THE UNIVERSITY OF DA NANG
VIETNAM-KOREA UNIVERSITY OF INFORMATION AND COMMUNICATION TECHNOLOGY
FACULTY OF COMPUTER SCIENCE

APPLICATION OF STATISTICAL AND MACHINE LEARNING MODELS IN VIETNAM'S ENERGY CONSUMPTION DEMAND FORECASTING

Students	**:**	**Nguyen Duc Trien**	**-**	**23AI050**
		Nguyen Ngoc Xuan Quynh	**-**	**23AI042**
		Huynh Xuan Hau	**-**	**23AI014**
Supervisor	**:**	**Dr. Phan Van Thanh**		

DaNang, December 20, 2025

THE UNIVERSITY OF DA NANG

VIETNAM-KOREA UNIVERSITY OF INFORMATION AND COMMUNICATION TECHNOLOGY

FACULTY OF COMPUTER SCIENCE

APPLICATION OF STATISTICAL AND MACHINE LEARNING MODELS IN VIETNAM'S ENERGY CONSUMPTION DEMAND FORECASTING

Students	**:**	**Nguyen Duc Trien**	**-**	**23AI050**
		Nguyen Ngoc Xuan Quynh	**-**	**23AI042**
		Huynh Xuan Hau	**-**	**23AI014**
Supervisor	**:**	**Dr. Phan Van Thanh**		

DaNang, December 20, 2025

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1

1.1. Project overview 1

1.2. Related works 2

CHAPTER 2. METHODOLOGY AND DATA 3

2.1. Data collection and Preparation 3

2.2. Forecasting model 3

2.2.1. Linear Regression 3

2.2.2. Polynomial Regression 4

2.2.3. Holt’s Exponential Smoothing (Double Exponential Smoothing).....4

2.2.4. Grey Model (GM(1,1)) optimized by PSO 5

2.2.5. Support Vector Regression (SVR) 7

2.3. Performance Evaluation Metrics 8

CHAPTER 3. RESULTS ANALYSIS 9

3.1. Performance Comparison 9

3.2. Forecasting Results 12

CONCLUSION 14

REFERENCES 15

CHAPTER 1. INTRODUCTION

1.1. Project overview

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, com- pounded 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 enegry consumption demand[5, 6, 7, 8, 9]. However, all researched papers are discovered under two main directions. The first is focused on the small-medium dataset such as gray theory system and the other direction is longer dataset such as machine learning model. In this study, based on two main direction was above mentioned, we tries to used five distinct models (Linear Regression, Polynomial Regression (degree 2), 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.

1.2. Related works

Research on national energy forecasting can be divided into two main approaches: classical time-series models and modern machine learning/deep learning models. In the classical category, Jamil (2020) applied ARIMA models to Pakistan's annual electricity consumption data, identifying ARIMA(1,2,1) as the best- fitting model. This showed ARIMA's ability to capture 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 optimality unclear.

Recently, 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 Duy Hieu 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 dataset—a long-term, national-level energy forecast with limited annual observations.

CHAPTER 2. METHODOLOGY AND DATA

2.1. Data collection and Preparation

The dataset used in this study is a univariate 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 dataset 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 obser- vations from 2017–2024 (about 20%) for evaluating forecasting performance.

2.2. Forecasting model

The dataset used in this study is a univariate 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 dataset 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 obser- vations from 2017–2024 (about 20%) for evaluating forecasting performance.

2.2.1. 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 (β_0 and β_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.
- β_0 is the intercept (the value of Y when $t = 0$).
- β_1 is the slope (the change in Y for a one-unit increase in t).
- t is the time variable.
- ϵ (epsilon) is the unavoidable error, accounting for random variance not explained by the model

After model fitting, the resulting prediction equation is:

$$\hat{Y} = \beta_0 + \beta_1 t \tag{2}$$

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

2.2.2. Polynomial Regression

Given that Vietnam's energy demand clearly exhibits a non-linear upward curve (growing faster over time), Polynomial Regression (Degree 2) is also evaluated. This is a special case of multiple linear regression where higher-order polynomial terms of time (e.g., $t^2, t^3, \ldots, t^k$) are introduced as predictor variables.

The general theoretical model for a k-th degree polynomial regression is:

$$Y = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k + \epsilon \tag{3}$$

The corresponding prediction equation is:

$$\hat{Y} = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k \tag{4}$$

Where k is the *degree* of the polynomial. For example, $k = 2$ represents a quadratic model, and $k = 3$ represents a cubic model.

To determine the optimal degree (k) of the polynomial (e.g., quadratic, cubic) and, importantly, to avoid overfitting, a strict selection process is employed. Instead of merely fitting to the training data, the model's degree is chosen by applying time-series cross-validation on the training set (1986-2016). The degree k that gives the lowest average validation error across all folds is chosen for the final model.

2.2.3. Holt's Exponential Smoothing (Double Exponential Smoothing)

While Simple Exponential Smoothing (SES) is only effective for time-series data without a trend, the dataset in this study exhibits a clear upward trend. To address this, Holt's (1957) Double Exponen- tial Smoothing method is employed. The 2004 publication by Holts 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 expo- nentially—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, α (alpha, for the level) and β (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:

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{5}$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \tag{6}$$

Equation 5 shows the Level as a weighted average of the current observation (y_t) and the one-step- ahead forecast from the previous period. Equation 6 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 calulated as:

$$F_{t+h} = l_t + h \cdot b_t \tag{7}$$

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, ϕ, 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 α and β are not chosen manually; they are optimized by the model to minimize the forecasting error on the training set.

2.2.4. 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 dataset is small, as in this study. The name GM(1,1) stands for "Grey Model, 1 variable, 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:

1. 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:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1,2,\ldots,n \tag{8}$$

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

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{9}$$

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 metaheuristic algorithm introduced by Kennedy and Eberhart (1995)[13], to optimize a critical input parameter.

- **Optimization Target:** Rather than replacing OLS, PSO is used to iteratively search for the optimal parameter λ (lambda val).
- **Role of λ:** 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 λ 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:

$$\min_{\lambda} F(\lambda) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{MAE}_k \tag{10}$$

An optimized λ 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 xˆ(1):

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} \tag{11}$$

Finally, to retrieve the forecast for the original, non-accumulated data xˆ(0), the model applies an inverse AGO (IAGO), which is a simple subtraction:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{12}$$

2.2.5. 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 ϵ-insensitive tube around the regression function. For any data point yi, if the prediction error is within this tube (i.e., $|yi - f(xi)| \leq \epsilon$), the loss is assumed to be zero. The model's loss function only penalizes data points that lie outside this ϵ-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 sys- tematic search for the optimal values of C (regularization) and γ (kernel coefficient) to achieve the best performance on the validation folds, producing a robust and well-generalized model.

2.3. Performance Evaluation Metrics

To compare the forecasting accuracy of the five 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.

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

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).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

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.

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

CHAPTER 3. RESULTS ANALYSIS

3.1. Performance Comparison

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

Table 1: Optimal parameters and hyperparameters of the model after training

Model	Parameters / Hyperparameters	Optimal value
Holt's (Additive)	α (Level smoothing coefficient)	1.0000
	β (Trend smoothing coefficient)	0.2440
SVR (RBF)	C (Regularization parameter)	5000
	γ (Kernel coefficient)	0.01
PSO-GM(1,1)	λ (Optimized Lambda)	0.3019
	a (Development coefficient)	-0.0870
	b (Grey input)	59.6558
Polynomial (Degree 2)	β_2 (t^2 coefficient)	0.9956
	β_1 (t^1 coefficient)	-3959.5338
	β_0 (Intercept)	3936834.2974
Linear Regression	β_1 (Slope / a)	24.8975
	β_0 (Intercept / b)	-49589.5742

And the error indexes of forecast model for training and testing dataset are summarized in Table 2 and Table 3, repectively.

Table 2: The error index of forecast model for training dataset

Model	MAE	RMSE	MAPE (%)
Polynomial (Degree 2)	**13.21**	**18.96**	**4.95**
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 dataset

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
Polynomial (Degree 2)	113.67	121.48	9.37
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 data and testing data at only 7.19% and 5.52%, respectively. This result demonstrated the highest reliability and forecasting accuracy among all evaluated models.

The second ranking is the SVR (tuned) model - also showed strong performance, ranking second with a very competitive MAPE of 7.90%. Conversely, the Polynomial Regression (Degree 2) and Linear Regression models both exhibited significantly higher forecasting errors. This indicates that simple regression models (even in polynomial form) 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 Fig. 1 and Fig.2 Based on the simulation results, this paper indicated that the clear superiority of the Statistic model named as Holt's Exponential Smoothing (Additive) model with a MAPE only 7.19% on the testing set (2017-2024), it substantially outperformed all counterparts in machine learning (SVR), grey systems (PSO-GM(1,1)), and classical regression (Polynomial, Linear)..

The failure of the Polynomial Regression model (ranked fourth with 9.37% MAPE) provides a critical insight. The rigorous Time-Series Cross-Validation (TSCV) process, documented in the accompanying notebook, correctly identified Degree 2 as the optimal polynomial (Validation MAE = 25.59).

Therefore, the Polynomial model's poor test performance was not a result of "underfitting". Instead, it demonstrates that the fundamental nature of a polynomial function is an inappropriate fit for describing or extrapolating the data's strong, accelerating exponential growth trend.

This contrast in performance is clearly illustrated when comparing the predictive performance of the models on the test set (Figure 2) with their fit on the training set (Figure 1).

Figure 1 shows that the Polynomial (Tier 2) model closely follows the historical data, achieving the lowest training MAPE (4.95%) of the models. This provides visual evidence that the Polynomial (Degree 2) model, although seemingly the best at training, is actually overfitting. It has learned the “noise” of the past data rather than just capturing the overall trend. In contrast, the Holt’s (Additive) (Additive) model, despite having a higher training error (5.52%), has a better ability to generalize, resulting in superior performance on the unseen test data (Figure 2).

Conversely, the Holt’s model, which is inherently designed to capture and project such trends, proved the most robust and accurate. The SVR model also performed exceptionally well (ranked second with 7.90% MAPE), demonstrating that modern, data-efficient machine learning models are a highly compet- itive alternative. Ultimately, however, the classical statistical model (Holt’s), when chosen correctly for the data’s structure, provided the most reliable forecast in this data-constrained (N=39) scenario.

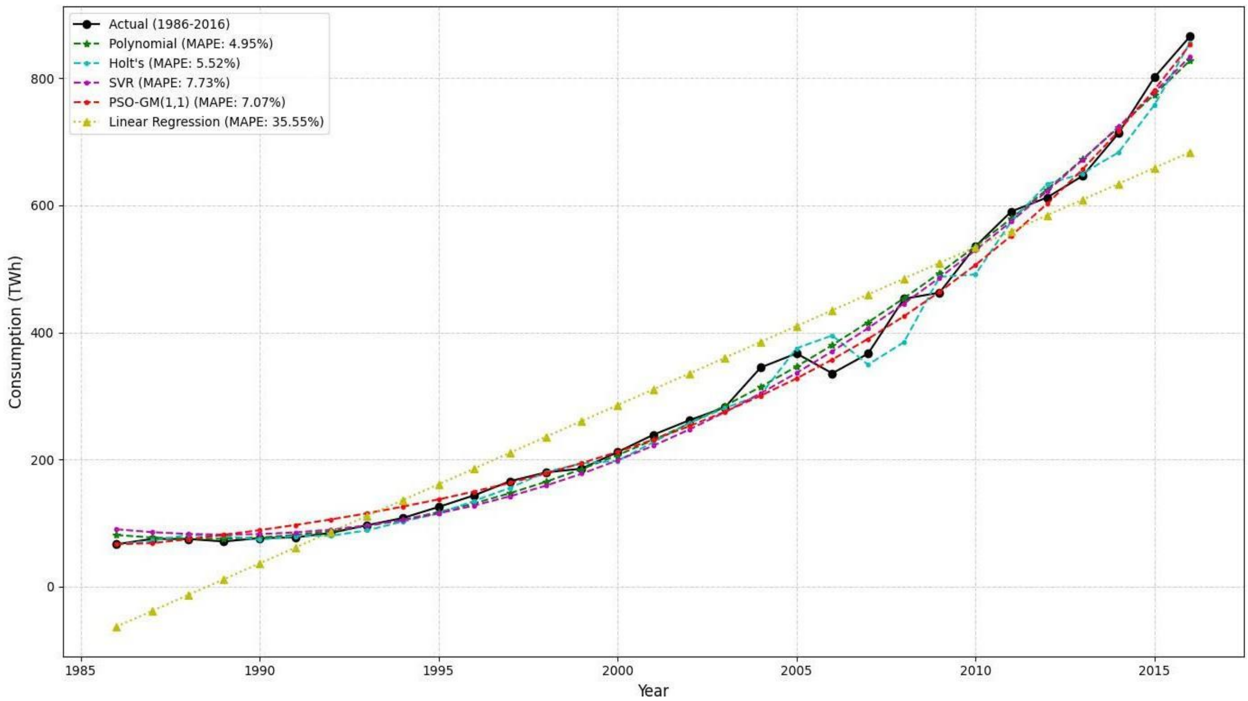

Figure 1: Forecasted value of forecast model for training data set (1986-2016).

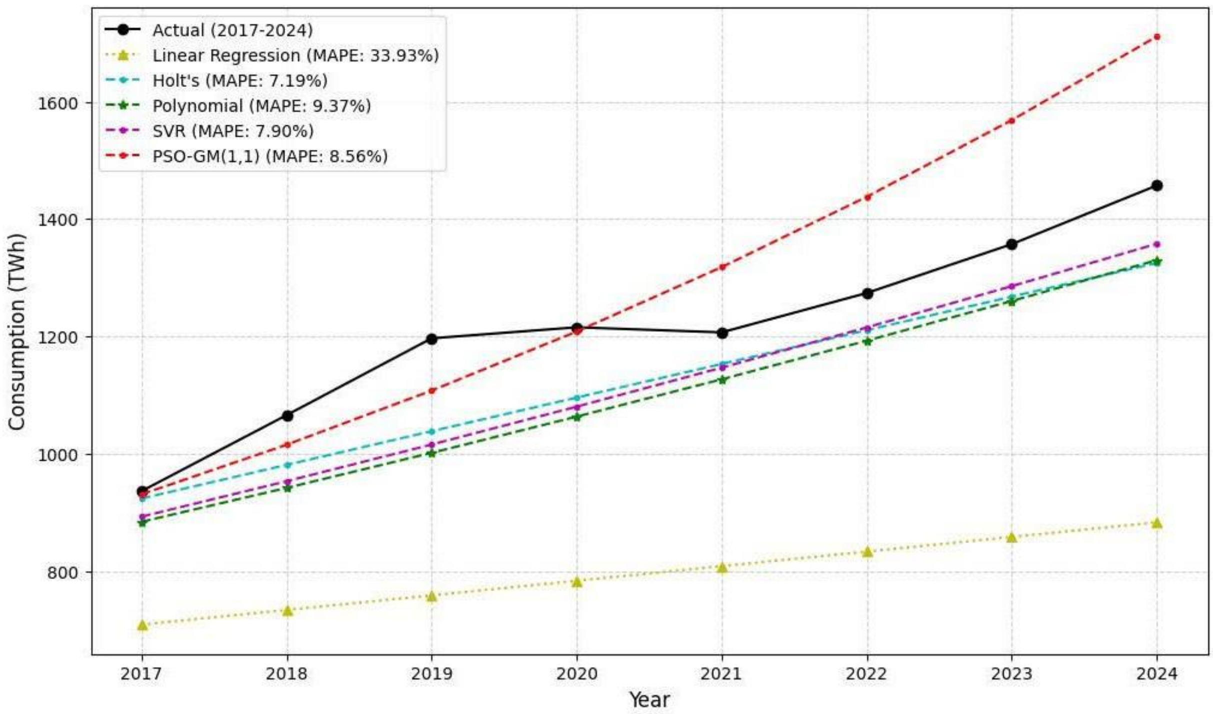

Figure 2: Forecasted value of forecast model for testing data set (2017-2024)

3.2. Forecasting Results

Based on the findings in the Section 4.1, the Holt's (Additive) model was strongly suggest to out - perform to forecast the energy demand 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 Fig 3.

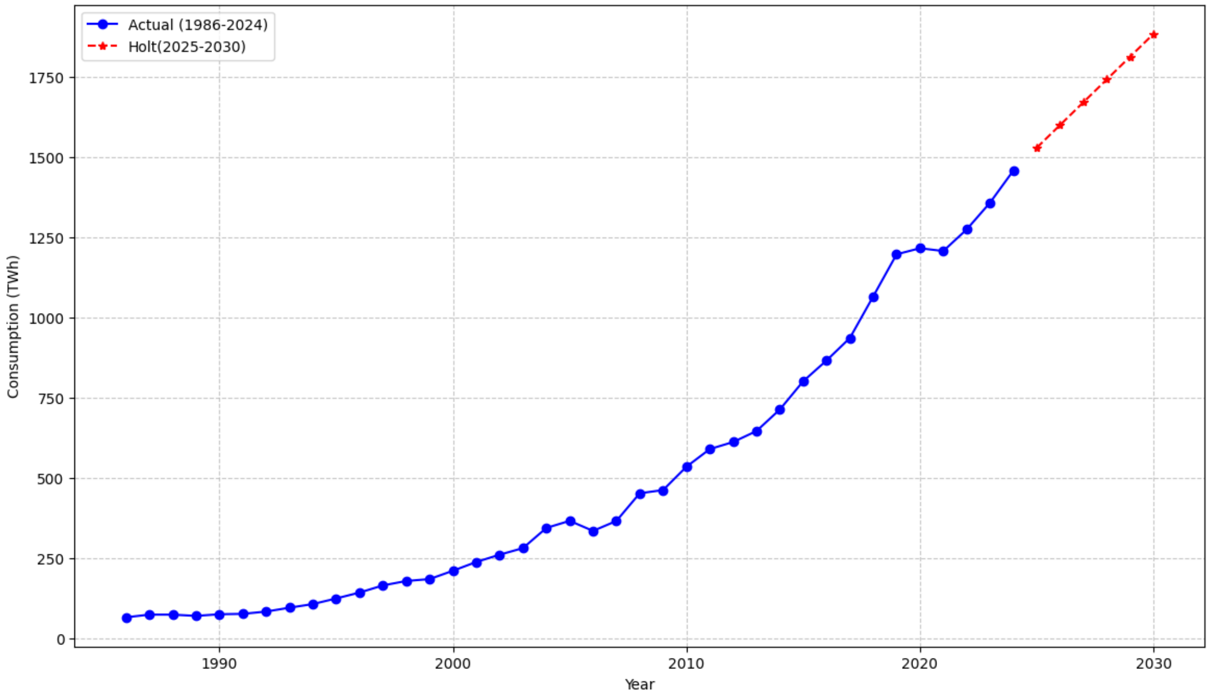

Figure 3: Model performance (2017–2024) and projected future trends (2025–2030).

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.

CONCLUSION

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 five forecasting models which are Linear Regression, Polynomial 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 dataset from 1986 to 2024, The empirical results robustly identified the Holt's Exponential Smoothing as the "champion model", 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 univariate 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. The secondly is highlighted the challenge of model selection with limited data (n=39). The rigorous Time-Series Cross-Validation, while correctly preventing the overfitting of a high-degree polynomial, resulted in a simpler (degree=2) model that ultimately underfitted the data. 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 disaggregate the forecast by sector or by energy source (coal, gas, hydro, renewables) to provide more specific and granular policy insights.

REFERENCES

[1] The World Bank. The World Bank in Viet Nam. https://www.worldbank.org/en/country/ vietnam/overview, 2024.

[2] Hannah Ritchie, Max Roser, and Pablo Rosado. Energy. *Our World in Data*, 2022. URL https://ourworldindata.org/energy.

[3] Prime Minister of Vietnam. Decision No. 500/QD-TTg: Approving the National Power Development Plan for the period 2021 - 2030, with a vision to 2050 (Power Development Plan 8). Decision 500/QD- TTg, Government of Vietnam, 2023.

[4] Vietnam Energy Partnership Group. COP26 outcomes & commitments. https://vepg.vn/ wp-content/uploads/2022/01/COP26-Outcomes-and-Way-forward.pdf, 2022.

[5] R. Jamil. Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply- demand analysis for the year 2030. *Renewable Energy*, 154:1–10, 2020. doi: 10.1016/j.renene.2020. 02.117.

[6] Zicheng Lin. A hybrid CNN-LSTM-Attention model for electric energy consumption forecasting. *Proceedings of the 2023 5th Asia Pacific Information Technology Conference*, 2023. URL https://api.semanticscholar.org/CorpusID:259139710.

[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 metaheuristic optimization algorithm and machine learning model for energy use forecast in non-residential buildings. *Scientific Reports*, 12: 1065, 2022. doi: 10.1038/s41598-022-04923-7.

[9] Nguyen Duy Hieu, Nguyen Vinh Anh, Hoang Anh, and Hoang Duc Chinh. Biogas electricity pro- duction forecasting in livestock farms using machine learning techniques: A case study in Vietnam. *Journal of Science and Technology*, 59(2A):165–170, 2023. doi: 10.57001/huih5804.2023.060.

[10] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 5th edition, 2012. ISBN 978-0-470-

54281-1.

[11] Julong Deng. Introduction to grey system theory. *The Journal of Grey System*, 1:1–24, 1989.

[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):5640–5644, 2010. doi: 10.1016/j.eswa.2010.02.048.

[13] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.

[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 Vapnik. *The nature of statistical learning theory*. Springer-Verlag, 1995.

[16] Rob J Hyndman and George Athanasopoulos. *Forecasting: Principles and practice*. OTexts, 2nd edition, 2018.

[17] Python Software Foundation. Python language reference, version 3.11, n.d. URL https://www. python.org/.