**PENDEKATAN REGRESI KUANTIL *SPLINE* DALAM**
***STATISTICAL DOWNSCALING* DENGAN**
**ANALISIS KOMPONEN UTAMA**
**(Studi Kasus : Curah Hujan Bulanan Kabupaten Pangkep)**

*QUANTILE SPLINE REGRESSION APPROACH IN*
*STATISTICAL DOWNSCALING WITH*
*PRINCIPAL COMPONENT ANALYSIS*
*(Case Study: Monthly Rainfall in Pangkep Regency)*

**ANDI SRI YULIANTI**
**H062222008**

**PROGRAM STUDI MAGISTER STATISTIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS HASANUDDIN**
**MAKASSAR**
**2024**

# PENDEKATAN REGRESI KUANTIL *SPLINE* DALAM *STATISTICAL DOWNSCALING* DENGAN ANALISIS KOMPONEN UTAMA
## (Studi Kasus : Curah Hujan Bulanan Kabupaten Pangkep)

Tesis
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

ANDI SRI YULIANTI
H062222008

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA**
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**
**UNIVERSITAS HASANUDDIN**
**MAKASSAR**
**2024**

**TESIS**

PENDEKATAN REGRESI KUANTIL *SPLINE* DALAM *STATISTICAL DOWNSCALING* DENGAN ANALISIS KOMPONEN UTAMA
(Studi Kasus : Curah Hujan Bulanan Kabupaten Pangkep)

ANDI SRI YULIANTI
H062222008

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin
pada tanggal 13 Maret 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

| Pembimbing Utama | Pembimbing Pendamping |
|---|---|
| **Dr. Anna Islamiyati, S.Si., M.Si.** | **Dr. Erna Tri Herdiani, S.Si., M.Si.** |
| NIP. 19770808 200501 2 002 | NIP. 19750429 200003 2 001 |
| Ketua Program Studi Magister Statistika | Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin |
| **Dr. Erna Tri Herdiani, S.Si., M.Si.** | **Dr. Eng. Amiruddin, M.Si.** |
| NIP. 19750429 200003 2 001 | NIP. 19720515 199702 1 002 |

## PERNYATAAN KEASLIAN TESIS
## DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "Pendekatan Regresi Kuantil *Spline* dalam *Statistical Downscaling* dengan Analisis Komponen Utama (Studi Kasus: Curah Hujan Bulanan Kabupaten Pangkep)" adalah benar karya saya dengan arahan dari komisi Pembimbing (Dr. Anna Islamiyati, S.Si., M.Si. dan Dr. Erna Tri Herdiani, S.Si., M.Si.). Karya Ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari tesis ini akan dipublikasikan di *Journal of Sciences, Islamic Republic of Iran* sebagai artikel dengan judul "*The Principal Component Linear Spline Quantile Regression Model In Statistical Downscaling For Rainfall Data*".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 13 Maret 2024

METERAI TEMPEL
42C61ALX078994189

Andi Sri Yulianti
NIM. H062222008

# UCAPAN TERIMA KASIH

*Assalamu'alaikum Warahmatullahi Wabarakatuh*

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas segala limpahan rahmat dan hidayah-Nya yang telah diberikan kepada penulis sampai saat ini. Shalawat dan salam senantiasa tercurahkan kepada baginda Rasulullah *Shallallahu 'Alaihi Wa sallam*, kepada para keluarga, tabi'in, tabi'ut tabi'in, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri-diri mereka. *Alhamdulillaahirobbil'aalamiin*, berkat rahmat dan kemudahan yang diberikan oleh Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tesis yang berjudul "Pendekatan Regresi Kuantil *Spline* dalam *Statistical Downscaling* dengan Analisis Komponen Utama (Studi Kasus: Curah Hujan Bulanan Kabupaten Pangkep)" sebagai salah satu syarat memperoleh gelar magister pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Penulis tidak akan sampai pada titik ini, jikalau tanpa dukungan dan bantuan dari pihak yang selalu ada, peduli dan menyayangi penulis. Oleh karena itu, penulis haturkan rasa terima kasih yang setulus-tulusnya serta penghargaan yang setinggi-tingginya untuk orang tua penulis, Ayahanda **Andi Tajuddin** dan Ibunda **Sitti Nurhayati P** yang telah memberikan dukungan penuh, pengorbanan, kesabaran hati, cinta dan kasih sayang, serta dengan ikhlas telah mengiri setiap langkah penulis dengan doa dan restunya. Kakakku **Andi Rahmadani** dan **Andi Desisulfiana**, terima kasih telah memberikan dorongan dan dukungan baik batin dan juga raga untuk penulis serta untuk keluarga besarku, terima kasih atas doa dan dukungannya selama ini.

Penghargaan yang tulus dan ucapan terima kasih dengan penuh keikhlasan juga penulis ucapkan kepada:

1. **Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc.**, selaku Rektor Universitas Hasanuddin berserta seluruh jajarannya.
2. **Bapak Dr. Eng. Amiruddin**, selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. **Ibu Dr. Anna Islamiyati, S.Si., M.Si.**, selaku Ketua Departemen Statistika ltas MIPA Universitas Hasanuddin dan juga menjadi Pembimbing a yang dengan penuh kesabaran telah meluangkan waktu dan kirannya di tengah berbagai kesibukan dan prioritasnya untuk

senantiasa memberikan arahan, dorongan, motivasi dan kemudahan kepada Penulis dalam menyelesaikan tesis ini.

4. **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.**, selaku Ketua Program Studi Magister Statistika Departemen Statistika Fakultas MIPA Universitas Hasanuddin dan juga menjadi Pembimbing Pertama yang dengan penuh kesabaran telah meluangkan waktu dan pemikirannya di tengah berbagai kesibukan dan prioritasnya untuk senantiasa memberikan arahan, dorongan, motivasi dan kemudahan kepada Penulis dalam menyelesaikan tesis ini.
5. **Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**, selaku Tim Penguji yang telah memberikan kritikan yang membangun dan arahan dalam penyempurnaan penyusunan tesis ini serta waktu yang telah diberikan kepada penulis.
6. **Ibu Prof. Dr. Nurtiti Sunusi, S.Si., M.Si.**, selaku Tim Penguji yang telah memberikan kritikan yang membangun dan arahan dalam penyempurnaan penyusunan tesis ini serta waktu yang telah diberikan kepada penulis.
7. **Ibu Prof. Dr. Dr. Georgina Maria Tinungki, M.Si.**, selaku Tim Penguji yang telah memberikan kritikan yang membangun dan arahan dalam penyempurnaan penyusunan tesis ini serta waktu yang telah diberikan kepada penulis.
8. **Bapak dan Ibu Dosen** Departemen Statistika Fakultas MIPA Universitas Hasanuddin yang dengan tulus dan ikhlas memberikan ilmu pengetahuan dan pengalaman yang dimiliki selama perkuliahan berlangsung.
9. Spesial untuk teman Penulis yaitu **Nurhidaya L.**, terima kasih atas waktunya selama $\pm$5.7. tahun telah menemani dan berjuang bersama dalam menempuh pendidikan sarjana dan magister. Terima kasih atas cerita dan kenangannya selama ini, selalu memberi dukungan, berbagi keceriaan, berbagi kesedihan, dan berbagi masalah, serta membantu Penulis selama peroses perkuliahan berlangsung dan dalam proses penyusunan tesis ini.
10. Teman seperjuangan di Program Studi Magister Statistika Angkatan 7, **Nurul, Nuge, Rifka**, **Kak Ani**, **Kak Haura**, **Kak Miftah**, **Kak Ikka**, **Hikmah**, **Lili**, **Mega**, **Isra**, **Kak Ika**, **Kak Cici**, **Fira**, dan lainnya terima kasih atas kebersamaannya selama ini.
11. Kepada diri sendiri **"Andi Sri Yulianti"**, terimakasih telah berusaha dan ...rja keras untuk menyelesaikan tahap ini untuk kedua kalinya.

12. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah *Subhanallahu Wa Ta'ala.*

Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini, untuk itu dengan segala kerendahan hati Penulis memohon maaf. Semoga tesis ini bermanfaat bagi pembaca dan perkembangan ilmu pengetahuan khususnya dalam dunia statistika.

*Wassalamu'alaikum Warahmatullahi Wabarakatuh*

Makassar, 13 Maret 2024

Andi Sri Yulianti

# ABSTRAK

ANDI SRI YULIANTI. **Pendekatan Regresi Kuantil *Spline* dalam *Statistical Downscaling* dengan Analisis Komponen Utama (Studi Kasus : Curah Hujan Bulanan Kabupaten Pangkep)** (dibimbing oleh Dr. Anna Islamiyati, S.Si., M.Si. dan Dr. Erna Tri Herdiani, S.Si., M.Si.)

Perubahan ekstrem dalam intensitas curah hujan dapat membawa konsekuensi negatif dalam berbagai aspek kehidupan. Salah satu sumber informasi untuk menganalisis curah hujan adalah dengan menggunakan *Global Circulation Model* (GCM) yang memiliki cakupan global. Metode *Statistical Downscaling* (SD) digunakan untuk mengaitkan informasi dari GCM dengan data curah hujan yang berskala lokal. Masalah multikolinearitas dan outlier pada data diatasi dengan menggunakan regresi kuantil dan analisis komponen utama dengan estimator nonparametrik yaitu *spline truncated*. Model tersebut diterapkan pada data curah hujan di Kabupaten Pangkep periode Januari 2008 sampai Desember 2022. Tujuan penelitian ini adalah untuk mendapatkan model regresi terbaik yang digunakan untuk meramalkan data curah hujan. Model terbaik adalah model dengan kriteria nilai *generalized cross validation* (GCV) minimum. Hasil analisis menunjukkan bahwa dua komponen utama telah menjelaskan 88.39% dari total variansi sampel dengan model optimal adalah model regresi pada kuantil 0.50 dengan 1 titik knot. Model tersebut dapat meramalkan curah hujan aktual dengan baik dengan tingkat korelasi sebesar 0.89.

Kata kunci : curah hujan, komponen utama, regresi kuantil, *spline*, *statistical downscaling*

# ABSTRACT

ANDI SRI YULIANTI. **Quantile Spline Regression Approach in Statistical Downscaling with Principal Component Analysis (Case Study: Monthly Rainfall in Pangkep Regency)** (supervised by Dr. Anna Islamiyati, S.Si., M.Si. and Dr. Erna Tri Herdiani, S.Si., M.Si.)

Extreme changes in rainfall intensity can bring negative consequences in various aspects of life. One source of information to analyze rainfall is by using the Global Circulation Model (GCM), which has global coverage. The Statistical Downscaling (SD) method is used to link information from the GCM with local-scale rainfall data. The problems of multicollinearity and outliers in the data are addressed using quantile regression and principal component analysis with a nonparametric estimator, namely truncated spline. This model was applied to rainfall data in Pangkep Regency for the period January 2008 to December 2022. The aim of this study is to obtain the best regression model used for forecasting rainfall data. The best model is the one with the minimum generalized cross-validation (GCV) criteria value. The analysis results show that two main components have explained 88.39% of the total sample variance, with the optimal model being the regression model at the 0.50 quantile with 1 knot point. This model can predict actual rainfall well with a correlation level of 0.89.

Keywords: rainfall, principal components, quantile regression, spline, statistical downscaling

# DAFTAR ISI

# DAFTAR TABEL

## DAFTAR GAMBAR

## DAFTAR LAMPIRAN

# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

Analisis regresi merupakan sebuah analisis statistika yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dengan variabel respon. Terdapat dua pendekatan estimasi model dalam analisis regresi, yaitu regresi parametrik dan regresi nonparametrik (Famalika dan Sihombing, 2022). Pada pendekatan regresi parametrik terdapat asumsi bahwa bentuk regresi mengikuti bentuk parametrik misalnya linier, kuadratik, kubik, polinomial, eksponensial dan lain-lain. Adapun model regresi nonparametrik memiliki bentuk kurva regresi yang diasumsikan tidak diketahui (Islamiyati, 2014). Beberapa estimator dalam pendekatan regresi nonparametrik telah banyak dikembangkan antara lain menggunakan *spline*, kernel, *wavelet*, *fourier* dan lainnya (Sanusi, Syam dan Adawiyah, 2019).

*Spline* merupakan salah satu estimator yang sering digunakan dengan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinu. Kelebihan dari *spline* yaitu adanya titik knot yang dapat memberikan fleksibilitas dalam perubah-perubahan yang terjadi pada kurva (Famalika dan Sihombing, 2022). Titik knot adalah titik perpaduan bersama yang menunjukkan perubahan pola perilaku data pada interval yang berbeda (Nisa dan Budiantara, 2016). Estimator *spline* terdiri dari beberapa, seperti *spline truncated*, *spline smoothing*, dan *penalized spline* (Aprilia dkk., 2020).

Estimator *spline* telah dikembangkan penggunaannya dalam regresi kuantil. Regresi kuantil merupakan salah satu metode yang digunakan untuk mengatasi data yang mengalami pelanggaran asumsi yaitu asumsi normalitas akibat adanya pencilan (Wahyudi dan Zain, 2014). Regresi kuantil dapat mengukur efek variabel prediktor tidak hanya di pusat sebaran data, tetapi juga pada bagian atas dan bawah ekor sebaran. Analisis ini sangat berguna dalam penerapannya jika nilai ekstrim merupakan permasalahan penting yang akan diatasi (Djuraidah dan Wigena, 2011). Penelitian tentang regresi kuantil *spline* telah banyak dilakukan diantaranya adalah Goldameir, Djuraidah dan Wigena [illegible] dengan menggunakan regresi kuantil pada *spline* terpenalti. Balami dan [illegible] (2019) dengan menggunakan regresi kuantil pada fungsi *spline* [illegible]. Aprilia dkk. (2020) menggunakan regresi kuantil *spline* kuadratik.

Pelanggaran asumsi lain yang sering terjadi pada data selain dari adanya pencilan, salah satunya adalah terjadinya korelasi antar variabel prediktor yang disebut dengan multikolinearitas. Pada penelitian Goldameir dkk. (2015), *spline* terpenalti digunakan untuk mengatasi masalah multikolinearitas yang didalamnya terdapat fungsi penalti. *Spline truncated* merupakan salah satu estimator lainnya yang memiliki kemampuan dalam menunjukkan perubahan pola perilaku data pada interval yang berbeda (Nisa dan Budiantara, 2016). *Spline truncated* telah dikembangkan dengan menggunakan Analisis Komponen Utama (AKU) dalam mengatasi masalah multikolinearitas (Islamiyati dkk., 2022). Penggunaan AKU bertujuan untuk mereduksi dimensi data yang saling berkorelasi menjadi dimensi data yang tidak saling berkorelasi lagi yaitu dengan variabel-variabel baru yang saling bebas (Mariana, 2013).

Regresi kuantil *spline truncated* dengan AKU akan diaplikasikan pada *Statistical Downscaling* (SD). SD adalah sebuah pendekatan statistik yang digunakan untuk mengetahui hubungan fungsional antara variabel iklim yang berskala global dengan variabel iklim yang berskala lokal. Hubungan fungsional antara variabel global atau variabel prediktor dengan variabel lokal atau variabel respon tersebut salah satunya dengan menggunakan model regresi. Data global yang dapat digunakan dalam SD adalah data *Global Circulation Model* (GCM). GCM berisi tentang informasi mengenai sirkulasi atmosfer yang menggambarkan berbagai macam subsistem iklim di bumi (Keller dkk., 2022). GCM berbentuk grid yang berskala besar atau petak wilayah dalam periode dan jangka waktu tertentu. Adapun data lokal yang dapat digunakan dalam SD untuk mengetahui perubahan iklim salah satunya adalah curah hujan (Sahriman dan Yulianti, 2023).

Iklim merupakan fenomena alam yang sangat penting terhadap berbagai aktivitas kehidupan manusia karena dapat memberi dampak di berbagai sektor. Unsur iklim yang memiliki keragaman dan fluktuasi paling tinggi di Indonesia adalah curah hujan. Intensitas curah hujan yang tidak berada dalam kondisi normal atau ekstrem dapat menimbulkan berbagai dampak buruk, seperti banjir, tanah longsor, angin puting beliung dan gelombang tinggi (Sunardi, Yudhana dan Muflih, 2020). Informasi mengenai cuaca dan iklim terkhusus pada curah hujan sangat dibutuhkan untuk mengurangi dampak buruk yang dapat ditimbulkan nelakukan sebuah analisis yang tepat. Oleh karena itu, penelitian ini mengenai regresi kuantil *spline truncated* dalam SD dengan ıakan analisis komponen utama untuk meramalkan curah hujan.

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, masalah yang dikaji dalam penelitian ini adalah :

1. Bagaimana bentuk estimasi parameter regresi kuantil *spline* dalam *statistical downscaling* dengan analisis komponen utama?
2. Bagaimana hasil dugaan curah hujan Kabupaten Pangkep menggunakan regresi kuantil *spline* dalam *statistical downscaling* dengan analisis komponen utama?

## 1.3 Batasan Masalah

Pada penelitian ini, penulis memberikan batasan masalah sebagai berikut:

1. Data curah hujan yang digunakan adalah data curah hujan bulanan Kabupaten Pangkep periode 2008-2022.
2. Nilai kuantil yang digunakan adalah 0.25, 050, dan 0.75
3. Estimator yang digunakan adalah spline *truncated.*

## 1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

1. Mendapatkan estimasi parameter regresi kuantil *spline* dalam *statistical downscaling* dengan analisis komponen utama
2. Mendapatkan hasil dugaan data curah hujan Kabupaten Pangkep menggunakan regresi kuantil *spline* dalam *statistical downscaling* dengan analisis komponen utama.

## 1.5 Manfaat Penelitian

Penelitian ini bermanfaat sebagai salah satu sumber referensi kepada pembaca tentang cara mengestimasi regresi kuantil *spline* dalam *statistical downscaling* dengan analisis komponen utama yang dapat dijadikan sebagai acuan dan literatur tambahan untuk penelitian selanjutnya yang memiliki relevansi dengan penelitian ini. Selain itu, dapat bermanfaat sebagai sumber informasi tentang curah hujan di Kabupaten Pangkep serta memberikan kontribusi terhadap pengembangan ilmu pengetahuan dan teknologi kedepannya.

# BAB II

# TINJAUAN PUSTAKA

## 2.1 Analisis Komponen Utama

Analisis Komponen Utama (AKU) merupakan sebuah metode yang digunakan untuk mengatasi masalah multikolinearitas. Adanya hubungan linier atau korelasi yang tinggi antar variabel prediktor dalam model regresi disebut dengan multikolinearitas. Salah satu cara untuk mengetahui keberadaan multikolinearitas dalam suatu model regresi adalah dengan menggunakan nilai *Tolerance* atau *Variance Inflation Factor* (VIF) pada Persamaan (2.1) berikut

$$VIF_j = \frac{1}{1-{R_j}^2} \tag{2.1}$$

dengan $j = 1,2,\dots,p$ dan $p$ adalah banyaknya variabel prediktor sedangkan ${R_j}^2$ adalah koefisien determinasi dari variabel prediktor yang diregresikan terhadap variabel prediktor lainnya. Nilai $VIF > 10$ mengindikasikan adanya multikolinearitas.

AKU akan mereduksi sekumpulan variabel yang saling berkorelasi menjadi variabel baru yang saling bebas atau tidak saling berkorelasi. Variabel baru yang terbentuk disebut komponen utama. Misalkan vektor acak $\boldsymbol{X}' = [X_1, X_2, \dots, X_p]$ dengan matriks ragam-peragam $\boldsymbol{\Sigma}$ dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ dan vektor eigen $\boldsymbol{e}' = (\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_p)$ maka kombinasi liniernya yaitu :

$$W_1 = \boldsymbol{e}_1'\boldsymbol{X} = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p$$
$$W_2 = \boldsymbol{e}_2'\boldsymbol{X} = e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p$$
$$\vdots$$
$$W_p = \boldsymbol{e}_p'\boldsymbol{X} = e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p \tag{2.2}$$

Kontribusi keragaman dari setiap komponen utama ke-*k* terhadap total keragaman dari matriks $\boldsymbol{\Sigma}$ adalah

$$proporsi = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \times 100\%, k = 1,2,\dots,p \tag{2.3}$$

Bila variabel yang diamati satuan ukurannya tidak sama, maka variabel tersebut akukan sehingga komponen utama ditentukan dari variabel baku atau orelasi. Total keragaman dari matriks korelasi yaitu

$$proporsi = \frac{\lambda_k}{p} \tag{2.4}$$

Menentukan komponen utama $W$ dengan banyaknya $W \leq p$ diharapkan dapat memuat semua informasi yang terdapat pada variabel asal sehingga data menjadi lebih sederhana. Metode yang digunakan untuk menentukan banyaknya komponen utama yaitu :

1. Berdasarkan proporsi kumulatif keragaman yang mampu dijelaskan, misalnya 75%.
2. Berdasarkan nilai eigen yang lebih besar atau sama dengan satu (Faizia, Prahutama dan Yasin, 2019).

## 2.2 Regresi Kuantil

Regresi kuantil merupakan suatu metode regresi dengan memisahkan atau membagi data menjadi kuantil-kuantil tertentu yang memiliki nilai dugaan yang berbeda. Regresi kuantil dapat digunakan untuk mengukur efek variabel prediktor tidak hanya di pusat sebaran data, tetapi juga pada bagian atas dan bawah ekor sebaran. Hal tersebut sangat berguna apabila masalah pencilan ingin diatasi dalam model regresi (Djuraidah dan Wigena, 2011). Pencilan merupakan kondisi amatan yang menyimpang jauh dari pengamatan lainnya. Pencilan dapat menyebabkan data tidak berdistribusi normal sehingga dapat berpengaruh terhadap pengambilan suatu kesimpulan pada sebuah penelitian (Sihombing dkk., 2022).

Misalkan terdapat data $\{x_{i1}, x_{i2}, \dots, x_{ip}, y_i\}$, $i = 1,2,\dots,n$ dan $j = 1,2,\dots,p$ adalah himpunan pasangan dari variabel acak yang berdistribusi secara independen dan identik dengan kuantil $\tau \epsilon (0,1)$. Data tersebut memiliki fungsi distribusi peluang bersyarat $F(y|x_i) = P(Y \leq y|x_i)$. Persamaan umum regresi kuantil linier khusus untuk kuantil bersyarat dari variabel respon $y_i$ yaitu:

$$y_i(\tau) = f(x_i) + \varepsilon_i(\tau)$$
$$f(x_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_p(\tau)x_{ip} \tag{2.5}$$

dengan

$y_i$ : variabel respon pada pengamatan ke-$i$ untuk $i = 1,2,\dots,n$

$\beta_0$ : intersep

$\beta_j$ : parameter dari variabel ke-$j$ untuk $j = 1,2,\dots,p$

$\tau$ : nilai kuantil $[0,1]$

iabel prediktor ke-$j$ dan pada pengamatan ke-$i$ untuk $i = 1,2,\dots,n$ dan

$1,2,\dots,p$

r ke-$i$ untuk $i = 1,2,\dots,n$

Jika model regresi kuantil disajikan dalam bentuk matriks, Persamaan (2.5) dapat ditulis seperti pada Persamaan (2.6):

$$\boldsymbol{y}(\tau) = \boldsymbol{X\beta}(\tau) + \boldsymbol{\varepsilon}(\tau)$$
$$\begin{bmatrix} y_1(\tau) \\ y_2(\tau) \\ \vdots \\ y_n(\tau) \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \vdots \\ \beta_p(\tau) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\tau) \\ \varepsilon_2(\tau) \\ \vdots \\ \varepsilon_n(\tau) \end{bmatrix} \tag{2.6}$$

Jika fungsi bersyarat dari kuantil ke−$\tau$ dengan variabel independen $X$ tertentu, maka fungsi bersyarat tersebut didefiniskan sebagai berikut:

$$Q_{y_i}(\tau | x_{i1}, x_{i2}, \dots, x_{ik}) = Q_y(\tau | X) = \boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \tag{2.7}$$

Koefisien $\boldsymbol{\beta}$ pada Persamaan (2.7) dapat diestimasi dengan meminimumkan fungsi objektif kuantil seperti berikut:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\,(\tau) \tag{2.8}$$

dengan $y_i = \{y_1, y_2, \dots, y_n\}$ merupakan sampel acak dengan variabel respon $Y$ dan $x_i \in \mathbb{R}^p$ merupakan vektor kovariat, sedangkan $\rho_\tau$ merupakan fungsi *loss* (Balami, 2017).

## 2.3 Estimasi Parameter Regresi Kuantil

Regresi kuantil diestimasi dengan meminimukan jumlah absolut *error* yang lebih dikenal dengan *least absolute deviation*. *Error* pada regresi kuantil diberi bobot yang berbeda yaitu pembobot $\tau$ untuk *error* non negatif dan pembobot $(1-\tau)$ untuk *error* negatif. Perkalian antara *error* dengan pembobot yang diberikan disebut fungsi loss $\rho_\tau$ yang dinyatakan sebagai berikut:

$$\rho_\tau(\varepsilon) = \sum_{i=1, \varepsilon_i \geq 0}^{n} \tau|\varepsilon_i| + \sum_{i=1, \varepsilon_i < 0}^{n} (1-\tau)|\varepsilon_i| \tag{2.9}$$

Hal tersebut membuat fungsi kuantil ke-$\tau$ dari variabel $y$ dengan syarat $x$ mempertimbangkan penduga $\beta_\tau$ sehingga permasalahan tersebut dapat diperoleh solusinya yang dinyatakan pada Persamaan (2.10) berikut :

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - Q_\tau(y|x)) \tag{2.10}$$

dengan :

$y_i$ : variabel respon ke-$i$

fungsi *loss*

indeks kuantil dengan $\tau \in (0,1)$

fungsi kuantil dari variabel $y$ dengan syarat $x$

Maka dari itu, fungsi kuantil bersyarat $Q_\tau(y|x)$ dapat didefinisikan sebagai berikut:

$$Q_\tau(y|x) = \boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \tag{2.11}$$

Nilai estimasi terhadap parameter $\beta$ dari regresi kuantil ke-$\tau$ diperoleh dengan meminimumkan fungsi *loss* dari Persamaan (2.10) sehingga didapatkan persamaan berikut

$$\hat{\beta}(\tau) = \min_\beta \sum_{i=1}^{n} \rho_\tau(\varepsilon) = \min_\beta \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}) \tag{2.12}$$

dengan $\rho_\tau(\varepsilon) = \begin{cases} \tau\varepsilon, & \varepsilon \geq 0 \\ (1-\tau)\varepsilon, & \varepsilon < 0 \end{cases}$

dengan mempertimbangkan estimasi $\beta(\tau)$, sehingga diperoleh solusi untuk permasalahan tersebut dapat dinyatakan sebagai berikut (Balami, 2017) :

$$\hat{\beta}(\tau) = \min_\beta \left\{ \tau \sum_{i=1,\varepsilon_i \geq 0}^{n} |y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| + (1-\tau) \sum_{i=1,\varepsilon_i < 0}^{n} |y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| \right\} \tag{2.13}$$

Pendugaan dalam regresi kuantil diperoleh dengan meyelesaikan masalah pemrograman linier. Salah satu metode yang dapat digunakan untuk pendugaan parameter regresi kuantil adalah dengan menggunakan metode simpleks. Metode simpleks banyak digunakan dalam aplikasi statistika dengan jumlah iterasi dapat meningkat secara eksponensial tergantung dari ukuran data. Algoritma simpleks memerlukan sebuah tabel simpleks yang dapat dilihat pada Tabel 2.1.

**Tabel 2.1.** Tabel simpleks

| $d_j$ | | | 0 | 0 | ... | 0 | $\tau$ | ... | $\tau$ | $(1-\tau)$ | ⋯ | $(1-\tau)$ | Rasio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $c_b$ | $v_b$ | $w_b$ | $x_0$ | $x_1$ | ... | $x_k$ | $q_1$ | ... | $q_n$ | $r_1$ | ⋯ | $r_n$ | |
| $\tau$ | $q_1$ | $b_1$ | $a_{ij}$ | | | | | | | | | | |
| $\tau$ | $q_2$ | $b_2$ | | | | | | | | | | | |
| ⋮ | ⋮ | ⋮ | | | | | | | | | | | |
| $\tau$ | $q_n$ | $b_n$ | | | | | | | | | | | |
| $z_j$ | | | | | | | | | | | | | |
| $d_j - z_j$ | | | | | | | | | | | | | |

an:

is koefisien fungsi tujuan.

m koefisien variabel basis.

$v_b$ : kolom dengan nama-nama variabel yang menjadi basis (variabel yang menyusun matriks identitas).

$w_b$ : kolom nilai ruas kanan dari kendala.

$z_j$ : baris dengan rumus $z_j = \Sigma d_i a_{ij}$ untuk $j = 1,2, \dots, n$.

Langkah-langkah untuk menduga parameter regresi kuantil sebagai berikut:

1. Mengubah masalah optimasi linier ke dalam bentuk stdanar.
2. Menentukan kolom kunci yaitu untuk masalah minimum pilih $d_j - z_j$ dengan hasil negatif terbesar sedangkan untuk maksimum dengan hasil positif yang terbesar.
3. Menentukan baris kunci, yaitu nilai rasio antara nilai ruas kiri $b_i$ dengan koefisien kolom kunci $a_{ij}$, pilih yang terkecil untuk masalah minimum dan maksimum. $Rasio = \frac{b_i}{a_{ij}}$ untuk rasio $> 0$.
4. Menentukan pivot dari perpotongan antara kolom kunci dan baris kunci yang dinamakan elemen kunci atau elemen penentu iterasi algoritma simpleks dan akan diubah nilainya menjadi 1
5. Selanjutnya, melakukan operasi baris dasar (OBD) berdasarkan pivot untuk baris lainnya, termasuk baris $d_j - z_j$ dengan nilai elemen-elemen yang termasuk di dalam kolom kunci dijadikan nol (selain elemen yang dijadikan pivot).
6. Proses iterasi untuk masalah masalah minimum berhenti jika semua nilai pada baris $d_j - z_j \geq 0$. Apabila masih ada $d_j - z_j < 0$, maka iterasi algoritma simpleks masih berlanjut (Hananingrum dan Achmad, 2021).

## 2.4 Regresi Nonparametrik *Spline Truncated*

Regresi nonparametrik adalah suatu metode yang digunakan apabila bentuk kurva regresi diasumsikan tidak diketahui. Regresi nonparametrik memiliki fleksibilitas yang tinggi dalam mengestimasi kurva regresi dibandingkan regresi parametrik (Islamiyati, 2014). Secara umum, model regresi nonparametrik yaitu :

$$y_i = f(x_i) + \varepsilon_i, \;\; i = 1,2, \dots, n \tag{2.14}$$

dengan $y_i$ merupakan variabel respon pada pengamatan ke-$i$, $f(x_i)$ merupakan ...gresi nonparametrik dan $\varepsilon_i$ merupakan *error* pada pengamatan ke-$i$.

...samaan (2.14) dengan fungsi $f(x_i)$ dapat dinyatakan sebagai fungsi ...salah satu estimator yang sering digunakan pada regresi nonparametrik dengan potongan-potongan polinomial yang memiliki sifat tersegmen dan kontinue adalah *spline truncated*. Fungsi *spline* berrorde $m$ dengan titik knot $k_1, k_2, \dots, k_r$ dapat ditulis dalam bentuk berikut:

$$f(x_i) = \sum_{g=0}^{m} \beta_g x_i^g + \sum_{h=1}^{r} \beta_{m+h} (x_i - k_h)_+^m \qquad (2.15)$$

maka diperoleh model regresi nonparametrk *spline truncated* sebagai berikut:

$$y_i = \sum_{g=0}^{m} \beta_g x_i^g + \sum_{h=1}^{r} \beta_{m+h} (x_i - k_h)_+^m + \varepsilon_i \qquad (2.16)$$

Sedangkan bentuk umum regresi *spline truncated* dengan prediktor sebanyak *p* yaitu :

$$\begin{aligned} y_i &= f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ip}) + \varepsilon_i \qquad (2.17) \\ &= \sum_{j=1}^{p} f(x_{ij}) + \varepsilon_i \,, \quad i = 1,2, \dots, n; \quad j = 1,2, \dots, p \end{aligned}$$

dengan $f(x_{ij}) = \sum_{g=0}^{m} \beta_{gj} x_{ij}^g + \sum_{h=1}^{r} \beta_{(m+h)j} (x_{ij} - k_{hj})_+^m$

$$(x_{ij} - k_{hj})_+^m = \begin{cases} (x_{ij} - k_{hj})^m, & x_{ij} \geq k_{hj} \\ 0, & x_{ij} < k_{hj} \end{cases}$$

Keterangan :

$y_i$ : variabel respon pada pengamatan ke-*i*

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

$\beta_{gj}$ : parameter polinomial pada prediktor ke-*j* dan orde ke-*g*

$\beta_{(m+h)j}$ : parameter *truncated* pada prediktor ke-*j* dan titik knot ke-$(m + h)$

$k_{hj}$ : nilai titik knot ke-*h* pada prediktor ke-*j*

$r$ : banyaknya titik knot

$m$ : orde polinomial *spline truncated*

$\varepsilon_i$ : *error* pada pengamatan ke-*i* (Nisa dan Budiantara, 2016)

## 2.5 Estimasi Fungsi *Spline Truncated*

Model regresi nonparametrik pada Persamaan (2.14) untuk setiap pengamatan ke-*i* adalah sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

a regresi dihampiri dengan fungsi *spline* yang memuat knot *k* seperti samaan (2.15), maka model regresi *spline* dalam bentuk matriks dapat sebagai berikut :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m & (x_1 - k_1)_+^m & \dots & (x_1 - k_r)_+^m \\ 1 & x_2 & x_2^2 & \dots & x_1^m & (x_1 - k_1)_+^m & \dots & (x_1 - k_r)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_1^m & (x_1 - k_1)_+^m & \dots & (x_1 - k_r)_+^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \\ \beta_{(m+1)} \\ \vdots \\ \beta_{(m+r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Model regresi *spline* juga dapat ditulis dalam bentuk vektor yaitu:

$$\boldsymbol{y} = \boldsymbol{X}[k]\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.18}$$

$\boldsymbol{y}$ merupakan vektor variabel respon pada pengamatan, $\boldsymbol{X}[k_1, k_2, \dots, k_r]$ merupakan matriks yang elemennya terdiri dari fungsi regresi dan fungsi *truncated* yang memuat titik knot *k*, $\boldsymbol{\beta}$ merupakan vektor koefisien regresi dan $\boldsymbol{\varepsilon}$ merupakan vektor yang elemennya adalah *error* pada pengamatan. Estimasi parameter $\boldsymbol{\beta}$ pada Persamaan (2.18) menggunakan metode kuadrat terkecil, yaitu dengan meminimumkan jumlah kuadrat *error* yaitu:

$$\begin{aligned} \min(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) &= \min\left(\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}' \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}\right) \\ &= \min\{(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})\} \end{aligned}$$

Jika dijabarkan terhadap jumlah kuadrat *error* maka diperoleh:

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon'\varepsilon &= (\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta}) \\ &= (\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}[k]'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}[k]'\boldsymbol{X}[k]\boldsymbol{\beta} \end{aligned} \tag{2.19}$$

Persamaan (2.19) jika diturunkan terhadap vektor $\boldsymbol{\beta}$ dengan ruas kanan disamakan dengan nol, maka:

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial\boldsymbol{\beta}} = \frac{\partial(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}[k]\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}[k]'\boldsymbol{X}[k])^{-1}\boldsymbol{X}[k]'\boldsymbol{y} \tag{2.20}$$

Berdasarkan Persamaan (2.20), diperoleh estimasi fungsi regresi *spline* dengan titik knot *k* berikut:

$$\begin{aligned} \hat{f}(x_i) &= \boldsymbol{X}[k]\,\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}[k](\boldsymbol{X}[k]'\boldsymbol{X}[k])^{-1}\boldsymbol{X}[k]'\boldsymbol{y} \end{aligned} \tag{2.21}$$

Kemudian, berdasarkan Persamaan (2.21) estimasi fungsi dari titik knot dapat dituliskan sebagai berikut :

$$\hat{f}(x_i) = \boldsymbol{A}[k]\boldsymbol{y}$$

dengan $\boldsymbol{A}[k] = \boldsymbol{X}[k](\boldsymbol{X}[k]'\boldsymbol{X}[k])^{-1}\boldsymbol{X}[k]'$ (Hidayat, Yuliani dan Sam, 2018)

## 2.6 Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan bersama yang menunjukkan perubahan perilaku pada data. Model regresi *spline* terbaik tergantung pada titik knot optimal. Metode yang sering dipakai untuk mencari titik knot optimal adalah *Generalized Cross Validation* (GCV). Nilai GCV yang memberikan titik knot optimal adalah nilai GCV yang minimum. Rumus GCV yaitu sebagai berikut:

$$GCV(k) = \frac{MSE(k)}{(n^{-1}trace[\boldsymbol{I}-\boldsymbol{A}(k)])^2} \tag{2.22}$$

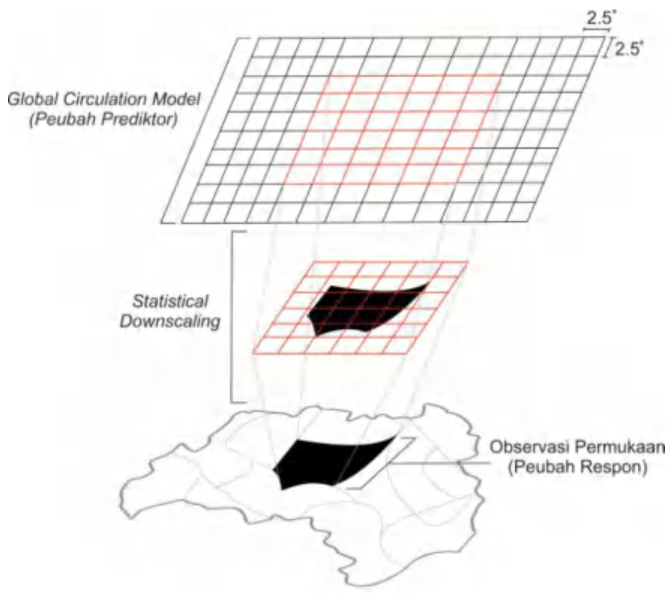
untuk $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2$ dengan $\boldsymbol{I}$ adalah matriks Identitas dan $n$ : jumlah pengamatan (Sanusi, Syam dan Adawiyah, 2019)

## 2.7 *Statistical Downscaling*

*Downscaling* merupakan suatu proses transformasi dari data yang berkala besar menjadi data berskala kecil. *Downscaling* bekerja dengan cara melakukan interpolasi pada variabel iklim global terhadap variabel iklim lokal (Cahyani, Wigena dan Djuraidah, 2016). *Downscaling* terdiri atas *Dynamical Downscaling* (DD) dan *Statistical Downscaling* (SD). DD dilakukan dengan menetapkan kondisi awal dan kondisi batas pada data global. DD jarang digunakan dalam studi perubahan iklim karna membutuhkan komputasi yang kompleks dengan periode waktu yang panjang dan membuat beberapa skenario. Adapun SD berdasarkan hubungan fungsional antara variabel iklim global dengan variabel iklim lokal (Keller dkk., 2022). Pendekatan SD disusun berdasarkan adanya hubungan antara variabel prediktor dengan grid beskala besar dan variabel respon dengan grid berskala kecil yang dinyatakan dalam model statistik (Zorita dan Storch, 1999). Hubungan tersebut terdiri atas dua kategori yaitu SD yang berbasis regresi dan SD yang berbasis klasifikasi tipe cuaca. SD berbasis regresi ...n untuk variabel lokal yang kontinu berdasarkan waktu seperti suhu, ...an kelembapan. Adapun SD berbasis klasifikasi tipe cuaca digunakan ...iabel lokal yang kontinu dan diskrit (Sailor dan Li, 1999).

Variabel global dalam SD memanfaatkan data *Global Circulation Model* (GCM) yang merupakan salah satu alat penting dalam studi keragaman dan perubahan iklim. GCM menggambarkan sejumlah subsistem dari iklim di bumi, seperti proses-proses di atmosfer, lautan, daratan, maupun mensimulasi kondisi iklim berskala global (Keller dkk., 2022). GCM mensimulasi variabel-variabel iklim global pada setiap grid (berukuran $\pm 2.5°$ atau $\pm 300\ km^2$) untuk lapisan (layer) atmosfer yang selanjutnya digunakan dalam menduga pola-pola iklim dalam jangka waktu tertentu (Upa dan Sahriman, 2021). GCM tidak dapat digunakan secara langsung sebagai variabel prediktor pada model regresi biasa karna merupakan data spasial dan temporal yang dapat menimbulkan terjadinya korelasi yang kuat antar grid dalam domain yang saling berdekatan (Goldameir, 2015)

Model SD akan memberikan hasil yang baik apabila hubungan antara variabel respon dengan prediktor memiliki korelasi yang tinggi untuk menjelaskan keragaman iklim lokal dengan baik, variabel prediktor dapat disimulasikan dengan baik oleh GCM, dan hubungan antara variabel respon dengan prediktor tidak berubah dengan adanya perubahan waktu serta tetap sama meskipun ada perubahan iklim di masa depan. Oleh karena itu, penentuan domain (lokasi dan jumlah grid) merupakan faktor penting yang dapat mempengaruhi kestabilan peramalan (Matulessy, 2015). Ilustrasi proses SD dapat dilihat pada Gambar 21.



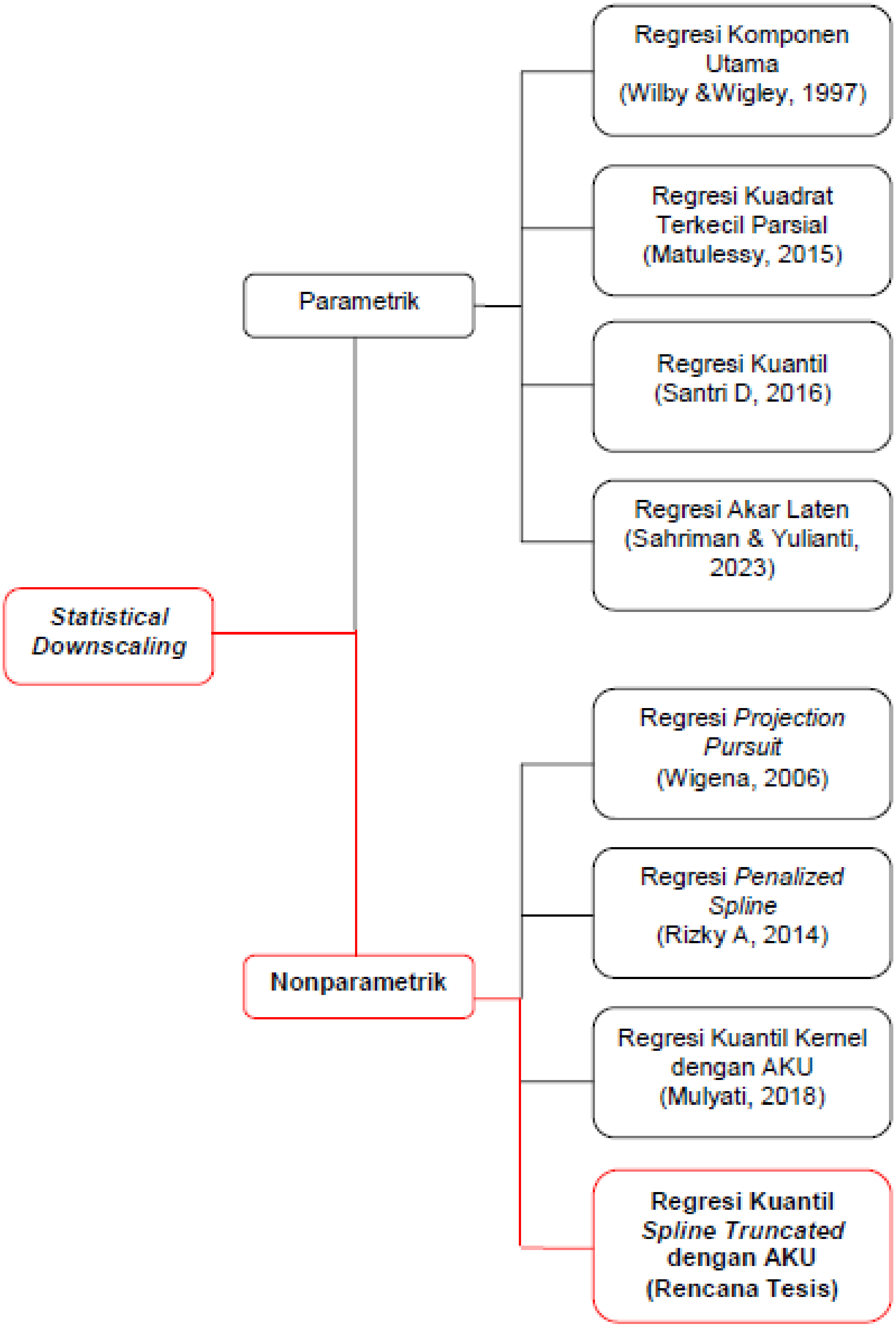

**Gambar 2.1** Ilustrasi proses *statistical downscaling*

## 2.8 Curah Hujan

Iklim merupakan fenomena alam yang sangat penting terhadap berbagai aktivitas kehidupan manusia karena dapat memberi dampak di berbagai sektor, misalnya kesehatan, pertanian, perekonomian dan industri. Unsur iklim yang memiliki keragaman dan fluktuasi paling tinggi di Indonesia adalah curah hujan. Ketinggian air hujan yang berada pada suatu tempat yang datar, tidak mengalir, tidak meresap, dan tidak menguap disebut dengan curah hujan. Satuan curah hujan dinyatakan dalam milimeter atau inci. Curah hujan dalam satu milimeter memiliki arti bahwa pada suatu tempat yang datar dengan luas satu meter persegi, tertampung air setinggi satu milimeter atau sebanyak satu liter air (Prawaka, Zakaria dan Tugiono, 2016)

Intensitas curah hujan yang beragam menjadikannya penting untuk dilakukan pencatatan. Hal tersebut berguna untuk memahami karakteristik curah hujan yang turun pada suatu wilayah tertentu. Pentingnya dalam meramalkan curah hujan selain untuk mengetahui debit air pada sumber air permukaan dapat juga digunakan untuk membantu memperkecil dampak bencana yang ditimbulkan seperti banjir, tanah longsor, angin puting beliung dan gelombang tinggi (Sunardi, Yudhana dan Muflih, 2020).

## 2.9 Kerangka Konseptual




**Gambar 2.2** Kerangka konseptual