

Medium-Term Forecasting for Electricity End Use

Hosam H. A. Mukhairez
Faculty of Information Technology
Islamic University of Gaza
Gaza, Palestine

Abstract

Electricity is the nervous system in our live on earth and it is involved in many electronic devices and systems. In this paper, the end use of electricity was forecasted in the medium-term scale. The medium-term forecasting is conducted using the following forecasting algorithms: Auto-Regressive Integrated Moving Average (ARIMA), Hybrid ARIMA, and Linear Regression with horizon value equals five (2018-2022). After evaluating tested algorithms using MAPE it is noticed that Hybrid ARIMA was the best forecasting algorithm to be used with MAPE 1.77%, while ARIMA MAPE equals 4.81% and linear regression MAPE equals 7.04%. The deviation of the forecasted 5 years (2018-2022) is decreasing by 1.21% with comparison to the last 5 years (2013-2017).

1. Introduction

Many systems, machines and electronic devices consume electric energy to generate desired output (i.e., light, heat, motion, etc.). During operation, some part of the energy-depending on the electrical efficiency - is consumed in unintended output, such as waste heat. Electricity has been generated in power stations since 1882[1]. The invention of the steam turbine in 1883 to drive the electric generator started a strong increase of world electricity consumption.

Global power generation growth picked up slightly to 2.2% vs 1.6% in 2015 but remained below the 10year average of 2.8% pa. The uptick reflects faster growth in the non-OECD, while OECD power generation continued to stagnate. Non-OECD power generation grew by 4.0%, up from 2.6% in 2015. Most of the acceleration came from China (5.4%) with a further significant boost from India (6.8%) OECD power generation has been essentially flat since 2010. In 2016 the main drag on OECD growth was a large decline in Japan (-3.2%) which offset modest growth elsewhere in the OECD[2].

Figure 1 presents the electricity generation (terawatt-hours) for regions and continents which is the focus of this research showing that the Asia Pacific is the most region in electricity end use rates.

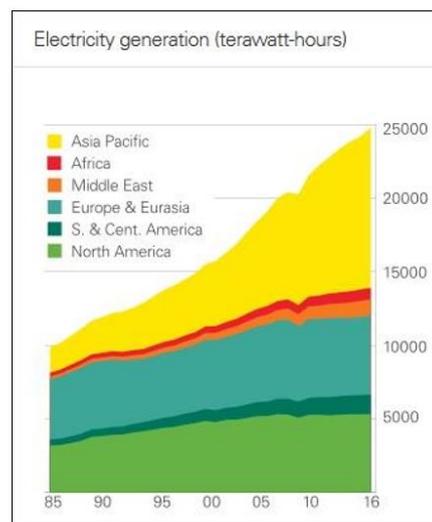


Figure 1. Electricity generation for regions and continents [2]

Forecasting task is understanding and recognizing the old behavior of an attribute to predict its future behavior or pattern. It is one of the oldest known predictive analytics techniques. The idea is to use historical data to make forecasts and prediction about future data. Forecasting is one of the most common techniques of time series data analysis. It's used to predict future trends in water demands, retail sales, economic indicators, weather forecasting, stock markets, and many other application scenarios[3].

Depending on the forecasting time horizon term, electricity end usage forecasts can be categorized, as suggested by [4] into: (1) Long-Term which forecasts decades, (2) Medium-Term which forecasts years to decade, (3) Short-Term which forecasts up to year. In this research, data was downloaded from BP Global¹ and presented in annual form starting from 1949 to 2017.

In this work MAPE is used as evaluation measure for selecting the best forecasting algorithms to forecast the next 5 years from 2018 to 2022.

The rest of this paper is organized as follows: Section 2 reviews some related works. Section 3

shows experimental process and results. Section 4 presents discussion. Section 5 the conclusion.

2. Related Work

De Giorgi et al. [5] proposed a hybrid statistical models which implemented for electricity system located in Apulia – South East of Italy. The conducted their research with different forecasting horizons, using historical output power data and performed by hybrid statistical models based on Least Square Support Vector Machines (LS-SVM) with Wavelet Decomposition (WD). They depended on short-term forecasting horizons starting from 1h to 24h. they noticed that the hybrid method based on LS-SVM and WD more accurate than other methods in most of cases.

Also, De Giorgi et al. [6] used the historical PV power data to forecast the photovoltaic. They used LSSVM and the GMDH models to predict short-term forecasting(24h) of the PV output power. They noticed that GLS-SVM method with the Dir-Rec have error rate 2.92% under different weather conditions with evident improvements respect to the traditional ANN (Artificial Neural Network).

Kaytez et al. [7] forecasted the Turkey's electricity consumption using LS-SVM, ANN and regression models depending on old data from 1970 to 2009. They used the metric Receiver operating characteristic (ROC) analysis to measure the accuracy of results. They found that LS-SVM model has accurate results than ANN and regression models.

Jurado et al. [8] suggested a hybrid model to forecast short-term of electricity load. They compared the prediction power of one machine learning (Neural Networks), two soft computing (fuzzy Inductive Reasoning, Random Forest) and one statistical technique (ARIMA). They observed that the performance of FIR (Fuzzy Inductive Reasoning) is out-performed than the other methodologies.

Chae et al. [9] proposed a very short-term forecasting model to predict the electricity usage. They built their model using Artificial Neural Network (ANN) model and Bayesian regularization algorithm. They claim that the accuracy is around 90% and 95% of daily total and daily peak energy usage.

Weron [10] reviewed available solutions to forecast electricity prices, their strengths and weaknesses, and the opportunities and threats that the forecasting tools offer or that may be encountered. He reviewed solution using the same datasets, the same robust error evaluation procedures, and statistical testing of the significance of one model's outperformance of another. In general, he listed some important titles like: how are we forecasting? available models in short, medium and long terms, their strengths and weaknesses.

Finally, he wrote some ideas of the future of 'electricity price forecasting'

3. Experimental Results

The three chosen forecasting algorithms is considered the most famous algorithms used in time series analysis field. These algorithms are Auto Regressive Integrated Moving Average (ARIMA), ARIMA combined with Neural Networks (Hybrid ARIMA) and Linear Regression forecast. These algorithms were run over datasets to calculate the Mean Absolute Percentage Error (MAPE) to detect which algorithms is the more accurate to be used in forecasting process depending on MAPE measure.

The data set is real data set collected from BP Global. The data set belongs to the yearly oil produced amount from most regions from the world starting from 1949 to 2017.

The original data consists of 2 columns (year, electricity end use 'Million Kilowatt-hours'). The data sets containing 68 years of indices data, was divided into training set, containing the first 60 years of data 1949 to 2009 (88.2% of the original data), and testing set holding the remaining data from 2010 to 2017 (11.8% of the original data). Models were trained on the training sets before being applied to the unseen data in the test sets. Table 1 gave a sample of training set for both attributes (consumption and revenue), and Table 2 gave sample of testing set for both attributes (consumption and revenue).

Table 1. Sample of the training set from 1949 to 2009

Year	Electricity End Use (Million Kilowatt-hours)
1949	254511.334
1950	291443.326
1951	330284.583
1952	356163.96
1953	396216.896
1954	424163.792
1955	496748.053
1956	546280.491
1957	575820.175
1958	587863.052

To choose the appropriate forecasting algorithms, I first evaluated and compared the results from each algorithm with the original data we have. Then three forecasting algorithms were run on the training set

and with horizon value 8 to reach 2017 to compare it with the actual data in testing set.

Table 2. Sample of the testing set from 2010 to 2017

Year	Electricity End Use (Million Kilowatt-hours)
2010	3886751.617
2011	3882600.217
2012	3832306.296
2013	3868329.758
2014	3903274.151
2015	3900159.908
2016	3902306.027
2017	3820031.572

3.1. Evaluating algorithms over dataset

First step, evaluation process of the selected three algorithms done over the datasets and compute the mean absolute percentage error MAPE for each algorithm.

$$MAPE = (((Actual - Predicted) / Actual) * 100)$$

This evaluation provides insights in recommending changes to existing models to reduce deviations in future forecasts.

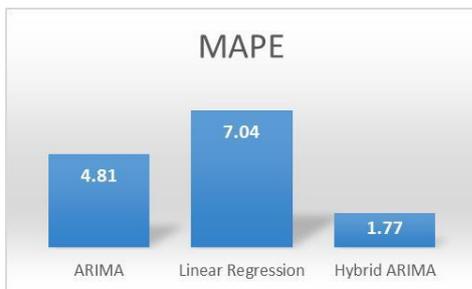


Figure 2. Calculated MAPE for three algorithms

Figure 2 presents the values of MAPE for the three selected forecasting algorithms. It's observed that Hybrid ARIMA was the best forecasting algorithm to be used over datasets (MAPE 1.77%).

Now after detecting the most appropriate algorithm, I'm going to forecast 5 years in advance to predict the electricity usage.

3.2. Forecasting electricity end use from 2018 to 2022 using Hybrid ARIMA

Figure 3 shows that the trend of electricity end use for the next 5 years is increasing in general.

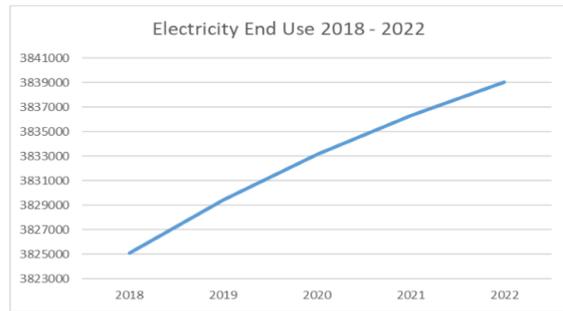


Figure 3. Forecasted results for Electricity end use from 2018 to 2022

3.3. Results Deviation

Table 3 Shows that the deviation of the sum of forecasted 5 years (2018 - 2022) of electricity end use with comparison to the last 5 years (2013 - 2017). The deviation is decreasing by 1.21%.

Table 3 Deviation of forecasted years compared to last 5 years

Year	Electricity End Use (Million Kilowatt-hours)
2013 - 2017	19394101.42
2018 - 2022	19162970
Deviation%	-1.21

Figure 4 presents the general chart of electricity end use from 1949 to 2022, which shows that the average curve after 2018 is lower than the average curve of the last 5 years (2013 - 2017).

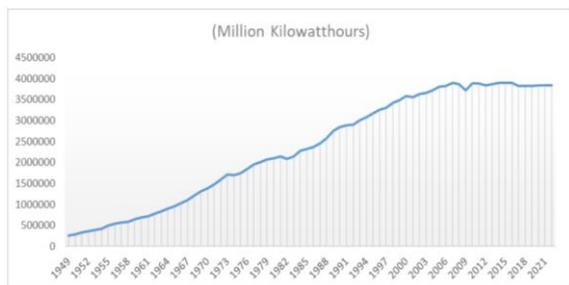


Figure 4. General view for electricity end use from 1949 to 2022

4. Discussion

From Figure 4, it is noticed that average electricity end use for the next five years will decrease, and this is maybe due to many people or

building units depends more on the solar energy to generate electricity and storing power sources in batteries. In general, the average of the next five years is decreasing by 1.21% with comparison to last five years.

5. Conclusion

In this paper, the electricity end use was forecasted in the medium-term future. Three main forecasting algorithms were used ARIMA, Hybrid ARIMA and Linear Regression. I applied them on dataset collected from BP Global. It's found that Hybrid ARIMA is the best algorithm to be used for forecasting electricity use with MAPE 1.77% and future deviation with -1.21%.

6. References

- [1] Generation, I.F.E., (2014). Electricity Generation.
- [2] Global, B., (2016). Electricity generation (terawatt-hours).
- [3] Aggarwal, C.C., (2015). Data mining: the textbook: Springer.
- [4] Billings, R.B. and C.V. Jones, (2011). Forecasting urban water demand. American Water Works Association.
- [5] De Giorgi, M.G., et al., (2015). Error analysis of hybrid photovoltaic power forecasting models: A case study of Mediterranean climate. Energy conversion and management, 100: p. 117-130.
- [6] De Giorgi, M., M. Malvoni, and P. Congedo, (2016). Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine. Energy. 107: p. 360-373.
- [7] Kaytez, F., et al., (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. International Journal of Electrical Power & Energy Systems. 67: p. 431-438.
- [8] Jurado, S., et al., (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy. 86: p. 276-291.
- [9] Chae, Y.T., et al., (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings. 111: p. 184194.
- [10] Weron, R., (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International journal of forecasting. 30(4): p. 1030-1081.