Time Series Forecasting using ARMA-GARCH Models: Sto. Niño, Marikina River Water Level Prediction

Alegarbes, Brynx Junil
Que, Neil Bryant
Tan, Jeremy Marcus

MATH 271.2

Ateneo de Manila University

11 July 2024

Outline

Outline

Background of the Study

Figure 1: Marikina River, Source: Philippine Star

- Flooding is a persistent issue in Marikina City, Philippines.

Background of the Study

Figure 1: Marikina River, Source: Philippine Star

- Moreover, heavy rainfall events like typhoons cause an increase in the water level along the Marikina River.

Background of the Study

Figure 1: Marikina River, Source: Philippine Star

- Thus, effective flood forecasting systems are crucial in minimizing the hazards posed by extreme rainfall events.

Marikina River Basin

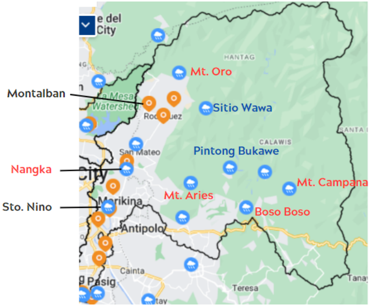


Figure 2: Map of the Marikina River Basin, Source: PAG-ASA

- The Sto. Niño water level station is located in the lower-lying section of the Marikina River Basin.

Marikina River Basin

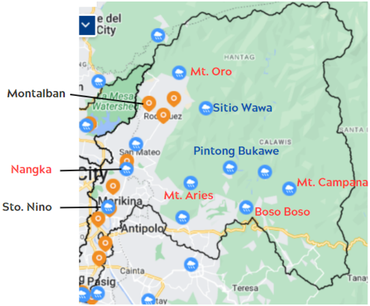


Figure 2: Map of the Marikina River Basin, Source: PAG-ASA

- Currently, Marikina City employs an alarm level system using the water level under Sto. Niño bridge (Serafica 2017).

Statement of the Problem

Due to the lack of forecasting capabilities along the MRB, this study aims to conduct a time series analysis in forecasting Marikina River water levels. Specifically, the project aims to:

1. Gather and analyze 2017 data of the Marikina River water levels at the Sto. Niño station from the Metropolitan Manila Development Authority (MMDA);

Statement of the Problem

Due to the lack of forecasting capabilities along the MRB, this study aims to conduct a time series analysis in forecasting Marikina River water levels. Specifically, the project aims to:

1. Gather and analyze 2017 data of the Marikina River water levels at the Sto. Niño station from the Metropolitan Manila Development Authority (MMDA);
2. Develop and implement linear time series models (AR, MA, and ARMA) in predicting the water levels;

Statement of the Problem

Due to the lack of forecasting capabilities along the MRB, this study aims to conduct a time series analysis in forecasting Marikina River water levels. Specifically, the project aims to:

3. Fit volatility models (ARCH and GARCH) to further capture the variation of the water levels; and

Statement of the Problem

Due to the lack of forecasting capabilities along the MRB, this study aims to conduct a time series analysis in forecasting Marikina River water levels. Specifically, the project aims to:

3. Fit volatility models (ARCH and GARCH) to further capture the variation of the water levels; and
4. Assess the performance of the best linear time series model combined with the best volatility model and identify the best model to capture water level dynamics and generate reliable forecasts.

Scope and Limitations

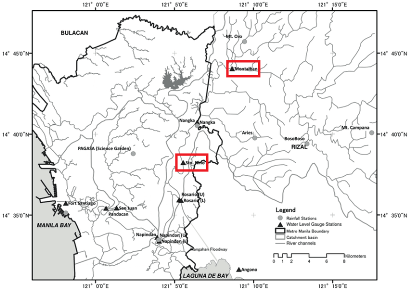


Figure 3: Marikina River Basin Stations, Source: MMDA

- The results of this analysis will be limited to the Sto. Niño Water Level Station of the MRB. There are other water level stations situated along the MRB which might require different models to analyze.

Scope and Limitations

- Because of the constantly changing physical characteristics of the Marikina River Basin, the time series models created may not be as accurate for extended forecasts beyond 2017.
- Only basic linear time series models (AR, MA, and ARMA) and basic volatility models (ARCH and GARCH) were used.

Scope and Limitations

- Other studies used different time series models for the mean equation, such as the SETAR-GARCH model, which gives interval predictions (Guo, Song, and Ma 2021).
- Related literature also used different volatility models such as multivariate GARCH (MGARCH) models to account for multiple hydrologic variables like rainfall in the prediction (Modarres and Ouarda 2013).

Outline

Dataset

Waterlevel_Sto_Niño	datetime
11.87	02/07/2017 13:00
11.85	02/07/2017 14:00
11.84	02/07/2017 15:00

Table 1: Snippet of the Dataset

- This study used hourly water level data (in meters) at the Sto. Niño station in the entire 2017, which was obtained from MMDA.

Dataset

Waterlevel_Sto_Niño	datetime
11.87	02/07/2017 13:00
11.85	02/07/2017 14:00
11.84	02/07/2017 15:00

Table 1: Snippet of the Dataset

- Further inspection reveals that there may be incorrect records.
- For instance, the data for most of January is constant, then it is followed by a sudden drop in water level.

Plot of the Dataset

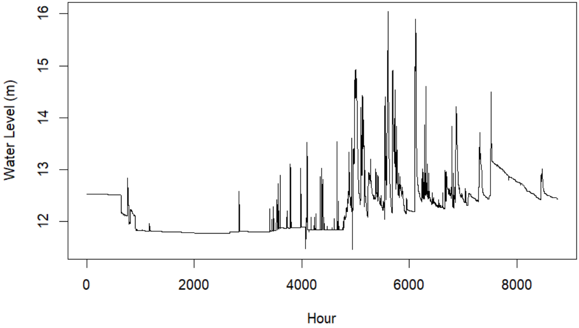


Figure 3: 2017 Sto. Niño water levels

- It is observed that the water level was typically low and stable during the first half of the year.

Plot of the Dataset

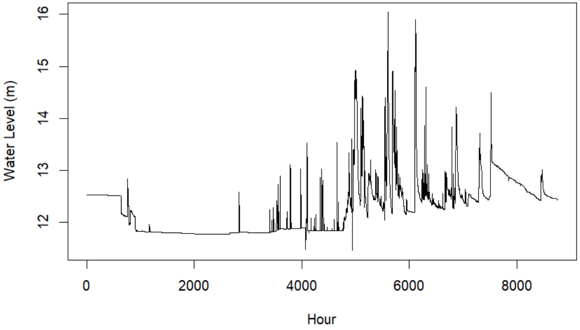


Figure 3: 2017 Sto. Niño water levels

- Spikes occur in the latter half of the year, which corresponds to the rainy and typhoon season in the Philippines.

Stationarity and Autocorrelation

Checking for Stationarity:

- Using the Augmented Dickey-Fuller Test, the time series data is stationary with a $p-$value of 0.01.

Stationarity and Autocorrelation

Checking for Stationarity:

- Using the Augmented Dickey-Fuller Test, the time series data is stationary with a $p-$value of 0.01.

Checking for Autocorrelation:

- Using the Ljung-Box Test, the time series data is serially correlated with $p-$value of less than 2.2×10^{-16}.

Autocorrelation Function (ACF)

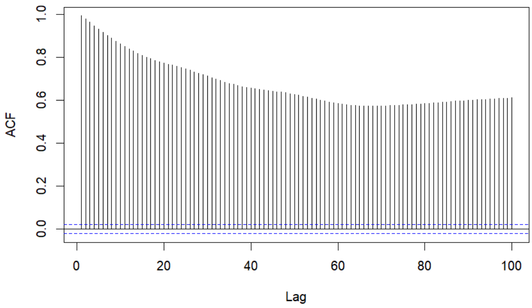


Figure 4: ACF Correlogram

- The ACF correlogram does not seem to cut off so no MA (and ARMA) models can be tested.

Partial Autocorrelation Function (PACF)

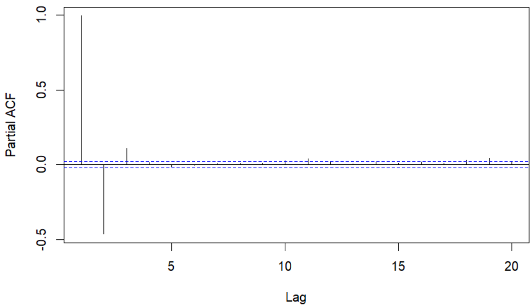


Figure 5: PACF Correlogram

- The lag orders where the PACF cuts off are 3 and 11. However, the lag order 2 was also checked for parsimony.

Identifying Candidate Models

Model	Log Likelihood	AIC	AICc	BIC
AR(2)	12367.56	−24727.13	−24727.12	−24698.81
AR(3)	12417.99	−24825.97	−24825.96	−24790.58
AR(11)	12431.05	−24836.10	−24836.06	−24744.09

Table 2: Model performance for candidate models

- The best and worst performing models based on the four metrics are shown in Table 2. To maintain parsimony, the AR(3) model will be chosen as the best model.

Best Mean Model

```
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1      ar2     ar3     mean
      1.5035  -0.6200  0.1070  12.2958
s.e.  0.0106   0.0181  0.0106   0.0658
```

Mean Model AR(3):

$$X_t = 0.1168 + 1.5035X_{t-1} - 0.6200X_{t-2} + 0.1070X_{t-3} + Y_t,$$

where $\phi_0 = \mu(1 - \phi_1 - \phi_2 - \phi_3)$.

ARCH Effect

- The ARCH (Autoregressive Conditional Heteroskedasticity) effect occurs when the variances of the residuals are not constant over time.

ARCH Effect

- The ARCH (Autoregressive Conditional Heteroskedasticity) effect occurs when the variances of the residuals are not constant over time.
- An ARCH model can be used to model and forecast the changing variances of the residuals. An ARCH(m) model is a function of the squared residuals of the past m values.

Mean Model Diagnostics

Checking for the ARCH effect:

- Using the Ljung-Box Test, the mean model residuals were not serially correlated with a $p-$value of 0.9028.

Mean Model Diagnostics

Checking for the ARCH effect:

- Using the Ljung-Box Test, the mean model residuals were not serially correlated with a $p-$value of 0.9028.
- Using the same test, the squares of the mean model residuals were serially correlated with a $p-$value less than 2.2×10^{-16}.

Mean Model Diagnostics

Checking for the ARCH effect:

- Using the Ljung-Box Test, the mean model residuals were not serially correlated with a $p-$value of 0.9028.
- Using the same test, the squares of the mean model residuals were serially correlated with a $p-$value less than 2.2×10^{-16}.
- Hence, the residuals of the mean model exhibit ARCH effect.

GARCH Model

GARCH Model:

- Generalized autoregressive conditionally heteroscedastic model

GARCH Model

GARCH Model:

- Generalized autoregressive conditionally heteroscedastic model
- Adds another parameter s to the ARCH model which takes into account the last s conditional variances (σ_t)

Grid Search for Best Volatility Model

- A grid search was used to find suitable values for m and s, the parameters of the GARCH model.
- Values of $1, \ldots, 4$ were assigned to m, while $0, \ldots, 4$ were assigned to s.

Grid Search Results

(m, s)	LLH	AIC	BIC	SIC	HQIC
(3,1)	16104.00	-3.6747	-3.6674	-3.6747	-3.6722
(1,0)	14599.49	-3.3318	-3.3270	-3.3318	-3.3302
(4,2)	13994.79	-3.1926	-3.1838	-3.1926	-3.1896
(3,4)	13620.50	-3.1070	-3.0973	-3.1070	-3.1037
(2,1)	13154.61	-3.0015	-2.9950	-3.0015	-2.9993

Table 3: Top 5 AR(3)-GARCH(m, s) models

- The best volatility model among all the metrics is GARCH(3,1).

Outline

Best Model Equation

The best model for the water level data is the AR(3)-GARCH(3,1) model with the equation as follows:

$$\begin{cases} X_t = 0.0065 + X_{t-1} - 0.0190X_{t-2} - 0.0038X_{t-3} + Y_t \\ Y_t = \sigma_t \epsilon_t \\ \sigma_t^2 = 3.078 \times 10^{-7} + 0.3933Y_{t-1}^2 + 10^{-8}Y_{t-2}^2 + 10^{-8}Y_{t-3}^2 + 0.9099\sigma_{t-1}^2 \end{cases}$$

where $\epsilon_t \overset{\text{iid}}{\sim} (0, 1)$.

Significance of the Best Model's Coefficients

Coefficient	$p-$value
μ	$< 2.2 \times 10^{-16}$
ϕ_1	$< 2.2 \times 10^{-16}$
ϕ_2	0.329
ϕ_3	0.849
α_0	NaN
α_1	$< 2.2 \times 10^{-16}$
α_2	1
α_3	1
β_1	$< 2.2 \times 10^{-16}$

Table 4: The p-value of the AR(3)-GARCH(3,1) coefficients

Refining the Model

Metric	AR(3)-GARCH(3,1)	AR(1)-GARCH(1,1)
LLH	16104.00	14039.95
AIC	-3.6747	-3.2043
BIC	-3.6674	-3.2003
SIC	-3.6747	-3.2043
HQIC	-3.6722	-3.2029

Table 5: Comparison between the original and refined models

- The insignificant coefficients can be removed, yielding an AR(1)-GARCH(1,1) model. Its comparison with the original model in terms of different metrics is found in Table 5.

Refining the Model

Metric	AR(3)-GARCH(3,1)	AR(1)-GARCH(1,1)
LLH	16104.00	14039.95
AIC	-3.6747	-3.2043
BIC	-3.6674	-3.2003
SIC	-3.6747	-3.2043
HQIC	-3.6722	-3.2029

Table 5: Comparison between the original and refined models

- Compared to the original, the refined model has slightly higher information criteria and lower log likelihood. Since the differences in the metrics were minimal, based on the principle of parsimony, the refined model was selected for forecasting.

Refining the Model

After removing the insignificant coefficients, the refined model is AR(1)-GARCH(1,1) with equation as follows:

$$\begin{cases} X_t = 0.0312 + 0.9495X_{t-1} + Y_t \\ Y_t = \sigma_t \epsilon_t \\ \sigma_t^2 = 0.0006 + 0.4580Y_{t-1}^2 + 0.5495\sigma_{t-1}^2 \end{cases}$$

where $\epsilon_t \overset{\text{iid}}{\sim} (0,1)$.

Significance of the Refined Model's Coefficients

Coefficient	$p-$value
μ	$< 2.2 \times 10^{-16}$
ϕ_1	$< 2.2 \times 10^{-16}$
α_0	$< 2.2 \times 10^{-16}$
α_1	$< 2.2 \times 10^{-16}$
β_1	$< 2.2 \times 10^{-16}$

Table 6: The p-value of the coefficients

- Now, the refined model's coefficients are all significant.

Checking Standardized Residuals

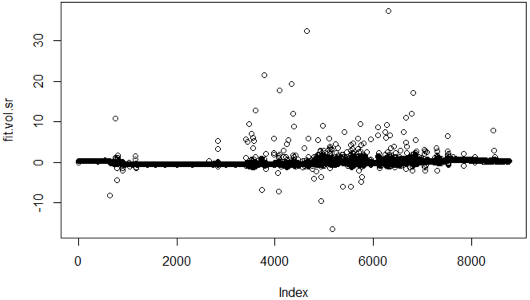


Figure 6: AR(1)-GARCH(1,1) standardized residuals.

- Using the Ljung-Box Test, the residuals are serially correlated since the $p-$value is less than 2.2×10^{-16}.

Checking Standardized Residuals

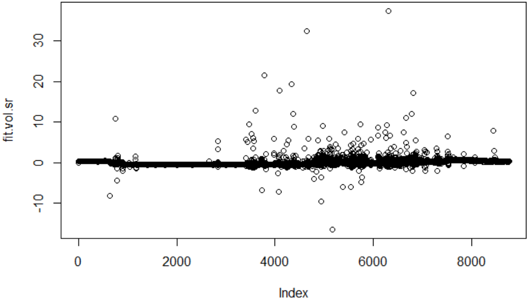


Figure 6: AR(1)-GARCH(1,1) standardized residuals.

- Using the Ljung-Box Test, the residuals are serially correlated since the $p-$value is less than 2.2×10^{-16}.
- Using the same test, the squared residuals are uncorrelated since the $p-$value is 0.9178.

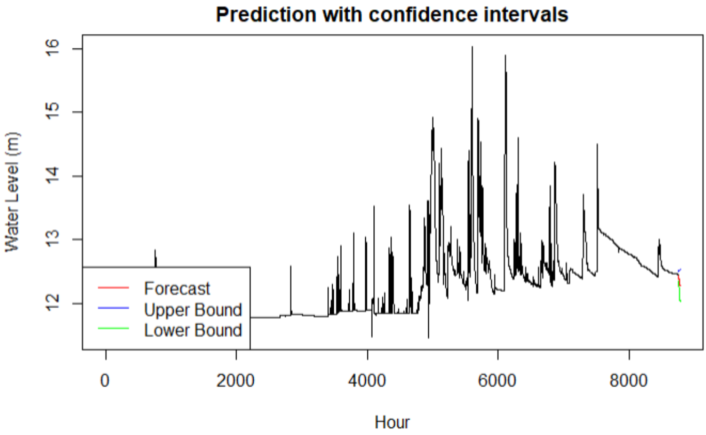
Correlated Standardized Residuals

- Because of the correlated standardized residuals, there are still patterns or structures in the data that the model has not fully captured.

Correlated Standardized Residuals

- Because of the correlated standardized residuals, there are still patterns or structures in the data that the model has not fully captured.
- The residuals are also not independent so it can lead to less accurate volatility forecasts as the model is not fully capturing the dynamics of the process.

Forecasting (24 Hours)



Forecasting (24 Hours)

Prediction	Mean Error	σ^2
12.4283	0.0384	0.0015
12.4173	0.0584	0.0021
12.4067	0.0760	0.0027
12.3967	0.0924	0.0033
12.3873	0.1079	0.0040
12.3783	0.1228	0.0046
12.3697	0.1372	0.0052
12.3616	0.1511	0.0059

Table 7: AR(1)-GARCH(1,1) Forecasts

Forecasting (24 Hours)

Prediction	Mean Error	σ^2
12.3539	0.1646	0.0065
12.3466	0.1777	0.0072
12.3396	0.1905	0.0078
12.3330	0.2029	0.0085
12.3268	0.2151	0.0091
12.3208	0.2269	0.0098
12.3152	0.2386	0.0105
12.3098	0.2499	0.0112

Table 8: AR(1)-GARCH(1,1) Forecasts

Forecasting (24 Hours)

Prediction	Mean Error	σ^2
12.3047	0.2611	0.0118
12.2999	0.2720	0.0125
12.2953	0.2827	0.0132
12.2909	0.2932	0.0139
12.2868	0.3036	0.0146
12.2829	0.3137	0.0154
12.2791	0.3237	0.0161
12.2756	0.3336	0.0168

Table 9: AR(1)-GARCH(1,1) Forecasts

Future Considerations

- Adding more GARCH or ARCH terms

Future Considerations

- Adding more GARCH or ARCH terms
- Including exogenous variables

Future Considerations

- Adding more GARCH or ARCH terms
- Including exogenous variables
- Exploring more complex GARCH-family models.

Time Series Forecasting using ARMA-GARCH Models: Sto. Niño, Marikina River Water Level Prediction

Alegarbes, Brynx Junil
Que, Neil Bryant
Tan, Jeremy Marcus

MATH 271.2

Ateneo de Manila University

11 July 2024