International Journal on Advanced Science Engineering Information Technology

Vol.11 (2021) No. 5
ISSN: 2088-5334

Simulation of Autoregressive Integrated Moving Average- Generalized Autoregressive Conditional Heteroscedasticity (ARIMA-GARCH) to Forecast Traffic Flow

Challiz D. Omorog [a,*]

[a] *College of Computer Studies, Camarines Sur Polytechnic Colleges, Nabua, Camarines Sur, 4434, Philippines*
*Corresponding author: *challizomorog@cspc.edu.ph*

Abstract—**Modeling the unprecedented traffic flow data generated by Intelligent Transportation Systems can boost the innovation-capacity of the transportation management systems to drive informed decision-making. Thus, this paper attempts to simulate traffic forecasting techniques that can be adopted in the Philippines to make fact-based decisions into accurate and effective traffic management schemes. In this research, a schematic framework is introduced organized into three stages (Preprocessing, Model Identification and Estimation, and Model Checking) sequentially arranged to comprehensively estimate the best-appropriate model to forecast traffic flow using ARIMA and GARCH models. The Model Identification and Estimation is the conditional stage in the framework that pre-determines if hybrid modeling is necessary based on the given datasets. Various accuracy metrics are also used to find the "best" model and select the optimal values for ARIMA and GARCH models. The proposed framework is simulated in R Programming using the vehicular traffic flow datasets at North Avenue, EDSA northbound, Manila, Philippines. The resulting models, consist of the best fit ARIMA (1,1,3) and GARCH (1,2), are combined as the hybrid model and compared using its prediction results. Based on the visual simulation data, the prediction accuracy result of the ARIMA model outperforms the combined ARIMA-GARCH model given the actual data. Conclusively, the simulation performance provides proof to suggest that the forecasting models are timely tools to predict future traffic flow and aid in making better traffic inventions and schemes.**

Keywords— **ARIMA; GARCH; hybrid algorithm; traffic prediction algorithm.**

Manuscript received 2 Feb. 2021; revised 21 Apr. 2021; accepted 1 Jun. 2021. Date of publication 31 Oct. 2021.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

The application of ICT in the transport sector or the so-called intelligent transportation system (ITS) is strong evidence that these technologies are seen as the top relevant solutions to the challenges faced by the transport system. ITS is an advanced mechanism that integrates ICT, and road transport, and traffic management [1]. Examples of ITS are the RFID [2], smart cameras for traffic signals [3], GPS [4], IoT-based real-time traffic monitoring systems [5], and other traffic control systems that promise traffic efficiency and mobility [6]. As ITS usage begins to multiply, the traffic data generated have also raised exponentially to big data [7]. Unfortunately, the aspect of historically analyzing the potential of this vast decaying relevancy of traffic data into actionable information remains unpopular [8].

Driven mostly by advanced countries, ITS is also becoming increasingly popular in many developing countries. The Philippines followed the rest of the world to solve its traffic woes. Its government installed ITS around the major thoroughfares of Metro Manila to usher congestion-free roadways. But regardless, traffic congestion has become more intricate and difficult [9], [10]. Manila's traffic ranked second-worst in the world according to TomTom Traffic Index since 2019.

Despite these challenges, the Philippine government seemed slow to recognize potential solutions overclouding the transport system. For decades, its traffic administrators focused persistently on expanding roadways and using ICT-based solutions to mitigate traffic volume. One emerging solution is the recent innovation in ITS, such as big data and analytics, to address the traffic volume challenges [11]. The key technology uses the voluminous traffic data in undertaking solutions to understand and predict the future traffic flow. Traffic prediction is perceived as the key solution to revolutionize the current transportation landscape [12], [13] ensuring benefits even to developing countries. However, to capture the full potential of these technologies, the government needs to implement holistic approaches and invoke huge cooperation from other agencies. Academia, too, has been challenged to collaborate with the government to find appropriate solutions in this significant opportunity [14].

One particular technique on big data analytics that is now gaining attention across numerous practical disciplines is the time series (TS) method. TS analysis is the process of measuring observations over time to predict future values. However, diversified predictions using various traffic datasets and approaches make it difficult to clearly recognize the prospects and limitations of TS [15]. In this paper, extensive TS experiments are carried out using the traffic data in Metro Manila. This paper aims to give insight into finding the best-appropriate traffic forecasting technique that can be adopted to make fact-based decisions into accurate and effective traffic management schemes in the Philippines. A schematic framework was also introduced using the Univariate ARIMA-GARCH model to predict traffic flow. The paper also applied different performance indexes for model adequacy to identify the best-fit model.

II. Materials and Method

A. The ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is fit for the non-stationary and non-seasonality data series. This simply means that an ARIMA model is considered when equally spaced TS of data exhibits patterns and random walk series (non-stationary series). Three (3) components characterize the ARIMA model: p (autoregressive parameter), d (no of times series is differenced), and q (moving average parameter) in general form as ARIMA(p,d,q). The mathematical form of ARIMA (p, d, q) is written as follows [16]:

$$\emptyset_P(B)(1-B)^d(y_t-\mu)=\theta_q(B)\varepsilon_t \quad (1)$$

where $\emptyset_P(B)=\left(1-\alpha_1 B-\alpha_2 B^2-\ldots-\alpha_p B^p\right)$, $\theta_q(B)=\left(1-\beta_1 B-\beta_2 B^2-\ldots-\beta_q B^q\right)$ are polynomial notations in terms of the Backward Shift Operator, B of p and q, and ε_t is the white noise series.

B. The GARCH Model

Since the ARIMA models do not model conditional heteroscedasticity, the GARCH with the p and q parameters solves this problem. The GARCH (p,q) model is indicated by the notation [17] :

$$h_t=\alpha_0+\sum_{i=1}^{q}\alpha_i a_{t-i}^2+\sum_{j=1}^{p}\beta_j h_{t-j} \quad (2)$$

where h_t is the conditional variance, and α_i and β_j are parameters of the model. In the equation, $\{a_t\}$ is a generalized autoregressive conditionally heteroscedastic model of order p, q, denoted by GARCH(p,q).

C. Accuracy Metrics of TS Forecast

To fully evaluate the potential of the three forecasting approaches, the performance criterion needs to be established to measure the discrepancy between the actual fully and predicted values. The ARIMA and GARCH performances are evaluated using the following (1-2) error criteria and (3) selection criteria, respectively:

1) Mean Absolute Error (MAE): measures how the prediction varies from actual values without considering their direction, given as [18]:

$$MAE=\frac{1}{n}\sum_{j=1}^{n}\left|y_j-\hat{y}_j\right| \quad (3)$$

2) Mean Squared Error (MSE): also measures the magnitude of error between actual and prediction values by taking the square root of the average squared errors. Doing so magnifies which model generally holds the highest error point. The formula is given as [18].

$$MSE=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j-\hat{y}_j\right)^2} \quad (4)$$

3) Akaike Information Criterion (AIC): used for selection and parameter estimation. According to this criterion, the best fit model should be the one with the smallest value based on the formula [19].

$$AIC=-2\log(L)+2m \quad (5)$$

D. The Proposed Hybrid Model

This paper proposes sequentially using ARIMA and GARCH models and then examining the hybrid's performance against its non-hybrid counterparts in forecasting traffic flow. The logic behind this scheme is to avoid biased estimation when modeling time-dependent volatility and nonlinearity TS data with ARIMA. Thus, an extension approach using GARCH was embedded in the framework of ARIMA to accommodate this problem, which is similar to the work of Yaziz [16], as shown in Figure 1 below.

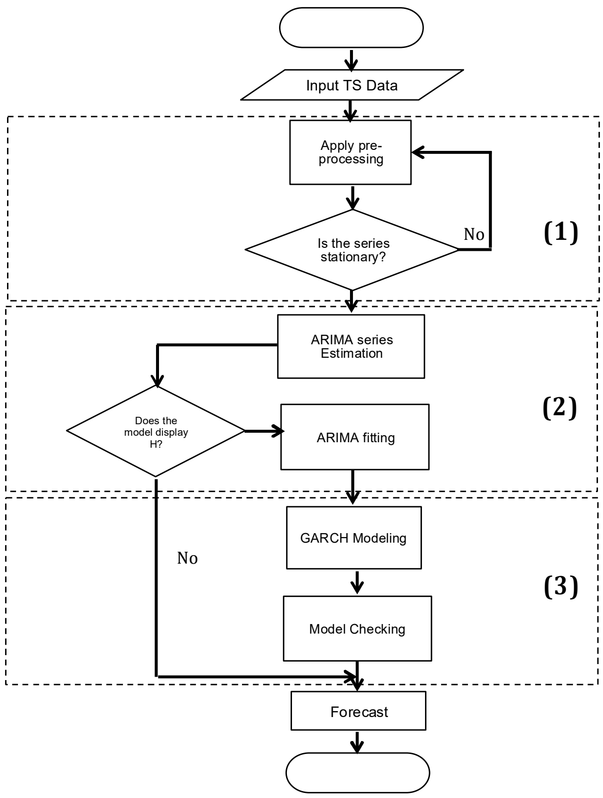

Fig. 1 Schematic representation of the Hybrid ARIMA-GARCH Model

GARCH is best used to model heteroscedasticity in time series data, such as the increasing and decreasing volatility. Therefore, combining these two models overcomes each component's limitations to better characterize the statistical features of the TS data, in this case, the traffic flow series. The flowchart in Figure 1 outlines the procedure fitting of the proposed hybrid model.

The hybrid model framework is divided into three (3) stages identified as (1) Preprocessing, (2) Model Identification and Estimation, and (3) Model Checking. The sequence of stages for the hybrid framework should be observed correspondingly, from the first stage until the last, especially when a cluster of volatility is detected in the TS plots (from stage 2). This is to avoid biased estimates and wrong forecasts. The framework's major highlight is particularly the GARCH model incorporated in the last stage- Model Checking. The addition of GARCH notation in the process is necessary to check whether conditional variance exists in the series. Moreover, this added process could potentially generate and reveal novel learning in the estimation given the time-varying volatility in the TS.

The first stage of the conceptual model is called the Preprocessing labeled as (1). The task is to make the TS data stationary. Here, the TS data is differenced *d* number of times depending on the complexity of the series, also referred to as the "Apply preprocessing" in the diagram. Therefore, the value of *d* is the number of differencing needed until the series is stationary.

After transforming the TS data to stationary, the next stage is the (2) Model Identification and Estimation or fitting a satisfactory ARIMA model. This means that the order values of *p* and *q* parameters are estimated from the ACF and PACF plots. Assuming that the optimal model is identified, the next step is to detect periods of conditional heteroscedasticity in the model. If there is no heteroscedastic series in the model, then no further modeling is needed except to plot the forecast. However, if there is heteroscedastic series in the model, proceed to the last phase as described in the framework to find a more appropriate model.

If the model still displays high and low volatility via successive lags, this suggests that the GARCH model fitting is appropriate. GARCH fitting is done in the third and last stage- (3) Model Checking. This approach requires the use of AIC to compare models with different orders *p* and *q*. Finally, the chosen model is subjected to model checking to ensure the correctness of the fitted model. If its P-value is statistically significant, residuals appear to be low and unstructured, and randomly distributed then the model fits well. After this point, the model can now be used for forecasting.

E. *Research Design*

This exploratory-qualitative research primarily deals with validating the best-fit model to forecast traffic flow using the TS dataset and analysis procedures. An exploratory technique was employed to describe the behavior of the models- ARIMA, GARCH, and the hybrid ARIMA-GARCH as predictive models. On the other hand, the qualitative method presents the results to understand which model is adequate and approximate to predict the future vehicular traffic volume.

F. *Research Setting*

The study relied on the synthetic traffic data provided by the Department of Transportation (DOTr) thru Advanced Science and Technology Institute- Department of Science and Technology (ASTI.DOST), which is the collection of the monthly vehicular traffic flow at North Avenue, EDSA northbound denoted by the circled star in the actual road web screenshot (from https://roadsafety.gov.ph/#!/map) displayed in Figure 2.

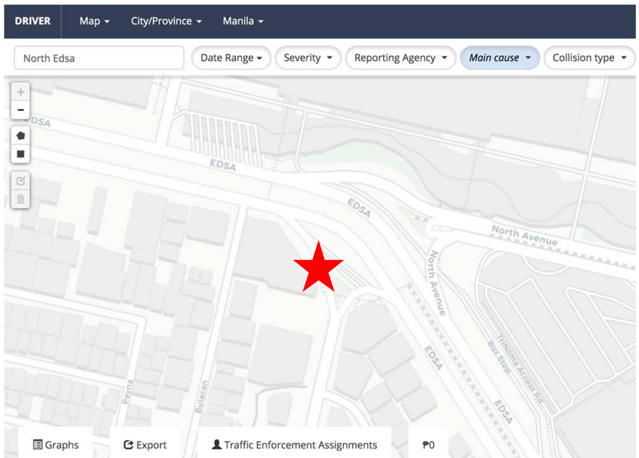

Fig. 2 North Avenue EDSA Northbound Screenshot

The sampling time period is about 15 years or 177 months. The data include vehicle flow, which denotes the total number of vehicles that traversed the defined roadway per month. Since vehicle flow varies depending on which day of the week it is, the dataset was also delimited to weekdays only. The modeling has adopted R programming to build and predict the traffic flow datasets. R is a powerful open-source programming language driven by the concept of data analysis with applications, which covers fields such as statistical analysis, data mining, graphical display, and so on [20].

G. *Research Procedures*

This paper adopted the schematic procedures to predict the proposed Hybrid ARIMA- GARCH model presented in Figure 2. The steps are as follows:

1) Processing Stage: The first process converts the traffic flow series to achieve stationarity in the mean and variance through differencing and log transformation, respectively [21]. This is important because most statistical methods are based on this assumption that time series modeling can only be meaningful when stationary time series is achieved [22]. We performed the differencing in R to convert the non-stationary into a stationary residual series using the mathematical test of Dickey and Fuller [16]. A stationary series is relatively easy to identify because its statistical property, such as its p-value, should not be greater than 0.05.

2) Model Identification and Estimation Stage: The estimation of the autocorrelation graph of the TS must be established in this process to determine the "best" appropriate ARIMA model. Finding the most appropriate values of p and q is typically identified using the Autocorrelation function (ACF) and partial autocorrelation function (PACF). In this paper, the ACF and PACF were performed both in the lag and differenced series. We follow the principle of parsimony in considering the best model, which means the model candidate with the fewest parameters is best [23]. The following rules were also considered in this paper [24]: If ACF cut off after lag n, PACF dies down: ARIMA (0,d,n) ➔identify MA(q); If ACF dies down, PACF cut off after lag n: ARIMA (n,d,0) ➔identify AR(q); and If ACF and PACF die down; mixed ARIMA model, need differencing.

A choice of 0 to 2 for p and q was tested for both the datasets (lag and differenced) for comparison purposes. We compare and understand its performance based on the evaluation criteria results of MAE and MSE. The criteria result with the lowest prediction error of MAE and MSE generated is the best prospective model candidate [16].

3) Model Checking Stage: Validating the model to determine if no significant heteroscedasticity exists is typically done through a test using the residuals of the TS dataset. Ljung-Box is a test that determines the existence of heteroscedasticity in the time series by verifying if autocorrelations are different from 0. This means that we accept the null hypothesis H_0 when the probability of the p-value is less than the chosen level (in our case 5%), which states that there is no heteroscedasticity in the series (the series is independent and uncorrelated) and skip the next process; otherwise, proceed with the modification process because fragments of serial correlation still exist in the series.

Another way to confirm if the residuals contain heteroscedasticity is that we check and plot the square residuals of ACF & PACF in R. Patterns in the residuals confirm volatility or conditional variance in the series when the plot shows a cluster of volatility at some points in time. ACF seems to die down or PACF cuts off after a certain lag even though some remaining lags are significant.

To model the conditional variance of the series, we need to fit the GARCH to the residuals of the ARIMA model candidates then calculate the log-likelihood using the logLik () function in R. Since GARCH (1,1) model is the simplest and most successful of the family of the volatility models [24], in this paper, the variance of the series was modeled using GARCH(1,1). After successfully building and fitting the ARIMA-GARCH model, AIC result was considered to check and confirm the best candidate model. The model with the lowest AIC was selected, as provided in the R result.

III. Results and Discussion

A. Preprocessing Stage

The first thing to do is to check for stationarity in the time series model based on the framework. Hence, the modeled non-stationary and stationary data are shown in Figure 3 and Figure 4, respectively. Augmented Dickey-Fuller (ADF) tests in R were performed to confirm stationarity in the time series dataset. The plotted graph in Figure 3 shows that a trend exists, which strongly indicates non-stationarity in the mean with a *p*-value of 0.568 of ADF test. Thus, the TS dataset needs to be differenced to remove trends and obtain a stationary series.

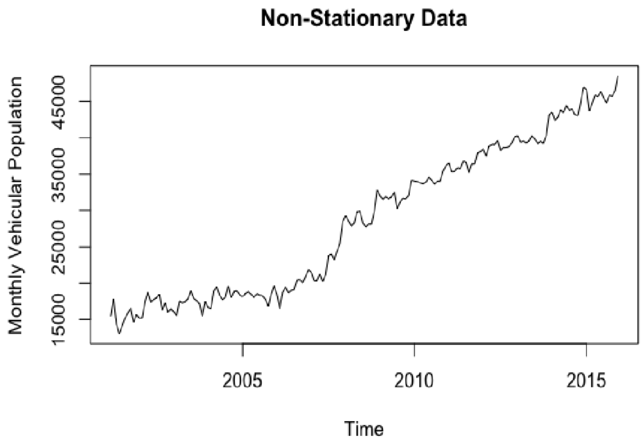

Fig. 3 Non-stationary Traffic Flow Data

The graph plotting in Figure 4 now illustrates a stationary series transformed through differencing. The result suggests stationarity since most of the data are located around the mean zero. The differenced series is also supported by the ADF test with a *p*-value of 0.01. Here, the value of $d = 1$ since the first order differencing performed in the TS series is already adequate.

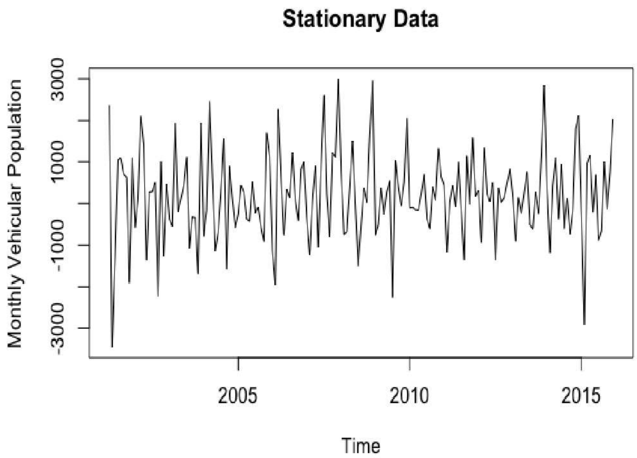

Fig. 4 Stationary Traffic Flow Data

B. Model Identification and Estimation

To successfully determine the best ARIMA model, it is necessary to carry out the ACF and PACF analysis, which include determining the order of *p*, and *q,* and making sure that the residuals of the model are random or do not exhibit seasonality. In this paper, we now set $d = 1$ with reference to the differencing performed in stage one to create a uniform and balanced comparison for all the models. The ACF and PACF plots are illustrated in Figures 5- 7.

The left graph in Figure 5 shows the ACF while on the right plots the PACF of the monthly traffic flow data. On these plots, the ACF slowly decreases (but not die down) while the PACF is significant in the first lag only. The plots suggest that it needs differencing since there are no significant or interpretable peaks at certain lags, which validates the differencing performed in stage one.

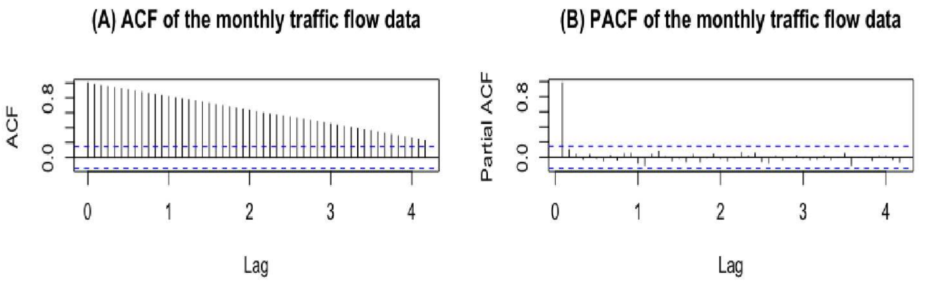

Fig. 5 (A) ACF and (B) PACF of the non-stationary monthly traffic flow data

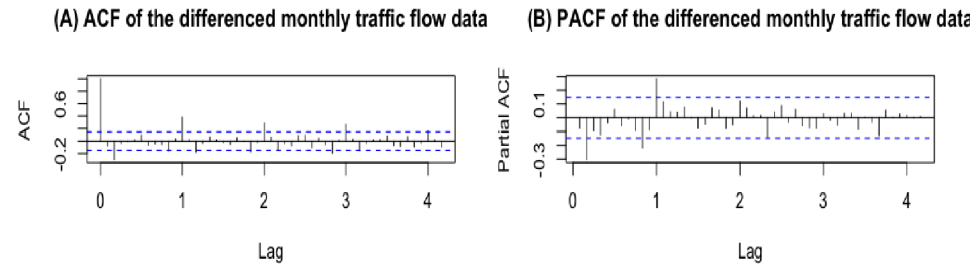

Fig. 6 (A) ACF and (B) PACF of the differenced (stationary) monthly traffic flow data

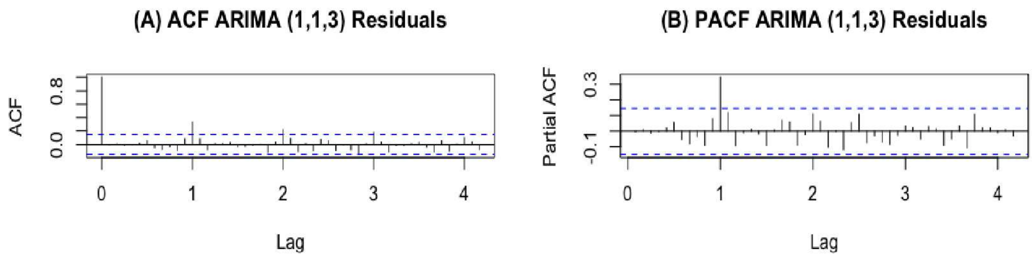

Fig. 7 (A) ACF and (B) PACF residuals of the selected ARIMA models

The graph in Figure 6-A shows regular spikes at every 12th period (months). These are the lags 1, 2, and 3, which corresponds to this being monthly data. According to ACF function, if the current and past values (lags) are consistently dependent, then the concept of correlation is positive. That is, when the autocorrelation bar is longer than the blue marker, then the correlation is considered significant and indicates a strong correlation between each value and the value occurring two points previously, so every 12 periods is correlated to the previous 12 periods. As can be noticed, spikes appear every December, and the spikes go beyond the 2 σ region. This suggests that the high volume of traffic is non- auto-correlated but data-dependent. Also, the ACF in Figure 6-A tails off quickly; hence it is stationary.

In terms of PACF in Figure 6-B, which suggests the order of conditional autocorrelation in the series, in this case, there are significant spikes in the plot concentrated in the early periods while lesser in the later periods. The plots both follow a geometric decay, which is exactly the expectation of the ACF and PACF plots for an ARMA process as explained in [24]. This indicated that it is sufficient to carry out an ARIMA fitting.

TABLE I
PERFORMANCE OF THE PROSPECTIVE ARIMA PREDICTION MODELS

Model	Ljung-Box (*p*)	MAE	MSE	AIC
ARIMA(1,1,0)	0.7645	805.66	1065.24	2976.03
ARIMA(2,1,0)	0.6006	770.60	1010.28	2959.51
ARIMA(3,1,0)	0.7637	769.14	1005.51	2959.86
ARIMA(0,1,1)	0.3638	791.74	1059.65	2974.21
ARIMA(0,1,2)	0.9794	758.96	998.33	2955.38
ARIMA(0,1,3)	0.8872	759.44	997.86	2957.22
ARIMA(1,1,1)	0.0953	809.40	1064.93	2977.93
ARIMA(1,1,2)	0.9121	759.26	997.95	2957.25
ARIMA(1,1,3)	0.9483	750.77	992.53	2957.64
ARIMA(2,1,1)	0.9697	764.34	999.66	2957.85
ARIMA(2,1,2)	0.8933	761.27	997.23	2959.01
ARIMA(2,1,3)	0.8935	761.16	997.24	2961.01
ARIMA(3,1,1)	0.9174	764.17	997.96	2959.26
ARIMA(3,1,2)	0.8954	761.26	997.23	2961.01
ARIMA(3,1,3)	0.899	752.49	992.57	2961.58

In fitting the ARIMA model, the idea of parsimony is considered in which the model can only be considered qualified on the basis of the following criteria: (1) low MSE to ensure optimum accuracy, (2) low MAE also to provide forecast accuracy.

Table 1 displays the various prospective ARIMA models selected for analysis, in this case, based on the given criteria. To make a clear comparison, the standard ARIMA model of the same order is performed with the same datasets. Based on the result in R, ARIMA (0,1,2), ARIMA (0,1,3), and ARIMA (3,1,3) models obtained promising values, but ARIMA (1,1,3) was considered best suited for estimation based on low MAE and MSE.

C. Model Checking

Since ARIMA only models linear time series, the residual of the data is analyzed to check if volatility or nonlinear behavior exists. This step shall ensure that all nonlinear structures are reflected and addressed accordingly. In order to model volatility, the residual of the chosen ARIMA model is analyzed and modeled using GARCH.

The ACF and PACF of ARIMA (1,1,3) residuals in Figure 7 show that significant lags exist in the plots. The time series plot of residuals shows some cluster of volatility, particularly at lag 1, evident in both the plots. These findings were ascertained by the Ljung-Box test result in column 2 of Table 2, particularly for ARIMA (1,1,3) with a *p*-value of 0.9483, which is > 5% level of significance. Hence, to model the volatility state of ARIMA (1,1,3) model to GARCH is justifiable.

The models with favorable results for the above criteria were further narrowed down for selection based on the lowest AIC because the lower the AIC value the better since it indicates stronger evidence of a better fit. Comparing the AICs in Table 2, the GARCH (1,2) model was chosen out of the prospective GARCH models since it has the least value of AIC and therefore best suited for estimation.

TABLE II
PERFORMANCE OF THE PROSPECTIVE GARCH PREDICTION MODELS

Model	Log-likelihood	AIC
GARCH(1,1)	-1106.61	1.125236
GARCH(1,2)	-1104.36	1.123964
GARCH(2,1)	-1106.97	1.126617
GARCH(2,2)	-1104.35	1.124977

For this analysis, ARIMA (1,1,3) was chosen out of the 4 prospective ARIMA models while GARCH(1,2) of the GARCH models. At this juncture, we will compare the results of ARIMA (1,1,3), and the combined ARIMA (1,1,3) and GARCH (1,2) model for forecasting.

After successfully building and identifying the best prospective model as shown in Tables 1 and 2, the information is used to visualize and compare the trends. To make a clear comparison, the selected ARIMA (1,1,3), and the combined ARIMA (1,1,3)-GARCH (1,2) were built and used for prediction together with the actual data as presented in Figure 8. The black line is the actual data with vehicular flow data from March 2001- June 2015 only. We reduced the dataset by six (6) months or only 97% of the data for model forecasting to visually appreciate the accuracy result of the prospective prediction models. As can be noted in the plot, the forecast of the two prospective models is for the next two-years only (from 2015), since we only consider 24 months prediction. The forecast result of ARIMA (1,1,3) bears a close resemblance to the actual value, whereas the hybrid ARIMA(1,1,3)-GARCH (1,2) shows that it failed to follow the actual data trend.

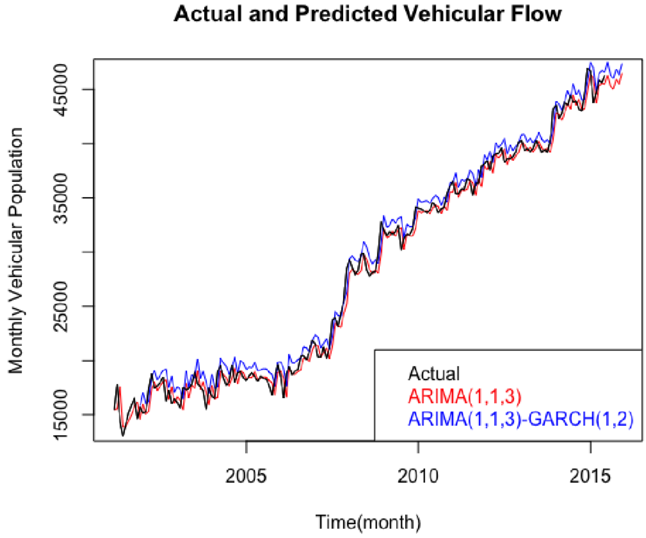

Fig. 8 Actual and Predicted Monthly Vehicular Population for North EDSA

IV. CONCLUSION

The large volume of traffic data available has been evocative and appealing to researchers and alike. Hence, this study mined the monthly vehicular traffic datasets at North Avenue, EDSA northbound, Manila, Philippines, to examine the performance of the hybrid univariate time series ARIMA-GARCH models as a prospective method to manage and forecast traffic flow. In consideration of the proposed framework and the Ljung-Box test results and ACF and PACF residual lags of the sample data, conditional variance or heteroscedasticity is evident. Hence, applying the GARCH method is justifiable before prediction.

In conclusion, the most efficient and suitable model is ARIMA (1,1,3) based on low MAE and MSE, and the actual proximity or prediction accuracy as against the actual data. The combined ARIMA-GARCH model may have relatively exhibited promising prediction projecting the conditional mean and variance simultaneously, yet the result is incomparable with the best-selected ARIMA model.

For future study, validating the hybrid model's accuracy is encouraged to quantify its forecasting potential further. Aggregating the traffic flow data to 5, 10, and 15 minutes intervals to better measure the traffic performance is a nice challenge. Another interesting direction would be to research the interdependency of adjacent or neighboring roads to traffic flow, inclement weather, or traffic accidents. Adding these factors can better quantify and improve prediction accuracy, possibly using multivariate time series.

Moreover, this research opens up timely conjectures alongside the available ITS facilities as effective forecasting tools for better traffic interventions and improvement. Likewise, it is highly suggested that the Philippines consider starting a variety of open access platforms on which real-time traffic datasets are available to support initiatives in scholarly works. Doing so will help entice and encourage many local and international researchers to develop more traffic-related technologies and studies.

REFERENCES

[1] A. Sumalee and H. W. Ho, "Smarter and more connected: Future intelligent transportation system," *IATSS Research*, vol. 42, no. 2. Elsevier B.V., pp. 67–71, Jul. 2018, doi: 10.1016/j.iatssr.2018.05.005.

[2] A. Atta, S. Abbas, M. A. Khan, G. Ahmed, and U. Farooq, "An adaptive approach: Smart traffic congestion control system," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 9, pp. 1012–1019, Nov. 2020, doi: 10.1016/j.jksuci.2018.10.011.

[3] W. C. Tchuitcheu, C. Bobda, and M. J. H. Pantho, "Internet of smart-cameras for traffic lights optimization in smart cities," *arXiv*, vol. 11. arXiv, p. 100207, Feb. 2020, doi: 10.1016/j.iot.2020.100207.

[4] N. Chan Soe and T. Lai Lai Thein, "Haversine Formula and RPA Algorithm for Navigation System," *Int. J. Data Sci. Anal.*, vol. 6, no. 1, p. 32, Feb. 2020, doi: 10.11648/j.ijdsa.20200601.14.

[5] M. Sarrab, S. Pulparambil, and M. Awadalla, "Development of an IoT based real-time traffic monitoring system for city governance," *Glob. Transitions*, vol. 2, pp. 230–245, Jan. 2020, doi: 10.1016/j.glt.2020.09.004.

[6] S. Felici-Castell, M. García-Pineda, J. Segura-Garcia, R. Fayos-Jordan, and J. Lopez-Ballester, "Adaptive live video streaming on low-cost wireless multihop networks for road traffic surveillance in smart cities," *Futur. Gener. Comput. Syst.*, vol. 115, pp. 741–755, Feb. 2021, doi: 10.1016/j.future.2020.10.010.

[7] H. Shin, J. Jung, and Y. Koo, "Forecasting the Video Data Traffic of 5G Services in South Korea," *Technol. Forecast. Soc. Change*, vol. 153, p. 119948, Apr. 2020, doi: 10.1016/j.techfore.2020.119948.

[8] Y. Lian, G. Zhang, J. Lee, and H. Huang, "Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles," *Accid. Anal. Prev.*, vol. 146, p. 105711, Oct. 2020, doi: 10.1016/j.aap.2020.105711.

[9] M. Rith, K. I. D. Z. Roquel, N. S. A. Lopez, A. M. Fillone, and J. B. M. M. Biona, "Towards more sustainable transport in Metro Manila: A case study of household vehicle ownership and energy consumption," *Transp. Res. Interdiscip. Perspect.*, vol. 6, p. 100163, Jul. 2020, doi: 10.1016/j.trip.2020.100163.

[10] Y. Boquet, "Managing Metro Manila," in *Springer Geography*, Springer, 2017, pp. 567–615.

[11] S. Kaffash, A. T. Nguyen, and J. Zhu, "Big data algorithms and applications in intelligent transportation system: A review and bibliometric analysis," *Int. J. Prod. Econ.*, vol. 231, p. 107868, Jan. 2021, doi: 10.1016/j.ijpe.2020.107868.

[12] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road Traffic Forecasting: Recent Advances and New Challenges," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2. Institute of Electrical and Electronics Engineers, pp. 93–109, Jun. 2018, doi: 10.1109/MITS.2018.2806634.

[13] M. A. Ansari and T. Arundhathi, "Street Traffic Forecasting Ongoing Advances and New Challenges," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 3, pp. 650–656, Mar. 2019, doi: 10.26438/ijcse/v7i3.650656.

[14] M. Lippi, M. Bertini, and P. Frasconi, "Short-term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, 2013, doi: 10.1109/TITS.2013.2247040.

[15] X. Fu, W. Luo, C. Xu, X. Zhao, and F. J. Hwang, "Short-Term Traffic Speed Prediction Method for Urban Road Sections Based on Wavelet Transform and Gated Recurrent Unit," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/3697625.

[16] S. R. Yaziz, N. A. Azizan, R. Zakaria, and M. H. Ahmad, "The Performance of Hybrid ARIMA-GARCH Modeling in Forecasting Gold Price," *20th Int. Congr. Model. Simulation, Adelaide, Aust.*, no. December, pp. 1–6, 2013.

[17] L. Seymour, P. J. Brockwell, and R. A. Davis, *Introduction to Time Series and Forecasting.*, vol. 92, no. 440. 1997.

[18] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," *L. Lambert Acad. Publ. Ger.*, p. 67, Feb. 2013.

[19] H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo, "A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics," *Appl. Intell.*, vol. 48, no. 8, pp. 2429–2440, Aug. 2018, doi: 10.1007/s10489-017-1095-9.

[20] B. K. C. Chan, "Data analysis using R programming," in *Simultaneous Mass Transfer and Chemical Reactions in Engineering Science*, Elsevier, 2020, pp. 39–60.

[21] M. Guidolin and M. Pedio, "Single-Factor Conditionally Heteroskedastic Models, ARCH and GARCH," in *Essentials of Time Series for Financial Applications*, Elsevier, 2018, pp. 151–228.

[22] S. Yousefzadeh-Chabok, F. Ranjbar-taklimie, R. Malekpouri, and A. Razzaghi, "A Time Series Model for Assessing the Trend and Forecasting the Road Traffic Accident Mortality," *PubMed Cent.*, no. 14, pp. 0–5, 2016, doi: 10.5812/atr.36570.Research.

[23] A. Coghlan, *A Little Book of R For Time Series Release 0.2*. Cambridge, UK, 2018.

[24] L. Pham, "Time Series Analysis with ARIMA – ARCH / GARCH model in R," 2013.