JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 5 No. 2 (2024)
https://doi.org/10.35877/454RI.jinav3266

Evaluating ARIMA Models for Short-Term Rainfall Forecasting in Polewali Mandar Regency

Ansari Saleh Ahmar[a,*] & Ali Mokhtar[b]

[a]*Department of Statistics, Universitas Negeri Makassar, Makassar, 90223, Indonesia*
[b]*Professional Engineer Program, University of Muhammadiyah Malang, Jl. Raya Tlogomas 246, Malang, 65144, Indonesia*

Abstract

This study aims to forecast rainfall in Polewali Mandar Regency using the ARIMA model. This is a quantitative study that uses secondary data, specifically monthly rainfall data (in mm) from January 2008 to December 2020, obtained from the NERC EDS Centre for Environmental Data Analysis. Two ARIMA models were tested: ARIMA(0,1,1)(0,1,1)[12] and ARIMA(1,1,1)(0,1,1)[12], with model selection based on the Akaike Information Criterion (AIC), which balances model fit and complexity. The AIC calculation revealed that the ARIMA(1,1,1)(0,1,1)[12] model had a lower AIC value (1677.33) compared to the ARIMA(0,1,1)(0,1,1)[12] model (1678.16), making ARIMA(1,1,1)(0,1,1)[12] the preferred model. Using this model, the forecasted rainfall for the next five months is as follows: January 2021: 279.8745 mm, February 2021: 238.2206 mm, March 2021: 237.1745 mm, April 2021: 349.3206 mm, and May 2021: 336.0976 mm. These forecasts provide valuable information for water resource management, agricultural irrigation planning, and disaster mitigation related to rainfall. The study emphasizes the importance of selecting the appropriate model to improve forecasting accuracy.

Keywords: Rainfall forecasting, ARIMA model, AIC, water resources, agriculture, disaster mitigation.

Received: 17 July 2024 Revised: 21 November 2024 Accepted: 16 December 2024

1. Introduction

Rainfall forecasting is a critical component in various sectors, including water resource management, agriculture, and disaster mitigation. Its significance is underscored by the necessity for accurate predictions to guide irrigation practices and optimize crop yields amidst varying seasonal conditions (Wu et al., 2023). Furthermore, accurate forecasting is vital for effective dam management, enabling operators to mitigate risks associated with both droughts and floods (Nguyen et al., 2019). Rainfall forecasts also play an essential role in land-use planning by identifying flood-prone areas and those susceptible to landslides, which is especially pertinent given the challenges posed by climate change (Zhu et al., 2021).

The traditional methods of rainfall forecasting typically utilize statistical analyses of historical data. While these approaches have proven effective in understanding general rainfall trends, they often fall short in addressing the dynamic changes induced by climate fluctuations (Huang et al., 2020). Advanced methods, such as remote sensing and numerical modeling, are increasingly leveraged to enhance forecasting accuracy. These approaches employ satellite data and sophisticated atmospheric simulations, thus enabling predictions that are both precise and timely (Zhu et al., 2018; Chen et al., 2024). Research indicates that the integration of remote sensing data into hydrological frameworks significantly improves forecast reliability and supports real-time decision-making (Perera et al., 2022; Sun et al., 2019).

With climate change yielding more erratic weather patterns and more intense rainfall events, the challenges of rainfall forecasting have escalated. Historical data may no longer adequately represent future precipitation trends, necessitating the adoption of cutting-edge techniques that incorporate Big Data and artificial intelligence (AI). These methods allow for the analysis of large datasets to discover complex relationships within the climate system, enhancing forecasting

* Corresponding author.
E-mail address: ansarisaleh@unm.ac.id

accuracy (Bhattacharya et al., 2019; , Aryastana et al., 2024). For instance, machine learning algorithms have demonstrated their efficacy in improving prediction accuracy by learning from extensive datasets of past rainfall events (Mardyansyah et al., 2022; , Küllahcı & Altunkaynak, 2024). Additionally, studies have explored the incorporation of remote sensing data into rainfall prediction systems, which offers a promising avenue for enhancing forecast quality amid changing climatic conditions (Viana et al., 2021; , Fletcher et al., 2019; Sudiatmika & Putra, 2024).

Integrating these innovative forecasting methods is not merely a response to evolving climate dynamics; it is pivotal for sustainable water resource management (Ahmar et. al. 2024). The advancements in remote sensing and AI-driven predictive models provide the necessary tools to anticipate and mitigate the adverse effects of climate variability on water systems (Ideki & Weli, 2019; , Sun et al., 2019; Nurman, Nusrang, & Sudarmin, 2022). Consequently, the current landscape of rainfall forecasting must adapt to leverage these technologies, ensuring that predictions remain robust and relevant in the face of increasing climate variability.

The SutteARIMA model represents a significant advancement in short-term rainfall forecasting, specifically designed to enhance predictive accuracy by integrating both seasonal and non-seasonal data components. Unlike conventional ARIMA models, which predominantly focus on linear relationships in historical data, SutteARIMA excels in capturing temporal variations, including seasonality and trends that are particularly relevant in environmental datasets such as rainfall Ahmar et al. (2023). Originally applied in economic contexts, its adaptation to meteorological forecasting showcases its versatility and effectiveness under varying climatic conditions. Research indicates that SutteARIMA demonstrates improved reliability specifically in regions experiencing high rainfall variability, thereby emerging as a crucial tool for modern meteorological applications, including water resource management and agricultural planning (Amini et al., 2023).

Incorporating cutting-edge technologies like Big Data and artificial intelligence (AI) into rainfall forecasting not only enhances predictive accuracy but also significantly bolsters preparedness and responsiveness to natural disasters. High-resolution rainfall forecasts generated by integrating these methodologies enable the establishment of early warning systems. Such systems empower vulnerable communities to take preemptive actions, thereby mitigating potential loss of life and reducing economic damage during adverse weather events (Gupta et al., 2024; Taillardat et al., 2019). For instance, deep learning techniques are employed alongside SutteARIMA to analyze extensive datasets, revealing complex nonlinear relationships that traditional methods might overlook (Mukaddim et al., 2024).

Furthermore, fostering interdisciplinary collaboration is critical in refining forecasting models that address the multifaceted impacts of climate change. Such collaborative efforts can integrate insights from meteorology, mathematics, and computer science to create scientifically robust and practically applicable forecasting tools (Nair et al., 2018; Swain et al., 2018). These innovations ensure that advancements in forecasting technology translate into accessible strategies for local governments and communities most affected by rainfall variability (Neal et al., 2022).

The ongoing integration of Big Data and AI technologies into rainfall forecasting represents a transformative approach, processing vast amounts of data from satellite imagery, weather stations, and other sensors to uncover intricate patterns and trends. Advanced machine learning models, including SutteARIMA, continue to evolve, capturing seasonal trends and cyclical variations with heightened precision (Melesse & Delele, 2024; Yang et al., 2018). By harnessing the capabilities of these sophisticated technologies, forecasters can enhance their predictive skill and adapt to the challenges posed by climate change, ultimately leading to more resilient and sustainable communities (Slater et al., 2022; , Basak, 2020).

The evolution of rainfall forecasting—from traditional models like ARIMA to advanced frameworks such as the SutteARIMA model combined with Big Data and AI—addresses growing challenges in water resource management and disaster preparedness. This technological progression not only enhances forecast precision but also supports proactive community measures in the face of climate variability, underscoring the vital role of effective forecasting methodologies in ensuring sustainable development (Khan et al., 2023).

2. Methods

This quantitative study utilizes secondary data obtained from the NERC Environmental Data Service (EDS) Centre for Environmental Data Analysis, accessible at https://catalogue.ceda.ac.uk/uuid/5fda109ab71947b6b7724077bf7eb753. The dataset includes monthly rainfall measurements (in millimeters) for Polewali Mandar Regency, covering the period from January 2008 to December 2020, totaling 156 months. This comprehensive dataset enables an in-depth analysis

of rainfall patterns and trends over an extended period, providing valuable insights into seasonal and annual fluctuations in the region. This information is crucial for water resource management, agricultural planning, and disaster preparedness, particularly in areas susceptible to climate variability.

To model and forecast rainfall accurately, the study employs the AutoRegressive Integrated Moving Average (ARIMA) model as its primary forecasting method. ARIMA is widely recognized for its effectiveness in time series analysis, particularly for data characterized by seasonality and temporal dependencies, making it well-suited for rainfall data. By analyzing historical rainfall patterns, the ARIMA model can identify underlying trends and seasonal cycles, allowing for the projection of future rainfall levels. This modeling approach is beneficial for predicting rainfall variability, enabling stakeholders to anticipate and mitigate the potential impacts of extreme weather conditions, such as floods or droughts, which may affect the local economy and public safety.

The selection of the ARIMA model in this study reflects its robustness in handling non-stationary data, a common characteristic of climate datasets. Additionally, ARIMA's ability to adjust for both seasonal and non-seasonal components allows for more precise and accurate forecasts. Thus, this research not only contributes to the understanding of rainfall patterns in Polewali Mandar Regency but also demonstrates the practical application of statistical modeling techniques in environmental and climate studies.

3. Result and Discussion

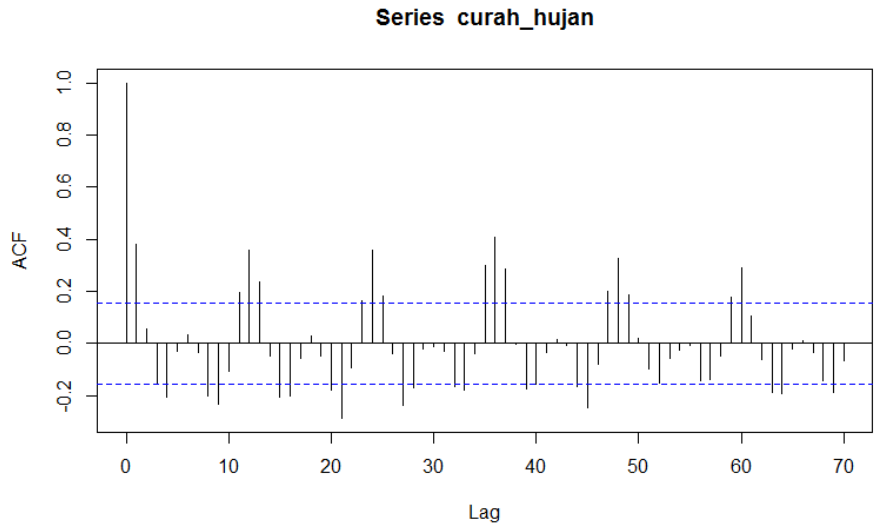
Figure 1 illustrates climate data for Polewali, West Sulawesi, Indonesia, covering the period from 2008 to 2020, located at coordinates 3.4315°S, 119.376°E, with an elevation of 54 meters above sea level. Based on the Köppen classification, Polewali is categorized under the tropical rainforest climate (Af). The displayed data includes monthly average temperature (in degrees Celsius) and monthly rainfall (in millimeters), presented in a matrix format. Temperature is shown with a brown gradient, where darker shades indicate higher temperatures, while rainfall is represented with a blue-green scale, where darker shades signify higher rainfall levels. Several months are marked in gray, indicating periods where data was unavailable.

Overall, the temperature in Polewali remains relatively stable throughout the year, ranging from 24.7°C to 26.0°C, reflecting the steady characteristics of a tropical climate. In contrast, rainfall exhibits significant seasonal variation, with peak rainfall occurring in certain months, such as May 2017, which recorded 500.3 mm, illustrated by dark blue shading. This pattern indicates a rainy season with high rainfall and a dry season with comparatively lower rainfall. This information is essential for climate change analysis, water resource management, and agricultural planning in Polewali Mandar, particularly in addressing climate challenges that may influence rainfall and temperature patterns in the region.

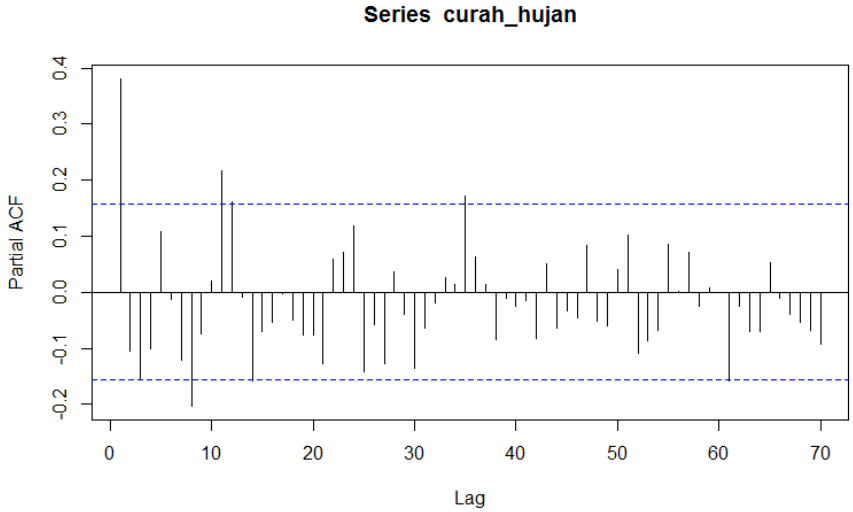
3.1. Plot the ACF and PACF

The initial step in data forecasting is the identification stage, which aims to determine whether the data is stationary. This is done by analyzing the time series plot, the Augmented Dickey-Fuller (ADF) Test, and the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. This stage involves plotting the data to observe its characteristics and patterns, as shown in Figure 2.

```
> adf.test(curah_hujan)
      Augmented Dickey-Fuller Test

data:  curah_hujan

Dickey-Fuller = -5.0083, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

> acf(curah_hujan, lag.max = 70)
```




```
> pacf(curah_hujan, lag.max = 70)
```



The time series plot in Figure 2 clearly indicates that the data is not stationary in terms of the mean, as the values show visible fluctuations and trends rather than oscillating around a constant average. This lack of stationarity in the mean suggests that the data contains underlying trends or periodic variations that may interfere with accurate forecasting if left unaddressed.

To achieve a stationary mean, a differencing process is applied. Differencing is a widely used transformation technique in time series analysis that involves subtracting each data point from the previous one, which effectively removes linear trends or seasonal patterns from the data. By stabilizing the mean through differencing, the transformed data becomes more suitable for time series models like ARIMA, which require stationarity in the mean to produce reliable forecasts.

In some cases, first-order differencing may suffice; however, if the data still exhibits trends, further differencing (such as seasonal differencing) might be necessary. Once stationarity is achieved, the data can be analyzed for autoregressive and moving average components, using the autocorrelation (ACF) and partial autocorrelation (PACF) plots to determine appropriate model parameters. This preparation ensures more accurate and meaningful forecasting results.

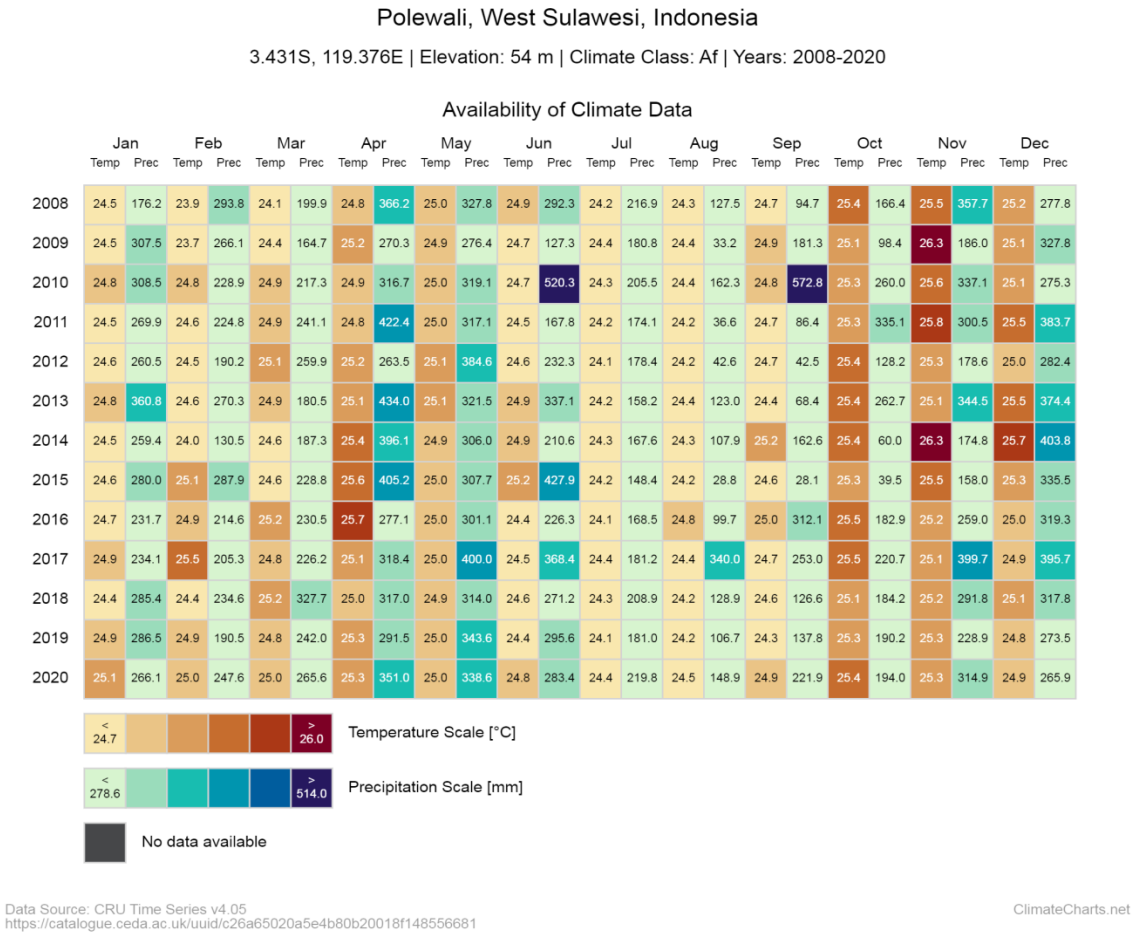


Figure 1. Rainfall Data in Polewali Mandar
Source: https://catalogue.ceda.ac.uk/uuid/5fda109ab71947b6b7724077bf7eb753

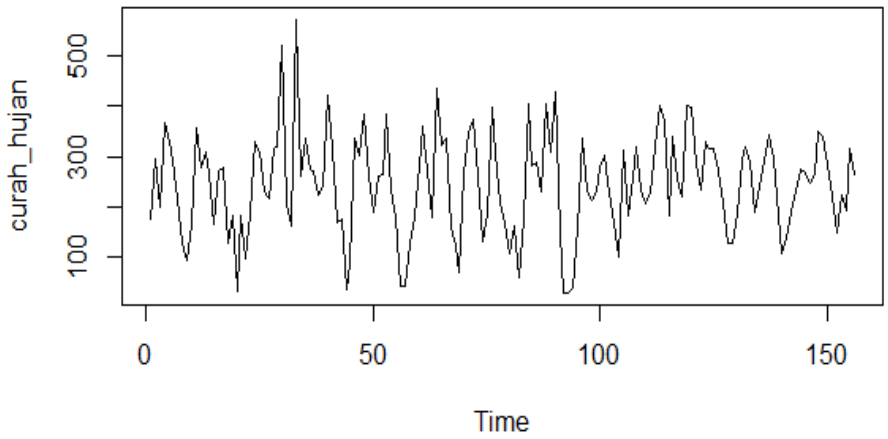
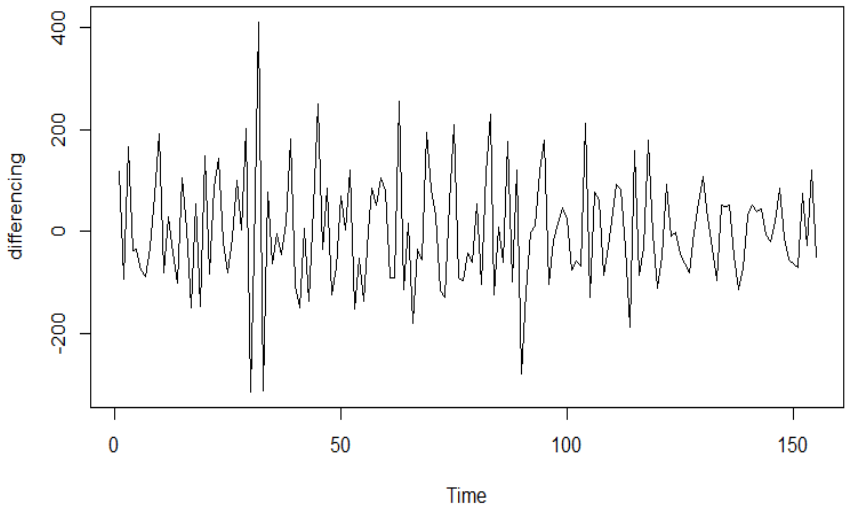


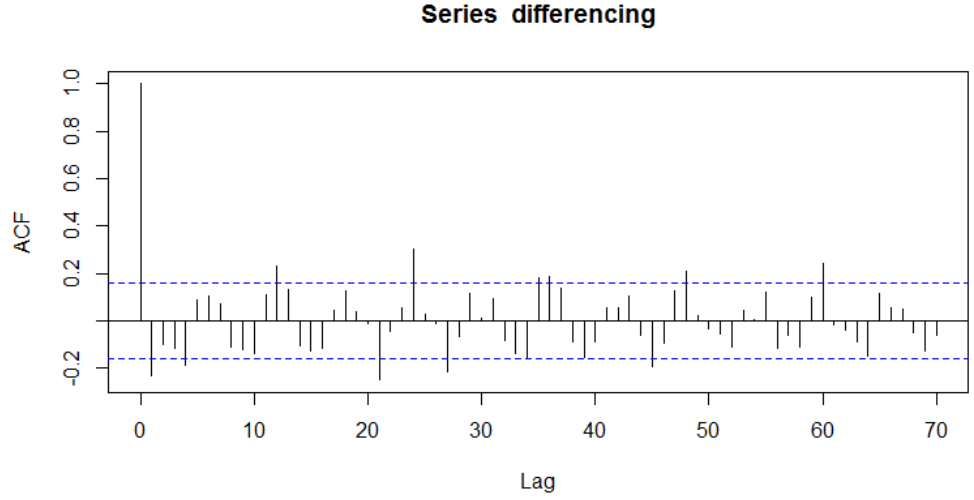
Figure 2. Time Series Plot of Rainfall Forecasting Data in Polewali Mandar

3.2. Differencing process for one non-seasonal (d = 1) of the rainfall data.

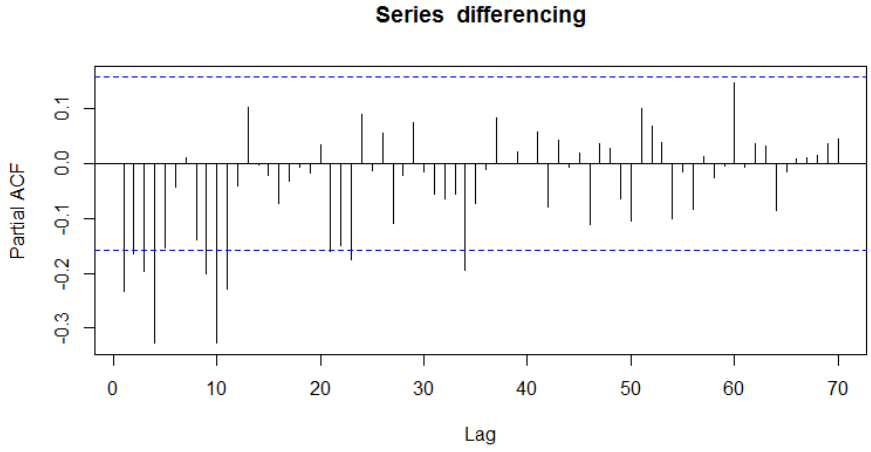
```
> differencing <- diff(curah_hujan, differences = 1)
> ts.plot(differencing)
```




```
> acf(differencing, lag.max = 70)
```




```
> pacf(differencing, lag.max = 70)
```



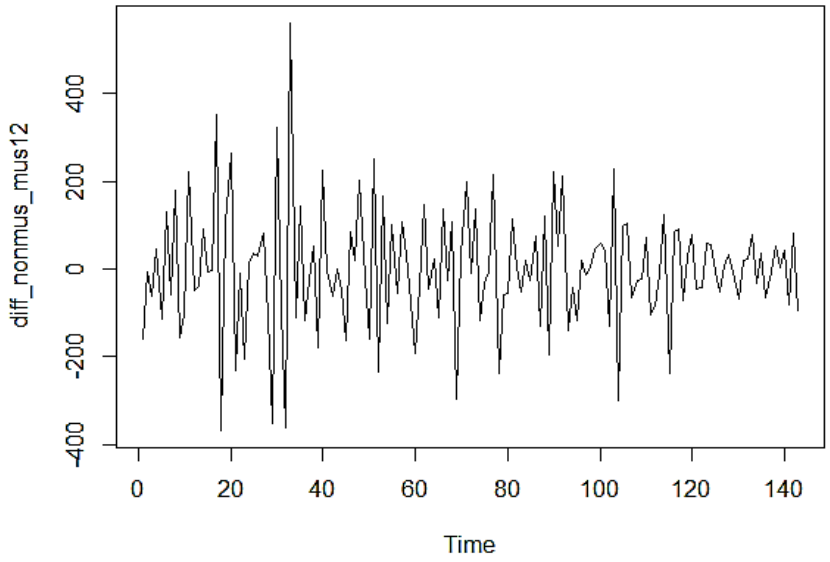
In the Partial Autocorrelation Function (PACF) plot reveals a slow decay at lags 12, 24, 36, and so on, indicating the presence of a seasonal pattern in the data. This pattern suggests that the data is not yet stationary in terms of its seasonal component, even after the initial differencing. The slow decay at these specific lag intervals points to recurring correlations every 12 periods, implying that the data exhibits seasonality on an annual basis.

To address this issue and achieve full stationarity, a seasonal differencing process is applied with a period of 12 (D = 1), effectively targeting the seasonal component in the data. Seasonal differencing involves subtracting each value from the corresponding value 12 periods earlier, thus removing the recurring seasonal patterns. By applying this transformation, the data is expected to stabilize further, eliminating the seasonal trends and making it more suitable for time series modeling, such as SARIMA, which requires both seasonal and non-seasonal stationarity.

Achieving stationarity through seasonal differencing ensures that the data meets the model assumptions, thereby enhancing the accuracy and reliability of any forecasts generated from the model. This preparation step is crucial in time series analysis, particularly when dealing with data that exhibits both trend and seasonal components.

3.3. Differencing process for one seasonal 12 (D = 1) from data that has been differenced for one non-seasonal.

```
> diff_nonmus_mus12 <- diff(differencing1, lag = 12)
> ts.plot(diff_nonmus_mus12)
```




```
> adf.test(diffnonmus_mus12)

        Augmented Dickey-Fuller Test

data:  diffnonmus_mus12
Dickey-Fuller = -7.0309, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

The Augmented Dickey-Fuller (ADF) test conducted on the data, which has been processed with both non-seasonal and seasonal differencing, produced a Dickey-Fuller statistic of -7.0309 with a lag order of 5 and a p-value of 0.01. As this p-value is below the significance threshold of 0.05, the alternative hypothesis is accepted, confirming that the data is stationary. This indicates that fluctuations in the dataset no longer exhibit strong trends or seasonal effects.

The time series plot of the seasonal differencing procedure, with a periodicity of 12 (D = 1), illustrates that the data is now stationary in terms of both mean and variance. This seasonal differencing process serves to remove recurring seasonal patterns, ensuring each data point aligns more consistently with a stable average, free from prominent trends or cycles. With stationarity achieved in both mean and variance, time series models such as ARIMA or SARIMA can be applied more effectively, as these models assume stationarity as a fundamental requirement.

Stationarity in the data is critical for generating accurate and stable forecasts. Models like ARIMA and SARIMA are specifically designed to capture autocorrelation in stationary data; if strong trends or seasonal patterns were present, the

model might yield biased or unreliable predictions. Consequently, the seasonal differencing process applied here is an essential step in preparing the data for modeling.

With the data now satisfying the stationarity requirement, the subsequent step involves determining the optimal model parameters, a process guided by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. Accurately identifying these parameters enables the development of a robust and reliable forecasting model, offering valuable insights into rainfall patterns and supporting informed decision-making in water resource management and disaster mitigation.

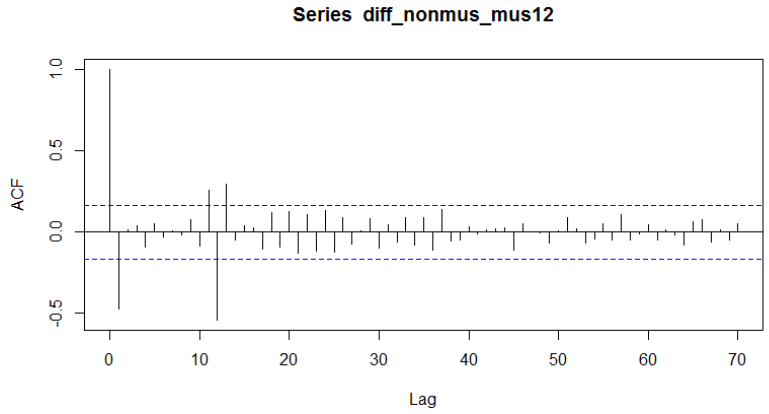
3.4. ARIMA Parameter Estimation

To estimate the parameters of the ARIMA model, the analysis begins by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the stationary data. This data has undergone both non-seasonal differencing (d = 1) and seasonal differencing (D = 1) to ensure that it meets the stationarity requirement. The ACF and PACF plots are essential tools for identifying the appropriate order of the ARIMA model, specifically the autoregressive (AR) and moving average (MA) components.

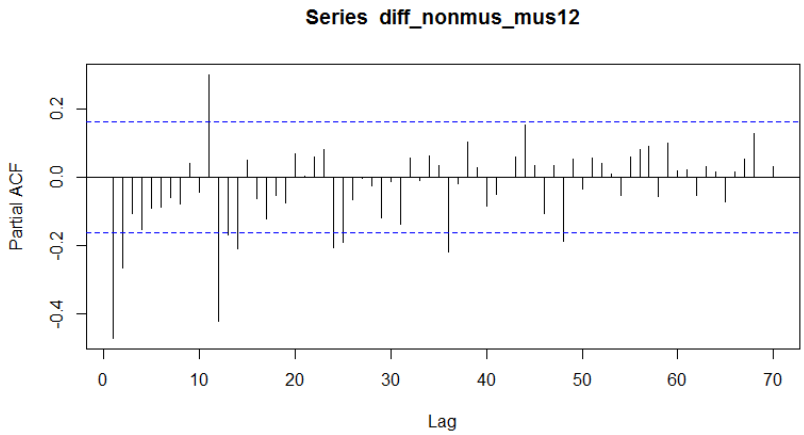
The ACF plot helps determine the lag at which the autocorrelation of the data becomes insignificant, indicating the order of the MA component. A slow decay in the ACF suggests the presence of a long-term dependence, while a rapid drop suggests a short-term dependence. The PACF plot, on the other hand, is used to identify the order of the AR component. Significant spikes at specific lags in the PACF indicate the presence of autoregressive terms at those lags.

By carefully analyzing these plots, one can estimate the optimal values for the AR (p), differencing (d), and MA (q) components, along with the seasonal parameters (P, D, Q) if applicable. This process is crucial in ensuring that the ARIMA model captures the underlying patterns in the data, leading to more accurate and reliable forecasts.

```
> acf(diff_nonmus_mus12, lag.max = 70)
```




```
> pacf(diff_nonmus_mus12, lag.max = 70)
```



Based on the ACF and PACF plots, the initial model options for the data are ARIMA(0,1,1)(0,1,1)[12] and ARIMA(1,1,1)(0,1,1)[12]. The ARIMA(0,1,1)(0,1,1)[12] model suggests no autoregressive (AR) component for the non-seasonal part but includes a seasonal moving average (MA) term, indicating a seasonal pattern with a period of 12. On the other hand, the ARIMA(1,1,1)(0,1,1)[12] model incorporates one non-seasonal autoregressive term (AR(1)) in addition to the seasonal components. These models are a good starting point for further refinement and validation.

3.5. Parameter estimation stage and diagnostic check of ARIMA model

3.5.1. Parameter estimation, white noise residual testing, and normally distributed residuals for the ARIMA(0,1,1)(0,1,1)[12] model.

```
> #Model ARIMA(0,1,1)(0,1,1) musiman = 12
> model1 <- arima(curah_hujan, order = c(0, 1, 1), seasonal = list(order = c(0,
1, 1), period = 12), method = "ML")
> summary(model1)

Call:
arima(x = curah_hujan, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1),
    period = 12), method = "ML")

Coefficients:
          ma1     sma1
      -0.7908  -1.0000
s.e.   0.0822   0.0873

sigma^2 estimated as 5615:  log likelihood = -836.08, aic = 1678.16

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE
Training set -1.357919 71.74437 50.76783 -16.64398 32.64694
                   MASE       ACF1
Training set 0.5886675 0.06226134

> #pemeriksaan diagnostik
> coeftest(model1)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ma1  -0.790765   0.082208  -9.619 < 2.2e-16 ***
sma1 -1.000000   0.087259 -11.460 < 2.2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Box.test(model1$residuals, type = "Ljung")

      Box-Ljung test

data:  model1$residuals
X-squared = 0.61643, df = 1, p-value = 0.4324

> shapiro.test(model1$residuals)

      Shapiro-Wilk normality test

data:  model1$residuals
W = 0.94489, p-value = 8.634e-06
```

```
> sf.test(model1$residuals)

      Shapiro-Francia normality test

data:  model1$residuals
W = 0.94106, p-value = 1.51e-05
```

The ARIMA(0,1,1)(0,1,1)[12] model was fitted to the rainfall data using the maximum likelihood estimation (MLE) method. The estimated coefficients for the non-seasonal moving average (MA) and seasonal moving average (SMA) components are as follows:

- MA(1) coefficient (ma1): -0.7908 with a standard error of 0.0822
- SMA(1) coefficient (sma1): -1.0000 with a standard error of 0.0873

a. Diagnostic Checking

The results from the z-test for the model coefficients show that both the MA(1) and SMA(1) coefficients are statistically significant, with p-values less than 2.2e-16, indicating that both components contribute significantly to explaining the data.

The Ljung-Box test for the residuals yields a p-value of 0.4324, which is greater than the significance level of 0.05. This indicates that the residuals are white noise, meaning there is no significant autocorrelation left in the residuals. The model has successfully captured the temporal dependencies in the data.

b. Interpretation

The ARIMA(0,1,1)(0,1,1)[12] model has successfully accounted for the autocorrelation in the data, with significant coefficients and residuals that show no significant autocorrelation. The Ljung-Box test confirms that the model has adequately fitted the data. The error metrics suggest that the model performs reasonably well, though further refinement could be considered, especially if improvements in accuracy are desired for future predictions.

3.5.2. Parameter estimation, white noise residual testing, and normally distributed residuals for the ARIMA $(1,1,1)(0,1,1)^{12}$ model.

```
> #Model ARIMA(1,1,1)(0,1,1), periode = 12
> model2 <- arima(curah_hujan, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12), method = "ML")
> summary(model2)

Call:
arima(x = curah_hujan, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12), method = "ML")

Coefficients:
          ar1      ma1     sma1
      0.2214  -1.0000  -1.0000
s.e.  0.0825   0.0614   0.0947

sigma^2 estimated as 5270: log likelihood = -834.67, aic = 1677.33

Training set error measures:
                    ME     RMSE      MAE       MPE
Training set -1.58521 69.50594 48.10014 -16.52771
                  MAPE     MASE         ACF1
Training set 31.41287 0.557735 -0.02436685

> #pemeriksaan diagnostik
> coeftest(model2)

z test of coefficients:

      Estimate Std. Error  z value  Pr(>|z|)
ar1   0.221421   0.082472   2.6848  0.007257 **
```

```
ma1  -0.999998   0.061433 -16.2778 < 2.2e-16 ***
sma1 -0.999997   0.094678 -10.5621 < 2.2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Box.test(model2$residuals, type = "Ljung")

      Box-Ljung test

data:  model2$residuals
X-squared = 0.094417, df = 1, p-value = 0.7586

> shapiro.test(model2$residuals)

      Shapiro-Wilk normality test

data:  model2$residuals
W = 0.92991, p-value = 6.359e-07

> sf.test(model2$residuals)

      Shapiro-Francia normality test

data:  model2$residuals
W = 0.92494, p-value = 1.553e-06
```

The ARIMA(1,1,1)(0,1,1)[12] model was fitted to the rainfall data (`curah_hujan`) using the maximum likelihood estimation (MLE) method. The estimated coefficients for the autoregressive (AR), moving average (MA), and seasonal moving average (SMA) components are as follows:

- AR(1) coefficient (ar1): 0.2214 with a standard error of 0.0825
- MA(1) coefficient (ma1): -1.0000 with a standard error of 0.0614
- SMA(1) coefficient (sma1): -1.0000 with a standard error of 0.0947

a. Diagnostic Checking

The results from the z-test for the model coefficients show that both the MA(1) and SMA(1) coefficients are highly statistically significant (p-values < 2.2e-16), while the AR(1) coefficient is also statistically significant with a p-value of 0.0073. This indicates that all components of the model are crucial in explaining the autocorrelation structure of the data.

The Ljung-Box test for residuals yields a p-value of 0.7586, which is greater than the significance level of 0.05. This suggests that the residuals are white noise, meaning no significant autocorrelation remains in the residuals. This indicates that the model has effectively captured the autocorrelation structure and no further model improvement is needed in this regard.

b. Interpretation

The ARIMA(1,1,1)(0,1,1)[12] model has successfully accounted for the temporal dependencies in the data, with statistically significant coefficients and no remaining autocorrelation in the residuals. The Ljung-Box test confirms the adequacy of the model in terms of autocorrelation. The error measures suggest that the model performs well overall, though further refinement could still be considered if more accurate forecasts are needed in the future.

In this study, the ARIMA(0,1,1)(0,1,1)[12] model yields an AIC value of 1678.16, while the ARIMA(1,1,1)(0,1,1)[12] model produces a slightly lower AIC of 1677.33. Although the difference in AIC between the two models is minimal, the model ARIMA(1,1,1)(0,1,1)[12] is deemed more efficient as it demonstrates a better fit to the data without introducing unnecessary complexity.

Therefore, ARIMA(1,1,1)(0,1,1)[12] is considered the more appropriate model based on its lower AIC value. While the AIC difference is marginal, a lower AIC indicates that the model has a better balance between data fit and the number of parameters used. The inclusion of an additional autoregressive parameter (AR(1)) in ARIMA(1,1,1)(0,1,1)[12],

compared to ARIMA(0,1,1)(0,1,1)[12], provides a slight improvement in model fit with a minimal increase in model complexity. Thus, based on the comparison of AIC values, the ARIMA(1,1,1)(0,1,1)[12] model is selected as the more suitable model for this dataset.

3.6. Forecasting Step

The forecasted monthly rainfall (in mm) for Polewali Mandar Regency for the next five months is presented in Table 1 and Figure 3.

Table 1. Forecasted Rainfall (mm) in Polewali Mandar Regency (Next 5 Months)

Months	Forecast
January 2021	279.8745
February 2021	238.2206
March 2021	237.1745
April 2021	349.3206
May 2021	336.0976

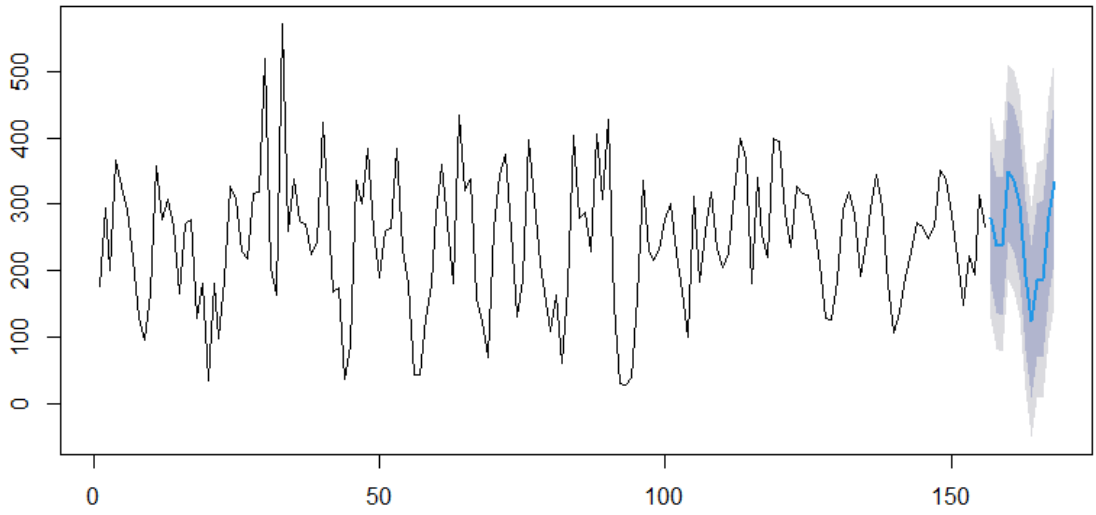


Figure 3. Rainfall Forecasting Results in Polewali Mandar (5 Future Periods)

4. Conclusion

In this study, two potential ARIMA models were considered for rainfall forecasting: ARIMA(0,1,1)(0,1,1)[12] and ARIMA(1,1,1)(0,1,1)[12]. After evaluating both models, the ARIMA(1,1,1)(0,1,1)[12] model was selected as the most suitable for predicting future rainfall values. This model was chosen based on its lower Akaike Information Criterion (AIC) value of 1677.33 compared to the ARIMA(0,1,1)(0,1,1)[12] model, which had an AIC value of 1678.16. The AIC is an essential metric for model selection, as it evaluates the trade-off between the goodness of fit and the complexity of the model. A lower AIC indicates a better model that balances model accuracy with simplicity. Although the difference in AIC values is small, the ARIMA(1,1,1)(0,1,1)[12] model provides a more efficient representation of the data without introducing unnecessary complexity.

The ARIMA(1,1,1)(0,1,1)[12] model incorporates one autoregressive (AR) term and one seasonal moving average (SMA) term, which allows it to capture both short-term fluctuations and seasonal patterns in the rainfall data. This model is particularly effective for time series data that exhibit both trend and seasonality, as is the case with rainfall data. The inclusion of the AR(1) term helps to capture the dependency of current values on past values, which improves

the model's ability to forecast future rainfall. The SMA(1) term accounts for the seasonal variation in rainfall, which is particularly important given the seasonal nature of rainfall patterns in the region.

Following the model selection, the forecasted rainfall values for the next five months were generated using the ARIMA(1,1,1)(0,1,1)[12] model. The forecasted values are as follows: 279.8745 mm (January 2021), 238.2206 mm (February 2021), 237.1745 mm (March 2021), 349.3206 mm (April 2021), and 336.0976 mm (May 2021).

These predictions are crucial for various applications, such as water resource management, agricultural planning, and disaster preparedness. The forecasted rainfall values provide an estimate of expected precipitation, which can help stakeholders plan for water storage, irrigation scheduling, and potential flood risks. For instance, the forecasted increase in rainfall during April and May suggests the potential for wetter conditions, which could have implications for flood risk management.

Moreover, while the ARIMA(1,1,1)(0,1,1)[12] model has proven to be effective for forecasting rainfall in the short term, it is important to recognize the inherent uncertainties in long-term weather predictions. As the forecast horizon extends, the accuracy of predictions typically decreases due to the unpredictable nature of climate systems. Thus, continuous monitoring and updating of the model with new data are essential to maintain forecast reliability. Furthermore, the use of additional forecasting methods, such as machine learning and Big Data analytics, could complement the ARIMA model by capturing more complex, nonlinear relationships in the data.

The last, the ARIMA(1,1,1)(0,1,1)[12] model provides a reliable and efficient approach to short-term rainfall forecasting. With the forecasted values for the next five months, stakeholders in agriculture, water resource management, and disaster preparedness can make more informed decisions. Nevertheless, future improvements to the model, along with integration of advanced techniques, will further enhance the accuracy and robustness of rainfall predictions.

References

Ahmar, A. S., Meliyana, S. M., Botto-Tobar, M., & Hidayat, R. (2024). The Comparison of Single and Double Exponential Smoothing Models in Predicting Passenger Car Registrations in Canada. *Daengku: Journal of Humanities and Social Sciences Innovation*, *4*(2), 367-371. https://doi.org/10.35877/454RI.daengku2639

Ahmar, A., Singh, P., Ruliana, R., Pandey, A., & Gupta, S. (2023). Comparison of arima, suttearima, and holt-winters, and nnar models to predict food grain in india. *Forecasting*, 5(1), 138-152. https://doi.org/10.3390/forecast5010006

Amini, A., Dolatshahi, M., & Kerachian, R. (2023). Effects of automatic hyperparameter tuning on the performance of multi-variate deep learning-based rainfall nowcasting. *Water Resources Research*, 59(1). https://doi.org/10.1029/2022wr032789

Aryastana, P., Dewi, L., & Wahyuni, P. (2024). A study of rainfall thresholds for landslides in badung regency using satellite-derived rainfall grid datasets. *International Journal of Advances in Applied Sciences*, 13(2), 197. https://doi.org/10.11591/ijaas.v13.i2.pp197-208

Basak, P. (2020). Forecasting of summer monsoon rainfall over gangetic west bengal, india utilising intrinsic mode functions, linear and neural regression. *Journal of Modeling and Optimization*, 12(1), 60-69. https://doi.org/10.32732/jmo.2020.12.1.60

Bhattacharya, B., Mazzoleni, M., & Ugay, R. (2019). Flood inundation mapping of the sparsely gauged large-scale brahmaputra basin using remote sensing products. *Remote Sensing*, 11(5), 501. https://doi.org/10.3390/rs11050501

Chen, X., Letu, H., Shang, H., Ri, X., Tang, C., Ji, D., … & Teng, Y. (2024). Rainfall area identification algorithm based on himawari-8 satellite data and analysis of its spatiotemporal characteristics. Remote Sensing, 16(5), 747. https://doi.org/10.3390/rs16050747

Fletcher, J., Parker, D., Turner, A., Menon, A., Martin, G., Birch, C., … & Bhat, G. (2019). The dynamic and thermodynamic structure of the monsoon over southern india: new observations from the incompass iop. *Quarterly Journal of the Royal Meteorological Society*, 146(731), 2867-2890. https://doi.org/10.1002/qj.3439

Gupta, A., Mitra, A., & Pandey, A. (2024). Prospects and status of forecasting monthly mean subregional rainfall during the indian summer monsoon using the coupled unified model. *Quarterly Journal of the Royal Meteorological Society*, 150(762), 2906-2919. https://doi.org/10.1002/qj.4741

Huang, Q., Qin, G., Zhang, Y., Tang, Q., Liu, C., Xia, J., … & Post, D. (2020). Using remote sensing data-based hydrological model calibrations for predicting runoff in ungauged or poorly gauged catchments. *Water Resources Research*, 56(8). https://doi.org/10.1029/2020wr028205

Ideki, O. and Weli, V. (2019). Analysis of rainfall variability using remote sensing and gis in north central nigeria. *Atmospheric and Climate Sciences*, 09(02), 191-201. https://doi.org/10.4236/acs.2019.92013

Khan, M., Mustafa, M., Hossain, M., Shams, S., & Julius, A. (2023). Short-term and long-term rainfall forecasting using arima model. *International Journal of Environmental Science and Development*, 14(5), 292-298. https://doi.org/10.18178/ijesd.2023.14.5.1447

Küllahcı, K. and Altunkaynak, A. (2024). Maximizing daily rainfall prediction accuracy with maximum overlap discrete wavelet transform-based machine learning models. *International Journal of Climatology*, 44(10), 3405-3426. https://doi.org/10.1002/joc.8530

Mardyansyah, R., Kurniawan, B., Soekirno, S., Nuryanto, D., & Satria, H. (2022). Artificial intelligent for rainfall estimation in tropical region : a survey. *Iop Conference Series Earth and Environmental Science*, 1105(1), 012024. https://doi.org/10.1088/1755-1315/1105/1/012024

Melesse, M. and Delele, G. (2024). Predicting and analyzing rainfall patterns in ethiopia using linear regression modeling. https://doi.org/10.20944/preprints202403.0679.v1

Mukaddim, A., Mohaimin, M., Hider, M., Karmakar, M., Nasiruddin, M., Alam, S., … & Anonna, F. (2024). Improving rainfall prediction accuracy in the usa using advanced machine learning techniques. *Journal of Environmental and Agricultural Studies*, 5(3), 23-34. https://doi.org/10.32996/jeas.2024.5.3.3

Nair, P., Chakraborty, A., Varikoden, H., Francis, P., & Kuttippurath, J. (2018). The local and global climate forcings induced inhomogeneity of indian rainfall. *Scientific Reports*, 8(1). https://doi.org/10.1038/s41598-018-24021-x

Neal, R., Guentchev, G., Arulalan, T., Robbins, J., Crocker, R., Mitra, A., … & Jayakumar, A. (2022). The application of predefined weather patterns over india within probabilistic medium-range forecasting tools for high-impact weather. *Meteorological Applications*, 29(3). https://doi.org/10.1002/met.2083

Nguyen, P., Shearer, E., Tran, H., Ombadi, M., Hayatbini, N., Palacios, T., … & Sorooshian, S. (2019). The chrs data portal, an easily accessible public repository for persiann global satellite precipitation data. *Scientific Data*, 6(1). https://doi.org/10.1038/sdata.2018.296

Nurman, S., Nusrang, M., & Sudarmin. (2022). Analysis of Rice Production Forecast in Maros District Using the Box-Jenkins Method with the ARIMA Model. *ARRUS Journal of Mathematics and Applied Science*, *2*(1), 36-48. https://doi.org/10.35877/mathscience731

Perera, H., Gunathilake, M., Panditharathne, R., Almahbashi, N., & Rathnayake, U. (2022). Statistical evaluation and trend analysis of ann based satellite products (persiann) for the kelani river basin, sri lanka. *Applied Computational Intelligence and Soft Computing*, 2022, 1-12. https://doi.org/10.1155/2022/2117771

Slater, L., Arnal, L., Boucher, M., Chang, A., Moulds, S., Murphy, C., … & Zappa, M. (2022). Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models. https://doi.org/10.5194/hess-2022-334

Sudiatmika, I. P. G. A., & Putra, I. M. A. W. (2024). Comparison of LSTM and GRU Models Performance in Forecasting Gold Prices: A Case Study Using Historical Data from Yahoo Finance. *ARRUS Journal of Engineering and Technology*, *4*(1). https://doi.org/10.35877/jetech2760

Sun, Y., Wendi, D., Kim, D., & Liong, S. (2019). Deriving intensity–duration–frequency (idf) curves using downscaled in situ rainfall assimilated with remote sensing data. *Geoscience Letters*, 6(1). https://doi.org/10.1186/s40562-019-0147-x

Swain, S., Nandi, S., & Patel, P. (2018). Development of an arima model for monthly rainfall forecasting over khordha district, odisha, india. 325-331. https://doi.org/10.1007/978-981-10-8636-6_34

Taillardat, M., Fougères, A., Naveau, P., & Mestre, O. (2019). Forest-based and semiparametric methods for the postprocessing of rainfall ensemble forecasting. *Weather and Forecasting*, 34(3), 617-634. https://doi.org/10.1175/waf-d-18-0149.1

Viana, J., Montenegro, S., Silva, B., Silva, R., Sriniva, R., Santos, C., … & Tavares, C. (2021). Evaluation of gridded meteorological datasets and their potential hydrological application to a humid area with scarce data for pirapama river basin, northeastern brazil. https://doi.org/10.21203/rs.3.rs-378302/v1

Wu, Z., Zheng, X., Chen, Y., Huang, S., Hu, W., & Duan, C. (2023). Urban flood loss assessment and index insurance compensation estimation by integrating remote sensing and rainfall multi-source data: a case study of the 2021 henan rainstorm. *Sustainability*, 15(15), 11639. https://doi.org/10.3390/su151511639

Yang, T., Yu, P., Lin, K., Kuo, C., & Tseng, H. (2018). Predictor selection method for the construction of support vector machine (svm)-based typhoon rainfall forecasting models using a non-dominated sorting genetic algorithm. *Meteorological Applications*, 25(4), 510-522. https://doi.org/10.1002/met.1717

Zhu, Z., Wright, D., & Yu, G. (2018). The impact of rainfall space-time structure in flood frequency analysis. *Water Resources Research*, 54(11), 8983-8998. https://doi.org/10.1029/2018wr023550

Zhu, Z., Yang, Y., Cai, Y., & Yang, Z. (2021). Urban flood analysis in ungauged drainage basin using short-term and high-resolution remotely sensed rainfall records. *Remote Sensing*, 13(11), 2204. https://doi.org/10.3390/rs13112204