Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 165
DOI: 10.11159/icsta25.165

Predictive Models for Household Electricity Consumption in Thailand Using Classification

Piyachat Leelasilapasart[1] and Orathai Polsen[2]
[1,2] Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road Wongsawang, Bangsue, Bangkok, Thailand
piyachat.l@sci.kmutnb.ac.th; orathai.p@sci.kmutnb.ac.th

***Abstract* -** Thailand relies on electricity as a key component of daily life and a crucial driver of the growth of business and industry. It serves as a significant indicator of the nation's economic performance. Although electricity demand continues to increase, the country's electricity production remains constant, with signs of decline. The downward trend is attributed to the decreasing availability of natural gas and other key resources used in electricity production. Since the electricity cannot be stored like regular inventory, relevant agencies must manage and plan production to meet demand adequately. The present study forecasts the residential electricity consumption of the entire country using the Box-Jenkins and classification method over 180 months from January 2010 to December 2024. The evaluation of model accuracy is evaluated using the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) criteria. The results of this study can be applied as a guideline for analyzing suitable forecasting models for electricity consumption and other related factors in the future.

***Keywords*:** Classification method; KNN Nearest Neighbour; Electricity Household Consumption; Predictive Model; Deep Learning.

1. Introduction

Thailand relies on electricity as a key factor in daily life and a crucial driver of business and industry development. It is one of the predominant economic indicators. As a developing country, Thailand requires a significant amount of electricity. In 2023, Thailand consumed 204,023 Gigawatt/Hours of electricity, an increase of 3.4% due to the economic recovery following the easing of the COVID-19 pandemic. Household electricity usage increased by 7.4%, partly due to hot weather, which boosted demand for air conditioning and the growing trend of working from home [1]. While electricity demand continues to rise, the country's electricity production remains constant and is expected to decline in the future. This is due to the decreasing availability of natural gas and other key raw materials for electricity generation, which could lead to electricity shortages. Furthermore, electricity cannot be stored like regular inventory, despite the development of energy storage technologies. However, various conditions still need to be considered, such as storage capacity, duration, and proper management to prevent energy loss. Additionally, electricity demand varies by region and time of day. Therefore, relevant agencies must ensure sufficient electricity to meet demand and plan production accordingly, considering the available electricity in each area and period [2]-[4].

According to the National Strategy (2018-2037), the strategy for environmentally friendly growth highlights electricity as a key element in national development. It emphasizes enhancing the country's energy security and promoting the use of environmentally friendly energy, with a focus on improving the efficiency of managing electricity demand and supply [5]. Thus, effective electricity management is crucial to promoting national development within the national strategy framework. Forecasting techniques based on statistical principles have been continuously developed in recent years. These techniques are valuable for planning and decision-making in both the short and long term. Forecasting is a statistical tool that provides future data to aid in planning. Therefore, it is widely accepted that forecasting techniques play a significant role in both the public and private sectors. Accurate and reliable forecasting is essential. If forecasts are significantly lower than actual demand, it could result in insufficient electricity production and supply, which would affect the population's well-being and the country's competitiveness. Conversely, excess electricity could result in wasted energy if forecasts are significantly higher than actual demand [6].

Time series prediction has garnered increasing attention recently, with various statistical techniques, including regression methods, decomposition models, and the Autoregressive Integrated Moving Average (ARIMA) model. Time series forecasting has become a critical research focus across various domains, including business, economics, engineering, medicine, social sciences, and politics. The ARIMA model has gained popularity in time series forecasting as a model-based stochastic approach to forecasting. Furthermore, alternative methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have become increasingly popular in time series forecasting. The stochastic nature of time series data has motivated the use of various data mining techniques, including indexing, classification, and clustering [8]. Clustering is an example of an unsupervised learning procedure that identifies groups of patterns based on their intrinsic structure. It aims to partition *n* observations into *k* disjoint groups, or clusters, where each observation belongs to the cluster with the nearest mean, serving as a cluster prototype. By doing so, the unforeseen patterns in the time series can be identified [9]. Applying K-Nearest Neighbours (KNN) for forecasting is an interesting approach, particularly in cases where the data exhibit complex structures that traditional statistical models struggle to capture accurately. KNN is an algorithm based on the principle of "nearest neighbors." It predicts unknown values by calculating the distance between data points and then using the *k* closest points from the training set to make a prediction. Combining K-Nearest Neighbors (KNN) with ARIMA is a compelling idea in the context of Hybrid Modeling. This approach often yields better results than using a single model alone by leveraging the strengths of both methods to work together. The ARIMA model performs well with linear data that exhibits temporal structure, such as trend and seasonality. In contrast, KNN is flexible with nonlinear patterns and repetitive structures in the data.

Given these concerns, the researcher is interested in analyzing household electricity consumption data in Thailand and developing a suitable forecasting model that incorporates both statistical techniques and a hybrid model. The research will utilize classification methods, beginning with the Box-Jenkins method to forecast the data series, then calculating the forecast errors, and subsequently training the K-Nearest Neighbors to predict these errors. The final forecast will be the combined prediction from ARIMA and the residual KNN prediction. The resulting model can predict household electricity consumption and serve as a guideline for setting measures and planning to allocate sufficient electricity to meet the population's needs.

This study compares the widely recognized Box-Jenkins approach with a hybrid model combining K-Nearest Neighbors (KNN) and ARIMA for forecasting energy consumption. The article is structured as follows: Section 2 presents the literature review, while Section 3 outlines the methodology and data used. Section 4 presents the results of the proposed model, and Section 5 concludes the study, providing suggestions for future improvements.

2. Methodology and Data

The secondary monthly time series data, spanning from January 2010 to December 2024, were collected from domestic online sources [10]. This dataset focuses on the electricity consumption of the whole country (classified by sector) and covers 180 months.

Table 1: Part of the data on residential electricity consumption in Thailand

Jan-2010	Feb-2010	March-2010	Apr-2010	May-2010	June-2010
2,427.749097	2,464.573006	2,872.615193	3,134.387275	3,304.331944	3,007.303068
Jul-2010	Aug-2010	Sep-2010	Oct-2010	Nov-2010	Dec-2010
2,906.591435	2,761.558473	2,701.776114	2,709.978791	2,436.586121	2,486.354158

For analysis, the most recent 20% of the data is used as the test set, while the earliest 80% is allocated for training. Thus, the training set comprises the first 168 records, and the evaluation set includes the final 12 records. We utilize Python to analyze the data employing two statistical methods as follows:

2.1. Box-Jenkins method

Box-Jenkins' ARIMA model is one of the most widely used forecasting methods. This model relies on the historical series of a variable for prediction, excluding other independent variables. This approach formulates the model using lagged values of the time series for the autoregressive (AR(p)) and the moving average (MA(q)). [11] - [12]

The model assumes a functional relationship between future time series values and current or past values, incorporating white noise into its predictive process, which relies on the historical value of the series. The ARIMA model decomposes historical data into an Autoregressive process. In the ARIMA (p, d, q) model, p represents the order of the autoregressive process, d indicates the order of the data's stationarity, and q represents the order of the moving average terms, as per the Box-Jenkins methodology [13]. Equation (1) presents the general form of the ARIMA model:

$$\emptyset_P(B)(1-B)^d x_t = \theta_q(B)\varepsilon_t\ , \tag{1}$$

where $\emptyset_P$ and θ_q are the AR and MA characteristic operators and are represented by

$$\emptyset_P(B) \ = \ 1-\emptyset_1 B-\emptyset_2 B^2-\cdots-\emptyset_p B^p \text{ , and}$$
$$\theta_q(B) = \ 1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\ ,$$

respectively. $(1-B)^d$ is the difference, while B is a backward shift operator that shifts the time series data back by one period, and α_t is the forecasting error at time *t*.

For seasonal time series data, a seasonal ARIMA model, denoted by ARIMA (p, d, q) (P, D, Q), is commonly used. Equation (2) presents the seasonal ARIMA model, which includes the seasonal AR and MA components, as well as the seasonal differencing parameter, D:

$$\emptyset_P(B)\omega_P(B^s)(1-B^s)^D y_t = \theta_q(B)\gamma_Q(B^s)\varepsilon_t\ , \tag{2}$$

where $\omega_P(B^s)$ and $\theta_q\gamma\_Q$ (B^s) are the seasonal AR and MA characteristic operators, respectively.

In conclusion, the ARIMA model offers a practical approach for analyzing non-stationary time series data, and incorporating seasonal components can significantly enhance its precision and forecasting capabilities.

2.2. K-Nearest Neighbour Classification ARIMA

Hybrid time series models integrate statistical forecasting approaches with machine learning techniques to capitalize on the strengths of both methods. Traditional statistical models effectively capture linear trends, whereas machine learning and neural networks can model complex nonlinear relationships. In contemporary forecasting, hybrid models have gained popularity due to consistently delivering more accurate predictions than standalone models across various applications [14]. For instance, the winning approach in the M4 Forecasting Competition was a hybrid model that integrated exponential smoothing with a recurrent neural network, substantially outperforming all individual methods [15]. In our study, we employed the K-Nearest Neighbors (KNN) algorithm to identify similar historical patterns and classify time series into different regimes, such as high, medium, and low demand. At a specific time point or within a defined time window, KNN selects the *k* most similar past instances based on distance metrics, such as Euclidean distance.

The hybrid model, which integrates K-Nearest Neighbors (KNN) and ARIMA, can be implemented through three primary strategies. First, a classification-based approach utilizes KNN to categorize the time series data into distinct groups or regimes, such as high, medium, and low demand. Separate ARIMA models are then constructed for each class, and forecasting is performed using the ARIMA model corresponding to the current classified group. Second, the residual correction approach begins by applying an ARIMA model to capture the primary linear trend and generate baseline forecasts. Subsequently, KNN is employed to learn and adjust the residuals by identifying and referencing similar residual patterns observed in historical data, thereby enhancing the forecast accuracy. Third, the ensemble prediction approach generates forecasts independently using ARIMA and KNN models. The final forecast is derived by aggregating the outputs of the two models, typically through averaging or weighted combinations. This ensemble technique leverages the complementary strengths of statistical and machine learning models to achieve improved predictive performance.

This study applied a hybrid forecasting approach that combines the ARIMA model with the K-Nearest Neighbors (KNN) algorithm through a residual correction mechanism. The procedure begins by constructing the primary forecast using the ARIMA model, formulated as follows:

$\hat{y}_t^{ARIMA}$ on the training set.

The residuals, which represent the discrepancy between the actual values and the ARIMA forecasts, are computed as:

$$e_t = y_t - \hat{y}_t^{ARIMA}.$$

Subsequently, the residuals are modeled using KNN to capture any remaining patterns not accounted for by the ARIMA model. The predicted residual at time *t*, denoted as $\hat{e}_t^{KNN} = \frac{1}{k}\sum e_i$, is calculated by averaging the residuals of the *k* =3 Most similar historical sequences, each of length 12 (i.e., a window size of 12 months):

$$\hat{e}_t^{KNN} = \frac{1}{k}\sum e_i$$

Finally, the corrected forecast is obtained by combining the ARIMA forecast with the estimated residual from the KNN model:

$$\hat{y}_t = \hat{y}_t^{ARIMA} + \hat{e}_t^{KNN}$$

This hybrid structure enables the model to benefit from ARIMA's ability to capture linear temporal dependencies and KNN's capacity to leverage local nonlinear residual patterns, thereby enhancing the overall forecasting accuracy.

2.3. Model Evaluation

In the context of forecasting, errors are inevitable. Several measures, including Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), are employed in this study to minimize these errors and ensure high accuracy. $RMSE$ is a quadratic scoring rule that also assesses the average magnitude of the error. $RMSE$ is a quadratic scoring rule that also assesses the average magnitude of the error. It is computed as the square root of the average of the squared differences between the predicted and actual values. It can be calculated as shown in Equation (3)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{2}$$

MAPE stands for Mean Absolute Percentage Error. It is a commonly used metric to measure the accuracy of a forecasting model. The equation is shown in Equation (4)

$$\text{MAPE} = \frac{1}{n}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100} \tag{2}$$

3. Result

To construct the forecasting model, the data is decomposed to find trends, seasonality, cyclicality, and randomness. The time series plot for the secondary data from domestic online sources presents monthly electricity consumption of the residential sector from January 2010 to December 2024, along with its decomposition, as shown in Figs. 1 and 2. It can be observed that the data exhibits trends and seasonality.

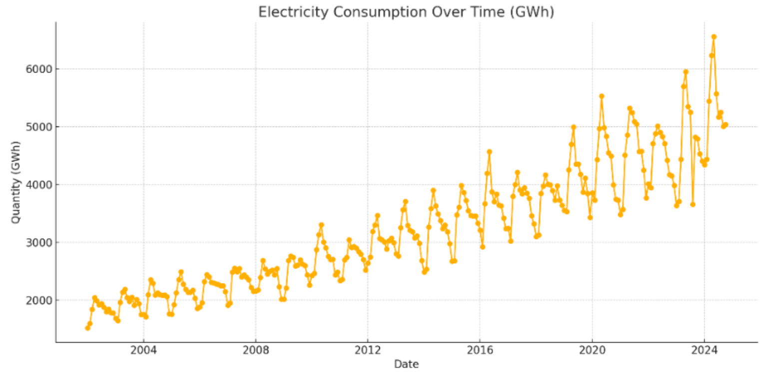

Fig. 1: The time series plot of monthly electricity consumption of the residential sector.

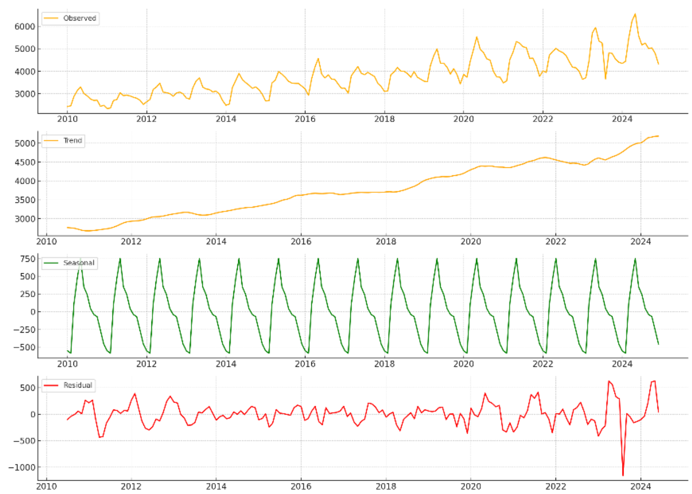

Fig. 2: Decomposition of time series data.

The data is divided into training and test sets, and based on that, the various errors are gauged. The accuracy of the K-Nearest Neighbour Classification ARIMA is compared with the standard ARIMA model. The RMSE and MAPE values are used as performance metrics for the proposed model. The lower the value, the more accurate the model.

Table 2 presents the values of RMSE and MAPE for the dataset using the ARIMA and K-Nearest Neighbour Classification ARIMA models.

Table 2: Model summary and accuracy value of forecasting

Forecasting methods	**Model**	**RMSE**	**MAPE**
Box-Jenkins	SARIMA (1,1,1)(1,1,1)[12]	391.26	5.94%
KNN-ARIMA	-	**378.99**	**5.59%**

As presented in Table 2, the RMSE obtained from the K-Nearest Neighbors Classification ARIMA model is lower than that of the standard ARIMA model, with values of 391.26 and 411.81, respectively. This indicates a slight improvement in terms of root mean squared error. Nevertheless, a comparison based on the MAPE suggests minimal performance difference between the two models. Therefore, it is impossible to draw a definitive conclusion regarding which model offers superior predictive accuracy.

The predicted values generated by the SARIMA$(1,1,1)(1,1,1)_{12}$ and the KNN-ARIMA hybrid models align closely with the actual observed values, as illustrated in Figs. 3 and 4, respectively. Both proposed models demonstrate low RMSE, indicating a high level of forecasting accuracy.

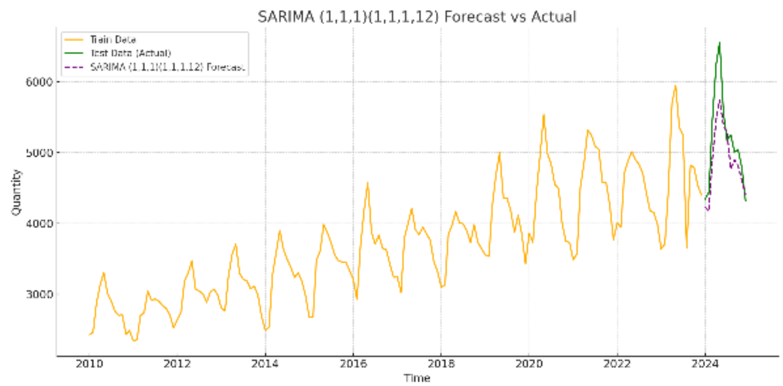

Fig. 4: Actual and predicted values of Household Electricity Consumption of Thailand using SARIMA$(1,1,1)(1,1,1)_{12}$

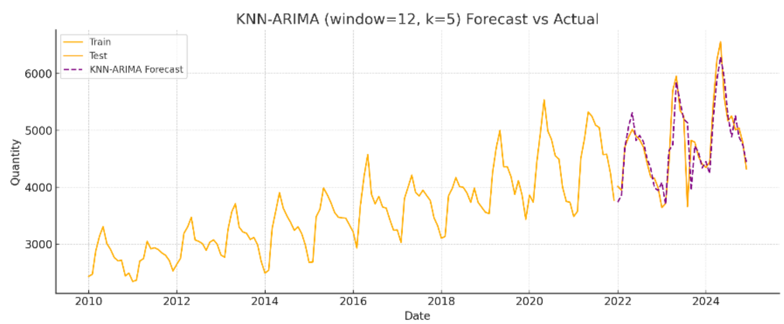

Fig. 5: Actual and predicted values of Household Electricity Consumption of Thailand using KNN-ARIMA

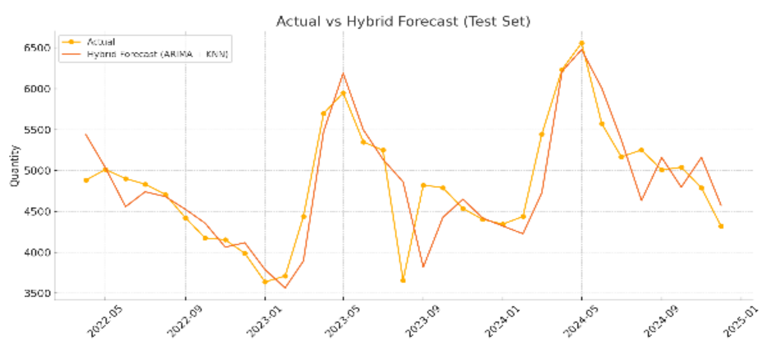

Fig. 6: Predicted values of Household Electricity Consumption of Thailand using KNN-ARIMA

4. Conclusion

To sum up, a hybrid forecasting model integrating K-Nearest Neighbors (KNN) with the Auto Regressive Integrated Moving Average (ARIMA) approach, referred to as the KNN-ARIMA model, is employed to predict household electricity consumption in Thailand. The main objective is to enhance the accuracy of short-term energy demand forecasting by combining the strengths of statistical time series modeling and machine learning techniques. The performance of the proposed model is assessed by comparing its forecasting errors against those of the conventional ARIMA model. Specifically, the results indicate a reduction in RMSE and MAPE, demonstrating a notable improvement in predictive accuracy.

Acknowledgements

This research was funded by the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok (fiscal year 2025, contract no. sci-680039). This support is gratefully acknowledged.

References

[1] Ministry of Energy. (2024, August 13). Energy Statistics of Thailand 2024 [Online]. Available: https://www.eppo.go.th/index.php/th/petroleum/gas/link-erc-pipeline/item/20753-%20energy-statistics-2567

[2] P. Kanjanasamranwong and S. Thongrong, "Forecasting Model in Electrical Power Consumption of the Provincial Electricity Authority in Southern Area 2," *Srinakharinwirot Science Journal*, vol. 28, no. 2, pp. 13–29, 2012.

[3] P. Chaimongkol, W. Surapat, and K. Hongsombat, "Effective Energy Management in Smart Grid Distribution System," *Kasetsart Engineering Journal*, vol. 27, no. 88, pp. 81–92, 2014.

[4] N. Kaewhawong, "Forecasting Electricity Consumption of Thailand by Using SARIMA and Regression Models with ARMA Errors", *Thai Journal of Science and Technology*, vol. 4, no. 1, 2015.

[5] Royal Thai Government Gazette. (2561, 13 October). The National Strategy Act B.E. 2561 - 2580 (2018 - 2037C.E.), volume 135, (in Thai), 2018.

[6] C. Hamzaçebi. "Primary energy sources planning based on demand forecasting: The case of Turkey," *Journal of Energy in Southern Africa*, 27(1), pp.1-10, 2016.

[7] J.M. Kohiro, R.O. Otieno, and C. Wafula, "Seasonal time series forecasting: a comparative study of ARIMA and ANN models." *Af J Sci Technol*. 2004;5(2):41–49.

[8] F. Petitjean, A. Ketterlin, P. Gançarski, A global averaging method for dynamic time warping, with applications to clustering, Pattern Recog. 44 (3) (2011) 678–693.

[9] P. Leelasilapasart, "*Implementation, Numerics, and Applications of Duality for Optimal Stochastic Switching*," Ph.D. dissertation, University of Technology Sydney, Sydney, Australia, 2020.

[10] Digital Government Development Agency. Electricity consumption of the whole country (classified by branch) [Online]. Available: https://data.go.th/en/dataset/dataset_11_371.

[11] P. E. George and G. M. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1976.

[12] P. Kamoljitprapa, O. Polsen, and U. Kabir Abdullahi, "Forecasting of Thai International Imports and Exports Using Holt-Winters' and Autoregressive Integrated Moving Average Models," *Journal of Applied Science and Emerging Technology (JASET)*, vol. 22, no. 3[2023]: e252955.

[13] P. Leelasilapasart and O. Polsen, "Prediction Model of Gold Fund in Thailand," In *Proceedings of the 2024 9th International Conference on Information and Education Innovations*, pp. 104-108, April 2024.

[14] L.B. Sina, C.A. Secco, M. Blazevic and K. Nazemi, "H*ybrid Forecasting Methods — A Systematic Review*." Electronics, 12(9), 2019.