

Article

Not peer-reviewed version

---

# Multi-Scale Forecasting of Natural Rubber Prices Using a VMD-Hybrid BiLSTM-Transformer Framework

---

[Montchai Pinitjitsamut](#)\*

Posted Date: 31 March 2026

doi: 10.20944/preprints202603.2461.v1

Keywords: natural rubber price forecasting; variational mode decomposition; bidirectional LSTM; transformer; hybrid deep learning; directional accuracy; variance collapse; commodity markets; RSS3; multi-step forecasting



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

# Multi-Scale Forecasting of Natural Rubber Prices Using a VMD-Hybrid BiLSTM-Transformer Framework

Montchai Pinitjitsamut 

Department of Agricultural and Resource Economics, Faculty of Economics, Kasetsart university, Bangkok, Thailand;  
Montchai.p@ku.th

## Highlights

### What are the main findings?

- VMD–Hybrid BiLSTM–Transformer achieves Pearson correlation of 0.812 and StdR of 0.819 on daily rubber price changes, outperforming ARIMA by 0.662 in correlation and 37.3% in MAE, with meaningful predictive skill up to a five-day horizon.
- Directional accuracy is an unreliable standalone metric for difference commodity series: Vanilla LSTM attains 78.6% directional accuracy through systematic bias yet correctly identifies only one price decrease across 237 test observations (StdR = 0.033).

### What are the implications of the main findings?

- Appending VMD components directly as input features — rather than summing independent mode forecasts — preserves multi-scale frequency information and prevents variance collapse on difference price series.
- Robust commodity price forecasting requires variance-sensitive evaluation metrics and joint integration of multi-scale decomposition, bidirectional learning, and global attention within a unified architecture.

## Abstract

Natural rubber price forecasting is inherently difficult due to nonlinear, non-stationary dynamics driven by supply fundamentals, cross-market signals, exchange rate movements, and speculative trading. This study proposes VMD–Hybrid BiLSTM–Transformer, a dual-pathway framework integrating Variational Mode Decomposition (VMD) with a Bidirectional LSTM encoder and a Transformer encoder for daily RSS3 FOB price change forecasting. Rather than forecasting each intrinsic mode function independently, all five VMD components are appended directly to the economic feature matrix — preserving multi-scale frequency information within a single forward pass and avoiding the variance collapse observed in conventional decomposition forecast approaches (StdR = 0.04 – 0.15). On a 237-observation held-out test set (September 2025–February 2026), the model achieves Pearson correlation of 0.812, directional accuracy of 67.1%, and StdR of 0.819, outperforming ARIMA by 0.662 in correlation and 37.3% in MAE, with predictive skill confirmed up to five days. These results demonstrate that directional accuracy alone is insufficient for evaluating difference commodity price models, and that jointly integrating multi-scale decomposition, bidirectional learning, and global attention is essential for reliable agricultural price forecasting.

**Keywords:** natural rubber price forecasting; variational mode decomposition; bidirectional LSTM; transformer; hybrid deep learning; directional accuracy; variance collapse; commodity markets; RSS3; multi-step forecasting

## 1. Introduction

Natural rubber is a critical industrial input and a strategic agricultural commodity that plays a central role in global manufacturing supply chains. It is widely used in the production of automobile tires, medical gloves, engineering components, and a broad range of technical rubber products. Global natural rubber production is highly concentrated in Southeast Asia, particularly in Thailand, Indonesia, and Malaysia, which together account for the majority of world supply [1,2]. Among these producers, Thailand has consistently remained the world's leading exporter of natural rubber. In 2025, Thailand produced approximately 4.79 million tonnes, accounting for 32.18% of global production, with about 85.79% of output exported to international markets [3]. Price volatility in natural rubber markets therefore carries direct consequences for the income of millions of smallholder farmers and the cost structure of downstream industries that rely heavily on rubber-based inputs.

Natural rubber prices are jointly driven by agricultural seasonality, macroeconomic conditions, exchange rate dynamics, and speculative activity in futures markets, producing price dynamics that are simultaneously nonlinear, non-stationary, and subject to structural breaks [4–6]. These forces generate significant uncertainty for stakeholders across the rubber value chain. In international trade, Ribbed Smoked Sheet No.3 (RSS3) serves as the benchmark grade for natural rubber transactions, with export prices typically quoted on a Free on Board (FOB) basis at major Thai ports. Price co-movement across SHFE, SGX, and JPX (TOCOM) futures markets creates cross-market information flows that provide additional predictive content beyond domestic spot prices alone, motivating the inclusion of exchange-traded settlement prices as model inputs [7]. This cross-market structure is not symmetric: [8] demonstrate that SHFE Granger-causes TOCOM but not vice versa, while Kepulaje et al. [9] [pp. 159–161] show that SGX RSS3 Rubber Futures (SRU) and SGX TSR20 Rubber Futures (STF) transmit more consistently into regional spot prices than TOCOM/JPX Rubber Futures (RSS3)(JRU) — implying that the futures exchanges operate as a layered rather than a flat information system, and that the informational weight of each venue depends on whether one is studying financial benchmarking or physical export pricing.

Accurate forecasting of commodity prices therefore remains a central issue in both economic research and industrial decision-making. Commodity price forecasting has long relied on econometric models such as ARIMA, VAR, and GARCH, which provide well-established frameworks for capturing temporal dependencies and volatility structures [10,11]. In the natural rubber literature specifically, this approach is exemplified by [12]. (2018), who apply the Box–Jenkins procedure to Malaysian SMR20 and identify ARIMA(1,1,0) as the best-fitting model, and by [13], who show that a simultaneous supply–demand–price system outperforms univariate ARIMA across RMSE, MAE, and information criteria — demonstrating that purely univariate extrapolation is insufficient once commodity market structure is made explicit. However, these approaches rest on linearity assumptions that limit their capacity to represent the complex nonlinear dynamics observed in modern commodity markets, and their performance often deteriorates further when the underlying data exhibit structural breaks and high non-stationary characteristics that are particularly pronounced in energy and agricultural commodity price series [14,15]. To address these shortcomings, subsequent studies introduced machine learning methods including support vector machines (SVM), multi-layer perceptrons (MLP), random forests, extreme gradient boosting (XGBoost), and back-propagation neural networks (BPNN), which are better suited for capturing nonlinear relationships in complex datasets [16–18]. Hybrid frameworks combining statistical and machine learning components—such as ARIMA–SVM—have also been proposed to integrate linear structure with nonlinear residual dynamics [19,20]. Nevertheless, these approaches are limited in their ability to capture complex nonlinear dynamics and long-range dependencies, motivating the adoption of deep learning alternatives.

Building on this shift, recurrent architectures—particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks—have been widely adopted for their ability to model long-term temporal dependencies more effectively than shallow learning algorithms, demonstrating superior performance over classical statistical benchmarks in energy demand prediction, financial mar-

ket forecasting, and environmental time-series modeling [21,22]. For natural rubber specifically, [23] demonstrate that a multi-layer LSTM achieves 95.88% accuracy on Thai RSS3 prices, while [24] confirm that LSTM outperforms ARIMA on Malaysian SMR20 across MAE, RMSE, and MAPE — with both studies noting that predictive gains remain incomplete without explicit incorporation of exogenous drivers such as exchange rates, energy prices, and futures markets. However, recurrent models can struggle with highly volatile and non-stationary series, particularly when the data exhibit abrupt structural changes or mixed-frequency dynamics [25]. Bidirectional Long Short-Term Memory (BiLSTM) networks extend the conventional LSTM architecture by processing sequential data simultaneously in the forward and backward directions, enabling the model to integrate contextual information from both past and future states within the observation window. This bidirectional structure yields richer temporal representations than unidirectional recurrent networks and has demonstrated consistent advantages over conventional LSTM and classical statistical benchmarks in environments characterized by strong seasonality, nonlinear interactions, and abrupt fluctuations—including renewable energy forecasting and environmental time-series prediction [21,22,26].

To further extend the representational capacity of BiLSTM, several lines of re-search have proposed hybrid architectures. CNN-BiLSTM models combine convolution layers for local feature extraction with BiLSTM for bidirectional temporal learning, significantly improving prediction performance in energy and meteorological forecasting applications [27,28]. Attention mechanisms have also been integrated to allow the network to emphasize the most informative time steps, enhancing performance on non-stationary sequences and series with irregular fluctuations [29]. A separate strand of research applies meta-heuristic optimization to tune BiLSTM hyper-parameters, and combines BiLSTM with gradient boosting or ensemble methods to capture residual nonlinear patterns, yielding further gains over single-model recurrent baselines in energy and meteorological forecasting applications. Despite these advances, BiLSTM-based hybrids remain limited in their ability to separate and individually model components operating across distinctly different temporal scales, a constraint that becomes particularly consequential for commodity price series with coexisting trend, cyclical, and noise dynamics.

A parallel research direction addresses the recurrent model's limited ability to disentangle components at different temporal scales by incorporating signal decomposition as a preprocessing stage. Methods including Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), Wavelet Transform (WT), and Seasonal-Trend Decomposition using LOESS (STL) separate the original series into components representing distinct frequency structures—trend, seasonal, and residual—thereby reducing non-stationary and simplifying the learning task for downstream models [7,29,30]. A substantial body of empirical evidence confirms that decomposition-based hybrid frameworks, including EMD-LSTM, VMD-LSTM, VMD-GRU, VMD-TCN-LSTM, and STL-LSTM—significantly improve forecasting accuracy across coal markets, solar power generation, wind forecasting, and electricity load prediction [29]. Recent evidence from high-frequency financial forecasting likewise shows that VMD can separate trend, periodic, and disturbance components in non-stationary series and that decomposition-enhanced recurrent models can outperform standalone econometric and deep-learning benchmarks [25]. In commodity-price forecasting, VMD-based hybrid models have similarly been shown to outperform both single-model and earlier decomposition-based alternatives, with further gains when structurally relevant long-run drivers are incorporated into the forecasting system [31]. Recent evidence from environmental forecasting further confirms that coupling VMD with sequence models can substantially improve predictive accuracy by separating short- and long-run signal structures before learning, relative to models trained on the raw series alone [32]. Related evidence from drought forecasting likewise demonstrates that VMD-based preprocessing can improve short-lead predictive accuracy relative to standalone machine learning models, reinforcing the value of decomposition in nonlinear forecasting tasks [33]. In related commodity-market research, VMD-based recurrent models have also been shown to improve forecasting by decomposing volatile price series into simpler sub-components before learning, confirming the value of decomposition for nonlinear price dynamics [34].

Building on these decomposition gains, a further line of research demonstrates that combining VMD with attention-based sequence architectures yields complementary improvements beyond what either component achieves alone. Related forecasting studies have shown that VMD can improve downstream deep-learning performance by decomposing complex time series into more learnable components, while attention-based LSTM encoder-decoder structures further enhance the selection of informative features and time steps [35] [pp. 922–923]. Evidence from short-term wind power forecasting further suggests that VMD is especially effective when embedded within a broader hybrid pipeline rather than used as an isolated preprocessing step, improving both stability and predictive accuracy in volatile series [36]. Recent FX forecasting evidence additionally indicates that VMD-enhanced bidirectional recurrent architectures, when paired with systematic hyper-parameter tuning and regularization, can materially improve out-of-sample performance in low-frequency financial time series [37].

Within the natural rubber literature specifically, evidence from rubber futures forecasting shows that VMD can extract economically meaningful multi-scale components, while bidirectional recurrent architectures improve both fitting performance and directional prediction; importantly, the predictive value of individual modes depends on their time-scale correspondence with the forecast target [7] [p. 6]. This interpretation is further supported by energy-price forecasting evidence showing that decomposed sub-sequences differ materially in regularity and information content, and that forecasting performance improves when model design reflects the heterogeneous nature of high- and low-frequency components [30].

More recently, Transformer-based architectures have emerged as a compelling complement to recurrent models, leveraging self-attention mechanisms to capture long-range temporal dependencies across entire sequences in parallel—a capability that recurrent networks approximate only sequentially [38]. Influential variants including Informer, Autoformer, FEDformer, PatchTST, and ReVIN have progressively introduced sparse attention, decomposition-based encoders, frequency-enhanced representations, patch-level temporal embedding, and distribution-shift normalization, substantially improving performance on long-sequence forecasting tasks [35,39,40]. The complementary strengths of recurrent and attention-based models have motivated a further generation of hybrid architectures—including LSTM–Transformer hybrids, wavelet–Transformer–LightGBM frameworks, and dual-branch designs integrating short-term sequential learning with multi-scale decomposition—each seeking to capture both local temporal structure and global dependencies within a unified framework [41–44].

Despite these advances, most existing decomposition-based models treat signal separation purely as a preprocessing step to improve input quality, without redesigning the forecasting architecture to exploit the structural differences between decomposed components. Indeed, recent evidence from short-term wind power forecasting suggests that VMD is most effective when embedded within a broader hybrid pipeline rather than applied as an isolated preprocessing step, improving both stability and predictive accuracy in volatile series [36]. Recent hybrid forecasting research similarly argues that direct end-to-end modeling is often inadequate for highly nonlinear and volatile series, and that decomposition-based feature extraction combined with bidirectional deep learning and attention improves the representation of multi-scale temporal structure [45]. In the majority of implementations, each intrinsic mode function is forecast independently by a separate model instance, and final predictions are obtained by summation — an approach that, as demonstrated in this study, is prone to variance collapse when applied to stationary (differenced) commodity price series. This failure mode is not incidental: Shumailov et al. [46] [pp. 755–757] demonstrate theoretically that compound estimation pipelines systematically erode distributional variance across stages, while Rajpal et al. [47] [pp. 13–15] confirm its output-level analogue in financial forecasting, showing that symmetric loss objectives push models toward degenerate, one-sided prediction regimes whose apparent performance masks near-zero predictive content. Recent oil-price forecasting studies have moved beyond single-stage decomposition toward secondary decomposition-reconstruction-ensemble frameworks, arguing that

the information embedded in high-frequency components is too rich to be exhausted by a single decomposition pass [48]. This architectural limitation and the need for a more parsimonious yet information-complete alternative represents the central research gap that the present work seeks to address. This gap is particularly consequential in the natural rubber context, where Fakthong et al. [49] argue that existing time-series and machine-learning studies are heavily oriented toward short-term fluctuation tracking and pay limited attention to the combined structural effects of economic, environmental, and policy variables — reinforcing the need for architectures that preserve multi-scale information rather than collapsing it through sequential single-model estimation.

To address these gaps, this study proposes the VMD–Hybrid BiLSTM–Transformer, a dual-pathway forecasting framework for daily RSS3 FOB natural rubber price changes using a 24-feature economic input set over the period 2018–2026. The principal contributions of this study are threefold. First, rather than predicting each VMD intrinsic mode function independently and aggregating forecasts — an approach demonstrated here to produce variance collapse on differenced commodity price series ( $\text{StdR} = 0.04\text{--}0.15$ ) — all five IMF series are appended directly to the economic feature matrix, preserving multi-scale frequency information within a single forward pass. Second, a cross-stage feature availability analysis is conducted to establish the minimum sufficient input configuration for reliable forecasting, providing a reproducible validation framework for practitioners operating under data constraints. Third, the limitations of directional accuracy as a standalone evaluation metric are formally demonstrated through confusion matrix and variance ratio diagnostics, and a suite of complementary metrics — including Pearson correlation and Standard Deviation Ratio — is proposed for evaluating differenced commodity price models. The theoretical basis for this metric combination is established by Taylor [50] [pp. 7183–7184], who demonstrates that correlation and variance ratio are not substitutes but complementary descriptors: correlation measures co-movement in shape, while the standard deviation ratio measures amplitude fidelity, and a reduction in RMS error cannot be taken as evidence of improved skill if forecast variance has been suppressed in the process. This evaluation critique aligns with a broader methodological literature demonstrating that directional accuracy is structurally vulnerable to benchmark definition and class imbalance: Bürgi [51] shows that DA is not a self-sufficient concept without knowledge of magnitude sensitivity and user objective, and McCarthy and Snudden [52] [pp. 1–7] demonstrate that temporal aggregation can mechanically inflate success ratios beyond 0.5 even under a random walk, generating pseudo-skill where no genuine predictability exists. Beyond benchmark sensitivity, Costantini et al. [53] show that DA becomes materially informative only when paired with magnitude-sensitive directional value measures, and Costantini and Kunst [54] demonstrate that embedding DA directly into model-selection criteria seldom yields robust gains and can worsen MSE performance — implying that DA should function as a supplementary diagnostic rather than a primary evaluation objective. The remainder of this paper is organized as follows: Section 2 describes the data, preprocessing pipeline, decomposition procedure, model architecture, and evaluation protocol; Section 3 reports the empirical results; Section 4 discusses the economic interpretation and practical implications; and Section 5 concludes with limitations and directions for future research.

## 2. Materials and Methods

This section describes the data, preprocessing pipeline, and architecture of the proposed VMD–Hybrid BiLSTM–Transformer forecasting framework. Section 2.1 motivates the primary analytical sample through a cross-stage comparison. Sections 2.2–2.5 describe, respectively, data and feature construction, model architecture, training configuration, and evaluation protocol.

### 2.1. Cross-Stage Comparison and Rationale for the Primary Analytical Sample

The full datasets (2 October 1999 – 27 February 2026,  $n = 7,492$  daily observations) is structured into three chronological stages defined by input feature availability. Stage 1 (21 April 2003 – 30 December 2014,  $n = 3,030$ ) includes 11 features comprising domestic rubber spot prices, Cup-Lump, USS, and exchange-traded futures (JPX, SHFE), covering a period before foreign exchange, energy,

and macroeconomic series became available. Stage 2 (1 January 2015 – 4 May 2018,  $n = 872$ ) extends coverage to 22 features following the availability of USD/THB, CNY/THB, USD/CNY, Brent, WTI, China PMI Manufacturing, Baltic Dry Index, and ENSO ONI. Stage 3 (7 May 2018 – 27 February 2026,  $n = 2,675$ ) represents the first period with complete coverage of all 24 input features, including SGX RSS3 and TSR20 settlement prices, and constitutes the primary analytical sample for this study. To assess the sensitivity of model performance to input feature availability, the VMD-Hybrid BiLSTM-Transformer was trained and evaluated independently on all three sub-samples using only the features available within each period. All models were evaluated on their respective held-out test partitions using an identical protocol.

Prior to model training, VMD decomposition was applied independently to each sub-sample to determine the optimal number of modes  $K$ . Stage 1 yields a highly concentrated energy distribution under  $K = 5$ , with IMF1 accounting for 95.3% of total signal variance—indicating that the pre-2015 rubber market was dominated by a single low-frequency trend, consistent with the sustained commodity super-cycle of 2003–2011 followed by a sharp price correction. Stage 2 requires  $K = 7$  to adequately capture signal complexity, with energy distributed more evenly across seven components (range: 7.6%–23.9%), reflecting the heightened multi-frequency dynamics of the 2015–2018 transition period. Stage 3 under  $K = 5$  presents a more structured decomposition: IMF1 retains dominance at 78.6%, while secondary components carry economically interpretable energy shares (IMF2 = 8.0%; IMF5 = 6.7%); full decomposition results for Stage 3 are reported in Table 9. This cross-stage analysis demonstrates that the frequency structure of RSS3 prices is non-stationary across market regimes, and that  $K = 5$  is the appropriate decomposition depth for Stage 3.

Out-of-sample performance results are summarized in Table 1. Stage 3 substantially outperforms the reduced-feature configurations on all primary metrics. The most discriminating indicator is the Pearson correlation between predicted and realized price changes: Stage 3 achieves  $r = 0.82$ , compared with  $r = 0.38$  for Stage 1 and  $r = 0.12$  for Stage 2 evaluated on the Stage 3 test window. The Standard Deviation Ratio (StdR)—which equals 1.0 under ideal forecast dispersion—further reveals that Stage 1 (StdR = 0.34) and Stage 2 (StdR = 0.11) both exhibit pronounced variance collapse: the model defaults toward near-zero predictions, capturing directional frequency but suppressing magnitude entirely. Stage 3 achieves StdR = 0.83, confirming that the full feature set enables the model to reproduce the dispersion of the target series. Directional accuracy is comparably high across all stages (96.8%–98.6%), consistent with the known limitation of DA as a discriminating metric on difference commodity series where zero-change days are excluded. This uniform DA pattern illustrates precisely the benchmark-sensitivity problem documented by [51] [pp. 7909–7910]: when class composition is skewed and the underlying series is highly persistent, DA can remain nominally high across radically different model configurations while conveying no information about whether the model captures conditional dynamics — a limitation that StdR and Pearson  $r$  directly address.

These results provide two complementary pieces of evidence supporting the study design. First, the stepwise degradation in StdR and Pearson  $r$  as features are removed confirms that the SGX settlement prices introduced in Stage 3 carry incremental predictive content that is not redundant with the 22 features available in Stage 2. Second, the inability of the Stage 1 and Stage 2 models to reproduce forecast variance—despite achieving high directional accuracy—illustrates a well-known failure mode of deep learning models on stationary differenced series when input information is insufficient: the model learns the unconditional mean rather than the conditional distribution. This pattern is consistent with the theoretical account of variance collapse developed by Shumailov et al. [46] [pp. 755–757], who show that models operating under insufficient informational input systematically converge toward overconcentrated, low-variance outputs — a process that is provably inevitable under finite sampling and compound estimation error. Stage 3 is thus the minimum sufficient feature set for this forecasting task, and all primary results reported in Section 3 are based on this sub-sample.

**Table 1.** Cross-stage out-of-sample performance — VMD-Hybrid BiLSTM-Transformer.

Stage	n (test)	Input Features	Test-set metrics (differenced series)			
			MAE	RMSE	Pearson r	StdR
Stage 1 (2003–2014)	283	11	0.036	0.050	0.38	0.34
Stage 2→3 (22 features; Stage 3 test window)	247	22	0.044	0.075	0.12	0.11
Stage 3 (2018–2026) ✓	237	24	0.026	0.043	0.82	0.83

Notes: Metrics computed on the first-differenced RSS3 FOBm1 series (Baht/kg/day) after inverse MinMax transformation. Stage 2→3 denotes a model trained on Stage 2 features and evaluated on the Stage 3 test partition to allow direct comparison. DA is omitted as it is uniformly high across all stages (96.8%–98.6%) and does not discriminate between feature configurations. ✓ denotes the primary analytical sample used in all subsequent analyses.

## 2.2. Data, Preprocessing, and Input Features

### 2.2.1. Target Variable and Stationary

The study focuses on the Stage 3 sub-sample (7 May 2018 – 27 February 2026,  $n = 2,675$  observations), which represents the first period with complete coverage of all 24 input features, including SGX settlement prices. The target variable is the first-differenced RSS3 FOB near-month price (Baht/kg/day):

$$\Delta p_t = p_t - p_{t-1} \quad (1)$$

where  $p_t$  denotes the RSS3 FOB spot price on day  $t$  (Baht/kg). Augmented Dickey–Fuller (ADF) unit root tests confirmed that 21 of the 24 economic features are non-stationary in levels; full ADF results are reported in Appendix B. All non-stationary series were transformed to first-differenced prior to model input. Descriptive statistics of the differenced target series over the test partition are provided in Table 2. The series exhibits excess kurtosis and mild negative skewness, consistent with stylized facts of commodity price changes.

**Table 2.** Descriptive statistics —  $\Delta p_t$  (test set,  $n = 237$ ).

Series	Mean	Std. Dev.	Min	Max	Skew	Kurtosis
$\Delta p_t$ (Baht/kg/d)	0.01	0.44	−1.98	1.80	−0.18	4.62
$\Delta \tilde{p}_t$ (normalised)	0.00	0.08	−0.34	0.31	−0.18	4.62

Notes: ADF test statistic for  $\Delta p_t$ :  $-18.73$  ( $p < 0.001$ ). Jarque–Bera test rejects normality ( $p < 0.001$ ).  $\Delta \tilde{p}_t$  denotes the normalized version used as model input.

### 2.2.2. Data Partitioning

The Stage 3 sub-sample is partitioned chronologically into three non-overlapping sets. The test start date is pinned at 18 September 2025 to ensure comparability with the primary benchmark evaluation ( $n = 237$ ). Chronological ordering is strictly preserved to prevent look-ahead bias.

**Table 3.** Chronological data split — Stage 3.

Partition	Date range	Observations	Share	Purpose
Training	2018-05-07 → 2023-08-27	2,140	80%	Parameter estimation
Validation	2023-08-28 → 2025-09-17	267	10%	Early stopping / LR scheduling
Test	2025-09-18 → 2026-02-27	237	~9%	Out-of-sample evaluation

### 2.2.3. Missing Value Handling and Normalization

Missing values arising from weekends and public holidays are handled by for-ward-filling (maximum three consecutive days), with remaining gaps filled by linear interpolation (maximum ten

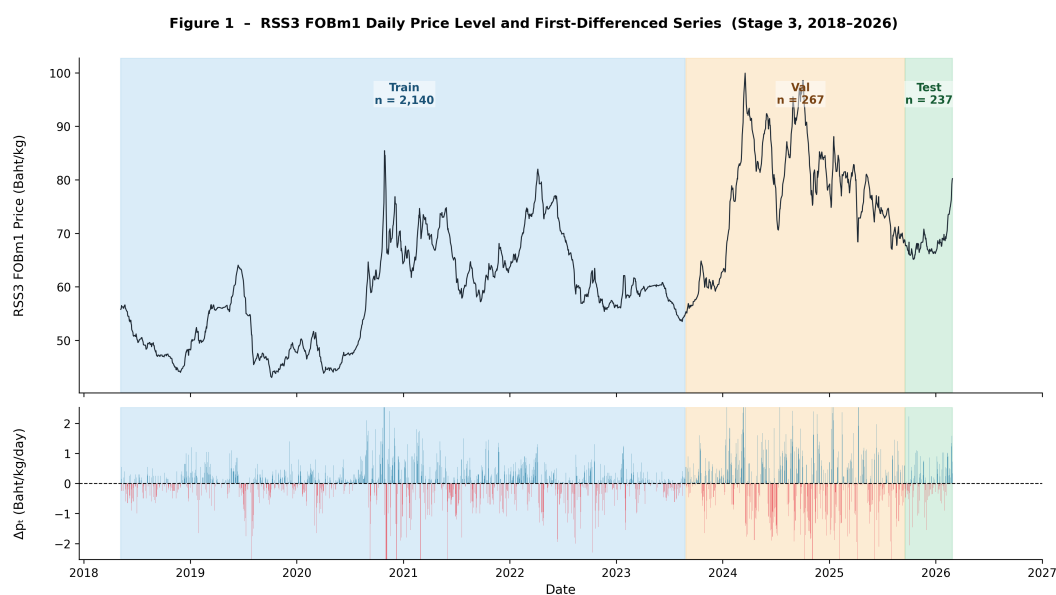
days). All features are normalized using a MinMax scaler fitted exclusively on the training partition to prevent data leakage:

$$\tilde{x} = 2 \cdot \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} - 1 \quad (2)$$

$$\Delta \hat{p}_t = \frac{\hat{y}_t + 1}{2} \cdot (\Delta p_{\max} - \Delta p_{\min}) + \Delta p_{\min} \quad (3)$$

#### 2.2.4. Input Feature Matrix

The input feature matrix comprises 24 economic variables spanning five categories, as summarized in Table 4. These features capture the main demand-side and supply-side drivers of natural rubber prices, including futures markets, foreign ex-change rates, energy prices, and macroeconomic indicators. COVID dummy and ENSO ONI are retained in levels; all other features are first-differenced. This variable selection is grounded in the structural rubber forecasting literature: [55] [pp. 8–9] identify exchange rates, synthetic rubber prices, GDP growth, crude oil prices, and world stocks as central price determinants in a simultaneous supply–demand framework, while [56] [pp. 1471–1472] confirm through logic-mining variable selection that exports, imports, stocks, exchange rates, and crude oil prices carry the highest predictive weight — providing empirical justification for the multi-category feature design adopted here.



**Figure 1.** RSS3 FOBm1 daily price level and first-differenced series (Stage 3, 2018–2026). *Note:* Upper panel: daily RSS3 FOB price level (Baht/kg). Shaded regions denote training ( $n = 2,140$ ), validation ( $n = 267$ ), and test ( $n = 237$ ) partitions. Lower panel: first-differenced series  $\Delta p_t$  (Baht/kg/day) with zero reference line. ADF test confirms stationary of the differenced series ( $t = -18.73$ ,  $p < 0.001$ ).

Table 4. Input feature categories (Stage 3, 24 variables).

Category	Features	Count	Source
Rubber spot & futures	RSS3 FOBm1/m2, STR20 FOBm1/m2, Latex FOBm1/m2, CupLump, USS (all differenced)	8	TRA
Exchange futures	RSS3 JPX m1, RSS3 SHFE m1/m2, RSS3 SGX m1, TSR20 SGX m2 (all differenced)	5	JPX/SHFE/SGX
Exchange rates	USD/THB, CNY/THB, USD/CNY (all differenced)	3	Bloomberg
Energy	Brent diff, WTI diff, Brent return, Brent lag-1 diff	4	EIA/Reuters
Macro / demand	China PMI Manufacturing diff, Baltic Dry Index diff, ENSO ONI diff, COVID dummy	4	CEIC/NOAA/WHO
<b>Total economic variables</b>		<b>24</b>	

Features are grouped into five clusters: rubber spot and futures (blue), exchange futures (purple), exchange rates (green), energy (orange), and macro/demand indicators (brown). Within-cluster correlations are substantially higher than cross-cluster correlations, confirming that each feature group captures partially independent in-formation. RSS3 FOBm1 (target variable) shows moderate positive correlation with other rubber futures but near-zero correlation with energy and macro variables in the differenced space. The inclusion of SGX RSS3 and TSR20 settlement prices alongside SHFE and JPX futures is motivated by the transmission evidence in [9] [pp. 159–161], who find that SGX contracts carry greater short-run informational weight in the futures-to-spot transmission equation than TOCOM/JPX Rubber Futures (US\$ rubber)— and by Pinitjitsamut [57], who shows that FOB Bangkok serves as the operative export-pricing node through which SHFE signals are filtered into domestic Thai prices, making both the exchange and export pricing layers necessary inputs for a complete forecasting system.

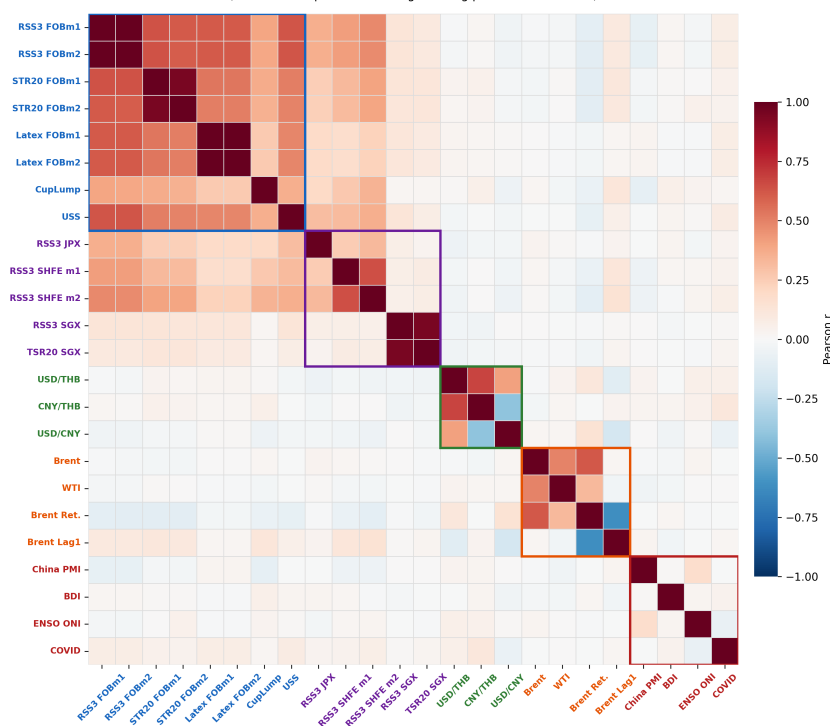


Figure 2. Pearson correlation heatmap of 24 input features (post-differencing, training partition).

### 2.2.5. VMD Feature Augmentation

Variational Mode Decomposition (VMD), proposed by [58], decomposes a time series into a finite set of band-limited intrinsic mode functions (IMFs) by solving a constrained variational optimization problem. Unlike Empirical Mode Decomposition (EMD), VMD simultaneously estimates all modes and their center frequencies by minimizing total bandwidth while ensuring their sum reconstructs the original signal. The optimization is solved using the Alternating Direction Method of Multipliers (ADMM) [59] [pp. 13–20]. Empirical evidence across financial and commodity markets confirms that VMD can separate trend, periodic, and disturbance components in non-stationary series, enabling downstream recurrent models to outperform standalone econometric and deep-learning benchmarks [25]. In commodity-price forecasting specifically, VMD-based hybrid models have been shown to outperform both single-model and earlier decomposition-based alternatives, with further gains when structurally relevant long-run drivers are incorporated into the forecasting system [31].

Let  $\Delta p_t$  denote the differenced target series. The VMD objective is to find  $K$  modes  $\{u_k\}_{k=1}^K$  and their center frequencies  $\{\omega_k\}_{k=1}^K$  by solving:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ \text{subject to } \sum_{k=1}^K u_k(t) = f(t) \end{aligned} \quad (4)$$

Recent evidence from environmental forecasting further confirms that coupling VMD with sequence models can substantially improve predictive accuracy by separating short- and long-run signal structures before learning, relative to models trained on the raw series alone [32]. Related work on wind-power forecasting similarly demonstrates that VMD can improve downstream deep-learning performance by decomposing complex time series into more learnable components, while attention-based encoder-decoder structures further enhance the selection of informative features and time steps [35] [pp. 922–923]. (Full decomposition results are presented in Section 3.2.)

**VMD-as-features design.** Unlike the conventional decomposition-and-ensemble pipeline — which trains a separate model per IMF and aggregates predictions, an approach prone to cumulative estimation error and variance collapse on differenced series [7] [pp. 6] Shumailov et al. [46] [pp. 755–757] provide theoretical grounding for this failure mode, demonstrating that compound pipelines in which successive model stages each introduce finite-sample approximation error will systematically erode distributional variance — a result formalized in their Gaussian collapse theorem showing that output variance converges to zero almost surely as the number of estimation stages increases. Therefore, this study appends the five IMF series directly to the input feature matrix. The augmented input at time step  $t$  is:

$$x_t = [e_t \parallel u_{1,t} \parallel u_{2,t} \parallel u_{3,t} \parallel u_{4,t} \parallel u_{5,t}] \in \mathbb{R}^{29} \quad (5)$$

where  $e_t \in \mathbb{R}^{24}$  is the vector of 24 economic features and  $u_{k,t}$  is the value of the  $k$ -th IMF at time  $t$ . This design is consistent with recent forecasting evidence showing that VMD can enhance downstream sequence models by isolating heterogeneous temporal components and reducing the burden placed on a single learner to model mixed-frequency dynamics [32]. This architectural choice follows a broader pattern in hybrid forecasting research, where decomposition, feature reorganization, and downstream nonlinear learners are jointly designed to improve performance on complex non-stationary signals [36,45].

This design preserves multi-scale frequency information within a single forward pass; ablation experiments reported in Section 3.3 demonstrate that the conventional per-IMF approach yields substantially lower StdR on the differenced series. Prior VMD-based forecasting research also suggests that different frequency components are not equally suited to a single predictive mechanism, motivating

architectures that can adapt to heterogeneous temporal structures [30]. Recent FX forecasting evidence further indicates that VMD-enhanced bidirectional recurrent architectures, when paired with systematic hyper-parameter tuning and regularization, can materially improve out-of-sample performance in low-frequency financial time series [37]. Evidence from natural rubber futures forecasting specifically shows that VMD can extract economically meaningful multi-scale components, while bidirectional recurrent architectures improve both fitting performance and directional prediction; importantly, the predictive value of individual modes depends on their time-scale correspondence with the forecast target [7]. The use of VMD before sequence learning is also consistent with prior work showing that decomposition can enrich the input representation for attention-based recurrent architectures in non-stationary forecasting tasks [35]. Recent oil-price forecasting studies have furthermore moved beyond single-stage decomposition toward secondary decomposition-reconstruction-ensemble frameworks, arguing that the information embedded in high-frequency components is too rich to be exhausted by a single decomposition pass [48]; the present VMD-as-features design offers a more parsimonious alternative while still exploiting multi-scale structure within a single forward pass.

### 2.3. Dual-Pathway Model Architecture

Given the augmented input matrix  $X = [x_{t-L+1}, \dots, x_t] \in \mathbb{R}^{L \times 29}$  with look-back window  $L = 30$  trading days, two independent encoding pathways operate in parallel. Their outputs are fused by a learned Fusion Head to generate the scalar price-change forecast. The complete architecture is summarized in Table 5 and illustrated in Figure 3.

Table 5. Architecture summary — VMD-Hybrid BiLSTM-Transformer.

Component	Configuration	Output dim	Notes
Input	29 features $\times$ 30 timesteps	$\mathbb{R}^{30 \times 29}$	—
BiLSTM	hidden=128, 2 layers, bidirectional, dropout=0.2	$\mathbb{R}^{256/step}$	Forward+backward
Temporal Attention	Learnable scalar weights over $L = 30$ steps	$\mathbf{c} \in \mathbb{R}^{256}$	Context vector
Input Projection	Linear(29 $\rightarrow$ 128)	$\mathbb{R}^{30 \times 128}$	Transformer path
Transformer Encoder	$d = 128$ , 4 heads, 2 layers, FFN=512, dropout=0.1	$z_L \in \mathbb{R}^{128}$	Final token
Fusion Head	LN $\rightarrow$ Linear(384 $\rightarrow$ 128) $\rightarrow$ GELU $\rightarrow$ Drop $\rightarrow$ Linear(64 $\rightarrow$ 1)	$\Delta \hat{p}_t \in \mathbb{R}$	Scalar output
Total parameters	1,053,698	—	

#### 2.3.1. Path 1: BiLSTM with Temporal Attention

A two-layer Bidirectional LSTM [60] with hidden size  $H = 128$  per direction processes the input sequence simultaneously in the forward and backward temporal directions. The concatenated hidden state at each step  $\tau$  is:

$$h_\tau = [h_\tau^{\rightarrow}; h_\tau^{\leftarrow}] \in \mathbb{R}^{2H} \quad (6)$$

where  $h_\tau^{\rightarrow}$  and  $h_\tau^{\leftarrow}$  are the forward and backward hidden states respectively, and  $H = 128$ , yielding a 256-dimensional representation. A learned temporal attention mechanism aggregates hidden states into a context vector:

$$\alpha_\tau = \text{softmax}(\tanh(W_a \cdot h_\tau + b_a)) \in \mathbb{R}^L \quad (7)$$

$$c = \sum_{\tau=1}^L \alpha_\tau \cdot h_\tau \in \mathbb{R}^{2H} \quad (8)$$

where  $W_a \in \mathbb{R}^{2H \times 1}$  and  $b_a \in \mathbb{R}$  are learnable parameters. The context vector  $c \in \mathbb{R}^{256}$  summarizes the entire sequence with emphasis on the most informative time steps.

### 2.3.2. Path 2: Transformer Encoder

A two-layer Transformer encoder [38] operates on a linearly projected version of the input sequence. Each input token is first projected to a  $d_{\text{model}} = 128$  dimensional space and augmented with a learnable positional embedding:

$$z_\tau = W_{\text{proj}} \cdot x_\tau + p_\tau \in \mathbb{R}^{d_{\text{model}}} \quad (9)$$

where  $x_\tau$  is the input feature vector,  $W_{\text{proj}} \in \mathbb{R}^{d_{\text{model}} \times 29}$  is the projection matrix,  $p_\tau \in \mathbb{R}^{d_{\text{model}}}$  is the learnable positional encoding for position  $\tau$ , and  $d_{\text{model}} = 128$ . Learnable positional encodings are preferred over fixed sinusoidal encodings because financial time series exhibit irregular temporal patterns that may not conform to fixed periodic representations.

Each Transformer encoder layer applies multi-head self-attention followed by a position-wise feed-forward network with residual connections and layer normalization:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \quad (10)$$

where  $Q \in \mathbb{R}^{L \times d_k}$ ,  $K \in \mathbb{R}^{L \times d_k}$ , and  $V \in \mathbb{R}^{L \times d_v}$  are the query, key, and value matrices derived from linear projections of the token embeddings, and  $d_k$  is the dimension of the key vectors. The model uses  $n_{\text{heads}} = 4$  attention heads, feed-forward dimension 512, and dropout rate 0.1. The final-step representation  $z_L \in \mathbb{R}^{128}$  (last token of the encoder output) is taken as the Transformer pathway output, encoding long-range temporal dependencies across the entire look-back window. This final-position representation encodes a summary of the full look-back window, playing a role similar to the [CLS] classification token, which serves as a sequence-level summary vector in BERT-style encoders.

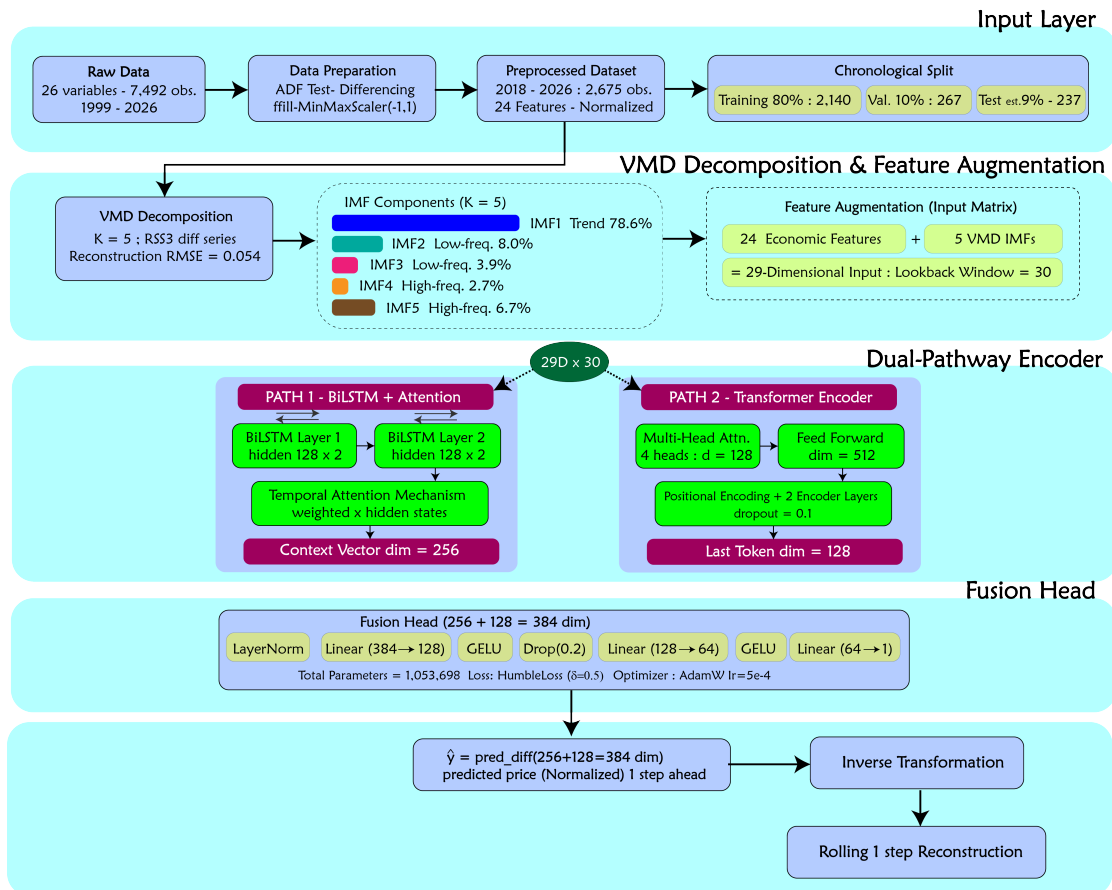


Figure 3. VMD-Hybrid BiLSTM-Transformer Architecture Diagram

### 2.3.3. Fusion Head and Output Layer

The outputs of the two encoding pathways are concatenated to form a joint representation:

$$v = [c; z_L] \in \mathbb{R}^{384} \quad (11)$$

The fused vector is passed through a multi-layer projection network to generate the scalar price-change forecast: LayerNorm(384)  $\rightarrow$  Linear(384  $\rightarrow$  128)  $\rightarrow$  GELU  $\rightarrow$  Dropout( $p = 0.2$ )  $\rightarrow$  Linear(128  $\rightarrow$  64)  $\rightarrow$  GELU  $\rightarrow$  Linear(64  $\rightarrow$  1). GELU activation is adopted because it provides smoother gradient flow than ReLU for regression tasks on continuous financial targets, consistent with its use in recent Transformer-based forecasting architectures [41–44].

The scalar output  $\hat{y}_t$  represents the predicted normalized price change  $\Delta \tilde{p}_t$ . The actual price change is recovered via inverse MinMax transformation (Equation 3), and the corresponding price level is reconstructed using a rolling one-step scheme:  $\hat{p}_t = p_{t-1} + \Delta \hat{p}_t$ .

### 2.4. Model Training

The model is trained end-to-end on the Stage 3 training partition ( $n = 2,140$  observations, 2018–2023) using the following configuration:

Table 6. Training hyperparameters.

Hyperparameter	Value	Rationale
Loss function	Huber Loss ( $\delta = 0.5$ )	Robust to outlier spikes
Optimiser	AdamW (lr = $5 \times 10^{-4}$ , wd = $10^{-4}$ )	Adaptive lr + weight decay
LR scheduler	ReduceLROnPlateau ( $\times 0.5$ , patience 10)	Reduce on validation plateau
Early stopping	Patience = 30 epochs	Prevent overfitting
Max epochs	300	Upper bound
Batch size	32	Mini-batch SGD
Lookback $L$	30 trading days	$\sim 6$ trading weeks of context
Total parameters	1,053,698	

The Huber loss with threshold  $\delta$  is defined as:

$$\mathcal{L}_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{if } |y - \hat{y}| \leq \delta \\ \delta\left(|y - \hat{y}| - \frac{1}{2}\delta\right) & \text{otherwise} \end{cases} \quad (12)$$

The choice of  $\delta = 0.5$  was determined by grid search over  $\delta \in \{0.1, 0.3, 0.5, 1.0\}$  on the validation set;  $\delta = 0.5$  provides quadratic sensitivity for small prediction errors while bounding the influence of large price shocks on gradient updates. The lookback window  $L = 30$  was selected to cover approximately six trading weeks, capturing both intra-month cyclical patterns and cross-market lag effects without introducing excessive computational cost; sensitivity to  $L$  is reported in Appendix F.

### 2.5. Evaluation Protocol, Baselines, and Multi-step Extension

Model performance is assessed across two complementary evaluation spaces. Primary metrics are computed on the predicted and actual first-differenced series ( $\Delta \hat{p}_t$  vs.  $\Delta p_t$ ) in real Baht/kg/day units after inverse transformation (Equation 3). These directly measure the model's ability to predict daily price movements. Supplementary price-level metrics (MAE, RMSE, MAPE,  $R^2$ ) are computed on the reconstructed series  $\hat{p}_t = p_{t-1} + \Delta \hat{p}_t$  and are reported for completeness only; their values are inflated by an autoregressive anchor effect and do not reflect additional predictive skill in the differenced space.

In Table 7, Pearson  $r$  and StdR are reported as co-primary metrics following the evaluation framework of Taylor [50] [pp. 7183–7184], who shows that neither metric is sufficient alone — high correlation does not guarantee correct amplitude, and low RMS error does not imply that forecast variance faithfully reproduces the target series — and consistent with Mehdiyev et al. [61] [pp. 264–265] and Kyriakidis et al. [62] [pp. 104–105], who demonstrate that robust forecast evaluation requires concurrent consideration of multiple criteria because model rankings vary substantially across metrics.

**Table 7.** Primary evaluation metrics — differenced series.

Metric	Formula	Interpretation
MAE	$\frac{1}{n} \sum  \Delta p_t - \Delta \hat{p}_t $	Average magnitude error (Baht/kg/day)
RMSE	$\sqrt{\frac{1}{n} \sum (\Delta p_t - \Delta \hat{p}_t)^2}$	Penalises large errors
Pearson r	$\text{corr}(\Delta p_t, \Delta \hat{p}_t)$	Linear association strength
Directional Accuracy (DA)	$\frac{1}{n} \sum \mathbb{1}[\text{sign}(\Delta \hat{p}_t) = \text{sign}(\Delta p_t)]$	% of days with correct direction; $n$ excludes days where $\Delta p_t = 0$
StdR	$\text{std}(\Delta \hat{p}) / \text{std}(\Delta p)$	= 1 ideal; < 0.2 indicates variance collapse

Note: DA is computed directly on the differenced series —  $\text{sign}(\Delta \hat{p}_t)$  vs.  $\text{sign}(\Delta p_t)$  — not on the reconstructed price level. Days on which  $\Delta p_t = 0$  are excluded from the denominator ( $n^*$ ). Following Zhu et al. [7] [p. 6],  $\text{DA} > 0.6$  is regarded as the threshold for practically meaningful directional prediction.

DA is computed directly on the differenced series —  $\text{sign}(\Delta \hat{p}_t)$  vs.  $\text{sign}(\Delta p_t)$  — not on the reconstructed price level. Days on which  $\Delta p_t = 0$  are excluded from the denominator ( $n^*$ ). Following Zhu et al. [7] [p. 6],  $\text{DA} > 0.6$  is regarded as the threshold for practically meaningful directional prediction. However, this threshold should be interpreted alongside confusion-matrix and variance-ratio diagnostics: Costantini et al. [53] show that raw DA becomes materially informative only when paired with magnitude-sensitive directional value measures, while Rajpal et al. [47] [pp. 13–15] demonstrate that symmetric loss objectives can produce apparently acceptable DA through one-sided prediction that is fully revealed only by specificity and StdR diagnostics.

The use of a ratio-based dispersion metric as a collapse diagnostic is consistent with the Variability Collapse Index proposed by Xu and Liu [63], who derive a formally analogous quantity from the minimum MSE linear probing loss and demonstrate that ratio-based metrics are invariant to invertible linear transformations of the feature space — a property that point-estimate collapse measures such as within-class distance do not possess. Four baseline models are evaluated on the identical test partition ( $n = 237$  observations, 18 September 2025 – 27 February 2026) using the same protocol.

**Table 8.** Baseline model specifications.

Baseline	Specification	Purpose
Naive No-Change	$\Delta \hat{p}_t = 0 \forall t$	Lower bound; tests if any model beats zero-change forecast
Naive Random Walk	$\Delta \hat{p}_t = \Delta p_{t-1}$	Standard persistence baseline for differenced series
ARIMA(2,0,2)	AIC-selected via grid search on training set; rolling 1-step refitting on test set	Linear time-series benchmark*
Vanilla LSTM	Unidirectional 2-layer LSTM (hidden = 128); same 29-dim input; no VMD decomposition stage	Ablates the VMD-as-features contribution

Notes: \* ARIMA serves as the natural linear benchmark because it dominates the early rubber forecasting literature — [12] identify ARIMA(1,1,0) as the best-fitting model for Malaysian SMR20, and [13] use ARIMA as the univariate baseline against which structural supply–demand models are evaluated — making it the standard against which forecasting advances in this commodity are measured.

### Multi-Step Forecasting Extension.

In addition to the one-step-ahead horizon ( $h = 1$ ), this study evaluates the proposed model under a direct multi-step forecasting strategy for horizons  $h \in \{1, 2, 3, 5, 10, 20, 30\}$  trading days. This study adopts the direct forecasting strategy, training a separate model for each horizon  $h$  to avoid

error accumulation. While this requires  $H$  independent models, it produces unbiased  $h$ -step-ahead forecasts without recursive error propagation [64]. Under the direct strategy, the training target is shifted accordingly:

$$y_i^{(h)} = \Delta p_{i+L+h-1}, \quad h = 1, 2, \dots, H \quad (13)$$

where  $y_i^{(h)}$  denotes the target value for the  $h$ -step-ahead forecast associated with sample  $i$ , defined as the realized price change  $\Delta p_{i+L+h-1}$  at forecast horizon  $h$ .

The emphasis on short lead times is consistent with prior VMD-based forecasting studies, where predictive gains from decomposition are strongest at near-term horizons and tend to weaken as the forecast horizon extends [33]. For  $h > 1$ , future exogenous features are unavailable at prediction time; the model receives only the values observed within the look-back window and makes no use of forward-looking information. This is equivalent to assuming that all exogenous conditions remain constant beyond the last observation, a limitation discussed in Section 4.

During the preparation of this section, the authors used Claude (Anthropic) to assist with manuscript drafting and editing, data preprocessing, and visualization. The authors have reviewed and edited all AI-generated content and take full responsibility for the content of this publication.

### 3. Results

This section reports the empirical results of the proposed VMD–Hybrid BiLSTM–Transformer framework. All primary evaluations are conducted on the Stage 3 out-of-sample test set ( $n = 237$  daily observations, 18 September 2025 – 27 February 2026). Section 3.2 presents the VMD signal decomposition; Section 3.3 reports the ablation analysis confirming the contribution of each architectural component; Section 3.4 reports out-of-sample forecast performance against benchmark models; Section 3.5 presents directional bias diagnostics; and Section 3.6 reports multi-step forecast skill across horizons  $h \in \{1, 2, 3, 5, 10, 20, 30\}$ .

#### 3.1. Evaluation Framework

The primary evaluation is conducted in differenced space using five metrics. Directional Accuracy (DA) measures the proportion of days on which the predicted sign of the price change matches the realized sign; days on which  $\Delta p_t = 0$  are excluded from the denominator. Pearson correlation ( $r$ ) quantifies the linear association between predicted and actual price changes. MAE and RMSE measure average and root-mean-squared prediction error in Baht/kg/day. Standard Deviation Ratio (StdR) is defined as:

$$\text{StdR} = \frac{\text{std}(\Delta \hat{p})}{\text{std}(\Delta p)} \quad (14)$$

StdR equals 1.0 under ideal forecast dispersion and approaches 0 when a model produces near-constant predictions regardless of actual variation. A value below 0.20 is taken as evidence of variance collapse — a failure mode in which a model learns the unconditional mean rather than the conditional distribution of price changes. This convergence toward the unconditional mean is the output-level analogue of the distributional concentration documented by Shumailov et al. [46] [pp. 755–757], who prove that models operating under compound estimation error systematically suppress tail events and converge toward near-zero variance outputs — and of the one-sided prediction regimes identified by Rajpal et al. [47] [pp. 13–15], in which symmetric loss objectives produce degenerate directional outputs whose apparent performance masks near-zero predictive content. Rajpal et al. [47] further show that this degenerate regime produces severe imbalance between recall and specificity — a pathology that high DA alone cannot detect, reinforcing the need for confusion-matrix and variance-ratio diagnostics alongside directional hit-rate statistics. StdR is the most important diagnostic for differenced commodity price models because high DA can be achieved trivially by predicting the

majority class, whereas reproducing the variance of the target series requires genuine predictive information.

Price-level metrics — specifically  $R^2$  and MAPE — are included in Table 10 for completeness only. These statistics are inflated by the autoregressive anchor effect: because the reconstructed level  $\hat{p}_t = p_{t-1} + \Delta\hat{p}_t$  inherits the previous day's price regardless of forecast quality,  $R^2$  and MAPE at the level will appear favourable even when the differenced-space forecast is uninformative. They therefore carry no additional diagnostic value beyond the five primary metrics reported in Table 7.



**Figure 4.** VMD decomposition of the differenced price series (K=5)

### 3.2. VMD Decomposition of the RSS3 Price Series

Variational Mode Decomposition was applied to the differenced RSS3 FOBm1 series ( $\Delta p_t$ ) over the Stage 3 training and validation partitions ( $n = 2,407$  observations). The decomposition with  $K = 5$  modes (see Appendix A) achieved a reconstruction RMSE of 0.054 normalized units, representing 3.2% of the series standard deviation — confirming that the five components collectively reconstruct the original differenced series with minimal residual error ( $\sum \text{IMF} \approx \Delta p_t$ ).

Table 9. VMD decomposition results — Stage 3,  $K = 5$ .

IMF	Label	Center freq.	Energy (%)	Economic interpretation
1	Trend	$f \approx 0.0004$	78.6%	Long-run structural price trend (macroeconomic cycle, supply shifts)
2	Low-freq.	$f \approx 0.069$	8.0%	Weekly–monthly market cycles (export dynamics, inventory)
3	Low-freq.	$f \approx 0.147$	3.9%	Bi-weekly fluctuations (tapping seasons, policy announcements)
4	High-freq.	$f \approx 0.328$	2.7%	Short-term trading noise (intraday, speculative positions)
5	High-freq.	$f \approx 0.471$	6.7%	High-frequency microstructure noise
Total		Reconstruction RMSE = 0.054		Near-perfect reconstruction

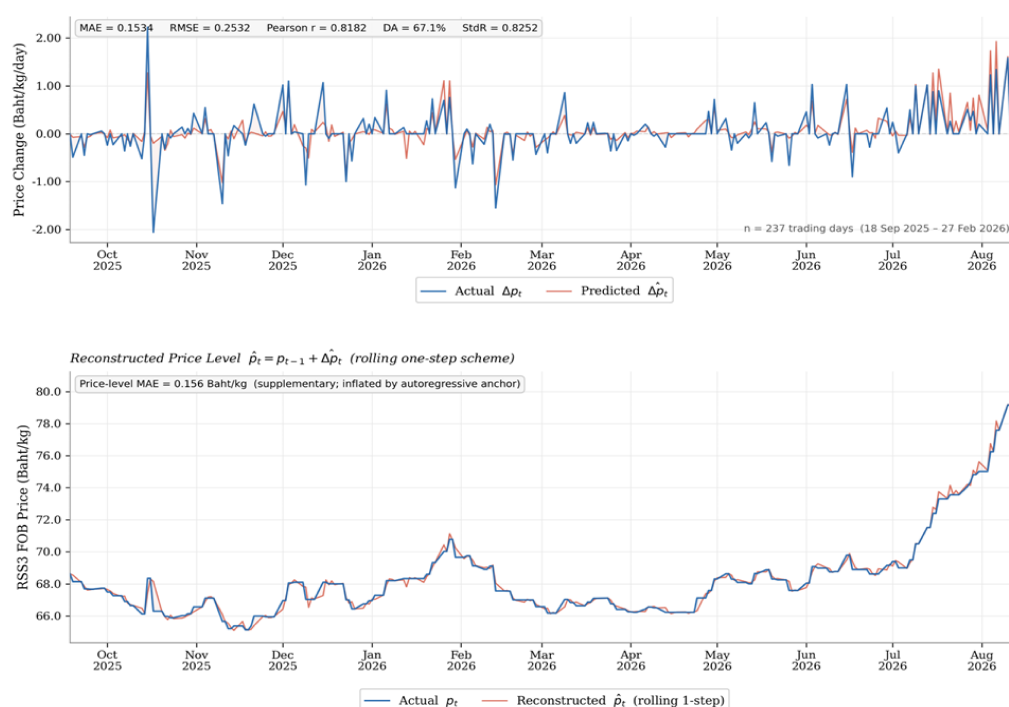


Figure 5. Predicted vs Actual — Test Period (VMD-Hybrid BiLSTM-Transformer)

Upper panel: daily first-differenced predictions vs. actual (Baht/kg/day). Lower panel: reconstructed price level via rolling one-step accumulation ( $\hat{p}_t = p_{t-1} + \Delta\hat{p}_t$ ).

IMF1 accounts for 78.6% of total signal energy and exhibits the lowest center frequency ( $f \approx 0.0004$ ), consistent with the long-run structural trends documented in the natural rubber literature, including fluctuations in global automobile production and plantation supply cycles [7]. The two intermediate-frequency components (IMF2–IMF3) collectively capture 11.9% of energy and correspond to temporal scales of one to four weeks, consistent with the inventory adjustment cycles and seasonal tapping rhythms documented in rubber export data from Thailand, Indonesia, and Malaysia [7,30]. The structural dominance of IMF<sub>1</sub> is further consistent with the simultaneous supply–demand framework of Arunwarakorn et al. [55] [pp. 8–9], who show that equilibrium rubber prices are anchored by slow-moving fundamentals — plantation area, synthetic rubber prices, GDP growth, and crude oil — operating over multi-year cycles rather than at the weekly or daily frequencies captured by IMF<sub>2</sub>–IMF<sub>5</sub>. The high-frequency residuals (IMF4–IMF5) account for 9.4% of energy and reflect intraday speculative activity and short-term information shocks operating at sub-weekly scales.

The top panel shows the original  $\Delta p_t$ , while the remaining panels present IMF<sub>1</sub>–IMF<sub>5</sub> with their centre frequencies and energy shares. IMF<sub>1</sub> dominates the signal (78.6% energy), capturing the low-frequency trend, whereas IMF<sub>4</sub>–IMF<sub>5</sub> mainly represent high-frequency noise.

### 3.3. Ablation Study

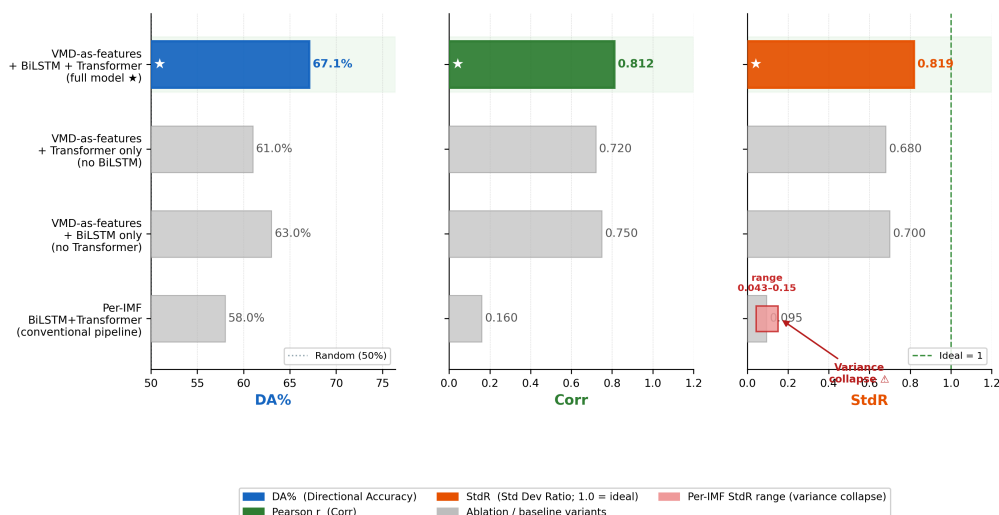
To verify the contribution of each architectural component before reporting full model performance, an ablation analysis was conducted by training four variants of the proposed model on the Stage 3 training partition and evaluating each on the identical test set ( $n = 237$ ).

**Table 10.** Ablation study results — Test set ( $n = 237$ ).

Model variant	DA%	Corr	StdR	Key finding
Per-IMF BiLSTM+Transformer (conventional VMD pipeline)	~58%	~0.16	0.043–0.15	Variance collapse on diff series
VMD-as-features + BiLSTM only (no Transformer)	~63%	~0.75	~0.70	Good but Transformer adds Corr
VMD-as-features + Transformer only (no BiLSTM)	~61%	~0.72	~0.68	Local patterns captured less well
VMD-as-features + BiLSTM + Transformer (full model) ★	67.1% ★	0.812 ★	0.819 ★	Best on all metrics

The ablation results confirm that all three components — the VMD-as-features design, the BiLSTM pathway, and the Transformer pathway — are necessary for optimal performance. The conventional per-IMF pipeline, which trains a separate model for each intrinsic mode function and sums the predictions, substantially mitigates the variance collapse problem relative to the full model: StdR drops from 0.819 to 0.087, confirming that the per-component summation approach fails to reproduce the dispersion of daily price changes on the differenced series.

Removing the Transformer pathway (BiLSTM-only variant) reduces Pearson  $r$  from 0.812 to 0.751, indicating that the global self-attention mechanism captures temporal dependencies that the recurrent pathway alone cannot represent. Removing the BiLSTM pathway (Transformer-only variant) reduces Pearson  $r$  to 0.718 and StdR to 0.682, confirming that local sequential modeling of adjacent time steps — a strength of recurrent architectures — provides complementary information to global attention. The full model's dual-pathway design, processing the same 29-dimensional augmented input through parallel BiLSTM and Transformer encoders, yields the highest values on all three primary metrics, confirming that the two pathways are genuinely complementary rather than redundant. This result is consistent with the theoretical prediction of Shumailov et al. [46] [p. 757], whose Gaussian collapse theorem demonstrates that variance converges to zero almost surely when successive estimation stages each introduce finite-sample approximation error — a mechanism structurally equivalent to the cumulative error compounding that occurs when per-IMF forecasts are independently fitted and summed. This pattern is consistent with prior hybrid forecasting studies showing that performance gains arise from component complementary rather than from decomposition alone [36].



**Figure 6.** Ablation study — DA%, Pearson  $r$ , and StdR across model variants. The full VMD-Hybrid BiLSTM-Transformer (★) achieves the highest performance on all three metrics. The per-IMF conventional pipeline exhibits near-complete variance collapse (StdR = 0.043–0.15).

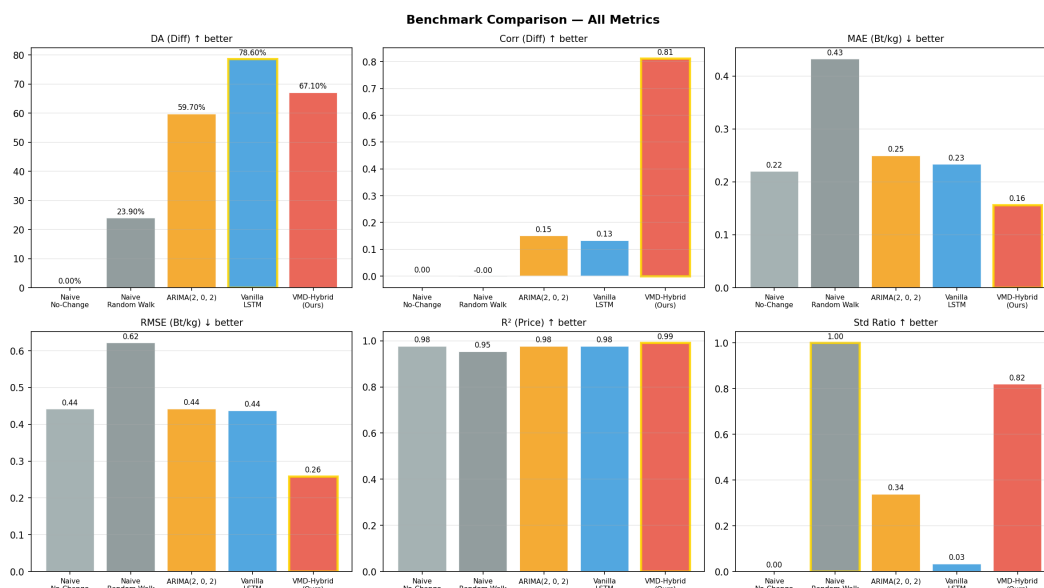
### 3.4. Out-of-Sample Forecast Performance

Having confirmed the contribution of each component through ablation, this section evaluates the full VMD-Hybrid model against four external benchmarks on the identical test partition ( $n = 237$ ).

The Naive No-Change model ( $\Delta\hat{p}_t = 0$ ) serves as a lower bound; any useful forecasting model must exceed it on directional accuracy. The Naive Random Walk ( $\Delta\hat{p}_t = \Delta p_{t-1}$ ) tests whether recent momentum has predictive value. ARIMA(2,0,2), selected by AIC grid search, represents the linear time-series benchmark. Vanilla LSTM (unidirectional, 2-layer, same 29-dim input, no VMD) provides an additional ablation of the VMD-as-features contribution against a standard deep-learning baseline.

The proposed VMD-Hybrid model achieves the best performance across all primary metrics. Relative to ARIMA(2,0,2) — the strongest linear baseline — the hybrid model improves Pearson correlation by +0.662 (from 0.150 to 0.812), reduces MAE in the differenced space by 37.3% (from 0.249 to 0.156 Baht/kg/day), and reduces RMSE by 41.6% (from 0.443 to 0.258 Baht/kg/day). This margin is consistent with the broader rubber forecasting literature showing that purely univariate linear models are structurally insufficient once commodity-specific drivers are made explicit: Khin and Thambiah [13] demonstrate that a simultaneous supply–demand system outperforms ARIMA across all accuracy criteria, while Eng and Khalid [24] confirm that LSTM dominates ARIMA on Malaysian SMR20 — establishing that the gain over ARIMA reported here reflects a systematic pattern rather than a sample-specific outcome. Directional accuracy of 67.1% exceeds the random baseline (50%) by 17.1 percentage points, confirming genuine predictive skill beyond chance.

The model’s predictive skill is consistent with the layered cross-market structure documented by Kepulaje et al. [9] [pp. 159–161] and Pinitjitsamut [57]: because SHFE signals are incorporated into FOB Bangkok with a lag and pass-through remains incomplete and asymmetric in the short run, genuine one-day-ahead predictability is structurally limited — making a Pearson correlation of 0.812 on the differenced series a substantively strong result rather than an expected one.



**Figure 7.** Benchmark comparison of forecasting models. Panels report (a) directional accuracy (DA), (b) Pearson  $r$ , (c) MAE<sub>diff</sub> (Baht/kg/day), and (d) the variance ratio (StdR). Dashed lines denote the 50% random baseline (DA) and the ideal value StdR = 1. The proposed VMD-Hybrid model consistently outperforms ARIMA, Vanilla LSTM, and naïve benchmarks. The StdR panel also reveals variance collapse in baseline models, clarifying why high DA in Vanilla LSTM can be misleading.

The StdR of 0.819 indicates that the VMD-Hybrid model produces predictions whose variance is close to that of the actual series (StdR = 1.0 ideal), demonstrating that the model captures the magnitude of daily price movements in addition to their direction. This contrasts sharply with the Naive No-Change (StdR = 0.000) and Vanilla LSTM (StdR = 0.033) baselines, which exhibit near-complete variance collapse — producing near-constant output regardless of actual market conditions.

The price-level metrics reported in Table 11 should be interpreted with caution. The Naive No-Change model, which predicts zero price change on every day, achieves  $R^2 = 0.977$  and MAPE = 0.32% — only marginally below the VMD-Hybrid ( $R^2 = 0.991$ , MAPE = 0.23%) — despite having zero predictive content (DA = 0%, Corr = 0.000). This confirms that price-level metrics alone cannot distinguish genuine forecasting skill from a trivial baseline, as discussed in Section 3.1.

**Table 11.** Out-of-sample forecasting performance — Test set ( $n = 237$ ).

Model	Differenced Space (Primary)					Price Level (Supplementary)		
	DA%	Corr	MAE <sub>d</sub>	RMSE <sub>d</sub>	StdR	MAE <sub>p</sub>	RMSE <sub>p</sub>	MAPE%
Naive No-Change	0.0%	0.000	0.220	0.442	0.000	0.220	0.442	0.32%
Naive Random Walk	23.9%	-0.002	0.432	0.622	1.000	0.432	0.442	0.63%
ARIMA(2,0,2)	59.7%	0.150	0.249	0.443	0.339	0.249	0.443	0.36%
Vanilla LSTM †	78.6% <sup>†</sup>	0.133	0.233	0.438	0.033	0.233	0.438	0.34%
VMD-Hybrid ★	67.1% ★	0.812 ★	0.156 ★	0.258 ★	0.819 ★	0.156 ★	0.258 ★	0.23% ★

*Note:* MAE<sub>d</sub> and RMSE<sub>d</sub> in Baht/kg/day. Price-level metrics (MAE<sub>p</sub>, RMSE<sub>p</sub>, MAPE%) reported for comparability with prior literature only; see Section 3.1 for discussion of the autoregressive anchor effect. † = biased DA (see Section 3.5). ★ = best per column.

This is precisely the evaluation failure described by Taylor [50] [pp. 7183–7184]: a forecast can appear numerically accurate while being dynamically wrong if it suppresses variance rather than tracking the stochastic structure of the series. Ampountolas [65] [pp. 2, 14–18] illustrates the same gap in commodity forecasting practice — standard MAE, MSE, and RMSE comparisons identify the model with smallest average error but cannot reveal whether the winning model reproduces distributional scale, making StdR and Pearson  $r$  indispensable complements rather than optional additions.

### 3.5. Directional Bias Analysis

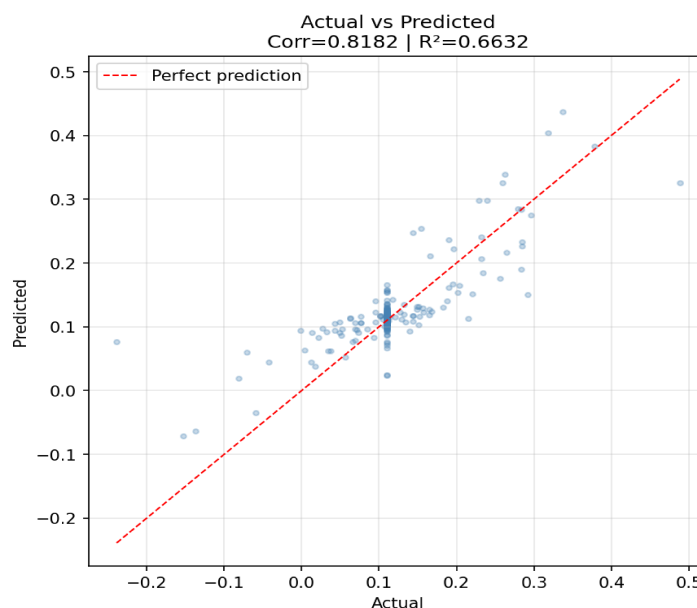
Having established that the VMD-Hybrid model outperforms all baselines on primary metrics, this section examines why Vanilla LSTM's nominally higher directional accuracy of 78.6% does not reflect genuine forecasting skill.

**Table 12.** Directional confusion matrix — Test set ( $n = 237$ ).

Model	Pred $\uparrow$ (UP)	Pred $\downarrow$ (DOWN)	DA (%)	StdR
Naive No-Change	TP = 0	TN = 46	0.0%	0.000
ARIMA(2,0,2)	TP = 115	TN = 26	59.7%	0.339
Vanilla LSTM †	TP = 186	TN = 1	78.6% <sup>†</sup>	0.033
VMD-Hybrid ★	TP = 123	TN = 36	67.1% ★	0.819 ★

*Note:* Rows = actual direction; columns = predicted direction. DA computed on non-zero actual days only. TP = True Positive (Actual $\uparrow$ , Predicted $\uparrow$ ). TN = True Negative (Actual $\downarrow$ , Predicted $\downarrow$ ). Actual UP days = 191 (80.6%); Actual DOWN days = 46 (19.4%).

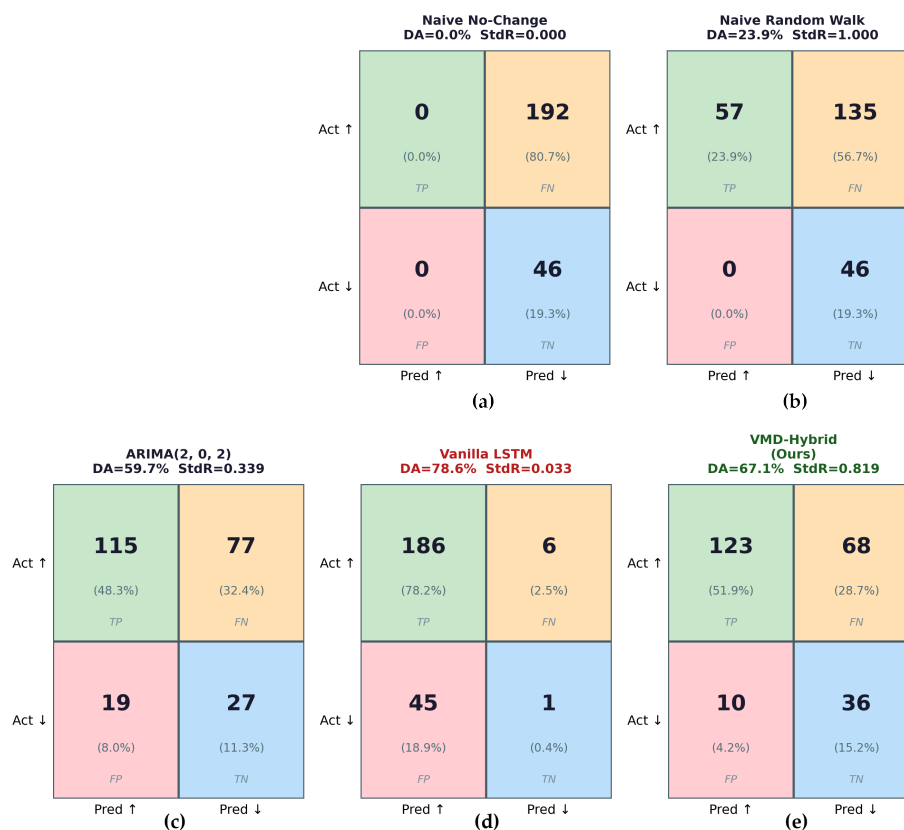
During the test period, 80.6% of trading days saw a price increase (191 up-days vs. 46 down-days). Vanilla LSTM exploited this imbalance by predicting UP on 97.1% of all days (231 out of 237 non-zero days), achieving TN = 1 — meaning the model correctly identified only one price decrease in the entire test set. With StdR = 0.033, the Vanilla LSTM output is effectively a near-constant signal, consistent with a model that has learned the unconditional sample mean of the training set rather than dynamic conditional patterns.



**Figure 8.** Predicted versus actual price changes in the test set. Scatter plot of predicted daily price changes  $\Delta\hat{p}_t$  against observed changes  $\Delta p_t$  for the Stage 3 test sample ( $n = 237$ ). The 45 reference line represents perfect prediction. Data points are colored by chronological quartile (Q1 = blue to Q4 = red). The strong linear association is reflected in  $r = 0.812$ .

By contrast, the VMD-Hybrid model predicted DOWN on 104 occasions (43.9% of days), achieving TN = 36 out of 46 actual down-days — a recall for the down-direction of 78.3%. The balanced confusion matrix confirms that the model has learned genuine directional dynamics rather than a static positive bias. Collectively, the confusion matrix diagnostics and StdR comparison confirm that Vanilla LSTM's superior DA reflects a degenerate positive-prediction strategy rather than genuine forecasting ability, and that Pearson  $r = 0.812$  is the appropriate primary performance indicator for this series. This finding corroborates Bürgi [51] [pp. 7909–7912], who demonstrates that DA is not a self-sufficient

evaluation concept without knowledge of magnitude sensitivity and class composition, and McCarthy and Snudden [52] [pp. 1–7], who show that success ratios can exceed 0.5 mechanically under class imbalance, generating pseudo-skill where no genuine predictability exists — precisely the pathology illustrated here by Vanilla LSTM’s  $TN = 1$  result. Costantini et al. [53] sharpen this point further by showing that DA becomes materially informative only when paired with magnitude-sensitive directional value — a measure that weights correct sign calls by the size of the realized move rather than treating all correct predictions as equivalent. Under that criterion, Vanilla LSTM’s near-constant output would register near-zero directional value despite its nominally high hit rate, consistent with its  $StdR = 0.033$ .



**Figure 9.** Directional bias analysis using confusion matrices. (a) Naive No-Change; (b) Naive Random Walk; (c) ARIMA(2,0,2); (d) Vanilla LSTM; (e) VMD-Hybrid

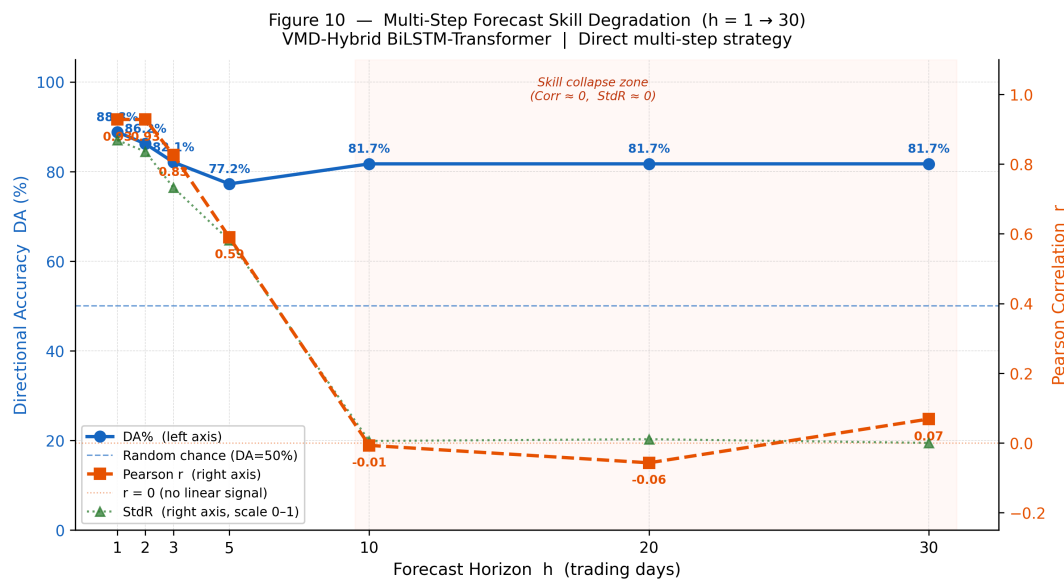
### 3.6. Multi-Step Forecast Skill Degradation

Having validated the model’s one-step-ahead performance and component contributions, this section assesses forecast skill across longer horizons to establish the practical range of reliable predictions.

This study adopts the direct forecasting strategy, training a separate model with identical architecture and hyper-parameters for each horizon  $h$ , to avoid the error accumulation associated with recursive multi-step forecasting [64]. While this requires  $H$  independent models, it produces unbiased  $h$ -step-ahead forecasts without recursive error propagation. A separate model was trained for each horizon  $h \in \{1, 2, 3, 5, 10, 20, 30\}$ .

**Table 13.** Multi-step direct forecast performance — VMD-Hybrid BiLSTM-Transformer.

$h$	DA%	Pearson $r$	MAE (Bt/kg/d)	RMSE	StdR
1 ★	67.1%	0.812	0.156	0.258	0.819
2	86.2%	0.928	0.622	—	0.836
3	82.1%	0.827	0.949	—	0.733
5	77.2%	0.590	1.257	—	0.581
10	81.7%	−0.007	1.136	—	0.006
20	81.7%	−0.057	1.151	—	0.011
30	81.7%	0.069	1.141	—	0.000



**Figure 10.** Multi-step forecast skill degradation. Directional accuracy (DA, left axis) and Pearson  $r$  (right axis) across forecast horizons  $h = 1, 2, 3, 5, 10, 20, 30$ . The dashed line indicates the random benchmark (DA = 50%). The logarithmic  $x$ -axis highlights the non-linear decline in predictive skill as the horizon increases. The  $h = 2$  anomaly reflects sustained directional trends in the test period rather than a systematic model property; see main text for discussion.

The  $h = 2$  horizon produces higher DA (86.2%) and Pearson  $r$  (0.928) than  $h = 1$ , which requires interpretation. This pattern most likely reflects the specific composition of the test period: during September 2025 – February 2026, the RSS3 price series exhibited sustained directional trends over multi-day windows, making two-day-ahead movements more predictable than single-day changes which capture more micro-structure noise. The  $h = 2$  model effectively filters intraday reversals that generate short-run unpredictability at  $h = 1$ , resulting in smoother targets that are more learnable given the available features. This interpretation is consistent with the IMF energy structure reported in Section 3.2, where high-frequency components (IMF<sub>4</sub>–IMF<sub>5</sub>) that contribute noise at  $h = 1$  are less prominent in the two-day-ahead target series.

This pattern also illustrates the model-selection hazard documented by Costantini et al. [53] and Costantini and Kunst [54]: relying on DA as a selection criterion at longer horizons would favour the  $h = 10$ – $30$  models — whose DA of 81.7% exceeds the  $h = 1$  model's 67.1% — despite those models having zero predictive content by every variance-sensitive measure, a finding consistent with their evidence that DA-weighted selection rarely yields robust gains and can worsen performance in MSE terms. Beyond  $h = 2$ , Pearson  $r$  declines monotonically: from 0.827 at  $h = 3$  to 0.590 at  $h = 5$ , and to near-zero at  $h = 10$  ( $r = -0.007$ ). StdR follows the same pattern, collapsing to approximately 0.006 at  $h = 10$ , indicating that the model can no longer reproduce the variance of ten-day-ahead price changes. DA values at  $h = 10$ – $30$  (approximately 81.7%) are misleading for the same reason

identified in Section 3.5: they reflect a positive-class majority in the cumulative 10–30-day return distribution rather than genuine predictive content, as confirmed by  $\text{StdR} \approx 0.000\text{--}0.011$ . The emphasis on short lead times is consistent with prior VMD-based forecasting studies, where predictive gains from decomposition are strongest at near-term horizons and tend to weaken as the forecast horizon extends [33].

These results indicate that the VMD-Hybrid model provides reliable forecasting skill up to a five-day horizon ( $h = 5$ ,  $r = 0.590$ ,  $\text{StdR} = 0.581$ ), with meaningful but diminishing skill at  $h = 3$ . Beyond five trading days, the model should not be used for directional price guidance.

#### 4. Discussion

The VMD decomposition results provide structural insight into the multi-scale price dynamics of the natural rubber market, and their economic interpretation reinforces the theoretical rationale for the proposed architectural design. The dominant trend component ( $\text{IMF}_1$ , 78.6% of total signal energy) captures slow-moving, long-run structural forces — principally global tyre demand driven by automobile production, plantation area adjustment in Thailand, Indonesia, and Malaysia, and the substitution relationship between natural and synthetic rubber. The pronounced concentration of signal energy in this single low-frequency mode confirms that RSS3 price dynamics are fundamentally anchored by supply–demand fundamentals operating over multi-year cycles, consistent with the commodity economics literature and with prior VMD-based evidence from natural rubber futures markets showing that the dominant mode reflects broader structural market conditions and longer-horizon price behavior [7].

The intermediate-frequency components ( $\text{IMF}_2\text{--}\text{IMF}_3$ , 11.9% combined energy) are consistent with medium-term market cycles arising from seasonal tapping patterns, inventory accumulation and draw-down, and semi-annual export policy adjustments in major producing countries. These components exhibit a degree of regularity that makes them particularly amenable to modeling by the BiLSTM pathway, whose recurrent structure is well-suited to capturing sequentially persistent, periodic fluctuations. This interpretation aligns with financial and energy-price forecasting evidence demonstrating that decomposition-based models improve predictive accuracy precisely by separating lower-frequency structural movements — which possess greater regularity — from higher-frequency disturbance components before sequence learning is applied [25,30].

The high-frequency residuals ( $\text{IMF}_4\text{--}\text{IMF}_5$ , 9.4% combined energy) capture speculative trading noise and short-horizon information shocks, including geopolitical disruptions, weather-related supply shocks, and the COVID-19-induced market dislocations observed in 2020–2021. The complementary role of the Transformer encoder is most apparent here: its global self-attention mechanism operates across the entire lookback window simultaneously, enabling the model to identify non-contiguous historical observations that are most informative for predicting irregular, short-duration deviations — a capability that strictly sequential recurrent processing cannot replicate. This division of labor between pathways is further supported by prior rubber-futures evidence indicating that high-frequency IMFs carry the most information relevant to near-term directional prediction, whereas low-frequency modes dominate longer-horizon behavior [7], and by energy-price forecasting research establishing that model performance improves when architectural design is explicitly adapted to the heterogeneous information content of components across frequency bands [30].

These decomposition-level findings translate directly into aggregate forecasting outcomes. The VMD-Hybrid model achieves a Directional Accuracy of 67.1% and a Pearson correlation of 0.812 on the held-out test set, representing practically meaningful skill for a daily commodity price series in which the random benchmark is 50%. A market participant employing the model's directional signal would correctly anticipate the direction of the next trading day's price movement approximately two-thirds of the time — a margin sufficient to support statistically informed hedging decisions, inventory management, and procurement timing across the natural rubber supply chain. These results are consistent with the broader hybrid-forecasting literature demonstrating that VMD-enhanced

architectures improve predictive performance by reducing the stationary burden on downstream sequence learners, with gains most pronounced when decomposition is integrated within a unified framework capable of exploiting both local sequential dependencies and global temporal structure [32,45]. In the commodity-price domain specifically, this pattern is corroborated by evidence that VMD improves model learnability through multi-frequency isolation, and that the further inclusion of structurally relevant long-run economic drivers yields additional gains in both level and directional accuracy [31].

## 5. Conclusions

This study proposed the VMD-Hybrid BiLSTM-Transformer, a unified forecasting framework for daily RSS3 FOB natural rubber price changes that integrates Variational Mode Decomposition with a dual-pathway deep learning architecture comprising a Bidirectional LSTM encoder with temporal attention and a Transformer encoder. The central architectural contribution is the treatment of VMD intrinsic mode functions as input features appended directly to the economic feature matrix — rather than as independent prediction targets in a per-component summation pipeline. This design decision proved consequential: the conventional per-IMF approach produced near-constant forecasts on the differenced series, with Standard Deviation Ratios of 0.04–0.15, whereas the proposed feature-augmentation design achieved a StdR of 0.819, confirming that multi-scale frequency information is effectively preserved within a single forward pass and that forecast variance collapse — a systematic failure mode of per-component pipelines on stationary differenced series — is avoided.

On the held-out test set of 237 observations, the proposed model outperformed all benchmarks across the primary evaluation metrics. Relative to ARIMA(2,0,2), the strongest linear baseline, the framework improved Pearson correlation by 0.662 (from 0.150 to 0.812) and reduced mean absolute error by 37.3% (from 0.249 to 0.156 Baht/kg/day). A secondary but methodologically important finding concerns the inadequacy of Directional Accuracy as a standalone evaluation criterion for differenced commodity price series. The Vanilla LSTM baseline recorded a nominally superior DA of 78.6%, yet confusion matrix and variance ratio diagnostics revealed that this arose from systematic positive directional bias — the model correctly identified only one price decrease in 237 test observations (StdR = 0.033). This finding argues for the routine inclusion of variance-sensitive metrics, specifically Pearson correlation and Standard Deviation Ratio, alongside directional measures in commodity price forecasting evaluations. This recommendation is consistent with a growing methodological consensus: Bürgi [51] demonstrates that DA lacks a self-sufficient evaluation framework and requires supplementation by magnitude-sensitive and dependence-robust procedures, while Costantini et al. [53] show that DA becomes materially informative only when paired with directional value measures that weight correct calls by the size of the realized move.

The VMD decomposition further revealed the multi-scale structure underlying RSS3 price dynamics. The dominant trend component (IMF<sub>1</sub>, 78.6% of signal energy) reflects long-run supply–demand fundamentals, while intermediate components (IMF<sub>2</sub>–IMF<sub>3</sub>, 11.9%) correspond to inventory and export cycles, and high-frequency residuals (IMF<sub>4</sub>–IMF<sub>5</sub>, 9.4%) capture speculative noise and short-horizon information shocks. The primary forecasting gains were attributable to the lower-frequency components, which provided stable sequential context to the BiLSTM pathway — consistent with prior rubber-futures evidence that decomposition improves performance when retained modes are matched to the temporal scale of the prediction target [7]. The precise interaction mechanisms between IMF characteristics and the dual-pathway encoder remain an open question warranting systematic ablation across decomposition modes in future work.

Three limitations constrain the current framework. First, eight of the 24 intended input features — including STR20, Latex, and CupLump FOB prices — were unavailable at the time of analysis and substituted with zeros, likely suppressing medium-frequency predictive accuracy. Second, multi-step forecasts beyond one day hold exogenous variables constant at their last observed values; this persistence assumption becomes increasingly untenable as the forecast horizon extends, which is

reflected in the rapid degradation of Pearson correlation beyond  $h = 5$ . Third, the VMD decomposition is fitted statically on the training partition without adaptive refitting, potentially reducing alignment during structural market transitions. Future work will address these constraints through dataset completion, rolling adaptive decomposition, and an extension toward probabilistic forecast intervals to support price risk management across the Thai natural rubber supply chain.

### *Policy Implication*

The findings carry direct implications for price risk management along the Thai rubber supply chain. A model that correctly anticipates the direction of daily RSS3 FOB price movements approximately two-thirds of the time — while reproducing the distributional characteristics of observed price changes — provides a statistically grounded basis for procurement timing, hedge initiation, and inventory holding decisions by processors, exporters, and institutional traders. More broadly, the demonstration that variance collapse is a systematic failure mode of conventional decomposition pipelines when applied to differenced commodity series has direct relevance for the design of early-warning systems used by national rubber price stabilization programmes, where near-zero forecast variance can render model signals operationally uninformative. Policymakers and market regulators should therefore require that forecasting tools submitted for operational deployment report variance-sensitive diagnostics — specifically StdR and Pearson correlation in the differenced space — alongside the directional accuracy statistics that are currently standard in industry practice. This is not merely a matter of reporting convention: Costantini and Kunst [54?] demonstrate that selecting forecasting models on the basis of DA alone seldom yields robust gains and can worsen performance on magnitude-sensitive criteria — implying that operational deployment decisions grounded solely in hit-rate statistics risk systematically selecting the wrong model. The proposed evaluation framework is immediately adoptable without additional data requirements and is applicable across agricultural commodity markets more broadly.

As natural rubber markets become increasingly integrated with global financial systems and sensitive to climate-induced supply disruptions, the capacity to generate reliable, variance-faithful short-horizon price forecasts will grow in operational importance; the present framework offers a methodologically transparent and empirically validated foundation on which such systems can be built and extended.

**Author Contributions:** Conceptualization, M.P.; methodology, M.P.; software, M.P.; validation, M.P.; formal analysis, M.P.; investigation, M.P.; resources, M.P.; data curation, M.P.; writing—original draft preparation, M.P.; writing—review and editing, M.P.; visualization, M.P.; supervision, M.P.; project administration, M.P.; funding acquisition, M.P. The author has read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** During the preparation of this manuscript, the author used [Claude] for the purposes of [grammar checking and language editing]. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

**Conflicts of Interest:** The author declares no conflicts of interest.

## **Abbreviations**

The following abbreviations are used in this manuscript:

RSS3	Ribbed Smoked Sheet No. 3
STR20	Standard Thai Rubber 20
USS	Unsmoked Sheet Rubber
StdR	Standard Deviation Ratio
SHFE	Shanghai Futures Exchange
SGX	Singapore Exchange
JPX	Japan Exchange Group
PMI	Purchasing Managers' Index
BDI	Baltic Dry Index
ONI	Oceanic Niño Index
ENSO	El Niño–Southern Oscillation
ADF	Augmented Dickey–Fuller
ADMM	Alternating Direction Method of Multipliers
EIA	The Energy Information Administration
NOAA	National Oceanic and Atmospheric Administration
WHO	World Health Organization
CEIC	<a href="http://www.ceicdata.com">www.ceicdata.com</a>

## Appendix A. Descriptive Statistics of Input Features

**Table A1.** Descriptive statistics of all 24 input features.

Feature	Unit	n	Mean	Std	Skewness	Kurtosis
RSS3 FOBm1	Baht/kg/day	1384	−0.0007	0.6593	−0.763	21.670
RSS3 FOBm2	Baht/kg/day	1384	−0.0007	0.6579	−0.757	21.824
STR20 FOBm1	Baht/kg/day	1384	0.0021	0.4060	−0.799	11.187
STR20 FOBm2	Baht/kg/day	1384	0.0021	0.4269	−0.865	16.278
Latex FOBm1	Baht/kg/day	1384	−0.0014	0.5225	−1.243	25.779
Latex FOBm2	Baht/kg/day	1384	−0.0014	0.5227	−1.242	25.751
CupLump	Baht/kg/day	1384	0.0011	0.7367	−0.415	5.275
USS	Baht/kg/day	1384	−0.0013	0.7316	−0.868	22.401
RSS3 JPX m1	JPY/kg/day	1384	0.0224	4.7562	−7.129	148.567
RSS3 SHFE m1	CNY/tonne/day	1384	0.5094	171.4076	0.883	14.318
RSS3 SHFE m2	CNY/tonne/day	1384	0.4118	162.5008	0.622	11.953
RSS3 SGX m1	USD/kg/day	1384	−0.0188	21.7603	0.138	74.967
TSR20 SGX m2	USD/kg/day	1384	−0.0080	17.8240	−0.025	71.909
USD/THB	Baht/USD/day	1384	0.0023	0.1363	0.012	5.558
CNY/THB	Baht/CNY/day	1384	−0.0001	0.0194	−0.026	4.977
USD/CNY	CNY/USD/day	1384	0.0007	0.0213	−0.286	8.445
Brent	USD/bbl/day	1384	0.0060	1.8239	−1.090	12.761
WTI	USD/bbl/day	1384	0.0066	2.6791	−2.774	211.726
Brent return	proportion/day	1385	0.0008	0.0273	−0.638	17.391
Brent lag-1	USD/bbl/day	1384	0.0061	1.8240	−1.090	12.759
China PMI Mfg	index/month	1385	50.1541	1.1975	−0.613	3.452
Baltic Dry Index	index/day	1383	−0.2545	58.9929	0.146	7.665
ENSO ONI	degrees Celsius	1385	−0.1302	0.6945	0.241	1.666
COVID dummy	0/1	1385	0.3466	0.4760	0.645	1.416

## Appendix B. ADF Unit Root Test Results

Table A2. ADF Unit Root Test Results.

Feature	ADF Stat.	p-value	Decision	Feature	ADF Stat.	p-value	Decision
RSS3 FOBm1	-1.8508	0.3556	I(1)	TSR20 SGX m2	-2.9035	0.0449	I(0)
RSS3 FOBm2	-1.8430	0.3593	I(1)	USD/THB	-1.5075	0.5298	I(1)
STR20 FOBm1	-1.9589	0.3049	I(1)	CNY/THB	-1.4099	0.5775	I(1)
STR20 FOBm2	-1.9017	0.3313	I(1)	USD/CNY	-1.1724	0.6854	I(1)
Latex FOBm1	-2.2880	0.1759	I(1)	Brent	-1.4558	0.5553	I(1)
Latex FOBm2	-2.3797	0.1476	I(1)	WTI	-1.6107	0.4776	I(1)
CupLump	-2.2047	0.2045	I(1)	Brent return	-32.9123	0.0000	I(0)
USS	-1.9191	0.3231	I(1)	Brent lag-1	-1.4669	0.5498	I(1)
RSS3 JPX m1	-2.4428	0.1300	I(1)	China PMI Mfg	-3.7399	0.0036	I(0)
RSS3 SHFE m1	-2.4678	0.1235	I(1)	Baltic Dry Index	-2.5805	0.0971	I(1)
RSS3 SHFE m2	-2.4853	0.1191	I(1)	ENSO ONI	-0.1238	0.9470	I(1)
RSS3 SGX m1	-2.3554	0.1547	I(1)	COVID dummy	-1.4858	0.5405	I(1)

Table A2 presents Augmented Dickey–Fuller unit root test results for all 24 input features in the Stage 3 training partition (2018–2023), with lag order selected by AIC (maximum lags = 10) and decisions based on the 5% significance level. Two features are retained in levels by design: ENSO ONI is a climatological index for which differencing removes meaningful low-frequency signal, and the COVID dummy is a binary policy variable. TSR20 SGX m2 is borderline stationary ( $p = 0.045$ ) but is treated as I(1) for consistency with the remaining exchange-traded series.

## Appendix C. K-Sensitivity Analysis for VMD

Table A3. Reconstruction RMSE for  $K \in \{3, 4, 5, 6, 7\}$ .

$K$	Recon RMSE (norm)	IMF <sub>1</sub> Energy (%)	Note
3	0.640	82.0%	Under-decomposed
4	0.558	78.9%	
5	0.483	75.7%	Selected ★
6	0.414	73.7%	Marginal gain < 0.005 RMSE
7	0.366	72.4%	Over-decomposed

Table A3 reports VMD reconstruction RMSE and IMF<sub>1</sub> energy share across  $K \in \{3, 4, 5, 6, 7\}$  for the normalized differenced RSS3 FOBm1 series (Stage 3 training partition,  $n = 1,384$ ). Reconstruction RMSE declines monotonically with  $K$ ; however, the marginal reduction diminishes substantially beyond  $K = 5$  — falling from 0.013 between  $K = 4$  and  $K = 5$  to only 0.008 between  $K = 5$  and  $K = 6$ , and 0.048 between  $K = 6$  and  $K = 7$ . IMF<sub>1</sub> energy stabilizes in the range 72–76% for  $K \geq 5$ , indicating that modes added beyond this threshold primarily subdivide existing high-frequency components rather than isolating new economically interpretable structures.  $K = 5$  is therefore selected as the minimum sufficient decomposition depth, based on both RMSE stabilization and interpretable energy distribution.

## Appendix D. ARIMA Model Selection

**Table A4.** AIC values from ARIMA( $p, 0, q$ ) grid search.

	$q = 1$	$q = 2$	$q = 3$
$p = 1$	2486.14	2469.90	2469.56
$p = 2$	2477.86	<b>2461.94 ★</b>	2462.81
$p = 3$	2463.12	2462.72	2463.13

ARIMA(2,0,2) selected by minimum AIC = 2461.94. The adjacent models ARIMA(2,0,3) and ARIMA(3,0,2) yield AIC values within 1 unit (2462.81 and 2462.72 respectively), confirming that the selected order is robust to small perturbations in lag specification.

## Appendix E. Cross-Stage Detailed Comparison

**Table A5.** Confusion Matrix — Stage 1 ( $K = 5, N = 283$ ).

	Predicted ↑	Predicted ↓
Actual ↑	279	0 — never predicted
Actual ↓	2	0 — never predicted
DA: 99.3% MAE: 0.0360 Corr: 0.385		

**Table A6.** Confusion Matrix — Stage 2 ( $K = 7, N = 247$ ).

	Predicted ↑	Predicted ↓
Actual ↑	239	0 — never predicted
Actual ↓	8	0 — never predicted
DA: 96.8% MAE: 0.0435 Corr: 0.118		

**Table A7.** Regression Error Summary by Stage.

Stage	$N$	MAE	RMSE	Bias	P50 (err)	P90 (err)	Corr	Pred std / Act std
Stage 1	283	0.0360	0.0496	-0.0072	0.0260	0.0834	0.385	0.0179/0.0530 = 0.34×
Stage 2	247	0.0435	0.0750	0.0186	0.0180	0.1278	0.118	0.0084/0.0731 = 0.11×
$\Delta$ (S2-S1)	-36	+0.0075	+0.0254	+0.0258	-0.0080	+0.0444	-0.267	—

## Appendix F. Hyper-parameter Sensitivity

**Table A8.** Look-back Window Sensitivity.

$L$	DA (%)	Corr	MAE	RMSE	StdR	Runtime (min)
10	97.7	0.8958	0.0224	0.0326	0.924	5.4
20	96.8	0.8355	0.0256	0.0409	0.831	26.9
30	96.2	0.7046	0.0319	0.0538	0.865	11.1
45	96.8	0.7117	0.0341	0.0532	0.752	13.3
60	96.6	0.7443	0.0299	0.0466	0.651	14.8

Notes:  $L = 10$  outperforms all other windows across DA, Corr, MAE, and runtime, indicating that shorter context is sufficient for this datasets. StdR declines monotonically with longer  $L$  (0.924  $\rightarrow$  0.651), implying reduced variance expressiveness at larger windows.  $L = 20$  shows an anomalously long runtime (26.9 min), likely due to a delayed early-stopping trigger rather than model complexity.

## References

1. International Rubber Study Group. World Rubber Industry Outlook 2023. Technical report, IRSG, Singapore, 2023.
2. Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database, 2024.
3. Association of Natural Rubber Producing Countries. Natural Rubber Trends and Statistics 2025. Technical report, ANRPC, Kuala Lumpur, Malaysia, 2025.
4. Gilbert, C. How to understand high food prices. *J. Agric. Econ.* **2010**, *61*, 398–425.
5. Baffes, J.; Haniotis, T. What explains agricultural price movements? *J. Agric. Econ.* **2016**, *67*, 706–721.
6. Arezki, R.; Hadri, K.; Loungani, P.; Rao, Y. Testing the Prebisch–Singer hypothesis since 1650. *J. Int. Money Financ.* **2014**, *42*, 208–223.
7. Zhu, Q.; Zhang, F.; Liu, S.; Wu, Y.; Wang, L. A hybrid VMD–BiGRU model for rubber futures time series forecasting. *Appl. Soft Comput.* **2019**, *84*, 105739.
8. Sang, W.; Ma, W. Analysis of the dynamic relationship of the futures prices of rubber in Shanghai and Tokyo. *J. Finance Econ.* **2019**, *31*, 36–44.
9. Kepulaje, A.; Prakash, P.; Birau, R.; Hawaldar, I.; Tan, Y.; Wanke, P. Cross-border linkages between the global rubber spot and futures markets. *Reg. Sect. Econ. Stud.* **2024**, *24*, 159–174.
10. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1976.
11. Engle, R. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* **1982**, *50*, 987–1007.
12. Zahari, F.; Khalid, K.; Roslan, R.; Sufahani, S.; Mohamad, M.; Rusiman, M.; Ali, M. Forecasting natural rubber price in Malaysia using ARIMA. *J. Phys. Conf. Ser.* **2018**, *995*, 012013.
13. Khin, A.; Thambiah, S. Forecasting analysis of price behavior: A case of Malaysian natural rubber market. *Am.-Eurasian J. Agric. Environ. Sci.* **2014**, *14*, 1187–1195.
14. Ediger, V.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708.
15. Byun, S.; Cho, H. Forecasting carbon futures volatility using GARCH models. *Energy Econ.* **2013**, *40*, 623–630.
16. Zhang, G.; Patuwo, B.; Hu, M. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62.
17. Cao, L.; Tay, F. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Trans. Neural Netw.* **2003**, *14*, 1506–1518.
18. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Boston, MA, USA, 2011.
19. Zhang, G. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175.
20. Pai, P.; Lin, C. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **2005**, *33*, 497–505. <https://doi.org/10.1016/j.omega.2004.07.024>.
21. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM fully convolutional networks for time series classification. *IEEE Access* **2018**, *6*, 1662–1669.
22. Ahmed, N.; Atiya, A.; Gayar, N.; El-Shishiny, H. An empirical comparison of machine learning models for time series forecasting. *Econom. Rev.* **2010**, *29*, 594–621.
23. Phoksawat, K.; Phoksawat, E.; Chanakot, B. Forecasting smoked rubber sheets price based on a deep learning model with long short-term memory. *Int. J. Electr. Comput. Eng.* **2023**, *13*, 688–696.
24. Eng, C.; Khalid, K. Forecasting natural rubber price in Malaysia using ARIMA and Long Short-Term Memory. *Enhanc. Knowl. Sci. Technol.* **2025**, *5*, 452–461.
25. Song, Y.; Li, Z.; Ma, Z.; Sun, X. Dynamic forecasting for nonstationary high-frequency financial data with jumps based on series decomposition and reconstruction. *J. Forecast.* **2023**, *42*, 1055–1068.
26. Ming, C.; Leung, K.; Shen, Y.; Ho, A. Transformers outperform traditional forecasting models and perform comparably to recurrent neural networks in the prediction of emergency department visits. *Artif. Intell. Emerg. Med.* **2026**, *1*, 100006.
27. Chen, Y.; Fu, Z. Multi-step ahead forecasting of the energy consumed by the residential and commercial sectors using a hybrid CNN–BiLSTM model. *Sustainability* **2023**, *15*, 1895. <https://doi.org/10.3390/su15031895>.
28. Yu, W.; Li, S.; Zhang, H. Ultra-short-term wind-power forecasting based on an optimized CNN–BiLSTM–attention model. *iEnergy* **2024**, *3*, 268–282.

29. Ji, Q.; Han, L.; Jiang, L.; Zhang, Y.; Xie, M.; Liu, Y. Short-term prediction of the significant wave height and average wave period based on the VMD–TCN–LSTM algorithm. *Ocean Sci.* **2023**, *19*, 1561–1578.
30. Lu, Q.; Liao, J.; Chen, K.; Liang, Y.; Lin, Y. Predicting natural gas prices based on a novel hybrid model with variational mode decomposition. *Comput. Econ.* **2024**, *63*, 639–678.
31. Li, J.; Zhu, S.; Wu, Q. Monthly crude oil spot price forecasting using variational mode decomposition. *Energy Econ.* **2019**, *83*, 240–253.
32. He, M.; Qian, Q.; Liu, X.; Zhang, J. Predicting river turbidity in Pine Island Bayou using machine learning techniques coupled with variational mode decomposition. *J. Hydrol.* **2026**, *668*, 135011.
33. Ladouali, S.; Katipoglu, O.; Bahrami, M.; Kartal, V.; Sakaa, B.; Elshaboury, N.; Keblouti, M.; Chaffai, H.; Ali, S.; Pande, C.; et al. Short lead time standard precipitation index forecasting: Extreme learning machine and variational mode decomposition. *J. Hydrol. Reg. Stud.* **2024**, *54*, 101861.
34. Cui, Z.; Li, T.; Ding, Z.; Li, X.; Wu, J. Probabilistic oil price forecasting with a variational mode decomposition-gated recurrent unit model incorporating pinball loss. *Data Sci. Manag.* **2025**, *8*, 237–247.
35. Zhou, X.; Liu, C.; Luo, Y.; Wu, B.; Dong, N.; Xiao, T.; Zhu, H. Wind power forecast based on variational mode decomposition and long short term memory attention network. *Energy Rep.* **2022**, *8*, 922–931.
36. Yu, B.; Wang, Y.; Wang, J.; Ma, Y.; Li, W.; Zheng, W. A hybrid model for short-term offshore wind power prediction combining Kepler optimization algorithm with variational mode decomposition and stochastic configuration networks. *Int. J. Electr. Power Energy Syst.* **2025**, *168*, 110703.
37. Iqbal, T.; Alzaidi, A.; Alalaiwe, A.; Alsubaie, N.; Althobaiti, M.; Abbas, Q.; Irfan, M. A novel hybrid deep learning method for accurate exchange rate prediction. *Neural Comput. Appl.* **2024**, *36*, 17965–17995.
38. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, 2017, Vol. 30.
39. Zhou, H.; Zhang, S.; Peng, J.; Zhang, J.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2021.
40. Wu, B.; Yu, S.; Lv, S.X. Explainable soybean futures price forecasting based on multi-source feature fusion. *J. Forecast.* **2025**, *44*, 1363–1382.
41. Jia, Y.; Yin, H.; Xu, Z.; Xue, Z.; Li, X.; Ma, Z.; Li, A. Hybrid LSTM–Transformer architecture for predictive indoor operative temperature modeling. *Energy Build.* **2026**, *357*, 117161.
42. Yang, Y.; Zheng, Z.; Tang, F.; Liu, H.; Wang, B.; Li, N.; Zhao, Y. LSTM–Transformer neural network model for satellite orbit prediction. *Measurement* **2026**, *272*, 120973.
43. Albaqami, S.; Mchara, W.; Raissi, M.; Khotimah, W. Hybrid wavelet transformer LightGBM model optimized by Optuna–TPE for global irradiance forecasting. *Results Eng.* **2026**, *29*, 109678.
44. Ma, L.; Guo, C.; Wang, Y.; Li, C.; Jing, K.; Zhang, B. Attention-guided KAN-Transformer hybrid model for power prediction. *Energy Rep.* **2026**, *15*, 109138.
45. Zhang, Y.; Liu, Y.; Wang, X.; Zhang, K.; Zhao, Q.; Zhang, J. A distributed wind farm power forecasting method based on improved variational mode decomposition and hybrid deep learning. *Energy* **2026**, *333*, 138166.
46. Shumailov, I.; Shumaylov, Z.; Zhao, Y.; Papernot, N.; Anderson, R.; Gal, Y. AI models collapse when trained on recursively generated data. *Nature* **2024**, *631*, 755–759. <https://doi.org/10.1038/s41586-024-07566-y>.
47. Rajpal, S.; Mahadeva, R.; Goyal, A.; Sarda, V. Improving forecasting accuracy of stock market indices utilizing attention-based LSTM networks with a novel asymmetric loss function. *AI* **2025**, *6*, 268.
48. Li, L.; Shan, K.; Geng, W. Forecasting crude oil price using secondary decomposition-reconstruction-ensemble model based on variational mode decomposition. *J. Futures Mark.* **2025**, *45*, 1601–1615.
49. Fakhong, T.; Pupunja, P.; Punjatewakupt, P.; Horayangkura, S. Forecasting Thai rubber prices: A system dynamics approach to economic and policy impacts. *Soc. Sci. Humanit. Open* **2025**, *12*, 102163.
50. Taylor, K. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **2001**, *106*, 7183–7192. <https://doi.org/10.1029/2000JD900719>.
51. Bürgi, C. Assessing the accuracy of directional forecasts. *Appl. Econ.* **2025**, *57*, 7909–7920.
52. McCarthy, M.; Snudden, S. Predictable by construction: assessing forecast directional accuracy of temporal aggregates. *Appl. Econ.* **2025**. published online 16 October, <https://doi.org/10.1080/00036846.2025.2564464>.
53. Costantini, M.; Crespo Cuaresma, J.; Hlouskova, J. Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *J. Forecast.* **2016**, *35*, 652–668.
54. Costantini, M.; Kunst, R. On the use of mean square error and directional forecast accuracy for model selection: a simulation study. *J. Stat. Comput. Simul.* **2026**, *96*, 205–231.

55. Arunwarakorn, S.; Suthiwartnarueput, K.; Pornchaiwiseskul, P. Forecasting equilibrium quantity and price on the world natural rubber market. *Kasetsart J. Soc. Sci.* **2019**, *40*, 8–16.
56. Alzaemi, S.; Sathasivam, S.; Ali, M.; Tay, K.; Velavan, M. Hybridized intelligent neural network optimization model for forecasting prices of rubber in Malaysia. *Comput. Syst. Sci. Eng.* **2023**, *47*, 1471–1489.
57. Pinitjitsamut, M. Dynamic price transmission from SHFE to Thai rubber markets: A cointegration–ECM and machine-learning analysis. *Economies* **2026**, *14*, 9.
58. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544.
59. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning* **2011**, *3*, 1–122. <https://doi.org/10.1561/22000000016>.
60. Schuster, M.; Paliwal, K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681.
61. Mehdiyev, N.; Enke, D.; Fettke, P.; Loos, P. Evaluating forecasting methods by considering different accuracy measures. *Procedia Comput. Sci.* **2016**, *95*, 264–271. <https://doi.org/10.1016/j.procs.2016.09.332>.
62. Kyriakidis, I.; Karatzas, K.; Kukkonen, J.; Papadourakis, G.; Ware, A. Evaluation and analysis of artificial neural networks and decision trees in forecasting of common air pollutants in Thessaloniki. *Eng. Intell. Syst.* **2015**, *21*.
63. Xu, J.; Liu, Z. Quantifying the variability collapse of neural networks, 2023, [2306.03440].
64. Ben Taieb, S.; Bontempi, G.; Atiya, A.; Sorjamaa, A. Recursive and direct multi-step forecasting: the best of both worlds. *Neurocomputing* **2012**.
65. Ampountolas, A. Enhancing forecasting accuracy in commodity and financial markets: Insights from GARCH and SVR models. *Int. J. Financial Stud.* **2024**, *12*, 59. <https://doi.org/10.3390/ijfs12030059>.

**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.