

Analisis Curah Hujan Ekstrem pada Kasus Elevasi Tinggi Air Muka Bendungan Bilibili Sulawesi Selatan dengan Pendekatan *Peaks Over Threshold*

Indri Febriani Hartono dan Sutikno

Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)

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

Abstrak—Sulawesi Selatan merupakan provinsi di Indonesia yang dipengaruhi oleh iklim tropis basah dengan rata-rata curah hujan 289 mm per tahun dan terletak di dekat laut serta dialiri banyak sungai. Salah satu sungai di Sulawesi Selatan adalah Sungai Jeneberang. Di aliran Sungai Jeneberang terdapat bendungan yaitu Bendungan Bilibili. Beberapa tahun terakhir, bendungan mengalami permasalahan kenaikan tinggi air muka yang diakibatkan oleh tingginya curah hujan yang terjadi di sekitar bendungan. Analisis curah hujan ekstrem menggunakan pendekatan *peaks over threshold* perlu dilakukan untuk mengetahui *return level* yaitu curah hujan tertinggi pada periode ulang tertentu. Data yang digunakan merupakan data curah hujan harian Stasiun Hujan Maritim Paotere dan Meteorologi Hasanuddin pada periode 2010 sampai 2018. Hasil penelitian menunjukkan bahwa Stasiun Hujan Maritim Paotere memiliki RMSE *return level non-declustering* lebih kecil daripada *declustering*. Sedangkan untuk Stasiun Hujan Meteorologi Hasanuddin, RMSE *return level declustering* lebih kecil daripada *non-declustering*. Hal ini menunjukkan bahwa kebaikan kinerja antara metode *non-declustering* dan *declustering* masih belum bisa bedakan secara jelas. Artinya metode *declustering* memiliki kinerja dengan baik, kalau data ekstrem memiliki kasus dependensi yang kuat. Periode ulang untuk *return level* yaitu 2 tahun sampai 20 tahun dengan estimasi *return level* untuk kedua stasiun hujan melampaui 100 mm yaitu curah hujan tinggi.

Kata Kunci—Curah Hujan Harian, *Extreme Value Theory*, *Peaks Over Threshold*, *Return Level*.

I. PENDAHULUAN

SULAWESI Selatan, provinsi yang beribukota di Makassar, dilalui oleh garis ekuator atau garis khatulistiwa yang terletak pada garis lintang 00. Sulawesi Selatan merupakan salah satu provinsi di Indonesia yang dipengaruhi oleh iklim tropis basah dengan rata-rata curah hujan 289 mm per tahun. Terletak di dekat laut dan dialiri oleh banyak sungai membuat Provinsi Sulawesi Selatan banyak dipengaruhi oleh kondisi hidrologinya. Terdapat 67 aliran sungai yang mengalir Sulawesi Selatan dengan jumlah aliran terbesar di Kabupaten Luwu, yakni 25 aliran sungai [1].

Salah satu sungai yang mengalir Provinsi Sulawesi Selatan adalah Sungai Jeneberang. Sungai Jeneberang adalah sungai utama di Daerah Aliran Sungai (DAS) Jeneberang di lengan selatan Pulau Sulawesi tepatnya di lereng barat dari pegunungan Gunung Lompobattang. Sungai Jeneberang melintasi Kota Makassar, Kabupaten Maros, Kabupaten Gowa, Kabupaten Takalar, Kabupaten Jeneponto, Kabupaten Bantaeng, Kabupaten Bulukumba, Kabupaten Selayar dan

Kabupaten Sinjai. Di aliran Sungai Jeneberang terdapat sebuah bendungan yaitu Bendungan Bilibili yang berada di Desa Bili-bili Kecamatan Bontomarannu, Kabupaten Gowa. Berdasarkan profil daerah irigasi Bilibili Kabupaten Gowa oleh Dinas Pengelolaan Sumber Daya Air (PSDA) Provinsi Sulawesi Selatan, Bendungan Bilibili dibangun untuk mencegah terjadinya banjir di Kota Makassar dan sekitarnya yaitu dengan mengendalikan debit air Sungai Jeneberang dari 2.200 m³/detik menjadi 1.200 m³/detik, serta di wilayah Kabupaten Gowa akibat luapan air Sungai Jeneberang di bagian hilir, sedangkan tujuannya untuk meningkatkan kesejahteraan penduduk Kota Makassar dan sekitarnya serta untuk mensuplai air irigasi di Kabupaten Gowa. Luas daerah tangkapan waduk sebesar 384.40 km² dengan kapasitas tampungan 375 juta m³ dan kapasitas tampungan efektif 346 juta m³.

Pada Januari 2019, ratusan rumah di bantaran Sungai Jeneberang, Kabupaten Gowa, dan Makassar, Sulawesi Selatan terendam banjir. Banjir ini diakibatkan karena dibukanya pintu-pintu air Bendungan Bili-Bili. Dibukanya pintu air bendungan sebagai akibat atas melonjaknya tinggi muka air Bendungan Bili-bili akibat curah hujan ekstrem.

Metode statistika yang dikembangkan berkaitan dengan analisis kejadian ekstrem adalah Teori Nilai Ekstrem atau *Extreme Value Theory* [2]. *Extreme Value Theory* bermanfaat dalam melihat karakteristik nilai ekstrem karena berfokus pada perilaku ekor (*tail*) distribusi dalam menentukan probabilitas nilai-nilai ekstrem. *Extreme Value Theory* banyak digunakan pada kasus lingkungan [3] dan pada kasus keuangan [4]. Ada dua metode yang digunakan dalam *Extreme Value Theory* adalah *Block Maxima* (BM) dari Nilai Ekstrem Terampat atau *Generalized Extreme Value* (GEV) dan Batas Ambang Nilai Atas atau *Peaks Over Threshold* (POT) dari Distribusi Pareto Terampat atau *Generalized Pareto Distribution* (GPD). Salah satu hal penting dalam interpretasi hasil *Extreme Value Theory* adalah *return level*. *Return level* merupakan nilai maksimum yang dilampaui satu kali dalam jangka waktu tertentu. Perhitungan *return level* menggunakan parameter distribusi *Extreme Value Theory*.

Berdasarkan uraian di atas, penelitian ini bertujuan untuk mengetahui karakteristik curah hujan ekstrem di Sulawesi Selatan, yaitu khususnya di sekitar Bendungan Bilibili. Selanjutnya diketahui model EVT dan nilai *return level* curah hujan ekstrem di sekitar Bendungan dengan pendekatan *Peaks Over Threshold*.

II. TINJAUAN PUSTAKA

A. Extreme Value Theory

Extreme Value Theory (EVT) merupakan metode statistika yang tepat digunakan untuk mendapatkan *return level* dari kejadian ekstrem. Tujuan utama dari EVT adalah untuk mengetahui atau memprediksi probabilitas statistik dari kejadian yang belum pernah atau jarang diamati. Dalam kaitannya dengan klimatologi, EVT dapat memprediksi terjadinya kejadian ekstrem pada data berekor panjang (*heavy-tail*). Ekor panjang yaitu ekor distribusi turun secara lambat apabila dibandingkan dengan distribusi normal. Implikasinya adalah peluang terjadinya nilai ekstrem akan lebih besar daripada pemodelan dengan distribusi normal.

Secara umum terdapat 2 cara untuk mengidentifikasi nilai ekstrem pada data [5]. 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*.

B. Pendekatan Peaks Over Threshold

Metode pendekatan *Peaks Over Threshold* (POT) merupakan pendekatan untuk mengetahui distribusi dari pengamatan yang melewati suatu nilai yang disebut *threshold* (u) [5]. Diketahui sebuah fungsi distribusi F dari variabel acak X , kemudian diestimasi fungsi distribusi F_u dari nilai x di atas suatu nilai *threshold* u . Teorema Picklands (1975) menyatakan bahwa untuk kelas besar yang mendasari fungsi distribusi F , fungsi distribusi $F_u(y)$, untuk u besar, diperkirakan oleh persamaan berikut:

$$F_u(y) \approx G_{\xi,\sigma}(y), u \rightarrow \infty$$

dimana,

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

Untuk $y \in [0, (x_F - u)]$ apabila $\xi \geq 0$ dan $y \in \left[0, -\frac{\sigma}{\xi}\right]$ apabila $\xi < 0$. $G_{\xi,\sigma}$ merupakan *Generalized Pareto Distribution* (GPD).

C. Penentuan Nilai Threshold

Nilai *threshold* adalah batas ambang yang menjadi patokan sebuah pengamatan bernilai ekstrem. Nilai-nilai yang berada di atas *threshold* merupakan nilai ekstrem. Nilai *threshold* yang terlalu kecil dapat mengakibatkan parameter yang bias dan apabila terlalu tinggi maka jumlah observasi semakin sedikit dan varians menjadi tinggi [5].

Ada beberapa cara dalam menentukan *threshold* diantaranya adalah dengan metode *Mean Residual Life Plot* (MRLP) dan metode persentase. Metode MRLP merupakan suatu metode dalam menentukan nilai *threshold* berdasarkan pada nilai rataan GPD. Metode persentase adalah metode yang lebih mudah digunakan dalam penentuan *threshold*. Langkah-langkah metode persentase sebagai berikut.

1. Mengurutkan data dari yang terbesar hingga terkecil.
2. Menghitung jumlah data ekstrem $k = 10\% \times N$ dimana k adalah jumlah data ekstrem dan N adalah jumlah sampel

data, sehingga data yang berada di urutan 1 hingga k merupakan nilai ekstrem.

3. Menentukan nilai *threshold* yaitu $u = k+1$.

Sepuluh persen dari data merupakan nilai kelebihan atau yang disebut dengan nilai ekstrem [6].

D. Autocorrelation Function

Plot Autocorrelation Function (ACF) berguna untuk memeriksa keacakan dari suatu set data [7]. Apabila data acak, nilai ACF yang ditunjukkan pada plot ACF berkisar di sekitar nol dan apabila data tidak acak, maka satu atau lebih dari nilai ACF secara signifikan bukan nol. Pada analisis nilai ekstrem, keacakan data mengindikasikan data ekstrem tidak dependen. Nilai ACF umumnya didapatkan dengan rumus pada persamaan berikut:

$$r_k = \frac{\sum_{i=1}^{N-k} (Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^N (Y_i - \bar{Y})^2}$$

E. Declustering dan Extremal Index

Kejadian yang terjadi dan bersifat ekstrem cenderung terjadi secara berurutan atau tidak acak (dependen). Data ekstrem cenderung berkelompok, sedangkan pengelompokan atau data yang berada dalam satu *cluster* yang sama mengindikasikan adanya dependensi, sehingga perlu dilakukan *declustering*. Teknik yang umum digunakan adalah *declustering* dengan pendekatan *run declustering* yang mengasumsikan data ekstrem yang independen dipisahkan oleh sebuah angka minimum dari data ekstrem yang dependen [5]. 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 ekstrem dan menentukan *run length* (r).
2. Menentukan *cluster*. *Cluster* dimulai ketika *threshold* (u) terlampaui untuk pertama kalinya dan berakhir setelah paling sedikit r pengamatan berturut-turut berada di bawah *threshold*.
3. Mengekstrak nilai maksimum *cluster* dan melanjutkan mengidentifikasi *cluster* berikutnya (menggunakan langkah 2). Prosedur berhenti bila data habis.

Aplikasi teknik *declustering* ini adalah untuk mengestimasi nilai *extremal index*. *Extremal Index* merupakan suatu indikator untuk mengukur berapa banyak *cluster* yang terbentuk sehingga data menjadi independen. Salah satu metode dalam mengestimasi *extremal index* adalah dengan persamaan 3 berikut.

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

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

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

F. Estimasi Parameter Generalized Pareto Distribution

Setelah ditentukan nilai *threshold* u , parameter dari GPD dapat diestimasi dengan metode *Maximum Likelihood Estimation* (MLE) [5]. Diketahui GPD memiliki *probability density function* seperti berikut.

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

Berdasarkan metode MLE maka fungsi *likelihood* untuk GPD dengan $\xi \neq 0$.

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

Fungsi \ln *likelihood* dari persamaan di atas adalah berikut:

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

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

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

Sehingga fungsi \ln *likelihood* dari persamaan di atas adalah

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

Persamaan \ln *likelihood* di atas, kemudian diturunkan terhadap parameter yang akan diestimasi dan disamakan dengan nol. Dari persamaan yang terbentuk, didapatkan persamaan yang tidak *close form* sehingga dibutuhkan analisis numerik dengan cara iterasi untuk memaksimalkan fungsi \ln *likelihood*. Metode *Newton Raphson* merupakan metode yang dapat digunakan untuk menyelesaikan persamaan yang tidak *close form*. Apabila $\mathbf{g}(\boldsymbol{\theta})$ adalah vektor dari turunan pertama dari $L(\xi, \sigma; \mathbf{x})$ dan $\mathbf{H}(\boldsymbol{\theta})$ adalah matriks *Hessian* yaitu matriks dari turunan kedua dari fungsi \ln *likelihood*. Persamaan umum *Newton Raphson* diperlihatkan sebagai berikut:

$$\boldsymbol{\theta}_{l+1} = \boldsymbol{\theta}_l - \mathbf{g}(\boldsymbol{\theta}_l) \mathbf{H}^{-1}(\boldsymbol{\theta}_l)$$

Iterasi berhenti apabila $|\boldsymbol{\theta}_{l+1} - \boldsymbol{\theta}_l| < \varepsilon$ dimana ε adalah bilangan yang sangat kecil.

G. Return Level

Return level merupakan nilai maksimum pada periode yang akan datang. Pada kasus ini, nilai *return level* dari data curah hujan akan memberikan informasi waktu kembalinya hujan ekstrem agar dapat diprediksi elevasi tinggi air muka dari bendungan. Persamaan di bawah ini merupakan estimasi *return level* untuk GPD.

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}} \left((m \hat{\xi}_u \theta)^\xi - 1 \right)$$

dengan,

u : Nilai *threshold*

\hat{x}_m : Nilai ekstrem yang terjadi pada jangka waktu m pengamatan

$\hat{\xi}_u$: $\frac{n_u}{n}$, $\hat{\sigma} : \frac{n_c}{n_u}$

n_c adalah banyaknya *cluster* yang terbentuk, n_u adalah banyaknya data yang melebihi *threshold* dan n adalah banyaknya data [5].

H. Gambaran Umum Bendungan Bilibili

Bendungan Bilibili adalah bendungan terbesar di Sulawesi Selatan yang terletak sekitar 30 km di sebelah timur Kota Makassar dan ke arah hulu pertemuan Sungai Jeneberang dengan Sungai Jenelata. Bendungan ini dibangun pada tahun 1978 setelah terjadinya banjir besar yang terjadi pada tahun 1976 yang menyebabkan hampir 2/3 Kota Makassar tergenang dan diresmikan pada tahun 1999. Bendungan Bilibili merupakan bendungan dengan pengelolaan terbaik ke-3 di seluruh Indonesia. Berdasarkan Profil Daerah Irigasi Bili-Bili Kabupaten Gowa, tujuan pembuatan Bendungan Bilibili adalah untuk mencegah terjadinya banjir di Kota Makassar dan sekitarnya serta di wilayah Kabupaten Gowa akibat luapan air Sungai Jeneberang dibagian hilir, sedangkan tujuannya untuk meningkatkan kesejahteraan penduduk Kota Makassar dan sekitarnya serta untuk mensuplai air irigasi di Kabupaten Gowa.

III. METODOLOGI PENELITIAN

A. 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 Stasiun Hujan Maritim Paotere dan Stasiun Hujan Sultan Hasanuddin pada periode 2000 sampai 2019. Data training yang digunakan pada penelitian ini adalah curah hujan harian pada periode 2010 sampai 2018, dan data testing yang digunakan adalah periode tahun 2019.

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian adalah curah hujan harian Stasiun Hujan Maritim Paotere dan Stasiun Hujan Sultan Hasanuddin di Sulawesi Selatan pada periode 2000 sampai 2019. Data yang digunakan merupakan data yang berurutan waktu dan merupakan curah hujan harian.

C. Langkah Analisis

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

1) *Melakukan Pra-Pemrosesan Data*

Melakukan pra-pemrosesan data dengan mengidentifikasi *missing value*, observasi pencilan (*outlier*), dan observasi yang tidak sesuai dengan data curah hujan di dua stasiun hujan yang dapat dilihat pada Tabel 1.

2) *Mengidentifikasi Karakteristik Curah Hujan Ekstrem*

Mengidentifikasi karakteristik curah hujan ekstrem di dua stasiun hujan dengan langkah sebagai berikut.

Tabel 1.
Identifikasi *missing value* data curah hujan

Stasiun Hujan	<i>N missing</i>	<i>%missing</i>
Maritim Paotere	598	16%
Meteorologi Hasanuddin	502	13%

Tabel 2.

Nilai mean, standar deviasi, minimum, dan maksimum curah hujan (mm) di stasiun hujan Meteorologi Hasanuddin dan Maritim Paotere

Stasiun Hujan	<i>N</i>	Mean	St. Dev	Min.	Maks.
Hasanuddin	2.922	9,675	18,911	0	162,5
Maritim Paotere	3.287	11,134	19,675	0	188,7

Tabel 3.

Jumlah pengamatan (*N*), nilai *threshold* (*u*), dan jumlah sampel ekstrem (*n_u*)

Stasiun Hujan	<i>N</i>	<i>u</i> (mm)	<i>n_u</i>
Maritim Paotere	3.287	31	328
Meteorologi Hasanuddin	2.922	30	286

Tabel 4.

Nilai *threshold*, *cluster*, *extremal index*, dan *r*

Stasiun Hujan	<i>Threshold</i>	<i>Cluster</i>	<i>Extremal Index</i>	<i>r</i>
Maritim Paotere	31	62	0,22	9
Meteorologi Hasanuddin	30	60	0,22	10

- 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 ekstrem dengan histogram dan *normality plot*.

3) *Pengambilan Sampel Data Ekstrem*

Pengambilan sampel data ekstrem dengan metode *Peaks Over Threshold*. Sampel data ekstrem didapatkan dengan cara menentukan *threshold* pada data dengan menggunakan metode presentase 10% untuk mendapatkan data ekstrem. Data diurutkan dari yang terbesar hingga yang terkecil.

4) *Mengidentifikasi Data Curah Hujan*

Mengidentifikasi data curah hujan pada masing-masing pos membentuk pola siklik atau linier tren melalui plot data itu sendiri. Selain itu juga melihat data ekstrem dependen melalui plot *autocorrelation function* (ACF).

5) *Pemeriksaan Kesesuaian Distribusi*

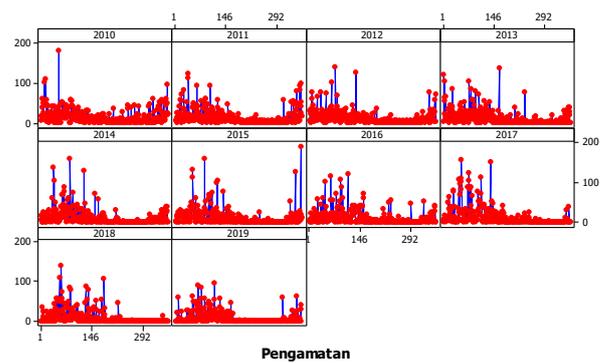
Pemeriksaan kesesuaian distribusi dapat dilakukan menggunakan *quantile plot* dan *probability plot*.

6) *Melakukan Declustering*

Melakukan *declustering* pada data ekstrem dependen melalui pendekatan *run-declustering*.

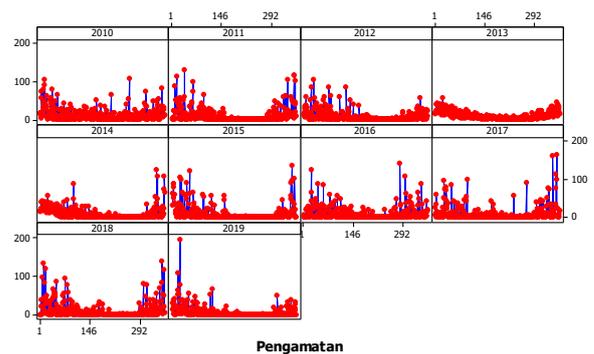
- a. Menentukan *r* (*run length*) yang optimum melalui estimasi *extremal index*.
- b. Melakukan *declustering* dengan menggunakan nilai *r* yang telah diperoleh dari hasil estimasi *extremal index* dengan langkah-langkah diantaranya; (1) Menggunakan nilai *threshold* (*u*) pada langkah ke-3 untuk mengidentifikasi nilai-nilai ekstrem; (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)

Stasiun Hujan Maritim Paotere



(a) Stasiun hujan maritim paotere

Stasiun Hujan Meteorologi Hasanuddin



(b) Stasiun hujan meteorologi hasanuddin

Gambar 1. Plot data curah hujan tahunan menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.

Mengekstrak nilai maksimum masing-masing *cluster* dan selanjutnya digunakan sebagai data baru untuk mengestimasi parameter GPD.

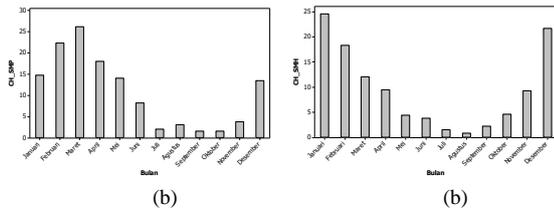
- c. Menghitung nilai estimasi parameter *Generalized Pareto Distribution* (GPD) dari data hasil *declustering* dan tanpa *declustering*.
- d. Melakukan uji hipotesis menggunakan *Kolmogorov-smirnov* terhadap data kedua stasiun hujan.
- e. Menghitung nilai *return level* curah hujan ekstrem di dua stasiun hujan kemudian membandingkan nilai RMSE untuk *return level* hasil *declustering* dan tanpa *declustering*.

IV. HASIL DAN PEMBAHASAN

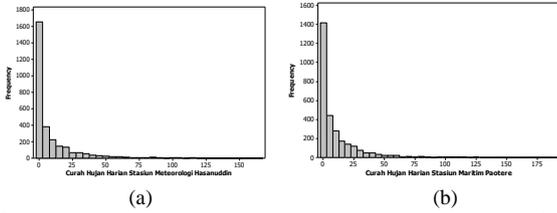
Pada bab analisis dan pembahasan ini dijelaskan mengenai pra-pemrosesan data, karakteristik data, identifikasi pola data *heavy tail* dan nilai ekstrem, pengambilan sampel ekstrem dengan pendekatan *peaks over threshold*, identifikasi dependensi data, proses *extremal index* dan *declustering*, estimasi parameter *Generalized Pareto Distribution*, uji *Kolmogorov-Smirnov*, dan *return level*.

A. *Pra-Pemrosesan Data Curah Hujan*

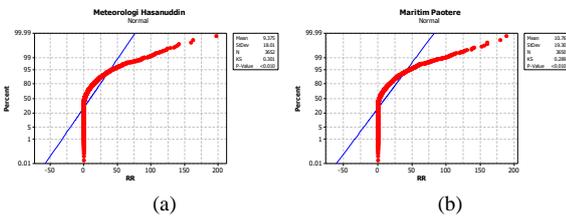
Pra-pemrosesan data sebelum dianalisis meliputi identifikasi *missing value*, imputasi data dan identifikasi data yang tidak wajar. *Missing value* atau *missing data* merupakan gangguan yang biasa ditemukan dalam data yang akan dianalisis, terutama pada data curah hujan harian yang tidak selalu lengkap. Data curah hujan harian yang didapatkan memiliki *missing value* dengan keterangan data tidak ada dan tidak dilakukan perhitungan pada hari tersebut. Berikut ini



Gambar 5. Pola curah hujan bulanan menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.



Gambar 6. Pola sebaran curah hujan harian menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.



Gambar 7. Plot normalitas data curah hujan harian menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.

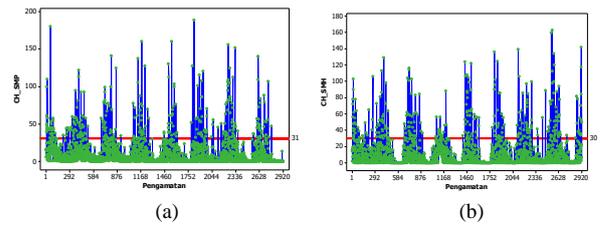
hasil identifikasi *missing value* dari data curah hujan harian di dua stasiun hujan di Sulawesi Selatan.

Data *missing* kemudian diatasi dengan melakukan imputasi dengan rata-rata curah hujan pada tanggal dan bulan yang sama dengan tanggal dan bulan data *missing* tersebut. Selanjutnya dilakukan identifikasi data yang tidak wajar. Data yang tidak wajar merupakan data curah hujan yang tidak sesuai dan tidak memungkinkan untuk terjadi. Data tidak wajar dapat terjadi dikarenakan adanya kesalahan imputasi. Identifikasi data yang tidak wajar diperlihatkan oleh plot yang menunjukkan curah hujan harian tiap tahunnya untuk kedua stasiun hujan.

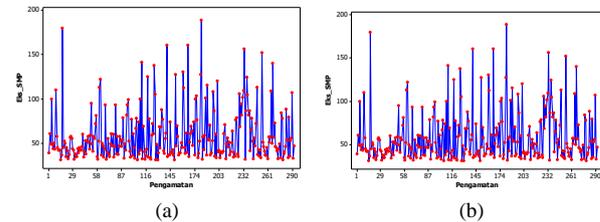
Gambar 1 menunjukkan bahwa curah hujan harian kedua stasiun hujan pada tahun 2013 menunjukkan pola yang berbeda dengan tahun lainnya. Hal ini disebabkan pada tahun 2013 seluruh data *missing* sehingga imputasi dilakukan sepanjang tahun. Berdasarkan pola, diambil kesimpulan bahwa data curah hujan harian untuk kedua stasiun hujan pada tahun 2013 dihapuskan. Selanjutnya data curah hujan harian dibagi menjadi data training dan testing untuk memudahkan penentuan model. Data training merupakan data curah hujan harian pada tahun 2010 sampai 2018 dan data testing merupakan data curah hujan harian pada tahun 2019.

B. Deskripsi Curah Hujan Harian

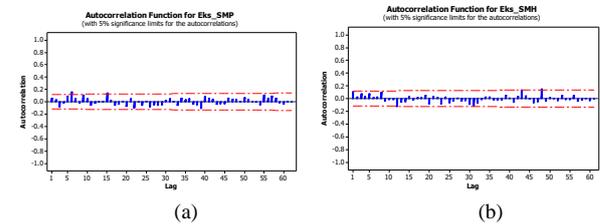
Tabel 2 menunjukkan bahwa curah hujan di Stasiun Hujan Maritim Paotere memiliki rata-rata yang lebih tinggi daripada curah hujan di Stasiun Hujan Meteorologi Hasanuddin. Demikian juga keragaman curah hujan di Stasiun Hujan Maritim Paotere memiliki keragaman curah hujan harian lebih tinggi daripada curah hujan di Stasiun Hujan Meteorologi Hasanuddin.



Gambar 2. Plot curah hujan harian dengan *threshold* menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.



Gambar 3. Plot curah hujan ekstrem menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.



Gambar 4. Plot ACF curah hujan ekstrem menurut stasiun hujan: (a) Maritim Paotere dan (b) Meteorologi Hasanuddin.

Rata-rata bulanan curah hujan harian dari kedua stasiun hujan memperlihatkan pola berbentuk huruf U dan memiliki satu puncak musim hujan (unimodal) sehingga membentuk pola monsoon yang ditunjukkan pada Gambar 2. Hasil ini sesuai dengan penelitian sebelumnya yang menyimpulkan bahwa Sulawesi Selatan berada pada region dengan pola curah hujan monsoon.

C. Identifikasi Pola Data Heavy Tail dan Nilai Ekstrem

Identifikasi data berpola *heavy tail* di Stasiun Hujan Maritim Paotere dan Meteorologi Hasanuddin dapat diketahui melalui histogram seperti pada Gambar 3.

Gambar 3 menunjukkan bahwa curah hujan di kedua stasiun hujan memiliki ekor distribusi turun lambat. Hal ini mengindikasikan adanya pola distribusi data *heavy tail* pada data curah hujan harian di Stasiun Hujan Maritim Paotere dan Meteorologi Hasanuddin. Adanya indikasi ini juga menunjukkan jika data curah hujan tidak berdistribusi normal, seperti yang disajikan pada Gambar 4. Berdasarkan plot normalitas dan *p-value* yang dihasilkan, diketahui bahwa data curah hujan harian dari kedua stasiun hujan tidak mengikuti distribusi normal. Oleh karena itu, analisis data selanjutnya menggunakan metode *extreme value theory*.

D. Pengambilan Sampel Ekstrem dengan POT

Pemilihan *threshold* dilakukan dengan metode persentase dengan persentase 10% pada data curah hujan harian yang diurutkan dari tertinggi sampai terendah di Stasiun Hujan Maritim Paotere dan Meteorologi Hasanuddin. Nilai *threshold* ditentukan dari data yang berada pada urutan $n_u + 1$, dimana n_u adalah jumlah sampel data ekstrem. Nilai *threshold* dan banyaknya data yang berada di atas *threshold*

Tabel 5.

Estimasi parameter GPD menurut stasiun hujan dan metode *non-declustering* dan *declustering*

Stasiun Hujan	Nilai	<i>Non-declustering</i>	<i>Declustering</i>
Maritim Paotere	$\hat{\xi}$	25,2890	78,3220
	$\hat{\zeta}$	0,0700	-0,4400
Meteorologi Hasanuddin	Dist. $\hat{\xi}$	Pareto 26,700	Beta 69,4240
	$\hat{\zeta}$	0,0013	-0,4900
	Dist.	Pareto	Beta

Tabel 6.

Perbandingan nilai aktual dan estimasi *return level*

Stasiun Hujan	Periode Ulang	Aktual	<i>Non-declustering</i>	<i>Declustering</i>
Maritim Paotere	3 bulan	89,1	48,67	75,095
	6 bulan	95,1	67,657	110,535
	9 bulan	95,1	79,267	126,824
	12 bulan	95,1	87,676	136,682
Meteorologi Hasanuddin	3 bulan	197	47,673	66,818
	6 bulan	197	66,353	97,218
	9 bulan	197	77,348	110,805
	12 bulan	197	85,122	118,876

pada masing-masing stasiun hujan yang ditunjukkan dalam Tabel 3.

Gambar 5 menunjukkan plot data curah hujan harian dengan garis *threshold* untuk kedua stasiun hujan dengan curah hujan ekstrem merupakan curah hujan harian yang berada di atas garis *threshold*. Diperlihatkan bahwa tiap tahunnya terhadap curah hujan yang melebihi *threshold* yang menandakan tiap tahun di kedua stasiun memiliki curah hujan harian esktrim.

E. Identifikasi Dependensi Data

Identifikasi dependensi data pada penelitian ini dilakukan dengan melihat plot dari data curah hujan ekstrem pada kedua stasiun hujan dengan pengamatan berurutan waktu mulai dari bulan pertama pengamatan yaitu Januari 2010 sampai akhir bulan Desember 2018.

Pola data curah hujan ekstrem berdasarkan Gambar 6 membentuk pola siklik yaitu pola naik dan turun di sekitar tren yang menunjukkan adanya dependensi data terhadap waktu untuk kedua stasiun hujan. Selain dengan plot data curah hujan, identifikasi dependensi data juga dilakukan menggunakan plot *autocorrelation function* dari data ekstrem curah hujan.

Berdasarkan Gambar 7, menunjukkan bahwa terdapat nilai *Autocorrelation Function* (ACF) yang melebihi batas atas fungsi autokorelasi yaitu pada Stasiun Hujan Maritim Paotere pada lag-7 dan beberapa lag setelahnya, pada Stasiun Hujan Meteorologi Hasanuddin yaitu pada lag-1. Sehingga plot *Autocorrelation Function* (ACF) curah hujan ekstrem kedua stasiun hujan memiliki korelasi yang tinggi dengan nilai ekstrem sebelum atau sesudahnya. Maka data ekstrem curah hujan di kedua stasiun hujan bersifat dependen.

Tabel 7. Nilai RMSE

Stasiun Hujan	Nilai RMSE	
	<i>Non-declustering</i>	<i>Declustering</i>
Maritim Paotere	25,949	28,151
Meteorologi Hasanuddin	128,649	100,545

Tabel 8. Nilai return level

Stasiun Hujan	Periode Ulang	Return Level	
		<i>Non-declustering</i>	<i>Declustering</i>
Maritim Paotere	2 th	108,448	156,519
	5 th	137,502	173,934
	10 th	160,751	183,153
	20 th	185,157	189,948
Meteorologi Hasanuddin	2 th	103,687	134,948
	5 th	128,254	148,235
	10 th	146,858	154,987
	20 th	165,479	159,794

F. Proses Extremal Index dan Declustering

Tabel 4 menunjukkan bahwa estimasi *extremal index* di kedua stasiun hujan adalah sama yaitu 0,22 dengan nilai *threshold*, jumlah *cluster*, dan nilai *r* yang hampir sama.

Setelah diketahui nilai *r* yang optimum, proses selanjutnya adalah melakukan declustering. Proses declustering untuk kedua stasiun hujan sebagai berikut.

1) Menentukan Threshold (*u*)

Pada analisis sebelumnya telah diketahui *threshold* untuk data curah hujan di Stasiun Hujan Maritim Paotere adalah 31 mm dan di Stasiun Meteorologi Hasanuddin adalah 30 mm.

2) Menentukan Nilai Run-Length (*r*)

Pada analisis ini nilai *r* didapatkan 9 untuk Stasiun Hujan Maritim Paotere dan 10 untuk Stasiun Meteorologi Hasanuddin.

3) Menentukan Cluster

Cluster pertama dimulai pada data pertama kali melebihi nilai *threshold* dan berakhir setelah minimal ada *r* data yang berada di bawah *threshold*.

- a. Pada data curah hujan di Stasiun Hujan Maritim Paotere *cluster* pertama terbentuk mulai pengamatan ke-1 sampai pengamatan ke-20. *Cluster* kedua mulai pengamatan ke-21 sampai pengamatan ke-31. Dimana pada *cluster* pertama dan kedua terdapat 9 pengamatan di bawah *threshold*. Pada data curah hujan di Stasiun Hujan Maritim Paotere terbentuk 62 *cluster*.
- b. Pada data curah hujan di Stasiun Hujan Meteorologi Hasanuddin, *cluster* pertama terbentuk mulai pengamatan ke-1 sampai pengamatan ke-18. *Cluster* kedua mulai pengamatan ke-19 sampai ke-37. Pada data curah hujan di Stasiun Hujan Meteorologi Hasanuddin terbentuk 60 *cluster*.

G. Estimasi Parameter GPD

Estimasi parameter menggunakan MLE kemudian dilanjutkan menggunakan Newton Rhapson. Tipe distribusi ditentukan dari besarnya nilai parameter bentuk (ξ). Apabila $\xi < 0$ maka data curah hujan ekstrem berdistribusi Beta. Apabila $\xi > 0$ maka berdistribusi Pareto dan jika $\xi = 0$ maka berdistribusi Eksponensial. Nilai estimasi parameter skala dan parameter bentuk GPD sebelum *declustering* dan

setelah dilakukan *declustering* seperti disajikan pada Tabel 5 sebagai berikut.

Tabel 5 menunjukkan hasil $\hat{\sigma}$ dan $\hat{\xi}$. Stasiun Hujan Maritim Paotere menunjukkan $\hat{\xi}$ positif sebelum dilakukan *declustering* yaitu 0,07 sehingga distribusi data curah hujan ekstrem adalah Pareto, sedangkan di stasiun hujan yang sama setelah dilakukan *declustering* estimasi parameter $\hat{\xi}$ negatif yaitu -0,44 sehingga distribusi data curah hujan ekstrem adalah Beta. Stasiun Hujan Meteorologi Hasanuddin menunjukkan $\hat{\xi}$ positif sebelum *declustering* dan $\hat{\xi}$ negatif setelah *declustering*, sehingga distribusi data curah hujan ekstrem pada Stasiun Hujan Meteorologi Hasanuddin sebelum *declustering* adalah Pareto dan setelah *declustering* adalah Beta.

H. Return Level

Nilai estimasi parameter GPD yang diperoleh digunakan untuk menghitung *return level*. *Return level* dapat digunakan untuk mengetahui nilai estimasi tertinggi pada periode waktu tertentu. Periode waktu yang digunakan adalah estimasi tahunan, hal ini dikarenakan permasalahan kenaikan elevasi di Bendungan Bilibili Sulawesi Selatan terjadi secara tahunan. Terlebih dahulu dilakukan perhitungan *return level* per-tiga bulan dari data *testing* yang kemudian akan dibandingkan dengan nilai aktual yaitu curah hujan harian tertinggi pada periode yang sama pada tahun 2019 yaitu data *training*. Perbandingan disajikan pada Tabel 6.

Penentuan metode yang sesuai dalam menghitung *return level* pada kedua stasiun hujan ditentukan dengan melihat nilai RMSE yang didapatkan dengan membandingkan hasil *return level* dengan data curah hujan harian aktual yaitu curah hujan harian tertinggi pada periode yang sama. Nilai RMSE diperlihatkan pada Tabel 7.

Return level untuk kedua stasiun hujan diduga melebihi 100 mm, yang dimana curah hujan lebih dari 100 mm dikategorikan sebagai curah hujan tinggi.

Semakin kecil nilai RMSE, maka semakin bagus model yang dihasilkan. Berdasarkan nilai RMSE yang ditunjukkan pada Tabel 8, untuk Stasiun Hujan Maritim Paotere, *return level* yang dihasilkan sebelum dilakukan *declustering* lebih baik dari setelah *declustering*. Sehingga hal ini menunjukkan bahwa nilai dugaan yang dihasilkan melalui *declustering* tidak sesuai, sehingga dalam perhitungan *return level* tidak memberikan hasil lebih baik.

Selanjutnya untuk Stasiun Hujan Meteorologi Hasanuddin, *return level* yang dihasilkan setelah *declustering* lebih bagus, hal ini menunjukkan nilai dugaan setelah *declustering* lebih mendekati nilai aktual. Nilai *return level* untuk curah hujan dalam periode tahunan di kedua stasiun hujan disajikan pada Tabel 8.

V. KESIMPULAN

Karakteristik curah hujan ekstrem dari dua Stasiun Hujan di sekitar Bendungan Bilibili Sulawesi Selatan memiliki pola *heavy tail* dan tidak berdistribusi normal. Pola curah hujan bulanan untuk kedua stasiun hujan mengikuti pola monsun. Data curah hujan harian dari kedua stasiun hujan menunjukkan adanya dependensi terhadap waktu karena memiliki nilai korelasi yang tinggi pada beberapa lag. *Declustering* dapat mengatasi dependensi data sehingga membuat data yang independen, namun pada salah satu stasiun hujan yaitu Stasiun Hujan Maritim Paotere, estimasi parameter dan *return level* yang dihasilkan setelah *declustering* tidak lebih baik dari sebelum dilakukan *declustering*. *Return level* atau curah hujan tertinggi diduga untuk periode ulang tahunan yaitu 2 tahun, 5 tahun, 10 tahun, 20 tahun. *Return level* untuk kedua stasiun hujan diduga melebihi 100 mm, yang dimana curah hujan lebih dari 100 mm dikategorikan sebagai curah hujan tinggi.

Permasalahan elevasi tinggi air muka Bendungan Bilibili dapat diminimalisir dampaknya dengan menduga curah hujan tertinggi. Analisis memberikan dugaan curah hujan tinggi yaitu lebih dari 100 mm pada setiap periode ulang, sehingga segera perlu dilakukan penanggulangan terhadap sistem bendungan agar tidak terjadi elevasi yang sangat tinggi yang dapat menyebabkan kebocoran pada bendungan.

Pengambilan kebijakan mengenai tinggi air muka Bendungan dapat dibantu dugaan curah hujan yang dihasilkan oleh analisis. Oleh karena itu pada analisis selanjutnya sebaiknya dilakukan analisis *declustering* pada stasiun hujan yang berbeda atau pos hujan yang lebih mendekati bendungan sehingga memungkinkan didapatkannya estimasi parameter dan *return level* yang optimal.

DAFTAR PUSTAKA

- [1] BPS Provinsi Sulawesi Selatan, *Sulawesi Selatan Dalam Angka 2019*. Makassar: BPS Provinsi Sulawesi Selatan, 2019.
- [2] R. Purwakianti, A. Rusgiyono, and A. Prahutama, "Aplikasi metode momen probabilitas terboboti untuk estimasi parameter distribusi Pareto terampat pada data curah hujan (Studi kasus: Data curah hujan di Kota Semarang tahun 2004-2013)," *J. Gaussian*, vol. 3, no. 4, pp. 821–830, 2014.
- [3] R. W. Katz, M. B. Parlange, and P. Naveua, "Statistics of extremes in hydrology," *Adv. Water Resour.*, vol. 25, no. 8–12, 2002.
- [4] A. J. McNeil, *Extreme Value Theory for Risk Manager*. Zurich: ETH Zentrum, 1998.
- [5] S. Coles, *An Introduction to Statistical Modelling of Extreme Values*. London: Springer-Verlag, 2001.
- [6] V. Chaves-Dermoulin and P. Lezaud, "Extreme value analysis: An introduction," *J. la Soc. Française Stat. Soc. Française den Stat. Soc. Math. Fr.*, vol. 154, no. 2, pp. 66–97, 2013.
- [7] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. New Jersey: John Wiley and Sons Inc, 2015.