

Mengenal Data Ekstrim dan Distribusinya

Indriati Bisono^{1,2}

Abstract: People adopt many definitions for extremes. One such definition is that extremes are simply the upper tail data. In this paper we briefly review the types of upper tail data, the limiting distribution of extreme values, the analysis and the model diagnostics. We also provide a simple example of fitting generalized extreme values (GEV) for climate extremes.

Keywords: Extreme value, generalized extreme value distribution, generalized pareto distribution, return level.

Pendahuluan

Ekstrim dapat diartikan sebagai 'besar', 'banyak', 'sering' atau sebaliknya 'amat jarang'. Kejadian yang dikategorikan sebagai ekstrim (*extreme event*) adalah kejadian yang jarang terjadi namun memberikan dampak yang besar dan tidak terduga datangnya (Embrechts *et al.* [7]). Bencana alam, misalnya gempa bumi, badai, dan banjir besar, adalah contoh kejadian ekstrim di bidang ekologi lingkungan yang bisa membawa dampak besar sosial dan ekonomi. Di negara-negara dimana asuransi sudah menjadi bagian yang tak terpisahkan, kejadian ekstrim seperti ini dapat mengakibatkan klaim asuransi yang sangat besar. Nilainya bisa mencapai 26% dari seluruh total klaim dalam setahun. Sehingga mau tak mau industri asuransi memberikan perhatian besar atas probabilitas munculnya kejadian ekstrim. Pengetahuan tentang kejadian ekstrim di bidang iklim sangat diperlukan oleh hidrologis, perencana tata kota atau desainer bangunan. Misal, saat akan membangun waduk yang dapat mengantisipasi perubahan iklim sampai 50 bahkan 100 tahun ke depan, diperlukan prediksi terjadinya kejadian ekstrim; seberapa besar atau seberapa sering banjir akan melanda daerah tersebut.

Dalam Statistika, data ekstrim secara sederhana didefinisikan sebagai data yang terletak di ujung distribusi, bisa di ujung kanan atau kiri. Analisa data ekstrim pada dasarnya mengamati kelakuan data di ujung ini. *Extreme event theory* (EVT) berkembang dengan pesat sejak diperkenalkan oleh Fisher dan Tippet [8] dan diformalkan oleh Gnedenko [9]. Saat ini EVT telah mencakup multivariate (Tawn [18]; Coles dan Tawn [1,2]), random vector

(Coles dan Walshaw [4]), ketergantungan spasial (Coles [3]; Schlather dan Tawn [15]), ketergantungan temporal (Hall dan Tajvidi [10]), dan penggunaan model bertingkat (*hierarchical model*) (Cooley [5]; Sang dan Gelfand [14]; Schliep *et al.* [16]). Aplikasinya merambah berbagai macam disiplin seperti hidrologi, biologi, ekologi, klimatologi, telekomunikasi dan bahkan olah raga.

Tujuan dari tulisan ini adalah memberikan gambaran selengkap nya tentang nilai ekstrim, distribusi dan pemodelannya, serta diagnosa model. Contoh sederhana pemodelannya diberikan di bagian akhir tulisan.

Metode Penelitian

Beberapa jenis nilai ekstrim, distribusinya, analisa dan diagnose model akan dibahas dalam bagian ini.

Nilai Ekstrim

Beberapa jenis nilai ekstrim diantaranya adalah blok maksima (*block maxima*), nilai di atas ambang (*values over threshold*), dan r -nilai terbesar dari data terurut (*r-largest ordered data*). Blok maksima adalah nilai maksimum dari observasi dalam jangka waktu tertentu, misal dari data produk cacat perhari hanya diambil nilai maksimum data selama satu bulan. Dengan demikian, hanya 12 observasi dalam setahun yang dipergunakan, 353 (=365-12) observasi lainnya terbuang sia-sia yang merupakan pemborosan data yang cukup besar. Ini adalah salah satu kekurangan dari menggunakan blok maksima. Mungkin saja dua atau lebih data terbesar dalam setahun terjadi di bulan yang sama, padahal dalam blok maksima hanya diambil satu nilai terbesar, dengan kata lain, ada resiko kehilangan informasi penting. Oleh karena itu sebagian orang menggunakan r -nilai terbesar. Tipe data ekstrim yang lain adalah nilai di atas ambang. Nilai ambang α ditetapkan terdahulu, kemudian semua data yang lebih besar dari α dikategorikan sebagai ekstrim.

¹ Fakultas Teknologi Industri, Jurusan Teknik Industri, Universitas Kristen Petra, Jl Siwalankerto 121-131 Surabaya 60238. Email: mlindri@petra.ac.id

² Faculty of Science, Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, 3010, Australia.

Diterima 18 Oktober 2011; revisi 15 November 2011; diterima untuk dipublikasikan 30 November 2011

Blok Maksima

Andai kita mempunyai variabel acak bebas X_1, X_2, \dots, X_n tiap variable X_i mempunyai fungsi distribusi sama yaitu $F(x)$. Selanjutnya kita pertimbangkan nilai maksimum $M_n = \max\{X_1, K, X_n\}$ maka

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \dots P(X_n \leq x) \\ &= F(x)^n \end{aligned}$$

Untuk n cukup besar $F(x)^n \rightarrow 0$, hal mana kurang menarik perhatian. Di lain pihak, standarisasi dari blok maksima menawarkan distribusi yang lebih menarik. Andai dapat ditemukan rangkaian $\{a_n\}$ dan $\{b_n > 0\}$, maka nilai standar M_n akan memiliki distribusi limit $H(x)$

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \rightarrow H(x)$$

Karakteristik $H(x)$ pertama-tama diteliti oleh Fisher dan Tippett [8], kemudian dilanjutkan oleh Gnedenko [9]. Jika $H(x)$ ada, maka ia akan mengikuti salah satu bentuk berikut:

$$\begin{aligned} H(x) &= \exp(-e^{-x}) \\ H(x) &= \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} \\ H(x) &= \begin{cases} \exp(-|x|^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases} \end{aligned}$$

Masing-masing bentuk di atas dikenal sebagai Gumbel, Frechet dan Weibull. Ketiganya tergabung dalam bentuk yang dikenal sebagai *generalised extreme value* (GEV)

$$G(x) = \exp\left(-\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)$$

dengan μ sebagai parameter lokasi, σ parameter skala dan ξ parameter bentuk (*shape*).

r-nilai Terbesar

Misal $X_{n,1} \geq X_{n,2} \geq \dots \geq X_{n,r}$ adalah r data terurut terbesar dari n sampel statistik yang bebas dan berdistribusi identik. Jika dapat ditemukan konstanta a_n and b_n adalah sehingga

$$\left(\frac{X_{n,1} - b_n}{a_n}, \dots, \frac{X_{n,r} - b_n}{a_n}\right) = (y_1, \dots, y_r)$$

konvergen dalam distribusi maka (y_1, \dots, y_r) akan memiliki fungsi densitas

$$\pi(y_1, \dots, y_r) = \sigma^{-r} \exp\left(-\left(1 + \xi \frac{y-\mu}{\sigma}\right)^{-1/\xi} - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^r \left(1 + \xi \frac{y_j-\mu}{\sigma}\right)\right)$$

Salah satu kelemahan r -data terurut terbesar adalah lebih rentan terhadap asumsi independen (i.i.d) sebab beberapa data terbesar yang berurutan mungkin saja berasal dari kejadian ekstrim yang sama.

Robinson dan Tawn [13] mengembangkan metode untuk menganalisa data ekstrim yang diraih oleh atlet atletik dalam sebuah kejuaraan dengan menggunakan tiga perolehan tertinggi. Contoh lain analisa data terurut terbesar dapat dilihat pada Weissman [20]; Smith [17] dan Tawn [19].

Exceedances over Threshold

Andai u nilai ambang yang ditentukan, dan $Y = X - u > 0$, maka fungsi distribusi kumulatif dari Y adalah

$$F_u(y) = P(Y \leq y | Y > 0) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

Saat u mendekati $\sup\{x: F(x) < 1\}$ maka $F_u(y) \approx G(y; \sigma, \xi)$ yang mana $G(y; \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}$

yaitu *Generalised Pareto Distribution* (GPD). Dalam aplikasi seringkali dipilih nilai ambang tinggi, yaitu 99,75 atau 95 persentil. Menyadari adanya ketergantungan data secara temporal, terutama pada data meteorologi, meteorologis dan hidrologis menggunakan *peak over threshold* (POT). Pada POT data dikelompokkan berdasar waktu tertentu misal pagi, siang, atau malam hari, selanjutnya hanya diambil data di atas nilai ambang dalam kelompok waktu yang sama. Hal ini dilakukan untuk menghindari autokorelasi. Sama halnya dengan *exceedances over threshold*, POT berdistribusi GPD.

Lebih jauh lagi, jumlah data di atas nilai ambang merupakan random variabel yang mungkin berdistribusi *Poisson*. Model yang mengikutkan sertakan keduanya; nilai diatas ambang dan banyaknya nilai di atas ambang, dikenal sebagai *Poisson-GPD* model (Davison and Smith [6]).

Estimasi Parameter

Setelah mengetahui jenis data ekstrim yang digunakan dan mengasumsikan distribusi dari data tersebut, misal GEV, langkah selanjutnya adalah mengestimasi parameter dari distribusi asumsi. Metode estimasi yang paling sering dipakai adalah *maximum likelihood estimation* (MLE). Andaikan y_1, \dots, y_n adalah n observasi yang independen dari variabel acak Y_1, \dots, Y_n yang masing-masing berdistribusi GEV dengan parameter tak tentu $\theta = \{\mu, \sigma, \xi\}$. Selain itu data diasumsikan tidak bergantung satu sama lain, maka

$$f(y_1, \dots, y_n | \theta) = f(y_1 | \theta) \dots f(y_n | \theta)$$

Inilah yang disebut dengan fungsi *likelihood*, seringkali ditulis $l(\theta | y)$. Log dari fungsi *likelihood*:

$$L(\theta | y) = K \prod_{i=1}^n \exp \left(- \left(1 + \xi \frac{y_i - \mu}{\sigma} \right)^{-1/\xi} \right) \prod_{i=1}^n \frac{1}{\sigma} \left(1 + \xi \frac{y_i - \mu}{\sigma} \right)^{\frac{1}{\xi} - 1} \quad (1)$$

Nilai parameter yang memaksimumkan *log-likelihood* (1) dapat ditemukan dengan teknik kalkulus biasa yaitu solusi turunan pertama terhadap tiap-tiap parameter sama dengan nol. Dengan kata lain, MLE mencari nilai parameter yang memaksimumkan probabilitas data.

Alternatif lain untuk mengestimasi parameter adalah dengan pendekatan Bayesian. Pada pendekatan ini parameter diasumsikan menganut sebuah distribusi $g(\theta)$ disebut distribusi prior. Dinamakan prior karena distribusi ini tidak terkondisi oleh data observasi. Distribusi dari parameter yang terkondisi oleh data observasi disebut distribusi posterior. Posterior distribusi $g(\theta | y)$ dapat diturunkan dengan menggunakan aturan Bayes untuk probabilitas bersyarat, yakni proporsional terhadap perkalian fungsi *likelihood* dan fungsi prior

$$g(\theta | y) \propto g(\theta) \prod_i f(y_i | \theta)$$

Distribusi prior dapat bersifat informatif, berdasarkan informasi dari pakar, atau non-informatif. Jika hal yang terakhir terjadi, umumnya digunakan distribusi *uniform* sebagai prior. Jika $f(y | \theta)$ fungsi yang kompleks seperti GEV maka $g(\theta | y)$ akan sangat kompleks dan sulit untuk diturunkan secara analitik. Namun hal ini dapat diatasi dengan berkembangnya metode Monte Carlo. Beberapa algoritma yang sering dipakai dalam Monte Carlo adalah Metropolis Hastings dan Gibbs sampler. Penjelasan rinci tentang metode ini dapat ditelusuri di Robert and Casella [12].

Teknik estimasi lain yang sering digunakan kala jumlah observasi sedikit adalah *L-moments*. *L-moments* adalah ekspektasi dari kombinasi linier dari data terurut yang cara kerjanya serupa dengan *method of moment*. *L-moments* merupakan metode estimasi yang tangguh untuk sampel berukuran kecil, bahkan kadang lebih unggul dibanding dengan MLE. Penjelasan secara terperinci dari metode ini dapat dilihat pada Hosking [11].

Diagnosa Model

Pemodelan data biasanya diawali dengan asumsi, data diasumsikan menganut distribusi tertentu.

Kemudian, dengan data di tangan parameter dari distribusi tersebut diestimasi. Konsekuensinya, asumsi awal harus diuji, yaitu kesejalaran data dengan model yang diasumsikan padanya. Dua teknik uji diagnosa berdasarkan grafik akan dijabarkan berikut ini.

Grafik Probabilitas dan QQ Plot

Misal y_1, \dots, y_n adalah rangkaian observasi nilai ekstrim (blok maksima) yang independen dengan fungsi distribusi G (tak diketahui). Jika y_i diasumsikan berdistribusi GEV, dan dengan MLE kita dapatkan estimasi dari distribusi ini, \hat{G} . Ketepatan estimasi \hat{G} dapat diuji dengan cara membandingkan model estimasi dengan distribusi empiris \tilde{G} .

Misal $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ adalah data terurut sedemikian hingga $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$, maka fungsi empiris dengan mudah didapatkan sebagai berikut,

$$P(Y \leq y_{(i)}) = \tilde{G}(y_{(i)}) = \frac{i}{n+1}$$

Distribusi empiris ini dapat digunakan mengestimasi distribusi data sesungguhnya (G). Jika asumsi $Y_i \sim GEV$ benar, maka $\hat{G} \approx \tilde{G}$ untuk tiap y_i . Akibatnya, grafik dari pasangan titik $(\tilde{G}(y_{(i)}), \hat{G}(y_{(i)}))$ untuk $i = 1, \dots, n$ akan terletak sepanjang garis diagonal (gradien 1). Ini yang dikenal dengan grafik probabilitas (*probability plot*). Grafik kuantil (*QQ plot*) di lain pihak adalah plot dari pasangan titik

$$\left(y_{(i)}, \hat{G}^{-1} \left(\frac{i}{n+1} \right) \right) \text{ dengan } \hat{G}^{-1} \left(\frac{i}{n+1} \right) = \hat{\mu} - \frac{\hat{\sigma}}{\xi} \left(1 - \left(-\log \frac{i}{n+1} \right)^{-\xi} \right)$$

Sama halnya dengan grafik probabilitas, jika asumsi model benar, maka titik-titik dalam *QQ plot* akan terletak di sepanjang garis diagonal.

The Mean Excess Plot

Grafik ini digunakan untuk mendiagnosa *threshold-exceedances*, bahkan seringkali digunakan sebelum pemodelan dilakukan. Tujuannya untuk menentukan nilai *threshold*.

Andai Y_1, Y_2, \dots, Y_n adalah n variabel acak dimana nilai lebih di atas titik ambang u_0 berdistribusi GPD. Mean dari nilai lebih untuk $\xi < 1$ adalah $E(Y) = \frac{\sigma}{1-\xi}$

Jika GPD valid untuk $Y > u_0$ maka seharusnya ia juga valid untuk $Y > u$, dimana $u > u_0$. Berdasarkan karakteristik GPD berlaku, $E(Y - u | Y > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi u}{1-\xi}$

Sehingga, untuk sebarang $u > u_0$, mean dari $Y > u$, $E(Y - u|Y > u)$ merupakan fungsi linier dari u dengan gradient $\xi/(1 - \xi)$. Tambahan lagi, $E(Y - u|Y > u)$ dapat diestimasi secara empiris dengan mean sampel dari kelebihan nilai dengan *threshold* u . Plot dari *threshold* u di sumbu x dan mean sampel di sumbu y inilah yang dikenal dengan *the mean excess plot* atau *the mean residual life plot*.

Hasil dan Pembahasan

Akhir-akhir ini sering dibicarakan adanya fenomena perubahan iklim (*climate change*). Musim kemarau semakin panjang, musim hujan juga menjadi lebih panjang, maximum temperatur makin tinggi minimum temperatur makin rendah. Tampaknya kejadian ekstrim lebih memberikan dampak terhadap perubahan iklim dibanding kejadian rata-rata. Pada seksi ini akan diberikan contoh memodelkan ekstrim data, yaitu temperatur dan curah hujan ekstrim dengan famili distribusi GEV dengan bantuan paket ‘extRemes’ dari *freeware* R.

Dua set data (maksimum curah hujan dan maksimum temperatur bulanan) diunduh dari *the Bureau’s online of the Australian Climate and Weather Extremes Monitoring System*. Data berasal dari stasiun Charleville Aero, salah satu tempat pencatatan dengan kualitas tinggi dan mempunyai data lengkap sejak 1942. Dari data bulanan hanya diambil data maksimum per tahun, plot anual maksimum dapat dilihat pada Gambar 1. Tampak masing-masing data set independen dan mempunyai varians konstan (Gambar 1). Oleh karenanya kita dapat mengasumsikan data independen dan menganut distribusi GEV. Nilai estimasi parameter dengan metode MLE, $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ untuk kedua data set disajikan pada Tabel 1; nilai dalam kurung adalah standar *error* dari masing-masing nilai estimasi.

Beberapa diagnostik plot untuk menilai keakuratan model disajikan pada Gambar 2. Titik-titik pada kedua *QQ* plot jatuh di sekitar garis diagonal, indikasi bahwa model GEV sesuai untuk kedua data set. Nilai estimasi $\hat{\xi}$ untuk data curah hujan adalah 0,07, *likelihood ratio test* untuk nilai ini (tidak disajikan) gagal menolak hipotesa nol $\xi=0$. Dengan kata lain distribusi Gumbel mungkin tepat untuk memodelkan data curah hujan ini. Parameter bentuk untuk model temperatur signifikan negatif, yaitu temperatur ekstrim mengikuti distribusi Weibull. Sebagai tambahan, GEV distribusi dapat memodelkan data temperatur selama setengah tahun (Nopember - April) dengan jauh lebih baik (tidak disajikan di sini). Hal ini dapat dimengerti mengingat temperatur sangat bersifat musiman (*seasonal*).

Tabel 1. Nilai estimasi dari parameter GEV

Variabel	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
Max temperatur	35,70 (0,19)	1,44 (0,14)	-0,41 (0,08)
Max curah hujan	391,38 (18,67)	133,03 (14,20)	0,07 (0,11)

Selanjutnya model estimasi dapat digunakan untuk menentukan besaran-besaran lain yang diperlukan, misalnya *return level*. p -year *return level* tidak lain adalah $1/p$ persentil dari distribusi kumulatif, seringkali diinterpretasikan sebagai nilai yang akan dilampaui sekali dengan probabilitas p dalam kurun waktu (*return period*) $1/p$ tahun. p -year *return level* dapat dengan mudah dihitung dari persamaan berikut

$$P(Z \leq y_p) = \exp\left(-\left(1 + \xi \frac{y_p - \hat{\mu}}{\hat{\sigma}}\right)^{-\frac{1}{\xi}}\right) = 1 - p$$

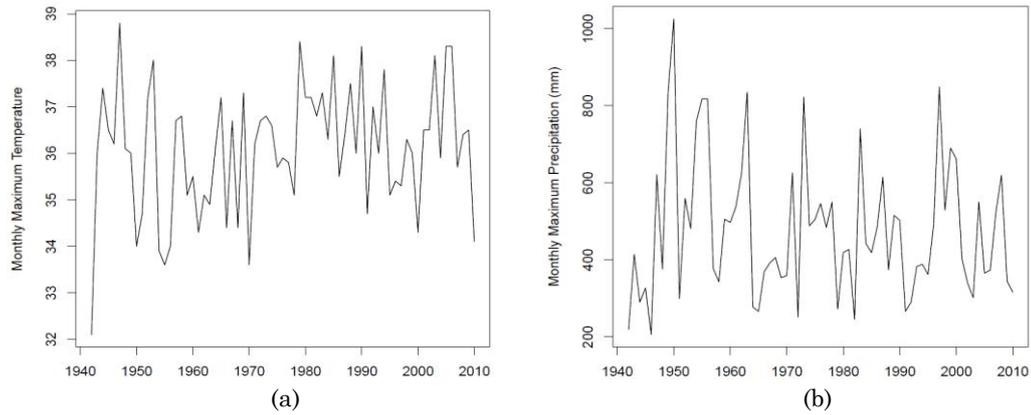
Invers dari persamaan di atas memberikan nilai p -years *return level*, yaitu

$$y_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\xi} (1 - (-\log(1 - p))^{-\xi}), & \text{untuk } \xi \neq 0 \\ \hat{\mu} - \hat{\sigma} \log(-\log(1 - p)), & \text{untuk } \xi = 0 \end{cases}$$

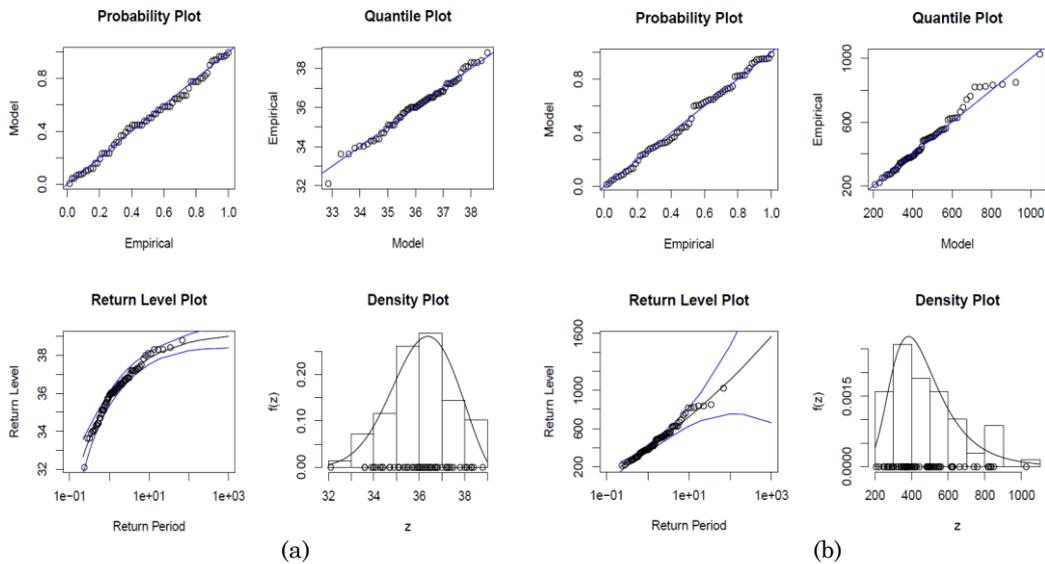
Sebagai contoh, untuk data temperatur, nilai estimasi dari parameter model adalah $\hat{\mu} = 35,70$; $\hat{\sigma} = 1,44$ dan $\hat{\xi} = -0,41$ substitusi nilai-nilai ini ke persamaan (1) dengan $p=0,01$ akan memberikan nilai $y_{0,01} = 38,68$. Artinya, dalam rentang waktu (*return period*) 100 tahun, temperatur di lokasi tersebut akan melewati 38,68°C dengan probabilitas 0,01. Grafik *return level* terhadap *return period* 10 s/d 1000 tahun, dengan 0.95 *confidence bands* untuk masing-masing data set dapat dilihat pada Gambar 3. Terlihat *confidence interval* dari *return level* untuk data curah hujan melebar dengan drastis, hal ini terjadi karena ketidak simetrian distribusi tidak diakomodasi dengan baik oleh metode delta (metode yang dipakai untuk menghitung *confidence interval* di sini). Alternatif lain untuk mendapatkan *confidence interval* yang lebih mengakomodasi ketidak-simetrian distribusi adalah dengan menggunakan *profile likelihood* atau menghitung probabilitas bersyarat melalui pendekatan Bayesian.

Simpulan

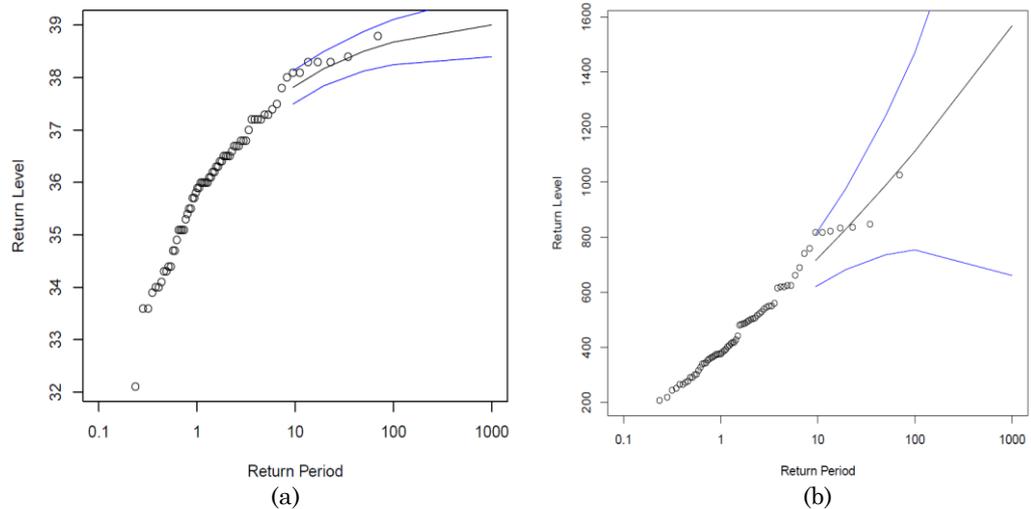
Ada beberapa tipe data ekstrim, diantaranya blok maksima, r -nilai terbesar, dan nilai-nilai di atas ambang batas. Seringkali data ekstrim, yang memenuhi kondisi tertentu, dapat tepat dimodelkan dengan distribusi GEV. Famili distribusi GEV memiliki tiga sub-famili yaitu Gumbel, Frechet dan Weibull. Distribusi ini dikarakterisasi oleh tiga parameter yaitu parameter lokasi, skala, dan bentuk.



Gambar 1. Data anual maksimum dari (a) temperatur dan (b) curah hujan.



Gambar 2. Diagnostik plot untuk data (a) temperatur dan (b) curah hujan.



Gambar 3. Grafik return level versus return periode untuk data (a) temperatur dan (b) curah hujan.

Estimasi dari ketiga parameter dapat dilakukan dengan MLE atau pendekatan Bayesian. Umumnya diantara ketiga parameter, estimasi parameter bentuk yang lebih sulit dilakukan, padahal parameter ini sangat menentukan. Khususnya jika

model estimasi untuk inferensi *return level* akan digunakan. Pada sesi aplikasi di atas, temperatur dan curah hujan maksimum berhasil dimodelkan dengan GEV, masing-masing menganut distribusi Weibull dan Gumbel.

Ucapan Terima Kasih

Artikel ini ditulis saat penulis menempuh studi di University of Melbourne dengan beasiswa dari Dirjen Dikti, kepadanya penulis mengucapkan terima kasih. Penulis juga mengucapkan terima kasih kepada mitra bestari (*reviewer*) atas masukannya sehingga artikel ini menjadi lebih baik.

Daftar Pustaka

1. Coles, S. G., and Tawn, J. A., Statistics of Coastal Flood Prevention, *Philosophical Transactions of the Royal Society of London, A*, 332, 1990, pp. 457–476.
2. Coles, S. G., and Tawn, J. A., Modelling Extreme Multivariate Events, *Journal of the Royal Statistical Society, Series B*, 53, 1991, pp. 377–392.
3. Coles, S. G., Regional Modelling of Extreme Storms via Max-Stable Processes, *Journal of the Royal Statistical Society, Series B*, 55, 1993, pp. 797–816.
4. Coles, S. G., and Walshaw, D., Directional Modeling of Extreme Wind Speeds, *Applied Statistics*, 43, 1994, pp. 139–157.
5. Cooley, D., Nychka, D., and Naveau, P., Bayesian Spatial Modeling of Extreme Precipitation Return Level. *Journal of the American Statistical Association*, 102, 2007, pp. 824–840.
6. Davison, A. C., and Smith, R. L., Models for Exceedances over High Thresholds (with discussion), *Journal of the Royal Statistical Society, Series B*, 52, 1990, pp. 393–442.
7. Embrechts, P., Kluppelberg, C., and Mikosch, T. *Modelling Extremal Events*, Springer, 1997.
8. Fisher, R. A., and Tippett, L. H. C., Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample, *Proc. Cambridge Philos. Soc.*, 24, 1928, pp. 180–190.
9. Gnedenko, B. V., Sur la Distribution Limite du Terme Maximum d'une s'erie al'eatoire, *The Annals of Mathematics*, 44, 1943, pp. 423–453.
10. Hall, P., and Tajvidi, N., Nonparametric Analysis of Temporal Trend when Fitting Parametric Models to Extreme-Value Data, *Statistical Science*, 15, 2000, pp. 153–167.
11. Hosking, J. R. M., and Wallis, J. R., Parameter and Quantile Estimation for the Generalized Pareto Distribution, *Technometrics*, 29, 1987, pp. 339–349.
12. Robert, C. P., and Casella, G., *Monte Carlo Statistical Methods*, Springer, 2004.
13. Robinson, M. E., and Tawn, J. A., Statistics for Exceptional Athletics Records, *Applied Statistics*, 44, 1995, pp. 499–511.
14. Sang, H., and Gelfand, A. E., Hierarchical Modeling for Extreme Values Observed over Space and Time, *Environ. Ecol. Stat.*, 16, 2009, pp. 407–426.
15. Schlather, M., and Tawn, J. A., Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference, *Biometrika*, 90, 2003, pp. 139–156.
16. Schliep, E. M., Cooley, D. Sain, S. R., and Hoeting, J. A., A Comparison Study of Extreme Precipitation from Six Different Regional Climate Models via Spatial Hierarchical Modeling, *Extremes*, 13, 2010, pp. 219–239.
17. Smith, R. L., Extreme Value Theory Based on The r-largest Annual Events, *Journal of Hydrology*, 86, 1986, pp. 27–43.
18. Tawn, J. A., Bivariate Extreme Value Theory: Models and Estimation. *Biometrika*, 75, 1988, pp. 397–415.
19. Tawn, J. A., An Extreme Value Theory Model for Dependent Observations, *Journal of Hydrology*, 101, 1988, pp. 227–250.
20. Weissman, I., Estimation of Parameters and Quantiles Based on The k-largest Observations. *Journal of the American Statistical Association*, 73, 1978, pp. 812–815.