

Perbandingan Estimasi *Return Level Declustering* dan *Non Declustering* pada Data Curah Hujan Ekstrem di Surabaya dan Mojokerto

Ainun Umami dan Sutikno

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data,

Institut Teknologi Sepuluh Nopember (ITS)

e-mail: sutikno@statistika.its.ac.id

Abstrak—Surabaya merupakan wilayah perkotaan yang rentan mengalami curah hujan tinggi sehingga sering terjadi banjir. Mojokerto yang dari tahun ke tahun mengalami peralihan fungsi lahan juga sering terjadi banjir. Untuk meminimalisir kerugian banjir, dibutuhkan pengetahuan terkait kejadian ekstrem menggunakan *Extreme Value Theory* (EVT). Terdapat dua pendekatan dalam EVT, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT). POT merupakan salah satu metode untuk mengidentifikasi nilai ekstrem dengan menggunakan nilai acuan yang disebut dengan *threshold*. Salah satu syarat EVT yaitu data harus independen. Namun pada kenyataannya data ekstrem seringkali bersifat dependen sehingga perlu dilakukan penanganan adanya dependensi data dengan menggunakan *Declustering*. Yang menarik pada EVT yaitu adanya *return level*. *Return level* merupakan nilai maksimum yang diharapkan akan terlampaui satu kali dalam jangka waktu tertentu. Pada penelitian ini akan dilakukan perbandingan estimasi *return level Declustering* dan POT (*Non Declustering*) pada data curah hujan ekstrem. Jumlah pos pengamatan yang digunakan sebanyak dua pos pengamatan di Surabaya dan lima pos pengamatan Mojokerto. Hasil penelitian ini menunjukkan bahwa di Surabaya metode yang lebih baik yaitu *Declustering*, sedangkan di Mojokerto metode yang lebih baik yaitu *Non Declustering*.

Kata Kunci—*Declustering*, *Mojokerto*, *Peaks Over Threshold*, *Return Level*, *Surabaya*.

I. PENDAHULUAN

SURABAYA merupakan wilayah perkotaan yang memiliki konstruksi material penyimpan panas sehingga dapat mengakibatkan terjadinya kenaikan suhu. Akibatnya rentan mengalami curah hujan yang tinggi sehingga dapat terjadi banjir. Berita yang disampaikan [1] mengatakan bahwa curah hujan dengan intensitas 110 mm mengguyur Surabaya. Intensitas curah hujan semakin tinggi terjadi pada tahun berikutnya yaitu sebesar 124 mm sehingga Surabaya kembali dikepung banjir. Tidak hanya di Surabaya, banjir juga terjadi di Mojokerto yang notabene dari tahun ke tahun mengalami peralihan fungsi lahan. Kondisi topografi di Mojokerto cenderung cekung ditengah-tengah dan tinggi dibagian Selatan dan Utara sehingga tidak menutup kemungkinan jika sering terjadi banjir. Tata kelola lahan yang buruk, karakteristik wilayah dan juga curah hujan yang cenderung meningkat dapat menyebabkan terjadinya banjir. Oleh karena itu, dibutuhkan pengetahuan terkait nilai-nilai ekstrem untuk meminimalkan dampak kerugian akibat curah hujan ekstrem dengan

menggunakan *Extreme Value Theory* (EVT). Terdapat dua pendekatan pada EVT, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT). Merujuk pada [2] menjelaskan bahwa penggunaan POT lebih efisien dibandingkan dengan pendekatan BM.

Salah satu syarat pada EVT yaitu urutan data harus independen. Namun pada kenyataannya data ekstrem seringkali bersifat dependen karena kejadian di alam cenderung terjadi berurutan. Oleh karena itu diperlukan penyelesaian yang dapat mengatasi dependensi data dengan menggunakan metode *Declustering*. Penelitian mengenai *Declustering* pernah dilakukan oleh [3] yang meneliti mengenai *Runs Declustering* dan menghasilkan kesimpulan bahwa proses *Declustering* pada data curah hujan ekstrem dependen dapat mengatasi dependensi data. Hal yang penting dalam EVT adalah *return level*. *Return level* diperoleh berdasarkan estimasi parameter. Pengertian dari *return level* merupakan nilai maksimum yang diharapkan akan terlampaui satu kali dalam jangka waktu tertentu. Informasi dari *return level* data curah hujan dapat digunakan untuk mengantisipasi adanya banjir.

Berdasarkan uraian yang telah dipaparkan, maka penelitian ini menganalisis data curah hujan menggunakan *Extreme Value Theory* dengan pendekatan *Peaks Over Threshold* untuk data curah hujan ekstrem di Surabaya dan Mojokerto. Selain itu, penelitian ini juga mengatasi dependensi data menggunakan *Declustering*. Diharapkan hasil penelitian ini dapat memberikan tambahan informasi untuk *climate early warning system* kepada masyarakat dan instansi yang terkait.

II. TINJAUAN PUSTAKA

A. *Extreme Value Theory* (EVT)

Extreme Value Theory (EVT) merupakan salah satu metode statistika untuk mempelajari perilaku ekor (*tail*) distribusi dari data ekstrem sehingga dapat menentukan probabilitas nilai-nilai ekstremnya [4]. Terdapat dua pendekatan untuk mengidentifikasi pergerakan nilai ekstrem, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT) [5].

B. *Peaks Over Threshold*

Peaks Over Threshold (POT) merupakan salah satu metode untuk mengidentifikasi nilai ekstrem dengan menggunakan acuan yang disebut *threshold* (u). Data yang berada di atas *threshold* akan diidentifikasi sebagai nilai ekstrem. POT

menghasilkan distribusi nilai ekstrem berupa *Generalized Pareto Distribution* (GPD). *Probability Density Function* (PDF) untuk GPD dapat ditulis seperti persamaan (1).

$$f_{GPD}(x|\xi, u, \sigma) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{(x-u)}{\sigma}\right), & \xi = 0 \end{cases} \quad (1)$$

Adapun *Cumulative Distribution Function* (CDF) dari GPD adalah sebagai berikut.

$$F_{GPD}(x|\xi, u, \sigma) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{(x-u)}{\sigma}\right), & \xi = 0 \end{cases} \quad (2)$$

C. Penentuan Nilai Threshold

Metode penentuan nilai *threshold* yang mudah dan sering digunakan adalah metode persentase. Tahapan dalam menentukan nilai *threshold* dengan menggunakan persentase didapatkan dengan cara sebagai berikut [6].

1. Mengurutkan data dari yang terbesar hingga yang terkecil.
2. Menghitung 1 persen dari jumlah data ekstrem

$$k = 1\% \times N \quad (3)$$

3. Menentukan nilai *threshold* (u), yaitu data yang berada pada urutan ke $-(k+1)$

D. Declustering

Sejumlah teknik *declustering* telah dikembangkan dengan prosedur yang mengarah pada identifikasi nilai maksimum pada tiap *cluster* dan diasumsikan identik independen (iid). *Declustering* bekerja dengan menentukan nilai *threshold* (u) kemudian dilakukan pembentukan *cluster* hingga mendapatkan sejumlah r (*Run Length*) pengamatan berada di bawah *threshold*. Prosedur *Declustering* pada pendekatan *Runs Declustering* adalah sebagai berikut.

1. Menentukan *threshold* (u) sehingga nilai-nilai yang melebihi u dianggap sebagai peristiwa ekstrem dan menentukan suatu nilai *run length* (r).
2. Menentukan *cluster* yang dimulai ketika *threshold* (u) terlampaui untuk pertama kalinya dan berakhir setelah paling sedikit r pengamatan berturut-turut jauh dibawah *threshold*.
3. Mengekstrak nilai maksimum *cluster* dan melanjutkan mengidentifikasi *cluster* berikutnya. Prosedur berhenti bila data habis

Nilai r yang terlalu rendah menyebabkan data sulit menjadi dependen, sedangkan nilai r yang terlalu tinggi data ekstrem yang diperoleh sedikit sehingga menghasilkan varians yang besar. Penentuan r yang optimal dapat menggunakan *extremal indeks* yang dituliskan pada persamaan (4).

$$q = \begin{cases} \min\left\{1, \frac{2 \sum_{i=1}^{N-1} \frac{\sigma_i^2}{T_i} a_i^{N-1}}{(N-1) \sum_{i=1}^{N-1} T_i^2}\right\} & \text{if } \max\{T_i : 1 \leq i \leq N-1\} \leq 2 \\ \min\left\{1, \frac{2 \sum_{i=1}^{N-1} \frac{\sigma_i^2}{(T_i-1)} a_i^{N-1}}{(N-1) \sum_{i=1}^{N-1} (T_i-1)(T_i-2)}\right\} & \text{if } \max\{T_i : 1 \leq i \leq N-1\} > 2 \end{cases} \quad (4)$$

dimana $T_i = S_{i+1} - S_i \quad i = 1, 2, K, N-1$

T_i : jarak antar pengamatan di atas *threshold*

N : jumlah pengamatan di atas *threshold*

S_i : urutan pengamatan ke- i di atas *threshold*

Estimator pada *extremal index* menghasilkan nilai yang disebut dengan *Runs Estimator*. Nilai *extremal index* dapat digunakan untuk menentukan r yang optimal, dimana $r = T_{(C)}$ dan $(C-1) = q(N-1)$. Pengembangan penelitian dilakukan oleh [7] menghasilkan *intervals estimator*. Ketika dilakukan perbandingan antara *intervals estimator* dan *runs estimator* untuk *extremal index*, menghasilkan kesimpulan bahwa dengan menggunakan *intervals estimator* menghasilkan nilai yang lebih *robust* dibandingkan dengan *runs estimator*. Penelitian [7] menunjukkan fakta bahwa *threshold optimal* untuk *interval estimator* biasanya lebih rendah dibandingkan *runs estimator*. Alasannya karena sensitivitas *intervals estimator* lebih tinggi daripada *runs estimator*, sehingga akan menghasilkan bias yang lebih besar jika dilakukan pada *minimum threshold*. Kesimpulan yang sama juga diperoleh pada *Markov chain* yang hasilnya tidak ditunjukkan pada penelitian [7]. Setiap pilihan r untuk *runs estimator* memiliki kinerja yang lebih unggul jika digunakan untuk *maximum threshold* dibandingkan menggunakan *intervals estimator*.

E. Estimasi Parameter GPD

Estimasi parameter GPD dapat ditaksir menggunakan *Maximum Likelihood Estimation* (MLE). Tahapan estimasi parameter dengan menggunakan MLE yaitu sebagai berikut.

1. Mengambil n sampel random x_1, x_2, K, x_n dengan cara mengambil nilai-nilai yang melebihi *threshold*.
2. Membuat *Probability Density Function* (PDF) seperti pada persamaan (1)
3. Membuat fungsi *likelihood*.
4. Membentuk fungsi \ln *likelihood* untuk distribusi GPD.
5. Memaksimumkan \ln *likelihood* dengan cara mencari turunan pertama terhadap parameter distribusi (ξ, s).
6. Bentuk umum tahap no. 5 disamakan dengan nol. Dan akan didapatkan estimasi parameter GPD.

Berdasarkan persamaan yang terbentuk, didapatkan estimasi parameter bentuk dan skala GPD yang *closed form* sehingga tidak diperlukan analisis numerik untuk mendapatkan nilai estimasi parameter GPD.

F. Pengujian Kesesuaian Distribusi

Pengujian kesesuaian distribusi dilakukan untuk memeriksa kesesuaian distribusi teoritis dengan distribusi empirik dengan

menggunakan uji *Kolmogorov-Smirnov*. Hipotesis uji *Kolmogorov-Smirnov* dituliskan sebagai berikut.

$$H_0 : F_n(x) = F_0(x) \text{ (Data mengikuti distribusi teoritis).}$$

$$H_0 : F_n(x) \neq F_0(x) \text{ (Data tidak mengikuti distribusi teoritis).}$$

$$\alpha = 0,05$$

Statistik uji yang digunakan adalah:

$$D_{hitung} = Maks |F_n(x) - F_0(x)| \tag{5}$$

Daerah penolakan untuk uji *Kolmogorov-Smirnov* adalah H_0 ditolak jika $D_{hitung} > D_\alpha$ pada tabel *Kolmogorov-Smirnov* satu sampel dengan taraf signifikan sebesar α .

G. Auto Correlation Function (ACF)

ACF merupakan metode untuk mengukur dependensi data runtun waktu. ACF dapat melihat korelasi antar lag, melihat pola data yang berbentuk tren maupun periodik, serta mengidentifikasi *white noise*. ACF diidentifikasi sebagai ukuran dependensi linear runtun waktu dari waktu t terhadap waktu $t+h$. ACF dapat ditulis sebagai :

$$r(h) = \frac{cov(x_t, x_{t+h})}{\sqrt{var(x_t)var(x_{t+h})}} \tag{6}$$

Penduga dari ACF adalah:

$$\hat{r}(h) = \frac{\sum_{t=h+1}^T (x_t - \bar{x})(x_{t-h} - \bar{x})}{\sqrt{\sum_{t=1}^t (x_t - \bar{x})^2} \sqrt{\sum_{t=h+1}^t (x_{t-h} - \bar{x})^2}} \tag{7}$$

H. Extremogram

Extremogram merupakan metode yang hanya dapat mengukur dependensi ekstrem untuk satu peubah runtun waktu saja. Secara sederhana bisa dikatakan bahwa konsep *extremogram* adalah menduga peluang terjadinya kejadian ekstrem dalam waktu $(t + h)$ dengan syarat ada kejadian ekstrem pada periode t . Adapun persamaan untuk *extremogram* yaitu sebagai berikut.

$$r_{A,B}(h) = \lim_{x \rightarrow \Psi} P(x^{-1}x_h \hat{I} B | x^{-1}x_0 \hat{I} A), h = 0,1,2,K \tag{8}$$

Penduga *extremogram* dapat ditulis sebagai berikut.

$$\hat{r}_{A,B}(h) = \frac{\sum_{t=1}^{n-h} \{a_m^{-1}x_{t+h} \hat{I} B, a_m^{-1}x_t \hat{I} A\}}{\sum_{t=1}^n I \{a_m^{-1}x_t \hat{I} A\}} \tag{9}$$

I. Return Level

Salah satu hal yang menarik dalam *Extreme Value Theory* (EVT) adalah menentukan *return level*. *Return level* merupakan nilai maksimum pada periode yang akan datang. Persamaan *return level* untuk GPD adalah sebagai berikut

$$RL_T \begin{cases} u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(T\hat{\xi}_u)^{\hat{\xi}} - 1 \right] & , \hat{\xi} \neq 0 \\ u + \hat{\sigma} \log(T\hat{\xi}_u) & , \hat{\xi} = 0 \end{cases} \tag{10}$$

T = periode waktu

$$\hat{\xi}_u = \frac{n_u}{n}$$

n_u = banyaknya pengamatan setiap tahun

Interpretasi nilai *return level* yaitu suatu level kejadian ekstrem yang terlampaui rata-rata sekali dalam suatu periode waktu tertentu dimana probabilitas terjadinya sebesar $\frac{1}{T}$.

J. Kriteria Pemilihan Model yang Sesuai

Kriteria pemilihan model yang sesuai dapat dilakukan dengan *Root Mean Square Error* (RMSE). *Error* merupakan jumlah yang diperkirakan dari kuantitas yang diperkirakan [8]. Persamaan untuk mencari nilai RMSE dapat ditulis sebagai berikut.

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (x_i - \hat{x}_i)^2} \tag{11}$$

dengan :

x_i : Nilai aktual

\hat{x}_i : Nilai dugaan

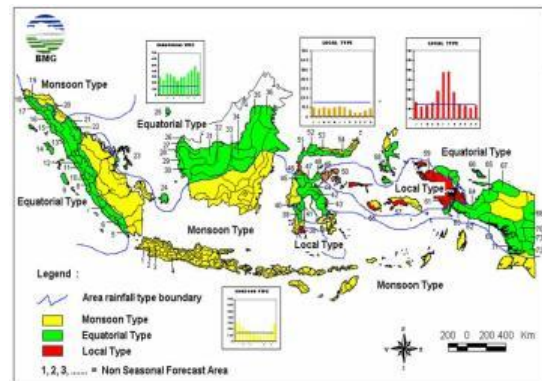
t : Banyaknya nilai yang diduga

K. Curah Hujan

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) diatas permukaan horizontal bila tidak terjadi evaporasi, *runoff*, dan infiltrasi [9]. Curah hujan 1 (satu) milimeter artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi satu milimeter atau tertampung air sebanyak satu liter. Intensitas curah hujan di wilayah Indonesia dibagi menjadi empat kategori sebagai berikut.

- a. Kategori ringan, sebesar 5-20 mm/hari.
- b. Kategori sedang, sebesar 20-50 mm/hari.
- c. Kategori lebat, sebesar 50-100 mm/hari.
- d. Kategori sangat lebat, lebih dari 100 mm/hari.

Berikut merupakan ilustrasi tipe curah hujan di Indonesia.



Gambar 1. Peta tipe hujan di Indonesia.

L. Zona Musim (ZOM)

Zona Musim adalah daerah yang pos hujan rata-ratanya memiliki perbedaan yang jelas antara periode Musim Hujan dan Musim Kemarau. Luas suatu wilayah ZOM tidak selalu sama dengan luas suatu wilayah administrasi pemerintahan. Dengan demikian, satu wilayah ZOM bisa terdiri dari beberapa

kabupaten dan sebaliknya satu wilayah kabupaten bisa terdiri dari beberapa ZOM. Pembagian wilayah ZOM di Jawa Timur akan ditampilkan pada Gambar 2.



Gambar 2. Peta Zona Musim Jawa Timur.

III. METODOLOGI PENELITIAN

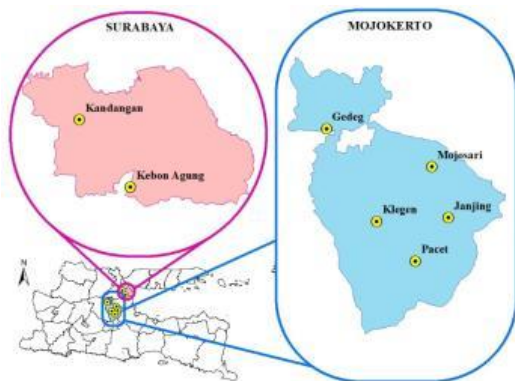
A. Sumber Data

Data penelitian ini adalah data sekunder yang diperoleh dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Stasiun Klimatologi Malang dan Dinas Pekerjaan Umum Sumber Dara Air Provinsi Jawa Timur. Data dari BMKG Stasiun Klimatologi Malang terdiri dari 1 Januari 2004 sampai 31 Desember 2013. Sedangkan data dari Dinas Pekerjaan Umum Sumber Dara Air Provinsi Jawa Timur terdiri dari 1 Januari 1999 sampai 31 Desember 2003 dan 1 Januari 2014 sampai 31 Desember 2018.

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel curah hujan harian. Pos pengamatan yang digunakan pada penelitian ini diwakili oleh satu pos pengamatan pada tiap zona musim di Surabaya dan Mojokerto. Letak dan Nama pos pengamatan selengkapnya disajikan pada Gambar 3.

Data dibagi menjadi dua, yaitu data *training* dan data *testing*. Data *training* digunakan untuk analisis dan data *testing* digunakan untuk validasi model. Data curah hujan harian tahun 2003 sampai 2016 akan digunakan sebagai data *training*, sedangkan untuk validasi digunakan data curah hujan harian pada tahun 2017. Validasi data digunakan sebagai pembandingan hasil nilai *return level* hingga 3 periode ulang.



Gambar 3. Peta Tujuh Stasiun di Surabaya dan Mojokerto.

C. Langkah Analisis

Langkah analisis data yang dilakukan pada penelitian ini dijelaskan sebagai berikut.

1. Melakukan pra pemrosesan data dengan mengidentifikasi *missing value*, *noisy*, dan *inconsistent*.
2. Mengeksplorasi data curah hujan harian.
3. Mengidentifikasi dependensi data menggunakan ACF.
4. Pengambilan data ekstrem menggunakan POT dan *delustering*.
5. Mengidentifikasi kesesuaian GPD menggunakan uji *Kolmogorov-Smirnov* dan mengidentifikasi dependensi data menggunakan *extremogram*.
6. Menaksir parameter GPD *Declustering* dan *Non Declustering* menggunakan *Maximum Likelihood Estimation (MLE)*.
7. Menentukan *Return Level Declustering* dan *Non Declustering*.
8. Membandingkan kebaikan estimasi *Return Level Declustering* dan *Non Declustering* berdasarkan RMSE.
9. Menarik kesimpulan.

IV. HASIL DAN PEMBAHASAN

A. Pra-Pemrosesan Data

Langkah awal dalam analisis data adalah tahap pra-pemrosesan data. Hal ini dilakukan karena dalam pengambilan data curah hujan seringkali melibatkan *missing value* akibat dari kerusakan alat pengukur hujan, pergantian alat, atau karena sebab teknis dan operasional lainnya.

Tabel 1.
Identifikasi *Missing Value* Data Curah Hujan

Pos Pengamatan	<i>Missing Value</i> (%)
Kandangan	1,132
Kebon Agung	1,898
Gedeg	0,000
Janjing	0,548
Klegan	0,000
Mojosari	0,000
Pacet	0,000

Berdasarkan Tabel 1 dapat diketahui yang memiliki persentase *missing value* terbesar terdapat pada Pos Kebon Agung sebesar 1,898 persen. Selain di Pos Kebon Agung, *missing value* juga terdapat di Pos Kandangan dan Pos Janjing yang masing-masing sebesar 1,132 persen dan 0,548 persen. *Missing value* selanjutnya diimputasi menggunakan rata-rata curah hujan harian di pos hujan pada tanggal dan bulan yang sama dengan *missing value* tersebut.

B. Eksplorasi Data Curah Hujan

Setelah dilakukan pra-pemrosesan data, maka dilakukan eksplorasi menggunakan statistika deskriptif yang bertujuan untuk mendapatkan gambaran umum data curah hujan.

Tabel 2 memberikan informasi bahwa rata-rata curah hujan harian tertinggi ada di Pos Pacet yaitu sebesar 6,678 mm per hari, sedangkan rata-rata curah hujan terendah tercatat di pos hujan Mojosari yaitu sebesar 5,064 mm per hari. Standar

deviasi curah hujan yang digunakan untuk melihat keragaman data pada masing-masing pos pengamatan. Pos Pacet memiliki keragaman curah hujan yang paling tinggi yaitu sebesar 15,543 dan Pos Gedeg memiliki keragaman curah hujan yang paling kecil sebesar 13,185. Sementara untuk curah hujan harian minimum untuk tujuh pos pengamatan adalah sama, yaitu sebesar 0 mm per hari yang berarti tidak terjadi hujan. Sedangkan curah hujan maksimum yang tertinggi terdapat pada Pos Pacet yaitu sebesar 225 mm per hari.

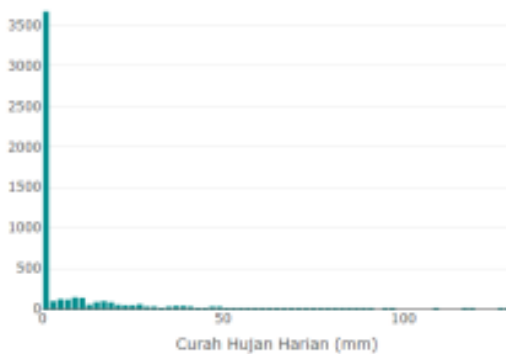
Tabel 2.
Statistika Deskriptif Curah Hujan Harian (mm)

Pos Pengamatan	Rata-Rata	StDev	Min	Max
Kandangan	6,137	13,808	0	130
Kebon Agung	6,175	13,991	0	114
Gedeg	5,372	13,185	0	126
Janjing	5,516	14,285	0	105
Klegen	5,130	13,584	0	120
Mojosari	5,064	13,836	0	162
Pacet	6,678	15,543	0	225

Tabel 3.
Skewness dan Kurtosis

Pos Pengamatan	Skewness	Kurtosis
Kandangan	3,29	13,74
Kebon Agung	3,28	12,90
Gedeg	3,66	16,61
Janjing	3,21	10,99
Klegen	3,52	14,10
Mojosari	3,97	19,94
Pacet	3,79	21,95

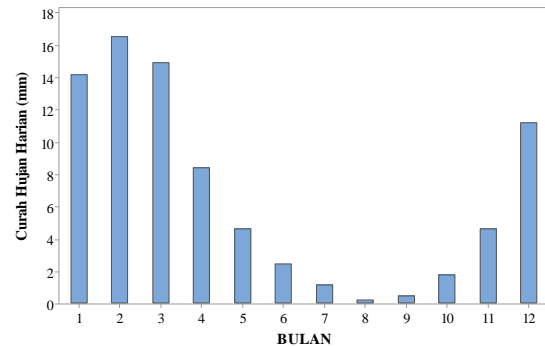
Tabel 3 memberikan informasi bahwa nilai *skewness* untuk masing-masing pos pengamatan bernilai lebih dari 0 (nol). Hal ini mengindikasikan distribusi curah hujan untuk semua pos pengamatan tidak simetris atau miring ke salah satu sisi (sisi kanan atau kiri). Sedangkan untuk nilai *kurtosis* digunakan untuk melihat seberapa runcing kurva distribusi data. Semakin besar nilai *kurtosis*nya, semakin runcing kurva yang mengindikasikan bahwa keragaman data cenderung lebih kecil. Pos Janjing memiliki nilai *skewness* terendah yaitu sebesar 3,21. Sedangkan yang memiliki nilai *kurtosis* tertinggi pada Pos Pacet sebesar 21,95 yang berarti pada Pos Pacet memiliki keragaman data yang terbesar dibandingkan dengan pos pengamatan lain.



Gambar 4. Histogram Curah Hujan Harian di Pos Pacet.

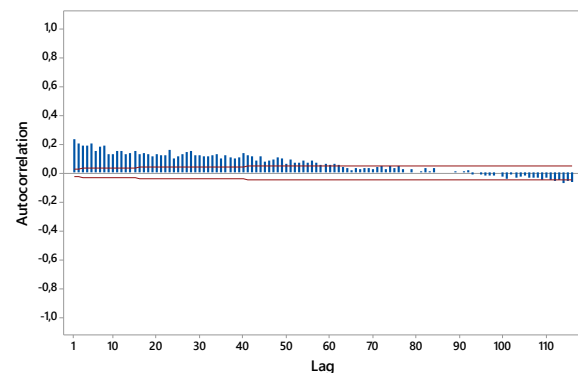
Pada Gambar 4 menunjukkan bahwa kurva distribusi data curah hujan harian pada Pos Pacet miring ke kanan dan

memperlihatkan tingginya frekuensi data menonjol di sekitar nol, sedangkan masih terdapat kejadian dengan curah hujan yang jauh lebih besar dari nol dengan frekuensi yang lebih kecil, sehingga mengindikasikan adanya pola data *heavy tail*.



Gambar 5. Bar Chart Curah Hujan Harian di Pos Pacet.

Seperti ditunjukkan pada Gambar 5 dapat diketahui bahwa pola rata-rata curah hujan harian di pos hujan Pacet adalah pola hujan monsun dimana pola hujan membentuk pola huruf U yang termasuk pada tipe curah hujan monsun yang hanya terdapat satu puncak musim hujan. Pada Pos Pacet puncak musim hujan terjadi pada bulan Februari. Sedangkan puncak terendah yaitu pada bulan Agustus, yang berarti musim kemarau terjadi pada bulan Agustus.



Gambar 6. Plot ACF Pos Pacet.

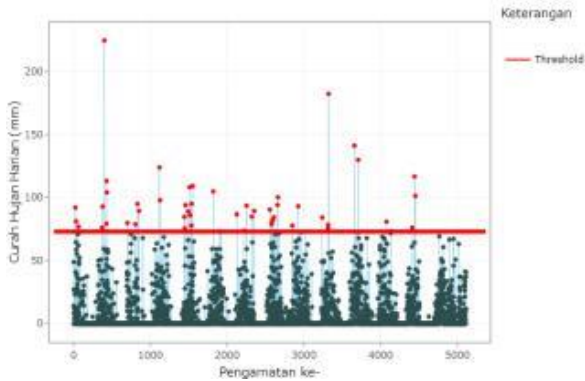
Gambar 6 menunjukkan beberapa lag yang melebihi batas atas fungsi autokorelasi. Sehingga menunjukkan data memiliki korelasi yang mengindikasikan bahwa data bersifat dependen. Indikasi adanya dependensi data juga ditemukan pada data curah hujan pada masing-masing pos pengamatan.

Berdasarkan eksplorasi data yang sudah dilakukan pada Pos Pacet menghasilkan kesimpulan bahwa data curah hujan memiliki pola *heavy tail*, tipe curah hujan monsoon tidak berdistribusi normal, dan bersifat dependen. Hal tersebut juga terjadi pada keenam pos pengamatan lainnya.

C. Peaks Over Threshold

Pada penelitian ini menggunakan EVT melalui pendekatan POT dengan cara menentukan nilai *threshold* terlebih dahulu menggunakan metode persentase, yaitu mengambil batas dari 1% keseluruhan data yang sudah diurutkan dari yang terbesar

hingga terkecil. Berikut merupakan gambar dari pengambilan data ekstrem POT.



Gambar 7. Pengambilan Sampel Ekstrem di Pos Pacet.

Informasi yang dapat diambil dari metode POT adalah informasi mengenai banyak pengamatan (N), *threshold* (u), dan banyak data ekstrem (n_u) untuk masing-masing pos pengamatan. Berikut merupakan informasi dari metode POT. Tabel 4 memberikan *Threshold* tertinggi terdapat pada Pos Pacet yaitu sebesar 73 mm. Dapat disimpulkan bahwa curah hujan di Pos Pacet relatif tinggi dibandingkan dengan keenam pos pengamatan lain. Sedangkan data ekstrem terbanyak yaitu pada Pos Kandangan, Pos Gedeg, dan Pos Mojosari yaitu sebanyak 51 *cluster*. Data ekstrem POT untuk analisis selanjutnya akan disebut dengan metode *Non Declustering*

Tabel 4. Nilai *Threshold* POT

Variable	N	u (mm)	n_u
Kandangan	5114	66	51
Kebon Agung	5114	69	50
Gedeg	5114	64	51
Janjing	5114	70	50
Klegen	5114	70	36
Mojosari	5114	72	51
Pacet	5114	73	48

D. Declustering

Pada penelitian ini, akan dilakukan *declustering* untuk mengatasi dependensi data. Dalam proses *declustering*, dibutuhkan nilai r (*run length*) untuk membentuk *cluster*. Nilai r yang optimum ditentukan melalui *extremal index*.

Tabel 5. *Extremal Index* pada Tujuh Pos Pengamatan

Pos Pengamatan	u	<i>Extremal Index</i>	c	r
Kandangan	66	0.658	34	49
Kebon Agung	69	0.817	41	53
Gedeg	64	0.746	39	1
Janjing	70	0.774	39	21
Klegen	70	0.972	35	19
Mojosari	72	0.745	39	43
Pacet	73	0.757	36	189

Pada Tabel 5 dapat diketahui nilai *extremal index* yang digunakan sebagai indikator untuk mengukur berapa banyak *cluster* (c) yang terbentuk sehingga mendapatkan data independen. Nilai r optimum yang terdapat pada Tabel 5

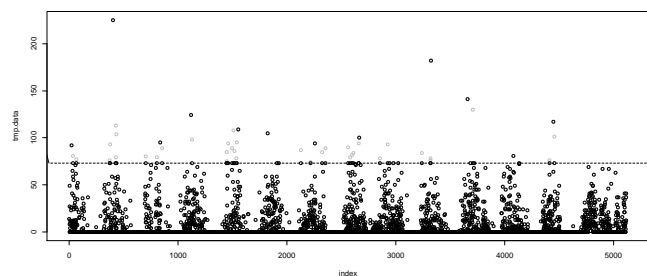
digunakan untuk proses selanjutnya yaitu melakukan *declustering*.

Tabel 6.

Hasil Estimasi <i>Runs Estimator</i> dan <i>Interval Estimator</i>			
Pos Pengamatan	Nilai	<i>Runs Estimator</i>	<i>Intervals Estimator</i>
Kandangan	<i>Extremal Index</i>	0.372	0.658
	c	19	34
	r	49	13
Kebon Agung	<i>Extremal Index</i>	0.420	0.817
	c	21	41
	r	53	12
Gedeg	<i>Extremal Index</i>	0.941	0.746
	c	48	39
	r	1	12
Janjing	<i>Extremal Index</i>	0.58	0.774
	c	29	39
	r	21	6
Klegen	<i>Extremal Index</i>	0.778	0.972
	c	28	35
	r	19	2
Mojosari	<i>Extremal Index</i>	0.529	0.745
	c	27	39
	r	43	12
Pacet	<i>Extremal Index</i>	0.250	0.757
	c	12	36
	r	189	10

Hasil *declustering* yang didapatkan menggunakan *software R* terdapat dua pendekatan. Pendekatan yang pertama yaitu *runs estimator* dan pendekatan yang kedua yaitu *intervals estimator*. Hasil pada kedua pendekatan tersebut, disajikan pada Tabel 6.

Jika dibandingkan antara Tabel 5 dan 6, ada beberapa nilai yang sama antara perhitungan manual dan menggunakan *R*. Ketika dilakukan perhitungan manual, dihasilkan nilai r yang sama dengan pendekatan *runs estimator*. Sedangkan pada nilai *extremal index* dan c pada perhitungan manual, memiliki nilai yang sama dengan metode *intervals estimator*. Hal ini membuktikan adanya teori yang dipaparkan pada subbab II.D. Adapun data ekstrem yang diperoleh dari *declustering* pada Pos Pacet disajikan pada Gambar 8.



Gambar 8. Data Ekstrem *Declustering* pada Pos Pacet.

Jika kedua pendekatan antara POT dan *declustering* dibandingkan, maka pendekatan POT umumnya memiliki jumlah data ekstrem yang lebih banyak dibandingkan pada pendekatan *declustering*. Oleh karena itu, data ekstrem POT

(yang selanjutnya akan disebut dengan *non declustering*) akan dibandingkan dengan *declustering*.

E. Identifikasi Distribusi dan Dependensi Data Ekstrem Declustering dan Non Declustering

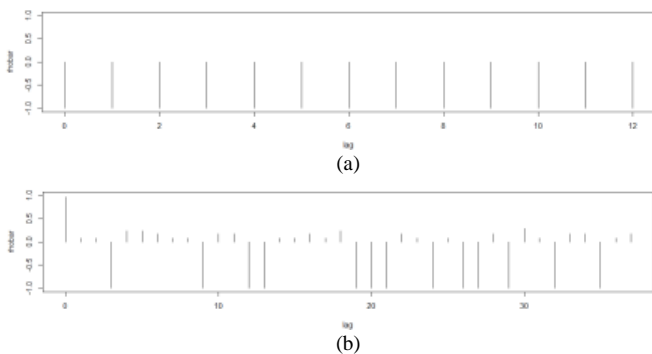
Berikut merupakan hasil identifikasi distribusi menggunakan uji *Kolmogorov-Smirnov* yang disajikan pada Tabel 7.

Tabel 7. Hasil KST Data Ekstrem Declustering dan Non Declustering

Pos Pengamatan	Declustering		Non Declustering	
	D_{hitung}	P-Value	D_{hitung}	P-Value
Kandangan	0,150	0,729	0,085	0,821
Kebon Agung	0,089	0,990	0,078	0,893
Gedeg	0,084	0,921	0,071	0,942
Janjing	0,120	0,746	0,165	0,114
Klegen	0,113	0,824	0,140	0,440
Mojosari	0,135	0,500	0,137	0,267
Pacet	0,138	0,875	0,084	0,850

Berdasarkan hasil pengujian *Kolmogorov-Smirnov* yang disajikan pada Tabel 7 dapat diketahui bahwa semua pos pengamatan pada data ekstrem *declustering* maupun *non declustering* memiliki nilai *P-Value* lebih besar dari α . Sehingga data ekstrem *declustering* dan *non declustering* pada semua pos pengamatan mengikuti distribusi teoritis yaitu GPD. Selain mengetahui kesesuaian data ekstrem yang mengikuti *Generalized Pareto Distribution*, perlu juga dilakukan identifikasi adanya dependensi data ekstrem *declustering* maupun *non declustering*.

Identifikasi dependensi pada data ekstrem dapat dilihat melalui plot *extremogram*. Berikut merupakan plot *extremogram* pada data ekstrem *declustering* dan *non declustering*.



Gambar 9. *Extremogram* Data Ekstrem (a) *Declustering* dan (b) *Non Declustering*.

Gambar 9 memperlihatkan plot *extremogram* pada data ekstrem Pos Pacet yang melalui proses *declustering* dan *non declustering*. Dapat diketahui bahwa Gambar 8 tidak membentuk suatu tren ataupun periode tertentu sehingga mengindikasikan bahwa data ekstrem *declustering* dan *non declustering* sudah independen.

F. Estimasi Parameter GPD

Berikut merupakan hasil estimasi parameter GPD dengan menggunakan metode MLE.

Parameter $\hat{\sigma}$ merupakan parameter skala yang menyatakan keragaman data. Pos pengamatan melalui proses *declustering* memiliki nilai keragaman data ekstrem yang terbesar yaitu pada Pos Kandangan, yang berarti curah hujan ekstrem pada Pos Kandangan memiliki *range* persebaran data terbesar, sehingga intensitas curah hujan terjadi lebih beragam dibandingkan dengan pos pengamatan yang lain. Sedangkan pos hujan yang memiliki keragamana data ekstrem terkecil adalah Pos Janjing yang berarti *range* persebaran data ekstrem pada Pos Janjing terkecil sehingga intensitas curah hujan tidak terlalu beragam dibandingkan dengan pos pengamatan yang lain. Jika dilihat pada parameter $\hat{\xi}$ yang melalui proses *non declustering* memiliki *range* persebaran data terbesar terdapat pada Pos Kebon Agung. Sedangkan yang memiliki *range* persebaran data terkecil terdapat pada Pos Klegen.

Tabel 8.

Pos Pengamatan	Hasil Estimasi Parameter GPD		
	Nilai	Declustering	Non Declustering
Kandangan	$\hat{\sigma}$	60,008	16,319
	$\hat{\xi}$	-0,933	0,000
	Tipe Distribusi	Pareto tipe 2	Ekspensial
Kebon Agung	$\hat{\sigma}$	41,143	24,188
	$\hat{\xi}$	-0,909	-0,501
	Tipe Distribusi	Pareto tipe 2	Pareto tipe 2
Gedeg	$\hat{\sigma}$	30,477	22,110
	$\hat{\xi}$	-0,413	-0,214
	Tipe Distribusi	Pareto tipe 2	Pareto tipe 2
Janjing	$\hat{\sigma}$	19,953	15,046
	$\hat{\xi}$	-0,532	-0,378
	Tipe Distribusi	Pareto tipe 2	Pareto tipe 2
Klegen	$\hat{\sigma}$	22,581	10,932
	$\hat{\xi}$	-0,372	0,000
	Tipe Distribusi	Pareto tipe 2	Ekspensial
Mojosari	$\hat{\sigma}$	20,314	16,010
	$\hat{\xi}$	-0,028	0,033
	Tipe Distribusi	Pareto tipe 2	Pareto
Pacet	$\hat{\sigma}$	41,939	20,672
	$\hat{\xi}$	-0,046	0,112
	Tipe Distribusi	Pareto tipe 2	Pareto

Parameter $\hat{\xi}$ merupakan parameter bentuk yang menyatakan bagaimana perilaku dari ekor kanan distribusi data. Semakin besar nilai parameter $\hat{\xi}$, ekor kanan cenderung turun secara landai (*heavy tail*). Semakin gemuk ekor distribusi data maka peluang terjadinya kejadian ekstrem semakin besar. Pada data ekstrem melalui proses *declustering* yang memiliki peluang terjadinya curah hujan ekstrem terbesar adalah Pos Mojosari. Hal tersebut mengartikan bahwa pada Pos Mojosari lebih sering terjadi hujan dengan curah hujan ekstrem

dibandingkan dengan pos pengamatan yang lain. Sedangkan pada data ekstrem yang melalui proses *non declustering* yang memiliki peluang terjadinya curah hujan ekstrem terkecil adalah Pos Kebon Agung.

Pada estimasi parameter melalui proses *declustering* menghasilkan distribusi Pareto tipe 2. Sedangkan estimasi parameter *non declustering* pada Pos Kandangan dan Pos Klegen menghasilkan distribusi Eksponensial, Pos Mojosari dan Pos Pacet menghasilkan distribusi Pareto. Sedangkan Pos Kebon Agung, Pos Gedeg, dan Pos Janjing menghasilkan distribusi Pareto tipe 2.

G. Return Level

Pada penelitian ini telah didapatkan nilai estimasi parameter skala dan bentuk yang sudah dibahas pada subbab IV.F. Nilai estimasi tersebut digunakan untuk mencari nilai estimasi *return level* yang disajikan pada Tabel 9. *Return level* merupakan nilai estimasi curah hujan ekstrem pada periode waktu tertentu. Perhitungan *return level* penelitian ini dilakukan untuk satu tahun yang terbagi dalam empat periode. Periode pertama terdiri dari Januari 2017 sampai Maret 2017, periode kedua dari Januari 2017 sampai Juni 2017, periode ketiga dari Januari 2017 sampai September 2017, sedangkan periode keempat dari Januari 2017 sampai Desember 2017. Periode estimasi tidak dilakukan hanya pada satu periode agar dapat dibandingkan antara nilai *return level* yang muncul dengan probabilitas yang ada.

Berdasarkan Tabel 9 dapat diketahui bahwa pada estimasi *return level declustering* lebih mendekati nilai aktual. Pada Pos Klegen dengan periode ke 2 diperoleh *return level* sebesar 125,381 mm per hari yang artinya peluang kejadian ekstrem melebihi 125,281 mm per hari sebesar 0,5 pada periode 3 bulan.

Tabel 9. Hasil Return Level

Pos Pengamatan	Periode Ulang (Bulan)	Aktual	Return Level Declustering (mm)	Return Level Non Declustering (mm)
Kandangan	2	85	130,129	173.609
	3	85	130,172	180.226
	4	98	130,194	184.921
Kebon Agung	2	92	114,134	115.478
	3	92	114,168	115.804
	4	112	114,186	115.998
Gedeg	2	117	132,894	141.937
	3	117	132,639	144.018
	4	117	134,097	145.388
Janjing	2	97	106,377	106.493
	3	97	106,595	106.960
	4	129	106,724	107.250
Klegen	2	125	125,381	142.084
	3	125	126,108	146.517
	4	125	126,561	149.662
Mojosari	2	118	193,960	190.175
	3	118	200,728	198.331
	4	118	205,482	204.187
Pacet	2	132	310,868	274.978
	3	132	323,232	292.969

4 132 331,863 306,240

H. Tingkat Kesalahan Ramalan

Metode yang memiliki nilai RMSE terkecil adalah metode yang lebih baik dalam menentukan *return level*. Hasil perbandingan nilai RMSE pada pos pengamatan di Surabaya dapat dilihat pada Tabel 10.

Tabel 10. RMSE Return Level Declustering dan Non Declustering

Pos Pengamatan	RMSE Declustering	RMSE Non Declustering
Kandangan	41,286	90,323
Kebon Agung	18,130	19,441
Rata-rata	29,708	54,882

Keterangan : *nilai RMSE yang lebih kecil pada Non Declustering

Tabel 10 merupakan perbandingan RMSE *return level* yang melalui *declustering* dan *non declustering* di Surabaya. Dapat diketahui bahwa rata-rata terkecil terdapat pada RMSE yang melalui proses *declustering*. Sehingga dapat disimpulkan bahwa di Surabaya metode estimasi *return level* yang lebih baik yaitu menggunakan *declustering*.

Tabel 11. RMSE Return Level Declustering dan Non Declustering

Pos Pengamatan	RMSE Declustering	RMSE Non Declustering
Gedeg	16,222	26,819
Janjing	15,013	14,859
Klegen	1,127	21,316
Mojosari	82,193	79,772
Pacet	190,183	159,90
Rata-rata	29,708	54,882

Hasil perbandingan nilai RMSE *return level* yang melalui *Declustering* dan *Non Declustering* pada pos pengamatan di Mojokerto dapat dilihat pada Tabel 11. Dapat diketahui bahwa rata-rata nilai RMSE yang melalui proses *Declustering* dan *Non Declustering* tidak jauh berbeda yang hanya selisih 0,414. RMSE yang terkecil pada pos pengamatan di Mojokerto yang memiliki nilai lebih kecil adalah RMSE yang melalui proses *Non Declustering*. Sehingga dapat disimpulkan bahwa di Mojokerto metode estimasi *return level* yang lebih baik yaitu menggunakan *Non Declustering*.

V. KESIMPULAN DAN SARAN

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan bahwa karakteristik curah hujan ekstrem di tujuh pos pengamatan memiliki pola distribusi *heavy tail* atau tidak berdistribusi normal. Pada estimasi parameter melalui proses *Declustering* menghasilkan distribusi Pareto tipe 2. Sedangkan estimasi parameter *Non Declustering* pada Pos Kandangan dan Pos Klegen menghasilkan distribusi Eksponensial, Pos Mojosari dan Pos Pacet menghasilkan distribusi Pareto. Sedangkan Pos Kebon Agung, Pos Gedeg, dan Pos Janjing menghasilkan distribusi Pareto tipe 2.Kebaikan estimasi *return level* curah hujan ekstrem di Surabaya menunjukkan hasil yang

lebih baik menggunakan *Delclustering*, sedangkan di Mojokerto yang lebih baik yaitu *Non Declustering*.

Pada penelitian ini, metode *Declustering* dan *Non Declustering* mampu mengatasi dependensi data, namun masih menghasilkan *return level* yang kurang baik. Sehingga pada analisis selanjutnya disarankan menggunakan data simulasi sehingga dapat diketahui apakah dengan menggunakan *Declustering* hanya mengatasi dependensi data saja dan tidak menghasilkan *return level* yang lebih baik. Selain itu dapat menggunakan variabel lain yang dapat memengaruhi prediksi curah hujan seperti kecepatan angin atau kelembapan udara.

DAFTAR PUSTAKA

- [1] ITS News, "Intensitas Curah Hujan Penyebab Surabaya Dikepung Banjir," www.its.ac.id, 2017. [Online]. Available: <https://www.its.ac.id/news/2017/11/26/intensitas-curah-hujan-penyebab-surabaya-dikepung-banjir/>. [Accessed: 14-Jun-2019].
- [2] D. Fawcett, Lee; Walshaw, "Improved Estimation for Temporally Clustered Extremes," *Environmetrics*, vol. 18, no. September 2006, p. 18, 2007.
- [3] R. S. Malika, "Declustering Peaks Over Threshold Pada Data Curah Hujan Ekstrem Dependen di Sentra Produksi Padi Jawa Timur," Surabaya, 2014.
- [4] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, 1 st editi., vol. 44, no. 4. London: Springer-Verlag, 2001.
- [5] A. J. Mcnail, "Extreme Value Theory for Risk Managers," Zurich, 1999.
- [6] Z. Li, *Study on Climate Change in Southwestern China*. Lanzhou, China: Springer-Verlag Berlin Heidelberg, 2015.
- [7] C. A. T. Ferro and J. Segers, "Inference for clusters of extreme values," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 65, no. 2, pp. 545-556, 2003.
- [8] Warto, "Discrete Wavelet Transform Untuk Mempercepat Prediksi Produksi Susu Sapi Perah Berbasis Neural Network," Masters thesis, Universitas Dian Nuswantoro Semarang, 2012.
- [9] Handoko, *Klimatologi dasar : landasan pemahaman fisika atmosfer dan unsur- unsur iklim*. Jakarta: Pustaka Jaya, 1994.