Forecasting Electricity Demand Using a Hybrid Statistical Model

Arvin Paul Sumobay [1, *]

[1] *Philippine Science High School, Central Mindanao Campus, Baloi, Lanao del Norte, Philippines*
[*] *Corresponding author. Email: apbsumobay@cmc.pshs.edu.ph*

ABSTRACT
In this study, a total of four models were utilized, the pure ARIMA and ANN models and two different methodologies that were used to combine ARIMA (Autoregressive Integrated Moving Average) and ANN (Artificial Neural Network) models, an additive methodology and a multiplicative methodology, to model the electricity consumption data of Cagayan de Oro City. Results from the evaluation of all four models, the two pure and two hybrid models, have shown that with a MAPE of $\mathbf{0.7411}$ and MSE of $\mathbf{9.43} \times \mathbf{1011}$, ANN was the best model that fit the electricity consumption data of Cagayan de Oro City. Out of the two hybrid models, the additive ARIMA-ANN hybrid model performed better than the multiplicative model, having a MAPE of $\mathbf{1.3896}$ and a MSE value of $\mathbf{1.67} \times \mathbf{1012}$ compared to evaluation values of MAPE = 1.4258 and MSE = $\mathbf{1.\ 68} \times \mathbf{1012}$. Out of the four, ARIMA performed the least, with a MAPE of 1.4334 and MSE of $\mathbf{1.70} \times \mathbf{1012}$. Results from forecasting using ANN have shown that in three years, the electricity consumption data of the city will increase, with an average monthly growth rate of 0.3123%.

Keywords: ARIMA, ANN, electricity

1. INTRODUCTION

Over the past decades, electricity consumption has been rapidly increasing mainly due to urbanization and the growing population. Electricity providing firms need proper planning and efficient management to prevent electricity shortages that could hinder the economic performance of developing cities [1]. Accurate electrical demand forecasting will greatly benefit electricity-providing firms, as it provides the data needed for the efficient management of sources especially for cities like Cagayan de Oro wherein forecasting demand is a challenge due to the continual urbanization [2]. In many other growing cities, empirical studies have emerged regarding electricity demand. However, in Cagayan de Oro, there are no previous studies that have dealt with forecasting the electricity demand of the city.

Statistically speaking, the electricity consumption data sampled in an equal time interval, is a collection of time series data. Forecasting times series data involves collecting and analyzing time-series data and developing a model that would describe the underlying relationship of the data [3]. The resulting model, a mathematical function that defines the data, would then be used to forecast the future values of the data. Two of the most utilized time series models are the Autoregressive Integrated Moving Average (ARIMA) model and the Artificial Neural Network (ANN) model.

In real-world situations, times series data are not purely linear or nonlinear. So, despite its advantages, pure models such as ARIMA and ANN may be incapable of forecasting real-world time series data accurately. In order to reduce the limitations of both ARIMA and ANN, and at the same time combine its advantages, various studies have emerged formulating ARIMA-ANN hybrid models in hypothesis that better and more accurate prediction would come up when the two respected models are combined.

A model with the combination of the ARIMA model and ANN model can efficiently produce a hybrid that predicts more accurately and can be more flexible than the pure models. However, every set of time series data has its own unique model that would fit its properties and saying that the hybrid models of ARIMA and ANN would also be the better fit for the electricity consumption dataset of Cagayan de Oro city would not be accurate. So, the aim of this study was to simulate a data under the hybrid ARIMA-ANN models, using both of the methodologies presented, and compare it to the existing pure models to ascertain whether a hybrid model would be the best fit for the data and which methodology would be best used in formulating a hybrid model [4], [5]. The models would forecast using the monthly electrical consumption of the city from year 2010-2017.

The objectives of the study are: (1) to formulate four models (ARIMA, ANN and two hybrid models combining ARIMA and ANN); (2) to compare the efficiency of the four models using the mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and the mean absolute error (MAE); and (3) to forecast the electricity demand of Cagayan de Oro city using the determined best fit model [6]–[8].

The existing hybrid models can be used as an efficient tool for forecasting electricity demand in Cagayan de Oro City. The implications and evaluations of the data may help electricity companies in making decisions for scheduling the operation of the existing and new power plants so that the supply of electricity can be made adequate to meet the future demands. It may also help the government in implementing programs and policies on energy conservation to avoid an energy crisis in the future.

2. METHOD

The additive model has always followed the assumption that a given time series data is a sum of the linear component Lt and the nonlinear component Nt. For the Afterwards, these values from the nonlinear component (et) was modelled using the best formulated ANN model with the ideal number of hidden nodes and then, the predicted values from the nonlinear model were obtained. In this model, the final prediction values of the additive model were obtained by adding the predicted values of the ARIMA model and the predicted values of the ANN model. The equation of the additive model was expressed in the equation below which was essentially the sum of the ARIMA model, Lt, and ANN model, Nt. The predicted value of the additive model is denoted as $\hat{y}$.

$$\hat{y} = L_t + N_t \tag{1}$$

Unlike the additive model, the multiplicative model in a given time series data was assumed to be a product of the linear component, Lt, and the nonlinear component Nt. In this model, the ARIMA model used in the additive model was also used to compose its linear component. Once again, this ARIMA model was utilized to predict values of the electricity consumption data. The residuals, however, were obtained by dividing the values predicted by the ARIMA to the electricity consumption data as shown in the equation below:

$$n_t = \frac{V_t}{L_t} \tag{2}$$

Afterwards, these nonlinear values from the nonlinear component (et) were modelled using the best formulated ANN model with the ideal number of hidden nodes and then, the predicted values from the model were obtained. In this multiplicative method, the final predictions were obtained by multiplying the predicted values of the ARIMA model and the predicted values of the ANN model. The multiplicative model is as shown in the equation.

$$y = L_t * N_t \tag{3}$$

After the formulation of the ARIMA, ANN, and the hybrid ARIMA-ANN models, these models were used to predict values of the electricity consumption data of Cagayan de Oro city. The predicted values were compared to the actual values of the electricity consumption data graphically. Along with the graphical comparison, the accuracy of the models was evaluated using five statistical linear model, the best fit ARIMA model that was obtained in the construction process using the Econometric Modeler toolbox was utilized to model the linear component, Lt, of the electricity consumption data. It was then used to predict values of the electricity consumption data.

After the linear component prediction of values from ARIMA, the residuals, which were assumed to be the nonlinear component series, was obtained by subtracting the predicted values of the ARIMA model (Lt) from the electricity consumption data (yt) as shown in the equation below. It is denoted as et given as:

$$et = yt - Lt \tag{4}$$

performance evaluation criteria: the mean square error (MSE), root mean square error (RMSE), the mean absolute percentage error (MAPE) and the mean absolute error (MAE), which were calculated using the software. The MSE, RMSE, MAPE and MAE values of the four formulated models were compared and the model with the least values of these criteria was determined as the most efficient model. The performances of the models were also summarized using the mean absolute deviation and standard deviation of each model. These were used to analyze the quality of the predictions.

3. RESULTS AND DISCUSSION

The monthly electricity consumption data of Cagayan de Oro City from the year 2010 - 2017 were used as the time series data for the formulation of the models. The data were composed of 96 data points. Based on the time series data plot (Figure 1), the electricity consumption of the city had an increasing trend after December 2011 and was non- stationary.

When the electricity consumption time series data were examined using the Augmented Dickey Fuller's test with significance level of 0.05, a p-value of 0.9580 was obtained, which meant that the time series was non-stationary. Table 2 shows the comparison of AIC and BIC values of the seven possible ARIMA models. Based on the comparison, ARIMA (3, 1, 1) has the least AIC and BIC values, thus, it is the best fit ARIMA model as it has the least performance error according to AIC and BIC.

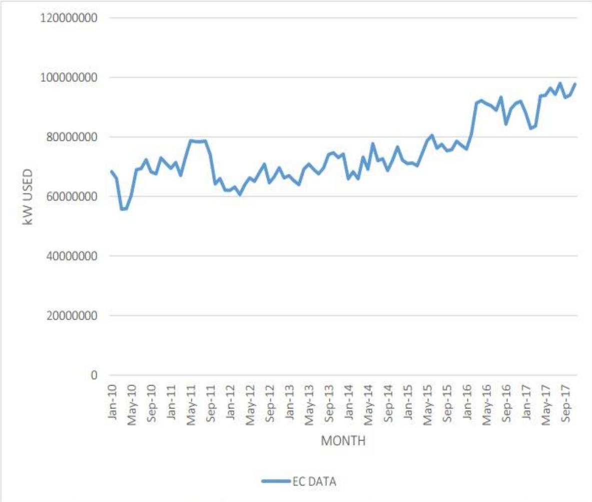

Figure 1 Time series data plot of the monthly electricity consumption (kW) of Cagayan de Oro City from January 2010 – December 2017

Table 1 Results from the Augmented Dickey-Fuller Test of the Electricity Consumption Time Series Data (d = 0) and the Differenced Data (d = 1) at α = 0.05

	Data	Lag order	Dickey-Fuller	p-value
$d = 0$	KWHUSED	4	1.3904	0.9580
$d = 1$	KWHUSEDDiff1	4	-5.8191	0.01

3.1. ARIMA Model

Table 2 shows the comparison of AIC and BIC values of the seven possible ARIMA models. Based on the comparison, ARIMA (3, 1, 1) has the least AIC and BIC values, thus, it is the best fit ARIMA model as it has the least performance error according to AIC and BIC.

Table 2 The Akaike Information Criterion and the Bayesian Information Criterion Values of the ARIMA Models with Different Set of Possible Parameters

ARIMA MODEL	AIC value	BIC value
ARIMA (1,1,1)	3186.7	3176.6
ARIMA (1,1,3)	3188.7	3173.5
ARIMA (3,1,1)	3184.9	3169.8
ARIMA (3,1,3)	3213.7	3193.6
ARIMA (3,1,12)	3229.9	3187.2
ARIMA (12,1,1)	3213.2	3177.1
ARIMA (12,1,3)	3249.1	3208.2

Using ARIMA (3, 1, 1), the coefficients of the model parameters were then obtained. Table 3 presents the coefficient of the model parameters of ARIMA (3,1,1) together with its statistical values.

Table 3 The Coefficients of the Parameters of ARIMA (3,1,1) for Model Building with Corresponding Statistical Values

	Value	Standard Error	t Statistics	p-value
Constant	$7.15\text{x}10^{3}$	$2.02\text{x}10^{4}$	0.3532	0.7239
AR{1}	-0.0347	0.1397	-0.2482	0.804
AR{2}	-0.163	0.1346	-1.2112	0.2258
AR{3}	-0.3022	0.1423	-2.1241	0.0337
MA{1}	-1	0.0751	-13.3164	$1.86\text{x}10^{-40}$
Variance	$2.00\text{x}10^{13}$	$2.25\text{x}10^{-5}$	$8.90\text{x}10^{17}$	0

After the identification of the best fit ARIMA model and the determination of the coefficients of its' parameters, the residuals were also examined. The model is determined to be homoscedastic and have insignificant errors during its performance, if the results of the examination of the residuals imply that it has a random and independent characteristic. Using the infer function in MATLAB, the residuals of ARIMA (3,1,1) were obtained and plots were graphed.

The graph of the residuals imply that the residuals were random and independent because no trend was observed in either graph, nor were there spikes outside the insignificant zone for both ACF and PACF plots. As seen in Figure 2, most of the residuals are in the range (-1,1), thus, the residuals do not violate the assumption of constant location and scale. This means that the residuals of ARIMA (3,1,1) are random, thus, the ARIMA model can be fitted in the value (Nist Sematech, 2016).

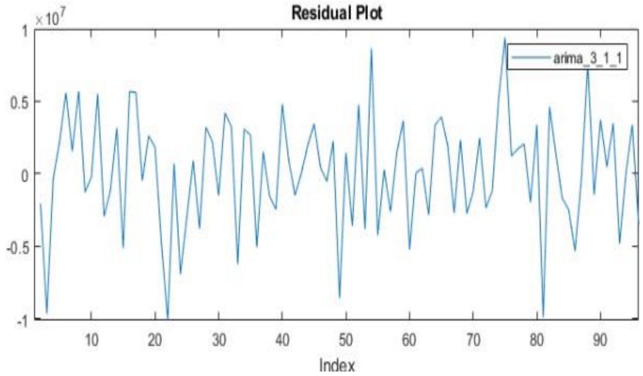

Figure 2 Plot of the residuals of the predicted monthly electricity consumption data using ARIMA (3,1,1)

After obtaining the residuals and confirming the homoscedasticity of the ARIMA model, the residuals were subtracted from the differenced electricity consumption data to get the actual predicted difference of the electricity consumption data of Cagayan de Oro city from January 2010 to November 2017. The difference was then subtracted to the initial electricity consumption data to obtain the final predicted value of ARIMA (3,1,1). The predicted values were imported into a Microsoft Excel file and was evaluated using several performance evaluation statistics. The predicted values of ARIMA (3,1,1) were plotted along with the initial electricity consumption (EC) data of Cagayan de Oro City from January 2010 to December 2017 for comparison (Figure 3). It is observed that the predicted values of ARIMA (3, 1, 1) is quite similar with the initial EC data and there is no obvious and significant difference between the two datasets.

A three-layered feedforward neural network model was developed for the prediction of the electricity consumption data using Bayesian regularization training algorithm with a Levenberg-Marquardt backpropagation. In the formulation of the artificial neural network, the electricity consumption data, with 96 data points, was divided into two sets. 68 data points out of 96 were used as the training set while 28 data points were used as the test set.

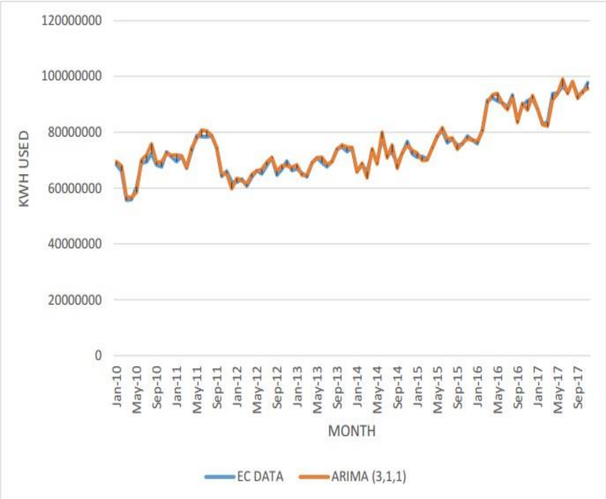

Figure 3 Plot of the monthly electricity consumption data of Cagayan de Oro City from January 2010 to December 2017 and the predicted data using ARIMA (3,1,1)

3.2. ANN Model

Prior to the training of the model, the number of neurons in the input and output layer was set as 1 and 1 respectively by default, and the number of delays was set as 1. The number of iterations that were used to train the data set was set to 1000. In the determination of the optimum number of hidden nodes, a trial and error method were used. The determination of the correct number of hidden neurons is important for the network performance in the training and forecasting stages. If too few parameters are selected, the network cannot capture the complex model dynamics, while a model with too many parameters neurons can be easily overfitted, thus, it will generate poor predictive performance [9].

Six possible values of hidden nodes were tested: 1, 2, 4, 5, 6, and 10. To determine the most optimum number of hidden nodes, six different ANN models with the corresponding number of hidden nodes were trained and their performance was evaluated using four statistical measures, the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

The results from the evaluation of the probable ANN models are presented in Table 4. The results show that there is that the ANN model has the least penalty errors when the number of hidden units is 4. Because of this, 4 was chosen as the number of hidden nodes. Thus, the number of each layer's neurons in the network was 1 -4-1, and with 68 data points as the training set, the best ANN model that was trained under the Bayesian regularization algorithm with Levenberg-Marquardt error backpropagation was the ANN(1x4x1).

Table 4 Values of the Evaluation Performance Criteria of the Possible ANN Models for the Selection of the Best ANN Model

ANN Model	MSE	RMSE	MAE	MAPE
ANN(1x1x1)	1.18×10^{12}	1,085,453.64	571,092.06	0.8390
ANN (1x2x1)	9.80×10^{11}	990,103.01	581,238.35	0.8651
ANN (1x4x1)	9.43×10^{11}	971,162.06	520,032.82	0.7411
ANN (1x5x1)	9.65×10^{11}	983,544.63	538,935.88	0.7706
ANN(1x6x1)	1.92×10^{12}	1,386,325.42	1,083,903.51	1.4839
ANN(1x10x1)	3.19×10^{12}	1,787,177.58	1,217,164.46	1.6648

As shown Figure 4a, the regression value for the training set was 0.91902, and the correlation coefficient value for the test set was 0.96683 (Figure 4b). These values indicate that ANN (1x4x1) model is an efficient model for predicting the true values of the electricity consumption data. This is because R values greater than 0.9 indicate a strong, positive correlation between the two values.

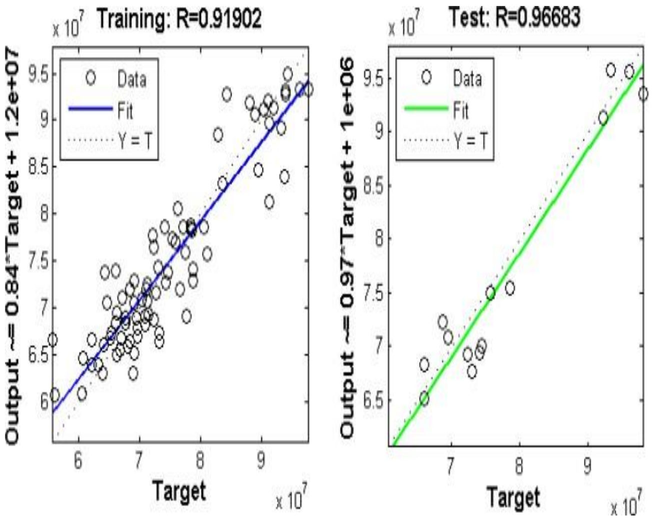

Figure 4 Regression plots of the predicted value by ANN(1x4x1) compared from the initial EC dataset in the training set (left) and test set (right)

In the training, the model exhibited only small testing sample errors, hence the errors were considered acceptable. The minimal number of errors is an assurance that the model is trained well. Hence, ANN (1x4x1) can be used to predict the actual values of the electricity consumption data. After the determination of ANN(1x4x1) as the best ANN model, the model was used to predict the values of the electricity consumption. Using the "Step-Ahead Prediction Network" function in MATLAB, the predicted values of ANN(1x4x1) model were obtained. The predicted values were imported into a Microsoft Excel file and were evaluated using performance evaluation statistics. The predicted values of ANN (1x4x1) were plotted along with the electricity consumption data (Figure 5).

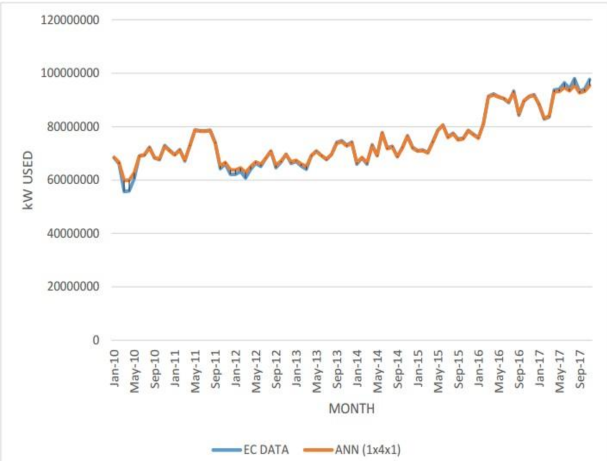

Figure 5 Plot of the monthly electricity consumption data of Cagayan de Oro City from January 2010 to December 2017 and the predicted data using ANN (1x4x1)

3.3. Hybrid Models

By fitting the ARIMA (3, 1, 1) model on the electricity consumption (EC) data of Cagayan de Oro City from January 2010 to December 2017, a set of predicted values was obtained and considered as the linear component for both hybrid models, denoted as Lt. The nonlinear component, denoted as Nt, was obtained using the different methodologies of each hybrid model. For the additive model, the predicted data of ARIMA (3, 1, 1) was subtracted from the initial EC data and the residuals obtained was modeled by ANN (1x4x1). For the multiplicative model, the residuals, which was the quotient of the predicted ARIMA values divided by the EC data was again modeled using ANN (1x4x1).

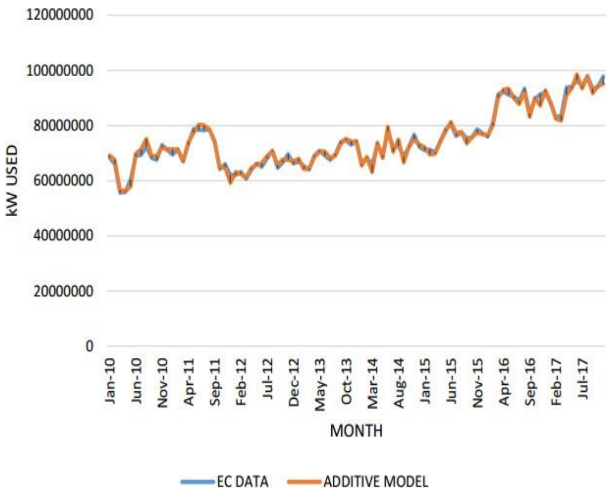

Figure 6 Plot of the monthly electricity consumption data of Cagayan de Oro City from January 2010 to December 2017 and the predicted data using additive ARIMA (3,1,1) – ANN (1x4x1)

Lt and Nt, were added together and the predicted values of the additive model were then imported into a Microsoft Excel file and plotted along with the EC data of Cagayan de Oro City (Figure 6). For the multiplicative hybrid model, Equation 6 was used to obtain the final predicted values of ARIMA (3, 1, 1)–ANN (1x4x1). The linear and nonlinear component were multiplied and the resulting predicted values were also imported into a Microsoft Excel file and plotted using one of the functions of the program. The plot of the predicted values of the multiplicative ARIMA (3,1,1)-ANN (1x4x1) is as shown Figure 7.

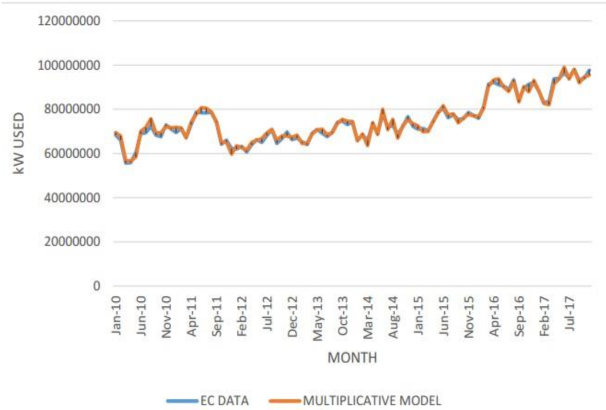

Figure 7 Plot of the monthly electricity consumption data of Cagayan de Oro City from January 2010 to December 2017 and the predicted data using multiplicative ARIMA (3,1,1) – ANN (1x4x1)

Figures 6 and 7 both show that the predicted data of both models closely resemble the initial EC data. However, visual representation is not enough to compare all formulated models. Thus, statistical evaluations were used to determine best fit model for the EC data of Cagayan de Oro City from January 2010 to December 2017.

3.4. Evaluation of the Models

After the formulation of the four models, the predicted values of each models were obtained following the methodologies of each model. Afterwards, the predicted values of each models were plotted along with the electricity consumption data for the purpose of comparison. Graphical analysis show that the predicted values by the four models are quite similar with the initial set of values as seen in Figure 8. However, statistical analysis is needed to accurately determine the quality of performances of all the models [2], [10], [11].

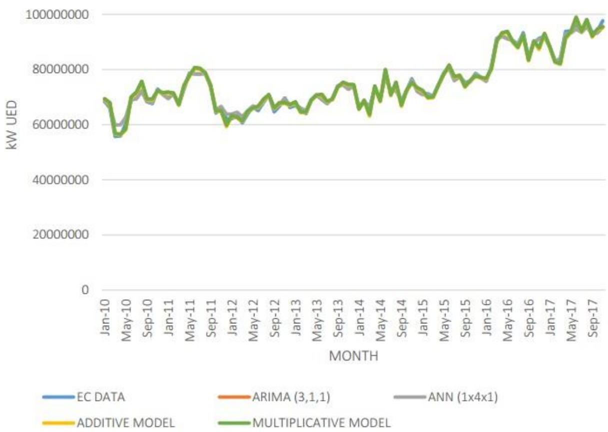

Figure 8 Plot of the predicted monthly electricity consumption data of the four models and the monthly electricity consumption data of Cagayan de Oro City from 2010-17

One of the techniques used to determine the accuracy of predictions is the use of deviation statistics such as mean absolute deviation (MAD) and standard deviation (SD). Results shows that the model with the least deviation is the ANN Model, with MAD of 9.43 × 1011 and SD of 971,162.06.

To further show the accuracy of the models, they were evaluated using five statistical performance evaluation criteria: mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE). Results show that the model that has the least penalty errors is the ANN (1x4x1) model, with a MAPE of 0.7411 and MSE of 9.43 × 1011. This implies that the ANN(1x4x1) model is the most efficient model in predicting actual values of the electricity consumption data. Because of this, the electricity consumption data are a nonlinearly generated time series, which can be seen directly in the time series plot of the data shown in Figure 1. Although there is an increasing trend observed in the electricity consumption data, there are several fluctuations in the data which might be caused by exogenous variables such as the impact of income, retail price of electricity, population and weather, number of customers, and many others. However, it is not considered in the formulation of the model itself. Also, due to ANN's capability of learning the patterns, it can predict well through all the time.

It can be seen also that among the formulated models, the ARIMA (3,1,1) model has the poorest performance, with a MAPE of 1.4334 and a MSE value of 1.70 × 1012. It can be seen that the ARIMA (3, 1, 1) worked well on the beginning phases but its performance declined in the last phases. This was because the 96 data points that were used in formulating the model is not sufficient enough to provide data in order to have a model that could accurately predict and forecast long term data with minimal errors. Improvement of the accuracy of the forecast must be obtained by utilizing a larger data [5], [12] and making use of other important factors such as meteorological and climatic variables.

The results from the evaluation of the hybrid models also show that both the additive ARIMA(3,1,1)-ANN(1x4x1) and multiplicative ARIMA(3,1,1)-ANN(1x4x1) models had better prediction results than ARIMA (3,1,1), with MAPE values of 1.3896 and 1.4258, respectively. This implied that there was an improvement in the prediction performance of the hybrid models after the nonlinear component was modeled by the ANN model. This implies that using ANN model to predict the nonlinear component of the time series based on each hybrid model's own methodology and then incorporating them into the linear component predicted by ARIMA model produce prediction values which are more accurate than those of the individual ARIMA model. Still, ANN(1x4x1) outperformed the hybrid models. This indicates that the formulation of an ARIMA model which produces more accurate prediction values with minimal errors is crucial in the formulation of the hybrid models.

Results from the evaluation of the models also show that the additive ARIMA(3,1,1)-ANN(1x4x1) with MAPE of and MSE of 1.3896 and 1.67 × 1012 , respectively, outperformed the multiplicative ARIMA(3,1,1)- ANN(1x4x1) (MAPE= 1.4258, MSE=1.68E+12). This implies it is better to predict the electricity consumption data if it is assumed to be the sum of linear components and nonlinear components. However, other related studies such as the study of Wang et al. (2013) have stated that multiplicative model exhibits better performance than the additive model. Because of this, it should not be generalized that the additive model is better than multiplicative model or the reversed situation in time series forecasting because each type of time series data has its own statistical properties that differ from the other type of time series data. Hence, further studies must be conducted regarding on the utilization of the two hybrid ARIMA- ANN models to clarify the conclusions regarding on the comparison of the two hybrid methodologies.

The results that were gathered and analyzed deviated from the general assumption that combining the linear ARIMA model and the nonlinear ANN model would predict more accurate results than the individual models. However, the individual ANN(1x4x1) outperformed all of the other models, including the hybrid models. Because of this, it is suggested that future studies regarding the utilization of the additive and multiplicative hybrid models should be done for further inspection regarding the quality of performance of the hybrid models. Also, factors that may affect the formulation of the models such as the number of data points in the dataset and exogenous variables should be considered.

3.5. Electricity Demand Forecasting

Since the ANN(1x4x1) model has the least penalty errors in predicting or fitting the electricity consumption data from January 2010 to December 2017, it was chosen to forecast future values of electricity demand of Cagayan de Oro City for the years 2018- 2025. Using MATLAB, ANN (1x4x1) was used to forecast the EC of Cagayan de Oro with 80% forecast interval and 95% forecast interval (Table C, Appendix C). The obtained forecasted data was then imported and plotted in Microsoft Excel. The plotted forecast data with 80% forecast interval and 95% interval is shown in Figures 9 and 10, respectively [10], [13].

The blue line in the above figures represents the forecasted data of ANN (1x4x1), however, it is impossible for a real-time series data to only have a linear behavior. So, a forecast interval was assumed with 80% and 95% confidence interval to predict the upper and lower limits of the forecasted data. The yellow line represents the upper limit of the possible forecast values while the gray line represents the lower limit. Hence, it is assumed that the real data will fluctuate within the bounds of the forecast intervals.

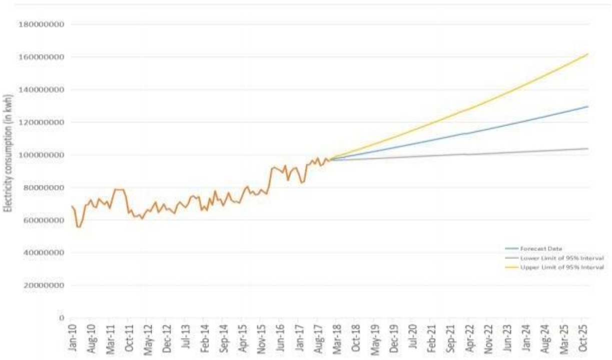

Figure 9 Plot of the forecast and the 95% forecast of ANN(1x4x1) of the monthly electricity consumption for January 2018 to December 2025

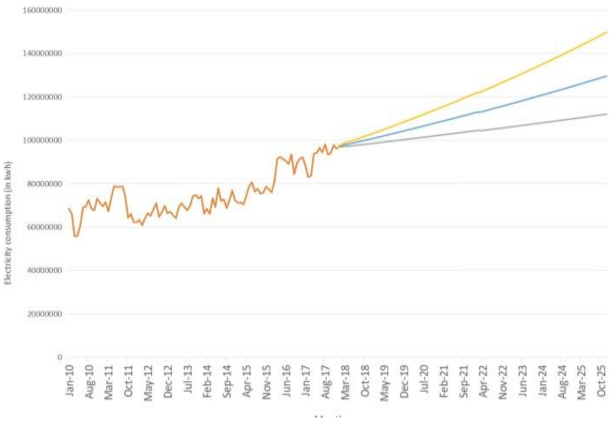

Figure 10 Plot of the forecast and the 80% forecast of ANN(1x4x1) of the monthly electricity consumption for January 2018 to December 2025

The summary of the forecasted data can be inferred that 8 years from December 2017, there will be an increase in the electricity consumption of Cagayan de Oro City at an average growth rate of 0.3062%.

It can also be noticed in the data that the yearly average growth rate did not just increase each year. The growth rate in the year 2019 decreased from 0.3794 to 0.3066 before increasing again to 0.3072 in 2020. This further proves that the forecasted data fluctuates and does not progress in linear movement but its values should be within the bounds of the forecast intervals. With these data, further planning and government policy measures in energy conservation must be considered explicitly by the Cagayan Electric Power and Load Corporation to meet the increasing demands in the future. However, the current data regarding on the planning and the current policy measures of CEPALCO were unknown to the researchers, due to the data privacy measures. Hence, the research cannot explicitly recommend activities to prevent any malfunctions in the operations of the corporation caused by the potential increase [14], [15].

4. CONCLUSION

In this study, four type of models, two individual stochastic time series models: autoregressive integrated moving average model (ARIMA) and artificial neural network, (ANN), and two hybrid models, additive ARIMA-ANN model and multiplicative ARIMA-ANN model, were formulated to predict the values of the electricity consumption of Cagayan de Oro City from January 2010 to December 2017. With a BIC value of 3148.9 and an AIC value of 3169.8, the best ARIMA model was ARIMA (3,1,1) which had an AR parameter of 3, 1 st differencing order and a MA parameter of 1. Then, the best ANN model was ANN (1x4x1), a neural network with one input neuron, four hidden neurons and an output layer with one neuron.

These two models were then utilized to create the additive and multiplicative hybrid ARIMA-ANN models and were also used to compare it to the generated hybrid models. Results from the comparison of the two hybrid models showed that the additive ARIMA(3,1,1)-ANN(1x4x1) model, with MAPE of 1.3896 and MSE of 1.67x1012, has a greater performance than that of the multiplicative model (MAPE= 1.4258, MSE=1.68x1012). Results from the evaluation of the four models showed that ANN(1x4x1) outperformed all the other formulated models with MAPE of 0.7411 and MSE of 9.43x1011. This proves that not all models that are identified as better are the best model at any given time series dataset since each dataset has its own statistical properties that may be or may not be modeled efficiently by any given methodology. The underperformance of the hybrid models may be accounted to the data limitation, wherein only 96 data points were utilized in the formulation of the models. A large dataset is very crucial especially in the formulation of the ARIMA model because it is a model that predicts more efficiently when there is a larger set of historical data. And while ANN had the best results, better results might still have been exhibited by ANN if a larger dataset was utilized in its training. Because the ARIMA clearly has been outperformed by ANN, it may have affected the results of the hybrid models. It is recommended to have a greater number of data points be utilized when forecasting using these models in the future.

The ANN (1x4x1) model was utilized to forecast the electricity demand of the city from January 2018 to December 2020. Results have shown that in three years, the electricity consumption data of the city will increase, with an average monthly growth rate of 0.3123%. Therefore, further planning and government policy measures in energy conservation must be considered explicitly by the Cagayan Electric Power and Load Corporation to meet the increasing demands in the future.

The forecasting process using ANN(1x4x1) only considered the electricity consumption data of Cagayan de Oro city. It is recommended that other external variables that affects or cause fluctuations on the electricity consumption data such as meteorological variables, population, number of power outages and electricity price should be considered in the formulation of the models. The formulated models may provide more accurate predictions and forecasts crucial to this type of research. It can also be said that the electricity demand of Cagayan de Oro city is more nonlinear that it is linear. This explains the best results exhibited by ANN. So, when forecasting using the hybrid ARIMA-ANN models, it is recommended to model these from datasets that are more or less equally linear and nonlinear. This type of data set may have a seasonal and increasing trend in its plotted graph. In conclusion, the best fit model for the electricity consumption data of Cagayan de Oro city from January 2010 – December 2017 was the ANN (1x4x1).

REFERENCES

[1] S. Nawaz, N. Iqbal, and S. Anwar, "Modelling electricity demand using the STAR (Smooth Transition Auto-Regressive) model in Pakistan," *Energy*, vol. 78, pp. 535–542, 2014.

[2] A. K. Singh, S. K. Ibraheem, M. Muazzam, and D. K. Chaturvedi, "An overview of electricity demand forecasting techniques," *Netw. Complex Syst.*, vol. 3, no. 3, pp. 38–48, 2013.

[3] H. G. K. Hummel, "Feedback model to support designers of blended learning courses," *... Rev. Res. Open Distrib. Learn.*, 2006.

[4] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[5] A. H. Nury, K. Hasan, and M. J. Alam, "Bin.(2015). Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh," *J. King Saud Univ.*

[6] F. Sowell, "Modeling long-run behavior with the fractional ARIMA model," *J. Monet. Econ.*, vol. 29, no. 2, pp. 277–302, 1992.

[7] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[8] S. C. Hillmer and G. C. Tiao, "An ARIMA-model-based approach to seasonal adjustment," *J. Am. Stat. Assoc.*, vol. 77, no. 377, pp. 63–70, 1982.

[9] Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy Build.*, vol. 111, pp. 184–194, 2016.

[10] T. Jakaša, I. Andročec, and P. Sprčić, "Electricity price forecasting—ARIMA model approach," in *2011 8th International Conference on the European Energy Market (EEM)*, 2011, pp. 222–225.

[11] M. Beccali, M. Cellura, V. Lo Brano, and A. Marvuglia, "Forecasting daily urban electric load profiles using artificial neural networks," *Energy Convers. Manag.*, vol. 45, no. 18–19, pp. 2879–2900, 2004.

[12] D. Whitebread, "Learning & Development: School Readiness: Part 3 - A better way?," *Nursery World*, vol. 2011, no. 12. Mark Allen Group, 2011.

[13] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. power Syst.*, vol. 20, no. 2, pp. 1035–1042, 2005.

[14] T. Al-Saba and I. El-Amin, "Artificial neural networks as applied to long-term demand forecasting," *Artif. Intell. Eng.*, vol. 13, no. 2, pp. 189–197, 1999.

[15] Z. Mohamed, P. S. Bodger, and D. J. Hume, "Forecasting Electricity Consumption: A Comparison of Models for New Zealand and the Maldives," 2004.