

Pemodelan Bayesian Hirarki Data Curah Hujan Ekstrem di Jakarta

Jupita Sari Ika Hanugraheni dan Nur Iriawan
Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: jupita.sih@gmail.com

Abstrak— Curah hujan ekstrem merupakan kejadian yang jarang terjadi namun dapat memberikan dampak yang merugikan bagi kehidupan. Banjir merupakan salah satu dampak dari curah hujan ekstrem. Salah satu wilayah yang paling sering terkena banjir adalah Jakarta. Hal ini mengakibatkan aktivitas manusia menjadi terganggu. Oleh karena itu, dibutuhkan pengetahuan terkait nilai ekstrem untuk meminimalkan dampak kerugian akibat curah hujan ekstrem. Pada penelitian ini, identifikasi curah hujan ekstrem dilakukan dengan metode *Peaks Over Threshold (POT)* dengan pola distribusi data ekstrem mengikuti *Generalized Pareto Distribution (GPD)*. Estimasi parameter *GPD* dilakukan dengan menggunakan *Model Bayesian Hirarki (MBH)* untuk mengatasi masalah keterbatasan data pengamatan ekstrem dan mengakomodasi hubungan antar perbedaan parameter shape dengan variabel prediktor (kovariat) di setiap tingkat struktur hirarki data. Distribusi prior yang digunakan adalah *improper non conjugate* dan *non informative prior*. Hasil estimasi parameter menunjukkan bahwa modus dari temperatur rata-rata harian, elevasi, longitude, dan latitude tidak berpengaruh signifikan terhadap perbedaan nilai parameter shape, sehingga tidak berpengaruh pada hasil prediksi *return level*. Pada penelitian selanjutnya, sebaiknya menggunakan *conjugate* dan *informative prior* dengan mempertimbangkan penggunaan atau penambahan kovariat lain.

Kata Kunci— Bayesian Hirarki, Curah Hujan, Ekstrem Value, *Peaks Over Threshold*, *Return Level*

I. PENDAHULUAN

Jakarta merupakan wilayah metropolitan yang cenderung lebih hangat dibanding dengan wilayah lain disekitarnya atau sering disebut sebagai *urban heat island*. *Urban heat island* mengakibatkan curah hujan total meningkat 51 persen [1]. Tata kelola lahan yang buruk dan curah hujan yang cenderung meningkat menyebabkan Jakarta sering terkena bencana banjir. Akibat kondisi ini, aktivitas manusia menjadi terganggu sehingga setiap pengambilan keputusan perlu mempertimbangkan kondisi cuaca, terutama curah hujan ekstrem. Pengetahuan terkait nilai-nilai ekstrem untuk meminimalkan dampak kerugian akibat curah hujan ekstrem dapat dilakukan dengan memodelkan nilai ekstrem dan menentukan *return level* (nilai maksimum) dalam periode waktu tertentu, sehingga dapat digunakan sebagai upaya mitigasi bencana.

Identifikasi nilai ekstrem dapat dilakukan dengan menggunakan *Extreme Value Theory (EVT)* dengan dua pendekatan, yaitu *Block Maxima (BM)* yang mengikuti distribusi *Generalized Extreme Value (GEV)* dan *Peaks Over Threshold (POT)* yang mengikuti distribusi *Generalized Pareto Distribution (GPD)*. Pendekatan POT dapat menggunakan data pengamat-

an lebih efisien dibanding pendekatan BM [2]. Efisiensi data diperlukan pada EVT karena kejadian ekstrem merupakan suatu kejadian yang jarang terjadi, sehingga data pengamatan ekstrem yang tersedia sangat terbatas. Selain itu, dalam suatu penelitian mengenai iklim jarang dilakukan pengklasifikasian data dalam tingkat yang berbeda. Struktur data hirarki mengandung pengertian bahwa unit-unit di tingkat yang lebih rendah tersarang (*nested*) atau mengelompok dalam unit-unit di tingkat yang lebih tinggi. Data curah hujan ekstrem harian akan tersarang dalam wilayah tempat stasiun pengamatan itu berada. Data tersebut merupakan data dengan struktur hirarki dua tingkat dengan unit observasi di tingkat pertama adalah hari dimana kejadian curah hujan ekstrem terjadi dan unit di tingkat kedua adalah stasiun pengamatan. Dengan demikian, curah hujan ekstrem dapat dimodelkan sebagai hasil dari kombinasi antara karakteristik curah hujan ekstrem harian dan wilayah tempat stasiun pengamatan berada. Pengelompokan data berstruktur hirarki dapat terjadi karena adanya kesamaan dalam anggota suatu komunitas tertentu, sehingga anggota-anggota dalam suatu komunitas yang sama umumnya memiliki kondisi yang mirip (*similar*) dibandingkan anggota-anggota komunitas yang berbeda [3].

Cooley dkk. [4] melakukan penelitian mengenai EVT dengan data yang terbatas dan non stasioner dengan mengaplikasikan Model Bayesian Hirarki (MBH) menggunakan pendekatan POT pada curah hujan di Colorado, US. Namun, pada proses estimasi parameter GPD, tahap proses tidak mempertimbangkan non stasioner daerah pengamatan. Di sisi lain, Amran [5] melakukan penelitian untuk mengatasi keterbatasan data pengamatan ekstrem melalui pendekatan non stasioner daerah pengamatan dengan menggunakan Model Bayesian Hirarki Spasial Non Stasioner Regional (MBHSNSR) pada data curah hujan di Kabupaten Malang. MBHSNSR sebagai metode integrasi MBH dengan metode penaksiran regional dapat menghasilkan estimasi parameter EVT yang lebih baik dibandingkan dengan metode penaksiran secara independen.

Berdasarkan uraian tersebut, penelitian ini mengkaji karakteristik curah hujan di wilayah Jakarta, untuk selanjutnya dilakukan identifikasi curah hujan ekstrem menggunakan POT, sedangkan MBH digunakan untuk memperoleh estimasi parameter untuk perhitungan *return level* dengan mengakomodasi efek dari temperatur rata-rata harian, elevasi, longitude, dan latitude. Hasil penelitian ini diharapkan dapat memberikan tambahan informasi *climate early warning system* bagi BMKG Jakarta, masyarakat dan instansi terkait lainnya.

II. TINJAUAN PUSTAKA

A. Extreme Value Theory (EVT)

Extreme Value Theory (EVT) merupakan metode statistika yang mempelajari perilaku ekor (*tail*) distribusi untuk dapat menentukan probabilitas nilai-nilai ekstrem. Kajian nilai ekstrem dibidang iklim memberikan informasi bahwa sebagian besar data iklim memiliki distribusi yang *heavy tail*. Ada dua pendekatan untuk identifikasi pergerakan nilai ekstrem, yaitu *Block Maxima* (BM) dan *Peaks Over Threshold* (POT).

B. Peaks Over Threshold (POT)

Metode *Peaks Over Threshold* (POT) merupakan salah satu metode EVT yang mengidentifikasi nilai ekstrem dengan berdasarkan nilai ambang batas yang disebut sebagai *threshold* (u). Data yang melebihi nilai *threshold* akan diidentifikasi sebagai nilai ekstrem. Data ekstrem yang diperoleh dari metode POT akan memiliki fungsi distribusi yang mendekati *Generalized Pareto Distribution* (GPD) [6]. *Cumulative Distribution Function* (CDF) dari GPD adalah sebagai berikut.

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

dengan $y=x-u$, σ adalah parameter skala, ξ adalah parameter bentuk (*shape*) atau *tail index*. Berdasarkan nilai dari ξ , GPD dapat dibedakan menjadi tiga tipe, yaitu distribusi eksponensial jika $\xi = 0$, distribusi Pareto jika nilai $\xi > 0$, dan distribusi Pareto tipe 2/Beta jika nilai $\xi < 0$.

Ada beberapa cara dalam menentukan *threshold* antara lain dengan metode persentil data dan *Mean Residual Plot* (MR-LP). Persentil yang direkomendasikan untuk penentuan nilai *threshold* adalah sekitar 95 persen dan 99 persen dari keseluruhan data yang telah diurutkan dari yang paling besar sampai yang paling kecil [7]. Rujukan [8] mendefinisikan curah hujan ekstrem sebagai curah hujan yang berada di atas *threshold* 95 persen, sedangkan curah hujan di atas *threshold* 99 persen merupakan curah hujan sangat ekstrem.

C. Pengujian Kesesuaian Distribusi

Pengujian kesesuaian distribusi dilakukan untuk memeriksa adanya kesesuaian distribusi teoritis dengan distribusi empirik. Pengujian kesesuaian distribusi secara visual dapat dilakukan dengan menggunakan *quantile plot* dan *probability plot*, sedangkan pengujian secara formal dapat dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*. Hipotesis untuk uji kesesuaian distribusi adalah sebagai berikut.

$H_0: F_n(y) = F_0(y)$, yang berarti bahwa data mengikuti distribusi teoritis $F_0(y)$.

$H_1: F_n(y) \neq F_0(y)$, yang berarti bahwa data tidak mengikuti distribusi teoritis $F_n(y)$.

Statistik uji yang digunakan adalah.

$$D_{hitung} = \text{Sup} | F_n(y) - F_0(y) | \quad (2)$$

Daerah penolakan atau daerah kritis untuk pengujian *Kolmogorov-Smirnov* adalah H_0 akan ditolak jika $D_{hitung} > D_\alpha$ pada tabel *Kolmogorov-Smirnov* satu sampel dengan taraf signifikansi sebesar α .

D. Metode Bayesian

Metode Bayesian memanfaatkan data sampel yang diperoleh dari populasi dengan menganggap parameter θ sebagai variabel random yang memiliki distribusi. Distribusi awal untuk parameter θ disebut sebagai *prior* atau $f(\theta)$. Selanjutnya, dis-tribusi *prior* digunakan untuk menentukan distribusi *posterior* $f(\theta|y)$ berdasarkan Teorema Bayes.

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \quad (3)$$

dimana $f(y)$ adalah konstanta densitas sebagai distribusi *prior*. Konsep Bayesian memberikan ilustrasi bahwa situasi yang berbeda-beda harus dapat ditangkap dan direpresentasikan dalam pemodelan, sehingga model mampu menerangkan fenomena permasalahannya melalui nilai parameter yang berdistribusi [9].

Pemilihan distribusi *prior* merupakan hal yang sulit dalam penerapan Metode Bayes. Hal ini dikarenakan tidak ada cara baku dalam menentukan distribusi *prior* untuk suatu parameter yang tidak diketahui, namun sesuai dengan permasalahan fisik yang ada. Ada empat macam distribusi *prior* yang dikenal dalam Metode Bayesian, yaitu *conjugate prior* atau *non conjugate prior*, *proper prior* atau *improper prior*, *informative prior* atau *uninformative prior*, dan *pseudo prior*.

Istilah yang berkaitan dengan Metode Bayes adalah *Markov Chain Monte Carlo* (MCMC) dan *Gibbs Sampling*. MCMC merupakan prosedur membangkitkan data parameter sesuai dengan proses *Markov Chain* dengan menggunakan simulasi *Monte Carlo* secara iteratif sehingga diperoleh distribusi *posterior* yang stasioner (*steady state*) [10]. *Gibbs sampling* adalah teknik untuk membangkitkan variabel random dari suatu distribusi *marginal* secara langsung tanpa harus menghitung fungsi kepadatan distribusi tersebut [11].

E. Model Linier Hirarki

Model Linier Hirarki (MLH) merupakan model regresi yang mengakomodasi keberadaan data hirarki. Tahap pembentukan model hirarki dua tingkat dimulai ketika θ_j merupakan himpunan parameter GPD di lokasi j dimana $\theta_j = (\mu_j, \sigma_j, \xi_j)$ dan $\gamma = (\mu_{\theta_j}, \sigma_{\theta_j}^2)$ ialah *mean* dan *variance* distribusi Gaussian parameter θ_j . Pada pembentukan model ini diduga μ_{ξ_j} dipengaruhi X_j yang merupakan kovariat di setiap lokasi j , sehingga menyebabkan ξ_j bervariasi di setiap lokasi, sehingga dibentuk model regresi.

$$\mu_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_p X_{pj} + e_j \quad (4)$$

dengan asumsi residual bersifat identik, independen, dan berdistribusi normal (IIDN).

F. Return Level

Return level merupakan nilai maksimum yang diharapkan akan dilampaui satu kali dalam jangka waktu tertentu [11]. Penentuan *return level* pada GPD melibatkan parameter ξ , σ , dan nilai *threshold* (u). *Return level* biasanya dinyatakan dalam satuan waktu tahunan untuk berbagai keperluan.

$$RL_T = \begin{cases} u + \frac{\sigma}{\xi} [(Tn_t \xi_u - 1)], & \xi \neq 0 \\ u + \sigma \log(Tn_t \xi_u), & \xi = 0 \end{cases} \quad (5)$$

dimana $\sigma > 0$, T ialah periode waktu dalam satu tahun dan n_t adalah banyaknya pengamatan setiap tahun.

G. Ukuran Kelayakan Model

Ukuran kelayakan model adalah suatu bentuk penilaian terhadap kesesuaian model yang digunakan dengan data pengamatan yang dimiliki [5]. Pada MBH dengan struktur yang kompleks, ukuran kelayakan model dapat dilakukan melalui *Deviance Information Criteria* (DIC) dan *Mean Square Error* (MSE). Persamaan (6) digunakan untuk menghitung DIC.

$$DIC = \widehat{D}(\theta) + 2p_D \tag{6}$$

dengan $\widehat{D}(\theta)$ merupakan *deviance* yang dihitung pada rata-rata *posterior* θ dan p_D adalah banyak parameter yang efektif dalam model.

MSE dihitung berdasarkan selisih antara nilai hasil penaksiran dan nilai sebenarnya. Persamaan untuk mencari nilai MSE dapat ditulis sebagai berikut.

$$MSE = E [(\hat{\theta} - \theta)^2] \tag{7}$$

H. Curah Hujan dan Curah Hujan Ekstrem

Curah hujan adalah jumlah air yang jatuh di permukaan tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) di atas permukaan horizontal bila tidak terjadi evaporasi, *runoff*, dan infiltrasi [13]. Selain itu, Supriatna [14] mengemukakan bahwa curah hujan ekstrem memiliki curah hujan dengan intensitas >50 mm per hari menjadi parameter hujan dengan intensitas lebat, sedangkan curah hujan ekstrem memiliki curah hujan >100 mm per hari.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data temperatur rata-rata harian dan curah hujan harian tahun 1986 sampai dengan tahun 2012 oleh BMKG. Unit observasi yang digunakan dalam penelitian ini adalah sembilan pos pengamatan di wilayah Jakarta, yang terdiri dari pos pengamatan Tangerang, Serang, Curug, Tanjung Priok, Kemayoran, Halim, Soekarno-Hatta Cengkareng, Darmaga Bogor, dan Pondok Betung.

B. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini disajikan dalam Tabel 1 berikut.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Definisi Operasional
y	Curah Hujan (mm)	Jumlah air yang jatuh ke di permukaan tanah datar selama periode tertentu
x_1	Modus Temperatur Udara (°C)	Nilai temperatur udara yang sering terjadi.
x_2	Elevasi atau Ketinggian (mdpl)	Posisi vertikal dari permukaan laut.
x_3	Longitude (derajat)	Lokasi berdasarkan garis bujur
x_4	Latitude (derajat)	Lokasi berdasarkan garis lintang.

C. Langkah Analisis

- Melakukan pra-pemrosesan data.
- Mendeskripsikan data curah hujan dan pola sebaran curah hujan.
- Mengidentifikasi distribusi data curah hujan untuk mengetahui pola data yang *heavy tail* dan memuat nilai ekstrem.

- Melakukan pengambilan sampel ekstrem dengan menggunakan metode POT. Proses diawali dengan menentukan nilai *threshold* menggunakan metode MRLP dan persentil 99 persen. Setelah *threshold* ditentukan maka diambil data curah hujan yang melebihi *threshold* untuk dianalisis.
- Mengidentifikasi data curah hujan pada masing-masing pos membentuk pola siklik atau linier tren melalui plot data itu sendiri.
- Pemeriksaan kesesuaian distribusi menggunakan *quantile plot*, *probability plot*, dan pengujian hipotesis dengan uji *Kolmogorov-Smirnov*.
- Mendapatkan penaksir parameter MBH dua tingkat. Proses ini diawali dengan menetapkan variabel respon dan distribusi populasinya. Kemudian menetapkan variabel prediktor dari model makro untuk membentuk model hirarki dua tingkat. Selanjutnya menyusun *Directed Acyclic Graph* (DAG), menentukan distribusi *prior*, *hyperprior*, dan *hyperparameter*, menentukan *initial value* dan melakukan proses iterasi untuk mendapatkan taksiran parameter. Setelah diperoleh taksiran parameter, maka dilakukan pengecekan spesifikasi model.
- Melakukan perhitungan *return level* untuk beberapa periode dan menginterpretasikan hasilnya.
- Menarik kesimpulan.

IV. ANALISIS DAN PEMBAHASAN

Analisis dan pembahasan data ekstrem dilakukan dengan menggunakan metode POT dan MBH pada data curah hujan di sembilan stasiun pengamatan di wilayah Jakarta.

A. Deskripsi Data Curah Hujan

Tahap pra-pemrosesan data dilakukan dengan melakukan identifikasi *missing value*, observasi pencilan (*outlier*), dan observasi yang *inconsistent*. Pada data pengamatan iklim *missing value* diisi dengan kode 9999, sedangkan kode 8888 pada data curah hujan merupakan indikasi terjadi hujan dengan nilai kurang dari 1 milimeter atau dikenal sebagai *trace* atau tidak terukur (TTU).

Tabel 2. Identifikasi *Missing Value* dan TTU Data Curah Hujan

Stasiun Pengamatan	TTU	% TTU	Missing Value	% Missing Value
Tangerang	50	0,51	0	0
Serang	896	9,09	0	0
Curug	1052	10,67	0	0
Tanjung Priok	298	3,02	0	0
Kemayoran	595	6,03	0	0
Halim	1601	16,23	304	3,18
Soekarno-Hatta	178	1,80	0	0
Cengkareng				
Darmaga Bogor	730	7,40	0	0
Pondok Betung	561	5,69	0	0

Semua stasiun pengamatan di wilayah Jakarta memiliki nilai TTU, hanya ada satu stasiun yang memiliki *missing value*, yaitu stasiun pengamatan Halim (Tabel 2). Selanjutnya, dilakukan analisis dengan menggunakan statistika deskriptif untuk mendapatkan gambaran secara umum data curah hujan yang digunakan.

Seperti yang ditunjukkan pada Tabel 3, stasiun pengamatan Serang merupakan stasiun pengamatan dengan nilai curah hujan rata-rata harian tertinggi, yaitu sebesar 10,631 mm per

hari. Namun, stasiun pengamatan Halim merupakan stasiun pengamatan dengan curah hujan harian tertinggi, yaitu sebesar 356 mm per hari.

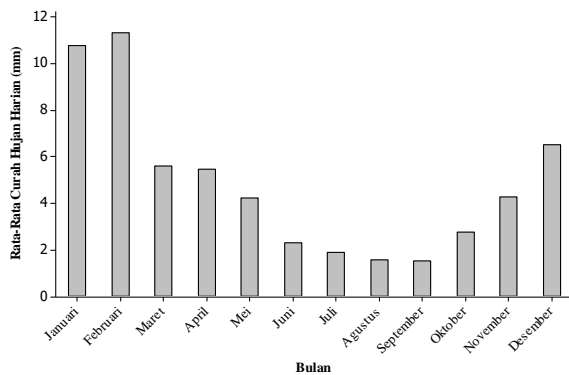
Tabel 3.
Statistika Deskriptif Curah Hujan Harian (mm)

Stasiun Pengamatan	Rata-Rata	Variance	Min	Median	Maks
Tangerang	4,805	162,381	0	0	286
Serang	10,631	366,185	0	1,3	240
Curug	6,421	228,321	0	0	339,8
Tanjung Priok	6,2	193,205	0	0	227,5
Kemayoran	6,618	254,344	0	0	305
Halim	5,009	195,726	0	0	356
Soekarno-Hatta Cengkareng	4,472	119,053	0	0	161
Darmaga Bogor	4,512	166,689	0	0	316,3
Pondok Betung	4,678	192,202	0	0	215,6

Keterangan:

Min= Minimum Maks= Maksimum

Berikut ini identifikasi untuk mengetahui pola curah hujan harian dengan menggunakan *bar chart*.



Gambar 2. Pola Curah Hujan Harian di Stasiun Pengamatan Tangerang

Seperti ditunjukkan pada Gambar 2 dapat diketahui bahwa pola curah hujan harian di stasiun pengamatan Tangerang adalah pola hujan monsun dimana pola hujan membentuk huruf U. Pola curah hujan di stasiun pengamatan lain di wilayah Jakarta juga memiliki pola hujan monsun dimana puncak curah hujan terjadi pada Bulan Desember, Januari, dan Februari.

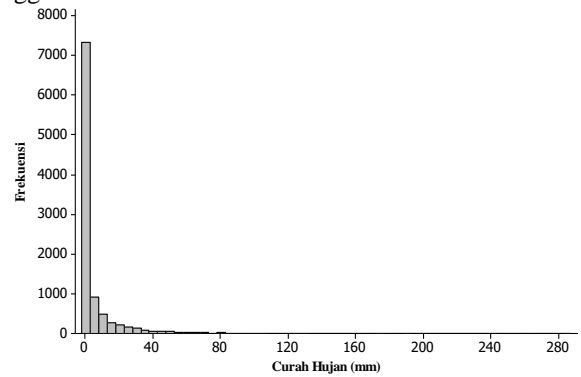
Pola distribusi data curah hujan diidentifikasi dengan nilai *skewness* dan *kurtosis*.

Tabel 4.
Skewness dan Kurtosis

Stasiun Pengamatan	Skewness	Kurtosis
Tangerang	5,45	54,12
Serang	4,24	25,98
Curug	3,88	23,16
Tanjung Priok	5,36	41,01
Kemayoran	6,39	78,38
Halim	4,75	39,63
Soekarno-Hatta Cengkareng	5,97	64,84
Darmaga Bogor	2,82	10,89
Pondok Betung	4,70	43,01

Tabel 4 memberikan informasi bahwa nilai *skewness* untuk masing-masing stasiun pengamatan di wilayah Jakarta bernilai lebih dari 0 (nol). Hal ini berarti bahwa distribusi curah hujan tidak simetris sehingga asumsi distribusi Gaussian pada data curah hujan tidak dapat digunakan. Hal ini juga didukung dari *histogram* curah hujan di stasiun pengamatan Tangerang memiliki ekor distribusi *heavy tail* (Gambar 3). Demikian juga

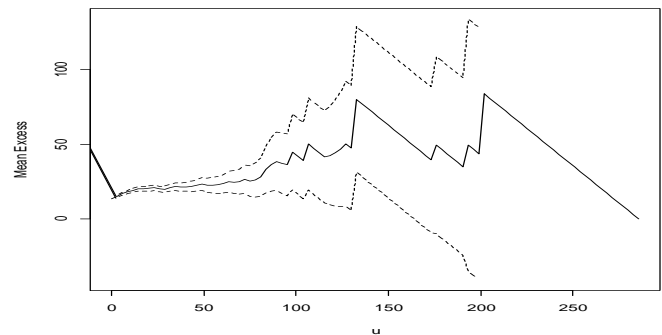
dengan *histogram* curah hujan harian di delapan stasiun pengamatan lain, sehingga dilakukan analisis lebih lanjut dengan menggunakan EVT.



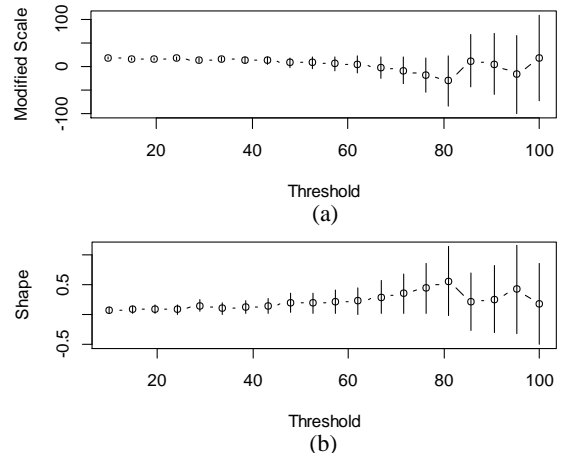
Gambar 3. Histogram Curah Hujan Harian di Stasiun Pengamatan Tangerang

B. Pengambilan Sampel Ekstrem dengan Metode POT

Penentuan nilai *threshold* dilakukan dengan menggunakan MRLP, berikut ini MRLP untuk data curah hujan di stasiun pengamatan Tangerang.



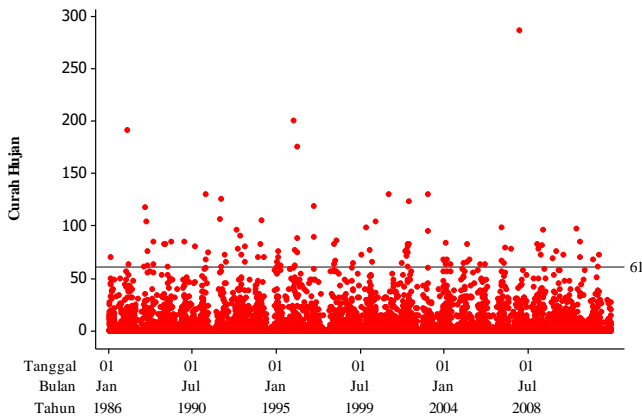
Gambar 4. MRLP Curah Hujan Harian di Stasiun Pengamatan Tangerang
Sumbu vertikal pada Gambar 4 menunjukkan nilai *mean excess* dan sumbu horizontal merupakan nilai *threshold* (u). Grafik MRLP cenderung berpola linier pada interval [10, 100]. Demikian juga grafik *modified scale* dan *shape parameter plot* (Gambar 5).



Gambar 5. (a) *Modified Scale* dan (b) *Shape Parameter Plot* Curah Hujan di Stasiun Pengamatan Tangerang

Oleh karena nilai *threshold* dengan metode MRLP dapat dianalisis keakuratannya melalui *diagnostic plot*, namun karena nilai yang dihasilkan berupa interval dan penentuan suatu nilai *threshold* bergantung subjektivitas peneliti, maka dilaku-

kan pertimbangan dengan menggunakan metode persentil, yaitu dengan cara mengambil *threshold* satu persen dari keseluruhan data yang telah diurutkan dari nilai terbesar sampai dengan nilai terkecil.



Gambar 6. Pengambilan Sampel Ekstrem di Stasiun Pengamatan Tangerang

Tabel 5 berikut ini menunjukkan nilai *threshold* dan banyak data curah hujan ekstrem harian di masing-masing stasiun pengamatan di wilayah Jakarta.

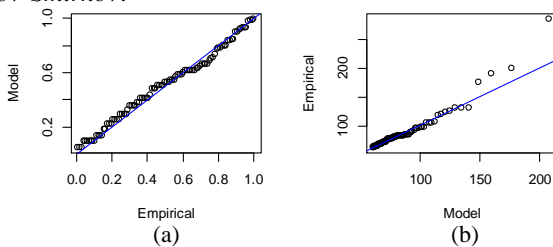
Tabel 5. Nilai *Threshold*

Stasiun Pengamatan	N	u (mm)	n_u
Tangerang	9.862	61,0	97
Serang	9.862	52,0	97
Curug	9.862	68,1	99
Tanjung Priok	9.862	71,0	99
Kemayoran	9.862	65,0	99
Halim	9.862	74,0	98
Soekarno-Hatta Cengkareng	9.862	61,6	99
Darmaga Bogor	9.862	87,0	98
Pondok Betung	9.862	71,3	99

Keterangan:

N= Banyak Pengamatan u = Nilai *Threshold* n_u = Banyak sampel ekstrim

Pemeriksaan distribusi secara visual dilakukan dengan menggunakan *probability plot* dan *quantile plot*, sedangkan pengujian formal dilakukan dengan menggunakan uji *Kolmogorov-Smirnov*.



Gambar 7. (a) *Probability Plot* dan (b) *Quantile Plot* GPD di Stasiun Pengamatan Tangerang

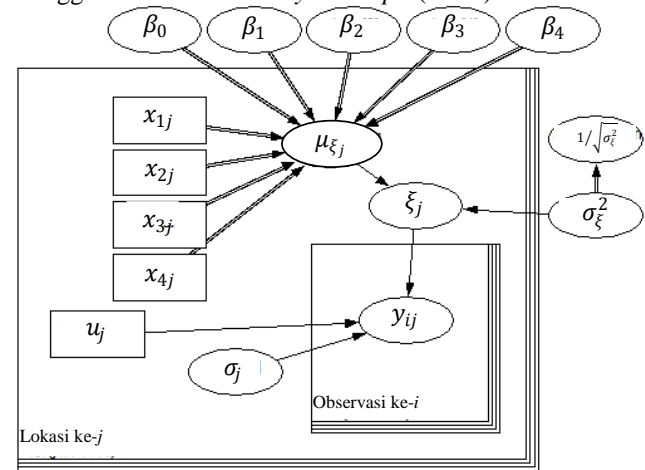
Secara visual data curah hujan ekstrem harian di wilayah Jakarta mengikuti GPD (Gambar 7). Uji *Kolmogorov-Smirnov* dilakukan dengan hipotesis nol menyatakan bahwa data curah hujan ekstrem mengikuti GPD pada taraf signifikansi $\alpha=0,05$. Seperti yang ditunjukkan pada Tabel 6, semua stasiun pengamatan memiliki nilai D_{hitung} yang lebih kecil dibandingkan dengan D_α dan $p-value$ yang lebih dari α sehingga gagal tolak H_0 dengan kesimpulan semua data curah hujan ekstrem harian di sembilan stasiun pengamatan di Jakarta mengikuti GPD. Sehingga proses analisis EVT dapat dilanjutkan.

Tabel 6. Uji *Kolmogorov-Smirnov* GPD

Stasiun Pengamatan	D_{hitung}	D_α	$p-value$	Keputusan
Tangerang	0,07778	0,13606	0,57328	Gagal Tolak H_0
Serang	0,06759	0,13606	0,74149	Gagal Tolak H_0
Curug	0,05786	0,13469	0,87549	Gagal Tolak H_0
Tanjung Priok	0,06465	0,13469	0,77779	Gagal Tolak H_0
Kemayoran	0,05702	0,13469	0,88600	Gagal Tolak H_0
Halim	0,07022	0,13537	0,69262	Gagal Tolak H_0
Soekarno-Hatta	0,06068	0,13469	0,83751	Gagal Tolak H_0
Cengkareng				
Darmaga Bogor	0,03579	0,13537	0,99920	Gagal Tolak H_0
Pondok Betung	0,08290	0,13469	0,47900	Gagal Tolak H_0

C. *Estimasi Parameter GPD dan Prediksi Return Level*

Pembentukan struktur parameter berdasarkan bentuk fungsi rangkaian variabel prediktor yang divisualisasikan dengan menggunakan *Directed Acyclic Graph* (DAG).



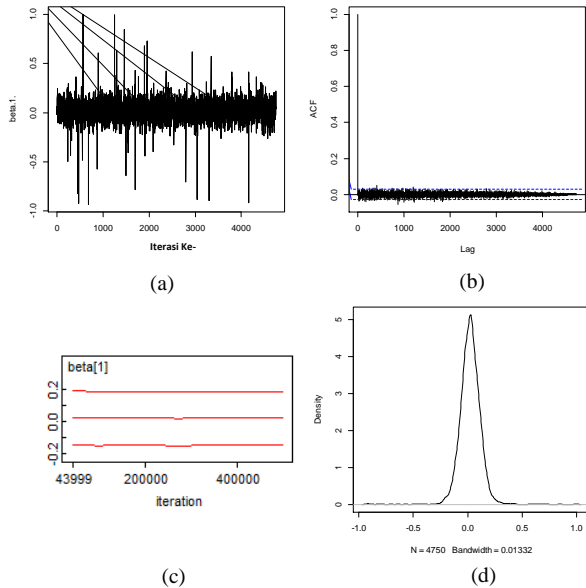
Gambar 8. DAG MBH Dua Tingkat Data Curah Hujan Ekstrem

Seperti ditunjukkan Gambar 8, diketahui y merupakan data curah hujan ekstrem dimana y ada sebanyak i untuk masing-masing stasiun pengamatan j . Berdasarkan hasil pengujian kesesuaian distribusi, curah hujan ekstrem di masing-masing stasiun (y_{ij}) berdistribusi GPD dengan tiga parameter, yaitu u , σ , dan ξ . Sehingga himpunan parameter GPD untuk masing-masing stasiun ke- j dapat ditulis sebagai $\theta_j = (\mu_j, \sigma_j, \xi_j)$. Parameter GPD untuk data curah hujan ekstrem harian ini merupakan parameter yang merepresentasikan parameter dari data curah hujan ekstrem di masing-masing stasiun pengamatan di wilayah Jakarta.

Nilai parameter u untuk masing-masing stasiun pengamatan adalah sama dengan nilai *threshold*. Sedangkan nilai ξ yang berbeda-beda di masing-masing stasiun pengamatan diindikasikan terjadi akibat adanya faktor-faktor lain, seperti x_1 (modus temperatur rata-rata), x_2 (elevasi/ketinggian), x_3 (*longitude*), dan x_4 (*latitude*). Sehingga ξ memiliki parameter $\gamma_{\xi_j} = (\mu_{\xi_j}, \sigma_{\xi_j}^2)$ yang merupakan *mean* dan *variance* distribusi *Gaussian* parameter ξ untuk stasiun ke- j . Dengan demikian, perbedaan nilai ξ untuk masing-masing stasiun pengamatan yang digambarkan sebagai μ_{ξ_j} merupakan fungsi regresi dari $x_1, x_2, x_3,$ dan x_4 . Parameter $\gamma_{\xi_j}, \beta_0, \beta_1, \beta_2, \beta_3$ dan β_4 merupakan parameter di tingkat ke dua atau disebut sebagai *hyperparameter* yang merupakan parameter distribusi *prior* ξ dan parameter regresi dari parameter *prior* ξ . Parameter *shape* (ξ) merupakan parame-

ter yang menyatakan bentuk ekor distribusi GPD (*index tail*), sehingga nilai ξ dapat menjelaskan pola perilaku pengamatan ekstrem di masing-masing stasiun pengamatan.

Penelitian ini menggunakan penentuan distribusi *prior non conjugate* dan *non informative* untuk parameter $\beta_0, \beta_1, \beta_2, \beta_3$ dan β_4 . Nilai taksiran parameter dalam MBH diperoleh dari distribusi *posterior* dimana proses pengambilan sampel dilakukan sebanyak 500.000 iterasi setelah proses *burn in* 25.000 iterasi dengan proses *thining* pada 100 iterasi untuk mendapatkan sampel parameter dari distribusi *posterior*.



Gambar 9. Plot Diagnostik Kekonvergenan (a) *Serial Plot*, (b) *Autocorrelation Plot*, (c) *Running Quantiles Plot*, dan (d) *Probability Density Plot*

Berdasarkan Gambar 9 diketahui bahwa parameter bangunan memenuhi sifat *Markov Chain*, yaitu *strongly ergodic* dan distribusi *posterior* telah mencapai kondisi *equilibrium* sebab Gambar 9 (a) dan (b) tidak menunjukkan pola tertentu. Selanjutnya dilakukan pengujian parameter hasil estimasi dengan menggunakan MBH seperti yang tertera pada Lampiran 1. Hal ini telah menjamin bahwa sampel yang diperoleh dari proses MCMC tersebut merupakan sampel dari distribusi *posterior* target parameter. Berdasarkan estimasi parameter dengan MBH diperoleh kesimpulan bahwa semua kovariat yang digunakan tidak signifikan mempengaruhi parameter ξ karena *interval credible* memuat nilai 0 (nol).

Validasi model dilakukan untuk mengetahui ketepatan model yang dihasilkan mampu *fit* terhadap data yang ada dan memenuhi asumsi dari distribusi yang digunakan dalam penentuan model. Pada penelitian ini, ketepatan model diukur dengan menggunakan DIC, MSE, dan SE. Berdasarkan hasil analisis diperoleh nilai DIC, MSE dan SE untuk model ini masing-masing adalah sebesar 7417, 196 dan 14.

Nilai estimasi parameter yang diperoleh dari pendekatan metode Bayesian Hirarki digunakan untuk menghitung nilai *return level*. Lampiran 2 memberikan informasi mengenai *return level*, yaitu suatu level kejadian ekstrem yang akan terlampaui rata-rata sekali dalam suatu periode waktu tertentu dimana probabilitas terjadinya sebesar $1/T$. Nilai *return level*

di stasiun pengamatan Tangerang pada periode satu tahun ke depan adalah sebesar 89,5 mm. Hal ini berarti bahwa ada kemungkinan 1/1 atau pasti terjadi curah hujan ekstrem yang melebihi level 89,5 mm rata-rata satu kali pada periode satu tahun. Demikian juga interpretasi untuk periode lima tahun, sepuluh tahun, dan lima belas tahun.

V. KESIMPULAN

Karakteristik curah hujan di wilayah Jakarta memiliki pola *heavy tail*. Pola ini mengindikasikan adanya data curah hujan ekstrem di wilayah Jakarta. Perilaku ekstrem ini selanjutnya diidentifikasi dengan menggunakan *threshold*. Hasil pengambilan nilai *threshold* memberikan nilai yang bervariasi untuk se-tiap wilayah. Dengan demikian, nilai parameter *scale* dan *sha-pe* juga akan bervariasi di setiap stasiun pengamatan. Nilai *shape* yang berbeda-beda di masing-masing stasiun pengamatan diindikasikan terjadi akibat adanya faktor-faktor eksternal. Berdasarkan hasil estimasi MBH diperoleh kesimpulan bahwa modus temperatur rata-rata, elevasi/ketinggian, *longitude*, dan *latitude* tidak berpengaruh signifikan terhadap parameter *sha-pe*, sehingga tidak mempengaruhi perhitungan nilai *return level*.

Pada penelitian selanjutnya dapat mempertimbangkan penggunaan atau penambahan kovariat lain dengan distribusi *conjugate prior* dan *informative prior* agar proses analisis Bayesian untuk memperoleh estimasi parameter dapat dilakukan lebih sederhana. Selain itu, hubungan antara parameter *shape* dan kovariat yang digunakan dapat diuji terlebih dahulu untuk mengetahui pola hubungannya linier atau tidak.

LAMPIRAN

Lampiran 1. Estimasi Parameter MBH

Parameter	Rata-Rata	Standar Deviasi	MC Error	2,50%	Median	97,50%
$\hat{\beta}_0$	-3,08	5,25	0,14	-13,22	-3,09	7,26
$\hat{\beta}_1$	0,02	0,08	0,00	-0,14	0,02	0,19
$\hat{\beta}_2$	-0,55	0,00	0,01	-3e-3	0,00	0,00
$\hat{\beta}_3$	-0,89	0,03	0,59	-0,05	0,00	0,05
$\hat{\beta}_4$	-0,45	0,57	0,01	-1,56	-0,46	0,71
$\hat{\sigma}_1$	20,82	2,83	0,02	15,76	20,65	26,87
$\hat{\sigma}_2$	18,42	2,60	0,02	13,76	18,26	23,93
$\hat{\sigma}_3$	15,41	2,08	0,01	11,70	15,28	19,84
$\hat{\sigma}_4$	24,30	3,40	0,02	18,32	24,08	31,55
$\hat{\sigma}_5$	23,54	3,45	0,02	17,38	23,34	30,88
$\hat{\sigma}_6$	22,96	3,20	0,02	17,28	22,76	29,83
$\hat{\sigma}_7$	26,62	3,31	0,01	20,72	26,41	33,68
$\hat{\sigma}_8$	17,00	2,55	0,01	12,45	16,84	22,44
$\hat{\sigma}_9$	22,03	2,83	0,01	16,98	21,85	28,08
$\hat{\xi}_1$	0,16	0,08	0,84	0,03	0,15	0,33
$\hat{\xi}_2$	0,08	0,10	0,00	-0,10	0,07	0,29
$\hat{\xi}_3$	0,18	0,08	0,00	0,04	0,18	0,36
$\hat{\xi}_4$	0,16	0,10	0,00	-0,03	0,15	0,37
$\hat{\xi}_5$	0,23	0,09	0,82	0,08	0,22	0,42
$\hat{\xi}_6$	0,23	0,08	0,89	0,07	0,22	0,40
$\hat{\xi}_7$	0,12	0,07	0,65	-0,01	0,12	0,27
$\hat{\xi}_8$	0,18	0,11	0,71	-0,01	0,17	0,43
$\hat{\xi}_9$	0,19	0,07	0,64	0,05	0,18	0,33

Keterangan Stasiun

No.	Stasiun Pengamatan
1	Tangerang
2	Serang
3	Curug
4	Tanjung Priok
5	Kemayoran
6	Halim
7	Soekarno-Hatta Cengkareng
8	Darmaga Bogor
9	Pondok Betung

[11] G. Casella dan E. I. George, "Explaining the Gibbs Sampler," *Journal of the American Statistical Association* 46(3) (1992) 167-174
 [12] M. Gilli dan E. Kellezi. *An Application of Extreme Value Theory for Measuring Risk*. Amsterdam: Elsevier Science (2003).
 [13] Handoko. "Klimatologi Dasar," Jakarta: Pustaka Jaya (1994).
 [14] Supriatna, J. "Diktat Diklat Meteorologi Publik," Jakarta: BMKG (2008).

Lampiran 2. Prediksi *Return Level*

Stasiun Pengamatan	u	$\hat{\sigma}$	$\hat{\xi}$	<i>Return Level</i> (mm)			
				1 Th	5 Th	10 Th	15 Th
Tangerang	61,0	19,8	0,20	89,5	137,7	163,6	180,5
Serang	52,0	18,5	0,03	76,2	108,2	122,5	131,0
Curug	68,1	15,6	0,15	90,3	124,6	142,1	153,2
Tanjung Priok	71,0	25,8	0,06	105,6	153,1	175,1	188,4
Kemayoran	65,0	21,8	0,30	99,2	165,2	205,0	232,2
Halim	74,0	22,2	0,22	107,3	164,9	196,9	218,0
Soekarno-Hatta Cengkareng	61,6	25,8	0,08	97,8	147,2	170,6	185,0
Darmaga Bogor	87,0	16,8	0,15	111,0	148,4	167,7	180,0
Pondok Betung	71,3	22,4	0,13	103,0	150,2	173,9	188,7

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang telah mengizinkan penulis untuk melakukan penelitian terhadap data curah hujan di sembilan stasiun pengamatan di wilayah Jakarta. Selain itu, penulis juga mengucapkan terimakasih kepada seluruh dosen dan karyawan di lingkungan Jurusan Statistika ITS yang telah memberikan banyak ilmu, pengalaman, dan bantuan kepada penulis.

DAFTAR PUSTAKA

[1] Goddard Space Flight Center. (2015, Oktober 10). [[NASA]] Satellite Confirms Urban Heat Island Increase Rainfall Around Cities. Tersedia. <http://www.sciencedaily.com/releases/2002/06/020619074019.htm>.
 [2] L. Fawcett dan D. Walshaw, "Improved Estimation for Temporally Clustered Extremes," *Environmetrics*, Vol. 18 (2007) 173-188.
 [3] Goldstein, H. "Multilevel Statistical Models". London: Edward Arnold (1995).
 [4] D. Cooley, D. Nychka, P. Naveau, "Bayesian Spatial Modeling of Extremes Precipitation Return Levels," *Journal of the American Assosiation*, Vol. 102 (2007) No. 479 824-840.
 [5] Amran, "Pemodelan Pengamatan Ekstrem Spasial Non Stasioner Menggunakan Bayesian Hirarki," Disertasi, Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya (2015).
 [6] M. Gilli dan E. Kellezi, *An Application of Extreme Value Theory for Measuring Risk*. Amsterdam: Elsevier Science
 [7] S. Coles. *An Introduction to Statistical Modelling of Extreme Value*. London: Springer-Verlag.
 [8] Li, Zongxing, *Study on Climate Changes in Southwestern China*. Berlin Heidelberg: Springer-Verlag (2015).
 [9] Iriawan, Nur, *Pemodelan dan Analisis Data-Driven*. Surabaya: ITS PRESS (2012).
 [10] P. Ismartini, Pengembangan Model Linier Hirarki dengan Pendekatan Bayesian untuk Pemodelan Data Pengeluaran Perkapita Rumah Tangga. Disertasi: Jurusan Statistika, Institut Teknologi Sepuluh Nopember (2013).