IMPLEMENTATION OF ARMA MODEL FOR BENGAWAN SOLO RIVER WATER LEVEL AT JURUG MONITORING POST

Sri Siswanti[1]; Retno Tri Vulandari[2*]; Setiyowati[3]

Informatics[1], Information technology[2], Information System[3]
STMIK Sinar Nusantara, Surakarta, Indonesia
www.sinus.ac.id
syswanty@sinus.ac.id[1], retnotv@sinus.ac.id[2*], setiyowati@sinus.ac.id[3]
(*) Corresponding Author

Abstract—The amount of annual rainfall in the Bengawan Solo watershed causes high water flow (water discharge) in several rivers. In addition, high flow rates significantly increased the water surface level at some observation sites. The Bengawan Solo River burst its banks in November 2016, causing flooding in several areas in eastern Solo. At that time, the river stage at the Jurug monitoring post passed ten. Therefore, a flood early warning system would be useful for predicting water levels in this context. Every day, one post on the Bengawan Solo River measures the water level. The time series data used in this study is the water level. Autoregressive Moving Average (ARMA) is a predictive method for measuring time set data. The assumption of homoscedasticity or constant error variance is used in this model. However, the study will use the ARMA model if the variance changes randomly. The study used 60 pieces of data from January to February 2018. This study can directly use ARMA because the data results are stationary based on ADF value 0.0036. After the first pause, the ACF and PACF are disconnected based on the correlogram pattern. This shows that the water level of the Bengawan Solo River in that period can appear on the Autoregressive Moving Average with orders p = 1 and q = 1 ARMA(1,1). Thus, the total average residue is 0.0668384, so the short error is 6.68384%.

Keywords: autoregressive moving average (ARMA), jurug monitoring post, time series analysis, water level.

Abstrak—Jumlah curah hujan tahunan di DAS Bengawan Solo menyebabkan aliran air (debit air) yang tinggi di beberapa sungai. Selain itu, laju aliran yang tinggi meningkatkan ketinggian permukaan air secara signifikan di beberapa lokasi pengamatan. Sungai Bengawan Solo meluap pada November 2016, menyebabkan banjir di beberapa daerah di timur Solo. Pada saat itu, tahap sungai di pos pemantauan Jurug melewati sepuluh. Oleh karena itu, sistem peringatan dini banjir akan berguna untuk memprediksi ketinggian air dalam konteks ini. Setiap hari, satu pos di Sungai Bengawan Solo mengukur tinggi airnya. Data time series yang digunakan dalam penelitian ini adalah ketinggian muka air. Autoregressive Moving Average (ARMA) adalah metode prediksi untuk mengukur data set waktu. Asumsi homoskedastisitas atau varians kesalahan konstan digunakan dalam model ini. Namun, penelitian ini akan menggunakan model ARMA jika variansnya berubah secara acak. Studi ini menggunakan 60 data dari Januari hingga Februari 2018. Penelitian ini dapat langsung menggunakan ARMA karena hasil datanya stasioner berdasarkan nilai ADF 0,0036. Setelah jeda pertama, ACF dan PACF terputus berdasarkan pola correlogram. Hal ini menunjukkan bahwa tinggi muka air Sungai Bengawan Solo pada periode tersebut dapat muncul pada Autoregressive Moving Average dengan order p=1 dan q=1 ARMA(1,1). Dengan demikian, total rata-rata residunya adalah 0,0668384, sehingga kesalahan singkatnya adalah 6,68384%.

Kata Kunci: autoregressive moving average (ARMA), pos pemantauan jurug, analisa runtun waktu, tinggi muka air.

INTRODUCTION

River basin is the most important information for water resource management. In addition, the information of its peak flow can be useful to design better flood control buildings (Hidayat et al., 2022). Meanwhile, the data of small

Vol. 20, No. 1 March 2024 | DOI: 10.33480/pilar.v20i1.5004

river stream is needed for useful water location planning, especially in long dry season. The average annual flow can give better pictures of water resource potential which can offer advantages from a river basin (Saidah & Hanifah, 2020). Water debit is water flow rate that passes a cross-section in the river per unit time. In technical reports, its flow rate usually appears in hydrograph (Biantoro et al., 2021). Hydrograph shows behaviour of water flow rate as a response of the changes of biogeophysical characteristic which happens because of the river basin management activity and seasonal or annual fluctuation, like local climate changes (Park et al., 2023).

The annual rainfall intensity in Bengawan Solo river basin causes its big water debit in some streams. In addition, the flows also increase the height of water level in some dam posts. The measurement of water level in every dam can certainly be useful to prevent flood (Trinugroho et al., 2022). At the end of 2016, Bengawan Solo River was overflowing and the flood covered East Solo areas. According to a report in Jurug monitoring post, the river stage passed over 10 and the flood covered 10 districts of East Solo. At this case, it needs an efficient model to predict the water level, especially flood stage, as flood early warning in order not to bring back the past disaster in Solo. This research uses the rate of water level, which is measured every day, at Jurug monitoring post as time series data.

Arrangement with a stationary model, such as Autoregressive Moving Average (ARMA), is possible with Autocorelation Function (ACF) and Partial Autocorelation Function (PACF). ARMA model has the assumption of homoscedacity or a constant error variance, but it turns into heteroscedacity if time series data from the water surface have a constant error variance every time (Nguyen, 2020).

MATERIALS AND METHODS

1. Data Processing Procedure

Procedures for processing data used in this study are as follows:

a. Data Collection
This study requires time series data and measurement results, including water level data from the Jurug monitoring post from 2009 to June 2018.

b. Input Data Processing
First, the ARMA-GARCH method was used to analyse the data, then the original data plot was used to identify the distribution pattern of the data, and then stationary tests were performed (Berutu et al., 2023).

c. Data Sharing (Load Data)
The data distribution consists of testing data, namely TMA data from January - February 2018 which is taken randomly, training data, namely the water level of the Bengawan Solo river in 2009 - June 2018.

2. Data Analysis

Stationary tests, identification of the ACF and PACF models, estimation of the ARMA model parameter, diagnostic tests, and then GARCH error variance models are all part of the process of creating the ARMA stationary model.

a. Stasionecity Test

The unit root test can be used to determine the data's stationarity. The hypothesis of the examination is written as

H0 : $\eta = 1$ (data has unit root)
H1 : $\eta < 1$ (data has no unit root)

Test statistics are the ratio of the estimated coefficient minus 1 compared to its standard deviation (Fauzi & Irviani, 2023). The Augmanted Dickey-Fuller (ADF), also known as the t-ratio, is formulated in the same way as in formula (1).

$$\text{ADF} = \frac{\eta - 1}{\sigma(\eta)} \quad (1)$$

with x_0 = 0, T is the sample size and is the x_t t-th observation data. H0 is rejected when the ratio t > t α ,(T − 1)

b. ACF and PACF Model Identification

The tools for identifying ARMA models are ACF and PACF. The autocorrelation function shows the magnitude of the correlation between observations at t-time and observations at previous times, while the PACF is a function that shows the magnitude of the partial correlation between observations at t-time and observations at previous times (Maulidiyah & Fauzy, 2023).

c. Stationary Model Parameter Estimation

ARMA model contains two components: the AR model and the MA model, where the order of AR is p and the order of MA is q (Gustiansyah, Rizki, & Apriyanti, 2023). Here is a stationary model according to (Safwandi, 2023).

(1) Autoregressive (AR)
A observation at time t that is expressed as a linear function against p at the prior time plus a random error e_t is autoregressive (AR). The general form of the p-order autoregressive model is formulated in formula (2).

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + e_t \text{.......(2)}$$

(2) Moving Average (MA)
A phenomenon that an observation at time t is represented as a linear combination of a number of random errors is called moving average (MA). The general form of the q-order moving average model is formulated in formula (3).

$$Y_t = e_t - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \cdots - \beta_a e_{t-q} \text{........(3)}$$

(3) Autoregressive Moving Average (ARMA)
The main model of Autoregressive Moving Average (ARMA) (p,q) is expressed in formula (4). ARMA is a combination of AR and MA.

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + e_t - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \cdots - \beta_a e_{t-q} \text{......................(4)}$$

d. Diagnostic Model
The diagnostic model is used to find out if the model is suitable for use. Error is an excellent indicator of model compatibility if it is homogeneous, autocorrelation-free, and has a low mean square error (MSE) value (Marheni & Triyanto, 2023). As a result, computations of the MSE value, variance homogeneity tests, and error autocorrelation tests are performed. Re-identifying and estimating is required if the error does not satisfy these three requirements, indicating that the model developed does not match the data (Dewi & Indah, 2022).

e. Model ARCH dan Model GARCH
The ARCH model and GARCH model (Sumiyati & Wilujeng, 2022) are models used for time series data that have high volatility and have inconstant error variances (heteroscedasticity). Engle developed the ARCH model with mean and variance modeled simultaneously. General form of the ARCH(p) model is formulated in formula (5).

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_p e_{t-p}^2 \text{.............(5)}$$

with σ_t^2 residual variances and e_{t-p}^2 residual squares of past periods. Bollerslev developed the ARCH model by taking into account residual variances of past periods. General form of the GARCH(p,q) model is formulated in formula (6).

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_p e_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + \cdots + \lambda_q \sigma_{t-q}^2 \text{ ....(6)}$$

with σ_{t-q}^2 residual variances of past periods.

f. Research Mindset
The operational steps to accomplish the research goals are listed below.
(1) To determine the distribution pattern of the data, create a data plot;
(2) Use a unit root test to perform a stationary test; if the data is stationary, the direct data can be modeled.
(3) The ln transformation is carried out if the data is not stationary. Next, run the unit through another root test.
(4) ACF and PACF graphs were used to identify the model after the stationary data. After that, it calculates the parameter's magnitude and draws conclusions from the data's stationary model.
(5) Diagnostic tests on errors produced by the model are then conducted following the formation of the stationary model. A mistake is the distinction between prediction data and actual data.
(6) If the assumption of homogeneity of variance is not met, then it means that the data has variable errors.
(7) Perform modeling for error variance correction using GARCH.
(8) Perform these simulation stages using MATLAB.

RESULTS AND DISCUSSION

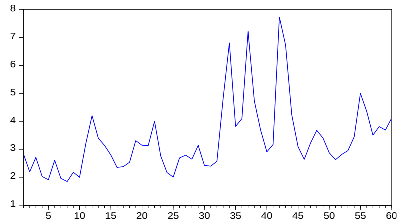
a. Description and Data Pattern

A time series that remains immobile is one whose characteristics remain constant regardless of the observation time of the sequence. Time series that exhibit a trend or seasonality are therefore non-stationary, and the presence of seasonality will impact the time series' value at different points in time. A white noise sequence, on the other hand, is stationary and does not change in appearance over time; it should appear relatively constant. A stagnant time series will typically not exhibit any predictable patterns over an extended period of time. Time plots will reveal that the series has a constant variance and is essentially level.

Table 1. Stationary Data Test

Null Hypothesis: TMA has a unit root		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3,906148	0,0036
Test critical values:	1% level	-3,546099	
	5% level	-2,911730	
	10% level	-2,593551	

(Research Results, 2018)



(Research Results, 2018)

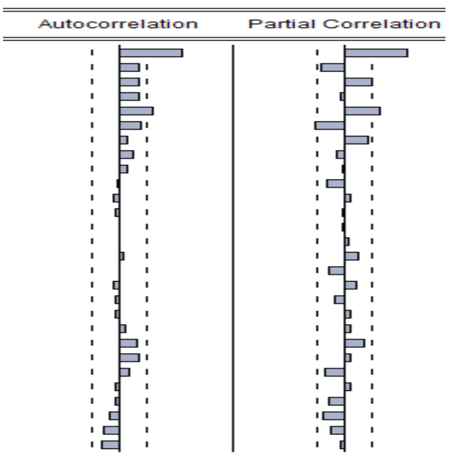
Figure 1. Time Series Plot of Water Level

Based on the research method, the first step to do is data description. Figure 1 shows the average data of water level is stationary in inconstant variance and its stationary and unit root test strengthens this circumstance. For the result of the test, the probability value of its Augmented Dickey-Fuller (ADF) is 0,0036. In this case, that probability value is smaller than significant level α = 0,05. Moreover, it can be proven by statistic value t that $|t|_{TMA} = 2{,}911730 > t_{(0{,}05;59)} = -1{,}671$, in which H_0 is successfully rejected. It means the data have no unit root (Table 1) and it turns to be stationary. As a result of stationary tendency toward its average, this research provides the average model firstly before its variance model.

b. Formation of Conditional Mean Model in Stationary Process

1) Identification Model

In this research, conditional mean model from stationary data can use ARMA. This research also applies ACF and PACF to identify ARMA comprehensively. In Figure 2, ACF and PACF value cut off after the first lag then ARMA will be fit for the conditional mean model.



(Research Results, 2018)

Figure 2. Water Level of ACF and PACF

2) Parameter Estimation

ARMA (1, 1) is a process of autoregressive order 1 and process of moving average order 1. It shows in the formula (7).

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} \ldots\ldots(7)$$

Then, it obtains auto covariance function, as in formulas (8) and (9).

For k = 0,

$$E(Y_t Y_t) = \gamma_0 = \phi\gamma_1 + \sigma_e^2 - \theta(\phi - \theta)\sigma_e^2 \ldots(8)$$

For k = 1,

$$E(Y_t Y_{t-1}) = \gamma_1 = \phi\gamma_0 - \theta\sigma_e^2 \ldots\ldots(9)$$

Applied with substitution of equation (2) to equation (1), it obtains formula as formulated in formula (10) and (11).

$$\gamma_0 = \frac{(1-2\theta\phi+\theta^2)}{(1-\phi^2)}\sigma_e^2 \ldots\ldots(10)$$

$$\gamma_1 = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\sigma_e^2 \ldots\ldots(11)$$

For k = 2, auto covariance function shows as in formula (12), (13), (14, (15) and (16).

$$\gamma_2 = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\phi\sigma_e^2 \ldots\ldots(12)$$

For k = k,

$$\gamma_k = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\phi^{k-1}\ \sigma_e^2 \ldots\ldots(13)$$

Mean while, auto correlation function for k = 1 is shown in the formula (14).

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2} \ldots\ldots(14)$$

For k = 2,

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)\phi}{1-2\theta\phi+\theta^2} \ldots\ldots(15)$$

For k = k,

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)\phi^{k-1}}{1-2\theta\phi+\theta^2} \ldots\ldots(16)$$

For the next one, the parameter estimation of ARMA (1,1) model shows a result in table 2.

Table 2. ARMA (1,1) Model

Variable	Coefficient
C	3.376777
AR(1)	0.075557
MA(1)	0.948131

(Research Results, 2018)

Table 2 shows $\hat{\phi}_1 = 0{,}075557, \hat{\theta}_1 =$ 0,0,948131 and the intercept value is 3,376777. As a result, ARMA(1,1) model is: $Y_t = 0{,}075557Y_{t-1} + 3{,}376777 + e_t -$ $0{,}0{,}948131e_{t-1}$ with Y_t is water level at time t and e_t is ARMA errors at time t

3) Diagnostic Test

In this auto correlation test, good conditional mean models have no auto correlation in the errors. Auto correlation shows the inter-correlation of its observations. Breusch-Godfrey test is a statistical test to find out the presence of auto correlation in the errors (Chaudhary et al., 2022). Table 3 shows this Breusch-Godfrey test of the research.

Table 3. Auto Correlation Test

Uji Breusch-Godfrey	0,2134
Error lag-1	0,7652
Error lag-2	0,7213
Error lag-3	0,4512
Error lag-4	0,2111
Error lag-5	0,1202
Error lag-6	0,0912
Error lag-7	0,2307
Error lag-8	0,3087

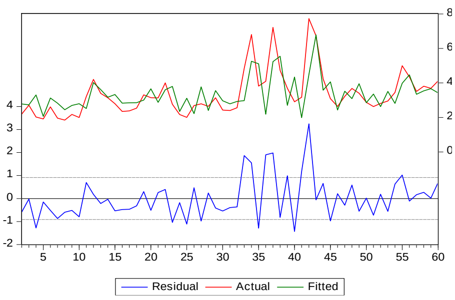
(Research Results, 2018)

Heteroscedacity has inconstant variability in the errors for each observation. Its data values tend to have rapid fluctuation. Meanwhile, volality enables to measure and draw fluctuation of data. Volality can be defined as data variance over the time series. Volality also shows the tendency of data to changes rapidly over the time so that error variance changes each time. In Table 4, White test can present the volality with data probability value less than $\alpha = 0,05$. As a result, it obtains $R^2 = 9,213801 > \chi^2_{0,05;2} = 5,991$ and H_0 is rejected which shows the heteroscedacity in the data.

Table 4. White's Heteroscedastic Test

Obs. R^2	9,213801
Probabilitas χ^2	0,0100

(Research Results, 2018)



(Research Results, 2018)
Figure 3. Residual, Actual, and Fitted

Based on Figure 3, it can be seen that the actual data with the fitted data has the same pattern. The red line indicates the actual data pattern, while the green line indicates the prediction data pattern. Actual data patterns are similar to fitted data pattern. Blue lines indicate residual or data errors, having the same pattern.

CONCLUSION

The water level of Bengawan Solo River on Januari – February 2018 uses ARMA (1,1) model because its result tends to be stationary toward the average but it has inconstant data variance. Based on research that has been done, the ARMA model (1.1) can be used to predict the water level of the Bengawan Solo river. The model was successfully used as an early warning to prevent flooding, this can be seen from the diagnostic test comparing actual data and fitted data in April and May 2018, having the same data pattern. For the following research, it will be great if there is data variance model using GARCH or TARCH model. Time series data is difficult to show in a certain model because its data fluctuation gets influence from many factors based on its individual characteristics such as TGARCH, MGARCH, and APARCH. Those models are fit for asymmetric time series data.

REFERENCE

Berutu, S. S., Budiati, H., & Gulo, F. (2023). Data Preprocessing Approach for Machine Learning-based Sentiment Classification. *Jurnal Informatics, Telecommunication, and Electronics Vol 15 No 4*, 34-43.

Biantoro, A. W., Iskendar, & Subekti. (2021). The Effects of Water Debit and Number of Blades on the Power Generated of Prototype Turbines Propeller as Renewable Electricity. *Jurnal Rekayasa Mesin Vol 12 No 1*, 54-64.

Chaudhary, A. A., Nazir, N., Riaz, A., Sadiq, N., & Riaz, N. (2022). Autocorrelation: what happens if the error or disturbance terms are correlated in time-series data. *Competitive Education Research Journal*, *3*(2), 154-163.

Dewi, W. I., & Indah, N. P. (2022). Pengaruh Perputaran Kas dan Modal Kerja terhadap Profitabilitas Perusahaan di Bursa Efek Indonesia. *Jurnal Ekonomi, Manajemen, dan Akuntansi (JEBM) Vol 24 No 3*, 624-629.

Fauzi, & Irviani, R. (2023). Analysis of The Effect of Macropredential and Macroprudential Indicators on The Stock Price Index of Islamic Banks in Indonesia Using The Error Correction Model. *Jurnal Syntax Admiration Vol 4 No 5*, 1-19.

Vol. 20, No. 1 March 2024 | DOI: 10.33480/pilar.v20i1.5004

Gustiansyah, M. A., Rizki, A., & Apriyanti, B. (2023). Aplikasi Model ARIMA dalam Peramalan Data Harga Emas Dunia. *Jurnal Statistika dan Aplikasinya Vol 7 No 1*, 84-92.

Hidayat, D., Legowo, S., & Farid, M. (2022). Clarification of Watershed Recharge in Cisadane River Basin throughGround Test. *Journal of the Civil Engineering Forum,*, 217-224.

Marheni, R., & Triyanto, E. (2023). Pengaruh Dana Alokasi Umum, Dana Alokasi Khusus dan Dana Bagi Hasil terhadap Belanja Modal di Kabupaten/Kota Provinsi Jawa Tengah. *Jurnal Cakrawala Ilmiah Vol 2 No 11*, 224-232.

Maulidiyah, W., & Fauzy, A. (2023). Perbandingan Metode Peramalan Double Exponential Smoothing dengan Damped Parameter dan Autoregressive Integrated Moving Average. *Emerging Statistics and Data Science Journal Vol 1 No 3*, 361-377.

Nguyen, X. H. (2020). Combining Statistical Machine Learning Models with ARIMA for Water Level forecasting: The case of the Red river. *Advances in Water Resources Journal Vol 142*, 47-52.

Park, K., Seong, Y., & Jung, Y. (2023). Development of Water Level Prediction Improvement Method Using Multivariate Time Series Data by GRU Model. *MDPI Journal Issu 15*, 1-14.

Safwandi. (2023). Time Series Model Using Autoregressive Integrated Moving Average (ARIMA) Method for Inflation in Indonesia. *Jurnal Investasi Islam Vol 8 No 1*, 13-25.

Saidah, H., & Hanifah, L. (2020). Analysis of Water System Condition for Monitoring of Critical Area in Jangkok Watershed. *Jurnal Tanah dan Sumberdaya Lahan Vol 7 No 2* , 237-248.

Sumiyati, & Wilujeng, P. R. (2022). Metode ARCH/GARCH untuk Memprediksi Hubungan Economic Uncertainty (COVID 19) dan Volatilitas Saham. *Jurnal Bisnis dan Akuntansi Vol 24 No 1*, 117-130.

Trinugroho, M. W., Arif, S. S., & Susanto, S. (2022). Changes in Rainfall Pattern in Bengawan Solo Sub-Watershed. *Journal of Soil Science and Argoclimatology Vol 19 No 2*, 125-136.