



TUGAS AKHIR - KS184822

**PERBANDINGAN ESTIMASI *RETURN LEVEL*  
*DECLUSTERING* DAN *NON DELCUSTERING*  
PADA DATA CURAH HUJAN EKSTREM  
DI SURABAYA DAN MOJOKERTO**

**AINUN UMAMI  
NRP 062115 4000 0090**

**Dosen Pembimbing  
Dr. Sutikno, S.Si, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**





**TUGAS AKHIR - KS184822**

**PERBANDINGAN ESTIMASI *RETURN LEVEL*  
*DECLUSTERING* DAN *NON DELCUSTERING*  
PADA DATA CURAH HUJAN EKSTREM  
DI SURABAYA DAN MOJOKERTO**

**AINUN UMAMI  
NRP 062115 4000 0090**

**Dosen Pembimbing  
Dr. Sutikno, S.Si, M.Si**

**PROGRAM STUDI SARJANA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**





**FINAL PROJECT - KS184822**

**COMPARISON OF THE RETURN LEVEL  
DECLUSTERING AND *NON DECLUSTERING*  
ESTIMATES FOR EXTREME RAINFALL DATA  
IN SURABAYA AND MOJOKERTO**

**AINUN UMAMI  
NRP 062115 4000 0090**

**Supervisor  
Dr. Sutikno, S.Si, M.Si**

**UNDERGRADUATE PROGRAMME  
DEPARTMENT OF STATISTICS  
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA 2019**



**LEMBAR PENGESAHAN**

**PERBANDINGAN ESTIMASI *RETURN LEVEL*  
*DECLUSTERING* DAN *NON DECLUSTERING*  
PADA DATA CURAH HUJAN EKSTREM  
DI SURABAYA DAN MOJOKERTO**

**TUGAS AKHIR**

Diajukan untuk Memenuhi Salah Satu Syarat  
Memperoleh Gelar Sarjana Statistika  
pada  
Program Studi Sarjana Departemen Statistika  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember

Oleh :

**Ainun Umami**

NRP. 062115 4000 0090

Disetujui oleh Pembimbing Tugas Akhir :

**Dr. Sutikno, S.Si, M.Si**  
NIP. 19710313 199702 1 001

(  )

Mengetahui,  
Ketua Departemen



  
**Dr. Suhartono** SA

NIP. 19710929 199512 1 001  
SURABAYA, JULI 2019

*(Halaman ini sengaja dikosongkan)*



**PERBANDINGAN ESTIMASI *RETURN LEVEL*  
*DECLUSTERING* DAN *NON DECLUSTERING*  
PADA DATA CURAH HUJAN EKSTREM  
DI SURABAYA DAN MOJOKERTO**

**Nama Mahasiswa** : Ainun Umami  
**NRP** : 062115 4000 0090  
**Departemen** : Statistika  
**Dosen Pembimbing** : Dr. Sutikno, S.Si, M.Si

**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*

*(Halaman ini sengaja dikosongkan)*

**COMPARISON OF THE RETURN LEVEL  
DECLUSTERING AND NON DECLUSTERING  
ESTIMATE FOR EXTREME RAINFALL DATA  
IN SURABAYA AND MOJOKERTO**

**Name** : Ainun Umami  
**NRP** : 062115 4000 0090  
**Department** : Statistics  
**Supervisor** : Dr. Sutikno, S.Si, M.Si

**Abstract**

*Surabaya is an urban area which is prone to heavy rainfall so frequent flooding. Mojokerto from year to year transitional land use also frequent flooding. To minimize flood losses, knowledge is needed regarding extreme events using Extreme Value Theory (EVT). There are two approaches in EVT, namely Block Maxima and Peaks Over Threshold. Peaks Over Threshold (POT) is one method to identify extreme values using a reference value called the threshold. One of the EVT requirements is that data must be independent. But in reality extreme data is often dependent so it is necessary to handle data dependencies using Declustering. It is interesting that the EVT is return level. Return level is the maximum expected value would be exceeded once in a given time period. In this research will be comparison of the return level Declustering and Non Declustering (POT) estimates was made on extreme rainfall data. The number of observation posts used was two observation posts in Surabaya and five observation posts in Mojokerto. The results of this research show that in Surabaya better method is Declustering, while in Mojokerto better method is Non Declustering.*

**Keywords:** *Declustering, Mojokerto, Peaks Over Threshold, Return Level, Surabaya*

*(Halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji syukur penulis panjatkan atas berkat dan rahmat yang diberikan Tuhan Yang Maha Esa sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Perbandingan Estimasi *Return Level Declustering* dan *Non Declustering* pada Data Curah Hujan Ekstrem Di Surabaya dan Mojokerto”** dengan lancar.

Penulis menyadari bahwa Tugas Akhir ini dapat terselesaikan tidak terlepas dari bantuan dan dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada :

1. Kedua orang tua dan keluarga besar penulis yang selalu memberikan doa dan dukungan kepada penulis dalam proses penyusunan Tugas Akhir ini.
2. Bapak Dr. Sutikno, S.Si, M.Si selaku dosen pembimbing Tugas Akhir yang telah meluangkan waktu, tenaga dan pikiran untuk membagikan ilmu, saran dan nasehat dalam penyusunan Tugas Akhir.
3. Bapak Dr. Bambang Widjanarko Otok, S.Si, M.Si dan Bapak Dr. Purhadi, M.Sc selaku dosen penguji Tugas Akhir yang telah banyak memberikan ilmu, saran, dan masukan kepada penulis untuk kesempurnaan Tugas Akhir.
4. Bapak Dr. Bambang Widjanarko Otok, S.Si, M.Si selaku dosen wali yang telah banyak memberikan saran dan arahan dalam proses belajar di Departemen Statistika ITS.
5. Bapak Dr. Suhartono selaku Kepala Departemen Statistika ITS dan Ibu Dr. Santi Wulan Purnami, S.Si, M.Si selaku Ketua Program Studi Sarjana yang telah memberikan fasilitas, sarana, dan prasarana.
6. Seluruh Dosen Statistika ITS yang telah memberikan ilmu dan pengetahuan yang tak ternilai harganya, serta segenap karyawan Departemen Statistika ITS.

7. Teman-teman Statistika ITS angkatan 2015 Vivacious, yang selalu memberikan dukungan kepada penulis selama ini.
8. Teman-teman seperjuangan bimbingan Tugas Akhir Pak Tikno yang selalu memberikan informasi dan mau untuk saya repotkan.
9. Serta semua pihak yang telah memberikan bantuan dan dukungan dalam pelaksanaan Tugas Akhir yang tidak bisa penulis sebutkan satu persatu..

Semoga seluruh kebaikan yang diberikan kepada penulis dibaloh oleh Allah SWT dengan balasan yang lebih baik. Aamiin

Penulis menyadari masih banyaknya kekurangan dalam pembuatan laporan Tugas Akhir ini, besar harapan bagi penulis untuk dapat menerima saran dan kritik yang bersifat membangun guna perbaikan di masa mendatang. Semoga laporan Tugas Akhir ini dapat bermanfaat bagi kita semua.

Surabaya, Juli 2019

Penulis

## DAFTAR ISI

	Halaman
<b>HALAMAN JUDUL</b> .....	i
<b>COVER PAGES</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	v
<b>ABSTRAK</b> .....	vii
<b>ABSTRACT</b> .....	ix
<b>KATA PENGANTAR</b> .....	xi
<b>DAFTAR ISI</b> .....	xiii
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xvii
<b>DAFTAR LAMPIRAN</b> .....	xix
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	5
1.4 Manfaat.....	5
1.5 Batasan Masalah.....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	7
2.1 Statistika Deskriptif.....	7
2.2 Pengujian Kesesuaian Distribusi.....	9
2.3 <i>Autocorrelation Function</i> .....	10
2.4 <i>Extremogram</i> .....	11
2.5 <i>Extreme Value Theory (EVT)</i> .....	11
2.5.1 <i>Peaks Over Threshold</i> .....	12
2.5.2 <i>Generalized Pareto Distribution</i> .....	13
2.5.3 Penentuan Nilai Threshold.....	15
2.5.4 <i>Declustering</i> .....	16
2.5.5 Return Level.....	19
2.6 Estimasi Parameter <i>Generalized Pareto</i> <i>Distribution</i> .....	20
2.7 Data Ekstrem Dependen.....	22
2.8 Kriteria Pemilihan Metode.....	24
2.9 Curah Hujan.....	24
2.10 Zona Musim (ZOM).....	26

<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>29</b>
3.1 Sumber Data.....	29
3.2 Variabel Penelitian.....	29
3.3 Langkah Analisis.....	31
<b>BAB IV ANALISIS DAN PEMBAHASAN .....</b>	<b>35</b>
4.1 Pra-Pemrosesan Data Curah Hujan.....	35
4.2 Eksplorasi Data Curah Hujan.....	36
4.2.1 Deskripsi Data Curah Hujan .....	36
4.2.2 Histogram Data Curah Hujan .....	38
4.2.3 Pola Data Curah Hujan .....	39
4.3 <i>Identifikasi Dependensi Data Curah Hujan</i> .....	40
4.4 <i>Extreme Value Theory</i> .....	41
4.3.1 <i>Peaks Over Threshold</i> .....	41
4.3.2 <i>Declustering</i> .....	43
4.5 Uji Kesesuaian GPD <i>Declustering</i> dan <i>Non Declustering</i> .....	46
4.6 Identifikasi Dependensi Pada Data Ekstrem <i>Declustering</i> dan <i>Non Declustering</i> .....	47
4.7 Estimasi Parameter GPD.....	48
4.8 Return Level.....	50
4.9 Tingkat Kesalahan Ramalan .....	54
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>57</b>
5.1 Kesimpulan .....	57
5.2 Saran .....	57
<b>DAFTAR PUSTAKA.....</b>	<b>59</b>
<b>LAMPIRAN .....</b>	<b>63</b>
<b>BIODATA PENULIS .....</b>	<b>127</b>



## DAFTAR GAMBAR

	Halaman
<b>Gambar 2.1</b> Contoh Histogram .....	9
<b>Gambar 2.2</b> Contoh Bar Char .....	9
<b>Gambar 2.3</b> Ilustrasi Pengambilan Sampel Data Ekstrem POT .....	12
<b>Gambar 2.4</b> Ilustrasi Penentuan Threshold dengan Metode Persentase .....	15
<b>Gambar 2.5</b> Ilustrasi Run Length pada Data Temperatur dengan nilai $r=1$ .....	16
<b>Gambar 2.6</b> Pola Data a) Dependen dan b) Independen .....	23
<b>Gambar 2.7</b> Pola Data (a) Siklik (b) Tren Naik (c) Tren Turun .....	24
<b>Gambar 2.8</b> Peta tipe hujan di Indonesia .....	26
<b>Gambar 2.9</b> Peta Zona Musim Jawa Timur .....	27
<b>Gambar 3.1</b> Peta Tujuh Stasiun di Surabaya dan Mojokerto .....	30
<b>Gambar 3.2</b> Diagram Alir Penelitian .....	33
<b>Gambar 4.1</b> Histogram Curah Hujan Harian di Pos Pacet...	38
<b>Gambar 4.2</b> Bar Chart Curah Hujan Harian di Pos Pacet ....	39
<b>Gambar 4.3</b> <i>Time Series Plot</i> Data Curah Hujan Harian Pos Pacet .....	40
<b>Gambar 4.4.</b> Plot ACF Pos Pacet .....	40
<b>Gambar 4.5</b> Pengambilan Sampel Ekstrem di Pos Pacet .....	42
<b>Gambar 4.6</b> Data Ekstrem <i>Delustering</i> pada Pos Pacet .....	46
<b>Gambar 4.7</b> <i>Extremogram</i> Data Ekstrem (a) <i>Declustering</i> (b) <i>Non Declustering</i> .....	48

*(Halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

	Halaman
<b>Tabel 3.1</b> Data Pos Hujan Di Surabaya dan Mojokerto.....	29
<b>Tabel 3.2</b> Struktur Data Penelitian.....	31
<b>Tabel 4.1</b> Identifikasi <i>Missing Value</i> Data Curah Hujan .....	35
<b>Tabel 4.2</b> Statistika Deskriptif Curah Hujan Harian (mm) ...	36
<b>Tabel 4.3</b> <i>Skewness</i> dan <i>Kurtosis</i> .....	37
<b>Tabel 4.4</b> Nilai Threshold POT .....	42
<b>Tabel 4.5</b> Extremal Index pada Tujuh Pos Pengamatan .....	43
<b>Tabel 4.6</b> Hasil Estimasi <i>Runs Estimator</i> dan <i>Interval Estimator</i> .....	45
<b>Tabel 4.7</b> Hasil KST Data Ekstrem <i>Declustering</i> dan <i>Non Declustering</i> .....	47
<b>Tabel 4.8</b> Hasil Estimasi Parameter GPD.....	49
<b>Tabel 4.9</b> Hasil <i>Return Level</i> .....	51
<b>Tabel 4.10</b> Nilai <i>Error</i> proses <i>Declustering</i> dan <i>Non Declustering</i> .....	53
<b>Tabel 4.11</b> RMSE <i>Return Level</i> pada Metode <i>Declustering</i> dan <i>Non Declustering</i> di Surabaya .....	54
<b>Tabel 4.12.</b> RMSE <i>Return Level</i> pada Metode <i>Declustering</i> dan <i>Non Declustering</i> di Mojokerto .....	54

*(Halaman ini sengaja dikosongkan)*

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1.</b> Data Curah Hujan Harian Pos Kandangan.....	63
<b>Lampiran 2.</b> Data Curah Hujan Harian Pos Kebon Agung.....	64
<b>Lampiran 3.</b> Data Curah Hujan Harian Pos Gedeg.....	65
<b>Lampiran 4.</b> Data Curah Hujan Harian Pos Janjing .....	66
<b>Lampiran 5.</b> Data Curah Hujan Harian Pos Klegen .....	67
<b>Lampiran 6.</b> Data Curah Hujan Harian Pos Mojosari.....	68
<b>Lampiran 7.</b> Data Curah Hujan Harian Pos Pacet .....	69
<b>Lampiran 8.</b> Estimasi Parameter GPD.....	70
<b>Lampiran 9.</b> Ilustrasi Penentuan <i>Threshold</i> Metode Persentase Pada Pos Pacet.....	75
<b>Lampiran 10.</b> Ilustrasi Menghitung <i>Extremal Index</i> Pada Pos Pacet.....	77
<b>Lampiran 11.</b> Ilustrasi Menentukan <i>Run Length</i> Pada Pos Pacet .....	79
<b>Lampiran 12.</b> Histogram Curah Hujan Harian .....	80
<b>Lampiran 13.</b> Bar Chart Curah Hujan Harian Per Bulan .....	82
<b>Lampiran 14.</b> <i>Time Series Plot</i> .....	84
<b>Lampiran 15.</b> Plot ACF .....	86
<b>Lampiran 16.</b> Pengambilan Sampel Ekstrem POT .....	88
<b>Lampiran 17.</b> Pengambilan Sampel Ekstrem <i>Declustering</i> .....	90
<b>Lampiran 18.</b> <i>Extremogram</i> Data Ekstrem <i>Declustering</i> ....	92
<b>Lampiran 19.</b> <i>Extremogram</i> Data Ekstrem <i>Non</i> <i>Declustering</i> .....	94
<b>Lampiran 20.</b> Syntax R Histogram Data Curah Hujan Harian .....	96

<b>Lampiran 21.</b>	<i>Syntax R Time Series Plot Data Curah Hujan Harian</i> .....	98
<b>Lampiran 22.</b>	<i>Syntax R Pengambilan Sampel Ekstrem POT</i> .....	100
<b>Lampiran 23.</b>	<i>Syntax R Declustering</i> .....	103
<b>Lampiran 24.</b>	<i>Syntax R Estimasi Parameter Declustering</i> .....	103
<b>Lampiran 25.</b>	<i>Syntax R Estimasi Parameter Non Declustering</i> .....	105
<b>Lampiran 26.</b>	<i>Syntax R Return Level Declustering</i> .....	107
<b>Lampiran 27.</b>	<i>Syntax R Return Level Non Declustering</i> ....	108
<b>Lampiran 28.</b>	<i>Output Declustering</i> .....	109
<b>Lampiran 29.</b>	<i>Output Estimasi Parameter GDP Declustering</i> .....	112
<b>Lampiran 30.</b>	<i>Output Estimasi Parameter GDP Non Declustering</i> .....	116
<b>Lampiran 31.</b>	<i>Output Return Level Declustering</i> .....	120
<b>Lampiran 32.</b>	<i>Output Return Level Non Declustering</i> .....	121
<b>Lampiran 33.</b>	<i>Tabel Kolmogorov Test Statistic</i> .....	123
<b>Lampiran 34.</b>	<i>Surat Pernyataan Data I</i> .....	125
<b>Lampiran 35.</b>	<i>Surat Pernyataan Data II</i> .....	126

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Surabaya merupakan wilayah perkotaan yang memiliki konstruksi material penyimpan panas, seperti beton yang digunakan untuk konstruksi bangunan dan aspal untuk pembuatan jalan. Material penyimpan panas tersebut dapat mengakibatkan terjadinya kenaikan suhu yang menyebabkan curah hujan yang tinggi sehingga terjadi banjir. Menurut Dinas Pekerjaan Umum Bina Marga dan Pematusan Surabaya dalam Alfionita (2018) mulai tahun 2011 hingga 2015 sudah lebih dari 150 titik area terkena dampak banjir di Surabaya. Data dari BPS (2017) menjelaskan bahwa secara geografis Surabaya memiliki luas wilayah 326,36 km<sup>2</sup>. Berita yang disampaikan ITS News (2017) mengatakan bahwa curah hujan dengan intensitas 110 mm mengguyur Surabaya sehingga terjadi banjir. 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, seperti lahan pertanian yang berubah fungsi menjadi lahan pemukiman, lahan industri, bangunan, dan jalan. Menurut BPS (2018) menjelaskan kondisi topografi di Mojokerto cenderung cekung ditengah-tengah dan tinggi dibagian selatan dan utara sehingga tidak menutup kemungkinan jika sering terjadi banjir. Berita mengenai banjir, dikabarkan melalui DetikNews (2016) yang memberitakan bahwa banjir merendam sebelas desa di empat kecamatan dengan ketinggian air lebih dari satu meter. Pada tahun 2019, di Mojokerto juga terjadi banjir dengan ketinggian air mencapai 20 sampai 30 cm yang memiliki potensi semakin bertambah (FaktualNews.co, 2019). Tata kelola lahan yang buruk, karakteristik wilayah dan curah hujan yang cenderung meningkat dapat menyebabkan terjadinya banjir.

Ir. Anggraheny.MSc, pakar banjir ITS Surabaya mengungkapkan bahwa salah satu penyebab banjir karena kurangnya tatanan pembangunan drainase dan minimnya ruang resapan air (Beritajatim.com, 2016). Jika pembangunan drainase dilakukan saat musim hujan, maka kerugian yang dihasilkan menjadi lebih besar. Untuk meminimalkan kerugian, dibutuhkan pengetahuan tentang kejadian hujan ekstrem pada periode mendatang, atau sering disebut dengan *Early Warning System*. Sehingga setiap pengambilan keputusan perlu mempertimbangkan kondisi cuaca, terutama curah hujan ekstrem.

Dalam ilmu statistik, metode yang digunakan untuk mengidentifikasi adanya kejadian ekstrem yaitu *Extreme Value Theory* (EVT). EVT berfokus pada perilaku ekor suatu distribusi dan digunakan untuk memodelkan kejadian-kejadian ekstrem (Dharmawan, 2012). Terdapat dua pendekatan yang digunakan dalam EVT, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT). Pendekatan BM menghasilkan distribusi nilai ekstrem berupa *Generalized Extreme Value* (GEV), sedangkan POT menghasilkan distribusi nilai ekstrem berupa *Generalized Pareto Distribution* (GPD). Pada *Block Maxima* memiliki kelemahan jika digunakan pada data curah hujan karena mengabaikan nilai-nilai yang memiliki kemungkinan menjadi nilai ekstrem pada tiap *block*. Sehingga Fawcett (2016) merekomendasikan POT karena dapat menggunakan data pengamatan yang lebih efisien dibandingkan dengan pendekatan BM. Penelitian mengenai perbandingan BM dan POT pernah dilakukan oleh (Daniela Jaruskova, 2006) yang menyimpulkan bahwa metode POT lebih sesuai daripada metode BM untuk data ekstrem yang dependen dalam mengestimasi *return level* pada data curah hujan di Moravia Utara, Republik Ceko. Hasil yang sama juga disampaikan oleh (Wahyudi; Sutikno, 2010). Penelitian tersebut menghasilkan kesimpulan bahwa pendekatan POT memberikan hasil yang lebih sesuai daripada pendekatan BM untuk memprediksi *return level* curah hujan di Ngawi. Penelitian lain juga dilakukan oleh (Kurniawati, 2013) yang melakukan



analisis data ekstrem non stasioner pada kasus curah hujan di Jawa Timur dengan pendekatan POT dengan melakukan pemodelan parameter skala pada GPD serta menentukan *return level*. Penelitian tersebut menghasilkan kesimpulan bahwa model parameter GPD diperoleh dengan melakukan transformasi data ekstrem dependen sesuai dengan pola data berupa siklik atau linear tren.

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 metode *Declustering*. Terdapat beberapa penelitian yang pernah dilakukan untuk menganalisis kasus data ekstrem menggunakan metode *Declustering*. Arapi (2011) yang melakukan perbandingan metode *Declustering* pada data tinggi gelombang laut. Pendekatan yang dilakukan oleh penelitian tersebut adalah *Standard Storm Length* (SSL), *Runs Declustering* (RL), *Declustering Algorithm* (DeCA), dan *Triple Annual Maximum* (TAM). Kesimpulan yang diperoleh dari penelitian tersebut adalah dengan metode *Declustering* melalui pendekatan *Runs Declustering* menunjukkan hasil dengan standar eror yang lebih kecil dan selang kepercayaan yang lebih sempit. Penelitian mengenai *Runs Declustering* juga pernah dilakukan oleh Malika (2014), penelitian tersebut menghasilkan kesimpulan bahwa *Declustering* dengan pendekatan *Runs Declustering* pada data curah hujan ekstrem dependen di sentra produksi padi Jawa Timur dapat mengatasi dependensi data sehingga menghasilkan data yang independen.

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. Pada pembuatan drainase, nilai *return level* akan memberikan informasi pembuatan drainase yang tepat dan volume drainase yang sesuai sehingga dapat mengurangi kerugian saat pembuatan

drainase. Selain itu informasi nilai *return level* data curah hujan dapat digunakan untuk mengantisipasi adanya banjir.

Berdasarkan uraian yang telah dipaparkan, maka penelitian ini menggunakan EVT dengan pendekatan POT untuk mengamati 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.

## 1.2 Rumusan Masalah

Permasalahan mengenai curah hujan ekstrem seringkali menggunakan data dependen karena kejadian di alam cenderung terjadi berurutan/membentuk pola tertentu. Namun pada EVT, asumsi yang harus dipenuhi yaitu data harus independen. Oleh karena itu diperlukan penyelesaian yang dapat mengatasi dependensi data menggunakan *Declustering* dan akan dibandingkan dengan proses *Non Declustering*. Setelah itu akan dilakukan prediksi curah hujan ekstrem kedepannya dengan menghitung *return level*.

Berdasarkan uraian tersebut, maka rumusan masalah yang diselesaikan dalam penelitian ini adalah sebagai berikut.

1. Bagaimana karakteristik curah hujan harian untuk masing-masing pos pengamatan di Surabaya dan Mojokerto?
2. Bagaimana hasil estimasi parameter GPD pada data curah hujan ekstrem yang melalui proses *Declustering* dan *Non Declustering* untuk masing-masing pos pengamatan di Surabaya dan Mojokerto?
3. Bagaimana kebaikan estimasi *return level* curah hujan ekstrem yang melalui proses *Declustering* dan *Non Declustering* untuk masing-masing pos pengamatan di Surabaya dan Mojokerto?

### 1.3 Tujuan

Berdasarkan latar belakang dan rumusan masalah, tujuan yang ingin dicapai dalam penelitian ini antara lain sebagai berikut.

1. Mengetahui karakteristik curah hujan harian untuk masing-masing pos pengamatan di Surabaya dan Mojokerto.
2. Mengetahui hasil estimasi parameter GPD pada data curah hujan ekstrem yang melalui proses *Declustering* dan *Non Declustering* untuk masing-masing pos pengamatan di Surabaya dan Mojokerto.
3. Membandingkan kebaikan estimasi *return level* curah hujan ekstrem yang melalui proses *Declustering* dan *Non Declustering* untuk masing-masing pos pengamatan di Surabaya dan Mojokerto.

### 1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Bagi pemerintah dan masyarakat Surabaya dan Mojokerto, dapat memberikan informasi mengenai intensitas curah hujan harian pada masa yang akan datang, sehingga dapat dibuat perkiraan dan mitigasi terjadinya banjir.
2. Bagi BMKG, dapat digunakan sebagai tambahan informasi *climate early warning system* dan digunakan sebagai metode alternatif untuk prakiraan iklim ekstrem.
3. Hasil penelitian dapat digunakan sebagai tambahan referensi untuk penelitian-penelitian selanjutnya.

### 1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut.

1. Variabel yang digunakan adalah curah hujan harian.
2. Data yang digunakan adalah data curah hujan harian yang dibagi menjadi data *training* dan data *testing*. Data *training* terdiri dari data curah hujan harian dari tahun 2003 sampai

2016. Sedangkan data *testing* yang digunakan pada penelitian ini adalah data curah hujan harian tahun 2017.

3. Pos pengamatan pada penelitian ini adalah Pos Kandangan, Pos Kebon Agung, Pos Gedeg, Pos Janjing, Pos Klegen, Pos Mojosari, dan Pos Pacet.
4. Metode estimasi parameter pada penelitian ini adalah *Maximum Likelihood Estimation* (MLE).

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Statistika Deskriptif

Statistika deskriptif adalah metode pengumpulan data dengan merangkum dan menggambarkan karakteristik dari data (Johnson & Bhattacharyya, 2010). Yang termasuk dalam statistika deskriptif yaitu ukuran pemusatan data, ukuran penyebaran data, dan penyajian data dalam bentuk grafik maupun diagram. Ukuran pemusatan data dan ukuran penyebaran data digunakan untuk mendeskripsikan data berupa data numerik. Yang sering digunakan pada ukuran pemusatan data pada kasus curah hujan harian adalah rata-rata. Rata-rata adalah hasil pembagian antara jumlah nilai setiap observasi dengan banyaknya data observasi yang dapat dituliskan dengan persamaan (2.1).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.1)$$

dengan,

- $\bar{x}$  : Rata-rata
- $x_i$  : Observasi ke- $i$  ;  $i=1,2,\dots,n$
- $n$  : Banyaknya observasi

Ukuran penyebaran data yang paling sering digunakan untuk kasus curah hujan harian adalah nilai maksimum, nilai minimum, dan standar deviasi. Nilai maksimum adalah nilai tertinggi/terbesar yang terdapat dalam segugus data. Nilai minimum adalah nilai terendah yang terdapat dalam sekumpulan data. Sedangkan standar deviasi merupakan akar dari kuadrat simpangan dari semua nilai data terhadap rata-rata yang ditunjukkan pada persamaan (2.2).

$$s = \sqrt{\frac{\sum_{i=1}^n x_i - \bar{x}^2}{n-1}} \quad (2.2)$$

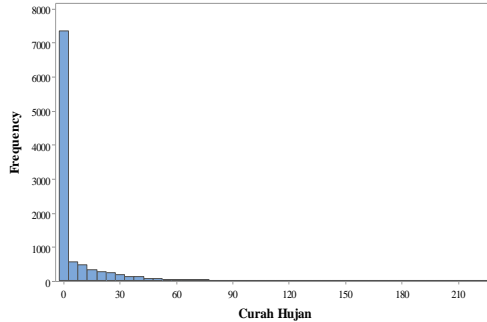
*Skewness* atau kemencengan merupakan derajat ketidasimetrisan atau dapat juga didefinisikan sebagai penyimpangan kesimetrisan dari suatu distribusi. Jika suatu kurva frekuensi dari suatu distribusi memiliki ekor kurva yang lebih panjang ke arah sisi kanan dibandingkan ke arah sisi kiri dari nilai maksimum tengah, maka distribusi ini dikenal dengan nama distribusi miring ke kanan, atau memiliki kemencengan positif. Untuk kondisi sebaliknya, distribusinya dikenal sebagai distribusi miring ke kiri atau memiliki kemencengan negatif. *Skewness* dapat dihitung dengan persamaan berikut.

$$\tau = \frac{\frac{1}{n} \sum_{i=1}^n x_i - \bar{x}^3}{\left[ \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \right]^2} \quad (2.3)$$

Kurtosis adalah derajat ketinggian puncak atau keruncingan suatu distribusi. Kurtosis dapat dihitung dengan persamaan sebagai berikut.

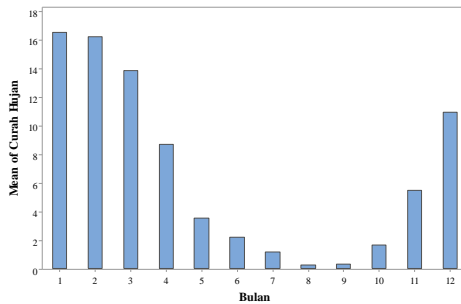
$$K = \frac{\frac{1}{n} \sum_{i=1}^n x_i - \bar{x}^4}{\left[ \frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \right]^2} \quad (2.4)$$

Ukuran yang digunakan untuk menyatakan keruncingan kurva distribusi atau kurtosis ini menggunakan momen keempat di sekitar nilai *mean*. Selain berupa angka, penyajian data yang digunakan pada data curah hujan harian dapat berupa histogram maupun bar *chart*. Histogram digunakan untuk menggambarkan adanya frekuensi distribusi data. Tampilan dari histogram pada data curah hujan dapat dilihat pada Gambar 2.1. Histogram merupakan grafik yang berbentuk batang dengan tinggi yang berbeda yang sisinya saling berdekatan. Tinggi setiap batang mewakili frekuensi nilai-nilai di kelas yang diwakili oleh setiap batang.



**Gambar 2.1** Contoh Histogram

Selain histogram, penyajian data curah hujan harian juga dapat berupa *bar chart*. Berikut merupakan tampilan dari *bar chart* pada data curah hujan.



**Gambar 2.2** Contoh Bar Char

## 2.2 Pengujian Kesesuaian Distribusi

Pengujian kesesuaian distribusi dilakukan untuk memeriksa kesesuaian distribusi teoritis dengan distribusi empirik. Pengujian kesesuaian distribusi secara formal dapat dilakukan dengan uji *Kolmogorov-Smirnov*. Pengujian ini dilakukan dengan menyesuaikan fungsi distribusi empiris (berdasarkan sampel)  $F_n(x)$  dengan distribusi teoritis tertentu (sesuai yang hipotesis)  $F_0(x)$ . Hipotesis uji distribusi dapat dituliskan sebagai berikut.

$H_0 : F_n(x) = F_0(x)$ , yang artinya data mengikuti distribusi teoritis.

$H_0 : F_n(x) \neq F_0(x)$ , yang artinya data tidak mengikuti distribusi teoritis.

Statistik uji yang digunakan adalah :

$$D_{hitung} = \text{Maks} |F_n(x) - F_0(x)| \quad (2.5)$$

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  $\alpha$ .  $F_n(x)$  adalah nilai peluang kumulatif (fungsi distribusi kumulatif) berdasarkan data sampel.  $F_0(x)$  adalah nilai peluang kumulatif (fungsi distribusi kumulatif) dibawah  $H_0$ ,  $P Z < Z_i$  (Daniel, 1989).

### 2.3 Autocorrelation Function

*Autocorrelation function* (ACF) merupakan suatu hubungan linear antara pengamatan  $x_i$  dengan pengamatan  $x_{i-k}$  pada data *time series* yang dipisahkan oleh waktu ke- $k$  (Wei, 2006). ACF digunakan untuk mengukur dependensi data *time series*. ACF juga dapat melihat korelasi antar lag, melihat pola data yang berbentuk tren maupun periodik. Rumus umumnya sebagai berikut (Rinaldi, Djuraidah, Wigena, Wayan Mangku, & Gunawan, 2018).

$$\rho_k = \frac{\text{COV } x_i, x_{i+k}}{\sqrt{\text{var } x_i} \sqrt{\text{var } x_{i+k}}} \quad (2.6)$$

Penduga dari ACF adalah sebagai berikut.

$$\hat{\rho}_k = \frac{\sum_{i=1}^n x_i - \bar{x} \quad x_{i+k} - \bar{x}}{\sqrt{\sum_{i=1}^n x_i - \bar{x}^2} \sqrt{\sum_{i=1}^n x_{i+k} - \bar{x}^2}} \quad (2.7)$$



## 2.4 Extremogram

*Extremogram* adalah metode yang mengukur dependensi ekstrem untuk satu peubah runtun waktu saja. Secara sederhana dapat dikatakan konsep *extremogram* adalah menduga peluang terjadinya kejadian ekstrem dalam waktu  $i+k$  dengan syarat ada kejadian ekstrem pada periode  $i$ . Misalkan  $X_i$  vektor acak yang sudah independen yang berada di ruang  $\mathbb{R}^d$ , serta diketahui himpunan Borel  $A$  dan  $B$  dalam  $\overline{\mathbb{R}^d}$  sedemikian hingga  $C = A \times \overline{\mathbb{R}^{d-k-1}} \times B$  nilainya terbatas dan menjauhi nol, maka *extremogram* didefinisikan untuk dua himpunan  $A$  dan  $B$  sebagai berikut.

$$\rho_k = \lim_{x \rightarrow \infty} P \left( x^{-1}x_k \in B \mid x^{-1}x_0 \in A \right), k = 0, 1, 2, \dots \quad (2.8)$$

jika limitnya ada. Karena himpunan  $A$  dan  $B$  nilainya terbatas dan menjauhi nol, maka kejadian  $x^{-1}x_0 \in A$  dan  $x^{-1}x_k \in B$  menjadi kejadian ekstrem yang ditandai dengan nilai peluang yang konvergen menuju 0 ketika  $x \rightarrow \infty$ . Untuk melakukan penduga *extremogram* dari persamaan (2.8), limit dari  $x$  diganti dengan  $x_u$  yang didefinisikan sebagai data diatas *threshold*. Nilai  $k$  merupakan nilai dari lag waktu. Penduga *extremogram* dapat ditulis sebagai.

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k} I_{x_u^{-1}x_{i+k} \in B, x_u^{-1}x_0 \in A}}{\sum_{i=1}^{n-k} I_{x_u^{-1}x_i \in A}} \quad (2.9)$$

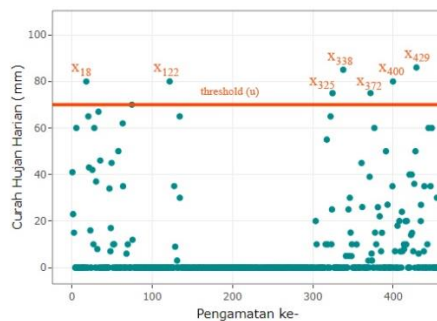
## 2.5 Extreme Value Theory (EVT)

*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. Kajian nilai ekstrem di bidang iklim memberikan

informasi bahwa sebagian besar data iklim memiliki ekor distribusi yang *heavy-tail*, yaitu ekor distribusi yang turun secara lambat bila dibandingkan dengan distribusi normal. Hal tersebut memberikan dampak peluang terjadinya nilai ekstrem lebih besar daripada distribusi normal. EVT menjadi salah satu aplikasi statistika yang penting untuk ilmu terapan selama lebih dari 50 tahun (Coles, 2001). Terdapat dua pendekatan untuk mengidentifikasi pergerakan nilai ekstrem, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT) (Mcneil, 1999).

### 2.5.1 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. Semakin tinggi nilai *threshold* maka fungsi distribusi akan mendekati *Generalized Pareto Distribution* (GPD) (Gilli & K ellezi, 2006).



**Gambar 2.3** Ilustrasi Pengambilan Sampel Data Ekstrem POT

Gambar 2.3 menunjukkan cara pengambilan data ekstrem menggunakan POT. Data  $x_{18}$ ,  $x_{122}$ ,  $x_{325}$ ,  $x_{338}$ ,  $x_{372}$ ,  $x_{400}$ , dan  $x_{429}$  adalah nilai yang berada di atas *threshold* ( $u$ ), sehingga ketujuh data yang berada di atas *threshold* merupakan data ekstrem yang

digunakan untuk analisis selanjutnya. POT mengandung tiga aspek, yakni :

1. Penentuan nilai *threshold* ( $u$ ).
2. Banyaknya kejadian nilai yang melebihi  $u$ , atas periode waktu tertentu merupakan proses *Poisson*.
3. Nilai-nilai yang melebihi  $u$  berdistribusi *Generalized Pareto*.

### 2.5.2 Generalized Pareto Distribution

Pickands (1975) dan Smith (1990) memperkenalkan model nilai ekstrem asimtotik yang menggambarkan distribusi nilai ekstrem dan dikenal dengan GPD. Keunggulan GPD dibandingkan GEV ialah penggunaan informasi sampel yang lebih efisien dalam proses penaksiran ekor distribusi daripada GEV (Lee Fawcett; David Walshaw, 2016). Misal  $X$  variabel random GPD, *Cumulative Distribution Function* (CDF) dinyatakan sebagai (Coles, 2001; Pickands, 1975).

$$F(x-u) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{(x-u)}{\sigma}\right), & \xi = 0 \end{cases} \quad (2.10)$$

Untuk mengetahui *Probability Density Function* (PDF) GPD pada  $\xi \neq 0$  dapat digunakan persamaan sebagai berikut.

$$\begin{aligned} \frac{dF(x-u)}{dx} &= \frac{d\left(1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}}\right)}{dx} \\ &= -\left(-\frac{1}{\xi}\right)\left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}-1}\left(\frac{\xi}{\sigma}\right) \\ &= \frac{1}{\sigma}\left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}-1} \end{aligned} \quad (2.11)$$

Dengan  $\xi$  ialah parameter *shape*,  $u$  ialah nilai *threshold*, dan  $\sigma$  adalah parameter *scale*. Jika  $\xi \leq 0$  maka  $0 \leq (x-u) < \infty$  dan  $0 \leq (x-u) \leq u + \frac{\sigma}{\xi}$ , jika  $\xi > 0$ . Nilai batas atas setiap pengamatan dalam distribusi ini ialah  $u + \frac{\sigma}{\xi}$ . Sedangkan untuk PDF GPD untuk  $\xi \neq 0$  adalah sebagai berikut.

$$\begin{aligned} \frac{dF(x-u)}{dx} &= \frac{d\left(1 - \exp\left(-\frac{(x-u)}{\sigma}\right)\right)}{dx} \\ &= -\left(-\frac{1}{\sigma}\right) \exp\left(-\frac{x-u}{\sigma}\right) \\ &= \frac{1}{\sigma} \exp\left(-\frac{x-u}{\sigma}\right) \end{aligned} \quad (2.12)$$

GPD dibedakan menjadi tiga tipe berdasarkan nilai parameter *shape*. Ketiga tipe GPD dapat diamati melalui bentuk penurunan fungsi distribusi pada bagian ekornya. Ketiga tipe GPD memiliki CDF seperti disajikan pada persamaan (2.13) sampai persamaan (2.15) sebagai berikut.

1. Tipe 1 berdistribusi Eksponensial, yang berarti penurunan secara eksponensial dari ekor distribusi terjadi jika nilai  $\xi = 0$ .

$$F(x-u) = 1 - \exp\left(-\frac{x-u}{\sigma}\right), -\infty < x-u < \infty \quad (2.13)$$

2. Tipe 2 berdistribusi Pareto, yang berarti ekor distribusi menurun secara perlahan dan tidak mempunyai batas atas jika nilai  $\xi > 0$ .

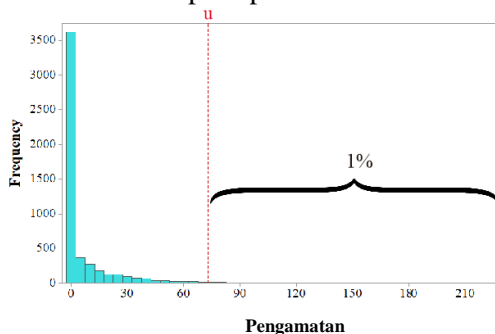
$$F(x-u) = \begin{cases} 0 & , x \leq u \\ 1 - \left(1 + \frac{x-u}{\sigma}\right)^{-\frac{1}{\xi}} & , x > u \end{cases} \quad (2.14)$$

3. Tipe 3 berdistribusi Pareto tipe 2, yang berarti distribusi ekor mempunyai batas atas jika nilai  $\xi < 0$ .

$$F_{x-u} = \begin{cases} 1 - \left(1 - \frac{x-u}{\sigma}\right)^{\frac{1}{\xi+1}}, & x < u \\ 1 & x \geq u \end{cases} \quad (2.15)$$

### 2.5.3 Penentuan Nilai Threshold

Sebelum mendapatkan nilai ekstrem dengan metode POT, terlebih dahulu dilakukan penentuan nilai *threshold* ( $u$ ) sebagai nilai acuan. Jika suatu nilai observasi melebihi nilai acuan, maka nilai tersebut diidentifikasi sebagai nilai ekstrem. Penentuan *threshold* bertujuan untuk mencari keseimbangan yang optimal agar diperoleh parameter *error* seminimal mungkin. Nilai *threshold* yang terlalu rendah akan menghasilkan parameter bias, sedangkan nilai *threshold* yang terlalu tinggi akan menyebabkan data ekstrem yang diperoleh akan sedikit sehingga tidak cukup untuk menaksir model, akibatnya *variance* yang dihasilkan besar (Coles, 2001). Secara visual penentuan *threshold* dengan metode persentase diilustrasikan seperti pada Gambar 2.4 sebagai berikut.



**Gambar 2.4** Ilustrasi Penentuan Threshold dengan Metode Persentase

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 (Li, 2015).

1. Mengurutkan data dari yang terbesar hingga yang terkecil.
2. Menghitung 1 persen dari banyak data ( $k$ ).

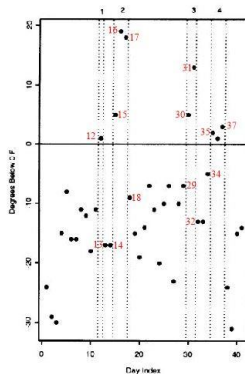
$$k = 1\% \times N \quad (2.16)$$

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

### 2.5.4 *Declustering*

Sejumlah teknik *Declustering* telah dikembangkan dengan prosedur yang mengarah pada identifikasi *cluster* maksima, yaitu observasi maksimum (puncak) pada masing-masing *cluster* yang terbentuk dan diasumsikan identik independen (iid). 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.



**Gambar 2.5** Ilustrasi Run Length pada Data Temperatur dengan nilai  $r=1$   
(Sumber: *An Introduction to Statistical Modeling of Extreme Value* Coles, 2001)

Berdasarkan ilustrasi pada Gambar 2.5 metode *Runs Declustering* ( $r=1$ ) pada Gambar 2.5 menjelaskan pembentukan *cluster* dimana *cluster* 1 dimulai ketika observasi pertama kali melewati *threshold* ( $u$ ) yaitu pada observasi ke-12, dan berhenti apabila terdapat minimal 1 pengamatan yang berada di bawah *threshold* yaitu observasi ke-13 dan 14. Sehingga anggota *cluster* 1 hanya terdiri atas pengamatan ke-12 saja. Pembentukan *cluster* selanjutnya yaitu *cluster* 2 dimulai ketika ada observasi yang melewati *threshold* lagi (oservasi ke-15) dan berhenti pada saat terdapat paling sedikit 1 observasi yang berada di bawah *threshold* (observasi ke-18), sehingga anggota *cluster* 2 terdiri atas pengamatan ke-15, 16, dan 17. Proses ini terus berlanjut sampai pembentukna semua *cluster* selesai dan data telah habis. Berdasarkan ilustrasi pada Gambar 2.5 menunjukkan bahwa proses *Declustering* dengan  $r=1$  dihasilkan 4 *cluster*. Nilai maksimum dari masing-masing *cluster*, yaitu observasi ke-12, 16, 31, dan 37 selanjutnya yang digunakan sebagai data baru untuk menentukan estimasi parameter *Generalized Pareto Distribution* (GPD). Terdapat isu yang penting mengenai pemilihan *threshold* dan panjang  $r$ . Jika pemilihan *threshold* dan  $r$  tidak tepat dapat menyebabkan bias atau varian tinggi.

Nilai  $r$  yang terlalu rendah menyebabkan data sulit menjadi dependen sedangkan nilai  $r$  yang terlalu tinggi maka data ekstrem yang diperoleh sedikit sehingga menghasilkan varians yang besar. Penentuan  $r$  yang optimal dapat menggunakan *extremal indeks*.

$$\theta_n u = \begin{cases} \min \left\{ 1, \frac{2 \left( \sum_{i=1}^{N-1} T_i \right)^2}{n_u - 1 \sum_{i=1}^{N-1} T_i^2} \right\} & \text{if } \max T_i : 1 \leq i \leq n_u - 1 \leq 2 \\ \min \left\{ 1, \frac{2 \left( \sum_{i=1}^{N-1} T_i - 1 \right)^2}{n_u - 1 \sum_{i=1}^{N-1} T_i - 1 T_i - 2} \right\} & \text{if } \max T_i : 1 \leq i \leq n_u - 1 > 2 \end{cases} \quad (2.17)$$

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

$T_i$  : jarak antar pengamatan di atas *threshold*

$n_u$  : banyak pengamatan di atas *threshold*

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

*Extremal indeks* merupakan suatu indikator untuk mengukur berapa banyak *cluster* yang terbentuk sehingga data menjadi independen. Salah satu metode dalam mengestimasi *extremal indeks* adalah dengan menggunakan persamaan 2.17 (Ferro & Segers, 2002). Estimator pada *Extremal Index* menghasilkan nilai yang disebut dengan *Runs Estimator*. Nilai *Runs Estimator* dapat digunakan untuk menentukan *run length* ( $r$ ) yang optimal untuk proses *Declustering*, dimana  $r = T_{(c)}$  dan  $c - 1 = \theta n_u - 1$

Pengembangan penelitian mengenai *external index* juga dilakukan oleh Ferro (2003) yang menghasilkan estimator *extremal index* berupa *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*. Hal ini terlihat pada *minimum threshold*, dimana *Intervals Estimator* memiliki jangkauan probabilitas lebih dekat dengan nilai nominal (Ferro & Segers, 2003).

Penelitian Ferro (2003) menunjukkan fakta bahwa *threshold optimal* untuk *Interval Estimator* biasanya lebih rendah dibandingkan *Runs Estimator*. Alasan karena sensitivitas *Interval 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 Ferro (2003). Setiap pilihan  $r$  untuk *Runs Estimator* memiliki kinerja yang unggul jika digunakan untuk *maximum threshold* daripada menggunakan *Intervals Estimator*.



### 2.5.5 Return Level

*Return level* adalah suatu level kejadian ekstrem yang diharapkan terlampaui rata-rata sekali dalam setiap periode waktu tertentu (Coles, 2001). Nilai *return level* dihitung dari nilai kuantil yang mempunyai probabilitas terjadi sebesar  $\left(\frac{1}{T}\right)$ , dimana  $T$  adalah panjang periode. Penentuan *return level* pada GPD melibatkan parameter *threshold* ( $u$ ),  $\sigma$ , dan  $\xi$ . Misal  $X$  variabel random dan  $u$  merupakan *threshold*, maka untuk  $x > u$

$$(i) \Pr(X > u) = \zeta_u \quad (2.18)$$

$$(ii) \Pr(X > x) = \zeta_u \left\{ 1 + \hat{\xi} \left( \frac{x-u}{\hat{\sigma}} \right) \right\}^{-\frac{1}{\hat{\xi}}} \quad (2.19)$$

$$(iii) \Pr(X > x | X > u) = \left\{ 1 + \hat{\xi} \left( \frac{x-u}{\hat{\sigma}} \right) \right\}^{-\frac{1}{\hat{\xi}}} \quad (2.20)$$

Berdasarkan Persamaan (2.19), rata-rata curah hujan ekstrem yang melebihi level  $u$  terjadi setiap  $T$  pengamatan adalah

$$\zeta_u \left\{ 1 + \hat{\xi} \left( \frac{x-u}{\hat{\sigma}} \right) \right\}^{-\frac{1}{\hat{\xi}}} = \frac{1}{T} \quad (2.21)$$

sehingga dapat diuraikan nilai *return level* pada  $T$  pengamatan seperti pada persamaan dibawah ini.

$$RL_T \begin{cases} u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ (T\zeta_u)^{\hat{\xi}} - 1 \right], & \hat{\xi} \neq 0 \\ u + \hat{\sigma} \log(T\zeta_u), & \hat{\xi} = 0 \end{cases} \quad (2.22)$$

dengan :

$T$  = periode waktu

$n_u$  = banyaknya data diatas *threshold*

$$\zeta_u = \frac{n_u}{n}$$

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

Misalnya nilai *return level* sebesar 70 mm perhari pada periode 10 tahun disuatu stasiun pengamatan, maka dapat dikatakan ada kemungkinan curah hujan ekstrem yang melebihi level 70 mm perhari sebesar  $\frac{1}{10}$  atau 0,1 rata-rata satu kali pada periode 10 tahun.

## 2.6 Estimasi Parameter *Generalized Pareto Distribution*

Estimasi parameter *Generalized Pareto Distribution* (GPD) dapat ditaksir menggunakan *Maximum Likelihood Estimation* (MLE). Hal utama yang harus dilakukan dalam mengestimasi parameter menggunakan MLE adalah memaksimalkan fungsi *likelihood Probability Density Function* (PDF) dari GPD dimana fungsi *likelihood* merupakan fungsi peluang bersama  $y_1, y_2, \dots, y_n$ , dimana  $y = x - u$  pada persamaan (2.11) dan persamaan (2.12).

$$f_{\sigma, \xi} = \begin{cases} \frac{1}{\sigma} \left( 1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi} - 1}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases} \quad (2.23)$$

Hasil *likelihood* pada persamaan (2.23) yaitu persamaan (2.24) sebagai berikut.

$$\begin{aligned} L(\sigma, \xi) &= f(y_1) f(y_2) \dots f(y_n) \\ &= \prod_{i=1}^n \frac{1}{\sigma} \left( 1 + \frac{\xi y_i}{\sigma} \right)^{-\frac{1}{\xi} - 1} \\ &= \sigma^{-n} \prod_{i=1}^n \left( 1 + \frac{\xi y_i}{\sigma} \right)^{-\frac{1}{\xi} - 1} \end{aligned} \quad (2.24)$$

Fungsi  $\ln$  *likelihood* dari persamaan diatas adalah sebagai berikut.

$$\begin{aligned} \ell(\sigma, \xi) &= \ln L(\sigma, \xi) \\ &= -n \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln\left(1 + \frac{\xi y_i}{\sigma}\right) \end{aligned} \quad (2.25)$$

Langkah selanjutnya adalah mendapatkan turunan pertama terhadap parameter  $\sigma$  dan  $\xi$ .

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \left(\frac{\xi y_i}{\sigma^2 + \sigma \xi y_i}\right) \quad (2.26)$$

$$\frac{\partial \ell}{\partial \xi} = \frac{1}{\xi^2} \sum_{i=1}^n \ln\left(1 + \frac{\xi y_i}{\sigma}\right) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} \quad (2.27)$$

Berdasarkan persamaan yang terbentuk pada persamaan (2.26) dan (2.27) diketahui bahwa turunan pertama fungsi  $\ln$  *likelihood* terhadap masing-masing parameter tidak *closed form*, sehingga diperlukan pendekatan secara numerik untuk menyelesaikan persamaan tersebut. Analisis numerik yang digunakan yaitu *Newton Raphson*.

Penggunaannya dilakukan iterasi-iterasi sampai mendapatkan hasil yang konvergen. Persamaan umum *Newton Raphson* sebagai berikut,

$$\theta_{i+1} = \theta_i - g \theta_i H^{-1} \theta_i \quad (2.28)$$

$g \theta$  adalah vektor gradien berukuran  $1 \times p$  dimana  $p$  adalah jumlah parameter.  $g \theta$  berisi turunan pertama PDF GPD terhadap parameter  $\sigma$  dan  $\xi$ .  $H \theta$  adalah matrik Hessian berukuran  $p \times p$  yang berisi turunan kedua terhadap parameter.

$$g \theta = \begin{bmatrix} \frac{\partial \ln L}{\partial \xi} & \frac{\partial \ln L}{\partial \sigma} \end{bmatrix} \quad (2.29)$$

$$H \theta = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \xi^2} & \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} \\ \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} & \frac{\partial^2 \ln L}{\partial \sigma^2} \end{bmatrix} \quad (2.30)$$

Turunan kedua dai fungsi  $\ln$  *likelihood* sebagai berikut.

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{n - \xi - 1}{\sigma^2} \sum_{i=1}^n \frac{2\sigma y_i + \xi y_i^2}{\sigma + \xi y_i^2} \quad (2.22)$$

$$\frac{\partial^2 \ell}{\partial \xi^2} = -\frac{2}{\xi^3} \sum_{i=1}^n \ln\left(1 + \frac{\xi y_i}{\sigma}\right) + \frac{2}{\xi^2} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{y_i^2}{\sigma + \xi y_i^2} \quad (2.23)$$

$$\frac{\partial^2 \ell}{\partial \xi \partial \sigma} = \frac{1}{\sigma} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i^2} \quad (2.24)$$

Iterasi *Newton Raphson* diawali dengan menentukan nilai  $\theta_0$  yang merupakan vektor elemennya berisi  $\hat{\xi}_0$  dan  $\hat{\sigma}_0$ , maka nilai estimasi awal tersebut disubstitusikan pada vektor gradien dan matrik Hessian. Nilai  $\hat{\sigma}_0$  didekati dengan standar deviasi data ekstrem sedangkan  $\hat{\xi}_0$  didapatkan dari substitusi persamaan (2.23) dan (2.24) ke persamaan (2.22). Hasil substitusi dijadikan sama dengan nol, sehingga didapatkan estimasi awal yaitu

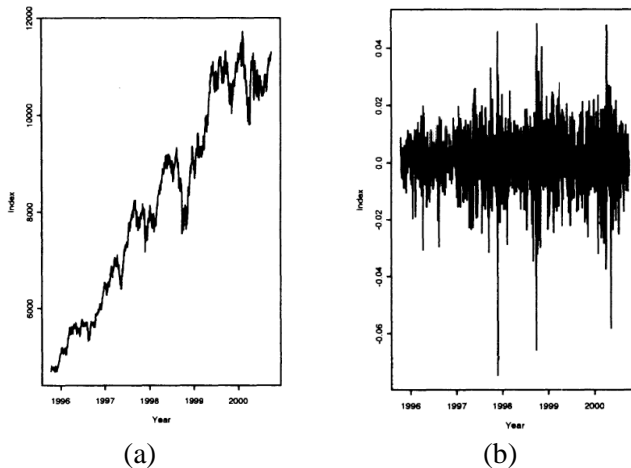
$$\hat{\xi}_0 = \frac{n^2 s - \sum_{i=1}^n x_i - u}{1 - n \sum_{i=1}^n x_i - u} \quad (2.25)$$

Iterasi berhenti apabila  $|\theta_{t+1} - \theta_t| < \varepsilon$  dimana  $\varepsilon$  adalah bilangan yang sangat kecil.

## 2.7 Data Ekstrem Dependen

Analisis nilai ekstrem dengan pendekatan POT tidak menggunakan semua data pengamatan dalam menentukan parameter GPD. Hanya data yang melebihi *threshold* yang digunakan dalam proses penaksiran GPD (Coles, 2001). Penentuan *threshold* pada penelitian ini menggunakan metode persentase. Sehingga jumlah data ekstrem yang dihasilkan menjadi lebih sedikit karena hanya data pengamatan yang melebihi *threshold* yang didefinisikan sebagai nilai ekstrem.

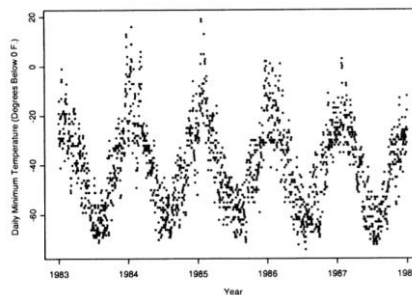
Terdapat banyak teori yang mengasumsikan data ekstrem yang bersifat independen dan identik (IID). Data ekstrem bersifat identik dan independen dikarenakan nilai-nilai ekstrem yang cenderung konstan selama beberapa pengamatan (Gilleland & Katz, 2006).



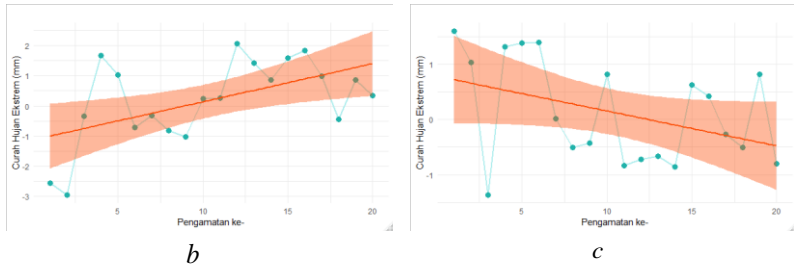
**Gambar 2.6** Pola Data a) Dependen dan b) Independen

(Sumber: *An Introduction to Statistical Modeling of Extreme Value* Coles, 2001)

Data yang dependen mengindikasikan jika data tersebut membentuk suatu pola tertentu, pola yang terbentuk bisa berupa pola siklik, tren naik, maupun tren turun. Berikut merupakan ilustrasi pola data siklik, tren naik, dan tren turun.



*a*



**Gambar 2.7** Pola Data (a) Siklik (b) Tren Naik (c) Tren Turun.

(Sumber: *An Introduction to Statistical Modeling of Extreme Value* Coles, 2001)

Ada beberapa cara yang digunakan untuk mengatasi dependensi data antara lain dengan melakukan transformasi pada data dan kemudian menentukan model parameter distribusinya, serta cara kedua yaitu *Declustering*, yang memfilter pengamatan dependen untuk mendapatkan satu set *threshold* yang independen.

## 2.8 Kriteria Pemilihan Metode

Kriteria pemilihan metode yang sesuai dapat dilakukan dengan *Root Mean Square Error* (RMSE). *Error* merupakan jumlah yang diperkirakan dari kuantitas yang diperkirakan (Warto, 2012). Persamaan untuk mencari nilai RMSE dapat ditulis sebagai berikut.

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t x_i - \hat{x}_i^2} \quad (2.26)$$

dengan :

$x_i$  : Nilai aktual

$\hat{x}_i$  : Nilai dugaan

$t$  : Banyaknya nilai yang diduga

## 2.9 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 (Handoko, 1994). 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 (BMKG, 2018), 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 , lebih dari 100 mm/hari.

Menurut BMKG berdasarkan distribusi data rata-rata curah hujan di Indonesia dibedakan menjadi 3 tipe, yaitu (Kadarsah, 2007) :

#### 1. Tipe Ekuatorial

Pola ekuatorial dicirikan oleh tipe curah hujan dengan bentuk bimodial (dua puncak musim hujan) yang biasanya terjadi sekitar bulan Maret dan Oktober atau pada saat terjadi ekuinoks, yaitu waktu atau peristiwa matahari berada dalam bidang katulistiwa bumi dimana peristiwa ini terjadi dua kali dalam setahun. Sebagian besar wilayah Sumatera dan Kalimantan, sering terjadi curah hujan yang mengikuti pola ini.

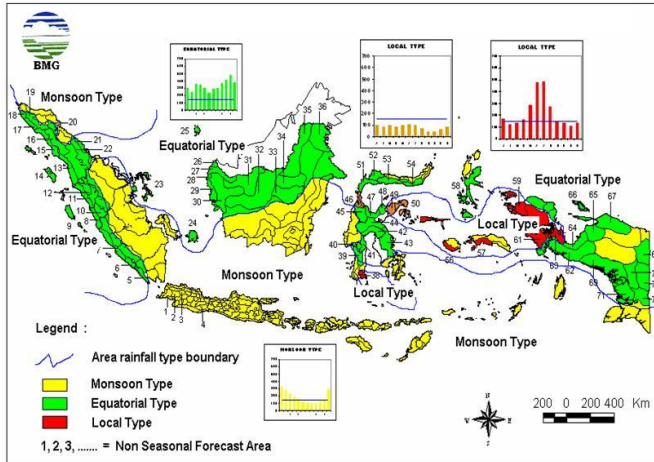
#### 2. Tipe Monsun

Curah Hujan dipengaruhi oleh tiupan angin monsun dan bersifat unimodal (satu puncak musim hujan, DJF (Desember-Januari-Februari) musim hujan, JJA (Juni-Juli-Agustus) musim kemarau). Tipe hujan ini terjadi di wilayah Indonesia bagian selatan, seperti di ujung Pulau Sumatera bagian selatan, Jawa, Bali, Nusa Tenggara dan Maluku Selatan.

#### 3. Tipe Lokal

Curah hujan dipengaruhi oleh kondisi lingkungan setempat, yakni adanya bentang perairan sebagai sumber penguapan dan pegunungan sebagai daerah tangkapan hujan. Pola Curah Hujan lokal memiliki distribusi hujan bulanan kebalikan dengan pola monsun, dicirikan oleh bentuk pola hujan unimodal (satu puncak hujan), tetapi bentuknya berlawanan dengan tipe monsun.

Berikut merupakan ilustrasi dari tipe curah hujan di Indonesia yang disajikan pada gambar 2.8.



**Gambar 2.8** Peta tipe hujan di Indonesia

Curah hujan kumulatif merupakan jumlah hujan yang terkumpul dalam rentang waktu kumulatif tersebut. Dalam periode musim, rentang waktunya adalah rata-rata panjang musim pada masing-masing Zona Musim (ZOM).

## 2.10 Zona Musim (ZOM)

Zona Musim adalah daerah yang pos hujan rata-ratanya memiliki perbedaan yang jelas antara periode Musim Hujan dan Musim Kemarau. Daerah-daerah yang pola hujannya tidak memiliki perbedaan yang jelas antara Musim Hujan dan Musim Kemarau disebut Non ZOM. 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 2.9.





**Gambar 2.9** Peta Zona Musim Jawa Timur  
(Sumber: <https://karangploso.jatim.bmkg.go.id/index.php/informa-si-iklim/peta-zona-musim-di-provinsi-jawa-timur>)

*(Halaman ini sengaja dikosongkan)*

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 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.

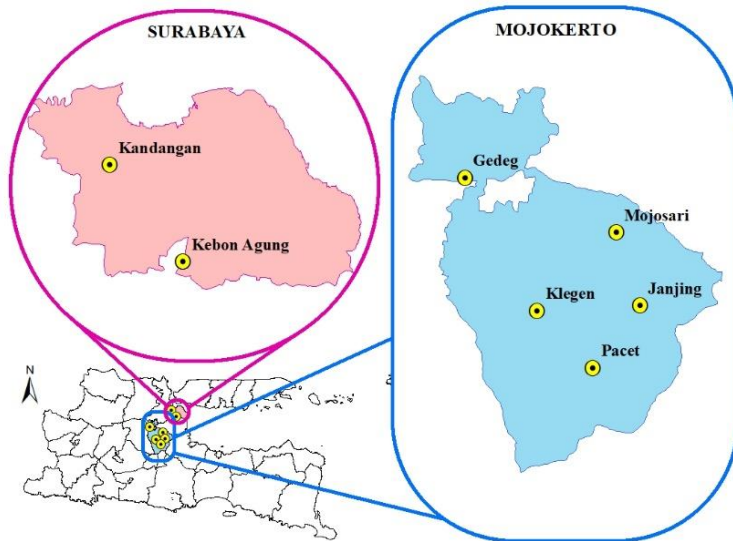
#### 3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel curah hujan harian. Nama pos pengamatan pada penelitian ini, selengkapnya disajikan pada Tabel 3.1. Sedangkan pada Gambar 3.1 merupakan peta persebaran dan letak pos pengamatan yang ada di Surabaya dan Mojokerto.

**Tabel 3.1** Data Pos Hujan Di Surabaya dan Mojokerto

(Sumber: <https://karangploso.jatim.bmkg.go.id/index.php/profil/sitemap/82-peralatan-observasi-klimatologi/55555575-lokasi-penakar-hujan-manual-ombrometer-di-jawa-timur>)

Wilayah	Variabel	Pos Pengamatan	Latitude ( $u$ )	Longitude ( $v$ )
Surabaya	$x_1$	Kandangan	-7.2533	112.655
	$x_2$	Kebon Agung	-7.3275	112.7108
	$x_3$	Klegen	-7.6003	112.4709
	$x_4$	Gedeg	-7.4563	112.3928
Mojokerto	$x_5$	Mojosari	-7.5154	112.5569
	$x_6$	Janjing	-7.5946	112.5821
	$x_7$	Pacet	-7.662	112.5313



**Gambar 3.1** Peta Tujuh Stasiun di Surabaya dan Mojokerto

Pos pengamatan yang digunakan pada penelitian ini diwakili oleh satu pos pengamatan pada tiap zona musim (ZOM) di Surabaya dan Mojokerto.

Berdasarkan Gambar 2.8, total ZOM di Surabaya dan Mojokerto sebanyak 8, diantaranya di Surabaya terdiri dari 3 ZOM dan di Mojokerto terdiri dari 5 ZOM. Namun karena ketersediaan data, di Surabaya terdapat 2 ZOM dan di Mojokerto terdapat 5 ZOM. Sehingga total pos pengamatan yang akan digunakan pada penelitian ini sebanyak 7 pos pengamatan.

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.

Struktur data pada penelitian ini ditunjukkan pada Tabel 3.2

**Tabel 3.2** Struktur Data Penelitian

<b>Tanggal</b>	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
1/1/1997	$x_{1,1}$	$x_{2,1}$	$x_{3,1}$	$x_{4,1}$	$x_{5,1}$	$x_{6,1}$	$x_{7,1}$
2/1/1997	$x_{1,2}$	$x_{2,2}$	$x_{3,2}$	$x_{4,2}$	$x_{5,2}$	$x_{6,2}$	$x_{7,2}$
3/1/1997	$x_{1,3}$	$x_{2,3}$	$x_{3,3}$	$x_{4,3}$	$x_{5,3}$	$x_{6,3}$	$x_{7,3}$
...	...	...	...	...	...	...	...
29/12/2016	$x_{1,5112}$	$x_{2,5112}$	$x_{3,5112}$	$x_{4,5112}$	$x_{5,5112}$	$x_{6,5112}$	$x_{7,5112}$
30/12/2016	$x_{1,5113}$	$x_{2,5113}$	$x_{3,5113}$	$x_{4,5113}$	$x_{5,5113}$	$x_{6,5113}$	$x_{7,5113}$
31/12/2016	$x_{1,5114}$	$x_{2,5114}$	$x_{3,5114}$	$x_{4,5114}$	$x_{5,5114}$	$x_{6,5114}$	$x_{7,5114}$

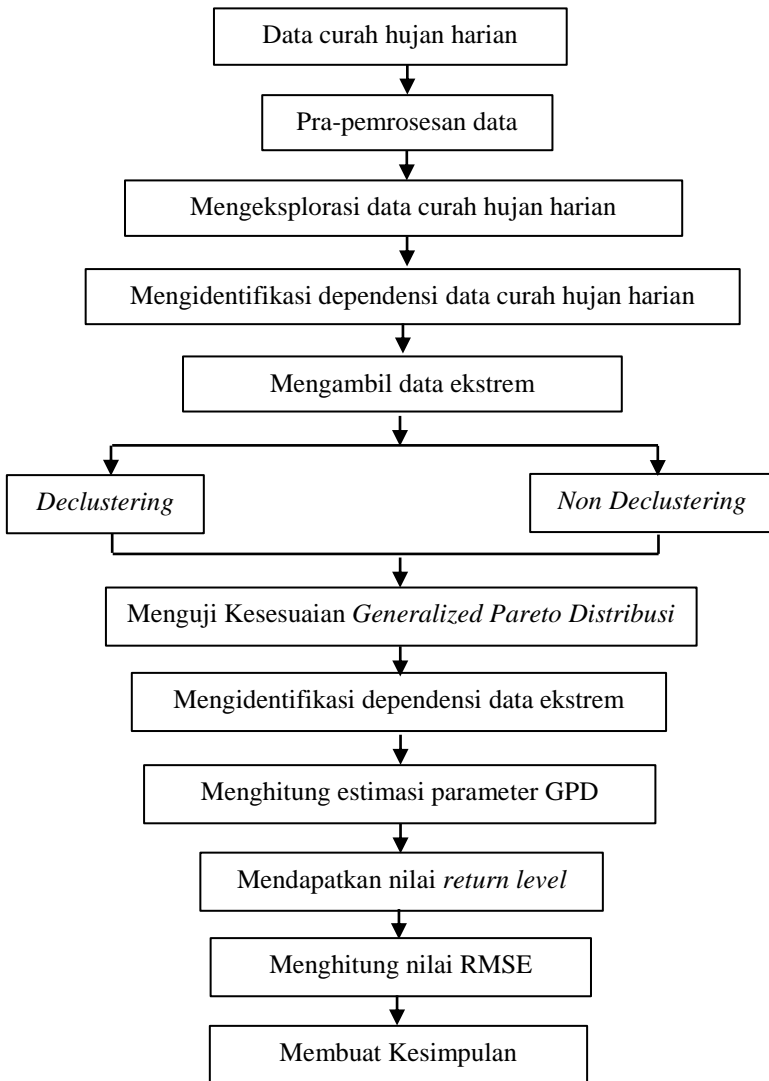
### 3.3 Langkah Analisis

Langkah-langkah analisis yang akan dilakukan berdasarkan tujuan penelitian ini yaitu.

1. Melakukan pra pemrosesan data dengan mengidentifikasi *missing value*, *noisy*, dan *inconsistent*.
2. Mengeksplorasi data curah hujan harian dengan satatitsika deskriptif disertai plot histogram dan *bar chart*.
3. Mengidentifikasi dependensi data curah hujan harian di masing-masing stasiun pengamatan.
4. Pengambilan data ekstrem untuk masing-masing pos pengamatan di Surabaya dan Mojokerto dengan metode *Peaks Over Threshold* (POT) adalah sebagai berikut.
  - a. Menentukan satu nilai *threshold* dengan menggunakan metode persentase.
    - i) Mengurutkan data dari yang terbesar hingga yang terkecil

- ii) Menghitung satu persen dari banyak data dengan menggunakan persamaan (2.8)
  - iii) Menentukan nilai *threshold*, yaitu data yang berada pada urutan ke  $-(k+1)$
  - b. Menghimpun data pengamatan yang nilainya melebihi *threshold*.
5. Melakukan *Declustering* melalui pedekaran *Runs Declustering*. Berikut merupakan langkah-langkah pada pendekatan *Runs Declustering*.
    - i. Menggunakan nilai *threshold* ( $u$ ) pada langkah kedua untuk mengidentifikasi nilai-nilai ekstrem.
    - ii. Menentukan *cluster* yang dimulai ketika *threshold* ( $u$ ) terlampaui untuk pertama kalinya dan berakhir setelah paling sedikit  $r$  pengamatan berturut-turut jauh dibawah *threshold*.
  6. Pemeriksaan kesesuaian *Generalized Pareto Distribution* menggunakan uji *Kolmogorov-Smirnov*.
  7. Mengidentifikasi dependensi data ekstrem pada masing-masing pos pengamatan membentuk pola siklik atau linear tren melalui plot *extremogram*.
  8. Mendapatkan estimasi parameter GPD data ekstrem melalui proses *Declustering* dan *Non Declustering* menggunakan *Maximum Likelihood Estimation* (MLE) untuk masing-masing stasiun pengamatan di Surabaya dan Mojokerto.
  9. Menghitung *Return Level* pada data ekstrem setelah *Declustering* dan *Non Declustering*.
  10. Membandingkan kebaikan estimasi *Return Level* pada data curah hujan ekstrem melalui proses *Decustering* dan *Non Declustering* berdasarkan RMSE untuk masing-masing stasiun pengamatan di Surabaya dan Mojokerto
  11. Menarik kesimpulan.

Langkah-langkah analisis dalam penelitian dapat ditunjukkan menggunakan diagram alir yang disajikan pada Gambar 3.2 berikut ini.



**Gambar 3.2** Diagram Alir Penelitian

*(Halaman ini sengaja dikosongkan)*



## BAB IV

### ANALISIS DAN PEMBAHASAN

Pada bab ini dilakukan analisis dan pembahasan mengenai *Extreme Value Theory* data curah hujan di tujuh pos pengamatan. Pembahasan diawali dengan tahap pra-pemrosesan data curah hujan, dilanjut dengan deskripsi curah hujan, lalu pengambilan data ekstrem menggunakan *Peaks Over Threshold* dan *Declustering*. Kemudian dilakukan uji kesesuaian GPD dan identifikasi dependensi data ekstrem. Jika data ekstrem sudah independen, maka dapat dilakukan analisis selanjutnya yaitu estimasi parameter GPD dan estimasi *return level*.

#### 4.1 Pra-Pemrosesan Data Curah Hujan

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. Oleh karena itu, perlu dilakukan tahap pra-pemrosesan data sebelum dilakukan analisis lebih lanjut. Berikut hasil identifikasi *missing value* dari data curah hujan harian pada tujuh pos pengamatan di Surabaya dan Mojokerto.

**Tabel 4.1** Identifikasi *Missing Value* Data Curah Hujan

<b>Pos Pengamatan</b>	<b><i>N</i></b>	<b><i>Missing Value</i> (%)</b>
Kandangan	5480	1,132
Kebon Agung	5480	1,898
Gedeg	5480	0,000
Janjing	5480	0,548
Klegen	5480	0,000
Mojosari	5480	0,000
Pacet	5480	0,000

Berdasarkan Tabel 4.1 dapat diketahui bahwa pos pengamatan 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 pada pos pengamatan di tanggal dan bulan yang sama dengan *missing value* tersebut. Sedangkan untuk Pos Gedeg, Pos Klegen, Pos Mojosari, dan Pos Pacet tidak ada *missing value*.

## 4.2 Eksplorasi Data Curah Hujan

Setelah dilakukan pra-pemrosesan data, maka dilakukan eksplorasi data menggunakan statistika deskriptif yang bertujuan untuk mendapatkan gambaran umum pada data curah hujan. Selain itu, untuk mengetahui karakteristik pola data curah hujan disajikan dalam bentuk histogram dan *bar chart*.

### 4.2.1 Deskripsi Data Curah Hujan

Deskripsi curah hujan di tujuh pos pengamatan perlu dilakukan sebagai informasi awal untuk mendapatkan gambaran umum dari data curah hujan. Data curah hujan yang digunakan terdapat pada Lampiran 1 sampai Lampiran 7. Berikut merupakan hasil statistika deskriptif pada data curah hujan di tujuh pos pengamatan yang disajikan pada tabel 4.2.

**Tabel 4.2** Statistika Deskriptif Curah Hujan Harian (mm)

Pos Pengamatan	<i>N</i>	Rata-Rata	StDev	Min	Max
Kandangan	5114	6,137	13,808	0	130
Kebon Agung	5114	6,175	13,991	0	114
Gedeg	5114	5,372	13,185	0	126
Janjing	5114	5,516	14,285	0	105
Klegen	5114	5,130	13,584	0	120
Mojosari	5114	5,064	13,836	0	162
Pacet	5114	6,678	15,543	0	225

Keterangan :

StDev= Standar Deviasi      Min=Minimum      Max=Maksimum

Tabel 4.2 memberikan informasi bahwa rata-rata curah hujan harian tertinggi berada di pos Pacet yaitu sebesar 6,678 mm per hari, sedangkan rata-rata curah hujan terendah tercatat di pos

Mojosari yaitu sebesar 5,064 mm per hari. Tabel 4.2 juga menunjukkan nilai 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 mm per hari dan Pos Gedeg memiliki keragaman curah hujan yang terkecil yaitu sebesar 13,185. Sementara untuk curah hujan minimum pada semua 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. Curah hujan maksimum yang tercatat pada tujuh pos pengamatan termasuk dalam kategori curah hujan sangat lebat. Hal ini dikarenakan, nilai curah hujan maksimum dari masing-masing pos pengamatan mencapai lebih dari 100 mm per hari.

Berdasarkan kajian mengenai data iklim, diketahui bahwa sebagian besar data iklim memiliki ekor distribusi yang *heavy tail*, yaitu ekor distribusi yang turun secara lambat bila dibandingkan dengan distribusi normal. Akibatnya, peluang terjadinya nilai ekstrem akan lebih besar dibandingkan dengan distribusi normal. Oleh karena itu, sebelum melakukan analisis dengan menggunakan *Extreme Value Theory* perlu dilakukan identifikasi data pola *heavy tail*. Pada Tabel 4.3 disajikan nilai *skewness* dan *kurtosis* untuk tujuh pos pengamatan.

**Tabel 4.3** *Skewness* dan *Kurtosis*

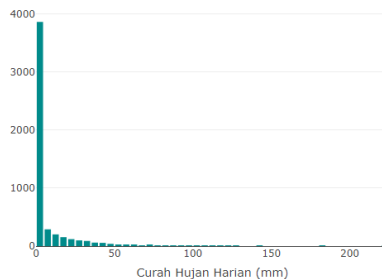
<b>Pos Pengamatan</b>	<b><i>Skewness</i></b>	<b><i>Kurtosis</i></b>
Kandangan	3,29	13,74
Kebon Agung	3,28	12,90
Gedeg	3,66	16,61
Janjing	3,21	10,99
Klegen	3,52	14,10
Mojosari	3,97	19,94
Pacet	3,79	21,95

Tabel 4.3 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.

#### 4.2.2 Histogram Data Curah Hujan

Adanya data ekstrem dan pola data *heavy tail* pada tujuh pos pengamatan juga dapat ditunjukkan dengan analisis secara visual pada histogram data curah hujan. Berikut merupakan histogram data curah hujan harian pada Pos Pacet.

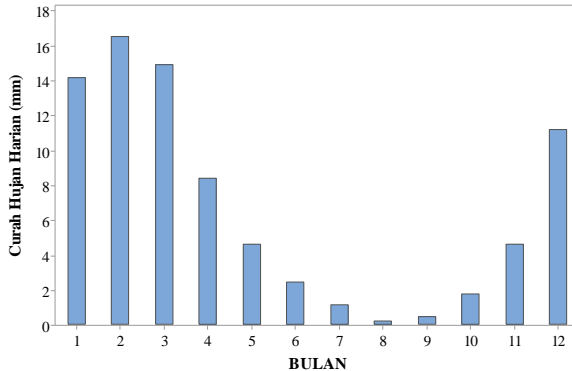


**Gambar 4.1** Histogram Curah Hujan Harian di Pos Pacet

Pada Gambar 4.1 menunjukkan bahwa kurva distribusi data curah hujan 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*. Hal tersebut juga terjadi pada ke enam pos lainnya yang dapat dilihat pada Lampiran 12. Berdasarkan alasan tersebut, ke tujuh pos pengamatan ini dikategorikan layak menjadi objek penelitian karena merupakan data *heavy tail*, sehingga dapat dilakukan pengambilan sampel ekstremnya.

### 4.2.3 Pola Data Curah Hujan

Selanjutnya, perlu dilakukan identifikasi untuk mengetahui pola curah hujan di tujuh pos pengamatan yang diteliti. Berikut merupakan pola rata-rata curah hujan harian tiap bulan pada Pos Pacet.

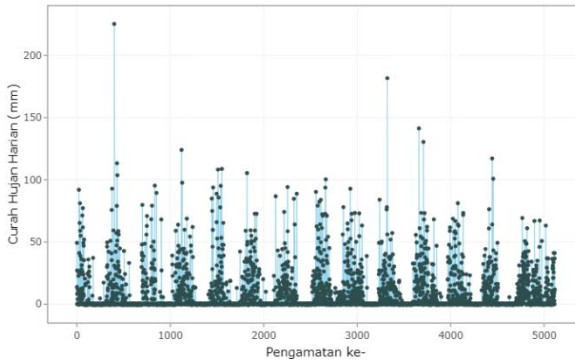


**Gambar 4.2** Bar Chart Curah Hujan Harian di Pos Pacet

Seperti ditunjukkan pada Gambar 4.2 dapat diketahui bahwa pola rata-rata curah hujan harian di pos hujan Pacet adalah pola hujan monsun dimana pola hujan membentuk pola huruf U. Pola rata-rata curah hujan harian untuk enam pos lain disajikan pada Lampiran 13 juga memiliki pola hujan monsun. Tipe curah hujan untuk pola mosoon adalah unimodal atau hanya terdapat satu puncak musim hujan. Biasanya, periode musim hujan terjadi pada Bulan Desember Januari Februari (DJF), Maret April Mei (MAM) merupakan periode transisi dari musim hujan ke musim kemarau, Juni Juli Agustus (JJA) adalah periode musim kemarau, sedangkan September Oktober November (SON) merupakan periode transisi dari musim kemarau menuju musim hujan. Pernyataan tersebut sesuai pada Gambar 4.2. Dapat diketahui bahwa Gambar 4.2 memiliki puncak tertinggi pada bulan Februari, yang berarti puncak musim hujan terjadi pada bulan Februari. Sedangkan puncak terendah yaitu pada bulan Agustus, yang berarti musim kemarau terjadi pada bulan Agustus.

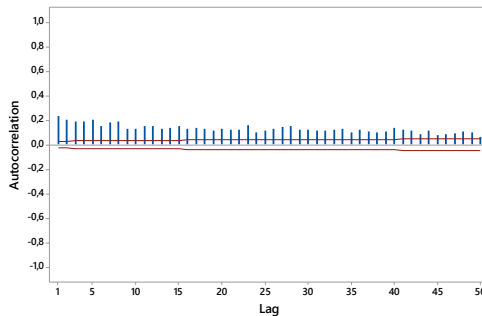
### 4.3 Identifikasi Dependensi Data Curah Hujan

Identifikasi dependensi data curah hujan pada penelitian ini dilakukan secara visual dengan melihat *time series plot* dan plot ACF. Dependensi pada data curah hujan pada masing-masing pos pengamatan dapat diketahui apabila plot tidak menunjukkan pola tertentu, seperti pola siklik, tren naik, maupun tren turun. Gambar 4.3 merupakan *time series plot* pada Pos Pacet.



**Gambar 4.3** *Time Series Plot* Data Curah Hujan Harian Pos Pacet

Hasil pemeriksaan secara visual pada Gambar 4.3 memperlihatkan bahwa *time series plot* pada Pos Pacet menunjukkan adanya pola siklik pada data curah hujan harian. Hal tersebut mengindikasikan adanya dependensi pada data curah hujan.



**Gambar 4.4.** Plot ACF Pos Pacet

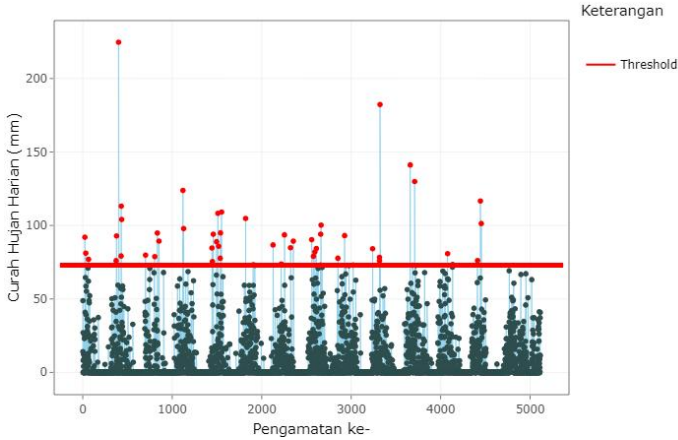
Identifikasi juga dapat dilakukan dengan melihat plot ACF. Berdasarkan plot ACF pada Gambar 4.4 dapat diketahui bahwa terbentuk pola siklik dan pada lag plot ACF melebihi batas atas dari ACF sehingga hal tersebut menjadi indikasi bahwa data curah hujan harian pada Pos Pacet bersifat dependen. Adapun plot ACF pada pos pengamatan yang lain dapat dilihat pada Lampiran 15.

#### **4.4 *Extreme Value Theory***

Pada penelitian ini dilakukan analisis pada data curah hujan menggunakan *Extreme Value Theory* dengan pendekatan *Peaks Over Threshold* (POT). Selain itu, digunakan juga metode *Declustering* untuk mengatasi dependensi data. Pada pendekatan POT dan *Declustering*, digunakan yang akan digunakan analisis selanjutnya terbatas disebabkan karena hanya data curah hujan yang intensitasnya melebihi nilai ambang batas ekstrem yang digunakan. Nilai ambang batas ekstrem dapat disebut dengan *threshold*.

##### **4.3.1 *Peaks Over Threshold***

Pengambilan data ekstrem untuk data curah hujan menggunakan *Peaks Over Threshold* dilakukan dengan cara menentukan nilai *threshold* terlebih dahulu. Pada penelitian ini, pemilihan *threshold* menggunakan metode persentase yaitu mengambil batas dari 1% keseluruhan data yang sudah diurutkan dari yang terbesar hingga terkecil. Berikut merupakan ilustrasi pengambilan sampel ekstrem pada Pos Pacet. Pada data curah hujan di Pos Pacet menghasilkan *threshold* 73 mm. Untuk *threshold* data curah hujan keenam pos pengamatan lainnya dapat dilihat pada Lampiran 16 dan penjelasan perhitungan pada Lampiran 19. Ilustrasi pengambilan data ekstrem POT disajikan pada Gambar 4.5.



**Gambar 4.5** Pengambilan Sampel Ekstrem di Pos Pacet

Tabel 4.5 memberikan informasi mengenai banyak pengamatan, nilai *threshold*, dan banyak sampel ekstrem untuk masing-masing pos pengamatan. Nilai *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 sampel ekstrem terbanyak yaitu pada Pos Kandangan, Pos Gedeg, dan Pos Mojosari.

**Tabel 4.4** Nilai Threshold POT

Pos Pengamatan	$N$	$u$ (mm)	$n_u$
Kandangan	5114	66	51
Kebonagung	5114	69	50
Gedeg	5114	64	51
Janjing	5114	70	50
Klegen	5114	70	36
Mojosari	5114	72	51
Pacet	5114	73	48

Setelah didapatkan sampel ekstrem, dapat dilakukan pemeriksaan kesesuaian distribusi untuk mengetahui adanya kesesuaian distribusi data curah hujan ekstrem dengan distribusi teoritisnya, yaitu *Generalized Pareto Distribution*.



### 4.3.2 Declustering

Berdasarkan hasil yang diperoleh pada analisis anak subbab 4.2.5 menunjukkan bahwa data curah hujan pada penelitian ini masih bersifat dependen, sehingga perlu dilakukan *Declustering* agar data curah hujan dapat menjadi data yang independen. Berikut merupakan hasil perhitungan manual *Extremal Index*, jumlah *cluster* yang terbentuk, dan *run length*.

Pada Tabel 4.5 dapat diketahui nilai *extremal index* yang digunakan sebagai indikator untuk mengukur berapa banyak *cluster*  $c$  yang terbentuk sehingga mendapatkan data independen. Semakin tinggi nilai *Extremal Index*, maka semakin banyak *cluster* yang terbentuk. Adapun *cluster* yang terbanyak terdapat pada Pos Kebon Agung yaitu terbentuk 41 *cluster*. Sedangkan *cluster* paling sedikit terdapat pada Pos Kandangan sebanyak 34 *cluster*. Tahapan perhitungan manual dapat dilihat pada Lampiran 10 sampai Lampiran 11.

**Tabel 4.5** Extremal Index pada Tujuh Pos Pengamatan

Pos Pengamatan	$u$	<i>Extremal Index</i>	$c$	$r$
Kandangan	66	0,658	34	49
Kebonagung	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

Nilai  $r$  optimum digunakan untuk proses selanjutnya yaitu *Declustering*. Adapun proses *Declustering* pada data curah hujan Pos Pacet adalah sebagai berikut.

1. Menentukan nilai *threshold* ( $u$ ) pada analisis yang sudah dilakukan pada subbab 4.3.1. Pada Pos Pacet memiliki nilai *threshold* sebesar 73 mm dan nilai-nilai yang berada di atas *threshold* dianggap sebagai nilai ekstrem.
2. Menentukan nilai  $r$ . Pada Tabel 4.5 diperoleh hasil *extremal index*. Nilai *extremal index* selanjutnya digunakan untu

menghitung nilai  $r$  yang optimum. Proses perhitungan dapat dilihat pada Lampiran 9 sampai Lampiran 11.

3. Menentukan *cluster*. *Cluster* pertama dimulai pada data pertama kali melebihi nilai *threshold* dan berakhir setelah minimal ada  $r$  data yang berada dibawah *threshold*. Pada Pos Pacet, *cluster* pertama terbentuk mulai pengamatan ke-2 sampai dengan ke-3. *Cluster* kedua terbentuk mulai pengamatan ke-16 sampai ke-23. Prosedur ini dilakukan terus menerus sampai pembentukan *cluster* selesai dan data telah habis. Pada data curah hujan Pos Pacet terbentuk 35 *cluster* dengan  $r = 4$ .
4. Mengidentifikasi nilai maksimum pada masing-masing *cluster*. Nilai maksimum ini selanjutnya digunakan untuk mengestimasi parameter *Generalized Pareto Distribution*.

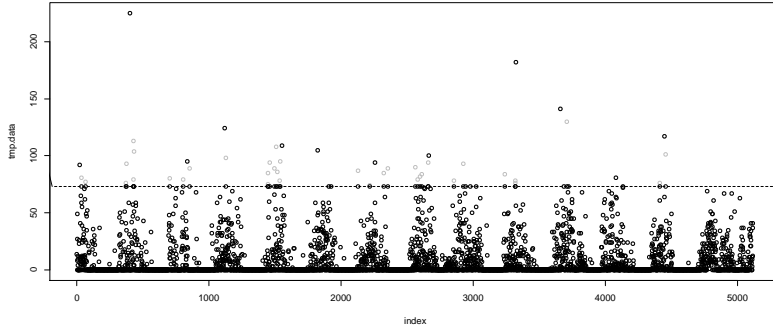
Penelitian ini membandingkan perhitungan *Declustering* menggunakan *software R* (Tabel 4.6) dan perhitungan secara manual (Tabel 4.5). Hasil *Declustering* yang didapatkan menggunakan *software R* dibagi menjadi dua pendekatan. pendekatan yang pertama yaitu *Runs Estimator* dan pendekatan yang kedua yaitu *Intervals Estimator*. Hasil pada kedua pendekatan tersebut, disajikan pada Tabel 4.6.

Jika dibandingkan antara Tabel 4.5 dan Tabel 4.6, ada beberapa nilai yang sama antara perhitungan manual dan menggunakan *software R*. Ketika dilakukan perhitungan manual, dihasilkan jumlah *cluster*  $c$  dan nilai *run length*  $r$  sama dengan pendekatan *Runs Estimator*. Sedangkan pada nilai *extremal index* pada perhitungan manual, memiliki nilai yang sama dengan pendekatan *Intervals Estimator*. Hal ini membuktikan adanya teori yang dipaparkan pada anak subbab 2.5.4.

**Tabel 4.6** Hasil Estimasi *Runs Estimator* dan *Interval Estimator*

Pos Pengamatan	Nilai	<i>Runs Estimator</i>	<i>Intervals Estimator</i>
Kandangan	<i>Extremal Index</i>	0,372	0,658
	<i>c</i>	19	34
	<i>r (run length)</i>	49	13
Kebon Agung	<i>Extremal Index</i>	0,420	0,817
	<i>c</i>	21	41
	<i>r (run length)</i>	53	12
Gedeg	<i>Extremal Index</i>	0,941	0,746
	<i>c</i>	48	39
	<i>r (run length)</i>	1	12
Janjing	<i>Extremal Index</i>	0,58	0,774
	<i>c</i>	29	39
	<i>r (run length)</i>	21	6
Klegen	<i>Extremal Index</i>	0,778	0,972
	<i>c</i>	28	35
	<i>r (run length)</i>	19	2
Mojosari	<i>Extremal Index</i>	0,529	0,745
	<i>c</i>	27	39
	<i>r (run length)</i>	43	12
Pacet	<i>Extremal Index</i>	0,250	0,757
	<i>c</i>	12	36
	<i>r (run length)</i>	189	10

Hasil perhitungan menggunakan *software R* juga menghasilkan Gambar 4.6. Gambar 4.6 menunjukkan bulatan-bulatan yang berwarna hitam yang berada di atas *threshold*. Bulatan-bulatan tersebut merupakan nilai maksimum pada masing-masing *cluster* yang terbentuk. Nilai tersebut selanjutnya akan digunakan untuk mengestimasi parameter *Generalized Pareto Distribution*.



**Gambar 4.6** Data Ekstrem *Declustering* pada Pos Pacet

Kesimpulan yang didapat pada metode *Extreme Value Theory* yaitu pada ukuran data pengamatan ekstrem umumnya jauh lebih sedikit dibanding ukuran data pengamatan yang dibawah *threshold*. 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 *Peaks Over Threshold* (yang selanjutnya akan disebut dengan metode *Non Declustering*) akan dibandingkan dengan metode *Declustering*.

#### 4.5 Uji Kesesuaian GPD *Declustering* dan *Non Declustering*

Uji Kesesuaian *Generalized Pareto Distribution* pada data ekstrem dilakukan untuk menunjukkan adanya kesesuaian distribusi teori, yaitu *Generalized Pareto Distribution*.

$H_0 : F_n x = F_0 x$ , (Data ekstrem mengikuti GPD)

$H_1 : F_n x \neq F_0 x$ , (Data ekstrem tidak mengikuti GPD).

$\alpha = 0,05$

Darah kritis :

Tolak  $H_0$  jika  $D_{hitung} > D_\alpha$  dengan  $D_\alpha$  adalah nilai kritis uji *Kolmogorov-Smirnov* atau  $P\text{-Value} > \alpha$ . Berikut ini hasil pengujian kesesuaian distribusi pada data curah hujan ekstrem untuk tujuh pos pengamatan.

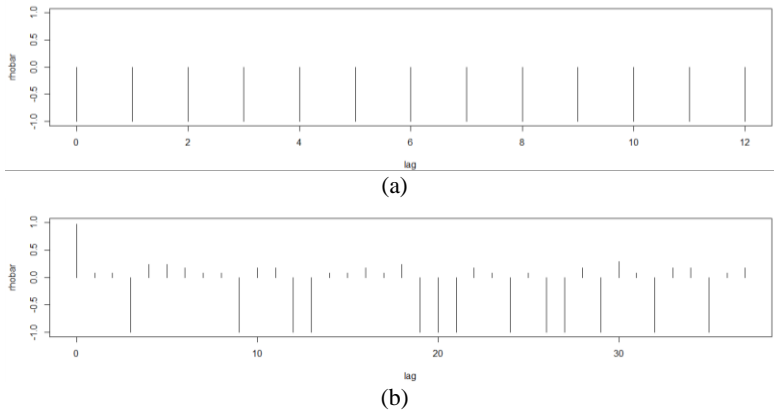
**Tabel 4.7** Hasil KST Data Ekstrem *Declustering* dan *Non Declustering*

Pos Pengamatan	<i>Declustering</i>		<i>Non Declustering</i>	
	$D_{hitung}$	$P-Value$	$D_{hitung}$	$P-Value$
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 4.7 dapat diketahui bahwa semua pos pengamatan pada data ekstrem *Declustering* maupun *Non Declustering* memiliki nilai  $P-Value$  lebih besar dari  $\alpha$ . Sehingga keputusan yang diambil yaitu gagal tolak  $H_0$  yang berarti data ekstrem *Declustering* dan *Non Declustering* pada semua pos pengamatan mengikuti distribusi teoritis yaitu *Generalized Pareto Distribution*. Sehingga dapat dilakukan analisis selanjutnya yaitu mengestimasi parameter *Generalized Pareto Distribution* pada data ekstrem *Declustering* dan *Non Declustering*. Estimasi parameter *Generalized Pareto Distribution* akan dibahas pada subbab 4.6.

#### 4.6 Identifikasi Dependensi Pada Data Ekstrem *Declustering* dan *Non Declustering*

Selain mengetahui kesesuaian data ekstrem yang mengikuti *Generalized Pareto Distribution*, perlu juga dilakukan identifikasi adanya dependensi data ekstrem *Declustering* maupun *Non Declustering*. Identifikasi awal dilakukan secara visual dengan melihat *time series plot*. Jika pada *time series plot* memiliki tren naik, tren turun, maupun siklik maka mengindikasikan bahwa data ekstrem bersifat depnden. Namun jika pada *time series plot* tidak memiliki tren naik, tren turun, maupun siklik maka ada indikasi bahwa data ekstrem sudah independen. Berikut merupakan *time series plot* pada data ekstrem *Declustering* dan *Non Declustering* pada Pos Pacet.



**Gambar 4.7** Extremogram Data Ekstrem (a) *Declustering* (b) *Non Declustering*

Berdasarkan Gambar 4.7 dapat diketahui bahwa data ekstrem *Declustering* dan *Non Declustering* pada Pos Pacet tidak memiliki *trend* sehingga dapat dikatakan bahwa data ekstrem melalui *Declustering* dan *Non Declustering* pada Pos Pacet sudah dependen.

#### 4.7 Estimasi Parameter GPD

Pada pendekatan *Declustering* dan *Non Declustering*, informasi yang tersedia menjadi terbatas disebabkan karena hanya curah hujan yang intensitasnya melebihi nilai ambang batas ekstrem yang dapat digunakan dalam proses penaksiran. Masalah terbatasnya data ekstrem yang digunakan dalam proses penaksiran parameter *Extreme Value Theory* menarik untuk diatasi guna memperoleh hasil estimasi parameter yang lebih. Pada subbab 4.6, menjelaskan tahap setelah mendapatkan data ekstrem *Declustering* dan *Non Declustering*. 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.

**Tabel 4.8** Hasil Estimasi Parameter GPD

<b>Pos Pengamatan</b>	<b>Nilai</b>	<b><i>Declustering</i></b>	<b><i>Non Declustering</i></b>
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

. Pos pengamatan yang memiliki keragaman 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  $\hat{\sigma}$  yang melalui proses *Non Declustering* memiliki *range* persebaran data terbesar terdapat pada Pos Kebon Agung. Sedangkan yang memiliki *range* persebaran data terbesar terdapat pada Pos Klegen.

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 terbesar adalah Pos Kebon Agung.

Berdasarkan uraian yang telah dipaparkan, dapat diketahui bahwa nilai 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.

#### 4.8 Return Level

Pada penelitian ini telah didapatkan nilai estimasi parameter skala dan bentuk yang sudah dibahas pada subbab 4.6. Nilai estimasi tersebut digunakan untuk mencari nilai estimasi *return level* yang disajikan pada Tabel 4.9. *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. Probabilitas dari *return level* yang diperoleh adalah  $\frac{1}{T}$ .

**Tabel 4.9** Hasil *Return Level*

Pos Pengamatan	Periode (3 Bulan)	Aktual	<i>Return Level</i>	<i>Return Level</i>
			<i>Declustering</i> (mm)	<i>Non Declustering</i> (mm)
Kandangan	2	85	130,129	173,609
	3	85	130,172	180,226
	4	98	130,194	184,921
Kebonagung	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

Perhitungan *return level* bergantung pada seberapa panjang periode yang ingin diprediksi nilai *return level*nya, dan seberapa besar peluang terjadinya kejadian ekstrem yang masih diperhitungkan oleh peneliti. Nilai estimasi *return level* akan dibandingkan dengan nilai aktual, dimana nilai aktual didapatkan dari data *testing* penelitian yaitu data curah hujan harian tahun 2017. Berikut merupakan hasil *return level* dalam satu tahun.

Penelitian ini menggunakan periode prediksi terbanyak yaitu satu tahun atau empat periode. Penentuan periode maksimum mempertimbangkan probabilitas yang diperoleh yaitu sebesar

$$\frac{1}{T} = \frac{1}{4} = 0,25 = 25\%$$

Ini adalah peluang maksimum yang dapat peneliti perhitungkan. Prediksi menggunakan nilai  $T$  sangat besar menghasilkan *return level* dengan peluang kejadian yang sangat kecil. Begitu pula penggunaan  $T$  yang sangat kecil menyebabkan kemungkinan terjadinya kejadian ekstrem yang lebih besar dengan peluang yang cukup signifikan menjadi terabaikan. Terlebih lagi berapapun nilai *return level* yang diperoleh memiliki peluang terjadinya pada periode manapun sejauh masih di dalam periode estimasi.

Berdasarkan Tabel 4.9 dapat diketahui bahwa pada estimasi *Return Level Declustering* lebih mendekati nilai aktual. Untuk melakukan validasi data, nilai *return level* dihitung selisihnya dengan nilai aktual untuk diperoleh nilai RMSE. Berdasarkan informasi BMKG, adanya perbedaan antara nilai prediksi dan aktual sebesar 25% sampai 30% dianggap masih cukup baik. Perbedaan nilai prediksi dan aktual disajikan pada Tabel 4.10.

Berdasarkan Tabel 4.10 terdapat empat pos pengamatan yang memiliki nilai *error* kurang dari 30%, antara lain Pos Kebon Agung, Pos Gedeg, Pos Janjing, dan Pos Klegen. Sedangkan ketiga pos lainnya mempunyai nilai *error* lebih dari 30%. Kondisi ini kemungkinan diakibatkan banyak faktor variabel lain yang perlu dikaji dalam penentuan prediksi curah hujan. Misalkan

kecepatan angin, kelembapan udara, ataupun variabel lain sehingga perlu penelitian lebih lanjut untuk menentukan prediksi curah hujan yang nilai prediksinya mendekati nilai aktual.

**Tabel 4.10** Nilai *Error* proses *Declustering* dan *Non Declustering*

Pos Pengamatan	Periode Ulang (3 Bulan)	<i> Error </i>	<i> Error </i>
		<i>Declustering</i> (%)	<i>Non-Declustering</i> (%)
Kandangan	2	45,129	88,609
	3	45,172	95,226
	4	32,194	86,921
Kebonagung	2	22,134	23,478
	3	22,168	23,804
	4	2,186	3,998
Gedeg	2	15,894	24,937
	3	15,639	27,018
	4	17,097	28,388
Janjing	2	9,377	9,493
	3	9,595	9,96
	4	22,276	21,75
Klegen	2	0,381	17,084
	3	1,108	21,517
	4	1,561	24,662
Mojosari	2	75,960	72,175
	3	82,728	80,331
	4	87,482	86,187
Pacet	6	178,868	142,978
	9	191,232	160,969
	12	199,863	174,24

Jika dikaji perbedaan nilai *error* antara *Declustering* dan *Non Declustering*, *error* pada metode *Declustering* banyak yang lebih kecil dibandingkan metode *Non Declustering*, namun untuk lebih meyakinkan, perlu dilakukan perbandingan nilai RMSE dari *return level* yang didapatkan berdasarkan proses *Declustering* dan *Non Declustering*.

#### 4.9 Tingkat Kesalahan Ramalan

Untuk menentukan metode mana yang lebih baik, perlu dilakukan perbandingan RMSE berdasarkan metode *Declustering* dan *Non Declustering*. Metode yang memiliki nilai RMSE terkecil adalah metode yang lebih baik dalam menentukan *return level*.

**Tabel 4.11** RMSE *Return Level* pada Metode *Declustering* dan *Non Declustering* di Surabaya

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

Tabel 4.11 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 4.12.** RMSE *Return Level* pada Metode *Declustering* dan *Non Declustering* di Mojokerto

Pos Pengamatan	RMSE <i>Declustering</i>	RMSE <i>Non Declustering</i>
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*
<b>Rata-rata</b>	<b>60,947</b>	<b>60,533</b>

Keterangan : \*) nilai RMSE lebih kecil pada Metode *Non Declustering*

Hasil perbandingan nilai RMSE *return level* yang melalui *Declustering* dan *Non Declustering* dapat dilihat pada Tabel 4.12. 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*.

*(Halaman ini sengaja dikosongkan)*

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Karakteristik curah hujan ekstrem di tujuh pos pengamatan memiliki pola distribusi *heavy tail* yang menunjukkan adanya data ekstrem, distribusi data tidak mengikuti distribusi normal, namun mengikuti distribusi *Generalized Pareto Distribution*
2. 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.
3. Kebaikan estimasi *return level* curah hujan ekstrem di Surabaya menunjukkan hasil yang lebih baik menggunakan *Declustering*, sedangkan di Mojokerto yang lebih baik yaitu *Non Declustering*.

#### 5.2 Saran

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 mempengaruhi prediksi curah hujan seperti kecepatan angin atau kelembapan udara.

*(Halaman ini sengaja dikosongkan)*



## DAFTAR PUSTAKA

- Beritajatim.com. (2016). Pakar ITS: Banjir Surabaya Barat Akibat Minimnya Drainase. Retrieved June 14, 2019, from beritajatim.com website: [http://m.beritajatim.com/peristiwa/260556/pakar\\_its:\\_banjir\\_surabaya\\_barat\\_akibat\\_minimnya\\_drainase.html](http://m.beritajatim.com/peristiwa/260556/pakar_its:_banjir_surabaya_barat_akibat_minimnya_drainase.html).
- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag, Berlin.
- Daniela Jaruskova, M. H. (2006). Peaks Over Threhsold Method In Comparison With Block-Maxima Method For Estimating High Return Levels Of Several Northern Moravia Precipitation and Discharges Series. *Journal of Hydrology and Hydromechanics*, 54, 309–319.
- Dharmawan, K. (2012). *Estimasi Nilai VaR Dinamis Indeks Saham Menggunakan Peak-Over Threshold dan Block Maxima*. 2(2), 1–12.
- FaktualNews.co. (2019). Imbas Luapan Sungai Lamong, Puluhan Rumah di Dawar Blandong Mojokerto Setiap Tahun Terendam Banjir. Retrieved June 14, 2019, from faktualnews.co website: <https://faktualnews.co/2019/01/18/imbas-luapan-sungai-lamong-puluhan-rumah-di-dawar-blandong-mojokerto-setiap-tahun-terendam-banjir/118957/>.
- Ferro, C. A. T., & Segers, J. (2002). *Automatic declustering of extreme values via an estimator for the extremal index*. Lancaster University.
- Ferro, C. A. T., & Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(2), 545–556.
- Gilleland, E., & Katz, R. W. (2006). Analyzing Seasonal To Interannual Extreme Weather and climate variability with the extremes toolkit (extRems) *18<sup>th</sup> Conference on Climate Variablity and Change*, Atlanta, GA, American Meteorological Society.

- Gilli, M., & K llezi, E. (2006). An application of extreme value theory for measuring financial risk. *Computational Economics*, 27(2–3), 207–228.
- Handoko. (1994). *Klimatologi dasar : landasan pemahaman fisika atmosfer dan unsur- unsur iklim*. Jakarta: Pustaka Jaya.
- Johnson, R. a., & Bhattacharyya, G. K. (2010). *Statistics Principles & Methods. 6<sup>th</sup> Edition*. John Wiley & Sons. New York.
- Kadarsah. (2007). Tiga Pola Curah Hujan Indonesia. Retrieved June 14, 2019, from <https://kadarsah.wordpress.com/2007/06/29/tiga-daerah-iklim-indonesia/%0D>.
- Kurniawati, Y. (2013). *Analisis Data Ekstrem Dependen (Non Stationary) Pada Kasus Curah Hujan Ekstrem di Jawa Timur Dengan Pendekatan Peaks Over Threshold*. Tugas Akhir S1: Jurusan Statistika, Institut Teknologi Sepuluh Nopember.
- Fawcett, L & Walshaw, D. (2007). Improved Estimation for Temporally Clustered Extremes. *Environmetrics*, 18(2),173-188.
- Li, Z. (2015). *Study on Climate Change in Southwestern China*. Springer Verlag, Berlin.
- Mcnail, A. J. (1999). Extreme Value Theory for Risk Managers. Internal Modeling CADII, Risk Books 93-113. ETH Zurich.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *J.R.Statist*, 3(1), 119–131.
- Rinaldi, A., Djuraidah, A., Wigena, A. H., Wayan Mangku, I., & Gunawan, D. (2018). Identification of Extreme Rainfall Pattern Using Extremogram in West Java. *IOP Conference Series: Earth and Environmental Science*, 187, 012064.
- Wahyudi. (2012). *Identifikasi Curah Hujan Ekstrem di Kabupaten Ngawi Menggunakan Generalized Extreme Value dan Generalized Pareto Distribution*. Tugas Akhir S1: Jurusan Statistika, Institut Teknologi Sepuluh Nopember.

- Warto. (2012). *Discrete Wavelet Transform Untuk Mempercepat Prediksi Produksi Susu Sapi Perah Berbasis Neural Network*. Masters thesis, Universitas Dian Nuswantoro Semarang
- Wei, W.W.S. (2004). *Time Series Analysis Univariate and Multivariate Methods*. Canada: Addison Wesley Publishing, Company. Inc.

*(Halaman ini sengaja dikosongkan)*

## LAMPIRAN

### Lampiran 1. Data Curah Hujan Harian Pos Kandangan

TAHUN	BULAN	TANGGAL	CURAH HUJAN
2003	1	1	7
2003	1	2	18
2003	1	3	32
2003	1	4	0
2003	1	5	12
2003	1	6	7
2003	1	7	0
2003	1	8	8
2003	1	9	0
2003	1	10	32
2003	1	11	12
2003	1	12	20
2003	1	13	15
2003	1	14	18
2003	1	15	15
2003	1	16	0
2003	1	17	0
2003	1	18	0
2003	1	19	0
2003	1	20	65
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	9.43
2017	12	23	10.29
2017	12	24	7.29
2017	12	25	6.71
2017	12	26	22.5
2017	12	27	14.79
2017	12	28	12.79
2017	12	29	14.43
2017	12	30	12.29
2017	12	31	12.14

**Lampiran 2.** Data Curah Hujan Harian Pos Kebon Agung

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	0
2003	1	2	21
2003	1	3	27
2003	1	4	0
2003	1	5	25
2003	1	6	9
2003	1	7	0
2003	1	8	0
2003	1	9	0
2003	1	10	14
2003	1	11	0
2003	1	12	25
2003	1	13	16
2003	1	14	12
2003	1	15	9
2003	1	16	0
2003	1	17	0
2003	1	18	0
2003	1	19	0
2003	1	20	40
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	0
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	8
2017	12	27	2
2017	12	28	0
2017	12	29	0
2017	12	30	49
2017	12	31	0

**Lampiran 3.** Data Curah Hujan Harian Pos Gedeg

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	10
2003	1	2	17
2003	1	3	62
2003	1	4	2
2003	1	5	2
2003	1	6	4
2003	1	7	0
2003	1	8	0
2003	1	9	0
2003	1	10	0
2003	1	11	0
2003	1	12	20
2003	1	13	0
2003	1	14	7
2003	1	15	0
2003	1	16	0
2003	1	17	0
2003	1	18	0
2003	1	19	0
2003	1	20	0
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	36
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	0
2017	12	27	0
2017	12	28	0
2017	12	29	0
2017	12	30	0
2017	12	31	0

**Lampiran 4.** Data Curah Hujan Harian Pos Janjing

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	41
2003	1	2	23
2003	1	3	15
2003	1	4	0
2003	1	5	0
2003	1	6	60
2003	1	7	0
2003	1	8	0
2003	1	9	0
2003	1	10	0
2003	1	11	0
2003	1	12	0
2003	1	13	0
2003	1	14	0
2003	1	15	0
2003	1	16	0
2003	1	17	0
2003	1	18	80
2003	1	19	0
2003	1	20	0
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	0
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	0
2017	12	27	0
2017	12	28	0
2017	12	29	4
2017	12	30	0
2017	12	31	13



**Lampiran 5.** Data Curah Hujan Harian Pos Klegen

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	20
2003	1	2	30
2003	1	3	20
2003	1	4	20
2003	1	5	0
2003	1	6	0
2003	1	7	0
2003	1	8	0
2003	1	9	0
2003	1	10	0
2003	1	11	0
2003	1	12	0
2003	1	13	0
2003	1	14	0
2003	1	15	0
2003	1	16	0
2003	1	17	0
2003	1	18	0
2003	1	19	0
2003	1	20	0
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	0
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	0
2017	12	27	0
2017	12	28	0
2017	12	29	0
2017	12	30	0
2017	12	31	0

**Lampiran 6.** Data Curah Hujan Harian Pos Mojosari

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	0
2003	1	2	22
2003	1	3	5
2003	1	4	0
2003	1	5	0
2003	1	6	34
2003	1	7	0
2003	1	8	0
2003	1	9	18
2003	1	10	3
2003	1	11	0
2003	1	12	10
2003	1	13	0
2003	1	14	32
2003	1	15	22
2003	1	16	0
2003	1	17	0
2003	1	18	5
2003	1	19	0
2003	1	20	0
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	0
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	0
2017	12	27	0
2017	12	28	0
2017	12	29	0
2017	12	30	0
2017	12	31	0

**Lampiran 7.** Data Curah Hujan Harian Pos Pacet

<b>TAHUN</b>	<b>BULAN</b>	<b>TANGGAL</b>	<b>CURAH HUJAN</b>
2003	1	1	14
2003	1	2	49
2003	1	3	27
2003	1	4	0
2003	1	5	8
2003	1	6	0
2003	1	7	0
2003	1	8	4
2003	1	9	12
2003	1	10	9
2003	1	11	0
2003	1	12	0
2003	1	13	0
2003	1	14	0
2003	1	15	0
2003	1	16	26
2003	1	17	23
2003	1	18	0
2003	1	19	11
2003	1	20	0
...	...	...	...
...	...	...	...
...	...	...	...
2017	12	22	0
2017	12	23	0
2017	12	24	0
2017	12	25	0
2017	12	26	0
2017	12	27	0
2017	12	28	0
2017	12	29	0
2017	12	30	0
2017	12	31	0

### Lampiran 8. Estimasi Parameter GPD

Berdasarkan kajian mengenai data iklim, diketahui bahwa sebagian besar data iklim memiliki ekor distribusi yang *heavy tail*, yaitu ekor distribusi yang turun secara lambat. Akibatnya, peluang terjadinya nilai ekstrem akan lebih besar. Ditinjau dari sisi resiko, diperoleh peluang kejadian ekstrem yang maksimum berarti menanggung resiko maksimum atas terjadinya kejadian ekstrem. Berdasarkan alasan tersebut, proses estimasi ini menggunakan PDF dari GPD yang berlaku untuk  $\xi \neq 0$ , yaitu

$$f_{\sigma, \xi} = \frac{1}{\sigma} \left( 1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi} - 1}$$

Fungsi *likelihood* untuk GPD adalah sebagai berikut.

$$L(\sigma, \xi) = \sigma^{-n} \prod_{i=1}^n \left( 1 + \frac{\xi y_i}{\sigma} \right)^{-\frac{1}{\xi} - 1}$$

Membentuk fungsi  $\ln$  *likelihood* untuk distribusi GPD seperti persamaan dibawah ini.

$$\ell(\sigma, \xi) = -n \ln(\sigma) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right)$$

Memaksimumkan  $\ln$  *likelihood* dengan cara mencari turunan pertama terhadap parameter  $\sigma$  dan  $\xi$ . Berikut adalah hasil dari turunan pertama dari  $\ln$  *likelihood*.

a. Turunan pertama terhadap parameter  $\sigma$ .

$$\begin{aligned} \frac{\partial \ell_{\sigma, \xi}}{\partial \sigma} &= \frac{\partial \left( -n \ln \sigma - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) \right)}{\partial \sigma} \\ &= -n \left( \frac{1}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \left( \frac{-\frac{\xi y_i}{\sigma^2}}{1 + \frac{\xi y_i}{\sigma}} \right) \end{aligned}$$

$$\begin{aligned}
&= -n \left( \frac{1}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \left( \frac{-\frac{\xi y_i}{\sigma^2}}{\sigma + \xi y_i} \right) \\
&= -n \left( \frac{1}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n - \left( \frac{\xi y_i}{\sigma^2} \right) \left( \frac{\sigma}{\sigma + \xi y_i} \right) \\
&= -\frac{n}{\sigma} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \left( \frac{\xi y_i}{\sigma^2 + \sigma \xi y_i} \right)
\end{aligned}$$

b. Turunan pertama terhadap parameter  $\xi$ .

$$\begin{aligned}
\frac{\partial \ell_{\sigma, \xi}}{\partial \xi} &= \frac{\partial \left( -n \ln \sigma - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) \right)}{\partial \xi} \\
&= \frac{1}{\xi^2} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{1}{\left( 1 + \frac{\xi y_i}{\sigma} \right)} \left( \frac{y_i}{\sigma} \right) \\
&= \frac{1}{\xi^2} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma \left( 1 + \frac{\xi y_i}{\sigma} \right)} \\
&= \frac{1}{\xi^2} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma \left( \frac{\sigma + \xi y_i}{\sigma} \right)} \\
&= \frac{1}{\xi^2} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i}
\end{aligned}$$

Berdasarkan persamaan pada turunan pertama fungsi  $\ln$  *likelihood* terhadap masing-masing parameter, menghasilkan persamaan yang tidak *closed form*. Sehingga diperlukan pendekatan secara numerik dengan menggunakan *Newton Raphson*. Yang diperlukan pada analisis numerik yaitu turunan kedua pada masing-masing parameter. Berikut merupakan turunan kedua pada masing-masing parameter.

a. Turunan kedua  $\sigma$  terhadap turunan pertama  $\sigma$ .

$$\begin{aligned}
 \frac{\partial^2 \ell_{\sigma, \xi}}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma} \left( \frac{\partial \ell_{\sigma, \xi}}{\partial \sigma} \right) \\
 &= \frac{\partial}{\partial \sigma} \left( -\frac{n}{\sigma} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \left( \frac{\xi y_i}{\sigma^2 + \sigma \xi y_i} \right) \right) \\
 &= -n \sigma^{-2} - 1 + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \xi y_i \sigma^2 + \sigma \xi y_i^{-1} \\
 &= \frac{n}{\sigma^2} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \xi y_i^{-1} \sigma^2 + \sigma \xi y_i^{-2} - 2\sigma + \xi y_i \\
 &= \frac{n}{\sigma^2} - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{\xi y_i}{\sigma^2 + \sigma \xi y_i} \frac{2\sigma + \xi y_i}{\sigma^2 + \sigma \xi y_i} \\
 &= \frac{n}{\sigma^2} - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{\xi (2\sigma y_i + \xi y_i^2)}{\sigma^4 + 2\sigma^3 \xi y_i + \sigma^2 \xi^2 y_i^2} \\
 &= \frac{n}{\sigma^2} - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{\xi (2\sigma y_i + \xi y_i^2)}{\sigma^2 (\sigma^2 + 2\sigma \xi y_i + \xi^2 y_i^2)} \\
 &= \frac{n}{\sigma^2} - \left( 1 + \frac{1}{\xi} \right) \left( \frac{\xi}{\sigma^2} \right) \sum_{i=1}^n \frac{2\sigma y_i + \xi y_i^2}{\sigma^2 + 2\sigma \xi y_i + \xi^2 y_i^2} \\
 &= \frac{n}{\sigma^2} - \left( \frac{\xi}{\sigma^2} + \frac{1}{\sigma^2} \right) \sum_{i=1}^n \frac{2\sigma y_i + \xi y_i^2}{\sigma^2 + 2\sigma \xi y_i + \xi^2 y_i^2} \\
 &= \frac{n - \xi - 1}{\sigma^2} \sum_{i=1}^n \frac{2\sigma y_i + \xi y_i^2}{\sigma + \xi y_i} \frac{1}{\sigma}
 \end{aligned}$$

b. Turunan kedua  $\xi$  terhadap turunan pertama  $\xi$ .

$$\begin{aligned}
 \frac{\partial^2 \ell_{\sigma, \xi}}{\partial \xi^2} &= \frac{\partial}{\partial \xi} \left( \frac{\partial \ell_{\sigma, \xi}}{\partial \xi} \right) \\
 &= \frac{\partial}{\partial \xi} \left( \frac{1}{\xi^2} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} \right) \\
 &= \left[ \left( -\frac{2}{\xi^3} \right) \left( \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) \right) + \left( \frac{1}{\xi^2} \right) \left( \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} \right) \right] - \\
 &\quad \left[ \left( -\frac{1}{\xi^2} \right) \left( \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} \right) + \left( 1 + \frac{1}{\xi} \right) \left( \sum_{i=1}^n \frac{-y_i^2}{\sigma + \xi y_i^2} \right) \right] \\
 &= -\frac{2}{\xi^3} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) + \frac{1}{\xi^2} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \\
 &\quad \frac{1}{\xi^2} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i^2}{\sigma + \xi y_i^2} \\
 &= -\frac{2}{\xi^3} \sum_{i=1}^n \ln \left( 1 + \frac{\xi y_i}{\sigma} \right) + \frac{2}{\xi^2} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \\
 &\quad \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i^2}{\sigma + \xi y_i^2}
 \end{aligned}$$

c. Turunan kedua  $\xi$  terhadap turunan pertama  $\sigma$ .

$$\begin{aligned}
 \frac{\partial^2 \ell_{\sigma, \xi}}{\partial \xi \partial \sigma} &= \frac{\partial}{\partial \xi} \left( \frac{\partial \ell_{\sigma, \xi}}{\partial \sigma} \right) \\
 &= \frac{\partial}{\partial \xi} \left( -\frac{n}{\sigma} + \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \left( \frac{\xi y_i}{\sigma^2 + \sigma \xi y_i} \right) \right) \\
 &= \frac{1}{\xi^2} \sum_{i=1}^n \frac{\left( \frac{\xi y_i}{\sigma^2} \right)}{\left( 1 + \frac{\xi y_i}{\sigma} \right)} - \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{-y_i}{\sigma + \xi y_i^2} \\
 &= \frac{1}{\xi^2} \sum_{i=1}^n \frac{\left( \frac{\xi y_i}{\sigma^2} \right)}{\left( \frac{\sigma + \xi y_i}{\sigma} \right)} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i^2} \\
 &= \frac{1}{\xi^2} \sum_{i=1}^n \frac{\xi y_i}{\sigma \sigma + \xi y_i} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i^2} \\
 &= \frac{1}{\xi^2} \left( \frac{\xi}{\sigma} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i^2} \\
 &= \frac{1}{\sigma} \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i} + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{y_i}{\sigma + \xi y_i^2}
 \end{aligned}$$



**Lampiran 9.** Ilustrasi Penentuan *Threshold* Metode Persentase Pada Pos Pacet

1. Mengurutkan Data dari yang terbesar hingga terkecil

<b>Data Curah Hujan (mm)</b>	
<b>Sebelum Sort</b>	<b>Setelah SORT</b>
14	225
49	182
27	141
...	...
0	0
0	0
0	0

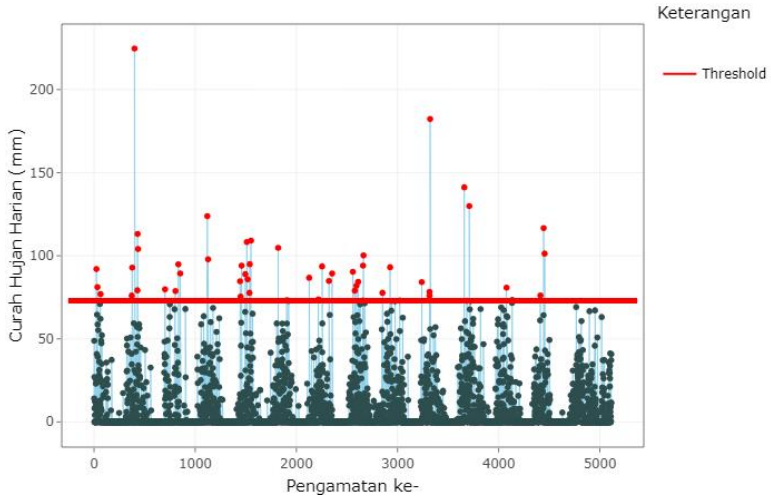
Jumlah data ( $N$ ) = 5114

2. Menghitung 1% dari jumlah data ( $k$ )
 
$$k = 1\% \times N$$

$$= 1\% \times 5114$$

$$= 51,14 \approx 51$$
3. Menentukan nilai *threshold* ( $u$ ), yaitu data urutan ke- ( $k+1$ )
 
$$k + 1 = 52$$

<b>NO</b>	<b>Setelah Sort</b>
1	225
2	182
3	141
...	...
50	73
<b>51</b>	<b>73</b>
52	73
...	...
5112	0
5113	0
5114	0



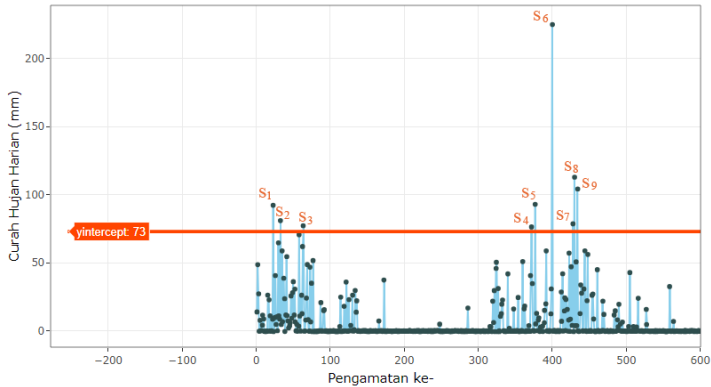
**Ilustrasi 1.** Penentuan *Threshold* Menggunakan Metode Persentase

$$\text{Threshold } (u) = 73 \text{ mm}$$

Semua pengamatan yang melebihi 73 mm diidentifikasi sebagai nilai ekstrem.

## Lampiran 10. Ilustrasi Menghitung *Extremal Index* Pada Pos Pacet

### 1. Mengidentifikasi $S_i$ (urutan pengamatan diatas *threshold*)



**Ilustrasi 2.** Urutan Pengamatan Diatas *Threshold*

Pada data curah hujan di Pos Pacet dengan nilai *threshold* ( $u$ ) 73 mm, terdapat 48 pengamatan yang berada di atas *threshold*. Sehingga terdapat  $S_1, S_2, S_3, \dots, S_{51}$ . Pada Ilustrasi 2. hanya menampilkan sebagian data pada Pos Pacet

### 2. Menentukan $T_i$ (Jarak antar pengamatan diatas *threshold*)

$$T_i = S_{i-1} - S_i$$

No	$S_i$	$T_i$
1	23	10
2	33	31
...	...	...
36	2663	189
...	...	...
46	4412	33
47	4445	10
48		

3. Menghitung *Eksternal Index*

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

$$\theta = \min \left[ 1, \frac{2 \left( \sum_{i=1}^{N-1} T_i - 1 \right)^2}{N-1 \sum_{i=1}^{N-1} T_i - 1 \quad T_i - 2} \right]$$

$$\theta = \min 1, 0.757463$$

$$\theta = 0.757463$$

**Lampiran 11.** Ilustrasi Menentukan *Run Length* Pada Pos Pacet

$$C - 1 = \theta N - 1$$

$$C - 1 = 0,757463 \cdot 48 - 1$$

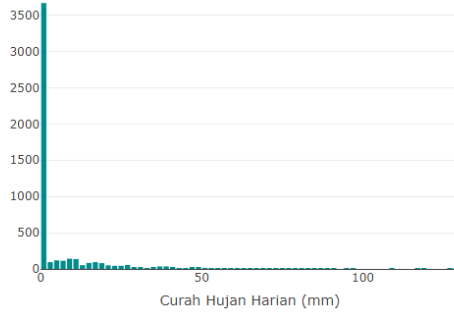
$$C - 1 = 36,60077 \approx 36$$

No	$S_i$	$T_i$
1	23	10
2	33	31
...	...	...
36	2663	189
...	...	...
46	4412	33
47	4445	10
48		

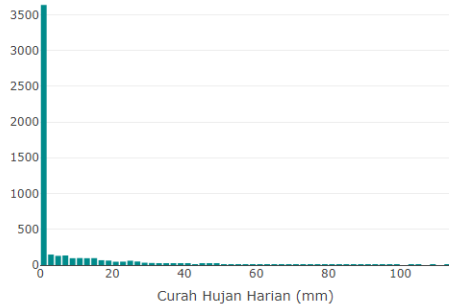
$$\begin{aligned} \text{Maka run length } r &= T_C \\ &= T_{36} \\ &= 189 \end{aligned}$$

**Lampiran 12. Histogram Curah Hujan Harian**

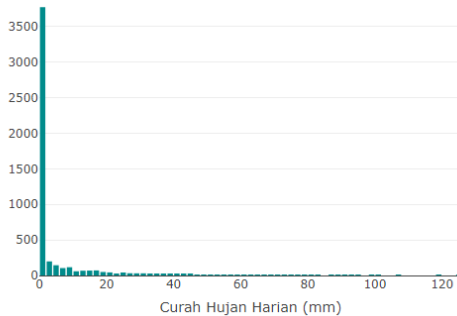
**Lampiran 12A. Histogram Curah Hujan Harian Pos Kandangan**



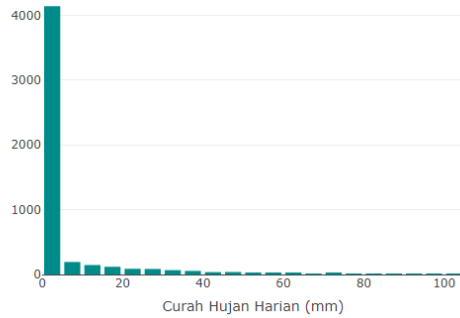
**Lampiran 12B. Histogram Curah Hujan Harian Pos Kebon Agung**



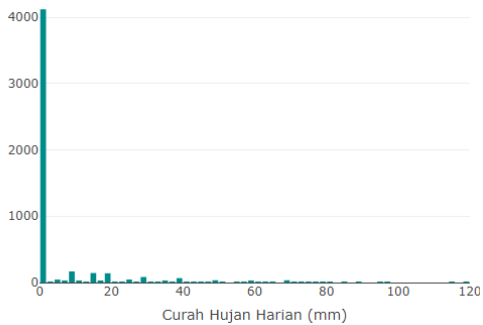
**Lampiran 12C. Histogram Curah Hujan Harian Pos Gedeg**



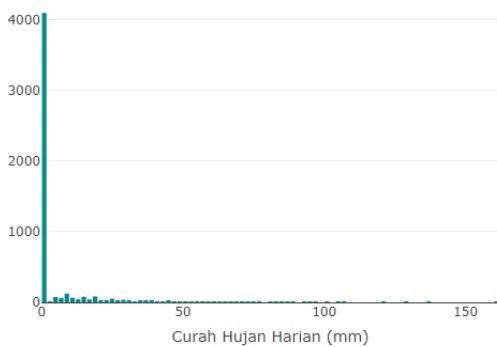
### Lampiran 12D. Histogram Curah Hujan Harian Pos Janjing



### Lampiran 12E. Histogram Curah Hujan Harian Pos Klegen

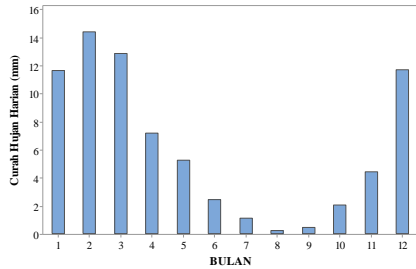


### Lampiran 12F. Histogram Curah Hujan Harian Pos Mojosari

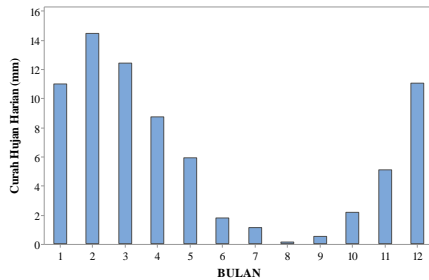


**Lampiran 13.** Bar Chart Curah Hujan Harian Per Bulan

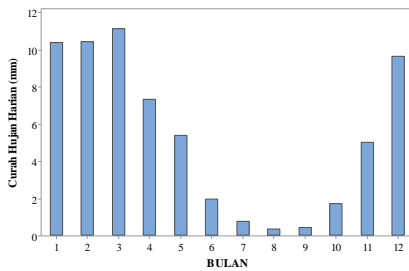
**Lampiran 13A.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Kandangan



**Lampiran 13B.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Kebon Agung

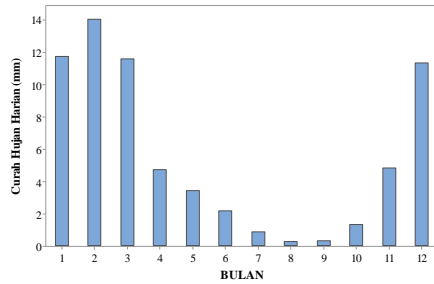


**Lampiran 13C.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Gedeg

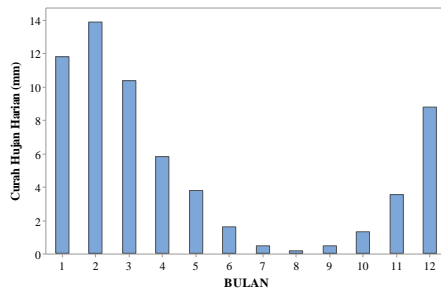




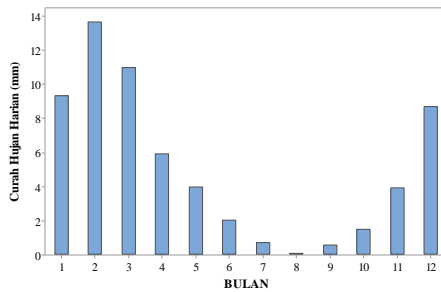
**Lampiran 13D.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Janjing



**Lampiran 13E.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Klegen

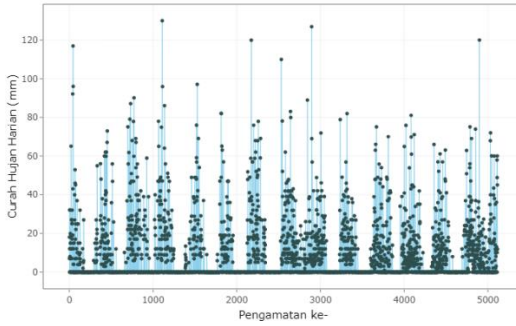


**Lampiran 13F.** *Bar Chart* Curah Hujan Harian Per Bulan Pos Mojosari

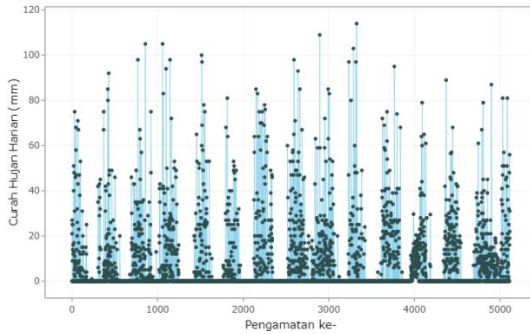


**Lampiran 14.** *Time Series Plot*

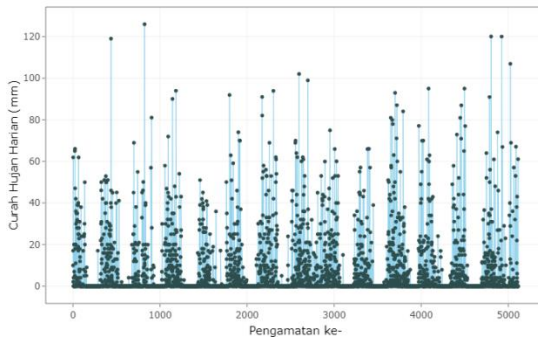
**Lampiran 14A.** *Time Series Plot* Pos Kandang

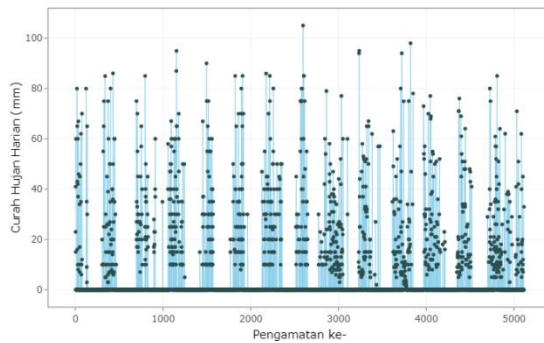
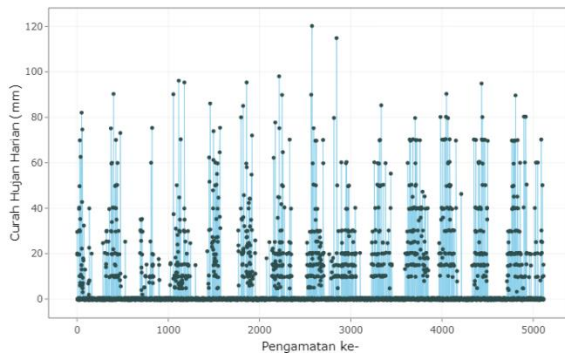
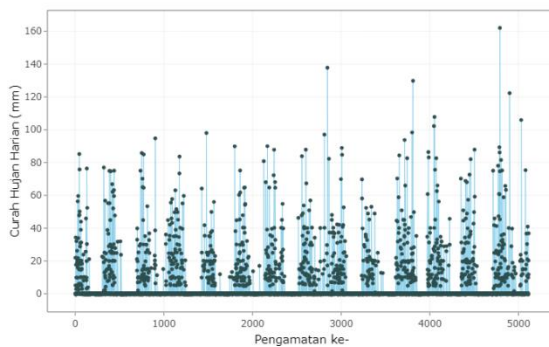


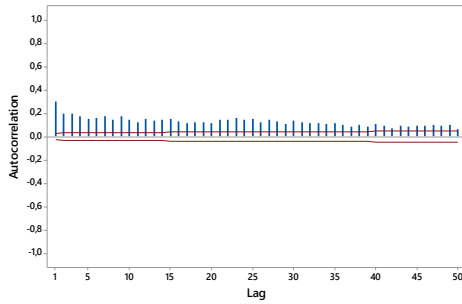
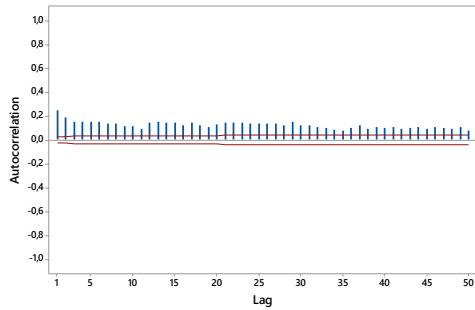
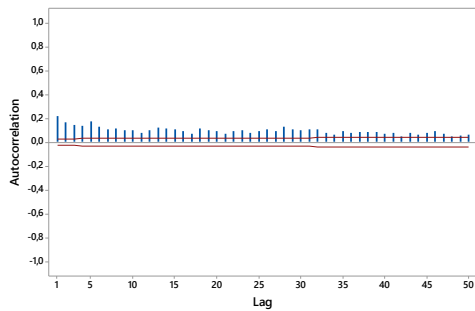
**Lampiran 14B.** *Time Series Plot* Pos Kebon Agung

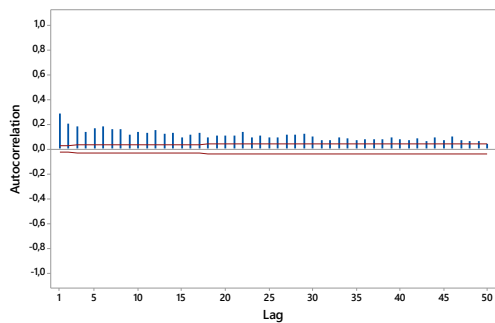
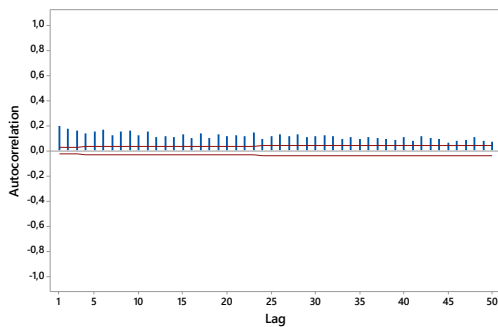
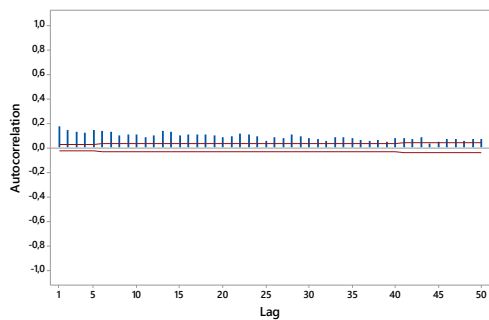


**Lampiran 14C.** *Time Series Plot* Pos Gedeg



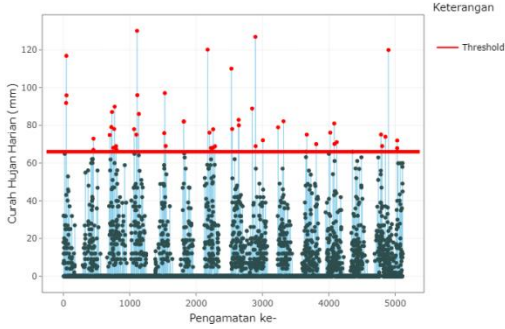
**Lampiran 14D. Time Series Plot Pos Janjing****Lampiran 14E. Time Series Plot Pos Klegen****Lampiran 14F. Time Series Plot Pos Mojosari**

**Lampiran 15. Plot ACF****Lampiran 15A. Plot ACF Pos Kandangan****Lampiran 15B. Plot ACF Pos Kebon Agung****Lampiran 15C. Plot ACF Pos Gedeg**

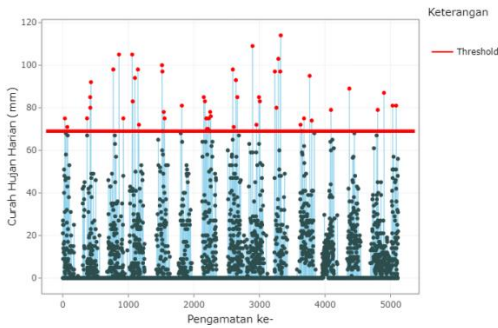
**Lampiran 15D.** Plot ACF Pos Janjing**Lampiran 15F.** Plot ACF Pos Klegan**Lampiran 15G.** Plot ACF Pos Mojosari

**Lampiran 16.** Pengambilan Sampel Ekstrem POT

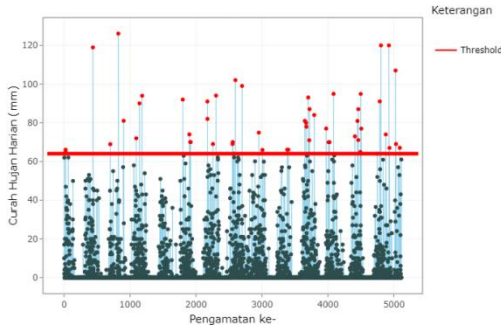
**Lampiran 16A.** Pengambilan Sampel Ekstrem Pos Kandangan



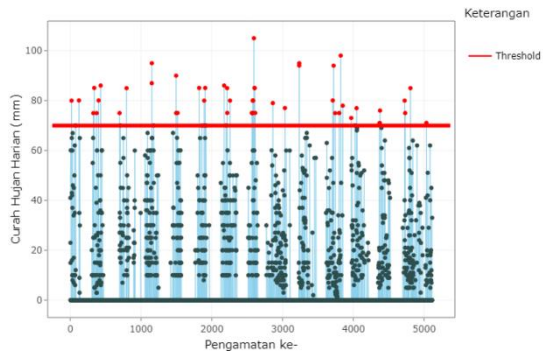
**Lampiran 16B.** Pengambilan Sampel Ekstrem Pos Kebon Agung



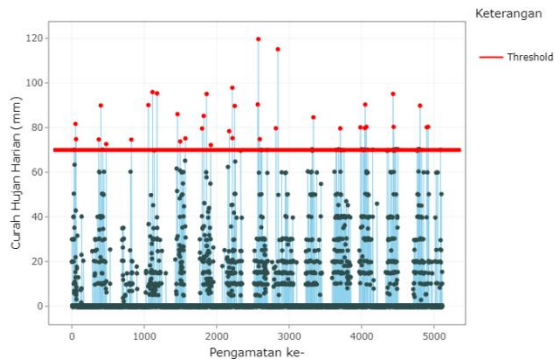
**Lampiran 16C.** Pengambilan Sampel Ekstrem Pos Gedeg



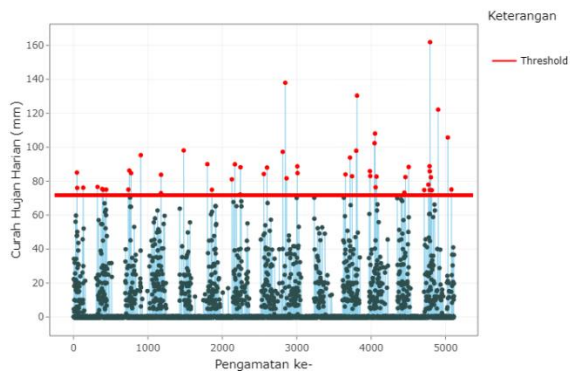
### Lampiran 16D. Pengambilan Sampel Ekstrem Pos Janjing

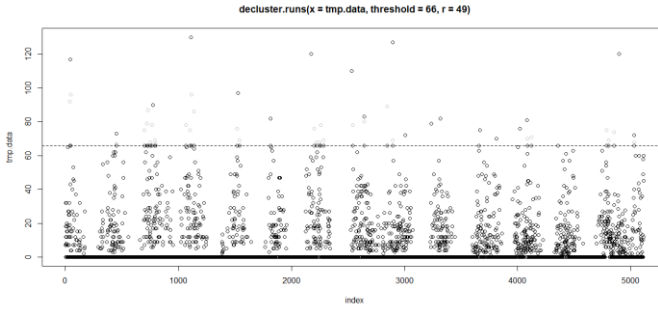
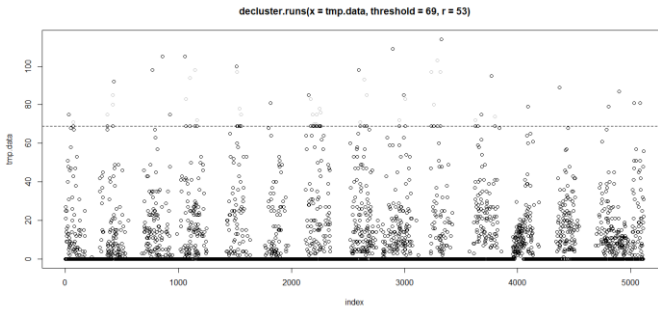
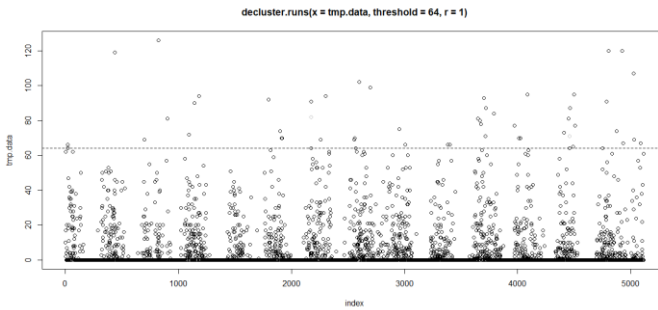


### Lampiran 16E. Pengambilan Sampel Ekstrem Pos Klegan

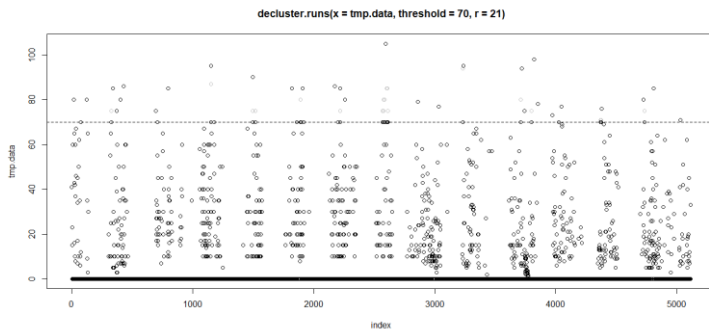
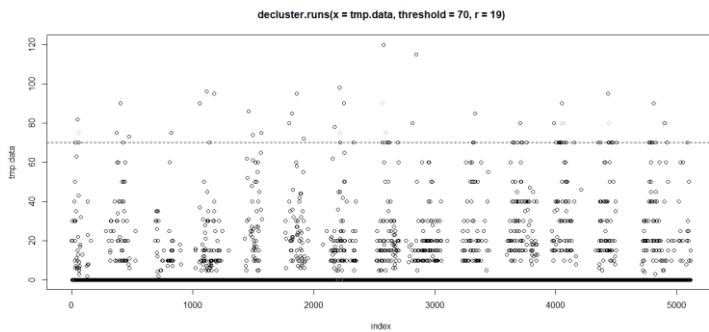
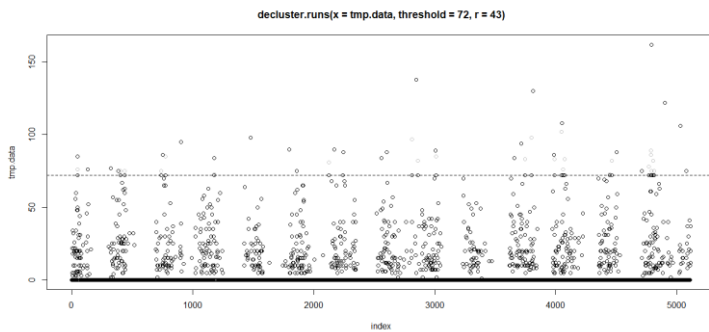


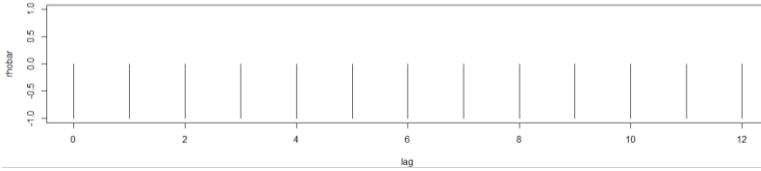
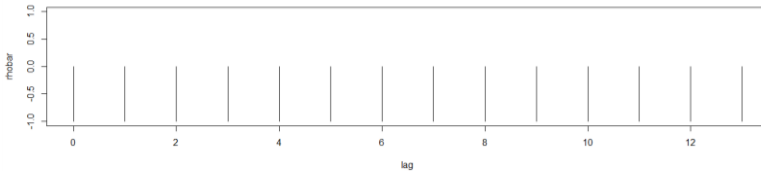
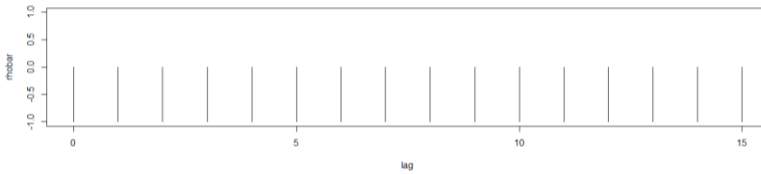
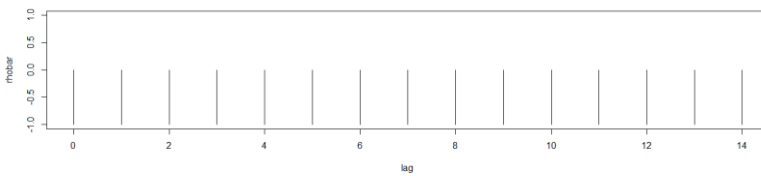
### Lampiran 16F. Pengambilan Sampel Ekstrem Pos Mojosari



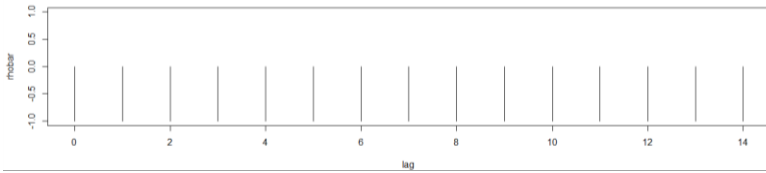
**Lampiran 17.** Pengambilan Sampel Ekstrem *Declustering***Lampiran 17A.** Plot Hasil *Declustering* Pos Kandangan**Lampiran 17B.** Plot Hasil *Declustering* Pos Kebon Agung**Lampiran 17C.** Plot Hasil *Declustering* Pos Gedeg



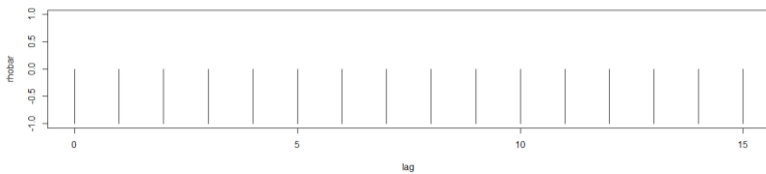
**Lampiran 17D.** Plot Hasil *Declustering* Pos Janjing**Lampiran 17E.** Plot Hasil *Declustering* Pos Klegen**Lampiran 17F.** Plot Hasil *Declustering* Pos Mojosari

**Lampiran 18.** *Extremogram Data Ekstrem Declsutering***Lampiran 18A.** *Extremogram Data Ekstrem Declsutering Pos Kandangan***Lampiran 18B.** *Extremogram Data Ekstrem Declsutering Pos Kebon Agung***Lampiran 18C.** *Extremogram Data Ekstrem Declsutering Pos Gedeg***Lampiran 18D.** *Extremogram Data Ekstrem Declsutering Pos Janjing*

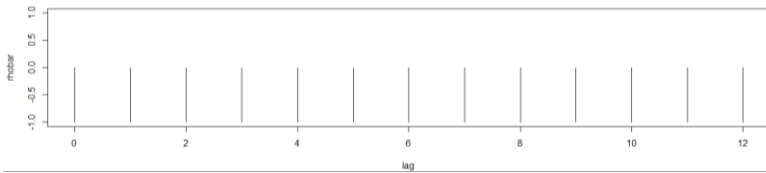
**Lampiran 18E.** *Extremogram Data Ekstrem Declsutering Pos Klegan*

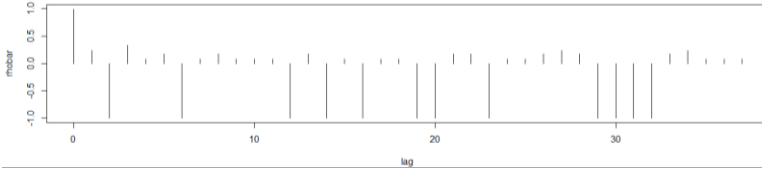
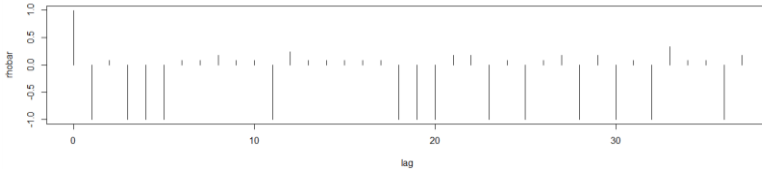
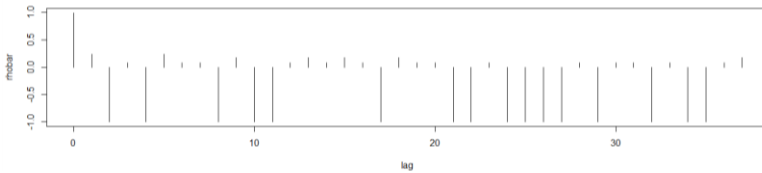
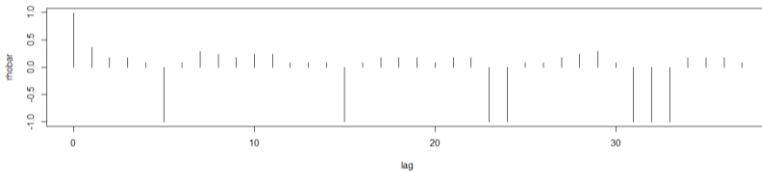


**Lampiran 18F.** *Extremogram Data Ekstrem Declsutering Pos Mojosari*

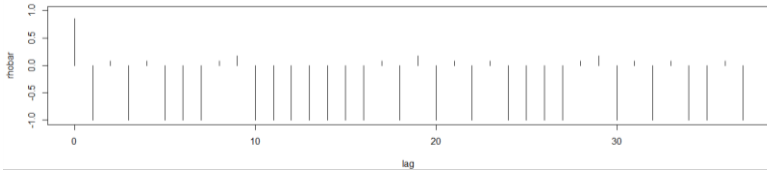


**Lampiran 18G.** *Extremogram Data Ekstrem Declsutering Pos Pacet*

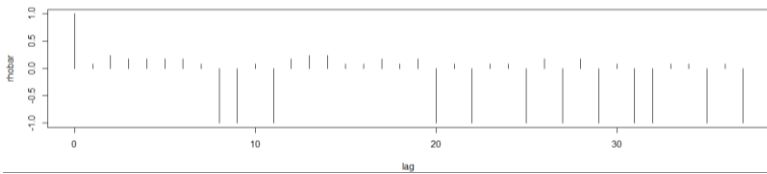


**Lampiran 19.** *Extremogram Data Ekstrem Non Declustering***Lampiran 19A.** *Extremogram Data Ekstrem Non Declustering*  
Pos Kandangan**Lampiran 19B.** *Extremogram Data Ekstrem Non Declustering*  
Pos Kebon Agung**Lampiran 19C.** *Extremogram Data Ekstrem Non Declustering*  
Pos Gedeg**Lampiran 19D.** *Extremogram Data Ekstrem Non Declustering*  
Pos Janjing

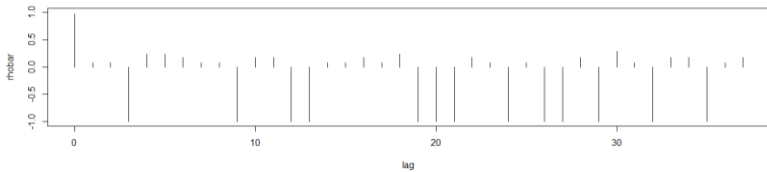
**Lampiran 19E.** *Extremogram Data Ekstrem Non Declsutering*  
Pos Klegan



**Lampiran 19F.** *Extremogram Data Ekstrem Non Declsutering*  
Pos Mojosari



**Lampiran 19G.** *Extremogram Data Ekstrem Non Declsutering*  
Pos Pacet



**Lampiran 20.** Syntax R Histogram Data Curah Hujan Harian

```

#PACKAGE YANG PERLU DI INSTALL
library(extRemes)
library(Kendall)
library(in2extRemes)
library(ggplot2)
library(tidyr)
library(dplyr)
library(ggpmisc)
in2extRemes()
theme_set(theme_minimal())
dev.off()

#INPUT DATA TRAINING
getwd()
setwd("E:/TUGAS AKHIR/Data Training")
data_training <- read.csv("data_training.csv", sep =
",", head=TRUE)

#HISTOGRAM
Hist1 <- function(method = "FD"){
  h1 <- hist(data_training$KANDANGAN, breaks = method,
            plot = FALSE)
  plot_ly(x=h1$mids, y=h1$counts,
          marker = list(color = 'darkcyan'))%>%
    layout(title="",
           xaxis=list(title="Curah Hujan Harian (mm)")%>%
             addBars(name = method))}
subplot(Hist1("Scott"), nrows = 1, shareX = TRUE)

Hist2 <- function(method = "FD"){
  h2 <- hist(data_training$KEBONAGUNG, breaks = method,
            plot = FALSE)
  plot_ly(x=h2$mids, y=h2$counts,
          marker = list(color = 'darkcyan')) %>%
    layout(title="",
           xaxis=list(title="Curah Hujan Harian (mm)")%>%
             addBars(name = method))}
subplot(Hist2("Scott"), nrows = 1, shareX = TRUE)

Hist3 <- function(method = "FD"){
  h3 <- hist(data_training$GEDEG, breaks = method,
            plot = FALSE)
  plot_ly(x=h3$mids, y=h3$counts,
          marker = list(color = 'darkcyan')) %>%
    layout(title="",

```

```

        xaxis=list(title="Curah Hujan Harian (mm)")>%
        add_bars(name = method)}
subplot(Hist3("Scott"), nrows = 1, shareX = TRUE)

Hist4 <- function(method = "FD"){
  h4 <- hist(data_training$JANJING, breaks = method,
            plot = FALSE)
  plot_ly(x=h4$mids, y=h4$counts,
          marker = list(color = 'darkcyan')) %>%
  layout(title="",
         xaxis=list(title="Curah Hujan Harian (mm)")>%
         add_bars(name = method)}
subplot(Hist4("Scott"), nrows = 1, shareX = TRUE)

Hist5 <- function(method = "FD"){
  h5 <- hist(data_training$KLEGEN, breaks = method,
            plot = FALSE)
  plot_ly(x=h5$mids, y=h5$counts,
          marker = list(color = 'darkcyan')) %>%
  layout(title="",
         xaxis=list(title="Curah Hujan Harian (mm)")>%
         add_bars(name = method)}
subplot(Hist5("Scott"), nrows = 1, shareX = TRUE)

Hist6 <- function(method = "FD"){
  h6 <- hist(data_training$MOJOSARI, breaks = method,
            plot = FALSE)
  plot_ly(x=h6$mids, y=h6$counts,
          marker = list(color = 'darkcyan')) %>%
  layout(title="",
         xaxis=list(title="Curah Hujan Harian (mm)")>%
         add_bars(name = method)}
subplot(Hist6("Scott"), nrows = 1, shareX = TRUE)

Hist7 <- function(method = "FD"){
  h7 <- hist(data_training$PACET, breaks = method,
            plot = FALSE)
  plot_ly(x=h7$mids, y=h7$counts,
          marker = list(color = 'darkcyan')) %>%
  layout(title="",
         xaxis=list(title="Curah Hujan Harian (mm)")>%
         add_bars(name = method)}
subplot(Hist7("Scott"), nrows = 1, shareX = TRUE)

```

## Lampiran 21. Syntax *R Time Series Plot* Data Curah Hujan Harian

```
#INPUT DATA TRAINING
getwd()
setwd("E:/TUGAS AKHIR/Data Training")
dataCH <- read.csv("data_training.csv", sep = ",", head=TRUE)

#Time Series Plot
T1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
y=dataCH$KANDANGAN))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(T1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

U1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
y=dataCH$KEBONAGUNG))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(U1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

V1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan, y=dataCH$GEDEG))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(V1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

W1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
y=dataCH$JANJING))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(W1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

X1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
y=dataCH$KLEGEN))+
  geom_line(color = "skyblue", size = 0.2)+
```



```

geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
theme_bw()
ggplotly(X1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

Y1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
y=dataCH$MOJOSARI))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(Y1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

Z1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan, y=dataCH$PACET))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_jitter(color="darkslategrey",width=0.8, size=0.8)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  theme_bw()
ggplotly(Z1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

```

## Lampiran 22. Syntax R Pengambilan Sampel Ekstrem POT

```

#INPUT DATA TRAINING
getwd()
setwd("E:/TUGAS AKHIR/Data Training")
dataCH <- read.csv("data_training.csv", sep = ",", head=TRUE)

#Menentukan Threshold 1% <- quantile(dataCH$KANDANGAN, 0.99,
na.rm = TRUE)
batas1 <- quantile(dataCH$KANDANGAN, 0.99, na.rm = TRUE)
batas2 <- quantile(dataCH$KEBONAGUNG, 0.99, na.rm = TRUE)
batas3 <- quantile(dataCH$GEDEG, 0.99, na.rm = TRUE)
batas4 <- quantile(dataCH$JANJING, 0.99, na.rm = TRUE)
batas5 <- quantile(dataCH$KLEGEN, 0.99, na.rm = TRUE)
batas6 <- quantile(dataCH$MOJOSARI, 0.99, na.rm = TRUE)
batas7 <- quantile(dataCH$PACET, 0.99, na.rm = TRUE)

#Plot Penentuan Threshold
POT1 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                           y=dataCH$KANDANGAN))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas1, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,6]>batas1,"red",
                           "darkslategrey"),size=1)+
  geom_hline(yintercept = batas1, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name ="Keterangan",
                     values = c("red", "darkgreen", "red"),
                     labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT1, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

POT2 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                           y=dataCH$KEBONAGUNG))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas2, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,7]>batas2,"red",
                           "darkslategrey"), size=1)+
  geom_hline(yintercept = batas2, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name ="Keterangan",
                     values = c("red", "darkgreen",
"orangered"),
                     labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT2, dynamicTicks = TRUE) %>%

```

```

layout(xaxis=list(autorange=T, range=c(1-5114)))

POT3 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                           y=dataCH$GEDEG))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas3, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,8]>batas3,"red",
                           "darkslategrey"), size=1)+
  geom_hline(yintercept = batas3, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name = "Keterangan",
                     values = c("red", "darkgreen",
                                "orangered"),
                     labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT3, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

POT4 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                           y=dataCH$JANJING))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas4, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,9]>batas4,"red",
                           "darkslategrey"), size=1)+
  geom_hline(yintercept = batas4, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name = "Keterangan",
                     values = c("red", "darkgreen",
                                "orangered"),
                     labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT4, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

POT5 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                           y=dataCH$KLEGEN))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas5, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,10]>batas5,"red",
                           "darkslategrey"),
             width=1, size=1)+
  geom_hline(yintercept = batas5, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name = "Keterangan",
                     values = c("red", "darkgreen",
                                "orangered"),
                     labels=c("Threshold", "", ""))+
  theme_bw()

```

```

ggplotly(POT5, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

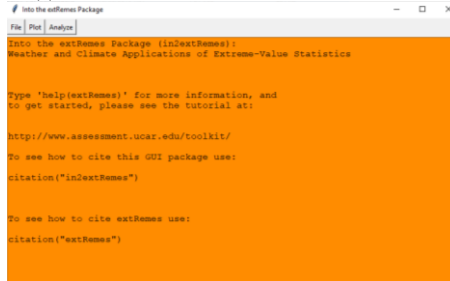
POT6 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                          y=dataCH$MOJOSARI))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas6, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,11]>batas6,"red",
                          "darkslategrey"),
             width=1, size=1)+
  geom_hline(yintercept = batas6, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name = "Keterangan",
                    values = c("red", "darkgreen",
                                "orangered"),
                    labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT6, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

POT7 <- ggplot(dataCH, aes(x=dataCH$Pengamatan,
                          y=dataCH$PACET))+
  geom_line(color = "skyblue", size = 0.2)+
  geom_hline(aes(yintercept=batas7, colour="Threshold"))+
  geom_jitter(color=ifelse(dataCH[,12]>batas7,"red",
                          "darkslategrey"),
             width=1, size=1)+
  geom_hline(yintercept = batas7, color="red", size=1.3)+
  labs( x="Pengamatan ke-", y="Curah Hujan Harian (mm)" )+
  scale_color_manual(name = "Keterangan",
                    values = c("red", "darkgreen",
                                "orangered"),
                    labels=c("Threshold", "", ""))+
  theme_bw()
ggplotly(POT7, dynamicTicks = TRUE) %>%
  layout(xaxis=list(autorange=T, range=c(1-5114)))

```

### Lampiran 23. Syntax R Declustering

```
#Declustering
library(in2extRemes)
in2extRemes()
par(mar=rep(2,4))
```



```
#Mengeksport Data Setelah Declustering
write.csv(DATA_TRAINING[["data"]], 'E:/TUGAS AKHIR/Data
Training/Hasil Declustering.csv')
```

### Lampiran 24. Syntax R Estimasi Parameter Declustering

```
#READ DATA EKSTREM DECLUSTERING
getwd()
setwd("E:/TUGAS AKHIR/Data Training")
data_DEC_KANDANGAN <- read.csv("DEC_KANDANGAN.csv", sep =
",", head=TRUE)
data_DEC_KEBONAGUNG <- read.csv("DEC_KEBONAGUNG.csv", sep =
",", head=TRUE)
data_DEC_GEDEG <- read.csv("DEC_GEDEG.csv", sep =
",", head=TRUE)
data_DEC_JANJING <- read.csv("DEC_JANJING.csv", sep =
",", head=TRUE)
data_DEC_KLEGEN <- read.csv("DEC_KLEGEN.csv", sep =
",", head=TRUE)
data_DEC_MOJOSARI <- read.csv("DEC_MOJOSARI.csv", sep =
",", head=TRUE)
data_DEC_PACET <- read.csv("DEC_PACET.csv", sep =
",", head=TRUE)

#ESTIMASI PARAMETER DECLUSTERING
EP_DEC_1 <- fevd(x = KANDANGAN, data = data_DEC_KANDANGAN,
threshold = 66, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
```

```

                                type = "GP", units = "mm",
                                time.units = "days", na.action = na.fail)
EP_DEC_1
EP_DEC_2 <- fevd(x = KEBONAGUNG, data = data_DEC_KEBONAGUNG,
                threshold = 69, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_2
EP_DEC_3 <- fevd(x = GEDEG, data = data_DEC_GEDEG,
                threshold = 64, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_3
EP_DEC_4 <- fevd(x = JANJING, data = data_DEC_JANJING,
                threshold = 70, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_4
EP_DEC_5 <- fevd(x = KLEGEN, data = data_DEC_KLEGEN,
                threshold = 70, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_5
EP_DEC_6 <- fevd(x = MOJOSARI, data = data_DEC_MOJOSARI,
                threshold = 72, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_6
EP_DEC_7 <- fevd(x = PACET, data = data_DEC_PACET,
                threshold = 73, threshold.fun = ~1,
                location.fun = ~1, scale.fun = ~1,
                shape.fun = ~1, use.phi = FALSE,
                type = "GP", units = "mm",
                time.units = "days", na.action = na.fail)
EP_DEC_7

```

## Lampiran 25. *Syntax R* Estimasi Parameter *Non Declustering*

```
#READ DATA EKSTREM NON DECLSTERING
getwd()
setwd("E:/TUGAS AKHIR/Data Training")
data_EXT1_KANDANGAN <- read.csv("EXT1_KANDANGAN.csv", sep =
",",head=TRUE)
data_EXT1_KEBONAGUNG <- read.csv("EXT1_KEBONAGUNG.csv", sep =
",",head=TRUE)
data_EXT1_GEGEG <- read.csv("EXT1_GEGEG.csv", sep =
",",head=TRUE)
data_EXT1_JANJING <- read.csv("EXT1_JANJING.csv", sep =
",",head=TRUE)
data_EXT1_KLEGEN <- read.csv("EXT1_KLEGEN.csv", sep =
",",head=TRUE)
data_EXT1_MOJOSARI <- read.csv("EXT1_MOJOSARI.csv", sep =
",",head=TRUE)
data_EXT1_PACET <- read.csv("EXT1_PACET.csv", sep =
",",head=TRUE)

#ESTIMASI PARAMETER NON DECLUSTERING
EP_NONDEC_1 <- fevd(x = KANDANGAN, data = data_EXT1_KANDANGAN,
threshold = 66, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_1
EP_NONDEC_2 <- fevd(x = KEBONAGUNG, data =
data_EXT1_KEBONAGUNG,
threshold = 69, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_2
EP_NONDEC_3 <- fevd(x = GEDEG, data = data_EXT1_GEGEG,
threshold = 64, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_3
EP_NONDEC_4 <- fevd(x = JANJING, data = data_EXT1_JANJING,
threshold = 70, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
```

```
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_4
EP_NONDEC_5 <- fevd(x = KLEGEN, data = data_EXT1_KLEGEN,
threshold = 70, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_5
EP_NONDEC_6 <- fevd(x = MOJOSARI, data = data_EXT1_MOJOSARI,
threshold = 72, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_6
EP_NONDEC_7 <- fevd(x = PACET, data = data_EXT1_PACET,
threshold = 73, threshold.fun = ~1,
location.fun = ~1, scale.fun = ~1,
shape.fun = ~1, use.phi = FALSE,
type = "GP", units = "mm",
time.units = "days", na.action = na.fail)
EP_NONDEC_7
```







```
RL_NONDEC_7
```

```
method = c("normal"),
do.ci = FALSE, verbose = FALSE,
qcov=B7, qcov.base = NULL)
```

## Lampiran 28. Output Declustering

### Lampiran 28A. Output Declustering Pos Kandangan

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index			
extremal.index	number.of.clusters		run.length
0.372549	19.000000		49.000000

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters		run.length
0.6587289	34.000000		13.000000

```
[1] "declustering performed for:"
[1] "KANDANGAN and assigned to KANDANGAN.u66r49dc"
```

### Lampiran 28B. Output Declustering Pos Kebon Agung

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index			
extremal.index	number.of.clusters		run.length
0.42	21.00		53.00

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters		run.length
0.8178007	41.000000		12.000000

```
[1] "declustering performed for:"
[1] "KEBONAGUNG and assigned to KEBONAGUNG.u69r53dc"
```

**Lampiran 28C. Output Declustering Pos Gedeg**

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index		
extremal.index	number.of.clusters	run.length
0.9411765	48.0000000	1.0000000

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters	run.length
0.7465916	39.0000000	12.0000000

```
[1] "declustering performed for:"
[1] "GEDEG and assigned to GEDEG.u64r1dc"
```

**Lampiran 28D. Output Declustering Pos Janjing**

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index		
extremal.index	number.of.clusters	run.length
0.58	29.00	21.00

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters	run.length
0.774105	39.0000000	6.0000000

```
[1] "declustering performed for:"
[1] "JANJING and assigned to JANJING.u70r21dc"
```

**Lampiran 28E. Output Declustering Pos Klegen**

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index		
extremal.index	number.of.clusters	run.length
0.7777778	28.0000000	19.0000000

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters	run.length
0.9728443	35.0000000	2.0000000

```
[1] "declustering performed for:"
[1] "KLEGEN and assigned to KLEGEN.u70r19dc"
```

**Lampiran 28F. Output Declustering Pos Mojosari**

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index		
extremal.index	number.of.clusters	run.length
0.5294118	27.0000000	43.0000000

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters	run.length
0.7454103	39.0000000	12.0000000

```
[1] "declustering performed for:"
[1] "MOJOSARI and assigned to MOJOSARI.u72r43dc"
```

**Lampiran 28G. Output Declustering Pos Pacet**

```
[1] "Declustering ..."
```

Runs Estimator for the Extremal Index		
extremal.index	number.of.clusters	run.length
0.25	12.00	189.00

Intervals Method Estimator for the Extremal Index  
NULL

theta.tilde used because there exist inter-exceedance times > 2.

extremal.index	number.of.clusters	run.length
0.7574632	36.0000000	10.0000000

```
[1] "declustering performed for:"
[1] "PACET and assigned to PACET.u73r189dc"
```

**Lampiran 29.** *Output* Estimasi Parameter GDP *Declustering***Lampiran 29A.** *Output* Estimasi Parameter GDP *Declustering*  
Pos Kandangan

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value:  79.07659

Estimated parameters:
      scale      shape
60.0088178 -0.9337679

Standard Error Estimates:
      scale      shape
6.614071  0.102463

Estimated parameter covariance matrix.
      scale      shape
scale 43.7459309 -0.67099239
shape -0.6709924  0.01049866

AIC = 162.1532

BIC = 164.0421
[1] "Model name: fit1"
```

**Lampiran 29B.** *Output* Estimasi Parameter GDP *Declustering*  
Pos Kebon Agung

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value:  79.9657

Estimated parameters:
      scale      shape
41.1430151 -0.9093123

Standard Error Estimates:
      scale      shape
7.1691010  0.1624859

Estimated parameter covariance matrix.
      scale      shape
scale 51.396009 -1.15915636
shape -1.159156  0.02640168

AIC = 163.9314
```

```
BIC = 166.0204
[1] "Model name: fit1"
```

**Lampiran 29C.** *Output Estimasi Parameter GDP Declustering*  
Pos Gedeg

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 156.138

Estimated parameters:
  scale      shape
30.4772267 -0.4134104

Standard Error Estimates:
  scale      shape
6.9891353 0.1771278

Estimated parameter covariance matrix.
      scale      shape
scale 48.848012 -1.14501099
shape -1.145011  0.03137425

AIC = 316.276

BIC = 319.6031
[1] "Model name: fit1"
```

**Lampiran 29D.** *Output Estimasi Parameter GDP Declustering*  
Pos Janjing

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 100.3793

Estimated parameters:
  scale      shape
19.9536586 -0.5320962

Standard Error Estimates:
  scale      shape
4.3952111 0.1542751

Estimated parameter covariance matrix.
      scale      shape
scale 19.3178809 -0.64615395
shape -0.6461539  0.02380082
```

```
AIC = 204.7585
```

```
BIC = 207.4931
```

```
[1] "Model name: fit1"
```

**Lampiran 29E. Output Estimasi Parameter GDP Declustering  
Pos Klegen**

```
[1] "Estimation Method used: MLE"
```

```
Negative Log-Likelihood Value: 104.8391
```

```
Estimated parameters:
```

```
  scale      shape  
22.5815770 -0.3728656
```

```
Standard Error Estimates:
```

```
  scale      shape  
5.4002389 0.1626493
```

```
Estimated parameter covariance matrix.
```

```
  scale      shape  
scale 29.1625799 -0.78579181  
shape -0.7857918  0.02645478
```

```
AIC = 213.6782
```

```
BIC = 216.3426
```

```
[1] "Model name: fit1"
```

**Lampiran 29F. Output Estimasi Parameter GDP Declustering  
Pos Mojosari**

```
[1] "Estimation Method used: MLE"
```

```
Negative Log-Likelihood Value: 139.3824
```

```
Estimated parameters:
```

```
  scale      shape  
20.31445731 -0.02889858
```

```
Standard Error Estimates:
```

```
  scale      shape  
4.7449938 0.1612699
```

```
Estimated parameter covariance matrix.
```

```
  scale      shape  
scale 22.5149658 -0.54460884  
shape -0.5446088  0.02600797
```



```
AIC = 282.7648
BIC = 285.8755
[1] "Model name: fit2"
```

**Lampiran 29G.** *Output* Estimasi Parameter GDP *Declustering*  
Pos Pacet

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 75.03169

Estimated parameters:
      scale      shape
41.9399150 -0.0469001

Standard Error Estimates:
      scale      shape
15.1307785  0.2603448

Estimated parameter covariance matrix.
      scale      shape
scale 228.940457 -2.9605026
shape -2.960503  0.0677794

AIC = 154.0634

BIC = 155.6086
[1] "Model name: fit1"
```

**Lampiran 30.** *Output* Estimasi Parameter GDP Non Declustering**Lampiran 30A.** *Output* Estimasi Parameter GDP Non Declustering Pos Kandangan

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value:  193.3924

Estimated parameters:
      scale      shape
1.631991e+01 1.772776e-08

Standard Error Estimates:
      scale      shape
2.9229006 0.1519071

Estimated parameter covariance matrix.
      scale      shape
scale  8.5433480 -0.30981292
shape -0.3098129  0.02307577

AIC = 390.7847

BIC = 394.6484
[1] "Model name: fit1"
```

**Lampiran 30B.** *Output* Estimasi Parameter GDP Non Declustering Pos Kebon Agung

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value:  184.2266

Estimated parameters:
      scale      shape
24.1882372 -0.5013332

Standard Error Estimates:
      scale      shape
4.1324205 0.1192431

Estimated parameter covariance matrix.
      scale      shape
scale 17.0768990 -0.46477936
shape -0.4647794  0.01421891

AIC = 372.4533
```

```
BIC = 376.2773
[1] "Model name: fit1"
```

**Lampiran 30C. Output Estimasi Parameter GDP Non  
Declustering Pos Gedeg**

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 197.9474

Estimated parameters:
  scale      shape
22.1100681 -0.2149236

Standard Error Estimates:
  scale      shape
4.6013495 0.1569304

Estimated parameter covariance matrix.
  scale      shape
scale 21.1724176 -0.61751751
shape -0.6175175 0.02462714

AIC = 399.8948

BIC = 403.7584
[1] "Model name: fit1"
```

**Lampiran 30D. Output Estimasi Parameter GDP Non  
Declustering Pos Janjing**

```
[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 166.6413

Estimated parameters:
  scale      shape
15.0466706 -0.3782803

Standard Error Estimates:
  scale      shape
2.5143294 0.1038225

Estimated parameter covariance matrix.
  scale      shape
scale 6.3218521 -0.23298647
shape -0.2329865 0.01077911
```

```
AIC = 337.2825
```

```
BIC = 341.1066
```

```
[1] "Model name: fit1"
```

**Lampiran 30E. Output Estimasi Parameter GDP Non  
Declustering Pos Klegen**

```
[1] "Estimation Method used: MLE"
```

```
Negative Log-Likelihood Value: 135.4055
```

```
Estimated parameters:
```

```
      scale      shape
1.093222e+01 9.999988e-09
```

```
Standard Error Estimates:
```

```
      scale      shape
1.9948048 0.1533139
```

```
Estimated parameter covariance matrix.
```

```
      scale      shape
scale 3.9792461 -0.21991029
shape -0.2199103 0.02350516
```

```
AIC = 274.811
```

```
BIC = 277.9781
```

```
[1] "Model name: fit1"
```

**Lampiran 30F. Output Estimasi Parameter GDP Non  
Declustering Pos Mojosari**

```
[1] "Estimation Method used: MLE"
```

```
Negative Log-Likelihood Value: 194.15
```

```
Estimated parameters:
```

```
      scale      shape
16.0104735 0.0335948
```

```
Standard Error Estimates:
```

```
      scale      shape
3.112511 0.134941
```

```
Estimated parameter covariance matrix.
```

```
      scale      shape
scale 9.6877228 -0.28063053
shape -0.2806305 0.01820908
```

```

AIC = 392.3
BIC = 396.1637
[1] "Model name: fit1"

```

**Lampiran 30G.** *Output* Estimasi Parameter GDP Non  
*Declustering* Pos Pacet

```

[1] "Estimation Method used: MLE"

Negative Log-Likelihood Value: 198.776

Estimated parameters:
      scale      shape
20.6723554  0.1123703

Standard Error Estimates:
      scale      shape
4.1408194  0.1398288

Estimated parameter covariance matrix.
      scale      shape
scale 17.1463853 -0.3503613
shape -0.3503613  0.0195521

AIC = 401.5521

BIC = 405.2945
[1] "Model name: fit1"

```

**Lampiran 31. Output Return Level Declustering****Lampiran 31A. Output Return Level Declustering Pos****Kandangan**

```
[1] "GP model fitted to KANDANGAN (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_KANDANGAN"
[1] "Return Levels for period units in years"
2-year level 3-year level 4-year level
130.1291      130.1720      130.1940
```

**Lampiran 31B. Output Return Level Declustering Pos Kebon****Agung**

```
[1] "GP model fitted to KEBONAGUNG (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_KEBONAGUNG"
[1] "Return Levels for period units in years"
2-year level 3-year level 4-year level
114.1337      114.1684      114.1863
```

**Lampiran 31C. Output Return Level Declustering Pos Gedeg**

```
[1] "GP model fitted to GEDEG (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_GEDEG"
[1] "Return Levels for period units in years"
2-year level 3-year level 4-year level
132.8937      133.6388      134.0966
```

**Lampiran 31D. Output Return Level Declustering Pos Janjing**

```
[1] "GP model fitted to JANJING (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_JANJING"
[1] "Return Levels for period units in years"
2-year level 3-year level 4-year level
106.3773      106.5952      106.7236
```

**Lampiran 31E. Output Return Level Declustering Pos Klegen**

```
[1] "GP model fitted to KLEGEN (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_KLEGEN"
[1] "Return Levels for period units in years"
2-year level 3-year level 4-year level
125.3807      126.1078      126.5608
```

**Lampiran 31F. Output Return Level Declustering Pos Mojosari**

```
[1] "GP model fitted to MOJOSARI (mm)"
Data are assumed to be stationary
[1] "Covariate data = data_DEC_MOJOSARI"
[1] "Return Levels for period units in years"
```

2-year level	3-year level	4-year level
193.9596	200.7276	205.4817

### **Lampiran 31G. Output Return Level Declustering Pos Pacet**

[1] "GP model fitted to PACET (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_DEC_PACET"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
310.8679	323.2317	331.8625

### **Lampiran 32. Output Return Level Non Declustering**

#### **Lampiran 32A. Output Return Level Non Declustering Pos Kandangan**

[1] "GP model fitted to KANDANGAN (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_KANDANGAN"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
173.6091	180.2263	184.9212

#### **Lampiran 32B. Output Return Level Non Declustering Pos Kebon Agung**

[1] "GP model fitted to KEBONAGUNG (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_KEBONAGUNG"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
115.4783	115.8038	115.9978

#### **Lampiran 32C. Output Return Level Non Declustering Pos Gedeg**

[1] "GP model fitted to GEDEG (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_GEDEG"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
141.9370	144.0181	145.3885

#### **Lampiran 32D. Output Return Level Non Declustering Pos Janjing**

[1] "GP model fitted to JANJING (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_JANJING"		
[1] "Return Levels for period units in years"		

2-year level	3-year level	4-year level
106.4927	106.9597	107.2501

**Lampiran 32E. Output Return Level Non Declustering Pos Klegen**

[1] "GP model fitted to KLEGEN (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_KLEGEN"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
142.0841	146.5168	149.6618

**Lampiran 32F. Output Return Level Non Declustering Pos Mojosari**

[1] "GP model fitted to MOJOSARI (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_MOJOSARI"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
190.1746	198.3315	204.1866

**Lampiran 32G. Output Return Level Non Declustering Pos Pacet**

[1] "GP model fitted to PACET (mm)"		
Data are assumed to be stationary		
[1] "Covariate data = data_EXT1_PACET"		
[1] "Return Levels for period units in years"		
2-year level	3-year level	4-year level
274.9776	292.9688	306.2402



Lampiran 33. Tabel *Kolmogorov Test Statistic*

<i>n</i>	<i>p</i> = .900	.950	.975	.990	.995
4	4	4	6	6	6
5	6	6	8	8	10
6	7	9	11	11	13
7	9	11	13	15	17
8	10	14	16	18	20
9	12	16	18	22	24
10	15	19	21	25	27
11	17	21	25	29	31
12	18	24	28	34	36
13	22	26	32	38	42
14	23	31	35	41	45
15	27	33	39	47	51
16	28	36	44	50	56
17	32	40	48	56	62
18	35	43	51	61	67
19	37	47	55	65	73
20	40	50	60	70	78
21	42	54	64	76	84
22	45	59	69	81	89
23	49	63	73	87	97
24	52	66	78	92	102
25	56	70	84	98	108
26	59	75	89	105	115
27	61	79	93	111	123
28	66	84	98	116	128
29	68	88	104	124	136
30	73	93	109	129	143
31	75	97	115	135	149
32	80	102	120	142	158
33	84	106	126	150	164
34	87	111	131	155	173
35	91	115	137	163	179
36	94	120	144	170	188
37	98	126	150	176	196
38	103	131	155	183	203
39	107	137	161	191	211
40	110	142	168	198	220

**Lampiran 33.** Tabel *Kolmogorov Test Statistic* (Lanjutan)

<i>n</i>	<i>p</i> = .900	.950	.975	.990	.995
41	114	146	174	206	228
42	119	151	181	213	235
43	123	157	187	221	245
44	128	162	194	228	252
45	132	168	200	236	262
46	135	173	207	245	271
47	141	179	213	253	279
48	144	186	220	260	288
49	150	190	228	268	296
50	153	197	233	277	305
51	159	203	241	285	315
52	162	208	248	294	324
53	168	214	256	302	334
54	173	221	263	311	343
55	177	227	269	319	353
56	182	232	276	328	362
57	186	240	284	336	372
58	191	245	291	345	381
59	197	251	299	355	391
60	202	258	306	364	402

### Lampiran 34. Surat Pernyataan Data I

#### SURAT KETERANGAN

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

Nama : Añun Umami  
NRP : 062115 4000 0090


Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : BMKB Stasiun Klimatologi Malang

Divisi/ bagian : Observasi & Informasi  
sejak tanggal 4 Maret 2019 sampai dengan 15 April 2019 untuk keperluan Tugas Akhir/ ~~Thesis~~ Semester ~~Gesat~~ Genap\* 2018/ 2019.

2. Tidak Keberatan/~~Keberatan~~\* nama perusahaan dicantumkan dalam Tugas Akhir/ Thesis mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/~~Keberatan~~\* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.

Malang, 5 Juli 2019  
Pimpinan Perusahaan



( Anang Suprayitno, S.Si )  
NIP. 19741114 1996003 1001

\*~~(coret yang tidak perlu)~~

**Lampiran 35. Surat Pernyataan Data II****SURAT KETERANGAN**

Saya yang bertanda tangan di bawah ini menerangkan bahwa :

1. Mahasiswa Statistika FMKSD-ITS dengan identitas berikut :

Nama : Ainun Umami

NRP : 062115 4000 0090

Telah mengambil data di instansi/perusahaan kami :

Nama Instansi : Dinas Pekerjaan Umum Sumber Daya Air  
Provinsi Jawa Timur

Divisi/ bagian : Bidang Perencanaan Sumber Daya Air

sejak tanggal 18 April sampai dengan 18 Mei untuk  
keperluan Tugas Akhir/ ~~Thesis~~ Semester ~~Gasal~~/Genap\* 2018/2019.

2. Tidak Keberatan/~~Keberatan~~\* nama perusahaan dicantumkan dalam Tugas Akhir/ Thesis mahasiswa Statistika yang akan di simpan di Perpustakaan ITS dan dibaca di lingkungan ITS.
3. Tidak Keberatan/~~Keberatan~~\* bahwa hasil analisis data dari perusahaan dipublikasikan dalam E journal ITS yaitu Jurnal Sains dan Seni ITS.

Surabaya, 2 Juli 2019  
Kasie Hidrologi dan Pemantauan  
Kualitas Air  
Bidang Perencanaan SDA



\*(coret yang tidak perlu)

## BIODATA PENULIS



*A*inun Umami, penulis buku Tugas

Akhir ini lahir di Jombang pada 7 Juli 1997. Penulis merupakan anak pertama dari pasangan Mukhamad Nurhasim dan Siti Alfiyah. Penulis menempuh pendidikan formalnya di SD N Semboro 4 Jember, MTs N 1 Surabaya, dan MA N Surabaya. Kemudian penulis melanjutkan pendidikan sebagai mahasiswa Departemen Statistika di Institut Teknologi Sepuluh Nopember pada tahun 2015. Selama di Kampus Perjuangan Institut Teknologi Sepuluh Nopember, pada tahun pertama hingga tahun ketiga penulis aktif dalam organisasi kemahasiswaan lingkup departemen yang pernah menjabat sebagai Manager *Human Resource Development* SCC HIMASTA-ITS 2017/2018 dan Sekretaris Departemen Dalam Negeri HIMASTA-ITS 2017/2018. Sedangkan lingkup Fakultas, penulis aktif sebagai Ketua Departemen Keuangan BEM FMKSD ITS 2018/2019. Penulis juga aktif mengikuti pelatihan seperti LKMM Pra-TD BEM FMIPA ITS 2015, LKMM TD HIMASTA-ITS 2016, dan LKMM TM FMIPA ITS 2017. Kemudian di tahun keempat, penulis aktif dalam kegiatan internasionalisasi seperti menjadi kandidat *ITS Goes Beyond at Singapore* pada Maret 2019. Tidak hanya itu, penulis juga pernah mengikuti beberapa *project* untuk belajar lebih dekat dengan dunia kerja sebagai *data entry, surveyor, validator*, sampai menjadi *Person In Charge* (PIC).

Melalui pengalaman-pengalaman tersebut, penulis mendapat banyak pelajaran berharga, terutama dalam hal manajerial serta komunikasi. Selain itu penulis juga banyak

belajar mengenai bagaimana mengatasi permasalahan-permasalahan dan mencari solusinya. Apabila pembaca ingin memberi kritik dan saran serta berdiskusi lebih lanjut terkait Tugas Akhir ini, dapat menghubungi penulis melalui email [ainunumami7797.au@gmail.com](mailto:ainunumami7797.au@gmail.com).