Forecasting Electricity Demand in Duke Energy Carolinas

By: Ina Shih-Hsuan Liao

Advisor: Luana Marangon Lima

December 15th, 2023

Table of Contents

1. Introduction

The growth in the population and socioeconomic and technological advancements have raised the electricity demand in the past few decades. The U.S. electricity consumption has increased from 0.4 trillion kilowatt-hours (kWh) to 4 trillion kWh (EIA, 2023). Accurate load forecasting has been critical to optimize the power system operation with increasing electricity demand.

Load forecast shows a strong relationship with weather variables on load demand, such as temperature, dew point, cooling degree days (CDD), heating degree days (HDD), wind speed, and holiday period. Research shows that the higher the degree days, the difference between the daily mean and standard temperatures for a location, the higher the energy used for space heating or cooling (De Rosa et al., 2014). To achieve better forecast results, all factors affecting the load demand can be considered inputs for forecast models.

Several load forecasting methods have been proposed to increase predictive accuracy. Statistical methods include linear regression, autoregressive moving average (ARIMA), exponential technique, and stochastic time series. These statistical methods are based on mathematical combinations of previous and current load values, and most of them could not achieve the desired forecasting accuracy due to some of their limitations (Mir et al. 2020; Deb et al. 2017). For example, linear regression methods depend on past historical data and poorly explain non-linear changes. In addition, ARIMA models do not consider external factors as input and generate results only based on past and current data.

To overcome the limitation of statistical methods, researchers have proposed artificial intelligence-based methods, including support vector machine (SVM), artificial neural network (ANN), and fuzzy logic. ANN is the most common and widely accepted method for predicted problems, and it provides the least forecasting error compared to traditional forecasting models for non-linear output. However, ANN also suffers from under-fitting or overfitting problems a slow learning rate, and it is restricted to providing insights into the relationship between the input variables and their determinants (Bedi & Toshniwal, 2018). In the past few years, researchers have combined ANN with various population-based optimization learning algorithms for adjusting network structure and network parameters to enhance the accuracy of load forecasts.

Table 1 shows a literature review of electricity load forecasting models based on two broad categories: statistical methods and artificial intelligence-based methods. The literature review findings show that statistical methods work well for long-term load forecasting, whereas artificial intelligence-based methods have better performance for short-term forecasting (Mir et al, 2020; Deb et al., 2017). On the other hand, hybrid models were widely used in forecasting and outperformed individual methods; however, none of the models clearly outperformed the others. Robust solutions are thus needed to analyze and forecast the electricity load patterns effectively.

The objective of this study is to inform the accuracy of existing forecasting models and provide insights for improvements. The study aims to answer the following questions:

(1) How would weather conditions and calendar events influence electricity demand?
(2) Does the proposed model have a lower forecasting error compared to models employed by balancing authorities or investor-owned utilities? If so, what are the primary factors that contribute to forecasting errors in their models?

The study proposes an ARIMA-ANN model to handle the nonlinear electricity demand. This study focused on the hourly forecast error of Duke Energy Carolinas (DUK), a vertically integrated electricity company that provides electricity within the Carolinas region. Load characterization is performed at three different levels: daily analysis, time-span analysis, and seasonal analysis to predict average and peak electricity demand for the season and day.

2. Methods and Data

2-1. Input Data

To investigate the factors that influence hourly forecast error, the dependent variables include temperature, dew point, CDD, HDD, wind speed, and calendar event. In contrast, the independent variable is the load demand. The hourly weather data of Charlotte, NC, between January 1st, 2012, and December 31st, 2022, was retrieved from the National Solar Renewable Energy Laboratory (NREL) National Solar Radiation Database. Charlotte was selected because it is near the border of South Carolina and North Carolina. CDD and HDD refer to "the difference between the daily mean temperature, calculated by the sum of the highest and lowest temperatures of the day divided by two, and 65°F, a standard temperature in the United States (EIA, 2023)". If the daily mean temperature is higher than 65°F, then the difference between the mean value and 65 is referred to as HDD; in other words, if the temperature is lower than 65°F, the difference represents CDD. Holidays and weekends are assigned as a binary variable to test if holidays and special occasions corresponded with higher forecast errors resulting from unexpected increased electricity demand. The DUK's hourly load forecast and actual demand are collected from the Energy Information Administration (EIA) Hourly Electric Grid Monitor from 2012 to 2022. DUK provides a day-ahead electricity demand forecast for every hour for the next day. Forecast errors are calculated by subtracting hourly electricity demand from forecasting demand. Positive errors represent over-forecast, whereas negative errors represent under-forecast. The raw and processed data files are uploaded on Git Hub (https://github.com/inaliao/Ind_Study/tree/main).

2-2. Data Preprocessing

Figure 1 shows the flowchart of the proposed ANN-ARAMI model. The multicollinearity analysis will determine whether independent variables are highly correlated. If multicollinearity is identified, one of the correlated variables will be removed to avoid bias in the model. The auto-correlation function (ACF) and the partial auto-correlation function (PACF) will then be used to determine if the autocorrelation exists between forecasting errors. If the autocorrelation is found, the first-order differencing approach, subtracting each observation from its previous one, will mitigate the autocorrelation in the data.

Time series analysis assumes the data to be stationary, meaning that the forecast error means that variances and autocorrelation structure remain constant over time. The decompose function will be used to separate seasonality, trend, and irregular components present in the time series data first, and the Mann-Kendall test, Augmented Dicker Fuller (ADF) test, and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test will be used to test the stationarity of the de-seasonalized time series. However, due to the high dimensionality of electricity demand, a data cluster will be used to identify the patterns in the electricity consumption data before conducting the time series analysis.

2-3. Data Clustering

Data clustering algorithms can be broadly divided into three classes: Partitioning-based clustering, Hierarchical clustering, and Density-based clustering algorithms. Each type of method has its way of defining groups and similarity measures. This study will use the K-mean algorithm to assign the months that follow a similar consumption pattern into groups based on a distance measure and detect outliers.

There are various distance measures, including Euclidean distance, Dynamic Time Wraping (DTW), and LB_Keogh distance. While Euclidean distance can be useful, it may not be the best option for time series data because it doesn't always account for distortions in the time domain. DTW distance, on the other hand, can help find an optimal non-linear alignment between time series, but it might work better for larger data sets. LB_Keogh distance can help overcome some of the storage issues that come with DTW distance by defining upper and lower bounds for the time series (Bedi & Toshniwal, 2019). This paper will use the elbow curve method to determine which will produce the best clustering result. This method involves running the K-mean algorithm with different values of k and calculating the sum of squared estimate of errors (SSE) for each one. Then, the elbow curve method works by executing iterative runs of the K-mean algorithm with a range of k values and computing the SSE of each k-value. A plot of k-value versus SSE will be generated to see where the error value changes most dramatically.

Data clustering also allows us to identify the load trend characterizations. In this paper, the load trend characterization will be performed at three levels: daily analysis, time-span analysis, and seasonal analysis. Daily analysis will demonstrate the change in the power demand pattern over the day in a week. The time-span analysis will determine the changes in electricity demand of each day in different time intervals: interval 1 (00:00 am to 06:00 am), interval 2 (06:00 am to 12:00 pm), interval 3 (12:00 pm to 06:00 pm), and interval 4 (06:00 pm to 00:00 am). Additionally, seasonal analysis will detect patterns in energy consumption during peak and off-peak hours. According to DUK, shoulder-peak periods are from 9:00 a.m. to noon and 5:00 p.m. to 8:00 p.m., Monday through Friday. The remaining time is considered off-peak hours. Data clustering can help us find sub-groups with similar demand trends, which will be beneficial in developing machine learning models in further analysis.

2-4. Proposed ANN-ARIMA Model

This paper will use ANN-ARIMA to forecast future electricity demand in DUK regions. To construct the ARIMA model, I will select parameter combinations with the lowest AIC and use them as inputs for the ANN model. There are two types of ANN: feed-forward neural network (FNN) and recurrent networks. While recurrent networks may have better prediction performance due to their ability to use internal memory to process arbitrary input sequences, this study will experiment with the simpler FNN and make any necessary adjustments accordingly.

This research will use a three-layer feed-forward network, where the inputs are connected to the output through a middle layer. Assuming the input data is univariate time series, meaning that the time series consists of single (scalar) observations recorded sequentially over equal time increments, the output is the value for the next period (Equation 1).

Equation 1 The output value in neural network.

$$S(t+1) = f(s(t), \ldots, s(t-n))$$

Where n is the number of past values of variables, and f is a nonlinear function approximated by a multilayer feed neural network.

To determine the ideal structure for the neural network, I will begin with a single hidden layer and conduct experiments to assess whether additional layers yield significant improvements. Moreover, the number of neurons in the hidden layer that yields the most accurate forecasts will be contingent on the Normalized Mean Square Error (NMSE). By utilizing this metric to evaluate forecasting models, I will identify the optimal configuration for the neural network in terms of both model selection and neuron count.

3. Expected Results and Deliverables

This project aims to develop a forecast model that will allow us to gain a deeper understanding of how weather conditions and calendar events impact electricity demand and ultimately help us reduce forecasting errors. The final deliverables include an ANN-ARAMI model and a report analyzing the additional factors that drive non-linear electricity growth, providing valuable insights to inform DUK's future advancements in forecasting models and infrastructure planning.

4. References

Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy*, 238, 1312-1326

Bedi, J., & Toshniwal, D. (2018). Empirical mode decomposition based deep learning for electricity demand forecasting. IEEE access, 6, 49144-49156.

De Rosa, M., Bianco, V., Scarpa, F., & Tagliafico, L. A. (2014). Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. *Applied energy*, 128, 217-229.

Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902-924.

Jiang, P., Li, R., Lu, H., & Zhang, X. (2020). Modeling of electricity demand forecast for power system. *Neural Computing and Applications*, 32, 6857-6875.

Jurado, S., Nebot, À., Mugica, F., & Avellana, N. (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. *Energy*, 86, 276-291.

Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., Refaat, S. S., & Oueslati, F. S. (2019). A novel approach based deep RNN using hybrid NARX-LSTM model for solar power forecasting. arXiv preprint arXiv:1910.10064.

Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y., & Imran, M. (2020). A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons. *Sustainability*, 12(15), 5931.

National Renewable Energy Laboratory. NSRDB: National Solar Radiation Database. Retrieved October 10th, 2023, from https://nsrdb.nrel.gov/data-viewer

Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. (2017). Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied soft computing*, 54, 246-255.

U.S. Energy Information Administration (n.d.). Units and calculators explained: Degree days. Retrieved October 10th, 2023, from https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php

U.S. Energy Information Administration (n.d.). Hourly Electricity Grid Monitor. Retrieved October 10th, 2023, from https://www.eia.gov/electricity/gridmonitor/dashboard/electric_overview/balancing_authority/D UK

U.S. Energy Information Administration (n.d.). Electricity explained: Use of electricity. Retrieved September 28, 2023, from https://www.eia.gov/energyexplained/electricity/use-of-electricity.php.

Wang, Y., Wang, J., Zhao, G., & Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy*, 48, 284-294.

Table 1 Summary table on electricity load forecasting literature review.

| Category	Models	Data	Results	Reference
statistical method & artificial intelligence-based method	ARIMA, random forest, neural network, fuzzy	Hourly electricity consumption in three building of the University Politecnica de Catalunya from 2011 to 2012.	Random forest, neural networks, and fuzzy inductive reasoning are proposed to perform short-term electric load forecasting (24 hours).	Jurado (2015)
statistical method	Seasonal ARIMA, residual modification	Electricity load data from 2006 to 2010 in northwestern China.	The proposed model performs better than the single seasonal ARIMA model.	Wang et al. (2012)
artificial intelligence-based method	Long Short Term Memory network (D-FED)	Electricity consumption data from 2013 to 2018 in Union Territory Chandigarh, India.	The D-FED can handle non-linear complexities, short-term and long-term dependencies of the electricity consumption time series data and has minimum prediction errors.	Bedi & Toshniwal (2019)
artificial intelligence-based method	Empirical mode decomposition--based deep learning approach	Electricity consumption data from 2013 to 2018 in Union Territory Chandigarh, India.	Empirical mode decomposition with the deep long short-term memory network outperforms other regression models for electricity demand time series forecasting.	Bedi & Toshniwal (2018)
artificial intelligence-based method	Nonlinear autoregressive exogenous network with short-term memory	2016-2019 weather data, including ambient temperature, wind speed, irradiation, and relative humidity.	Hybrid models are a combination of two or more prediction models, which enhance accuracy since the feature of each model will be transferred.	Massaoudi et al (2019)
artificial intelligence-based method	Integrated model that combines adaptive Fourier decomposition method	New South Wales and Queensland over February, 2014, and from South Australia over August, 2013.	This paper mainly focuses on short-term electricity demand, and thus the methods might not be applicable for long-term prediction.	Jiang et al. (2020)
artificial intelligence-based method	Empirical mode decomposition--based deep learning approach	the data sets of year 2013 from New SouthWales (NSW), Tasmania (TAS), Queensland (QLD), South Australia (SA) and Victoria (VIC)	Deep learning algorithms show their advantages in dealing with nonlinear features when the forecasting horizon increases.	Qiu et al. (2017)
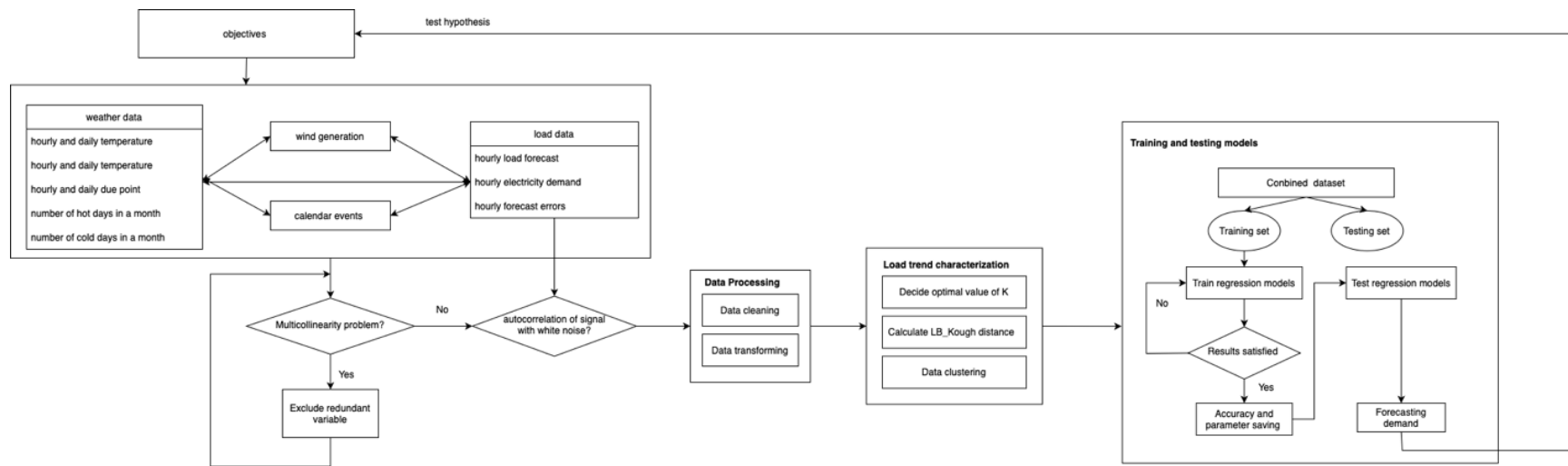
Figure 1 Flowchart of the proposed method.