



TUGAS AKHIR - SS 091324

Declustering Peaks Over Threshold Pada Data Curah Hujan Ekstrim Dependen di Sentra Produksi Padi Jawa Timur

Rosna Malika
NRP 1310 100 073

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

JURUSAN STATISTIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2014



TUGAS AKHIR - SS091324

***DECLUSTERING PEAKS OVER THRESHOLD* PADA DATA
CURAH HUJAN EKSTRIM DEPENDEN DI SENTRA
PRODUKSI PADI JAWA TIMUR**

Rosna Malika
NRP 1310 100 073

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

JURUSAN STATISTIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2014



FINAL PROJECT - SS091324

***DECLUSTERING PEAKS OVER THRESHOLD OF
DEPENDENT EXTREME RAINFALL IN CENTER OF PADDY
PRODUCTION IN EAST JAVA***

Rosna Malika
NRP 1310 100 073

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

DEPARTMENT OF STATISTICS
Faculty Of Mathematics And Science
Sepuluh Nopember Institute Of Technology
Surabaya 2014

Declustering Peaks Over Threshold
Pada Data Curah Hujan Ekstrim Dependen
di Sentra Produksi Padi Jawa Timur

Nama : Rosna Malika
NRP : 1310100073
Jurusan : Statistika FMIPA – ITS
Pembimbing : Dr. Sutikno, S.Si, M.Si.

ABSTRAK

Iklm ekstrim merupakan fenomena langka, namun dapat memicu dampak merugikan salah satunya bagi sektor pertanian. Curah hujan yang ekstrim perlu diidentifikasi untuk mengurangi kerugian pertanian. Extreme Value Theory (EVT) merupakan salah satu metode statistika yang digunakan untuk mempelajari perilaku nilai-nilai ekstrim. Salah satu yang menjadi penting dalam model extreme value adalah return level. Return level merupakan nilai maksimum yang diharapkan akan dilampaui satu kali dalam jangka waktu tertentu. Sebelum menentukan nilai return level, terlebih dahulu harus diketahui nilai parameter distribusi EVT yang dipilih, dengan syarat urutan data harus independen. Namun pada kenyataannya, seringkali ditemukan data ekstrim yang dependen, akibatnya akan diperoleh nilai return level yang tidak valid. Salah satu cara untuk mengatasi hal tersebut dengan metode declustering. Penelitian ini menganalisis data curah hujan ekstrim dependen dengan pendekatan Peaks Over Threshold (POT) serta mengatasi dependensi data dengan declustering. Hasil penelitian menunjukkan bahwa metode declustering berhasil mengatasi dependensi data, namun tidak menghasilkan return level yang lebih baik daripada non declustering. Hal ini ditunjukkan nilai RMSE return level declustering yang sama dengan non declustering. Oleh karena itu, pada analisis selanjutnya sebaiknya menggunakan metode pemodelan parameter Generalized Pareto Distribution.

Kata Kunci: *Curah Hujan, Extreme Value Theory (EVT), Peaks Over Threshold, Generalized Pareto Distribution, Declustering, Return Level.*

(Halaman ini sengaja dikosongkan)

***Declustering Peaks Over Threshold Of Dependent
Extreme Rainfall In Center Of Paddy Production
In East Java***

Name : Rosna Malika
NRP : 1310100073
Department : Statistics FMIPA – ITS
Supervisor : Dr. Sutikno, S.Si, M.Si.

ABSTRACT

Extreme climate is a rare phenomenon that lead bad impact for the agricultural sector. Extreme rainfall event need to be identified to minimize losses due to the incident. Extreme Value Theory (EVT) is one of the statistical methods that used to study the behavior of extreme value. It is usually more convenient to interpret extreme value models in terms of return levels. To determine the value of return level, must be known parameter values of distribution first. Each of extreme value models derived so far has been obtained through mathematical arguments that assume an underlying process consisting of a sequence of independent random variables. However, in the real case this condition is often violated, then its reducing the validity of conclusions. This research was conducted in five districts as the center of paddy production in East Java. In this study the method of approach to the Peaks Over Threshold (POT) is used to determine the return level, and its also used declustering method to solve the problem of dependent data. The results showed that the order of the data become more independent when declustering method is applied. But the RMSE of return level by using declustering method have same value as RMSE of return level with non declustering method. Furthermore, it will be better if the data analyze with approach to modelling of Generalized Pareto Distribution parameter.

Key Word : *Rainfall, Extreme Value Theory (EVT), Peaks Over Threshold, Generalized Pareto Distribution, Declustering, Return Level.*

(Halaman ini sengaja dikosongkan)

LEMBAR PENGESAHAN

DECLUSTERING PEAKS OVER THRESHOLD PADA DATA CURAH HUJAN EKSTREM DEPENDEN DI SENTRA PRODUKSI PADI JAWA TIMUR

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Kelulusan
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

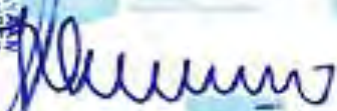
ROSNA MALIKA
NRP 1310 100 073

Disetujui oleh Pembimbing Tugas Akhir

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



Mengetahui
Ketua Jurusan Statistika FMIPA-ITS


Dr. Muhammad Mashuri, MT
NIP. 19620408 198701 1 001

SURABAYA, JANUARI 2014

KATA PENGANTAR

Allah mengajarkan kita untuk bersyukur, satu kata yang jauh lebih luas maknanya daripada terimakasih. Alhamdulillah, hanya kata itu yang pantas penulis ucapkan atas kemurahan Allah *azza wa jalla* yang tiada henti sehingga penulis dapat menyelesaikan buku Tugas Akhir yang berjudul “***Declustering Peaks Over Threshold Pada Data Curah Hujan Ekstrim Dependen di Sentra Produksi Padi Jawa Timur***” dengan baik.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih banyak kepada :

1. Bapak Dr. Sutikno, S.Si, M.Si selaku dosen pembimbing yang dengan sabar memberikan bimbingan, ilmu, saran, dan waktu yang beliau luangkan untuk membimbing penulis.
2. Bapak Dr.Bambang Wijarnako Otok, M.Si, dan Ibu Ir.Sri Pingit Wulandari, M.Si selaku dosen penguji atas kritik dan saran yang membangun dalam penyusunan Tugas Akhir ini.
3. Bapak Dr. Muhammad Mashuri, MT selaku Ketua Jurusan dan Ibu Dra.Lucia Aridinanti, MT selaku Ketua Prodi S1 Statistika ITS yang telah memberikan banyak fasilitas untuk kelancaran penyelesaian Tugas Akhir ini.
4. Ibunda, kakak, dan adik tercinta yang senantiasa mencurahkan segala rasa cinta dan semangatnya kepada penulis. Penulis yakin, dalam selaksa doa yang mereka panjatkan, ada nama penulis yang mereka sebut, agar nama itu meraih kesuksesan, kebahagiaan, dan keteguhan iman dalam naungan cinta-Nya. Serta tak lupa almarhum Ayahanda yang selalu dan selamanya akan penulis rindukan yang telah mendidik penulis dengan sepenuh hati demi mendapatkan kekayaan nurani. *Rabbighfirlil wa li walidayya...*
5. Keluarga besar FORSIS-ITS dan HIMASTA-ITS, wadah pertama kali bagi penulis untuk belajar berorganisasi, belajar menjadi *leader* dan *follower*, serta belajar bekerja secara *teamwork*. Rekan-rekan SOSMAS HIMASTA-ITS 12/13 yang keren dan luar biasa. Persaudaraan tulus dari teman-teman PH

FORSIS-ITS 12/13 yang seru, kocak dan menakjubkan, *The Great Team. Jazaakumullaahu khairan katsiiran...*

6. Teman-teman “amazing” Σ21 Statistika 2010, *we will always be amazing, guys*. Rekan-rekan seperjuangan PW109, yang saling mendukung dan memberi semangat demi kelancaran menuju upacara wisuda bersama-sama, dan teman-teman *Fast Track* Statistika angkatan 13/14, para calon ilmuwan Statistika masa depan.
7. Ida, Lala, Sari, Flashy, Hani, dan Umi atas segala sumbangan keceriaan, suka, dan duka dalam hari-hari penulis sejak tahun pertama kuliah, dan Mbak Yuli Kurniawati yang telah membantu proses pengolahan dan analisis data.
8. Keluarga BK DANUS JMMI 13/14, hadiah terindah buat penulis di semester akhir hidupnya sebagai mahasiswa S1. Sebuah pengalaman yang tak terduga bisa bekerja sama dengan teman-teman yang seru dari berbagai jurusan di ITS. *Jazaakumullaahu khairan katsiiran. Barakallah!*

Tak mampu penulis menyebut satu persatu semua pihak yang telah membantu dan mendukung. Tapi yang jelas penulis juga memohon ketulusan maaf atas segala kesalahan diri penulis, juga pada pembaca semua. Sungguh itu murni datangnya dari penulis yang masih miskin ilmu dan pengalaman. Oleh karena itu, penulis mengharapakan kritik dan saran dari pembaca agar kedepannya lebih baik. Kini atas nikmat-Nya buku ini hadir, semoga bisa bermanfaat bagi kita semua. Amin.

Surabaya, Januari 2014

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iv
ABSTRAK	v
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA	7
2.1 Statistika Deskriptif	7
2.2 <i>Extreme Value Theory</i> (EVT).....	9
2.3 <i>Peaks Over Threshold</i> (POT)	10
2.4 Penentuan Nilai <i>Threshold</i>	13
2.5 <i>Autocorrelation Function</i> (ACF).....	15
2.6 Estimasi Parameter <i>Generalized Pareto Distribution</i> (GPD)	16
2.7 Pemeriksaan Kesesuaian Distribusi	17
2.8 <i>Extreme of Dependent</i>	18
2.9 <i>Extremal Index</i>	22
2.10 <i>Return Level</i>	24
2.11 <i>Root Mean Square Error</i> (RMSE).....	25
2.12 Curah Hujan.....	25
2.13 Penelitian Sebelumnya	27

BAB III METODOLOGI PENELITIAN	29
3.1 Sumber Data	29
3.2 Variabel Penelitian.....	29
3.3 Langkah Analisis Data.....	29
BAB IV ANALISIS DAN PEMBAHASAN	33
4.1 Pra-Pemrosesan Data Curah Hujan.....	33
4.2 Deskripsi Curah Hujan di Lima Pos Pengamatan.....	36
4.3 Identifikasi Pola Data <i>Heavy Tail</i> dan Nilai Ekstrim.....	38
4.4 Pengambilan Sampel Ekstrim dengan <i>Peaks Over Threshold</i>	40
4.5 Identifikasi Dependensi Data Curah Hujan Ekstrim di Lima Pos Pengamatan	41
4.6 Pemeriksaan Kesesuaian Distribusi <i>Generalized Pareto</i>	45
4.7 Proses <i>Extremal Index</i> dan <i>Declustering</i>	47
4.8 Estimasi Parameter GPD di Lima Pos Pengamatan.....	49
4.9 <i>Return Level</i> Curah Hujan Ekstrim di Lima Pos Pengamatan	50
BAB V PENUTUP	53
5.1 Kesimpulan	53
5.2 Saran	53
DAFTAR PUSTAKA	55
LAMPIRAN	59

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Contoh Histogram 8
Gambar 2.2	Contoh Diagram Batang Rata-Rata Curah Hujan per Bulan..... 9
Gambar 2.3	Ilustrasi Pengambilan Data Sampel dengan Metode POT 10
Gambar 2.4	PDF Tipe Distribusi Eksponensial 11
Gambar 2.5	PDF Tipe Distribusi Pareto 12
Gambar 2.6	PDF Tipe Distribusi Beta 12
Gambar 2.7	<i>Mean Residual Life Plot</i> untuk Data Curah Hujan di Inggris..... 15
Gambar 2.8	Pola Data Menurut Runtun Waktu : Data <i>Non-Stationary</i> (a) dan Data <i>Stationary</i> (b) 19
Gambar 2.9	Pola Linear Trend 19
Gambar 2.10	Pola Data Temperatur Siklik 19
Gambar 2.11	Ilustrasi Metode <i>Run-declustering</i> Pada Data Temperatur dengan Nilai $r = 1$ 21
Gambar 2.12	Sebaran Pola Hujan di Indonesia..... 26
Gambar 3.1	Diagram Alir Analisis Data 32
Gambar 4.1	Plot Data Curah Hujan di Pos Cawak..... 34
Gambar 4.2	Plot Data Curah Hujan di Pos Cawak yang sudah dibersihkan dari Pengamatan Aneh (<i>clean</i>)..... 35
Gambar 4.3	Pola Curah Hujan Harian di Pos Cawak..... 37
Gambar 4.4	Histogram Curah Hujan Harian di Pos Cawak... 38
Gambar 4.5	Pengujian <i>Normality Probability Plot</i> Curah Hujan Harian di Pos Cawak..... 38

Gambar 4.6	Histogram Curah Hujan di Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun.....	39
Gambar 4.7	<i>Normality Probability Plot</i> Curah Hujan di Pos Pasewaran, Pos Karang Kedawung, Pos Ngale dan Pos Karangbinangun	39
Gambar 4.8	Pengambilan Sampel Ekstrim di Pos Cawak.....	40
Gambar 4.9	Pola Data Curah Hujan Harian Pos Cawak	42
Gambar 4.10	Pola Data Curah Hujan Harian di Pos Pasewaran Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun.....	42
Gambar 4.11	Pola Data Ekstrim Curah Hujan Harian di Pos Cawak	43
Gambar 4.12	Pola Data Ekstrim Curah Hujan Harian di Pos Pasewaran (a), Pos Karang Kedawung (b), Pos Ngale (c), dan Pos Karangbinangun (d).....	43
Gambar 4.13	Plot <i>Autocorrelation Function</i> Data Ekstrim Curah Hujan di Pos Cawak.....	44
Gambar 4.14	Plot <i>Autocorrelation Function</i> Data Ekstrim Curah Hujan Harian di Pos Pasewaran(a), Pos Karang Kedawung(b), Pos Karangbinangun(c) dan Pos Ngale (d).....	45
Gambar 4.15	<i>Probability Plot</i> dan <i>Quantile Plot</i> di Pos Cawak dengan GPD.....	46
Gambar 4.16	Plot <i>Autocorrelation Function</i> Data Curah Hujan Pos Cawak dengan <i>Declustering</i> $r = 8$	48

DAFTAR TABEL

	Halaman
Tabel 4.1 Identifikasi <i>Missing Value</i> Data Curah Hujan	33
Tabel 4.2 Nilai Rata-Rata, Standard Deviasi, Nilai Minimum, dan Nilai Maksimum Curah Hujan Harian (mm).....	36
Tabel 4.3 Jumlah Pengamatan (N), Nilai <i>Threshold</i> (u), dan Jumlah Sampel Ekstrim (n_u).....	41
Tabel 4.4 Uji <i>Kolmogorov-Smirnov Generalized Pareto Distribution</i>	46
Tabel 4.5 <i>Extremal Index</i> per Pos Pengamatan.....	47
Tabel 4.6 Estimasi Parameter GPD Pos Cawak, Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun	49
Tabel 4.7 Nilai <i>Return Level</i> (mm) di Lima Pos Pengamatan.....	50
Tabel 4.8 Nilai RMSE <i>Return Level</i>	51

(Halaman ini sengaja dikosongkan)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Iklim ekstrim merupakan fenomena yang langka namun dapat memicu dampak yang seringkali merugikan bagi kehidupan. Dua unsur utama iklim adalah suhu dan curah hujan. Indonesia sebagai daerah tropis ekuatorial mempunyai variasi suhu yang kecil, sementara variasi curah hujannya cukup besar (Hermawan, 2010). Oleh karena itu curah hujan merupakan unsur iklim yang paling sering diamati dibandingkan dengan suhu. Dampak yang mungkin timbul dari iklim ekstrim adalah tingginya curah hujan yang terus mengguyur dan berlangsung berkepanjangan. Jika ini terjadi akan berakibat timbulnya berbagai bencana alam dan sejumlah kerentanan, seperti kerentanan pangan dan ekosistem.

Pertanian merupakan salah satu sektor yang sangat merasakan dampak akibat iklim ekstrim tersebut. Iklim ekstrim, seperti curah hujan yang sangat tinggi serta kekeringan yang sangat lama, mengakibatkan kerusakan tanaman pangan dan kegagalan panen, sehingga produktivitasnya menurun. Pada skala yang ekstrim, berkurangnya hasil panen dapat mengancam ketahanan pangan sehingga menyebabkan pasokan pangan menjadi tidak pasti. Oleh karena itu dibutuhkan pengembangan aplikasi ilmu pengetahuan untuk mengetahui perilaku iklim ekstrim, sehingga produksi padi dapat dimaksimalkan dan kerugian akibat gagal panen juga bisa diminimalkan.

Menaksir dan memodelkan perilaku kejadian-kejadian ekstrim sering menjadi kajian yang menarik dalam bidang ekonomi, *finance*, hidrologi, dan klimatologi. *Extreme Value Theory* (EVT) merupakan salah satu metode statistika yang digunakan untuk mempelajari perilaku nilai-nilai ekstrim tersebut. Metode ini berfokus pada perilaku ekor (*tail*) suatu distribusi untuk dapat menentukan probabilitas nilai-nilai ekstrimnya, juga dapat meramalkan terjadinya kejadian ekstrim pada data *heavy*

tail yang tidak dapat dilakukan dengan pendekatan tradisional lainnya. Dalam mengidentifikasi pergerakan nilai ekstrim terdapat dua pendekatan yang digunakan yaitu dengan metode *Block Maxima* (BM) yang mengikuti distribusi *Generalized Extreme Value* (GEV) dan metode *Peaks Over Threshold* (POT) yang mengikuti distribusi *Generalized Pareto Distribution* (GPD) (Coles, 2001). Metode *Block Maxima* merupakan metode yang mengidentifikasi nilai ekstrim berdasarkan nilai tertinggi data observasi yang sudah dikelompokkan berdasarkan periode tertentu. Dalam metode BM data amatan yang dimasukkan dalam sampel adalah pengamatan yang memiliki nilai tertinggi, karena nilai maksimum tersebut merupakan nilai ekstrim data dalam satu periode tertentu. Metode *Peaks Over Threshold* menentukan nilai ekstrim dengan menggunakan besaran patokan yang disebut *threshold*. Semua data amatan yang berada di atas (melebihi) nilai *threshold* diidentifikasi sebagai nilai ekstrim. Wahyudi (2012) membandingkan dua pendekatan antara BM dan POT. Penelitian tersebut menghasilkan kesimpulan bahwa pendekatan POT memberikan hasil yang lebih sesuai daripada dengan pendekatan BM untuk curah hujan di Ngawi. Penelitian lain dilakukan oleh Jaruskova dan Hanek (2006) membandingkan metode POT dan BM untuk mengestimasi *return level* pada data curah hujan di Moravia Utara, Republik Ceko. Penelitian tersebut menghasilkan kesimpulan bahwa metode POT lebih sesuai untuk data ekstrim yang dependen.

Salah satu yang menjadi penting dalam interpretasi model *extreme value* adalah menghitung *return level*. Menurut Gilli dan Kellezi (2003) *return level* merupakan nilai maksimum yang diharapkan akan dilampaui satu kali dalam jangka waktu tertentu. Nilai *return level* dapat memberikan informasi yang penting dalam bidang ekonomi, keuangan, dan asuransi, serta pertanian. Pada bidang asuransi, *return level* digunakan untuk mengukur seberapa besar resiko yang ditanggung perusahaan asuransi dari klaim yang dianggap ekstrim atau melebihi rata-rata klaim lainnya sebagai salah satu upaya untuk menjaga portofolio

perusahaannya. Dalam bidang ekonomi dan keuangan, nilai *return level* dapat digunakan untuk mengetahui tingkat resiko saham yang diperlukan oleh para investor untuk menentukan saat yang tepat untuk berinvestasi. Sedangkan pada bidang pertanian nilai *return level* akan memberikan informasi waktu tanam yang sesuai untuk varietas tanaman pangan tertentu sehingga dapat ditentukan varietas tanaman yang sesuai dan adaptif terhadap keadaan yang ekstrim tersebut. Sebelum menentukan nilai *return level*, terlebih dahulu harus diketahui nilai parameter distribusi EVT yang dipilih. Syarat utama dalam menaksir parameter distribusi EVT adalah urutan data harus independen. Namun pada kenyataannya seringkali ditemukan data ekstrim yang dependen, sehingga melanggar syarat tersebut. Beberapa hal yang mempengaruhi stabilitas estimasi *return level* yaitu data yang dependen dan seasonal (Jaruskova dan Hanek, 2006). Oleh karena itu diperlukan penyelesaian adanya dependensi data, sehingga akan didapatkan nilai *return level* yang sesuai.

Beberapa penelitian yang pernah dilakukan untuk menganalisis kasus data ekstrim yang dependen adalah dengan memodelkan parameter distribusi EVT yang dipilih. Kurniawati (2013) melakukan analisis data ekstrim dependen pada kasus curah hujan di Jawa Timur dengan pendekatan *Peaks Over Threshold* (POT). Penelitian tersebut melakukan pemodelan parameter skala (*scale*) pada *Generalized Pareto Distribution* (GPD), serta menentukan *return level* curah hujan di Jawa Timur. Model parameter GPD diperoleh dengan melakukan transformasi data ekstrim dependen sesuai dengan pola data berupa siklik atau linier tren. Penelitian sejenis juga pernah dilakukan oleh Lopes, Nascimento, dan Gamerman (2011) yang menganalisis kejadian ekstrim dependen dengan memodelkan parameter GPD dependen terhadap waktu. Penelitian tersebut menggunakan metode *Dynamic Linear Model* (DLM) dengan pendekatan algoritma *Markov Chain Monte Carlo* (MCMC) untuk mendapatkan model parameter. Pada mulanya analisis dilakukan pada data simulasi

yang kemudian diaplikasikan pada tiga jenis data *financial time series*.

Selain memodelkan parameter, untuk mengatasi data ekstrim yang dependen juga bisa dilakukan dengan metode *declustering* yaitu mem-*filter* data dependen untuk mendapatkan satu set *threshold* yang independen. Secara khusus, beberapa aturan empiris digunakan untuk mendefinisikan kelompok (*cluster*) data ekstrim, nilai maksimal dalam setiap *cluster* diidentifikasi dan mengikuti *Generalized Pareto Distribution*, dengan asumsi independen. Salah satu cara sederhana untuk menentukan *cluster* umumnya dikenal dengan istilah *run-declustering*. Jaruskova dan Hanek (2006) melakukan penelitian dengan membandingkan metode POT dan BM untuk mengestimasi *return level* pada data curah hujan ekstrim di Moravia Utara, Republik Ceko. Pada penelitian tersebut diidentifikasi bahwa data curah hujan yang dianalisis memiliki kecenderungan bersifat dependen sehingga dianalisis dengan metode *declustering* terlebih dahulu. Penelitian lain juga dilakukan oleh Soukissian dan Arapi (2011) yang melakukan perbandingan metode *declustering* pada data tinggi gelombang laut. Metode *declustering* yang digunakan pada penelitian tersebut dengan 4 pendekatan yaitu *Standard Storm Length* (SSL), *Runs declustering* (RL), *DeClustering Algorithm* (DeCA), dan *Triple Annual Maximum* (TAM). Kesimpulan yang diperoleh dari penelitian tersebut yaitu dengan metode *declustering* melalui pendekatan *run-declustering* menunjukkan hasil dengan standar error yang lebih kecil dan selang kepercayaan yang lebih sempit.

Berdasarkan uraian tersebut, maka dalam tugas akhir ini mengkaji data curah hujan ekstrim dependen dengan pendekatan *Peaks Over Threshold* (POT) dan mengatasi dependensi data dengan metode *declustering*. Penelitian ini juga mengkaji curah hujan ekstrim di beberapa tempat Jawa Timur sebagai provinsi yang memberikan kontribusi produksi padi terbesar di Indonesia (BPS, 2013) yaitu di Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi. Salah satu hasil penelitian ini, juga

dapat menentukan *return level* dalam periode waktu ulang tertentu. Selanjutnya sebagai salah satu cara penanganan dampak kerugian produksi pertanian akibat iklim ekstrim dan dapat menentukan waktu tanam yang sesuai.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, permasalahan yang diselesaikan dalam penelitian ini adalah sebagai berikut.

1. Bagaimana karakteristik curah hujan ekstrim di Jawa timur khususnya Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi?
2. Bagaimana hasil estimasi parameter GPD dan *return level* curah hujan di lima kabupaten tersebut melalui *declustering* pada data curah hujan ekstrim?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah, maka tujuan penelitian ini adalah:

1. Mengetahui karakteristik curah hujan ekstrim di Jawa Timur khususnya Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi.
2. Mengetahui hasil estimasi parameter GPD dan *return level* curah hujan di lima kabupaten tersebut melalui *declustering* pada data curah hujan ekstrim.

1.4 Manfaat Penelitian

Manfaat penelitian ini antara lain sebagai berikut.

1. Bagi Departemen Pertanian, penelitian ini dapat dijadikan sebagai acuan prakiraan cuaca ekstrim sehinggaantisipasi kerugian akibat gagal panen dapat diminimalkan dan dapat digunakan untuk merancang strategi inovasi teknologi dan varietas tanaman pangan baru yang lebih adaptif terhadap iklim ekstrim.
2. Bagi BMKG, sebagai masukan metode alternatif untuk prakiraan iklim ekstrim.

3. Bagi Pengembangan keilmuan, dapat dijadikan pengetahuan dalam mengidentifikasi kejadian ekstrim terhadap cuaca dan iklim dan dapat dijadikan pustaka untuk penelitian berikutnya.

1.5 Batasan Masalah

Batasan yang digunakan dalam penelitian ini antara lain sebagai berikut.

1. Data yang digunakan berupa data curah hujan di Kabupaten Jember, Lamongan, Ngawi, Banyuwangi, dan Bojonegoro tahun 1981 sampai tahun 2010.
2. Metode *declustering* yang digunakan adalah dengan pendekatan *run-declustering*.

BAB II

TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Statistika merupakan metode pengumpulan data, analisis, interpretasi, dan penyimpulan hasil analisis (Johnson dan Bhattacharya, 1977). Ilmu statistika terbagi menjadi statistika deskriptif dan statistika inferensia. Pengertian statistika deskriptif adalah metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data, sehingga memberikan informasi yang berguna (Walpole, 1995). Penggunaan statistika deskriptif yakni memberikan informasi mengenai data dan tidak mengambil suatu kesimpulan (inferensi). Yang termasuk dalam statistika deskriptif antara lain: ukuran pemusatan data, ukuran penyebaran data, dan penyajian data misal: grafik dan diagram.

Ukuran pemusatan data dan ukuran penyebaran data merupakan suatu alat yang dapat digunakan untuk mendefinisikan ukuran-ukuran numerik yang menjelaskan karakteristik data. Ukuran pemusatan data yang sering digunakan untuk mendeskripsikan data adalah rata-rata. Rata-rata adalah hasil pembagian antara jumlahan 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 ; $i = 1, 2, \dots, n$

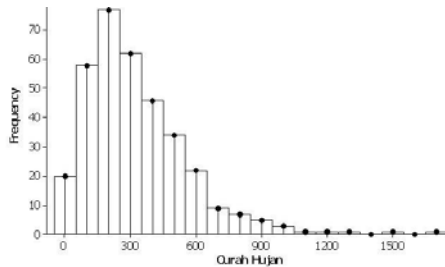
n : Banyaknya observasi

Ukuran penyebaran data yang sering digunakan untuk mendeskripsikan data adalah nilai maksimum, nilai minimum, dan varians. Nilai maksimum merupakan nilai tertinggi/terbesar yang terdapat dalam segugus data dan nilai minimum adalah nilai

terendah yang terdapat dalam sekumpulan data. Sedangkan varians (s^2) merupakan kuadrat simpangan dari semua nilai data terhadap rata-rata yang dituliskan dengan persamaan (2.2).

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

Secara visual penyebaran data dapat disajikan melalui histogram yang menggambarkan frekuensi distribusi data seperti pada Gambar 2.1.

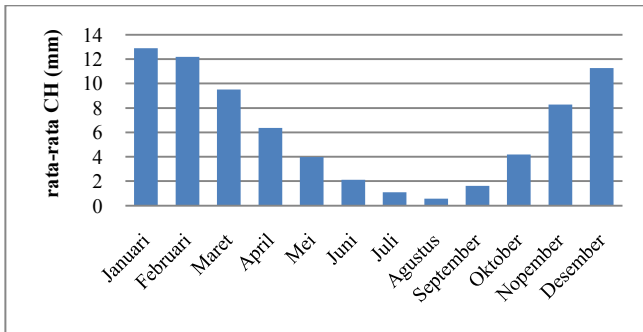


Gambar 2.1 Contoh Histogram

Histogram adalah 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 (Aczel dan Sounderpandian, 2008).

Selain histogram, penyajian data dalam statistika deskriptif juga bisa melalui diagram atau grafik, salah satunya adalah diagram batang. *Bar chart* (yang menggunakan persegi panjang horizontal atau vertikal) sering digunakan untuk menampilkan data kategori di mana tidak ada penekanan pada persentase dari total yang diwakili oleh masing-masing kategori (Aczel dan Sounderpandian, 2008). Diagram batang dengan menggunakan *bar* horizontal dan *bar* vertikal pada dasarnya sama, sesuai dengan tujuan tertentu. Sebagai contoh, jika kita ingin menulis nama masing-masing kategori di dalam persegi panjang yang mewakili kategori tersebut, maka grafik batang horizontal

mungkin lebih sesuai. Jika ingin menekankan ketinggian kolom yang berbeda sebagai ukuran jumlah, kita menggunakan grafik batang vertikal seperti pada Gambar 2.2.



Gambar 2.2 Contoh Diagram Batang Rata-Rata Curah Hujan per Bulan.

2.2 *Extreme Value Theory (EVT)*

EVT merupakan salah satu metode statistika yang digunakan untuk mempelajari bagaimana perilaku ekor (*tail*) distribusi tersebut. Metode ini berfokus pada perilaku ekor (*tail*) suatu distribusi untuk dapat menentukan probabilitas nilai-nilai ekstrimnya. EVT telah lama diterapkan dalam berbagai bidang ilmu, antara lain: hidrologi, klimatologi, dan teori reliabilitas. Kajian mengenai ekor distribusi menunjukkan bahwa sebagian besar data iklim memiliki ekor distribusi yang *heavy-tail*, yaitu ekor distribusi turun secara lambat bila dibandingkan dengan distribusi normal. Dampaknya adalah peluang terjadinya nilai ekstrim akan lebih besar daripada pemodelan dengan distribusi normal. EVT dapat meramalkan terjadinya kejadian ekstrim pada data *heavy-tail* yang tidak dapat dilakukan dengan pendekatan tradisional lainnya. Metode ini mampu menjelaskan kerugian kejadian ekstrim dan jika terjadi mempunyai konsekuensi kerugian yang sangat besar (bernilai maksimum) dan tidak dapat dimodelkan dengan pendekatan biasa.

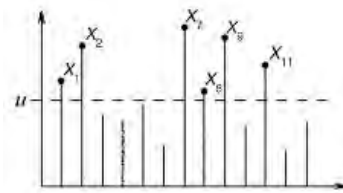
Extreme Value Theory muncul sebagai salah satu aplikasi statistika yang paling penting untuk ilmu terapan selama lebih

dari 50 tahun (Coles, 2001). Dalam disiplin ilmu lainnya, metode ini juga banyak digunakan misalnya untuk penyesuaian portofolio perusahaan asuransi, penilaian resiko di pasar keuangan, dan untuk prediksi lalu lintas telekomunikasi. Beberapa aplikasi pemodelan nilai ekstrim yang telah diterbitkan antara lain pemodelan gelombang laut (Dawson, 2000), strategi manajemen (Dahan dan Mendelson, 2001), pengolahan data biomedis (Roberts, 2000), dan termodinamika gempa bumi (Lavenda dan Cipollone, 2000).

Dalam mengidentifikasi pergerakan nilai ekstrim terdapat dua pendekatan yang digunakan. Pendekatan pertama yaitu dengan metode *Block Maxima* (BM) yaitu mengambil nilai maksimum dalam satu periode sedangkan pendekatan kedua yaitu dengan metode *Peaks Over Threshold* (POT) yaitu mengambil nilai-nilai yang melewati suatu nilai *threshold* (McNeil, 1999 & Gilli dan Kellezi, 2003).

2.3 *Peaks Over Threshold* (POT)

POT merupakan salah satu metode untuk mengidentifikasi nilai ekstrim dengan menggunakan nilai patokan yang disebut *threshold* (u). Data yang berada di atas nilai patokan tersebut akan diidentifikasi sebagai nilai ekstrim.



Gambar 2.3 Ilustrasi Pengambilan Data Sampel dengan Metode POT

Gambar 2.3 menunjukkan cara pengambilan data ekstrim menggunakan POT. Data x_1 , x_2 , x_7 , x_8 , x_9 , dan x_{10} adalah nilai yang berada di atas *threshold* (u), sehingga keenam data yang berada di atas *threshold* merupakan nilai ekstrim yang akan digunakan untuk analisis selanjutnya. Semakin tinggi nilai *threshold* maka

data ekstrim akan semakin mengikuti distribusi General Pareto. Berikut ini adalah *Probability Density Function Generalized Pareto Distribution* (GPD).

$$f(x-u) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}-1} & , \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{(x-u)}{\sigma}\right) & , \xi = 0 \end{cases} \quad (2.3)$$

dimana $0 \leq (x-u) \leq \infty$ jika $\xi \geq 0$ dan $u \leq (x-u) < u - (\sigma/\xi)$ jika $\xi < 0$.

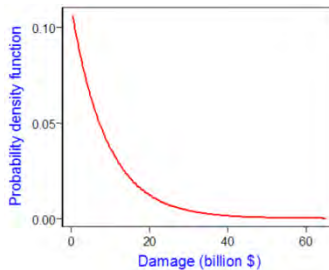
Adapun *Cumulative Distribution Function* (CDF) dari *Generalized Pareto Distribution* (GPD) adalah sebagai berikut:

$$F(x-u) = \begin{cases} 1 - \left(1 + \frac{\xi(x-u)}{\sigma}\right)^{-\frac{1}{\xi}} & , 0 \leq (x-u) < u - \frac{\sigma}{\xi} & \text{jika } \xi < 0 \\ 0 \leq (x-u) < \infty & \text{jika } \xi > 0 \\ 1 - \exp\left(-\frac{(x-u)}{\sigma}\right) & , 0 \leq (x-u) < \infty & \text{jika } \xi = 0 \end{cases} \quad (2.4)$$

dengan σ adalah parameter skala (*scale*) dan ξ adalah parameter bentuk (*shape*)/*tail index*

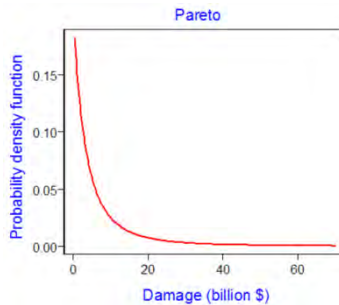
Generalized pareto distribution juga dibedakan menjadi tiga tipe jika dilihat dari nilai parameter bentuk (ξ) yaitu:

1. Tipe 1 berdistribusi Ekspensial jika nilai $\xi = 0$ (Gambar 2.4).



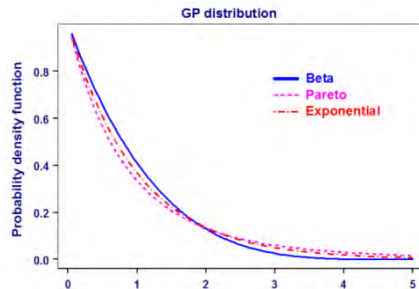
Gambar 2.4 PDF Tipe Distribusi Ekspensial.

2. Tipe 2 berdistribusi Pareto jika nilai $\xi > 0$, seperti pada Gambar 2.5 berikut



Gambar 2.5 PDF Tipe Distribusi Pareto.

3. Tipe 3 berdistribusi Pareto tipe 2/Beta jika nilai $\xi < 0$, seperti pada Gambar 2.6 berikut.



Gambar 2.6 PDF Tipe Distribusi Beta.

Nilai ξ pada GPD juga menjelaskan jika $\xi < 0$ distribusi nilai ekstrim memiliki batas atas tertentu, jika $\xi > 0$ distribusi nilai ekstrim tidak memiliki batas atas, sedangkan jika $\xi = 0$, distribusi nilai ekstrim memiliki batasan yang tidak terbatas (Coles, 2001). Semakin besar nilai ξ , maka distribusi akan memiliki ekor yang semakin berat (*heavy tail*), sehingga akan berdampak peluang terjadinya nilai ekstrim akan semakin besar. Oleh karena itu, diantara tipe GPD yang memiliki ekor yang paling gemuk adalah tipe 2 (distribusi pareto).

2.4 Penentuan Nilai *Threshold*

Untuk mendapatkan nilai ekstrim dengan metode *Peaks Over Threshold*, terlebih dahulu menentukan nilai *threshold* (u) sebagai nilai acuan. Jika suatu nilai observasi melebihi nilai patokan maka nilai tersebut diidentifikasi sebagai nilai ekstrim. Penentuan *threshold* pada umumnya mencari keseimbangan yang optimal agar diperoleh parameter error seminimal mungkin. Penentuan nilai-nilai ekstrim dengan *threshold* (u) sebagai batasannya adalah hal yang sulit seperti halnya dengan penentuan ukuran blok dalam metode *Block Maxima*. Permasalahan tersebut bisa menghasilkan taksiran parameter yang bias dan nilai varians yang besar. Hal ini terjadi jika nilai *threshold* terlalu rendah atau terlalu tinggi. Nilai *threshold* yang terlalu rendah akan menghasilkan parameter yang bias sedangkan nilai *threshold* yang terlalu tinggi maka data ekstrim yang diperoleh akan sedikit sehingga tidak cukup untuk menaksir model, akibatnya menghasilkan varians yang besar (Coles, 2001).

Ada beberapa cara dalam menentukan *threshold* diantaranya adalah dengan metode *Mean Residual Life Plot* (MRLP) dan metode persentase. Metode penentuan nilai *threshold* yang lebih mudah dan sering digunakan adalah metode persentase. Penentuan nilai *threshold* dengan menggunakan metode persentase didapatkan dengan cara sebagai berikut.

1. Mengurutkan data dari yang terbesar hingga yang terkecil.
2. Menghitung 10% dari jumlah data (k)

$$k = 10\% \times N$$

dimana N = jumlah data.

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

Penentuan nilai *threshold* dengan metode presentase diilustrasikan pada Lampiran J.

Berdasarkan Chaves-Dermoulin (2004) dalam Wahyudi (2012) merekomendasikan bahwa data yang berada dalam *threshold* yaitu sekitar 10% dari keseluruhan data yang sudah diurutkan dari yang paling besar sampai terkecil. Hal ini

dikarenakan berdasarkan kajian analisis sensitivitas, menyatakan bahwa apabila ada sedikit pergeseran terhadap *threshold* maka taksiran parameter yang dihasilkan tidak akan terpengaruh oleh pergeseran tersebut.

Selain metode presentase, pemilihan *threshold* juga bisa dilakukan menggunakan metode *Mean Residual Life Plot* (MRLP) yang didasarkan pada nilai rata-rata GPD. Jika X merupakan variabel yang mengikuti GPD dengan parameter σ dan ξ maka

$$E(X) = \frac{\sigma}{1-\xi} \quad (2.5)$$

Jika sebuah model valid untuk u_o maka untuk nilai u yang lebih dari u_o model juga akan valid. Rataan dengan nilai *threshold* (u) dan u_o dapat dinyatakan sebagai berikut :

$$E(X - u_o | X > u_o) = \frac{\sigma_{u_o}}{1-\xi} \quad (2.6)$$

$$E(X - u | X > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_o} + \xi(u - u_o)}{1-\xi}$$

Sehingga $E(X - u | X > u)$ adalah fungsi linear dari u . Berdasarkan uraian ini, maka untuk menentukan nilai *threshold* dapat dilakukan sebagai berikut :

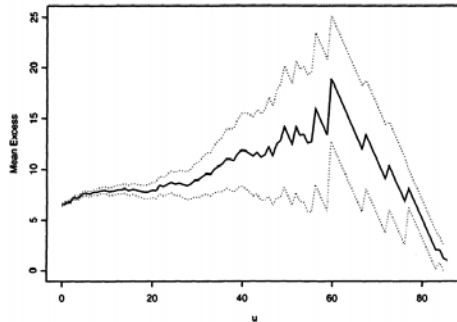
1. Membuat MRLP dengan koordinat titik sebaran dengan persamaan (2.7) sebagai berikut.

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{mak} \right\} \quad (2.7)$$

dengan n_u banyaknya nilai yang melebihi u .

2. Memilih titik *threshold* jika plot mulai membentuk pola linear disekitar nilai u .

Interpretasi *Mean Residual Life Plot* tidak selalu mudah. Gambar 2.7 menunjukkan *Mean Residual Life Plot* dengan *confident interval* 95% pada data curah hujan.



Gambar 2.7 Mean Residual Life Plot untuk data curah hujan di Inggris (Sumber : *An Introduction to Statistical Modeling of Extreme Value*, Coles.S, 2001)

Berdasarkan Gambar 2.7 dari $threshold(u) = 0$ grafik terlihat tidak stabil sampai $u \approx 60$. Setelah titik $u = 60$ grafik mendekati pola linear sehingga diambil nilai $threshold(u) = 60$. Namun hanya ada 6 titik observasi yang melebihi nilai $threshold$ sehingga belum cukup data untuk memodelkan *extreme value*. Oleh karena itu dengan menggeser nilai $threshold(u) = 30$ maka data ekstrim yang diperoleh akan cukup memberikan informasi dalam analisis *extreme value*. Dari Gambar 2.7 juga masih menunjukkan pola linear setelah melewati nilai $threshold(u) = 30$.

2.5 Autocorrelation Function (ACF)

Fungsi Autokorelasi atau *autocorrelation function* (ACF) merupakan suatu hubungan linier antara pengamatan X_t dengan pengamatan X_{t-k} pada data *time series* yang dipisahkan oleh waktu ke- k (Wei, 2004). Rumus umumnya adalah:

$$\rho_k = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{\text{var}(X_t)}\sqrt{\text{var}(X_{t+k})}} \quad (2.8)$$

Dalam analisis *extreme value*, plot *autocorrelation function* (ACF) merupakan salah satu cara yang digunakan untuk mengetahui adanya data ekstrim yang dependen selain dengan melihat plot dari data itu sendiri. Apabila nilai ACF dari suatu

data kurang dari batas bawah atau lebih dari batas atas fungsi *autocorrelation*, maka dikatakan bahwa data bersifat dependen.

2.6 Estimasi Parameter *Generalized Pareto Distribution*

Penaksir parameter metode *Generalized Pareto Distribution* (GPD) dapat ditaksir dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Secara umum GPD memiliki *probability density function* seperti pada Persamaan (2.3). Berdasarkan metode MLE maka fungsi *likelihood*-nya untuk GPD dengan $\xi \neq 0$

$$\begin{aligned}
 L(u, \xi, \sigma | x_1, x_2, \dots, x_n) &= f(x_1)f(x_2)\dots f(x_n) \\
 &= \frac{1}{\sigma} \left(1 + \frac{\xi(x_1 - u)}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \frac{1}{\sigma} \left(1 + \frac{\xi(x_2 - u)}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \dots \frac{1}{\sigma} \left(1 + \frac{\xi(x_n - u)}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \\
 &= \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^n \left(1 + \frac{\xi(x_i - u)}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \\
 &= \sigma^{-n} \prod_{i=1}^n \left(1 + \frac{\xi(x_i - u)}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \tag{2.9}
 \end{aligned}$$

Fungsi *ln likelihood* dari Persamaan 2.9 adalah sebagai berikut

$$\ln L(u, \xi, \sigma | x_1, x_2, \dots, x_n) = -n \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi(x_i - u)}{\sigma}\right) \tag{2.10}$$

Sedangkan untuk GPD dengan $\xi = 0$ fungsi *likelihood*-nya adalah

$$\begin{aligned}
 L(u, \sigma | x_1, x_2, \dots, x_n) &= \prod_{i=1}^n f(x_i, u, \sigma) \\
 &= \prod_{i=1}^n \frac{1}{\sigma} \exp\left(-\frac{(x_i - u)}{\sigma}\right) \\
 &= \left(\frac{1}{\sigma}\right)^n \exp\left(-\frac{1}{\sigma} \sum_{i=1}^n (x_i - u)\right) \tag{2.11}
 \end{aligned}$$

Sehingga fungsi *ln likelihood* dari persamaan 2.11 adalah sebagai berikut

$$\ln L(u, \sigma | x_1, x_2, \dots, x_n) = -n \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^n (x_i - u) \tag{2.12}$$

Selanjutnya dari persamaan \ln *likelihood* yang diperoleh kemudian diturunkan terhadap parameter yang akan ditaksir dan disamakan dengan nol. Berdasarkan persamaan yang terbentuk, diperoleh persamaan yang tidak *close form* maka dibutuhkan analisis numerik lebih lanjut dengan cara iterasi untuk memaksimumkan fungsi \ln *likelihood*.

Salah satu analisis numerik yang digunakan untuk menyelesaikan persamaan yang tidak *close form* adalah metode *Newton Raphson*. Apabila $\mathbf{g}(\boldsymbol{\theta})$ adalah vektor dari turunan pertama dari $L(\xi, \sigma; x)$ dan $\mathbf{H}(\boldsymbol{\theta})$ adalah matriks *Hessian* yaitu matriks dari turunan kedua dari fungsi \ln *likelihood*. Persamaan umum *Newton Raphson* sebagai berikut.

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \mathbf{g}(\boldsymbol{\theta}_t) \mathbf{H}^{-1}(\boldsymbol{\theta}_t) \quad (2.13)$$

Iterasi berhenti apabila $|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t| < \varepsilon$ dimana ε adalah bilangan yang kecil sekali.

2.7 Pemeriksaan Kesesuaian Distribusi

Pemeriksaan kesesuaian distribusi dilakukan untuk menunjukkan adanya kesesuaian distribusi teoritis. Pemeriksaan distribusi dapat dilakukan dengan dua cara yaitu melalui *quantile plot* dan *probability plot* atau dengan uji *Kolmogorov-Smirnov*.

a. *Quantile Plot dan Probability Plot*

Pemeriksaan distribusi dengan *quantile plot* dan *probability plot* dilakukan karena dengan melihat pola sebaran nilai-nilai ekstrim yang mengikuti garis linier. Adapun langkah-langkah membuat *quantile plot* dan *probability plot* sebagai berikut :

1. Mengurutkan data nilai $(y-u)$ dari yang terkecil sampai terbesar $y_{(1)}, y_{(2)}, y_{(3)}, \dots, y_{(k)}$ dan menentukan nilai $p_{(i)}$ untuk setiap urutan $y_{(i)}$ dengan rumus

$$p_{(i)} = \frac{i}{k+1} \quad (2.14)$$

2. Membuat *quantile plot* dengan koordinat titik-titik sebaran $(G^{-1}(p), y_i)$ dimana nilai $G^{-1}(p)$ adalah sebagai berikut

$$G^{-1}(p) = \begin{cases} \frac{\hat{\sigma}}{\xi} \left((1-p)^{-\xi} - 1 \right), & \xi \neq 0 \\ -\hat{\sigma} \ln(1-p), & \xi = 0 \end{cases} \quad (2.15)$$

3. Membuat *probability plot* dengan koordinat titik-titik sebaran ($p_{(i)}$, $G(y)$). Jika *quantile plot* dan *probability plot* mengikuti garis lurus atau linear maka distribusi tersebut sudah sesuai.

b. Uji Kolmogorov-Smirnov

Selain melalui *quantile plot* dan *probability plot*, pemeriksaan distribusi juga dilakukan dengan uji *Kolmogorov-Smirnov*. Pengujian ini dilakukan dengan menyesuaikan fungsi distribusi empiris (berdasarkan sampel) $F_n(x)$ dengan distribusi teoritis tertentu (sesuai yang dihipotesiskan) $F_0(x)$.

Uji Hipotesis :

$H_0 : F_n(x) = F_0(x)$.(Data mengikuti distribusi teoritis $F_0(x)$)

$H_1 : F_n(x) \neq F_0(x)$.(Data tidak mengikuti distribusi teoritis $F_0(x)$)

Statistik Uji:

$$D = \text{Maks} |F_n(x) - F_0(x)| \quad (2.16)$$

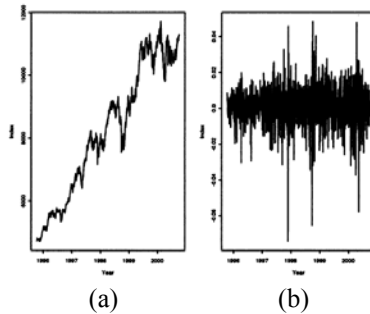
Daerah Kritis :

Tolak H_0 jika $D_{\text{hitung}} > D_\alpha$ pada tabel *Kolmogorov-Smirnov* satu sampel dengan taraf signifikansi (α). $F_n(x)$ 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).

2.8 Extreme of Dependent

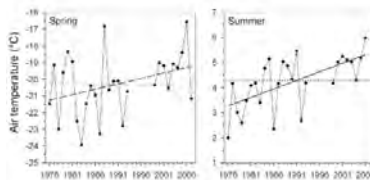
Banyak teori yang diterapkan sejauh ini mengasumsikan data ekstrim bersifat independen dan identik (IID), hal ini terjadi ketika melihat nilai-nilai ekstrim yang cenderung konstan selama beberapa pengamatan atau stasioner (Gilleland dan Katz, 2006). Hal tersebut memungkinkan terdapat parameter dengan variasi fungsi waktu untuk data series yang non stationer. Gambar 2.8

merupakan contoh bentuk data yang *stationary* dan *non stationary*.

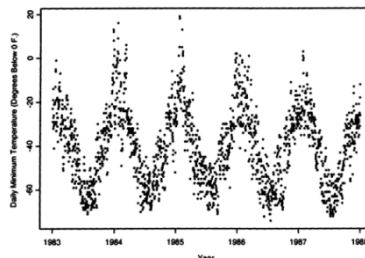


Gambar 2.8 Pola data menurut runtun waktu: Data *Non Stationary* (a) dan Data *Stationary* (b).

Data yang non stationer mengindikasikan jika data tersebut dependen dan membentuk suatu pola tertentu, pola yang terbentuk bisa berupa pola siklik, tren naik, maupun tren turun. Bentuk pola linear trend dan siklik dapat dilihat pada Gambar 2.9 dan 2.10 sebagai berikut



Gambar 2.9 Pola Linier Trend



Gambar 2.10 Pola Data Temperatur Siklik (Sumber :*An Introduction to Statistical Modeling of Extreme Value*, Coles.S, 2001)

Ada beberapa cara yang digunakan untuk mengatasi data ekstrim yang dependen (non-stasioner) antara lain dengan melakukan transformasi pada data dan kemudian menentukan model parameter distribusinya, serta cara yang kedua yaitu dengan melakukan *declustering* pada data.

1. Model Parameter

Terdapat dua model untuk parameter $\mu(t)$ dan $\sigma(t)$ baik yang linear maupun untuk siklus *trend* dengan koefisien μ dan σ . Model untuk series *trend* mengikuti deret Fourier yang pertama, sehingga didapatkan model sebagai berikut :

1. $\mu(t) = \mu_0 + \mu_1 \frac{t}{n+1}$ dan $\sigma(t) = 1$
2. $\mu(t) = 0$ dan $\log \sigma(t) = \sigma_0 + \sigma_1 \frac{t}{n+1}$
3. $\mu(t) = \mu_0 + \mu_1 \cos\left(\frac{2\pi t}{N}\right) - \mu_2 \sin\left(\frac{2\pi t}{N}\right)$ dan $\sigma(t) = 1$
4. $\mu(t) = 0$ dan $\log \sigma(t) = \sigma_0 + \sigma_1 \cos\left(\frac{2\pi t}{N}\right) - \sigma_2 \sin\left(\frac{2\pi t}{N}\right)$

dengan:

n : banyaknya observasi atau data

N : periode observasi di setiap siklus yang dihasilkan oleh deret Fourier.

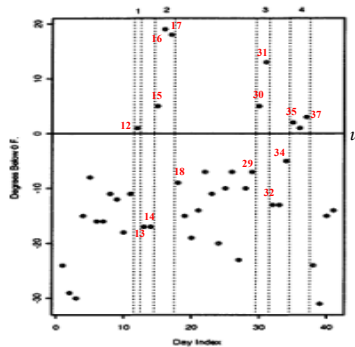
2. Metode *Declustering*

Kejadian-kejadian di alam yang bersifat ekstrim cenderung terjadi secara berurutan (dependen). Hal tersebut merupakan sebuah fakta yang melanggar asumsi utama dalam EVT, yaitu bahwa variabel yang diteliti harus independen. Untuk tujuan tersebut, maka dalam analisis EVT dilakukan identifikasi kelompok independen atau yang sering disebut dengan *decluster* data ekstrim dependen. Metode yang paling banyak digunakan untuk menangani masalah ini adalah *declustering*, yang memfilter pengamatan dependen untuk mendapatkan satu set *threshold* yang independen. Sejumlah teknik *declustering* telah dikembangkan, yaitu prosedur yang mengarah pada identifikasi *cluster*

maksima, yaitu observasi maksimum (puncak) pada masing-masing *cluster* yang terbentuk dan diasumsikan independen. Salah satu pendekatan metode *declustering* adalah teknik *run declustering*. Soukissian dan Arapi (2011) melakukan metode *declustering* pada data tinggi gelombang laut dengan 4 pendekatan yaitu *Standard Storm Length (SSL)*, *Runs declustering (RL)*, *DeClustering Algorithm (DeCA)*, dan *Triple Annual Maximum (TAM)*. Kesimpulan yang diperoleh dari penelitian tersebut yaitu dengan metode *declustering* dimana pendekatan *Runs declustering (RL)* menunjukkan hasil dengan standar eror yang lebih kecil dan selang kepercayaan yang lebih sempit.

Secara garis besar, prosedur dalam melakukan teknik *run-declustering* adalah sebagai berikut.

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



Gambar 2.11 Ilustrasi Metode *Run-declustering* pada Data Temperatur dengan nilai $r=1$ (Sumber :*An Introduction to Statistical Modeling of Extreme Value, Coles.S, 2001*)

Berdasarkan ilustrasi metode *run-declustering* ($r=1$) pada Gambar 2.11 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 (obs ke 15) dan berhenti pada saat terdapat paling sedikit 1 observasi yang berada di bawah *threshold* (obs ke 18), sehingga anggota *cluster* 2 terdiri atas pengamatan ke-15, 16, dan 17. Proses ini terus berlanjut sampai pembentukan semua *cluster* selesai dan data telah habis. Berdasarkan lustrasi pada Gambar 2.11 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 akan 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 varians tinggi. Dengan demikian, beberapa metode harus diterapkan untuk beberapa pilihan u dan r dan sensitivitas hasil harus diteliti dengan seksama.

2.9 *Extremal Index*

Pendekatan distribusi asimtotik pada data yang melebihi nilai *threshold* mengasumsikan bahwa data bersifat independen identik (iid), namun pada kenyataannya asumsi tersebut tidak sesuai dengan data-data dalam kasus riil. Oleh karena itu untuk mengatasi masalah dependensi data, cara yang sering digunakan adalah dengan *declustering*. Menurut Coles (2001), *declustering* yaitu melakukan *decluster* data ekstrim dependen untuk mendapatkan satu set data yang independen. Metode ini bekerja dengan menentukan suatu *threshold* (u) kemudian pembentukan *cluster* hingga sejumlah r pengamatan berada di bawah *threshold*.

Seperti yang disebutkan oleh Coles (2001), pemilihan nilai r yang tidak sesuai akan menyebabkan bias dan varians yang tinggi. Nilai r yang terlalu rendah akan menyebabkan data sulit menjadi independen sedangkan nilai r yang terlalu tinggi maka data ekstrim yang diperoleh akan sedikit sehingga menghasilkan varians yang besar. Untuk menentukan estimasi r (*run length*) yang optimal, dapat menggunakan *extremal index* (Ferro dan Segers, 2003). *Extremal index* merupakan suatu indikator untuk mengukur berapa banyak *cluster* yang terbentuk sehingga data menjadi independen.

Data yang independen identik (iid) berlaku bahwa $\Pr\{M_n \leq u_n\} = F(u_n)^n$, sedangkan pada data yang dependen hal tersebut masih tetap berlaku dengan $\Pr\{M_n \leq u_n\} = F(u_n)^{\theta n}$, dimana $\theta \in (0, 1]$ disebut *extremal index*. Jika data independen, maka $\theta=1$; namun hal ini tidak berlaku untuk semua data. Ada kemungkinan juga ditemukan data yang dependen namun memiliki *extremal index* $\theta = 1$ (Coles, 2001). Salah satu metode dalam mengestimasi *extremal index* adalah dengan menggunakan persamaan 2.17. (Ferro dan Segers, 2003).

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

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

T_i : jarak antar pengamatan di atas *threshold*

N : jumlah pengamatan di atas *threshold*

S_i : urutan pengamatan ke- i diatas *threshold*

Nilai *extremal index* (θ) dapat digunakan untuk menentukan *run length* (r) yang optimal untuk proses *declustering*, dimana $r = T_{(C)}$ dan $(C-1) = \theta (N-1)$. Ilustrasi penentuan nilai r dengan *extremal index* selengkapnya disajikan pada Lampiran K.

2.10 Return Level

Hal yang menarik dalam kajian *extreme value theory* dalam bidang klimatologi selain untuk menaksir parameter adalah menentukan *return level*. Nilai maksimum pada periode yang akan datang disebut *return level*. Nilai *return level* dapat memberikan informasi yang sangat penting dalam bidang ekonomi, keuangan, dan asuransi, serta pertanian. Pada bidang asuransi, *return level* digunakan untuk mengukur seberapa besar resiko yang ditanggung perusahaan asuransi dari klaim yang dianggap ekstrim atau melebihi rata-rata klaim lainnya sebagai salah satu upaya untuk menjaga portofolio perusahaannya. Dalam bidang ekonomi dan keuangan, nilai *return level* dapat mengetahui tingkat resiko saham yang diperlukan oleh para investor untuk menentukan saat yang tepat untuk berinvestasi. Sedangkan pada bidang pertanian nilai *return level* akan memberikan informasi waktu tanam yang sesuai untuk varietas tanaman pangan tertentu sehingga dapat ditentukan varietas tanaman yang sesuai dan adaptif terhadap keadaan yang ekstrim tersebut. Penentuan *return level* pada GPD melibatkan parameter ξ , σ , dan nilai *threshold* (u). Persamaan *return level* untuk GPD adalah sebagai berikut.

$$x_m = u + \frac{\hat{\sigma}}{\xi} \left((m \zeta_u \theta)^\xi - 1 \right) \quad (2.18)$$

dengan :

u : nilai *threshold*

x_m : nilai ekstrim yang terjadi pada jangka waktu m -pengamatan

$$\zeta_u : \frac{n_u}{n} \quad \hat{\theta} : \frac{n_c}{n_u}$$

ξ : parameter bentuk (*shape*)

$\hat{\sigma}$: taksiran parameter skala (*scale*)

Nilai x_m atau nilai ekstrim yang terjadi satu kali pada jangka waktu m pengamatan. Nilai $\zeta_u \theta$ dapat ditaksir dengan $\hat{\zeta}_u \hat{\theta} = n_c/n$ dengan n_c adalah banyaknya cluster yang terbentuk,

n_u adalah banyaknya data yang melebihi *threshold* dan n adalah banyaknya data (Coles, 2001).

2.11 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) digunakan untuk mengetahui akar kesalahan rata-rata kuadrat dari setiap metode. Pada penelitian ini RMSE digunakan sebagai kriteria pemilihan metode yang sesuai.

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (x_i - \hat{x}_i)^2} \quad (2.19)$$

dengan :

- x_i : Nilai aktual
- \hat{x}_i : Nilai dugaan
- t : Banyaknya nilai yang diduga

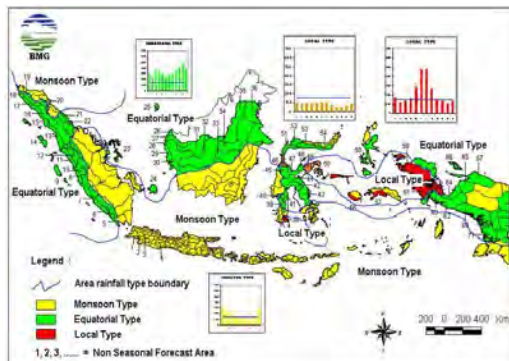
2.12 Curah Hujan

Curah hujan menunjukkan jumlah air hujan yang turun pada suatu daerah dalam waktu tertentu. Curah hujan diukur dalam jumlah harian, bulanan, dan tahunan. Menurut BMKG berdasarkan distribusi data rata-rata curah hujan bulanan, umumnya wilayah Indonesia dibagi menjadi 3 (tiga) pola hujan, yaitu :

1. Pola hujan monsun, yang wilayahnya memiliki perbedaan yang jelas antara periode musim hujan dan periode musim kemarau kemudian dikelompokkan dalam Zona Musim (ZOM), tipe curah hujan yang bersifat unimodial (satu puncak musim hujan). Zona Musim (ZOM) di bagi menjadi 4 yaitu
 - a. Bulan Desember, Januari, dan Februari (DJF) terjadi musim hujan.
 - b. Bulan Maret, April, dan Mei (MAM) pada bulan tersebut terjadi transisi dari musim hujan menuju musim kemarau
 - c. Bulan Juni, Juli, dan Agustus (JJA) musim kemarau

- d. Bulan September, Oktober, dan November (SON) transisi dari musim kemarau ke musim hujan.
2. Pola hujan equatorial, yang wilayahnya memiliki distribusi hujan bulanan bimodial dengan dua puncak musim hujan maksimum dan hampir sepanjang tahun masuk dalam kriteria musim hujan. Pola ekuatorial dicirikan oleh tipe curah hujan dengan bentuk bimodial (dua puncak hujan) yang biasanya terjadi sekitar bulan Maret dan Oktober atau pada saat terjadi ekinoks.
 3. Pola hujan lokal, yang wilayahnya memiliki distribusi hujan bulanan kebalikan dengan pola monsun. Pola lokal dicirikan oleh bentuk pola hujan unimodial (satu puncak hujan), tetapi bentuknya berlawanan dengan tipe hujan monsun.

Pada kondisi normal, daerah yang bertipe hujan monsun akan mendapatkan jumlah curah hujan yang berlebih pada saat monsun barat (DJF) dibanding saat monsun timur (JJA). Pengaruh monsun di daerah yang memiliki pola curah hujan ekuator kurang tegas akibat pengaruh insolasi pada saat terjadi ekinoks, demikian juga pada daerah yang memiliki pola curah hujan lokal yang lebih dipengaruhi oleh efek orografi. Gambar 2.12 berikut merupakan pola curah hujan dari BMKG



Gambar 2.12 Sebaran Pola Curah Hujan di Indonesia
(Sumber: BMKG, 2008)

Rata-rata curah hujan di Indonesia untuk setiap tahunnya tidak sama. Namun masih tergolong cukup banyak, yaitu rata-rata 2000-3000 mm/tahun. Begitu pula antara tempat yang satu dengan tempat yang lain rata-rata curah hujannya tidak sama. Ada beberapa daerah yang mendapat curah hujan sangat rendah dan ada pula daerah yang mendapat curah hujan tinggi:

1. Daerah yang mendapat curah hujan rata-rata per tahun kurang dari 1000 mm, meliputi 0,6% dari luas wilayah Indonesia, di antaranya Nusa Tenggara, dan 2 daerah di Sulawesi (Lembah Palu dan Luwuk).
2. Daerah yang mendapat curah hujan antara 1000-2000 mm per tahun di antaranya sebagian Nusa Tenggara, daerah sempit di Merauke, Kepulauan Aru, dan Tanibar.
3. Daerah yang mendapat curah hujan antara 2000-3000 mm per tahun, meliputi Sumatera Timur, Kalimantan Selatan, dan Kalimantan Timur, sebagian besar Jawa Barat dan Jawa Tengah, sebagian Irian Jaya, Kepulauan Maluku dan sebagian besar Sulawesi.

Daerah yang mendapat curah hujan tertinggi lebih dari 3000 mm per tahun meliputi dataran tinggi di Sumatera Barat, Kalimantan Tengah, dataran tinggi Irian bagian tengah, dan beberapa daerah di Jawa, Bali, Lombok, dan Sumba.

2.13 Penelitian Sebelumnya

Beberapa penelitian sebelumnya mengenai *Extreme Value Theory* dengan studi kasus data ekstrim dependen antara lain: Lopes, Nascimento, dan Gamerman (2006) yang menganalisis kejadian ekstrim dependen dengan memodelkan parameter GPD berupa fungsi yang dependen terhadap waktu. Penelitian tersebut menggunakan metode *Dynamic Linear Model* (DLM) dengan pendekatan algoritma *Markov Chain Monte Carlo* (MCMC) untuk mendapatkan model parameter. Pada mulanya analisis dilakukan pada data simulasi yang kemudian diaplikasikan pada tiga jenis data *financial time series*. Jaruskova dan Hanek (2006) melakukan penelitian dengan membandingkan metode POT dan

BM untuk mengestimasi *return level* pada data curah hujan ekstrim di Moravia Utara, Republik Ceko. Pada penelitian tersebut diidentifikasi bahwa data curah hujan yang dianalisis memiliki kecenderungan bersifat dependen sehingga dianalisis dengan metode *declustering* terlebih dahulu. Penelitian lain juga dilakukan oleh Soukissian dan Arapi (2011) yang melakukan perbandingan metode *Block Maxima* dengan *declustering* pada data tinggi gelombang laut. Metode *declustering* yang digunakan pada penelitian tersebut dengan 4 pendekatan yaitu *Standard Storm Length* (SSL), *Runs declustering* (RL), *DeClustering Algorithm* (DeCA), dan *Triple Annual Maximum* (TAM).

Kajian terkait data ekstrim dependen di Indonesia dengan pendekatan *Extreme Value Theory* juga telah dilakukan Kurniawati (2013) yang melakukan analisis data ekstrim dependen pada kasus curah hujan di Jawa Timur dengan pendekatan *Pekas Over Threshold*. Penelitian tersebut melakukan pemodelan untuk parameter skala (*scale*) pada *Generalized Pareto Distribution* serta menentukan *return level* curah hujan di Jawa Timur. Model parameter GPD diperoleh dengan melakukan transformasi trigonometri pada data ekstrim dependen sesuai dengan pola data yang memiliki pola siklik.

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Data yang diperoleh adalah data curah hujan harian di beberapa kabupaten di Jawa Timur yaitu di Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi pada periode 1981 sampai 2010.

3.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian adalah curah hujan harian yang diambil dari satu pos tertentu dimasing-masing kabupaten yaitu Pos Karang Kedawung di Kabupaten Jember, Pos Cawak di Kabupaten Bojonegoro, Pos Karangbinangun di Kabupaten Lamongan, Pos Pasewaran di Kabupaten Banyuwangi, dan Pos Ngale di Kabupaten Ngawi pada periode 1981 sampai 2010. Pemilihan pos tersebut didasarkan pada penelitian Nihayatin (2013) yang melakukan perbandingan uji homogenitas pada data curah hujan di Jawa Timur sebagai pra-pemrosesan kajian perubahan iklim. Pos hujan yang dipilih tersebut telah memenuhi homogenitas runtun waktu.

3.3 Metode Analisis Data

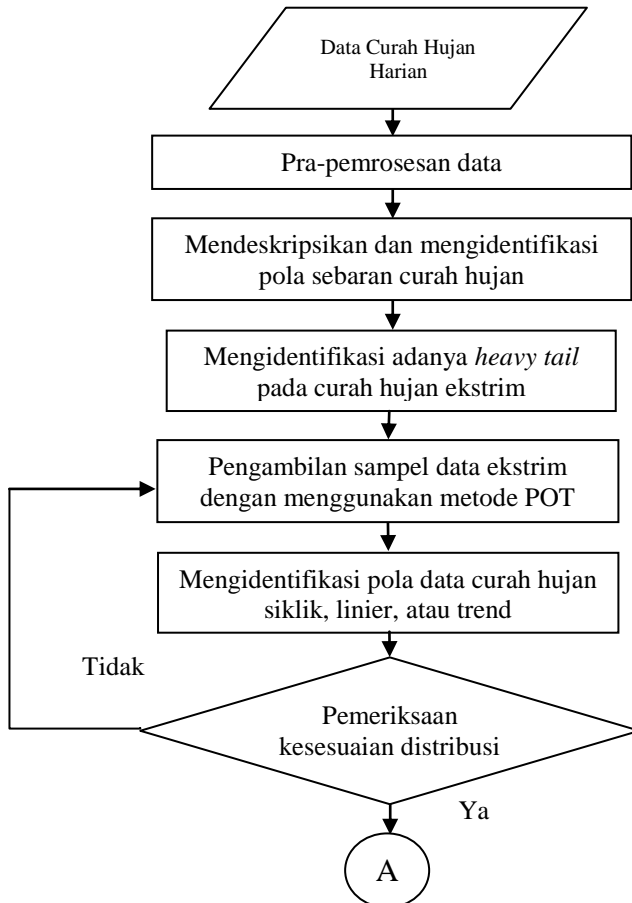
Metode analisis yang digunakan dalam penelitian sebagai berikut.

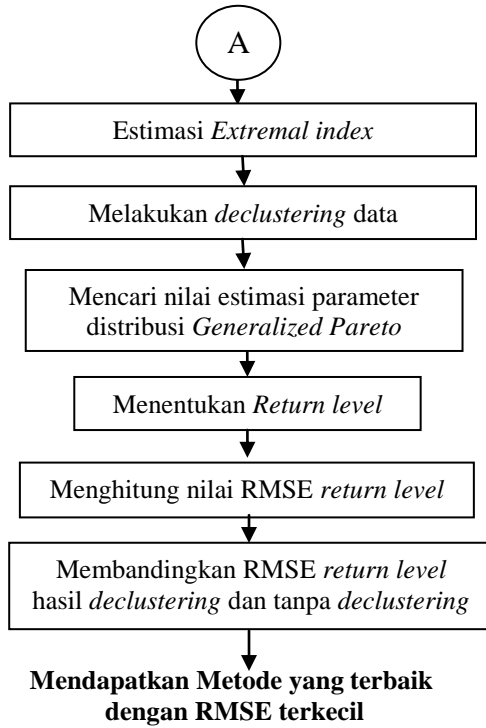
1. Melakukan pra-pemrosesan data dengan mengidentifikasi *missing value*, observasi pencilan (*outlier*), dan observasi yang tidak sesuai dari data curah hujan di Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi.
2. Mengidentifikasi karakteristik curah hujan ekstrim di Jawa Timur khususnya Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi dengan langkah sebagai berikut.

- a. Mendeskripsikan data curah hujan dengan statistika deskriptif dan pola sebaran curah hujan.
- b. Mengidentifikasi distribusi data curah hujan di masing masing kabupaten untuk mengetahui adanya distribusi data *heavy tail* dan nilai ekstrim dengan histogram dan *normality plot*.
3. Pengambilan sampel data ekstrim dengan metode *Peaks Over Threshold*. Sampel data ekstrim didapatkan dengan cara menentukan *threshold* pada data dengan menggunakan metode presentase 10% untuk mendapatkan data ekstrim. Data diurutkan dari yang terbesar hingga yang terkecil.
4. Mengidentifikasi data curah hujan pada masing-masing pos membentuk pola siklik atau linier tren melalui plot data itu sendiri. Selain itu juga melihat data ekstrim dependen melalui plot *autocorrelation function* (ACF).
5. Pemeriksaan kesesuaian distribusi menggunakan *quantile plot*, *probability plot* dan pengujian hipotesis dengan uji *Kolmogorov-Smirnov*. Jika data tidak memenuhi kesesuaian distribusi *Generalized Pareto Distribution* maka kembali ke langkah-3 dengan melakukan pengambilan sampel ulang pada stasiun pengamatan curah hujan yang berbeda di kabupaten tertentu.
6. Melakukan *declustering* pada data ekstrim dependen melalui pendekatan *run-declustering*.
 - a. Menentukan r (*run length*) yang optimum melalui estimasi *extremal index* pada persamaan (2.17).
 - b. Melakukan *declustering* dengan menggunakan nilai r yang telah diperoleh dari hasil estimasi *extremal index* dengan langkah-langkah sebagai berikut.
 - Menggunakan nilai *threshold* (u) pada langkah ke-3 untuk mengidentifikasi nilai-nilai ekstrim
 - Menentukan *cluster*. *Cluster* dimulai ketika *threshold* (u) terlampaui untuk pertama kalinya dan berakhir setelah paling sedikit r pengamatan berturut-turut jatuh di bawah *threshold*.

- Mengekstrak nilai maksimum masing-masing *cluster* dan selanjutnya digunakan sebagai data baru untuk mengestimasi parameter GPD.
- 7. Menghitung nilai estimasi parameter *Generalized Pareto Distribution* (GPD) dari data hasil *declustering* dan tanpa *declustering*.
- 8. Menghitung nilai *return level* curah hujan ekstrim di lima kabupaten kemudian membandingkan nilai RMSE untuk *return level* hasil *declustering* dan tanpa *declustering*.

Tahapan analisis data selengkapnya disajikan pada Gambar 3.1.





Gambar 3.1 Diagram Alir Analisis Data

BAB IV

ANALISIS DAN PEMBAHASAN

Pada bagian ini dilakukan analisis data curah hujan ekstrim dengan menggunakan metode *declustering* di lima kabupaten di Jawa Timur yaitu: Kabupaten Bojonegoro, Banyuwangi, Jember, Lamongan, dan Ngawi. Pembahasan diawali dengan pra-pemrosesan data curah hujan, deskripsi curah hujan, kemudian dibahas pula pengambilan sampel ekstrim dengan metode *peaks over threshold*, pemeriksaan kesesuaian distribusi, dan proses *declustering*. Di bagian akhir dibahas estimasi parameter serta menentukan *return level* pada masing-masing pos pengamatan.

4.1 Pra-Pemrosesan Data Curah Hujan

Langkah awal dalam analisis data yang berkaitan dengan iklim seperti curah hujan adalah penyiapan data. Dalam pencatatan data curah hujan di suatu lokasi terkadang terjadi perpindahan alat pengukur curah hujan, pergantian alat, dan sebagainya, sehingga banyak stasiun pencatat curah hujan yang memiliki data *missing* cukup banyak dan seringkali ditemukan data yang tidak sesuai termasuk di stasiun pengamatan curah hujan dalam penelitian ini. Oleh karena itu perlu dilakukan pra-pemrosesan data sebelum dianalisis meliputi identifikasi *missing value* dan imputasi data.

Missing value atau *missing data* merupakan gangguan yang biasa ditemukan dalam data yang akan dianalisis. Berikut ini hasil identifikasi *missing value* dari data curah hujan di lima pos pengamatan di Jawa Timur.

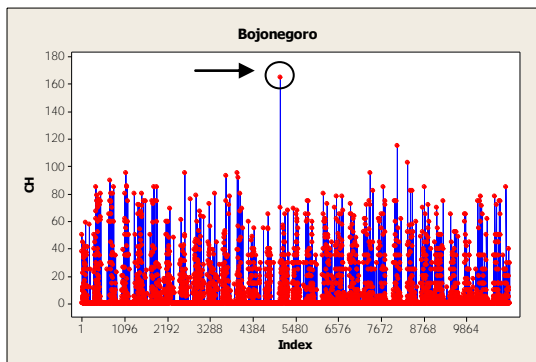
Tabel 4.1 Identifikasi *Missing Value* Data Curah Hujan

	Pasewaran	Karang Kedawung	Ngale	Cawak	Karang- binangun
N tidak sesuai	203	203	203	203	203
% <i>missing</i>	0,84	0,00	5,16	0,28	0,54

Tabel 4.1 menunjukkan bahwa ditemukan banyak data *missing* di pos pengamatan Pasewaran, Ngale, Cawak, dan

Karangbinangun, dimana Pos Ngale di Kabupaten Ngawi memiliki data *missing* yang paling banyak yaitu sebanyak 565 data *missing* atau sekitar 5,16%. Sedangkan di Pos Karang Kedawung tidak ditemukan data *missing* sama sekali, namun masih terdapat beberapa pengamatan yang tidak sesuai yaitu sebanyak 203 pengamatan. Jumlah yang sama juga ditemukan di pos pengamatan yang lain. Pengamatan tidak sesuai yang dimaksud adalah pengamatan yang dicatat pada tanggal-tanggal yang tidak mungkin terjadi, yaitu tanggal 31 pada bulan Februari, April, Juni, September, November, dan tanggal 29 Februari pada tahun-tahun non-kabisat.

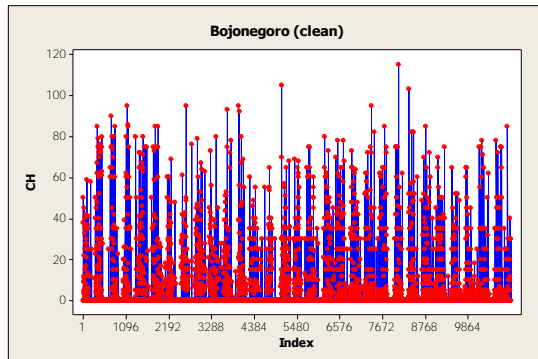
Penanganan yang dilakukan pada pengamatan tidak sesuai adalah dengan menghapus pengamatan. Sedangkan untuk mengatasi *missing value* dilakukan proses imputasi data dengan menggantikan data *missing* dengan menggunakan rata-rata curah hujan pada tanggal dan bulan yang sama dengan tanggal dan bulan data *missing* tersebut. Sebelum dilakukan analisis lebih lanjut, maka data yang sudah bersih dari pengamatan tidak sesuai dan *missing value* dicek ulang untuk mengidentifikasi pengamatan yang aneh atau menyimpang. Berikut ini disajikan plot data curah hujan di pos pengamatan Cawak Kabupaten Bojonegoro.



Gambar 4.1 Plot Data Curah Hujan di Pos Cawak.

(Catatan: Yang diberikan tanda merupakan data pencilon yang perlu diklarifikasi ulang ke pengumpul data)

Gambar 4.1 menunjukkan bahwa terdapat satu pengamatan yang aneh di Pos Cawak, yaitu data curah hujan tercatat 165 mm. Pengamatan ini terjadi pada tanggal 11 Desember 1994. Pada hari-hari sebelum dan setelah tanggal 11 di bulan dan tahun yang sama, rata-rata curah hujan tercatat sekitar 10-20 mm (hujan sedang) bahkan banyak hari tidak terjadi hujan (curah hujan 0 mm), sehingga pengamatan pada tanggal 11 Desember 1994 ini dianggap aneh karena nilai curah hujannya terlalu tinggi bahkan tergolong kategori hujan yang sangat lebat (>100 mm) yang dapat menyebabkan banjir di wilayah tersebut. Oleh karena itu penanganan yang dilakukan untuk mengatasi kasus tersebut adalah dengan melakukan penyesuaian data.



Gambar 4.2 Plot Data Curah Hujan di Pos Cawak yang sudah dibersihkan dari pengamatan aneh (*clean*).

Gambar 4.2 menunjukkan plot data curah hujan di Pos Cawak yang sudah dilakukan penyesuaian. Berdasarkan Gambar 4.2 sudah tidak ditemukan data yang tampak tidak sesuai sehingga dapat dilanjutkan analisis selanjutnya. Identifikasi data yang kurang sesuai di empat pos lainnya dapat dilihat di Lampiran B. Penanganan yang sama juga dilakukan di empat pos pengamatan lainnya untuk pengamatan-pengamatan curah hujan yang dianggap tidak sesuai atau kurang realistis.

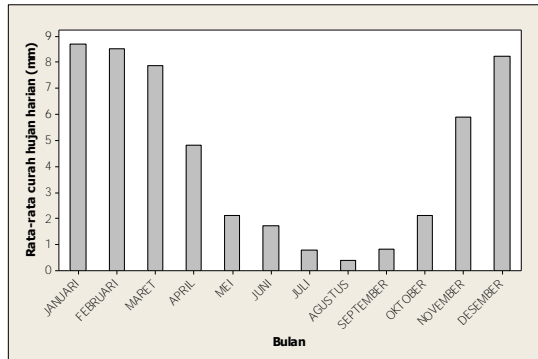
4.2 Deskripsi Curah Hujan di Lima Pos Pengamatan

Analisis deskriptif dilakukan untuk mendapatkan gambaran secara umum data sebagai informasi awal karakteristik pola data curah hujan yang digunakan. Hasil statistik deskriptif curah hujan di lima pos pengamatan disajikan pada Tabel 4.2.

Tabel 4.2 Nilai Rata-Rata, Standard Deviasi, Nilai Minimum, dan Nilai Maksimum Curah Hujan Harian (mm)

Pos Pengamatan	Rata-Rata	Std Deviasi	Minimal	Maksimal
Cawak	4,315	12,398	0	115
Pasewaran	6,532	16,742	0	153
Karang Kedawung	5,149	12,756	0	119
Karangbinangun	4,152	11,261	0	124
Ngale	5,352	13,199	0	127

Tabel 4.2 menunjukkan bahwa rata-rata curah hujan harian di lima pos pengamatan yang tertinggi ada di Pos Pasewaran Kabupaten Banyuwangi yaitu sebesar 6,532 mm/hari, sedangkan rata-rata curah hujan harian terendah tercatat di Pos Karangbinangun Kabupaten Lamongan yaitu sebesar 4,152 mm/hari. Rata-rata curah hujan harian di Pos Ngale, Pos Cawak, dan Pos Karang Kedawung berturut-turut adalah 5,352 mm/hari, 4,315 mm/hari, dan 5,149 mm/hari. Selain rata-rata, Tabel 4.2 juga menunjukkan nilai standard deviasi curah hujan di masing-masing pos pengamatan dimana Pos Pasewaran Kabupaten Banyuwangi memiliki keragaman curah hujan yang paling tinggi yaitu sebesar 16,742 dan Pos Karangbinangun Kabupaten Lamongan memiliki keragaman curah hujan yang paling kecil sebesar 11,261. Nilai standard deviasi dari ketiga pos lainnya yaitu Pos Cawak, Pos Ngale, dan Pos Karang Kedawung berturut-turut yaitu 12,398; 13,199; dan 12,756. Sementara untuk nilai curah hujan minimum pada masing-masing pos adalah sama yaitu 0 mm/hari. Sedangkan curah hujan maksimum yang tercatat di Pos Ngale 127 mm/hari, Pos Cawak 115 mm/hari, Pos Karangbinangun 124 mm/hari, Pos Karang Kedawung 119 mm/hari, dan Pos Pasewaran 153 mm/hari.



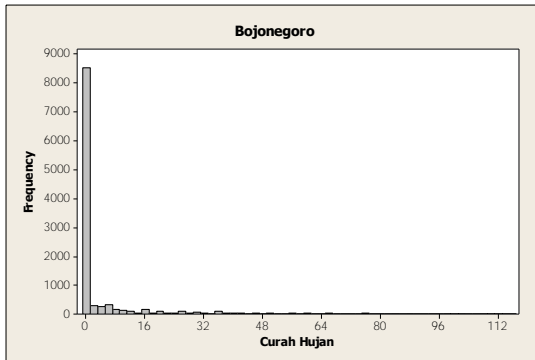
Gambar 4.3 Pola Curah Hujan Harian di Pos Cawak

Perlu dilakukan identifikasi untuk mengetahui pola curah hujan di lima pos pengamatan yang diteliti. Gambar 4.3 merupakan diagram batang yang terbentuk dari rata-rata curah hujan harian mulai bulan Januari 1981 sampai Desember 2010 di Pos Cawak Kabupaten Bojonegoro. Berdasarkan Gambar 4.3 dapat diketahui bahwa pola curah hujan di Pos Cawak membentuk huruf U dan memiliki satu puncak musim hujan (unimodal) yang terjadi pada bulan Januari sehingga membentuk pola monsun. Selain itu, Gambar 4.3 juga menunjukkan puncak musim kemarau terjadi pada bulan Agustus karena jumlah curah hujannya paling rendah diantara bulan-bulan lainnya.

Pola curah hujan harian keempat pos lainnya juga memiliki pola curah hujan harian yang sama yaitu pola monsun dan disajikan pada Lampiran C. Puncak musim hujan pada Pos Pasewaran terjadi pada bulan Januari dengan rata-rata curah hujan sebesar 17,513 mm/hari. Sedangkan Pos Ngale, Pos Karang Kedawung, dan Pos Karangbinangun memiliki satu puncak musim hujan yang sama pada bulan Februari dengan rata-rata curah hujan sebesar 10,493 mm/hari untuk Pos Ngale, 10,002 mm/hari untuk Pos Karang Kedawung, dan 8,593 mm/hari untuk Pos Karangbinangun.

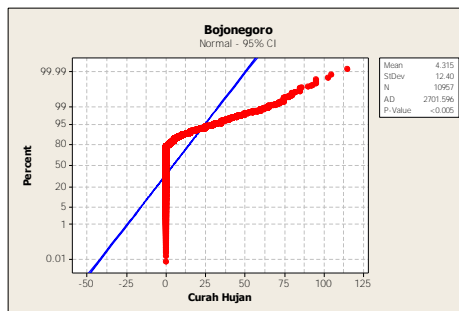
4.3 Identifikasi Pola Data *Heavy Tail* dan Nilai Ekstrim

Identifikasi data berpola *heavy tail* dapat diketahui melalui histogram seperti disajikan pada Gambar 4.4.



Gambar 4.4 Histogram Curah Hujan Harian di Pos Cawak

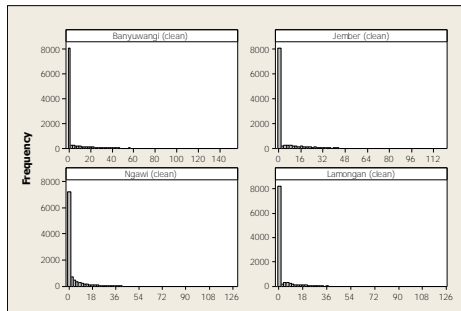
Histogram pada Gambar 4.4 menunjukkan bahwa curah hujan di Pos Cawak memiliki ekor distribusi turun lambat. Hal ini mengindikasikan adanya pola distribusi data *heavy tail* pada data curah hujan harian di Pos Cawak. Adanya indikasi ini juga menunjukkan jika data curah hujan tidak berdistribusi normal sehingga menggunakan metode *extreme value theory*. Hal ini dapat dilihat juga dengan menggunakan *normality probability plot* pada Gambar 4.5 berikut.



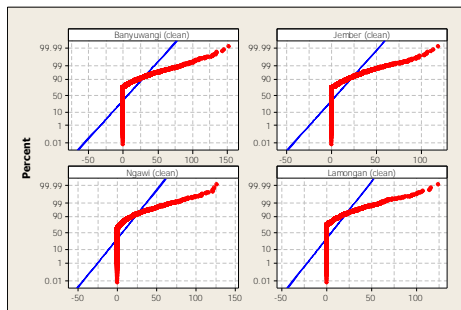
Gambar 4.5 Pengujian *Normality Probability Plot* Curah Hujan Harian di Pos Cawak

Gambar 4.5 menunjukkan bahwa sebagian besar sebaran titik-titik (merah) tidak mengikuti garis linear (biru) sehingga curah hujan tidak berdistribusi normal. Dari hasil histogram dan *normality probability* plot menunjukkan data curah hujan harian yang tercatat di Pos Cawak berpola *heavy tail* (mengandung nilai ekstrim) dan tidak mengikuti distribusi normal.

Histogram dan *normality probability* plot keempat pos pengamatan yang lain disajikan pada Gambar 4.6 dan 4.7.



Gambar 4.6 Histogram Curah Hujan di Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun.



Gambar 4.7 *Normality Probability* Plot Curah Hujan di Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun

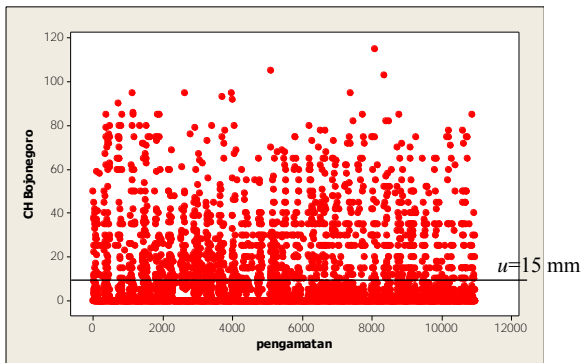
Histogram pada Gambar 4.6 menunjukkan bahwa curah hujan harian di empat pos lainnya terlihat ekor distribusinya turun secara lambat sehingga mengindikasikan adanya pola distribusi data *heavy tail*. Sementara pada Gambar 4.7 menunjukkan bahwa

dari keempat pos pengamatan menunjukkan sebaran data curah hujan yang tidak mengikuti distribusi normal, sebagian besar sebaran titik merah tidak mengikuti garis linear.

4.4 Pengambilan Sampel Ekstrim dengan *Peaks Over Threshold*

Pengambilan sampel ekstrim dengan metode *peaks over threshold* dilakukan dengan menentukan nilai *threshold* terlebih dahulu. Pada penelitian ini pemilihan *threshold* dilakukan dengan metode presentase yaitu mengambil batas dari 10% dari keseluruhan data yang sudah diurutkan dari nilai terbesar sampai terkecil.

Pemilihan *threshold* dengan persentase 10% pada data curah hujan harian di Pos Cawak Kabupaten Bojonegoro didapatkan nilai *threshold* sebesar 15 mm dan pengambilan sampel ekstrim dilakukan dengan mengambil data yang hanya melebihi nilai *threshold* yang disajikan pada Gambar 4.8. Proses penentuan nilai *threshold* dan pengambilan sampel ekstrim di Pos Cawak secara lebih jelas disajikan pada Lampiran J.



Gambar 4.8 Pengambilan Sampel Ekstrim di Pos Cawak

Pengambilan sampel ekstrim di empat pos lainnya disajikan dalam Lampiran D. Nilai *threshold* dan banyaknya data yang berada di atas *threshold* pada masing-masing pos pengamatan disajikan dalam Tabel 4.3.

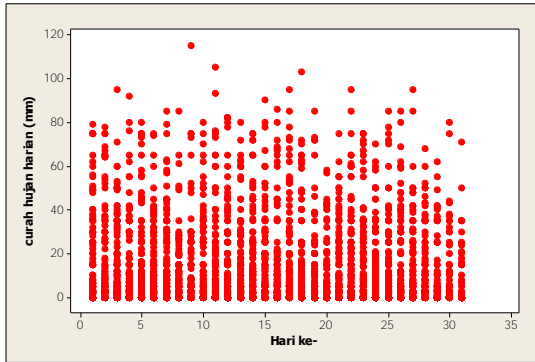
Tabel 4.3 Jumlah Pengamatan (N), Nilai *Threshold* (u), dan Jumlah Sampel Ekstrim (n_u)

Pos Pengamatan	N	u (mm)	n_u
Cawak	10.957	15	985
Pasewaran	10.957	23	1.095
Karang Kedawung	10.957	19	1.051
Karangbinangun	10.957	14	1.081
Ngale	10.957	17	1.087

Jumlah pengamatan curah hujan pada masing-masing pos pengamatan berjumlah 10.957 pengamatan. Pada Pos Ngale nilai *threshold*-nya sebesar 17 mm dan jumlah pengamatan yang berada diatas *threshold* sebanyak 1.087 pengamatan. Pos Cawak memiliki nilai *threshold* sebesar 15 mm dan jumlah pengamatan yang berada diatas *threshold* sebanyak 985 pengamatan. Pos Karang Kedawung nilai *threshold*-nya sebesar 19 mm dan banyak pengamatan yang berada diatas *threshold* yaitu 1.051 pengamatan. Pos Karangbinangun memiliki nilai *threshold* sebesar 14 mm dan jumlah pengamatan yang berada diatas *threshold* sebanyak 1.081 pengamatan. Sedangkan Pos Pasewaran memiliki data ekstrim sebanyak 1.095 pengamatan dengan nilai *threshold* 23 mm.

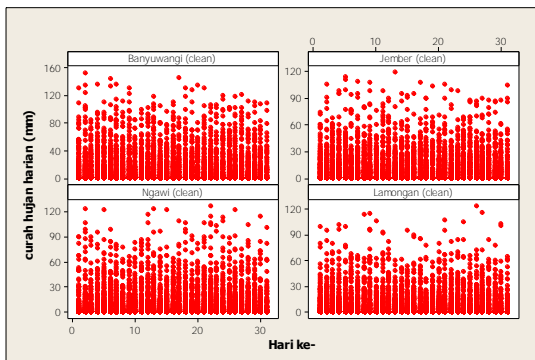
4.5 Identifikasi Dependensi Data Curah Hujan Ekstrim di Lima Pos Pengamatan

Identifikasi dependensi data pada penelitian ini dilakukan dengan melihat plot dari data curah hujan harian serta plot *autocorrelation function*. Dependensi data curah hujan pada masing-masing pos pengamatan dapat diketahui apabila data curah hujan harian tidak menyebar secara random atau membentuk suatu pola tertentu. Pola tersebut dapat berupa tren naik, tren turun, atau berpola siklik. Pada Gambar 4.9 akan ditunjukkan pola data curah hujan harian di Pos Cawak Kabupaten Bojonegoro sebagai berikut.



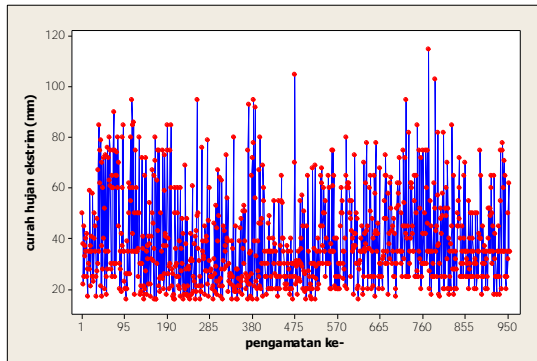
Gambar 4.9 Pola Data Curah Hujan Harian Pos Cawak

Pola data curah hujan berdasarkan Gambar 4.9 membentuk pola siklik. Hal tersebut mengindikasikan bahwa data curah hujan harian di Pos Cawak bersifat dependen terhadap waktu. Sementara itu hal yang sama juga ditunjukkan pada pola data curah hujan harian di empat pos lainnya yang disajikan pada Gambar 4.10. Indikasi adanya dependensi data juga ditemukan pada data curah hujan harian di masing-masing pos dimana pola yang terbentuk adalah pola siklik.

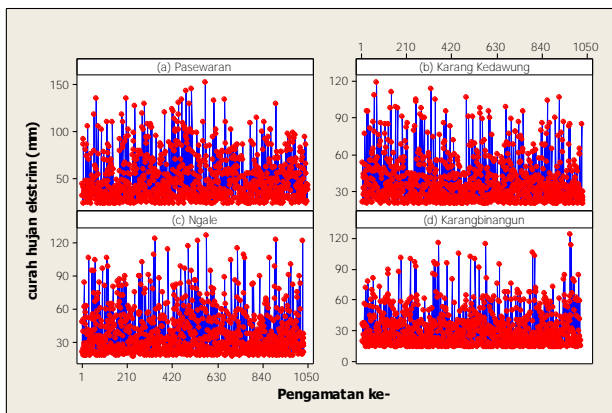


Gambar 4.10 Pola Data Curah Hujan Harian di Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun

Selain pola data curah hujan harian yang membentuk pola siklik, pola dari data ekstrimnya juga menunjukkan pola siklik. Hal tersebut dapat dilihat dari *time series* plot data ekstrim curah hujan di masing-masing pos pengamatan pada Gambar 4.11 untuk Pos Cawak, sedangkan pola data ekstrim untuk keempat pos lainnya dapat dilihat pada Gambar 4.12.



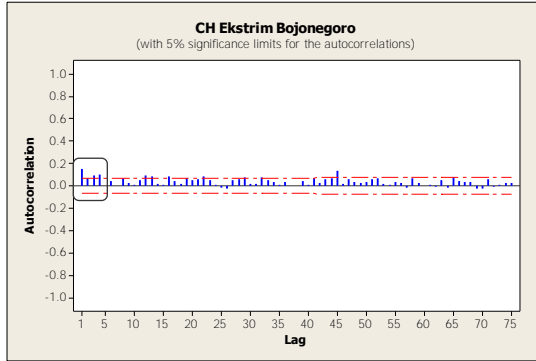
Gambar 4.11 Pola Data Ekstrim Curah Hujan Harian di Pos Cawak



Gambar 4.12 Pola Data Ekstrim Curah Hujan Harian di Pos Pasewaran (a), Pos Karang Kedawung (b), Pos Ngale (c), dan Pos Karangbinangun (d)

Selain dengan plot data curah hujan, identifikasi dependensi data juga dilakukan menggunakan plot *autocorrelation*

function dari data ekstrim curah hujan. Plot *autocorrelation function* (ACF) pada data ekstrim curah hujan di Pos Cawak disajikan pada Gambar 4.13. Apabila nilai ACF pada lag tertentu melebihi batas atas dan kurang dari batas bawah fungsi autokorelasi maka disimpulkan adanya dependensi data pada lag tersebut.

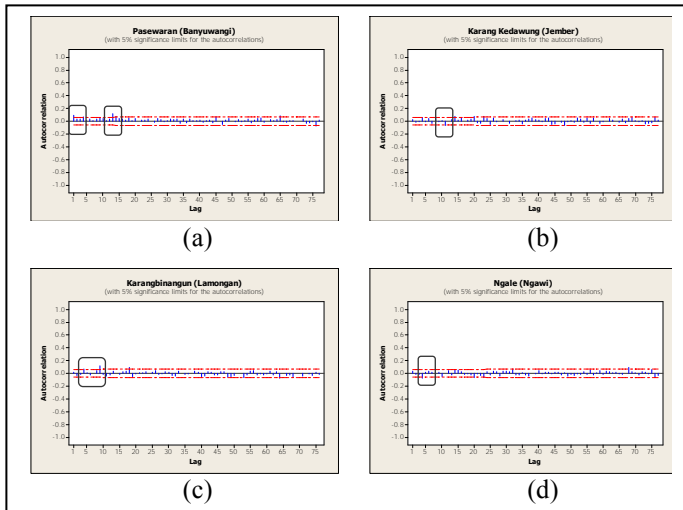


Gambar 4.13 Plot *Autocorrelation Function* Data Ekstrim Curah Hujan di Pos Cawak

(Catatan: Tanda kotak hitam menunjukkan lag-lag yang keluar dari batas atas atau batas bawah yang mengindikasikan adanya dependensi atau autokorelasi)

Gambar 4.13 menunjukkan ada beberapa nilai ACF yang melebihi batas atas fungsi autokorelasi yaitu pada lag-1, lag-3, dan lag-4, serta beberapa lag yang lebih tinggi lainnya. Maka data ekstrim curah hujan di Pos Cawak bersifat dependen.

Demikian pula pada data ekstrim curah hujan di empat pos lainnya terdapat beberapa nilai ACF pada lag tertentu yang keluar dari batas atas dan batas bawahnya. Sehingga menunjukkan data ekstrim memiliki korelasi yang tinggi dengan nilai ekstrim tahun sebelumnya atau sesudahnya. Plot ACF data ekstrim di empat pos pengamatan lainnya dapat dilihat pada Gambar 4.14



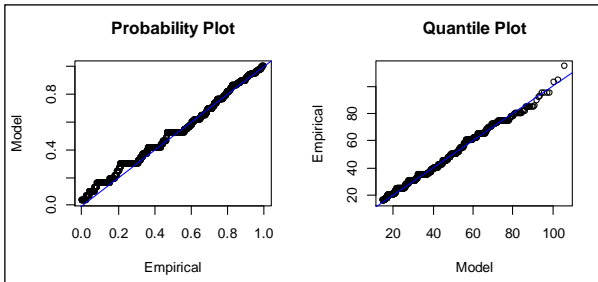
Gambar 4.14 Plot *Autocorrelation Function* Data Ekstrim Curah Hujan Harian di Pos Pasewaran (a), Pos Karang Kedawung (b), Pos Karangbinangun (c), dan Pos Ngale (d)
(Catatan: Tanda kotak hitam menunjukkan lag-lag yang keluar dari batas atas atau batas bawah yang mengindikasikan adanya dependensi atau autokorelasi)

4.6 Pemeriksaan Kesesuaian Distribusi *Generalized Pareto*

Pemeriksaan kesesuaian distribusi pada sampel ekstrim curah hujan dilakukan untuk menunjukkan adanya kesesuaian distribusi *generalized pareto*. Pada penelitian ini, pemeriksaan distribusi dilakukan dengan dua cara yaitu secara visual melalui *probability* plot dan *quantile* plot, serta melalui uji *Kolmogorov Smirnov*.

Gambar 4.15 menunjukkan *probability* plot dan *quantile* plot di Pos Cawak Kabupaten Bojonegoro. Pada kedua plot tersebut sebagian besar sebaran titik-titik mengikuti garis linear. Begitu pula pada Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun menunjukkan bahwa sebaran titik-titik pada *probability* plot dan *quantile* plot telah mengikuti garis linear, sehingga dapat disimpulkan bahwa data ekstrim

curah hujan di semua pos pengamatan mengikuti distribusi *generalized pareto*. *Probability* plot dan *quantile* plot di empat pos pengamatan lainnya dapat dilihat pada Lampiran E.



Gambar 4.15 *Probability* Plot dan *Quantile* Plot di Pos Cawak dengan GPD

Pemeriksaan kesesuaian distribusi untuk mengetahui bahwa sampel ekstrim curah hujan berdistribusi *Generalized Pareto Distribution* (GPD) juga dilakukan dengan uji *Kolmogorov-Smirnov* sebagai berikut.

Pengujian hipotesis :

$H_0: F_n(x) = F_0(x)$ (Data mengikuti *Generalized Pareto Distribution*)

$H_1: F_n(x) \neq F_0(x)$ (Data tidak mengikuti *Generalized Pareto Distribution*)

$\alpha = 0,02$

Daerah kritis :

Tolak H_0 jika $D_{hitung} > D_\alpha$ dimana D_α adalah nilai kritis untuk uji *Kolmogorov-Smirnov* dengan taraf signifikansi sebesar α .

Tabel 4.4 Uji *Kolmogorov-Smirnov Generalized Pareto Distribution*

Pos Pengamatan	D_{hitung}	D_α	Keputusan
Cawak	0,0453	0,0492	Gagal Tolak H_0
Pasewaran	0,0304	0,0469	Gagal Tolak H_0
Karang Kedawung	0,0339	0,0473	Gagal Tolak H_0
Ngale	0,0333	0,0473	Gagal Tolak H_0
Karangbinangun	0,0461	0,0476	Gagal Tolak H_0

Berdasarkan Tabel 4.4 menunjukkan bahwa D_{hitung} pada semua pos pengamatan memiliki nilai yang lebih kecil dibandingkan D_u sehingga kesimpulan yang dapat diambil adalah semua data ekstrim curah hujan di lima pos pengamatan mengikuti *Generalized Pareto Distribution*.

4.7 Proses *Extremal Index* dan *Declustering*

Salah satu syarat utama dalam menaksir parameter GPD adalah data harus independen. Berdasarkan hasil yang diperoleh pada analisis subbab 4.5 menunjukkan bahwa data curah hujan ekstrim dalam penelitian ini memiliki dependensi terhadap waktu sehingga melanggar asumsi tersebut. Oleh karena itu, sebelum mendapatkan estimasi parameter, dalam analisis ini akan dilakukan *declustering* untuk memfilter pengamatan dependen. Dalam proses *declustering* dibutuhkan suatu nilai r (*run length*) untuk membentuk *cluster*. Nilai r yang optimum ditentukan melalui *extremal index*. Berikut ini hasil estimasi *extremal index* di lima pos pengamatan curah hujan dalam Tabel 4.5.

Tabel 4.5 *Extremal Index* per Pos Pengamatan

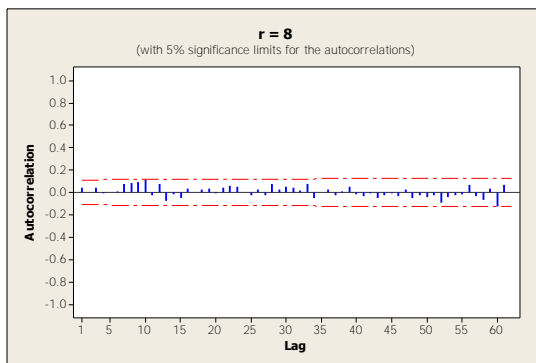
Pos Pengamatan	u (mm)	<i>Extremal Index</i>	r
Cawak	15	0,3234*	8*
Pasewaran	23	0,1534	10
Karang Kedawung	19	0,2748	8
Karangbinangun	14	0,2955	8
Ngale	17	0,2958	8

* perhitungan *extremal index* dan penentuan nilai r untuk Pos Cawak dapat dilihat di Lampiran K

Setelah diketahui nilai r yang optimum, proses selanjutnya adalah melakukan *declustering*. Secara spesifik, berikut ini dijelaskan proses *declustering* pada data curah hujan di salah satu pos pengamatan yaitu Pos Cawak Kabupaten Bojonegoro.

1. Menentukan *threshold* (u). Pada analisis sebelumnya ditentukan nilai *threshold* (u) untuk data curah hujan di Pos Cawak adalah 15 mm dan nilai-nilai yang berada diatas u dianggap sebagai nilai ekstrim.

2. Menentukan nilai r . Pada analisis ini digunakan nilai $r = 8$ yang diperoleh dari hasil *extremal index*.
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 data curah hujan di Pos Cawak, *cluster* pertama terbentuk, mulai pengamatan ke-2 sampai dengan ke-3. *Cluster* kedua terbentuk mulai pengamatan ke-14 sampai ke-35, dimana antara *cluster* pertama dan *cluster* kedua ada sebanyak 10 pengamatan di bawah *threshold*. Prosedur ini dilakukan terus menerus sampai pembentukan *cluster* selesai dan data telah habis. Pada data curah hujan di Pos Cawak terbentuk 307 *cluster* dengan $r=8$.
4. Mengidentifikasi nilai maksimum pada masing-masing *cluster*. Nilai maksimum ini selanjutnya digunakan untuk mengestimasi parameter GPD.
5. Setelah proses *declustering* selesai, kemudian mengidentifikasi dependensi data curah hujan hasil *declustering*. Gambar 4.16 menunjukkan plot ACF data curah hujan di Pos Cawak yang sudah melalui proses *declustering*.



Gambar 4.16 Plot *Autocorrelation Function* Data Curah Hujan Pos Cawak dengan *Declustering* $r = 8$

Plot *autocorrelation function* pada Gambar 4.16 menunjukkan tidak ada nilai ACF yang keluar dari batas atas dan batas bawah sehingga mengindikasikan bahwa data curah hujan

hasil proses *declustering* sudah independen. Plot ACF hasil *declustering* pada pos pengamatan lainnya disajikan pada Lampiran F. Identifikasi dependensi data curah hujan pada empat pos pengamatan lainnya juga menunjukkan hasil yang sama yaitu menunjukkan data curah hujan sudah independen setelah dilakukan *declustering*.

4.8 Estimasi Parameter GPD di Lima Pos Pengamatan

Estimasi parameter GPD di lima pos pengamatan dilakukan dengan dua cara, yaitu tanpa proses *declustering* dan dengan melalui proses *declustering*. *Declustering* pada cara yang kedua dilakukan dengan menggunakan nilai r yang diperoleh dari hasil *extremal index* pada pembahasan subbab 4.7 sebelumnya.

Nilai estimasi parameter skala dan parameter bentuk GPD, serta jumlah *cluster* yang terbentuk dan tipe distribusi data ekstrim curah hujan di lima pos pengamatan disajikan pada Tabel 4.6 sebagai berikut.

Tabel 4.6 Estimasi Parameter GPD Pos Cawak, Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun

Pos Pengamatan	Nilai	<i>Non-declustering</i>	<i>Declustering</i>
Cawak	$\hat{\sigma}$	30,0778	73,0761
	$\hat{\xi}$	-0,2818	-0,7284
	n_c	-	307
	Tipe Distribusi	Beta	Beta
Pasewaran	$\hat{\sigma}$	28,7968	78,0252
	$\hat{\xi}$	-0,1038	-0,5917
	n_c	-	177
	Tipe Distribusi	Beta	Beta
Karang Kedawung	$\hat{\sigma}$	20,0262	42,9709
	$\hat{\xi}$	-0,0457	-0,3903
	n_c	-	302
	Tipe Distribusi	Beta	Beta

Tabel 4.6 Estimasi Parameter GPD Pos Cawak, Pos Pasewaran, Pos Karang Kedawung, Pos Ngale, dan Pos Karangbinangun (Lanjutan)

Pos Pengamatan	Nilai	<i>Non-declustering</i>	<i>Declustering</i>
Ngale	$\hat{\sigma}$	21,695	48,3195
	$\hat{\xi}$	-0,0317	-0,3941
	n_c	-	321
	Tipe Distribusi	Beta	Beta
Karangbinangun	$\hat{\sigma}$	17,7387	42,4165
	$\hat{\xi}$	0,0146	-0,348
	n_c	-	295
	Tipe Distribusi	Pareto	Beta

4.9 *Return level* Curah Hujan Ekstrim di Lima Pos Pengamatan

Nilai estimasi parameter GPD yang diperoleh digunakan untuk menghitung *return level*. Sedangkan *return level* dapat digunakan untuk mengetahui nilai estimasi curah hujan pada periode waktu tertentu. Nilai *return level* untuk curah hujan selama 1 tahun mendatang lima pos pengamatan disajikan pada Tabel 4.7.

Tabel 4.7 Nilai *Return Level* (mm) di Lima Pos Pengamatan

Pos Pengamatan	Periode Ulang	Aktual	<i>Non-Declustering</i>	<i>Declustering</i>
Cawak	3 bln	78	62,7905	64,6179
	6 bln	78	73,2482	84,7191
	9 bln	78	78,4827	92,5454
	12 bln	85	81,8501	96,8516
Pasewaran	3 bln	106	79,7064	50,0559
	6 bln	106	95,0278	83,3174
	9 bln	106	103,4922	100,1516
	12 bln	106	109,2853	108,715

Tabel 4.7 Nilai *Return Level* (mm) di Lima Pos Pengamatan (Lanjutan)

Pos Pengamatan	Periode Ulang	Aktual	<i>Non-Declustering</i>	<i>Declustering</i>
Karang Kedawung	3 bln	40	60,6546	52,6145
	6 bln	60	73,0183	70,7418
	9 bln	60	80,071	79,2820
	12 bln	80	84,9963	84,5718
Karangbinangun	3 bln	105	53,1844	47,0653
	6 bln	105	65,9425	66,1015
	9 bln	105	73,4658	75,2846
	12 bln	105	78,8308	81,0574
Ngale	3 bln	122	62,8333	56,2739
	6 bln	122	76,7115	76,1948
	9 bln	122	84,6895	85,5605
	12 bln	122	90,2882	91,3543

Nilai RMSE digunakan untuk menentukan metode yang sesuai dalam menghitung *return level* di lima pos pengamatan. Nilai RMSE dari *return level* hasil *declustering* dibandingkan dengan RMSE dari *return level* tanpa dilakukan *declustering*. Nilai RMSE dihitung berdasarkan nilai aktual dan nilai dugaan *return level* dari masing-masing kabupaten. Metode yang memiliki nilai RMSE terkecil adalah metode yang sesuai dalam menentukan *return level*. Hasil perbandingan nilai RMSE disajikan pada Tabel 4.8 berikut.

Tabel 4.8 Nilai RMSE *Return Level*

Pos Pengamatan	<i>Non-Declustering</i>	<i>Declustering</i>
Cawak	8,125	12,002
Pasewaran	14,395	30,355
Karang Kedawung	15,999	12,915
Karangbinangun	38,372	39,767
Ngale	44,579	46,597

Tabel 4.8 menunjukkan bahwa nilai RMSE yang dihasilkan dari *return level* melalui *declustering* cenderung sama dibandingkan nilai *return level* tanpa *declustering*. Hal ini

menunjukkan bahwa nilai dugaan yang dihasilkan melalui *declustering* tidak sesuai dengan return level aktual, sehingga dalam perhitungan *return level* memberikan hasil yang sama dengan hasil estimasi tanpa *declustering*. Selain itu, hal ini dapat ditunjukkan dengan pengujian *mean* dua populasi sebagai berikut:

Hipotesis:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

dimana : μ_1 = rata-rata RMSE *return level* dari data tanpa *declustering*

μ_2 = rata-rata RMSE *return level* dari data hasil *declustering*

$$\alpha = 0,05$$

$$t_{\text{hitung}} = -1,26$$

$$t_{\text{tabel}} = 3,495$$

$t_{\text{hit}} < t_{\text{tabel}}$ sehingga kesimpulan yang diperoleh : gagal tolak H_0 . Maka rata-rata RMSE *return level* dari data tanpa *declustering* sama dengan rata-rata RMSE *return level* dari data hasil *declustering*. Ada dugaan bahwa *declustering* merupakan tahap awal dalam analisis memodelkan parameter GPD sehingga hanya mampu mengatasi dependensi data namun tidak menghasilkan *return level* yang lebih baik. Oleh karena itu, perlu dibahas dan dianalisis lebih lanjut dengan pendekatan model parameter untuk menghasilkan nilai *return level* yang lebih sesuai.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, maka dapat disimpulkan bahwa:

1. Karakteristik curah hujan ekstrim di Jawa Timur khususnya Kabupaten Jember, Bojonegoro, Lamongan, Banyuwangi, dan Ngawi memiliki pola *heavy tail* (mengandung nilai ekstrim) dan tidak berdistribusi normal, namun mengikuti distribusi General Pareto. Selain itu data ekstrim di lima pos pengamatan masing-masing kabupaten bersifat dependen karena berpola siklik dan memiliki nilai korelasi yang tinggi pada lag-lag tertentu.
2. Estimasi parameter dengan *declustering* menghasilkan suatu nilai estimasi parameter bentuk dan skala GPD. Proses *declustering* pada data ekstrim dapat mengatasi dependensi data sehingga menghasilkan data yang independen. Namun, hasil *return level* dengan menggunakan metode *declustering* menghasilkan nilai RMSE yang sama dengan RMSE *return level* data tanpa *declustering*.

5.2 Saran

Pada penelitian ini metode *declustering* berhasil mengatasi dependensi data, namun tidak menghasilkan *return level* yang lebih baik daripada data tanpa *declustering*. Pada analisis selanjutnya sebaiknya dilakukan analisis *declustering* pada pos pengamatan yang berbeda atau menggunakan data simulasi sehingga ada kemungkinan akan menghasilkan estimasi parameter dan *return level* yang lebih baik. Selain itu juga dapat melakukan analisis dengan menggunakan metode pemodelan parameter GPD untuk memperoleh estimasi parameter dan *return level* yang lebih baik.

(Halaman ini sengaja dikosongkan)

DAFTAR PUSTAKA

- Aczel, Sounderpandian. (2008). *Complete Business Statistics*. United States of America: McGraw-Hill Primis
- Badan Pusat Statistika. (2013). *Data Strategis 2013*. Jakarta : Badan Pusat Statistika.
- Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. London: Springer-Verlag.
- Dahan, E., Mendelson, H. (2001). An extreme value model of concept testing. *Management Science* 47, 102-116.
- Daniel, W. (1989). *Statistika Nonparametrik Terapan*. Jakarta: PT Gramedia.
- Dawson, T. H. (2000). Maximum wave crests in heavy seas. *Journal of Offshore Mechanics and Arctic Engineering - Transactions of the AMSE* 122, 222-224.
- Ferro, C.A.T., Segers, J. (2003). Inference for Clusters of Extreme Values. *Journal R. Stat. Society*, B65:545-556.
- Gilli, M., Kellezi, E. (2003). *An Application of Extreme Value Theory for Measuring Risk*. Elsevier Science.
- Gilleland, E., Katz, R.W. (2006). Analyzing Seasonal to Interannual Extreme Weather and Climate Variability with the Extremes Toolkit (extRemes). *18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual Meeting*. Atlanta.
- Hermawan, E. (2010). *Kondisi Iklim Indonesia Saat ini dan Prediksinya dalam Beberapa Bulan Mendatang Berbasis Data Iklim Global*. Bandung: Lembaga Penerbangan dan Antariksa Nasional (LAPAN).
- Jaruskova, D., Hanek, M. (2006). Peaks Over Threshold 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.
- Johnson, R.A., Bhattacharyya, G.K. (1977). *Statistical Concepts and Methods*. New York : John Wiley & Sons.
- 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.
- Lavenda, B. H., Cipollone, E. (2000). Extreme value statistics and thermodynamics of earthquakes: aftershock sequences. *Annali di geofisica* 43, 967-982.
- Lopes, H.F., Nascimento, F. F., Gamerman, D. (2006). Generalized Pareto models with time-varying tail behavior. Retrieved September 22, 2013, from <http://www.ime.unicamp.br/sinape/sites/default/files/extendedsinape.pdf>
- McNeil, A.J. (1999). *Extreme Value Theory for Risk Managers*. Zurich: Departement Mathematic ETH Zentrum.
- Nihayatin, L.N. (2013). *Perbandingan Uji Homogenitas Pada Data Curah Hujan Di Jawa Timur Sebagai Pra-Pemrosesan Kajian Perubahan Iklim*. Tugas Akhir S1: Jurusan Statistika, Institut Teknologi Sepuluh Nopember
- Roberts, S. J. (2000). Extreme value statistics for novelty detection in biomedical data processing. *IEEE Proceedings - Science Measurement and Technology* 147, 363-367.
- Soukissian, T. H., Arapi, P. M. (2011). The Effect of Declustering in the r-Largest Maxima Model for the Estimation of H_s -Design Values. *The Open Ocean Engineering Journal*, 4, 34-43.

- 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.
- Walpole, R. E. (1995). *Pengantar Statistika*. Jakarta: PT Gramedia Pustaka Utama.
- Wei, W.W.S. (2004). *Time Series Analysis Univariate and Multivariate Methods*. Canada: Addison Wesley Publishing, Company. Inc

(Halaman ini sengaja dikosongkan)

BIOGRAFI PENULIS



Rosna Malika, lahir di Lamongan pada tanggal 23 Mei 1992, anak kedua dari 3 bersaudara dari pasangan Suparyo (Alm) dan Suswati. Ia menempuh jenjang pendidikan formal TK Aisyah Bustanul Athfal Parengan, SDN Sekaran 2, SMPN 1 Maduran, dan SMAN 2 Lamongan, kemudian tahun 2010 ia diterima di Jurusan Statistika ITS melalui jalur SNMPTN, dan sekarang sedang menempuh pendidikan magister S2 Statistika di ITS dengan memperoleh beasiswa *fasttrack*. Semasa kuliah Rosna menempa dirinya

dengan aktif dibeberapa organisasi dan kegiatan kemahasiswaan. Sebagai Koordinator putri BK DANUS JMMI-ITS 2013/2014, Koordinator putri Departemen Tablighul Islam FORSIS-ITS 2012/2013, Sekretaris Departemen Sosial Masyarakat HIMASTA-ITS 2012/2013, Tim Sekretaris HIMASTA-ITS 2011/2012, Koordinator putri Tim Keuangan FORSIS-ITS 2011/2012, Ketua Tim Soal *Statistics Competition* (STATION) 2012, dan kepanitiaan kegiatan kampus lainnya. Rosna juga memiliki pengalaman menjadi asisten dosen mata kuliah Pengantar Metode Statistika dan Program Komputer serta pernah mengikuti *Internship Program* di PT. Petrokimia Gresik dan PT. AIA Financial. Apabila pembaca memiliki saran, kritik, atau ingin berdiskusi dengan penulis, dapat dihubungi melalui email: rosna23@gmail.com.

DAFTAR LAMPIRAN

	Halaman
Lampiran A	Data Curah Hujan (1981-2010)..... 59
Lampiran B	Identifikasi Data yang Kurang Sesuai 64
Lampiran C	Pola Curah Hujan Harian 66
Lampiran D	Pengambilan Sampel Ekstrim dengan <i>Peaks Over Threshold</i> 68
Lampiran E	<i>Probability & Quantile Plot</i> 70
Lampiran F	Plot ACF Hasil <i>Declustering</i> 71
Lampiran G	<i>Extremal Index</i> dan <i>Declustering</i> 73
Lampiran H	Estimasi Parameter GPD..... 79
Lampiran I	<i>Return Level</i> 84
Lampiran J	Ilustrasi Penentuan Nilai <i>Threshold</i> dengan Metode Presentase Pada Data Curah Hujan di Pos Cawak..... 89
Lampiran K	Ilustrasi Menghitung <i>Extremal Index</i> dan Menentukan <i>Run Length (r)</i> di Pos Cawak 91

(Halaman ini sengaja dikosongkan)

LAMPIRAN

LAMPIRAN A: DATA CURAH HUJAN (1981-2010)

A1. Pos Cawak Kabupaten Bojonegoro

Tahun	Bulan	Tanggal	Curah Hujan (mm)
1981	1	1	0
1981	1	2	50
1981	1	3	38
1981	1	4	9
1981	1	5	0
1981	1	6	0
1981	1	7	0
1981	1	8	9
1981	1	9	0
1981	1	10	0
1981	1	11	0
1981	1	12	3
1981	1	13	5
1981	1	14	22
1981	1	15	0
...
...
...
2010	12	25	0
2010	12	26	0
2010	12	27	0
2010	12	28	0
2010	12	29	0
2010	12	30	30
2010	12	31	0

A2. Pos Pasewaran Kabupaten Banyuwangi

Tahun	Bulan	Tanggal	Curah Hujan (mm)
1981	1	1	44
1981	1	2	32
1981	1	3	20
1981	1	4	24
1981	1	5	15
1981	1	6	87
1981	1	7	92
1981	1	8	0
1981	1	9	0
1981	1	10	0
1981	1	11	0
1981	1	12	14
1981	1	13	41
1981	1	14	32
1981	1	15	1
...
...
...
2010	12	23	7
2010	12	24	0
2010	12	25	0
2010	12	26	0
2010	12	27	78
2010	12	28	6
2010	12	29	0
2010	12	30	0
2010	12	31	0

A3. Pos Karang Kedawung, Kabupaten Jember

Tahun	Bulan	Tanggal	Curah Hujan (mm)
1981	1	1	0
1981	1	2	21
1981	1	3	53
1981	1	4	11
1981	1	5	8
1981	1	6	44
1981	1	7	41
1981	1	8	4
1981	1	9	0
1981	1	10	32
1981	1	11	0
1981	1	12	0
1981	1	13	0
1981	1	14	22
1981	1	15	0
...
...
...
2010	12	23	0
2010	12	24	0
2010	12	25	0
2010	12	26	0
2010	12	27	3
2010	12	28	0
2010	12	29	0
2010	12	30	0
2010	12	31	0

A4. Pos Ngale, Kabupaten Ngawi

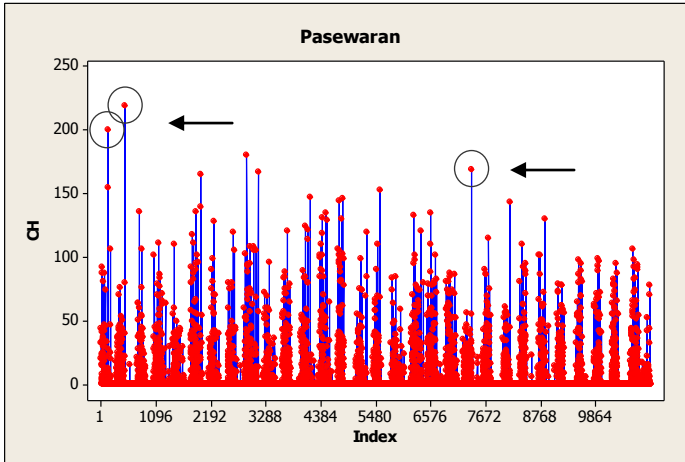
Tahun	Bulan	Tanggal	Curah Hujan (mm)
1981	1	1	0
1981	1	2	0
1981	1	3	0
1981	1	4	4
1981	1	5	12
1981	1	6	1
1981	1	7	1
1981	1	8	21
1981	1	9	2
1981	1	10	0
1981	1	11	0
1981	1	12	11
1981	1	13	2
1981	1	14	4
1981	1	15	6
...
...
...
2010	12	23	15
2010	12	24	113
2010	12	25	0
2010	12	26	0
2010	12	27	0
2010	12	28	1
2010	12	29	11
2010	12	30	0
2010	12	31	0

A5. Pos Karangbinangun, Kabupaten Lamongan

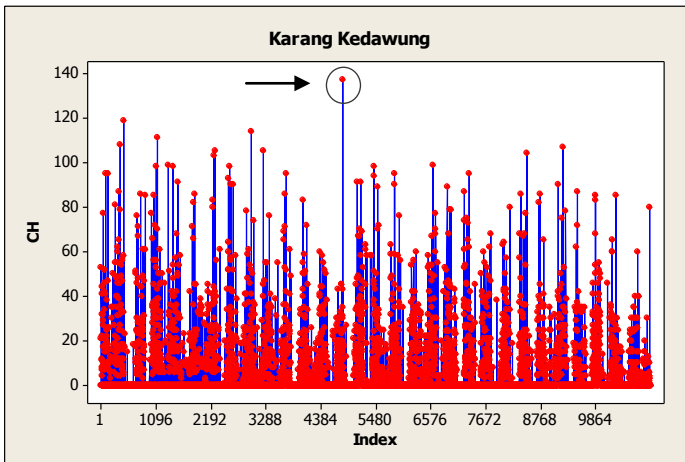
Tahun	Bulan	Tanggal	Curah Hujan (mm)
1981	1	1	12
1981	1	2	20
1981	1	3	37
1981	1	4	26
1981	1	5	0
1981	1	6	0
1981	1	7	0
1981	1	8	38
1981	1	9	3
1981	1	10	12
1981	1	11	4
1981	1	12	4
1981	1	13	19
1981	1	14	3
1981	1	15	7
...
...
...
2010	12	23	36
2010	12	24	25
2010	12	25	0
2010	12	26	3
2010	12	27	0
2010	12	28	0
2010	12	29	0
2010	12	30	7
2010	12	31	0

LAMPIRAN B : IDENTIFIKASI DATA YANG KURANG SESUAI

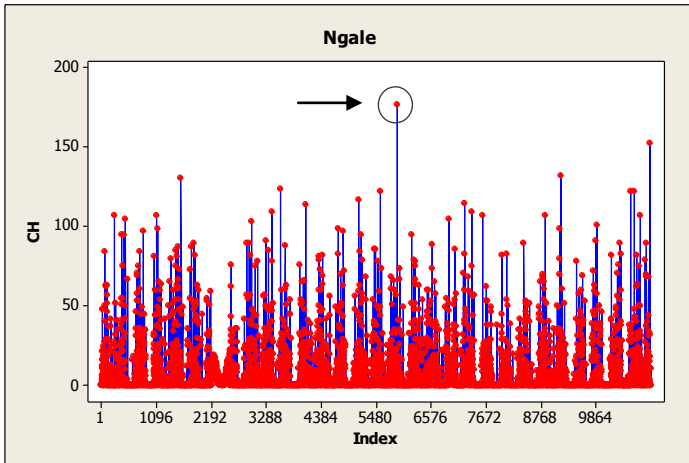
B1. Pos Pasewaran, Kabupaten Banyuwangi



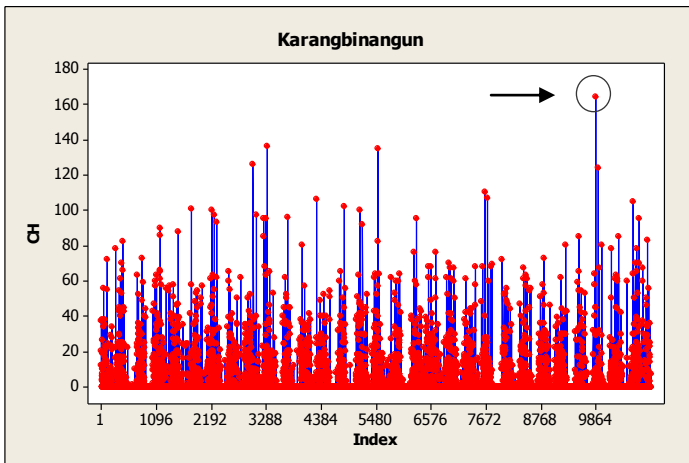
B2. Pos Karang Kedawung, Kabupaten Jember

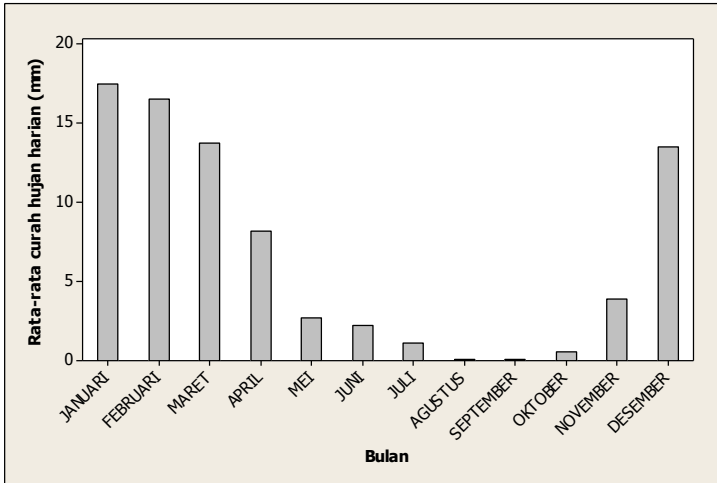
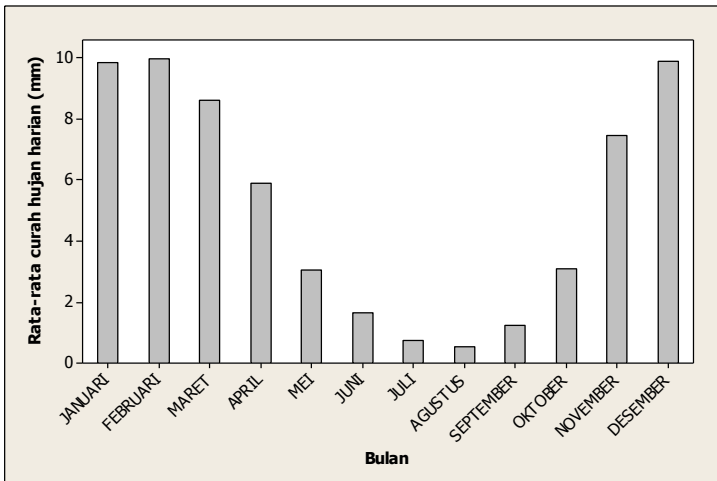


B3. Pos Ngale, Kabupaten Ngawi

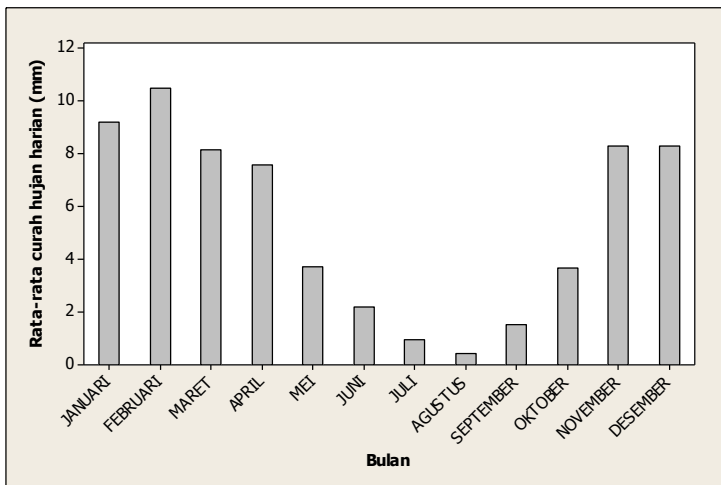


B4. Pos Karangbinangun, Kabupaten Lamongan

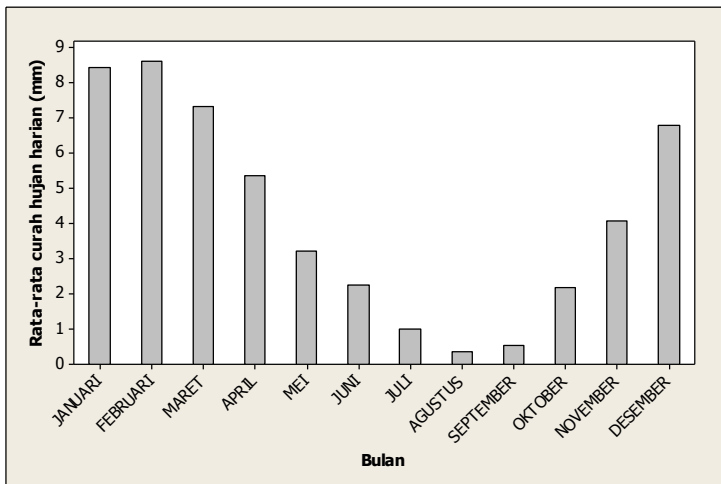


LAMPIRAN C : POLA CURAH HUJAN HARIAN**C1. Pos Pasewaran, Kabupaten Banyuwangi****C2. Pos Karang Kedawung, Kabupaten Jember**

C3. Pos Ngale, Kabupaten Ngawi

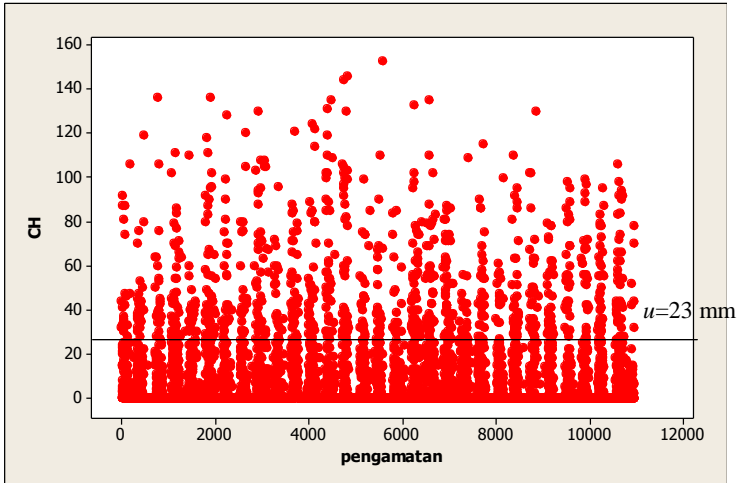


C4. Pos Karangbinangun, Kabupaten Lamongan

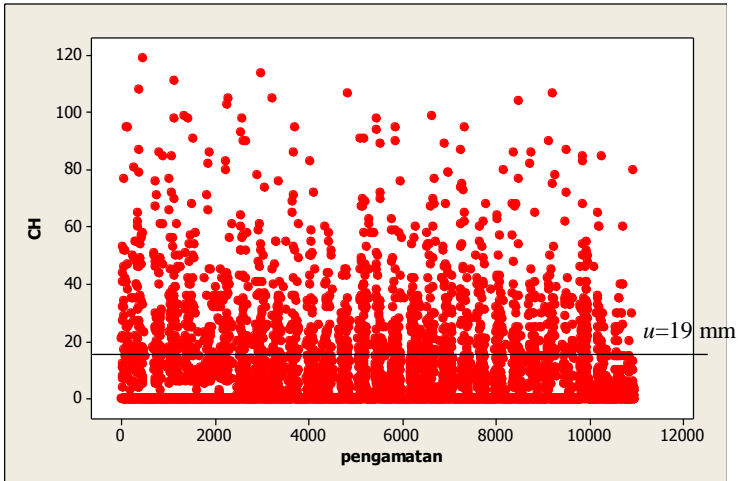


LAMPIRAN D : PENGAMBILAN SAMPEL EKSTRIM DENGAN *PEAKS OVER THRESHOLD*

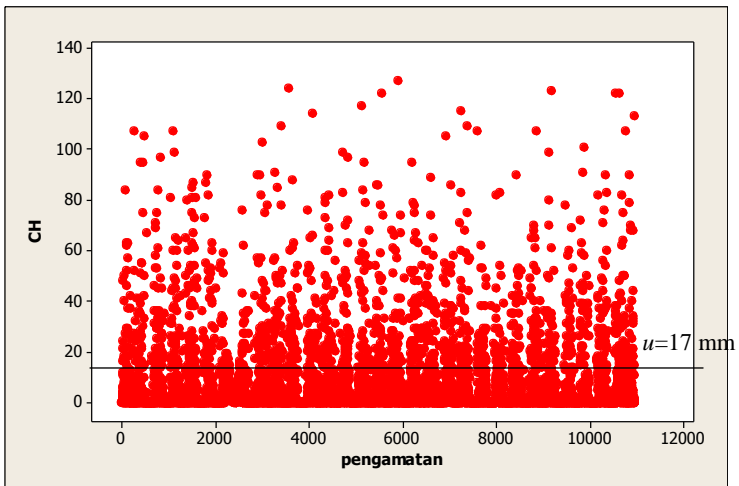
D1. Pos Pasewaran, Kabupaten Banyuwangi



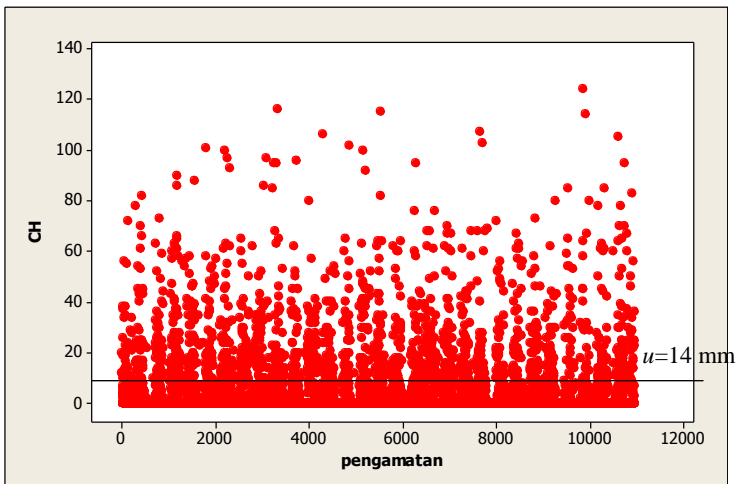
D2. Pos Karang Kedawung, Kabupaten Jember

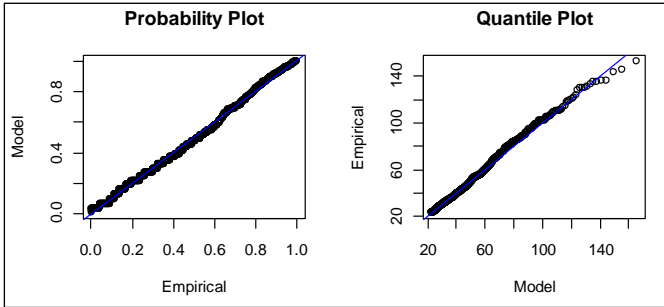
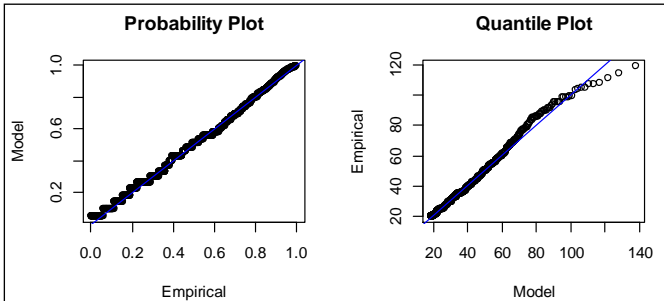
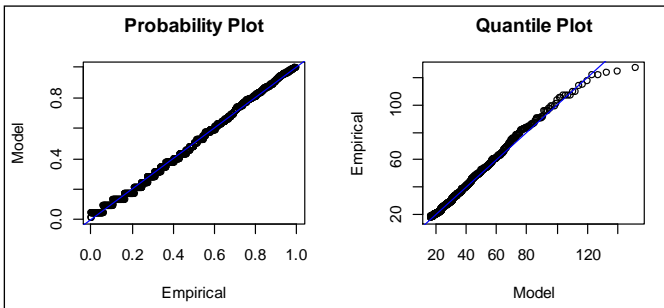


D3. Pos Ngale, Kabupaten Ngawi

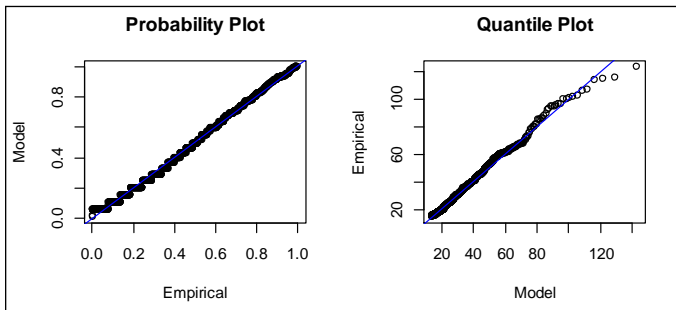


D4. Pos Karangbinangun, Kabupaten Lamongan



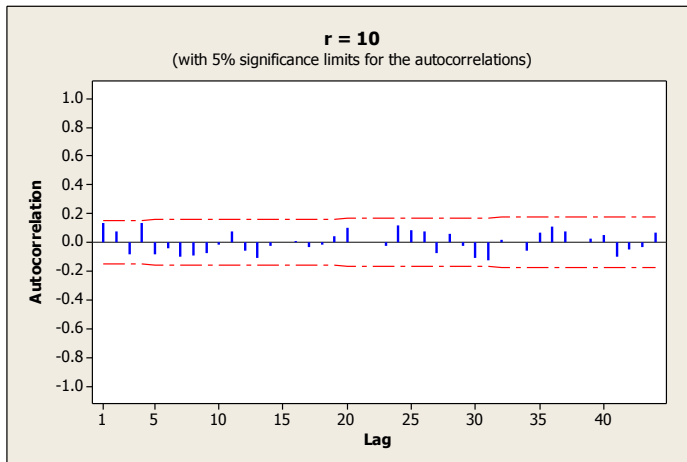
LAMPIRAN E : PROBABILITY & QUANTILE PLOT**E1. Pos Pasewaran, Kabupaten Banyuwangi****E2. Pos Karang Kedawung, Kabupaten Jember****E3. Pos Ngale, Kabupaten Ngawi**

E4. Pos Karangbinangun, Kabupaten Lamongan

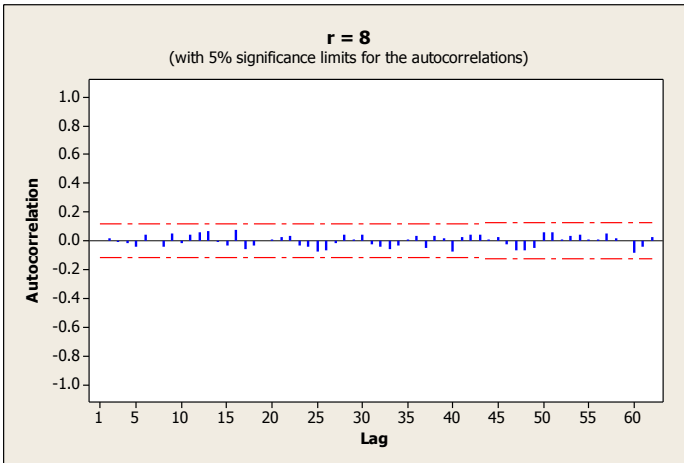


LAMPIRAN F : PLOT ACF HASIL *DECLUSTERING*

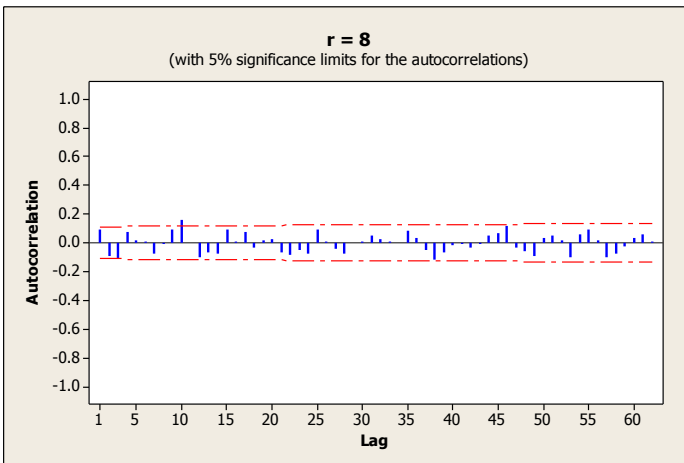
F1. Pos Pasewaran, Kabupaten Banyuwangi



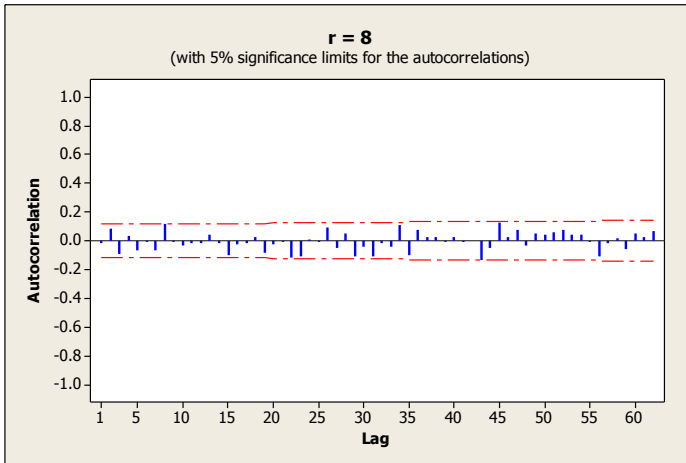
F2. Pos Karang Kedawung, Kabupaten Jember



F3. Pos Ngale, Kabupaten Ngawi

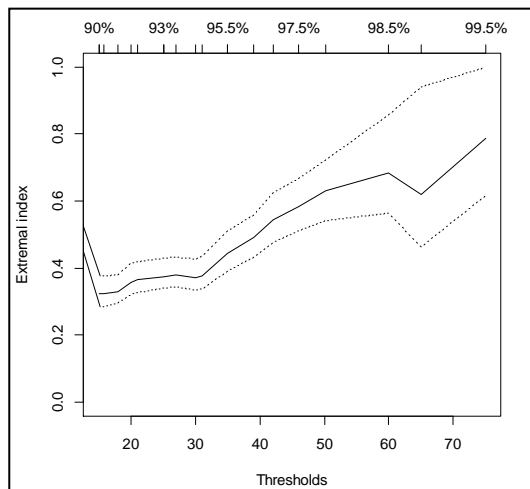


F4. Pos Karangbinangun, Kabupaten Lamongan



LAMPIRAN G : *EXTREMAL INDEX DAN DECLUSTERING*

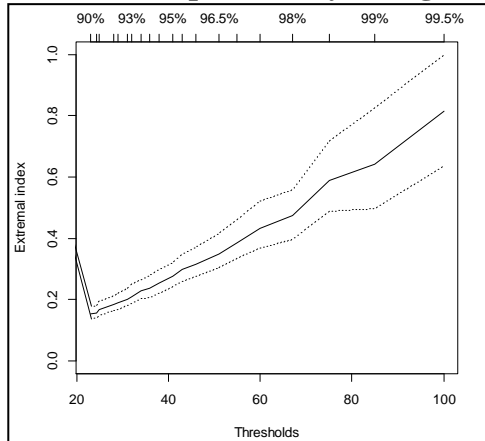
G1. Pos Cawak, Kabupaten Bojonegoro



	Thresholds	No. of Clusters	Run Length	Extremal Index	
	90%	15.00	286	8	0.32342 (0.28579, 0.37652)
	90.5%	15.00	286	8	0.32342 (0.28579, 0.37652)
	91%	15.81	286	8	0.32342 (0.28579, 0.37652)
	91.5%	18.00	285	8	0.32986 (0.29601, 0.37869)
	92%	20.00	253	9	0.35799 (0.32065, 0.41702)
	92.5%	21.00	250	9	0.3649 (0.32654, 0.41904)
	93%	25.00	237	9	0.37519 (0.33962, 0.42871)
	93.5%	25.00	237	9	0.37519 (0.33962, 0.42871)
	94%	27.00	238	9	0.3796 (0.34252, 0.43385)
	94.5%	30.00	192	11	0.37224 (0.33408, 0.42603)
	95%	31.00	188	11	0.37778 (0.33835, 0.43523)
	95.5%	35.00	178	11	0.44288 (0.39231, 0.51093)
	96%	35.00	178	11	0.44288 (0.39231, 0.51093)
	96.5%	39.00	168	12	0.4903 (0.43314, 0.55949)
	97%	42.00	157	12	0.54327 (0.47669, 0.62545)
	97.5%	46.00	149	11	0.58273 (0.51216, 0.66777)
	98%	50.18	133	10	0.6295 (0.54175, 0.72163)
	98.5%	60.00	93	15	0.68246 (0.56248, 0.85651)
	99%	65.00	57	25	0.61969 (0.46484, 0.94078)
	99.5%	75.00	30	32	0.78775 (0.61673, 1)

	[1]	"Declustering ..."			
	[1]	"declustering performed for:"			
	[1]	"CH and assigned to CH.u15r8dcbyYEAR"			
	[1]	"307 clusters using threshold of 15 and r = 8"			

G2. Pos Pasewaran, Kabupaten Banyuwangi



	Thresholds	No. of Clusters	Run Length	Extremal Index
90%	23.00000	154	10	0.1534 (0.13774, 0.17776)
90.5%	24.29483	152	10	0.157 (0.14073, 0.18024)
91%	25.00000	156	10	0.16692 (0.14946, 0.19443)
91.5%	28.00000	160	10	0.18352 (0.16378, 0.21355)
92%	29.00000	161	10	0.19054 (0.16827, 0.22196)
92.5%	31.00000	150	11	0.20199 (0.18162, 0.23637)
93%	32.00000	152	11	0.21108 (0.18769, 0.2513)
93.5%	34.00000	153	11	0.22908 (0.20399, 0.26455)
94%	36.00000	146	12	0.23762 (0.2077, 0.27845)
94.5%	38.00000	140	13	0.25393 (0.22156, 0.30035)
95%	41.00000	141	12	0.27687 (0.24299, 0.32214)
95.5%	43.00000	139	12	0.29912 (0.25965, 0.35045)
96%	46.00000	126	14	0.31517 (0.27602, 0.37234)
96.5%	51.00000	125	14	0.34998 (0.30503, 0.4174)
97%	55.00000	119	14	0.38536 (0.33547, 0.46009)
97.5%	60.00000	109	15	0.43186 (0.36882, 0.52203)
98%	67.00000	97	15	0.47332 (0.39692, 0.55949)
98.5%	75.00000	91	13	0.58917 (0.48981, 0.71591)
99%	85.00000	64	14	0.64278 (0.49757, 0.82525)
99.5%	100.0000	41	14	0.81354 (0.63717, 1)

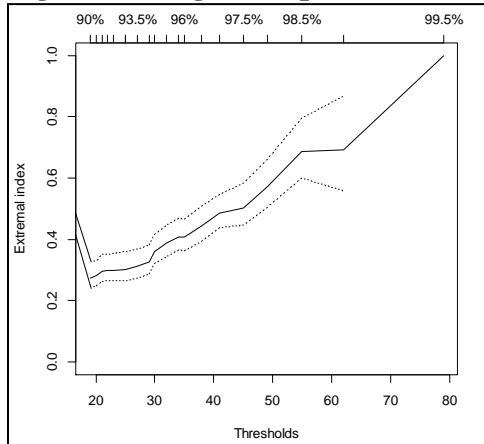
[1] "Declustering ..."

[1] "declustering performed for:"

[1] "CH and assigned to CH.u23r10dcbyYEAR"

[1] "177 clusters using threshold of 23 and r = 10"

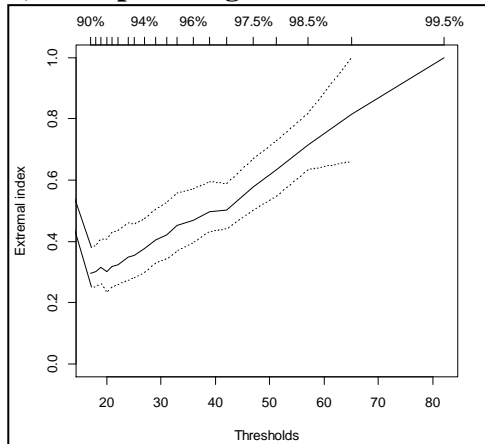
G3. Pos Karang Kedawung, Kabupaten Jember



Thresholds	No. of Clusters	Run Length	Extremal Index	
90%	19.00	281	8	0.27485 (0.24114, 0.3257)
90.5%	20.00	249	9	0.28311 (0.24957, 0.331)
91%	21.00	253	9	0.29584 (0.26144, 0.351)
91.5%	22.00	251	9	0.29883 (0.26456, 0.35305)
92%	23.00	250	9	0.29904 (0.26518, 0.35433)
92.5%	25.00	209	11	0.30221 (0.26444, 0.36013)
93%	25.00	209	11	0.30221 (0.26444, 0.36013)
93.5%	27.00	203	11	0.31319 (0.27372, 0.36849)
94%	29.00	201	11	0.32716 (0.28721, 0.38232)
94.5%	30.00	200	11	0.36158 (0.32081, 0.41595)
95%	32.00	196	11	0.38817 (0.34463, 0.44641)
95.5%	34.00	190	11	0.4071 (0.36509, 0.46798)
96%	35.00	163	13	0.40835 (0.36447, 0.46708)
96.5%	38.00	155	13	0.44454 (0.39374, 0.50774)
97%	41.00	149	13	0.48671 (0.43886, 0.54725)
97.5%	45.00	123	16	0.50263 (0.44623, 0.58237)
98%	49.18	119	16	0.57088 (0.50564, 0.66226)
98.5%	55.00	104	13	0.68744 (0.59931, 0.79409)
99%	62.09	73	20	0.69245 (0.55884, 0.86874)
99.5%	79.00	52	0	1 (NA, NA)

[1] "Declustering ..."				
[1] "declustering performed for:"				
[1] "CH and assigned to CH.u19r8dcbyYEAR"				
[1] "302 clusters using threshold of 19 and r = 8"				

G4. Pos Ngale, Kabupaten Ngawi



Thresholds	No. of Clusters	Run Length	Extremal Index	
90%	17.00	304	8	0.29584 (0.25137, 0.37684)
90.5%	18.00	260	9	0.30268 (0.25117, 0.38832)
91%	19.00	262	9	0.31693 (0.26512, 0.40824)
91.5%	20.00	256	9	0.30147 (0.23475, 0.40903)
92%	21.00	255	9	0.31742 (0.25105, 0.42991)
92.5%	22.00	229	10	0.32529 (0.25978, 0.43482)
93%	24.00	229	10	0.34932 (0.27363, 0.46111)
93.5%	25.00	227	10	0.35597 (0.28223, 0.45919)
94%	27.00	220	10	0.37567 (0.30037, 0.47484)
94.5%	29.00	219	10	0.40512 (0.32958, 0.50411)
95%	31.00	214	10	0.42281 (0.34263, 0.52622)
95.5%	33.00	210	10	0.45219 (0.36885, 0.55731)
96%	36.00	190	11	0.46973 (0.39609, 0.57324)
96.5%	39.00	174	12	0.49803 (0.43246, 0.59448)
97%	42.00	151	14	0.50245 (0.44172, 0.58861)
97.5%	47.00	147	14	0.5784 (0.50209, 0.66891)
98%	51.18	132	14	0.63235 (0.54773, 0.72849)
98.5%	57.00	112	15	0.71443 (0.63391, 0.81793)
99%	65.00	85	15	0.81613 (0.66243, 0.99879)
99.5%	82.00	50	0	1 (NA, NA)

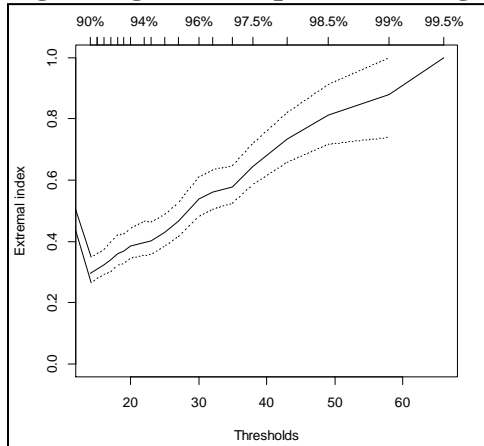
[1] "Decustering ..."

[1] "declustering performed for:"

[1] "CH and assigned to CH.u17r8dcbyYEAR"

[1] "321 clusters using threshold of 17 and r = 8"

G5. Pos Karangbinangun, Kabupaten Lamongan



Thresholds	No. of Clusters	Run Length	Extremal Index	
90%	14	277	8	0.29547 (0.26138, 0.34937)
90.5%	15	282	8	0.30984 (0.27902, 0.36041)
91%	15	282	8	0.30984 (0.27902, 0.36041)
91.5%	16	283	8	0.3248 (0.29262, 0.37373)
92%	17	280	8	0.34021 (0.30111, 0.39889)
92.5%	18	277	8	0.36047 (0.32325, 0.42035)
93%	19	247	9	0.36852 (0.32823, 0.42481)
93.5%	20	250	9	0.38539 (0.34614, 0.44385)
94%	22	230	10	0.39697 (0.35551, 0.46587)
94.5%	23	225	10	0.4032 (0.35753, 0.46395)
95%	25	218	10	0.42908 (0.38575, 0.48892)
95.5%	27	218	10	0.46605 (0.41718, 0.52633)
96%	30	213	10	0.5394 (0.48173, 0.61197)
96.5%	32	189	11	0.55994 (0.50571, 0.63443)
97%	35	180	11	0.57679 (0.52447, 0.64705)
97.5%	38	162	11	0.64531 (0.58485, 0.72043)
98%	43	146	10	0.7328 (0.65997, 0.82023)
98.5%	49	125	10	0.81071 (0.71829, 0.91142)
99%	58	87	7	0.87859 (0.73973, 1)
99.5%	66	52	0	1 (NA, NA)

[1] "Declustering ..."				
[1] "declustering performed for:"				
[1] "CH and assigned to CH.u14r8dcbyYEAR"				
[1] "295 clusters using threshold of 14 and r = 8"				

LAMPIRAN H : ESTIMASI PARAMETER GPD

H1. Pos Cawak, Kabupaten Bojonegoro

TANPA DECLUSTERING

Likelihood ratio test (5% level) for $x_i=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 76.38334 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 2.336143e-18
 Convergence successfull!

[1] "Threshold = 15"

[1] "Number of exceedances of threshold = 954"

[1] "Exceedance rate (per year)= 32.8973281722054"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 30.0778440	1.08945438
Shape (xi)	: -0.2818331	0.01865372

[1] "Negative log-likelihood: 3932.34082592176"

Parameter covariance:

	[,1]	[,2]
[1,]	1.18691084	-0.0168304151
[2,]	-0.01683042	0.0003479614

DECLUSTERING r=8

Likelihood ratio test (5% level) for $x_i=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 32.4445 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 1.226454e-08
 Convergence successfull!

[1] "Threshold = 15"

[1] "Number of exceedances of threshold = 296"

[1] "Exceedance rate (per year)= 10.2071374622357"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 73.076064	0.95092188
Shape (xi)	: -0.728404	0.00722797

[1] "Negative log-likelihood: 1324.52014715237"

Parameter covariance:

	[,1]	[,2]
[1,]	0.904252420	-6.282158e-03
[2,]	-0.006282158	5.224355e-05

H2. Pos Pasewaran, Kabupaten Banyuwangi

TANPA DECLUSTERING

Likelihood ratio test (5% level) for $x_i=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 8.334235 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 0.003890486
 Convergence successful!

[1] "Threshold = 23"

[1] "Number of exceedances of threshold = 1050"

[1] "Exceedance rate (per year)= 36.2077511329305"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 28.796821	1.29463691
Shape (xi)	: -0.103829	0.03284874

[1] "Negative log-likelihood: 4469.26114876404"

Parameter covariance:

	[,1]	[,2]
[1,]	1.67608473	-0.03362536
[2,]	-0.03362536	0.00107904

DECLUSTERING $r=10$

Likelihood ratio test (5% level) for $x_i=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 53.23697 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 2.956419e-13
 Convergence successful!

[1] "Threshold = 23"

[1] "Number of exceedances of threshold = 171"

[1] "Exceedance rate (per year)= 5.89669089879154"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 78.0252466	6.18946899
Shape (xi)	: -0.5917317	0.05076595

[1] "Negative log-likelihood: 814.880124062991"

Parameter covariance:

	[,1]	[,2]
[1,]	38.3095263	-0.307832640
[2,]	-0.3078326	0.002577181

H3. Pos Karang Kedawung, Kabupaten Jember

TANPA DECLUSTERING

Likelihood ratio test (5% level) for $\xi=0$ does not reject Exponential hypothesis.
 likelihood ratio statistic is 1.750061 < 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 0.1858691
 Convergence successfull!

[1] "Threshold = 19"

[1] "Number of exceedances of threshold = 1032"

[1] "Exceedance rate (per year)= 35.5870468277946"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 20.02625201	0.91253040
Shape (xi)	: -0.04573256	0.03330667

[1] "Negative log-likelihood: 4077.75514763166"

Parameter covariance:

	[,1]	[,2]
[1,]	0.83271174	-0.023062942
[2,]	-0.02306294	0.001109334

DECLUSTERING $r=8$

Likelihood ratio test (5% level) for $\xi=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 27.40963 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 1.646095e-07
 Convergence successfull!

[1] "Threshold = 19"

[1] "Number of exceedances of threshold = 295"

[1] "Exceedance rate (per year)= 10.1726538897281"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 42.9709565	3.26218817
Shape (xi)	: -0.3903523	0.05379689

[1] "Negative log-likelihood: 1289.27855280701"

Parameter covariance:

	[,1]	[,2]
[1,]	10.6418717	-0.158957340
[2,]	-0.1589573	0.002894105

H4. Pos Ngale, Kabupaten Ngawi

TANPA DECLUSTERING

Likelihood ratio test (5% level) for $x_i=0$ does not reject Exponential hypothesis.
 likelihood ratio statistic is 0.7840884 < 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 0.3758937
 Convergence successful!

[1] "Threshold = 17"

[1] "Number of exceedances of threshold = 1033"

[1] "Exceedance rate (per year)= 35.6215304003021"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 21.69504101	1.01304571
Shape (xi)	: -0.03168812	0.03483267

[1] "Negative log-likelihood: 4178.89379649743"

Parameter covariance:

	[,1]	[,2]
[1,]	1.02626161	-0.026967417
[2,]	-0.02696742	0.001213315

DECLUSTERING r=8

Likelihood ratio test (5% level) for $x_i=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 31.21275 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 2.312423e-08
 Convergence successful!

[1] "Threshold = 17"

[1] "Number of exceedances of threshold = 309"

[1] "Exceedance rate (per year)= 10.6554239048338"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 48.3195500	3.54261323
Shape (xi)	: -0.3940615	0.05140022

[1] "Negative log-likelihood: 1385.50546463371"

Parameter covariance:

	[,1]	[,2]
[1,]	12.5501085	-0.164989665
[2,]	-0.1649897	0.002641983

H5. Pos Karangbinangun, Kabupaten Lamongan

TANPA DECLUSTERING

Likelihood ratio test (5% level) for $\xi=0$ does not reject Exponential hypothesis.
 likelihood ratio statistic is 0.1753394 < 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 0.6754095
 Convergence successfull!

[1] "Threshold = 14"

[1] "Number of exceedances of threshold = 1020"

[1] "Exceedance rate (per year)= 35.1732439577039"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 17.73867435	0.83916951
Shape (xi)	: 0.01464998	0.03546515

[1] "Negative log-likelihood: 3968.20645942745"

Parameter covariance:

	[,1]	[,2]
[1,]	0.70420546	-0.022053043
[2,]	-0.02205304	0.001257777

DECLUSTERING $r=8$

Likelihood ratio test (5% level) for $\xi=0$ does not accept Exponential hypothesis.
 likelihood ratio statistic is 29.25349 > 3.841459 1 df chi-square critical value.
 p-value for likelihood-ratio test is 6.35018e-08
 Convergence successfull!

[1] "Threshold = 14"

[1] "Number of exceedances of threshold = 288"

[1] "Exceedance rate (per year)= 9.93126888217523"

[1] "Maximum Likelihood Estimates:"

	MLE	Std. Err.
Scale (sigma)	: 42.4164813	3.02702378
Shape (xi)	: -0.3480247	0.04518774

[1] "Negative log-likelihood: 1267.06489141415"

Parameter covariance:

	[,1]	[,2]
[1,]	9.1628730	-0.120011073
[2,]	-0.1200111	0.002041932

LAMPIRAN I : RETURN LEVEL

11. Pos Cawak, Kabupaten Bojonegoro

TANPA DECLUSTERING

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 62.7905

0.25-year return level: 95% confidence interval approximately (60.16883,65.06326)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 73.2482

0.5-year return level: 95% confidence interval approximately (70.05289,75.79417)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 78.4827

0.75-year return level: 95% confidence interval approximately (75.00027,81.21798)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 81.8501

1-year return level: 95% confidence interval approximately (78.18299,84.75115)

DECLUSTERING $r=8$

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 64.6179

0.25-year return level: 95% confidence interval approximately (64.52205,65.30366)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 84.7191

0.5-year return level: 95% confidence interval approximately (84.58437,85.53243)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 92.5454

0.75-year return level: 95% confidence interval approximately (92.39552,94.28865)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 96.8516

1-year return level: 95% confidence interval approximately (96.69341,99.98295)

I2. Pos Pasewaran, Kabupaten Banyuwangi

TANPA DECLUSTERING

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 79.7064

0.25-year return level: 95% confidence interval approximately (76.75366,82.90201)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 95.0278

0.5-year return level: 95% confidence interval approximately (91.19794,99.54359)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 103.4922

0.75-year return level: 95% confidence interval approximately (98.96933,109.12401)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 109.2853

1-year return level: 95% confidence interval approximately (104.18094,115.85987)

DECLUSTERING $r=10$

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 50.0559

0.25-year return level: 95% confidence interval approximately (49.7372,54.1598)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 85.3174

0.5-year return level: 95% confidence interval approximately (84.58335,92.52574)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 100.1516

0.75-year return level: 95% confidence interval approximately (99.24289,107.71553)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 108.715

1-year return level: 95% confidence interval approximately (107.70538,116.17062)

13. Pos Karang Kedawung, Kabupaten Jember

TANPA DECLUSTERING

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 60.6546

0.25-year return level: 95% confidence interval approximately (58.3263,63.20009)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 73.0183

0.5-year return level: 95% confidence interval approximately (69.86586,76.75617)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 80.071

0.75-year return level: 95% confidence interval approximately (76.26764,84.80637)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 84.9963

1-year return level: 95% confidence interval approximately (80.64551,90.5789)

DECLUSTERING $r=8$

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 52.6145

0.25-year return level: 95% confidence interval approximately (50.4161,56.39494)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 70.7418

0.5-year return level: 95% confidence interval approximately (67.35823,75.2861)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 79.282

0.75-year return level: 95% confidence interval approximately (75.3403,83.94562)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 84.5718

1-year return level: 95% confidence interval approximately (80.28462,89.34063)

I4. Pos Ngale, Kabupaten Ngawi

TANPA DECLUSTERING

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 62.8333

0.25-year return level: 95% confidence interval approximately (60.23449,65.68589)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 76.7115

0.5-year return level: 95% confidence interval approximately (73.13917,80.99136)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 84.6895

0.75-year return level: 95% confidence interval approximately (80.33665,90.17653)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 90.2882

1-year return level: 95% confidence interval approximately (85.2761,96.8039)

DECLUSTERING $r=8$

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 56.2739

0.25-year return level: 95% confidence interval approximately (52.93908,60.49299)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 76.1948

0.5-year return level: 95% confidence interval approximately (71.16846,81.15538)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 85.5605

0.75-year return level: 95% confidence interval approximately (79.73898,90.63869)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 91.3543

1-year return level: 95% confidence interval approximately (85.04081,96.54773)

15. Pos Karangbinangun, Kabupaten Lamongan

TANPA DECLUSTERING

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 53.1844

0.25-year return level: 95% confidence interval approximately (50.83913,55.77992)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 65.9425

0.5-year return level: 95% confidence interval approximately (62.60816,69.96531)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 73.4658

0.75-year return level: 95% confidence interval approximately (69.33605,78.70045)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 78.8308

1-year return level: 95% confidence interval approximately (74.02711,85.10587)

DECLUSTERING $r=8$

Estimating CIs for GPD 0.25-yr. return level.

Using 365.25 days per year.

Estimated 0.25-yr. return level = 47.0653

0.25-year return level: 95% confidence interval approximately (44.33238,50.79322)

Estimating CIs for GPD 0.5-yr. return level.

Using 365.25 days per year.

Estimated 0.5-yr. return level = 66.1015

0.5-year return level: 95% confidence interval approximately (62.88781,70.88408)

Estimating CIs for GPD 0.75-yr. return level.

Using 365.25 days per year.

Estimated 0.75-yr. return level = 75.2846

0.75-year return level: 95% confidence interval approximately (71.50742,80.35419)

Estimating CIs for GPD 1-yr. return level.

Using 365.25 days per year.

Estimated 1-yr. return level = 81.0574

1-year return level: 95% confidence interval approximately (76.92687,86.30907)

LAMPIRAN J : Ilustrasi Penentuan Nilai *Threshold* dengan Metode Presentase Pada Data Curah Hujan di Pos Cawak

1. Mengurutkan data dari yang terbesar hingga terkecil

Data Curah Hujan (mm)	
sebelum <i>sort</i>	setelah <i>sort</i>
0	115
50	105
38	103
9	95
0	95
0	95
0	95
9	93
0	92
⋮	⋮
0	0
0	0
30	0
0	0

Jumlah data (N) = 10957

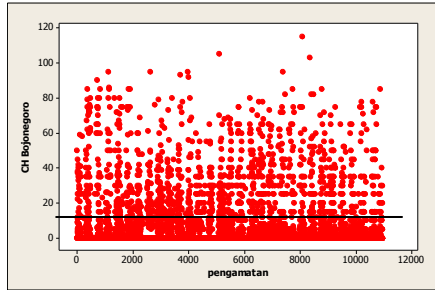
2. Menghitung 10% dari jumlah data (k)

$$\begin{aligned}
 k &= 10\% \times N \\
 &= 0,1 \times 10957 \\
 &= 1095,7 \approx 1096
 \end{aligned}$$

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

$$\begin{aligned}
 k &= 1096 \\
 (k+1) &= 1097
 \end{aligned}$$

No.	Data
1	115
2	105
3	103
4	95
5	95
6	95
7	95
8	93
9	92
⋮	⋮
1096	15
1097	15
⋮	⋮
10956	0
10957	0



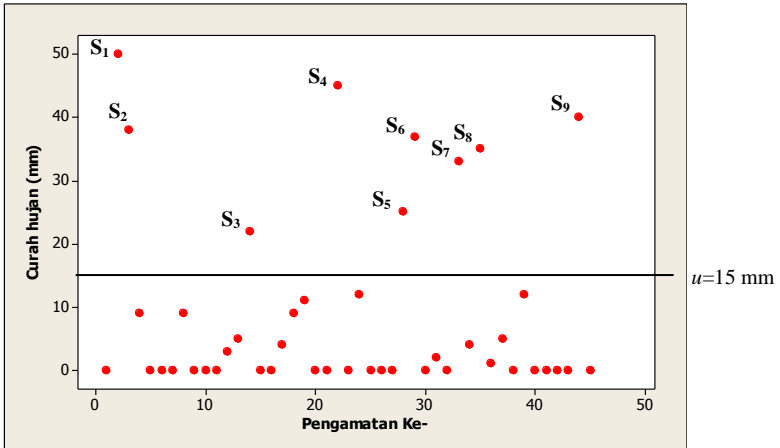
$Threshold (u) = 15 \text{ mm}$
 Semua pengamatan dengan curah hujan lebih dari 15 mm diidentifikasi sebagai nilai ekstrim.

4. Tabel frekuensi nilai ekstrim curah hujan (mm) di Pos Cawak

Curah Hujan	f	Curah Hujan	f	Curah Hujan	f	Curah Hujan	f	Curah Hujan	f
16	24	33	4	50	32	68	3	93	1
17	17	34	1	51	4	69	3	95	4
18	28	35	96	52	11	70	6	103	1
19	13	36	14	53	3	71	2	105	1
20	70	37	15	54	3	72	10	115	1
21	25	38	10	55	16	73	5		
22	12	39	13	56	2	74	1		
23	11	40	31	57	2	75	23		
24	8	41	7	58	5	76	2		
25	98	42	21	59	2	77	5		
26	13	43	8	60	27	78	2		
27	11	44	2	61	3	79	9		
28	21	45	30	62	13	82	3		
29	8	46	7	63	3	85	8		
30	72	47	1	64	2	86	1		
31	9	48	13	65	25	90	1		
32	16	49	6	67	4	92	1	Total	985

LAMPIRAN K : Ilustrasi Menghitung *Extremal Index* dan Menentukan *Run Length (r)* di Pos Cawak

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



Pada data curah hujan di Pos Cawak dengan nilai *threshold* (u) 15 mm terdapat 985 pengamatan yang berada diatas *threshold*, sehingga terdapat $S_1, S_2, S_3, \dots, S_{985}$. (Ilustrasi pada gambar hanya menampilkan sebagian data)

$S_1 = 2$	$S_5 = 28$	$S_9 = 44$
$S_2 = 3$	$S_6 = 29$	$S_{10} = 52$
$S_3 = 14$	$S_7 = 33$	\vdots
$S_4 = 22$	$S_8 = 35$	$S_{985} = 10956$

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

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

$T_1 = S_2 - S_1 = 1$	$T_5 = S_6 - S_5 = 1$	$T_9 = S_{10} - S_9 = 8$
$T_2 = S_3 - S_2 = 11$	$T_6 = S_7 - S_6 = 4$	$T_{10} = S_{11} - S_{10} = 9$
$T_3 = S_4 - S_3 = 8$	$T_7 = S_8 - S_7 = 2$	\vdots
$T_4 = S_5 - S_4 = 6$	$T_8 = S_9 - S_8 = 9$	$T_{984} = S_{985} - S_{984} = 6$

Maksimum ($T_1, T_2, T_3, \dots, T_{984}$) = 194

3. Menghitung *extremal index* (θ) dengan persamaan sebagai berikut.

$$\theta = \min \left\{ 1, \frac{2 \left(\sum_{i=1}^{N-1} (T_i - 1) \right)^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1)(T_i - 2)} \right\} = \min \{ 1 ; 0,3234 \}$$

$$= 0,3234$$

4. Menentukan *run length* (r) untuk analisis *declustering*

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

$$(C-1) = 0,3234 (985-1)$$

$$(C-1) = 318,22$$

$$C = 319,22 \approx 319$$

$$\begin{aligned} \text{Maka } \textit{run length} (r) &= T_C \\ &= T_{319} \\ &= 8 \end{aligned}$$