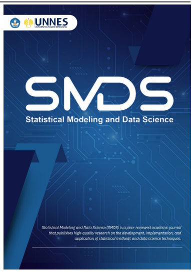
Vol. 1, No. 1 (2026) |

OPEN ACCESS

SMDS

Statistical Modeling and Data Science

https://journal.unnes.ac.id/journals/smds



Comparison of hybrid models VMD-LSTM and VMD-RNN for monthly forecasting of international tourist arrivals in Indonesia by entry point

Alfina Akmalina[1*] Elya Gita Astuti[1] Iqbal Kharisudin[2]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Indonesia*

[2] *Department of Statistics and Data Science, Universitas Negeri Semarang, Indonesia*

[*] *Corresponding author, email: alfinaakmalina@students.unnes.ac.id*

Abstract

Accurate forecasting of international tourist arrivals is crucial for strategic planning, but the nonlinear and volatile nature of tourism data presents significant modeling challenges. This study evaluates the effectiveness of a hybrid decompose-and-ensemble approach for estimating international tourist arrivals to Indonesia via three main entry points: air, sea, and land. Unlike traditional frequency component decomposition methods, the proposed hybrid model–Variational Mode Decomposition-Long Short-Term Memory (VMD-LSTM) and Variational Mode Decomposition-Recurrent Neural Network (VMD-RNN)–integrates all Intrinsic Mode Functions (IMFs) resulting from Variational Mode Decomposition (VMD) into a single multivariate dataset to project all components simultaneously. These models were evaluated using monthly time-series data from January 2008 to March 2026. Contrary to the common assumption that decomposition always improves accuracy, this study demonstrates the theoretical duality of high-order multivariate decomposition ($K = 12$) based on data volatility. For air and sea routes with stable seasonal trends, a single Recurrent Neural Network (RNN) model consistently delivered the best performance, with test set R^2 values of 95.98% and 86.90%, respectively. The decline in the accuracy of the hybrid model on these stable routes is caused by the accumulation of reconstruction error variance during the linear combination of the 12 Intrinsic Mode Functions (IMF) components. Conversely, on purely chaotic and noise-laden land routes, the multivariate configuration successfully reversed the trend by achieving test R^2 scores above 75%. This study concludes that high-order decomposition acts as a reliable noise filter in volatile environments but becomes a source of error variance inflation in stable data streams.

Keywords: *Deep Learning, Recurrent Neural Network, Time Series Modeling, Tourism Forecasting, Variational Mode Decomposition*

1. Introduction

Tourism is an important part of Indonesia's economy, contributing a significant share of the country's foreign-exchange earnings and driving inclusive economic growth and new job creation [1]. A press

ISSN xxxx-xxxx (Online)/

release by the Ministry of Tourism shows an estimated growth in foreign exchange earnings of 19.3% in 2024 compared to the previous year. The sector is also expected to contribute 4.5% to the national Gross Domestic Product (GDP) in 2024 [2]. The latest data from the Badan Pusat Statistik (BPS) confirms the positive trend and significant scale of this sector. In the fourth quarter of 2024, international tourist arrivals totaled 3.37 million, a 10.3 percent increase from the same period the previous year. Furthermore, the sector's socio-economic impact is reflected in its employment figures. In 2024, the tourism sector employed 25.01 million people, a 2.5 percent increase from 24.41 million in 2023 [3].

Tourist time series data are inherently historical and exhibit two key characteristics: non-stationarity and non-linearity [4]. These properties mean that neither the relationships between variables nor the statistical moments of the series remain constant over time. This complexity poses significant challenges for conventional forecasting methods. Econometric models such as Vector Autoregressive (VAR) and Vector Error Correction Model (VECM) aim to capture causal relationships between variables. Meanwhile, time series models such as Autoregressive Integrated Moving Average (ARIMA) rely on historical patterns including trends and seasonality [5]. However, both approaches share a fundamental limitation: they assume linearity and stationarity, making them ill-suited for the complex non-linear dynamics commonly found in tourism data [6].

Artificial intelligence (AI)-based models have emerged as stronger alternatives. These include machine learning approaches such as Support Vector Regression (SVR) and deep learning architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [7]. A key advantage of AI-based models is their ability to capture complex non-linear relationships directly from data without imposing rigid statistical assumptions. When dealing with long-term temporal dependencies, deep learning architectures such as LSTM have proven to be highly effective [8]. Nevertheless, the recognized limitations of single AI models have driven the development of hybrid approaches that combine multiple models to compensate for individual weaknesses [9].

The decomposition-ensemble strategy has become popular with this hybrid approach. This paradigm is based on the understanding that various components of complex time series data consist of trends, seasonality, and non-linear cycles. Rather than applying a single model to the entire series, this approach decomposes the data into simpler parts and assigns each to the model best suited for it. This collaborative system has been proven to produce forecasts that are much more accurate and robust. Essentially, the decomposition-ensemble paradigm follows a 'divide and conquer' strategy: complex, unstable series are broken into simpler, more stable signals called Intrinsic Mode Functions (IMFs), each predicted separately before the results are recombined.

In selecting the most effective decomposition technique, the literature offers various alternatives, such as Empirical Mode Decomposition (EMD) and its variant Ensemble EMD (EEMD), Singular Spectrum Analysis (SSA), and Wavelet Transform. However, a comparative study by Zhang et al. [8] found that Variational Mode Decomposition (VMD) is the best decomposition method, surpassing methods such as SSA, Wavelet, and EEMD. Unlike sequential methods such as EEMD, which are prone to mode-mixing and error accumulation, VMD solves a simultaneous optimization problem that extracts all signal components at once. This makes VMD more noise-resistant and yields cleaner, more stable sub-signals.

The VMD–deep learning combination has also shown strong results in domains beyond tourism. Wang et al. [10] demonstrated state-of-the-art air quality forecasting using a VMD–Graph Attention

Network–Bidirectional LSTM hybrid (VMD-GATBiLSTM). In the study by Aksan et al. [11], electricity load was successfully forecasted using the VMD-CNN-LSTM (Variational Mode Decomposition-Convolutional Neural Network-Long Short-Term Memory) model. This cross-disciplinary success reinforces the argument that the VMD-Deep Learning framework is a mature, robust, and highly promising methodological approach for addressing the challenges in tourism demand forecasting. Motivated by these challenges, this study evaluates the effectiveness of two hybrid decomposition-based deep learning models – VMD-LSTM and VMD-RNN – against their standalone counterparts, LSTM and RNN, for forecasting international tourist arrivals to Indonesia via air, sea, and land entry points.

The primary methodological distinction of this study lies in its multivariate IMF integration strategy. Unlike the conventional decomposition-ensemble paradigm — in which each IMF is modeled independently by a dedicated submodel before linear recombination — this study concatenates all IMFs derived from VMD into a single multivariate input matrix, enabling the network to jointly learn inter-component frequency relationships in a single forward pass. While Zheng et al. [12] demonstrated the effectiveness of this multivariate feeding strategy for wave height forecasting, its application to tourism demand forecasting across structurally distinct entry points remains unexplored.

The main scientific contribution of this study is the empirical demonstration of the theoretical duality of high-order multivariate decomposition under varying data volatility regimes. Contrary to the prevailing assumption in tourism forecasting literature that decomposition-ensemble strategies consistently improve accuracy [4], [8]. This study provides statistically validated evidence that VMD preprocessing degrades performance on stable, high-volume data streams such as air and sea routes — due to reconstruction error variance accumulation — while significantly outperforming single models on purely chaotic data such as land routes. This volatility-conditional behavior of decomposition has not been previously characterized in the tourism forecasting literature.

The remainder of this paper is organized as follows: Section 2 details the research methodology, including data preprocessing and model architectures. Section 3 presents the empirical results and performance evaluation. Section 4 discusses the implications of these findings, and finally, Section 5 concludes the paper.

2. Methodology

2.1. Research Design

This study employs a quantitative methodology centered on a comparative analysis of baseline deep learning models and hybrid VMD decomposed approaches for forecasting international tourist arrivals to Indonesia via air, sea, and land entry points.

The research framework is systematically illustrated in Figure 1 and divided into several interconnected stages. The first stage encompasses data collection and preprocessing, during which historical time-series data is collected and normalized, then split chronologically for neural network training and evaluation. At this stage, the signal is routed to a VMD module where the original complex series is decomposed into a set of more stable sub-signals.

The determination of VMD decomposition parameters, particularly the number of IMF K and the penalty factor α, is carried out through automated hyperparameter optimization using the Optuna framework with the Minimum Average Envelope Entropy as the objective function. This approach ensures that the selected parameters yield the most informative IMFs, with minimal noise and no mode mixing. The detailed optimization procedure is presented in Sub-section 3.4.

Prior to hybrid modeling, the baseline RNN and LSTM models are hyperparameter-tuned using the Optuna framework to identify optimal architectural configurations, including the number of layers, neuron units, dropout rate, batch size, and training epochs. The optimal configurations are subsequently applied uniformly to the hybrid models to ensure a controlled and objective comparison. Full configuration details are reported in Sub-section 3.3.

During the hybrid modeling phase, all IMFs derived from VMD are concatenated into a single multivariate dataset rather than being modeled independently per component. This multivariate feeding strategy enables the network to simultaneously forecast all IMF components in a single forward pass, thereby jointly capturing inter-component frequency relationships and temporal dynamics. The detailed procedure for IMF concatenation and multivariate input construction is presented in Sub-section 3.5.

The final stage involves model evaluation, in which the predicted values across all IMF components are linearly aggregated to reconstruct the final forecasted arrival series. Performance is assessed using MAE, MSE, MAPE, and R^2 metrics. The forecasting performance of the hybrid VMD-RNN and VMD-LSTM models is then benchmarked against their standalone counterparts to determine the most effective forecasting approach for highly volatile tourism demand data.

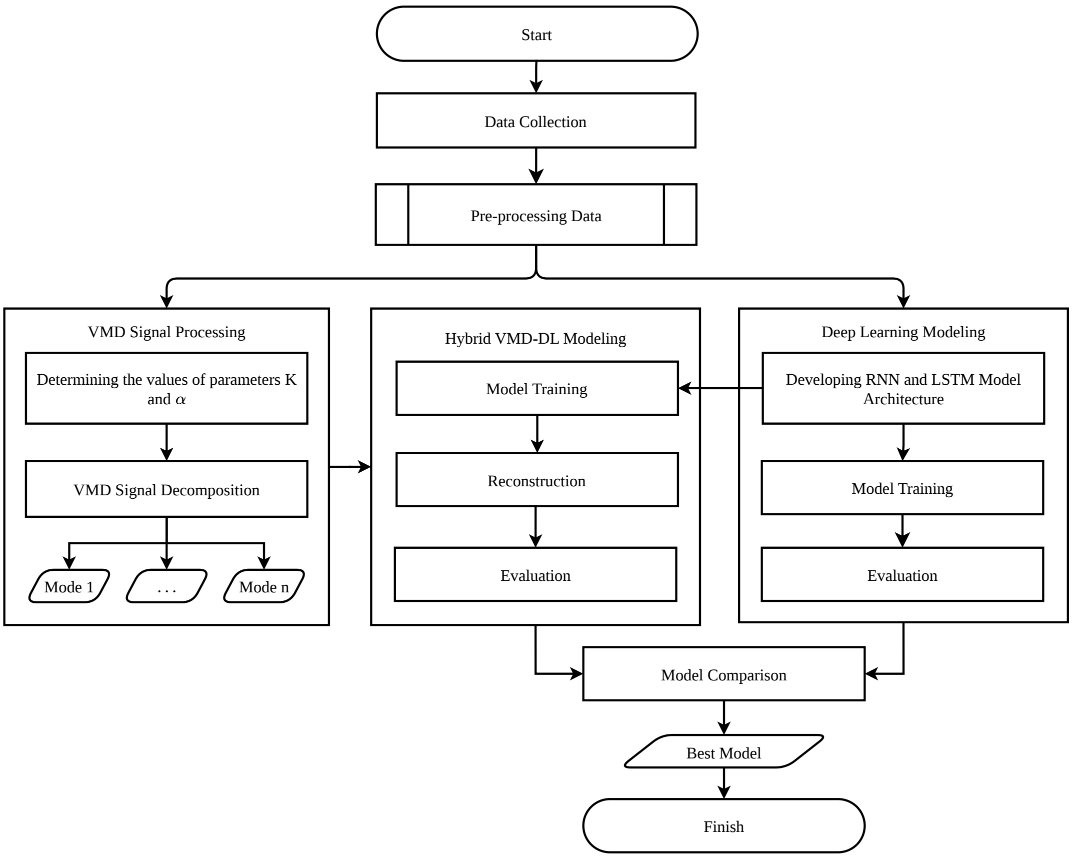


Figure 1. Research design framework for forecasting international tourist arrivals.

2.2. Data Description

This study uses monthly univariate time-series data on international tourist arrivals by entry point in Indonesia, sourced from the Badan Pusat Statistik (BPS) and supported by data from the Ministry of Tourism and Creative Economy. The dataset covers the period from January 2008 to March 2026, comprising 219 monthly observations for each of the three entry points: air (AIR), sea (SEA), and land (LAND). The extended observation window through March 2026 is deliberately chosen to encompass the full trajectory of post-pandemic recovery, providing the model with sufficient exposure to both the structural disruption caused by COVID-19 and the subsequent rebound dynamics necessary for robust generalization.

After inspection, no missing values were found in the dataset, allowing the data proceeded directly to the preprocessing stage without imputation. However, it is important to acknowledge a structural break across all three series during 2020–2021, corresponding to the global COVID-19 pandemic. This period is characterized by a near-complete cessation of international tourist arrivals, producing extremely low-value observations that significantly deviate from pre-pandemic patterns. Rather than treating this period as an anomaly to be removed, it is retained in its entirety to preserve the temporal integrity of the series and to evaluate model robustness under conditions of extreme distributional shift.

2.3. Data Preprocessing

A systematic preprocessing pipeline is implemented to prepare raw data for the modeling stage, incorporating procedures for handling data discontinuities and scaling data to a range suitable for neural network training. To maintain temporal continuity, the process begins by handling missing values, filling each gap with linear interpolation to estimate values from surrounding data points.

This is followed by data normalization using Min-Max Scaling, which transforms the data into the range $[0, 1]$ to improve model convergence and reliability. The transformation follows the formula:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{1}$$

where X represents the original value, and X_{min} and X_{max} are the minimum and maximum values observed in the dataset. Crucially, these scaling parameters are determined solely from the training data and subsequently applied to both the training and testing sets to ensure valid model evaluation [13].

Finally, a strict chronological data splitting is implemented to preserve the temporal order of observations and prevent data leakage. The first 70% of the time-ordered data is allocated to the training set for model learning and optimization, while the remaining 30% is reserved as a test set for final performance evaluation [14].

2.4. Model Architecture

2.4.1. Variational Mode Decomposition

Variational Mode Decomposition (VMD) is an adaptive decomposition technique designed for non-stationary signals [15]. This technique decomposes complex original signals into K sub-signals, each amplitude- and frequency-modulated and referred to as an IMF. Suppose a signal $P_{agg}(t)$ represents a sequence of N discrete values collected at regular intervals, which generally represents a sequence of non-stationary power usage [16]. Here is how the signal will be decomposed using VMD:

$$P_{agg}(t) = \sum_{k=1}^{K} IMF_k(t) + r(t), \tag{2}$$

where $IMF_k(t)$ is the k-th IMF, K is the total number of IMF, and $r(t)$ is the residual. The mathematical formulation can be expressed as follows:

$$IMF_k(t) = A_k(t)\cos(\phi_k(t)), \quad A_k(t) \geq 0, \tag{3}$$

where $\phi_k(t)$ is the phase and $A_k(t)$ is the slowly varying envelope corresponding to the k-th IMF; it also has an instantaneous frequency ω_k that varies slowly, is largely compact, and is non-decreasing around the center frequency ω_k.

The VMD approach uses the Alternate Direction Method of Multipliers (ADMM) optimization technique to perform various calculations simultaneously to identify K-IMFs and their associated center frequencies. According to ADMM, VMD can decompose the input signal $P_{agg}(t)$ into IMF_k and ω_k using the following equations [17]:

$$I\hat{M}F_k^{n+1}(\omega) = \frac{\hat{P}_{agg}(\omega) - \sum_{i \neq k} I\hat{M}F_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

and

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |I\hat{M}F_k(\omega)|^2 d\omega}{\int_0^\infty |I\hat{M}F_k(\omega)|^2 d\omega}, \tag{5}$$

where n is the iteration number, λ is the Lagrangian multiplier, $I\hat{M}F_k^{n+1}(\omega)$, $\hat{P}_{agg}(\omega)$, and $\hat{\lambda}(\omega)$ correspond to the Fourier transform of $IMF_k^{n+1}(t)$, $P_{agg}(t)$, $IMF_k(t)$, and $\lambda(t)$, respectively.

2.4.2. Recurrent Neural Network

A recurrent neural network (RNN) is a specific type of Artificial Neural Network (ANN) structure designed to handle time series data [18]. Unlike typical feedforward neural networks, which treat each input as unrelated, RNNs have connections that form directed cycles or loops. This repetitive architecture enables RNNs to have internal memory that stores data from previously received inputs. RNNs consist of three main layers: the input layer, the hidden layers, and the output layer.

Mathematically, the calculation of the hidden state (h_t) at time t can be formulated as follows:

$$h_t = \tanh(W_h \cdot h_{t-1} + W_x \cdot x_t + b_h), \tag{6}$$

and the output y_t at time t is calculated as [19]:

$$y_t = W_y \cdot h_t + b_y, \tag{7}$$

where h_t is the hidden state at time t, h_{t-1} is the hidden state of the past, x_t is the input at time t, W_h, W_x, W_y are weight matrices whose values are learned during the training process, b_h, b_y are bias vectors, and $\tanh$ is the hyperbolic tangent activation function [20].

2.4.3. Long Short-Term Memory

Long Short-Term Memory (LSTM) [21] is an improved variant of Recurrent Neural Network (RNN). This architecture is specifically designed to overcome the weaknesses of traditional RNNs, especially the problem of gradient decay, which makes it difficult to process data with long-term dependencies. The LSTM's ability to handle long-term delays is managed by three control gates, which use activation functions such as sigmoid and $\tanh$ to update and regulate the state of memory cells.

Assuming x_t is the input, the following functions are calculated by the LSTM cell:

$$i_t = \sigma(V_i x_t + W_i h_{t-1} + b_i), \tag{8}$$
$$f_t = \sigma(V_f x_t + W_f h_{t-1} + b_f), \tag{9}$$
$$\tilde{c}_t = \tanh(V_c x_t + W_c h_{t-1} + b_c), \tag{10}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \tag{11}$$
$$o_t = \sigma(V_o x_t + W_o h_{t-1} + b_o), \tag{12}$$
$$h_t = o_t \odot \tanh(c_t), \tag{13}$$

where i_t, f_t, c_t and o_t represent the input, forget, cell, and output gates, respectively. Meanwhile, V_*, W_*, and b_* represent learnable parameters, while h_* represents hidden states. Furthermore, σ is the sigmoid activation function, and $\odot$ denotes element-wise multiplication [22].

2.5. Model Performance Evaluation

In evaluating the accuracy of point forecasting results, four common evaluation metrics are used, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2) [23]. The calculation formulas for each evaluation metric used are as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{14}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{15}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{17}$$

where n is the total number of observations, y_i represents the actual observed value at time i, $\hat{y}_i$ represents the predicted value at time i, and $\bar{y}$ is the mean of the actual observed values.

3. Empirical Results/Experiments

3.1. Descriptive Statistics

Time series data on international tourist arrivals by entry point from 2008 to 2026 are shown in Figure 2, where air arrivals are marked in blue, sea arrivals in orange, and land arrivals in green.

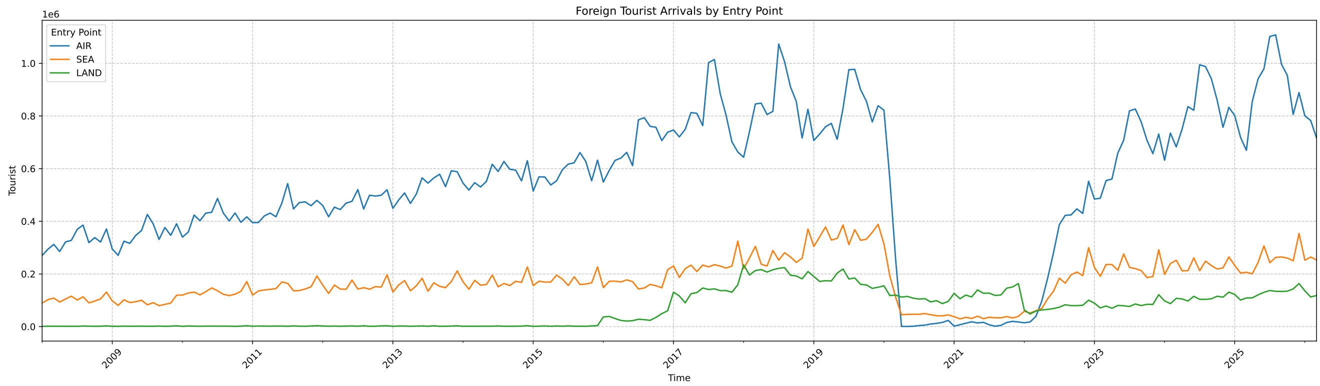


Figure 2. Graph of monthly international tourist arrivals by entry point (2008–2026)

The air arrival data records the highest volume and shows strong growth. This growth is interrupted by a drastic decline around 2020 due to the global pandemic, followed by a significant recovery. Meanwhile, the sea and land arrival data register smaller volumes but follow a similar trajectory, including the noticeable decline in 2020. Additionally, all three modes display consistent seasonal patterns characterized by regular fluctuations.

Table 1 presents the descriptive statistics for the three entry point categories. The air series dominates in both volume and variability, recording a mean of 540,505 arrivals per month and a standard deviation of 269,130, with values ranging from a minimum of 506 during the pandemic trough to a maximum of 1,108,133. The sea series records a mean of 174,453 with a standard deviation of 81,209, while the land series has the lowest mean at 67,192 despite exhibiting a coefficient of variation comparable to the air series, reflecting its structurally volatile and irregular nature. The wide gap between minimum and maximum values across all three categories confirms that the data experienced significant distributional shifts during the observation period, primarily attributable to the COVID-19 structural break. This high degree of non-stationarity and volatility underscores the necessity of employing adaptive decomposition methods prior to deep learning modeling.

Table 1. Descriptive statistics of monthly international tourist arrivals by entry point (2008–2026).

	AIR	SEA	LAND
count	219	219	219
mean	540505	174453	67192
std	269130	81209	68696
min	506	29300	1043
25%	395061	122938	1873
50%	551646	167909	60278
75%	747908	226831	120040
max	1108133	388495	234618

3.2. Data Preprocessing

As described in Section 2.2, the 219 monthly observations contain no missing values. The sequential data is chronologically split, allocating 70% for training and 30% for testing. It is then normalized using a Min-Max scaler to a range of 0 to 1 to ensure training stability.

3.3. Deep Learning Modeling

The optimal architectures for the RNN and LSTM networks are determined through an automated hyperparameter tuning process using the Optuna framework [24]. To construct the most effective network, the optimization algorithm explores a specific search space for the model parameters. The number of recurrent layers is set to vary between 1 and 2, while the number of neurons within each layer is tuned between 8 and 64 units. To mitigate the risk of overfitting, a dropout layer is incorporated, with its rate optimized from 0.1 to 0.4 in increments of 0.1. Furthermore, the batch size for the training process is selected from the categorical options of 8, 16, or 32.

During the evaluation phase, each model configuration is trained for a maximum of 200 epochs. An early stopping mechanism halts training when validation performance stops improving, helping maintain time efficiency and retaining the best model weights. The primary objective of the Optuna optimization is to find the exact combination of these hyperparameters that minimizes the Mean Squared Error on the test dataset. Once identified, these best-performing parameters are directly implemented for both the baseline architectures and the hybrid forecasting models.

Table 2. Optimal hyperparameter configurations for LSTM and RNN models.

	Model	Time step	Layers	Dense	Dropout rate	Epochs	Batch size
Air	LSTM	3	1	[46]	0.3	200	8
	RNN	3	2	[16, 17]	0.3	200	16
Sea	LSTM	3	2	[46, 54]	0.2	200	16
	RNN	3	2	[24, 22]	0.3	200	32
Land	LSTM	3	2	[31, 40]	0.4	200	8
	RNN	3	2	[32, 26]	0.4	200	16

Table 2 summarizes the optimal configurations obtained from the tuning process for the LSTM and RNN models across the air, sea, and land arrival datasets. A consistent time step of 3 and a maximum training limit of 200 epochs are applied uniformly across all configurations.

For the air dataset, the LSTM model achieves optimal performance using a single layer with 46 neurons and a batch size of 8. The RNN model, on the other hand, requires a two-layer architecture with 16 and 17 neurons, and a batch size of 16. Both models for this specific dataset share a dropout rate of 0.3.

Regarding the sea dataset, both networks perform best with a two-layer structure. The LSTM model is configured with 46 and 54 neurons, a dropout rate of 0.2, and a batch size of 16. Meanwhile, the RNN model utilizes 24 and 22 neurons, a dropout rate of 0.3, and a larger batch size of 32.

Finally, for the land dataset, the LSTM and RNN models both adopt a two-layer structure and share a higher dropout rate of 0.4. The LSTM model uses 31 and 40 neurons and a batch size of 8, whereas the RNN model uses 32 and 26 neurons and a batch size of 16.

3.4. VMD Signal Decomposition

After DL modeling, the cleaned time-series signals are decomposed using VMD. The number of IMF K and the penalty factor α are critical hyperparameters, as an excessively small K induces mode mixing while an excessively large K produces spurious IMFs.

Both parameters were optimized using the Optuna framework with Minimum Average Envelope Entropy as the objective function. Following [25] and [26], this metric represents signal sparsity, whereby its minimization guarantees focused IMFs with minimal noise and freedom from mode mixing. The search space was defined as $K \in \{3, \ldots, 12\}$ and $\alpha \in [100, 3000]$, explored via the Tree-structured Parzen Estimator (TPE) algorithm over 20 trials. Optimization results indicated an optimal K in the range of 10-12. A value of $K = 12$ was subsequently adopted by [4], which established this value for a comparable monthly tourism dataset, further supported by its proximity to the empirically derived optimum. The final parameters applied were $K = 12$, $\alpha = 715$, and a convergence tolerance of 10^{-7}.

Figure 3. VMD signal decomposition.

Decomposition results are presented in Figure 3. Each signal was successfully resolved into 12 IMFs representing a frequency hierarchy from low to high. IMF 1 captures the long-term trend, reflecting gradual growth from 2008 to 2019, a sharp decline in 2020, and a progressive recovery thereafter. IMFs 2-4 isolate multi-year seasonal cycles, IMFs 5-8 represent short-term seasonal variations, and IMFs 9-12 contain noise components and incidental local variations [4].

A cross-route comparison reveals that the AIR series is dominated by low-IMF components with the largest amplitude, the SEA series exhibits a more balanced inter-component profile, and the LAND series carries the highest high-frequency energy ratio, indicating the greatest structural volatility and consistent with its largest coefficient of variation reported in Table 1. Thus, the 12 IMFs from each entry point will proceed to the model training stage using LSTM and RNN, and then to the prediction process.

3.5. Hybrid VMD-Deep Learning Modeling

The hybrid modeling phase integrates the VMD decomposition results from Sub-section 3.4 with the optimized deep learning architectures (RNN and LSTM) established in Table 2. Rather than forecasting each IMF independently, this study adopts a multivariate feeding strategy in which all $K = 12$ IMFs extracted from each entry point are concatenated into a single multivariate matrix. Following [12], who demonstrated the effectiveness of feeding decomposed multivariate IMFs simultaneously into recurrent architectures, this study adopts an analogous strategy by restructuring the 12 univariate VMD-derived IMFs into a single multivariate input matrix, enabling the network to jointly learn inter-component frequency relationships in a single forward pass.

The combined IMF matrix is normalized using Min-Max scaling applied jointly across all components to the range [0, 1], ensuring consistent scale across IMFs of different frequency characteristics. The scaled data is then split chronologically with 70% allocated to training and the remaining 30% reserved for testing, preserving the temporal order of observations to prevent data leakage [14].

Multivariate input sequences are constructed using a sliding-window approach with a lookback window of 3 time steps, yielding input tensors of shape $(\text{samples} \times 3 \times 12)$. The network simultaneously receives all 12 IMF values at each time step and generates predictions for all 12 components concurrently through an output Dense layer of dimension $K = 12$. To ensure a controlled comparison, the hybrid networks adopt the exact same layer configurations, neuron counts, dropout rates, and batch sizes as their standalone counterparts identified in Table 2, isolating VMD preprocessing as the sole experimental variable. All models are trained with MSE as the loss function, the Adam optimizer, and an early stopping mechanism with a patience of 15 epochs that restores the best model weights upon convergence.

The final forecasted arrival series is reconstructed by summing the predicted values across all 12 IMF components, where the aggregated output is subsequently inverse-transformed to restore the original data scale prior to evaluation.

4. Discussion

4.1. Performance Analysis Across Entry Points

Analysis of the experimental results reveals important findings that challenge conventional paradigms in tourism time series forecasting. Based on the comprehensive evaluation metrics presented in Table 3, 4, 5, the single neural network model, specifically RNN, consistently demonstrates better generalization capabilities compared to both the LSTM model and the VMD based hybrid approach across all entry points.

Table 3. Model Evaluation Metrics - AIR Series

Set	Model Type	MAE	MSE	MAPE (%)	R2 (%)
train	Prophet	86715.12	2.66E+10	186510.16	41.81
	SARIMA	44864.13	4.38E+09	30055.14	90.40
	CNN	57473.03	7.86E+09	750.56	82.61
	GRU	46597.41	4.93E+09	299.40	89.09
	RNN	50268.32	6.27E+09	577.11	86.13
	LSTM	48047.37	5.45E+09	410.00	87.95
	VMD-RNN	68636.43	1.07E+10	1262.17	75.42
	VMD-LSTM	33527.14	2.43E+09	642.53	94.41
test	Prophet	314210.92	1.57E+11	257555.14	-18.32
	SARIMA	50804.43	4.31E+09	10931.99	96.75
	CNN	93962.41	1.27E+10	476.20	90.00
	GRU	52913.49	4.42E+09	112.65	96.51
	RNN	59042.15	5.09E+09	142.64	95.98
	LSTM	53937.86	4.45E+09	110.13	96.48
	VMD-RNN	83785.76	1.05E+10	313.23	91.94
	VMD-LSTM	57813.43	5.43E+09	104.30	95.84

Table 4. Model Evaluation Metrics - SEA Series

Set	Model Type	MAE	MSE	MAPE (%)	R2 (%)
train	Prophet	33020.04	3.01E+09	2894.22	47.15
	SARIMA	19726.67	7.89E+08	1352.51	86.17
	CNN	24704.35	1.30E+09	17.18	77.17
	GRU	22402.48	1.23E+09	16.41	78.48
	RNN	22620.78	1.31E+09	17.14	77.02
	LSTM	23475.00	1.31E+09	17.96	77.09
	VMD-RNN	19360.77	8.78E+08	16.38	84.40
	VMD-LSTM	14740.65	4.91E+08	11.87	91.27
test	Prophet	119852.66	2.05E+10	19727.24	-139.67
	SARIMA	22398.79	7.83E+08	1982.14	90.85
	CNN	34444.62	2.01E+09	43.82	75.19
	GRU	24530.96	1.09E+09	18.08	86.53
	RNN	23462.70	1.06E+09	17.02	86.90
	LSTM	25428.39	1.10E+09	21.54	86.40
	VMD-RNN	21628.53	6.90E+08	28.74	91.69
	VMD-LSTM	27754.42	1.26E+09	42.50	84.83

Table 5. Model Evaluation Metrics - LAND Series

Set	Model Type	MAE	MSE	MAPE (%)	R2 (%)
train	Prophet	27030.11	1.32E+09	52336.89	76.23
	SARIMA	5392.41	1.37E+08	2380.96	97.54
	CNN	10660.99	2.49E+08	246.82	95.57
	GRU	10616.37	2.42E+08	243.40	95.71
	RNN	6467.58	1.87E+08	56.06	96.67
	LSTM	8781.00	2.76E+08	103.41	95.09
	VMD-RNN	7198.61	1.63E+08	66.28	97.13
	VMD-LSTM	5057.12	6.88E+07	67.68	98.78
test	Prophet	160822.34	2.83E+10	16835.12	-3982.47
	SARIMA	10776.87	3.29E+08	1147.70	52.48
	CNN	11984.62	3.47E+08	12.48	51.74
	GRU	12263.93	3.99E+08	13.71	44.51
	RNN	11740.81	3.46E+08	12.72	51.96
	LSTM	13350.43	4.54E+08	15.04	36.94
	VMD-RNN	9563.82	1.71E+08	10.32	76.18
	VMD-LSTM	8875.90	1.73E+08	10.39	75.93

For the AIR Series Table 3, the evaluation metrics show that both conventional time-series models and deep learning architectures are capable of capturing stable seasonal trends with very high accuracy. On the test data, the standard SARIMA, GRU, LSTM, and RNN models demonstrated highly competitive generalization capabilities, with R^2 values ranging from 95.98% to 96.75%. Interestingly, with the $K = 12$ configuration, although the VMD-LSTM achieved an R^2 value of 95.84%, this model was still unable to outperform its single-model counterpart (LSTM: 96.48%). In fact, the VMD-RNN experienced a fairly significant drop in performance, with the R^2 value falling to 91.94%. This phenomenon aligns with the argument put forth by Zeng et al. [27], who state that in data environments with regular seasonal patterns and a short lookback window (3 time steps), excessive parameter complexity risks misinterpreting local noise as false deterministic patterns.

In the SEA Series Table 4, an interesting empirical shift occurred. Unlike the low-mode architecture in previous models, the $K = 12$ configuration successfully isolated the fundamental structural cycles of maritime volatility during the training phase (training set), where VMD-LSTM led with an R^2 of 91.27%. However, in the test set phase, there was a drastic reversal of results: the single RNN model emerged as the best architecture with the highest R^2 of 86.90% and the lowest MAPE of 17.02%. In contrast, the VMD-LSTM's performance dropped to an R^2 of 84.83%, with MAPE surging to 42.50%. This localized failure of the hybrid architecture indicates that when high-frequency IMFs encounter unexpected trend variations in out-of-sample data, the alignment of the sub-signals deviates sharply from the target vector.

The most positive impact of using the $K = 12$ configuration is clearly evident in the LAND Series Table 5, which is known to have the highest level of structural volatility due to land-boundary anomalies. Under these conditions, the hybrid model successfully overcomes the limitations experienced by single-model approaches. In the test data, VMD-RNN and VMD-LSTM dramatically outperformed all single models, recording dominant R^2 scores of 76.18% and 75.93%, respectively, and reducing the MAPE to around 10.3%. By comparison, single models such as RNN (51.96%), LSTM (36.94%), and SARIMA (52.48%) struggled significantly to capture the data patterns. This demonstrates that when data is heavily contaminated by high-frequency noise and sudden shocks, extracting 12 IMFs effectively isolates the chaotic residual variations into high-order IMF, allowing the neural network to optimally model the cleaner primary sub-signals.

The visualizations comparing predicted versus actual values reinforce the quantitative findings discussed earlier. For the AIR route in Figure 4, the single RNN model tracked the recovery trend following the pandemic most accurately through 2026 without any lag. Meanwhile, the multivariate hybrid architecture showed widening fluctuations around the trend inflection point. This widening fluctuation provides direct visual evidence of variance inflation, which is caused by forcing twelve IMF components to reconstruct a seasonal signal that was originally stable.

A similar pattern appears on the SEA route in Figure 5. The single RNN model acted as a macro filter that ignored local noise while consistently tracking the central path of the actual data. This filtering behavior explains its lower error variance on the test set shown in Table 4. In contrast, the prediction curve of the hybrid VMD and LSTM architecture appeared overfitted to subtle seasonal fluctuations, causing the predictions to deviate when applied to new validation data.

This pattern reverses entirely on the LAND route in Figure 6. This specific route is characterized by highly chaotic and low volume data with irregular extreme spikes caused by border crossing dynamics. All single models including RNN, LSTM, and CNN failed to anticipate the amplitude of the spikes and produced largely flat estimates. However, the prediction curve of the hybrid VMD model with twelve components closely tracked these random fluctuations. This accuracy confirms that the VMD decomposition effectively isolates temporary noise in highly volatile data environments.

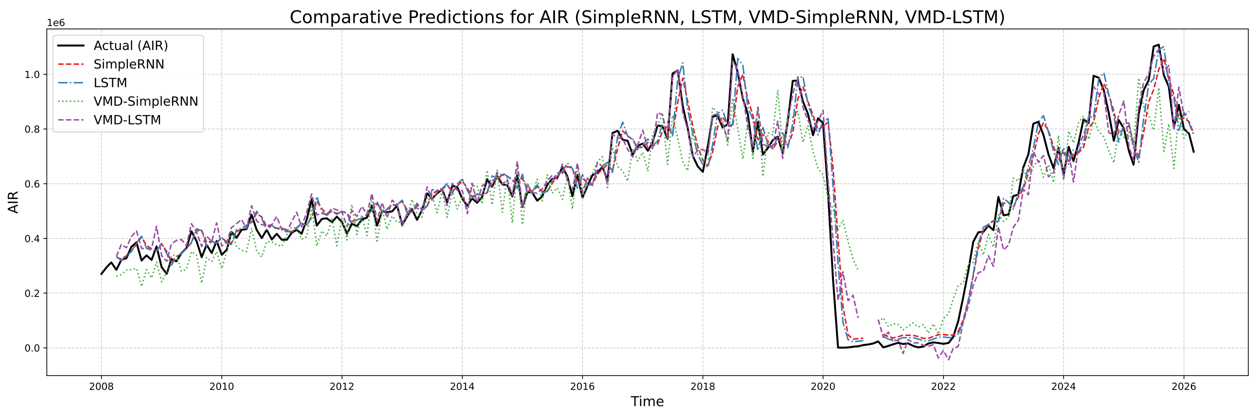


Figure 4. Comparison of predicted versus actual international tourist arrivals on the Air route.

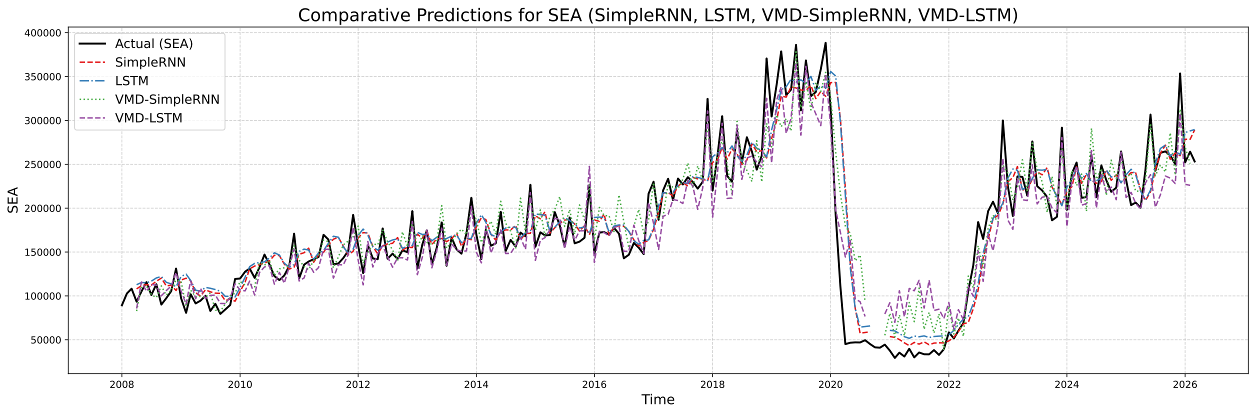


Figure 5. Comparison of predicted versus actual international tourist arrivals on the Sea route.

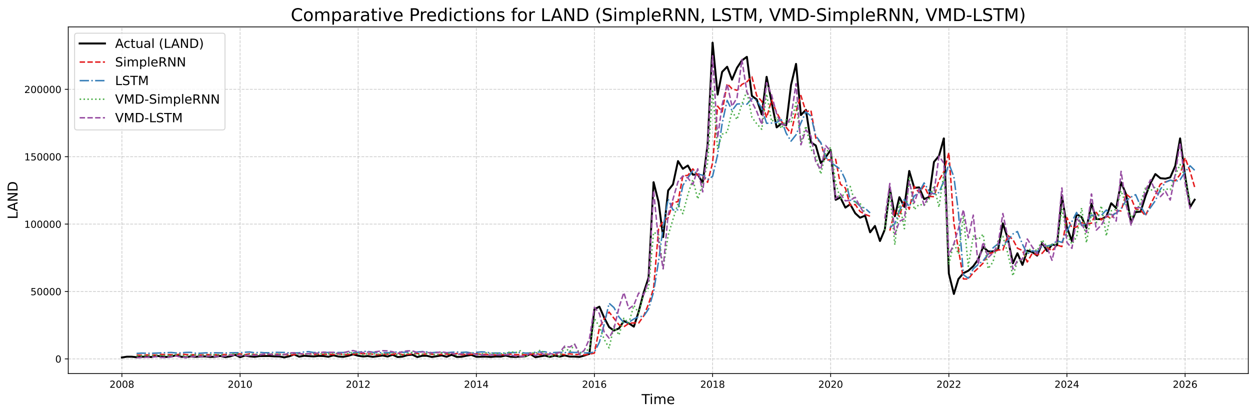


Figure 6. Comparison of predicted versus actual international tourist arrivals on the Land route.

A critical analysis of the MAPE metric in Figure 7 reveals highly conspicuous statistical inconsistencies in the AIR Series, where the MAPE values of some models skyrocketed to extreme levels, reaching thousands or even hundreds of thousands of percent (Table 3). These massive spikes do not reflect a fundamental failure of the models to recognize the direction of global tourism trends, but rather result from the intrinsic mathematical weakness of the MAPE formula itself when confronted with extreme minimum values in the actual data.

This massive inflation is caused by an intrinsic mathematical weakness in the MAPE calculation itself (as defined in Equation 16) when dealing with near-zero actual values. Based on the descriptive statistics in Table 1, the AIR route recorded its lowest-ever visit volume—just 506 visits—due to the impact of the COVID-19 pandemic in 2020. When the actual values (y_i) are very close to zero, the absolute error ($|y_i - \hat{y}_i|$) between the model's predictions and the actual data is amplified asymmetrically because the division is performed by a very small denominator. This formula's tendency to overforecast at the pandemic's trough point drastically inflates the percentage values in Figure 7, thereby distorting an objective assessment of the model's performance. Therefore, when tourism data is contaminated by near-zero anomaly shocks, the use of variance-based metrics such as R^2 is far more valid and stable than the MAPE metric, which is highly susceptible to denominator bias.

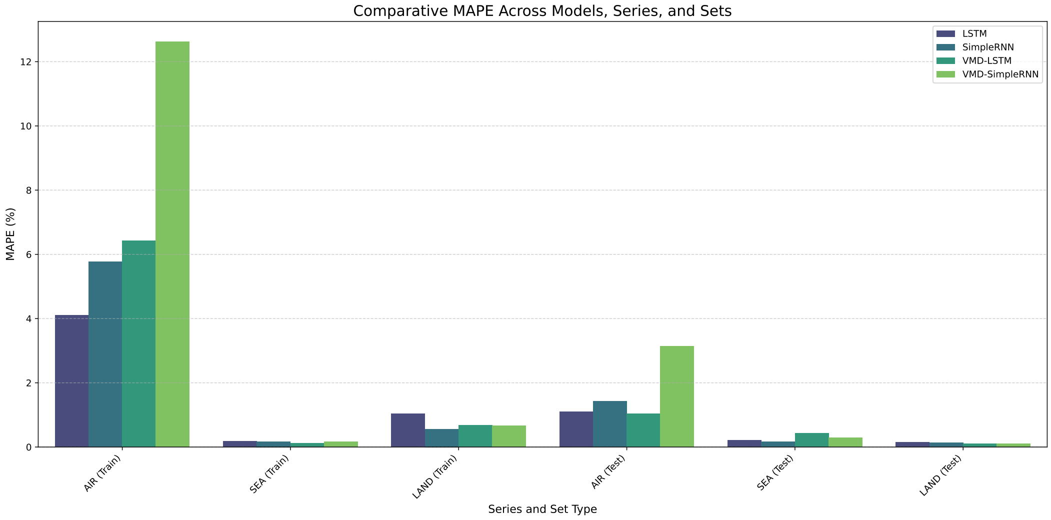


Figure 7. Comparison of MAPE values across models, series, and datasets.

The comparison of the Coefficient of Determination (R^2) values shown in Figure 8 confirms that the generalization characteristics of the models vary in contradictory ways across the three main input routes. For the AIR (AIR Series) and SEA (SEA Series) routes, the bar charts show a very stable dominance of altitude during both the training and testing phases, where most models were able to exceed the 0.8 threshold and approach 1.0. This pattern visually demonstrates that air and sea routes have data structures with well-established and regular macro-seasonal components, making them highly trackable for both single deep learning architectures and hybrid models. In contrast, a very drastic shift in performance was clearly evident in the LAND Test group. While single models such as LSTM and standard RNN experienced a massive drop in bar height to below the 0.6 corridor due to their inability to cope with the distortion caused by random data shocks, the bars for the hybrid VMD models (VMD-RNN and VMD-LSTM) demonstrated absolute superiority by consistently soaring above the 0.75 line. The visualization of bar height variations in Figure 8 provides conclusive empirical evidence: the multivariate feeding strategy proved to act as a crucial noise filter in a purely chaotic tourism data environment (land route), but tended to be redundant and trigger error accumulation in data streams that already had strong and stable sectoral trends (air and sea routes).

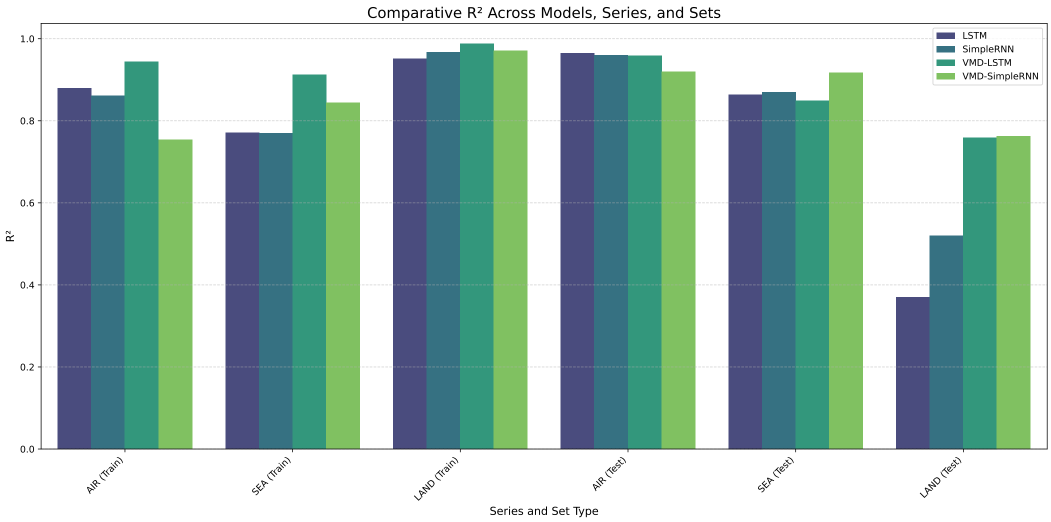


Figure 8. Comparison of Coefficient of Determination (R^2) values across models, series, and datasets.

4.2. Statistical Significance Analysis Using the Diebold-Mariano Test

To statistically validate the observed performance differences among competing models, the Diebold-Mariano (DM) test [28] was employed. The DM test is a widely adopted statistical procedure for comparing the predictive accuracy of two competing forecasting models based on their loss differential series [29]. Unlike simple metric comparisons, the DM test formally tests the null hypothesis that two models have equal predictive accuracy, providing a rigorous basis for determining whether observed performance differences are statistically significant or merely attributable to sampling variation. In this study, MSE is used as the loss function, and a two-sided test is applied at a significance level of $\alpha = 0.05$. A total of 84 pairwise comparisons are conducted across all model combinations and entry point series (28 pairs $\times$ 3 series). The results are reported in Table 6 and reveal several important patterns that both corroborate and nuance the metric-based findings discussed above.

Across all three entry points, Prophet consistently registers the largest negative DM statistics against all competing models (all $p < 0.0001$), confirming its categorical inferiority for this type of volatile, structurally disrupted tourism data. This finding is consistent with Prophet's known limitation in handling abrupt structural breaks, as the pandemic-induced collapse in 2020 fundamentally violates the smooth trend and seasonality assumptions underlying its generative model.

On the AIR route, the DM test reveals that RNN, LSTM, GRU, and VMD-LSTM are statistically equivalent to one another (all $p > 0.05$), suggesting that for high-volume air arrival data with relatively stable seasonal patterns, additional architectural complexity — whether through gating mechanisms (LSTM, GRU) or VMD preprocessing (VMD-LSTM) — does not yield statistically significant predictive gains over a simple recurrent architecture. Notably, VMD-RNN is statistically outperformed by RNN ($p = 0.0064$), LSTM ($p = 0.0033$), and GRU ($p = 0.0035$), providing direct statistical evidence that the VMD decomposition strategy degraded rather than improved forecasting accuracy for this series. This outcome is consistent with the error accumulation argument advanced in the main analysis — the linear recombination of 12 independently predicted IMFs amplifies total reconstruction error, particularly for a series with strong and stable low-frequency trend components that a direct recurrent model can already capture effectively [30].

Table 6. Diebold-Mariano Test Results for Pairwise Model Comparisons Across Entry Points

Model 1	Model 2	AIR		SEA		LAND	
		DM	Winner	DM	Winner	DM	Winner
SARIMA	Prophet	-6.10^{***}	SARIMA	-6.18^{***}	SARIMA	-15.39^{***}	SARIMA
SARIMA	VMD-LSTM	-0.73	—	-1.79	—	0.82	—
SARIMA	VMD-RNN	-2.82^{**}	SARIMA	0.54	—	0.82	—
SARIMA	RNN	-0.74	—	-1.38	—	-0.01	—
SARIMA	LSTM	0.09	—	-1.65	—	-0.64	—
SARIMA	CNN	-4.46^{***}	SARIMA	-3.76^{***}	SARIMA	-0.04	—
SARIMA	GRU	0.14	—	-1.59	—	-0.39	—
Prophet	VMD-LSTM	6.03^{***}	VMD-LSTM	6.28^{***}	VMD-LSTM	15.31^{***}	VMD-LSTM
Prophet	VMD-RNN	5.82^{***}	VMD-RNN	6.35^{***}	VMD-RNN	15.25^{***}	VMD-RNN
Prophet	RNN	6.05^{***}	RNN	6.01^{***}	RNN	15.31^{***}	RNN
Prophet	LSTM	6.08^{***}	LSTM	6.02^{***}	LSTM	15.22^{***}	LSTM
Prophet	CNN	5.84^{***}	CNN	5.84^{***}	CNN	15.27^{***}	CNN
Prophet	GRU	6.07^{***}	GRU	5.99^{***}	GRU	15.28^{***}	GRU
VMD-LSTM	VMD-RNN	-2.17^{*}	VMD-LSTM	3.37^{**}	VMD-RNN	0.02	—
VMD-LSTM	RNN	0.25	—	0.60	—	-1.24	—
VMD-LSTM	LSTM	0.81	—	0.50	—	-1.81	—
VMD-LSTM	CNN	-3.30^{**}	VMD-LSTM	-2.03^{*}	VMD-LSTM	-1.43	—
VMD-LSTM	GRU	0.77	—	0.50	—	-1.53	—
VMD-RNN	RNN	2.83^{**}	RNN	-1.27	—	-1.26	—
VMD-RNN	LSTM	3.06^{**}	LSTM	-1.48	—	-1.81	—
VMD-RNN	CNN	-1.14	—	-3.55^{***}	VMD-RNN	-1.48	—
VMD-RNN	GRU	3.04^{**}	GRU	-1.41	—	-1.56	—
RNN	LSTM	1.45	—	-1.19	—	-1.18	—
RNN	CNN	-5.43^{***}	RNN	-4.53^{***}	RNN	-0.12	—
RNN	GRU	1.28	—	-0.85	—	-0.94	—
LSTM	CNN	-4.94^{***}	LSTM	-4.47^{***}	LSTM	1.21	—
LSTM	GRU	0.12	—	0.25	—	1.40	—
CNN	GRU	5.13^{***}	GRU	4.42^{***}	GRU	-0.91	—

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. A positive DM statistic indicates Model 2 outperforms Model 1.
— indicates no statistically significant difference between the two models ($p \geq 0.05$).

On the SEA route, a similar pattern of statistical equivalence emerges among RNN, LSTM, GRU, VMD-LSTM, and VMD-RNN (all pairwise $p > 0.05$), with the exception that VMD-RNN statistically outperforms CNN ($p = 0.0007$) and VMD-LSTM statistically outperforms CNN ($p = 0.0472$). SARIMA performs comparably to most deep learning models on this route, reflecting the relatively more regular seasonal structure of sea arrival data. These results collectively suggest that on the SEA route, the choice among recurrent architectures — with or without VMD preprocessing — does not produce statistically distinguishable outcomes, and model selection may therefore be guided by computational efficiency considerations.

The LAND route presents the most striking pattern: with the sole exception of Prophet, no statistically significant differences are detected among any of the competing models (all pairwise $p > 0.05$). This blanket statistical equivalence across SARIMA, CNN, GRU, RNN, LSTM, VMD-LSTM, and VMD-RNN provides strong evidence that the extreme volatility and structural uncertainty characterizing land arrival data fundamentally limits the discriminative power of any modeling approach tested in this study. Under such conditions, the marginal gains from architectural sophistication or decomposition preprocessing are statistically indistinguishable from noise, reinforcing the conclusion that land entry point forecasting represents a fundamentally harder prediction problem that warrants dedicated future investigation.

Taken together, the DM test results confirm that hybrid VMD models do not achieve statistically superior forecasting accuracy over their direct recurrent counterparts on any of the three entry point series. This finding constitutes the primary empirical contribution of this study — challenging the prevailing assumption in the tourism forecasting literature that decompose-and-ensemble strategies inherently improve predictive performance [4].

5. Conclusion

This study compares the effectiveness of a single deep learning architecture against a hybrid framework based on multivariate decomposition for forecasting international tourist arrivals to Indonesia via air, sea, and land routes. The experimental results challenge the common assumption that strategies combining decomposition and ensemble methods consistently improve forecasting accuracy.

The main scientific contribution of this study is revealing the dual nature of high order multivariate decomposition. On stable and high volume data such as air and sea routes, combining all component signals into a single input matrix actually triggers an accumulation of reconstruction error variance. This accumulation makes single models like RNN superior. Conversely, on highly chaotic and noisy data such as land routes, this multivariate framework proves effective as a feature filter. It isolates random fluctuations into higher order IMFs and enables the network to map underlying trends more accurately.

Thus, architectural complexity is not an absolute guarantee of higher accuracy. Simpler models are more robust at preserving the integrity of macro trends on regular routes, while multivariate hybrid models excel only when extreme volatility dominates the data. Future research should explore adaptive decomposition parameter selection using metaheuristic algorithms, along with integrating attention mechanisms or probabilistic forecasting frameworks to enhance model flexibility against future structural shifts in the data.

References

[1] I. M. Hasibuan, S. Mutthaqin, R. Erianto, and I. Harahap, "Kontribusi sektor pariwisata terhadap perekonomian nasional," *Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah*, vol. 8, 2023. DOI: 10.30651/jms.v8i2.19280.

[2] K. Wisnubroto, *Indonesia.go.id - geliat sektor pariwisata pacu pertumbuhan ekonomi*, Accessed: Dec. 03, 2025, 2025.

[3] Kompassindo, *Bps: Inflasi, pariwisata, dan transportasi alami tren positif di awal 2025 - kompassindo*, Accessed: Dec. 03, 2025, 2025.

[4] C. Zhang, M. Li, S. Sun, L. Tang, and S. Wang, "Decomposition methods for tourism demand forecasting: A comparative study," *Journal of Travel Research*, vol. 61, no. 7, pp. 1682–1699, 2022. DOI: 10.1177/00472875211036194.

[5] H. Song, R. T. Qiu, and J. Park, "A review of research on tourism demand forecasting: Launching the annals of tourism research curated collection on tourism demand forecasting," *Annals of Tourism Research*, vol. 75, pp. 338–362, 2019. DOI: 10.1016/j.annals.2018.12.001.

[6] M. Abdou, E. Musabanganji, and H. Musahara, "Tourism demand modelling and forecasting: A review of literature," *African Journal of Hospitality, Tourism and Leisure*, vol. 10, no. 4, pp. 1370–1393, 2021. DOI: 10.46222/ajhtl.19770720-168.

[7] Y. Zhang, W. H. Tan, and Z. Zeng, "Tourism demand forecasting based on a hybrid temporal neural network model for sustainable tourism," *Sustainability*, vol. 17, no. 5, p. 2210, 2025. DOI: 10.3390/SU17052210.

[8] Y. Zhang, G. Li, B. Muskat, and R. Law, "Tourism demand forecasting: A decomposed deep learning approach," *Journal of Travel Research*, vol. 60, no. 5, pp. 981–997, 2021. DOI: 10.1177/0047287520919522.

[9] R. Law, G. Li, D. K. C. Fong, and X. Han, "Tourism demand forecasting: A deep learning approach," *Annals of Tourism Research*, vol. 75, pp. 410–423, 2019. DOI: 10.1016/j.annals.2019.01.014.

[10] X. Wang et al., "Air quality forecasting using a spatiotemporal hybrid deep learning model based on vmd-gat-bilstm," *Scientific Reports*, vol. 14, no. 1, p. 17 841, 2024. DOI: 10.1038/s41598-024-68874-x.

[11] F. Aksan, V. Suresh, P. Janik, and T. Sikorski, "Load forecasting for the laser metal processing industry using vmd and hybrid deep learning models," *Energies*, vol. 16, no. 14, p. 5381, 2023. DOI: 10.3390/en16145381.

[12] Z. Zheng et al., "Multivariate data decomposition based deep learning approach to forecast one-day ahead significant wave height for ocean energy generation," *Renewable and Sustainable Energy Reviews*, vol. 185, p. 113 645, Oct. 2023, ISSN: 1364-0321. DOI: 10.1016/J.RSER.2023.113645.

[13] W. Zhang, M. K. Lim, M. Yang, X. Li, and D. Ni, "Using deep learning to interpolate the missing data in time-series for credit risks along supply chain," *Industrial Management & Data Systems*, vol. 123, no. 5, pp. 1401–1417, 2023. DOI: 10.1108/IMDS-08-2022-0468.

[14] H. Han, J. Peng, J. Ma, S. L. Liu, and H. Liu, "Research on load forecasting based on ceemdan se vmd and selfattention tcn fusion model," *Scientific Reports*, vol. 15, pp. 14 530–, 1 Apr. 2025, ISSN: 2045-2322. DOI: 10.1038/s41598-025-98224-4.

[15] S. Wu and H. Cai, "Short-term power load prediction of vmd-lstm based on issa optimization," *Applied Sciences*, vol. 15, no. 9, p. 5037, 2025. DOI: 10.3390/app15095037.

[16] S. M. Shin, A. Rasheed, P. Kil-Heum, and K. C. Veluvolu, "Fast and accurate short-term load forecasting with a hybrid model," *Electronics*, vol. 13, no. 6, p. 1079, 2024. DOI: 10.3390/electronics13061079.

[17] I. Jrhilifa, H. Ouadi, A. Jilbab, N. Mounir, and A. Ouaguid, "A vmd-deep learning approach for individual load monitoring and forecasting for residential buildings energy management," *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, p. 100 624, 2024. DOI: 10.1016/j.prime.2024.100624.

[18] M. Tayebi and S. E. Kafhali, "Performance analysis of recurrent neural networks for intrusion detection systems in industrial-internet of things," *Franklin Open*, vol. 12, p. 100 310, 2025. DOI: 10.1016/j.fraope.2025.100310.

[19] D. H. Hopfe, K. Lee, and C. Yu, "Short-term forecasting airport passenger flow during periods of volatility: Comparative investigation of time series vs. neural network models," *Journal of Air Transport Management*, vol. 115, p. 102 525, 2024. DOI: 10.1016/j.jairtraman.2023.102525.

[20] J. N. Munyao, L. A. Oluoch, H. Iftikhar, and P. C. Rodrigues, "Recurrent neural networks for hierarchical time series forecasting: An application to the s&p 500 market value," *Physica A: Statistical Mechanics and Its Applications*, vol. 678, p. 130 869, Nov. 2025, ISSN: 0378-4371. DOI: 10.1016/J.PHYSA.2025.130869.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 8 Nov. 1997, ISSN: 08997667. DOI: 10.1162/NECO.1997.9.8.1735.

[22] J. Korstanje, *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*. Apress Media LLC, Jul. 2021, ISBN: 9781484271506. DOI: 10.1007/978-1-4842-7150-6.

[23] D. Chen, H. Li, S. Sun, J. Bai, and Z. Huang, "Point and interval forecasting approach for short-term urban subway passenger flow based on residual term decomposition and fuzzy information granulation," *Applied Soft Computing*, vol. 166, p. 112 187, 2024. DOI: 10.1016/j.asoc.2024.112187.

[24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2623–2631, Jul. 2019. DOI: 10.1145/3292500.3330701.

[25] J. Ding, L. Huang, D. Xiao, and X. Li, "Gmpso-vmd algorithm and its application to rolling bearing fault feature extraction," *Sensors*, vol. 20, p. 1946, 7 Mar. 2020, ISSN: 1424-8220. DOI: 10.3390/S20071946.

[26] T. Liang, H. Lu, and H. Sun, "Application of parameter optimized variational mode decomposition method in fault feature extraction of rolling bearing," *Entropy*, vol. 23, p. 520, 5 Apr. 2021, ISSN: 1099-4300. DOI: 10.3390/E23050520.

[27] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 11 121–11 128. DOI: 10.1609/aaai.v37i9.26317.

[28] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, pp. 253–263, 3 1995, ISSN: 15372707. DOI: 10.1080/07350015.1995.10524599.

[29] D. Harvey, S. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, vol. 13, pp. 281–291, 2 Jun. 1997, ISSN: 0169-2070. DOI: 10.1016/S0169-2070(96)00719-4.

[30] Q. Qin and L. Li, "A vmd-based four-stage hybrid forecasting model with error correction for complex coal price series," *Mathematics*, vol. 13, no. 18, p. 2912, 2025. DOI: 10.3390/math13182912.