ITM Web of Conferences **88**, 01002 (2026) https://doi.org/10.1051/itmconf/20268801002
AIBDCS 2026

Short-Term Electricity Load Forecasting for the Panama Power System Using an ARIMA Baseline

Zexin Wang*

School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, 999077, China

Abstract. The forecasting of short-term load (STLF) is an important part of the functioning of a power system, as it aids in discovering the dispatch plans and taking the strain off. This paper researches hourly short-term electricity load forecasting of the Panama power system based on a univariate Autoregressive Integrated Moving Average (ARIMA) model as a clean statistical reference. The national demand in 20152020 was chosen as the dataset obtained in the publicly available Kaggle repository. It was tested on a 24-hour-ahead forecasting problem with a train-test split that was based on chronology. The ARIMA (2,1,2) model gained the reliability of the held-out test window of 48.88 MWh, RMSE of 60.50 MWh, and MAPE of 4.32%. Although this error rate is widely similar to classical statistical thresholds in short-horizon load forecasting papers, it is nevertheless significantly larger than the accuracy commonly found by the more advanced machine learning and deep learning models, specifically when using exogenous variables. The results hence result in ARIMA being an efficient base of clarity and benchmarking, with the development being more expressive models to route the non-linearities in electricity demand.

1 Introduction

The short-term load forecasting (STLF) has become a significant topic in the power systems literature, with the reason being the significance of the approach in operational planning and the reliability of systems.

Early research on short-term power load forecasting mainly relied on traditional statistical methods and time series models. These methods have certain advantages in handling linear series, but their predictive performance is often limited when faced with nonlinear characteristics, seasonal variations, or random disturbances, as demonstrated by Alam et al., who compared multiple time-series models and reported performance degradation under nonlinear load patterns [1].

In the 21st century, machine learning technology has gradually been introduced into the field of short-term load forecasting. Models such as support vector machines can effectively study the nonlinear relationship between historical load data and external influencing factors. For example, Toufiq et al. incorporated weather variables such as temperature and humidity into machine learning models and showed that these approaches significantly outperform traditional statistical models in prediction accuracy [2].

Nevertheless, the ARIMA model still occupies an irreplaceable benchmark position. Its good interpretability, computational efficiency, and robustness under limited data make it not only the basis for comparing subsequent advanced algorithms, but also a key module in hybrid models for capturing linear components.

This report describes the work of building an ARIMA model based on the Kaggle "Short-term Electricity Load Forecasting (Panama)" dataset for short-term load forecasting. The report aims to establish a reproducible forecasting workflow, visualize time-series patterns and forecast performance, and evaluate accuracy. Make ARIMA a baseline for comparison with more advanced models.

2 Method

2.1 Dataset and target definition

The data used is the release of Kaggle “Short-term electricity load forecasting (Panama)”, which offers a pre-processed continuous table of hourly observations of January 2015 to June 2020 [3]. The rows are identified by timestamps (datetime), and they have national demand (in MWh). The analysis presented in this report has employed the univariate target series to develop the baseline forecasting model, which only relies on the previous demand values.

The CSV file was imported into Python, where the datetime field was converted to timestamps and set as the index. Records were then sorted chronologically to preserve the time sequence for modelling and evaluation. Because the dataset was already provided as a single continuous file, no merging was required. The modelling workflow assumes an hourly sampling frequency consistent with the dataset description, and the time-series library automatically inferred this hourly frequency from the index.

* Corresponding author: wzx1553@gmail.com

2.2 ARIMA model specification

In this report, the basic model is an Autoregressive Integrated Moving Average model, which is abbreviated as ARIMA (p, d, q). ARIMA is a combination of autoregressive (order p), differencing (order d) to overcome non-stationarity, and moving-average (order q) to explain the residual autocorrelation. ARIMA has often been used as a reference in comparing STLFs because it can be interpreted and is computationally cheap [4].

ARIMA (2,1,2) was fitted on the training series, and a 24-step-ahead point forecast was obtained. The forecast result was matched with the timestamps of the held-out 24-hour test.

The ARIMA order used in this implementation was chosen according to both past empirical data on the load forecasting literature and according to data analysis in Panama. Previous studies have extensive experience in discovering the low-order ARIMA structures and their combinations around (2,1,2) to become a good base de-similarity in the prediction of electricity demand with a short horizon. Apparently, according to this empirical scale, several candidate models were trained, and the predictions of these models were checked and assessed against the training data. The most appropriate specification between the estimated ones was ARIMA (2,1,2) as indicated by the best validation specification, MAPE, and therefore, it was selected to execute the final exercise, 24-hour ahead forecasting. To address the non-stationarity of the demand series, a d=1 differencing order was used.

2.3 Evaluation metrics

Analysis is based on a simple and open forecasting, both in terms of a train/test split and numerical assessment. A single baseline model is constructed based on the national series of electricity demand to forecast the future 24 hours at hourly forecasts. To prevent leakage of data, the dataset is divided chronologically, with all the observations except the last 24 hours of training and the latter making the test set. Forecasting Time-ordered splitting of this nature is natural in forecasting research, as any random shuffling would destroy the time dependence and provide an invalid evaluation [5].

Measured the accuracy of forecasts using five standard error measures of performance, which are complementary measures of performance. They are MAE, RMSE, MAPE, MSE, and R^2. These indicators are typically presented in the load forecasting research and reviews, and they enable the comparison of baseline performance with other approaches [6].

3 Results and discussion

3.1 Time-series overview

Fig. 1 plots the complete series of national demand from 2015-2020 at the hourly levels. The series exhibits good time-periodicities, which are in line with the behaviour of the load of electricity, such as repetitive daily cycles and extended run variations. The plot also suggests that the levels of demand are not constant with time, and this prompts differencing in the ARIMA specification. There are occasional sudden deviations and sharp drops seen in the long horizon view, which could be an indication of abnormal operating conditions, data artifacts, or infrequent events. These irregularities are referred to as difficulties in load forecasting since they can corrupt fitted parameters, and augment error in the short horizon if they happen close to the forecast source [7]. After first-order differencing (d = 1), the series becomes more stationary and fluctuates around a stable mean, which satisfies the stationarity requirement of the ARIMA model.

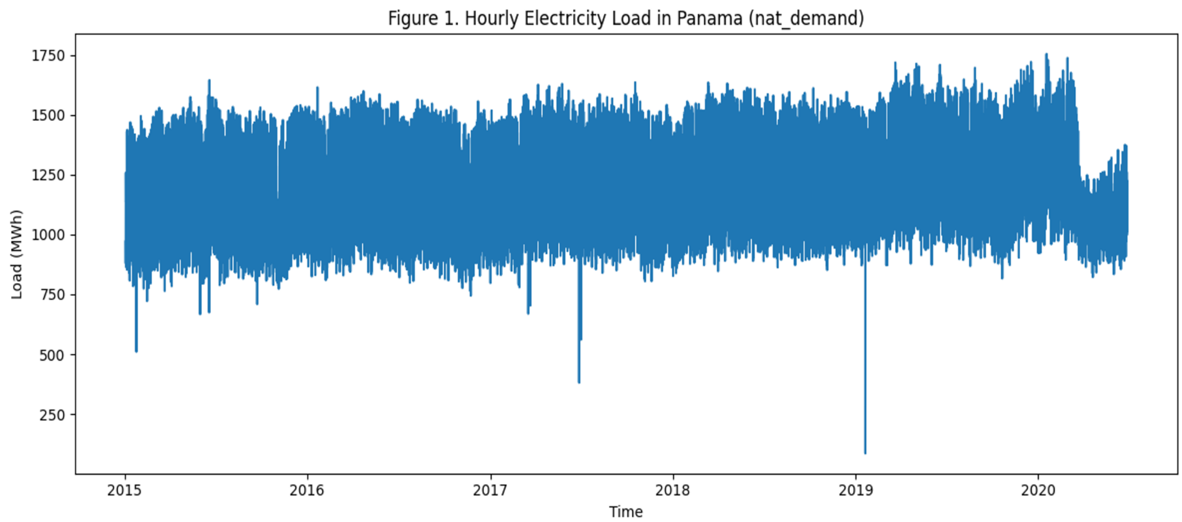

Fig. 1. Hourly Electricity Load in Panama (Photo/Picture credit: Original).

Fig. 1 has a visual form that justifies the application of a statistical baseline since there was a high autocorrelation in the series and repetitions that can be modelled using the past demand. Simultaneously, the number of drivers considered by the dataset (weather and calendar variables) indicates that further-developed models may result in greater performance of conditioning on exogenous variables. This is consistent with the companion study that presented the data as the baseline of the forecasting model training and testing on a larger scale than the operator baselines [8].

3.2 Forecast-versus-actual behaviours

Fig. 2 compares the 24-hour ahead forecast generated by the ARIMA (2,1,2) model with the actual load observed in the test period. The horizontal axis represents the date, and the vertical axis represents the load. The blue line shows the last 72 hours of the training data, the orange line represents the actual load during the 24-hour test window, and the green line indicates the model's forecast for the same period. The Fig. encompasses the remaining 72 hours of the training series to offer details of the recent dynamics to the effects up to the test period. The projection takes the general trend of demand and repeats short-term movement, meaning that the model captures recent trends of autocorrelation.

Nevertheless, the prediction is less accurate than the actual one and does not entirely coincide with the sharp troughs and peaks of intraday. This is expected of univariate linear models, which can be under-reactive to rapid variations that require drivers that are not present in the target history in isolation. Weather and calendar variations can cause shifts in magnitude and the time of the various peaks in STLF, and models with exogenous regressors or non-linear structures tend to be more accurate in volatile times. This is a key drive in the use of feature-based machine learning in the present-day STLF and deep learning sequence models [9].

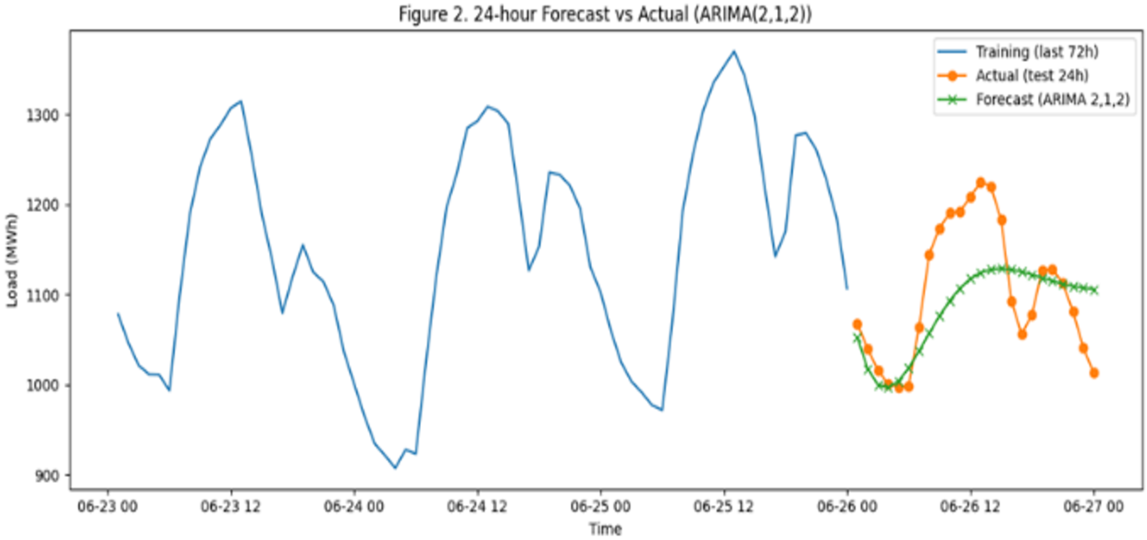

Fig. 2. 24-hour Forecast vs Actual (ARIMA (2,1,2)) (Photo/Picture credit: Original).

Since this report has deliberately excluded the exogenous variables in order to offer a clean baseline, the forecast behaviour given in Fig. 2 is quite reasonable and expected. In practice, an ARIMA base is commonly taken to have a minimum performance baseline before using more elaborate methods. Comparative analysis of random forest and other machine learning models outperforms ARIMA in instances of adequate explanatory variables, especially when non-linear load-weather associations are modelled [4].

3.3 Quantitative accuracy

Table 1 shows MAE, RMSE, MAPE, MSE, and R2 of the 24-hour forecast. The held-out test set and the ARIMA forecast were used to compute these computations.

Results derived by means of the MAPE of about 4.32% suggest that the baseline has relatively low relative error with regard to the 24-hour period, particularly considering the simplicity of the method. The value of MAE and RMSE also indicates that the absolute deviations are still moderate relative to the overall magnitude of the plotted hourly demand. The difference between the RMSE and MAE suggests that certain hours have more significant errors, as it is expected by the smoothing of the forecasts in Fig. 2.

Table 1. Forecasting accuracy for 24-hour horizon (ARIMA (2,1,2))

Metric	Value
MAE (MWh)	48.8804
RMSE (MWh)	60.4989
MAPE (%)	4.3210
MSE (MWh²)	3660.12
R^2	0.94

3.4 Discussion

Methodologically, these findings align with the fact that ARIMA is a powerful tool in short-term forecasting, where the preponderant structure of the series is autocorrelation and seasonality. Past surveys suggest that no single forecasting method can be best in all circumstances since the effectiveness is influenced by the forecast horizon, data richness, and load pattern stability [10]. Deep learning models, which include LSTM architectures, are commonly favoured when nonlinear dependence in time is substantial, and attention can emphasize important time steps in sequence modelling [11]. Combination techniques that enhance robustness and predictive capability in various regimes have also been introduced that combine gradient boosting with neural networks [12].

However, there are some restrictions that should be accepted. To start with, the model uses a univariate series of loads and does not thus take into consideration other potentially informative exogenous variables, including weather conditions or the calendar effect. These covariates may be of great importance in improving predictors. Second, the data utilized in this work have a narrow time range and might not reflect the shifts in the structure of demand in the end and infrequent events.

Future modelling might hence incorporate weather and calendar characteristics to further extend the modelling framework to evaluate longer predictive horizons, e.g., the 168-hour weekly forecasting window of operational pre-dispatch planning [8], and compare the ARIMA with more sophisticated machine-learning models on the same evaluation protocol. The latter would enable a more detailed evaluation of forecasting performance.

4 Conclusion

Short-term electricity load forecasting pipeline (using Kaggle Panama data) and an evaluation of a univariate ARIMA baseline were used in this report on an out-of-sample 24-horizon. The workflow used was preparation, time-ordered train-test split, ARIMA (2,1,2) model, and forecast generation and performance evaluation, on MAE, RMSE, and MAPE. It is determined that under an ARIMA background, it is possible to track short-time-periodic demand movement with a reasonable degree of precision, with MAE = 48.8804 MWh, RMSE = 60.4989 MWh, and MAPE = 4.3210%.

In the analysis, it can be seen that the univariate ARIMA baseline tends to flatten steep variations intra-day as it only uses values of the past loads and fails to consider exogenous variations like weather and the calendar effects. All in all, the findings indicate that the ARIMA baseline can offer a decent and understandable reference to predict the electricity load in the short term. Future studies can enhance performance on forecasting using exogenous factors like weather, calendar, and can also use more sophisticated machine learning or deep learning algorithms.

References

1. M. S. Alam, J. Uddin, M. Z. Islam, and M. I. Hossain, Sensors **24** (2024)
2. S. Toufiq, M. T. Islam, and M. R. Islam, Energies **18** (2025)
3. Kaggle, "Short-term electricity load forecasting (Panama)," 2020. Available: https://www.kaggle.com/datasets/ernestojaguilar/shortterm-electricity-load-forecasting-panama
4. S. Noureen, S. Atique, V. Roy, and S. Bayne, IRJET **6** 1812–1821 (2019)
5. N. Ahmad, Y. Ghadi, M. Adnan, and M. Ali, IEEE Access **10** 71054–71090 (2022)
6. H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, and M. I. Ibrahem, Energies **16** (2023)
7. B. Biswal, S. Deb, S. Datta, T. S. Ustun, and U. Cali, Energy Reports **12** 3654–3670 (2024)
8. E. Aguilar Madrid and N. Antonio, Information **12** (2021)
9. S. Muzaffar and A. Afshari, Energy Procedia **158** 2922–2927 (2019)
10. K. Misiurek, T. Olkuski, and J. Zyśk, Energies **18** (2025)
11. J. Lin, J. Ma, J. Zhu, and Y. Cui, Int. J. Electr. Power Energy Syst. **137** (2022)
12. M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, Energy **214** (2021)