Electricity Load Forecasting (Panama Case Study)

Mid-term Peak Load Forecasting Using SARIMA and STL-LSTM Techniques (Panama Case Study)

Mohsen Kaveh
Department of Electrical and Computer Engineering
Western University
London, Canada
mkaveh@uwo.ca

***Abstract*—Accurately predicting peak demand is crucial for energy suppliers to meet the consumer loads connecting to the power grid. This study explores the performance of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Seasonal Trend decomposition using Loess (STL) along with long short-term memory (LSTM) as peak demand forecasting models. The study validates the models using demand data from Panama's power system between 2015 and 2018, taking into account historical demand, holidays, and weather variables as input features. The results indicate that both SARIMA and STL-LSTM models can predict future demand with less than 8% mean absolute percentage error (MAPE). Since the threshold for defining highly accurate demand prediction models for MAPE is 10%, both proposed models are highly accurate. Therefore, the proposed models can help energy suppliers make informed decisions and plan for future demand.**

Keywords—load forecasting, machine learning, SARIMA, STl-LSTM, electricity demand prediction

I. Introduction

Electricity is a crucial component in every country, and over the years, energy consumption has been consistently increasing worldwide, mainly due to population growth and economic development [1]. Balancing the supply and demand of electricity in real-time has become more challenging for power systems, especially during the peak periods when demand is at its highest, and consumer patterns fluctuate throughout the day and year. Modern power systems typically employ an expensive mix of reserve power plants and demand response programs to mitigate the peak demand. Power system planners are now concerned about accurately forecasting the peak load to ensure an uninterrupted power supply. Accurate short to mid-term forecasts will enable planners to determine the required generation capacity for the power grid.

To anticipate forthcoming energy demand, various methods are utilized, including statistical, artificial intelligence, and hybrid [2]. Statistical methods, like regression analysis and time series models, are traditionally used to recognize patterns in past data, which provides a strong foundation for predicting future loads. In recent times, artificial intelligence methods, such as artificial neural networks and support vector machines, have gained popularity for their ability to model intricate, nonlinear relationships and adjust to changing patterns. Finally, hybrid methods combine the advantages of both statistical and artificial intelligence methods, utilizing the robustness of conventional techniques and the flexibility of modern computational methods.

The energy sector has yet to fully explore the potential of deep learning in peak load forecasting, which is an important research topic. This raises the question of how deep learning methods can improve support for these complex patterned areas [3]. This study aims to create a forecasting model using LSTM to predict mid-term daily peak demand and compare it to time series models like SARIMA.

The rest of this paper is organized as follows. Section II describes the dataset used for this research. Section III explains SARIMA and STL-LSTM models. Section IV shows the results of both models. Finally, the results will be discussed in the conclusion.

II. Dataset

The dataset used in this study was sourced from the National Secretary of Energy of Panama and made available on the Kaggle website [4]. The dataset includes 24-hour daily load consumption data from January 2015 to June 2020. The dataset comprises several features, with the time of day being the most significant. Additionally, the calendar data of school periods and holidays also influence power consumption levels. Lastly, weather variables such as temperature, relative humidity, precipitation, and wind speed, for three primary cities in Panama, namely Tocumen, San Miguelito, and David, were considered.

III. Methodology

In order to forecast mid-term load accurately, a dataset of 24-hour daily recorded Panama loads from January 2015 to the end of 2018 was utilized. This dataset was first converted to the maximum daily load and then divided into three sections, with 94% used for training, 8% for evaluation, and the remaining month of data for testing. The training, evaluation, and testing samples consist of 1374, 58, and 30 samples respectively. Mean absolute percentage error (MAPE) is the commonly used metric for evaluating demand forecasting models, with a score below 10% indicating high accuracy [5].

A. SARIMA

Seasonal Autoregressive Integrated Moving Average (SARIMA) is a time series model that is widely used to capture seasonal patterns in data. SARIMA is an extension of ARIMA and is effective in handling data with trends, seasonality, and non-stationary. The model is equipped with parameters for seasonal differences, as well as autoregressive and moving average components. It can also be improved by incorporating exogenous variables. SARIMA's efficiency has been demonstrated in various publications, such as its ability to predict mid-term load forecasting. SARIMA's advantage lies in its capacity to handle multiple seasonal periods, making it ideal for modeling data with numerous seasonal cycles [6].

The study utilized the SARIMA model to analyze the normalized training data. To be specific, the non-seasonal component was fitted with an order of (1,1,1) while the seasonal component was fitted with a seasonal order of (1,1,1,7) with a 7-day cycle. The model was then utilized to forecast values for the training, validation, and testing datasets. The validation data was utilized to optimize the model parameters while the testing data was utilized to evaluate the model's predictive accuracy. The model's performance was assessed by comparing its predictions to the actual values. This methodology was employed to ensure a comprehensive analysis of the data.

B. STL

STL (Seasonal Trend decomposition using Loess) is a statistical technique that breaks down a time series into three parts: seasonal, trend, and remainder [2]. It utilizes locally weighted regression, commonly known as Loess, to determine the seasonal and trend components, while the remainder is derived by subtracting the estimated seasonal and trend components from the initial time series. The trend component signifies the long-term trend of the time series, while the seasonal component represents the cyclical variations. Lastly, the remainder component captures the unpredictable fluctuations.

In the given research, the STL approach has been applied to break down the data into three distinct parts, which were analyzed individually for the training, validation, and testing sets. These derived components were utilized to predict upcoming values of the time series.

C. LSTM

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that has gained popularity due to its ability to model long-term dependencies in sequential data. Its architecture is more complex than traditional RNNs, incorporating gates and memory cells that enable LSTMs to selectively remember or forget information from previous time steps, thereby preventing the vanishing gradient problem. Moreover, LSTMs can store information for extended periods, making them ideal for tasks that require modeling long-term dependencies. Several load forecasting studies have demonstrated the efficacy of LSTMs [7].

To prepare the dataset for LSTM analysis, it is first normalized between 0 and 1 using Min-Max-Scaler. A look-back window of 1 is then applied, where the input sequence is a sliding window of time steps and the output is a single value. The data is reshaped to match the input shape of the LSTM layer. The model contains a hidden layer with 64 units, followed by a dense output layer with a linear activation function. The Adam optimizer and mean squared error loss function are used to compile the model. The model is trained and validated using prepared training and validation datasets, respectively. During the training process, the model's loss and validation loss can be assessed to evaluate its performance.

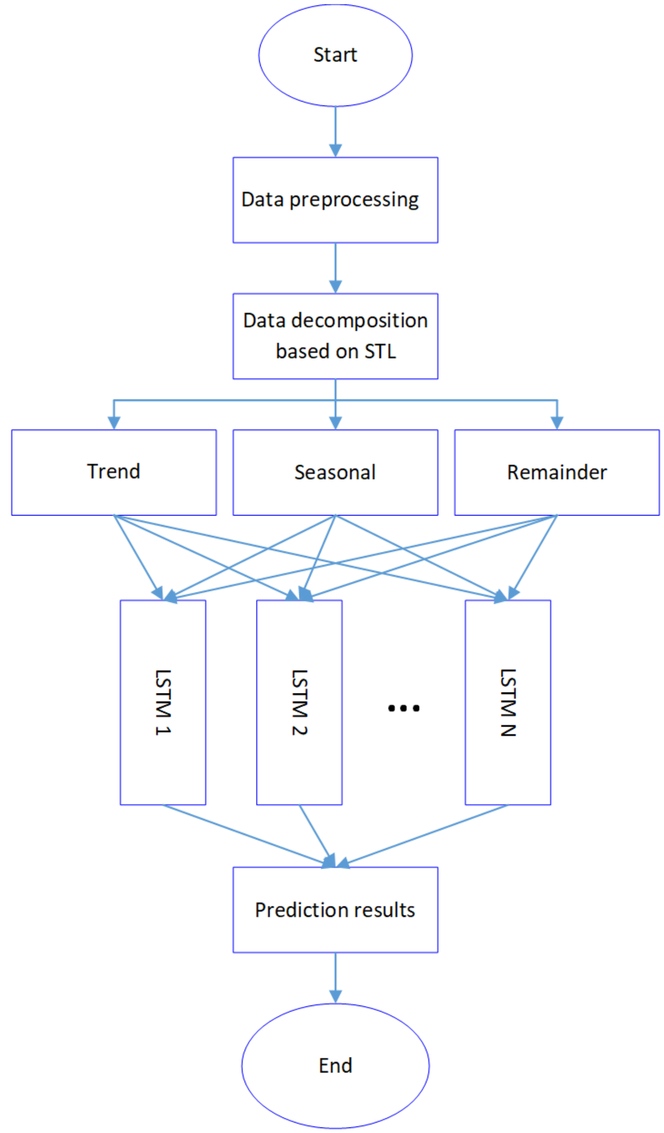

Fig. 1. The flow chart of the STL-LSTM model

Fig. 1. shows the flow chart of the STL_LSTM model in general which N is equal to 64 in this study.

IV. Results

Since the daily peak daily load forecasting is the goal of this research, the data will be processed to calculate the maximum load of each day. Fig. 2. shows that peak daily data has a seasonal feature which is important for the SARIMA model.

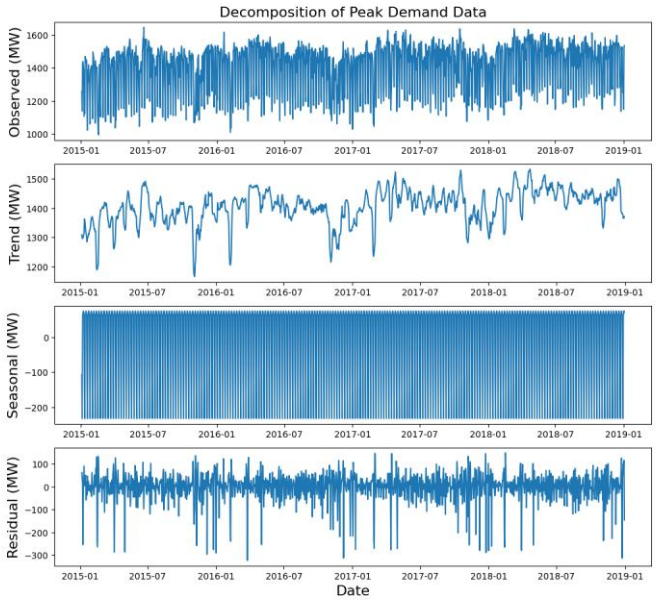

Fig. 2. Observed, trend, seasonality, and residual plots of peak demand

A. *SARIMA model results for daily peak demand forecasting*

The cross-validation score for the SARIMA model is 1.75, and Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) for the testing set are 120.27, 63.79, and 5.07 respectively. MAPE is lower than 10% and shows the model's highly accurate prediction.

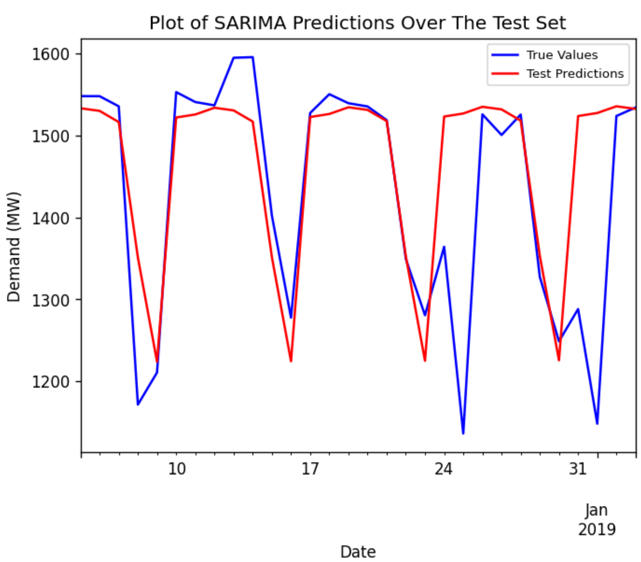

Fig. 3. Prediction results of the SARIMA model for daily peak demand

Fig. 3. indicates that SARIMA had a reasonable daily peak demand data prediction.

B. *STL-LSTM model results for daily peak demand forecasting*

LSTM model trained for 120 iterations, and training loss per iteration is as follows.

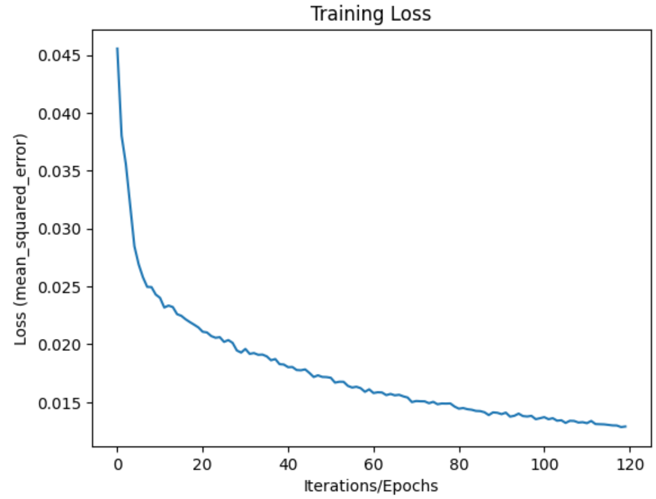

Fig. 4. Training loss of LSTM model over iteration number

After predicting testing set by STL-LSTM, RMSE, MAE, and MAPE are 140.06, 98.69, and 7.35 respectively. MAPE is lower than 10% and shows the model's highly accurate prediction. MAPE for STL-LSTM model used in [2] was slightly better than this study, 5.9%. This is because it was used for short-term prediction not mid-term peak demand forecasting like this study.

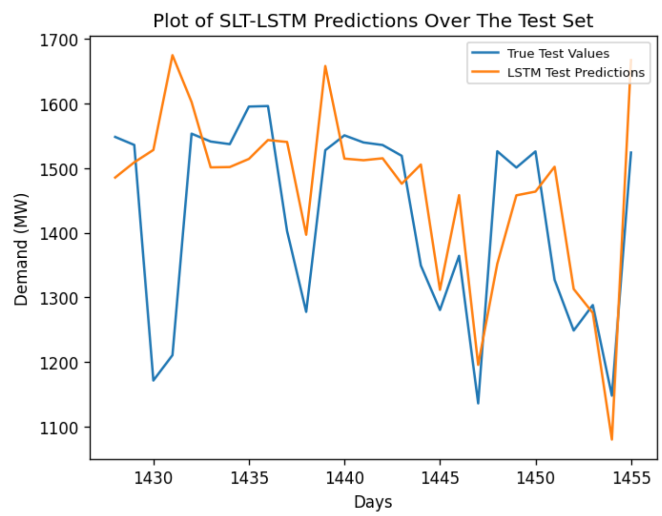

Fig. 5. Training loss of LSTM model over iteration number

Fig. 5. shows that STL-LSTM had a reasonable prediction for daily peak demand data.

V. Conclusions

This paper aimed to examine the effectiveness of SARIMA and STL-LSTM models for predicting daily peak demand in Panama. The study utilized data from January 2015 to the end of 2018, as well as information on holidays and weather patterns in the country. The SARIMA model utilized a basic ARIMA model with a 7-day seasonality, while the STL model was followed by an LSTM model with 64 neuron units. Both models had a MAPE of less than 8%, indicating high accuracy. Future research could explore the potential benefits of incorporating other machine learning models, such as XGBoost, into the LSTM model.

REFERENCES

[1] N. Ahmad, Y. Ghadi, M. Adnan and M. Ali, "Load Forecasting Techniques for Power System: Research Challenges and Survey," in *IEEE Access*, vol. 10, pp. 71054-71090, 2022.

[2] M. Fan, Y. Hu, X. Zhang, H. Yin, Q. Yang and L. Fan, "Short-term Load Forecasting for Distribution Network Using Decomposition with Ensemble prediction," *2019 Chinese Automation Congress (CAC)*, Hangzhou, China, 2019, pp. 152-157.

[3] J. Zhu, H. Dong, W. Zheng, S. Li, Y. Huang, and L. Xi, "Review and prospect of data-driven techniques for load forecasting in integrated energy systems," *Applied Energy*, vol. 321, p. 119269, 2022,

[4] S. Shahane, "Electricity Load Forecasting (Panama case study)." [Online]. Available: https://www.kaggle.com/datasets/saurabhshahane/electricity-load-forecasting?resource=download&select=continuous+dataset.csv. [Accessed: 10-Feb-2023].

[5] E. Vivas, H. Allende-Cid, and R. Salas, "A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score," *Entropy*, vol. 22, no. 12, p. 1412, Dec. 2020

[6] H. Musbah and M. El-Hawary, "SARIMA Model Forecasting of Short-Term Electrical Load Data Augmented by Fast Fourier Transform Seasonality Detection," *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, Edmonton, AB, Canada, 2019, pp. 1-4.

[7] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," in *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.