Computer Modeling in
Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.086196

ARTICLE

A New Hybrid Framework Based on Grey and Neuro-Fuzzy Inference System for Energy Demand Forecasting in Vietnam

Xuan Kien Pham[1], Van Dat Nguyen[2,*], Van Thanh Phan[3,*] and Duc Trien Nguyen[4,*]

[1]Faculty of Management Information Systems, Ho Chi Minh University of Banking, 36 Ton That Dam, Nguyen Thai Binh Ward, District 1, Ho Chi Minh City, Vietnam

[2]Faculty of Business Administration, Ho Chi Minh University of Banking, 36 Ton That Dam, Nguyen Thai Binh Ward, District 1, Ho Chi Minh City, Vietnam

[3]Faculty of Digital Economy and E-Commerce, Vietnam-Korea University of Information and Communication Technology, The University of Danang, 470 Tran Dai Nghia, Ngu Hanh Son Ward, Danang, Vietnam

[4]Faculty of Computer Science, Vietnam-Korea University of Information and Communication Technology, The University of Danang, 470 Tran Dai Nghia, Ngu Hanh Son Ward, Danang, Vietnam

*Corresponding Authors: Van Dat Nguyen. Email: datnv@hub.edu.vn; Van Thanh Phan. Email: pvthanh@vku.udn.vn; Duc Trien Nguyen. Email: triennd.23ai@vku.udn.vn

Received: 26 May 2026; Accepted: 01 July 2026; Published: 27 July 2026

ABSTRACT: Accurate energy consumption forecasting faces two major challenges: limited historical data and complex consumption patterns. To address these challenges, this study proposes a new hybrid framework named the Decomposition-based Grey-Neuro-Fuzzy Architecture (DeGNA). The model first uses the Denton method to convert limited annual records into high-frequency monthly data. Next, it applies STL decomposition to separate the data into trend, seasonal and residuals components. A rolling-window GM(1,1) model is then used to predict the main growth trend, while a GWO-optimized ANFIS model uses economic indicators (IIP and FDI) to forecast complex seasonal changes. This study evaluates the performance of the DeGNA model using Vietnam's energy data from 2009 to 2024. The results show that DeGNA achieves high accuracy with a MAPE of 4.11%, performing competitively alongside modern machine learning (XGBoost) and deep learning (LSTM, BiLSTM) models. More importantly, a 5-fold cross-validation reveals that while autoregressive models such as SARIMAX perform well during stable growth periods, DeGNA provides more consistent performance under macroeconomic disruptions. While SARIMAX struggles with sudden demand shocks (maximum error of 22.23), DeGNA maintains a much tighter error distribution (maximum 15.39) and the lowest variance (±1.09%). Therefore, DeGNA provides a highly reliable tool for national energy planning under data constraints.

KEYWORDS: Data limitation; temporal disaggregation; hybrid time-series modeling; ANFIS; meta-heuristic optimization; emerging economies

1 Introduction

Rapid growth in developing countries makes energy security very important. Vietnam is a prime example. Due to increased manufacturing and foreign investment, Vietnam's electricity demand increased by an average of 11% during the 10-year period before the COVID-19 pandemic [1]. To support this growth, Vietnam's electricity demand is projected to reach approximately 505 billion kWh by 2030, with peak capacity demand of around 90,500 MW [2]. However, planning for this growth is challenging because there isn't enough detailed historical data. Because of this missing data, accurate energy forecasting becomes crucial.

It helps prevent power shortages, guides investments in the power grid, and supports the goal of net-zero emissions by 2050.

Recently, many researchers used various methods to forecast energy, but each faces limitation in environments with limited data. Classical statistical models are easy to understand and theoretically reliable, but they need long, continuous historical records to work effectively [3]. In addition, artificial intelligence and deep learning models are known for the ability to detect complex patterns. However, they require large amounts of data. When data is limited, they risk over-fitting and act like "black boxes", making it difficult for policymakers to trust or utilize their results [3]. As a result, depending on these standard models in data-limited situations presents a challenging trade-off between the amount of data needed and the accuracy of predictions.

To overcome these limitations, hybrid decomposition-based frameworks are often used. However, existing models fail to address Vietnam's specific challenges: the need to break down limited annual data, model nonlinear seasonal patterns, and handle extreme demand shocks. To bridge this gap, this study proposes DeGNA, a novel architecture that combines the Denton method for temporal disaggregation, STL decomposition, GM(1,1), and GWO-ANFIS. Evaluated on Vietnam's primary energy consumption data (2009–2024), DeGNA not only demonstrates competitive average accuracy but also reduces the maximum absolute forecasting error by 44.5% compared to SARIMAX.

2 Literature Review

In energy forecasting, the lack of high-frequency data poses a significant methodology challenge. While many regional studies rely on data-heavy models, these approaches often fail to address structural data constraints or misrepresent actual economic trends [4,5]. To overcome this, the Denton proportional method [6] has been applied as an econometric solution, disaggregating annual statistics into high-frequency series while preserving temporal consistency.

Even with high-frequency data, traditional standalone models (e.g., classical statistical methods or standard deep learning) struggle to capture complex non-linear seasonality. Consequently, the literature has increasingly shifted towards hybrid architectures [7]. Integrating Seasonal-Trend decomposition using Loess (STL) has proven highly effective in isolating structural signals, allowing specialized models to process trend and seasonal components separately [8].

Within these hybrid frameworks, fuzzy-based systems like the Adaptive Neuro-Fuzzy Inference System (ANFIS) are widely adopted to manage demand uncertainty, often outperforming standard neural networks [9]. However, classical gradient-based ANFIS training is stuck in local optima. To address this, recent studies have integrated meta-heuristic algorithms to tune neuro-fuzzy parameters by effectively balancing global exploration and local exploitation [10].

Despite these advances, recent hybrid forecasting studies still exhibit critical limitations across three distinct design philosophies. While decomposition-based ensembles effectively separate structural from residual dynamics [8,11], their seasonal components are modeled using linear or standard ML methods that cannot adaptively capture nonlinear economic-driven patterns. Similarly, grey system approaches retain strong performance under limited sample sizes [12], but they frequently constrain the seasonal structure to predefined functional forms or tune components in isolated stages, or handle trend and seasonality jointly without structural decomposition [13]. Furthermore, although ANFIS-metaheuristic hybrids successfully apply global search to parameter optimization [14,15], they typically lack a preceding decomposition stage to isolate the underlying trend from complex seasonality.

To systematically illustrate these methodological gaps, Table 1 summarizes the existing frameworks and shows their limitations. Ultimately, no single framework currently combines all these solutions to simultaneously solve data shortages, track long-term growth, and handle complex seasonality. This gap directly motivates the proposed DeGNA architecture.

Table 1: Summary of related hybrid energy forecasting frameworks and their methodological limitations.

Study	Method	Data/Problem	Methodological Gap
Zhang & Li (2021) [8]	STL + SARIMA/SVR/ANN/LSTM ensemble	Monthly electricity consumption, China	Assumes sufficient high-frequency data; no exogenous variables; seasonal component modeled with linear ML methods
Tang et al. (2022) [12]	GM(1,1) + seasonal index (linear/exp/log trend forms)	Monthly electricity, industrial sector, western China	Seasonal index fixed to predefined functional forms, not adaptively learned
Athanasopoulou et al. (2021) [11]	Linear regression (seasonality) + LSTM	Half-hourly electricity demand, UK	Designed for data-rich, short-term horizons; linear seasonality cannot capture nonlinear economic-driven patterns
Wu et al. (2023) [14]	ANFIS–ELM optimized via Developed Parasitism–Predation Algorithm	Short-term (hourly) electricity demand	No signal decomposition; trend and seasonal patterns learned end-to-end without structural separation
Bakare et al. (2024) [15]	ANFIS + Gene Expression Programming	Monthly industrial load, Uganda	No structural decomposition; exogenous inputs excluding macroeconomic drivers
Hao & Ma (2024) [13]	Grey model + stacked LSTM, grid-search-tuned	Annual energy consumption (coal, electricity, gasoline), Henan Province, China	Annual-only input, without decomposition, no economic context

3 Methodology and Proposed Approach

3.1 The Denton Proportional Method

Advanced forecasting models often require large, high-frequency datasets to capture the temporal patterns. To overcome the limitation of annual energy consumption records, this study uses the Denton proportional disaggregation method [16]. This method converts low-frequency annual observations into monthly estimates by using a related indicator while preserving consistency with the original annual totals.

Let A_T be the annual energy consumption for year T (where $T = 1, \ldots, M$), and $\mathrm{I_t}$ be the corresponding monthly indicator for month t $(t = 1, \ldots, 12M)$. The method aims to estimate the monthly energy series X_t, under the constraint that the monthly estimates sum to the observed annual value for each year. This constraint is formally expressed as:

$$\sum_{t \in T} X_t = A_T, \text{ for } T = 1, \ldots, M \tag{1}$$

Specifically, X_t is estimated by minimizing the first differences in the ratio between the interpolated series and the indicator ($\mathrm{X_t/I_t}$). This approach preserves the natural variability of the high-frequency indicator while avoiding the excessive smoothing introduced by linear interpolation. The objective function is defined as:

$$\min_{\mathrm{X_t}} \sum_{\mathrm{t=2}}^{\mathrm{12M}} \left(\frac{\mathrm{X_t}}{\mathrm{I_t}} - \frac{\mathrm{X_{t-1}}}{\mathrm{I_{t-1}}} \right)^2 \tag{2}$$

Subject to the aggregation constraint defined in Eq. (1), this approach solves via quadratic programming and distributes annual residuals without distorting the underlying trend. This property makes the method well suited for modeling energy consumption, which generally follows changes in economic activity.

3.2 Seasonal and Trend Decomposition Using Loess (STL)

Energy consumption data are complex, as it is a combination of long-term macroeconomic growth and the short-term cyclic fluctuations. So putting this complex, changing data straight into prediction models can decrease the accuracy. To separate these different dynamics, the Seasonal and Trend decomposition using Loess (STL) is applied to the high-frequency monthly datasets generated in the previous disaggregation step [17].

The STL method is a strong, filtering-based algorithm that decomposes the original time series into three additive components. Mathematically, the original sequence X_t at time t is expressed as:

$$X_t = T_t + S_t + R_t \tag{3}$$

where:

- T_t represents the trend component, capturing the low-frequency, long-term progression of energy demand driven by economic factors.
- S_t denotes the seasonal component, isolating the periodic cyclic variations (such as the 12-month climate-driven cycle).
- R_t is the stochastic residual component, which contains unpredictable, non-linear noise and market shocks.

3.3 Rolling-Window Grey Forecasting Model—GM(1,1)

The baseline Grey Model, GM(1,1) [18,19], is well known for effectively managing small datasets. After applying STL decomposition, GM(1,1) is used to forecast the deterministic trend component (T_t). However, standard grey models are static, their accuracy tends to decline in long-term forecasts because they give equal importance to all past observations. To resolve this, a rolling-window approach with a fixed size of 12 steps is applied [20]. This 12-month window is selected because it matches the annual cycle of macroeconomic and energy data. By using exactly one year of data, this setup helps the model capture short-term trends and avoid noise from outdated information.

Mathematically, let the trend sequence within the current window be defined as

$$X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \right), n \geq 4 \tag{4}$$

The process within each window proceeds through the following steps:

Step 1: Accumulated Generating Operation (AGO). To weaken the randomness of the raw data, the first-order AGO sequence $x^{(1)}$ is generated:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, 3, \ldots, n \tag{5}$$

Step 2: Grey Differential Equation. The core GM(1,1) model is established based on a first-order differential equation. The data matrix B and the data vector Y are constructed as follows:

$$\boldsymbol{B} = \begin{bmatrix} -0.5 \times (x^{(1)}(2) + x^{(1)}(1) & 1 \\ -0.5 \times (x^{(1)}(3) + x^{(1)}(2) & 1 \\ \vdots & \vdots \\ -0.5 \times (x^{(1)}(n) + x^{(1)}(n-1) & 1 \end{bmatrix}; \boldsymbol{Y} = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{6}$$

Step 3: Parameter Estimation. The parameter vector $\hat{a}$, consisting of the development coefficient a and the grey action quantity b, is estimated using the Ordinary Least Squares (OLS) method:

$$\hat{\boldsymbol{a}} = \begin{bmatrix} a \\ b \end{bmatrix} = \left(\boldsymbol{B}^T \boldsymbol{B}\right)^{-1} \boldsymbol{B}^T \boldsymbol{Y} \tag{7}$$

Step 4: Time Response Function. By solving the differential equation, the predicted accumulated value at step $k + 1$ is given by the time response function:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a}, k = 1, 2, \ldots \tag{8}$$

Step 5: Inverse AGO (I-AGO) and Rolling Mechanism. The predicted trend value is then reconstructed by subtracting the previous accumulated value:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{9}$$

To forecast the next month (step $n + 1$), the formula is applied for $k = n$. When this prediction is generated, the temporal window moves forward by one month. The oldest observation $x^{(0)}(1)$ is discarded, the newly observed value is appended, and the parameters $\hat{a}$ are continuously re-estimated. This sliding mechanism ensures the trend projection is adaptive to the latest changes in the time series.

3.4 Adaptive Neuro-Fuzzy Inference System (ANFIS) Architecture

While GM(1,1) is effective for linear trend, it is unsuited to capture volatile, non-linear fluctuations. To address this limitation, the Adaptive Neuro-Fuzzy Inference System (ANFIS) is used [21]. ANFIS effectively maps complex, non-linear relationships by combining the learning capabilities of artificial neural networks with the reasoning strengths of fuzzy logic. This study used a first-order Takagi-Sugeno ANFIS architecture [22], as illustrated in Fig. 1.

Layer 1: Fuzzification Layer. This layer contains adaptive nodes that map the input variables into fuzzy membership degrees. For each node i the output represents the degree to which the input x belongs to the fuzzy set A_i (e.g., "Low", "High"). The corresponding output is expressed as:

$$O_{1,i} = \mu_{A_i}(x), \text{ for } i = 1, 2, \ldots \tag{10}$$

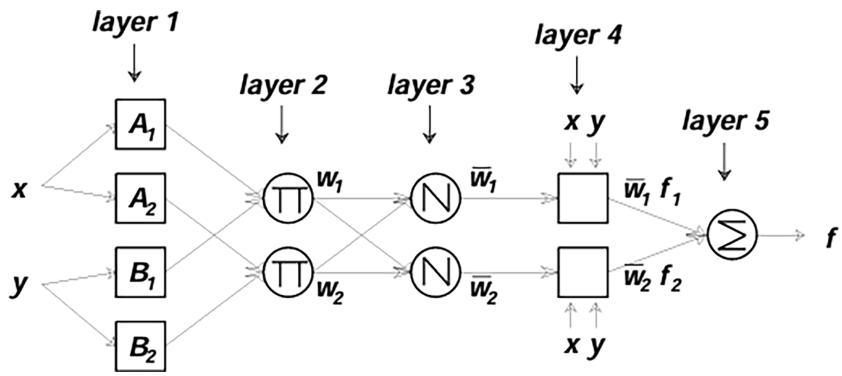

Figure 1: Architecture of the first-order Takagi-Sugeno ANFIS model. The diagram illustrates the five processing layers and highlights the specific parameters targeted for optimization.

This study employs Gaussian membership functions owing to their smoothness and strong capability in representing nonlinear patterns. The formulation of the Gaussian function is given by:

$$\mu_{A_i}(x) = \exp\left(-\frac{1}{2}\left(\frac{x - c_i}{\sigma_i}\right)^2\right) \tag{11}$$

where $\{c_i, \sigma_i\}$ are the center and width.

Layer 2: Rule Layer. This layer is composed of fixed nodes represented by the Π operator. Each node determines the firing strength of a fuzzy rule by multiplying the incoming membership degrees, corresponding to the fuzzy AND operation:

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), i = 1, 2 \tag{12}$$

Layer 3: Normalization Layer. This layer contains fixed nodes denoted by N. Each node normalizes the firing strength of the corresponding rule by dividing it by the sum of the firing strengths from all rules:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1, 2 \tag{13}$$

Layer 4: Defuzzification (Consequent) Layer. This is the second adaptive layer. Each node computes the contribution of its corresponding fuzzy rule through a linear function. The output is defined as:

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{14}$$

where $\{p_i, q_i, r_i\}$ are consequent parameters.

Layer 5: Output Layer. The final layer consists of a single fixed node, denoted by σ. It generates the overall predicted output by aggregating the outputs received from the preceding layer:

$$O_{5,i} = \text{Overall Output} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{15}$$

Traditional ANFIS training typically uses a hybrid method combining gradient descent and least squares. However, because the search space is highly complex, gradient-based methods frequently get trapped in local optima. To address this limitation, the proposed framework employs the Grey Wolf Optimizer (GWO) to optimize both the premise parameters (Layer 1) and the consequent parameters (Layer 4).

3.5 Grey Wolf Optimizer

Training the ANFIS model can be formulated as a global optimization problem, where the objective is to minimize the prediction error. Let Θ denote the complete parameter vector, consisting of both the nonlinear premise parameters and the linear consequent parameters:

$$\Theta = \sigma_{ij}, c_{ij}, p_k, q_k, r_k \tag{16}$$

here, σ and c represent the width and center of the Gaussian membership functions, while p, q, r are the consequent weights.

The optimization objective is to determine the optimal set Θ^* that minimizes the cost function, defined as the Mean Squared Error (MSE):

$$\Theta^* = \arg\min_{\Theta} \left(\frac{1}{N} \sum_{t=1}^{N} \left(y_{\text{actual}}^{(t)} - y_{\text{forecast}}^{(t)} (\Theta) \right)^2 \right) \tag{17}$$

where N is the sample size, $y_{\text{actual}}^{(t)}$ is the observed value, and $y_{\text{forecast}}^{(t)}$ is the model output.

Developed by Mirjalili et al. (2014), the GWO [23] is a meta-heuristic algorithm inspired by the social hierarchy and hunting behavior of grey wolves. The best solution is designated as the leader (α), followed by the second (β) and third (δ) best solutions, while all remaining candidates are classified as followers (ω). The positions of the ω wolves are updated using the guidance provided by the three leading wolves, thereby simulating the hunting process. The distance and movement toward each leader are calculated as follows:

$$\boldsymbol{D}_{\alpha} = \left| \boldsymbol{C}_1 \cdot \boldsymbol{X}_{\alpha} - \boldsymbol{X}(t) \right| \tag{18}$$

$$\boldsymbol{D}_{\beta} = \left| \boldsymbol{C}_2 \cdot \boldsymbol{X}_{\beta} - \boldsymbol{X}(t) \right| \tag{19}$$

$$\boldsymbol{D}_{\delta} = \left| \boldsymbol{C}_3 \cdot \boldsymbol{X}_{\delta} - \boldsymbol{X}(t) \right| \tag{20}$$

$$\boldsymbol{X}_1 = \boldsymbol{X}_{\alpha} - \boldsymbol{A}_1 \cdot \boldsymbol{D}_{\alpha} \tag{21}$$

$$\boldsymbol{X}_2 = \boldsymbol{X}_{\beta} - \boldsymbol{A}_2 \cdot \boldsymbol{D}_{\beta} \tag{22}$$

$$\boldsymbol{X}_3 = \boldsymbol{X}_{\delta} - \boldsymbol{A}_3 \cdot \boldsymbol{D}_{\delta} \tag{23}$$

The final position of a wolf in the next iteration is determined by averaging the positions indicated by the three leaders:

$$\boldsymbol{X}(t+1) = \frac{\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3}{3} \tag{24}$$

In these equations, $\boldsymbol{A}$ and $\boldsymbol{C}$ are coefficient vectors. Throughout the optimization process, these parameters are adaptively updated to enable a gradual transition from exploration to exploitation.

3.6 Forecasting Framework Proposed

While individual forecasting models possess specific strengths, they often fail to capture the full series of energy demand when used in isolation. Building upon the strengths of temporal disaggregation, neural forecasting, and heuristic optimization, this study develops a hybrid framework termed the Decomposition-based Grey-Neuro-Fuzzy Architecture (DeGNA).

The DeGNA framework is systematically designed to address data scarcity as well as model linear economic growth and highly non-linear seasonal fluctuations. The complete pipeline is illustrated in Fig. 2 and proceeds through these fundamental steps:

Step 1: Temporal Disaggregation: Annual primary energy consumption data, alongside monthly reference indicator, are fed into the Denton proportional module. This mathematical transformation generates a high-frequency monthly energy series without data loss.

Step 2: Structural Decomposition: To separate the underlying economic growth from short-term market noise, the high-frequency energy series undergoes STL decomposition. This process splits the data into three components: underlying trend, periodic seasonality, and residuals.

Step 3: Underlying Trend Forecasting: The underlying trend component is processed by the Rolling-window GM(1,1) module (with a window size of n = 12). This step efficiently projects the baseline growth trajectory while adapting to recent economic shifts.

Step 4: Non-Linear Forecasting: At the same time, an enriched feature is constructed by combining the seasonality, the residuals, and exogenous variables. This input is fed into the GWO-ANFIS engine. The GWO algorithm globally optimizes the premise and consequent parameters of the fuzzy network to accurately forecast the highly complex, non-linear variations.

Step 5: Final Forecast: In the final step, the predicted trend from the GM(1,1) module and the predicted non-linear variation from the GWO-ANFIS module are additively combined. This reconstruction yields the final, high-precision energy demand forecast.

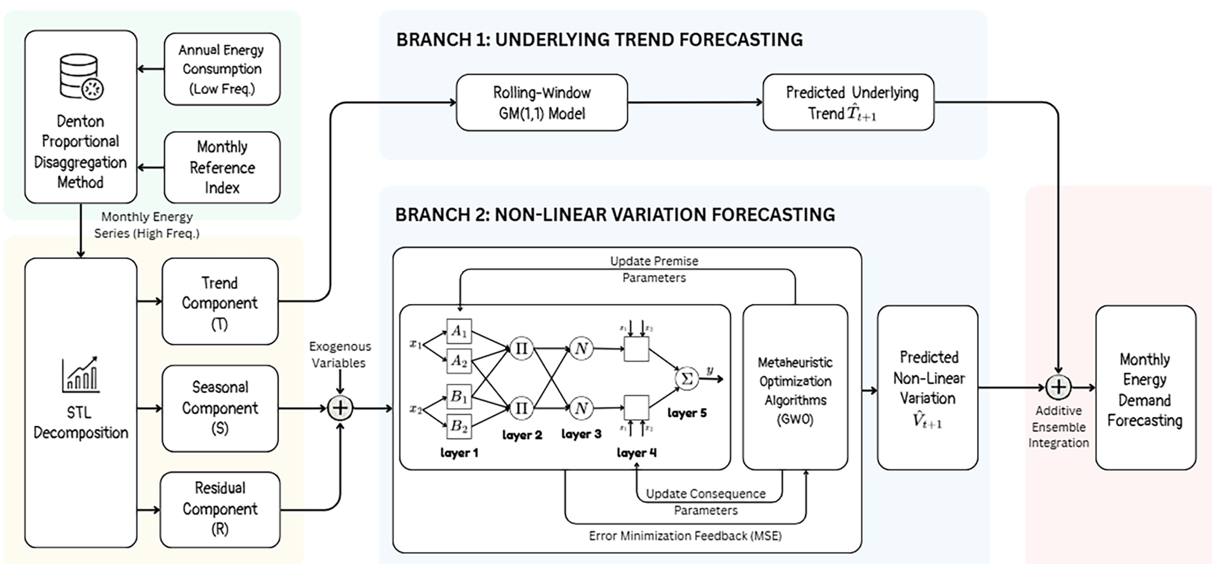

Figure 2: Flowchart of the proposed DeGNA framework, illustrating the data processing pipeline from Denton disaggregation to the final forecast.

4 Simulation Experimental

4.1 Data Preparation

This study uses macroeconomic and energy data from Vietnam from January 2009 to December 2024. The main target variable is total energy consumption, which is collected annually from Our World in Data [24]. Because forecasting models need higher-frequency data to train effectively, we use the monthly commercial electricity output from the Ministry of Industry and Trade (MOIT) [25] as a reference series. Electricity use is a reliable indicator of total energy demand, making it highly suitable for temporal disaggregation.

Additionally, we selected two primary input variables for the forecasting model: the Index of Industrial Production (IIP) and registered Foreign Direct Investment (FDI). We collected the IIP from the General Statistics Office [26] because the industrial sector accounts for most of Vietnam's energy consumption. FDI data comes from the Ministry of Planning and Investment [27], since the country's manufacturing activities rely heavily on foreign capital, which directly affects energy demand. The descriptive statistics for all datasets are summarized in Table 2.

Table 2: Descriptive statistics of the raw datasets.

	Variable	N	Mean	Std. Dev.	Min	Max	Skewness
Annual Data	Primary Energy (Billion kWh)	16	933.54	322.95	462.68	1457.18	0.0904
Monthly Indicators	Industrial Production Index (IIP)	192	101.46	46.04	36.90	176.20	0.1474
	Foreign Direct Investment (Million USD)	192	1376.15	594.62	300.00	3671.00	0.8436
	Commercial electricity (Billion kWh)	192	13.99	5.3143	5.0620	25.2000	0.1773

As mentioned, advanced forecasting models—especially neuro-fuzzy—require large, high-frequency datasets to find patterns effectively. However, our total energy data is annual, while the economic indicators are monthly. To fix this difference in data frequency, we applied the Denton proportional disaggregation method. Using the commercial electricity output as our reference indicator, this process converts the annual energy data into a monthly series. The Denton method is ideal for this task because it ensures that the new monthly values sum exactly to the original annual total while preserving the seasonal trends in the electricity data.

4.2 Benchmark Models Configuration

Statistical Models (ARIMA and SARIMAX) [28]. For the traditional statistical models, the specific parameters included auto-regressive order (p), differencing degree (d), and moving average order (q), along with their seasonal counterparts (P, D, Q, m) for SARIMAX were determined by using an automated stepwise search algorithm to minimize the Akaike Information Criterion (AIC) [29]. Based on this search process, the ARIMA model was configured with the order ($p = 0, d = 1, q = 0$). The SARIMAX model was configured with a non-seasonal order of $(2, 1, 2)$ and an additional seasonal component ($P = 1, D = 0, Q = 2, m = 12$) to capture the 12-month cycle in the data, with IIP and FDI included as exogenous variables.

Grey Forecasting Model—GM(1,1). The baseline GM(1,1) model is well known for handling small datasets effectively [18,19]. To adapt it for continuous monthly forecasting, we applied a rolling-window approach [20] with a fixed size of 12 steps. We chose this 12-month window because it perfectly matches the annual cycle of our macroeconomic and energy data. By using exactly one year of data, this setup helps the model capture short-term trends and avoid noise from outdated information.

Machine Learning & Deep Learning Models. To evaluate the performance against modern data-driven approaches, we designed specific architectures for deep learning (LSTM [30], BiLSTM [31]) and implemented an advanced tree-based machine learning model (XGBoost [32]). For the deep learning baselines, both LSTM and BiLSTM architectures consist of a single 32-unit recurrent layer, a Dropout layer (rate = 0.1) to avoid over-fitting, and a 16-unit ReLU Dense layer. Both were compiled with the Adam optimizer (learning rate = 0.001) and trained for up to 300 epochs, utilizing a 15% validation split and an early stopping mechanism (patience = 20). Furthermore, an Extreme Gradient Boosting (XGBoost) model was introduced to represent state-of-the-art machine learning. To strictly prevent over-fitting on the constrained datasets, its complexity was restricted using a shallow tree structure (max depth = 3), a learning rate of 0.05, 300 estimators, alongside both L1 and L2 regularization penalties.

Neuro-Fuzzy Systems. To evaluate the proposed architecture, three first-order Takagi-Sugeno ANFIS [21] variations (Gaussian membership, 4 rules) were implemented: Standard ANFIS, ANFIS-GWO, and DeGNA. The number of fuzzy rules was deliberately limited to four to reduce model complexity and mitigate over-fitting. Given four input features (lag-1, lag-12, IIP, and FDI), increasing the number of rules would substantially expand the parameter space. Restricting the model to 4 rules keeps the number of trainable parameters proportional to the available sample size (approximately 192 monthly observations), helping maintain a balance between model flexibility and generalization. Standard ANFIS was optimized via gradient descent (learning rate: 0.01, 500 epochs). Conversely, to prevent local optima, both ANFIS-GWO and DeGNA utilized the Grey Wolf Optimizer (80 search agents, 200 iterations, MSE objective), selecting the best global optimum from 30 independent runs. This GWO configuration was determined empirically through preliminary experimentation. Furthermore, the 30 independent runs act as a multi-start strategy to neutralize the stochastic sensitivity inherent to population-based search. While ANFIS-GWO uses standard historical features, DeGNA processes an enriched input space—integrating STL seasonality, residuals, and macroeconomic indicators (IIP, FDI)—to effectively capture complex non-linear variations. Additionally, Section 5.4 provides a dedicated sensitivity analysis quantifying the robustness of DeGNA's performance to these specific parameter choices.

4.3 Evaluation Framework and Statistical Tests

To fully test the stability and effectiveness of the proposed DeGNA framework, a strict testing process was established, including data validation strategies, stationarity checks, standard error metrics, and non-parametric statistical tests.

Data Splitting and Cross-Validation Strategy. To evaluate the models, we kept the natural order of the data by splitting it into 80% for training and 20% for testing. The first 80% of the data was used to train the models and find the best parameters, while the last 20% was kept for testing. Additionally, to ensure the models are stable over time, we used a Time-Series Walk-Forward Cross-Validation. This method used an "expanding window" with 5 folds. In each step, we moved the testing period forward by 12 months.

Stationarity Tests. To ensure reliable forecasting, verifying time-series stationarity is essential. This study employs a dual-validation framework utilizing the Augmented Dickey-Fuller (ADF) [33,34] and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [35] tests. The ADF test evaluates the presence of a unit root (stationarity confirmed if $p < 0.05$), while the KPSS test assumes stationarity as its null hypothesis (stability confirmed if $p > 0.05$). Consequently, a series is considered strictly stationary only if it satisfies both criteria. This cross-verification dictates whether further differencing is necessary prior to modeling.

Forecasting Evaluation Metrics. To validate the predictive accuracy of the models, four metrics were calculated: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2) [36,37]. They are mathematically defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{25}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{26}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \tag{27}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{28}$$

where n is the number of predictions, y_i represents the actual observed value, $\widehat{y}_i$ is the forecasted value, and $\overline{y}$ denotes the mean of the actual values.

Statistical Significance. Standard error metrics alone cannot determine whether performance disparities stem from true algorithmic superiority or random sample variance. To rigorously validate the proposed framework, we employed the Diebold–Mariano (DM) test [38], incorporating the Harvey et al. (1997) [39] modification to correct for small-sample biases in finite time series. If the resulting p-value is below 0.05, the difference in performance is considered statistically significant.

5 Empirical Results & Analysis

5.1 Data Preprocessing & Stationarity Analysis

Temporal Disaggregation Analysis. To resolve the frequency mismatch between annual energy consumption data and monthly indicators, this study employed the Denton proportional temporal disaggregation method. Monthly commercial electricity output was selected as the reference indicator, given its near-perfect linear correlation with primary energy at the annual level ($r = 0.9977$). However, it should be noted that residual uncertainty remains, particularly at seasonal turning points where the reference indicator may not perfectly capture total primary energy dynamics. This is an inherent limitation of any disaggregation approach in the absence of directly observed monthly data.

The requirement of this process is the preservation of the original low-frequency data. Notably, the computational results confirm that the Denton method perfectly maintains the annual totals, with a maximum error of 2.27×10^{-13}. This clearly ensures no data loss during transformation, providing a highly reliable high-frequency dataset for the next forecasting phase. It should be noted, however, that the resulting monthly series constitutes synthetic (pseudo-monthly) data derived from annual totals, not directly observed records. Downstream forecasting results should therefore be interpreted with this disaggregation uncertainty in mind. As illustrated in Fig. 3, the Denton method preserves within-year seasonal patterns more faithfully than linear interpolation.

To verify the reliability of the high-frequency data, we compared the Denton method against a standard linear approach. Table 3 shows the detailed quantitative results of this comparison.

As presented in Table 3, both approaches perfectly preserve the annual totals ($error = 0.000$). While linear interpolation shows a smaller maximum year-boundary jump (4.4239 vs. 25.9040), this smoothness completely ignores the reference indicator, resulting in a lower pattern fidelity ($r = 0.9741$) and failing to reflect actual consumption dynamics. In contrast, the Denton method is directly designed to balance temporal distribution with structural preservation. It minimizes the Sum of Squared Differences (SSD) to 0.0413, outperforming the baseline by approximately 677. Notably, the Denton-derived series achieves a near-perfect Pearson correlation ($r = 0.9974$) with the reference indicator. This confirms that the Denton method successfully captures the seasonal variations, which logically explains its larger year-boundary jumps.

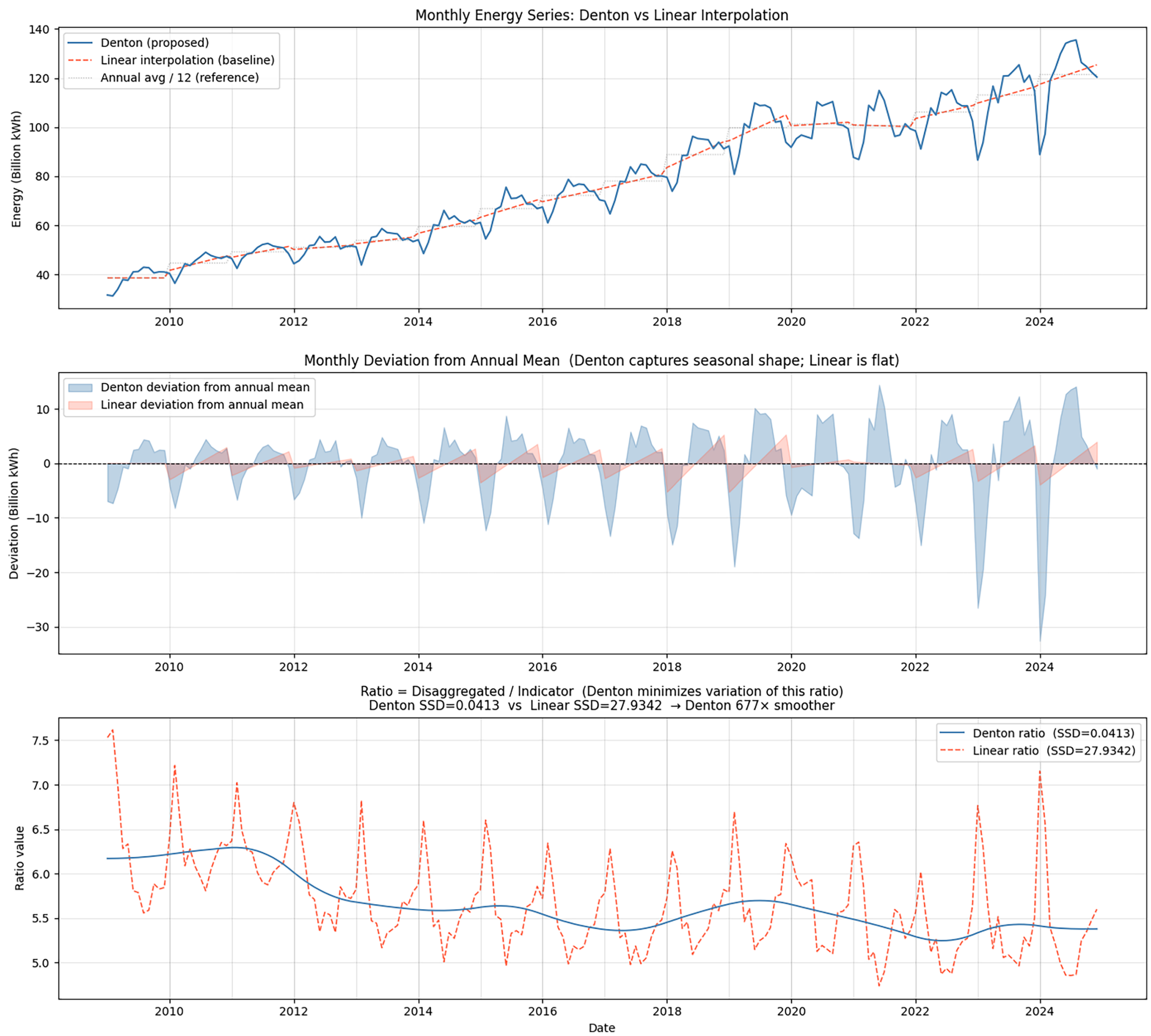

Figure 3: A visual comparison of temporal disaggregation methods. The subplots display the within-year trends generated by the Denton method and standard linear interpolation, clearly demonstrating the Denton method's advantage in preserving seasonal patterns and maintaining structural smoothness.

Table 3: Quantitative comparison of temporal disaggregation methods.

Evaluation Metric	Denton Method	Linear Interpolation
Annual Constraint Error	0.0000	0.0000
SSD of Ratio Diff	0.0413	27.9342
Pattern Fidelity	0.9974	0.9741
Max Year-Boundary Jump	25.9040	4.4239

Stationarity Assessment. Following the successful temporal disaggregation of the datasets, the next phase is to verify the stationarity of the generated high-frequency series. As outlined in Section 4.3, a dual-testing approach using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests was conducted. As detailed in Table 4, the evaluation shows that the monthly energy consumption data are non-stationary. The data get an ADF p-value of 0.9914 and a KPSS p-value of 0.0100. This non-stationarity is visually evident in the data's clear upward trend and the slow decay of its Autocorrelation Function (ACF) plot (Fig. 4). To stabilize the series, first-order differencing ($d = 1$) was applied. Subsequent tests confirmed that the transformed data (dEnergy) successfully achieved stationarity, as the ADF p-value dropped to 0.0000 and the KPSS p-value rose to 0.1000.

Table 4: Unit root and stationarity test results for energy demand.

Variable	Test	Raw Level Statistic (p-Value)	Differenced Statistic (p-Value)
Energy Demand	*ADF*	0.9914	0.0000
	KPSS	0.0100	0.1000

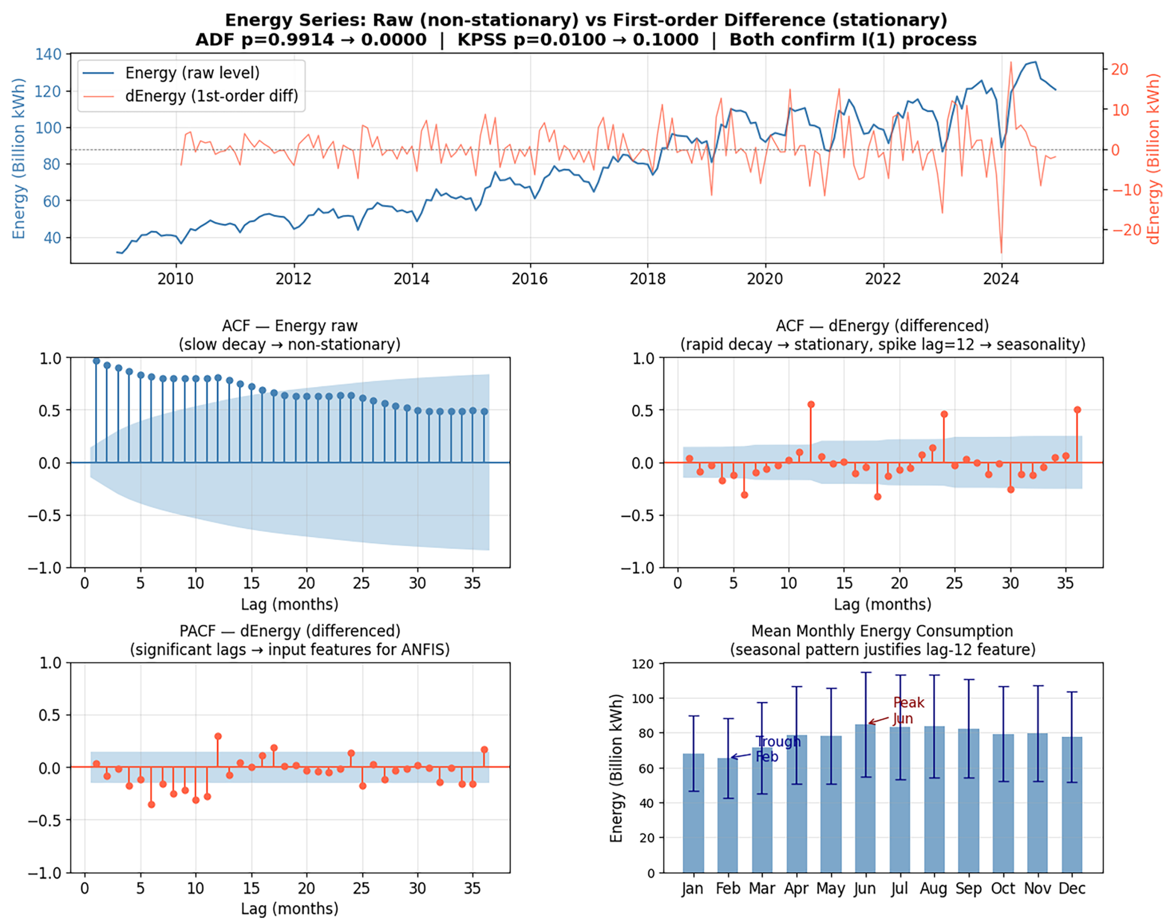

Figure 4: Stationarity transformation and exploratory data analysis of the monthly energy demand.

Beyond confirming stationarity, the Exploratory Data Analysis (EDA) uncovered key structural patterns in the data. While the auto-correlation function (ACF) of the differenced series decays rapidly, it shows a significant spike at lag 12, clearly indicating a strong annual cycle. This seasonality is further supported by the monthly consumption profile, which peaks in June and reaches a trough in February. These empirical findings strongly justify the inclusion of seasonal features, specifically 12-month lags, in the forecasting models.

5.2 *Forecasting Performance Comparison*

To provide a comprehensive evaluation for the DeGNA framework, we used the stationary, high-frequency monthly datasets derived from the previous step. The model's performance on the test set was compared against four benchmarks: traditional statistical methods (ARIMA, SARIMAX), the classic grey forecasting model (GM(1,1)), machine learning models like XGBoost, deep learning architectures (LSTM, BiLSTM), and existing neuro-fuzzy baselines (Standard ANFIS, ANFIS-GWO). Table 5 presents the forecasting performance of all models on the held-out test set.

Table 5: Forecasting performance comparison of the proposed DeGNA framework against baseline models on the test set.

	Model	**RMSE**	**MAE**	**MAPE**	R^2	**Training Time (s)**
	GM(1,1)	14.7308	12.9028	11.6312	−0.3219	0.040
ANFIS Variants	ANFIS_Std	8.5706	6.4498	5.8797	0.5525	24.282
	ANFIS_GWO	6.4590	5.2143	4.7928	0.7459	177.586
Statistical Models	ARIMA	8.4778	6.2375	5.8496	0.5622	0.520
	SARIMAX	5.6947	4.0175	3.7049	0.8024	200.110
ML and DL Models	XGBoost	7.4868	4.9202	4.6715	0.6585	0.194
	LSTM	7.1728	5.1385	4.8603	0.6866	12.847
	BiLSTM	7.2877	5.3193	5.0242	0.6765	12.000
Proposed Model	DeGNA	5.7620	4.5071	4.1052	0.7977	168.790

Compared with the deep learning baselines, DeGNA achieves consistently lower forecasting errors. Specifically, it reduces the RMSE to 5.76 and the MAPE to 4.11%, achieving lower point-forecasting errors on this test set than LSTM (RMSE: 7.17; MAPE: 4.86%) and BiLSTM (RMSE: 7.28; MAPE: 5.02%), and the advanced tree-based XGBoost model (RMSE: 7.49; MAPE: 4.67%). These results suggest that the hybrid grey-neuro-fuzzy architecture remains effective under limited-data conditions. The visual fit of all models on the test set is shown in Fig. 5.

Furthermore, an internal evaluation of the neuro-fuzzy models confirms the efficiency of the proposed architecture. First, replacing standard gradient-based training with the GWO meta-heuristic yields substantial improvements: specifically, ANFIS-GWO achieves a 24.6% reduction in RMSE (from 8.57 to 6.46) and a decrease in MAPE from 5.88% to 4.79%. These results indicate that GWO effectively addresses the non-convex optimization problem encountered during ANFIS training. Finally, the proposed architecture—DeGNA performs even better than ANFIS-GWO. Specifically, it reduces the RMSE by 10.8% (from 6.46 to 5.76) and the MAPE by 14.2% (from 4.79% to 4.11%). This finding suggests that decomposing the original series into trend, seasonal, and residual components improves forecasting accuracy beyond optimization alone.

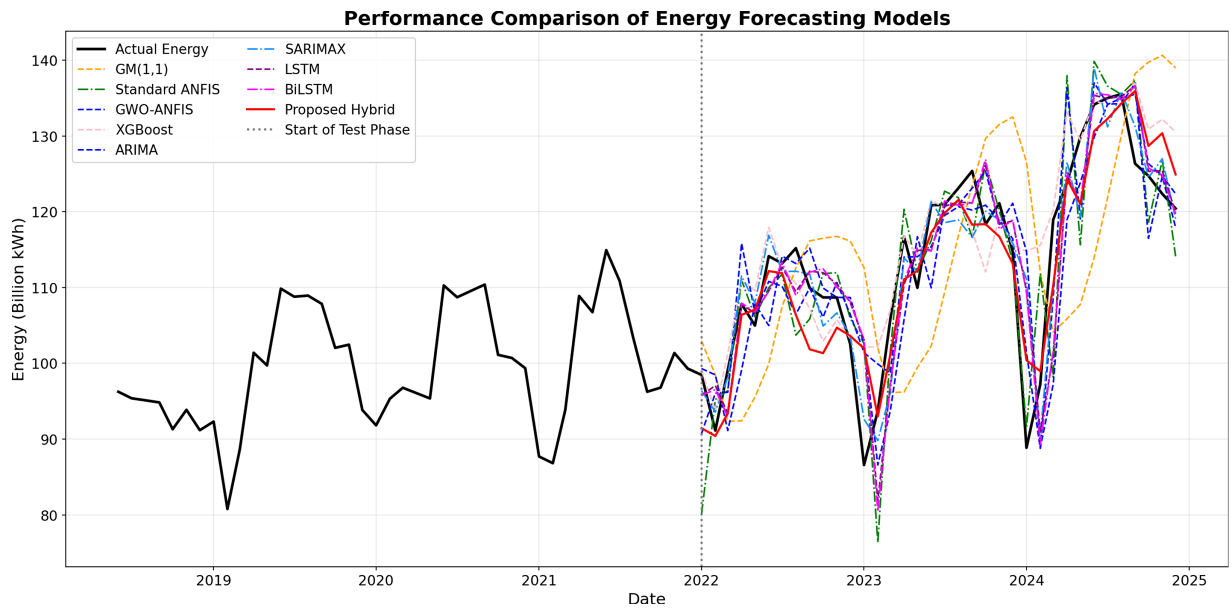

Figure 5: Predicted vs. actual monthly energy demand on the test set, comparing the visual fit of the proposed DeGNA framework against benchmark statistical and deep learning models.

Regarding computational complexity, Table 5 also reports the execution time for the training phase. The proposed DeGNA framework requires approximately 168.79 s to complete its rigorous global optimization process. While this computational overhead is higher than that of deep learning baselines (e.g., LSTM at 12.85 s) or shallow tree models (XGBoost at 0.19 s), this represents an acceptable trade-off. The additional training time is a necessary investment in global parameter optimization, allowing DeGNA to achieve higher point-forecasting accuracy and greater structural stability than these computationally lighter deep learning models. Furthermore, the framework remains highly tractable; in fact, DeGNA executes faster than the automated stepwise parameter search required for SARIMAX (200.11 s). Given that national energy demand forecasting operates on a monthly planning horizon, a training phase of under three minutes demonstrates practical for monthly energy planning applications.

Although SARIMAX achieves a slightly lower average forecasting error (MAPE of 3.70% compared to 4.11% for DeGNA), its performance deteriorates under extreme demand conditions. As shown in Table 6 and Fig. 6, SARIMAX exhibits substantially larger prediction errors during these periods, with a maximum error of 22.23, which is 44.5% higher than DeGNA's worst-case error of 15.39. In addition, DeGNA yields a lower error standard deviation (3.59 vs. 4.04) and produces no extreme outliers. These findings indicate that while SARIMAX is slightly more accurate on average under normal conditions, DeGNA provides more consistent forecasts under highly volatile demand conditions.

Table 6: Error distribution comparison.

Metric	SARIMAX	DeGNA
Max Absolute Error	22.2339	15.3872
95th Percentile Error	9.8758	10.0685
Std of Absolute Errors	4.0361	3.5899

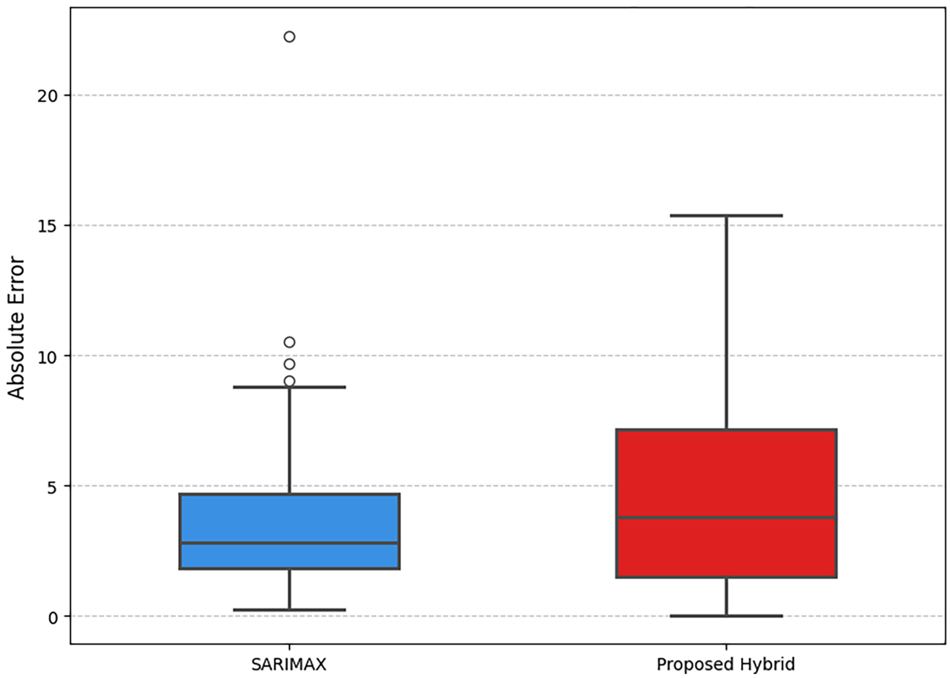

Figure 6: Boxplot of absolute forecasting errors on the test set.

5.3 Robustness and Significance Analysis

To ensure that DeGNA's results in Section 5.2 are not just due to luck, we performed further tests. This validation includes two parts: first, checking whether the differences between models are statistically significant, and second, testing the model's stability over different time periods using cross-validation.

Table 7 summarizes the results of the Diebold–Mariano (DM) test with the small-sample correction and the Wilcoxon signed-rank test. Notably, DeGNA demonstrates a clear statistical advantage over foundational benchmarks, confirming high significance ($p < 0.01$) against GM(1,1). Furthermore, the framework remains reliable when compared to Standard ANFIS and ARIMA. With p-values generally under 0.10–and strictly p less than 0.05 in the DM test for ANFIS and the Wilcoxon test for ARIMA.

Table 7: Statistical significance tests against the proposed DeGNA framework.

	Diebold–Mariano Test			**Wilcoxon Signed-Rank Test**		
	DM stat	***p*-Value**	**Result**	**W stat**	***p*-Value**	**Result**
GM (1,1)	−4.657	0.000	***	33.0	0.000	***
ANFIS_Std	−2.327	0.026	**	231.0	0.056	*
ANFIS_GWO	−1.320	0.195	ns	269.0	0.161	ns
ARIMA	−1.988	0.055	*	226.0	0.047	**
SARIMAX	+0.061	0.952	ns	384.0	0.788	ns
XGBoost	−1.238	0.224	ns	343.0	0.565	ns
LSTM	−1.524	0.137	ns	311.0	0.369	ns
BiLSTM	−1.668	0.104	ns	296.0	0.285	ns

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$, ns = not significant. DM test applies the Harvey–Leybourne–Newbold (1997) small-sample correction.

In contrast, the statistical tests indicate no significant differences in average point-forecasting accuracy between DeGNA and advanced data-driven baselines like XGBoost, LSTM, and BiLSTM. Consequently, we acknowledge that DeGNA performs comparably to these models on this specific test split. Nevertheless, achieving this comparable accuracy remains a substantial advantage, as it proves the proposed framework can deliver highly competitive results without demanding massive training datasets, GPU acceleration, or suffering from black-box opacity. Furthermore, while their average errors on the test set are statistically similar, DeGNA possesses a distinct structural advantage in mitigating temporal variance, which will be demonstrated in the subsequent cross-validation analysis.

Finally, the comparison with SARIMAX highlights a crucial trade-off between average accuracy and structural stability. It is important to note that while DeGNA yields a slightly higher average MAPE than SARIMAX (4.11% vs. 3.70%), the Diebold–Mariano test confirms this difference is not statistically significant (DM stat = +0.061, p = 0.952). The preference for the DeGNA framework is therefore justified not by marginal gains in average accuracy, but by its capacity to mitigate extreme forecasting failures. Specifically, SARIMAX's maximum absolute error reaches 22.23, which is 44.5% higher than DeGNA's worst-case error of 15.39. Furthermore, DeGNA maintains a lower error standard deviation (3.59 vs. 4.04), indicating superior temporal stability across structurally disrupted periods. By avoiding these extreme deviations, DeGNA proves to be significantly more robust when the underlying data distribution shifts.

Cross-validation results summarized in Table 8, deep learning models (LSTM and BiLSTM) yield the highest mean errors alongside the largest fold-to-fold variance (MAPE > 8.0 ± 3.0%), which is consistent with their well-known sensitivity to limited training data. While XGBoost and SARIMAX achieve lower mean MAPE (6.06% and 6.51%, respectively), both exhibit notable instability across folds (standard deviations of ±1.71% and ±2.81%). In contrast, DeGNA records the lowest standard deviation (±1.09%), indicating the most consistent forecasting performance across different validation periods.

Table 8: Summary of walk-forward cross-validation performance (Mean ± Std) across 5 sequential folds.

Model	RMSE	MAE	MAPE
ARIMA	11.848 ± 3.057	10.375 ± 3.239	9.372 ± 2.474
SARIMAX	7.868 ± 2.783	6.638 ± 2.629	6.511 ± 2.809
XGBoost	7.711 ± 2.300	6.624 ± 2.269	6.062 ± 1.714
LSTM	10.213 ± 3.859	8.726 ± 3.667	8.472 ± 3.751
BiLSTM	10.502 ± 3.519	9.233 ± 3.466	8.900 ± 3.613
DeGNA (Proposed)	9.114 ± 1.600	7.681 ± 1.713	7.043 ± 1.089

The per-fold results in Table 9 provide additional insight into the source of this stability. During Folds 1–3 (2020–2022)—a period marked by the COVID-19 demand shock and subsequent recovery—DeGNA consistently outperforms SARIMAX (average MAPE: 6.26% vs. 7.99%), as its STL-based decomposition effectively separates structural disruptions from the underlying trend. Conversely, in Folds 4–5 (2023–2024), a period characterized by rapid and near-monotonic demand growth driven by post-pandemic manufacturing expansion and rising FDI inflows, SARIMAX's auto-regressive structure adapts more efficiently to the accelerating regime (average MAPE: 4.31% vs. 8.21%). This pattern indicates that DeGNA's decomposition architecture is particularly well-suited to structurally volatile periods, while purely statistical models retain an advantage under smooth, trend-dominant conditions. For national energy planning in an emerging economy facing recurring external shocks, DeGNA's stable performance may be more operationally reliable than models with lower average errors but larger worst-case failures.

Table 9: Per-fold MAPE (%) across the walk-forward cross-validation window. Bold values indicate the best-performing model for each fold.

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
ARIMA	7.34	7.08	7.70	12.86	11.87
SARIMAX	8.07	11.35	4.54	**4.38**	**4.23**
XGBoost	**4.82**	**4.49**	4.69	8.10	8.22
LSTM	14.56	10.81	**4.00**	6.47	6.52
BiLSTM	14.95	10.80	4.68	7.38	6.69
DeGNA (Proposed)	5.56	6.65	6.58	8.77	7.65

5.4 Sensitivity Analysis

To assess the robustness of DeGNA with respect to its key hyper-parameters, a one-at-a-time sensitivity analysis was performed using the same 5-fold walk-forward cross-validation procedure described in Section 4.3. The analysis focused on the rolling window size, the number of ANFIS rules, and the GWO population and iteration settings. For each experiment, all parameters were fixed at their default values (window size equal 12, number of rules equal 4, number of wolves equal 80, and maximum iterations equal 200), while only one parameter was varied within a reasonable range. The resulting test MAPE and its standard deviation were then averaged across the five validation folds.

As shown in Table 10 and Fig. 7, the mean MAPE remained relatively stable across all tested parameter ranges, varying by only 0.13–0.97 percentage points. This variation was smaller than the cross-fold standard deviation (1.09–1.64 percentage points), and the uncertainty ranges largely overlapped across all parameter settings.

Table 10: Summary of sensitivity analysis results for key empirical hyper-parameters.

Parameter	Tested Values	MAPE (%) Mean ± Std
Rolling window size	6	7.00 ± 1.12
	9	7.01 ± 1.10
	12 (used)	7.04 ± 1.09
	15	7.08 ± 1.08
	18	7.12 ± 1.09
Number of ANFIS rules	2	6.37 ± 1.24
	3	6.45 ± 1.25
	4 (used)	7.04 ± 1.09
	5	6.44 ± 1.22
	6	6.74 ± 1.24
GWO population size (n_wolves)	20	6.07 ± 1.64
	40	6.48 ± 1.41
	60	6.18 ± 1.10
	80 (used)	7.04 ± 1.09
	100	6.94 ± 1.57

(Continued)

Table 10 (continued)

Parameter	Tested Values	MAPE (%) Mean ± Std
GWO max iterations	50	6.86 ± 1.44
	100	7.20 ± 1.10
	150	6.58 ± 1.26
	200 (used)	7.04 ± 1.09
	300	6.79 ± 1.36

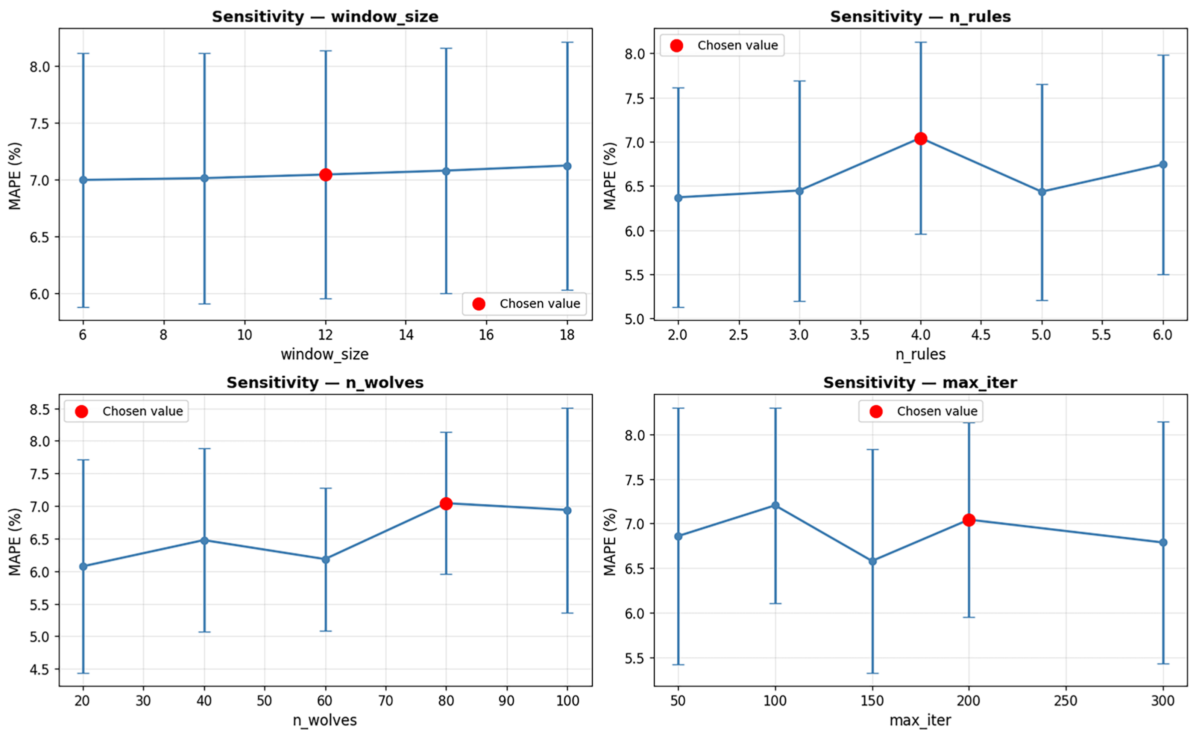

Figure 7: Sensitivity analysis of key empirical hyperparameters. The solid blue line represents the mean MAPE across 5 folds, while the error bars indicate the standard deviation. The red dot highlights the final hyper-parameter value selected for the DeGNA framework.

Among the tested parameters, the rolling window size exhibited the lowest sensitivity, with the mean MAPE ranging from 7.00% to 7.12% for window sizes between 6 and 18 months. The selected value of 12 months is also consistent with the annual seasonal cycle identified in the earlier time-series analysis (Fig. 4). The number of fuzzy rules showed similarly stable behavior, with the mean MAPE varying between 6.37% and 7.04% for configurations ranging from 2 to 6 rules. The selected configuration of four rules provides a reasonable balance between model complexity and the available sample size ($n \approx 192$ monthly observations).

For the GWO parameters, increasing the population size from 20 to 80 wolves reduced the cross-fold standard deviation from ±1.64% to ±1.09%, indicating more consistent optimization performance across folds. Further increases beyond 80 wolves provided limited additional improvement, suggesting that this setting offers a practical trade-off between stability and computational cost. Similarly, the maximum number of iterations showed only modest variation in mean MAPE (6.58%–7.20%), indicating that 200 iterations are sufficient to achieve stable optimization performance.

Overall, the sensitivity analysis suggests that DeGNA's forecasting performance is relatively robust to moderate changes in its hyper-parameter settings, supporting the reliability of the reported configuration.

6 Conclusion

Forecasting national energy demand under data-scarce conditions presents a critical challenge for developing economies. This study addresses the data limitation by developing DeGNA, a hybrid forecasting framework for long-term macroeconomic prediction. By combining STL, GM(1,1), GWO-ANFIS, and key economic indicators (IIP and FDI), DeGNA successfully captures complex consumption patterns without requiring extensive historical datasets.

Empirical evaluations confirm the robustness of the proposed framework. On the test set, DeGNA achieved a MAPE of 4.11%, statistically outperforming foundational models such as GM(1,1) ($p < 0.01$) and ARIMA ($p < 0.05$). Statistical significance testing demonstrates that DeGNA performs comparably to advanced ML/DL models (XGBoost, LSTM, and BiLSTM). However, evaluation on a single test split does not fully reflect model stability under varying temporal conditions. The 5-fold walk-forward cross-validation results (Tables 8 and 9) provide a complementary assessment of generalization performance. Although SARIMAX and XGBoost obtain lower average MAPEs (6.51% and 6.06%, respectively), they exhibit substantially greater variability across folds (±2.81% and ±1.71%), indicating less stable generalization performance. In contrast, DeGNA demonstrates superior performance precisely during structurally disrupted periods, where its STL decomposition isolates demand shocks from the underlying trend—a scenario frequently encountered in emerging economies.

Under stable, trend-dominant conditions, purely autoregressive models such as SARIMAX may retain a marginal average accuracy advantage. However, for emerging economies subject to recurring structural shocks, DeGNA's stability evidenced by the lowest cross-validation variance (±1.09%) offers greater operational reliability. In real-world energy management, especially in Vietnam, forecasting stability may be more valuable than marginal improvements in average accuracy because large prediction errors can disproportionately affect long-term planning decisions. By minimizing extreme forecasting failures, DeGNA provides policymakers with a dependable tool to optimize grid investments, mitigate supply shortage risks, and safeguard national energy security. Although the proposed framework achieved strong predictive performance, several limitations should be acknowledged. First, the monthly observations were generated using the Denton interpolation method rather than directly measured, resulting in a pseudo-monthly dataset. As a result, errors introduced during disaggregation may also affect the subsequent forecasting results. Although this risk was mitigated by anchoring the disaggregation to a highly correlated reference indicator ($r = 0.9974$) and by employing strict out-of-sample validation, these limitations should still be acknowledged. Secondly, the temporal scope is limited to a relatively short data period (2009–2024), which may restrict the models' exposure to long-term multi-decadal economic cycles. Lastly, this research serves as a single-country case study focused exclusively on Vietnam. While it is highly representative of rapidly growing emerging economies, the specific macroeconomic dynamics (IIP, FDI) may not perfectly generalize to developed nations with different energy consumption patterns.

As official high-frequency macroeconomic datasets and longer historical time series become available, future studies could further validate and refine this framework using those ground-truth records. Furthermore, future research should expand the application of the DeGNA framework to a multi-country analysis, comparing its robustness across different developing and developed economic regions. Incorporating variables such as renewable-energy penetration or climate indicators represents a promising extension for improving long-term forecasting accuracy.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The research article was contributed by: Conceptualization, Xuan Kien Pham, Van Thanh Phan, Duc Trien Nguyen, and Van Dat Nguyen; methodology, Van Thanh Phan and Duc Trien Nguyen; software, Van Thanh Phan, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; validation, Van Thanh Phan, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; formal analysis, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; investigation, Van Thanh Phan and Duc Trien Nguyen; resources, Van Thanh Phan and Duc Trien Nguyen; data curation, Duc Trien Nguyen, Xuan Kien Pham, and Van Dat Nguyen; writing—original draft preparation, Van Thanh Phan, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; writing—review and editing, Van Thanh Phan, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; visualization, Van Thanh Phan, Duc Trien Nguyen, Van Dat Nguyen, and Xuan Kien Pham; supervision, Van Thanh Phan; project administration, Van Thanh Phan, Van Dat Nguyen, and Xuan Kien Pham. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets supporting the findings of this study are publicly available on Zenodo at https://doi.org/10.5281/zenodo.21028899. Raw data were originally sourced from Our World in Data [24], MOIT [25], GSO [26], and MPI [27].

Ethics Approval: Not applicable.

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

References

1. Nguyen MP, Ponomarenko T, Nguyen N. Energy transition in Vietnam: a strategic analysis and forecast. Sustainability. 2024;16(5):1969. doi:10.3390/su16051969.
2. Prime Minister of Vietnam. Decision No. 500/QD-TTg: approving the national power development plan for 2021–2030, with a vision to 2050. Hanoi, Vietnam: Government of Vietnam; 2023 [cited 2026 Jan 1]. Available from: https://vanban.chinhphu.vn/?pageid=27160&docid=207889.
3. Kontopoulou VI, Panagopoulos AD, Kakkos I, Matsopoulos GK. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Fut Internet. 2023;15(8):255. doi:10.3390/fi15080255.
4. Steinbuks J. Assessing the accuracy of electricity production forecasts in developing countries. Int J Forecast. 2019;35(3):1175–85. doi:10.1016/j.ijforecast.2019.04.009.
5. Bhattacharyya SC, Timilsina GR. Modelling energy demand of developing countries: are the specific features adequately captured? Energy Policy. 2010;38(4):1979–90. doi:10.1016/j.enpol.2009.11.079.
6. Denton FT. Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. J Am Stat Assoc. 1971;66(333):99–102. doi:10.1080/01621459.1971.10482227.
7. Dong Q, Huang R, Cui C, Towey D, Zhou L, Tian J, et al. Short-term electricity-load forecasting by deep learning: a comprehensive survey. Eng Appl Artif Intell. 2025;154(7):110980. doi:10.1016/j.engappai.2025.110980.
8. Zhang X, Li R. A novel decomposition and combination technique for forecasting monthly electricity consumption. Front Energy Res. 2021;9:792358. doi:10.3389/fenrg.2021.792358.
9. Acakpovi A, Ternor AT, Asabere NY, Adjei P, Iddrisu AS. Time series prediction of electricity demand using adaptive neuro-fuzzy inference systems. Math Probl Eng. 2020;2020(2):4181045. doi:10.1155/2020/4181045.
10. Wang H, Chen B, Sun H, Li A, Zhou C. AnFiS-MoH: systematic exploration of hybrid ANFIS frameworks *via* metaheuristic optimization hybridization with evolutionary and swarm-based algorithms. Appl Soft Comput. 2024;167(2):112334. doi:10.1016/j.asoc.2024.112334.

11. Athanasopoulou ME, Deveikyte J, Mosca A, Peri I, Provetti A. A hybrid model for forecasting short-term electricity demand. In: Proceedings of the Second ACM International Conference on AI in Finance; 2021 Nov 3–5; Virtual. p. 1–6. doi:10.1145/3490354.3494371.
12. Tang T, Jiang W, Zhang H, Nie J, Xiong Z, Wu X, et al. GM(1,1) based improved seasonal index model for monthly electricity consumption forecasting. Energy. 2022;252(2):124041. doi:10.1016/j.energy.2022.124041.
13. Hao Y, Ma X. A hybrid grey system model based on stacked long short-term memory layers and its application in energy consumption forecasting. Processes. 2024;12(8):1749. doi:10.3390/pr12081749.
14. Wu C, Li J, Liu W, He Y, Nourmohammadi S. Short-term electricity demand forecasting using a hybrid ANFIS–ELM network optimised by an improved parasitism–predation algorithm. Appl Energy. 2023;345(1):121316. doi:10.1016/j.apenergy.2023.121316.
15. Bakare MS, Abdulkarim A, Shuaibu AN, Muhamad MM. A hybrid long-term industrial electrical load forecasting model using optimized ANFIS with gene expression programming. Energy Rep. 2024;11(1):5831–44. doi:10.1016/j.egyr.2024.05.045.
16. Sax C, Steiner P. Temporal disaggregation of time series. R J. 2013;5(2):80–7. doi:10.32614/rj-2013-028.
17. Cleveland R. STL: a seasonal-trend decomposition procedure based on loess. 1990 [cited 2026 Apr 7]. Available from: https://www.semanticscholar.org/paper/STL-%3A-A-Seasonal-Trend-Decomposition-Procedure-on-Cleveland/585bf445ec84c1d9621b2726bdcce9f544b515c8.
18. Deng JL. Control problems of grey systems. Syst Control Lett. 1982;1(5):288–94. doi:10.1016/S0167-6911(82)80025-X.
19. Deng JL. Introduction to grey system theory. J Grey Syst. 1989;1(1):1–24.
20. Liu L, Wang Q, Liu M, Li L. An intelligence optimized rolling grey forecasting model fitting to small economic dataset. Abstr Appl Anal. 2014;2014(4):641514. doi:10.1155/2014/641514.
21. Jang JSR. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern. 1993;23(3):665–85. doi:10.1109/21.256541.
22. Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern. 1985;SMC-15(1):116–32. doi:10.1109/TSMC.1985.6313399.
23. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw. 2014;69:46–61. doi:10.1016/j.advengsoft.2013.12.007.
24. Vietnam Country Profile. Energy. Our world in data. [cited 2026 Apr 3]. Available from: https://ourworldindata.org/profile/energy/vietnam.
25. Ministry of Industry and Trade of Vietnam (MOIT). Report on power production and business activities. 2024 [cited 2026 Jan 1]. Available from: https://moit.gov.vn/thong-ke/bao-cao-tong-hop.
26. General Statistics Office of Vietnam. Statistical data: index of industrial production (IIP). 2024 [cited 2026 Jan 1]. Available from: https://www.nso.gov.vn/so-lieu-thong-ke/.
27. Ministry of Planning and Investment (MPI). Situation of foreign direct investment (FDI) in Vietnam. 2024 [cited 2026 Jan 1]. Available from: https://www.mpi.gov.vn/portal/Pages/solieudautunuocngoai.aspx.
28. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2016.
29. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19(6):716–23. doi:10.1109/TAC.1974.1100705.
30. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
31. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process. 1997;45(11):2673–81. doi:10.1109/78.650093.
32. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13–17; San Francisco, CA, USA. p. 785–94. doi:10.1145/2939672.2939785.
33. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc. 1979;74(366):427. doi:10.2307/2286348.

34. Said SE, Dickey DA. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika. 1984;71(3):599–607. doi:10.1093/biomet/71.3.599.
35. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root How sure are we that economic time series have a unit root? J Econ. 1992;54(1–3):159–78. doi:10.1016/0304-4076(92)90104-Y.
36. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22(4):679–88. doi:10.1016/j.ijforecast.2006.03.001.
37. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021;7(3):e623. doi:10.7717/peerj-cs.623.
38. Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat. 1995;13(3):253–63. doi:10.1198/073500102753410444.
39. Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. Int J Forecast. 1997;13(2):281–91. doi:10.1016/S0169-2070(96)00719-4.