

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perubahan iklim global yang dipicu oleh peningkatan emisi gas rumah kaca telah menyebabkan perubahan signifikan pada pola curah hujan di berbagai belahan dunia, termasuk Indonesia (Ainurrohmah & Sudarti, 2022). Curah hujan ekstrem, yang ditandai dengan intensitas tinggi dan durasi singkat, dapat memicu bencana alam seperti banjir dan tanah longsor. Kejadian ini berpotensi mengancam berbagai sektor seperti pertanian, infrastruktur, dan kesejahteraan masyarakat, terutama di wilayah tropis yang rentan terhadap dampak perubahan iklim (Nurlatifah dkk., 2023).

Sulawesi Selatan memiliki iklim tropis dengan pola musim hujan dan kemarau yang cukup jelas. Curah hujan di wilayah ini dipengaruhi oleh angin muson, sirkulasi atmosfer global, dan topografi yang beragam, sehingga menghasilkan distribusi curah hujan yang tidak merata (BMKG, 2024). Beberapa wilayah di Sulawesi Selatan, termasuk Kabupaten Pangkep, mengalami curah hujan tinggi saat puncak musim hujan pada bulan Desember hingga Februari. Dalam beberapa dekade terakhir, perubahan iklim telah menyebabkan peningkatan kejadian ekstrem, seperti hujan dengan intensitas tinggi dalam durasi singkat, yang sering kali memicu banjir dan tanah longsor.

Kabupaten Pangkep, yang perekonomiannya bergantung pada sektor pertanian dan perikanan, menjadi salah satu daerah yang terdampak (BPS, 2024). Curah hujan ekstrem sering mengganggu panen padi dan produksi garam, yang bergantung pada pola curah hujan yang teratur. Selain itu, pada bulan September 2024, petani rumput laut di Desa Bulu Cindea, Kecamatan Bungoro, Kabupaten Pangkep, menghadapi tantangan serius akibat curah hujan ekstrem (Thahirah, 2024). Angin kencang dan gelombang tinggi menyebabkan gagal panen dan penurunan kualitas rumput laut, yang berdampak langsung pada perekonomian petani setempat. Oleh karena itu, prediksi curah hujan ekstrem menjadi langkah penting untuk membantu masyarakat mengantisipasi dampaknya.

Curah hujan ekstrem adalah kondisi ketika jumlah curah hujan jauh melebihi atau berada di bawah rata-rata normal dalam suatu periode tertentu. Berdasarkan BMKG (2011), curah hujan ekstrem didefinisikan sebagai curah hujan harian melebihi 100 mm atau bulanan lebih dari 400 mm. *World Meteorological Organization* (WMO) juga mengklasifikasikan curah hujan ekstrem sebagai kejadian cuaca yang berada di persentil tertinggi dari distribusi historis curah hujan suatu wilayah. Fenomena ini dipengaruhi oleh berbagai faktor seperti topografi, kelembapan udara, serta interaksi antara lautan dan atmosfer, sehingga menjadikannya sulit untuk diprediksi secara akurat.

Prediksi curah hujan ekstrem dapat dilakukan dengan memanfaatkan data informasi iklim global yang dihasilkan oleh *Global Circulation Model* (GCM), yaitu simulasi numerik yang menggambarkan berbagai variabel iklim global seperti suhu, tekanan udara, dan curah hujan (Legionosuko dkk., 2019). GCM merupakan alat penduga utama iklim dan cuaca yang berfungsi sebagai sumber informasi untuk

menilai pengaruh perubahan iklim (Wilby dan Wigley, 1997). Namun, data GCM memiliki resolusi spasial yang rendah ($\pm 2,5^\circ$ atau $\pm 300 \text{ km}^2$), sehingga kurang memadai untuk digunakan pada skala lokal. Oleh karena itu, dibutuhkan model peramalan curah hujan lokal dengan mempertimbangkan sirkulasi atmosfer global dari luaran GCM. Hal ini memerlukan metode yang mampu mentransformasi data GCM menjadi informasi yang relevan untuk wilayah tertentu, salah satunya melalui pendekatan *Statistical Downscaling* yang dapat mempertimbangkan wilayah bersifat lokal dengan resolusi lebih tinggi dari data luaran GCM (Sahrman dan Yulianti, 2021).

Statistical Downscaling (SD) adalah metode yang digunakan untuk menghubungkan data iklim berskala besar dari GCM dengan data iklim lokal (Sahrman & Annisa, 2023). Metode ini berfungsi untuk mentransformasi data GCM menjadi informasi yang lebih relevan pada skala lokal dengan menggunakan model regresi untuk menentukan hubungan antara variabel global dan lokal. Pendekatan yang umum digunakan dalam SD adalah regresi linear biasa, yang memodelkan hubungan antara variabel prediktor dan respon dengan asumsi bahwa hubungan tersebut berdistribusi normal (Halik dkk., 2012). Pada regresi linear biasa, pemodelan hubungan antara variabel prediktor dan respon hanya dilakukan pada rata-rata distribusi data, yang setara dengan kuantil 0.50 (Amarrullah dkk., 2023). Namun, karena curah hujan ekstrem sering kali memiliki distribusi yang asimetris dan bervariasi di berbagai tingkat intensitas, regresi linear tidak cukup untuk menangkap variabilitas pada tingkat distribusi yang lebih tinggi. Oleh karena itu, regresi kuantil dinilai cocok untuk menduga curah hujan ekstrem karena analisis ini dapat menangani sebaran data yang asimetris pada berbagai tingkat kuantil, termasuk kuantil yang lebih tinggi, seperti kuantil 0.70 hingga 0.90 yang dianggap relevan dalam menduga intensitas hujan ekstrem (Santri & Hanike, 2020).

Regresi kuantil adalah metode yang digunakan untuk menduga hubungan antara variabel prediktor (data iklim global) dan variabel respon (data curah hujan lokal) dengan fokus pada kuantil-kuantil tertentu dari distribusi data, bukan hanya pada rata-rata seperti dalam regresi linear biasa. Namun, dalam penerapannya pada data berdimensi tinggi, regresi kuantil menghadapi kendala berupa adanya multikolinearitas antara variabel prediktor, yang dapat memengaruhi akurasi estimasi model (Ciuperca, 2016). Multikolinearitas adalah kondisi ketika terdapat hubungan linear yang tinggi antara dua atau lebih variabel prediktor dalam suatu model regresi (Upa dkk., 2021). Efek utama dari multikolinearitas adalah ketidakstabilan estimasi koefisien regresi, yang berarti bahwa nilai koefisien regresi ($\hat{\beta}$) dapat berubah drastis dengan sedikit perubahan pada data. Hal ini juga dapat menyebabkan variansi dari penduga menjadi besar, sehingga model menjadi tidak efisien dalam melakukan estimasi (Amarrullah dkk., 2023). Untuk mengatasi kendala ini, diterapkan regresi kuantil *Absolute Shrinkage and Selection Operator*.

Metode *Absolute Shrinkage and Selection Operator* (LASSO) adalah teknik regularisasi yang menambahkan *penalty* L1 pada fungsi objektif, yang dapat menyusutkan koefisien variabel prediktor menjadi nol (Fauzan dkk., 2024). Dalam

konteks curah hujan ekstrem, LASSO dapat membantu memilih variabel iklim global yang paling relevan dalam memprediksi curah hujan di tingkat lokal. Berbeda dengan *Ridge* yang hanya mengecilkan koefisien dan *Elastic Net* yang cenderung mempertahankan variabel berkorelasi. Walaupun regresi kuantil LASSO efektif dalam seleksi variabel, metode ini memiliki keterbatasan karena menerapkan *penalty* yang sama untuk semua variabel tanpa mempertimbangkan tingkat kepentingannya dalam model (Santri & Hanike, 2020). Akibatnya, variabel yang memiliki pengaruh besar tetap mengalami penyusutan yang signifikan, dan berpotensi menurunkan akurasi model. Untuk mengatasi keterbatasan ini, *Adaptive LASSO* diperkenalkan sebagai pengembangan dari metode LASSO dengan menambahkan bobot *penalty* yang berbeda untuk setiap variabel. Penambahan bobot *penalty* yang berbeda dalam *Adaptive LASSO* memungkinkan model untuk menjadi lebih *adaptive* dalam memilih variabel yang signifikan, serta mengurangi pengaruh variabel yang kurang relevan. Hal tersebut meningkatkan kemampuan model untuk memprediksi dengan akurat dan efisien, terutama dalam kasus data dengan banyak variabel yang berkorelasi tinggi atau variabilitas besar (Hui Zou, 2012). Dengan adanya bobot ini, variabel yang lebih signifikan akan diberikan *penalty* yang lebih kecil, sehingga tetap dipertahankan dalam model, sedangkan variabel yang kurang signifikan akan menerima *penalty* yang lebih besar, yang cenderung membuatnya dieliminasi dari model (Rahmawati & Subanar, 2023).

Pada penelitian terdahulu yang dilakukan oleh Winkowski (2019), menunjukkan bahwa regresi kuantil mampu menangkap karakteristik data pada berbagai tingkat kuantil, terutama di bagian ekor distribusi yang merepresentasikan kejadian ekstrem. Selanjutnya, Rachmawati dkk., (2021) membahas bahwa Metode LASSO telah terbukti unggul dalam menangani multikolinearitas dan seleksi variabel secara otomatis, dengan cara memberikan *penalty* terhadap koefisien regresi, sehingga hanya variabel yang signifikan yang dipertahankan dalam model. Penelitian oleh Takada & Fujisawa (2023) juga menemukan bahwa penerapan *Penalty Adaptive* dalam analisis hidrologi dapat meningkatkan akurasi model serta kemampuannya dalam menangkap variabilitas ekstrem. Penggunaan regresi kuantil LASSO dengan *Penalty Adaptive* secara bersamaan menghasilkan kombinasi metode yang menjadi dasar yang baik untuk pendugaan curah hujan ekstrem.

Analisis data curah hujan di Kabupaten Pangkep menunjukkan adanya pola ekstrem dengan fluktuasi yang tinggi, terutama pada musim penghujan. Data historis mengindikasikan bahwa curah hujan ekstrem terjadi dengan intensitas yang tinggi dalam periode waktu yang singkat, yang sering kali mengakibatkan bencana hidrometeorologi seperti banjir dan longsor. Tantangan utama dalam pemodelan curah hujan ekstrem ini adalah keterbatasan data iklim lokal. Hal ini bisa diatasi dengan penggunaan data luaran GCM. Namun, data prediktor dari GCM sering kali memiliki hubungan linear yang kuat antar gridnya (multikolinearitas). Oleh karena itu, pendekatan regresi kuantil LASSO dengan metode *Penalty Adaptive* diajukan sebagai solusi, karena mampu mengatasi multikolinearitas dengan melakukan seleksi variabel sehingga menghasilkan estimasi yang lebih stabil dan akurat dalam memprediksi curah hujan ekstrem.

1.2 Batasan Masalah

Berdasarkan permasalahan yang telah diuraikan sebelumnya, batasan masalah pada penelitian ini adalah sebagai berikut.

1. Data curah hujan yang digunakan adalah data curah hujan di tiga stasiun pengamatan di Kabupaten Pangkep (Bungoro, Ma'rang, dan Labakkang) periode Januari 1999 – Desember 2023.
2. Nilai kuantil ekstrem yang digunakan terdiri dari 0.70, 0.80, dan 0.90.
3. Reduksi dimensi menggunakan *Partial Least Squares* (PLS).

1.3 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Memperoleh penduga model *Statistical Downscaling* dengan Regresi Kuantil *Least Absolute Shrinkage and Selection Operator* menggunakan metode *Penalty Adaptive* pada data curah hujan di Kabupaten Pangkep.
2. Memperoleh nilai dugaan curah hujan ekstrem di Kabupaten Pangkep berdasarkan model *Statistical Downscaling* dengan Regresi Kuantil *Least Absolute Shrinkage and Selection Operator* menggunakan metode *Penalty Adaptive*.

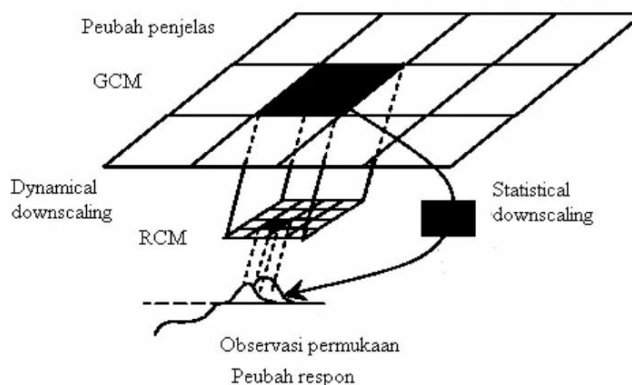
Berdasarkan tujuan penelitian di atas, maka setelah melakukan penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Sebagai sumber pengetahuan mengenai penerapan Regresi Kuantil *Least Absolute Shrinkage and Selection Operator* dengan metode *Penalty Adaptive* dalam pemodelan pendugaan curah hujan ekstrem.
2. Sebagai sumber pengetahuan dan informasi yang berguna bagi pengambil keputusan di Kabupaten Pangkep, dalam perencanaan mitigasi risiko cuaca ekstrem serta pengelolaan sumber daya air yang lebih efektif.

1.4 Teori

1.4.1 *Global Circulation Model*

Global Circulation Model (GCM) adalah alat utama dalam memahami sistem iklim karena mampu menyediakan informasi mengenai perubahan iklim dari masa lalu, masa kini, hingga proyeksi di masa depan. GCM dirancang untuk mensimulasikan pola sirkulasi udara di atmosfer serta perubahan iklim dalam jangka panjang. Salah satu kelemahan GCM dalam melakukan prediksi adalah tingkat akurasi yang relatif rendah (Santri & Hanike, 2020). Hal ini disebabkan oleh kompleksitas proses fisika yang menentukan curah hujan, yang sangat rumit, sensitif, dan bersifat nonlinear, sehingga tidak sepenuhnya dapat dimodelkan dengan baik oleh GCM (Fadhli dkk., 2019). Secara umum, ilustrasi GCM dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi Data GCM (Berliana & Sutikno, 2007)

Setiap model GCM memiliki variasi dalam hal resolusi spasial, jumlah lapisan, serta persamaan yang digunakan untuk menghasilkan parameter atmosfer. GCM dapat mempresentasikan keadaan iklim berskala global (resolusi rendah) dengan baik, namun tidak dalam berskala lokal/regional (resolusi tinggi). Oleh karena itu, perlu dilakukan teknik *Statistical Downscaling* untuk menduga variabel iklim skala lokal dengan tingkat akurasi tinggi untuk mengaplikasikannya pada skala lokal.

1.4.2 Statistical Downscaling

Statistical downscaling (SD) adalah proses *downscaling* yang didasarkan pada hubungan grid berskala besar, seperti GCM, sebagai variabel prediktor, dan grid berskala lokal sebagai variabel respon (Tweedie dkk., 2020). Teknik ini memanfaatkan hubungan statistik antara variabel prediktor skala besar, seperti output dari GCM, dan variabel respon skala lokal, misalnya data curah hujan pengamatan di stasiun meteorologi lokal.

Teknik SD memungkinkan pengembangan model yang dapat menghasilkan prediksi dengan resolusi yang lebih tinggi dari data asli GCM, sehingga lebih relevan dan sesuai dengan kondisi lokal. Pendekatan utama dalam SD melibatkan model regresi, dengan variabel respon (y) seperti curah hujan lokal, dihubungkan dengan variabel prediktor (x) yang diambil dari GCM. Model ini sering kali dinyatakan dalam Persamaan (1):

$$y = f(x) \quad (1)$$

dengan y peubah skala lokal respon (curah hujan) berukuran $n \times 1$, x adalah peubah prediktor (presipitasi GCM) berukuran $n \times k$, n merupakan banyaknya pengamatan/waktu (bulan), dan k menunjukkan banyaknya *grid* pada skala global GCM.

Pendekatan SD memanfaatkan data dari GCM untuk memprediksi iklim pada skala lokal. Dua faktor utama yang sangat penting dalam pendekatan ini adalah pemilihan variabel prediktor yang tepat dan penentuan jumlah grid yang digunakan. Model SD dapat menghasilkan prediksi yang akurat jika beberapa syarat terpenuhi, yaitu (1) terdapat korelasi yang kuat antara variabel respon dan prediktor sehingga dapat menjelaskan keragaman iklim dengan baik, (2) variabel prediktor dapat

disimulasikan dengan baik oleh GCM, dan (3) hubungan antara variabel respon dan prediktor tidak berubah seiring waktu atau dipengaruhi oleh perubahan iklim (Berliana & Sutikno, 2007).

1.4.3 Multikolinearitas

Multikolinearitas adalah kondisi ketika terdapat korelasi yang tinggi antara variabel bebas dalam sebuah model regresi. Kondisi ini dapat menyebabkan ketidakakuratan dalam estimasi koefisien regresi dan menurunkan kualitas prediksi (Azizah dkk., 2021). Dalam konteks prediksi curah hujan dengan data GCM, multikolinearitas sering terjadi karena data GCM memiliki dimensi tinggi dengan korelasi yang kuat antara variabelnya. Berbagai metode dapat digunakan untuk mengatasi masalah ini, termasuk teknik seleksi variabel, LASSO, dan analisis komponen utama.

Salah satu ukuran yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah *Variance Inflation Factor* (VIF) (Indahwati dkk., 2014). Ukuran ini berfungsi sebagai indikator dalam mengidentifikasi multikolinearitas pada regresi linear yang melibatkan lebih dari dua variabel prediktor. Nilai VIF untuk koefisien ke- j dihitung dengan Persamaan (2):

$$VIF_j = \frac{1}{1-R_j^2} \quad (2)$$

dengan $j = 1, 2, \dots, k$ sedangkan R_j^2 adalah koefisien determinasi dari variabel prediktor ke- j yang diregresikan terhadap variabel prediktor lainnya. Menurut (Sriningsih dkk., 2018), nilai $VIF > 10$ mengindikasikan adanya multikolinearitas sehingga dilakukan reduksi dimensi, dan jika terjadi masalah multikolinearitas pada model regresi maka dibutuhkan sebuah metode untuk menangani multikolinearitas tersebut.

1.4.4 Regresi Kuantil

Regresi kuantil adalah metode statistik yang dikembangkan oleh (Koenker & Bassett, 1978) untuk mengestimasi hubungan antara variabel prediktor dan variabel respon pada berbagai titik distribusi kondisional dari variabel respon. Berbeda dengan regresi linear yang berfokus pada estimasi nilai rata-rata, regresi kuantil memungkinkan estimasi pada kuantil tertentu (misalnya, median atau kuantil lainnya) dari distribusi variabel respon, sehingga lebih fleksibel dalam menangani data distribusi yang tidak normal. Regresi kuantil merupakan pengembangan dari regresi linear yang memungkinkan estimasi kuantil kondisional dari variabel respon y terhadap variabel prediktor x (Idris dkk., 2018).

Misalkan diberikan data $\{y_1, y_2, y_3, \dots, y_n\}$ dan τ adalah fungsi kumulatif atau kuantil ke- τ dari y maka:

$$F_Y(y) = F(y) = P(Y \leq y) = \tau \quad (3)$$

dengan $\tau \in (0,1)$ sehingga kuantil ke- τ dari y atau fungsi kuantil ($Q_Y(\tau)$) adalah *invers* dari Persamaan (3) yaitu $Q_Y(\tau) = F_Y^{-1}(\tau) = \inf\{y: F_Y(y) > \tau\}$. Pada regresi linear, $E(y|x_i) = X_i'\beta$ sementara dalam regresi kuantil $Q_Y(\tau|x_i) = X_i'\beta_\tau$ yang dapat diuraikan menjadi Persamaan (4):

$$y_i(\tau) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_j(\tau)x_{ij} + \dots + \beta_k(\tau)x_{ik} + \varepsilon_i(\tau) \quad (4)$$

disederhanakan seperti pada Persamaan (5):

$$y_i = \beta_{\tau 0} + \sum_{j=1}^k \beta_{\tau j} x_{ij} + \varepsilon_{\tau i}, i = 1, \dots, n \quad (5)$$

dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$

$y_i(\tau)$: variabel respon ke- i pada kuantil ke- τ

β_0 : koefisien konstanta

τ : nilai kuantil

$\beta(\tau)$: penduga parameter pada kuantil ke- τ

x_{ij} : pengamatan ke- i pada variabel prediktor ke- j

$\varepsilon_i(\tau)$: eror ke- i pada kuantil ke- τ

Dalam bentuk matriks regresi kuantil dapat dituliskan pada Persamaan (6):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_{\tau 0} \\ \beta_{\tau 1} \\ \vdots \\ \beta_{\tau k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{\tau 1} \\ \varepsilon_{\tau 2} \\ \vdots \\ \varepsilon_{\tau n} \end{bmatrix} \quad (6)$$

Bentuk umum dari matriks pada Persamaan (6) dapat dituliskan sebagai berikut:

$$y(\tau) = X\beta(\tau) + \varepsilon(\tau) \quad (7)$$

Dengan

$y(\tau)$: vektor kolom berukuran $n \times 1$ dari variabel respon y

X : matriks berukuran $n \times (k + 1)$ dengan baris n merupakan pengamatan pada

kolom k variabel prediktor ke- j dengan $j = 1, 2, \dots, k$

β : vektor kolom berukuran $(k + 1) \times 1$ dari parameter β_j dengan $j = 1, 2, \dots, k$

ε : vektor kolom berukuran $n \times 1$ dari eror ε_i

1.4.5 Regularisasi LASSO L_1

Least Absolute Shrinkage and Selection Operator (LASSO) adalah metode regularisasi yang dikenalkan oleh (Tibshirani, 1996). Teknik ini digunakan untuk mengatasi masalah *overfitting* dalam model regresi, terutama saat jumlah variabel prediktor sangat banyak. Prinsip LASSO adalah meminimalkan jumlah kesalahan prediksi dengan menambahkan *penalty* terhadap besarnya koefisien regresi. Koefisien yang bernilai nol menunjukkan bahwa metode LASSO melakukan seleksi variabel, sehingga hanya variabel prediktor yang paling kuat pengaruhnya yang dipertahankan. Hal ini dapat membantu mengurangi variasi dalam hasil prediksi, dengan mengurangi variabel-variabel yang kurang relevan (Tibshirani, 1996).

Misal $(x_i, y_1), \dots, (x_n, y_n)$ merupakan pasangan data sebanyak n untuk pemodelan $x_{i1}, x_{i2}, \dots, x_{ip}$ sebagai variabel prediktor, $y_i \in \mathbb{R}$ sebagai variabel respons, dan $i = 1, \dots, n$. Dugaan koefisien regresi kuantil regularisasi LASSO ($\hat{\beta}_{\tau}^{LASSO}$) diperoleh dari solusi Persamaan (8):

$$\min \sum_{i=1}^n \rho_{\tau}(y_i - x_i^T \beta) \quad (8)$$

Dengan batasan $\sum_j^p |\beta_{\tau,j}| \leq \lambda_1, \lambda_1 \geq 0$ adalah parameter LASSO. Solusi $\hat{\beta}_{\tau}^{LASSO}$ dapat juga dituliskan dalam persamaan *Lagrange* pada Persamaan (9):

$$\min \sum_{i=1}^n \rho_{\tau}(\mathbf{y}_i - \mathbf{x}_i^T \boldsymbol{\beta}_{\tau}) + \lambda_1 \sum_{j=1}^p |\beta_{\tau,j}| \quad (9)$$

Koefisien LASSO λ_1 adalah parameter *penalty* yang mengontrol seberapa besar regularisasi diterapkan. Jika $\lambda_i = 0$, maka $\hat{\beta}_{\tau}^{LASSO}$ sama dengan dugaan parameter regresi kuantil tanpa regularisasi. Jika $\lambda_i = \infty$, maka $\hat{\beta}_{\tau}^{LASSO}$ sama dengan nol. Regresi kuantil LASSO merupakan gabungan antara konsep regresi kuantil dan *penalty* LASSO (Ciuperca, 2016). Tujuan utama dari regresi kuantil LASSO adalah untuk melakukan estimasi kuantil dengan menambahkan *penalty* LASSO agar model tetap sederhana dan mampu menangani jumlah variabel prediktor lebih banyak dibandingkan pengamatan. Persamaan regresi kuantil LASSO pada Persamaan (10):

$$\min \left(\sum_{\{i: y_i \geq x_i^T \beta\}} \tau |y_i - x_i^T \beta| + \sum_{\{i: y_i < x_i^T \beta\}} (1 - \tau) |y_i - x_i^T \beta| + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (10)$$

dengan

- τ : kuantil yang dipilih dalam regresi kuantil LASSO.
- λ : parameter *penalty* yang mengatur kekuatan regularisasi.
- β_j : koefisien regresi yang perlu diestimasi.
- $x_i^T \beta$: nilai prediksi untuk pengamatan ke- i .
- y_i : nilai pengamatan (variabel respon) untuk pengamatan ke- i .
- p : jumlah variabel prediktor dalam model.

1.4.6 Metode *Penalty Adaptive*

Salah satu pengembangan dari metode LASSO adalah *penalty adaptive* yang diperkenalkan oleh Hui Zou pada tahun 2006 (Rahmawati & Subanar, 2023). *Adaptive* merupakan metode regresi yang menyempurnakan LASSO biasa dengan menambahkan bobot *adaptive* (*adaptive weight*) pada setiap koefisien regresi dalam tahap *penalty*. Dengan penggunaan bobot ini, *penalty* L1 menjadi lebih fleksibel sehingga memungkinkan metode ini untuk mempertahankan variabel penting dan mengeliminasi variabel yang tidak relevan secara lebih akurat dibandingkan LASSO standar (Zou, 2006).

Hal ini memungkinkan metode *penalty adaptive* untuk (1) mengidentifikasi variabel yang penting (signifikan) dengan lebih baik, dan (2) mengurangi bias yang terjadi pada LASSO biasa, terutama saat terjadi korelasi tinggi antar variabel prediktor. Bobot *adaptive* dihitung berdasarkan *invers* dari nilai absolut dugaan awal koefisien regresi. Variabel yang koefisiennya besar akan diberi *penalty* lebih kecil, dan sebaliknya, variabel yang koefisiennya kecil akan diberi *penalty* lebih besar (Takada & Fujisawa, 2023). Perhitungan bobot *adaptive* ditentukan melalui Persamaan (11):

$$w_j = \frac{1}{|\hat{\beta}_{LASSO,j}|^{\gamma}} \quad (11)$$

dengan:

w_j : bobot untuk koefisien β_j ,

$\hat{\beta}_{lasso,j}$: dugaan awal koefisien regresi dari model LASSO,

γ : parameter yang mengontrol tingkat *penalty adaptive* (umumnya $\gamma = 1$)

Bobot ini kemudian dimasukkan ke dalam fungsi *penalty*, menghasilkan bentuk fungsi objektif *adaptive* pada LASSO pada Persamaan (12):

$$\min \left(\sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p w_j |\beta_j| \right) \quad (12)$$

dengan penggunaan *penalty* berbobot ini, penambahan *penalty adaptive* pada LASSO mampu mengurangi bias dan memberikan hasil yang lebih baik dibandingkan LASSO standar.

1.4.7 Reduksi Dimensi menggunakan *Partial Least Squares*

Partial Least Square (PLS) merupakan *soft* model yang dapat menjelaskan struktur keragaman data. Metode ini berfungsi untuk mereduksi dimensi data dengan menemukan komponen-komponen utama yang paling relevan, yang dapat menghubungkan variabel prediktor dengan respon secara optimal, sehingga mengatasi masalah multikolinearitas.

Pada dasarnya, PLS memodelkan hubungan variabel Y dengan variabel X berdasarkan variabel internal. Variabel X dibagi kedalam t_h dan loading p_h , yang dinyatakan pada Persamaan (13):

$$X = t_1 p_1 + t_2 p_2 + t_3 p_3 + \dots + t_h p_h + E_h \quad (13)$$

dengan

X : variabel bebas

t_h : vektor skor variabel X

p_h : vektor muatan variabel X

E_h : matriks residual variabel X

Variabel Y juga dibagi dalam skor u_h dan loading q_h yang dinyatakan pada Persamaan (14):

$$Y = u_1 q_1 + u_2 q_2 + u_3 q_3 + \dots + u_h q_h + E_h \quad (14)$$

dengan

X : variabel tak bebas

u_h : vektor skor variabel Y

q_h : vektor muatan variabel Y

E_h : matriks residual variabel Y

Pemilihan jumlah komponen yang optimal dalam PLS dilakukan dengan memperhatikan nilai proporsi variansi yang dijelaskan oleh model. Proporsi variansi yang dijelaskan oleh setiap komponen h dapat dihitung menggunakan Persamaan (15):

$$PV_h = \frac{\lambda_h}{\sum_{i=1}^H \lambda_i} \times 100\% \quad (15)$$

dengan PV_h : proporsi variansi yang dijelaskan oleh komponen ke- h

λ_h : nilai eigen dari komponen ke- h

H : jumlah total komponen yang diekstrak

Jika nilai proporsi variansi yang dihasilkan pada model PLS dengan satu komponen mencapai lebih besar dari atau sama dengan 80%, maka komponen pertama sudah cukup untuk menjelaskan variasi data. Sebaliknya, jika proporsi variansi kurang dari 80%, maka dibutuhkan model regresi PLS dengan dua komponen untuk menangkap variasi yang lebih baik (Indahwati dkk., 2014). Skor komponen laten untuk setiap sampel dalam data *training* dan *testing* dihitung menggunakan Persamaan (16) dan Persamaan (17):

$$t_h = XW_h \quad (16)$$

$$u_h = YQ_h \quad (17)$$

dengan

t_h : skor komponen laten dari variabel X pada komponen ke- h

u_h : skor komponen laten dari variabel Y pada komponen ke- h

W_h : matriks bobot dari variabel X

Q_h : matriks bobot dari variabel Y

Setelah skor komponen laten diperoleh, regresi kuantil LASSO dengan *penalty adaptive* dapat diterapkan pada skor komponen laten tersebut untuk mendapatkan model yang optimal.

1.4.8 Validasi Model

Validasi model adalah langkah krusial dalam proses pemodelan untuk memastikan bahwa model yang kita bangun tidak hanya bekerja baik pada data pelatihan, tetapi juga generalisasi dengan baik pada data baru (Tweedie dkk., 2020). Dalam konteks *adaptive* LASSO, *cross-validation* digunakan untuk menentukan nilai optimal dari parameter *penalty* λ , yang mengontrol seberapa kuat *penalty* LASSO diterapkan pada koefisien regresi.

Kriteria kebaikan model dalam validasi dapat dinilai menggunakan *Root Mean Square Error Prediction* (RMSEP). Ukuran ini digunakan untuk mengevaluasi kinerja suatu model, terutama dalam konteks validasi prediksi. Nilai RMSEP yang lebih rendah menunjukkan bahwa model memiliki kemampuan prediksi yang akurat, dengan kesalahan prediksi (residual) yang relatif kecil. Rumus untuk menghitung RMSEP diberikan pada Persamaan (18):

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

keterangan:

y_i : nilai aktual

$\hat{y}_i$: nilai prediksi dari model

n : jumlah pengamatan dalam data uji

Nilai RMSEP memberikan gambaran yang jelas tentang seberapa baik model dapat memprediksi data *testing*. Ukuran ini menjadi indikator utama dalam mengevaluasi apakah model mengalami *overfitting* pada data *training* atau cukup fleksibel untuk menggeneralisasi pola pada data *testing*.

1.4.9 Curah Hujan

Curah hujan adalah ketinggian air hujan yang terkumpul pada permukaan datar tanpa adanya penguapan, peresapan, atau aliran. Satuan yang digunakan untuk mengukur curah hujan adalah milimeter (mm), di mana 1 mm menunjukkan bahwa 1 liter air hujan telah jatuh pada area seluas 1 meter persegi (Nurlatifah dkk., 2023). Menurut Badan Meteorologi, Klimatologi, dan Geofisika (BKMKG), curah hujan dibagi menjadi empat kelompok, yaitu:

1. Curah hujan rendah: 0-20 mm, 21-50 mm, dan 51-100 mm.
2. Curah hujan menengah: 101-150 mm, 151-200 mm, dan 201-300 mm.
3. Curah hujan tinggi: 301-500 mm.
4. Curah hujan sangat tinggi: > 500 mm.

Sulawesi Selatan memiliki tiga pola curah hujan utama, yaitu pola monsun, ekuatorial, dan lokal. Wilayah bagian barat dan selatan, termasuk Kabupaten Pangkep, umumnya mengikuti pola monsun, dengan curah hujan tinggi (> 300 mm/bulan) terjadi pada awal dan akhir tahun, serta curah hujan rendah pada pertengahan tahun (< 50 mm/bulan) (Hartono & Sutikno, 2020). Produksi garam di Kabupaten Pangkep sangat dipengaruhi oleh pola curah hujan. Metode produksi garam yang umum digunakan adalah penguapan air laut secara alami dengan bantuan sinar matahari. Curah hujan yang tinggi dapat menghambat proses penguapan, sehingga menurunkan produktivitas garam.

BAB II METODE PENELITIAN

2.1 Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari data presipitasi yang dihasilkan oleh GCM melalui proyek *Climate Model Intercomparison Project* (CMIP5), dengan satuan mm/bulan. Data tersebut tersedia secara online melalui situs web yang disediakan oleh KNMI Belanda, yaitu <http://www.climatexp.knmi.nl/>. Selain itu, data curah hujan rata-rata di Kabupaten Pangkep, Sulawesi Selatan, untuk periode Januari 1999 – Desember 2023 diperoleh dari stasiun Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) wilayah IV Makassar.

2.2 Deskripsi Variabel

Variabel yang digunakan dalam penelitian ini:

1. Curah hujan lokal (Y), merujuk pada rata-rata curah hujan (mm/bulan) di tiga stasiun pengamatan di Kabupaten Pangkep (Bungoro, Ma'rang, dan Labakkang) periode Januari 1999 – Desember 2023.
2. Prediktor GCM (X), mencakup 64 prediktor yang berasal dari data grid domain GCM, dan variabel *dummy* sebanyak 3 variabel. Struktur data dapat dilihat pada Tabel 1 dan tabulasi data pada Lampiran 1. Adapun nilai variabel *dummy* yang dapat dilihat pada Lampiran 2.

Tabel 1. Struktur data

Pengamatan	y	X_1	X_2	...	X_{64}	D_1	D_2	D_3
1	y_1	$x_{1(1)}$	$x_{1(2)}$	...	$x_{1(64)}$	$D_{1(1)}$	$D_{1(2)}$	$D_{1(3)}$
2	y_2	$x_{2(1)}$	$x_{2(2)}$	...	$x_{2(64)}$	$D_{2(1)}$	$D_{2(2)}$	$D_{2(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
298	y_{298}	$x_{298(1)}$	$x_{298(2)}$	...	$x_{298(64)}$	$D_{298(1)}$	$D_{298(2)}$	$D_{298(3)}$
299	y_{299}	$x_{299(1)}$	$x_{299(2)}$	...	$x_{299(64)}$	$D_{299(1)}$	$D_{299(2)}$	$D_{299(3)}$
300	y_{300}	$x_{300(1)}$	$x_{300(2)}$	...	$x_{300(64)}$	$D_{300(1)}$	$D_{300(2)}$	$D_{300(3)}$

2.3 Metode Analisis

Berikut ini adalah langkah-langkah analisis data yang dilakukan dalam penelitian ini.

1. Deskripsi data curah hujan bulanan sebagai informasi awal keragaman data amatan.
2. Membagi data menjadi data *training* (80%) mencakup periode Januari 1999 – Desember 2012, dan data *testing* (20%) mencakup periode Januari 2013 – Desember 2023 untuk pengembangan dan validasi model.
3. Identifikasi multikolinearitas antar variabel prediktor GCM menggunakan VIF dengan Persamaan (2).
4. Reduksi dimensi peubah prediktor menggunakan *Partial Least Squares* (PLS).
 - Menentukan jumlah komponen laten (KL) berdasarkan proporsi variansi yang dijelaskan.

- Memilih jumlah komponen laten yang dianggap mampu menjelaskan sebagian besar keragaman data.
 - Menghitung skor komponen laten untuk data *training* dan data *testing*.
5. Memodelkan *statistical downscaling* menggunakan regresi kuantil LASSO dengan metode *penalty adaptive*
- Menentukan kuantil yang akan diprediksi (τ), yaitu $\tau = 0.70, 0.80$, dan 0.90 .
 - Menyusun data model dengan peubah respon (Y) dan peubah prediktor (X) yang telah direduksi.
 - Membangun model regresi kuantil LASSO dengan metode *penalty adaptive*.
 - Melakukan estimasi parameter model untuk masing-masing kuantil.
 - Melakukan prediksi terhadap data *training* dan data *testing*.
6. Validasi model dengan menggunakan rumus RMSEP seperti pada Persamaan (18) pada data *testing* periode Januari 2013 – Desember 2023.