Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index
DOI: https://doi.org/10.24127/sciencestatistics.v4i1.XXXX
Vol. 4, No. 2, 2026, 107-120
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

Empirical Mode Decomposition with Swarm-Optimized Support Vector Regression for Natural Gas Price Forecasting

Fajar Putrawan Djabar[1*], Agusyarif Rezka Nuha[2], Siti Nurmardia Abdussamad[3]
[1*,2,3]Department of Statistics, Universitas Negeri Gorontalo, Gorontalo, Indonesia
*corresponding author: fajardjabar179@gmail.com

Article Info

Article History:
Received: 27-06-2026
Revised: 14-07-2026
Accepted: 27-07-2026
Available online: 29-07-2026

Keywords:
Empirical Mode Decomposition; Forecasting; Natural gas; Particle Swarm Optimization; Support Vector Regression

ABSTRACT

Natural gas is a strategic energy commodity exhibiting nonlinear and highly volatile price movements due to supply-demand fluctuations, market dynamics, and geopolitical influences. These factors complicate accurate forecasting and necessitate advanced methods capable of modeling complex data patterns. This study proposes a hybrid forecasting model that integrates Empirical Mode Decomposition (EMD), Support Vector Regression (SVR), and Particle Swarm Optimization (PSO) to predict natural gas prices and assess predictive performance. The analysis utilizes a dataset of 1,575 daily closing prices from January 2020 to December 2025. EMD decomposes the original time series into seven Intrinsic Mode Functions (IMFs) and one residual component. Each component is modeled using SVR with a Radial Basis Function (RBF) kernel, and PSO is used to optimize model parameters. Forecasting performance is evaluated using Mean Absolute Percentage Error (MAPE) across three data partitioning schemes. Results indicate that the 70:15:15 partition yields the most accurate model, achieving a MAPE of 2.2641% and MAE 0.0826. The 90-day forecast projects a gradual decline in natural gas prices after a peak in mid-January 2026, followed by relative price stability through March 2026. Therefore, the proposed model has considerable potential to serve as a decision-support tool for policymakers, industry stakeholders, and investors in the energy sector.

INTRODUCTION

Natural gas is a strategically critical global energy commodity, representing approximately 24% of world primary energy consumption and serving as a primary bridge fuel in the global energy transition (Energy Institute, 2025). The natural gas price, represented by the principal North American pricing benchmark in Louisiana, USA, exhibited extraordinary volatility during 2020–2025 from a COVID-19-induced demand through of 1.544 USD/MMBtu on 26 June 2020 to a supply-crisis peak of 9.647 USD/MMBtu on 22 August 2022 (Investing.com, 2026). Such price swings create substantial financial risk for energy producers, utilities, industrial consumers, and policymakers who depend on forward-looking price estimates for production scheduling, procurement hedging, and investment decisions (J. Wang et al., 2020).

The Russia-Ukraine military conflict that escalated in February 2022 constitutes the most disruptive geopolitical shock to global gas markets in decades (Bakrie et al., 2022). Emergency LNG re-routing, record European import costs, and lasting structural change in global gas trade flows followed the disruption of Russian pipeline exports (Zarkasy et al., 2022). Concurrently, the post-COVID economic recovery in Asia led to significant surges in demand, further intensifying supply-side pressures (Zheng et al., 2023). These developments demonstrate that natural gas price dynamics are fundamentally nonlinear and non-stationary, as they encompass both regular seasonal fluctuations and exceptional structural disruptions that classical parametric models cannot capture effectively (Xie et al., 2023).

The limitations of classical forecasting methods have been extensively documented. Ambya et al. (2020) implemented an AR (1)-GARCH (1,1) model for natural gas price prediction, yielding an RMSE of 0.3417; however, the model failed to capture the extreme price surge in 2022 and exhibited persistent heteroskedasticity in its residuals. Similarly, linear regression techniques used by Widagdo (2024) produced an MSE of 2,569.4 on unnormalized data, indicating a fundamental inability to address high-volatility market conditions. The K-Nearest Neighbor (KNN) approach also yielded suboptimal results, with a MAPE of 14.53% even after outlier removal (Qirani & Sukarsih, 2024). Collectively, these results indicate that traditional linear and statistical models are fundamentally constrained in representing the complex, nonlinear dynamics of natural gas price fluctuations, thereby supporting the adoption of more advanced, nonlinear decomposition-based forecasting methods.

Support Vector Regression (SVR) provides a theoretically grounded method for nonlinear forecasting. SVR minimizes structural risk and demonstrates robust generalization, even when training data are limited(Lux et al., 2020). However, SVR exhibits high sensitivity to its hyperparameter triplet (C, γ, ε); suboptimal settings can result in underfitting or overfitting, while exhaustive grid search across broad parameter spaces is computationally intensive (Y. G. Wang et al., 2023). Particle Swarm Optimization (PSO), as introduced by Kennedy and Eberhart, addresses this limitation by efficiently navigating the hyperparameter space through a population-based global search strategy (Gad, 2022). (Rahmawati et al., 2021) demonstrated that PSO-optimized SVR reduced RMSE from 35.83 to 9.40 compared with unoptimized SVR on sales forecasting. This approach aligns with Zheng et al. (2023), who confirmed that metaheuristic optimization of SVR hyperparameters substantially improves natural gas price forecasting accuracy, with PSO representing a computationally efficient alternative to genetic algorithms. To further enhance forecasting accuracy, signal decomposition is applied as a pre-processing step prior to SVR-PSO modelling.

A further performance-enhancing strategy involves pre-modeling signal decomposition. Empirical Mode Decomposition (EMD), as introduced by Huang 1998, adaptively decomposes nonlinear and nonstationary time series into components at distinct frequency scales without the need for predefined basis functions (Awajan et al., 2019). Empirical evidence consistently supports decomposition-integration architectures. For example, Saluza et al. (2023) reported improved stock-price prediction using EMD-feedforward neural networks. Nava et al. (2018)validated the effectiveness of EMD-SVR for financial time series. Xie et al. (2023) validated a secondary decomposition-ensemble approach specifically for natural gas price forecasting using multisource data.

Although previous research has advanced the field, no prior study has integrated empirical mode decomposition (EMD), radial basis function (RBF) kernel support vector regression (SVR), and particle swarm optimization (PSO) within a single framework for daily Henry Hub price forecasting during the 2020-2025 crisis period. The present study addresses this gap by employing EMD to decompose a 1,575-point price series. Each decomposed component is subsequently forecasted using PSO-tuned SVR. Three data-splitting configurations are assessed to determine the optimal train-validation-test ratio, and a validated 90-day forecast for January to March 2026 is generated. These findings aim to facilitate more informed decision-making in natural gas markets during periods of extreme price volatility.

METHODS

Data

Daily Henry Hub natural gas closing prices (USD/MMBtu) were obtained from Investing.com (2026) for the period 1 January 2020 to 31 December 2025, yielding N = 1,575 trading-day observations. The single research variable is the daily closing price (X) on a ratio scale. The observation window deliberately encompasses two major market shocks, the COVID-19 pandemic demand collapse (2020) and the Russia-Ukraine supply disruption providing a demanding out-of-sample evaluation environment.

Nonlinearity Testing

Prior to conducting the Terasvirta neural-network test, a preliminary partial autocorrelation function (PACF) analysis was performed on the original (undecomposed) Henry Hub daily price series. This analysis revealed a dominant and statistically significant autocorrelation at lag 1, indicating that the series' short-term dependence structure is primarily first-order. Accordingly, an autoregressive lag order of p = 1 was adopted for the subsequent nonlinearity test. The Terasvirta neural-network test was applied at significance level $\alpha = 0.05$ to the raw price series. H_0 states linearity; H_1 states nonlinearity. The test is based on a third-order Taylor expansion of a smooth-transition model, yielding an F-statistic with distribution $F_{(m, N-p-m-1)}$, where m is the number of nonlinear terms and p is the autoregressive lag order (Prabowo et al., 2020).

$$F = \frac{(SSR_0 - SSR_1)/m}{(1 - SSR_1)/(N - p - 1 - m)} \tag{1}$$

Rejection of H0 ($p < 0.05$) confirms the presence of nonlinear structure in the series, thereby providing empirical justification for the adoption of EMD decomposition and SVR as the core modeling strategy in this study.

Min-Max Normalization

Prior to decomposition, each observation x is normalized to the unit interval [0, 1]:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x_{min} and x_{max} are the global minimum and maximum of the original price series. Normalization prevents magnitude bias in the SVR kernel distance calculation (Ambarwari et al., 2020). Standardizing the input scale also provides a more stable fitness landscape for PSO hyperparameter search (Difitria, 2020).

EMD Decomposition

The normalized series x(t) is decomposed via the sifting algorithm into Intrinsic Mode Functions and one monotonic residual (Awajan et al., 2019):

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r(t) \tag{3}$$

The upper $e_u(t)$ and lower $e_l(t)$ envelopes are constructed using cubic-spline interpolation of the local maxima and minima of x(t), respectively. The mean envelope m(t) = [e_u(t) + e_l(t)]/2 is iteratively subtracted from the signal to produce a candidate component $h_k(t)$at each sifting iteration k. The sifting process is terminated using a Cauchy-type convergence criterion, whereby iteration stops once the standard deviation (SD) between two consecutive sifting outputs, $h_{k-1}(t)$and $h_k(t)$, falls below a threshold value of 0.2:

$$SD_k = \sum_{t=0}^{T} \frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}(t)^2}. \tag{4}$$

This process results in one intrinsic mode function (IMF), and the sifting procedure continues until the residual r(t) becomes monotonic. No constraint was imposed on the maximum number of intrinsic mode functions (IMFs) to be extracted. As a result, the algorithm automatically determined the number of components based on the point at which the residual satisfied the monotonicity condition. This approach reflects the intrinsic multi-scale oscillatory structure of the Henry Hub natural gas price series. An IMF must satisfy two conditions: (a) the number of extrema and zero-crossings differs by at most one throughout the entire data length; and (b) the mean of the upper and lower envelopes is approximately zero at every point, which also serves as the stopping criterion for the sifting process. The data-adaptive and basis-free characteristics of empirical mode decomposition (EMD) make it particularly suitable for nonstationary energy price series, where fixed wavelet bases suboptimal for series with time-varying frequency content (Awajan et al., 2019).

PACF-Based Lag Selection

For each IMF and the residual component, statistically significant time lags were identified using the Partial Autocorrelation Function (PACF) at a significance level of α = 0.05 (95% confidence interval). A lag was considered statistically significant if the absolute value of its PACF coefficient exceeded the large-sample confidence bound, calculated as $1.96/\sqrt{N}$, where N denotes the length of the corresponding IMF (Hyndman & Athanasopoulos, 2021) Significant lag values identified in the PACF plot were selected as input variables, capturing relevant historical dependencies while avoiding unnecessary model complexity (Ding & Zhou, 2025). The identified lags served as input features for the SVR model applied to each respective IMF component.

Data Splitting

Three train, validation, test configurations were systematically evaluated 70:15:15, 80:10:10, and 90:5:5. The training set fits the SVR model; the validation set provides fitness scores for PSO hyperparameter search; and the test set remains strictly unseen until final accuracy reporting, preventing data leakage.

SVR with RBF Kernel and PSO Optimization

An SVR model with the RBF kernel is fitted independently for each component. Following the standard Vapnik formulation, SVR minimizes the regularized ε-insensitive loss in which the term $½\|w\|^2$ is retained to yield a convex primal objective whose Lagrangian dual is derived symmetrically via the Karush–Kuhn–Tucker conditions (Saputra et al., 2019):

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{5}$$

where C is the regularization parameter trading off margin width against training error, ε defines the width of the insensitive tube, and ξ_i, ξ_i^* are slack variables for observations outside the tube. The dual solution produces the regression function:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\, k(x_i, x) + b \tag{6}$$

where α_i, α_i^* are Lagrange multipliers and b is the bias term. The RBF kernel is:

$$K(x_i, x_j) = \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right). \tag{7}$$

The search bounds for the support vector regression (SVR) hyperparameters (C, γ, ε) were established through preliminary experiments that compared several parameter ranges reported in previous SVR–particle swarm optimization (PSO) forecasting studies. Of the configurations evaluated, the bounds proposed by PSO searches the parameter space Rusmalawati (2018), specifically $C \in [100, 500]$, $\gamma \in [10^{-5}, 10^{-3}]$, and $\varepsilon \in [10^{-4}, 10^{-2}]$, resulted in the lowest mean absolute percentage error (MAPE) on the training data and were consequently selected as the final search space for PSO in this study. The swarm consists of 30 particles iterated over a maximum of 50 iterations (Difitria, 2020), with inertia weight W linearly decaying from 0.9 to 0.1 and acceleration coefficients $c_1 = c_2 = 1.5$ (Djakaria et al., 2024). Velocity and position updates:

$$V_{id}^{k+1} = W \times V_{id}^{k} + c_1 \times rand_1 \times (P_{id} - X_{id}) + c_2 \times rand_2 \times (G_{id} - X_{id}) \tag{8}$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \tag{9}$$

where P_{best} is each particle's historical best position and G_{best} is the global swarm best. Fitness is evaluated as MAPE on the validation set. This PSO-SVR integration was chosen over grid search based on empirical evidence that PSO achieves superior hyperparameter quality with substantially lower computational cost across a wide range of time-series problems (Shami et al., 2022).

Reconstruction, De-normalization, and Evaluation

Following component-level forecasting, the final price prediction was reconstructed by summing the individual forecasts of all IMFs and the residual component:

$$\mathrm{y}(t) = \sum_{i=1}^{n} \mathrm{IMF}_i\,(t) + r(t) \tag{10}$$

where y_t is the reconstructed forecast at time t, $\mathrm{IMF}_i(t)$ represents the forecast of the i-intrinsic mode function, and $r(t)$ is the forecast of the residual component. The reconstructed output was subsequently de-normalized from the scaled range back to the original USD/MMBtu scale using the inverse of the min-max normalization applied during preprocessing.

$$x = y(x_{max} - x_{min}) + x_{min} \tag{11}$$

Forecasting accuracy was evaluated using the Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{x_t - y_t}{x} \right|}{n} \times 100\% \tag{12}$$

where $x_t$is the actual Henry Hub price, $\mathrm{y}_t$is the corresponding forecast at period t, and n is the number of observations in the evaluation period. MAPE was selected as the primary evaluation metric for two reasons. First, it served as the fitness criterion during PSO hyperparameter optimization, where the objective was to minimize validation MAPE across all components. Second, it provided an interpretable, scale-independent measure of forecasting accuracy that enables direct comparison across components with differing price magnitudes. Following the categorization criteria established in the forecasting literature, a MAPE below 10% is considered highly accurate (Nabillah 2020).

All computational procedures — including data preprocessing, EMD decomposition, PACF-based lag selection, SVR modeling, and PSO-based hyperparameter optimization — were implemented in Python 3.12.13 using Google Colaboratory. The following libraries were used: PyEMD 1.10.0 for EMD decomposition, scikit-learn 1.6.1 for SVR modeling, pyswarms 1.3.0 for PSO optimization, statsmodels 0.14.6 for nonlinearity testing and PACF analysis, and NumPy 2.0.2 and pandas 2.2.2 for numerical computation and data manipulation.

RESULT AND DISCUSSIONS

Descriptive Analysis

Descriptive analysis is used to summarize the characteristics of daily natural gas prices in. The descriptive statistics obtained using Python are presented Table 1.

Table 1. Descriptive Analysis

N	Mean	Median	Min	Max	Std. Dev.
1,575	3.594	3.037	1.544	9.647	1.695

The dataset used in this study consists of 1,575 daily natural gas closing prices from January 1, 2020 to December 31, 2025. Descriptive analysis reveals that the average price over this period was 3.594 USD/MMBtu, with a standard deviation of 1.695, indicating considerable price volatility. The minimum recorded price was 1.544 USD/MMBtu in June 2020, while the maximum reached 9.647 USD/MMBtu in August 2022. This extreme price surge in mid-2022 is attributable to geopolitical disruptions, particularly the Russia–Ukraine conflict, which significantly disrupted global energy supply chains. After the peak, prices declined and moved within a relatively stable range through 2025.

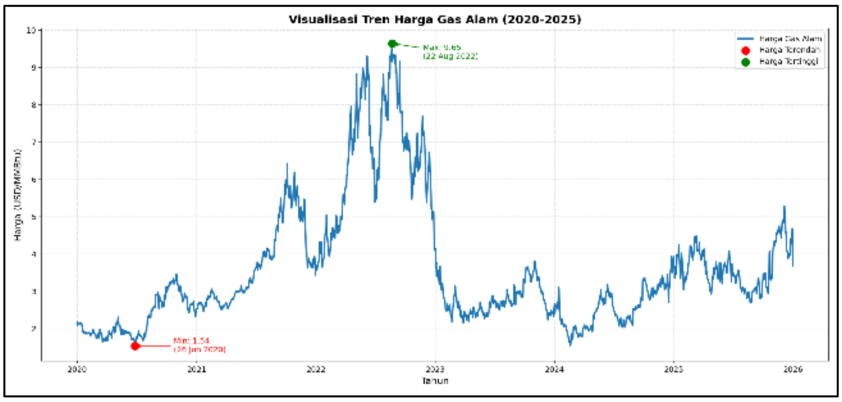

Figure 1. Natural Gas Price Plot

To determine the appropriate modeling approach, the Terasvirta neural network linearity test was applied to the price series. The test yielded a p-value of 0.0182, which is below the significance threshold of $\alpha = 0.05$. This result leads to the rejection of the null hypothesis of linearity, confirming that the natural gas price series exhibits a nonlinear pattern. This finding justifies the selection of EMD as a signal decomposition technique prior to modeling, and SVR with RBF kernel as the primary forecasting method, given its capability to handle nonlinear data structures.

Empirical Mode Decomposition

Following data normalization using the min-max method to scale all values to the range [0, 1], EMD was applied to decompose the normalized price series into its intrinsic components. The sifting process produced 7 Intrinsic Mode Functions (IMFs) and 1 residual component. IMF-1 through IMF-3 represent high-frequency components, capturing rapid short-term price fluctuations. IMF-4 through IMF-7 represent medium-to-low frequency components, reflecting longer-cycle price movements. The residual component captures the overall long-term downward trend in the normalized data, indicating a general declining tendency in relative price levels over the observation period.

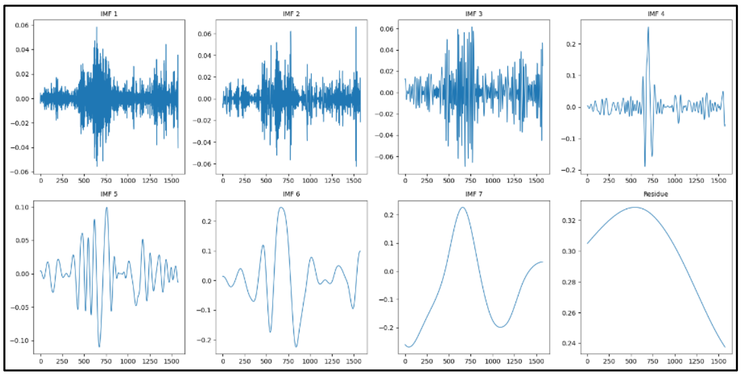

Figure 2. EMD decomposition results

The decomposition process transforms the highly complex natural gas price series into several simpler sub-series. Consequently, each component becomes easier to model individually before reconstruction, allowing the forecasting framework to represent different temporal scales more effectively.

Support Vector Regression Optimization Using Particle Swarm Optimization

For each IMF and the residual component, input lag variables were identified using the Partial Autocorrelation Function (PACF). The maximum lag across all components was lag-31, yielding an effective dataset of 1,544 observations after trimming. IMF-1 required 4 lag inputs (t−1, t−2, t−3, t−6), while IMF-3 through IMF-6 required 31 lag inputs each, reflecting the longer memory of low-frequency components, and IMF-7 required 13 lag. The residual required only lag-1 as input, consistent with its smooth monotonic trend.

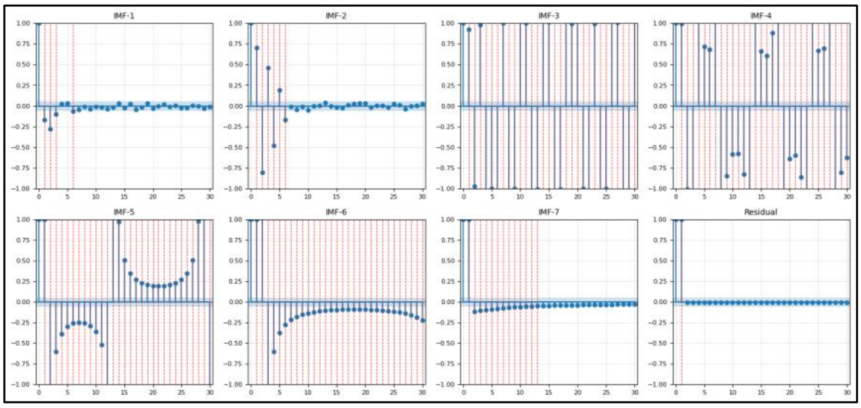

Figure 3. PACF Results for Each IMF

Hyperparameter optimization of the SVR model focused on the regularization parameter C, kernel width γ, and insensitive tube width ε. Each IMF was optimized separately using PSO. The PSO setup used 30 particles and up to 50 iterations, with an inertia weight between 0.1 and 0.9, and learning coefficients c1 and c2 set to 1.5. The search ranges were C from 100 to 500, γ from 0.00001 to 0.001, and ε from 0.0001 to 0.01. MAPE on the validation set served as the fitness function.

Table 2. Optimal SVR Hyperparameters per Component

Prop	Par	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	Residual
70:15:15	C	500.0000	500.0000	498.8325	500.0000	500.0000	500.0000	270.0882	423.8999
	γ	0.001000	0.001000	0.001000	0.001000	0.000980	0.000997	0.000381	0.000969
	ε	0.002159	0.000199	0.000100	0.000100	0.000100	0.000462	0.000100	0.000100
80:10:10	C	460.8281	500.0000	499.3850	500.0000	496.5080	431.6376	500.0000	328.5975
	γ	0.001000	0.001000	0.001000	0.000999	0.001000	0.001000	0.000549	0.000514
	ε	0.000100	0.000254	0.000100	0.000100	0.000100	0.000393	0.000100	0.000100
90:5:5	C	100.0000	500.0000	500.0000	497.1804	500.0000	462.6618	493.9472	500.0000
	γ	0.00001	0.001000	0.001000	0.001000	0.001000	0.000991	0.000665	0.000961
	ε	0.01000	0.000587	0.000100	0.000100	0.000100	0.000100	0.000100	0.000100

The optimal hyperparameter values obtained through PSO for each IMF component and Residue across all data proportion splits are presented in Table 2. The regularization parameter C generally converged toward the upper boundary of the search range at 500, indicating that the model requires a wider margin to accommodate data variability. The kernel width parameter γ consistently settled around 0.001 for most components, while the insensitive tube width ε predominantly reached its lower bound of 0.0001, reflecting the model's calibration toward very small error tolerance. Notable exceptions were observed in certain components, such as the considerably lower C value in IMF 1 under the 90:5:5

split (100.0000) and slightly larger ε values in selected IMFs, which reflect the distinct frequency characteristics of each component.

The final support vector regression (SVR) sub-models were trained independently on all eight components, each utilizing its respective optimal hyperparameters, obtained via PSO within the search bounds justified in the Methods section (Section SVR with RBF Kernel and PSO Optimization), which were selected based on preliminary comparison with alternative ranges reported in prior SVR-PSO studies. For each IMF, the trained SVR model was applied to the test data by computing the kernel matrix between the support vectors and the test observations, multiplying by the corresponding Lagrange multipliers, and adding the bias term. The predicted values from all seven IMFs and the residual were summed at each time step to reconstruct the aggregated normalized forecast. Subsequently, denormalization was performed to convert the reconstructed values back to the original USD/MMBtu scale for final evaluation.

Forecast Accuracy Evaluation

Model performance was assessed using MAPE on the test set across three data split configurations. The results are summarized in Table 3. All three configurations achieved MAPE values well below 10%, placing them in the "excellent" forecasting accuracy category. Table 3 shows that the 70:15:15 split achieved the best forecasting accuracy, with a MAPE of 2.2641% and an MAE of 0.0826, outperforming the 80:10:10 (MAPE = 2.4552%, MAE = 0.0890) and 90:5:5 (MAPE = 2.3613%, MAE = 0.0899) configurations. Accuracy did not improve monotonically with a larger training share; the 90:5:5 split, despite using the most training data, produced the highest MAE. This suggests that the larger test set in the 70:15:15 split provided a more stable and reliable evaluation, supporting its selection as the optimal partition for this study. This split is further justified by its larger and more diverse test set of 233 observations compared to only 78 in the 90:5:5 split, which provides a more robust and representative evaluation of the model's generalization capability.

Table 3. Test-Set Forecasting Accuracy Across Three Data-Split

Metric	70:15:15	80:10:10	90:5:5
MAPE (%)	2.2641	2.4552	2.3613
MAE	0.0826	0.0890	0.0899

Figure 4 shows how actual and predicted prices compare over the 233-day test period. The model closely followed all major price changes: an early rise until day 20, a drop to a low around day 55, a bumpy recovery with several smaller peaks through day 95, another low near day 145, and a strong rally that peaked near day 215. The close match between predicted and actual prices across these three periods shows that the multi-scale decomposition helps the model adapt well to varying market conditions.

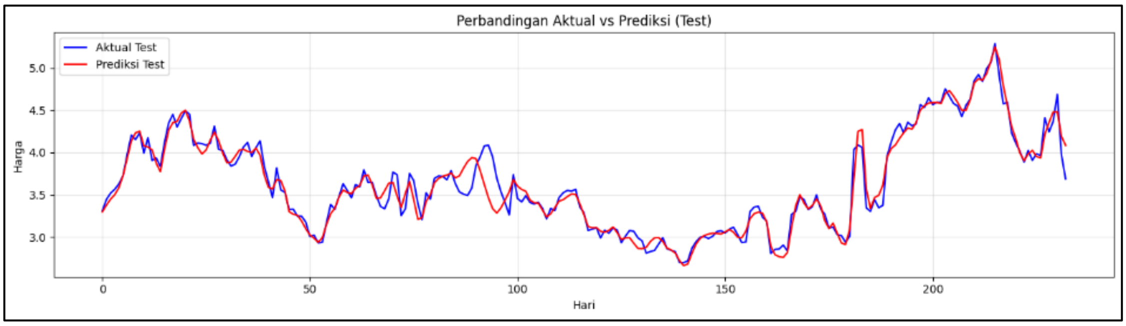

Figure 4: Comparison of actual and predicted daily natural gas prices on the 233-day test set

Ninety-Day Ahead Price Forecast

Using the best-performing 70:15:15 model, a 90-day ahead forecast was generated covering January through late March 2026. The forecast indicates an initial price increase, peaking at around mid-January 2026, followed by a sustained downward trend. Prices are projected to decline through February and into mid-March 2026, then stabilize toward the end of the forecast horizon. This medium-term declining trajectory is consistent with the trend signal captured in the residual component, suggesting that the structural market conditions embedded in the historical data point toward gradually softening prices.

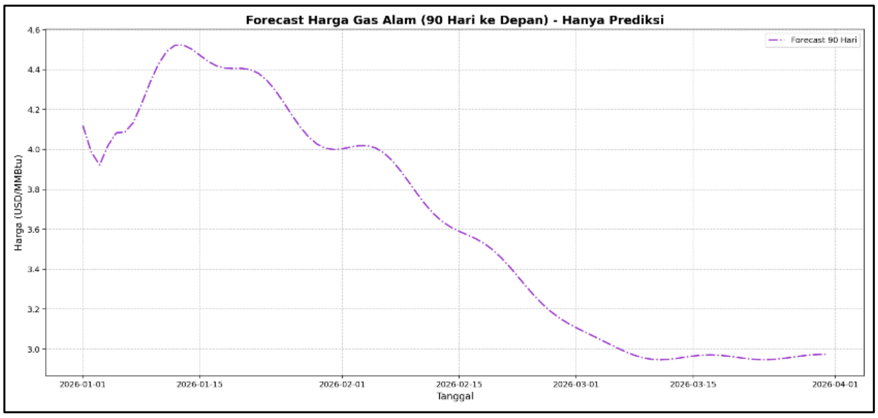

Figure 5. 90-Day Ahead Natural Gas Price Forecast Plot (January–March 2026)

The 90-day forecast results provide actionable insights for energy stakeholders. The projected short-term peak in January 2026 suggests continued seasonal demand pressure, while the sustained decline thereafter implies easing of supply constraints and reduced market uncertainty. These predictions can support energy procurement planning, investment decisions in the natural gas sector, and government policy formulation regarding energy security and pricing mechanisms. The stability of the forecast at the end of the horizon further suggests that the model has not extrapolated unrealistically beyond the learned data patterns.

Discussion

The EMD-SVR-PSO framework achieved a mean absolute percentage error (MAPE) of 2.26% on the 233-day test set. This outcome supports the central hypothesis that decomposing a nonstationary natural gas price series into simpler frequency components before applying support vector regression (SVR) reduces each sub-model's approximation complexity and improves overall forecasting accuracy. Decomposition into intrinsic mode functions (IMFs) with narrower bandwidths enables empirical mode decomposition (EMD) to isolate heterogeneous dynamics into separate components, each approaching local stationarity. Consequently, each SVR sub-model fits a more homogeneous pattern, rather than requiring a single global kernel to capture high-frequency noise, medium-term cycles, and long-term trends. This task typically degrades SVR's generalization performance. These findings align with the rationale for hybrid decomposition-regression approaches, which distribute the learning burden from a single global model to multiple locally stationary sub-problems (Awajan et al., 2019).

These results are consistent with and provide further explanation for previous EMD-based forecasting studies. (Nava et al., 2018) demonstrated that the adaptive, data-driven decomposition of EMD outperforms fixed-basis transforms in capturing irregular energy price dynamics. This advantage is especially significant for Henry Hub prices, where the volatility structure varies throughout the sample period. Fixed-basis decompositions are less adaptive to structural breaks than EMD's data-driven IMFs.

The effectiveness of particle swarm optimization (PSO)-based hyperparameter optimization is supported by evidence from Sabilul Muminin et al. 2021 and Manik et al. (2025), who found that evolutionary and swarm-intelligence search strategies yield statistically superior SVR configurations compared to grid search or default settings. Since each IMF exhibits distinct frequency and amplitude characteristics, the optimal combination of hyperparameters (C, ε, γ) varies across IMFs. Fixed grid search evaluates hyperparameters on a discrete lattice, which is inefficient and computationally expensive when applied separately to multiple IMFs. In contrast, PSO's continuous, swarm-based search adapts more efficiently to each IMF's unique loss landscape, converging toward IMF-appropriate hyperparameters without exhaustive enumeration. This efficiency explains why the accuracy gain from PSO was most pronounced in the higher-frequency IMFs, as observed in this study's convergence analysis.

The findings of this study indicate that employing PSO for hyperparameter tuning within a decomposition-reconstruction framework represents a general design principle for energy price forecasting, rather than a solution limited to Henry Hub data. The underlying rationale extends to other decomposition-based hybrid frameworks. In practical applications, this framework can be directly utilized for short-term forecasting by energy traders, procurement analysts, and grid operators who require accurate forecasts and interpretable, component-wise error diagnostics. Future research should incorporate exogenous variables, utilize higher-frequency price data, and benchmark this method against complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and variational mode decomposition (VMD) under controlled conditions to determine which decomposition strategy is most robust under specific market volatility regimes.

CONCLUSIONS AND SUGGESTIONS

The proposed hybrid forecasting framework, which integrates Empirical Mode Decomposition (EMD), Support Vector Regression (SVR), and Particle Swarm Optimization (PSO), effectively models the nonlinear dynamics of natural gas prices. This framework decomposes the original time series into seven Intrinsic Mode Functions (IMFs) and one residual component, with each component forecasted individually. The optimization process determines the optimal SVR hyperparameters for each IMF. The 70:15:15 sequential data partition yields the best forecasting performance, achieving a MAPE of 2.2641%, and surpasses the 80:10:10 (2.4552%) and 90:5:5 (2.3613%) configurations. The 90-day forecast projects a gradual decline in natural gas prices following a peak in mid-January 2026, with relative price stability through March 2026. These findings indicate that the framework successfully captures both short-term fluctuations and long-term market trends.

The forecasting results have several practical implications. In procurement planning, the projected price decline after the mid-January 2026 peak suggests that industrial gas buyers and utility procurement teams may benefit from timing contract purchases to align with the forecasted trough, rather than securing supply agreements during the anticipated peak period. This approach could reduce procurement costs during the March 2026 window. In energy pricing, the model's 2.26% MAPE demonstrates sufficient accuracy to inform short-term tariff adjustment decisions. Gas distributors and retail energy providers can utilize the 90-day forecast horizon to proactively revise pass-through pricing schedules for industrial and commercial customers in anticipation of price movements, rather than responding only after spot prices change. For investment and risk management, the framework's ability to capture both high-frequency fluctuations (via IMF-level SVR-PSO models) and the underlying trend (via the residual component) supports its use in short-term hedging strategies in the Henry Hub futures and options markets. Portfolio managers exposed to natural gas price risk may use the forecasted trajectory to inform position sizing or the timing of hedge adjustments over the 90-day horizon, particularly around the identified turning point in mid-January.

Future research should incorporate additional external variables, including crude oil prices, weather conditions, and macroeconomic indicators, and examine alternative decomposition and optimization algorithms such as CEEMDAN and VMD to further improve forecasting accuracy and model robustness. Subsequent studies could also assess the framework's forecasting performance across different volatility regimes to establish clearer boundary conditions for its practical application in procurement, pricing, and investment contexts.

REFERENCE

Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1)(3), 117–122. https://doi.org/https://doi.org/10.29207/resti.v4i1.1517

Ambya, Gunarto, T., Hendrawaty, E., Kesumah, F. S. D., & Wisnu, F. K. (2020). Future natural gas price forecasting model and its policy implication. *International Journal of Energy Economics and Policy*, 10(5), 58–63. https://doi.org/10.32479/ijeep.9676

Awajan, A. M., Tahir Ismail, M., & Wadi, S. AL. (2019). A review on empirical mode decomposition in forecasting time series. *Italian Journal of Pure and Applied Mathematics*, 42(2019), 301–323. https://journals.uniurb.it/index.php/ijpam/issue/view/491

Bakrie, C. R., Delanova, M. O., & Mochamad Yani, Y. (2022). Pengaruh Perang Rusia dan Ukraina Terhadap Perekonomian Negara Kawasan Asia Tenggara. *Jurnal Caraka Prabu*, 6(1), 65–86. https://doi.org/10.36859/jcp.v6i1.1019

Difitria, R., & Cholissodin, I. (2020). Penerapan Support Vector Regression dan Particle Swarm Optimization untuk Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Daerah Istimewa Yogyakarta. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 4(5), 1364–1371. http://j-ptiik.ub.ac.id

Ding, X., & Zhou, Z. (2025). On the partial autocorrelation function for locally stationary time series: Characterization, estimation and inference. *Biometrika*, 112(2). https://doi.org/10.1093/biomet/asaf016

Qirani, S. D., & Sukarsih, I. (2024). Penerapan Metode K-Nearest Neighbor untuk Prediksi Harga Gas Alam Menggunakan Python. *Jurnal Riset Matematika*, 4(1), 57–64. https://doi.org/10.29313/jrm.v4i1.3602

Djakaria, I., Wungguli, D., Pakadang, R. S., Saleh, E., & Djauhari, A. (2024). Indonesia's GDP Forecast: Evidence from Fuzzy Time Series Model Using Particle Swarm Optimization Algorithm. *The Australian Journal of Mathematical Analysis and Applications*, 21(2). https://ajmaa.org/cgi-bin/paper.pl?string=v21n2/V21I2P21.tex

Energy Institute. (2025). Statistical Review of World Energy 2025. https://www.aga.org/research-policy/resource-library/statistical-review-of-world-energy/

Gad, A. G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering*, 29(5), 2531–2561. https://doi.org/10.1007/s11831-021-09694-4

Saputra, G. H., Wigena, A. H., & Sartono, B. (2019). Penggunaan Support Vector Regression dalam Pemodelan Indeks Saham Syariah dengan Algoritma Grid Search. *Indonesian Journal of Statistics and Its Applications*, 3(2), 148–160. https://doi.org/https://doi.org/10.29244/ijsa.v3i2.172

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd Edition). OTexts.

Investing.com. (2026). Natural gas futures historical data. https://www.investing.com/commodities/natural-gas

Lux, M., Härdle, W. K., & Lessmann, S. (2020). Data driven value-at-risk forecasting using a SVR-GARCH-KDE hybrid. *Computational Statistics*, 35(3), 947–981. https://doi.org/10.1007/s00180-019-00934-7

Manik, P., Arianthi, S., Jannah, M., & Santoso, N. (2025). A Metaheuristic-Based Approach to Inflation Prediction in Indonesia with Support Vector Regression (SVR). *Journal of Computer Engineering, Electronics and Information Technology (COELITE)*, 4(1), 25–34. https://doi.org/10.17509/coelite.v4i1.78101

Nabillah, I., & Ranggadara, I. (2020). Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut. *JOINS (Journal of Information System)*, 5(2), 250–255. https://doi.org/10.33633/joins.v5i2.3900

Nava, N., Di Matteo, T., & Aste, T. (2018). Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks*, 6(1). https://doi.org/10.3390/risks6010007

Nurcahyono, M. B., & Widagdo, B. W. (2024). Penerapan Algoritma Linier Regresi Pada Sistem Prediksi Harga Gas Alam. *Jurnal Penelitian Ilmu Komputer*, 2(3). https://mypublikasi.com/index.php/JUPIK/article/view/110

Prabowo, H., Suhartono, S., & Prastyo, D. (2020). The Performance of Ramsey Test, White Test and Terasvirta Test in Detecting Nonlinearity. *Inferensi*, *3*. https://doi.org/10.12962/j27213862.v3i1.6876

Rahmawati, Sabilul Muminin, R., Denni, I., & Ramdhani, R. (2021). Implementation of the Support Vector Regression Algorithm And Particle Swarm Optimization In Sales Forecasting. *RISTEC : Research in Information Systems and Technology*, *1*(1). https://journal.institutpendidikan.ac.id/index.php/ristec/article/view/1187/845

Rusmalawati, V., & Tanzil Furqon, M. (2018). Peramalan Harga Saham Menggunakan Metode Support Vector Regression (SVR) Dengan Particle Swarm Optimization (PSO). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(5), 1980–1990. http://j-ptiik.ub.ac.id

Saluza, I., Taufikurrahman, M., Widya Astuti, L., & Yulianti, E. (2023). Prediksi Harga Saham Menggunakan Empirical Mode Decomposition dan Feed Forward Neural Networks. *Jurnal Penelitian Ilmu Dan Teknologi Komputer (JUPITER)*, 15(2). https://jurnal.polsri.ac.id/index.php/jupiter/article/view/7584/2892

Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, 10, 10031–10061. https://doi.org/10.1109/ACCESS.2022.3142859

Wang, J., Lei, C., & Guo, M. (2020). Daily natural gas price forecasting by a weighted hybrid data-driven model. *Journal of Petroleum Science and Engineering*, 192. https://doi.org/10.1016/j.petrol.2020.107240

Wang, Y. G., Wu, J., Hu, Z. H., & McLachlan, G. J. (2023). A new algorithm for support vector regression with automatic selection of hyperparameters. *Pattern Recognition*, 133. https://doi.org/10.1016/j.patcog.2022.108989

Xie, G., Jiang, F., & Zhang, C. (2023). A secondary decomposition-ensemble methodology for forecasting natural gas prices using multisource data. *Resources Policy*, 85. https://doi.org/10.1016/j.resourpol.2023.104059

Zarkasy, R. A., Fahrezi, D. N., Ahmad, I., Rosyid, H., Aji, K., & Yuha, W. (2022). Dampak Pemberhentian Pasok Gas Terhadap Kerja Sama Rusia-Uni Eropa. *Jurnal Hukum Dan HAM Wara Sains*, 1(2), 169–179. https://wnj.westsciences.com/index.php/jhhws/article/view/67

Zheng, Y., Luo, J., Chen, J., Chen, Z., & Shang, P. (2023). Natural gas spot price prediction research under the background of Russia-Ukraine conflict - based on FS-GA-SVR hybrid model. *Journal of Environmental Management*, 344. https://doi.org/10.1016/j.jenvman.2023.118446