mathematics

MDPI

Article

MOECSO-Based Framework for Crude Oil Price Forecasting

Lihong Zhao [1,2], Zhihui Chen [2], Naiqi Wu [1] and Liping Bai [1,*]

[1] Department of Engineering Science, Macau University of Science and Technology, Macao 999078, China
[2] School of Intelligent and Connected Systems, Guangzhou City Polytechnic, Guangzhou 511370, China
* Correspondence: lipbai@must.edu.mo

Abstract

Multi-model ensembles and multi-objective evolutionary algorithms provide a systematic approach to reconciling competing criteria in time-series forecasting. However, most existing methods are tailored to specific tasks and lack essential mathematical details. This study introduces a general multi-objective ensemble framework based on a Multi-Objective Enhanced Crisscross Optimization (MOECSO) algorithm, exemplified through Brent crude oil price forecasting. Initially, ensemble-weight selection is framed as a bi-objective optimization problem, where the two objectives penalize Mean Absolute Error (MAE) and the Sample Standard Deviation of the Validation Residuals (SSDVRs), both assessed on the original United States Dollar (USD) scale under a leakage-free rolling-origin protocol. Subsequently, a Variational Mode Decomposition (VMD) reconstruction operator is defined, which adaptively reconstructs the raw series by integrating intrinsic mode functions with weights derived from their entropy and center-frequency characteristics, while adhering to nonnegativity and normalization constraints. Furthermore, horizontal and vertical crossover operators, along with a hypervolume–ideal-distance archive rule, are introduced, collectively forming a comprehensive MOECSO scheme for bi-objective ensemble weighting. Utilizing a public Brent crude oil dataset, the proposed ensemble demonstrates superior performance compared to robust statistical, machine-learning, and deep-learning benchmarks in terms of MAE, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), while also reducing error dispersion and enhancing robustness during crisis periods. Diebold–Mariano (DM) and superior predictive ability tests with multiple-comparison control validate that these improvements are statistically significant. In summary, this paper presents a mathematically transparent framework for constructing and analyzing multi-objective ensembles in univariate time-series forecasting.

Keywords: VMD; ensemble learning; crisscross; multi-objective optimization; crude oil price forecasting

MSC: 68T01; 68T07; 68T09

Check for updates

Academic Editor: Raymond Lee

Received: 30 January 2026
Revised: 21 February 2026
Accepted: 25 February 2026
Published: 27 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Crude oil price serves as a crucial indicator of energy security and the macroeconomy. Long-term price dynamics reflect shifts in supply and demand, geopolitical developments, financial cycles, and the energy transition [1]. The interplay among these factors renders the oil-price series complex, nonlinear, and multi-scale [2]. Consequently, this study aims to develop a forecasting model assessed on an extensive dataset, which achieves high predictive accuracy and robust out-of-sample performance amid multiple simultaneous sources of uncertainty [3]. Enhanced forecasts, evaluated over a lengthy historical period,

can assist governments in optimizing energy security and macroeconomic policies, enable energy firms and financial institutions to hedge and manage risk, and provide more reliable inputs for planning related to the green transition and emission reduction.

Most studies adopt one of the two primary strategies. The first strategy employs linear statistical models, such as the AutoRegressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), which assume quasi-stationarity and employ simple structures to characterize trends and volatility [4]. While these models are interpretable and straightforward to implement, they exhibit poor performance in the presence of structural breaks and significant nonstationarity, resulting in accumulated bias in noisy and multiscale environments [5]. The second strategy involves nonlinear AI models, including Back Propagation (BP), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM), which leverage flexible function classes to capture complex dynamics and typically achieve reduced in-sample error. However, their extensive parameter spaces heighten sensitivity to noise and scale mixing, thereby compromising out-of-sample stability [6]. Recent findings indicate that forecast stability, defined as the sensitivity of forecasts to minor changes in the information set or estimation window, represents a distinct criterion from average accuracy. This distinction underscores the need for explicit dispersion and stability objectives in rolling-origin evaluation [7]. A prevalent compromise integrates decomposition techniques, such as Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), wavelets, and Variational Mode Decomposition (VMD), with these forecasting models in a decomposition–prediction–reconstruction pipeline [8]. In numerous studies, reconstruction continues to rely on fixed or heuristic fusion rules, model selection is based on single-objective criteria, and the maintenance of the Pareto front is not explicitly addressed, collectively limiting the robustness and reproducibility.

To overcome the limitations in signal reconstruction, model search, and Pareto-front maintenance, we reorganize three core components of the decomposition–ensemble workflow. First, at the objective level, ensemble-weight selection is formulated as a bi-objective optimization problem that minimizes both average forecast error and error dispersion under a leakage-free rolling-origin protocol, thereby establishing a robust mathematical foundation for subsequent design. Second, at the reconstruction level, we define a VMD-based operator that adaptively reconstructs a raw time series by integrating intrinsic mode functions with weights derived from entropy and center frequency, while adhering to non-negativity and normalization constraints. Third, at the evolutionary-search level, we propose an elite-guided Crisscross Optimization method that employs tailored horizontal and vertical crossovers on weight vectors, along with a hybrid archive criterion that combines hypervolume (HV [9]) with distance to an ideal point, to maintain a well-distributed Pareto front. All components are presented in a fully explicit form. The main contributions of this paper are summarized as follows.

- An adaptive VMD-based reconstruction operator is proposed, which maps a normalized input series to a reconstructed signal by integrating Intrinsic Mode Functions (IMFs). This integration employs weights derived from entropy-based relevance scores and normalized center frequencies, adhering to non-negativity and unit-sum constraints. Each IMF is characterized by a brief vector of statistical indicators, which includes the absolute Pearson and Spearman correlations with the target, the maximal information coefficient, and the energy ratio. These indicator vectors undergo min-max normalization across IMFs and are combined using entropy weighting to generate relevance scores. Subsequently, the relevance scores are refined through a frequency penalty and normalized to produce the final fusion weights. This operator

is model-agnostic and can function as a generic pre-processing module within other decomposition-ensemble frameworks.

- Elite-guided horizontal and vertical crossover operators for ensemble weighting are introduced. These two crossover operators directly manipulate simplex-constrained ensemble-weight vectors. The horizontal operator generates convex combinations of the current solution, a peer, and a Pareto-elite individual. In contrast, the vertical operator updates a limited subset of weight dimensions to enhance promising patterns. Both operators incorporate annealed Gaussian perturbations, ensure feasibility through non-negativity enforcement and unit-sum normalization, and promote information exchange between elite and non-elite regions of the population throughout the search process.
- Hybrid hypervolume–ideal-distance archive rule. A novel hybrid archive maintenance rule is introduced that concurrently considers HV to enhance coverage and diversity, as well as the distance to the ideal point to improve convergence. The resulting archive provides a well-distributed, high-quality approximation of the Pareto front. This update integrates a fixed-reference-point strategy and explicit termination criteria to ensure stable selection and reproducible optimization.

The rest of this work is arranged as follows. In Section 2, the methodological foundations relevant to this study are reviewed, including decomposition–ensemble forecasting pipelines, ensemble-weight learning and forecast combination, and multi-objective evolutionary optimization with Pareto-front maintenance. In Section 3, the proposed methodology is described in detail, covering the notation and problem formulation, the entropy–frequency VMD reconstruction operator, and the Multi-Objective Enhanced Crisscross Optimization (MOECSO) algorithm with elite-guided horizontal and vertical crossover operators and an archive update scheme based on a hybrid elite criterion. In Section 4, the experimental design is specified, including the Brent dataset, the leakage-free rolling-origin protocol, evaluation metrics, statistical tests, competing baselines, and the definition of stress-period sub-samples. In Section 6, the empirical results are presented and discussed, with emphasis on full-sample performance, robustness under market stress, and supporting statistical evidence. In Section 7, the main findings are summarized, and directions for future research are outlined.

2. Related Work

2.1. Decomposition–Ensemble Forecasting as a Structured Pipeline

Nonstationarity and multi-scale behavior are prevalent in economic and energy time series [10]. Common methodologies for crude oil price forecasting encompass wavelet-based decompositions, as well as variants of EMD and CEEMDAN, and VMD, followed by component-wise modeling and subsequent aggregation [11]. Although these approaches demonstrate their efficacy in practice, they generally treat decomposition and reconstruction as modular heuristics, rendering them sensitive to design choices such as mode selection, scale cutoffs, and arbitrary denoising rules. Huang et al. propose a VMD–EMD–Transformer pipeline for crude oil price forecasting problems, incorporating a second decomposition stage to address residual components prior to sequence modeling [12]. Wang et al. develop an ensemble-driven LSTM framework based on CEEMDAN decomposition and meta-heuristic optimization, reporting enhanced robustness in crude oil price forecasting [13].

2.2. Ensemble-Weight Learning for Forecast Combination

Forecast combination that combines various methods has a rich history, encompassing methods ranging from simple averaging and linear regression-based stacking to Bayesian

model averaging and various data-driven weighting rules [14]. A prevalent method-combined forecast strategy involves estimating weights within a validation window to enhance the accuracy and robustness compared to individual-based models. Recent studies investigate adaptive and time-varying weights that are updated across rolling origins to monitor regime shifts and mitigate model drift [15]. A recent study on crude oil price forecasting by Liu et al. develops a robust time-varying weight combined forecasting model, demonstrating improved robustness during critical periods such as COVID-19, consistent with per-origin weight updates under rolling evaluation [16]. Despite these advancements, several practical challenges persist: (1) preventing the overfitting of weights to short validation segments, (2) accommodating heterogeneous base models with varying bias–variance profiles, and (3) establishing selection criteria that yield stable combinations in noisy, nonstationary market conditions [17].

2.3. Multi-Objective Evolutionary Search and Pareto-Front Maintenance

Multi-objective evolutionary algorithms, including Non-dominated Sorting Genetic Algorithm II/III (NSGA-II/III), Strength Pareto Evolutionary Algorithm 2 (SPEA2), and Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D), provide systematic approaches to balance competing criteria and generate sets of Pareto-optimal trade-offs. Recent variants enhance elitism and sorting techniques to improve both convergence and diversity preservation. For example, an elitist non-dominated sorting crisscross algorithm has been proposed, demonstrating performance improvements on benchmark multi-objective suites [18]. In forecasting applications, multi-objective formulations are often motivated by the need to balance average predictive accuracy with stability or robustness criteria that quantify error variability or stabilization behavior. This motivation appears in recent forecasting studies that incorporate stability-aware weighting or variance-stabilization diagnostics [19]. In multi-objective evolutionary search, archive update and elite selection are highly design-dependent components, and recent work highlights the diverse roles of archives and the need to customize update and truncation rules for different goals and stages [20]. This design freedom is closely related to the interplay between diversity preservation and convergence promotion. The HV is frequently used to encourage well-spread non-dominated sets, while distance-to-ideal measures are commonly adopted to quantify convergence toward the ideal solution. However, their combined use is rarely specified within a unified, schedule-driven, and fully reproducible selection framework [21].

3. Methods

3.1. Notation and Problem Setting

Let $\{y_t\}_{t=1}^{T}$ denote a univariate time series, representing daily Brent crude oil spot prices in USD. One-step-ahead forecasting is evaluated under a rolling-origin walk-forward protocol. For each origin τ, a rolling window of length W is used, and the inner validation set is $\mathcal{V}_\tau = \{t : \tau - V + 1 \le t \le \tau\}$, while the standardization statistics are computed on $F_\tau = \{t : \tau - W + 1 \le t \le \tau - V\}$ so that $\mathcal{V}_\tau$ is excluded from fitting-time normalization. Forecasts on $\mathcal{V}_\tau$ are inverse-transformed to USD before residual evaluation.

Let $\hat{y}_t^{(m)}$ be the one-step-ahead forecast of base model m using information up to time $t-1$, for $m = 1, \ldots, M$. Collect base forecasts as $\hat{\mathbf{y}}_t = \left(\hat{y}_t^{(1)}, \ldots, \hat{y}_t^{(M)}\right)^\top \in \mathbb{R}^M$. Ensemble fusion is conducted on the probability simplex $\Delta^M = \{\mathbf{w} \in \mathbb{R}^M : w_m \ge 0, \sum_{m=1}^{M} w_m = 1\}$. For any $\mathbf{w} \in \Delta^M$ and $t \in \mathcal{V}_\tau$, with horizon $h = 1$ and origin $\tau = t - 1$, define

$$\hat{y}_t^{\text{ens}}(\mathbf{w}) = \sum_{m=1}^{M} w_m \hat{y}_t^{(m)}, \qquad e_t(\mathbf{w}) = y_t - \hat{y}_t^{\text{ens}}(\mathbf{w}). \tag{1}$$

Two objectives are jointly minimized on Δ^M using residuals on the USD scale:

$$f_1(\mathbf{w}) = \frac{1}{|\mathcal{V}_\tau|} \sum_{t \in \mathcal{V}_\tau} |e_t(\mathbf{w})|, \qquad f_2(\mathbf{w}) = \sqrt{\frac{1}{|\mathcal{V}_\tau| - 1} \sum_{t \in \mathcal{V}_\tau} \Big(e_t(\mathbf{w}) - \bar{e}(\mathbf{w})\Big)^2}, \tag{2}$$

where $\bar{e}(\mathbf{w}) = \frac{1}{|\mathcal{V}_\tau|} \sum_{t \in \mathcal{V}_\tau} e_t(\mathbf{w})$. Here, f_1 is MAE and f_2 is the sample standard deviation of validation residuals, both in USD.

The Pareto set is

$$\mathcal{P}^* = \Big\{ \mathbf{w} \in \Delta^M : \quad \nexists\, \mathbf{w}' \in \Delta^M \text{ with } f_i(\mathbf{w}') \le f_i(\mathbf{w})\ \forall i \in \{1,2\} \text{ and } f_j(\mathbf{w}') < f_j(\mathbf{w}) \text{ for some } j \Big\}. \tag{3}$$

Proposition 1. *If $\mathcal{V}_\tau$ is finite, then the bi-objective optimization over Δ^M admits at least one Pareto-optimal solution, hence $\mathcal{P}^*$ is non-empty.*

Proof. The feasible set Δ^M is non-empty, closed, and bounded, hence compact in $\mathbb{R}^M$. For finite $\mathcal{V}_\tau$, each $e_t(\mathbf{w})$ is affine in $\mathbf{w}$ and both f_1 and f_2 are continuous on Δ^M. A standard existence result for Pareto minimizers of continuous vector objectives over compact feasible sets implies that at least one Pareto-optimal solution exists [22]. □

Proposition 2. *For any $\mathbf{w} \in \Delta^M$, $\hat{y}_t^{\text{ens}}(\mathbf{w})$ is a convex combination of $\{\hat{y}_t^{(m)}\}_{m=1}^M$. Moreover, the absolute error satisfies*

$$|e_t(\mathbf{w})| \le \sum_{m=1}^{M} w_m\, |e_t^{(m)}|, \qquad e_t^{(m)} = y_t - \hat{y}_t^{(m)}. \tag{4}$$

Averaging over $t \in \mathcal{V}_\tau$ yields

$$\text{MAE}(\mathbf{w}) \le \sum_{m=1}^{M} w_m\, \text{MAE}_m. \tag{5}$$

In contrast, MAPE involves a time-varying normalization term and therefore is not written as an analogous linear upper bound; instead, it is evaluated using the stabilized denominator $|y_t| + \varepsilon$.

3.2. Overall Framework

Figure 1 presents a three-stage forecasting pipeline assessed using a leakage-free rolling-origin protocol. For each origin, all preprocessing, decomposition, scoring, and model fitting occur exclusively within the training window, while the held-out fold is designated solely for out-of-sample evaluation.

- Data processing and reconstruction

  The Brent price series is represented by y_t. The training window is standardized, and the VMD technique is utilized to derive K IMFs, arranged based on their center frequencies. Subsequently, an innovation, denoted as Innovation #1, is implemented, introducing an entropy-frequency reconstruction operator. This operator assigns a relevance score to each IMF by employing a multi-indicator assessment that integrates entropy-based aggregation and a frequency regularization term. The resultant non-negative weights are normalized and utilized for the reconstruction of a singular input signal $x_t^{(\text{rec})}$ in the standardized domain.
- Heterogeneous base predictors

  After reconstructing the input series $x_t^{(\text{rec})}$, a diverse set of base models is trained. Feature augmentation, like differencing or returns, is selectively applied to certain

learning-based models, as illustrated by dashed links in Figure 1, while ARIMA is fitted without augmentation. Each base learner generates one-step-ahead forecasts at every origin, collectively forming a forecast vector.

- Bi-objective simplex ensemble via MOECSO

  The final forecast is derived by optimizing the convex-combination weights $\mathbf{w} \in \Delta^M$ across the base forecasts. The validation objectives include (1) MAE and (2) SSDVR. The proposed MOECSO solver consists of two primary components. The first component is an elite-guided horizontal and vertical crossover operator (Innovation #2), while the second component is an external-archive maintenance mechanism that utilizes a hybrid elite criterion (Innovation #3). The update process is executed through non-dominated sorting, followed by scoring, filling, and truncation. The algorithm generates a Pareto archive of candidate weight vectors, from which a single operational solution is chosen to produce the ensemble forecast.

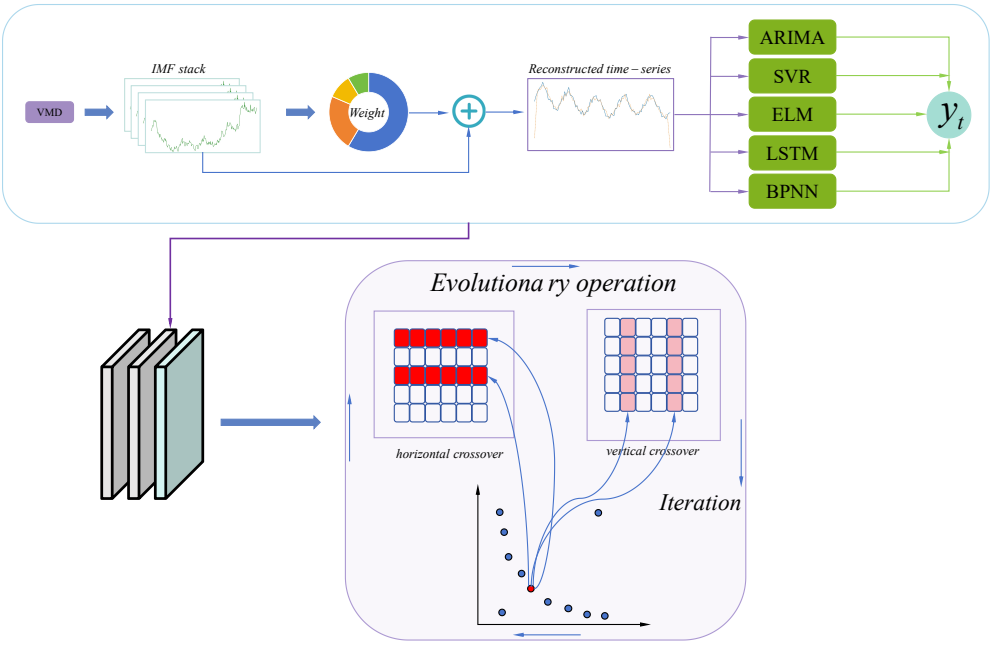


Figure 1. Framework of the proposed forecasting pipeline. The raw series is standardized and decomposed by VMD. IMFs are adaptively reweighted to reconstruct an informative input sequence, which is then fed into multiple base forecasters, including ARIMA, SVR, ELM, LSTM, and BPNN. An evolutionary search with horizontal and vertical crossover performs Pareto-guided weight fusion, while the external archive is updated using the Hypervolume Ideal Distance hybrid elite criterion to balance convergence and coverage.

3.3. Adaptive VMD Reconstruction Based on Multi-Indicator Fusion

This subsection formalizes the adaptive VMD reconstruction used to generate the input series for all the base models. Figure 2 illustrates the overall entropy–frequency adaptive reconstruction pipeline. VMD is applied to the standardized input series $\{z_t\}_{t=1}^{T}$ to extract K band-limited components $\mathrm{IMF}_k(t)$, ordered by their center frequencies ω_k from low to high. For each IMF, a compact set of relevance indicators is computed and rescaled across k. These indicators are then fused via entropy-based weighting to obtain a per-IMF relevance score s_k. To reduce the contribution of noisy high-frequency modes, a frequency-aware penalty p_k is further introduced based on the normalized center frequency $\widehat{f}_k$. The preliminary score is defined as $v_k = s_k p_k$. The scores are clipped at zero and

normalized to sum to one, yielding fusion weights $\{w_k\}_{k=1}^{K}$ [23]. The reconstructed signal used in the subsequent stages is the resulting convex combination in Equation (6):

$$x_t^{(\mathrm{rec})} = \sum_{k=1}^{K} w_k \, \mathrm{IMF}_{k,t} \tag{6}$$

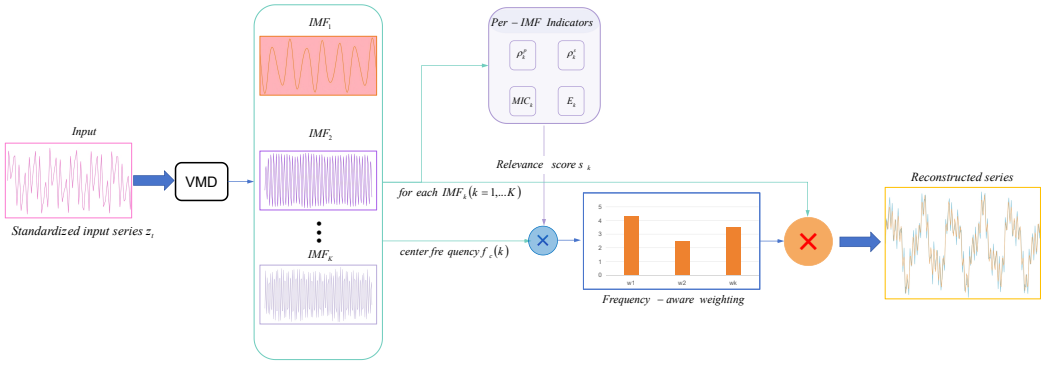


Figure 2. Entropy–frequency adaptive VMD reconstruction. VMD decomposes the standardized series into K IMFs with center frequencies $f_c(k)$. Per-IMF indicators are fused via entropy weighting to obtain s_k, combined with a frequency-aware penalty to form simplex-normalized weights ($w_k \geq 0$, $\sum_k w_k = 1$), and the final series is reconstructed by $x_t^{(\mathrm{rec})} = \sum_k w_k \, \mathrm{IMF}_{k,t}$.

3.3.1. VMD Decomposition

VMD is applied to the standardized input series $\{z_t\}_{t=1}^{T}$ to extract K band-limited components $\{\mathrm{IMF}_k(t)\}_{k=1}^{K}$ such that the original series can be represented by their superposition:

$$z_t \approx \sum_{k=1}^{K} \mathrm{IMF}_{k,t}, \qquad t = 1, \ldots, T \tag{7}$$

where $\mathrm{IMF}_{k,t}$ denotes the k-th mode at time t and K is the number of modes. The components are ordered from low to high according to their center frequencies ω_k. Following the decomposition in Equation (7), a compact set of relevance indicators is computed for each IMF and rescaled across k before fusion.

3.3.2. Indicators

Four complementary relevance indicators are computed for each mode IMF_k with respect to the standardized reference series $\{z_t\}_{t=1}^{T}$:

- **Pearson correlation**

$$\rho_k^P \triangleq \frac{\mathrm{cov}(\mathrm{IMF}_k, z)}{\sigma_{\mathrm{IMF}_k} \sigma_z} \in [-1, 1] \tag{8}$$

  where, the sample covariance $\mathrm{cov}(\cdot, \cdot)$ and sample standard deviation σ in Equation (8) are computed over $t = 1, \ldots, T$.
- **Spearman correlation**

$$\rho_k^S \triangleq \rho^P\big(\mathrm{rank}(\mathrm{IMF}_k), \, \mathrm{rank}(z)\big) \in [-1, 1] \tag{9}$$

  where the Pearson correlation in Equation (9) is applied to the ranked samples. Ties are handled by standard average ranking in practice.
- **Maximum Information Coefficient [24]**

$$\mathrm{MIC}_k \triangleq \frac{I(\mathrm{IMF}_k; z)}{\min\big\{H(\mathrm{IMF}_k), \, H(z)\big\}} \in [0, 1] \tag{10}$$

where $I(\cdot;\cdot)$ is the empirical mutual information and $H(\cdot)$ is the empirical entropy computed from discretized samples in Equation (10). MIC captures potentially nonlinear dependence. A mild upward bias may occur when the sample size is small.

- **Energy ratio [25]**

$$E_k \triangleq \frac{\|\mathrm{IMF}_k\|_2^2}{\sum_{\ell=1}^{K} \|\mathrm{IMF}_\ell\|_2^2} \in [0,1], \qquad \sum_{k=1}^{K} E_k = 1 \tag{11}$$

where the energy ratio E_k in Equation (11) measures the relative energy contribution of IMF_k.

3.3.3. Multi-Indicator Fusion via Entropy Weighting

To treat positive and negative dependencies symmetrically, absolute correlations are used, and four raw indicators for each mode IMF_k are defined as $u_k^{(1)} = |\rho_k^P|$, $u_k^{(2)} = |\rho_k^S|$, $u_k^{(3)} = \mathrm{MIC}_k$, and $u_k^{(4)} = E_k$. Because these indicators have different scales, each indicator $u_k^{(j)}$ is min–max normalized across k to the unit interval $[0,1]$ before fusion. Indicators that are constant across IMFs are excluded from weighting to avoid degenerate normalization.

Entropy-based weights are then computed to emphasize indicators that provide higher cross-IMF discrimination [26]. For each indicator j, a probability mass function over IMFs is formed in Equation (12) as

$$p_k^{(j)} = \frac{u_k^{(j)} + \varepsilon}{\sum_{\ell=1}^{K} \left(u_\ell^{(j)} + \varepsilon\right)}, \qquad \varepsilon = 10^{-12} \tag{12}$$

The entropy and the corresponding entropy weight are defined by

$$H_j = -\sum_{k=1}^{K} p_k^{(j)} \log p_k^{(j)}, \qquad \xi_j = \frac{1 - H_j}{\sum_{h \in \mathcal{J}} (1 - H_h)} \tag{13}$$

where $\mathcal{J} \subseteq \{1, \ldots, 4\}$ denotes the set of non-constant indicators after normalization in Equation (13). The fused relevance score of IMF_k is finally computed as

$$s_k = \sum_{j \in \mathcal{J}} \xi_j \, u_k^{(j)} \in [0,1] \tag{14}$$

where, if $\mathcal{J}$ is empty, s_k in Equation (14) is set to $\frac{1}{K}$ to allocate relevance uniformly across IMFs.

3.3.4. Frequency-Aware Scoring and Reconstruction

To suppress noisy high-frequency modes while retaining informative components, we combine the relevance score $s_k \in [0,1]$ with a frequency-aware penalty derived from the center frequency of each IMF_k. The center frequency is estimated in practice by the spectral centroid of the one-sided power spectrum:

$$f_c(k) = \frac{\sum_{f \geq 0} f \left|\widehat{\mathrm{IMF}}_k(f)\right|^2}{\sum_{f \geq 0} \left|\widehat{\mathrm{IMF}}_k(f)\right|^2} \tag{15}$$

where, $\widehat{\mathrm{IMF}}_k(f)$ in Equation (15) is obtained by FFT, and mirror padding can be applied to mitigate boundary effects.

The frequencies are normalized across modes within each rolling origin so that the normalized values lie in $[0, 1]$, and a monotonically decreasing exponential penalty is used. The relevance–frequency fusion score in Equation (16) is then defined by

$$v_k = s_k \exp\big(-\beta \widehat{f}_k\big), \qquad \beta > 0, \quad k = 1, \ldots, K \tag{16}$$

with the smallest-frequency mode exempted by setting its penalty to one. Finally, v_k is truncated at zero and normalized to obtain nonnegative reconstruction weights $\{w_k\}_{k=1}^{K}$ with $\sum_{k=1}^{K} w_k = 1$, and the reconstructed series in Equation (17) is

$$x_t^{(\mathrm{rec})} = \sum_{k=1}^{K} w_k \, \mathrm{IMF}_{k,t} \tag{17}$$

3.3.5. Complexity and Implementation

All quantities used in the reconstruction, including s_k and the frequency-based penalty, are computed on the fit segment F_τ at each rolling origin to avoid any leakage into the inner validation fold. The penalty strength is fixed at $\beta = 5$ for all rolling origins, and the test fold is strictly excluded from any hyperparameter selection.

In terms of computational cost, Pearson and Spearman correlations and the energy ratio scale linearly with the segment length, yielding $O(KN)$ per rolling origin for K modes and N samples. Computing MIC for all modes scales as $O\big(K \cdot \mathrm{MIC}(N)\big)$ and is empirically close to $O(KN \log N)$ in our implementation. For the Brent experiments, the per-origin training window length is $W = 8000$, making MIC tractable relative to the FFT-based operations required by the VMD-related steps.

Definition 1. *The operator R_θ for reconstructing the entropy–frequency VMD is defined as follows:*

$$R_\theta : \{y_t\} \mapsto \{x_t^{(\mathrm{rec})}\} \tag{18}$$

where the mapping in Equation (18) is implemented by first decomposing the standardized input series z_t into VMD components $\{\mathrm{IMF}_k(t)\}_{k=1}^{K}$, each accompanied by a characteristic center frequency. Each IMF is associated with a relevance score s_k and a frequency penalty p_k. To reconstruct the series z_t, the process involves calculating $\mathrm{IMF}_k(t)$, s_k, and p_k, merging relevance and frequency details into initial scores, transforming them into positive unit-sum weights by clipping and renormalization, and ultimately creating the reconstruction $x^{(\mathrm{rec})}(t)$ through the weighted summation of the VMD components.

Proposition 3. *The reconstruction induced by R_θ under simplex normalization is feasible and continuous. Given VMD outputs $\{\mathrm{IMF}_k(t)\}_{k=1}^{K}$, we define the fusion weights by Equation (19) with $\varepsilon > 0$:*

$$\pi_k(v) = \frac{\max\{v_k, 0\}}{\sum_{i=1}^{K} \max\{v_i, 0\} + \varepsilon}, \qquad k = 1, \ldots, K \tag{19}$$

where $\pi_k(v) \geq 0$ for all k. Moreover, the weights satisfy the unit-sum constraint in Equation (20), and, therefore, $\pi(v) \in \Delta^K$:

$$\sum_{k=1}^{K} \pi_k(v) = 1 \tag{20}$$

Furthermore, the mapping $v \mapsto \pi(v)$ is continuous on $\mathbb{R}^K$ because it is a composition of continuous operations and its denominator is bounded away from zero by $\varepsilon > 0$. Consequently, the reconstruction in Equation (21) varies continuously with respect to v:

$$x^{(\mathrm{rec})}(t) = \sum_{k=1}^{K} \pi_k(v)\,\mathrm{IMF}_k(t) \tag{21}$$

Remark 1. *The interpretation of R_θ as a data-driven low-pass reconstruction suggests that it functions as a filter preserving a blend of VMD components without imposing strict frequency cut-offs or relying on a single indicator. The weights w_k are determined by both the statistical significance indicated by entropy-based scores s_k and the normalized frequency denoted by p_k. Consequently, high-frequency elements are de-emphasized only when they exhibit weak correlation with the target, leading to more reliable reconstructions amidst the presence of noisy high-frequency elements. This approach is independent of any specific model and can serve as a general pre-processing step in various decomposition-ensemble frameworks.*

3.4. Bi-Objective Ensemble-Weight Optimization via MOECSO

At each rolling origin τ, ensemble-weight selection is formulated as a bi-objective optimization problem over the simplex Δ^M. Over the inner validation set $\mathcal{V}_\tau$, the ensemble forecast is obtained as a convex combination of the base one-step-ahead forecasts, and the validation residual is defined as the difference between the observed price and the ensemble forecast on the original USD scale. The two objectives jointly minimize (1) MAE and (2) SSDVR, thereby balancing accuracy and stability.

MOECSO updates candidate weights on the simplex through two elite-guided crossover operators. The horizontal crossover facilitates whole-vector recombination, promoting extensive exploration of Δ^M. In contrast, the vertical crossover modifies only a small subset of coordinates, enabling fine-grained local refinement around the promising solutions. Objective evaluation follows the notation and problem setting defined above, and standardization is applied solely for base-model fitting within each origin.

To approximate the Pareto front, an external archive is maintained and updated using a hybrid criterion that balances diversity and convergence, measured by HV and the distance to the ideal point, respectively. An annealed schedule emphasizes HV in early iterations to enhance coverage and reduce premature convergence, and gradually shifts toward the ideal-point distance in later iterations to strengthen convergence. In terms of computational cost, both crossover and simplex repair incur an $O(M)$ cost per individual, leading to an $O(NM)$ cost per generation for a population size of N.

3.4.1. Horizontal Crossover with Elite Guidance

The new candidate is formulated as an elite-guided linear recombination of the current solution, a randomly selected peer, and a Pareto-elite individual, augmented by Gaussian noise. When $\alpha_E = 0$, the operator is simplified as a standard horizontal crossover [27], which exclusively mixes information between w and w^{peer}. Introducing a positive α_E provides explicit elite guidance, directing each update toward solutions along the current Pareto front while maintaining contributions from both the incumbent and the peer. This elite-guided crossover enhances convergence to high-performing regions of the weight space, mitigates the risk of deviating from the front when objectives are noisy, and sustains diversity through the inclusion of peer and noise components. A schematic illustration of the elite-guided crossover operators is provided in Figure 3. The horizontal update is defined as

$$\tilde{w} = w + \beta_H(w^{\mathrm{peer}} - w) + \alpha_E(e - w) + \sigma_t \varepsilon \tag{22}$$

where $\beta_H \in [0, 1]$ in Equation (22) regulates horizontal mixing and is independent of the VMD frequency-penalty parameter β.

Meanwhile, to balance exploration and exploitation, a progress variable $p = \frac{g}{T_{\max}-1} \in [0,1]$ is introduced, which increases linearly with the generation counter as follows:

$$\alpha_E = \alpha_{E,0} p, \quad \sigma_t = \sigma_0 (1-p) \tag{23}$$

where $g = 0, 1, \ldots, T_{\max} - 1$ in Equation (23) represents the generation index. Here, $\alpha_{E,0}$ denotes the maximum strength of elite guidance, and σ_0 determines the initial magnitude of Gaussian noise. Initially, the noise is higher and elite guidance is weaker in the early iterations, whereas in later iterations the noise decreases gradually and elite guidance strengthens.

As a safeguard, a mild elementwise clipping step truncates each component of $\tilde{w}$ to the interval $[-5, 5]$ before enforcing non-negativity and unit-sum normalization, which mitigates rare numerical spikes when M is large or when previous weights are extreme while remaining sufficiently loose to leave typical perturbations essentially unaffected. The horizontal update performs full-vector recombination and applies the same simplex-repair procedure via non-negativity enforcement and unit-sum normalization, yielding a computational complexity of $O(M)$ per individual and $O(NM)$ per generation.

3.4.2. Vertical Crossover with Elite Guidance

Complementing horizontal crossover that promotes global exploration, vertical crossover [27] performs localized refinement by modifying only a small subset of dimensions while preserving the overall structure of the current solution, which is particularly effective in later iterations when subtle weight adjustments can escape shallow local optima more reliably than large global moves (see Figure 3). Let $w \in \Delta^M$ denote the current solution. Two distinct indices $i \neq j$ are uniformly sampled from $\{1, \ldots, M\}$ without replacement, and an elite e is uniformly drawn from the Pareto front F_1 excluding w. With $r \sim U(0,1)$ and $\varepsilon_i, \varepsilon_j \sim N(0,1)$, only the selected dimensions are updated as follows in Equation (24):

$$\begin{aligned} \tilde{w}_i &= w_i + r(w_j - w_i) + \alpha_E(e_i - w_i) + \sigma_t \varepsilon_i \\ \tilde{w}_j &= w_j + r(w_i - w_j) + \alpha_E(e_j - w_j) + \sigma_t \varepsilon_j \end{aligned} \tag{24}$$

For all other indices $m \notin \{i, j\}$, $\tilde{w}_m = w_m$, so each move perturbs only two coordinates. The coefficient r induces a swap-like linear recombination between the selected components, the elite-guidance term $\alpha_E(e - w)$ pulls them toward the elite solution, and the annealed noise term $\sigma_t \varepsilon$ improves robustness against premature convergence. To mitigate rare numerical spikes that may occur when M is large or when the previous weights are extreme, $\tilde{w}$ is first clipped elementwise to the interval $[-5, 5]$, and then repaired to satisfy feasibility by enforcing non-negativity followed by unit-sum normalization. Although the raw update touches only a constant number of coordinates, the simplex-repair step requires $O(M)$ time per individual, leading to an $O(NM)$ cost per generation for a population of size N.

3.4.3. Archive Update and Selection

This subsection describes how the Pareto archive is refreshed at each generation to retain a compact, non-redundant set of competitive solutions. At each generation, the previous archive A is merged with the offspring set O produced by the horizontal and vertical operators to form the candidate set $C = A \cup O$. To remove redundancy, near-identical duplicates under floating-point tolerance are screened in the objective space using an infinity-norm tolerance: letting $\mathbf{f}(\mathbf{w}) = (f_1(\mathbf{w}), f_2(\mathbf{w}))$ denote the bi-objective vector, if $\|\mathbf{f}(x) - \mathbf{f}(y)\|_\infty \leq \epsilon_{\text{dup}}$ with $\epsilon_{\text{dup}} = 10^{-9}$ by default, only one representative is retained.

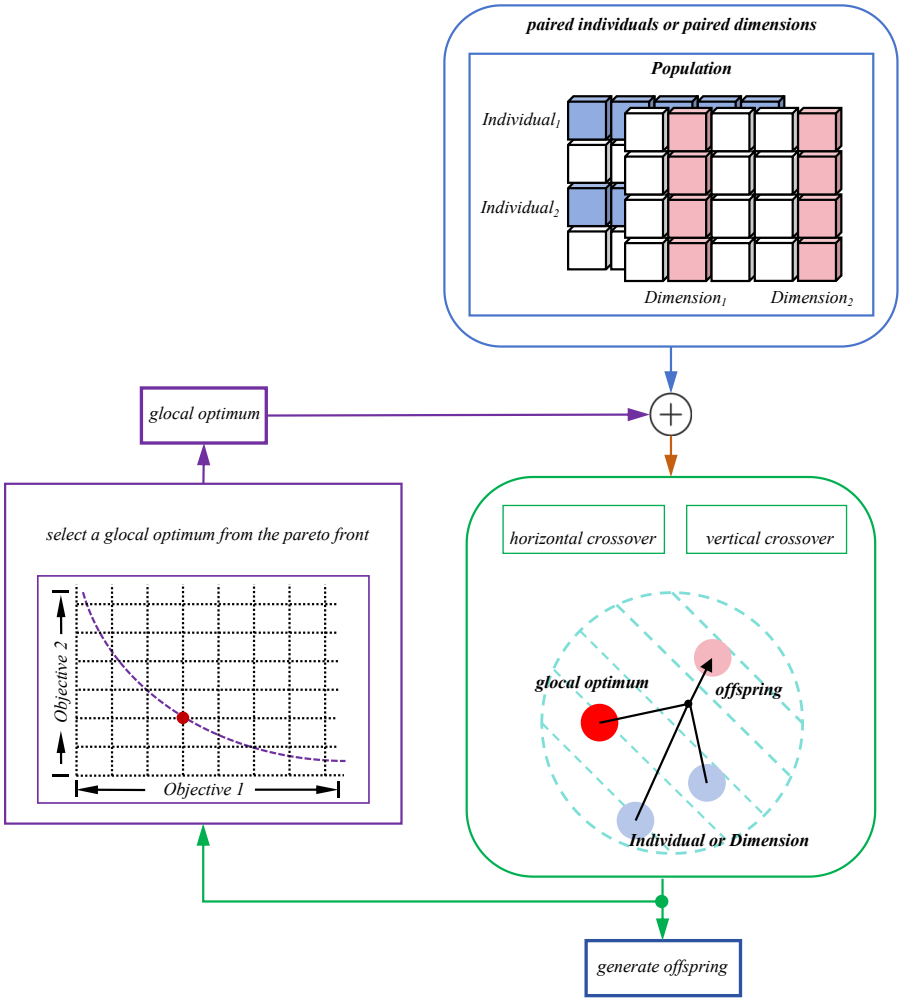


Figure 3. Schematic of elite-guided crossover in MOECSO for ensemble-weight optimization. At each generation, an elite solution is selected from the current Pareto archive (global optimum under the adopted criterion) and used to guide two operators: horizontal crossover (whole-vector recombination) and vertical crossover (partial coordinate update). The resulting offspring are repaired onto the simplex to satisfy non-negativity and unit-sum constraints.

The core of the archive update is the hybrid elite criterion, which ranks candidates by combining two measures: HV to promote diversity, and distance-to-ideal to promote convergence. Figure 4 provides a schematic illustration of this hypervolume–ideal-distance hybrid elitism and the role of the adaptive mixing weight η_g. For each objective $i \in \{1, 2\}$, the HV reference point and the ideal point with safety margins are defined as $r_i = \max_{w \in C} f_i(w) + \delta_r$ and $z_i^{\text{ideal}} = \min_{w \in C} f_i(w) - \delta_{\text{ideal}}$, where $\delta_r > 0$ and $\delta_{\text{ideal}} > 0$ are both set to 10^{-6}. Let $r = (r_1, r_2)$ and $z^{\text{ideal}} = (z_1^{\text{ideal}}, z_2^{\text{ideal}})$. Given the current archive $A \subseteq C$, the HV of a candidate w is

$$\Delta HV(w) = HV(A \cup \{w\}; r) - HV(A; r) \tag{25}$$

where $HV(\cdot; r)$ in Equation (25) denotes the HV measured with respect to the reference point r, which is chosen to be dominated by all candidates in C. To make the distance comparable across the objectives, each objective is normalized by its scale relative to the ideal point: define $s_i = \max_{w \in C} |f_i(w) - z_i^{\text{ideal}}|$ and use $\max(s_i, \hat{\sigma})$ with $\hat{\sigma} = 10^{-12}$ to prevent division by zero. The distance from w to z^{ideal} in Equation (26) is

$$D_{\text{ideal}}(w) = \left(\sum_{i=1}^{2} \left(\frac{|f_i(w) - z_i^{\text{ideal}}|}{\max(s_i, \hat{\sigma})} \right)^q \right)^{1/q}, \qquad q \in \{1, 2, \infty\} \tag{26}$$

At generation g, each candidate w receives the archive selection score

$$J_g(w) = \eta_g \, \Delta HV(w) + (1 - \eta_g)\big(-D_{\text{ideal}}(w)\big) \tag{27}$$

where, candidates with larger $J_g(w)$ in Equation (27) are preferred. Ties are broken using crowding distance as in NSGA-II [28]. The mixing weight $\eta_g \in [0, 1]$ evolves as $\eta_g = \eta_1 + (\eta_0 - \eta_1)\exp(-\kappa g)$, when η_g is larger and gradually shifts the emphasis toward convergence to the ideal point as η_g becomes smaller.

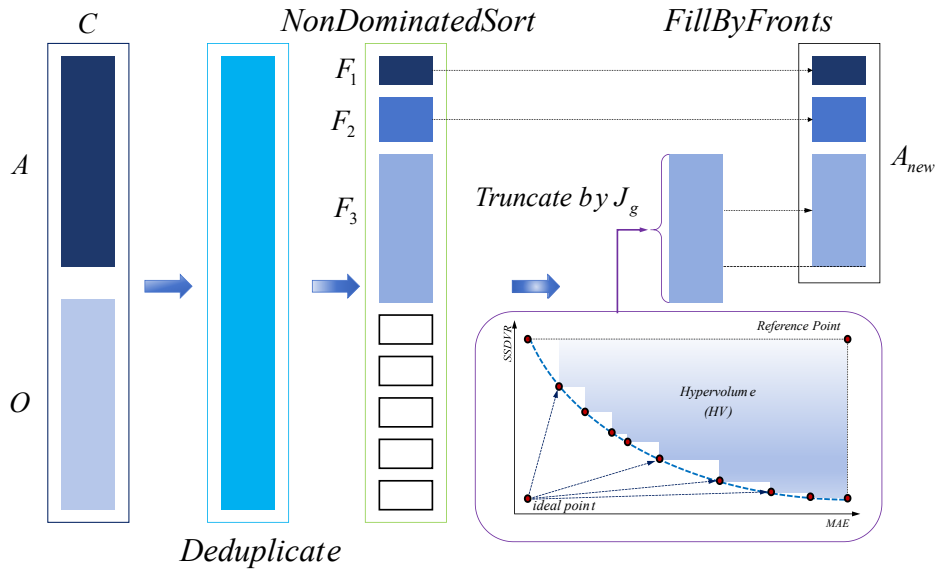


Figure 4. Schematic of the hybrid elitism used for archive selection, combining HV contribution and normalized distance to the ideal point. The blue curve illustrates a Pareto front in the objective space; the upper-right point is the HV reference point, and the lower-left point is the ideal point. The mixing weight η_g decays with the generation index, gradually shifting the selection preference from diversity-oriented coverage to convergence toward the ideal point.

Archive filling and truncation are performed sequentially over the non-dominated fronts $F_1, F_2, \ldots$. The new archive is initialized as $A_{\text{new}} \leftarrow \varnothing$, and fronts are processed in increasing rank: if $|A_{\text{new}}| + |F_k| \le A_{\max}$, all solutions in F_k are appended to A_{new}, otherwise, solutions in F_k are sorted in a descending order of $J_g(\mathbf{w})$, the top $(A_{\max} - |A_{\text{new}}|)$ solutions are appended to A_{new}, and the procedure terminates. Finally, the archive is updated by setting $A \leftarrow A_{\text{new}}$.

For reproducibility, all scores are computed using the shared reference point r and the common ideal point z^{ideal} determined from C, and the schedule parameters (η_0, η_1, κ) are fixed throughout the experiments, making the scoring function $J_g(\mathbf{w})$ fully determined at each generation. In the bi-objective case, non-dominated filtering for the first front can be implemented via sorting in $O(|F_1| \log |F_1|)$ time, and HV evaluation together with HV computation can also be carried out in $O(|F_1| \log |F_1|)$ time using sorting-based routines, while the normalized distance-to-ideal computation is linear in the number of candidates. Overall, this design maintains both computational efficiency and reproducibility of the archive update process.

3.5. Overall Evolutionary Process and Termination

Algorithm 1 outlines the MOECSO optimization loop. Beginning with an initial feasible population, candidate weight vectors are iteratively refined through the elite-guided horizontal and vertical crossover operators. Following each update, candidates undergo a repair process to ensure compliance with non-negativity and unit-sum constraints, after which they are evaluated on the validation set $\mathcal{V}_\tau$ using the bi-objective functions f_1 and f_2. Archive maintenance and selection are conducted according to the hybrid elite criterion. The process concludes at $T_{\max}$, with the final archive A_T being returned as the Pareto set.

Algorithm 1: MOECSO

Input: base forecasts $\{\hat{y}_t^{(m)}\}_{m=1}^{M}$ on $\mathcal{V}_\tau$; population size N; archive capacity $A_{\max}$; iteration cap $T_{\max}$; hyperparameters Θ.
Output: final archive A_T; selected weights $\mathbf{w}^* \in \Delta^M$.

Initialize:
  Sample N feasible weights on Δ^M to form P_0
  Evaluate (f_1, f_2) on $\mathcal{V}_\tau$ for all $\mathbf{w} \in P_0$
  Set A_0 by non-dominated sorting of P_0 and truncate to $A_{\max}$ if needed
  Set $g \leftarrow 0$
Main loop:
while $g < T_{\max}$ **do**
    $Q_g \leftarrow \emptyset$
  for *each* $\mathbf{w} \in P_g$ **do**
      $\mathbf{w}^{\text{peer}} \leftarrow \text{PeerSelect}(P_g)$
      $\mathbf{e} \leftarrow \text{EliteSelect}(A_g)$
      $\mathbf{w}_h \leftarrow \text{HorizontalCrossover}(\mathbf{w}, \mathbf{w}^{\text{peer}}, \mathbf{e})$
      $\mathbf{w}_v \leftarrow \text{VerticalCrossover}(\mathbf{w}, \mathbf{e})$
      $Q_g \leftarrow Q_g \cup \{\mathbf{w}_h, \mathbf{w}_v\}$
  end
    Evaluate (f_1, f_2) on $\mathcal{V}_\tau$ for all $\mathbf{w} \in Q_g$
    $C_g \leftarrow A_g \cup Q_g$
    $C_g \leftarrow \text{Deduplicate}(C_g)$
    Update $(r, z^{\text{ideal}}, s_1, s_2)$ from C_g and set η_g by the annealing schedule
    $\{F_1, F_2, \ldots\} \leftarrow \text{NonDominatedSort}(C_g)$
    $A_{g+1} \leftarrow \text{FillByFronts}(\{F_1, F_2, \ldots\}, A_{\max})$
  if *truncation is required within the last accepted front* **then**
      Rank that front by $J_g(\mathbf{w})$ and keep the top solutions to reach $|A_{g+1}| = A_{\max}$
  end
    $P_{g+1} \leftarrow \text{SelectPopulation}(C_g, N)$ $g \leftarrow g + 1$
end
Return:
  $\mathbf{w}^* \leftarrow \text{SelectFinalWeight}(A_{T_{\max}})$
  Output $A_{T_{\max}}$ and $\mathbf{w}^*$

4. Experimental Setup

4.1. Data and Leakage-Free Evaluation

Daily Europe Brent spot price in USD from the U.S. Energy Information Administration table Europe Brent Spot Price FOB [29] spans 20 May 1987 to 8 December 2025. Trading day observations are used without imputation. The task is a univariate one-step-ahead forecasting.

A leakage-free rolling-origin protocol is used [1]. After burn-in $N_0 = 8000$, each origin τ uses a fixed training window $W = 8000$ ending at τ and forecasts $\hat{y}_{\tau+1}$ with $h = 1$ and $S = 1$. The last $V = 500$ points in the window are used for validation, and the rest for fit. Z-score parameters μ_{F_τ} and σ_{F_τ} are derived from the fit segment only, standardize fit and validation, and forecasts are inverse-transformed to USD for error evaluation.

Ensembling uses simplex weights:

$$\Delta^M = \left\{ w \in \mathbb{R}_+^M : \sum_{m=1}^{M} w_m = 1 \right\}, \qquad \widehat{y}_{\tau+1} = \sum_{m=1}^{M} w_m \, \widehat{y}_{m,\tau+1}. \tag{28}$$

At each origin, $w(\tau)$ is optimized on validation by MOECSO with population 120 and 150 iterations to minimize MAE and SSDVR in USD, then base learners refit on the full window, and the out-of-sample forecast uses the selected weights. No future observations enter scaling, fitting, or weight search.

Fixed configurations include ARIMA [30], SVR [31], ELM [32], LSTM [33], and BPNN [34]. The seed equals 42 unless stated otherwise. Calendar-period statistics in Table 1 support regime description only. Stress robustness restricts test origins to T^*_{stress} under the same protocol and W, and metrics aggregate out-of-sample points based on their variability, measured as standard deviation.

Table 1. Descriptive statistics of the Brent price series in USD.

Sample	Mean	Std. Dev.	Min	Max
Full sample	50.8	32.6	9.1	144.0
Pre-2000	18.1	4.0	9.1	41.4
2000–2007	42.1	18.8	16.5	95.9
2008–2014	95.5	22.2	33.7	144.0
2015–2019	57.1	11.6	26.0	86.1
2020–2025	74.4	20.0	9.1	133.2

4.2. Evaluation Metrics and Statistical Tests

A leakage-free rolling-origin protocol evaluates out-of-sample accuracy on the original USD scale. With y_t and $\hat{y}_{m,t}$ denoting the realized and predicted Brent prices, the error is $e_{m,t} = y_t - \hat{y}_{m,t}$. Accuracy metrics are

$$\begin{aligned} \text{MAE}_m &= \frac{1}{T_{\text{OOS}}} \sum_{t=1}^{T_{\text{OOS}}} |e_{m,t}|, \\ \text{RMSE}_m &= \sqrt{\frac{1}{T_{\text{OOS}}} \sum_{t=1}^{T_{\text{OOS}}} e_{m,t}^2}, \\ \text{MAPE}_m &= \frac{100}{T_{\text{OOS}}} \sum_{t=1}^{T_{\text{OOS}}} \frac{|e_{m,t}|}{|y_t| + \varepsilon}, \end{aligned} \tag{29}$$

where T_{OOS} is the total number of test points and $\varepsilon > 0$ avoids division by zero.

Significance is assessed by loss-based tests on squared errors. For baseline b and proposed method p, define $d_t = e_{p,t}^2 - e_{b,t}^2$ and $\bar{d} = \frac{1}{T_{\text{OOS}}} \sum_{t=1}^{T_{\text{OOS}}} d_t$. The Diebold–Mariano statistic [35] is

$$\text{DM} = \frac{\sqrt{T_{\text{OOS}}}\, \bar{d}}{\sqrt{\hat{\sigma}_{\bar{d}}^2}}, \tag{30}$$

where $\hat{\sigma}_{\bar{d}}^2$ is the Newey–West HAC long-run variance with Bartlett weights [36] and $L = \lfloor T_{\text{OOS}}^{1/3} \rfloor$. Negative DM implies lower mean squared error for the proposed method.

To correct for data-snooping across multiple competitors, Hansen's SPA test [37] is applied with MOECSO as the reference (0). For competitor k, let $d_{k,t} = \ell(e_{0,t}) - \ell(e_{k,t})$ and $\bar{d}_k = \frac{1}{T_{\mathrm{OOS}}} \sum_{t=1}^{T_{\mathrm{OOS}}} d_{k,t}$. The studentized statistic is

$$T_{\mathrm{SPA}} = \max_{1 \le k \le K_c} \frac{\sqrt{T_{\mathrm{OOS}}}\, \bar{d}_k}{\hat{\omega}_k}, \quad \hat{\omega}_k = \sqrt{\mathrm{LRV}(d_{k,t})}, \tag{31}$$

where, $\mathrm{LRV}(\cdot)$ is estimated by the same Newey–West rule and p-values from a circular block bootstrap.

SPA p-values are adjusted by Benjamini–Hochberg [38]: for ordered $p_{(1)} \le \cdots \le p_{(m)}$,

$$q_{(i)} = \min_{j \ge i} \left\{ \frac{m}{j} p_{(j)} \right\}, \quad i = 1, \ldots, m, \tag{32}$$

and significance is declared at 5% when $q < 0.05$.

4.3. Computational Complexity and Scalability

Let W denote the rolling window length, V the inner validation length, K the number of VMD modes, M the number of base learners, N the population size, and G the number of MOECSO iterations. The per-origin runtime is approximated by

$$T_{\mathrm{origin}} \approx T_{\mathrm{VMD}}(W, K) + T_{\mathrm{feat}}(W, K) + T_{\mathrm{fit}}(W, M) + T_{\mathrm{MO}}(V, M, N, G). \tag{33}$$

With FFT-based implementations, $T_{\mathrm{VMD}}(W, K) = O(KW \log W)$ up to an inner iteration constant. Feature extraction across all modes costs $O(KW)$ for correlation and energy indicators, whereas MIC is empirically close to $O(KW \log W)$. Base learner training contributes $T_{\mathrm{fit}}(W, M) = \sum_{m=1}^{M} C_m(W)$. The multi-objective weight search evaluates simplex weighted combinations over V points and M learners with cost $O(VM)$ per evaluation, so

$$T_{\mathrm{MO}}(V, M, N, G) = O(GNVM). \tag{34}$$

Overall, runtime grows approximately linearly with M, N, G, and V, and grows close to $W \log W$ due to decomposition and indicator extraction. Inference is a lightweight simplex aggregation.

To complement asymptotic complexity, Table 2 reports an empirical benchmark under the same rolling origin protocol. The wall clock time is measured per origin and decomposed into decomposition, fitting, and weight search. Decomposition dominates the total runtime, especially for CEEMDAN, while the MOECSO search adds a small overhead under the current budget. A lightweight ridge projection baseline provides negligible optimization time but remains less accurate than MOECSO. A small-scale scalability check shows an increase in total runtime with larger W and larger M, which aligns with the $O(KW \log W)$ and $O(GNVM)$ terms.

4.4. Competitive Baselines and Implementation Details

It follows from Table 3 that the competing baselines exhibit a clear trade-off between training cost and online inference latency. ARIMA has the smallest size proxy at 0.007×10^3, and its per-origin training time is 2.5296 s while its inference time is 1.4291 ms/step. SVR requires 6.1088 s per origin, and its inference time is 4.1378 ms/step. Among the neural forecasters, LSTM is the most computationally demanding with 50.497×10^3 parameters, a training time of 22.8054 s per origin, and an inference time of 14.1226 ms/step. BPNN also incurs nontrivial costs with 16.129×10^3 parameters, 16.0802 s per origin, and 9.8429 ms/step.

Table 2. Summary of empirical runtime, baseline comparison, and scalability under the rolling origin protocol. All times are per origin in seconds.

Setting	Decomp.	Fit	WeightOpt.	Total
Runtime breakdown				
No decomposition	0.0000	3.7673	0.0699	6.1076
CEEMDAN decomposition	20.1390	3.2230	0.0608	24.9326
VMD decomposition	2.4183	1.1502	0.0511	4.3772
Baseline comparison under VMD				
EqualWeight	2.4183	1.1502	0.0000	4.3278
RidgeProj	2.4183	1.1502	0.0001	6.0366
MOECSO	2.4183	1.1502	0.0556	5.2857
Scalability in W under VMD and MOECSO				
$W = 2000$				1.3713
$W = 4000$				3.1184
$W = 8000$				6.7830
Scalability in M under VMD and MOECSO				
$M = 3$				6.1792
$M = 5$				7.2466

The proposed MOECSO pipeline concentrates its computational burden in the offline stage. Its end-to-end per-origin training time is 79.4172 s because it integrates VMD-based reconstruction, refitting all base learners, and weight optimization at each rolling origin. This added complexity does not translate into online latency. Once the base forecasts are available, MOECSO introduces only 0.0200 ms/step additional overhead for simplex-weighted aggregation. This overhead is of a magnitude smaller than the inference time of LSTM, which is 14.1226 ms/step, and it remains substantially lower than that of SVR, which is 4.1378 ms/step. These results indicate that MOECSO is well-suited to deployment settings that allow heavier offline updates while requiring fast one-step forecasts in real time.

Table 3. Model-complexity proxies and average per-origin computational cost under the rolling-origin protocol.

Model	Parameters ($\times 10^3$)	Train Time (s)	Inference Time (ms/Step)
ARIMA	0.007	2.5296	1.4291
SVR	1.314	6.1088	4.1378
ELM	8.060	0.0709	0.0378
LSTM	50.497	22.8054	14.1226
BPNN	16.129	16.0802	9.8429
MOECSO	–	79.4172	0.0200

Note: The training time is reported per rolling origin, and inference time corresponds to a one-step forecast. For MOECSO, inference time includes only the aggregation overhead.

4.5. Robustness Under Market Stress

Table 4 and Figure 5 jointly indicate that the proposed MOECSO ensemble remains operationally robust under pronounced market stress while preserving stable tracking of the price trajectory. Stress-period evaluation follows the same leakage-free rolling-origin protocol as the full-sample experiment. At each origin, preprocessing, model refitting, and simplex-weight optimization use only the current training window, and the one-step-ahead forecast is recorded as a strictly out-of-sample prediction [39]. A consistent per-origin budget is enforced across methods. LSTM and BPNN are trained for 25 epochs

at each origin, and simplex weights are optimized at every origin using MOECSO with the population size $P = 120$ and $G = 150$ iterations.

Across all three crisis episodes, the ensemble delivers consistent one-step-ahead out-of-sample forecasts under the same fixed-length rolling-origin setting. During the 2008–2009 Global Financial Crisis segment, the average errors remain moderate with MAE being 1.99339 and Root Mean Square Error (RMSE) being 2.65644. The tail-focused statistic MAE_{95}, equal to 4.92702, shows that extreme deviations are contained relative to the magnitude of the regime shift. This behavior is also visible in Figure 5, where the ensemble forecast follows the overall decline and subsequent stabilization without persistent drift, and the largest discrepancies concentrate around abrupt transitions rather than accumulating over time.

The stress results further highlight the practical advantage of MOECSO in adapting to different disruption patterns without retuning the model family. The COVID-19 shock yields the largest dispersion, as reflected by RMSE of 11.68760 and MAE_{95} of 10.30360, consistent with the sudden collapse and rapid rebound in 2020. The forecast path still captures directional movements and avoids long-lasting bias. The 2022 energy-price spike exhibits a higher average MAE equal to 3.30720, together with a smaller MAPE equal to 2.97267, suggesting that the ensemble maintains relative accuracy at elevated price levels. Overall, the combination of stable trajectory tracking in Figure 5 and controlled tail errors in Table 4 supports that MOECSO provides a resilient fusion mechanism. By optimizing simplex weights at each rolling origin, MOECSO can shift emphasis among the base learners as market conditions change, thereby reducing sustained mis-calibration and limiting worst-case deviations during crisis-driven regime shifts.

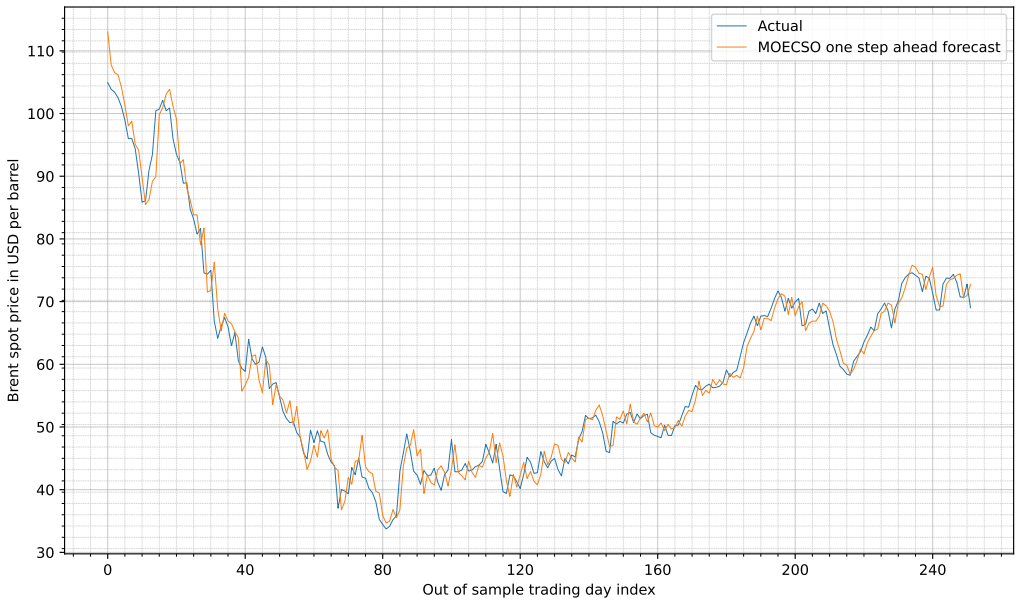


Figure 5. Example out-of-sample forecast trajectory under the fixed-length rolling-origin protocol during the 2008–2009 Global Financial Crisis segment.

Table 4. Stress-period out-of-sample forecasting performance of the MOECSO ensemble under the fixed-length rolling-origin protocol.

Stress Period	N	MAE	RMSE	MAPE (%)	MAE_{95}
GFC 2008–2009	252	1.99339	2.65644	3.54952	4.92702
COVID-19 2020	126	3.96927	11.68760	16.03590	10.30360
Energy shock 2022	125	3.30720	4.45515	2.97267	9.30593

5. Ablation Study

5.1. Sensitivity Analysis

Table 5 reports a single factor sensitivity study around the default configuration under the same rolling origin protocol. The overall pattern indicates moderate robustness for most factors, since varying α, archive update interval, crossover intensity, and noise annealing strength around the default values produces only small changes in MAE, RMSE, and MAPE. The mode number K shows a clearer effect, where $K = 8$ provides the best performance among the tested options, while $K = 10$ causes a pronounced accuracy collapse with MAE increasing to 5.974 and RMSE increasing to 7.704. The penalty parameter α also exhibits a non monotonic trend, with $\alpha = 1000$ yielding slightly lower MAE than $\alpha = 2000$, whereas $\alpha = 4000$ substantially degrades all metrics. Overall, the sensitivity results suggest that the pipeline behavior remains stable under moderate perturbations, but extreme settings such as overly large K can severely harm forecasting accuracy.

Table 5. Single factor sensitivity around the default configuration. Lower is better. Results are reported as mean ± standard deviation over five seeds.

Factor	Setting	MAE	RMSE	MAPE
VMD modes K	$K = 6$	2.604 ± 0.115	3.616 ± 0.128	4.258 ± 0.198
VMD modes K	$K = 8$	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
VMD modes K	$K = 10$	5.974 ± 0.173	7.704 ± 0.291	10.869 ± 0.392
VMD penalty α	1000	2.469 ± 0.211	3.431 ± 0.302	4.205 ± 0.489
VMD penalty α	2000	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
VMD penalty α	4000	3.654 ± 0.543	4.711 ± 0.745	6.158 ± 1.153
Frequency penalty β	3	2.473 ± 0.196	3.403 ± 0.273	4.141 ± 0.362
Frequency penalty β	5	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
Frequency penalty β	7	2.502 ± 0.199	3.442 ± 0.274	4.197 ± 0.358
Archive update interval	5	2.485 ± 0.195	3.419 ± 0.270	4.164 ± 0.356
Archive update interval	10	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
Archive update interval	20	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
Noise annealing strength	low	2.493 ± 0.177	3.431 ± 0.247	4.186 ± 0.323
Noise annealing strength	mid	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
Noise annealing strength	high	2.485 ± 0.195	3.419 ± 0.270	4.164 ± 0.355
Crossover intensity	0.2	2.485 ± 0.195	3.419 ± 0.270	4.164 ± 0.356
Crossover intensity	0.4	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
Crossover intensity	0.6	2.484 ± 0.195	3.419 ± 0.270	4.164 ± 0.356

5.2. Component-Level Ablation of Fusion and Reconstruction

Table 6 isolates the contribution of fusion baselines, component selection, reconstruction weighting, and the MOECSO improvement module under the same rolling origin evaluation. The results indicate the dominance of short-horizon persistence, since the naive last value baseline achieves MAE 1.306 and RMSE 1.886, while the best non-naive single learner selected on validation remains much weaker with MAE 3.265 and RMSE 4.577. Equal averaging is not competitive, suggesting that unfocused fusion amplifies weak learners, whereas simplex constrained fusion yields clear gains, with Dirichlet random simplex search reaching MAE 2.488 and RMSE 3.422, and MOECSO further improving to MAE 2.484 and RMSE 3.419 under the capped naive constraint. Component selection and reconstruction show limited separation between single-indicator and multi-indicator variants in this configuration, both yielding MAE 2.484 and RMSE 3.419, whereas retaining all components without indicator selection attains MAE 2.456 and RMSE 3.380. In contrast, reconstruction weighting emerges as a key factor, because equal-weight reconstruction

degrades sharply, with MAE 3.714 and RMSE 4.660, while entropy-weighted reconstruction restores performance to MAE 2.484 and RMSE 3.419. Finally, the proposed MOECSO improvement module provides an additional gain under the same search budget, with the variant without the module producing MAE 2.526 and RMSE 3.457 and the full variant reducing the error to MAE 2.484 and RMSE 3.419.

Table 6. Ablation of fusion and reconstruction components under the rolling origin protocol. Lower is better. Mean ± standard deviation over five seeds. Abbreviations: NW denotes naive weight constraint; MI denotes multi-indicator selection; EW-R denotes equal-weight reconstruction; Ent-R denotes entropy-weighted reconstruction; Imp denotes the proposed MOECSO improvement module.

Group	Setting	MAE	RMSE	MAPE
A	Best single, naive	1.306 ± 0.000	1.886 ± 0.000	1.959 ± 0.000
A	Best single, non-naive	3.265 ± 0.293	4.577 ± 0.553	5.542 ± 0.510
A	Equal averaging	3.953 ± 0.288	5.319 ± 0.317	6.989 ± 0.534
A	Dirichlet simplex search	2.488 ± 0.192	3.422 ± 0.276	4.198 ± 0.371
A	MOECSO fusion, NW	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
B	All components, no selection	2.456 ± 0.199	3.380 ± 0.277	4.113 ± 0.366
B	Single-indicator selection	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
B	Multi-indicator selection	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
C	MI + EW-R	3.714 ± 0.264	4.660 ± 0.271	6.402 ± 0.488
C	MI + Ent-R	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356
D	MOECSO without Imp	2.526 ± 0.149	3.457 ± 0.215	4.291 ± 0.290
D	MOECSO with Imp	2.484 ± 0.195	3.419 ± 0.270	4.163 ± 0.356

5.3. Impact of Input Representation

Table 7 studies the role of the input representation within the MOECSO ensemble under the strict rolling origin one-step-ahead protocol. Across the first three settings, the downstream pipeline uses the same base learners and MOECSO weight search configuration, and the only change lies in the series fed into this pipeline, namely, the raw series, the CEEMDAN-reconstructed series, or the VMD-reconstructed series. Under this controlled MOECSO-based comparison, the raw series yields the lowest error among the three representations, with MAE 1.659 and RMSE 2.665. For decomposition-based inputs, VMD reconstruction is markedly more effective than CEEMDAN reconstruction in the same MOECSO pipeline, lowering MAE from 2.672 to 1.736 and RMSE from 3.549 to 2.457. In direct comparison with these three MOECSO-based variants, the full proposed framework attains the lowest MAE, RMSE, and MAPE overall, indicating that the proposed end-to-end design delivers the best forecasting performance under the same protocol.

Table 7. Effect of input representation within the MOECSO ensemble under the rolling origin protocol. Lower is better. The first three settings use the same base learners and the same MOECSO weight search configuration and differ only in the input series, namely, no decomposition, CEEMDAN reconstruction, and VMD reconstruction. Results are reported for a single run, and rolling origins are subsampled with a step size of 20. The last row reports the full proposed framework under the same protocol.

Setting	MAE	RMSE	MAPE (%)
No decomposition	1.659	2.665	2.285
CEEMDAN	2.672	3.549	4.270
VMD	1.736	2.457	2.577
MOECSO	1.322	1.893	2.063

6. Results and Discussion

Table 8 and Figure 6 report full-sample out-of-sample performance under the leakage-free rolling-origin protocol. MOECSO achieves MAE 1.32224 and RMSE 1.89303, remaining close to the best single baseline while outperforming ARIMA and BPNN. ARIMA yields MAE 13.91880 and RMSE 18.88150, consistent with limited flexibility under nonlinear and regime-varying dynamics. BPNN yields MAE 4.27716 and RMSE 5.07485, suggesting that standalone neural fitting without adaptive fusion is insufficient in this setting. LSTM attains the lowest single-model MAE and RMSE (1.31287 and 1.87689, respectively), while the numerical gap relative to MOECSO remains small, and the DM, SPA, and FDR results do not support a decisive statistical advantage under the rolling-origin protocol. MOECSO optimizes simplex constrained fusion weights under a joint accuracy and stability objective, so a slight increase in point error can occur when lower dispersion is prioritized across rolling origins. LSTM is a single learner that can be numerically favored when the local dynamics match its inductive bias, while MOECSO is still favored in practice because it reduces reliance on a single model class under nonstationary price dynamics [40]. Figure 6 shows close trajectory tracking and no persistent drift on the representative test segment.

The rolling-origin evaluation produces a simplex weight vector $w(\tau) \in \Delta^M$ at each origin, so ensemble weights form a time-indexed outcome that reflects how forecast reliance shifts across market conditions. Weight distributions across calendar regimes and crisis-defined stress sets characterize the location and dispersion of $w(\tau)$, offering economic interpretation for decision-making under distinct market environments. To summarize whether the ensemble behaves as a concentrated selector or a diversified combiner, a concentration index is computed:

$$H(\tau) = \sum_{m=1}^{M} w_m(\tau)^2. \tag{35}$$

A larger $H(\tau)$ indicates weight concentration on fewer models, while a smaller $H(\tau)$ indicates diversified averaging, linking the selected weights to an interpretable accuracy versus stability trade-off. Because the objectives penalize MAE and SSDVR jointly, regime and crisis shifts that change the feasible MAE to SSDVR trade-off induce systematic changes in $w(\tau)$ and in $H(\tau)$, so the weights carry statistical and economic implications rather than acting as fixed coefficients. The stress-period evidence in Table 4 supports this interpretation, since elevated dispersion and tail deviations in the 2020 interval align with stronger emphasis on stability, which is consistent with shifts in optimized weight profiles across origins.

Table 8. Out-of-sample forecasting performance and statistical evidence under the rolling-origin protocol with the DM, Hansen SPA, and FDR procedures.

Model	MAE	RMSE	MAPE (%)	DM	SPA	FDR q-Value
ARIMA	13.91880	18.88150	22.55630	−6.80099	1.00000	1.00000
SVR	1.35782	1.98366	2.14820	−2.35666	0.99000	1.00000
ELM	1.35894	1.96014	2.09847	−1.72803	0.95500	1.00000
LSTM	1.31287	1.87689	1.96490	0.88809	0.20000	0.80000
BPNN	4.27716	5.07485	5.91480	−11.61650	1.00000	1.00000
MOECSO	1.32224	1.89303	2.06393	–	–	–

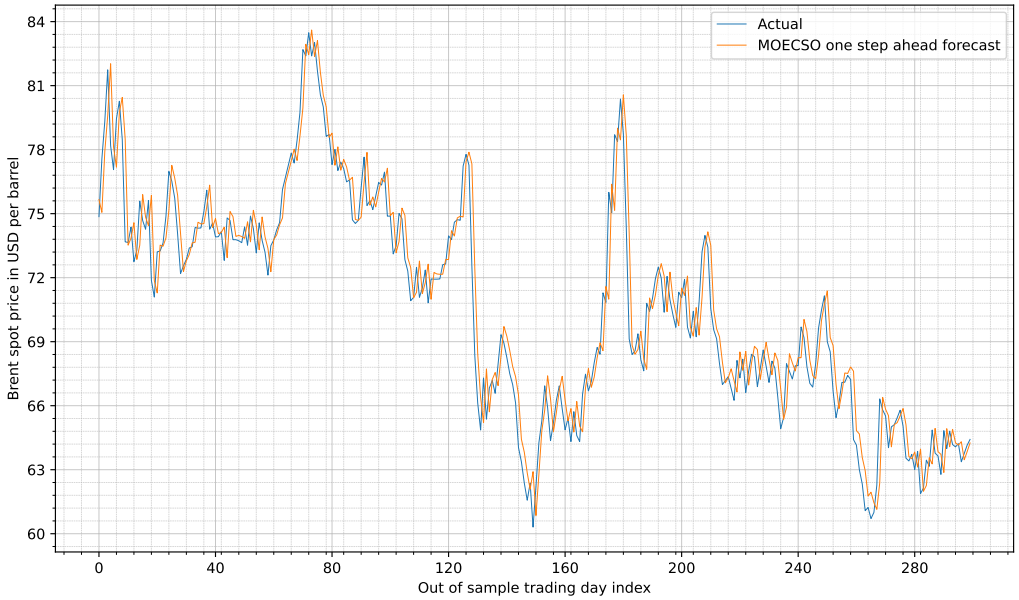


Figure 6. Actual Brent price and MOECSO one-step-ahead forecasts on a representative test segment.

7. Conclusions

This paper presents a leakage-aware decomposition-ensemble forecasting framework for Brent crude oil prices, wherein the final prediction is derived as a convex combination of heterogeneous base forecasters, constrained by simplex conditions. The framework consists of two main components. First, an entropy-frequency VMD reconstruction operator transforms multi-mode decompositions into a single task-adaptive input signal. Second, a bi-objective ensemble weight search is addressed using the multi-objective evolutionary algorithm MOECSO. This optimization operates on a probability simplex and simultaneously aims to achieve complementary validation objectives that balance mean accuracy with residual stability. To enhance search reliability, two algorithmic mechanisms are integrated. An elite-guided horizontal and vertical crossover facilitates efficient exploration, while a hybrid archive update ensures stable maintenance of a Pareto set.

The proposed method is evaluated using a leakage-free rolling-origin protocol, with performance measured by standard error metrics. Inference is supported by the DM and SPA tests, incorporating FDR control. In addition to full-sample comparisons, robustness is assessed on stress-period subsamples corresponding to the 2008–2009 Global Financial Crisis, the 2020 COVID-19 pandemic, and the 2022 energy shock. The analysis reports both mean error criteria and the tail-focused diagnostic MAE_{95}. Results from the stress periods indicate that forecast errors cluster around regime shifts and volatility bursts, suggesting that robustness should be evaluated in terms of both average accuracy and tail risk.

Several limitations indicate directions for future research. First, while VMD-based reconstruction enhances robustness, the reconstruction operator could benefit from incorporating uncertainty-aware weighting or regime-conditioned scoring. Second, the current ensemble is based on fixed model families. Integrating modern sequence and mixer architectures and assessing their marginal benefits under the same leakage-free protocol would enhance generalizability. Third, weight optimization could explicitly focus on transaction-relevant objectives, such as directional accuracy during periods of high volatility, drawdown-aware loss, or asymmetric penalties, and could be extended to generate calibrated predictive intervals. Finally, repeated refitting remains computationally intensive. Implementing warm-starting, incremental updates, or surrogate-assisted optimization could decrease runtime while maintaining statistical rigor.

https://doi.org/10.3390/math14050814

Author Contributions: Conceptualization, L.B. and N.W.; methodology, L.Z., Z.C. and L.B.; software, Z.C.; validation, Z.C.; formal analysis, L.Z.; investigation, Z.C. and N.W.; resources, L.B.; data curation, Z.C.; writing—original draft, Z.C.; writing—review & editing, L.Z., N.W. and L.B.; visualization, Z.C.; supervision, L.B. and N.W.; project administration, L.B.; funding acquisition, L.B. and N.W. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the Science and Technology Development Fund in Macau (No. 0027/2022/AGJ and No. 0082/2025/AFJ) and the Guangdong Provincial Science and Technology Plan Project (No. 2023A0505020007).

Data Availability Statement: The data used in the experimentation of this article are publicly available from the U.S. Energy Information Administration (EIA). Specifically, the crude oil spot price data can be accessed at: https://www.eia.gov/dnav/pet/hist/RBRTEd.htm (accessed on 8 December 2025).

Conflicts of Interest: The authors declare no conflicts of interest. The funders provided financial support for the project but had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Hewamalage, H.; Bergmeir, C.; Bandara, K. Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Min. Knowl. Discov.* **2023**, *37*, 788–832. [CrossRef] [PubMed]
2. Guan, K.; Gong, X. A new hybrid deep learning model for monthly oil prices forecasting. *Energy Econ.* **2023**, *128*, 107136. [CrossRef]
3. Iftikhar, H.; Qureshi, M.; Canas Rodrigues, P.; Usman Iftikhar, M.; Linkolk López-Gonzales, J.; Iftikhar, H. Daily Crude Oil Prices Forecasting Using a Novel Hybrid Time Series Technique. *IEEE Access* **2025**, *13*, 98822–98836. [CrossRef]
4. Xu, Z.; Mohsin, M.; Ullah, K.; Ma, X. Using econometric and machine learning models to forecast crude oil prices: Insights from economic history. *Resour. Policy* **2023**, *83*, 103614. [CrossRef]
5. Lu, W.; Huang, Z. Crude Oil Prices Forecast Based on Mixed-Frequency Deep Learning Approach and Intelligent Optimization Algorithm. *Entropy* **2024**, *26*, 358. [CrossRef]
6. Memon, B.A. Forecasting Fossil Energy Price Dynamics with Deep Learning: Implications for Global Energy Security and Financial Stability. *Algorithms* **2025**, *18*, 776. [CrossRef]
7. Godahewa, R.; Bergmeir, C.; Baz, Z.E.; Zhu, C.; Song, Z.; García, S.; Benavides, D. On Forecast Stability. *Int. J. Forecast.* **2025**, *41*, 1539–1558. [CrossRef]
8. Lin, S.; Wang, Y.; Wei, H.; Wang, X.; Wang, Z. Hybrid Method for Oil Price Prediction Based on Feature Selection and XGBOOST-LSTM. *Energies* **2025**, *18*, 2246. [CrossRef]
9. Zitzler, E.; Thiele, L. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Trans. Evol. Comput.* **1999**, *3*, 257–271. [CrossRef]
10. Iftikhar, H.; Zafar, A.; Turpo-Chaparro, J.E.; Canas Rodrigues, P.; Lopez-Gonzales, J.L. Forecasting Day-Ahead Brent Crude Oil Prices Using Hybrid Combinations of Time Series Models. *Mathematics* **2023**, *11*, 3548. [CrossRef]
11. Zhang, S.; Luo, J.; Wang, S.; Liu, F. Oil price forecasting: A hybrid GRU neural network based on decomposition-reconstruction methods. *Expert Syst. Appl.* **2023**, *218*, 119617. [CrossRef]
12. Huang, L.; Yang, X.; Lai, Y.; Zou, A.; Zhang, J. Crude Oil Futures Price Forecasting Based on Variational and Empirical Mode Decompositions and Transformer Model. *Mathematics* **2024**, *12*, 4034. [CrossRef]
13. Wang, X.; Yang, P.; Zhou, X.; Wang, Z. Crude Oil Price Forecasting: An Ensemble-Driven Long Short-Term Memory Model Based on CEEMDAN Decomposition and ALS-PSO Optimization. *Energy Sci. Eng.* **2023**, *11*, 4054–4076. [CrossRef]
14. Wang, X.; Hyndman, R.J.; Li, F.; Kang, Y. Forecast combinations: An over 50-year review. *Int. J. Forecast.* **2023**, *39*, 1518–1547. [CrossRef]
15. Chen, Q.; Hong, Y.; Li, H. Time-varying forecast combination for factor-augmented regressions with smooth structural changes. *J. Econom.* **2024**, *240*, 105693. [CrossRef]
16. Liu, L.; Zhou, S.; Jie, Q.; Du, P.; Xu, Y.; Wang, J. A Robust Time-Varying Weight Combined Model for Crude Oil Price Forecasting. *Energy* **2024**, *299*, 131352. [CrossRef]
17. Yuan, J.; Li, J.; Hao, J. A dynamic clustering ensemble learning approach for crude oil price forecasting. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106408. [CrossRef]
18. Chen, Z.; Lan, T.; He, D.; Cai, Z. Elitist Non-Dominated Sorting Crisscross Algorithm for Multi-Objective Optimization with Application in Neural Architecture Search. *Mathematics* **2025**, *13*, 1258. [CrossRef]

19. Sun, Y.; Qu, Z.; Liu, Z.; Li, X. Hierarchical Multi-Scale Decomposition and Deep Learning Ensemble Framework for Enhanced Carbon Emission Prediction. *Mathematics* **2025**, *13*, 1924. [CrossRef]
20. Zhang, K.; Zhao, S.; Zeng, H.; Chen, J. Two-Stage Archive Evolutionary Algorithm for Constrained Multi-Objective Optimization. *Mathematics* **2025**, *13*, 470. [CrossRef]
21. Miguel, F.M.; Frutos, M.; Méndez, M.; Tohmé, F.; González, B. Comparison of MOEAs in an Optimization-Decision Methodology for a Joint Order Batching and Picking System. *Mathematics* **2024**, *12*, 1246. [CrossRef]
22. Miettinen, K. *Nonlinear Multiobjective Optimization*; International Series in Operations Research & Management Science; Springer: Boston, MA, USA, 1998; Volume 12. [CrossRef]
23. Li, L.; Shan, K.; Wenyuan, G. Forecasting Crude Oil Price Using Secondary Decomposition-Reconstruction-Ensemble Model Based on Variational Mode Decomposition. *J. Futures Mark.* **2025**, *45*, 1601–1615. [CrossRef]
24. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Lander, E.S. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [CrossRef] [PubMed]
25. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
26. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
27. Meng, A.B.; Chen, Y.C.; Yin, H.; Chen, S.Z. Crisscross optimization algorithm and its application. *Knowl.-Based Syst.* **2014**, *67*, 218–229. [CrossRef]
28. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.A.M.T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
29. U.S. Energy Information Administration. Europe Brent Spot Price FOB (Dollars per Barrel). 2025. Available online: https://www.eia.gov/dnav/pet/hist/RBRTED.htm (accessed on 8 December 2025).
30. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015. [CrossRef]
31. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems 9*; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161.
32. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
33. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
34. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
35. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [CrossRef]
36. Newey, W.K.; West, K.D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **1987**, *55*, 703–708. [CrossRef]
37. Hansen, P.R. A Test for Superior Predictive Ability. *J. Bus. Econ. Stat.* **2005**, *23*, 365–380. [CrossRef]
38. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [CrossRef]
39. Liang, X.; Luo, P.; Li, X.; Wang, X.; Shu, L. Crude oil price prediction using deep reinforcement learning. *Resour. Policy* **2023**, *81*, 103363. [CrossRef]
40. Dağkurs, B.; Atacak, I. Deep learning-based novel ensemble method with best score transferred-adaptive neuro fuzzy inference system for energy consumption prediction. *Peerj Comput. Sci.* **2025**, *21*, e2680. [CrossRef]

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