

Fitness Function Selection in PSO-VMD Decomposition: Implications for LSTM-Based Oil Price Forecasting

Abdelkader AOUAR

akaouar.eco@gmail.com

Research Article

Keywords: Crude oil price forecasting, Variational Mode Decomposition, Long Short-Term Memory, Time series forecasting

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11097467/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Fitness Function Selection in PSO-VMD Decomposition: Implications for LSTM-Based Oil Price Forecasting

Abdelkader AOUAR^{1*}

^{1*}Algeria.

Corresponding author(s). E-mail(s): akaouar.eco@gmail.com;

Abstract

Accurate crude oil price forecasting remains a challenging task due to the nonlinear, non-stationary, and highly volatile nature of financial time series. To address these challenges, this study proposes a hybrid PSO-VMD-LSTM framework that integrates Particle Swarm Optimization (PSO), Variational Mode Decomposition (VMD), and Long Short-Term Memory (LSTM) networks. In the proposed framework, PSO is employed to optimize the key VMD parameters, namely the number of decomposition modes (K) and the penalty factor (α), while four alternative fitness functions—Envelope Entropy, Sample Entropy, Permutation Entropy, and Reconstruction Error—are systematically evaluated. The decomposed intrinsic mode functions are then used as inputs to the LSTM model for crude oil price forecasting. Experimental results demonstrate that the choice of the fitness function has a significant impact on forecasting performance. Among all evaluated criteria, Reconstruction Error achieved the best results with an RMSE of 1.5444, an MAE of 1.1020, a MAPE of 1.37%, and an R^2 of 0.9859, substantially outperforming the entropy-based alternatives. The findings indicate that preserving the information contained in the original signal is more effective than entropy-based optimization for VMD parameter selection in crude oil price forecasting. Overall, the proposed PSO-VMD-LSTM framework provides an accurate and robust solution for nonlinear financial time series forecasting and offers a promising optimization strategy for decomposition-based predictive models.

Keywords: Crude oil price forecasting, Variational Mode Decomposition, Long Short-Term Memory, Time series forecasting

JEL Classification: C22 , C45 , C61 , Q41

1 Introduction

Crude oil price forecasting remains one of the most challenging problems in time series analysis, owing to the pronounced non-stationarity, sharp volatility, and complex geopolitical and economic drivers that are difficult to capture with traditional statistical methods [1].

In recent years, hybrid models that combine signal decomposition with deep learning have emerged as one of the most effective approaches, most notably the framework built on Variational Mode Decomposition (VMD) coupled with Long Short-Term Memory (LSTM) networks [1].

Because the effectiveness of VMD depends critically on the tuning of two key parameters—the number of modes (K) and the quadratic penalty factor (α)—and because manual selection of these parameters is unreliable and difficult to generalize, numerous studies have employed metaheuristic optimization algorithms such as Particle Swarm Optimization (PSO) to automate this tuning process [2].

Despite its importance, many published studies adopt envelope entropy as the optimization criterion, largely owing to its successful application in other domains, such as mechanical bearing fault diagnosis, without sufficiently evaluating its suitability for financial time series, which exhibit fundamentally different statistical characteristics, including volatility clustering, abrupt price jumps, and nonlinear dynamics.

Consequently, the choice of the fitness function has received considerably less attention than the optimization algorithm itself, and systematic comparisons among alternative fitness functions remain scarce in the literature.

2 Literature Review

Crude oil price forecasting has long attracted considerable attention owing to its crucial role in global economic activity and the substantial challenges posed by the nonlinear, nonstationary, and highly volatile nature of oil price dynamics. Early forecasting studies relied predominantly on classical statistical and econometric models, particularly the Autoregressive Integrated Moving Average (ARIMA) model, which is based on the Box–Jenkins methodology and provides a flexible framework for modeling linear time series [3]. Despite their widespread adoption and satisfactory performance in short-term forecasting, ARIMA models are inherently limited in their ability to represent the nonlinear dynamics and volatility clustering that characterize crude oil prices [3]. Comparative studies have consistently shown that neither ARIMA nor the GARCH family of models can adequately capture all characteristics of oil price movements across different market conditions. For example, Yahaya and Etuk concluded that no single model within the ARIMA or GARCH families consistently outperforms the others, highlighting the persistent difficulty of modeling the nonlinear behavior of crude oil prices. To partially overcome these limitations, hybrid ARIMA–GARCH frameworks were introduced, where ARIMA models the conditional mean while GARCH captures the conditional heteroskedasticity of the residuals [3, 4]. Such hybrid approaches have demonstrated improved performance in forecasting oil price return volatility [4, 5], and subsequent studies have further confirmed their effectiveness in jointly modeling the conditional mean and volatility of international crude oil prices [6]. Nevertheless,

these approaches continue to rely fundamentally on assumptions of linearity and stationarity that rarely hold for real-world crude oil price series, thereby limiting their capability to characterize the complex nonlinear structures and stochastic fluctuations observed in practice [6, 7].

The inability of conventional statistical models to adequately describe the intrinsic complexity of crude oil price dynamics motivated the emergence of decomposition-based forecasting frameworks, commonly referred to as the decomposition-then-forecast paradigm. Rather than modeling the original series directly, these approaches first decompose it into a set of simpler sub-series with more regular statistical properties before applying forecasting models to each component individually [7]. Among the earliest and most influential decomposition techniques is Empirical Mode Decomposition (EMD), proposed by Huang et al., which adaptively decomposes a signal into a finite number of Intrinsic Mode Functions (IMFs) without requiring predefined basis functions, unlike conventional Fourier- or wavelet-based methods [8]. Despite its flexibility and fully data-driven nature, EMD suffers from the well-known mode-mixing problem, whereby oscillations with substantially different frequencies may coexist within a single IMF, or similar frequency components may be distributed across multiple IMFs, thereby reducing the physical interpretability of the decomposition [8, 9]. This limitation becomes particularly pronounced for financial time series, which are inherently noisy, nonlinear, and strongly nonstationary [9].

To alleviate the mode-mixing problem, Wu and Huang introduced Ensemble Empirical Mode Decomposition (EEMD), which repeatedly adds finite-amplitude white noise to the original signal, performs EMD decomposition on each realization, and averages the resulting IMFs across the ensemble [10]. Although EEMD significantly mitigates mode mixing, its effectiveness depends heavily on manually selected parameters such as the ensemble size and noise amplitude, while residual noise may still remain in the reconstructed signal [10, 11]. These limitations subsequently motivated the development of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), which introduces adaptive noise at successive decomposition stages to achieve more accurate signal reconstruction and further suppress mode mixing relative to both EMD and EEMD [11]. Owing to these improvements, CEEMDAN has been successfully incorporated into numerous hybrid forecasting frameworks for financial applications, including crude oil price prediction, where it has been combined with sample entropy to assess the complexity of decomposed components and integrated with LSTM networks optimized through adaptive-learning-strategy particle swarm optimization [12]. Although these methods substantially improved decomposition quality, they remained constrained by the empirical and iterative nature of the sifting process, which lacks a rigorous mathematical formulation and exhibits sensitivity to noise and sampling errors [13].

To overcome these theoretical limitations, Dragomiretskiy and Zosso proposed Variational Mode Decomposition (VMD), a mathematically rigorous decomposition technique formulated as a constrained variational optimization problem [13, 14]. Unlike EMD-based methods, VMD estimates all modes concurrently through a non-recursive optimization process, forcing each mode to be compactly concentrated around its own center frequency. This formulation yields greater robustness to noise, improved

decomposition stability, and substantially reduced mode mixing [13, 15]. Furthermore, because VMD can be interpreted as an adaptive Wiener filter bank, it provides a theoretically sound mechanism for separating signal components with different frequency characteristics while overcoming many of the intrinsic shortcomings associated with empirical decomposition techniques [16]. Consequently, VMD has become one of the most widely adopted signal decomposition methods in modern forecasting frameworks for nonlinear time series, particularly in applications involving financial and energy markets.

Despite its superior theoretical foundation and decomposition capability, the performance of VMD is highly sensitive to two user-defined parameters: the number of modes (K) and the quadratic penalty factor (α). Inappropriate parameter selection may lead to either under-decomposition, where important frequency information is overlooked, or over-decomposition, where redundant modes are generated, both of which adversely affect subsequent forecasting performance [17]. Since manually determining these parameters is subjective and difficult to generalize across different datasets, numerous studies have adopted metaheuristic optimization algorithms to automate the parameter selection process. Among these, Particle Swarm Optimization (PSO), originally proposed by Kennedy and Eberhart [18], remains one of the most widely used approaches because of its simplicity and computational efficiency. More recently, other optimization techniques, including Genetic Algorithms (GA), Sparrow Search Algorithm (SSA), Newton–Raphson-Based Optimization (NRBO), Artificial Lemming Algorithm (ALA), and Improved Hippopotamus Optimization (IHO), have also been successfully employed to search for the optimal VMD parameter combination [17, 19–24]. Regardless of the optimization algorithm employed, however, the search process fundamentally depends on the adopted fitness function, which determines how candidate parameter combinations are evaluated and ultimately governs the quality of the resulting decomposition.

Among the objective functions proposed in the literature, envelope entropy has become by far the most widely adopted criterion for VMD parameter optimization. Originally introduced in the field of mechanical fault diagnosis, envelope entropy quantifies the sparsity of the signal envelope and tends to decrease when periodic transient features become more pronounced, making it particularly effective for analyzing vibration signals generated by defective mechanical components [17, 19]. Consequently, numerous studies have adopted minimum envelope entropy as the optimization objective when determining the optimal values of K and α , including IPSO-VMD for rolling bearing fault diagnosis [17], GMP SO-VMD [19], SSA-VMD combined with RCMDE [21], and other enhanced optimization frameworks [20]. More recently, this criterion has also been incorporated into hybrid deep learning frameworks based on NRBO-VMD-BiLSTM for predicting dissolved gas concentrations in transformer oil [22]. Although these studies consistently report improved decomposition quality and predictive performance, the widespread adoption of envelope entropy has largely been inherited from mechanical vibration analysis, where the optimization objective is closely related to the extraction of transient fault features.

Besides envelope entropy, several alternative objective functions have been explored for optimizing VMD parameters. Sample entropy and permutation entropy are among

the most frequently investigated alternatives, as both quantify the complexity and regularity of decomposed modes. Permutation entropy, in particular, has been employed directly as the fitness function in VMD optimization for applications such as natural gas pipeline leak detection using the Salp Swarm Algorithm [25] and wind turbine gearbox fault diagnosis using an immunized fruit fly optimization algorithm [26]. Likewise, multi-scale permutation entropy has been integrated with VMD and PSO-SVM for rolling bearing fault diagnosis [27]. Other studies have proposed reconstruction error as an optimization criterion, which directly evaluates the discrepancy between the original signal and the reconstructed signal obtained by summing all decomposed modes. Compared with entropy-based measures, reconstruction error is more directly related to preserving the information content of the original time series and may therefore be better aligned with forecasting objectives. In addition, several researchers have introduced composite objective functions that simultaneously optimize multiple decomposition characteristics. Examples include combining permutation entropy, the Pearson correlation coefficient, and inter-mode distinguishability within an improved hippopotamus optimization framework [24], as well as integrating envelope entropy with Rényi entropy through a multi-island genetic algorithm [28]. These studies collectively demonstrate that the choice of fitness function is itself an important research problem and may substantially influence the quality of the resulting decomposition.

Despite the diversity of proposed objective functions, direct comparative investigations remain remarkably scarce. One of the few systematic comparisons was conducted by Chen et al., who evaluated four entropy-based fitness functions—envelope entropy, information entropy, permutation entropy, and sample entropy—within the same VMD optimization framework for processing hydrodynamic data collected from underwater gliders. Their results indicated that envelope entropy produced the highest reconstruction quality among the evaluated criteria [29]. Although this study provides valuable methodological evidence demonstrating that different fitness functions can significantly influence decomposition performance, its conclusions were derived from physical hydrodynamic signals whose statistical properties differ substantially from those of financial time series. Consequently, whether the superiority of envelope entropy extends to crude oil price forecasting remains an open question.

This question has become increasingly important because envelope entropy has recently been transferred directly from mechanical signal analysis to financial forecasting applications. A notable example is the recently proposed ALA-VMD-CASA framework, in which the Artificial Lemming Algorithm adaptively optimizes the VMD parameters by minimizing envelope entropy before forecasting Brent crude oil prices through a hybrid deep learning architecture. Using data spanning 2010–2025, the proposed framework achieved superior forecasting performance compared with several benchmark models, including ARIMA, LSTM, and Transformer architectures [23]. Nevertheless, the study adopted envelope entropy without conducting a systematic comparison against alternative fitness functions. Given that envelope entropy was originally designed to characterize transient impulsive features in mechanical vibration signals rather than the volatility clustering, structural breaks, and nonlinear dependencies that characterize financial markets, its widespread adoption in crude oil forecasting

appears to be driven more by precedent than by empirical evidence demonstrating its superiority for financial time series.

Although numerous studies have successfully combined VMD with metaheuristic optimization algorithms to enhance decomposition quality and forecasting performance, the existing literature reveals a notable methodological imbalance. Most optimization frameworks—including those based on PSO, GA, SSA, NRBO, ALA, and IHO—primarily focus on developing increasingly sophisticated optimization algorithms while adopting a single predefined fitness function, most commonly envelope entropy, without providing empirical evidence demonstrating its superiority over alternative objective functions within the same forecasting framework [17, 19–24]. Consequently, improvements in forecasting performance reported by these studies cannot be attributed solely to the optimization algorithm itself, since the influence of the adopted fitness function is rarely isolated or systematically evaluated.

Moreover, the limited number of studies that have directly compared multiple fitness functions have been conducted almost exclusively in non-financial domains. For example, Chen et al. compared four entropy-based objective functions within a unified VMD optimization framework using hydrodynamic data from underwater gliders and demonstrated that different fitness functions can lead to substantially different decomposition outcomes [29]. However, hydrodynamic signals differ fundamentally from crude oil price series, which exhibit pronounced nonlinearity, volatility clustering, structural breaks, and rapidly changing market dynamics. Consequently, the superiority of a particular fitness function in physical signal processing cannot be assumed to generalize to financial time series forecasting.

Furthermore, studies adopting alternative objective functions—such as permutation entropy [25, 26], multi-scale permutation entropy [27], reconstruction error, or composite criteria [24, 28]—have typically evaluated these functions independently and under different experimental settings, optimization algorithms, datasets, and forecasting models. Such heterogeneous experimental designs make it impossible to determine whether the observed performance differences arise from the fitness function itself or from other methodological factors. As a result, the literature still lacks a unified and controlled comparative evaluation of representative fitness functions under identical experimental conditions, particularly in the context of crude oil price forecasting.

To address this gap, the present study performs a systematic comparison of several representative fitness functions for VMD parameter optimization within a unified forecasting framework. By maintaining the same decomposition method, optimization algorithm, forecasting model, dataset, and evaluation protocol while varying only the fitness function, this study isolates the effect of the optimization criterion itself on decomposition quality and forecasting accuracy. This experimental design provides a fair and rigorous assessment of the suitability of different fitness functions for crude oil price forecasting and offers empirical evidence that may guide future research in selecting appropriate optimization criteria for VMD-based forecasting models.

3 Methodology

The proposed forecasting framework consists of three sequential stages: signal decomposition, parameter optimization, and time-series forecasting. First, the original crude oil price series is pre-processed and transformed into a supervised learning dataset. Second, Variational Mode Decomposition (VMD) is employed to decompose the normalized signal into multiple intrinsic mode functions (IMFs), while Particle Swarm Optimization (PSO) is used to determine the optimal decomposition parameters. Finally, each decomposed component is independently modeled using a Long Short-Term Memory (LSTM) network, and the individual predictions are aggregated to generate the final oil price forecast.

3.1 Data Pre-processing

The daily crude oil price series is first inspected to ensure data quality. Missing observations, if any, are appropriately handled before model development. Since neural networks are sensitive to the scale of input variables, the original series is normalized using Min-Max normalization to map all observations into the interval $[0, 1]$:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where x_i denotes the original observation, while $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of the training dataset, respectively.

After normalization, the one-dimensional time series is transformed into a supervised learning problem using a sliding-window approach. Given a look-back window of length L , the input-output pairs are constructed as

$$X_t = [x_t, x_{t+1}, \dots, x_{t+L-1}], \quad (2)$$

$$Y_t = x_{t+L}, \quad (3)$$

where X_t represents the input sequence and Y_t denotes the corresponding prediction target.

Finally, the dataset is divided into training and testing subsets. The training set is employed for model learning and parameter optimization, whereas the testing set is reserved exclusively for performance evaluation.

3.2 Signal Decomposition Using Variational Mode Decomposition

Although crude oil prices exhibit highly nonlinear and non-stationary characteristics, these complex fluctuations can be represented as the superposition of several oscillatory components with different frequency bands. Therefore, Variational Mode Decomposition (VMD) is adopted to separate the original signal into a finite number of intrinsic mode functions (IMFs), each corresponding to a specific frequency range.

Given an input signal $x(t)$, VMD decomposes it into K intrinsic mode functions according to

$$x(t) = \sum_{k=1}^K u_k(t), \quad (4)$$

where $u_k(t)$ denotes the k -th IMF.

Unlike empirical decomposition techniques, VMD formulates signal decomposition as a constrained variational optimization problem that simultaneously estimates all modes and their corresponding center frequencies. The optimization objective minimizes the sum of bandwidths of all decomposed modes while satisfying the reconstruction constraint:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{subject to} & \quad \sum_{k=1}^K u_k(t) = x(t), \end{aligned} \quad (5)$$

where

- $u_k(t)$ is the k -th intrinsic mode function,
- ω_k denotes its center frequency,
- $\delta(t)$ is the Dirac delta function,
- $*$ denotes the convolution operator,
- ∂_t represents the partial derivative with respect to time.

The constrained optimization problem is converted into an unconstrained optimization problem by introducing a quadratic penalty term and a Lagrange multiplier. The augmented Lagrangian is expressed as

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| x - \sum_{k=1}^K u_k \right\|_2^2 + \left\langle \lambda, x - \sum_{k=1}^K u_k \right\rangle, \end{aligned} \quad (6)$$

where α is the quadratic penalty parameter and λ denotes the Lagrange multiplier.

The optimization problem is subsequently solved using the Alternating Direction Method of Multipliers (ADMM), which iteratively updates each mode and its center frequency until convergence. During each iteration, the center frequency of the k -th mode is estimated by

$$\omega_k = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \quad (7)$$

where $\hat{u}_k(\omega)$ denotes the Fourier transform of the corresponding IMF.

The decomposition quality strongly depends on two key parameters: the number of decomposition modes (K) and the quadratic penalty factor (α). Selecting inappropriate values may lead to either under-decomposition or over-decomposition, thereby reducing forecasting performance. Therefore, rather than assigning these parameters empirically, they are automatically optimized using Particle Swarm Optimization, as described in the following subsection.

3.3 PSO-based Optimization of VMD Parameters

The decomposition performance of VMD is highly dependent on the selection of the number of modes (K) and the quadratic penalty factor (α). Instead of manually tuning these parameters, Particle Swarm Optimization (PSO) is employed to automatically determine their optimal values. In the proposed framework, each particle represents a candidate solution defined by the parameter pair (K, α) , while the quality of each candidate is evaluated through a multi-objective fitness function.

In this study, the search variables of the PSO algorithm are the VMD parameters, namely the number of decomposition modes (K) and the penalty factor (α). Each particle therefore represents a candidate pair (K, α) , and the optimization process aims to identify the parameter combination that minimizes the proposed fitness function.

Initially, a population of particles is randomly generated within the predefined search space. During each iteration, particles update their velocities and positions according to both their personal best solution and the global best solution found by the swarm. The velocity update is defined as

$$v_i^{t+1} = wv_i^t + c_1r_1(p_i - x_i^t) + c_2r_2(g - x_i^t), \quad (8)$$

where

- v_i^t is the velocity of particle i at iteration t ;
- x_i^t denotes the current particle position;
- p_i is the personal best position;
- g is the global best position;
- w is the inertia weight;
- c_1 and c_2 are the cognitive and social learning coefficients;
- $r_1, r_2 \sim U(0, 1)$ are uniformly distributed random numbers.

The particle position is subsequently updated according to

$$x_i^{t+1} = x_i^t + v_i^{t+1}. \quad (9)$$

For each candidate solution, VMD is performed using the corresponding (K, α) values, and the decomposition quality is evaluated using four complementary criteria.

- **Envelope Entropy (H_{env}):** This entropy evaluates the sparsity of the Hilbert envelope associated with each IMF. Lower values indicate that the decomposed modes possess more concentrated energy distributions and clearer oscillatory structures.
- **Sample Entropy (H_{samp}):** Sample entropy measures the complexity and irregularity of the decomposed components. It is computed as

$$H_{smp} = -\ln\left(\frac{A^{m+1}(r)}{B^m(r)}\right), \quad (10)$$

where m denotes the embedding dimension, r is the similarity tolerance, and $A^{m+1}(r)$ and $B^m(r)$ represent the numbers of matched vector pairs of dimensions $m+1$ and m , respectively.

- **Permutation Entropy (H_{perm}):** Permutation entropy quantifies the ordinal complexity of each IMF according to

$$H_{perm} = -\sum_{i=1}^n p_i \ln(p_i), \quad (11)$$

where p_i denotes the probability associated with the i -th ordinal pattern.

- **Reconstruction Error (RMSE):** To preserve the information contained in the original signal, the reconstruction error between the original time series and the reconstructed signal is calculated as

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2}, \quad (12)$$

where x_t and $\hat{x}_t$ denote the original and reconstructed signals, respectively.

Since these criteria characterize different aspects of decomposition quality, they are integrated into a unified optimization objective defined as

$$Fitness = \sum_{i=1}^M w_i E_i + \lambda RMSE, \quad (13)$$

where E_i denotes the normalized entropy measure, w_i is the corresponding weighting coefficient satisfying

$$\sum_{i=1}^M w_i = 1,$$

and λ is a regularization coefficient controlling the trade-off between information complexity and reconstruction accuracy.

The parameter combination producing the minimum fitness value is selected as the optimal VMD configuration and subsequently employed for signal decomposition.

3.4 LSTM-based Forecasting

After obtaining the optimized intrinsic mode functions, each IMF is considered an independent feature sequence and modeled using a Long Short-Term Memory (LSTM) neural network. Owing to its gated memory mechanism, the LSTM effectively captures both short-term dynamics and long-term temporal dependencies while mitigating the vanishing gradient problem commonly encountered in conventional recurrent neural networks.

At each time step t , the forget gate determines the proportion of historical information retained in the memory cell:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (14)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function.

The input gate determines which new information should be incorporated into the cell state:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (15)$$

while the candidate memory state is computed as

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c). \quad (16)$$

The cell state is then updated according to

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (17)$$

where $\odot$ denotes element-wise multiplication.

Next, the output gate is calculated as

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (18)$$

and the hidden state is obtained by

$$h_t = o_t \odot \tanh(C_t). \quad (19)$$

Each decomposed IMF is independently learned by the LSTM network, producing a corresponding prediction. The individual forecasts are subsequently combined to reconstruct the final crude oil price prediction, as described in the following subsection.

4 Results and Discussion

The proposed PSO-VMD-LSTM framework was evaluated using four alternative fitness functions for optimizing the VMD parameters. Specifically, Particle Swarm Optimization (PSO) was employed to determine the optimal number of decomposition modes (K) and penalty factor (α) according to each fitness criterion. The resulting VMD decomposition was subsequently used to construct the LSTM forecasting model. The predictive performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The comparative results are presented in Table 1.

The results clearly demonstrate that the selection of the VMD fitness function has a substantial impact on the forecasting accuracy of the proposed framework. Among all investigated criteria, the Reconstruction Error objective consistently achieved the best predictive performance, yielding the lowest RMSE, MAE, and MAPE values, together with the highest coefficient of determination. Compared with the best entropy-based criterion (Permutation Entropy), the Reconstruction Error reduced the RMSE from 3.9702 to 1.5444, representing an improvement of approximately 61%, while increasing the R^2 value from 0.9070 to 0.9859. These findings indicate that preserving the

Table 1 Performance comparison of PSO-optimized VMD-LSTM models using different VMD fitness functions.

Fitness Function	K	α	RMSE	MAE	MAPE (%)	R^2
Envelope Entropy	9	1513.57	15.1358	12.5370	14.86	-0.3519
Sample Entropy	2	2975.65	4.0805	2.8906	3.49	0.9017
Permutation Entropy	2	2937.90	3.9702	2.8074	3.40	0.9070
Reconstruction Error	9	117.73	1.5444	1.1020	1.37	0.9859

information contained in the original signal during decomposition is more beneficial for crude oil price forecasting than solely maximizing signal regularity.

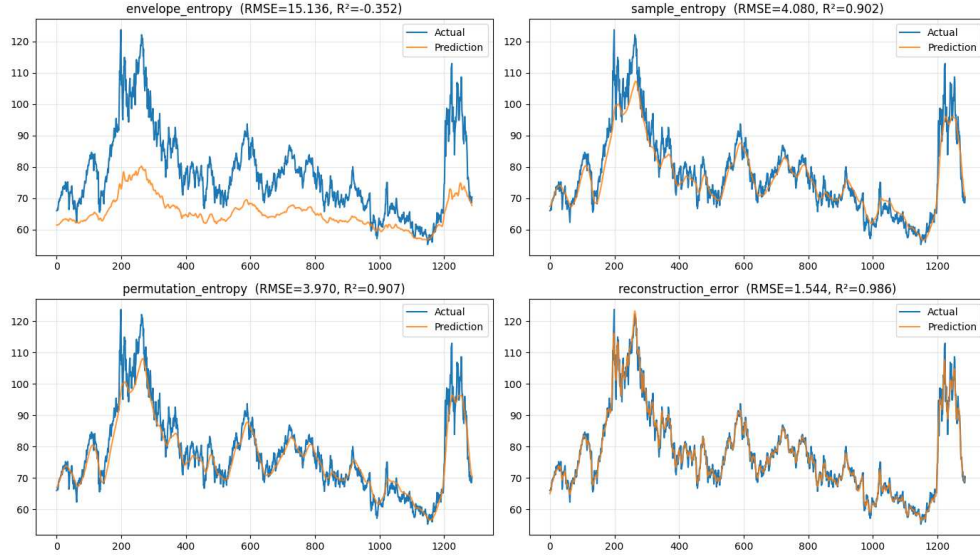


Fig. 1 Convergence behavior of PSO and comparative RMSE performance obtained using the four VMD fitness functions.

Figure 1 illustrates both the convergence characteristics of the PSO optimization process and the forecasting performance obtained using the four fitness functions. Several important observations can be drawn from these experimental results.

First, the Reconstruction Error criterion clearly outperformed all entropy-based objectives. This suggests that minimizing the reconstruction loss enables VMD to preserve the intrinsic characteristics of crude oil price fluctuations more effectively, producing decomposed components that retain essential information for the LSTM predictor. Consequently, the proposed framework achieves substantially lower prediction errors and a significantly higher coefficient of determination.

Second, the optimization process selected a relatively large number of decomposition modes ($K = 9$) when Reconstruction Error was adopted, whereas both Sample Entropy and Permutation Entropy converged to only two modes. This result indicates

that crude oil prices exhibit complex multi-scale temporal dynamics that cannot be adequately represented by only a few intrinsic mode functions. A richer decomposition allows the model to capture long-term trends, medium-frequency oscillations, and short-term market fluctuations simultaneously, thereby improving the learning capability of the LSTM network.

Furthermore, the optimal penalty parameter obtained using Reconstruction Error ($\alpha = 117.73$) is considerably smaller than those selected by the entropy-based criteria. This relatively low penalty factor relaxes the bandwidth constraint imposed during decomposition, allowing each mode to preserve more useful information from the original signal. Conversely, the entropy-based objectives favored much larger α values, resulting in smoother decompositions that inevitably removed part of the informative market dynamics together with the noise.

It is also noteworthy that Envelope Entropy produced the weakest forecasting performance despite selecting the same number of modes ($K = 9$) as the Reconstruction Error criterion. This observation demonstrates that the decomposition quality depends not only on the number of modes but also on the optimization objective itself. Simply increasing the decomposition resolution is insufficient if the fitness function does not preserve the predictive information embedded in the original time series.

Overall, the experimental results demonstrate that the Reconstruction Error criterion constitutes the most suitable optimization objective for PSO-based VMD parameter selection in crude oil price forecasting. By directly minimizing information loss during signal decomposition, it generates more informative intrinsic mode functions that significantly enhance the predictive capability of the LSTM model. These findings confirm the effectiveness of the proposed PSO-VMD-LSTM framework and highlight the importance of selecting an appropriate fitness function for decomposition-based forecasting models.

5 Conclusion

This study presented a hybrid PSO-VMD-LSTM framework for improving crude oil price forecasting by combining Particle Swarm Optimization (PSO), Variational Mode Decomposition (VMD), and Long Short-Term Memory (LSTM) networks. In the proposed approach, PSO was employed to optimize the VMD parameters, while four alternative fitness functions were systematically investigated to identify the most appropriate decomposition criterion.

The experimental results demonstrated that the Reconstruction Error fitness function consistently achieved the best forecasting performance, yielding the lowest RMSE, MAE, and MAPE values together with the highest coefficient of determination ($R^2 = 0.9859$). These findings indicate that preserving the original signal information during decomposition is more effective than entropy-based optimization criteria for modeling the complex nonlinear behavior of crude oil prices. Furthermore, the optimized decomposition with $K = 9$ modes provided a richer representation of multi-scale market dynamics, enabling the LSTM model to learn more informative temporal patterns.

Overall, the proposed PSO-VMD-LSTM framework provides an effective and reliable solution for decomposition-based financial time series forecasting and demonstrates the importance of selecting an appropriate optimization objective for VMD parameter estimation.

Despite these encouraging results, the proposed model was evaluated using historical crude oil price data only and did not explicitly incorporate external factors such as macroeconomic indicators, geopolitical events, or market sentiment. Future research will investigate the integration of such exogenous variables, as well as the application of the proposed optimization strategy to other financial and energy market forecasting problems.

Declarations

- Funding: Not applicable.
- Code availability The custom code developed for this study is available from the corresponding author on reasonable request.

References

- [1] Li T, Qian Z, Deng W, Zhang D. Forecasting crude oil prices based on variational mode decomposition and random sparse Bayesian learning. *Applied Soft Computing*. 2021;113:108032. <https://doi.org/10.1016/j.asoc.2021.108032>.
- [2] Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm. *Energy*. 2021;229:120797. <https://doi.org/10.1016/j.energy.2021.120797>.
- [3] Yaziz SR, Ahmad MH, Nian LC, Muhammad N. The performance of hybrid ARIMA-GARCH modeling in forecasting gold price. In: *Proceedings of the 20th International Congress on Modelling and Simulation*; 2013. .
- [4] Dritsaki C. The performance of hybrid ARIMA-GARCH modeling and forecasting oil price. *International Journal of Energy Economics and Policy*. 2018;8(3):14–21.
- [5] Mohammadi H, Su L. International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*. 2010;32(5):1001–1008. <https://doi.org/10.1016/j.eneco.2010.04.009>.
- [6] Lu W, Huang Z. Crude oil prices forecast based on mixed-frequency deep learning approach and intelligent optimization algorithm. *Entropy*. 2024;26(5):358. <https://doi.org/10.3390/e26050358>.

- [7] Moshiri S, Foroutan F. Forecasting nonlinear crude oil futures prices. *The Energy Journal*. 2006;27(4). DOI not confirmed via search – verify before submission.
- [8] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A*. 1998;454:903–995. <https://doi.org/10.1098/rspa.1998.0193>.
- [9] Wu Z, Huang NE. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London A*. 2005;460:1597–1611. <https://doi.org/10.1098/rspa.2003.1221>.
- [10] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*. 2009;1(1):1–41. <https://doi.org/10.1142/S1793536909000047>.
- [11] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2011. p. 4144–4147.
- [12] Wang J, et al. Crude oil price forecasting: An ensemble-driven long short-term memory model based on CEEMDAN decomposition and ALS-PSO optimization. *Energy Science & Engineering*. 2023;<https://doi.org/10.1002/ese3.1561>.
- [13] Dragomiretskiy K, Zosso D. Variational mode decomposition. *IEEE Transactions on Signal Processing*. 2014;62(3):531–544. <https://doi.org/10.1109/TSP.2013.2288675>.
- [14] Dragomiretskiy K, Zosso D. Two-dimensional variational mode decomposition. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*. vol. 8932 of *Lecture Notes in Computer Science*. Cham: Springer; 2015. p. 197–208.
- [15] Li C, Liu Y, Liao Y. An Improved Parameter-Adaptive Variational Mode Decomposition Method and Its Application in Fault Diagnosis of Rolling Bearings. *Shock and Vibration*;2021(1):2968488. <https://doi.org/https://doi.org/10.1155/2021/2968488>.
- [16] Wang Y, Markert R, Xiang J, Zheng W. An optimized VMD method and its applications in bearing fault diagnosis. *Measurement*. 2020;166:108185. <https://doi.org/10.1016/j.measurement.2020.108185>.
- [17] Zhang Q, Chen S, Fan ZP. Bearing fault diagnosis based on improved particle swarm optimized VMD and SVM models. *Advances in Mechanical Engineering*. 2021;13(6). <https://doi.org/10.1177/16878140211028451>.

- [18] Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 – International Conference on Neural Networks. vol. 4; 1995. p. 1942–1948.
- [19] Ding J, Huang L, Xiao D, Li X. GMPSO-VMD algorithm and its application to rolling bearing fault feature extraction. *Sensors*. 2020;20(7):1946. <https://doi.org/10.3390/s20071946>.
- [20] Tian Z, Ke Z, Liu H, Zhao B, Zhu X, Tang Y, et al. Bearing fault diagnosis method based on an improved particle swarm optimization–variational mode decomposition (PSO-VMD) multidimensional index and its application in a crane amplitude-changing mechanism. *Mechanical Sciences*. 2025;16:431–443. <https://doi.org/10.5194/ms-16-431-2025>.
- [21] Wang X, et al. Fault diagnosis method of rolling bearing based on SSA-VMD and RCMDE. *Scientific Reports*. 2024;14:30637. <https://doi.org/10.1038/s41598-024-81262-9>.
- [22] Wang N, Li W, Li X. Optimized NRBO-VMD-AM-BiLSTM Hybrid Architecture for Enhanced Dissolved Gas Concentration Prediction in Transformer Oil Soft Sensors. *Sensors*. 2025;25(16).
- [23] He J, Ma C, Wang S, Zhai Y, Tang Q. Entropy-Driven Adaptive Decomposition and Linear-Complexity Score Attention: An AI-Powered Framework for Crude Oil Financial Market Forecasting. *Entropy*. 2026;28(4).
- [24] Li J, Lv D, Dong B, Xi K. A Study on the Universality of a Variational Mode Parameter Optimization Method Based on an Improved Hippopotamus optimization Algorithm. New York, NY, USA: Association for Computing Machinery; 2025. Available from: <https://doi.org/10.1145/3704558.3707060>.
- [25] Lu J, Yue J, Zhu L, Wang D, Li G. An improved variational mode decomposition method based on the optimization of salp swarm algorithm used for denoising of natural gas pipeline leakage signal. *Measurement*. 2021;185:110107. <https://doi.org/10.1016/j.measurement.2021.110107>.
- [26] Zhou J, Guo X, Wang Z, Du W, Wang J, Han X, et al. Research on fault extraction method of variational mode decomposition based on immunized fruit fly optimization algorithm. *Entropy*. 2019;21(4):400. <https://doi.org/10.3390/e21040400>.
- [27] Ye M, Yan X, Jia M. Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM. *Entropy*. 2021;23(6):762. <https://doi.org/10.3390/e23060762>.
- [28] Liang T, Lu H, Sun H. Application of parameter optimized variational mode decomposition method in fault feature extraction of rolling bearing. *Entropy*. 2021;23(5):520. <https://doi.org/10.3390/e23050520>.

- [29] Lyu G, Luo C, Wu S, Wang C, Ma W, Yang M, et al. An enhanced variational mode decomposition method for processing hydrodynamic data of underwater gliders. *Measurement*. 2025;244:116468. <https://doi.org/https://doi.org/10.1016/j.measurement.2024.116468>.