International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.11, No.1/2, May 2022

# ANALYSIS AND FORECASTING OF ELECTRICITY DEMAND IN DAVAO DEL SUR, PHILIPPINES

Samuel John E. Parreño

Department of Teacher Education, University of Mindanao Digos College, Philippines

## ABSTRACT

*The Box-Jenkins method or the Autoregressive integrated moving average model is a famous technique for forecasting time series data. The study aims to analyze the historical data and develop an appropriate model that will forecast the electricity demand in Davao del Sur province in the Philippines. The dataset used in the paper was provided by the Davao del Sur Electric Cooperative, Inc. (DASURECO, Inc.) upon request through the electronic Freedom of Information (eFOI) website. The annual data cover the years 2000 to 2021. The data series was transformed using the Box-Cox transformation, and differencing was performed to address nonstationarity and nonconstant variance. The best model was selected among the tentative models by selecting the model with the least Akaike Information Criterion (AIC) value. The chosen model has undergone diagnostic checking and found that the residuals behave like white noise. The results show that ARIMA (0,1,0) with drift is the statistically valid model and is suitable for predicting electricity demand. The forecast shows that the electricity demand in the region will continue to increase, and it is predicted that by 2026, the demand will reach 505,246.4 megawatt-hours.*

## KEYWORDS

*ARIMA, machine learning, electricity demand, time series forecasting*

## 1. INTRODUCTION

Electricity plays a crucial role in modern living, and people rely heavily on it in performing daily activities, whether at home, school, or work [1]. Compared to other forms of energy, electricity has many advantages, for instance, efficient transfer, easy control, reliability, convenience, versatility, and cleanliness.

The Philippines saw a steady increase in electricity consumption in the past years. In 2019, the Department of Energy (DOE) reported that the country consumed 106,041 gigawatt-hours (GWh) of electricity, a 6.3 percent increase from 99,765 GWh of the previous year [2]. In Davao City, the Davao Light and Power Corporation also observed a growing electricity demand in the same year. It is reported that the electricity consumption of residential consumers significantly increased by 17 percent [3]. With the dependence and fast-growing demand, forecasting electricity consumption has become necessary. Policymakers, power utility owners, and private investors find forecasting electricity essential. The forecasts of electricity demand ensures that enough electricity is being produced to meet future needs. The accuracy of demand load forecasts can help save operating costs and increase power supply dependability [4]. For instance, the generation of electricity heavily relies on other resources like water, coal, oil, natural gas, biomass, and nuclear energy. With consumption forecasts, power stations can plan the number of resources needed to generate electricity. With this, underestimation and overestimation leading to power shortages and overspending of resources can be avoided.

Several prediction techniques were introduced to accurately forecast the electricity demand, and they can be categorized mainly into statistical (traditional) and Artificial Intelligence-based algorithms. One of the well-known statistical techniques in forecasting time series data is the Box-Jenkin's method or the Autoregressive integrated moving average (ARIMA) model. The model has been widely used in various fields as it is known to be an effective model for forecasting time series data [5]. In 2019, Delima forecasted the Philippine electricity consumption for 2018-2022 [6]. Parreño forecasted the electricity consumption for 2021-2030 [7]. The results found that the ARIMA model has high precision predictions and is appropriate for forecasting electricity consumption. Therefore, this paper applied the Box-Jenkin's method or the ARIMA model to forecast electricity demand in Davao del Sur. Moreover, there are few to no studies that forecast the electricity demand in Davao del Sur.

This paper primarily aims to examine the possibility to develop a model that will efficiently forecast the electricity demand in Davao del Sur. Specifically, this paper shall a) investigate the characteristics of the available historical data of electricity demand in Davao del Sur, b) develop an appropriate model for the electricity demand in Davao del Sur using the Box-Jenkins method, and c) forecast the electricity demand in Davao del Sur using the model found in b). This study is being carried out with the aim that the model could be used as a basis for planning electricity generation to meet the increasing electricity demand in Davao del Sur.

This paper is organized as follows: Section 2 presents the studies and literatures related to this paper. Section 3 describes the dataset used, forecasting methodology, and statistical software and packages used in this paper. Section 4 discusses the performance of the forecasting model and simulation results. Finally, the conclusions and recommendations are given in Section 5.

## 2. LITERATURE REVIEW

This section presents different models used to forecast electricity demand.

The performance of ARIMA and seasonal ARIMA (SARIMA) models in forecasting the electricity consumption of a healthcare building were compared in [8]. The models were built from a dataset consisting of 132 observations. It was revealed that the SARIMA(0,1,0)(0,1,1)12 model has better performance than the ARIMA(2,1,3) model since its MAPE and RMSE values were smaller than the ARIMA model.

The ARIMA and seasonal ARIMA model was also used in forecasting electricity demand in Ghana. The study found that SARIMA(0,1,1)(0,0,2) was appropriate for special load tariff data, ARIMA(1,1,1) was appropriate for non-special load tariff data, SARIMA(1,1,0)(1,0,0) was appropriate for prepaid load tariff data, and SARIMA(0,1,1)(1,1,3) was appropriate for regional load tariff data. The models were chosen based on their sum of squared estimates (SSE). The models were then validated using out-of-sample data with an error criterion of $\leq$1 GW [9].

A study applied the ARIMA models to forecast the electricity consumption of the Philippines. It applied model splitting; the first 43 data points were used to build the model, while the remaining 5 data points were used to evaluate the model. It considered the model with the least AIC value and implemented diagnostic checks and forecast evaluation to determine if the model was reliable. The results of the analysis revealed that ARIMA (0,2,1) was the appropriate model to forecast electricity consumption. The study predicted that by 2030, the electricity consumption of the Philippines would be 163,639.9 GWh. [7]

Other methods were also used to forecast electricity consumption. In [10], a modified grey prediction model was applied to forecast the electricity consumption of industries in China. The particle swarm optimization algorithm was utilized to generate the optimal parameters of the model. The novel prediction model applied these changes to the conditions to overcome the challenge of having a fixed structure and poor adaptability of the model in transforming the raw data. The results revealed that the novel modified grey prediction model performs better than other models. The vector error correction model (VECM) with the self-adapting screening (SAS) method was proposed in [11] to forecast and determine the effects of external economic factors on the monthly electricity consumption of China from 2000-2014. The X-12- ARIMA was implemented first to filter the seasonal patterns in the data series. The VECM was then applied to explore the correlations of the external economic factors to electricity consumption. Then, the self-adapting screening method was utilized to determine which external economic factors are influential and can cause of contradiction among data quantity and data length. Finally, in [12], a seasonal GM(1,1) model was implemented to forecast industrial electricity consumption in China. The model was built from seasonal historical data from 2010 to 2016. The performance of seasonal GM(1,1) was compared to GM(1,1), PSO-GM(1,1), and APL-SFGM(1,1) models. It was revealed that the prediction accuracy of seasonal GM(1,1) was better than other models. It is predicted that by 2020, the electricity consumption of industries in China will be 107.645 TWh.

# 3. METHOD

## 3.1. Dataset

The dataset used in this paper was provided by the Davao del Sur Electric Cooperative, Inc. (DASURECO, Inc.) upon the author's request through the electronic Freedom of Information (eFOI) website. It is in megawatt-hours (MWh) and consists of annual electricity demand from 2000 to 2021.

## 3.2. Autoregressive Moving Average (ARIMA) model

The ARIMA model was used to forecast the electricity demand for 2022 to 2026. The general ARIMA model is given by:

$$\Phi(B) = (1-B)^d y_t = \delta + \Theta(B)\epsilon_t \quad (1)$$

where $\delta$ is the constant term, $\epsilon$ is the white noise, $\Theta(B) = 1 - \Phi B - \Phi_2 B^2 - \cdots - \Phi_p B^p$ and $\Theta(B) = 1 - \theta B - \theta_2 B^2 - \cdots - \theta_3 B^3$ [13].

## 3.3. Autocorrelation Function

The autocorrelation calculates the correlation coefficient for each set of ordered pairs of within variable values separated by *k* period or lags. It helps determine the order of the moving average (MA) component. The autocorrelation coefficient at lag *k* is given by

$$\mathrm{ACF}(Lk) = \frac{\frac{1}{n-1}\sum_{i=1}^{n-k}((x_{i+1}-\bar{x})(Lk_i-\bar{x}))}{\sigma_x \sigma_x} \quad (2)$$

## 3.4. Partial Autocorrelation Function

The partial autocorrelation calculates the correlation among the current observation and the previous observation, given that both observations are correlated to observations at other times. It also helps determine the order of the autoregressive (AR) component. The *k*th partial autocorrelation of $y_t$ and $y_{t-k}$ is given by

$$PACF = (y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1})}{\sigma_{y_t | y_{t-1}, \dots, y_{t-k+1}} \sigma_{y_{t-k} | y_{t-1}, \dots, y_{t-k+1}}} \quad (3)$$

### 3.5. Akaike Information Criterion

The Akaike Information Criterion (AIC) was used to select the best model among the tentative models. The AIC value is a way to measure the goodness-of-fit of the model while penalizing the model for data over-fitting. The model with the least AIC value indicates superior goodness-of-fit and lesser tendency to overfit and was considered the most appropriate model. The AIC is given by:

$$AIC = 2k - 2 \ln L \quad (4)$$

where $k$ is the number of model parameters and $L = L(\hat{\theta})$ is the maximum value of the likelihood function of the model.

### 3.6. Statistical Software

All calculations in this paper were done using the R statistical software. The following packages were used: ‘tseries’ to test the stationarity, ‘stats’ to perform a test of normality, ‘forecast’ to fit the ARIMA model to the series and to check the residuals, and ‘lmtest’ to test the coefficients of the fitted model.

## 4. RESULTS AND DISCUSSIONS

The electricity demand in Davao del Sur from 2000 to 2021 is shown in Figure 1. The presence of an upward trend can be observed in the time series plot. A 2% decrease in electricity demand is observed from 2019 to 2020, and a 9% increase is observed from 2020 to 2021. These changes can be attributed to the adjustments made in response to the COVID-19 pandemic. Several lockdowns and quarantines were imposed in the region in 2020. In addition, different guidelines were imposed by departments, agencies, and local government units on different sectors to manage the spread of the virus that led to the decrease in electricity demand. The guidelines include changing the air condition temperature, suspending free WIFIs, and prohibiting the use of hand blowers or jet dryers. Business establishments were also forced to close temporarily. In 2021, several restrictions were eased, lockdowns and quarantines were lifted, and some businesses reopened. These events have caused an increase in electricity demand.

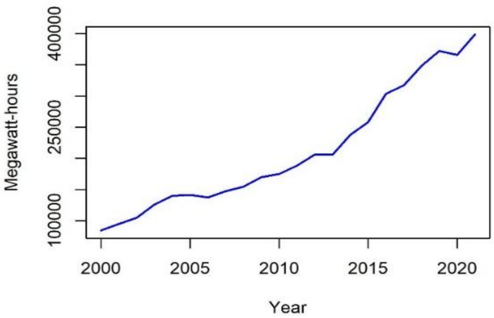


Figure 1. Electricity demand in Davao del Sur.

Figure 2 shows the autocorrelation function (ACF) plot of the series. A sinusoidal and exponential decay can be observed in the plot, which means that the series is nonstationary. In order to address this, transformation and differencing will be applied to the series.

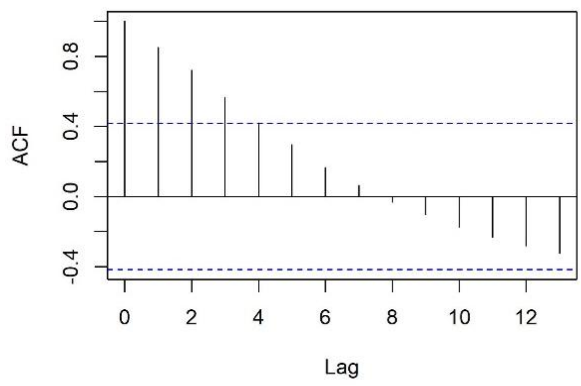


Figure 2. ACF plot of the series.

A problem with the variance in the original data series is also observed in Figure 1. Specifically, an issue of heteroscedasticity in the residuals is observed when trying to fit a model to the data. Box-Cox transformation with the optimal lambda of 0.52896 is applied to the series to address the issue. First differencing is also applied to the series to remove the linear trend and achieve stationarity. Figure 3 shows the first-differenced series.

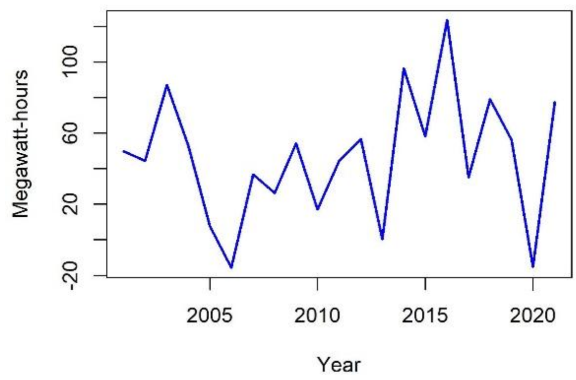


Figure 3. Electricity demand in Davao del Sur after transformation and differencing.

Figure 4 shows the ACF and PACF plots of the Box-Cox transformed and first-differenced series. It can be observed that the lags are close to zero and within the limits, indicating that the transformed and differenced series is stationary. Also, the autocorrelation function cuts off at lag zero and the partial autocorrelation function tails off. Hence, we have $q$=0 and $p$=0.

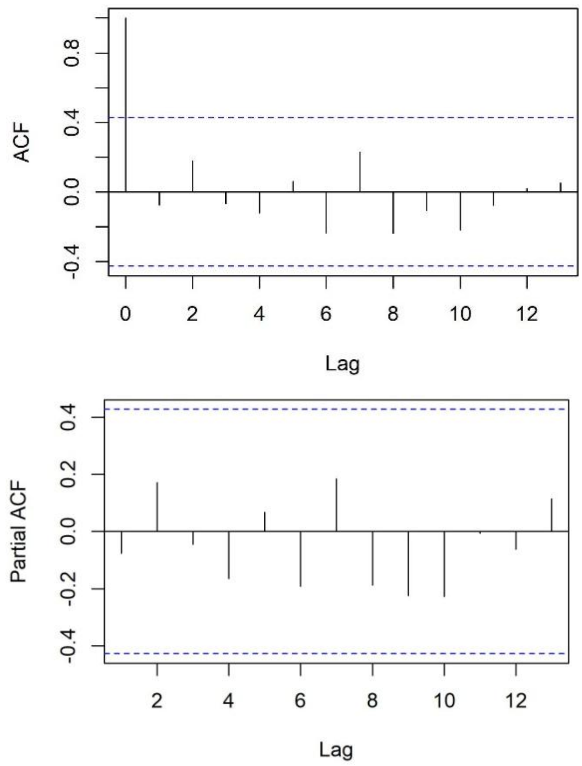


Figure 4. ACF and PACF plot of the transformed and first-differenced series.

Table 1 displays the result of the search for the optimal model. The tentative models and their corresponding AIC values are shown in the table. It can be observed that ARIMA (0,1,0) with drift has the smallest AIC value of 213.2992. Hence, ARIMA (0,1,0) with drift is the suitable model.

Table 1. Tentative models and their AIC values.

| Model | AIC |
|---|---|
| ARIMA (0,1,0) | 232.0635 |
| **ARIMA (0,1,0) with drift** | **212.6326** |
| ARIMA (0,1,1) with drift | 214.5418 |
| ARIMA (1,1,0) with drift | 214.5102 |
| ARIMA (1,1,1) with drift | 216.2781 |

The and (quasi-) Wald tests of estimated coefficients are performed to determine if the model is capable of making accurate predictions. It shows that the drift of 46.3075 has a standard error of 7.5855, a *p*-value of 6.1047, and a *p*-value of less than 0.01. These results indicate that the model can make accurate predictions.

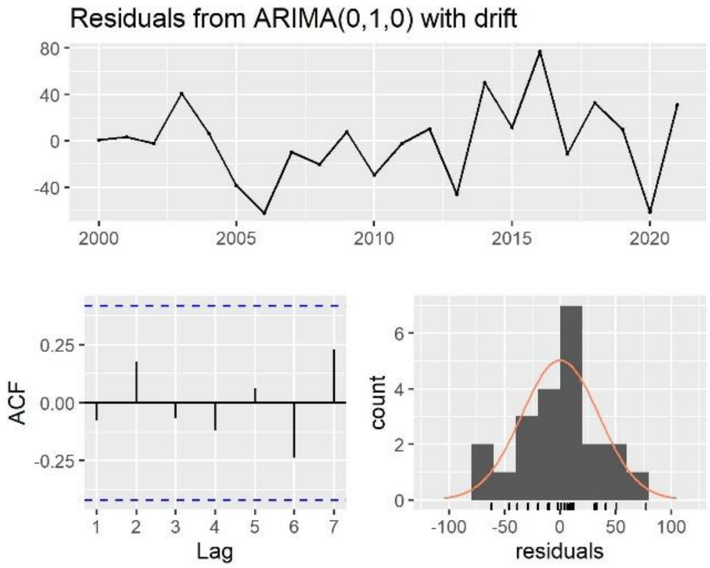


Figure 5. Residual diagnostics.

The residual plots are shown in Figure 5. It can be observed that there are no patterns left in the residuals. Moreover, the result of the KPSS test shows that the residuals are stationary ($p > 0.1$). This result indicates that the model does not produce false relationships. Also, the spikes in the ACF plot are within acceptable limits. It can also be observed that the residuals closely resemble a normal distribution. In order to confirm this Shapiro-Wilk test is performed, it has a test statistic of $w$=0.97311 and a $p$-value of 0.7817. Hence, the residuals are approximately normally distributed. The results indicate that the residuals behave like white noise. This indicates that the model has captured all the information in the time series.

Further, the Ljung-Box test is performed to formally test the model's lack of fit. The test has a test statistic of 1.5283, with degrees of freedom of 3 and a $p$-value of 0.6758. The result indicates that the model does not show a lack of fit.

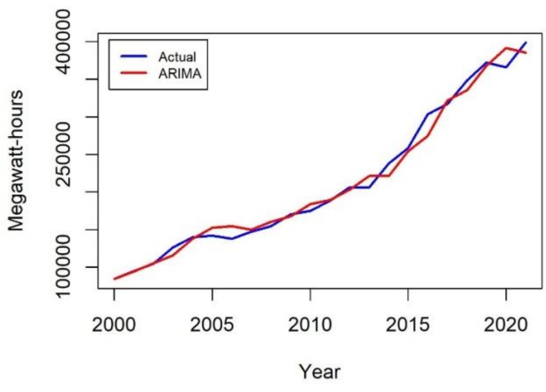


Figure 6. Fitted values from ARIMA (0,1,0) with drift compared to actual data.

The plot of the actual and fitted values of the model is shown in Figure 6. It can be observed that the selected model is sufficient as the predicted values are close to the actual values.

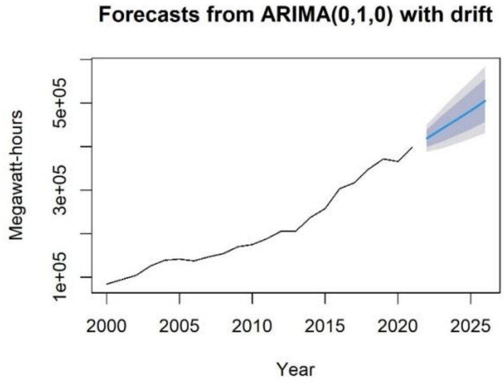


Figure 7. Forecasted electricity demand in Davao del Sur from 2022 to 2026.

The electricity demand forecast for 2022 to 2026 and 95% confidence interval values computed using the ARIMA (0,1,0) model with drift are shown in Figure 7. Table 2 shows the forecasted values and the 95% confidence interval values. It can be observed that the electricity demand in Davao del Sur will continue to increase in the following years.

Table 2. Forecasted electricity demand with 95% confidence interval

| Year | Forecast | Lower 95% | Upper 95% |
|---|---|---|---|
| 2022 | 419026.7 | 388506.8 | 450631.2 |
| 2023 | 439869.1 | 396007.1 | 485894.4 |
| 2024 | 461187.3 | 406521.2 | 519089.6 |
| 2025 | 482980.2 | 418712.8 | 551551.1 |
| 2026 | 505246.4 | 432078.0 | 583780.0 |

## 5. CONCLUSIONS AND RECOMMENDATIONS

In this paper, the appropriate ARIMA model was developed to forecast electricity demand in Davao del Sur. The dataset has been transformed using Box-Cox transformation with the optimal lambda of 0.52896. Then, first-differencing was applied to make the dataset stationary. The ACF and PACF plots were used to identify the order of the ARIMA model. Models with drift terms were also considered. In selecting the appropriate model, the AIC values of tentative models were taken into account. Results of the simulation revealed that ARIMA (0,1,0) with drift was the statistically valid model to forecast electricity demand. The forecast revealed that the electricity demand in Davao del Sur will continue to increase in the following years. By 2026, the electricity demand will reach 505,246.4 MWh. It can be as low as 432,078.0 MWh and as high as 583,780.0 MWh.

Due to limited data, only 22 observations were used to build the model, out-of-sample testing was not performed. However, the use of AIC values and residual analysis was able to address the issue. Also, the ARIMA model was based on the annual observations of electricity demand. Hence, the model was not able to capture the seasonality of the electricity demand in Davao del Sur. With this, future researchers may consider the monthly electricity consumption in Davao del Sur to better understand its behaviour and seasonality. Also, future researchers may use other models such as SVR and hybrid models, and compare their results with the results presented in this study.

## REFERENCES


[1] Zohuri, Bahman (2016). Application of compact heat exchangers for combined cycle driven efficiency in next generation nuclear power plants, New York: Springer.

[2] Lagare, Jordeene (2020, May 25). 2019 PH power consumption surges. The Manila Times. https://www.manilatimes.net/2020/05/25/business/business-top/2019-ph-power-consumption-surges/727067

[3] Fenequito, Amando, Jr. (2020, May 22). Higher usage led to increase in electric bills in Davao City. Manila Bulletin. https://mb.com.ph/2020/05/22/higher-usage-led-to-increase-in-electric-bills-in-davao-city/?amp

[4] Kaur, Navroop & Sood, Sandeep (2017). An energy-efficient architecture for the internet of things (IoT). IEEE Systems Journal, vol. 11, no. 2, pp. 796-805.

[5] Yang, Qiuying, Wang, Jie, Ma, Hongli, & Wang, Xihao (2020). Research on COVID-19 based on ARIMA model$^{\Delta}$—taking Hubei, China as an example to see the epidemic in Italy. Journal of Infection and Public Health, vol. 13, no. 10, pp. 1415-1418.

[6] Delima, Allemar Jhone (2019). Application of time series analysis in projecting Philippines' electricity consumption. International Journal of Machine Learning and Computing, vol 9, no. 5, pp. 694-699.

[7] Parreño, Samuel John (in press). Forecasting electricity consumption in the Philippines using ARIMA models. International Journal of Machine Learning and Computing.

[8] Kaur, Harveen and Ahuja, Sachin (2019). SARIMA modelling for forecasting electricity consumption of a health care building. International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 12, pp. 2795-2799.

[9] Andoh, Prince Yaw, Sekyere, Charles Kofi, Mensah, L. D. & Dzebere, D. E. (2021). Forecasting electricity demand in Ghana with the SARIMA model. Journal of Applied Engineering and Technological Science, vol. 3, no. 1, pp. 1-9.

[10] Ding, Song, Hipel, Keith & Dang, Yao-guo (2018). Forecasting China's electricity consumption using a new grey predicition model. Energy, vol. 149, pp. 314-328.

[11] Guo, Hongye, Chen, Qixin, Xia, Qing, Kang, Chongqing & Zhang, Xian (2018). A monthly electricity consumption forecasting method based on vector error regression model and self-adaptive screening method. International Journal of Electrical Power and Energy Systems, vol. 95, pp. 427-439.

[12] Wang, Zheng-Xin, Li, Qin & Pei, Ling-Ling (2018). A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic. Energy, vol. 154, pp. 522-534.

[13] Montgomery, Douglas, Jennings, Cheryl & Kulachi, Murat (2015). Introduction to time series and forecasting, 1st ed. New Jersey: John Wiley & Sons.


## AUTHOR

**Samuel John E. Parreño**

He is a faculty of the University of Mindanao Digos College. He received his bachelor's degree from the University of Mindanao Digos College, Digos City, in 2018. He is currently pursuing his master's degree in applied mathematics at the University of Southeastern Philippines, Davao City. He placed first in the Licensure Examination for Teachers in 2018. Mr. Parreño is a member of the Philippine Association of Researchers and Statistical Software Users, Inc. His research interest is in data analytics, mathematical modelling, and machine learning.