

# Estimasi Parameter *Generalized Pareto Distribution* Pada Kasus Identifikasi Perubahan Iklim di Sentra Produksi Padi Jawa Timur

Yustika Desi Wulan Sari dan Sutikno

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111

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

**Abstrak**—Jawa Timur memiliki potensi besar dalam sektor pertanian, khususnya produksi padi. Lima kabupaten penghasil padi terbesar di Jawa Timur adalah Kabupaten Banyuwangi, Bojonegoro, Jember, Lamongan, dan Ngawi. Produksi padi rentan terhadap keberagaman iklim terutama iklim ekstrim dan terjadinya perubahan iklim. Dalam upaya meminimalkan kerugian akibat iklim, maka perlu mempelajari karakteristik iklim ekstrim dan identifikasi adanya perubahan iklim. Penelitian ini dilakukan untuk mengetahui cara mengestimasi parameter bentuk dan skala menggunakan *Maximum Likelihood Estimation (MLE)*, mengeksplorasi karakteristik curah hujan di wilayah penelitian, dan mengidentifikasi perubahan iklim. Metode yang digunakan menganalisis curah hujan ekstrim adalah *Extreme Value Theory*. Salah satu pendekatan untuk mengidentifikasi nilai ekstrim adalah *Peaks Over Threshold* yang mengikuti distribusi *Generalized Pareto Distribution (GPD)*. Data Curah hujan dibagi menjadi dua periode yaitu periode 1 (1981-1990) dan periode 2 (1991-2010). Estimasi parameter bentuk dan skala didapatkan melalui MLE yang selanjutnya diselesaikan dengan Newton Raphson, karena menghasilkan persamaan yang tidak *closed form*. Penelitian ini menghasilkan estimasi parameter bentuk dan skala distribusi GPD, serta *confidence interval (1- $\alpha$ )100%* dengan  $\alpha$  sebesar 5%. Di samping itu disimpulkan bahwa terjadinya perubahan iklim di kelima kabupaten, khususnya pada musim kemarau dan transisi

**Kata Kunci**—*Extreme Value Theory, Generalized Pareto Distribution, Maximum Likelihood Estimation, Peaks Over Threshold, Perubahan iklim*

## I. PENDAHULUAN

Jawa Timur merupakan provinsi yang memiliki potensi besar dalam bidang pertanian. Sektor pertanian merupakan salah satu kontributor tertinggi PDRB di Jawa Timur sebesar 15,39%. Berbagai macam jenis komoditas yang diproduksi antara lain padi, jagung, kedelai, ubi, dan kacang-kacangan. Produksi terbesar di Jawa Timur adalah tanaman padi. Menurut BPS pada tahun 2011 sebesar 30,74% Jawa Timur menyumbang hasil produksi padi terhadap Jawa dan sebesar 16,09% berkontribusi terhadap nasional. Pada tahun 2011 jumlah padi yang dihasilkan sebesar 10.576.543 ton/tahun dengan luas lahan panen 1.926.796 ha (BPS, 2012). Beberapa kabupaten yang menghasilkan produksi padi dalam jumlah yang besar yaitu Banyuwangi, Ngawi, Bojonegoro, Lamongan, dan Jember. Pada tahun 2011, hasil produksi padi di Kabupaten Banyuwangi sebesar 695.962 ton, Kabupaten Ngawi sebesar

674.224 ton, Kabupaten Bojonegoro menghasilkan 675.697 ton, Kabupaten Lamongan sebesar 601.505 ton dan Kabupaten Jember sebesar 813.514 ton.

Salah satu yang mempengaruhi produksi padi adalah keragaman iklim terutama iklim ekstrim dan perubahan iklim. Menurut *World Meteorological Organization (WMO)*, iklim dapat dikatakan sebagai rata-rata cuaca atau secara ilmiah merupakan bentuk statistik deskriptif dari rata-rata dan variabilitas dalam suatu periode. Kondisi iklim Indonesia dipengaruhi oleh fenomena El Nino/La Nina dan Dipole Mode berdasarkan informasi dari BMKG. Fenomena El Nino/La Nina bersumber dari wilayah timur Indonesia dan Dipole Mode bersumber dari barat Indonesia. Keadaan curah hujan pertahun di Jawa Timur mempunyai karakteristik kurang dari 1.750 mm, diantara 1.750 hingga 2.000 mm dan lebih dari 2.000 mm.

Metode *Extreme Value Theory (EVT)* dapat digunakan dalam mengidentifikasi adanya perubahan iklim dengan melihat perubahan distribusi antar periode waktu, serta besar parameter distribusi EVT-nya. Oleh karena itu metode pendugaan parameter distribusi EVT mempunyai peran penting dalam kajian tersebut. Salah satu penelitian yang telah mengaplikasikan teori tersebut adalah Rahayu [1] yang dilakukan identifikasi perubahan iklim menggunakan metode *Block Maxima (BM)* dengan mengestimasi parameter *Generalized Extreme Value (GEV)* menggunakan *Maximum Likelihood Estimation (MLE)* dan *Probability Weighted Moments (PWM)*. Kesimpulan lain yang diperoleh penelitian tersebut adalah estimasi parameter menggunakan PWM akan menghasilkan persamaan yang *closed form*. Penelitian perubahan iklim juga dapat dilakukan dengan melakukan simulasi untuk curah hujan di Republik Ceko yang berlokasi di tengah Eropa dilakukan dengan meningkatkan nilai *threshold* yang dilakukan oleh Kyseli & Beranova [2]. Peningkatan nilai *threshold* bertujuan untuk mengestimasi *multi year return level* jumlah curah hujan.

Tujuan penelitian ini adalah mengestimasi parameter *Generalized Pareto Distribution (GPD)* menggunakan MLE, mengeksplorasi karakteristik curah hujan di sentra produksi padi, dan mengidentifikasi adanya perubahan iklim di wilayah tersebut melalui pendekatan EVT-POT.

## II. TINJAUAN PUSTAKA

Metode yang digunakan dalam penelitian ini adalah *Peaks Over Threshold*, metode persentase 10%, *Maximum Likelihood Estimation, confidence interval parameter*

100(1- $\alpha$ )% dan uji kesesuaian distribusi *Kolmogorov Smirnov*.

**A. Peaks Over Threshold**

Metode *Peaks Over Threshold* (POT) adalah metode lain EVT yang digunakan untuk mengidentifikasi nilai ekstrim. Metode POT menggunakan nilai patokan atau yang biasa disebut dengan *threshold*. Metode POT yang mengikuti distribusi *Generalized Pareto Distribution* (GPD) diperkenalkan pertama kali oleh Picklands(1975) selanjutnya dipelajari oleh Davidson (1984). Teorema Pickland-Dalkema dan de Haan menyatakan bahwa apabila semakin tinggi nilai *threshold* maka distribusinya akan mengikuti GPD

Gambar 1 menunjukkan cara pengambilan data ekstrim menggunakan POT. Data  $x_1, x_2, x_7, x_8, x_9$  dan  $x_{11}$  adalah nilai yang berada di atas *threshold* ( $u$ ), sehingga keenam data yang berada di atas *threshold* merupakan nilai ekstrim. Semakin tinggi nilai *threshold* maka data ekstrim akan semakin mengikuti distribusi General Pareto. Berikut ini adalah *Probability Density Function* GPD.

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

Dimana  $0 \leq x < \infty$  jika  $\xi \geq 0$  dan  $0 \leq x < -\sigma/\xi$  jika  $\xi < 0$ .

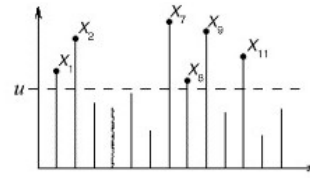
GPD memiliki dua parameter yaitu parameter bentuk ( $\xi$ ) dan parameter skala ( $\sigma$ ). Terdapat tiga tipe distribusi dalam GPD. Tipe 1 berdistribusi Eksponensial jika  $\xi = 0$ . Tipe 2 berdistribusi Pareto jika  $\xi > 0$  dan tipe 3 berdistribusi Beta jika  $\xi < 0$ .

Semakin besar nilai  $\xi$  maka distribusi akan memiliki ekor yang semakin gemuk. Sehingga peluang terjadinya nilai ekstrim pun semakin besar. Menurut Kotz & Nadarajah [3] Apabila  $\xi < 0$  maka kejadian tersebut memiliki *short tail* dan jika  $\xi > 0$  maka kejadian tersebut memiliki *long tail*.

**B. Penentuan Nilai Threshold**

Langkah awal dalam menganalisis menggunakan POT adalah menentukan nilai *threshold*. Nilai *threshold* adalah batas ambang patokan dalam menentukan nilai ekstrim. Nilai-nilai yang berada di atas *threshold* merupakan nilai ekstrim. Coles [4] menjelaskan bahwa penentuan nilai *threshold* sangat penting karena apabila *threshold* terlalu kecil maka akan mengakibatkan parameter yang bias dan apabila terlalu tinggi maka jumlah observasi semakin sedikit dan varians akan tinggi.

Terdapat beberapa cara dalam menentukan nilai *threshold* diantaranya adalah *Mean Residual Life Plot* (MRLP), metode prosentase, dan *Sample Mean Excess Function* (SMEF). Metode MRLP merupakan suatu metode dalam menentukan nilai *threshold* berdasarkan pada nilai rataan dari GPD. SMEF adalah metode lain yang menentukan nilai *threshold* dengan menampilkan grafik. SMEF adalah estimasi dari fungsi rata-rata dari kelebihan data(data ekstrim). Metode yang lebih mudah digunakan dalam penentuan *threshold* adalah metode prosentase. 10% dari data merupakan nilai kelebihan atau yang disebut dengan nilai ekstrim menurut Chaves-Dermoulin & Embrechts [5]. Langkah-langkah metode prosentase sebagai berikut.



Gambar 1 Pengambilan Sampel pada POT

Mengurutkan data dari yang terbesar hingga yang terkecil.

1. Menghitung jumlah data ekstrim  $n = 10\% \times N$  dimana  $k$  adalah jumlah data ekstrim dan  $N$  adalah jumlah sampel data. Sehingga data yang berada di urutan 1 hingga  $n$  merupakan nilai ekstrim.
2. Menentukan nilai *threshold* ( $u$ ) yaitu  $u = n + 1$ .

**C. Maximum Likelihood Estimation**

*Maximum Likelihood Estimation* (MLE) merupakan salah satu metode estimasi yang memaksimalkan fungsi likelihood untuk mendapatkan estimasi parameternya. Fungsi *likelihood* didapatkan dari perkalian *probability density function* sampel random. Persamaan fungsi *likelihood* sebagai berikut.

$$L(\xi, \sigma | x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(\xi, \sigma; x_i) \quad (2)$$

Persamaan fungsi *ln likelihood* sebagai berikut.

$$\ln L(\xi, \sigma | x_1, x_2, \dots, x_n) = \ln \prod_{i=1}^n f(\xi, \sigma; x_i) \quad (3)$$

Nilai estimasi didapatkan apabila persamaan turunan pertama membentuk persamaan yang *closed form*. Apabila persamaan yang terbentuk tidak *closed form* maka diperlukan analisis numerik lanjutan untuk penyelesaiannya.

**D. Newton Raphson**

Salah satu analisis numerik yang digunakan untuk menyelesaikan persamaan yang tidak *closed form* adalah Newton Raphson. Apabila  $g(\theta)$  adalah vektor dari turunan pertama dari  $\ln L(\xi, \sigma; x)$  dan  $H(\theta)$  adalah matriks dari turunan kedua dari fungsi  $\ln likelihood$ .  $H(\theta)$  disebut juga matriks Hessian. Persamaan umum Newton Raphson yang didapatkan dari penurunan deret Taylor sebagai berikut.

$$\theta_{l+1} = \theta_l - g(\theta_l)H^{-1}(\theta_l) \quad (5)$$

Iterasi berhenti apabila  $|\theta_{l+1} - \theta_l| < \epsilon$ .

**E. Confidence Interval Estimasi Parameter**

Setelah mendapatkan estimasi GPD, selanjutnya mencari selang kepercayaan sebagai batas atas dan batas bawah dari estimasi. *Confidence interval* dengan tingkat kepercayaan 100(1- $\alpha$ )% untuk estimasi parameter bentuk sebagai berikut.

$$\hat{\xi} - Z_{\alpha/2} \cdot SE(\hat{\xi}) \leq \xi \leq \hat{\xi} + Z_{\alpha/2} \cdot SE(\hat{\xi}) \quad (6)$$

Dimana  $SE(\hat{\xi})$  didapatkan dari  $\sqrt{var(\hat{\xi})}$ .

$$\sqrt{var(\hat{\xi})} = -E \left[ \left( \frac{\partial^2 \ln L(\xi; x)}{\partial \xi^2} \right)^{-1} \right] \quad (7)$$

**F. Uji Kesesuaian Distribusi**

Pengujian distribusi dapat dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*. Pengujian ini

dilakukan dengan menyesuaikan fungsi distribusi sampel (empirik)  $F_n(x)$  dengan distribusi teoritis tertentu  $F_0(x)$ . Pengujian hipotesis sebagai berikut.

$H_0 : F_n(x) = F_0(x)$  (data telah mengikuti distribusi teoritis  $F_0(x)$ )

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

Dimana  $F_n(x)$  adalah suatu fungsi distribusi kumulatif yang belum diketahui. Statistik uji untuk kesesuaian distribusi adalah

$$D_{hitung} = \sup |F_n(x) - F_0(x)| \quad (8)$$

Menurut Daniel [6] untuk mendapatkan kesimpulan maka membandingkan  $D_{hitung}$  dengan  $D_{1-\alpha}$  pada tabel *Kolmogorov-Smirnov* dengan taraf signifikansi ( $\alpha$ ). Tolak  $H_0$  apabila  $D_{hitung} > D_{1-\alpha}$ .

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Data ini berupa data pengamatan curah hujan harian yang diambil di setiap pos pengukuran di lima kabupaten dari tahun 1981 sampai 2010, meliputi: Kabupaten Banyuwangi, Bojonegoro, Jember, Ngawi, dan Kabupaten Lamongan.

#### B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah curah hujan harian. Dalam satu tahun akan dibagi menjadi empat triwulan dimana masing-masing triwulan mewakili musim kemarau, masa transisi kemarau-hujan, musim hujan, masa transisi hujan-kemarau. Musim hujan ditandai oleh Bulan Desember, Januari dan Februari (DJF). Bulan Maret, April, Mei (MAM) sebagai masa transisi hujan-kemarau. Musim kemarau diwakili oleh Bulan Juni, Juli, Agustus (JJA). Masa transisi kemarau-hujan diwakili oleh Bulan September, Oktober, November (SON).

#### C. Metode Analisis Data

Langkah-langkah analisis data yang dilakukan dalam penelitian ini ialah sebagai berikut :

1. Membuat diagram batang data curah hujan untuk mengetahui pola sebaran hujan di masing-masing kabupaten.
2. Mengidentifikasi adanya ekor distribusi gemuk data curah hujan harian melalui histogram.
3. Mengidentifikasi ekor distribusi gemuk data ekstrim menggunakan *heavy tail plot*.
4. Membagi data menjadi 4 triwulan yaitu DJF, MAM, JJA, SON.
5. Membagi data menjadi dua periode. Periode 1 (1981-1990) dan periode 2 (1991-2010) di masing-masing bagian.
6. Menentukan *threshold* di masing-masing triwulan di setiap periode.
7. Pengambilan data ekstrim yaitu data yang melebihi nilai *threshold*.
8. Mengestimasi parameter bentuk dan skala menggunakan MLE.
9. Membuat *confidence interval* 100(1- $\alpha$ )% parameter bentuk.

10. Melakukan uji kesesuaian distribusi menggunakan uji *Kolmogorov Smirnov*.

11. Mengidentifikasi perubahan iklim melalui estimasi parameter bentuk dan *confidence interval* nya.

### IV. HASIL DAN PEMBAHASAN

#### A. Parameter Generalized Pareto Distribution Menggunakan Maximum Likelihood Estimation

Salah satu metode mengestimasi parameter *Generalized Pareto Distribution* (GPD) adalah *Maximum Likelihood Estimation* (MLE). Cara kerja metode ini adalah memaksimalkan fungsi *likelihood* yang merupakan fungsi peluang bersama  $x_1, x_2, \dots, x_n$ .

Fungsi *likelihood* dari *probability density GPD* untuk  $\xi \neq 0$  adalah sebagai berikut.

$$L(\xi, \sigma | x_1, x_2, \dots, x_n) = \sigma^{-n} \prod_{i=1}^n \left(1 + \frac{\xi x_i}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)} \quad (9)$$

Fungsi *ln likelihood* dari persamaan 9 adalah sebagai berikut.

$$\ln L(\xi, \sigma | x_1, x_2, \dots, x_n) = -n \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi x_i}{\sigma}\right) \quad (10)$$

Langkah selanjutnya setelah mendapatkan fungsi *ln likelihood* adalah mendapatkan turunan pertama terhadap parameternya yaitu  $\xi$  dan  $\sigma$ .

$$\frac{\partial \ln L}{\partial \xi} = \frac{1}{\xi^2} \sum_{i=1}^n \ln \left(1 + \frac{\xi x_i}{\sigma}\right) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \frac{x_i}{\sigma + \xi x_i} \quad (11)$$

$$\frac{\partial \ln L}{\partial \sigma} = \sigma^{-1} \left(-n + (1 + \xi) \sum_{i=1}^n \frac{x_i}{\sigma + \xi x_i}\right) \quad (12)$$

Selanjutnya membuat persamaan turunan pertama menjadi sama dengan nol hingga terbentuk persamaan yang *closed form* untuk mendapatkan estimasi parameter sebagai berikut.

$$\hat{\xi} = \frac{\sum_{i=1}^n \ln \left(1 + \frac{\hat{\xi} x_i}{\hat{\sigma}}\right)}{(1 + \hat{\xi} - n \hat{\xi}) \sum_{i=1}^n x_i} \quad \hat{\sigma} = \frac{\sum_{i=1}^n x_i}{n^2} \quad (13)$$

Persamaan 13 merupakan persamaan yang tidak *closed form* karena masih terdapat parameter di dalam persamaan akhirnya. Salah satu penyelesaian persamaan yang tidak *closed form* adalah metode Newton Raphson.

Penggunaan metode Newton Raphson dilakukan dengan melakukan iterasi-iterasi hingga didapatkan hasil yang konvergen. Persamaan umum Newton Raphson sebagai berikut.

$$\theta_{l+1} = \theta_l - \mathbf{g}(\theta_l) \mathbf{H}^{-1}(\theta_l) \quad (14)$$

$\mathbf{g}(\theta)$  adalah vektor gradien berukuran  $1 \times p$  dimana  $p$  adalah jumlah parameter.  $\mathbf{g}(\theta)$  berisi turunan pertama *probability density function* GPD terhadap parameternya.  $\mathbf{H}(\theta)$  adalah matriks Hessian berukuran  $p \times p$  yang berisi turunan kedua terhadap parameter.

$$\mathbf{g}(\theta) = \begin{bmatrix} \frac{\partial \ln L}{\partial \xi} & \frac{\partial \ln L}{\partial \sigma} \end{bmatrix} \quad (15)$$

$$\mathbf{H}(\theta) = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \xi^2} & \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} \\ \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} & \frac{\partial^2 \ln L}{\partial \sigma^2} \end{bmatrix} \quad (16)$$

Turunan kedua dari fungsi *ln likelihood* sebagai berikut.

$$\frac{\partial^2 \ln L}{\partial \xi^2} = 2\xi^{-3} \left[ \xi \sum_{i=1}^n \frac{x_i}{\sigma + \xi x_i} - \sum_{i=1}^n \ln \left( 1 + \frac{\xi x_i}{\sigma} \right) \right] + \left( 1 + \frac{1}{\xi} \right) \sum_{i=1}^n \frac{x_i^2}{(\sigma + \xi x_i)^2} \quad (17)$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = \sigma^{-2} \left[ n - (1 + \xi) \sum_{i=1}^n \frac{x_i (2\sigma + \xi x_i)}{(\sigma + \xi x_i)^2} \right] \quad (18)$$

$$\frac{\partial^2 \ln L}{\partial \xi \partial \sigma} = \xi^{-1} \left[ (1 + \xi) \sum_{i=1}^n \frac{x_i}{(\sigma + \xi x_i)^2} - \sigma^{-1} \sum_{i=1}^n \frac{x_i}{\sigma + \xi x_i} \right] \quad (19)$$

Iterasi Newton Raphson diawali dengan menentukan nilai  $\theta_0$ .  $\theta_0$  merupakan vektor yang elemennya berisi  $\xi_0$  dan  $\hat{\sigma}_0$ . Maka nilai estimasi awal tersebut disubstitusikan pada vektor gradien dan matriks Hessian. Nilai  $\hat{\sigma}_0$  didekati dengan standar deviasi data ekstrim (*s*) sedangkan  $\xi_0$  didapatkan dari substitusi persamaan 13 untuk  $\sigma$  ke persamaan 11. Hasil substitusi dijadikan sama dengan nol. Estimasi awal parameter bentuk sebagai berikut.

$$\hat{\xi}_0 = \frac{n^2 s - \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i - n \sum_{i=1}^n x_i} \quad (20)$$

Iterasi berhenti apabila  $|\theta_{n+1} - \theta_n| < \epsilon$ .

Fungsi *Likelihood* untuk  $\xi = 0$  dari *probability density function* GPD adalah sebagai berikut.

$$L(\sigma | x_1, x_2, \dots, x_n) = \sigma^{-n} e^{-\sum_{i=1}^n \frac{x_i}{\sigma}} \quad (21)$$

Fungsi *ln likelihood* dari persamaan 21 adalah sebagai berikut.

$$\ln L(\sigma | x_1, x_2, \dots, x_n) = -n \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^n x_i \quad (22)$$

Estimasi parameter skala  $\hat{\sigma}$  diperoleh dengan membuat persamaan turunan pertama fungsi *ln likelihood* menjadi sama dengan nol.

$$\hat{\sigma} = \bar{x} \quad (23)$$

**B. Karakteristik Curah Hujan di Wilayah Penelitian**

Gambaran umum karakteristik curah hujan dalam penelitian ini menggunakan deskripsi curah hujan, diagram batang, histogram, dan *heavy tail plot* untuk masing-masing kabupaten.

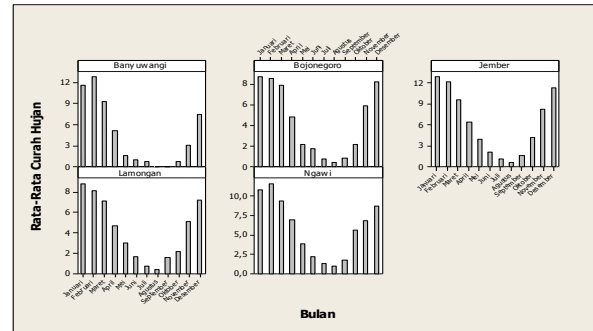
Tabel 1 menunjukkan bahwa rata-rata curah hujan tertinggi berada di Kabupaten Jember sebesar 6,144 mm/hari, sedangkan rata-rata curah hujan terendah terjadi di Kabupaten Lamongan sebesar 4,225 mm/hari. Keragaman curah hujan tertinggi terletak di Kabupaten Ngawi. Frekuensi tertinggi curah hujan terletak pada Kabupaten Ngawi yaitu sebesar 221 mm/hari.

Gambar 2 menjelaskan rata-rata curah hujan di tiap bulan dari tahun 1981 hingga tahun 2010. Pola curah hujan di kelima kabupaten berpola monsun karena berbentuk kurva U. Namun puncak curah hujan di kelima kabupaten tersebut berbeda. Puncak curah hujan di Kabupaten Banyuwangi dan Ngawi terdapat pada Bulan Februari, sedangkan ketiga kabupaten lainnya yaitu Kabupaten Bojonegoro, Jember, dan Lamongan memiliki puncak curah hujan di bulan Januari.

Histogram digunakan untuk mengetahui adanya indikasi terjadinya ekor gemuk (*heavy tail*). Apabila terdapat indikasi tersebut maka terdapat kemungkinan adanya data ekstrim.

Tabel 1.  
Nilai Rata-Rata, Standar Deviasi, Minimum dan Maksimum Curah Hujan di Tiap Kabupaten

Kabupaten	Rata-Rata (mm/hari)	St.Dev	Min (mm/hari)	Maks (mm/hari)
Banyuwangi	4,314	12,685	0	213
Bojonegoro	4,320	12,456	0	165
Jember	6,144	14,343	0	157
Lamongan	4,225	11,384	0	140
Ngawi	5,847	14,846	0	221



Gambar 2 Pola Curah Hujan di Wilayah Penelitian

Data yang diplotkan dalam histogram adalah data curah hujan harian dari tahun 1981-2010. Terdapat lima histogram yang setiap histogram menunjukkan pola untuk setiap kabupaten. Histogram kelima kabupaten ditampilkan dalam Gambar 3.

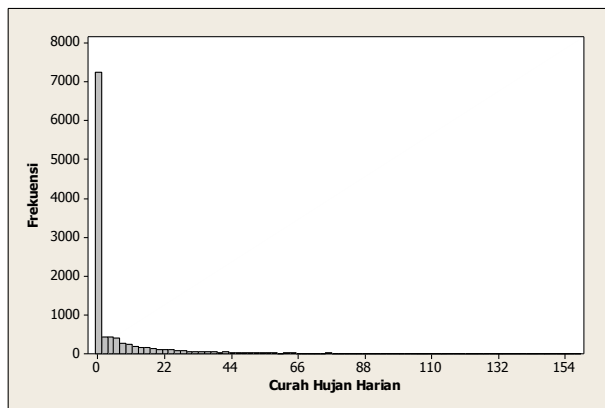
Gambar 3 adalah histogram curah hujan harian di Kabupaten Jember dari tahun 1981 hingga tahun 2010. Gambar 3 menunjukkan bahwa ekor distribusi turun secara lambat. Hal ini menunjukkan bahwa adanya indikasi *heavy tail* atau ekor gemuk di Kabupaten Jember. Histogram keempat kabupaten lainnya juga menunjukkan bahwa terjadi ekor distribusi yang turun secara lambat.

*Heavy tail* mengindikasikan adanya data ekstrim. Data ekstrim di masing-masing triwulan di tiap periode didapatkan melalui penentuan nilai *threshold*. Data ekstrim yang didapatkan diplotkan untuk mengetahui pola ekor distribusinya.

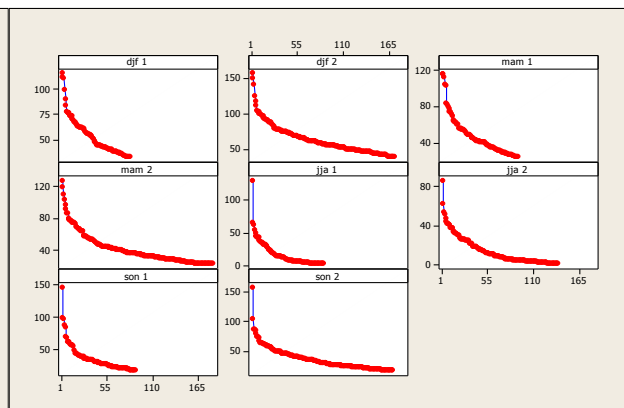
Gambar 4 adalah data curah hujan ekstrim Kabupaten Jember yang diplotkan menggunakan *time series plot* tanpa memperhatikan urutan waktu. Tabel 1 menunjukkan bahwa rata-rata curah hujan dan standar deviasi tertinggi terdapat di Kabupaten Jember, sehingga adanya kemungkinan indikasi *heavy tail* terjadi. Informasi yang didapatkan dari Gambar 4 adalah ekor distribusi data curah hujan ekstrim yang turun secara lambat di setiap triwulan di Kabupaten Jember. hal ini juga terjadi pada keempat wilayah penelitian lainnya.

**C. Penentuan Nilai Threshold**

Penentuan nilai *threshold* didapatkan melalui metode persentase 10%. Dimana data diurutkan dari yang terbesar hingga yang terkecil. 10% dari data teratas merupakan data ekstrim sehingga nilai *threshold* pun dapat ditentukan yaitu data urutan ke  $n + 1$  dimana  $n$  adalah jumlah data ekstrim.



Gambar 3 Histogram Curah Hujan Kabupaten Jember



Gambar 4 Heavy Tail Plot Curah Hujan Ekstrem Kabupaten Jember

Tabel 3 Estimasi Parameter GPD Menurut Periode dan Wilayah

Tri-wulan	Nilai	Banyuwangi		Bojonegoro		Jember		Lamongan		Ngawi	
		P1	P2	P1	P2	P1	P2	P1	P2	P1	P2
DJF	$\hat{\sigma}$	17,276	19,073	30,243	21,151	24,694	28,883	17,769	20,919	27,104	27,556
	$\hat{\xi}$	0,081	0,014	-0,489	-0,076	-0,158	-0,153	-0,096	-0,179	0,003	-0,105
	CI 90% $\xi$	-0,107 ; 0,268	-0,119 ; 0,147	-0,642 ; -0,337	-0,154 ; 0,001	-0,336 ; 0,019	-0,248 ; -0,057	-0,285 ; 0,094	-0,305 ; -0,055	-0,175 ; 0,182	-0,218 ; 0,008
	Tipe Distribusi	Pareto	Pareto	Beta	Beta	Beta	Beta	Beta	Beta	Pareto	Beta
MAM	$\hat{\sigma}$	15,442	26,584	40,469	30,686	26,671	19,167	21,367	21,076	25,810	24,777
	$\hat{\xi}$	0,249	-0,146	-0,469	-0,346	-0,125	0,011	-0,055	-0,056	-0,002	-0,001
	CI 90% $\xi$	0,044 ; 0,456	-0,250 ; -0,042	-0,628 ; -0,311	-0,435 ; -0,257	-0,296 ; 0,045	-0,122 ; 0,145	-0,262 ; 0,152	-0,164 ; 0,052	-0,163 ; 0,159	-0,103 ; 0,102
	Tipe Distribusi	Pareto	Beta	Beta	Beta	Beta	Pareto	Beta	Beta	Beta	Beta
JJA	$\hat{\sigma}$	20,218	6,379	20,203	14,201	11,117	12,804	5,926	9,904	10,494	17,303
	$\hat{\xi}$	0,125	0,566	-0,027	0,046	0,276	0,081	0,609	0,243	0,494	0,119
	CI 90% $\xi$	-0,263 ; 0,513	0,274 ; 0,857	-0,260 ; 0,207	-0,132 ; 0,223	0,024 ; 0,528	-0,099 ; 0,261	0,235 ; 0,984	0,056 ; 0,430	0,2 ; 0,789	-0,028 ; 0,268
	Tipe Distribusi	Pareto	Pareto	Beta	Pareto	Pareto	Pareto	Pareto	Pareto	Pareto	Pareto
SON	$\hat{\sigma}$	23,238	21,768	21,802	33,010	18,197	20,699	21,129	25,029	15,902	20,582
	$\hat{\xi}$	-0,108	-0,048	-0,227	-0,281	0,073	-0,034	0,031	-0,153	0,153	-0,049
	CI 90% $\xi$	-0,370 ; 0,154	-0,2 ; 0,105	-0,390 ; -0,064	-0,381 ; -0,180	-0,105 ; 0,250	-0,135 ; 0,067	-0,198 ; 0,260	-0,280 ; -0,025	-0,040 ; 0,346	-0,154 ; 0,055
	Tipe Distribusi	Beta	Beta	Beta	Beta	Pareto	Beta	Pareto	Beta	Pareto	Beta

Jumlah keseluruhan data di setiap periode disimbolkan dengan N, sehingga jumlah data ekstrim adalah 10% dari jumlah data keseluruhan. Namun, dalam prakteknya terkadang  $n$  tidak tepat 10% dari N karena adanya nilai curah hujan yang sama setiap harinya.

Tabel 2 menunjukkan nilai *threshold* di Kabupaten Jember. *Threshold* triwulan DJF untuk periode 1 adalah 32 dengan jumlah data ekstrim sebanyak 83 data dari 902 data pengamatan curah hujan harian. Nilai *threshold* pada triwulan JJA di periode 2 bernilai 0. Itu berarti bahwa pada triwulan tersebut jarang turun hujan sehingga observasi yang teramati dominan tercatat 0 mm/hari.

Langkah yang sama juga dilakukan untuk menentukan nilai *threshold* pada keempat kabupaten lainnya. Masing-masing triwulan di setiap periode memiliki nilai *threshold* yang berbeda-beda.

D. Estimasi Parameter Generalized Pareto Distribution

Setelah didapatkan data ekstrim di masing-masing triwulan di setiap periode, data tersebut diolah menggunakan program R. Tujuannya untuk mengetahui tipe distribusi masing-masing bagian di setiap periode melalui estimasi parameter. Estimasi parameter menggunakan MLE dan dilanjutkan menggunakan Newton Raphson.

Tabel 2 Nilai *Threshold* Kabupaten Jember

Triwulan	N	Periode 1	Periode 2
		threshold	n
DJF	902	32	83
	920	32	172
	1840	32	38
MAM	920	23	92
	1840	23	183
	1840	23	22
JJA	920	2	87
	1840	2	140
	1840	2	0
SON	910	16	90
	1820	16	170
	1820	16	16

Tipe distribusi ditentukan dari besarnya nilai parameter bentuk ( $\xi$ ). Apabila  $\xi < 0$  maka data curah hujan ekstrim berdistribusi Beta. Apabila  $\xi > 0$  maka berdistribusi Pareto dan jika  $\xi = 0$  maka berdistribusi Eksponensial.

Tabel 3 menunjukkan hasil  $\hat{\sigma}$  dan  $\hat{\xi}$ . Kabupaten Jember menunjukkan  $\hat{\xi}$  negatif yaitu -0,158 di periode 1 dan -0,153 di periode 2, sehingga distribusi data curah hujan ekstrim di triwulan DJF adalah Beta. Triwulan JJA menunjukkan  $\hat{\xi}$  positif yaitu 0,276 di periode 1 dan 0,081 di periode 2, sehingga berdistribusi Pareto. Namun di triwulan MAM, periode 1 dan periode 2 memiliki nilai yang berlawanan.

**Tabel 4** Uji Kesesuaian Distribusi

Kabupaten	Triwulan (Periode)	D <sub>hitung</sub>		Kabupaten	Triwulan (Periode)	D <sub>tabel</sub>	
		D <sub>hitung</sub>	D <sub>tabel</sub>			D <sub>hitung</sub>	D <sub>tabel</sub>
Banyuwangi	DJF (1)	0,061*	0,144	Lamongan	DJF (1)	0,073*	0,148
	(2)	0,047*	0,103		(2)	0,057*	0,101
	MAM (1)	0,069*	0,143		MAM (1)	0,06*	0,143
	(2)	0,042*	0,102		(2)	0,043*	0,103
	JJA (1)	0,138*	0,210		JJA (1)	0,108*	0,147
	(2)	0,061*	0,147		(2)	0,055*	0,125
Bojonegoro	SON (1)	0,124*	0,212	SON (1)	0,1*	0,145	
	(2)	0,086*	0,131	(2)	0,062*	0,101	
	DJF (1)	0,139*	0,148	DJF (1)	0,045*	0,145	
	(2)	0,08*	0,103	(2)	0,029*	0,104	
	MAM (1)	0,125*	0,144	MAM (1)	0,044*	0,142	
	(2)	0,07*	0,1	(2)	0,036*	0,102	
Jember	JJA (1)	0,12*	0,194	JJA (1)	0,081*	0,148	
	(2)	0,103*	0,128	(2)	0,061*	0,128	
	SON (1)	0,066*	0,144	SON (1)	0,054*	0,145	
	(2)	0,075*	0,101	(2)	0,048*	0,104	
	DJF (1)	0,085*	0,149	*mengikuti GPD			
	(2)	0,05*	0,104				
MAM (1)	0,042*	0,142					
(2)	0,036*	0,101					
JJA (1)	0,099*	0,146					
(2)	0,098*	0,115					
SON (1)	0,054*	0,143					
(2)	0,032*	0,104					

Nilai  $\xi$  periode 1 sebesar -0,125 dan periode 2 bernilai positif yaitu 0,011, sehingga periode 1 triwulan MAM Kabupaten Jember berdistribusi Beta dan periode 2 berdistribusi Pareto. Hal yang sama juga terjadi di triwulan SON. Nilai  $\xi$  di periode 1 positif dan periode 2 bernilai negatif, sehingga distribusi data periode 1 adalah Pareto dan Beta di periode 2. Keempat Kabupaten lainnya juga mengalami hal yang serupa. Terdapat beberapa triwulan yang memiliki tipe distribusi yang sama antar periodenya, beberapa lainnya memiliki tipe distribusi yang berbeda.

Tabel 3 menunjukkan bahwa pada Kabupaten Jember nilai  $\hat{\xi}$  periode 1 pada triwulan MAM tidak berada di dalam *confidence interval* 90% periode 2. Hal yang sama terjadi juga pada triwulan JJA. Keempat Kabupaten lainnya juga mengalami hal yang sama dimana estimasi parameter bentuk di salah satu periode tidak terdapat dalam *confidence interval* 90% di periode lainnya.

**E. Uji Kesesuaian Distribusi**

Uji *Kolmogorov Smirnov* digunakan untuk menguji distribusi data curah hujan ekstrim. Tujuannya untuk mengetahui apakah data tersebut telah mengikuti distribusi GPD.  $\alpha$  yang digunakan sebesar 5%. Hasil uji kesesuaian distribusi disajikan dalam Tabel 4.

Perhitungan uji kesesuaian distribusi menggunakan program Minitab dengan makro Minitab untuk mendapatkan  $D_{hitung}$  dan tabel *Kolmogorov Smirnov* untuk mendapatkan  $D_{tabel}$ .  $D_{tabel}$  digunakan untuk acuan  $D_{hitung}$  dalam mengambil keputusan. Perhitungan  $D_{tabel}$  dilakukan secara manual manual karena jumlah data lebih dari 100 yang tidak terdapat dalam tabel *Kolmogorov Smirnov*. Formula  $D_{tabel}$  adalah  $1,36/n$  untuk  $\alpha = 5\%$ . Tolak  $H_0$  apabila  $D_{hitung}$  lebih dari  $D_{tabel}$ .

Tabel 4 menunjukkan bahwa seluruh triwulan di tiap periode pada Kabupaten Jember telah mengikuti distribusi GPD. Hal yang sama juga terjadi pada keempat kabupaten

lainnya yaitu data curah hujan ekstrim telah mengikuti distribusi GPD.

**F. Identifikasi Perubahan Iklim**

Perubahan iklim terjadi apabila salah satu dari dua syarat berikut terpenuhi. Syarat pertama yaitu adanya perbedaan distribusi antara periode 1 dengan periode 2. Syarat kedua yaitu estimasi parameter bentuk periode 1 berada di luar *confidence interval*  $100(1-\alpha)\%$  periode 2 atau sebaliknya. Identifikasi perubahan iklim disajikan dalam Tabel 5.

Tabel 5 menunjukkan bahwa terjadi perubahan iklim di kelima kabupaten tersebut. Perubahan iklim pada triwulan DJF yang mewakili musim hujan, terjadi di Kabupaten Bojonegoro dan Ngawi. Perubahan iklim triwulan MAM yang mewakili musim transisi hujan-kemarau terjadi di tiga kabupaten yaitu Kabupaten Banyuwangi, Bojonegoro, dan Jember. Perubahan iklim di triwulan JJA (musim kemarau) terjadi di kelima kabupaten tersebut. Perubahan iklim di triwulan SON yang mewakili musim transisi kemarau-hujan terjadi di tiga kabupaten, yaitu di Kabupaten Jember, Lamongan, dan Ngawi.

**Tabel 5** Identifikasi Perubahan Iklim

Kabupaten	DJF	MAM	JJA	SON
Banyuwangi	-	√	√	-
Bojonegoro	√	√	√	-
Jember	-	√	√	√
Lamongan	-	-	√	√
Ngawi	√	-	√	√

√ = terjadi perubahan iklim

**V. KESIMPULAN DAN SARAN**

Estimasi parameter GPD untuk  $\xi \neq 0$  menggunakan MLE harus diselesaikan dengan Newton Raphson karena persamaan turunan pertama tidak *closed form*, sedangkan untuk  $\xi = 0$  turunan pertama *closed form* sehingga didapatkan estimator parameter skala. Pola curah hujan di Indonesia adalah monsun yang berbentuk kurva U dan unimodal memiliki satu puncak. Perubahan iklim terjadi di setiap kabupaten di kabupaten. Perubahan paling banyak terjadi pada musim kemarau dan musim transisi.

Adapun saran yang diberikan untuk penelitian selanjutnya berdasarkan penelitian yang telah dilakukan adalah adanya metode estimasi parameter lain yang dapat menyelesaikan persamaan *probability density function GPD* sehingga didapatkan persamaan umum  $\hat{\xi}$  dan  $\hat{\sigma}$ . Selain itu adanya penyelesaian apabila terjadi kasus dimana data yang digunakan *dependent* atau tidak random.

**DAFTAR PUSTAKA**

- [1] Rahayu, A., 2011, *Estimasi dan Pengujian Distribusi Generalized Extreme Value (GEV) (Studi Kasus : Identifikasi Perubahan Iklim di Jawa)*, Institut Teknologi Sepuluh Nopember Surabaya.
- [2] Kysely, J., dan Beranova, R., 2000, *Climate-Change Effects on Extreme Precipitation in Central Europe : Uncertainties of Scenario Based on Regional Climates Models*, *Theor Apply Climatol*, 95,361-374.
- [3] Kotz, S. dan Nadarajah, S., 2000, *Extreme Value Distributions Theory and Applications*, Imperial College Press.
- [4] Coles, S., 2001, *An Introduction to Statistical Modeling of Extreme Values*, London : Springer-Verlag.
- [5] Chaves-Dermoulin, V., dan Embrechts, P., 2002, *Smooth External Models for Operational Risk*, *Financial Valuation and Risk Management Working Paper Series*, 135.
- [6] Daniel, W. 1989. *Statistika Nonparametrik Terapan*. Jakarta: PT Gramedia.