E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Implementation Of Arima Model In The Analysis Of City Temperature Average

Muhamat Abdul Rohim[1*], Agung Muliawan[2], Ferry Wiranto[3]

ITSM, System and Technology Information, Faculty of Science, Technology and Industry, Institute Technology and Science Mandala, Indonesia[1]
ITSM, Sofware Engineering, Faculty of Science, Technology and Industry, Institute Technology and Science Mandala, Indonesia[2]
ITSM, Sofware Engineering, Faculty of Science, Technology and Industry, Institute Technology and Science Mandala, Indonesia[3]

Corresponding Author: Muhamat Abdul Rohim (muhamatabdulrohim@itsm.ac.id)

ARTICLE INFO

Date of entry:
30 April 204
Revision Date:
05 September 2024
Date Received:
31 April 2023

ABSTRACT

This study analyzes the daily average temperature data of Delhi city from 2013 to 2017 using the Autoregressive Integrated Moving Average (ARIMA) model to model and predict temperature trends. The temperature data processed in this study is non-stationary, so differentiation is applied to achieve stationarity. Two ARIMA models were evaluated: ARIMA (1,1,1) and ARIMA (1,1,1)(1,0,1). The ARIMA (1,1,1) model is effective in capturing short-term patterns, while the ARIMA (1,1,1)(1,0,1) model performs better in handling seasonal components. The findings show that the ARIMA (1,1,1)(1,0,1) model provides more accurate prediction results by accounting for seasonal fluctuations in temperature data. This research is expected to serve as a reference for preventive measures related to temperature changes, as temperature variations can affect public health, infrastructure, and quality of life in rapidly growing cities like Delhi. Understanding temperature trends and making accurate predictions helps in city planning, resource management, and developing adaptation strategies for climate change, which is crucial for mitigating negative impacts and planning for a more sustainable future.

Keywords: ARIMA, Analysis, Average Temperature, Delhi, India.

Cite this as: Rohim, M. A., Muliawan, A., & Wiranto, F. (2024). Implementation Of Arima Model In The Analysis Of City Temperature Averag. *Journal of Informatics Development*, *3*(1), 18–26. https://doi.org/10.30741/jid.v3i1.1419

INTRODUCTION

As the capital of India and one of the fastest-growing cities in the world, Delhi faces significant challenges related to climate change and rapid urbanization. Temperature changes in Delhi can impact public health, infrastructure, and overall quality of life. This article utilizes daily temperature data from the "Daily Climate Time Series Data" dataset provided by Sumanth Rao, accessible via Kaggle, covering the period from 2013 to 2017. The dataset provides information on daily maximum, minimum, and average temperatures, allowing for in-depth analysis of temperature trends, seasonal patterns, and temperature variations in the city. A study by Kumar and Garg (2018) shows that the urban heat island effect in Delhi exacerbates the impacts of climate change, while

Reddy and Reddy (2016) emphasize the importance of understanding temperature trends for better urban planning. Kothawale and Rupa Kumar (2011) also note significant changes in temperature patterns in India, which need to be considered in the context of urbanization. By analyzing this data, it is hoped that valuable insights can be gained for city planning, resource management, and climate change adaptation strategies in Delhi.

The Autoregressive Integrated Moving Average (ARIMA) method is used in this analysis to model and predict temperature trends based on historical data. ARIMA, developed by Box and Jenkins (1976), is an effective technique for forecasting time series data by combining autoregressive, differencing, and moving average elements. This method is suitable for capturing patterns in temperature data, such as long-term trends and seasonal fluctuations. Research by Hyndman and Athanasopoulos (2018) shows that ARIMA can provide accurate forecasts for time series data with trends and seasonality, making it a useful tool in urban climate planning and management. By applying the ARIMA method to Delhi's temperature data, this analysis aims to provide deeper insights into temperature trends and assist in planning responses to climate change.

Several studies have demonstrated the effective application of ARIMA models in predicting climate variables in various contexts. For example, Dimri et al. (2020) used the Seasonal ARIMA (SARIMA) approach to analyze temperature and precipitation, finding that the model can effectively capture both seasonal fluctuations and long-term trends. Meanwhile, Yan et al. (2022) applied ARIMA to analyze flood risk in reservoirs due to climate change, providing important insights into climate prediction-based flood risk management. In the data-poor Mediterranean region, Al Sayah et al. (2021) utilized ARIMA in conjunction with remote sensing to evaluate climate change, demonstrating how the model can be adapted to data-limited conditions.

Rosmiati et al. (2021) developed an ARIMA model to strengthen marine climate prediction skills, especially in the context of science education for pre-service teachers. In the economic context, Xu et al. (2024) used ARIMA to analyze the impact of climate change on decision-making in the insurance industry, emphasizing the importance of data-driven prediction for risk assessment. In addition, Deshmukh et al. (2024) compared the performance of ARIMA with LSTM deep learning models in predicting climate parameters in the Vidarbha region of India, showing that while ARIMA excels in linear and seasonal patterns, LSTM is more accurate in capturing more complex patterns.

These studies show that ARIMA models are not only effective in predicting climate trends but also make important contributions to decision-making in various sectors, including risk management, education, and economics. This shows the relevance of using ARIMA models in this study to predict temperature trends in Delhi. Based on the previous explanation, the researcher aims to analyze the existing temperature data of Delhi using the ARIMA model. The objective of this study is to understand the temperature trends in Delhi and the patterns generated. The results of this research could be used to examine the stationarity of temperature patterns in Delhi.

METHOD

The research stages are illustrated in Figure 1, which provides a detailed overview of the process flow. The figure outlines each step, from literature review, data collection, and data analysis. This

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

visual representation aims to clarify the research workflow and highlight the logical sequence of activities undertaken in this study.

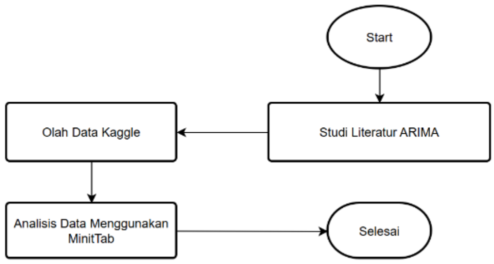

Figure 1. Research Steps
Source : Researcher (2024)

1. **Literature Review**

At this stage, the researcher conducted a comprehensive literature review on the application and implementation steps of the ARIMA model. The review focused on understanding the theoretical foundation, parameter estimation, and best practices for applying ARIMA in time series analysis. Additionally, this stage involved evaluating the suitability of the ARIMA model for analyzing temperature trends in Delhi, considering its ability to handle non-stationary data and seasonal variations. The insights gained from the literature review were crucial in guiding the model selection and adaptation process for this research.

The Autoregressive Integrated Moving Average (ARIMA) model is one of the well-known time series model families and was originally used in economics (Box & Jenkins, 2015). This model predicts future values in time series data using a combination of three main components: Autoregressive (AR), Integrated (I), and Moving Average (MA). Each of these components is represented by a parameter in the ARIMA model denoted as ARIMA(*p*, *d*, *q*), where:

p: Order of the autoregressive component,
d: Degree of differencing required to achieve stationarity,
q: Order of the moving average component.

The ARIMA model can be expressed in the following equation:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Where:
y_t is the actual value at time tt,
c is a constant,
$\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients,
ϵ_t is the error term at time tt,
$\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients,
$\epsilon_{t-1}, \epsilon_{t-2}, \ldots$ are the error term values at the previous lag.

The autoregressive component measures the relationship between the current value and the previous value at a given lag. This component is represented by the parameter p, which indicates how many lags are used in the model. For example, in AR(1), only one lag is considered:

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t$$

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

The integrated component indicates how many times the data needs to be differentiated to achieve stationarity. Differentiation is done by subtracting the previous value from the current value, and this is represented by the parameter d. For example, if the data needs to be differentiated once, then:

$$y_t' = y_t - y_{t-1}$$

The moving average component takes into account the dependence between the current value and the error at the previous lag. This component is represented by the parameter q, which indicates how many error lags are considered in the model. For example, in MA(1), only the error at the first lag is used:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1}$$

In the application of the ARIMA model, the Box-Jenkins method is used for systematic model identification, estimation, and validation. This process includes :

1. Identification: Determining the optimal values for the parameters p, d, and q based on the ACF and PACF plots.
2. Estimation: Estimation of the model coefficients using statistical estimation methods, such as least squares or maximum likelihood estimation.
3. Validation: Evaluating the model through residual analysis to ensure that the model residuals do not show significant autocorrelation.

The ARIMA model has high flexibility and has been widely applied in various fields, such as epidemiology (Alabdulrazzaq et al., 2021), climatology (Krispin, 2019), and economics (Hamilton, 2020), making it suitable for predicting time series trends in this study.

2. Data Collection

In this phase, the researcher gathered temperature data from Kaggle, specifically focusing on the average temperature in Delhi, India. To create a comprehensive dataset, the researcher combined multiple sources of average temperature data available on Kaggle into a single, unified dataset.

3. Data Analysis

At this stage, the researcher imported the temperature data obtained from Kaggle into MiniTab for further analysis. MiniTab, a powerful statistical software, is widely used for data analysis and problem-solving across various fields, such as business, industry, and research. It is designed to assist users in performing a wide range of statistical analyses, from basic descriptive statistics to complex predictive modeling. In this study, MiniTab was utilized to analyze trends, evaluate model performance, and interpret results, enabling the researcher to make informed decisions based on the statistical outputs.

RESULTS AND DISCUSSION

The first step involves observing and analyzing the Initial Data Plot generated by MiniTab. This step is essential for understanding the general patterns, trends, and potential seasonality present in the data. The plot provides a visual representation of the average temperature over time, enabling the researcher to identify key characteristics such as trend direction, fluctuations, and any apparent anomalies. The initial analysis helps in determining the appropriate model adjustments needed for further statistical evaluation. Below is the Initial Data Plot generated by MiniTab.

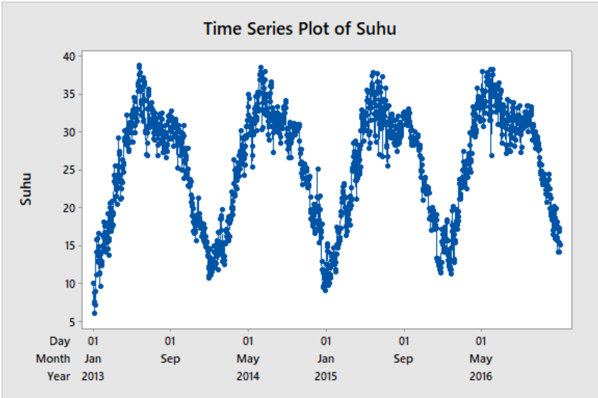

Figure 2. Plot Data
Source : Data processing (2024)

In Figure 2, the data clearly exhibits seasonal characteristics, demonstrated by the relatively regular repeating pattern observed each year, specifically at lag 12. This indicates the presence of a seasonal component in the temperature trends, making it necessary to account for seasonality in the modeling process. Following this observation, an analysis of data stationarity is performed, focusing on whether the mean, variance, and autocorrelation structure of the series remain constant over time. Achieving stationarity is essential for accurate modeling with the ARIMA framework, as it relies on this assumption to effectively capture the underlying patterns in the data.

To assess stationarity, the data is processed to generate Autocorrelation Function (ACF) plots and Box-Cox transformation plots. The ACF plot helps identify any persistent correlation at different lags, while the Box-Cox transformation evaluates potential adjustments for stabilizing variance. The result of the ACF data processing is displayed in Figure 3.

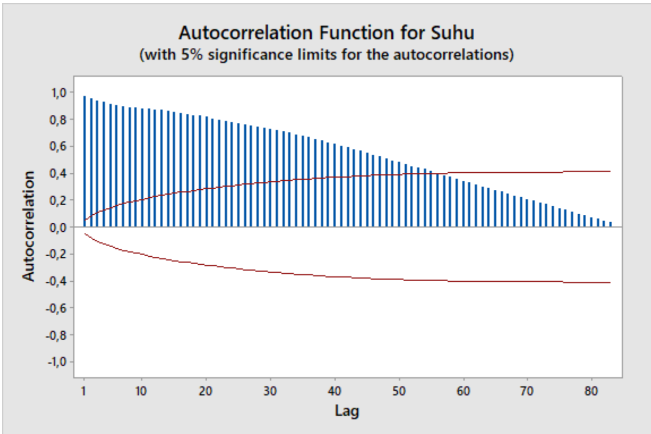

Figure 3. ACF Plot
Source : Data processing (2024)

Based on Figure 3, the ACF plot exhibits a slow decay towards zero (Dying Down), which indicates that the data is not stationary. This gradual decline suggests the presence of a trend or autocorrelation over time, signaling that the series does not maintain constant variance. Additionally, the ACF plot reveals that the first 2/3 of the lags fall outside the confidence interval, further confirming that the data is not stationary in terms of the mean. To address potential non-stationarity in variance, a Box-Cox transformation is applied. This transformation aims to stabilize the variance and make the series more suitable for ARIMA modeling. The results of the Box-Cox plot analysis are presented in Figure 4.

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

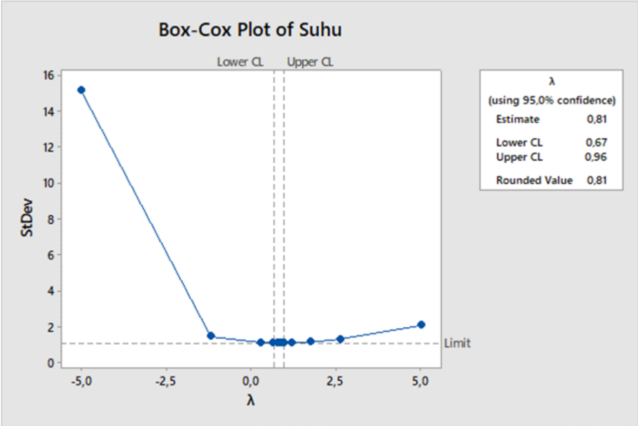

Figure 4. Box-Cox Plot
Source : Data processing (2024)

The condition for data to be stationary in variance, according to the Box-Cox method, is to have a Rounded Value (λ) of 1. However, as seen in Figure 4, the Rounded Value is 0.81, indicating that the data is not stationary in terms of variance. This suggests that the series exhibits non-constant variance, which can affect the accuracy of the ARIMA model if not properly addressed.

To handle this issue, a Variance Stabilizing Transformation (VST) was applied to the data. This transformation aims to reduce variance fluctuations and bring the data closer to stationarity. The Box-Cox test results after performing one round of VST are presented below, showing the adjusted λ value and its impact on the variance of the series.

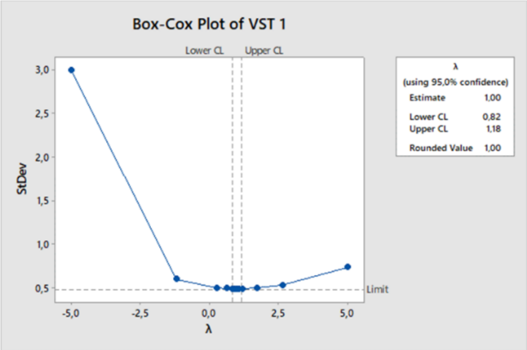

Figure 5. Box-Cox Plot of VST 1
Source : Data processing (2024)

Based on Figure 5, the Rounded Value is now 1, indicating that the data has become stationary in terms of variance after applying the Variance Stabilizing Transformation (VST). This adjustment ensures that the variance remains constant over time, making the data more suitable for ARIMA modeling.

When the data is stochastic and not stationary in terms of the mean, the differencing process is applied. In the ARIMA(p,d,q) model, this process determines the value of “d”, which represents the number of differencing operations required to achieve stationarity in the mean. The value of "d" is established based on how many times the differencing must be applied until the resulting data shows no significant autocorrelation, confirming that it is now stationary in terms of the mean.

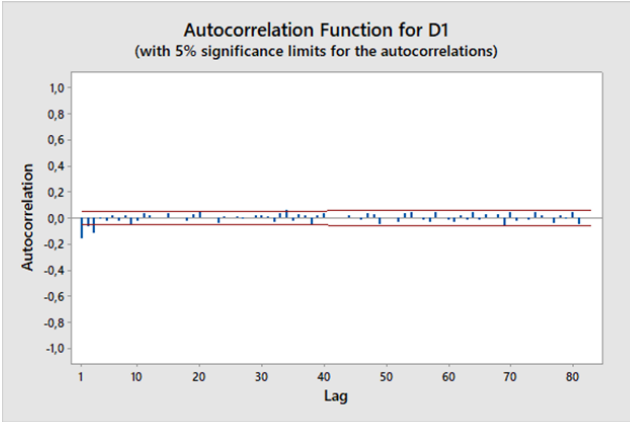

Figure 6. ACF Plot
Source : Data processing (2024)

In Figure 6, the ACF plot of the data after one round of differencing shows that the ACF is significant at lag 1, indicating that there is still some level of autocorrelation at this lag. This suggests that the differencing has helped in achieving partial stationarity, but there may still be dependencies that need to be addressed by the autoregressive components in the ARIMA model.

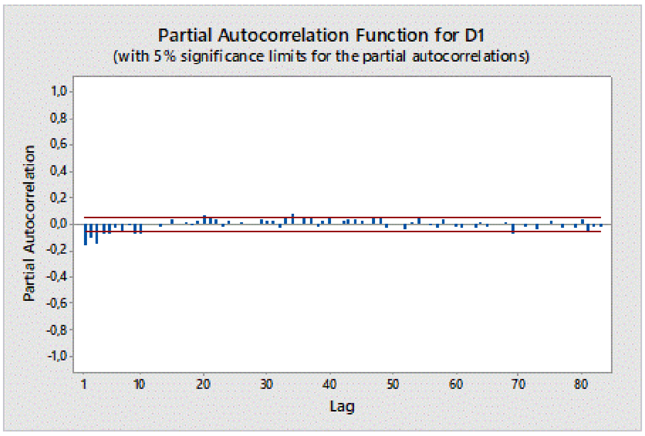

Figure 7. ACF Plot
Source : Data processing (2024)

In Figure 7, the PACF plot of the data after one round of differencing also shows that the PACF is significant at lag 1, indicating a strong correlation at this lag. This observation suggests the potential inclusion of an AR(1) component in the final ARIMA model, as it captures the immediate dependencies in the data that persist even after differencing. Both plots confirm that one round of differencing has effectively reduced non-stationarity in the mean, allowing the ARIMA model to be fine-tuned with appropriate p, d, and q parameters.

Table 1. Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.5859	0.0478	12.25	0.000
SAR 12	0.727	0.243	2.99	0.003
MA 1	0.8134	0.0342	23.82	0.000
SMA 12	0.686	0.257	2.67	0.008
Constant	0.00025	0.00109	0.23	0.815

Source : Data processing (2024)

Table 2. Residual Sums of Squares

DF	SS	MS
1455	729.256	0.501207

Source : Data processing (2024)

Table 3. Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-square	14.51	32.32	57.26	72.78
DF	7	19	31	43
P-Value	0.043	0.029	0.003	0.003

Source : Data processing (2024)

As discussed in Figure 2, the data exhibits seasonal characteristics with a significant pattern at lag 12, indicating the need for a seasonal ARIMA model. During the differencing process, it was observed that both the PACF and ACF remained significant at lag 1 after one round of differencing, leading to the selection of the ARIMA(1,1,1) model for the non-seasonal component. No differencing was applied to the seasonal component, resulting in the final model being ARIMA(1,1,1)(1,0,1).

The estimation results calculated using MiniTab software are presented in Tables 1, 2, and 3. The coefficients of the model components are detailed in Table 1, the AR(1) coefficient is 0.5859 with a standard error of 0.0478, yielding a t-value of 12.25 and a p-value of 0.000, indicating that the AR term is highly significant. The SAR(12) coefficient is 0.727 with a standard error of 0.243, resulting in a t-value of 2.99 and a p-value of 0.003, confirming the significance of the seasonal autoregressive term. The MA(1) coefficient is 0.8134 with a standard error of 0.0342, producing a t-value of 23.82 and a p-value of 0.000, which also demonstrates strong significance. The SMA(12) coefficient is 0.686 with a standard error of 0.257, giving a t-value of 2.67 and a p-value of 0.008, indicating that the seasonal moving average term is statistically significant. The constant term is 0.00025 with a standard error of 0.00109, yielding a t-value of 0.23 and a p-value of 0.815, suggesting that it is not statistically significant.

The analysis of variance (ANOVA) results are shown in Table 2, where the degrees of freedom (DF) is 1455, the sum of squares (SS) is 729.256, and the mean square (MS) is 0.501207. This indicates a good fit of the model to the data, as the residual variance is relatively low. The Ljung-Box test results in Table 3 assess the adequacy of the model by testing the independence of residuals at lag 12, the Chi-square value is 14.51 with DF 7 and a p-value of 0.043, indicating marginal significance. At lag 24, the Chi-square value is 32.32 with DF 19 and a p-value of 0.029, suggesting that some autocorrelation remains in the residuals. At lags 36 and 48, the Chi-square values are 57.26 and 72.78, respectively, with p-values of 0.003 at both lags, which indicates significant autocorrelation. These results suggest that while the ARIMA(1,1,1)(1,0,1) model effectively captures both the non-seasonal and seasonal components, some autocorrelation remains in the residuals at higher lags. This suggests potential areas for model refinement, such as incorporating additional seasonal terms or re-evaluating differencing strategies.

CONCLUSIONS

Based on the research conducted, it can be concluded that the initial data obtained from Kaggle is non-stationary, necessitating the use of a Variance Stabilizing Transformation (VST) to achieve variance stationarity. After the transformation, the chosen ARIMA model for analysis was ARIMA(1,1,1), with an additional seasonal component that did not require differencing, resulting in the final model ARIMA(1,1,1)(1,0,1). This model effectively captured the average temperature trends in Delhi, addressing both short-term patterns and seasonal components. Several recommendations are proposed for future research. First, the data used in this study was solely obtained from Kaggle. It would be beneficial to incorporate more recent field data to provide results that are more relevant to current temperature conditions. Second, in addition to the ARIMA method, it is advisable to explore other methods to diversify the analysis of average temperature trends, thus offering a broader understanding of temperature patterns in Delhi.

REFERENCES

Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., Alghannam, B. A., Al-Hassan, A. A., & Al-Anzi, F. S. (2021). On the accuracy of ARIMA-based prediction of COVID-19 spread. Results in Physics, 27, 104509.

Al Sayah, M. J., Abdallah, C., Khouri, M., Nedjai, R., & Darwich, T. (2021). A framework for climate change assessment in Mediterranean data-sparse watersheds using remote sensing and ARIMA modeling. Theoretical and Applied Climatology, 143, 639-658.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.

Box, G., & Jenkins, G. (2015). Time Series Analysis: Forecasting and Control. Wiley.

Deshmukh, K. H., Bamnote, G. R., & Deshmukh, P. S. (2024). Predicting climatic parameters using the ARIMA time series and LSTM deep learning models for Vidarbha region. International Research Journal on Advanced Engineering Hub (IRJAEH), 2(05), 1255-1264.

Dimri, T., Ahmad, S., & Sharif, M. (2020). Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth System Science, 129, 1-16.

Hamilton, J. D. (2020). Time Series Analysis. Princeton University Press.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts. Retrieved from https://otexts.com/fpp3/

Krispin, R. (2019). Hands-On Time Series Analysis with R: Perform Time Series Analysis and Forecasting Using R. Packt Publishing Limited.

Kothawale, D. R., & Rupa Kumar, K. (2011). On the recent trends of temperature and rainfall in India. Journal of Earth System Science, 120(4), 405-423. https://doi.org/10.1007/s12040-011-0140-3

Kumar, S., & Garg, R. (2018). Analysis of urban heat island effect in Delhi, India. Journal of Environmental Management, 216, 229-240. https://doi.org/10.1016/j.jenvman.2018.04.039

Rahmalia, N. (2021, January 14). Kaggle, komunitas belajar data science yang bisa jadi sumber uang. Glints. Retrieved from https://glints.com/id/lowongan/kaggle-adalah/#.YLAdR6gzbDc

Rao, S. (2019). Daily Climate Time Series Data. Kaggle. Retrieved from https://www.kaggle.com/sumanthvrao/daily-climate-time-series-data

Reddy, M. S., & Reddy, M. (2016). Trends and variability of temperature in Delhi: A statistical analysis. Climate Dynamics, 46, 3231-3243. https://doi.org/10.1007/s00382-015-2806-8

Rosmiati, R., Liliasari, S., Tjasyono, B., & Ramalis, T. R. (2021). Development of ARIMA technique in determining the ocean climate prediction skills for pre-service teacher. In Journal of Physics: Conference Series (Vol. 1731, No. 1, p. 012072). IOP Publishing.

Xu, H., Ge, Z., & Ao, W. (2024). Research on climate change prediction based on ARIMA model and its impact on insurance industry decision-making. Frontiers in Computing and Intelligent Systems, 8(1), 1-5.

Yan, B., Mu, R., Guo, J., Liu, Y., Tang, J., & Wang, H. (2022). Flood risk analysis of reservoirs based on full-series ARIMA model under climate change. Journal of Hydrology, 610, 127979.