Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Exploratory Data Analysis and Energy Intensity Forecasting at Water Treatment Plant

Rosnalini Mansor[1,*], Pang Hwee[1], Nur Afiqah Mohd Said[1], Saiful Azlan Mat Radhi[2], Bahtiar Jamili Zaini[1], Mohamad Shukri Abdul Hamid[1], Malina Zulkifli[1], Zahayu Md Yusof[1]

[1] School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
[2] Syarikat Air Darul Aman Sdn. Bhd., No. 892, Jalan Sultan Badlishah Bandar Alor Setar, 05000 Alor Setar, Kedah, Malaysia

ARTICLE INFO

Article history:
Received 2 February 2025
Received in revised form 4 August 2025
Accepted 23 September 2025
Available online 31 October 2025

Keywords:
Energy intensity; Exploratory data analysis; Forecasting; Univariate time series; Water treatment plant

ABSTRACT

Energy intensity optimization in water treatment plants (WTPs) is essential for ensuring sustainable operations and cost-effective resource management. In Malaysia, WTPs consume substantial energy to maintain water treatment and distribution, yet inefficiencies in energy usage remain a concern. This study integrates Exploratory Data Analysis (EDA) and Univariate Time Series (UTS) forecasting to analyze and predict energy intensity trends at four WTPs in Northern Kedah Region One. The primary objective is to enhance energy efficiency by identifying consumption patterns and selecting the most suitable forecasting model for energy intensity prediction. The methodology involved data collection on electricity consumption and water production from January 2021 to October 2023, followed by EDA to detect patterns, anomalies, and relationships in energy usage. Several UTS models, including Naïve, Moving Average, Simple Exponential Smoothing, and ARIMA, were applied to forecast energy intensity. The results highlight significant variations in energy intensity among the WTPs, with Jenun Baru exhibiting the lowest energy intensity, indicating greater efficiency, while Jeneri recorded the highest. Furthermore, findings demonstrate that no single forecasting model is universally optimal, as performance varies based on data characteristics. This study underscores the importance of incorporating EDA in forecasting to improve forecasting model accuracy and support informed decision-making in WTP operations. The insights derived from this research can guide policymakers and industry practitioners in implementing energy-saving strategies and optimizing water treatment processes. Future research should explore multivariate time series models that incorporate external factors such as weather conditions and operational changes to enhance forecasting precision and energy efficiency.

1. Introduction

Malaysia is one of Asia's highest energy per capita consumers in terms of total consumption and intensity. The country's final energy consumption rose from 13 million tons of oil equivalent (toe) in 1990 to approximately 41 million in 2010, reflecting an average annual growth rate of 6%. Rahman *et al.* [1] explained that despite aggressive energy efficiency initiatives over the past 20 years, Malaysia has not significantly improved energy consumption and conservation. Ritchie *et al.* [2]

* *Corresponding author.*
E-mail address: rosnalini@uum.edu.my (Rosnalini Mansor)

describe that according to Our World in Data, Malaysia has improved energy efficiency, but progress has been slow. The country faces challenges in terms of energy consumption and conservation. Rahman *et al.* [1] identified that the lower-than-expected results from previous energy efficiency programs prompted the Malaysian government to launch the National Energy Efficiency Action Plan (NEEAP) for the 2016-2025 period, considering socio-cultural, policy, financial, and administrative barriers.

By looking at this fact, each level and sector in Malaysia should support the agenda together since it is not only about money but also sustainability and environmental responsibility. According to the study by Pakharuddin *et al.* [3], water treatment plants (WTPs) play a vital role in providing safe drinking water to communities by improving water quality. They process raw water from rivers, lakes, or groundwater to remove impurities and contaminants. The treatment involves coagulation, filtration, disinfection, and pH adjustment. Then, treated water is distributed through pipes to homes and businesses for domestic and non-domestic usage. Since the WTPs' role is very significant in all aspects of life, the management of operation WTPs should take note of monitoring procedures to ensure efficient and safe operation.

Although one of the monitoring procedures in WTPs is forecasting activities, inaccurate, incomplete, and anomalous data will make the results meaningless and cause high forecasting errors. Ismail *et al.* [4] explained that Exploratory Data Analysis (EDA) is the systematic, thorough data analysis to find significant patterns, relationships, and insights. Hence, the EDA is the best option for the preprocessing stage in forecasting. Furthermore, EDA bridges raw data, meaningful information, and actionable knowledge in WTPs.

The study by Tukey [5] described that EDA constitutes a fundamental phase in research analysis. Since Tukey's groundbreaking research in 1977, Komorowski *et al.* [6] explained that EDA has grown significantly in popularity for data set analysis. Examining the data for distribution, anomalies, and outliers is the primary goal of EDA, which helps guide the hypothesis's specific testing and as prior knowledge before further analysis. EDA seeks to support the analyst's ability to recognize natural patterns. Hence, some researchers in previous studies applied EDA in their data research profiling to show the significance of EDA's role in their studies, such as in energy profiling for university buildings by Usman *et al.* [7], in wastewater study by Xiao *et al.* [8] and in electricity load demand by Ismail et *al.,* [4]. Therefore, this aligns with the initial step of univariate time series forecasting procedures: plot data and identify the existence of the time series components described by Bowerman *et al.* [9] based on data patterns.

Time series forecasting predicts the future value(s) based on historical data. Univariate time series forecasting only considers the time factor in its analysis. The time series data is the data value in sequence time. Maciel [10] and Mansor *et al.* [11] explained that the data could be in interval-valued time series (ITS), fuzzy-valued time series (FTS) and point-valued time series (PTS). Since the data in this study is numeric or crisp data from the WTP, this study applied the forecasting method suitable with PTS done by Mansor and Zaini [12].

However, the research by Othman *et al.* [13], Biswas and Yek [14] and Labo [15] stated there are some other issues in WTPs, such as issues in energy consumption measuring and monitoring data and energy-saving technologies. The study by Biswas and Yek [14] revealed that water treatment plants consume large amounts of energy to operate the treatment process, which can contribute to greenhouse gas emissions and operational costs. Moreover, Labo [15] clarified that energy intensity optimization is essential for WTPs because of increased energy costs. The optimization of energy intensity will be accomplished by integrating energy recovery from the WTPs process and energy-saving technologies.

Despite numerous studies on energy intensity and efficiency in various sectors, including solar energy [16], energy-water efficiency [13], and the cement industry [17], limited research has specifically focused on integrating EDA with univariate time series forecasting to optimize energy intensity at water treatment plants (WTPs). Most existing studies primarily address energy consumption measurement or energy-saving [13-15] without providing a comprehensive approach to predicting and optimizing energy intensity using historical data patterns. Additionally, while some studies focus on energy efficiency strategies and technologies, the specific integration of EDA and univariate time series models in WTPs remains underexplored [18].

Therefore, this study aims to bridge this gap by developing an integrated forecasting model that utilizes EDA and univariate time series methods, offering a novel approach to decision-making and planning in WTP operations. By addressing this gap, the study contributes to enhancing energy efficiency, reducing operational costs, and promoting sustainable practices within the water treatment sector.

2. Methodology

This study presents Exploratory Data Analysis (EDA) and Univariate Time Series (UTS) forecasting results for four water treatment plants (WTPs) in Northern Kedah Region One, Kedah state, providing comprehensive insights into energy-water efficiency. In Malaysia, WTPs are managed at the state level, with each state having its water authority responsible for managing, operating, and maintaining WTPs. These authorities adhere to national standards set by the Ministry of Natural Resources, Environment, and Climate Change (NRECC) and the National Water Services Commission (SPAN). They implement management practices such as water quality monitoring, process optimization, and energy efficiency initiatives to align with Malaysia's sustainability goals. In Kedah, Syarikat Air Darul Aman (SADA) manages water treatment and supply, managing 36 WTPs across six regions: Northern Kedah Region One, Northern Kedah Region Two, Central Region, East Region, Southern Region, and Langkawi Region. This study focuses on the WTPs in Northern Kedah Region One; Jenun Baru, Jenun Lama, Jeneri, and Pokok Sena.

As outlined in Figure 1, the research methodology was systematically applied to each WTP in Northern Kedah Region One from Phase 1 to Phase 4. In Phase 1, data on electricity consumption (EC) and water production (WP) were collected from January 2021 to October 2023 to establish a foundational dataset for the analysis. Phase 2 included EDA to analyze energy intensity (EI) trends over the same period. In Phase 3, the research involved partitioning the data, modelling using data from January 2021 to December 2022 and evaluating the models with 10 data points from January 2023 to October 2023.

Various UTS forecasting models were used, such as Simple Average, Naive, Moving Models (MA3, MA4, MA5), Exponential Smoothing (SES), and Box-Jenkins Model. Finally, Phase 4 involved evaluating model performance, identifying the best UTS forecasting model and forecasting energy intensity for 10 months. Figure 1 shows the connection between all the phases by integrating EDA and UTS forecasting. The following subsection explains each phase.

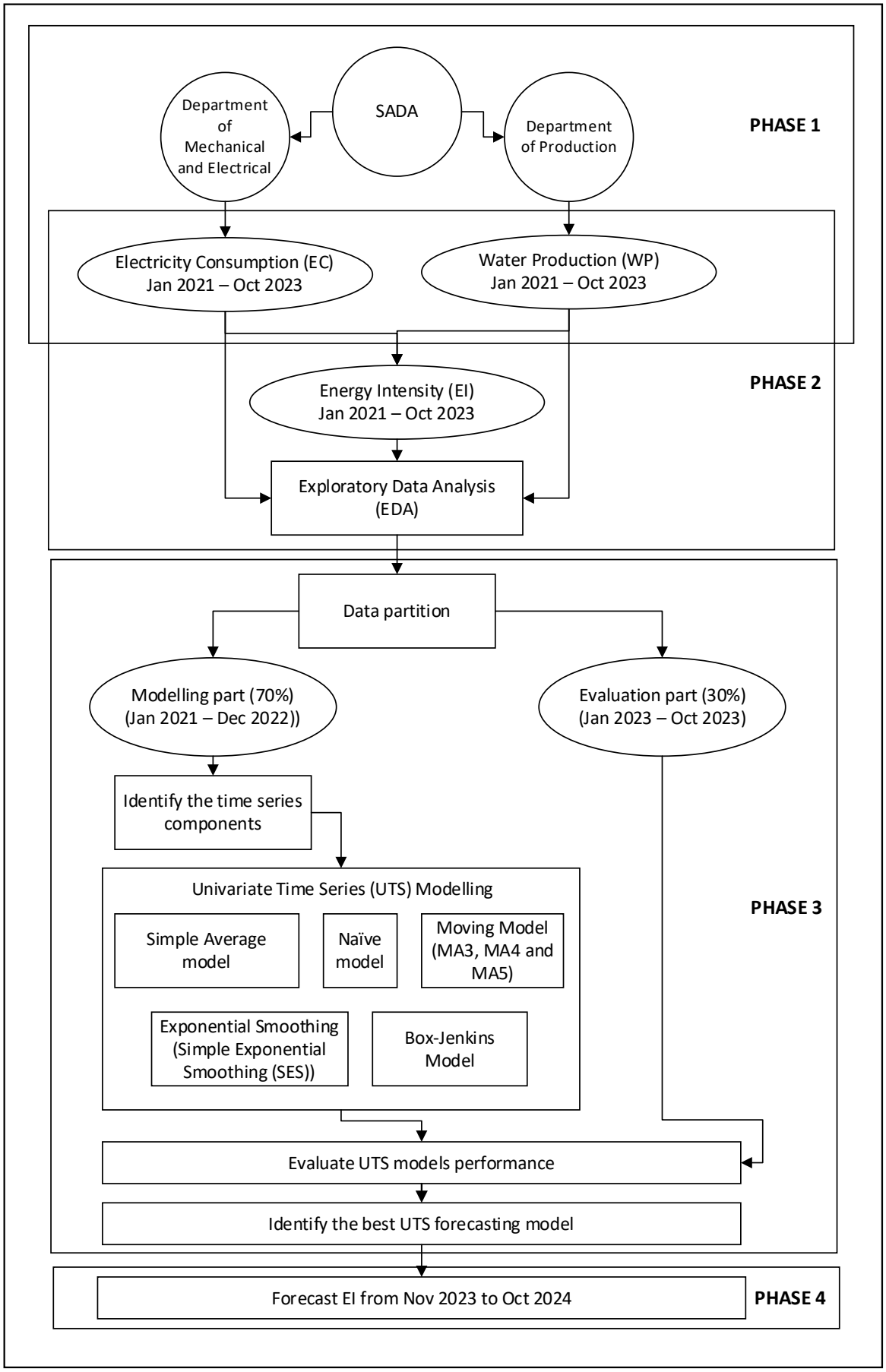

Fig .1. Research methodology

2.1. Phase 1: Collecting the data

The study's first phase involves gathering electricity consumption (EC) and water production (WP) data from the Department of Mechanical and Electrical and the Department of Production at SADA. The dataset covers January 2021 to October 2023, focusing on the relationship between EC and WP as primary inputs for calculating energy intensity (EI). This phase establishes a strong data foundation for subsequent analysis by collecting comprehensive and accurate records from relevant departments.

2.2. Phase 2: Exploring the data

Next, from the data in Phase 1, data EI (EI value) was created from data on EC and WP. EDA was conducted by grouping the data into three analyses. Then, EDA was executed from these three data sets. The role of EDA in this study is to give a clear picture of the time series data by showing the trend and movement pattern and the summary statistics like maximum and minimum values, central tendency measures, and measures of dispersion. Also, any significant differences in the pattern were investigated.

Majid *et al.* [19] described that the EI value is obtained by dividing the energy consumption by the total water production. Kilowatt-hours per cubic metre (kWh/m^3) are standard energy intensity units expressing the energy used to produce one water unit. Othman *et al.* [13] and Liu *et al.* [20] explained that the EI can be represented as in Eq. (1)

$$Energy\ Intensity = \frac{Energy\ Consumption\ (kWh)}{Water\ Production\ (m^3)} \quad (1)$$

And can be simply by Eq. (2)

$$EI = \frac{EC}{WP} \quad (2)$$

2.3. Phase 3: Identifying the best univariate time series (UTS) model

Phase 3 focuses on the EI data set only. A few univariate EI forecasting models are suggested for the UTS analysis under the time series pattern discovered in Phase 2. In this phase, the data has been divided into modelling and evaluation parts before determining the optimal model. When performing UTS forecasting analysis, identifying the time series components, whether trend, seasonality, cyclic patterns, and residuals, should be based on the modelling part of the data. For some reason, in Phase 2 findings, the data points for modelling and evaluation will be presented in the findings section soon.

The role of data partition in this phase is that the data in the modelling part is used to build the forecasting. By analyzing the time series components in the modelling part, we ensure that the model captures the underlying patterns and structures in the historical data. Identifying and understanding these components accurately is crucial for developing a reliable model. Since the results from Phase 2 show no trend in the modelling part of the data, which is no significant up or down trend movement and no significant seasonality, the suitable UTS forecasting models are presented in Table 1. Models 1 to 6 were executed using calculations in Microsoft Excel and Model 7 using RStudio. The steps in R studio are presented in Table 2

Table 1
List of univariate time series forecasting models

No.	Model	Forecasting model
1	Naïve	$EI_{t+1} = y_t$
2	Recursive Simple Average (RSA)	$EI_{t+1} = \frac{\sum(all\ data\ values)}{t}$
3	Moving Average (MA3)	$EI_{t+1} = \frac{\sum(most\ recent\ 3\ data\ values)}{3}$
4	Moving Average (MA4)	$EI_{t+1} = \frac{\sum(most\ recent\ 4\ data\ values)}{4}$
5	Moving Average (MA5)	$EI_{t+1} = \frac{\sum(most\ recent\ 5\ data\ values)}{5}$
6	Simple Exponential Smoothing (SES)	$EI_{t+1} = \alpha y_t + (1-\alpha)F_t\ ,\quad 0 \le \alpha \le 1$
7	Box-Jenkins ARIMA(p,d,q)	$\phi_p(B)(1-B)^d EI_t = c + \theta_q(B)\varepsilon_t$

Meanwhile, the data is reserved for validating the UTS models' performance in the evaluation part. It acts as a hold-out set to assess how well the model can predict unseen data. Competition of four forecasting errors of measurements was executed as a statistical tool to evaluate the UTS forecasting model performance. The error measurements that were used were mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). The model exhibiting the lowest forecasting error across these metrics was designated the most accurate and declared the best EI forecasting model in this study.

Table 2
Steps to apply the ARIMA model in Rstudio.

Step	Process	Code
1	Read the data from the EXCEL file	>library(readxl) >xls.file <- file.path("your_file_path.xlsx") >EI <- read_excel(xls.file) >View(EI)
2	Transform the data into time series data	>ts_EI <- ts(EI, start = c(2021,1), end = c(2022,12), frequency = 12)
3	Check Autocorrelation of the time series data	>library(ggplot2) >library(fpp) >library(forecast) >ggAcf(ts_EI) + ggtitle("ACF OF EI at WTP ") >ggPacf(ts_EI) + ggtitle("PACF OF EI at WTP ")
4	Check stationary of the time series data using ADF test	>adf.test(ts_EI) *Note:* H_0 is rejected if p-value < α (0.05), the data series is stationary H_0 is not rejected if p-value > α (0.05), the data series is non-stationary.
5	Differencing is required when the time series data is non-stationary, and check again the ADF test after differencing	>EI_diff1 <- diff(ts_EI, differences = 1, lag = 1) >EI_diff1 >adf.test(EI_diff1)
6	Generate ARIMA model and select the best model automatically	>EImodel <- auto.arima(ts_EI, seasonal = FALSE, ic = "aic", trace = TRUE)

7	Generate the fitted value based on the best ARIMA model	fit_EI <- arima(ts_EI, order = (0,0,0)) > fitted(fit_EI)
8	Forecast the values based on the best ARIMA model	>EIForecast = forecast(EImodel, level = c(95), h = 1*10)

2.4. Phase 4: Forecasting 12 months ahead of energy intensity

In the final phase, the best-performing UTS model forecasts energy intensity (EI) from November 2023 to October 2024. The forecast results will inform strategic energy management decisions at the Jenun Baru Water Treatment Plant. By optimizing energy usage, the study aims to enhance the plant's overall energy-water efficiency, contributing to cost savings and supporting broader sustainability goals in water resource management.

3. Findings and Discussion

Since this study's primary analysis is EDA and UTS forecasting, this section is divided into two sections: one focused on EDA and the other on UTS forecasting results.

3.1 Exploratory data analysis (EDA) results

Table 3 and Figure 2 show the statistical values and graphical representation of electricity consumption across the four water treatment plants (WTPs) in Northern Kedah Region One, revealing significant variations in usage patterns. Jenun Baru WTP exhibits the highest electricity consumption, with an average of 850,699 kWh and the highest recorded maximum and minimum consumption levels, suggesting a consistently high demand. However, in contrast, Jeneri WTP has the lowest average electricity consumption (304,102 kWh) and the smallest range, indicating more stable and lower energy usage. The coefficients of variation values show that all WTPs have relatively stable consumption patterns, with variations below 4%. The time series graph confirms these findings, showing that Jenun Baru and Jenun Lama WTPs have consistently higher consumption trends. At the same time, Jeneri and Pokok Sena WTPs operate at comparatively lower levels. The observed trends and variations highlight potential opportunities for optimizing energy efficiency, particularly at high-consumption plants like Jenun Baru.

Table 3
Summary of electricity consumption at WTPs in Northern Kedah Region One

Statistics	Jenun Baru	Jenun Lama	Jeneri	Pokok Sena
Max	891375	624683	315190	494032
Min	724820	542681	275674	437395
Mean	850698.8235	588500.1765	304101.5588	470017.5294
Median	855786.5	594829	306849	470370.5
Range	166555	82002	39516	56637
Standard Deviation	32854.5880	21460.4132	9689.6863	14605.3505
Coefficient of Variation	3.86%	3.65%	3.19%	3.11%

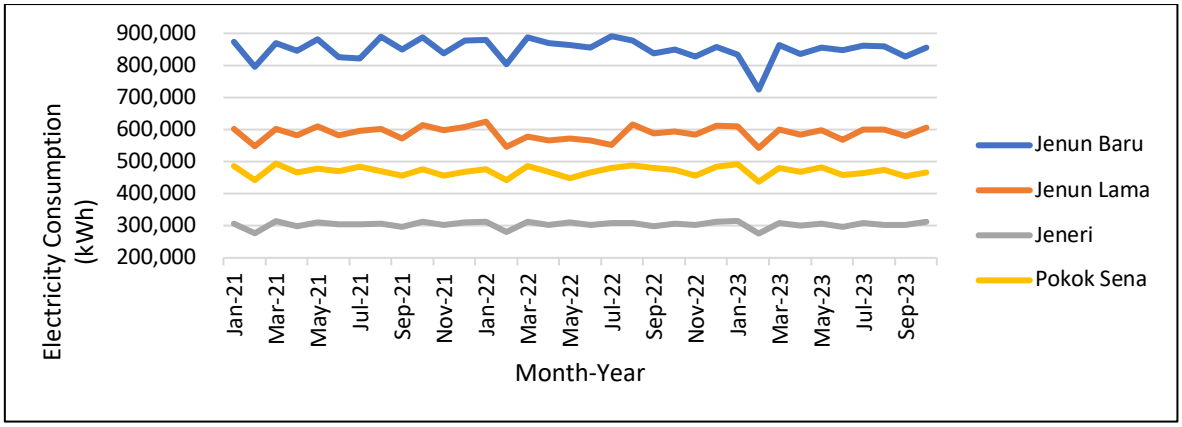

Fig. 2. Electricity consumption (kWh) movement by month at WTPs in Northern Kedah Region One.

Table 4 and Figure 3 show the statistical analysis and graphical trends of water production, revealing that Jenun Baru WTP consistently produces the highest volume of water, with an average production of 2,253,292 m³ and the largest range, indicating significant fluctuations in production levels. In contrast, Jeneri WTP has the lowest mean water production (540,099 m³) and the smallest range, suggesting a more stable but lower output. The coefficients of variation values for all WTPs remain below 3.6%, demonstrating relatively consistent production patterns over time. The time series graph further illustrates that Jenun Baru maintains the highest production levels, followed by Jenun Lama, Pokok Sena, and Jeneri. These findings highlight the substantial variation in production capacity among the WTPs, emphasizing the need for optimized resource allocation and operational efficiency, particularly at high-production facilities like Jenun Baru.

Table 4
Summary of water production at WTPs in Northern Kedah Region One

Statistics	Jenun Baru	Jenun Lama	Jeneri	Pokok Sena
Max	2351277	1264682	567932	1048696
Min	2047237	1089133	486035	923452
Mean	2253292.2941	1196623.3382	540098.8953	1007923.3824
Median	2264584.5	1201578	545552.5	1018494
Range	304040	175549	81897	125244
Standard Deviation	2253292.2941	1196623.3382	540098.8953	1007923.3824
Coefficient of Variation	3.09%	3.54%	3.50%	3.02%

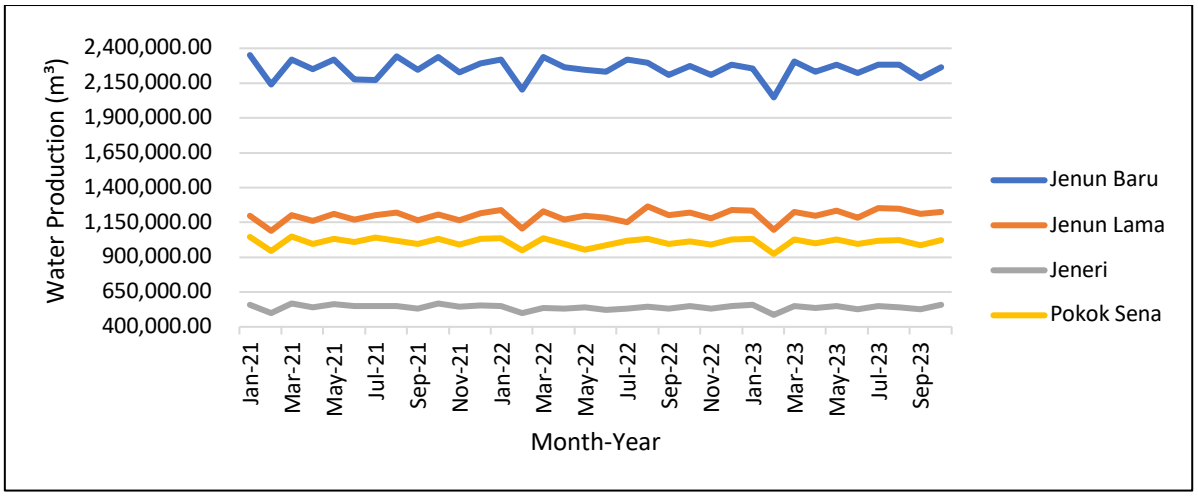

Fig. 3. Water production movement by month at WTPs in Northern Kedah Region One.

The results presented in Table 5 and Figure 4 provide an in-depth analysis of energy intensity across four water treatment plants (WTPs) in Northern Kedah Region One. The statistical values in summary indicate that Jeneri WTP has the highest average energy intensity (0.56317 kWh/m³), followed by Jenun Lama (0.49189 kWh/m³), Pokok Sena (0.46635 kWh/m³), and Jenun Baru (0.37747 kWh/m³). Notably, Jenun Baru WTP exhibits the lowest energy intensity values, suggesting greater energy efficiency in water production than other WTPs. Additionally, the coefficients of variation values, ranging from 1.34% to 2.08%, indicate relatively low variability in energy intensity over time, implying stable operational efficiency at each WTP.

The time series analysis in Figure 4 further illustrates monthly fluctuations in energy intensity, highlighting the trends and variations across the WTPs from January 2021 to September 2023. Jeneri WTP consistently demonstrates the highest energy intensity, exceeding the average of all WTP in Region One, 0.47472 kWh/m³, suggesting a higher energy requirement per cubic meter of water produced. Conversely, Jenun Baru WTP consistently operates below the energy intensity average in Region One WTPs, reinforcing its position as the most energy-efficient facility among the four. The fluctuations observed in Jenun Lama and Pokok Sena WTPs indicate potential operational adjustments, seasonal effects, or external factors impacting energy consumption.

These findings significantly affect energy management and optimization in water treatment operations. The relatively stable energy intensity values suggest consistent performance; however, the variations across WTPs indicate opportunities for improvement. Adopting energy-efficient technologies, optimizing pump operations, and implementing real-time monitoring could help enhance energy efficiency, particularly for high-energy-consuming WTPs like Jeneri. Future research could further investigate the factors contributing to energy intensity differences, including plant design, water source quality, and operational protocols, to develop targeted interventions for reducing energy consumption while maintaining water production efficiency.

Table 5
Summary of energy intensity at WTPs in Northern Kedah Region One

Statistics	Jenun Baru	Jenun Lama	Jeneri	Pokok Sena
Max	0.38463	0.51342	0.58269	0.48174
Min	0.35405	0.47024	0.55013	0.45340
Mean	0.37747	0.49189	0.56317	0.46635
Median	0.37833	0.49418	0.56142	0.46653
Range	0.03059	0.04318	0.03255	0.02834
Standard Deviation	0.00554	0.01024	0.00886	0.00626
Coefficient of Variation	1.47%	2.08%	1.57%	1.34%

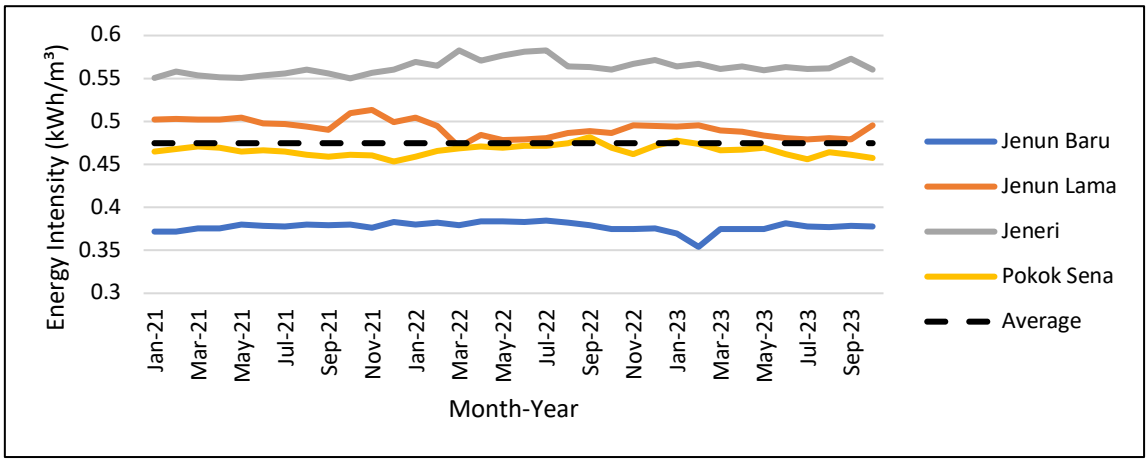

Fig. 4. Monthly energy intensity movement at WTPs in Northern Kedah Region One.

3.2 Univariate Time Series Forecasting

This section's results focus only on selecting the best energy intensity univariate time series forecasting model.

3.2.1 Jenun Baru WTP

Table 6 and Table 7 show the results from the model performance evaluation for the Jenun Baru WTP, indicating the ARIMA (3,0,0) model consistently outperforms others based on forecasting error measures, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Generalized RMSE (GRMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD). The ranking analysis in Table 6 further supports this conclusion, as ARIMA (3,0,0) attains the lowest total rank, signifying its superior predictive accuracy. The results suggest that the ARIMA model is robust and reliable for capturing the underlying patterns in energy intensity data at the Jenun Baru WTP, making it suitable for short- and medium-term forecasting.

Meanwhile, the graphical representation in Figure 5 confirms that the ARIMA model successfully captures fluctuations in energy performance while maintaining consistency with historical trends. The alignment between actual, fitted, and forecasted values further validates the model's effectiveness. Given its ability to produce precise forecasts with minimal error, the ARIMA (3,0,0) model is recommended for energy performance forecasting at Jenun Baru WTP over the next 12 months.

Table 6
Jenun Baru WTP's forecasting models performance based on the evaluation part of the data

Model	MSE	RMSE	GRMSE	MAPE	MAD
Naïve	7.88336E-05	6.21473E-09	0.002500359	1.559542111	0.004697804
SES	7.39611E-05	5.47024E-09	0.002799733	1.546789084	0.004618383
SA	7.24236E-05	5.24518E-09	0.002174912	1.331995068	0.000297836
MA3	6.43373E-05	4.13929E-09	0.001760306	1.559170082	0.004306377
MA4	6.87058E-05	4.72048E-09	0.003359110	1.639138239	0.004228836
MA5	7.37036E-05	5.43222E-09	0.002220925	1.699094235	0.004043294
ARIMA (3,0,0)	6.30609E-05	3.97667E-09	0.002125014	1.251786036	0.000673090

Table 7
Total performance rank for each Jenun Baru WTP's forecasting model

Model	MSE	RMSE	GRMSE	MAPE	MAD	Total Rank
Naïve	7	7	5	5	7	31
SES	6	6	6	3	6	27
SA	4	4	3	2	1	14
MA3	2	2	1	4	5	14
MA4	3	3	7	6	4	23
MA5	5	5	4	7	3	24
ARIMA (3,0,0)	1	1	2	1	2	7

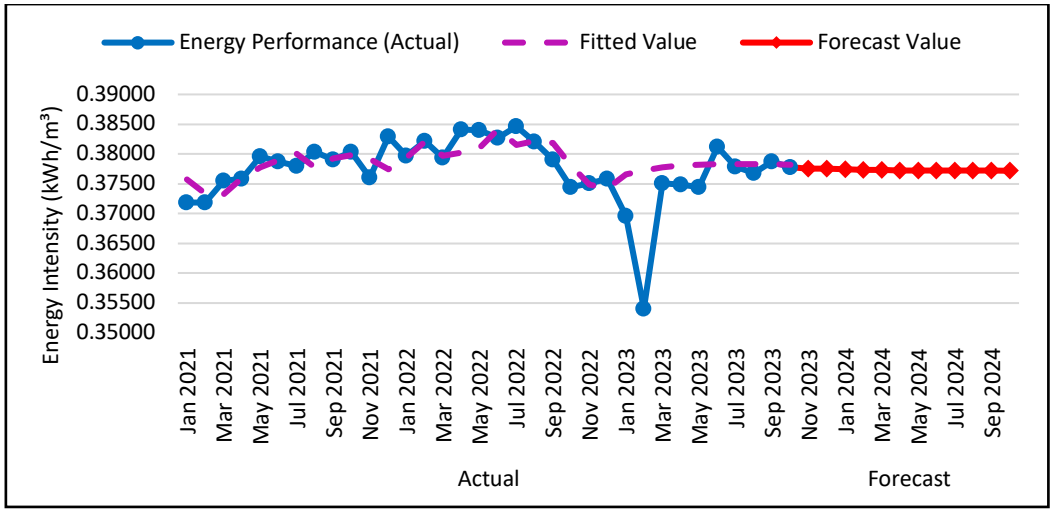

Fig. 5. Forecast values of energy intensity at Jenun Baru WTP.

3.2.2 Jenun Lama WTP

The evaluation of forecasting models for the Jenun Lama Water Treatment Plant (WTP) in Table 8 and the ranking of models in Table 9 demonstrate that the Naïve model consistently outperforms the others across most error measures, achieving the best rankings in MSE, RMSE, GRMSE, and MAPE, with a total rank of 11. The SES model also shows strong performance, particularly in MSE and RMSE, securing the second-best total rank of 14. In contrast, the ARIMA (0,1,0) model performs poorly across all metrics, resulting in the highest (worst) total rank of 27. The remaining models, including SA, MA3, MA4, and MA5, exhibit moderate performance with total ranks ranging from 17 to 24. These findings indicate that simpler models like the Naïve and SES approaches may provide better forecasting accuracy for this specific dataset, suggesting that model complexity does not always correlate with improved performance in forecasting water treatment plant data.

Figure 6 shows the energy performance (kWh/m³) at Jenun Lama WTP from January 2021 to September 2024 using the Naïve model, which was identified as the best-performing model. The actual and fitted values align closely, demonstrating the Naïve model's strong ability to capture trends in the historical data. The forecasted values remain stable from November 2023 onward, reflecting the Naïve model's characteristic of projecting the most recent observed value forward, suggesting a consistent energy performance prediction for the next year.

Table 8
Model performance in the evaluation part of data of Jenun Lama WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD
Naïve	3.41599×10^{-05}	1.1669×10^{-09}	0.002283388	0.762077890	0.007218521
SES	3.42203×10^{-05}	1.1710×10^{-09}	0.002387421	0.782550458	0.007027391
SA	8.53528×10^{-05}	7.28511×10^{-09}	0.004833000	1.617653934	0.001454915
MA3	4.24106×10^{-05}	$1.79866 \text{Ex} 10^{-09}$	0.003097700	1.008064037	0.005617362
MA4	5.03547×10^{-05}	2.53559×10^{-09}	0.004886312	1.220180968	0.004775290
MA5	5.51967×10^{-05}	3.04667×10^{-09}	0.005746474	1.344089451	0.004026547
ARIMA (0,1,0)	0.000112806	1.27252×10^{-08}	0.004320566	1.784897380	0.003059562

Table 9
Total ranking for each error measure based on the models in Jenun Lama WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD	Total Rank
Naïve	1	1	1	1	7	11
SES	2	2	2	2	6	14
SA	6	6	5	6	1	24
MA3	3	3	3	3	5	17
MA4	4	4	6	4	4	22
MA5	5	5	7	5	3	25
ARIMA (0,1,0)	7	7	4	7	2	27

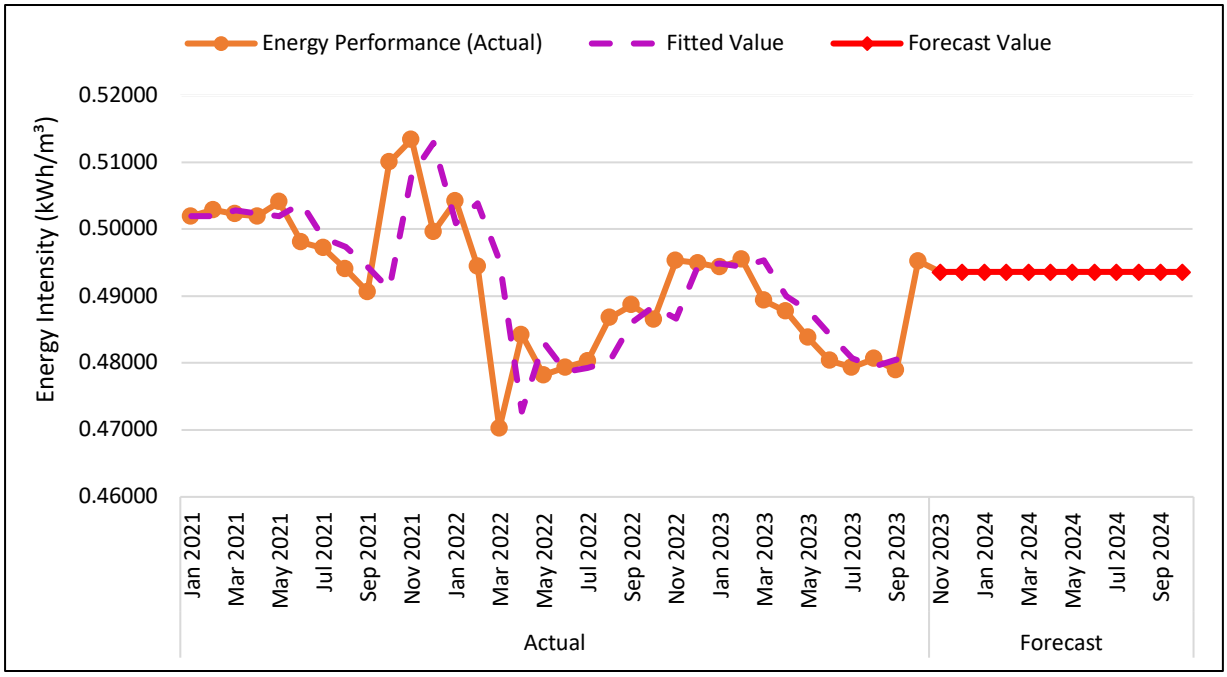

Fig. 6. Forecast values of energy intensity at Jenun Lama WTP

3.2.3 Jenari WTP

The performance evaluation of models for the Jeneri Water Treatment Plant (WTP) data, as shown in Tables 10 and 11, reveals that the Simple Average (SA) model performs best among all models, achieving the lowest overall total rank of 9. The SA model ranks first across multiple error measures, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Generalized RMSE (GRMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD). The Naïve model, on the other hand, ranks the lowest with a total rank of 31, suggesting it is the least effective model for forecasting energy performance at Jeneri WTP.

The graph in Figure 7 depicts the actual, fitted, and forecasted energy performance values at Jeneri WTP. The fitted values (purple dashed line) closely follow the actual energy performance (grey line) during the actual period, indicating a good model fit. The forecasted values demonstrate a stable energy performance trend from November 2023 to September 2024, suggesting consistent energy efficiency at the Jeneri WTP.

Table 10
Model performance in the evaluation part of data of Jeneri WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD
Naïve	4.32553E-05	1.87102E-09	0.003365079	0.941204166	0.003348295
SES	3.82266E-05	1.46127E-09	0.002766663	0.866673058	0.003085001
SA	1.51630E-05	2.29916E-10	0.001948176	0.50106597	0.000119461
MA3	2.14641E-05	4.60707E-10	0.001026281	0.558675451	0.002089332
MA4	2.20014E-05	4.84063E-10	0.001840002	0.575292179	0.001834129
MA5	2.03188E-05	4.12855E-10	0.001527988	0.567291851	0.001549863
ARIMA (0,1,0)	7.48387E-05	5.60082E-09	0.007226013	0.144336113	0.008245380

Table 11
Total ranking for each error measure based on the models in Jeneri WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD	Total Rank
Naïve	6	6	6	7	6	31
SES	5	5	5	6	5	26
SA	1	1	4	2	1	9
MA3	3	3	1	3	4	14
MA4	4	4	3	5	3	19
MA5	2	2	2	4	2	12
ARIMA (0,1,0)	7	7	7	1	7	29

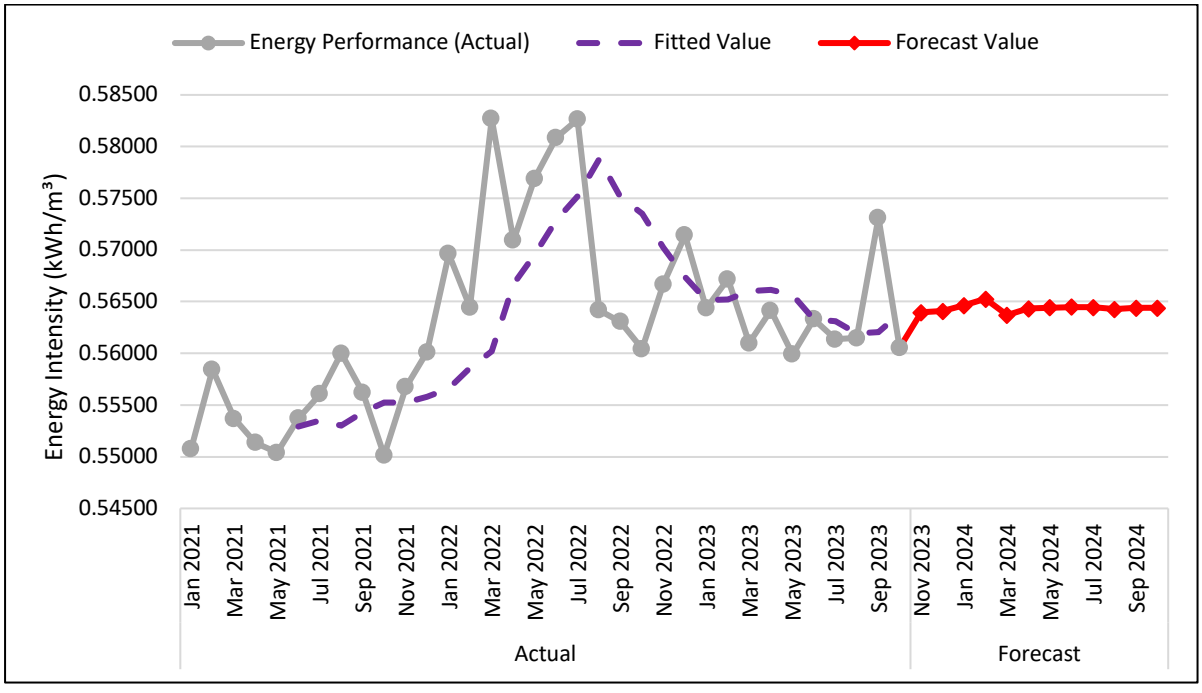

Fig. 7. Forecast values of energy intensity at Jeneri WTP.

3.2.4 Pokok Sena WTP

The performance of forecasting models in the evaluation part of the data for Pokok Sena Water Treatment Plant (WTP) is presented in Tables 12 and 13. According to Table 12, the Moving Average model with a 4-period (MA4) consistently achieved the lowest error values across multiple metrics, including MSE, RMSE, GRMSE, MAPE, and MAD. The overall ranking in Table 13 further emphasizes the superiority of MA4 with a total rank of 13, outperforming all other models. The Naïve model, while showing decent performance in MSE and RMSE, ranked relatively low overall, with a total rank of 23, indicating that it was not the best choice for this dataset. The Simple Average (SA) model had the highest total rank of 26, reflecting weaker performance across the evaluation metrics.

The graph in Figure 8 illustrates the forecasted energy performance values at the Pokok Sena WTP. The fitted values closely follow the trend of the actual energy performance up until the forecast period. The forecast values remain relatively stable from November 2023 to September 2024, suggesting a steady energy performance projection for the upcoming year.

Table 12
Model performance in the evaluation part of data of Pokok Sena WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD
Naïve	2.98512E-05	8.91093E-10	0.004110754	1.059115109	0.00490736
SES	3.02351E-05	9.14162E-10	0.003015721	1.041311055	0.004733165
SA	4.42487E-05	1.95795E-09	0.003634927	1.173411981	0.00072991
MA3	3.55971E-05	1.26716E-09	0.00318190	1.052627944	0.003851459
MA4	2.79646E-05	7.82021E-10	0.003691837	0.981256741	0.003855224
MA5	3.1552E-05	9.95526E-10	0.003235215	0.97787381	0.003894832
ARIMA (1,0,0)	3.69801E-05	1.36753E-09	0.003845884	1.100105909	0.001373633

Table 13
Ranking for each error measure based on the models in Pokok Sena WTP

Model	MSE	RMSE	GRMSE	MAPE	MAD	Total Rank
Naïve	2	2	7	5	7	23
SES	3	3	1	3	6	16
SA	7	7	4	7	1	26
MA3	5	5	2	4	3	19
MA4	1	1	5	2	4	13
MA5	4	4	3	1	5	17
ARIMA (1,0,0)	6	6	6	6	2	26

Overall, Figure 9 highlights the varying energy intensity patterns among the Northern Kedah Region One WTPs.Jeneri exhibits the highest energy intensity, indicating lower efficiency, while Jenun Baru maintains the lowest, suggesting higher operational efficiency. Jenun Lama and Pokok Sena show moderate energy intensity with fluctuations. The forecasted values (red markers) suggest stable trends across all WTPs, emphasizing the need for tailored energy management strategies to optimize energy consumption

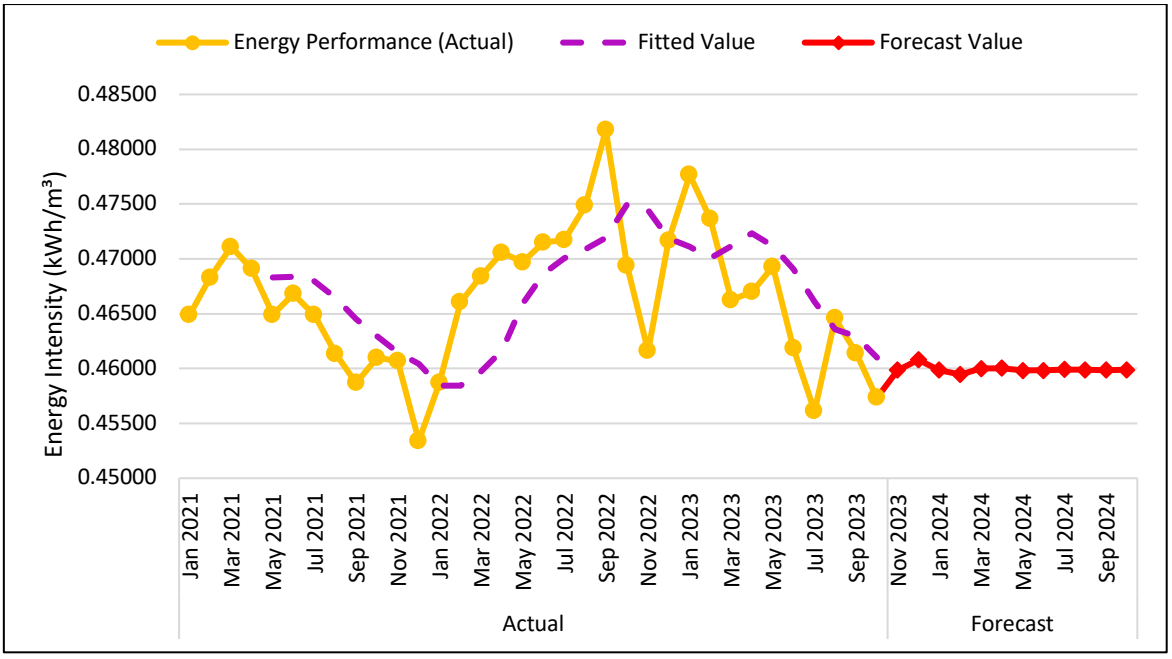

Fig. 8. Forecast values of energy intensity at Pokok Sena WTP.

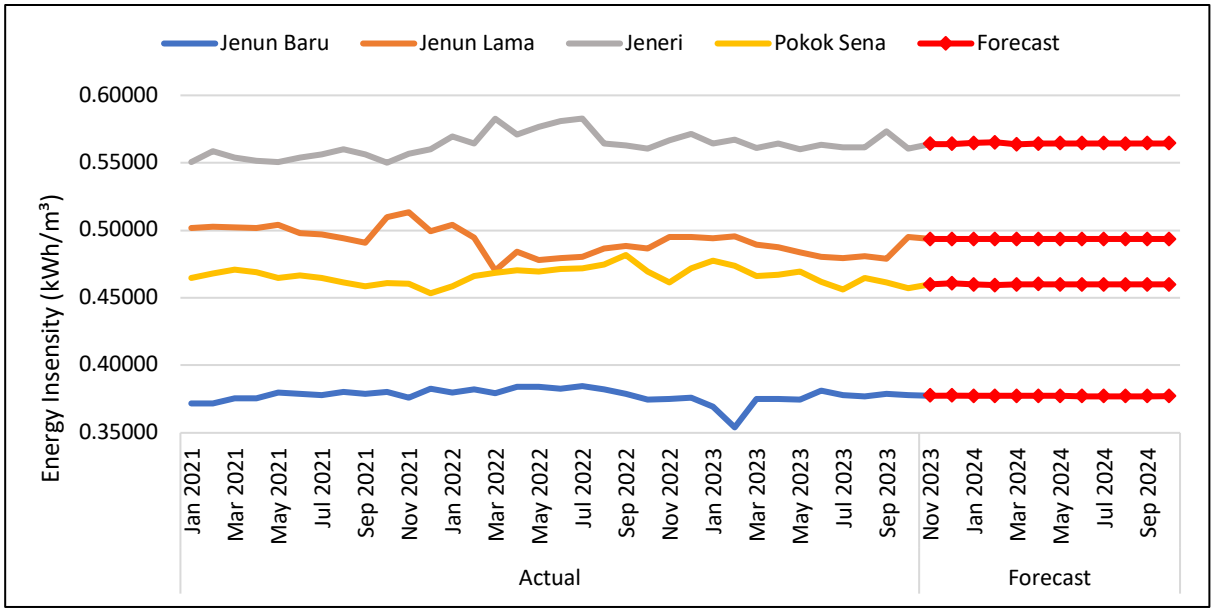

Fig. 9. Energy intensity in Northern Region One Kedah WTPs

4. Conclusions

This study has demonstrated the integration of Exploratory Data Analysis (EDA) and Univariate Time Series (UTS) forecasting models in evaluating and predicting energy intensity at four water treatment plants (WTPs) in Northern Kedah Region One. The findings provide valuable insights into energy consumption patterns, water production efficiency, and the overall energy-water intensity trends at these facilities. Through systematic data exploration and forecasting, the study highlights opportunities for optimizing energy usage, thereby contributing to cost savings and sustainable resource management in water treatment operations.

A key takeaway from this study is that there is no single "best" forecasting model applicable to all datasets. Different models perform optimally depending on the characteristics of the data, including trends, seasonality, and variability. For instance, the ARIMA model yielded superior predictive accuracy for Jenun Baru WTP, while the Naïve model was the best choice for Jenun Lama WTP. Similarly, the Simple Average model performed optimally for Jeneri WTP, whereas the Moving Average model (MA4) was most suitable for Pokok Sena WTP. These results reinforce the necessity of model selection based on data-specific attributes rather than relying on a one-size-fits-all approach.

The study also underscores the importance of rigorous exploratory data analysis before model selection and forecasting. By examining data distributions, trends, and anomalies, EDA serves as a crucial preliminary step in ensuring the effectiveness of forecasting models. The integration of EDA with time series forecasting enables a more data-driven and systematic approach to energy management at WTPs, enhancing operational efficiency and sustainability. The comparative results from this study show that different forecasting models yield varying levels of accuracy depending on the dataset's characteristics, emphasizing the need for careful model selection tailored to specific energy intensity patterns at different WTPs.

Furthermore, the comparative analysis of energy intensity (EI) among Jenun Baru, Jenun Lama, Jeneri, and Pokok Sena WTPs reveals significant variations in energy efficiency across the facilities. Jenun Baru WTP exhibited the lowest energy intensity, indicating greater efficiency in converting electricity into water production, whereas Jeneri WTP had the highest EI, suggesting higher energy consumption per unit of water treated. These differences highlight the impact of operational factors, infrastructure, and management practices on energy performance. Understanding these variations is crucial for implementing targeted energy-saving strategies at each facility, ensuring optimal resource utilization and improved sustainability.

Future research can explore the incorporation of multivariate time series models that consider additional external factors influencing energy intensity, such as weather conditions, operational changes, and policy interventions. Furthermore, machine learning-based forecasting techniques may provide alternative approaches for improving predictive accuracy and robustness in energy efficiency studies. By continuously refining forecasting methodologies and leveraging advanced analytics, the optimization of energy consumption at water treatment plants can be further enhanced, aligning with broader sustainability and energy conservation objectives.

Acknowledgement
This research received support from the Universiti Utara Malaysia (UUM) through the Generation Research Grant Scheme (S/O code: 21447). The authors also wish to express their appreciation to the UUM Research and Innovation Management Centre (RIMC) for their assistance. Additionally, special thanks go to Syarikat Air Darul Aman (SADA) Sdn. Bhd, Kedah, for providing all the necessary data for this study.

References

[1] Abd Rahman, Nurul Asra, Syahrul Nizam Kamaruzzaman, and Farid Wajdi Akashah. "Scenario and strategy towards energy efficiency in Malaysia: a review." In *MATEC Web of Conferences*, vol. 266, p. 02012. 2019. https://doi.org/10.1051/matecconf/201926602012
[2] Ritchie, Hannah, Pablo Rosado, and Max Roser. "Energy." Our World in Data, January 4, 2024.
[3] Pakharuddin, N. H., M. N. Fazly, SH Ahmad Sukari, K. Tho, and W. F. H. Zamri. "Water treatment process using conventional and advanced methods: A comparative study of Malaysia and selected countries." In *IOP Conference Series: Earth and Environmental Science*, vol. 880, no. 1, p. 012017. 2021. https://doi.org/10.1088/1755-1315/880/1/012017

[4] Ismail, Suzilah, Malina Zulkifli, Rosnalini Mansor, Muhammad Mat Yusof, and M. I. Ismail. "The role of exploratory data analysis (EDA) in electricity forecasting." *Pertanika Journal of Social Sciences & Humanities* 24 (2016): 93-100

[5] Tukey, John Wilder. (1977). *Exploratory Data Analysis*. Addison-Wesley Publishing Company.

[6] Komorowski, Matthieu, Dominic C. Marshall, Justin D. Salciccioli, and Yves Crutain. "Exploratory Data Analysis." *Secondary Analysis of Electronic Health Records*, September (2016): 185–203. https://doi.org/10.1007/978-3-319-43742-2_15

[7] Mohammed Usman, Abdullahi, Akmal Nizam Mohammed, Mohd Faizal Mohideen, Mas Fawzi Mohd Ali, Kamil Abdullah, and Juntakan Taweekun. "Energy Profiling for Residential College Buildings." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 81, no. 2 (2021): 139–145. https://doi.org/10.37934/arfmts.81.2.139145

[8] Xiao, Feng, Thomas R. Halbach, Matt F. Simcik, and John S. Gulliver. "Input Characterization of Perfluoroalkyl Substances in Wastewater Treatment Plants: Source Discrimination by Exploratory Data Analysis." *Water Research* 46, no. 9 (2012): 3101–3109. https://doi.org/10.1016/j.watres.2012.03.027

[9] Bowerman, Bruce L, Richard T O'Connell, and Anne B Koehler. (2005). *Forecasting, Time Series, and Regression*. South-Western College.

[10] Maciel, Leandro. "Financial Interval Time Series Modelling and Forecasting Using Threshold Autoregressive Models." *International Journal of Business Innovation and Research* 19, no. 3 (2019): 285. https://doi.org/10.1504/ijbir.2019.100323

[11] Mansor, Rosnalini, Bahtiar Jamili Zaini, and Norhayati Yusof. 2019. "Prediction Stock Price Movement Using Subsethood and Weighted Subsethood Fuzzy Time Series Models." *AIP Conference Proceedings* 2138, January (2019): 050018. https://doi.org/10.1063/1.5121123

[12] Mansor, Rosnalini, and Bahtiar Jamili Zaini. "Forecasting Using Point-Valued Time Series and Fuzzy-Valued Time Series Models." *International Journal of Membrane Science and Technology* 10, no. 2 (2023): 244–250. https://doi.org/10.15379/ijmst.v10i2.1168

[13] Othman, Shahidah, Rosnalini Mansor, and Fakhrurazi Ahmad. "Weighted Subsethood Fuzzy Time Series towards Energy-Water Efficiency for Water Treatment Plant." *Environment and Ecology Research* 10, no. 2 (2022): 182–192. https://doi.org/10.13189/eer.2022.100207

[14] Biswas, Wahidul K., and Pauline Yek. "Improving the Carbon Footprint of Water Treatment with Renewable Energy: A Western Australian Case Study." *Renewables: Wind, Water, and Solar* 3, no. 1 (2016). https://doi.org/10.1186/s40807-016-0036-2

[15] Labo, W. "Ways to Improve Water Treatment Plant Efficiency." World Pumps 2017 no. 9 (2017): 32–33. https://doi.org/10.1016/s0262-1762(17)30297-3

[16] Suad Hassan Danook, Khamis J Jassim, and Adnan M Hussein. "Efficiency Analysis of TiO2/Water Nanofluid in Trough Solar Collector." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 67, no. 1 (2020): 178–185.

[17] Oak, Hena."Factors Influencing Energy Intensity of Indian Cement Industry." *International Journal of Environmental Science and Development* 8, no. 5 (2017): 331–336. https://doi.org/10.18178/ijesd.2017.8.5.973

[18] Niu, Kunyu, Jian Wu, Lu Qi, and Qianxin Niu. "Energy Intensity of Wastewater Treatment Plants and Influencing Factors in China" Science of the total environment 670, June (2019): 961–970. https://doi.org/10.1016/j.scitotenv.2019.03.159

[19] Majid, Aman, Iliana Cardenes, Conrad Zorn, Tom Russell, Keith Colquhoun, René Bañares-Alcantara, and Jim W. Hall. "An Analysis of Electricity Consumption Patterns in the Water and Wastewater Sectors in South East England, UK." *Water* 12, no. 1,225 (2020): 1-17. https://doi.org/10.3390/w12010225

[20] Liu, Feng, Alain Ouedraogo, Seema Manghee, and Alexander Danilenko. (2012). "A Primer on Energy Efficiency for Municipal Water and Wastewater Utilities." *Openknowledge.worldbank.org*, February.