

Article

Not peer-reviewed version

---

# Application of Statistical and Machine Learning Models in Vietnam's Energy Consumption Demand Forecasting

---

[Van Thanh Phan](#)\*, [Duc Tien Nguyen](#)\*, [Ngoc Xuan Quynh Nguyen](#), [Xuan Hau Huynh](#)

Posted Date: 26 March 2026

doi: 10.20944/preprints202603.2045.v1

Keywords: energy consumption demand forecasting; Vietnam; statistical model; machine learning model; Holt's model; grey model; optimization; accuracy



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Application of Statistical and Machine Learning Models in Vietnam's Energy Consumption Demand Forecasting

Van Thanh Phan \*, Duc Trien Nguyen \*, Ngoc Xuan Quynh Nguyen and Xuan Hau Huynh

Vietnam-Korea University of Information and Communication Technology, The University of Danang, 470 Tran Dai Nghia, Ngu Hanh Son ward, Danang city, Vietnam

\* Correspondence: pvthanh@vku.udn.vn (V.T.P.); triennnd.23ai@vku.udn.vn (D.T.N.); Tel.: 0915377063 (V.T.P.)

## Abstract

Energy consumption demand forecasting plays a critical role in the planning and development of the nation's energy security, which underpins the 8-year Power Development Plan (PDP8) and Vietnam's ambitious Net-Zero 2050 commitment. However, this task becomes more difficult while challenges the big data environment is filled with a lot of noise and high fluctuation data. In order to deal with the problem, this paper using four distinct models which are Linear Regression, Holt's (Additive), PSO-GM (1,1), and Support Vector Regression (SVR) to conduct a rigorous comparative analysis to identify the most accurate forecasting model. The performance evaluated by MAE, RMSE, MAPE indexes based on the Vietnam's total primary energy demand data from 1986 to 2024. To check the accuracy of forecasting model, this study splits the length of data was into two period time, first time for the training data (1986- 2016) and next time for the testing data set (2017-2024). The results decisively identified that the Holt's model achieving significantly outperforming all counterparts with the lowest error metrics (MAE = 89.33, RMSE = 99.50, and a MAPE of 7.19%). This model is strongly suggested to forecast the Vietnam's energy demand in the period time 2025 to 2030. Based on this model, the Vietnam's energy demand will reach 1528.08 TWh and 1882.55 TWh in 2025 and 2030, respectively. Furthermore, this study provides empirical evidence that simpler, well-chosen statistical models can surpass complex alternatives in small-sample scenarios, offering a reliable quantitative baseline for policymakers to navigate infrastructure development and decarbonization challenges.

**Keywords:** energy consumption demand forecasting; Vietnam; statistical model; machine learning model; Holt's model; grey model; optimization; accuracy

## 1. Introduction

Vietnam is currently navigating a period of profound economic transformation. With a robust economic performance, including a high GDP growth of 7.1% in 2024 [1], the nation's demand for energy has increased. This is powerfully illustrated by the total national energy demand, which increased from approximately 535.32 TWh in 2010 to 1,457.18 TWh in 2024 [2], representing a stark increase of more than 172% in just 15 years. Accurate forecasting of this total demand is no longer just a technical exercise; it is a critical component of national energy security. The forecasts underpin the nation's multi-billion dollar infrastructure strategy as outlined in the Power Development Plan 8 (PDP8) [3], which must balance this rising demand against ambitious international climate commitments (e.g., COP26 Net-Zero 2050) [4].

However, forecasting this trajectory is fraught with challenges. The rapid, non-linear growth, compounded by recent global shocks and the inherent volatility of a transitioning economy, makes it difficult to find a model with high predictive accuracy. The academic literature reflects this challenge, presenting a wide spectrum of modeling choices. On one end, numerous recent studies

champion the use of complex, data-hungry deep learning models like: Long-Short term memory (LSTM) or other neural networks, often claiming superior performance.

In recent years, many scholars have used different methodologies and dynamic approaches to forecast energy consumption demand [5–9]. However, all researched papers are discovered under two main directions, The first is focused on dealing with the small-medium datasets such as Grey theory system and the other direction is longer datasets such as statistical and machine learning model. In this study, based on two main direction was above mentioned, we tries to used four distinct models (Linear Regression, Holt's (Additive), GM(1,1), and SVR) to forecast the Vietnam's total energy consumption demand data from 1986 to 2024. The primary contribution is to identify the most accurate and practical "champion model" for this common data-constrained scenario, thereby providing a reliable forecast for national energy demand from 2026 to 2030.

## 2. Related Works

Research on national energy forecasting can be divided into two main approaches which are classical time-series models and modern machine learning/deep learning models. In term of the classical category, Jamil (2020) applied ARIMA models to forecast the Pakistan's annual electricity consumption demand, the result of study identified that ARIMA(1,2,1) model is the best- fitting model in this cases. This model is strongly suggested to use in dealing with the long-term trends in low-frequency datasets [5]. However, the study only evaluated ARIMA in isolation, without comparing it to other methods like Holt's trend method or Support Vector Regression (SVR), leaving its true optimal unclear. In the other way, research has shifted toward complex deep learning and hybrid models. For example, Lin et al. (2020) used CNN-LSTM-Attention [6], and Ngo et al. (2022) proposed a WIO-SVR hybrid [8]. These methods perform well in short-term load forecasting and capture high-frequency, non-linear fluctuations. Despite their accuracy, they typically perform best with large datasets, which can be challenging for planning where annual data are limited.

In order to solving the small-data set, some studies have explored alternative approaches. Nguyen Dinh Tien (2025) applied Grey model variants to forecast renewable energy, with DGM (1,1) achieving a MAPE of just 6.76% [7]. Similarly, Nguyen et al. (2023) showed that Polynomial Regression could outperform deep learning models (MLP, RNN) when forecasting bio-energy output with limited data [9]. However, these studies are limited in scope: Grey Model research lacked direct comparison with machine learning and statistical models, and the polynomial regression study focused only on the bio-energy sector. This review highlights a clear research gap: there is a need for a comprehensive comparison of classical models (e.g., Holt), machine learning models (e.g., SVR), and small-data specialists (e.g., GM(1,1)) on the same practical datasets-a long-term, national-level energy forecast with limited annual observations.

## 3. Methodology and Data

### 3.1. Data Collection and Preparation

The data set used in this study is a uni-variate annual time-series of Vietnam's total primary energy consumption, spanning 39 years from 1986 to 2024. The data was collected from the publicly available "Our World in Data" database [2]. The energy unit is measured in terawatt-hours (TWh) and represents the aggregate demand across all sectors of the economy.

A data set of 39 observations is considered a small sample. For validation purposes, the data was divided into two parts: 31 observations from 1986–2016 (approximately 80%) for training, and 8 observations from 2017–2024 (about 20%) for evaluating forecasting performance.

### 3.2. Forecasting Models

To model the relationship between the independent variable, time ( $t$ ), and the dependent variable, energy demand ( $Y$ ), this study employs four distinctive models to compare and find the

“best-fit” model that describes the historical data and forecast the future data. Each model is described below.

### 3.2.1. Linear Regression

The first model, Linear Regression, assumes the relationship between  $t$  and  $Y$  is a simple straight line. It used the Ordinary Least Squares (OLS) method, a technique independently developed by Legendre (1805) and Gauss (1809)[10].

OLS operates by finding the parameters ( $\beta_0$  and  $\beta_1$ ) for a line that minimizes the sum of the squared residuals—the differences between the observed actual  $Y$  values and the  $\hat{Y}$  values predicted by the model.

The theoretical statistical model is represented as:

$$Y = \beta_0 + \beta_1 t + \epsilon \quad (1)$$

where:

$Y$  is the observed energy demand.

$\beta_0$  is the intercept (the value of  $Y$  when  $t = 0$ ).

$\beta_1$  is the slope (the change in  $Y$  for a one-unit increase in  $t$ ).

$t$  is the time variable.

$\epsilon$  (epsilon) is the unavoidable error, accounting for random variance not explained by the model.

After model fitting, the resulting prediction equation is:

$$\hat{Y} = \beta_0 + \beta_1 t \quad (2)$$

where ( $\hat{Y}$ ) is the predicted value for Energy demand.

### 3.2.2. Holt's Exponential Smoothing (Double Exponential Smoothing)

While Simple Exponential Smoothing (SES) is only effective for time-series data without a trend, the data set in this study exhibits a clear upward trend. To address this, Holt's (1957) Double Exponential Smoothing method is employed. The 2004 publication by Holt's model is a formal reprint of this foundational work.

The main idea is to take a weighted average of past observations, with weights that decay exponentially- so recent values (like 2024) receive much more emphasis than older ones (like 1986). Holt's innovation was to apply this smoothing principle to two distinct components: the level ( $l_t$ ) and the trend ( $b_t$ ) of the time series.

The model is controlled by two smoothing parameters,  $\alpha$  (alpha, for the level) and  $\beta$  (beta, for the trend), both ranging from 0 to 1. The update equations at each time step  $t$  for the Additive trend version—which was selected as the champion model—are as follows:

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (3)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (4)$$

Equation 3 shows the level as a weighted average of the current observation ( $y_t$ ) and the one-step-ahead forecast from the previous period. Equation 4 shows the trend as a weighted average of the most recent change in level ( $l_t - l_{t-1}$ ) and the previous trend estimate ( $b_{t-1}$ ).

The final forecast for  $h$  steps ahead (e.g., forecasting 2025, where  $h = 1$ ) is then calculated as:

$$F_{t+h} = l_t + h \cdot b_t \quad (5)$$

To ensure a careful model selection, two variants were tested: (1) the standard additive trend model and (2) the damped additive trend model (which introduces a damping parameter,  $\phi$ , to flatten

the trend). As showed in the results (Section 4.1, Table 1), the standard additive trend model (MAPE 7.19%) performed better than the Damped variant (MAPE 7.52%) on the test set. Therefore, the standard Additive model was selected. The parameters  $\alpha$  and  $\beta$  are not chosen manually; they are optimized by the model to minimize the forecasting error on the training set.

**Table 1. Optimal parameters and hyper-parameters of the model after training (on the datasets 1986-2016).**

| Model             | Parameters / Hyperparameters           | Optimal value |
|-------------------|--|---------------|
| Holt's (Additive) | $\alpha$ (Level smoothing coefficient) | 1.0000        |
|                   | $\beta$ (Trend smoothing coefficient)  | 0.2440        |
| SVR (RBF)         | C (Regularization parameter)           | 5000          |
|                   | $\gamma$ (Kernel coefficient)          | 0.01          |
| PSO-GM(1,1)       | $\lambda$ (Optimized Lambda)           | 0.3019        |
| Linear Regression | $\beta_1$ (Slope / a)                  | 24.8975       |
|                   | $\beta_0$ (Intercept / b)              | -49589.5742   |

### 3.2.3. Grey Model (GM(1,1)) Optimized by PSO

The foundational model used is the Grey Model (GM(1,1)), introduced by Deng (1989) [11]. It is designed for situations where information is incomplete or where the data set is small, as in this study. The name GM (1,1) stands for "Grey Model, 1 variable, the 1st order equation."

The key idea is that the model does not work directly on the raw (and possibly noisy) data  $X(0)$ .

Instead, it first transforms the data and then models the transformed series through a two-step process:

Accumulated Generating Operation (AGO): The model first transforms the raw data  $X(0)$  into a new, smoother, and monotonically increasing sequence  $X(1)$  using the 1-AGO. This new sequence is defined as:

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k=1,2,\dots,n \quad (6)$$

Differential Equation Assumption: The sequence  $X(1)$  is then assumed to follow a first-order differential equation that captures the underlying trend:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (7)$$

where  $a$  is the development coefficient and  $b$  is the Grey input.

In the traditional GM(1,1) model, the parameters  $a$  and  $b$  are estimated using the Ordinary Least Squares (OLS) method. However, this method is sensitive to data fluctuations and may not always produce optimal results, leading to larger forecasting errors [12]. To overcome this deficiency and improve predictive precision, this study proposes a hybrid PSO-GM(1,1) model. The core idea is to employ Particle Swarm Optimization (PSO), a powerful meta-heuristic algorithm introduced by Kennedy and Eberhart (1995) [13], to optimize a critical input parameter.

Optimization Target: Rather than replacing OLS, PSO is used to interactively search for the optimal parameter  $\lambda$  (lambda val).

Role of  $\lambda$ : This parameter is important because it is used to create the background sequence  $z_1$ , which then affects the OLS estimation of  $a$  and  $b$ .

Objective Function: The optimal  $\lambda$  is determined by applying time-series cross-validation on the training set. PSO seeks to minimize the objective (cost) function  $F(\lambda)$ , formally defined as the average Mean Absolute Error (MAE) across the folds:

$$\min_{\lambda} F(\lambda) = \frac{1}{K} \sum_{k=1}^K MAE_k \quad (8)$$

An optimized  $\lambda$  produces a better background sequence, which improves the OLS estimation of  $a$  and  $b$  achieves much better performance than the traditional GM(1,1) model [14].

Once the optimized  $a$  and  $b$  coefficients are found, the solution to the differential equation provides the forecasting formula for the accumulated sequence  $\hat{x}^{(1)}$ :

$$\hat{x}^{(1)}(k+1) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (9)$$

Finally, to retrieve the forecast for the original, non-accumulated data  $\hat{x}^{(0)}$ , the model applies an inverse AGO (IAGO), which is a simple subtraction:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (10)$$

### 3.2.4. Support Vector Regression (SVR)

Support Vector Regression (SVR) is a powerful machine learning model based on the principles of statistical learning theory, introduced by Vapnik (1995) [15].

SVR's objective is fundamentally different from OLS. Instead of minimizing the squared error for all data points, SVR operates on the principle of Structural Risk Minimization. Its goal is to find a function  $f(x)$  that is as "flat" as possible while accurately fitting the data.

It achieves this by defining an  $\epsilon$ -insensitive tube around the regression function. For any data point  $y_i$ , if the prediction error is within this tube (i.e.,  $|y_i - f(x_i)| \leq \epsilon$ ), the loss is assumed to be zero. The model's loss function only penalizes data points that lie outside this  $\epsilon$ -tube. The data points on the boundary or outside the tube are called Support Vectors, because they are the only points that determine the final regression function. This makes SVR robust to outliers and good at generalizing from small datasets, since it ignores points inside the tube and avoids fitting the noise.

To capture complex non-linear relationships in our energy data, SVR uses the kernel trick. This method applies a kernel function (such as polynomial or Radial Basis Function (RBF)) to implicitly map the data into a higher-dimensional space. In this space, a simple linear regression can model patterns that are non-linear in the original space. In this study, the RBF kernel was chosen for its effectiveness. To find the best hyperparameters, GridSearchCV was used along with TimeSeriesSplit on the training data (1986–2016). This time-series cross-validation keeps the data in order, preventing "look-ahead bias." The procedure conducted a systematic search for the optimal values of  $C$  (regularization) and  $\gamma$  (kernel coefficient) to achieve the best performance on the validation folds, producing a robust and well-generalized model.

### 3.3. Performance Evaluation Metrics

To compare the forecasting accuracy of the four selected models, their performance was evaluated based on the in-sample data (1986–2016) and testing data set (2017–2024). Three of the most common and robust statistical metrics were calculated. In all cases, a lower value indicates a more accurate model [16].

Where  $n$  is the number of observations in the test set,  $y_i$  is the actual, observed value (Actual Demand), and  $\hat{y}_i$  is the forecast value by the model (Predicted Demand).

Root Mean Square Error (RMSE): This is the square root of the average of the squared prediction errors. By squaring the errors, RMSE gives significantly more weight to large errors. It is one of the most popular metrics as it effectively penalizes models that produce large, unacceptable deviations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

Mean Absolute Error (MAE): This metric measures the average absolute magnitude of the errors. Unlike RMSE, it treats all errors linearly (it does not square them), making it less sensitive to large outliers. MAE provides a clear, interpretable measure of the average error in the original units of the data (TWh).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

Mean Absolute Percentage Error (MAPE): This is one of the most useful metrics for interpretation as it is “scale-independent.” It measures the average error as a percentage of the actual value. A MAPE of 5% means that, on average, the model’s forecast is off by 5%. This allows for an intuitive comparison of accuracy, regardless of the data’s scale.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

The model that yields the lowest values across these three metrics, particularly for MAPE, will be identified as the better model and used for the final forecast in the subsequent section.

## 4. Results Analysis and Policy Implication

### 4.1. Performance Comparison

This section presents a comparison of the four forecasting models. The models were trained rely on the historical yearly data from 1986–2016 and tested data from 2017–2024. To simulate the algorithm of four forecasting models were mentioned in the session 3, this study using the Python language [17], All parameters or hyper-parameters of forecasting models was illustrated in the Table 1.

And the error indexes of forecast model for training and testing data set are summarized in Table 2 and Table 3, respectively.

**Table 2. The error index of forecast model for training datasets.**

| Model             | MAE   | RMSE  | MAPE (%) |
|-------------------|-------|-------|----------|
| Holt’s (Additive) | 16.07 | 23.78 | 5.52     |
| PSO-GM(1,1)       | 14.31 | 18.54 | 7.07     |
| SVR               | 16.34 | 19.55 | 7.73     |
| Linear Regression | 61.21 | 73.61 | 35.55    |

**Table 3. The error index of forecast model for testing datasets.**

| Model             | MAE    | RMSE   | MAPE (%) |
|-------------------|--------|--------|----------|
| Holt’s (Additive) | 89.33  | 99.50  | 7.19     |
| SVR               | 95.30  | 104.73 | 7.90     |
| PSO-GM(1,1)       | 111.65 | 141.03 | 8.56     |
| Linear Regression | 417.83 | 429.08 | 33.93    |

From Table 2 and Table 3, the results on the test set clearly indicate that the Holt’s (Additive) model delivered superior performance. Specifically, this model recorded the lowest MAPE for the training and testing data at only 7.19% and 5.52%, respectively. This result demonstrated the highest reliability and forecasting accuracy among all evaluated models.

Follow by is the SVR (tuned) model and improvement of GM (1,1) by PSO algorithm with the MAPE equal 7.90% and 8.56%, respectively. By contrast, the linear regression model exhibited significantly higher forecasting errors with the MAPE is 33.93%. This indicates that simple regression models failed to effectively capture the complex characteristics of the time-series data, rendering them unsuitable for this forecasting task. More detailed was illustrated in the Figure 1 and Figure 2.

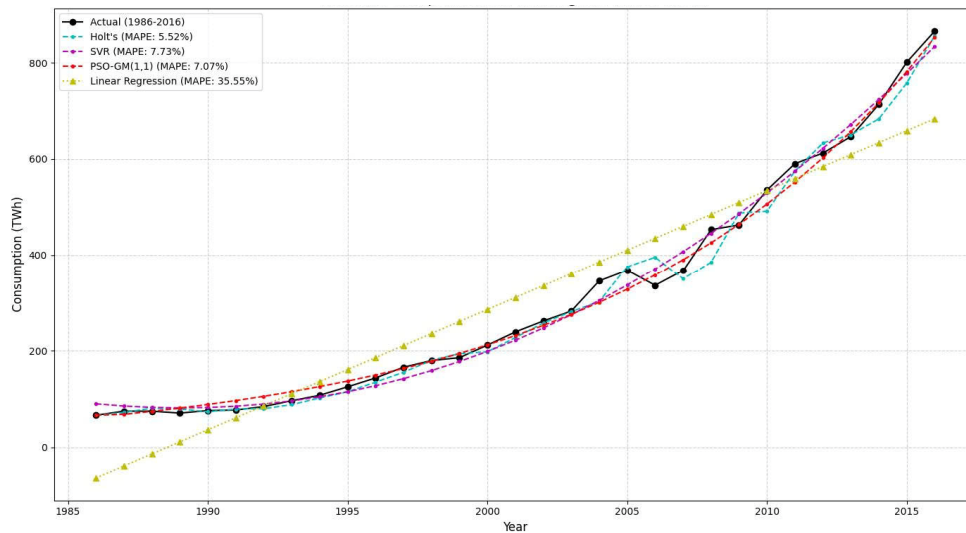


Figure 1. Forecasted value of forecast model for training data set (1986-2016).

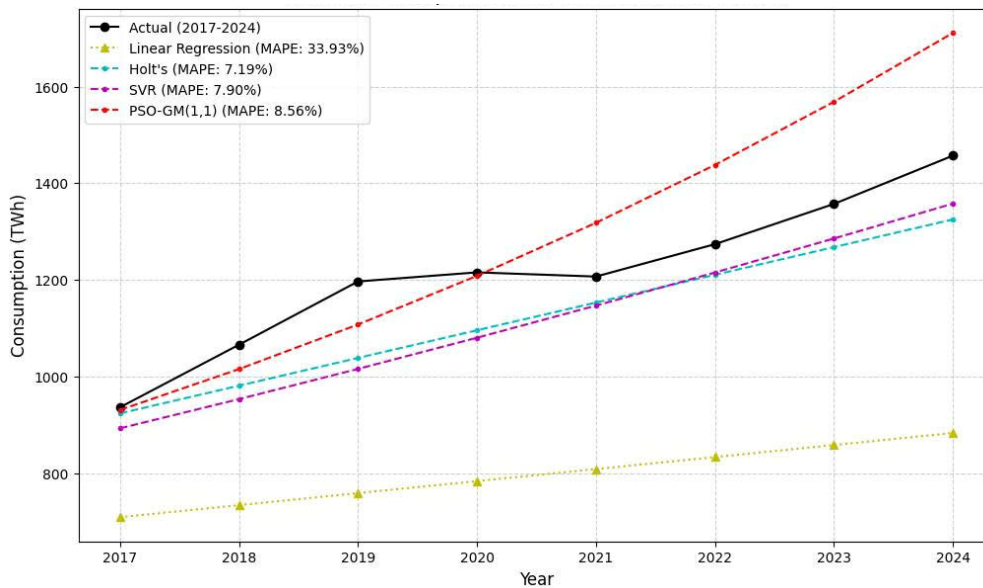


Figure 2. Forecasted value of forecast model for testing data set (2017-2024).

### 5. Forecasting Results and Policy Implication

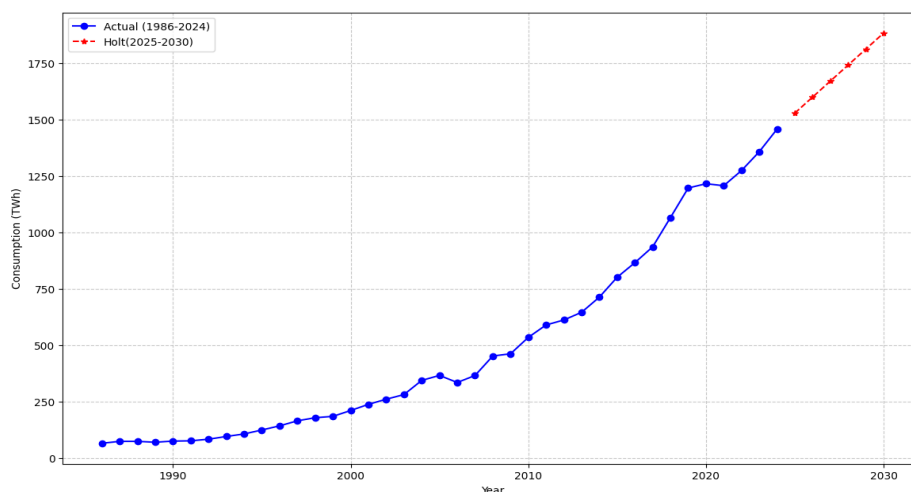
Based on the findings in the Section 4.1, the Holt’s (Additive) model was strongly suggest to out—perform the energy consumption demand forecasting in Vietnam during the period time 2025 to 2030. All forecast result was show in Table 4:

Table 4. Energy consumption demand forecasting during the period time 2025 to 2030.

| Year | Forecasted Value (TWh) |
|------|------------------------|
| 2025 | 1528.08                |
| 2026 | 1598.97                |
| 2027 | 1669.87                |
| 2028 | 1740.76                |
| 2029 | 1811.66                |
| 2030 | 1882.55                |



The result in the Table 4 indicated that the energy consumption demand will be significantly increase in the future. Specifically, It will be reach 1528.08 TWh by 2025, rising to 1882.55 TWh by 2030. These results provide a critical foundation for formulating long-term energy policies and strategic planning based on high-reliability forecast data. Moreover, Holt's Exponential Smoothing model can be further extended and adjusted flexibly as new data become available in the future. Nevertheless, these results should be further validated with updated data for the 2025–2030 period to ensure the model's long-term stability and robustness. More visualization of the future scenarios of the energy consumption demand in Vietnam was show in Figure 3.



**Figure 3. Model performance (2017–2024) and projected future trends (2025–2030).**

Figure 3 illustrates the projected line continues the accelerating trend observed in the historical data, providing a stable, medium-term outlook for national energy planning

Compare with This projection highlights the immense scale of the national challenge, particularly when contextualized with the Power Development Plan 8 (PDP8). PDP8 targets a commercial electricity demand of 505.2 TWh by 2030. Our forecast, which covers the total energy required to generate that electricity plus all other sectors (like transportation and industry), underscores the enormous primary energy supply needed to fuel this growth.

## 6. Conclusions

Within the context of dynamic economic and fluctuation data, the selection of suitable model to forecast the energy consumption demand become more important task in planning and making decision about Vietnam's national energy security in the future. Awareness of the important role, this paper using four forecasting models which are Linear Regression, Holt's, PSO-GM(1,1), and Support Vector Regression—SVR to conduct and find out the the most accurate forecasting model in this cases. Based on the Vietnam's total energy demand data set from 1986 to 2024, The empirical results robustly identified the Holt's Exponential Smoothing as the better model in this situation, achieving the lowest error metrics across all three indices for in and out -of sample. This demonstrates its superior capability in capturing the non-linear and accelerating growth trend inherent in Vietnam's energy data. Based on this model, the forecast indicates that Vietnam's total energy demand will continue its strong upward trajectory, projected to reach 1528.08 TWh by 2025 and exceed 1883 TWh by 2030.

Besides the achievements, this study has the following limitations and this is also the future direction in the next research: Firstly, This study just utilized a uni-variate time-series model. However, the energy consumption demand was significant impact by external exogenous variables. Future research should incorporate multivariate factors such as number of population, GDP growth, foreign direct investment (FDI), and urbanization rates to build more complex and accurate

multivariate forecasting models (e.g., ARIMAX, VAR, or ML models like Random Forest). The secondly is highlighted the challenge of model selection with limited data (n=39). Similarly, while the tuned SVR model was competitive, its performance (MAPE 7.90%) was still outperformed by the stability of the classical Holt's (Additive) model (MAPE 7.19%). This suggests a potential limitation in the generalization capability of these specific regression and ML models in this particular low-data scenario. Future work should explore other data-efficient models (e.g., ARIMA, Prophet) that might balance flexibility and stability differently. Finally, this study focused on total primary energy demand. A valuable future direction is to dis-aggregate the forecast by sector (industrial, residential, commercial) or by energy source (coal, gas, hydro, renewable) to provide more specific and granular policy insights.

**Author Contributions:** The research articles paper was contributed by: Conceptualization, VT, Phan, DT Nguyen, and NXQ, Nguyen.; methodology, VT, Phan, DT Nguyen.; software, VT, Phan, DT Nguyen; NXQ, Nguyen, and XH Huynh; validation, VT, Phan, DT Nguyen, and NXQ, Nguyen; formal analysis, NXQ, Nguyen, and XH Huynh; investigation, VT, Phan, DT Nguyen; resources, VT, Phan, DT Nguyen; data curation, DT Nguyen; NXQ, Nguyen, and XH Huynh; writing—original draft preparation, VT, Phan, DT Nguyen; NXQ, Nguyen, and XH Huynh; writing—review and editing, VT, Phan; visualization, DT Nguyen; NXQ, Nguyen, and XH Huynh; supervision, VT, Phan; project administration, VT, Phan; funding acquisition, VT, Phan. All authors have read and agreed to the published version of the manuscript.”.

**Funding:** This research received no external funding.

**Data Availability Statement:** We encourage all authors of articles published in MDPI journals to share their research data. In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. Where no new data were created, or where data is unavailable due to privacy or ethical restrictions, a statement is still required. Suggested Data Availability Statements are available in section “MDPI Research Data Policies” at <https://www.mdpi.com/ethics>.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. The World Bank (2024), the World Bank in Viet Nam. [available online] <https://www.worldbank.org/en/country/vietnam/overview>, 2024.
2. Hannah Ritchie, Max Roser, and Pablo Rosado. Energy (2022), Our World in Data, [available online] <https://ourworldindata.org/energy>.
3. Prime Minister of Vietnam. Decision No. 500/QĐ-TTg (2023): Approving the National Power Development Plan for the period 2021–2030, with a vision to 2050 (Power Development Plan 8).
4. Vietnam Energy Partnership Group. COP26 outcomes & commitments (2022) [available online] <https://vepg.vn/wp-content/uploads/2022/01/COP26-Outcomes-and-Way-forward.pdf>.
5. R. Jamil , “Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030”, *Renewable Energy*, (2020),154:1–10. DOI: 10.1016/j.renene.2020. 02.117.
6. Zicheng Lin , “A hybrid CNN-LSTM-Attention model for electric energy consumption forecasting (2023)”, *Proceedings of the 2023 5th Asia Pacific Information Technology Conference*, (2023) pp.167-173, DOI:10.1145/3588155.358818.
7. Nguyen Dinh Tien , “Using grey models for forecasting Vietnam’s renewable energy consumption”, *The VMOST Journal of Social Sciences and Humanities*, (2025) DOI: 10.31276/VMOSTJOSSH.2025.0020.
8. Ngoc Tri Ngo, Thi Thu Ha Truong, Ngoc Son Truong, Anh Duc Pham, Nhat To Huynh, Tuan Minh Pham, and Vu Hong Son Pham , “Proposing a hybrid meta-heuristic optimization algorithm and machine learning model for energy use forecast in non-residential buildings”, *Scientific Reports*, (2022), 12: 1065, DOI: 10.1038/s41598-022-04923-7.

9. Nguyen Duy Hieu, Nguyen Vinh Anh, Hoang Anh, and Hoang Duc Chinh (2023), "Biogas electricity production forecasting in livestock farms using machine learning techniques: A case study in Vietnam", *Journal of Science and Technology*, 59(2A), pp. 165–170, DOI: 10.57001/huih5804.2023.060.
10. Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining (2012), *Introduction to linear regression analysis*. John Wiley & Sons, 5th edition, ISBN 978-0-470-54281-1.
11. Julong Deng , "Introduction to Grey system theory", *The Journal of Grey system*, (1989), pp. 1–24.
12. Yuhong Wang, Yaoguo Dang, Yueqing Li, and Sifeng Liu , "An approach to increase prediction precision of GM (1,1) model based on optimization of the initial condition", *Expert Systems with Applications*, 37(8): (2010) pp. 5640–5644, 2010. DOI: 10.1016/j.eswa.2010.02.048.
13. James Kennedy and Russell Eberhart (1995) Particle swarm optimization. In *Proceedings of ICNN'95—International Conference on Neural Networks*, volume 4, pp. 1942–1948. IEEE.
14. Elvis Twumasi, Emmanuel Asuming Frimpong, Daniel Kwegyir, and Denis Folitse , "Improvement of Grey system model using particle swarm optimization", *Journal of Electrical Systems and Information Technology*, 8(1):12, (2021). DOI: 10.1186/s43067-021-00036-9.
15. Vladimir N Vapnikv (1995), *The nature of statistical learning theory*, Springer-Verlag
16. Rob J Hyndman and George Athanasopoulos (2018), *Forecasting: Principles and practice*, OTexts, 2nd edition.
17. Python Software Foundation, *Python language reference*, version 3.11, n.d. [available online] <https://www.python.org/>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.