

PEMODELAN SPATIAL EXTREME VALUE DENGAN REGULARISASI FUSED LASSO DAN FUSED RIDGE

(STUDI KASUS CURAH HUJAN EKSTREM DI SUMATERA UTARA)

Adellia Purnomo Putri

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya, Indonesia

Alamat e-mail: adelliapurnomo.21044@mhs.unesa.ac.id

A'yunin Sofro

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya, Indonesia

*Alamat e-mail: ayuninsofro@unesa.ac.id

Abstrak

Curah hujan ekstrem merupakan salah satu fenomena hidrometeorologi yang berpotensi menimbulkan bencana seperti banjir dan tanah longsor, khususnya di Provinsi Sumatera Utara yang memiliki karakteristik geografis dan topografi yang kompleks. Penelitian ini bertujuan untuk memodelkan curah hujan ekstrem secara spasial menggunakan pendekatan *Extreme Value Theory* (EVT) melalui model *Generalized Pareto Distribution* (GPD) dengan metode *Peaks Over Threshold* (POT), serta membandingkan performa regularisasi *fused lasso* dan *fused ridge* dalam menstabilkan estimasi parameter spasial. Data yang digunakan berupa data curah hujan harian tahun 2014–2024 dari lima stasiun BMKG di Provinsi Sumatera Utara. Penentuan threshold dilakukan menggunakan kombinasi metode *quantile selection* dan *Mean Residual Life Plot* (MRL). Estimasi parameter dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE). Evaluasi model dilakukan menggunakan *Takeuchi Information Criterion* (TIC). Hasil penelitian menunjukkan bahwa kedua metode regularisasi mampu memodelkan curah hujan ekstrem spasial dengan baik dan menghasilkan nilai *return level* yang meningkat seiring bertambahnya periode ulang. Nilai *return level* tertinggi pada kedua model terdapat di wilayah Tapanuli Tengah. Hasil evaluasi model menunjukkan bahwa regularisasi *fused ridge* menghasilkan nilai TIC yang lebih kecil dibandingkan dengan *fused lasso*, yaitu sebesar 100913,9 sehingga dipilih sebagai model terbaik dalam memodelkan spasial curah hujan ekstrem di Provinsi Sumatera Utara. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan pemodelan nilai ekstrem spasial serta mendukung mitigasi risiko bencana hidrometeorologi berbasis data.

Kata Kunci: *Spatial Extreme Value, Generalized Pareto Distribution, Gaussian Copula, fused lasso, fused ridge.*

Abstract

Extreme rainfall is a hydrometeorological phenomenon that can cause disasters such as floods and landslides, especially in North Sumatera Province, which has a complex geographical and topographical character. The purpose of this study is to model extreme rainfall spatially using the Extreme Value Theory (EVT) approach through the Generalized Pareto Distribution (GPD) model with the Peaks Over Threshold (POT) method, as well as to compare the performance of regularization using fused lasso and fused ridge in stabilizing spatial parameter estimation. The data used is daily rainfall data from 2014 to 2024 from five BMKG stations in North Sumatera Province. Thresholds were determined using a combination of the percentile method and the Mean Residual Life Plot (MRL). The parameter estimates were calculated using Maximum Likelihood Estimation (MLE). Model evaluation was performed using the Takeuchi information criterion (TIC). The results show that both regularization methods are capable of modeling spatial extreme rainfall effectively and produce return levels that increase with longer return periods. The highest return levels for both models were observed in Central Tapanuli. Based on the TIC values, fused ridge outperformed fused lasso with a TIC value of 100,913.9 and was therefore selected as the best model for spatial extreme rainfall modeling in North Sumatra Province. This study is expected to contribute to the development of spatial extreme value modeling and support data-driven hydrometeorological disaster risk mitigation.

Keywords: *Spatial Extreme Value, Generalized Pareto Distribution, Gaussian Copula, fused lasso, fused ridge.*

PENDAHULUAN

Pemodelan fenomena ekstrem spasial telah menjadi tantangan utama dalam statistik lingkungan,

terutama ketika fenomena tersebut berkaitan dengan risiko bencana yang tersebar secara geografis (Davison et al., 2012). Salah satu penekatan yang berkembang pesat adalah *Spatial Extreme Value*

(SEV) *Modelling*, yang dirancang untuk memahami probabilitas kejadian langka dengan dampak besar pada beberapa lokasi sekaligus (de Fondeville & Davison, 2022). Model ini berawal dari pendekatan distribusi nilai ekstrem yang kemudian diaplikasikan melalui *Extreme Value Theory* (EVT) untuk mengukur dependensi spasial antar lokasi geografis (Chen et al., 2015). EVT berfokus pada perilaku ekor (*tail*) suatu distribusi guna menentukan probabilitas nilai-nilai ekstremnya. Dua metode utama dalam EVT adalah *Generalized Extreme Value* (GEV) dengan pendekatan *Block Maxima* (BM) dan *Generalized Pareto Distribution* (GPD) dengan pendekatan *Peaks Over Threshold* (POT). Menurut de Fondeville & Davison (2022), metode GPD lebih unggul dibandingkan GEV karena menggunakan seluruh nilai ekstrem dalam pengelolaannya sehingga mendapatkan lebih banyak informasi dari data. Selain itu, dalam penelitian Fisher et al. (2013) ditemukan bahwa GPD terbukti memberikan estimasi periode pengembalian (*return level*) yang lebih akurat dibandingkan GEV. Oleh karena itu, penelitian ini menggunakan *Generalized Pareto Distribution* (GPD) dengan pendekatan *Peaks Over Threshold* (POT) untuk kerangka SEV karena kemampuannya dalam memanfaatkan lebih banyak data ekstrem dan menghasilkan estimasi *return level* yang lebih akurat.

Namun, dalam konteks spasial, asumsi independensi antarlokasi tidak lagi valid. Ketergantungan spasial sering muncul karena kedekatan geografis, kesamaan topografi, atau pengaruh atmosfer regional, yang jika diabaikan akan menghasilkan estimasi yang bias dan kurang dapat diandalkan (Wadsworth & Tawn, 2012). Meskipun model ketergantungan spasial dapat menggambarkan hubungan spasial antarlokasi dalam data ekstrem, pada pemodelan GPD secara spasial, estimasi parameter pada lokasi yang berdekatan dapat menunjukkan perbedaan yang tidak wajar akibat tingginya variabilitas data ekstrem dan keterbatasan jumlah observasi ekstrem pada setiap lokasi. Kondisi ini dapat menghasilkan estimasi yang kurang stabil dan kurang merepresentasikan kemiripan karakteristik geografis antarlokasi. Padahal, secara geografis, lokasi-lokasi tersebut berdekatan dan seharusnya memiliki karakteristik yang mirip. Dalam hal ini, metode regularisasi menjadi penting. Regularisasi tidak hanya membantu menstabilkan estimasi, tetapi juga

memungkinkan pemodelan parameter spasial GPD yang lebih halus dan masuk akal secara geografis (Shao et al., 2024). Metode seperti *fused ridge* dan *fused lasso*, telah diperkenalkan untuk mengatasi tantangan tersebut. *Fused lasso* unggul dalam mendeteksi perubahan mendadak atau diskontinuitas (Tibshirani & Saunders, 2005). Sementara *fused ridge* unggul dalam mereduksi variansi estimasi tanpa menghilangkan struktur spasial (Shao et al., 2024). Selain itu, penelitian Sass et al. (2021) menemukan bahwa *fused penalty* lebih unggul daripada *spatial* GEV dan GPD biasa dalam estimasi *return level*. Penelitian ini akan membandingkan dua metode regularisasi tersebut dengan evaluasi kinerja model dilakukan menggunakan *Takeuchi Information Criterion* (TIC) untuk menilai kualitas pemodelan spasial ekstrem secara objektif.

Aplikasi dari pendekatan ini sangat relevan untuk fenomena curah hujan ekstrem di wilayah daratan dan pegunungan Sumatera Utara yang dikenal rawan bencana hidrometeorologi seperti banjir bandang dan tanah longsor (BNPB, 2024). Curah hujan merupakan salah satu indikator kunci dalam studi klimatologi tropis dan memegang peran penting dalam penilaian risiko kebencanaan darat, khususnya dalam konteks pengelolaan DAS (Daerah Aliran Sungai), perencanaan infrastruktur, dan mitigasi banjir (Ward et al., 2013). Dalam era perubahan iklim, kejadian hujan lebat yang terjadi secara berturut-turut mengalami peningkatan baik dalam frekuensi maupun intensitas, yang berkorelasi langsung dengan tingginya risiko bencana hidrologis ((IPCC), 2021). Oleh karena itu, analisis spasial terhadap curah hujan ekstrem menjadi sangat krusial untuk memahami pola geografis risiko dan memperkuat sistem mitigasi berbasis data.

Secara spasial, penelitian ini menggunakan data dari lima stasiun meteorologi utama di Sumatera Utara yakni Stasiun Meteorologi FL Tobing mewakili wilayah Tapanuli Tengah, Stasiun Meteorologi Aek Gondang mewakili wilayah Padang Lawas Utara, Stasiun Meteorologi Kualanamau mewakili wilayah Deli Serdang, Stasiun Meteorologi Maritim Belawan mewakili wilayah Medan, dan Stasiun Meteorologi Binaka mewakili wilayah Kepulauan Nias yang dipilih karena keandalan data harian dari BMKG. Kelima lokasi ini mencakup ragam topografi pesisir, pulau, dataran rendah, dan pegunungan di provinsi tersebut. Klimanya sering mengalami hujan

berturut-turut akibat gangguan atmosfer tropis seperti Madden-Julian Oscillation (MJO) dan sistem konvergensi lokal (pertemuan massa udara dari berbagai arah di suatu wilayah) yang menyebabkan udara naik ke atas yang tertahan oleh topografi pegunungan Bukit Barisan bagian barat dan tengah (Prasetyo et al., 2018). Selain itu, data kejadian banjir daerah Sumatera Utara selama 2014 - 2024 menunjukkan bahwa mayoritas bencana yang terjadi ialah banjir besar (DIBI, n.d.). Ini menjadikan Sumatera Utara sebagai wilayah yang sesuai untuk studi spasial ekstrem menggunakan pendekatan GPD dan pemodelan curah hujan. Secara keseluruhan, pendekatan pemodelan spasial ekstrem terhadap curah hujan di Provinsi Sumatera Utara yang tidak hanya relevan dari segi kebencanaan hidrometeorologi, tetapi juga secara ilmiah memberikan kontribusi penting dalam mengisi kesenjangan pengetahuan terkait karakteristik spasial curah hujan ekstrem di wilayah tropis.

Dengan demikian, penelitian ini bertujuan untuk membandingkan kinerja metode regularisasi *fused lasso* dan *fused ridge* dalam estimasi parameter skala model GPD pada data curah hujan ekstrem di Provinsi Sumatera Utara. Pendekatan ini digunakan untuk mengatasi ketidakstabilan estimasi parameter pada lokasi yang berdekatan secara geografis. Kinerja kedua metode dievaluasi berdasarkan stabilitas parameter dan kualitas model menggunakan Takeuchi Information Criterion (TIC). Hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan pemodelan nilai ekstrem spasial serta mendukung upaya mitigasi risiko bencana hidrometeorologi di Indonesia.

KAJIAN TEORI

EXTREME VALUE THEORY (EVT)

Extreme Value Theory (EVT) merupakan salah satu metode statistika yang digunakan untuk mempelajari tentang bagaimana perilaku ekor (tail) sebuah distribusi untuk data univariate. Metode statistika EVT berfokus pada perilaku ekor (tail) suatu distribusi untuk menentukan probabilitas nilai-nilai ekstremnya. Metode ini dapat menjelaskan kerugian fenomena ekstrem yang jika terjadi mempunyai konsekuensi kerugian yang sangat besar (bernilai maksimum) dan tidak dapat dimodelkan dengan pendekatan biasa. Dalam mengidentifikasi

pergerakan nilai ekstrem terdapat dua pendekatan yang digunakan (Omari & Nguni, 2021). Pendekatan pertama yakni metode *Block Maxima* (BM) dengan sistem yang mengambil nilai maksimum dalam satu periode. Sedangkan pendekatan kedua yakni metode *Peaks Over Threshold* (POT) dengan sistem yang mengambil nilai-nilai yang melewati suatu nilai *threshold*.

GENERALIZED PARETO DISTRIBUTION (GPD)

Menurut Lee & Kim (2019), distribusi GPD didefinisikan sebagai persamaan berikut.

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

$d = 1, 2, 3, \dots, n.$

GPD memiliki 2 parameter yaitu ξ merupakan parameter bentuk dan σ merupakan parameter skala. y_i menotasikan nilai ekstrem yang berada diatas *threshold* u yang kemudian dikurangi dengan besar *threshold* u .

Terdapat 3 tipe distribusi dalam GPD diantaranya tipe 1 berdistribusi eksponensial jika $\xi = 0$, tipe 2 berdistribusi pareto jika $\xi > 0$, dan tipe 3 berdistribusi beta jika $\xi < 0$.

Adapun *probability density function* (pdf) dari GPD ialah :

$$h(y_d; \xi, \sigma) = \begin{cases} \frac{1}{\sigma} \left(1 + \frac{\xi y_d}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{y_d}{\sigma}\right), & \xi = 0 \end{cases} \quad (2)$$

Generalized Pareto Distribution (GPD) didasarkan pada pendekatan *Peaks Over Threshold* (POT), yaitu metode yang mengidentifikasi nilai-nilai ekstrem dengan menetapkan suatu nilai ambang batas (*threshold*, u). Semua data yang melebihi *threshold* tersebut disebut sebagai data *exceedance*, dan hanya data inilah yang digunakan dalam proses estimasi parameter model GPD. Dalam penelitian ini penentuan *threshold* (u), dilakukan dengan mengombinasikan 2 metode pemilihan *threshold* yang berbeda. Metode pertama menggunakan metode estimasi *threshold* dinamis, yakni metode memilih dalam rentang persentil (*quantile selection*) 50 - 95 dari kumpulan data intensitas curah hujan sebagai *threshold* dengan mempertimbangkan nilai parameter skala dan nilai *p-value* uji *Anderson Darling* (Gao et al.,

2021). Metode kedua menggunakan bantuan *mean residual life plot* karena *mean residual life plot* linier terhadap u (Anifah, 2014). *Mean residual life plot* didasari oleh nilai rata-rata GPD yang dianalogikan seperti menentukan ukuran blok yang akan dipilih sehingga terdapat keseimbangan antara bias dan keragaman. Koordinat titik sebaran *mean residual life plot* ditentukan berdasarkan:

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (y_i - u) \right) : u < y_{maks} \right\} \quad (3)$$

dengan $y_1, y_2, \dots, y_{(n_u)}$ adalah observasi sebanyak n_u yang melebihi nilai u ; y_{maks} merupakan nilai terbesar dari y_d ; dan n_u adalah banyak nilai yang melebihi u (Far & Wahab, 2016).

PARAMETER SPATIAL GPD

Dalam analisis *Spatial Extreme Value* (SEV), transformasi dari model *univariate* menjadi model spasial merupakan langkah penting untuk menangkap dependensi antarlokasi. Setelah mengidentifikasi data *exceedance* melalui *threshold* yang telah ditentukan, langkah selanjutnya ialah memodelkan parameter GPD secara spasial menggunakan struktur dependensi yang sesuai. Model GPD spasial dikembangkan dengan memperluas konsep *univariate* GPD ke konteks spasial melalui pemodelan parameter yang bervariasi secara spasial (Sass et al., 2021). Untuk data *exceedance* di lokasi s_i , parameter GPD dimodelkan sebagai fungsi dari koordinat spasial:

- Parameter *Scale* (skala)

$$\sigma(s_i) = \beta^T X(s_i) \quad (4)$$

- Parameter *Shape* (bentuk)

$$\xi(s_i) \quad (5)$$

diasumsikan konstan

di mana $X(s_i)$ adalah vektor kovariat spasial yang biasanya mencakup koordinat longitude dan latitude pada lokasi ke- i (Sass et al., 2021). Pemodelan GPD spasial ini penting mengingat kejadian ekstrem, seperti curah hujan ekstrem atau kecepatan angin ekstrem, sangat bergantung pada lokasi geografis.

MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Estimasi parameter *univariate* pada distribusi GPD menggunakan metode *Maximum Likelihood Estimation* (MLE). MLE merupakan salah satu metode estimasi yang memaksimumkan fungsi *likelihood* untuk mendapatkan estimasi parameter nya (Jakata &

Chikobvu, 2019). Untuk MLE, misalkan data *exceedance* pada setiap lokasi berjumlah n , yang diasumsikan mengikuti distribusi GPD maka fungsi *log-likelihood* dari data $y_d = y_1, y_2, y_3, \dots, y_n$ pada satu lokasi (tanpa regularisasi), ditunjukkan sebagai berikut.

$$\ell(\sigma, \xi | y_d) = -n \ln \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{d=1}^n \ln \left(1 + \frac{\xi y_d}{\sigma} \right) \quad (6)$$

Namun, estimasi parameter spasial dapat menghadapi tantangan utama berupa *overfitting* dan ketidakstabilan numerik, terutama ketika jumlah lokasi pengamatan besar. Dalam konteks *spatial extremes*, parameter cenderung bervariasi secara halus (*smooth*) antarlokasi yang berdekatan karena sifat kontinuitas fenomena alam. Tanpa regularisasi yang tepat, estimasi metode *Maximum Likelihood Estimation* (MLE) dapat menghasilkan parameter yang bervariasi secara tidak realistis antarlokasi tetangga (Tibshirani & Saunders, 2005). Sehingga, regularisasi menjadi kunci untuk mencapai keseimbangan antara model yang baik dan *smoothness* spasial yang realistis. Pendekatan *fused penalty* khususnya dirancang untuk mengatasi masalah ini dengan mengenakan *penalty* pada perbedaan parameter antarlokasi yang berdekatan (Sass et al., 2021). Pendekatan ini memungkinkan pemodelan dependensi spasial sambil tetap menjaga kontrolabilitas dan interpretabilitas pada estimasi parameter *univariate* GPD yang diregularisasi.

FUSED PENALTY

Berdasarkan penelitian Sass et al. (2021), lokasi spasial direpresentasikan sebagai graf tidak berarah G yang terdiri dari m titik (lokasi) dan sisi E .

$$W_{p,q} = \frac{1}{\|s_p - s_q\|} \quad (7)$$

merupakan bobot invers jarak $W_{p,q}$ yang ditentukan oleh jarak spasial antara s_p dan s_q dengan titik s_i merepresentasikan lokasi spasial dan sisi $(s_p, s_q) \in E$ menunjukkan lokasi bertetangga. Hubungan ketetanggaan antar lokasi ditentukan menggunakan *Minimum Spanning Tree* (MST) sebagaimana digunakan pada artikel Sass et al. (2021), sehingga seluruh lokasi tetap terhubung dengan jumlah edge minimum.

1. Estimasi MLE dengan Regularisasi *Fused Lasso*

Fused lasso merupakan metode regularisasi yang menggabungkan dua komponen penalti, yakni

penalti L1 terhadap nilai koefisien (seperti pada lasso biasa), serta penalti L1 terhadap perbedaan antar-koefisien yang berdekatan secara spasial atau indeks (Tibshirani & Saunders, 2005). Dengan menggunakan regularisasi *fused lasso*, akan diperoleh estimasi parameter spasial yang halus (tidak berubah drastis antarlokasi yang berdekatan), model tetap mempertahankan kemampuan untuk mendeteksi diskontinuitas parameter di lokasi yang memang menunjukkan perilaku ekstrem yang berbeda.

Secara umum, fungsi penalti *fused lasso* dapat dituliskan sebagai :

$$P(\theta) = \lambda \sum_{(s_p, s_q) \in E} W_{p,q} (\theta(s_p) - \theta(s_q)) \quad (8)$$

di mana:

- $\theta(s)$ = parameter yang akan diestimasi di lokasi s
- $W_{p,q}$ = bobot spasial dari lokasi yang bertetangga (persamaan nomor 7)
- λ = parameter regularisasi
- E = pasangan lokasi bertetangga

Nilai λ sangat berpengaruh terhadap hasil estimasi. Nilai λ yang terlalu kecil menghasilkan estimasi yang terlalu bervariasi, sementara nilai λ yang terlalu besar menyebabkan parameter menjadi terlalu seragam (*oversmoothing*). Oleh karena itu, perlu dipilih nilai λ yang optimal menggunakan beberapa pendekatan seperti *k-fold cross-validation*, *Bayesian Information Criterion* (BIC), dan *Takeuchi Information Criterion* (TIC) (Zhang et al., 2019).

Dalam pendekatan regularisasi *fused lasso*, *log-likelihood* pada persamaan (6) diperluas dengan menambahkan penalti spasial *fused lasso* (persamaan nomor 8) terhadap perbedaan nilai parameter skala (σ) antarlokasi yang bertetangga (Sass et al., 2021):

$$\ell_{pen}(\sigma) = \ell_s(\sigma_s, \hat{\xi} | y_d) - P(\sigma) \quad (9)$$

$$\begin{aligned} & \ell_{pen}(\sigma) \\ &= \sum_{s=1}^m \sum_{d=1}^n \ell_s(\sigma_s, \hat{\xi} | y_d) \\ & - \lambda_\sigma \sum_{(s_p, s_q) \in E} W_{p,q} (\sigma(s_p) - \sigma(s_q)) \end{aligned} \quad (10)$$

di mana:

- $\ell_s(\sigma_s, \hat{\xi} | y)$ adalah *log-likelihood* (persamaan nomor 6) dari data d pada lokasi s
- E adalah kumpulan pasangan lokasi yang bertetangga

- λ_σ merupakan parameter regularisasi untuk parameter skala

Estimasi parameter ($\hat{\sigma}$) diperoleh dengan memaksimalkan $\ell_{pen}(\sigma)$.

$$\hat{\sigma} = \arg \max_{\sigma} \ell_{pen}(\sigma) \quad (11)$$

Karena bentuk fungsi pada persamaan nomor 11 tidak memiliki solusi analitik tertutup (*closed-form*), digunakan algoritma optimasi numerik. Metode yang dapat digunakan ada *Alternating Direction Method of Multipliers* (ADMM), *Proximal Gradient Descent*, dan *Newton Raphson* dengan modifikasi pada bagian penalti (Sass et al., 2021). Sementara untuk parameter bentuk (ξ), akan diestimasi menggunakan MLE biasa tanpa regularisasi karena diasumsikan konstan (tidak dimodelkan secara spasial).

2. Estimasi MLE dengan Regularisasi *Fused Ridge*

Fused ridge merupakan metode regularisasi yang menggabungkan dua komponen penalti, yakni penalti *ridge* terhadap nilai koefisien (seperti pada regresi *ridge*), serta penalti terhadap perbedaan antar-koefisien yang berdekatan secara spasial untuk halusnya estimasi (Ishwaran & Rao, 2005). Dengan menggunakan regularisasi *fused ridge*, akan diperoleh estimasi parameter spasial yang lebih stabil, di mana model tetap mempertahankan fleksibilitas untuk menangkap perbedaan struktur antar lokasi. Dibandingkan regularisasi *ridge* biasa, *fused ridge* lebih sesuai digunakan dalam konteks spasial karena mempertimbangkan kedekatan antarlokasi, serta dapat digunakan untuk mengontrol kehalusan perubahan parameter antarlokasi (Sass et al., 2021).

Secara umum, fungsi penalti *fused ridge* dapat dituliskan sebagai :

$$P(\theta) = \lambda \sum_{(s_p, s_q) \in E} W_{p,q} (\theta(s_p) - \theta(s_q))^2 \quad (12)$$

di mana:

- $\theta(s)$ = parameter yang akan diestimasi di lokasi s
- $W_{p,q}$ = bobot spasial dari lokasi yang bertetangga (persamaan nomor 7)
- λ = parameter regularisasi
- E = pasangan lokasi bertetangga

Nilai λ sangat berpengaruh terhadap hasil estimasi. Nilai λ yang terlalu kecil menghasilkan estimasi yang terlalu bervariasi, sementara nilai λ yang terlalu besar menyebabkan parameter menjadi

terlalu seragam (*oversmoothing*). Oleh karena itu, perlu dipilih nilai λ yang optimal menggunakan beberapa pendekatan seperti *k-fold cross-validation*, *Bayesian Information Criterion* (BIC), dan *Takeuchi Information Criterion* (TIC) (Zhang et al., 2019).

Dalam pendekatan regularisasi *fused ridge*, *log-likelihood* pada persamaan (6) diperluas dengan menambahkan penalti spasial *fused ridge* (persamaan nomor 12) terhadap perbedaan nilai parameter skala (σ) antarlokasi yang bertetangga (Sass et al., 2021):

$$\ell_{pen}(\sigma) = \ell_s(\sigma_s, \xi | y_d) - P(\sigma) \quad (13)$$

$$\begin{aligned} \ell_{pen}(\sigma) &= \sum_{s=1}^m \sum_{d=1}^n \ell_s(\sigma_s, \xi | y_d) \\ &\quad - \lambda_\sigma \sum_{(s_p, s_q) \in E} W_{p,q} (\sigma(s_p) - \sigma(s_q))^2 \end{aligned} \quad (14)$$

di mana:

- $\ell_s(\sigma_s, \xi | y)$ adalah *log-likelihood* (persamaan nomor 2.6) dari data d pada lokasi s
- E adalah kumpulan pasangan lokasi yang bertetangga
- λ_σ merupakan parameter regularisasi untuk parameter skala

Estimasi parameter ($\hat{\sigma}$) diperoleh dengan memaksimalkan $\ell_{pen}(\sigma)$.

$$\hat{\sigma} = \arg \max_{\sigma} \ell_{pen}(\sigma) \quad (15)$$

Karena bentuk fungsi pada persamaan nomor 15 tidak memiliki solusi analitik tertutup (*closed-form*), digunakan algoritma optimasi numerik. Metode yang dapat digunakan ada *Alternating Direction Method of Multipliers* (ADMM), *Proximal Gradient Descent*, dan Newton Raphson dengan modifikasi pada bagian penalti (Sass et al., 2021). Sementara untuk parameter bentuk (ξ) akan diestimasi menggunakan MLE biasa tanpa regularisasi karena diasumsikan konstan (tidak dimodelkan secara spasial).

UJI KESESUAIAN DISTRIBUSI

Untuk mendapatkan model GPD yang baik dalam *Spatial Extreme Value*, langkah yang paling penting ialah menguji kesesuaian data dengan distribusi GPD. Dilakukan validasi kesesuaian distribusi dengan 2 pengujian yakni; Uji *Anderson-Darling* untuk mengecek apakah distribusi GPD cocok dengan data dan Uji *Likelihood Ratio Test* untuk menguji kesesuaian distribusi.

1. Uji Anderson-Darling

Uji ini digunakan untuk mengukur kesesuaian antara distribusi teoritik dan distribusi empiris dari data dan merupakan pengembangan dari uji *Kolmogorov-Smirnov*, namun dengan penekanan lebih besar pada bagian ekor distribusi sehingga lebih sensitif terhadap deviasi yang terjadi di ekor distribusi (Stephens, 1974). Penggunaan uji ini mengacu pada studi Choulakian & Stephens (2001) yang secara khusus mengembangkan uji *Anderson-Darling* untuk distribusi GPD. Dalam pengujian ini, data yang telah difit ke model GPD ditransformasikan ke distribusi *uniform* [0,1] menggunakan cdf dari GPD dengan parameter hasil estimasi yang kemudian digunakan untuk menghitung statistik *Anderson-Darling* dari data hasil transformasi tersebut.

Hipotesis yang diuji:

H_0 : Data *exceedance* mengikuti distribusi GPD dengan parameter tertentu

H_1 : Data *exceedance* tidak mengikuti distribusi GPD

Statistik uji:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln(U_{(i)}) + \ln(1 - U_{(n+1-i)})] \quad (16)$$

dengan n = banyaknya data *exceedence*

$U_{(i)}$ = data yang telah ditransformasikan dan diurutkan

Digunakan tingkat signifikansi sebesar $\alpha = 5\%$, maka diambil keputusan data *exceedance* mengikuti distribusi GPD dengan parameter tertentu jika $A^2 < 2,492$ di mana 2,492 merupakan nilai kritis yang diperoleh dari tabel *Anderson-Darling*.

2. Uji Likelihood Ratio Test

Uji ini merupakan metode umum dalam statistika dalam membandingkan dua model yang saling berhubungan, yakni model penuh dan model tereduksi. Sehingga, uji *likelihood ratio test* dapat dilakukan untuk mengetahui tipe distribusi yang tepat dari GPD, dengan pengujian hipotesis sebagai berikut (Ramadani, 2015).

Hipotesis yang diuji:

H_0 : $\xi = 0 \rightarrow$ Tipe distribusi Eksponensial

H_1 : $\xi \neq 0 \rightarrow$ Tipe distribusi Pareto atau tipe distribusi Beta

Statistik uji:

$$\Lambda = 2[\ell_{GPD} - \ell_{EXP}] \quad (17)$$

dengan $\ell_{GPD} = \log\text{-likelihood}$ model GPD dengan parameter $\hat{\sigma}$ dan $\hat{\xi}$

$\ell_{EXP} = \log\text{-likelihood}$ model GPD dengan parameter $\hat{\sigma}$

Digunakan tingkat signifikansi sebesar $\alpha = 5\%$, maka diambil keputusan data tipe distribusi Pareto atau tipe distribusi Beta jika $\Lambda > 3,841$ di mana 3,841 merupakan nilai kritis distribusi chi-squared ($\chi^2_{\alpha=0,95;df=1}$).

RETURN LEVEL

Return level merupakan salah satu *output* penting dalam analisis nilai ekstrem karena menggambarkan tingkat kejadian ekstrem yang diharapkan terlampaui sekali dalam periode waktu tertentu. Untuk data univariat, *return level* x_T dihitung sebagai nilai yang dilampaui dengan probabilitas $\frac{1}{T}$, di mana T adalah *return period*. Berdasarkan model GPD, *return level* univariat dirumuskan sebagai berikut (Coles, 2001) :

$$x_T = u + \frac{\sigma}{\xi} \left[(T \cdot \Pr(X > u))^{\xi} - 1 \right], \quad (18)$$

untuk $\xi \neq 0$

dengan u adalah *threshold*, $\Pr(X > u)$ adalah probabilitas data yang melebihi *threshold*.

CURAH HUJAN

Curah hujan adalah jumlah air hujan yang jatuh di permukaan tanah datar selama periode tertentu, yang diukur dalam satuan tinggi milimeter (mm) di atas permukaan horizontal dengan asumsi tidak terjadi penguapan, peresapan, atau aliran air (BMKG, n.d.). Menurut BMKG, (n.d.), curah hujan 1 mm didefinisikan sebagai air hujan setinggi 1 mm yang tertampung pada tempat datar seluas 1 m² tanpa adanya penguapan, aliran, atau peresapan.

Di wilayah tropis seperti Indonesia, curah hujan cenderung tinggi dan sangat dipengaruhi oleh proses konveksi serta pembentukan awan hujan panas. Daerah pegunungan umumnya menerima curah hujan lebih tinggi dibandingkan dengan dataran rendah. Topografi pegunungan memiliki pengaruh besar terhadap curah hujan. Hal ini dipengaruhi oleh tingkat penguapan dan pembentukan awan untuk menghasilkan hujan (Agus et al., 2020).

Data curah hujan sangat penting untuk berbagai sektor, seperti pertanian untuk menentukan musim tanam, pengelolaan sumber daya air, mitigasi bencana, dan perencanaan tata ruang dan pembangunan wilayah (Adminhandal, 2024).

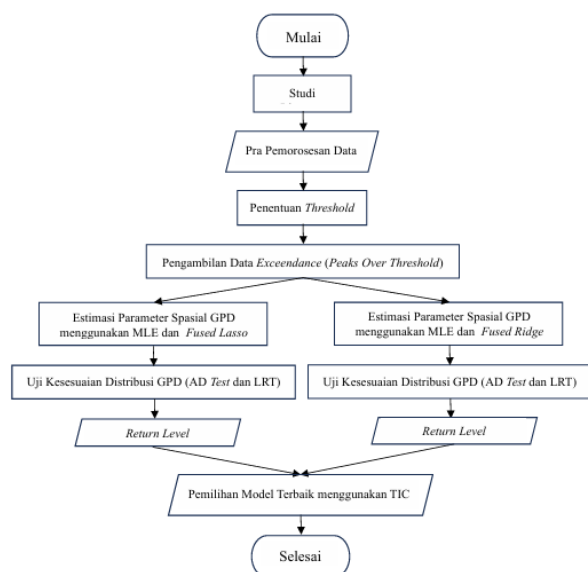
Sehingga, dapat dikatakan bahwa curah hujan merupakan salah satu parameter cuaca yang sangat penting dan berperan besar dalam kehidupan manusia serta lingkungan.

METODE

DATA PENELITIAN

Data yang digunakan merupakan data sekunder. Data sekunder merupakan data yang berasal dari jurnal dan referensi terkait yang sudah pernah dipublikasikan oleh pihak tertentu. Data yang akan digunakan dalam penelitian ini yakni data curah hujan tahun 2014 – 2024 dari stasiun pengamatan FL Tobing, Aek Gondang, Kualanamu, Maritim Belawan, dan stasiun pengamatan Binaka di Provinsi Sumatera Utara. Data ini diperoleh dari web resmi BMKG (<https://www.bmkg.go.id/>). Kelima stasiun pengamatan dipilih berdasarkan representasi spasial dan peran klimatologis dari masing-masing wilayah. Stasiun Meteorologi FL Tobing mewakili wilayah Tapanuli Tengah, Stasiun Meteorologi Aek Gondang mewakili wilayah Padang Lawas Utara, Stasiun Meteorologi Kualanamu mewakili wilayah Deli Serdang, Stasiun Meteorologi Maritim Belawan mewakili wilayah Medan, dan Stasiun Meteorologi Binaka mewakili wilayah Kepulauan Nias.

DIAGRAM ALIR PENELITIAN



Gambar 1 Diagram Alir Penelitian

Penelitian ini akan dilakukan melalui tahapan pra-pemrosesan data, penentuan threshold,

pengambilan data *exceedance*, estimasi parameter menggunakan metode *Maximum Likelihood Estimation* (MLE) dengan regularisasi *fused lasso* dan *fused ridge*, uji kesesuaian distribusi, serta evaluasi model menggunakan *Takeuchi Information Criterion* (TIC). Selain itu, akan dilakukan perhitungan *return level* untuk menggambarkan risiko kejadian curah hujan ekstrem pada berbagai periode ulang.

HASIL DAN PEMBAHASAN

Pada bagian ini akan diberikan hasil dari analisis data yang telah dilakukan mulai dari pra-pemrosesan data untuk mengumpulkan serta mengetahui karakteristik data, kemudian menjelaskan prosedur untuk mendapatkan nilai estimasi parameter menggunakan regularisasi *fused lasso* dan *fused ridge*. Proses tersebut dilakukan untuk memprediksi *joint return period* curah hujan di Provinsi Sumatera Utara. Selain itu, dilakukan evaluasi fungsi distribusi gabungan dari regularisasi *fused lasso* dan *fused ridge* untuk memperoleh hasil terbaik dengan menggunakan metode TIC. Seluruh proses pengolahan data, analisis, dan visualisasi pada penelitian ini dilakukan menggunakan perangkat lunak *RStudio*.

PRA-PEMROSESAN DATA

Pra-pemrosesan data dilakukan dengan mengumpulkan data sekunder curah hujan dari BMKG pada 5 Stasiun Meteorologi di Provinsi Sumatera Utara Tahun 2014 – 2024 yang kemudian diidentifikasi karakteristik data menggunakan statistika deskriptif.

Tabel 1 Hasil Statistika Deskriptif

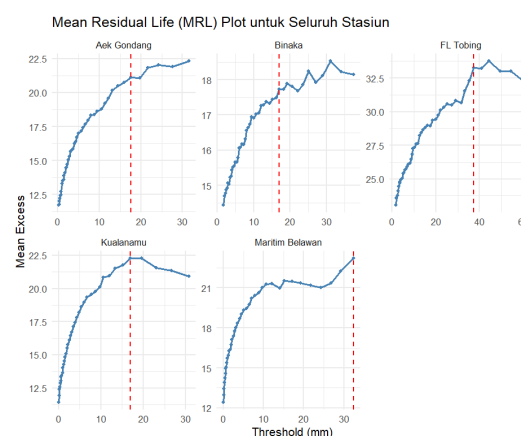
Stasiun Pengamatan	Min.	Rata – Rata	Maks.	Std. Dev.
Aek Gondang	0	5,76	172	13,6
Binaka	0	8,17	136	14,2
FL Tobing	0	12,5	242	23,7
Kualanamu	0	4,96	132	12,8
Maritim Belawan	0	5,68	159	14,0

Berdasarkan Tabel 1, dapat disimpulkan stastistika deskriptif untuk variabel curah hujan, dengan jumlah hari hujan pada tahun 2014-2024 sebanyak 2032 hari untuk stasiun pengamatan Aek Gondang; 2313 hari untuk stasiun pengamatan Binaka; 1886 hari untuk stasiun pengamatan FL tobing; 1527 hari untuk stasiun pengamatan Kualanamu; dan 1587 hari untuk stasiun pengamatan Maritim Belawan. Perhitungan rata-rata curah hujan didapatkan dengan membagi jumlah curah hujan dengan jumlah hari hujan. Analisis pada tabel di atas menunjukkan

data curah hujan memiliki rata-rata yang relatif rendah dan standar deviasi yang tinggi. Hal ini berarti data memiliki persebaran yang cukup berbeda dari rata-rata curah hujan.

PENENTUAN THRESHOLD

Pendekatan kuantitatif dengan persentil memberikan nilai awal *threshold* berdasarkan proporsi data tertinggi, sedangkan metode MRL digunakan sebagai validasi visual untuk memeriksa kestabilan rata-rata residual di atas kandidat *threshold* tersebut.



Gambar 2 Penentuan *Threshold*

Pendekatan *Mean Residual Life* (MRL) pada Gambar 2 digunakan sebagai validasi visual untuk menilai kestabilan rata-rata residual di atas kandidat *threshold*. Berdasarkan hasil tersebut, penentuan nilai *threshold* akhir dilakukan dengan mempertimbangkan pendekatan persentil yang memberikan nilai awal secara kuantitatif. Nilai *threshold* yang diperoleh untuk masing-masing stasiun pengamatan disajikan pada Tabel 2.

Tabel 2 Hasil Penentuan Nilai *Threshold*

Stasiun Pengamatan	Persentil	Threshold	MRL
Aek Gondang	90 %	17,6	21,121
Binaka	84 %	17,0	17,725
FL Tobing	90 %	37,4	33,271
Kualanamu	91 %	17,0	22,286
Maritim Belawan	95 %	32,5	23,239

Berdasarkan Tabel 2, diperoleh nilai *threshold* yang berbeda untuk setiap stasiun pengamatan. Perbedaan tersebut menunjukkan bahwa karakteristik curah hujan ekstrem pada masing-masing lokasi tidak seragam sehingga diperlukan nilai *threshold* yang disesuaikan dengan kondisi lokal. Nilai *threshold* tersebut selanjutnya digunakan untuk mengidentifikasi data yang melebihi *threshold* (*exceedances*) pada setiap stasiun pengamatan. Stasiun pengamatan Aek Gondang memiliki 400 data; stasiun pengamatan Binaka memiliki 642 data;

stasiun pengamatan FL Tobing memiliki 401 data; stasiun pengamatan Kualanamu memiliki 360 data; dan stasiun pengamatan Maritim Belawan memiliki 200 data. Jumlah *exceedance* yang diperoleh bervariasi antarstasiun, tergantung pada karakteristik hujan ekstrem di masing-masing lokasi.

ESTIMASI PARAMETER SPASIAL GPD

Untuk nilai estimasi parameter bentuk (ξ) yang diasumsikan konstan secara spasial, diperoleh menggunakan metode MLE melalui bantuan *package R*, yaitu memaksimumkan persamaan 6 terhadap seluruh data *exceedance*, sehingga diperoleh satu nilai parameter bentuk yang berlaku sama pada seluruh stasiun pengamatan. Berdasarkan hasil dari estimasi, seluruh stasiun memiliki nilai parameter bentuk (ξ) yang sama, yakni 0,043.

Fused Lasso

Untuk mendapatkan nilai parameter skala dilakukan dengan memaksimumkan fungsi yang diperluas tersebut sebagaimana dijelaskan pada persamaan 11, sehingga diperoleh nilai estimasi parameter skala (σ) pada tiap stasiun pengamatan, sebagai berikut.

Tabel 3 Nilai Parameter Skala (σ) dari Regularisasi *Fused Lasso*

Stasiun Pengamatan	Parameter skala (σ)
Aek Gondang	23,847
Binaka	23,563
FL Tobing	23,847
Kualanamu	23,331
Maritim Belawan	23,331

Nilai parameter skala (σ) dari regularisasi *fused lasso* cenderung hampir sama pada setiap stasiun pengamatan, dan hal ini menunjukkan bahwa penalti *fused lasso* berhasil mengurangi perbedaan nilai parameter antarlokasi yang bertetangga, sehingga menghasilkan estimasi parameter yang lebih homogen secara spasial.

Fused Ridge

Untuk mendapatkan nilai parameter skala dilakukan dengan memaksimumkan fungsi yang diperluas tersebut sebagaimana dijelaskan pada persamaan 15, sehingga diperoleh nilai estimasi parameter skala (σ) pada tiap stasiun pengamatan, sebagai berikut.

Tabel 4 Nilai Parameter Skala (σ) dari Regularisasi *Fused Ridge*

Stasiun Pengamatan	Parameter skala (σ)
Aek Gondang	20,585
Binaka	17,109
FL Tobing	30,342
Kualanamu	21,746
Maritim Belawan	22,310

Nilai parameter skala (σ) dari regularisasi *fused ridge* berbeda pada setiap stasiun yang menunjukkan adanya perbedaan karakteristik variasi curah hujan ekstrem di masing-masing lokasi pengamatan. Hal ini mengindikasikan bahwa regularisasi *fused ridge* mampu menghaluskan perbedaan parameter pada lokasi yang bertetangga tanpa menyatukan nilainya sepenuhnya, sehingga keragaman spasial antarstasiun tetap dapat dipertahankan.

UJI AD DAN LRT

Nilai estimasi parameter bentuk (ξ) memiliki nilai positif ($\xi > 0$) dan relatif kecil yakni sebesar 0,043. Nilai tersebut mendekati 0 ($\xi \approx 0$) untuk seluruh stasiun, hal ini menunjukkan bahwa distribusi GPD yang diperoleh cenderung mendekati distribusi eksponensial. Validasi terhadap asumsi distribusi GPD dilakukan melalui uji Anderson-Darling (AD) pada persamaan 16 dan Likelihood Ratio Test (LRT) pada persamaan 17.

Tabel 5 Hasil Uji AD dan Uji LRT

Metode	Stasiun Pengamatan	Nilai Hitung AD	Nilai Hitung LRT	Nilai Kritis <i>Chi-Square</i>	Kesimpulan Uji LRT
<i>Fused Lasso</i>	Aek Gondang	5,045	-8,157	3,8415	Gagal Tolak H_0
	Binaka	23,878	-57,809	3,8415	Gagal Tolak H_0
	FL Tobing	16,361	-35,893	3,8415	Gagal Tolak H_0
	Kualanamu	0,954	-5,245	3,8415	Gagal Tolak H_0
	Maritim Belawan	0,358	-1,686	3,8415	Gagal Tolak H_0
<i>Fused Ridge</i>	Aek Gondang	0,491	1,524	3,8415	Gagal Tolak H_0
	Binaka	0,196	-0,273	3,8415	Gagal Tolak H_0
	FL Tobing	1,042	-2,139	3,8415	Gagal Tolak H_0
	Kualanamu	1,104	-3,226	3,8415	Gagal Tolak H_0
	Maritim Belawan	0,486	-1,412	3,8415	Gagal Tolak H_0

Berdasarkan hasil **Error! Reference source not found.**, diperoleh bahwa nilai statistik AD pada masing-masing stasiun relatif kecil. Hal ini menunjukkan bahwa distribusi GPD mampu menunjukkan data ekstrem dengan cukup baik. Selain itu, hasil uji LRT menunjukkan bahwa nilai statistik LRT pada sebagian besar stasiun dibandingkan dengan nilai kritis distribusi *Chi-Square* pada taraf signifikansi 5%. Dengan demikian, hipotesis nol yang menyatakan bahwa data mengikuti distribusi eksponensial ($\xi = 0$) gagal ditolak di seluruh stasiun, sehingga dapat disimpulkan bahwa parameter bentuk mendekati nol sehingga distribusi GPD yang diperoleh cenderung mendekati distribusi eksponensial.

RETURN LEVEL

Hasil perhitungan *return level* univariat untuk stasiun lainnya serta untuk periode ulang yang berbeda

dihitung menggunakan prosedur yang sama dengan bantuan package R.

Tabel 6 Hasil *Return Level*

Stasiun Pengamatan	<i>Fused Lasso</i>			<i>Fused Ridge</i>		
	5 Tahun	10 Tahun	25 Tahun	5 Tahun	10 Tahun	25 Tahun
Aek Gondang	228,757	251,884	283,528	199,872	219,835	247,150
Binaka	225,642	248,494	279,761	168,498	185,091	207,794
FL Tobing	248,557	271,684	303,327	306,068	335,494	375,756
Kualanamu	223,590	246,217	277,176	209,556	230,646	259,502
Maritim Belawan	239,090	261,717	292,676	230,044	251,680	281,284

Berdasarkan hasil tersebut, diperoleh bahwa nilai *return level* cenderung meningkat seiring dengan bertambahnya periode ulang.

PERBANDINGAN TIC

Tujuan dari subbab ini ialah mengevaluasi kinerja model yang dihasilkan dari masing-masing metode regularisasi. Evaluasi ini bertujuan untuk menentukan model terbaik dalam memodelkan data ekstrem curah hujan. Pada penelitian ini, pemilihan model dilakukan menggunakan *Takeuchi Information Criterion* (TIC), yang merupakan pengembangan dari *Akaike Information Criterion* (AIC) dan lebih sesuai digunakan pada kondisi model *penalty* (Konishi & Kitagawa, 1996). Nilai TIC yang lebih kecil menunjukkan model dengan keseimbangan yang lebih baik antara kompleksitas model dan kecocokan terhadap data.

Tabel 7 Hasil Perhitungan Nilai TIC

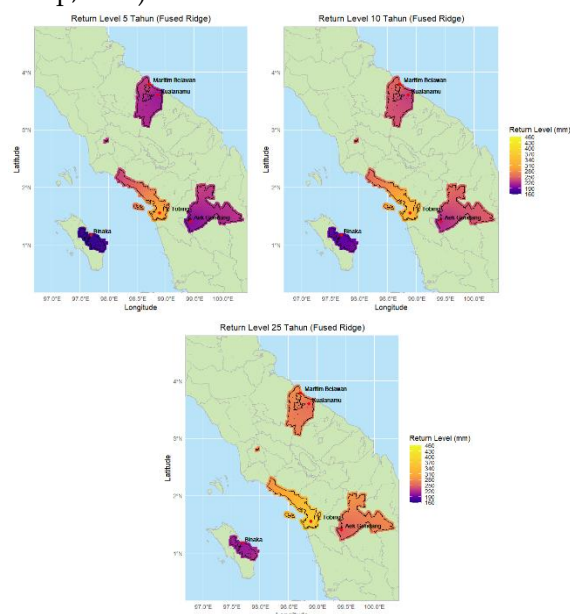
Metode Regularisasi	Nilai TIC
<i>Fused Lasso</i>	103245,1
<i>Fused Ridge</i>	100913,9

Berdasarkan hasil Tabel 7, diperoleh bahwa nilai TIC untuk metode *fused lasso* dan *fused ridge* relatif berdekatan. Hal ini mengindikasikan bahwa kedua metode regularisasi memberikan performa model yang hampir serupa dalam memodelkan data ekstrem curah hujan pada masing-masing stasiun. Tetapi model dengan nilai TIC yang lebih kecil sebesar 100913,9, yakni model GPD dengan estimasi parameter menggunakan metode regularisasi *fused ridge*, dianggap sebagai model yang lebih baik diantara model yang dibandingkan. Hasil ini juga sejalan dengan analisis sebelumnya pada estimasi parameter, di mana kedua metode *fused ridge* tetap dapat menjaga variasi dalam menghasilkan nilai parameter.

VISUALISASI SPASIAL

Peta *return level* univariat ditampilkan untuk beberapa periode ulang, yaitu 5 tahun, 10 tahun, dan 25 tahun. Visualisasi dilakukan menggunakan

metode interpolasi *Inverse Distance Weighting* (IDW) untuk menghasilkan gambaran distribusi spasial yang lebih kontinu antarstasiun pengamatan (Li & Heap, 2011).



Gambar 3 Peta *Return Level* Periode 5, 10, dan 25 Tahun

Secara umum, nilai *return level* cenderung meningkat seiring dengan bertambahnya periode ulang. Hal ini menunjukkan bahwa kejadian curah hujan ekstrem dengan intensitas yang lebih tinggi memiliki peluang yang lebih kecil untuk terjadi, namun tetap mungkin terjadi dalam jangka waktu yang lebih panjang. Selain itu, stasiun pengamatan FL Tobing menunjukkan nilai *return level* yang relatif lebih tinggi dibandingkan dengan stasiun lainnya, yang mengindikasikan potensi risiko curah hujan ekstrem yang lebih besar pada lokasi tersebut berdasarkan estimasi univariat yang dihasilkannya.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis, model *Generalized Pareto Distribution* (GPD) dengan parameter skala yang diestimasi menggunakan *Maximum Likelihood Estimation* (MLE) serta regularisasi *fused lasso* dan *fused ridge* mampu memodelkan karakteristik curah hujan ekstrem pada beberapa lokasi di Provinsi Sumatera Utara. Dengan parameter bentuk yang diasumsikan konstan sebesar $\xi = 0,043$, diperoleh bahwa nilai *return level* meningkat seiring bertambahnya periode ulang pada seluruh stasiun pengamatan. Nilai *return level* tertinggi secara

konsisten diperoleh pada Stasiun Meteorologi FL Tobing, sedangkan nilai terendah bervariasi bergantung pada metode regularisasi yang digunakan.

Evaluasi model menggunakan *Takeuchi Information Criterion* (TIC) menunjukkan bahwa model GPD dengan regularisasi *fused ridge* memberikan performa yang lebih baik dibandingkan dengan *fused lasso*. Hal ini ditunjukkan oleh nilai TIC yang lebih kecil dengan selisih sebesar 2331,2. Dengan demikian, model GPD dengan regularisasi *fused ridge* direkomendasikan sebagai model terbaik untuk prediksi spasial curah hujan ekstrem di Provinsi Sumatera Utara.

SARAN

Dari kesimpulan yang telah diperoleh, prediksi curah hujan dapat dilakukan dengan pengembangan model prediktif dan regularisasi. Penelitian selanjutnya dapat mempertimbangkan penggunaan kombinasi regularisasi dengan model ekstrem spasial lainnya, seperti model *Generalized Extreme Value* (GEV), *max-stable process* (MSP), atau metode prediksi lainnya agar diperoleh perbandingan hasil yang lebih beragam. Selain itu, penggunaan data dengan karakteristik spasial yang lebih heterogen perlu dipertimbangkan agar pengaruh regularisasi terhadap estimasi parameter dapat terlihat lebih jelas.

Penelitian ini mengasumsikan parameter bentuk GPD bernilai konstan pada seluruh lokasi untuk menyederhanakan proses estimasi. Oleh karena itu, penelitian selanjutnya disarankan untuk memodelkan parameter bentuk secara spasial sehingga variasi karakteristik ekor distribusi antarlokasi dapat dipresentasikan dengan lebih baik dan menghasilkan model yang lebih fleksibel.

DAFTAR PUSTAKA

- (IPCC), I. P. on C. C. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.
- Adminhandal. (2024). *Studi Hidrologi: Dasar Pengelolaan Sumber Daya Air dan Mitigasi Bencana*. Kreasi Handal Selaras. <https://www.handalselaras.com/studi-hidrologi-dasar-pengelolaan-sumber-daya-air-dan-mitigasi-bencana/>
- Agus, L. Y., Dinar, D. A. P., & Sarino. (2020). Pengaruh Perbedaan Topografi Terhadap Variasi Intensitas Curah Hujan Pada Prediksi Kehilangan Tanah. *Sriwijaya University Institutional Repository*. <http://repository.unsri.ac.id/id/eprint/7594>
- Anifah, U. (2014). *Pemilihan Threshold pada Pemodelan Nilai Ekstrem Curah Hujan di Sentra Produksi Padi Jawa Timur dengan MRLP-Baru (New Technique for Mean Residual Life Plot)* [Institut Teknologi Sepuluh Nopember]. <https://repository.its.ac.id/1301/2/1310100057-Undergraduate-Thesis.pdf>
- BMKG, W. I. (n.d.). *Daftar Istilah Klimatologi*.
- BNPB. (2024). *Data Bencana Indonesia 2024*.
- Chen, J., Lei, X., Zhang, L., & Peng, B. (2015). Using Extreme Value Theory Approaches to Forecast the Probability of Outbreak of Highly Pathogenic Influenza in Zhejiang, China. *PLOS One*, 10. <https://doi.org/10.1371/journal.pone.0118521>
- Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(4), 478–484. <https://doi.org/10.1198/00401700152672573>
- Coles, S. (2001). *An Introduction to Statistical Modelling of Extreme Values*. Great Britain : Springer-Verlag London Berlin Heidelberg.
- Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2), 161–186. <https://doi.org/10.1214/11-STS376>
- de Fondeville, R., & Davison, A. C. (2022). Functional peaks-over-threshold analysis. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 84(4), 1392–1422. <https://doi.org/10.1111/rssb.12498>
- DIBI. (n.d.). *Data Informasi Bencana Indonesia*. https://dibi.bnpb.go.id/superset/dashboard/1/?standalone=0&expand_filters=0
- Far, S. S., & Wahab, A. K. A. (2016). Evaluation of Peaks-Over-Threshold Method. *Ocean Science*. <https://doi.org/10.5194/os-2016-47>
- Fisher, R. A., Tippett, L. H. C., Gumbel, E. J., & Weibull, W. (2013). *An Introduction to Extreme Value Statistics*. *An Introduction to Extreme Value Statistics*.
- Gao, J., Ma, P., Du, J., & Huang, X. (2021). Spatial distribution of extreme precipitation in the Tibetan Plateau and effects of external forcing factors based on Generalized Pareto Distribution. *Water Science and Technology: Water Supply*, 21(3), 1253–1262. <https://doi.org/10.2166/ws.2020.365>
- Ishwaran, H., & Rao, J. S. (2005). Spike and Slab Variable Selection: Frequentist and Bayesian Strategies. *The Annals of Statistics*, 33, 730–773.

- <https://doi.org/10.1214/009053604000001147>
- Jakata, O., & Chikobvu, D. (2019). Modelling extreme risk of the South African Financial Index (J580) using the generalised Pareto distribution. *Journal of Economic and Financial Sciences*, 12(1), 1-7. <https://doi.org/10.4102/jef.v12i1.407>
- Konishi, S., & Kitagawa, G. (1996). Generalised Information Criteria in Model Selection. *Biometrika*, 875-890. <https://doi.org/10.1093/biomet/83.4.875>
- Lee, S., & Kim, J. H. T. (2019). Exponentiated generalized Pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics - Theory and Methods*, 48(8), 2014-2038. <https://doi.org/10.1080/03610926.2018.1441418>
- Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6(3-4), 228-241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>
- Omari, C., & Ngunyi, A. (2021). The Predictive Performance of Extreme Value Analysis Based-Models in Forecasting the Volatility of Cryptocurrencies. *Journal of Mathematical Finance*, 11. <https://doi.org/10.4236/jmf.2021.113025>
- Prasetyo, B., Irwandi, H., & Pusparini, N. (2018). Karakteristik Curah Hujan Berdasarkan Ragam Topografi di Sumatera Utara. *Jurnal Sains & Teknologi Modifikasi Cuaca*. <https://doi.org/10.29122/jstmc.v19i1.2787>
- Ramadani, I. R. (2015). *Spatial Extreme Value Modeling Dengan Max-Stable Processes Model Smith Dan Brown-Resnick*.
- Sass, D., Li, B., & Reich, B. J. (2021). Flexible and Fast Spatial Return Level Estimation Via a Spatially Fused Penalty. *Journal of Computational and Graphical Statistics*, 30(4), 1124-1142. <https://doi.org/10.1080/10618600.2021.1938584>
- Shao, X., Hazra, A., Richards, J., & Huser, R. (2024). Flexible Modeling of Nonstationary Extremal Dependence using Spatially-Fused LASSO and Ridge Penalties. *Technometrics*, 0(0), 1-34. <https://doi.org/10.1080/00401706.2024.2388549>
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730-737. <https://doi.org/10.1080/01621459.1974.10480196>
- Tibshirani, R., & Saunders, M. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society*, 91-108. <https://doi.org/10.1111/j.1467-9868.2005.00490.x>
- Wadsworth, J. L., & Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 253-272. <https://doi.org/10.1093/biomet/asr080>
- Ward, P. J., Jongman, B., Weiland, S. F., Bouwman, A., Beek, R. van, Bierkens, M. F. P., Ligtoet, W., & Winsemius, H. C. (2013). Assessing flood risk at the global scale: model setup, results, and sensitivity. *Environmental Research Letter*. <https://doi.org/10.1088/1748-9326/8/4/044019>
- Zhang, Z., Yan, J., & Zhang, X. (2019). A Comparison of AIC and TIC in Regularized Copula-Based Extreme Models. *Computational Statistics & Data Analysis*, 179-193.