujan adalah milimeter.

## 2.2 Metode Penelitian

Prosedur analisis data penentuan nilai ambang batas curah hujan sebagai berikut :

1. Eksplorasi data curah hujan menggunakan *histogram* dan *density plot*.
2. Penentuan ambang batas dengan dua metode, yaitu :

(a) *Measure of Surprise*

Dalam Manurung (2018) tahapan untuk menentukan ambang batas dengan MOS yaitu :

1. Menentukan beberapa ambang batas dugaan ( $u_i$ ) dengan  $u_i < \dots < u_s$ .
2. Pada setiap ambang batas dugaan ( $u_i$ ) :
  - i. Gunakan 10000 rantai Markov pada algoritma MCMC Metropolis-Hasting.
  - ii. Menduga parameter GPD pada  $\theta = (\sigma, \xi)$  setiap rantai dengan metode Bayes. Hasil dari butir ini adalah matriks  $2 \times 10000$ .
  - iii. Ambil contoh acak parameter dari rantai yang dihasilkan sebanyak 5000 contoh. Contoh acak yang diambil hanya rantai ke 2501-10000 karena 2500 rantai pertama dianggap sebagai *burn-in period*.
  - iv. Hitung nilai  $p_{post}$ , kemudian akan diperoleh nilai *surprise*.
3. Ulangi butir (2) sebanyak 30 kali sehingga nilai-nilai *surprise* membentuk matriks  $s \times 30$ .
4. Membuat *boxplot* antara setiap ambang batas dugaan ( $u_i$ ) dengan nilai-nilai *surprise*, dan selanjutnya menentukan ambang batas yang paling sesuai.

(b) *Mean Residual Life Plot*

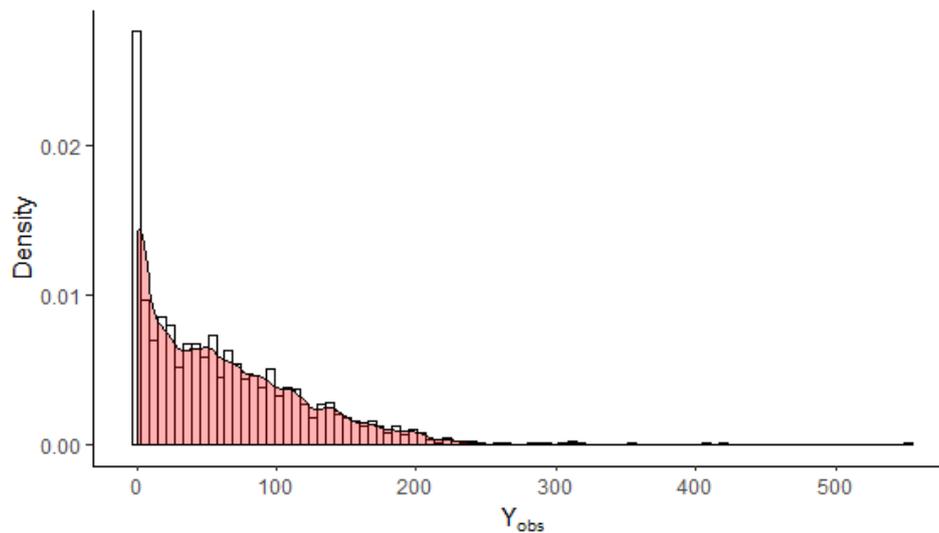
Dalam Coles et al. (2001) tahapan untuk menentukan ambang batas dengan MRLP yaitu :

1. Menentukan selang calon ambang batas dari MRLP untuk digunakan di tahap selanjutnya.
  2. Dengan selang calon ambang batas yang didapatkan dari MRLP, lalu memilih selang calon ambang batas yang lebih kecil dari grafik TC untuk digunakan di tahap selanjutnya dengan melihat kekonstanan parameter pada grafik TC.
  3. Memilih satu calon ambang batas dari ujung selang terkecil, lalu periksa jumlah sisa data yang melampaui.
  4. Apabila jumlah sisa data yang melampaui itu cukup (tidak terlalu banyak dan sedikit), maka dilanjutkan dengan memeriksa sebaran sisa data yang melampaui calon ambang batas tersebut dengan *probability* dan *quantile plot* serta uji formal yaitu uji *Kolmogorov-Smirnov*. Jika sisa data dari ambang batas tersebut sudah menyebar GPD, maka calon ambang batas tersebut digunakan sebagai ambang batas curah hujan ekstrem.
3. Membandingkan hasil pendugaan ambang batas metode MOS dan MRLP dengan validasi silang, lalu menentukan ambang batas terbaik.

### 3 Hasil dan Pembahasan

#### 3.1 Eksplorasi Data

Eksplorasi data curah hujan dasarian pada stasiun Pondok Betung dengan *histogram* dan *density plot*. Gambar 1 menunjukkan *histogram* dan *density plot* dari data curah hujan dasarian di Pondok Betung. *Histogram* tersebut menjulur ke kanan, sehingga dapat dikatakan terdapat nilai-nilai curah hujan yang ekstrem.



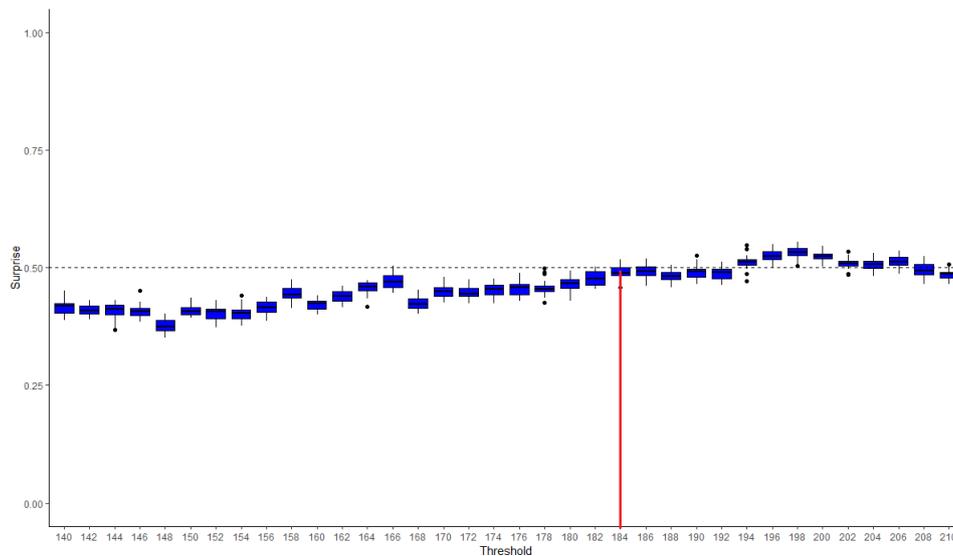
Gambar 1: Histogram curah hujan dasarian periode 1977-2009

Data dasarian menunjukkan titik-titik yang dapat dicurigai sebagai permulaan GPD yaitu nilai puncak global dan lokal dari sebaran data. Nilai puncak global adalah nilai tertinggi pada sebaran peluang sedangkan nilai puncak lokal adalah nilai yang lebih besar daripada nilai sebelum dan sesudahnya secara relatif di lokasi tertentu pada sebaran peluang (Manurung (2018)). Puncak global diperkirakan pada nilai 50 mm, lalu puncak lokal terdapat pada nilai 140 mm. Nilai-nilai tersebut berpotensi menjadi nilai ambang batas.

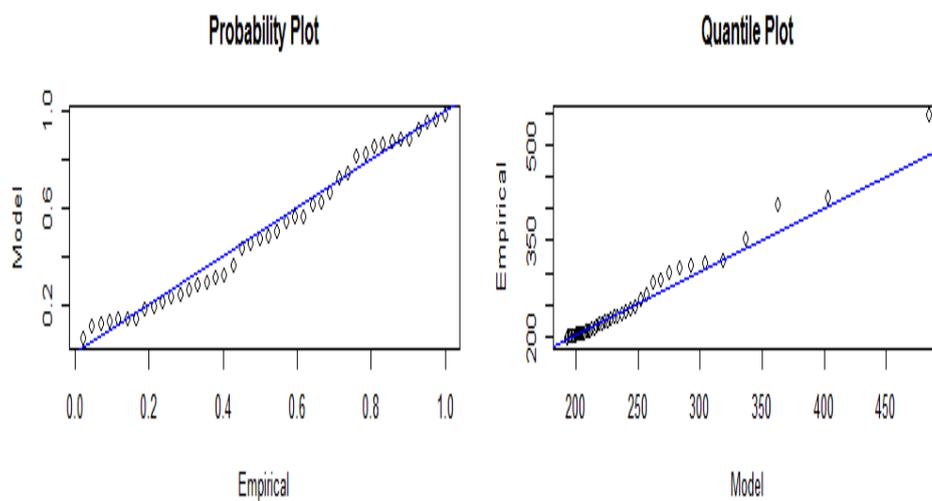
#### 3.2 Penentuan Ambang Batas dengan Metode MOS

Selang calon  $u$  yang digunakan adalah 0 hingga 220. Curah hujan ekstrem diperkirakan akan dimulai dari 140, sehingga selang awal diperkecil menjadi 140 hingga 220 dengan interval 2. Pada  $u = 140$  terdapat 135 sisa amatan, pada  $u = 220$  terdapat 22 sisa amatan yang melampaui  $u$  tersebut. Nilai  $u$  dapat ditentukan dengan *boxplot* dan garis putus-putus yang menunjukkan batas terendah saat *surprise* akan mulai stabil.

Gambar 2 menunjukkan *boxplot* pada  $u = 184$  telah menyentuh garis 0.5. *Boxplot* setelah  $u = 184$  menunjukkan pola yang cenderung konstan naik. Gambar 3 menunjukkan bahwa sisa data setelah  $u = 184$  sudah menyebar GPD, maka sisa data pada  $u > u_0$  juga menyebar GPD. Pengujian *Kolmogorov-Smirnov* menghasilkan  $p$ -value yaitu 0.9984, sehingga dapat dikatakan terima  $H_0$ . Dengan demikian, curah hujan yang mencapai dan melampaui  $u = 184$  dapat dikatakan bahwa curah hujan tersebut ekstrem.



Gambar 2: Grafik MOS data curah hujan dasarian



Gambar 3: *Probability* dan *quantile plot* untuk  $u = 184$

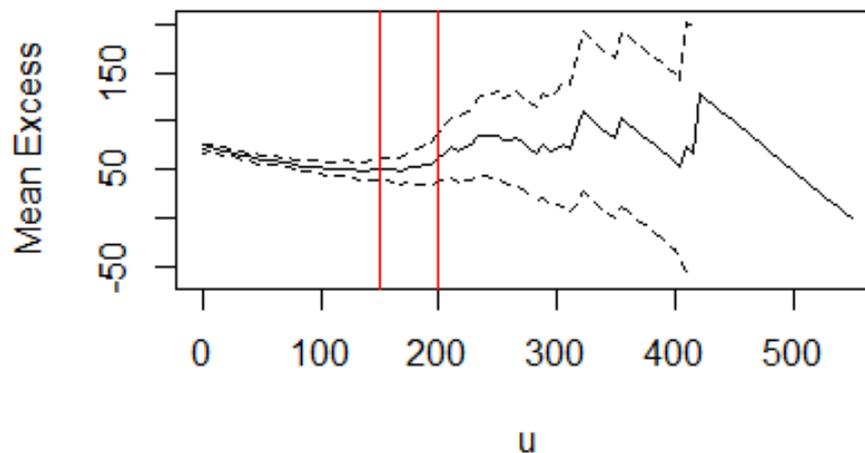
### 3.3 Penentuan Ambang Batas dengan Metode MRLP

Penentuan ambang batas dengan MRLP harus memperhatikan jumlah amatan yang melampaui ambang batas. Apabila memilih ambang batas yang terlalu rendah maka akan menghasilkan pendugaan parameter yang bias. Memilih ambang batas yang terlalu tinggi maka data yang melampaui ambang batas akan sangat sedikit sehingga menghasilkan ragam yang besar.

Penentuan garis linear yang konsisten setelah ambang batas juga harus diperhatikan. Saat menentukan suatu nilai awal ambang batas pada grafik dan setelah nilai ambang batas tersebut konsisten sampai nilai akhir ambang batas dimana setelah nilai tersebut garis menjadi tidak konsisten, maka nilai awal hingga nilai akhir ambang batas tersebut dapat dijadikan selang untuk menentukan satu nilai ambang batas yang akan digunakan.

Gambar 4 menunjukkan bahwa pada  $u = 0$  hingga  $u = 140$  grafik diawali dengan

garis yang linear. Pada  $u = 150$  hingga  $u = 200$  garis cenderung konsisten. Pada  $u = 150$  terdapat 109 sisa amatan, dan  $u = 200$  terdapat 35 sisa amatan yang melampaui. Jumlah amatan tersebut cukup untuk melakukan pendugaan parameter GPD.  $u = 150$  hingga  $u = 200$  adalah selang yang lebih kecil dan dapat digunakan dalam penentuan ambang batas selanjutnya.



Gambar 4: Grafik MRLP data curah hujan dasarian

Selanjutnya melakukan pemilihan  $u$  menggunakan TC. Pada TC akan disajikan dua buah grafik yaitu grafik  $u$  dengan reparameterisasi parameter skala  $\sigma^*$  (modified scale), dan dugaan parameter bentuk  $\hat{\xi}$  (shape). TC berguna untuk memperkecil selang. Dengan menggunakan selang yang lebih kecil yang didapatkan dari MRLP, maka didapatkan selang  $u$  yang lebih kecil lagi untuk menentukan  $u$ . Penentuan  $u$  dapat diketahui dengan memeriksa kekonstanan kedua grafik parameter tersebut.

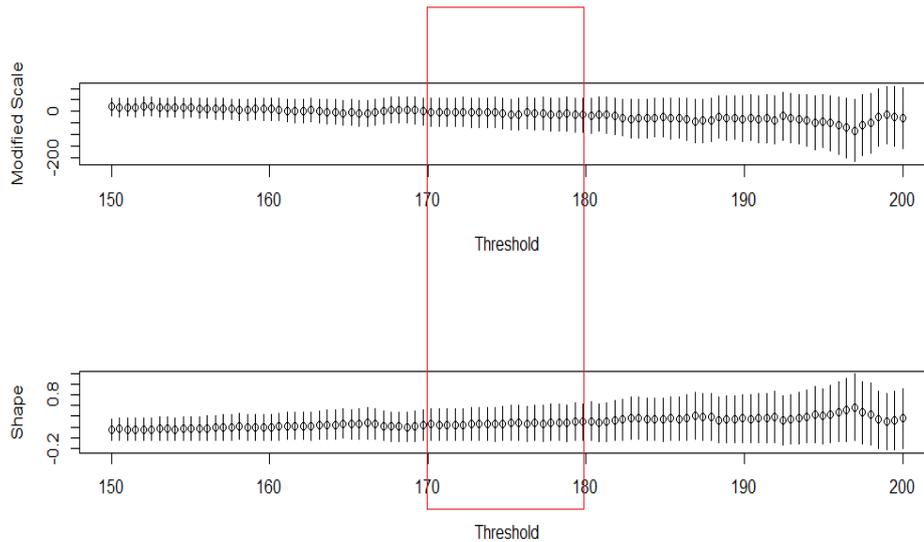
Gambar 5 menunjukkan pada  $u = 170$  hingga  $u = 180$  memiliki kekonstanan yang hampir sama. Langkah selanjutnya dengan selang yang lebih kecil tersebut akan diperiksa sebaran data yang melampaui  $u$  yang ditentukan menyebar GPD atau tidak. Jika sisa data pada  $u_0$  sudah menyebar GPD, maka sisa data pada  $u > u_0$  akan menyebar GPD juga.

Pada selang  $u = 170$  hingga  $u = 180$  dapat ditentukan bahwa  $u_0 = 170$ . Pemeriksaan  $u_0 = 170$  menyebar GPD atau tidak dengan menggunakan *probability* dan *quantile plot*. Selanjutnya, melakukan uji formal dengan uji *Kolmogorov-Smirnov*. Uji ini memiliki hipotesis nol yaitu  $F(x) = F_0(x)$ .  $F(x)$  merupakan sebaran empirik sedangkan  $F_0(x)$  merupakan sebaran teoritis yaitu GPD.

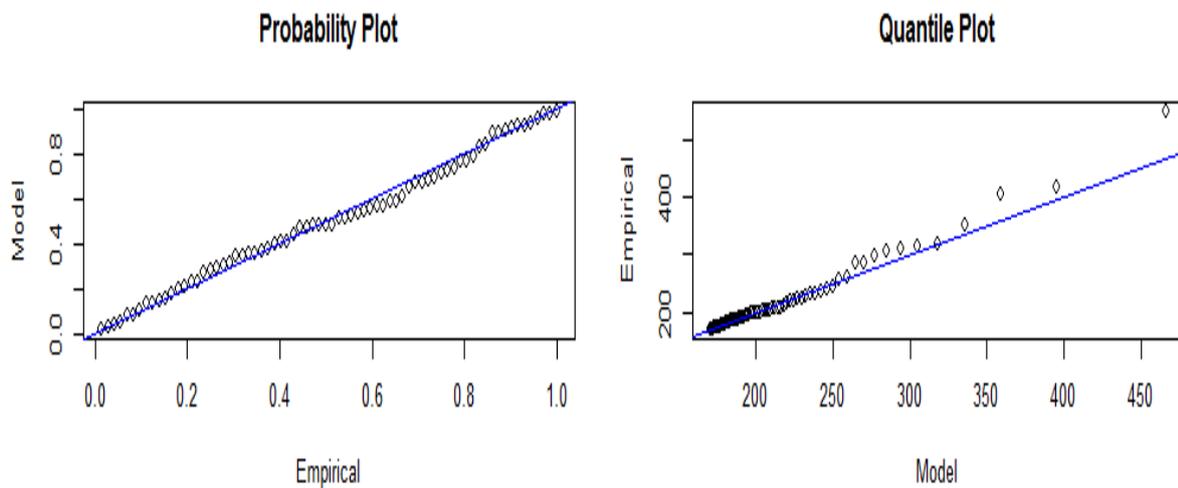
Gambar 6 menunjukkan bahwa sisa data setelah  $u_0 = 170$  sudah menyebar GPD, maka sisa data pada  $u > u_0$  juga menyebar GPD. Pengujian *Kolmogorov-Smirnov* menghasilkan *p-value* yaitu 0.9951, sehingga dapat dikatakan terima  $H_0$ . Dari metode MRLP, dihasilkan nilai ambang batas yang akan digunakan adalah 170. Jika nilai curah hujan mencapai dan melampaui  $u = 170$ , maka dapat dikatakan curah hujan tersebut ekstrem.

### 3.4 Perbandingan antara Metode MOS dan MRLP

Perbedaan metode MOS dan MRLP dapat dilihat dari proses menentukan nilai  $u$ . Prosedur MOS lebih singkat dibandingkan MRLP. Penentuan selang  $u$  dengan metode MOS dapat dikatakan juga bersifat subjektif, tapi penentuan  $u$  yang akan digunakan



Gambar 5: Grafik TC untuk  $u$  150 hingga 200



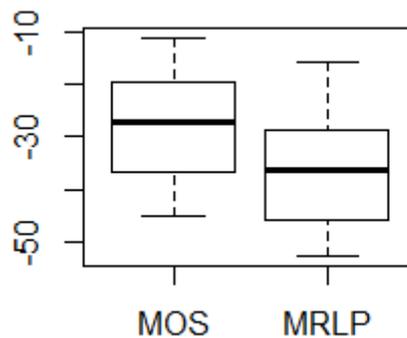
Gambar 6: *Probability* dan *quantile plot* untuk  $u = 170$

lebih mudah dan dapat dikatakan lebih objektif. Jika *boxplot* yang sudah mencapai dan melebihi garis putus-putus pada  $u$  tertentu, maka calon  $u$  tersebut dapat dikatakan sebagai ambang batas.

Pada metode MRLP, penentuan calon-calon  $u$  pada grafik-grafik yang digunakan cukup sulit untuk ditentukan. Proses penentuan pada MRLP sangat subjektif dan cukup sulit, dilanjutkan penentuan selang  $u$  dengan grafik TC dan melakukan pemeriksaan GPD pada  $u$  tersebut. Kedua metode tersebut menghasilkan nilai  $u$  yang berbeda.

Pemeriksaan  $u$  yang lebih baik menggunakan nilai *log-likelihood* dari hasil validasi silang. Gambar 7 menunjukkan *boxplot* dari MOS letaknya lebih tinggi dibandingkan *boxplot* MRLP. Demikian juga hasil nilai *log-likelihood* dari validasi silang didapatkan nilai untuk  $u = 184$  yaitu -26.733 dan  $u = 170$  yaitu -35.631. Dengan demikian, metode

MOS dapat dikatakan lebih baik dibandingkan dengan MRLP karena nilai *log-likelihood* MOS lebih besar dan mendekati nol.



Gambar 7: *Boxplot* nilai *log-likelihood* MOS dan MRLP data dasarian

## 4 Simpulan

Ambang batas curah hujan ekstrem pada Stasiun Pondok Betung dengan metode MOS dan MRLP masing-masing yaitu 184 dan 170. Prosedur MOS lebih singkat dan lebih mudah menentukan ambang batas yang digunakan karena lebih objektif. Hasil validasi silang menunjukkan bahwa nilai *log-likelihood* MOS lebih besar daripada MRLP sehingga MOS lebih baik dari MRLP.

## Daftar Pustaka

- Coles, S., Bawa, J., Trenner, L. and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, Vol. 208, Springer.
- Manurung, A. (2018). *Pendugaan ambang batas bayes pada sebaran pareto terampat dengan measure of surprise [Tesis]*, Institut Pertanian Bogor, Bogor(ID).
- Soleh, A. M. (2015). *Pemodelan linier sebaran gamma dan pareto terampat dengan regularisasi L1 pada statistical down scaling untuk pendugaan curah hujan bulanan. [Disertasi]*, Institut Pertanian Bogor, Bogor(ID).
- Vicari, I. (2010). Measures of surprise and threshold selection in extreme value statistics, *Technical report*.
- Weaver, W. (1948). Probability, rarity, interest, and surprise, *The Scientific Monthly* **67**(6): 390–392.
- Widyanto, I. (2013). *Penentuan ambang batas curah hujan ekstrim dengan mean residual life dan threshold choice (Studi kasus : curah hujan periode 1977-2010 di stasiun pondok betung) [Skripsi]*, Institut Pertanian Bogor, Bogor(ID).