

## Peramalan Curah Hujan Di Kota Samarinda Menggunakan Vector Error Correction Model

### *Rainfall Forecasting In Samarinda City Using Vector Error Correction Model*

Ilyas Astafira<sup>1a)</sup>, Meiliyani Siringoringo<sup>2</sup>, M. Fathurahman<sup>3</sup>

<sup>1,2,3</sup>Program Studi Statistika, FMIPA, Universitas Mulawarman, Indonesia

<sup>1, 2</sup>Laboratorium Statistika Ekonomi dan Bisnis, Jurusan Matematika, FMIPA, Universitas Mulawarman, Indonesia

<sup>3</sup>Laboratorium Statistika Komputasi, Jurusan Matematika, FMIPA, Universitas Mulawarman, Indonesia

<sup>a)</sup>Corresponding author: ilyasstat20@gmail.com

#### ABSTRACT

The Vector Error Correction Model (VECM) was one of the multivariate time series models that was a development of the Vector Autoregressive (VAR). VECM could be used to forecast non-stationary time series variables that had cointegration relationships. This study used monthly data of rainfall, minimum air temperature, and maximum air humidity variables from January 2015 to December 2023 to form the VECM model. The purpose of this study was to obtain a VECM model for rainfall in the city of Samarinda and to forecast rainfall in the city of Samarinda using VECM. The results of the study showed that the VECM model that formed was VECM(1) with two cointegration relationships. The rainfall forecasted results with VECM(1) showed a downward trend until April 2024 and a horizontal pattern from May to December, with forecasting accuracy of 32.369% measured using MAPE, which indicating quite good forecasting.

**Keywords:** rainfall, stationarity, cointegration, var, vecm

#### 1. Pendahuluan

Peramalan adalah perkiraan munculnya sebuah kejadian di masa depan berdasarkan data yang ada di masa lampau yang dilakukan secara sistematis. Metode peramalan dibagi menjadi 2 berdasarkan sifatnya yaitu peramalan kualitatif dan kuantitatif. Metode peramalan kualitatif adalah pendekatan subjektif yang mengandalkan opini, intuisi, dan pengalaman dari para ahli yang terlibat dalam peramalan, sedangkan metode peramalan kuantitatif menggunakan model matematis dan statistika untuk membuat prediksi. Salah satu teknik yang umum digunakan dalam metode peramalan kuantitatif adalah analisis runtun waktu. Analisis runtun waktu adalah kumpulan metode statistika yang digunakan untuk menganalisis data runtun waktu. Data runtun waktu sendiri adalah serangkaian nilai-nilai pengamatan yang diukur dalam interval waktu tertentu (Kusdarwati, dkk., 2022).

Metode dalam analisis runtun waktu terbagi menjadi dua berdasarkan banyaknya variabel, yaitu metode univariat dan multivariat. Metode multivariat dalam analisis runtun waktu digunakan untuk menganalisis adanya hubungan antara beberapa variabel. Jenis-jenis hubungan dalam data runtun waktu dibagi menjadi dua yaitu hubungan jangka pendek dan hubungan jangka panjang. Hubungan jangka pendek adalah keterkaitan antara beberapa variabel pada interval waktu yang singkat seperti harian atau bulanan, sedangkan hubungan jangka panjang adalah keterkaitan beberapa variabel pada periode yang lama seperti tahunan yang dapat menggambarkan tren atau pola siklik (Shin, 2019).

Salah satu model runtun waktu multivariat yang cukup terkenal adalah *Vector Autoregressive* (VAR). Model VAR adalah jenis model deret waktu multivariat yang merupakan perluasan model autoregresif untuk menangkap hubungan diantara beberapa variabel. Model VAR mengharuskan semua variabel stasioner. Jika variabel dalam VAR tidak stasioner, maka kointegrasi dapat terjadi jika terdapat kombinasi linier diantara variabel tersebut yang stasioner. Kointegrasi adalah hubungan jangka panjang antar variabel meskipun variabel tersebut tidak stasioner. Jika terdapat kointegrasi dalam model VAR, maka model yang digunakan adalah *Vector Error Correction Model* (VECM) (Juselius, 2019). VECM merupakan model runtun waktu multivariat yang menaksir hubungan jangka panjang dan jangka pendek antara variabel nonstasioner yang terkointegrasi. (Winarno, dkk., 2021).

Metode VECM merupakan metode ekonometrika sehingga penerapan metode ini kebanyakan pada data ekonomi seperti data inflasi (P. Das, 2019), namun penerapan metode ini pada data non-ekonomi, seperti data iklim (curah hujan dan suhu), masih jarang ditemukan. Penelitian sebelumnya yang menerapkan VECM pada data curah hujan dilakukan oleh Matulesy (2023) yang bertujuan untuk memodelkan curah hujan di Kabupaten Manokwari menggunakan VECM dan ARIMA menggunakan data bulan Januari 2012 hingga Desember 2017. Hasil penelitian menunjukkan VECM lebih unggul dari model ARIMA dalam pemodelan curah hujan yang ditunjukkan oleh metrik validasi yang lebih rendah seperti *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE). Selanjutnya penelitian di India

yang dilakukan oleh Das dan Srinivasan (2019) bertujuan untuk menganalisis hubungan antara variasi suhu dan curah hujan akibat perubahan kondisi iklim dan juga meramalkan tingkat variabel-variabel ini di masa depan. Penelitian ini menggunakan VECM untuk menganalisis data suhu dan curah hujan selama 106 tahun, yang menunjukkan hubungan antara kedua variabel tersebut yaitu suhu dan curah hujan. Yollanda, dkk., (2023) melakukan penelitian yang bertujuan untuk memprediksi sirkulasi Cross Equatorial Northerly Surge (CENS) dan Madden-Julian Oscillation (MJO) serta curah hujan di Indonesia menggunakan VECM. Hasil penelitian tersebut mengungkapkan bahwa CENS aktif berhubungan dengan peningkatan curah hujan di wilayah Indonesia bagian selatan, sedangkan CENS tidak aktif berhubungan dengan curah hujan yang lebih sedikit. Model VECM memberikan prediksi akurat yang untuk CENS, MJO, dan intensitas curah hujan di 14 kluster, dengan tingkat akurasi model yang diukur menggunakan nilai MAPE sebesar 1,851 untuk CENS dan 0,061 untuk MJO.

Berdasarkan uraian yang telah dipaparkan, maka akan dilakukan peramalan curah hujan di Kota Samarinda menggunakan metode VECM. Pada penelitian ini variabel terikat yang digunakan sama seperti penelitian sebelumnya yaitu curah hujan, akan tetapi variabel bebas dan lokasi yang digunakan dalam penelitian ini berbeda. Selain itu penelitian ini menggunakan data yang terbaru yang dipublikasi oleh Badan Pusat Statistik Kota Samarinda.

## 2. Tinjauan Pustaka

Analisis runtun waktu adalah metode statistika yang digunakan untuk menganalisa data runtun waktu yang dikumpulkan pada interval waktu tertentu. Peramalan adalah metode analisis runtun waktu yang melibatkan penggunaan model statistik untuk memprediksi nilai pada masa depan berdasarkan nilai yang diamati sebelumnya. Peramalan bergantung pada data historis yang dikumpulkan secara berkala (Wei, 2019).

### 2.1. Stasioneritas

Stasioneritas adalah sifat data runtun waktu ketika rata-rata dan variansinya konstan sepanjang waktu. Jika suatu data runtun waktu tidak stasioner, maka data runtun waktu tersebut dapat diubah menjadi data runtun waktu yang stasioner secara berurutan yaitu transformasi untuk menjadi stasioner dalam variansi dan *differencing* untuk menjadi stasioner dalam rata-rata. Transformasi yang dimaksud disini adalah transformasi *Box-Cox* pada Persamaan (1):

$$y_t = \begin{cases} z_t^\lambda, & \lambda \neq 0 \\ \ln(z_t), & \lambda = 0 \end{cases} \quad (1)$$

Setelah data dilakukan transformasi *Box-Cox*, maka dapat dilanjutkan dengan pengujian stasioner dalam rata-rata menggunakan uji ADF sebagai berikut.

$H_0$  : Data stasioner dalam rata-rata

$H_1$  : Data tidak stasioner dalam rata-rata

Statistik hitung yang digunakan pada Persamaan (2):

$$\Delta z_t = \gamma z_{t-1} + \sum_{i=1}^p \Delta z_{t-i} \quad (2)$$

$$\tau = \frac{\gamma}{SE(\gamma)}$$

dengan  $\tau$  adalah statistik hitung,  $\gamma$  adalah parameter  $z_{t-1}$ ,  $SE(\gamma)$  adalah simpangan baku  $\gamma$  dan  $\Delta z_t$  adalah *differencing* data runtun waktu  $z_t$  dari Persamaan (3):

$$\Delta z_t = z_t - z_{t-1} \quad (3)$$

Jika data runtun waktu tidak stasioner dari pengujian ADF, maka perlu dilakukan *differencing* menggunakan Persamaan (4) (Wei, 2019).

### 2.2. Pengujian Dependensi

Pengujian dependensi dilakukan untuk mengetahui apabila terdapat hubungan antar variabel menggunakan uji *Bartlet test of sphericity*. Pengujian ini dilakukan dengan menguji kesamaan matriks korelasi dengan matriks identitas yang bertujuan menunjukkan adanya korelasi antar variabel. Prosedur pengujian *Bartlet test of sphericity* adalah sebagai berikut:

$H_0$  :  $\mathbf{R} = \mathbf{I}$  (Tidak terdapat korelasi antar variabel)

$H_1$  :  $\mathbf{R} \neq \mathbf{I}$  (Terdapat korelasi antar variabel)

Statistik Hitung dapat dilihat pada Persamaan (4):

$$\chi_{hitung}^2 = -\left(n - 1 - \frac{2m + 5}{6}\right) \ln(\mathbf{R}) \quad (4)$$

dengan:

$$\mathbf{R} = [r_{ij}]_{m \times m}$$

$$r_{ij} = \frac{\sum_{t=1}^n (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)}{\sqrt{\sum_{t=1}^n (z_{it} - \bar{z}_i)^2 \sum_{t=1}^n (z_{jt} - \bar{z}_j)^2}}$$

- $z_{it}$  = nilai pengamatan variable ke- $i$  pada waktu ke- $t$
- $\mathbf{R}$  = matriks korelasi
- $\bar{z}_i$  = rata-rata pengamatan variabel ke- $i$
- $n$  = banyaknya data
- $m$  = banyaknya variabel
- $r_{ij}$  = korelasi antara variabel ke- $i$  dan variabel ke- $j$

Jika nilai  $\chi_{hitung}^2$  lebih besar dari  $\chi_{m(m-1)/2;\alpha}^2$  dengan  $m$  adalah banyaknya variabel dan  $\alpha$  adalah taraf signifikansi, maka diputuskan terdapat korelasi antar variabel (Denis, 2021).

### 2.3. Model Vector Autoregressive

Pemodelan VAR dapat dilanjutkan jika hasil pengujian dependensi menunjukkan hasil yang signifikan. Model VAR adalah model yang digunakan untuk menganalisis variabel runtun waktu berdasarkan variabel itu sendiri dan variabel lain dari masa lampau. Persamaan model VAR dapat dilihat pada Persamaan (5) (Mills, 2019):

$$\mathbf{z}_t = \Phi_0 + \sum_{h=1}^p \Phi_h \mathbf{z}_{t-h} + \mathbf{a}_t \tag{5}$$

dengan:

- $\mathbf{z}_t$  = matriks pengamatan ke- $t$ ,  $t = 1, 2, \dots, n$
- $\Phi_h$  = matriks parameter ke- $h$ ,  $h = 0, 1, 2, \dots, p$
- $\mathbf{a}_t$  = vektor *residual* ke- $t$ ,  $t = 1, 2, \dots, n$

Penaksiran model VAR Persamaan (5) dapat dilihat pada Persamaan (6):

$$\Xi = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} \tag{6}$$

dengan:

$$\begin{aligned} \mathbf{Y} &= [\mathbf{z}_{p+1} \quad \mathbf{z}_p \quad \mathbf{z}_{p-1} \quad \dots \quad \mathbf{z}_n]^T \\ \mathbf{Z} &= \begin{bmatrix} 1 & \mathbf{z}_p & \mathbf{z}_{p-1} & \dots & \mathbf{z}_1 \\ 1 & \mathbf{z}_{p+1} & \mathbf{z}_p & \dots & \mathbf{z}_2 \\ 1 & \mathbf{z}_{p+2} & \mathbf{z}_{p+1} & \dots & \mathbf{z}_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{z}_{n-1} & \mathbf{z}_{n-2} & \dots & \mathbf{z}_{n-p} \end{bmatrix}, h > 1 \\ \Phi_h &= \begin{bmatrix} \Phi_{1,1,h} & \Phi_{1,2,h} & \dots & \Phi_{1,m,h} \\ \Phi_{2,1,h} & \Phi_{2,2,h} & \dots & \Phi_{2,m,h} \\ \vdots & \vdots & \dots & \dots \\ \Phi_{m,1,h} & \Phi_{m,2,h} & \dots & \Phi_{m,m,h} \end{bmatrix}, h > 1 \\ \mathbf{z}_t &= [z_{1,t} \quad z_{2,t} \quad \dots \quad z_{m,t}] \\ \Xi &= [\Phi_0 \quad \Phi_1 \quad \dots \quad \Phi_p]^T \\ \Phi_0 &= [\Phi_{0,1} \quad \Phi_{0,2} \quad \dots \quad \Phi_{0,m}]^T \end{aligned}$$

Setelah dilakukan penaksiran model VAR, maka dapat dilakukan pemilihan *lag* optimum yang bertujuan menentukan model VAR yang tidak akan mengalami *underfitting* dan *overfitting*. Pemilihan *lag* yang optimum dapat dilakukan menggunakan pengujian *Likelihood Ratio Test* (LRT) sebagai berikut (Reinsel et al., 2022):

$$\begin{aligned} H_0 &: \Phi_h = \mathbf{0} \\ H_1 &: \Phi_h \neq \mathbf{0} \end{aligned}$$

Statistik hitung dapat dilihat pada Persaman (7):

$$\Lambda(h) = -(n - P - 1,5 - mh) \ln \left( \frac{|\sigma_{a,h}|}{|\sigma_{a,h-1}|} \right) \tag{7}$$

dengan:

- $\sigma_{a,h}$  = matriks kovariansi *residual* model VAR( $h$ )
- $|\sigma_{a,h}|$  = determinan matriks kovariansi *residual* model VAR( $h$ )
- $P$  = orde maksimum model VAR yang akan diujikan
- $\mathbf{a}_t$  = vektor *residual* ke- $t$ ,  $t = 1, 2, \dots, n$
- $n$  = banyaknya pengamatan
- $m$  = banyaknya variabel
- $h$  = *lag* yang diujikan

Daerah kritis statistik hitung pada Persamaan (7) adalah  $\Lambda(h) > \chi_{\alpha, m^2}^2$  dengan  $\chi_{\alpha, m^2}^2$  adalah nilai kritis distribusi *chi-square* pada taraf signifikansi  $\alpha$  dan derajat bebas  $m^2$ .

**2.4. Pengujian Stabilitas**

Pengujian stabilitas model dilakukan setelah pemilihan *lag* optimum karena pengujian ini dilakukan pada model VAR dengan *lag* optimum (Wei, 2005). Model VAR dikatakan stabil jika semua nilai  $K$  yang memenuhi Persamaan (4) memiliki nilai kurang dari satu:

$$K^p \mathbf{I}_p - \sum_i^p \Phi_i K^{p-i} = 0 \tag{8}$$

dengan  $\Phi_i$  adalah parameter model VAR ke- $i$  dan  $\mathbf{I}_p$  adalah matriks identitas berukuran  $p \times p$ .

**2.5. Pengujian Kointegrasi**

*Lag* optimum terpilih selanjutnya akan digunakan dalam pengujian kointegrasi. Salah satu metode yang dapat digunakan untuk menguji kointegrasi adalah uji Johansen. Terdapat dua pendekatan dalam metode ini yaitu *trace statistic* dan *maximum eigenvalue statistic*. Pengujian kointegrasi dengan *trace statistic* menggunakan semua nilai *eigen* sehingga dapat menangkap lebih banyak informasi dari hubungan kointegrasi dibanding *maximum eigenvalue* yang hanya menggunakan nilai *eigen* yang tertinggi. Adapun pengujian *trace statistic* adalah sebagai berikut (Onatski & Wang, 2019):

$H_0 : r = r_0$  (Banyaknya vektor kointegrasi adalah  $r_0$ )

$H_1 : r > r_0$  (Banyaknya vektor kointegrasi lebih besar dari  $r_0$ ),

Statistik hitung yang digunakan dapat dilihat pada Persamaan (9):

$$L(r_0) = -n \sum_{i=r_0+1}^m \ln(1 - \lambda_i) \tag{9}$$

dengan  $n$  adalah banyaknya data,  $m$  adalah banyaknya variabel,  $r_0$  adalah *rank* matriks  $\Pi$  yang diujikan dan  $\lambda_i$  adalah nilai *eigen* matrix  $\Pi$  dari Persamaan (10).  $H_0$  ditolak jika statistik hitung ( $L(r_0)$ ) lebih besar dari nilai kritis ( $\lambda_{r_0, \alpha}$ ). Uji Johansen dengan pendekatan *trace statistic* dilakukan secara berurutan, dimulai dari tidak ada vektor kointegrasi ( $r_0 = 0$ ) sampai dengan  $r_0 = m$  ( $m$  adalah banyaknya variabel) dan pengujian ini dihentikan ketika terdapat  $H_0$  yang gagal ditolak pada nilai  $r_0$  tertentu.

**2.6. Vector Error Correction Model**

VECM merupakan model runtun waktu multivariat yang digunakan untuk menganalisis data runtun waktu yang tidak stasioner dan terkointegrasi. Asumsi dalam model VECM adalah semua variabel harus bersifat tidak stasioner dan adanya kointegrasi diantara variabel. Persamaan model VECM secara umum adalah sebagai berikut:

$$\mathbf{z}_t = \Pi \mathbf{z}_{t-p} + \Gamma_0 + \sum_{h=1}^{p-1} \Gamma_h \mathbf{z}_{t-h} + \mathbf{e}_t \tag{10}$$

dengan:

$$\Gamma_h = - \sum_{h_1=1}^h \Phi_{h_1}$$

$p$  = *lag* optimum model VAR

$\mathbf{e}_t$  = vektor *residual* waktu ke- $t$

$\Phi_{h_1}$  = matriks parameter model VAR dari Persamaan (5)

$\Gamma_h$  = matriks parameter model VECM ke- $h$  untuk  $h = 0, 1, 2, \dots, p - 1$

$\Pi$  = matriks kointegrasi VECM (Mills, 2019).

**2.7. Pengujian Diagnostik Residual**

Pengujian diagnostik *residual* dapat dilakukan setelah model VECM telah ditaksir. Pengujian diagnostik *residual* terdiri dari pengujian non-otokorelasi dan normalitas *residual*. Non-otokorelasi *residual* merupakan asumsi penting dalam analisis runtun waktu yang berarti bahwa *residual* dari model tidak bergantung satu sama lain. Salah satu metode yang umum digunakan untuk menguji otokorelasi adalah metode Ljung-Box yang digunakan untuk mendeteksi keberadaan otokorelasi sampai pada *lag* tertentu. Pengujian hipotesis multivariat Ljung-Box adalah sebagai berikut (Huang & Petukhina, 2022):

$H_0$  : Tidak terdapat otokorelasi pada *residual*

$H_1$  : Terdapat otokorelasi pada *residual*

Statistik hitung dapat dilihat pada Persamaan (11):

$$Q_m(h) = n(n + 2) \sum_{h=1}^M \frac{1}{n - h} \text{tr}(\hat{\Gamma}_h \hat{\Gamma}_h^{-1} \hat{\Gamma}_h' \hat{\Gamma}_h^{-1}) \tag{11}$$

dengan:

- $n$  = banyaknya data
- $h$  = lag yang sedang diujikan
- $M$  = banyaknya lag
- $\hat{\Gamma}_h$  = taksiran matriks kovariansi silang pada lag ke- $h$ ,  $h = 0, 1, 2, \dots, M$

Daerah kritis dari Persamaan (11) adalah  $Q_m(h) > \chi_{\alpha, hm^2}^2$  dengan  $\chi_{\alpha, hm^2}^2$  adalah nilai kritis distribusi *chi-square* pada taraf signifikansi  $\alpha$  dan derajat bebas  $hm^2$ .

Setelah dilakukan pengujian otokorelasi *residual*, maka bisa dilanjutkan dengan pengujian normalitas *residual* untuk mengetahui apabila *residual* berdistribusi normal menggunakan *Probability Plot Correlation Coefficient* (PPCC). Langkah-langkah pengujian normalitas dengan PPCC adalah sebagai berikut:

1. Mengurutkan data sampel dari yang terkecil hingga terbesar
2. Menghitung kuantil distribusi normal
3. Menghitung PPCC menggunakan Persamaan (12)

$$r_Q = \frac{\sum_{t=1}^n (z_{(t)} - \bar{z})(q_{(t)} - \bar{q})}{\sqrt{\sum_{t=1}^n (z_{(t)} - \bar{z})^2 \sum_{t=1}^n (q_{(t)} - \bar{q})^2}} \tag{12}$$

Setelah dilakukan perhitungan PPCC, maka dapat dilakukan pengujian hipotesis PPCC sebagai berikut:

- $H_0$  : *Residual* berdistribusi normal
- $H_1$  : *Residual* tidak berdistribusi normal

$H_0$  ditolak jika statistik uji PPCC yang dihitung dari Persamaan (12) kurang dari nilai kritis PPCC (Winarno, dkk., 2021).

### 2.8. Akurasi Peramalan

Keakuratan peramalan merujuk pada sejauh mana prediksi yang dibuat tentang suatu peristiwa atau data di masa depan mendekati kenyataan. Keakuratan peramalan dapat diukur dengan berbagai metode yang masing-masing memberikan informasi tentang seberapa baik model atau metode peramalan yang digunakan. Salah satu metode yang umum digunakan untuk mengukur akurasi peramalan adalah *Mean Absolute Error Percentage* (MAPE) yang dapat dilihat pada Persamaan (13) (Rafferty, 2023):

$$MAPE = \sum_{t=1}^n \left| \frac{z_t - \hat{z}_t}{z_t} \right| \times 100\% \tag{13}$$

dengan  $z_t$  adalah nilai aktual runtun waktu,  $\hat{z}_t$  adalah nilai prediksi data runtun waktu, dan  $n$  adalah banyaknya data. Interpretasi nilai MAPE dapat dilihat pada Tabel 1.

Tabel 1. Interpretasi Nilai MAPE

Rentang nilai MAPE	Interpetasi
$MAPE < 10\%$	Sangat baik
$10\% \leq MAPE < 20\%$	Baik
$20\% \leq MAPE < 50\%$	Cukup
$MAPE \geq 50\%$	Buruk

### 3. Metodologi Penelitian

Variabel yang digunakan dalam penelitian ini adalah Curah Hujan, Suhu Udara Minimum, Kelembapan Udara Maksimum yang diambil dari Publikasi Badan Pusat Statistik Provinsi Kalimantan Timur tahun 2015-2023 yang dapat diakses pada link berikut: <https://kaltim.beta.bps.go.id/id>. Perhitungan pada analisis data akan dilakukan menggunakan bantuan *software* R. Langkah-langkah peramalan metode VECM adalah sebagai berikut:

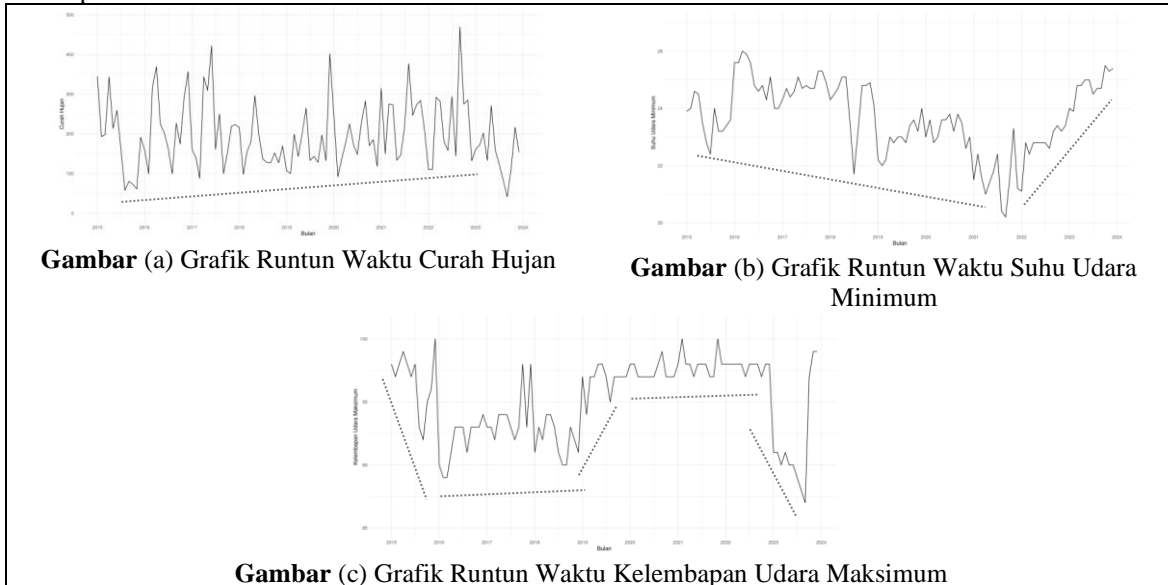
1. Melakukan analisis statistika deskriptif menggunakan grafik runtun waktu.
2. Melakukan pengujian dependensi menggunakan uji *Bartlet of Sphericity* menggunakan Persamaan (4).
3. Melakukan pengujian stasionaritas variansi menggunakan tranformasi *Box-Cox* dari Persamaan (1) dan Stasioeritas dalam rata-rata menggunakan uji ADF dari Persamaan (2).
4. Melakukan penaksiran model VAR menggunakan OLS dari Persamaan (6).
5. Melakukan pemilihan lag optimum model VAR menggunakan uji LRT dari Persamaan (7).
6. Melakukan pengujian stabilitas model VAR menggunakan Persamaan (8).

7. Melakukan pengujian kointegrasi menggunakan uji Johansen menggunakan Persamaan (9).
8. Melakukan penaksiran model VECM menggunakan OLS dari Persamaan (6).
9. Melakukan pengujian diagnostik model VECM menggunakan uji multivariat Ljung-Box dari Persamaan (11) dan PPCC dari Persamaan (12).
10. Melakukan prediksi dan peramalan VECM menggunakan model VECM pada Persamaan (10).
11. Melakukan perhitungan akurasi peramalan menggunakan MAPE dari Persamaan (13).

**4. Hasil dan Pembahasan**

**4.1. Statistika Deskriptif**

Analisis statistika deskriptif dilakukan menggunakan grafik runtun waktu pada ketiga variabel yaitu Curah Hujan, Suhu Udara Minimum, Kelembapan Udara Maksimum. Grafik runtun dari ketiga variabel dapat dilihat pada Gambar 1.



**Gambar 1.** Grafik Runtun Waktu ketiga variabel.

Berdasarkan Gambar 1a terlihat bahwa variabel curah hujan sepanjang periode Januari 2015 - Desember 2023 menunjukkan adanya tren naik atau turun yang cukup kecil karena tidak ada peningkatan atau penurunan curah hujan yang signifikan selama periode itu yang ditunjukkan oleh garis putus-putus. Selain itu fluktuasi curah hujan juga secara visual tidak konstan sepanjang periode tersebut.

Berdasarkan Gambar 1b dapat dilihat bahwa grafik runtun waktu dari variabel suhu udara minimum tidak stasioner dalam rata-rata karena terdapat beberapa tren pada variabel tersebut seperti tren naik dari tahun 2015 sampai tahun 2018 yang ditunjukkan oleh garis putus-putus. Besarnya fluktuasi secara visual cukup stabil dari waktu ke waktu.

Berdasarkan Gambar 1c dapat dilihat bahwa grafik runtun waktu dari variabel kelembapan udara maksimum tidak stasioner dalam rata-rata karena terdapat beberapa tren contohnya tren turun dari tahun 2015 sampai tahun 2016 yang ditunjukkan oleh garis putus-putus. Fluktuasi pada grafik runtun waktu kelihatan konsisten karena naik turunnya data yang konsisten sepanjang Januari 2015 - Desember 2023.

**4.2. Pengujian Dependensi**

Korelasi antar variabel dapat diuji menggunakan uji *Bartlett Test of Sphericity* dengan hipotesis nol ( $H_0$ ) tidak terdapat korelasi diantara variabel. Hasil pengujian ini dapat dilihat pada Tabel 2:

**Tabel 2.** Hasil Pengujian Dependensi

Statistik Hitung	Nilai Kritis	Kesimpulan
47,328	7,815	$H_0$ ditolak

Berdasarkan Tabel 2, dapat dilihat bahwa  $H_0$  ditolak yang berarti terdapat hubungan yang signifikan diantara variabel. Hal ini mengindikasikan bahwa terdapat korelasi yang cukup pada data untuk dilakukan analisis lanjutan pada keterkaitan antar variabel

**4.3. Pengatiasian Stasioneritas**

Pengatiasian stasioneritas akan dilakukan dalam dua cara yaitu variansi dan rata-rata secara berurutan. Stasioneritas yang pertama akan diatasi adalah stasioneritas dalam variansi menggunakan transformasi *Box-Cox* dan hasilnya dapat dilihat pada Persamaan (14):

$$\begin{aligned}
 y_{1,t} &= z_{1,t}^{0,288} \\
 y_{2,t} &= z_{2,t}^{4,539} \\
 y_{3,t} &= z_{3,t}
 \end{aligned}
 \tag{14}$$

Pada Persamaan (14),  $y_{i,t}$  merupakan hasil transformasi  $z_{i,t}$  menggunakan nilai parameter Box-Cox  $\lambda_i = \{0,288 ; 4,539 ; 1\}$  untuk  $i = 1,2,3$  dari Persamaan (1). Setelah dilakukan transformasi *Box-Cox*, maka akan dilakukan pengujian ADF pada data transformasi untuk mengetahui apakah data transformasi tersebut sudah stasioner dalam rata-rata atau belum. Hasil pengujian ADF dapat dilihat pada Tabel 3.

**Tabel 3.** Pengujian ADF pada data awal

Variabel	Statistik Hitung	p-value
$y_{1,t}$	-0,610	0,418
$y_{2,t}$	-0,331	0,507
$y_{3,t}$	0,003	0,614

Berdasarkan Tabel 3 dapat dilihat bahwa semua variabel tidak ada yang stasioner dikarenakan semua nilai *p-value* lebih besar dari 0,05. Hal ini berarti  $H_0$  yang menyatakan bahwa masing-masing variabel tidak stasioner gagal ditolak. Agar ketiga variabel tersebut menjadi stasioner, maka perlu dilakukan *differencing* menggunakan Persamaan (3). Setelah dilakukan *differencing* maka hasil pengujian ADF adalah sebagai berikut:

**Tabel 4.** Pengujian ADF pada data *differencing*

Variabel	Statistik Hitung	p-value
$\Delta y_{1,t}$	-9,946	0,01
$\Delta y_{2,t}$	-8,964	0,01
$\Delta y_{3,t}$	-10,412	0,01

Berdasarkan Tabel 4 dapat dilihat bahwa semua *p-value* memiliki nilai kurang dari 0,05 sehingga berdasarkan  $H_0$ , semua variabel dalam *differencing* sudah stasioner dalam rata-rata.

**4.4. Pemodelan Vector Autoregressive**

Pemilihan *lag* optimum pada VAR akan dilakukan setelah dilakukan penaksiran model VAR dari *lag* 1-10 menggunakan Persamaan (5). Metode yang digunakan untuk pemilihan *lag* optimum adalah LRT menggunakan Persamaan (8) yang hasilnya dapat dilihat pada Tabel 5.

**Tabel 5.** Pemilihan Lag Optimum

<i>h</i>	<i>M(p)</i>	<i>p - value</i>	<i>h</i>	<i>M(p)</i>	<i>p - value</i>
1	34,003	0,0001	6	6,028	0,7371
2	15,336	0,0821	7	4,423	0,8814
3	22,857	0,0065	8	17,385	0,0430
4	13,102	0,1580	9	4,738	0,8565
5	11,492	0,2434	10	8,192	0,5142

Berdasarkan Tabel 5 dapat dilihat bahwa nilai *lag* optimum yang terpilih adalah 2 dikarenakan dari beberapa *lag* yang tidak signifikan, *lag* 2 merupakan *lag* yang terkecil sehingga dapat disimpulkan model VAR yang terpilih adalah VAR(2).

**4.5. Pengujian Stabilitas**

Pengujian stabilitas dilakukan pada model VAR dengan *lag* optimum terpilih yang dalam hal ini adalah model VAR(2) dan dilakukan menggunakan Persamaan (6). Adapun nilai akar-akar stabilitas model VAR pada *lag* optimum yaitu VAR(2) dapat dilihat pada Tabel 6.

**Tabel 6.** Akar-Akar Stabilitas model VAR(2)

Akar	Nilai	Akar	Nilai
$K_1$	0,511	$K_4$	0,511
$K_2$	0,481	$K_5$	0,481
$K_3$	0,319	$K_6$	0,025

Berdasarkan Tabel 6, dapat dilihat bahwa semua nilai akar-akar stabilitas model ( $K_i$ ) VAR(2) memiliki nilai yang kurang dari 1, maka model VAR(2) sudah stabil menurut Wei (2019).

**4.6. Pengujian Kointegrasi**

Berdasarkan pengujian ADF sebelumnya dapat disimpulkan semua variabel tidak stasioner, maka pengujian kointegrasi dapat dilanjutkan yang hasilnya dapat dilihat pada Tabel 7.

**Tabel 7.** Hasil Pengujian Kointegrasi

$H_0$	Statistik Hitung	Nilai Kritis	Keputusan
$r_0 = 0$	65,78	31,52	$H_0$ ditolak
$r_0 = 1$	27,54	17,95	$H_0$ ditolak

$r_0 = 2$	6,04	8,18	$H_0$ gagal ditolak
-----------	------	------	---------------------

Berdasarkan Tabel, dapat dilihat bahwa diperoleh dua vektor kointegrasi karena  $H_0$  gagal ditolak pada  $r_0 = 2$ .

**4.7. Pemodelan VECM**

Model VECM yang terbentuk berdasarkan pemilihan lag optimum adalah VECM(1) karena VAR(2 – 1)=VECM(1). Hasil Penaksiran VECM(1) dapat dilihat pada Persamaan (15):

$$\begin{aligned}
 \Delta y_{1,t} &= -1,690 - 0,7521 ECT_{1,t} + 2,377 \times 10^{-4} ECT_{2,t} - 0,7632 \Delta y_{1,t-1} \\
 &\quad + 6,289 \times 10^{-7} \Delta y_{2,t-1} + 0,0524 \Delta y_{3,t-1} \\
 \Delta y_{2,t} &= 3,376 \times 10^6 - 4,314 \times 10^4 ECT_{1,t} - 2,516 \times 10^{-1} ECT_{2,t} - 6,053 \times 10^3 \Delta y_{1,t-1} \\
 &\quad - 2,260 \times 10^{-1} \Delta y_{2,t-1} - 1,033 \times 10^4 \Delta y_{3,t-1} \\
 \Delta y_{3,t} &= 2,580 \times 10^1 - 8,264 \times 10^{-1} ECT_{1,t} - 1,852 \times 10^{-6} ECT_{2,t} - 2,507 \times 10^{-2} \Delta y_{1,t-1} \\
 &\quad - 1,415 \times 10^{-6} \Delta y_{2,t-1} - 3,952 \times 10^{-1} \Delta y_{3,t-1}
 \end{aligned} \tag{15}$$

dengan persamaan  $ECT_{1,t}$  dan  $ECT_{2,t}$  adalah sebagai berikut:

$$\begin{aligned}
 ECT_{1,t} &= y_{1,t} - 0,003 y_{3,t} \\
 ECT_{2,t} &= y_{1,t} - 1,211 \times 10^5 y_{3,t}.
 \end{aligned}$$

Pada Persamaan (7) dapat dilihat bahwa koefisien curah hujan periode sebelumnya ( $\Delta y_{1,t-1}$ ) pada persamaan curah hujan ( $\Delta y_{1,t}$ ) sebesar 0,763 menunjukkan bahwa sebanyak 76,3% variasi curah hujan periode sebelumnya memengaruhi Curah Hujan periode sekarang. Selain itu, koefisien kelembapan periode sebelumnya ( $\Delta y_{3,t-1}$ ) pada persamaan Curah Hujan ( $\Delta y_{1,t}$ ) sebesar 0,0524 menunjukkan bahwa sebanyak 5,24% variasi Kelembapan Maksimum periode sebelumnya memengaruhi Curah Hujan periode sekarang. Selanjutnya koefisien Kelembapan Maksimum periode sebelumnya ( $\Delta y_{3,t-1}$ ) pada Persamaan Kelembapan Maksimum ( $\Delta y_{3,t}$ ) sebesar 0,395 menunjukkan bahwa sebanyak 39,5% variasi pada Kelembapan Maksimum periode sekarang ( $\Delta y_{3,t}$ ) dijelaskan oleh Kelembapan Maksimum periode sebelumnya. Koefisien  $ECT_{1,t}$  sebesar 0,752 pada persamaan Curah Hujan ( $\Delta y_{1,t}$ ) menunjukkan bahwa 75,2% variasi pada Curah Hujan dijelaskan oleh kombinasi linier  $ECT_{1,t}$  dari Persamaan (15). Sama halnya untuk koefisien  $ECT_{2,t}$  pada Persamaan Curah Hujan sebesar  $2,377 \times 10^{-4}$  yang menunjukkan bahwa 0,002% variasi pada Curah Hujan dijelaskan oleh kombinasi linier  $ECT_{2,t}$  dari Persamaan (15). Begitu seterusnya untuk semua koefisien di Persamaan (15).

**4.8. Pengujian Diagnostik**

Pengujian independensi residual dari hasil penaksiran model VECM(1) untuk mengetahui apabila residual saling bebas atau tidak. Pengujian akan dilakukan dengan uji multivariat independensi residual yang hasilnya dapat dilihat pada Tabel 8.

**Tabel 8.** Statistik hitung dan p-value Pengujian Multivariat Ljung-Box.

lag	Q(m)	derajat bebas	p-value	lag	Q(m)	derajat bebas	p-value
6	30,562	54	1	48	356,894	432	1
12	74,987	108	0,99	54	402,364	486	1
18	125,784	162	0,98	60	472,963	540	0,98
24	174,493	216	0,98	66	540,16	594	0,94
30	227,135	270	0,97	72	603,835	648	0,89
36	277,828	324	0,97	78	666,484	702	0,83
42	318,88	378	0,99	84	729,53	756	0,75

Berdasarkan Tabel 8, dapat dilihat bahwa semua nilai p-value dari pengujian multivariat independensi residual lebih besar dari 0,05 sehingga dapat disimpulkan tidak terdapat otokorelasi residual pada VECM(1). Setelah dilakukan pengujian independensi residual, maka akan dilakukan pengujian normalitas residual menggunakan PPCC. Berikut adalah pengujian normalitas residual menggunakan PPCC yang hasilnya dapat dilihat pada Tabel 9:

**Tabel 9.** Hasil Pengujian Normalitas Residual

Statistik Hitung	Nilai Kritis	Keputusan
0,993	0,982	$H_0$ gagal ditolak

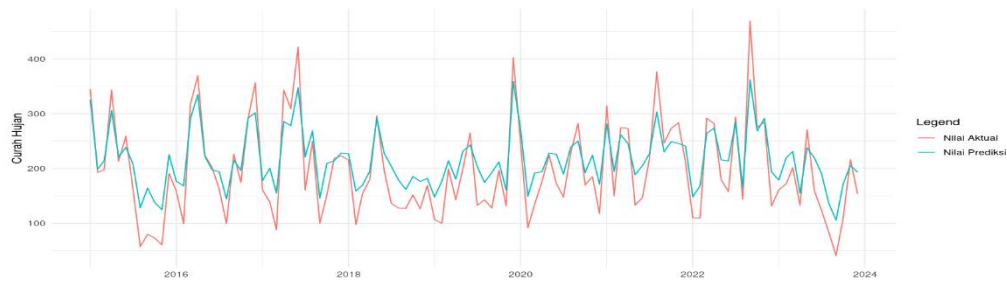
Berdasarkan Tabel 9 dapat dilihat bahwa  $H_0$  gagal ditolak yang berarti residual VECM(1) sudah berdistribusi normal.

**4.9. Prediksi dan Peramalan**

Prediksi dilakukan dengan memperkirakan data curah hujan bulanan 2015-2023 menggunakan

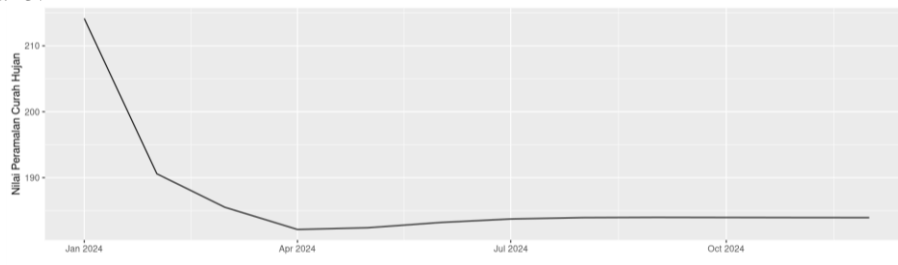


VECM(1) dari Persamaan (15) yang grafiknya dapat dilihat pada Gambar 2.



**Gambar 2.** Grafik Prediksi VECM(1)

Berdasarkan Gambar 2 dapat dilihat bahwa pola data pada grafik hasil prediksi dari variabel curah hujan di Kota Samarinda pada Januari 2015 - Desember 2023 dengan menggunakan model VECM(1) hampir mengikuti pola data aktual. Setelah dilakukan prediksi model, maka akan dilakukan perhitungan keakuratan model menggunakan MAPE dan diperoleh MAPE sebesar 32,369% yang termasuk peramalan cukup baik menurut Tabel 1. Setelah dilakukan prediksi, maka akan dilakukan peramalan untuk memperkirakan data curah hujan selama 12 bulan pada tahun 2024 menggunakan VECM(1) dari Persamaan (15) yang grafiknya dapat dilihat pada Gambar 3.



**Gambar 3.** Grafik Peramalan VECM(1)

Berdasarkan grafik peramalan curah hujan pada Gambar 3, dapat dilihat untuk variabel curah hujan secara visual menunjukkan adanya tren penurunan curah hujan dari Januari sampai April 2024 dan setelahnya menjadi mendatar yang berarti tidak ada perubahan pada prediksi curah hujan.

## 5. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan Tingkat akurasi hasil peramalan dari VECM pada data curah hujan di Kota Samarinda berdasarkan nilai MAPE adalah sebesar 32,369% yang termasuk ke dalam peramalan yang cukup baik. Selain itu, Hasil peramalan curah hujan di Kota Samarinda menunjukkan tren penurunan dari periode Januari 2024 - April 2024 dan setelah itu membentuk pola horizontal sampai Desember 2024. Penelitian berikutnya dapat mengkaji Panel VECM yaitu metode VECM yang diterapkan pada data panel.

## 6. Daftar Pustaka

- Das, D., & Srinivasan, R. (2019). Variation of temperature and rainfall in India. *International Journal of Advances in Engineering & Technology*, 6(4), 1803.
- Das, P. (2019). *Econometrics in Theory and Practice: Analysis of Cross Section, Time Series and Panel Data with Stata 15.1*. Springer Nature Singapore. <https://books.google.co.id/books?id=rK6tDwAAQBAJ>
- Denis, D. J. (2021). *Applied Univariate, Bivariate, and Multivariate Statistics: Understanding Statistics for Social and Natural Scientists, With Applications in SPSS and R*. Wiley. <https://books.google.co.id/books?id=zIwIEAAAQBAJ>
- Huang, C., & Petukhina, A. (2022). *Applied Time Series Analysis and Forecasting with Python*. Springer International Publishing. <https://books.google.co.id/books?id=TfaVEAAAQBAJ>
- Juselius, K. (2019). *The Cointegrated VAR Model: Methodology and Applications*. OUP Oxford. <https://books.google.co.id/books?id=N6LmgE8v1IEC>
- Kusdarwati, H., Effendi, U., & Handoyo, S. (2022). *Analisis Deret Waktu Univariat Linier: Teori dan Terapannya dengan Rstudio*. Universitas Brawijaya Press. <https://books.google.co.id/books?id=XGimEAAAQBAJ>
- Matulesy, E. R. (2023). The Application of VECM (Vector Error Correction Model) for Rainfall Modeling in Manokwari Regency. *Indonesian Journal of Multidisciplinary Science*, 2(6), 2712–2720. <https://doi.org/10.55324/ijoms.v2i6.483>

- Mills, T. C. (2019). *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*. Elsevier Science. <https://books.google.co.id/books?id=1daEDwAAQBAJ>
- Onatski, A., & Wang, C. (2019). Extreme canonical correlations and high-dimensional cointegration analysis. *Journal of Econometrics*, 212(1), 307–322. <https://doi.org/10.1016/j.jeconom.2019.04.032>
- Rafferty, G. (2023). *Forecasting Time Series Data with Prophet: Build, improve, and optimize time series forecasting models using Meta's advanced forecasting tool*. Packt Publishing. <https://books.google.co.id/books?id=HVG3EAAAQBAJ>
- Reinsel, G. C., Velu, R. P., & Chen, K. (2022). *Multivariate Reduced-Rank Regression: Theory, Methods and Applications*. Springer New York. <https://books.google.co.id/books?id=kQafEAAAQBAJ>
- Shin, Y. (2019). *Time Series Analysis in the Social Sciences: The Fundamentals*. University of California Press. <https://books.google.co.id/books?id=z7QwDwAAQBAJ>
- Wei, W. W. S. (2005). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). Pearson Education.
- Wei, W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. Wiley. <https://books.google.co.id/books?id=9naCDwAAQBAJ>
- Winarno, S., Usman, M., Warsono, Kurniasari, D., & Widiarti. (2021). Application of Vector Error Correction Model (VECM) and Impulse Response Function for Daily Stock Prices. *Journal of Physics: Conference Series*, 1751(1). <https://doi.org/10.1088/1742-6596/1751/1/012016>
- Yollanda, M., Harjupa, W., Devianto, D., Andarini, D. F., Nauval, F., Saufina, E., Purwaningsih, A., Harjupa, W., Trismidianto, Harjana, T., Risyanto, Rahmatia, F., Pratama, R., & Satiadi, D. (2023). Prediction of CENS, MJO, and Extreme Rainfall Events in Indonesia Using the VECM Model. In *Springer Proceedings in Physics* (Vol. 290). Springer Nature Singapore. [https://doi.org/10.1007/978-981-19-9768-6\\_35](https://doi.org/10.1007/978-981-19-9768-6_35)