

Article

VMD-Enhanced Multi-Model Ensemble Framework for Tourism Demand Forecasting: A Feature Engineering Approach

Cuicui Yu ¹, Wangyu Shen ^{2,3} , Erlong Zhao ⁴, Shouyang Wang ^{1,3,5,6,7}  and Pei Du ^{8,*}

¹ School of Economics and Management, Xidian University, Xi'an 710126, China; cuicuiyu2013@163.com (C.Y.); sywang@amss.ac.cn (S.W.)

² School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China; shenwy2024@shanghaitech.edu.cn

³ School of Entrepreneurship and Management, ShanghaiTech University, Shanghai 201210, China

⁴ School of Statistics, Xi'an University of Finance and Economics, Xi'an 710100, China; elzhao@xaupe.edu.cn

⁵ Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

⁶ School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

⁷ Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China

⁸ School of Business, Jiangnan University, Wuxi 214122, China

* Correspondence: dupeidp@jiangnan.edu.cn

Highlights

What are the main findings?

- This research indicates that combining VMD-based multi-scale feature construction with a validation-driven weighted ensemble strategy effectively optimizes the contributions of different base models and significantly enhances the ability to capture complex dynamic patterns in tourist volume time series.
- The empirical findings further demonstrate that the proposed three-stage VMD-enhanced “decomposition-features-ensemble” forecasting framework consistently outperforms multiple mainstream time series forecasting models across various evaluation metrics and statistical tests, exhibiting superior accuracy, robustness, and overall effectiveness.

What are the implications of the main findings?

- From a theoretical perspective, the proposed three-stage VMD-enhanced ensemble forecasting framework redefines the role of signal decomposition in tourism forecasting, offering a more robust, computationally efficient, and interpretable forecasting paradigm for limited samples and non-stationary time series, with strong methodological transferability.
- From a practical perspective, the proposed framework enhances prediction accuracy and stability in high-volatility demand scenarios such as tourist flow forecasting while also providing a powerful and scalable solution for advanced forecasting tasks in domains including energy, finance, and climate science.



Academic Editor: Sonia Leva

Received: 28 January 2026

Revised: 14 July 2026

Accepted: 21 July 2026

Published: 5 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Abstract

This paper proposes a novel variational mode decomposition (VMD)-enhanced ensemble framework for tourism demand forecasting, which achieves competitive predictive performance. The core innovation is a structured hybrid methodology that first employs VMD to decompose complex temporal signals, from which multiple interpretable features—encoding trend, volatility, and momentum characteristics—are extracted. These enriched features are then processed by an intelligently weighted ensemble of three diverse models: support vector regression (SVR), gradient boosting regression (GBR), and random forest

(RF). An adaptive weighting mechanism dynamically optimizes the contribution of each base model, enhancing robustness and accuracy. Comprehensive experiments demonstrate the framework's superior effectiveness. Compared to strong contemporary benchmarks such as CNN-LSTM, our approach achieves significant error reduction: an RMSE of 0.837 (16.29% improvement), an MAE of 0.584 (12.19% improvement), and a MAPE of 15.42% (13.38% improvement). These results validate that the integration of VMD-based feature engineering with an adaptive multi-model ensemble effectively captures complex temporal dynamics and mitigates individual model limitations.

Keywords: time series forecasting; tourism demand; feature engineering; VMD

1. Introduction

Tourism demand forecasting is widely recognized as a fundamental issue in tourism management and smart tourism, and it has been extensively utilized to support critical decision-making processes [1–3]. Tourists' travel choices and consumption behaviors exhibit significant temporal and spatial dynamics [4], resulting in demand series that often display pronounced non-stationarity and structural shifts. In particular, both the mean and variance of tourism demand tend to drift due to seasonal transitions, policy adjustments, and evolving supply-side conditions [5]. Moreover, abrupt level shifts, sharp peaks, and volatility clustering frequently occur in response to holiday periods, major events, or unexpected shocks [6]. Such fluctuations are typically driven by the superposition of multiple temporal components [7]: low-frequency components associated with long-term trends and seasonal cycles coexist with mid- to high-frequency components linked to intraweek regularities and short-term disturbances, and cross-scale interference and coupling may further occur. Additionally, external mechanisms, including heterogeneous tourist behavior, information diffusion, and policy intervention, can give rise to nonlinear response [3], thereby amplifying inter-scale coupling. These complexities impose increasingly stringent requirements on forecasting models in terms of accuracy, robustness, and interpretability.

As a well-established strategy for handling non-stationarity and multi-scale fluctuation patterns in tourism demand series, decomposition-ensemble forecasting frameworks have been widely employed and have shown consistently strong performance across diverse tourism demand forecasting settings [4,8,9]. Within this line of research, variational mode decomposition (VMD) extracts band-limited modes by solving a variational optimization problem [10]. Relative to conventional methods such as empirical mode decomposition, VMD is generally viewed as having a firmer theoretical grounding, greater robustness against mode mixing, and more stable decompositions, which has contributed to its widespread adoption in the forecasting literature [3,11,12]. Nevertheless, much of the extant work leverages decomposed information via a “decompose-forecast separately-reconstruct” pipeline [3,13]. Under this paradigm, component-level forecast errors can accumulate and propagate during reconstruction, ultimately weakening overall robustness. Moreover, decomposed modes are often not exploited in a structured, integrative manner; instead, they are treated as independent sub-series and modeled in isolation. Because the extracted modes are not necessarily strictly independent, such naive component-wise forecasting may ignore cross-scale coupling and interdependencies, preventing models from fully benefiting from information shared across temporal scales.

In recent years, deep learning approaches have been increasingly applied to time-series forecasting to capture nonlinearities and complex dependency structures, with strong performance reported in domains such as tourism demand prediction [14–16]. Neverthe-

less, deep models are typically highly parameterized and entail substantial training and inference costs, which in turn raise demands on computational resources, hyperparameter optimization, and data availability [17]. These demands can pose practical constraints in applied settings characterized by limited sample sizes and the need for rapid deployment. By contrast, conventional machine-learning models can deliver comparable or even superior predictive performance on small- to medium-sized datasets at markedly lower computational cost, making them more suitable for operational use [17]. Meanwhile, despite the proliferation of advanced forecasting techniques, empirical performance is often sensitive to shifts in data distributions, noise intensity, and forecasting horizons; accordingly, methodological advantages are rarely invariant across contexts [3]. Prior research in tourism forecasting further indicates that the intrinsic complexity of sequential data reduces the likelihood that any single method will remain consistently optimal across problems [18,19], thereby motivating the use of ensemble learning to improve robustness and predictive accuracy.

In response to these considerations, a VMD-enhanced ensemble model is proposed as a VMD-augmented ensemble forecasting framework. Within this framework, complex temporal signals are first decomposed using variational mode decomposition (VMD), after which a 20-dimensional set of interpretable features is constructed from both time- and frequency-domain perspectives, such that trend, volatility, and momentum characteristics are jointly represented. The resulting enriched representations are then used for prediction through a dynamically weighted ensemble in which three base learners, i.e., support vector regression (SVR), gradient boosting regression (GBR), and random forest (RF), are integrated, and their contributions are adaptively determined. Finally, the proposed framework is benchmarked against seven representative time-series forecasting methods, and its effectiveness is further evaluated within a multiple-comparison statistical testing framework, through which the robustness of the observed performance differences is assessed.

The main contributions of this study are threefold: (1) A highly interpretable multi-scale feature fusion strategy is proposed, through which time-domain statistical descriptors and frequency-domain features derived from VMD components are organically integrated to form a 20-dimensional enhanced feature space, enabling a comprehensive characterization of tourist flow dynamics across temporal scales. (2) A validation-driven weighted ensemble strategy is designed, where the best base learner is dynamically selected on the validation set and an “emphasize-the-best” weighting scheme is adopted, thereby allowing better adaptation to scenario-specific characteristics while retaining the robustness of ensemble learning and highlighting the predictive strength of superior models. (3) A three-stage tourist volume forecasting framework “decomposition-features-ensemble” is further established and is empirically examined using two representative scenic areas in Beijing, where effectiveness and robustness are systematically validated via a multiple statistical testing framework.

The remainder of the paper is structured as follows. Section 2 reviews the existing literature on decomposition-ensemble methods and machine learning applications. Section 3 elaborates the proposed methodology, including the framework, feature engineering, and ensemble strategy. Section 4 presents the experimental setup, comparative results, and analysis. Finally, Section 5 concludes the study and discusses future work.

2. Literature Review

2.1. Application of Decomposition-Ensemble Methods in Tourism Forecasting

To address the complex characteristics of time series, decomposition-ensemble forecasting has been developed to partition an original prediction task into multiple subproblems and to aggregate the resulting forecasts [13,20], whereby predictive accuracy can

be improved under a “divide-and-conquer” paradigm [21]. In tourism forecasting, the decomposition techniques that have been adopted can be broadly classified into three categories depending on how the underlying signal is processed: filtering, denoising, and multi-scale decomposition [7,22].

In filtering-based decomposition, an observed series is typically separated into a trend component and a cyclical component, such that information conveyed by long-run evolution can be distinguished from business-cycle fluctuations. Representative methods include the Hodrick–Prescott (HP) filter and the Hamilton filter [23]. Using Australian tourist arrivals, Bosupeng [23] compared these two filters and reported that more effective out-of-sample gains were achieved by the Hamilton filter, which also outperformed the HP filter when applied to the original arrivals series.

Denoising-based methods aim to mitigate the masking effect of noise on the series' frequency components and structural information through a decomposition-filtering-reconstruction procedure; the reconstructed signal is then fed into forecasting models to enhance accuracy. In tourism forecasting, singular spectrum analysis (SSA) and wavelet decomposition are two representative denoising approaches. SSA achieves component separation and reconstruction via embedding, singular value decomposition, grouping, and diagonal averaging, and it often treats components associated with small eigenvalues as noise to be removed [24,25], while wavelet decomposition, grounded in time-frequency analysis, expands the series across multiple resolutions, enabling simultaneous characterization of long-run properties and local fluctuations [26]. For example, Kummong and Supratid [26] proposed a hybrid forecasting framework based on discrete wavelet decomposition, extracting “latent features” at different frequency resolutions to improve forecasting performance for Thailand's tourism series. Besides, Hassani et al. [24] applied SSA to denoise monthly U.S. inbound arrivals and achieved higher forecasting accuracy, with subsequent evidence reporting similar gains for monthly international arrivals across European countries [27].

Multi-scale decomposition, by contrast, treats a time series as a superposition of components operating over distinct frequency bands. By following a decomposition-component forecasting-forecast combination workflow, such methods aim to capture nonlinear and nonstationary dynamics in tourism demand while representing short-term fluctuations, medium-term shocks, and long-term trends within a unified framework. For instance, Chen et al. [28] proposed a multi-series structural time series (STS) approach in which a seasonal series is re-stacked into season-specific sub-series so that inter-season dependence can be modeled within a state-space structure without requiring a fully multivariate specification. Zhang et al. [7] showed that STL-based decomposition can be strengthened when it is coupled with a deep-learning module that performs data-driven feature and lag selection, thereby alleviating overfitting while improving both short- and long-horizon forecasts. Li et al. [22] employed EEMD to separate monthly arrivals into intrinsic components at multiple frequencies, which were forecasted individually and then recombined; more stable gains than non-decomposed benchmarks were reported across markets. More recently, Xie et al. [4] developed a CEEMDAN-based decomposition-ensemble framework in which adaptive noise decomposition is complemented by component reconstruction guided by complexity analysis, and improvements in both point and interval forecasting were demonstrated using Hong Kong tourism demand data. Among decomposition techniques, VMD offers a theoretically grounded variational formulation that yields band-limited modes with well-separated center frequencies, conferring advantages in interpretability over empirically driven alternatives such as EMD and CEEMDAN [10,11].

Despite continued progress, most decomposition-ensemble forecasting studies still operationalize decomposed outputs in two relatively narrow ways: modes are either forecast

separately and additively reconstructed or appended as auxiliary inputs to downstream predictors. In both cases, a key gap persists: multi-scale structure is seldom encoded as a joint representation that can be learned across scales, so that reconstruction pipelines remain vulnerable to the accumulation of component-wise errors, while “components-as-features” approaches may fail to make inter-scale dependencies explicit. This omission is nontrivial in tourism demand, where shocks frequently manifest concurrently across temporal scales. To address this gap, a compact and interpretable cross-scale representation is constructed by fusing time-domain descriptors with frequency-domain characteristics extracted from decomposed components, after which a VMD-augmented ensemble framework is built on this unified feature space so that cross-scale co-movements can be exploited more directly and robustly in forecasting.

2.2. Application of Machine Learning Methods in Tourism Forecasting

In recent years, as machine-learning models have matured and computing resources have expanded, tourism demand forecasting has increasingly shifted toward AI paradigms that are more data-driven and better suited to nonlinear patterns [16,29,30]. Broadly, two streams can be distinguished by their modeling mechanisms and representation capacity: (i) machine-learning approaches that rely on feature engineering and (ii) deep-learning approaches designed to capture long-range dependencies.

In feature-engineered machine learning, historical signals and exogenous inputs are first transformed into learnable features, after which relatively shallow nonlinear learners are used to approximate the mapping from features to demand, so that nonlinearity and structural change can be better represented. Support vector regression (SVR) has often been adopted as a benchmark, where kernels project data into high-dimensional spaces to fit complex relations and have been shown to outperform ARIMA-like baselines out of sample [6]. Related variants, such as fuzzy time series, genetic-algorithm tuning, BPNN, and RVFL, have also been introduced to improve flexibility and generalization [3,8]. For efficiency, extreme learning machines (ELM) and kernelized ELM have been used to reduce training costs via minimal iteration or closed-form solutions, with further gains reported when search-based behavioral data are incorporated [31].

Deep learning, by contrast, learns higher-order temporal representations directly from sequences through gated recurrent networks or attention mechanisms and is particularly effective at modeling both short- and long-term dependencies under strong seasonality. LSTM-based RNNs mitigate vanishing gradients via gating and have therefore been found to outperform SVR and conventional neural baselines in multiple settings [14]. Building on this, Bayesian BiLSTM combines bidirectional dependence learning with automated hyperparameter tuning to improve stability [32], while image-based encoding with CNNs has been used to capture local shapes and nonlinear swings around peaks and turning points [33]. More recently, hybrid designs have been emphasized, including decomposition coupled with Temporal Fusion Transformer for robust multivariate forecasting [34] and mixed-frequency BiLSTM-MIDAS models that align high-frequency searches with low-frequency arrivals to better exploit timely information [35]. Nevertheless, deep architectures typically require large training samples and entail substantial computational costs, and their internal representations remain difficult to inspect—a practical concern in tourism management contexts where forecast interpretability and operational deployability are essential [17].

Nevertheless, there remains scope to further improve computational efficiency and interpretability. To this end, we propose a machine-learning ensemble framework designed for practical deployment, with inputs comprising statistical, trend, volatility, and momentum features. Because these features have clear behavioral and time-series interpretations,

the resulting model is more interpretable and auditable than deep-learning “black-box” alternatives that are typically large in parameter size and opaque in mechanism. Moreover, by adopting a validation-driven strategy for dynamic model selection and weighted ensembling, the influence of each base learner is adaptively adjusted based on observed performance, thereby improving forecasting stability and robustness across datasets and prediction horizons.

2.3. Data Design and Out-of-Sample Evaluation in Tourism Forecasting

A further stream of recent tourism forecasting research emphasizes that the informational design of a forecasting task can be as important as the algorithm itself. Studies using multisource big data, short-video information, online news, search-query denoising, Google Trends, and improved mixed-data sampling models show that variables should be selected according to behavioral relevance, data timeliness, and measurement frequency rather than convenience alone [36–41]. This evidence is particularly relevant for tourism demand, where changes in attention, search, mobility, or media signals may precede realized changes in visitor volume.

The same logic applies to the temporal and spatial organization of the series. Recent studies have shown that spatiotemporal features, local and nested spatiotemporal models, multi-attraction hourly data, time- and feature-varying structures, and seasonality-cycle representations can materially change predictive performance [42–47]. In the general forecasting literature, feature-based model averaging, feature-based performance prediction, interpretable multi-horizon forecasting, and Bayesian forecast combinations using time-varying features also support the view that model choice should be conditioned on data characteristics and forecast horizons [48–51].

Finally, tourism forecasting should be evaluated under leakage-free out-of-sample conditions whenever the intended use is operational decision-making. Real-time forecasting assessments and the COVID-19 tourism forecasting competition show that predictive performance can change substantially when models face structural breaks, recovery dynamics, and data restrictions [52–54]. These findings motivate the present study’s use of a chronological holdout design and a direct comparison with earlier deep-learning and decomposition-based benchmarks.

3. Materials and Methods

3.1. The Framework

To systematically characterize the non-stationarity and the multi-scale fluctuation structure of tourism-demand series, a variational mode decomposition-based enhanced ensemble forecasting model is proposed, which is referred to as VMD-enhanced ensemble. The method follows a four-staged pipeline comprising decomposition, feature extraction, ensemble learning, and evaluation. The framework of our VMD-enhanced ensemble method is illustrated in Figure 1.

In stage I, the original series is decomposed via VMD, where the signal is separated into $K = 4$ band limited modal components, so that information spanning low-frequency long-run movements and high-frequency short-run disturbances can be explicitly represented. This is achieved by constructing and solving a variational optimization problem through which the center frequencies and bandwidths of the modes are simultaneously estimated, thereby enabling adaptive band partitioning and noise suppression.

In stage II, multiscale feature engineering is performed to obtain a unified input representation. Based on the decomposed modes, a 20-dimensional feature vector is constructed comprising time-domain statistical, trend, volatility, and momentum descriptors, together with scale-related features derived from the VMD components. After feature concatenation

and standardization, heterogeneous information across different magnitudes and frequency bands is mapped into a comparable and learnable feature space.

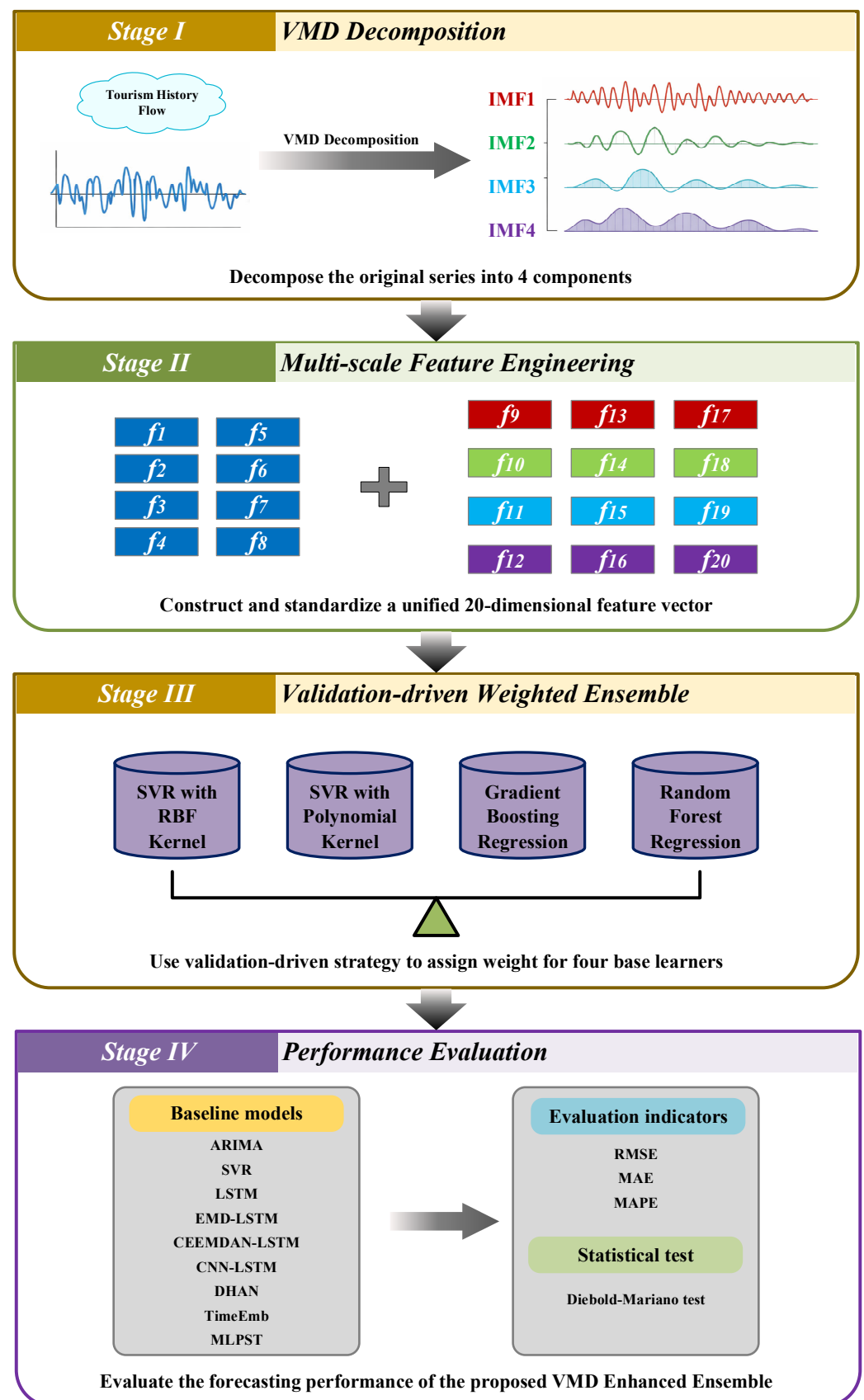


Figure 1. The framework of our VMD-enhanced ensemble method.

In stage III, heterogeneous base learners are trained to form a candidate set of forecasts, after which their errors are evaluated on a validation set. A validation-driven weighting strategy that highlights the best performing learner is then adopted, so that the final forecast is produced as a weighted linear combination of the base learner outputs, with stability and generalization being jointly considered. The framework is designed to balance predictive accuracy with computational practicality and interpretability, two considerations that are increasingly emphasized in the forecasting literature [55,56].

In stage IV, multi-step forecasting performance is assessed using RMSE, MAE, and MAPE, and multiple statistical testing procedures are further employed to examine the significance and robustness of the improvement. Meanwhile, six representative benchmark models are implemented, covering conventional time series methods, machine learning approaches, and deep learning paradigms, thereby enabling controlled comparative experiments.

3.2. VMD-Based Decomposition

To enable effective separation and characterization of components operating at different temporal scales in the frequency domain, variational mode decomposition is applied to the training set visitor flow series in an adaptive manner. The central idea of VMD is that the original signal is decomposed into $K = 4$ band limited modes [57], while the center frequency of each mode is adaptively estimated, so that each mode exhibits a finite bandwidth distribution around its own center frequency, and an adaptive frequency band partition is thus achieved [10,58]. Within the specific paradigm of tourism demand forecasting, the four intrinsic modes yielded by the decomposition demonstrate profound domain-specific interpretability. Specifically, they delineate the long-term trajectory that encapsulates the destination's underlying appeal and intrinsic evolutionary dynamics; effectively capture multi-scale seasonal regularities, including monthly cyclicity and intra-week periodicities; and isolate high-frequency exogenous shocks precipitated by policy realignments, localized holiday spikes, or random volatility. This physically meaningful deconstruction provides rigorous theoretical scaffolding for the formulation of the multi-scale augmented feature space.

In implementation, each intrinsic mode function is treated as an amplitude modulated and frequency modulated signal, which can be written as follows:

$$r_k(t) = m_k(t) \cos(\phi_k(t)) \quad (1)$$

where $m_k(t)$ denotes the instantaneous amplitude, and $\phi_k(t)$ is a nondecreasing phase function that corresponds to the instantaneous frequency. Building on this representation, an analytic signal is constructed, and frequency shifting is performed, such that the spectrum of each mode is translated to the baseband for bandwidth measurement, namely

$$[(\delta(t) + \frac{j}{\pi t}) * r_k(t)] e^{-j\omega_k t} \quad (2)$$

where $\delta(\cdot)$ is the Dirac delta function, j represents the imaginary unit, ω_k denotes the estimated center frequency of the k -th mode, and the symbol $*$ signifies the convolution operator.

Under the above formulation, decomposition is cast as a constrained variational optimization problem, in which the sum of the bandwidths of the demodulated baseband modes is minimized, subject to the requirement that the modes reconstruct the original signal, that is,

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \sum_{k=1}^K \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t} \right\|_2^2 \\ \text{s.t.} & \sum_{k=1}^K u_k(t) = y(t). \end{aligned} \quad (3)$$

Here, $\partial_t(\cdot)$ measures the smoothness of the baseband signal, which is equivalent to a bandwidth measure, and it therefore encourages each u_k to form a more compact band limited distribution in the frequency domain. $\|\cdot\|_2^2$ represents the squared L2 norm, which measures the smoothness of the shifted baseband signal and mathematically corresponds to the bandwidth of each individual decomposed mode in the frequency domain.

To solve the constrained problem, a quadratic penalty term and a Lagrange multiplier are introduced, so that an augmented Lagrangian formulation is obtained, where α is the penalty factor and $\lambda(t)$ is the Lagrange multiplier:

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| y(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), y(t) - \sum_{k=1}^K u_k(t) \right\rangle. \end{aligned} \quad (4)$$

The alternating direction method of multipliers is then used, where $\{u_k\}$, $\{\omega_k\}$, and λ are updated in an alternating manner, and the saddle point of the augmented Lagrangian is approached iteratively until a preset convergence criterion is met.

3.3. Multi-Scale Feature Engineering

While decomposition provides a multi-scale signal representation, the transformation of such information into model learnable inputs relies on systematic feature engineering. By following the principle of recent window characterization and multi-scale component alignment, an enhanced feature vector is constructed at the target time t , which integrates time domain statistical information with frequency domain scale information, so that the joint state of visitor flow across different temporal scales can be represented, and both fitting efficiency and generalization can be improved.

Let the original visitor flow series be $\{x_t\}_{t=1}^T$. For each target time t , a historical window of length $L = 10$ is defined as

$$W_t = [x_{t-L}, x_{t-L+1}, \dots, x_{t-1}] \quad (5)$$

which covers the recent observations prior to prediction, such that short-term dynamics are captured, while redundant noise that may be induced by an overly long window is mitigated. Based on W_t , eight time domain features $\{f_i(t)\}_{i=1}^8$ calculated directly from the raw, un-decomposed visitor volume series are constructed. These features are designed to jointly characterize the local scale, dispersion, boundary extremes, robust location, trend velocity, and structural momentum of tourist arrivals prior to signal decomposition. These features are defined in a unified form as follows:

$$\begin{aligned} f_1(t) &= \text{mean}(W_t), \\ f_2(t) &= \text{std}(W_t), \\ f_3(t) &= \max(W_t), \\ f_4(t) &= \min(W_t), \\ f_5(t) &= \text{median}(W_t), \\ f_6(t) &= \text{polyfit}(\{(i, x_{t-L+i})\}_{i=1}^L, 1)_{[0]}, \\ f_7(t) &= \text{std}(\Delta W_t), \\ f_8(t) &= x_{t-1} - x_{t-L}. \end{aligned} \quad (6)$$

Here, $f_6(t)$ is taken as the slope of a first-order polynomial fit over the window, so that recent trend movements are captured parsimoniously, whereas $f_7(t)$ and $f_8(t)$ quantify,

respectively, the intensity of local fluctuations and the net directional change within the same period.

In addition to these time domain descriptors, the remaining twelve features are designed to summarize the frequency scale information carried by the VMD modes $\{u_k(t)\}_{k=1}^K$ with $K = 4$. For each mode, an aligned window of equal length is constructed as $U_{k,t} = [u_k(t-L), u_k(t-L+1), \dots, u_k(t-1)]$. Three component-level statistics are then extracted, namely the component mean, component standard deviation, and the most recent component value, which together reflect the scale specific level, variability, and immediate state. These twelve features are defined jointly as:

$$\mu_k(t) = \text{mean}(U_{k,t}), \sigma_k(t) = \text{std}(U_{k,t}), \uparrow_k(t) = u_k(t-1), k = 1, \dots, 4 \quad (7)$$

$$\begin{aligned} f_9(t) &= \mu_1(t), f_{10}(t) = \sigma_1(t), f_{11}(t) = \uparrow_1(t), \\ f_{12}(t) &= \mu_2(t), f_{13}(t) = \sigma_2(t), f_{14}(t) = \uparrow_2(t), \\ f_{15}(t) &= \mu_3(t), f_{16}(t) = \sigma_3(t), f_{17}(t) = \uparrow_3(t), \\ f_{18}(t) &= \mu_4(t), f_{19}(t) = \sigma_4(t), f_{20}(t) = \uparrow_4(t). \end{aligned} \quad (8)$$

Finally, the above time domain and component features are concatenated to form a 20-dimensional enhanced feature vector, $\mathbf{F}_t = [f_1(t), f_2(t), \dots, f_{20}(t)]^\top$, after which Z score standardization is applied to each feature dimension to eliminate magnitude differences and to improve numerical stability:

$$\tilde{\mathbf{F}}_{t,j} = \frac{\mathbf{F}_{t,j} - \mu_j}{\sigma_j}, j = 1, \dots, 20 \quad (9)$$

where μ_j and σ_j denote the mean and standard deviation of the j th feature computed from the training data.

3.4. Base-Learner Training and Validation-Driven Weighted Ensemble

In forecasting tasks, a single learner may fail to maintain consistent advantages across attractions and fluctuation regimes. Accordingly, an ensemble framework is constructed to enhance robustness and generalization, where four complementary regression learners are trained in parallel and subsequently combined through a validation-driven weighting strategy. Specifically, support vector regression with an RBF kernel, polynomial kernel support vector regression, and gradient boosting regression were chosen to model nonlinear interactions and to improve robustness to noise through aggregation. Recent theoretical and empirical reviews of forecast combination strategies indicate that weighted ensembles with validation-driven weighting frequently outperform both single-model selection and equal-weight averaging across heterogeneous datasets [59,60].

Let $\tilde{\mathbf{F}}_i \in \mathbb{R}^{20}$ denote the standardized feature vector for sample i and let y_i be the corresponding ground truth. Each base learner $m \in \{1, 2, 3, 4\}$ is trained independently on the training set, and the learning process can be written in an empirical risk minimization form:

$$\theta_m^* = \underset{\theta_m}{\operatorname{argmin}} \sum_{i \in \mathcal{D}_{\text{train}}} \mathcal{L}(y_i, f_m(\tilde{\mathbf{F}}_i; \theta_m)) \quad (10)$$

where $\mathcal{L}(\cdot)$ is the loss function, and squared loss is adopted.

The validation set is used to evaluate generalization error without accessing the test set and to provide a basis for weight assignment. Predictions on the validation set are produced as:

$$\hat{y}_{m,i}^{\text{val}} = f_m(\tilde{\mathbf{F}}_i^{\text{val}}; \theta_m^*), m \in \{1, 2, 3, 4\}, i \in \mathcal{D}_{\text{val}} \quad (11)$$

The root mean squared error is then computed for each model:

$$\text{RMSE}_m = \sqrt{\frac{1}{n_{\text{val}}} \sum_{i \in \mathcal{D}_{\text{val}}} (y_i^{\text{val}} - \hat{y}_{m,i}^{\text{val}})^2} \quad (12)$$

where n_{val} is the size of the validation set. The best base learner is identified as:

$$m^* = \operatorname{argmin}_{m \in \{1,2,3,4\}} \text{RMSE}_m \quad (13)$$

To combine base learner outputs, a validation-driven weighting strategy that highlights the best is adopted. Let $M = 4$ denote the number of base learners, and let $\mathbf{w} = (w_1, \dots, w_M)$ denote the weight vector, which is defined as:

$$w_m = \begin{cases} 0.5, & m = m^*, \\ \frac{0.5}{M-1}, & m \neq m^*, \end{cases} \quad \text{s.t.} \quad \sum_{m=1}^M w_m = 1, w_m \geq 0 \quad (14)$$

Accordingly, for any test sample at time t , the ensemble prediction is computed as

$$\hat{y}_t = \sum_{m=1}^M w_m f_m(\tilde{\mathbf{F}}_t; \theta_m^*) = 0.5 f_{m^*}(\tilde{\mathbf{F}}_t; \theta_{m^*}^*) + \sum_{m \neq m^*} \frac{0.5}{M-1} f_m(\tilde{\mathbf{F}}_t; \theta_m^*) \quad (15)$$

By emphasizing the learner with the lowest validation error while retaining the stabilizing effect of combining multiple hypotheses, the proposed ensemble is designed to balance adaptivity and robustness under nonstationarity. The relatively compact look-back window and moderate feature dimensionality also help mitigate the overfitting risk that can accompany small-to-moderate training samples, consistent with the findings of Cerqueira et al. [61] on the interaction between dataset size and model complexity in time series forecasting.

3.5. Performance Evaluation and Statistical Tests

To evaluate the forecasting performance of the proposed VMD-enhanced ensemble under a strict and comparable experimental setting and to examine whether its improvements over benchmark methods are statistically significant, RMSE, MAE, MAPE, and R^2 are used to quantify predictive accuracy. Let the ground truth and predictions be $\{y_i\}_{i=1}^n$ and $\{\hat{y}_i\}_{i=1}^n$.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (16)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (17)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \quad (18)$$

For benchmarking, the proposed framework is compared with six representative models that cover traditional linear time series methods, classical machine learning methods, and deep learning or decomposition combined deep learning approaches, namely ARIMA, SVR, LSTM, EMD LSTM, CEEMDAN LSTM, CNN LSTM, DHAN, TimeEmb and MLPST. All benchmarks are implemented under the same data split and the same training, validation, and testing protocol, and final evaluation is conducted on the identical test set to ensure fair comparability.

To test whether the accuracy difference between two forecasting methods is significant, the Diebold–Mariano test [62] is employed. Let the benchmark and the proposed method

yield errors, and the loss differential series is constructed as $\{d\}$ and the test statistic is computed as:

$$DM = \frac{\bar{d}}{\sqrt{\text{Var}(d)/n}} \quad (19)$$

To further strengthen robustness, paired t tests and Wilcoxon signed rank tests are also reported so that both approximately normal and non-normal error difference settings are covered. In addition, Cohen's d effect size is used to quantify the practical magnitude of improvement:

$$d = \frac{\mu_b - \mu_p}{\sigma_{\text{pooled}}} \quad (20)$$

where μ_b and μ_p are the sample means of the benchmark and the proposed method under the same error metric, and σ_{pooled} is the pooled standard deviation. By jointly reporting error metrics, significance tests, and effect sizes, the effectiveness and advantages of the proposed framework are assessed in terms of both statistical meaning and practical relevance.

4. Results

4.1. Dataset Construction and Partitioning

This study utilized hourly visitor volume data retrieved from the official Beijing tourism portal, Visit Beijing (<https://www.visitbeijing.com.cn/>), focusing on two representative scenic spots: the Forbidden City and the Ming Tombs. These sites were specifically selected as case studies to establish a comparative experimental framework covering two distinct demand patterns: ultra-high volume with high volatility and moderate volume with stability. This dual-case approach facilitates a comprehensive assessment of the VMD-enhanced ensemble framework's generalizability and robustness when confronted with tourism demand of diverse scales and structural complexities. The dataset covers a period from 17 November 2022 to 6 September 2024, totaling 660 days, which include the post-pandemic recovery period to evaluate the model's ability to handle external shocks and capture multi-scale fluctuation regularities. The hourly records for each attraction were aggregated by calendar day to form daily visitor-volume series with 660 observations. The sample was split chronologically: the first 528 daily observations (17 November 2022 to 27 April 2024) were used for model training and validation, while the last 132 observations (28 April 2024 to 6 September 2024) were reserved as the held-out testing period. This testing period was not used for model estimation, VMD parameter setting, feature standardization, ensemble weighting, or hyperparameter selection. The data distribution is presented in Table 1 and Figure 2.

Table 1. Statistical description of the dataset.

Scenic Spots	Max	Std	Med	Skewness	Kurtosis	Min	Mean
Forbidden City	12.75	3.11	4.85	0.20	−0.34	0.00	4.86
Ming Tombs	6.15	0.77	1.45	1.97	5.16	0.25	1.45

Consistent with this literature, the empirical design is not restricted to choosing a more complex forecasting algorithm. The daily visitor volume is used as the target because scenic-area staffing and capacity decisions are typically made at daily frequency, while the raw hourly observations are first aggregated to reduce intraday reporting noise and to retain operationally interpretable demand patterns. The 10-day lookback window was selected to cover a full weekly cycle plus short holiday spillovers, balancing information sufficiency against the overfitting risk of longer windows in a 660-day sample. The feature vector

therefore combines domain-guided descriptors, including level, dispersion, extremes, trend slope, volatility, and momentum, with VMD-derived component summaries, so that model inputs reflect both researcher intuition about tourist behavior and data-driven multiscale structure [39,42,45–51].

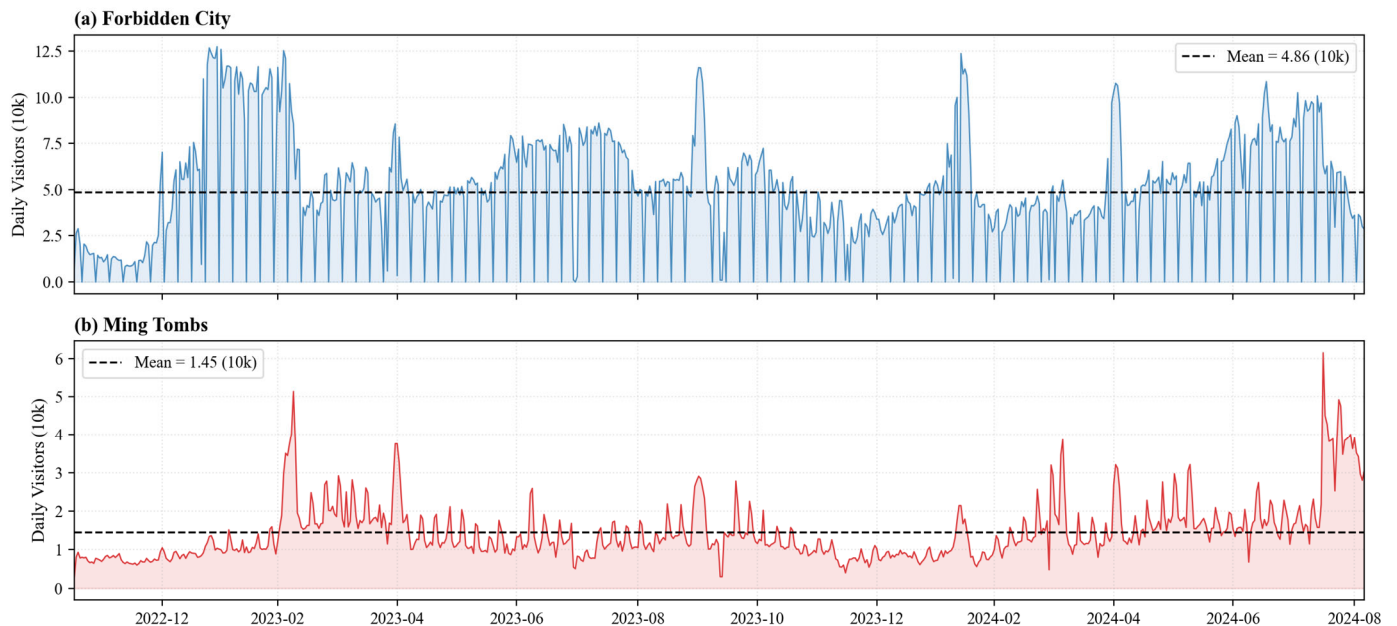


Figure 2. Daily visitor volume time series analysis.

Table 1 reveals that the Forbidden City has much higher and more volatile demand (mean = 4.86, std = 3.11) with a near-symmetric distribution (skewness = 0.20), while the Ming Tombs show lower, stabler demand (mean = 1.45, std = 0.77) with a right-skewed and heavy-tailed distribution (skewness = 1.97, kurtosis = 5.16). The zero minimum of the Forbidden City further highlights its exposure to extreme external disruptions, providing a stringent test for model robustness. These differences make the two datasets complementary for evaluating model robustness across heterogeneous tourism demand regimes.

4.2. Performance Comparison of Various Methods

To verify the performance of different prediction methods, this paper conducts a comprehensive comparison of ten prediction methods. Specifically, it includes nine benchmark models (ARIMA, SVR, LSTM, EMD-LSTM, CEEMDAN-LSTM, CNN-LSTM, DHAN, TimeEmb, MLPST) and the model proposed in this paper (VMD-enhanced ensemble). The performance of each method is compared by predicting the daily visitor volume of the Forbidden City and the Ming Tombs. RMSE, MAE, and MAPE are selected as evaluation indicators.

All models use consistent preprocessing and a fixed lookback window of 10 for fair comparison [63,64]. ARIMA is set to order (2,1,2) with rolling one-step prediction. SVR employs the RBF kernel with $C = 100$, $\gamma = \text{"scale"}$, and $\epsilon = 0.01$. LSTM has input size 1, hidden size 64, two layers, and dropout 0.2. EMD-LSTM and CEEMDAN-LSTM decompose signals via Python 3.8.13. The predictions of each IMF are summed to obtain the final forecast. CNN-LSTM uses 1D convolution, pooling, and an LSTM layer. DHAN applies dual attention for feature fusion. TimeEmb uses 16-dimensional positional encoding. MLPST adds normalized week, month, and year as extra features. For our VMD-enhanced ensemble, the VMD parameters are set as follows: $K = 4$, $\alpha = 2000$, $\tau = 0$, $DC = 0$, $\text{init} = 1$, $\text{tol} = 1 \times 10^{-7}$. We build a 20-dimensional feature vector and

use GradientBoostingRegressor with n estimators = 100, learning rate = 0.05, and max depth = 4.

The prediction results of different models for the daily visitor volume data of the Forbidden City and the Ming Tombs are shown in Table 2.

Table 2. Prediction results of tourist flow for each model.

Models	Forbidden City			Ming Tombs		
	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
ARIMA	2.665	1.914	19.812	0.587	0.353	16.089
SVR	1.819	1.315	23.665	0.938	0.684	32.801
LSTM	2.713	2.108	22.828	0.607	0.369	16.117
EMD-LSTM	2.076	1.792	26.641	0.605	0.378	16.51
CEEMDAN-LSTM	2.245	1.880	27.229	0.689	0.455	19.378
CNN-LSTM	1.089	0.853	14.649	0.659	0.418	17.898
DHAN	2.584	1.992	21.874	0.625	0.392	17.095
TimeEmb	1.462	1.158	19.707	0.583	0.350	15.889
MLPST	1.350	1.067	17.988	0.643	0.409	17.316
VMD-Enhanced Ensemble	0.939	0.687	11.428	0.465	0.284	11.351

As shown in Table 2, the proposed VMD-enhanced ensemble model achieves the best performance in predicting the daily visitor volume at the Forbidden City, with RMSE, MAE, and MAPE values of 0.939, 0.687, and 11.428%, respectively. Compared with the best-performing benchmark model CNN-LSTM (RMSE = 1.089, MAE = 0.853, MAPE = 14.649%), the VMD-enhanced ensemble model reduces RMSE, MAE, and MAPE by 13.77%, 19.46%, and 21.99%, respectively. For the daily visitor volume prediction at the Ming Tombs, the VMD-enhanced ensemble model still performs the best, with RMSE, MAE, and MAPE values of 0.465, 0.284, and 11.351%, respectively. Relative to the top-performing benchmark model TimeEmb (RMSE = 0.583, MAE = 0.350, MAPE = 15.889%), it reduces RMSE, MAE, and MAPE by 20.24%, 18.86%, and 28.56%, respectively.

These results can also be interpreted relative to prior real-time and competition-based tourism forecasting studies. Earlier prospective forecasting assessments and the COVID-19 tourism forecasting competition show that forecast performance deteriorates sharply when structural breaks, recovery dynamics, and data restrictions are not handled carefully and that no single class of models dominates across all destinations and horizons [52–54]. Compared with strong sequence-learning and decomposition-deep-learning benchmarks, the proposed framework reduces MAPE by 21.99% at the high-volatility Forbidden City and 28.56% at the more stable Ming Tombs. This comparison positions the findings alongside recent tourism studies that emphasize spatiotemporal information, mixed-frequency signals, and feature-varying model behavior as key sources of predictive improvement [36–40,42–46,54].

Based on the prediction results of the ten models for the visitor flow of the two scenic areas, the comparison of RMSE, MAE, and MAPE between VMD-enhanced ensemble and the nine benchmark models are shown in Figure 3.

For the Forbidden City, VMD-enhanced ensemble achieves an RMSE of 0.939, an MAE of 0.687, and a MAPE of 11.428%. In comparison, the benchmark models yield RMSE values ranging from 1.089 to 2.713, MAE values from 0.853 to 2.108, and MAPE values from 14.649% to 27.229%. For the Ming Tombs, the proposed model achieves an RMSE of 0.465, an MAE of 0.284, and a MAPE of 11.351%, while the benchmark models produce RMSE values between 0.583 and 0.938, MAE values between 0.350 and 0.684, and MAPE values between 15.889% and 32.801%.

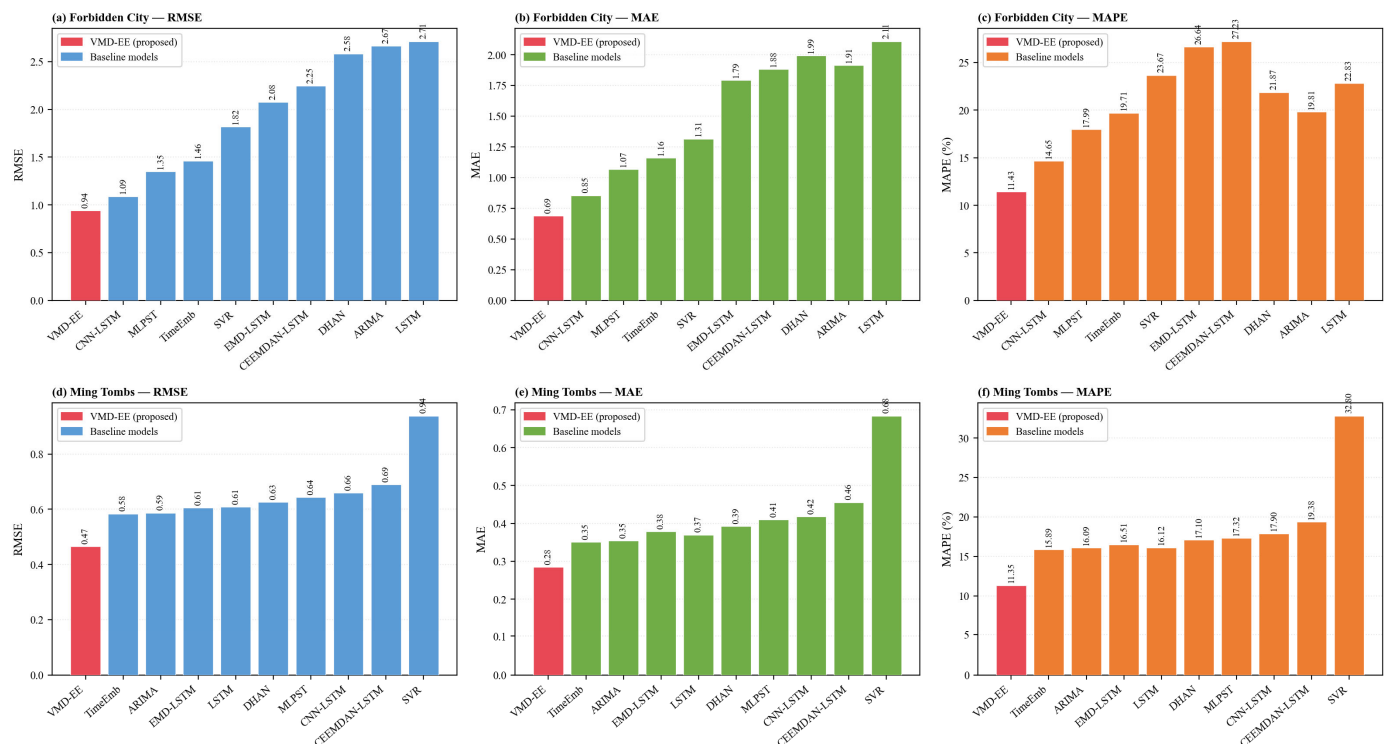


Figure 3. Comparison of prediction performance of ten models.

Overall, the VMD-enhanced ensemble model has the smallest prediction error across all evaluation metrics. Among the nine benchmark models, CNN-LSTM performs best on the Forbidden City with an RMSE of 1.089, an MAE of 0.853, and a MAPE of 14.649%, while TimeEmb performs best on the Ming Tombs with an RMSE of 0.583, an MAE of 0.350, and a MAPE of 15.889%. Compared with CNN-LSTM on the Forbidden City, the proposed model reduces RMSE, MAE, and MAPE by 13.77%, 19.46%, and 21.99%, respectively. Compared with TimeEmb on the Ming Tombs, it reduces RMSE, MAE, and MAPE by 20.24%, 18.86%, and 28.56%, respectively.

Compared with the classical methods SVR and LSTM on the Forbidden City, VMD-enhanced ensemble reduces RMSE by 48.38% and 65.39%, MAE by 47.76% and 67.41%, and MAPE by 51.71% and 49.94%, respectively. Compared with the decomposition methods EMD-LSTM and CEEMDAN-LSTM on the Forbidden City, it reduces RMSE by 54.77% and 58.17%, MAE by 61.66% and 63.46%, and MAPE by 57.11% and 58.04%, respectively.

The prediction curves of the VMD-enhanced ensemble, ARIMA, and the optimal benchmark models (CNN-LSTM for the Forbidden City and TimeEmb for the Ming Tombs) for the daily visitor volume of the two scenic areas were compared with the actual values. The results are shown in Figure 4. Figure 4 reports the 132 daily observations in the held-out testing period from 28 April 2024 to 6 September 2024.

According to Figure 4, in general, the predicted values of the VMD-enhanced ensemble model are the closest to the actual values, followed by CNN-LSTM for the Forbidden City and TimeEmb for the Ming Tombs. The predicted values of the ARIMA model differ the most from the actual values. Based on the above analysis, it can be concluded that the VMD-enhanced ensemble model has the best overall performance.

Figure 3. Predicted vs. Actual: VMD-EE and Selected Baselines

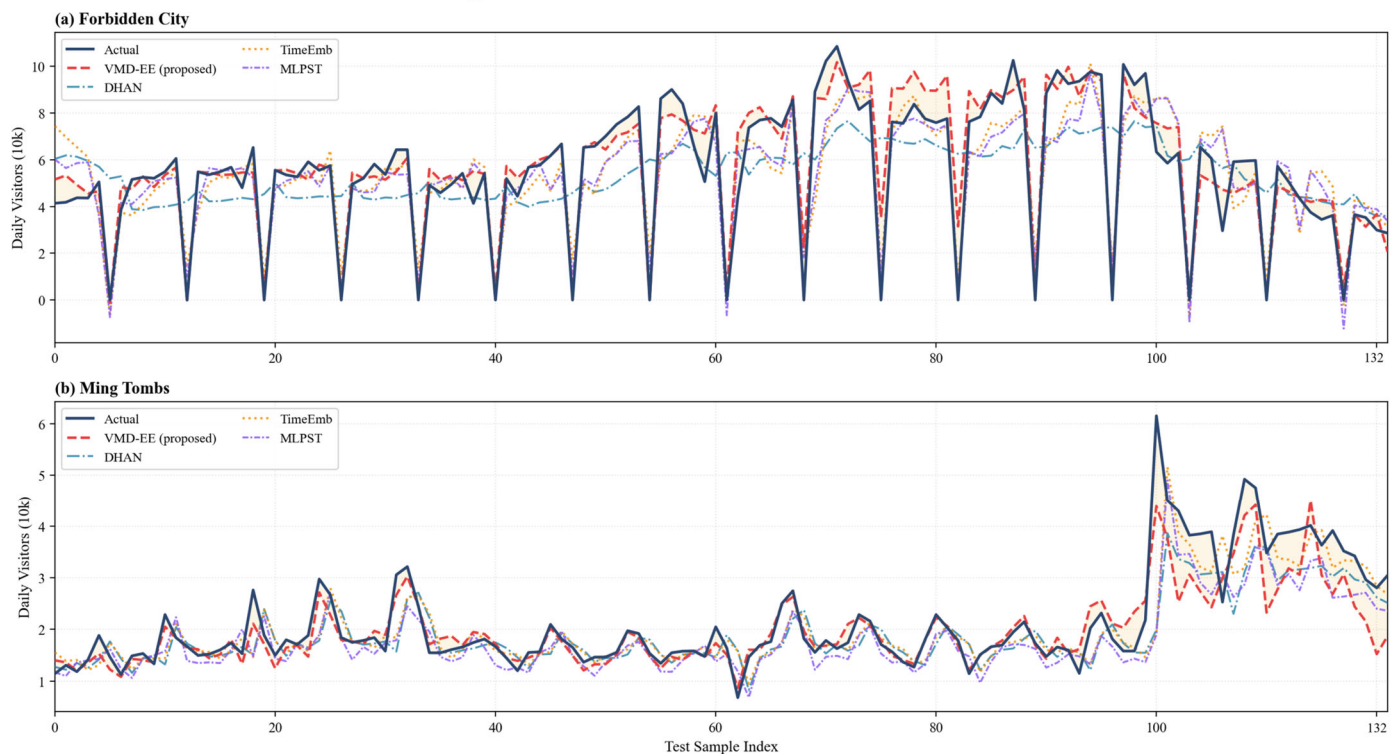


Figure 4. Comparison of predicted values and actual values.

4.3. Model Robustness Analysis

To statistically validate the effectiveness and robustness of the proposed VMD-enhanced ensemble model, the Diebold–Mariano (DM) test is employed. The DM test is a widely accepted statistical method. It detects significant differences in prediction accuracy between two forecasting models, especially for time series data. It accounts for autocorrelation in forecast errors and does not assume normally distributed errors. This makes it suitable for tourism demand data, which often exhibit heteroscedasticity and serial correlation. In this study, the DM test is conducted based on MAE as the loss function, and the significance level is set at 0.05. A p -value below 0.05 indicates that the proposed model significantly outperforms the baseline model at the 95% confidence level.

The DM test is performed pairwise between the proposed VMD-enhanced ensemble model and each of the nine benchmark models, separately for the Forbidden City and Ming Tombs datasets. The resulting p -values are presented in Table 3.

Table 3. Results of DM tests based on MAE.

Models	Forbidden City		Ming Tombs	
	p		p	
TimeEmb	0.000	***	0.0422	**
CNN-LSTM	0.0097	***	0.0002	***
MLPST	0.000	***	0.0003	***
SVR	0.000	***	0.000	***
EMD-LSTM	0.000	***	0.0081	***
CEEMDAN-LSTM	0.000	***	0.000	***
DHAN	0.000	***	0.0024	***
ARIMA	0.000	***	0.0382	**
LSTM	0.000	***	0.0154	**

Note: *** and ** represent significance levels of 1% and 5%, respectively.

As shown in Table 3, the proposed VMD-enhanced ensemble model achieves statistically significant improvements over all benchmark models at both scenic spots, with all p -values below 0.05. For the Forbidden City, eight out of nine benchmark models yield p -values of 0.000 ($p < 0.001$), and even the strongest benchmark CNN-LSTM shows a p -value of 0.0097 ($p < 0.01$). For the Ming Tombs, six benchmark models achieve p -values below 0.001, while the best-performing benchmark TimeEmb yields a p -value of 0.0422 ($p < 0.05$).

These results confirm that the proposed model is not only numerically superior (as shown in Table 2) but also statistically robust and generalizable across both ultra-high-volatility (Forbidden City) and stable (Ming Tombs) demand patterns. Overall, the VMD-enhanced ensemble model is a reliable and effective tool for tourism demand forecasting.

5. Discussion

The empirical findings also allow us to directly answer the research questions proposed in this study. First, regarding whether the proposed decomposition-feature-ensemble framework improves tourism demand forecasting accuracy, the results clearly show that VMD-enhanced ensemble achieves the lowest RMSE, MAE, and MAPE on both datasets. Second, regarding whether VMD-based multi-scale feature construction enhances predictive performance, the comparison with non-VMD and decomposition-based benchmark models indicates that integrating time-domain statistical features with VMD-derived component features provides a more informative representation of tourist-flow dynamics. Third, regarding whether the framework is robust across heterogeneous demand patterns, the proposed model performs consistently well on both the high-volume and highly volatile Forbidden City dataset and the lower-volume but more right-skewed Ming Tombs dataset. These findings confirm that the proposed framework is not only accurate but also robust and transferable across different tourism demand regimes.

Methodologically, the evaluation is an ex-post study based on historical data and assessed with a chronologically held-out testing period. All operations are confined to the training and validation windows before the testing period is evaluated, which prevents temporal data leakage. While VMD cannot anticipate entirely unprecedented exogenous shocks, a limitation requiring exogenous indicators to resolve, its core value lies in its structural isolation capacity. When a shock enters the historical lookback window, VMD segregates high-frequency volatility into a specific mode (IMF4), preventing raw noise from contaminating the stable, low-frequency trend components (IMF1). Because the training period from November 2022 through mid-2024 contains diverse post-pandemic shifts, holiday surges, and weather anomalies, their signatures are successfully captured by IMF4 and time domain descriptors. The ensemble model can thus condition its inference on these recent shock patterns, preserving high predictive precision under highly nonstationary regimes.

6. Conclusions

Accurate prediction of tourism demand is of vital importance for optimizing tourism management and resource allocation. The main objective is to explore whether the “decomposition-feature-ensemble” three-stage framework can improve prediction accuracy. This is tested on nonlinear and non-stationary tourist flow data. For this purpose, this study proposes an enhanced ensemble prediction model based on VMD, namely VMD-enhanced ensemble, and conducts a comprehensive comparison between it and nine benchmark models (ARIMA, SVR, LSTM, EMD-LSTM, CEEMDAN-LSTM, CNN-LSTM, DHAN, TimeEmb, MLPST). To effectively verify the accuracy and robustness of the method, predictions were made for the daily visitor volume at the Forbidden City and the Ming Tombs. The study found that the framework of “decomposition-feature-ensemble” can fully extract the

features of tourist flow data, thereby significantly improving the prediction accuracy and is an effective tool for achieving accurate tourist flow forecasting.

From a theoretical perspective, this study redefines the role of VMD in time series forecasting. It positions VMD as a structured information extraction mechanism. This mechanism provides forecasting models with feature representations that have multi-scale interpretability. This is empirically supported by the comparative results on two distinct demand patterns: the ultra-high-volume and highly volatile Forbidden City dataset, and the moderate-volume and stable Ming Tombs dataset. The results show that integrating frequency-domain information (from VMD) with time-domain statistical features enables better characterization of trends, volatility, and dynamic variations. This integration is further exploited through ensemble learning (GradientBoostingRegressor). Specifically, the proposed VMD-enhanced ensemble model achieves RMSE values of 0.939 (Forbidden City) and 0.465 (Ming Tombs), outperforming all nine benchmark models, including CNN-LSTM and TimeEmb. The DM test further confirms that these improvements are statistically significant, with all p -values below 0.05. This finding extends the theoretical application of decomposition methods in the forecasting literature and offers a more robust, interpretable, and methodologically transferable paradigm for forecasting non-stationary time series.

In terms of practical implications, the empirical results indicate that the proposed forecasting framework exhibits clear advantages in demand forecasting scenarios characterized by high volatility and uncertainty, achieving improved predictive accuracy while maintaining forecast stability. As shown in Table 2, the proposed model reduces MAPE from 14.649% (CNN-LSTM) to 11.428% on the Forbidden City and from 15.889% (TimeEmb) to 11.351% on the Ming Tombs, representing relative error reductions of 21.99% and 28.56%, respectively. This is particularly important for real-world applications that rely on forecasts to support resource allocation and operational decision-making. In the context of tourist flow forecasting, the framework enables managers to better capture demand dynamics, thereby supporting decisions related to staffing, capacity management, and service optimization. Moreover, owing to its structurally generalizable design, the applicability of the framework is not confined to tourism but can be extended to complex systems such as energy load forecasting, financial time series analysis, and climate-related forecasting, providing a solution that balances accuracy, robustness, and scalability across multiple domains. The successful validation on two heterogeneous datasets further demonstrates the framework's strong generalizability across different demand regimes.

Although the proposed framework has achieved excellent performance in terms of prediction, it still has some limitations that point to future research directions. Firstly, the representativeness of the data samples is limited. Using only data from two scenic areas may not cover the diverse fluctuation patterns of tourist flows across different regions, different types, and different scales of scenic spots, meaning that the generalizability of the conclusions requires more empirical support. Secondly, other commonly used predictive techniques such as advanced econometrics, alternative machine learning models, and deep learning architectures can also be explored to better capture the complex non-linear patterns in tourist flow data. Finally, future studies will conduct systematic and comprehensive comparisons with previous research frameworks, alternative forecasting technologies, and specific technical parameters to further validate and contextualize the methodological superiority of our approach. In addition, the empirical design is retrospective: the holdout period is evaluated after the historical series has been collected, and the framework should not be interpreted as predicting unprecedented shocks before they occur. Future research should incorporate exogenous leading indicators, such as policy announcements, weather warnings, mobility data, search intensity, and event calendars, to support genuinely forward-looking shock-sensitive forecasting.

Author Contributions: Conceptualization, C.Y. and P.D.; methodology, C.Y.; software, C.Y. and E.Z.; validation, E.Z. and W.S.; data curation, E.Z.; writing—original draft preparation, C.Y. and W.S.; writing—review and editing, C.Y., W.S., and P.D.; visualization, C.Y. and W.S.; supervision, S.W. and P.D.; project administration, P.D.; funding acquisition, S.W. and P.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research work was supported by the National Natural Science Foundation of China under Grants No. 72201201 and No. 71988101 and by the Fundamental Research Funds for the Central Universities under Grant No. QTZX23085, Shaanxi Province Natural Science Basic Research Program No. 2025JC-YBQN-1003.

Data Availability Statement: The data for this study can be obtained from the official Beijing tourism portal, Visit Beijing (<https://www.visitbeijing.com.cn/>).

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Zhang, Y.; Li, G.; Muskat, B.; Law, R.; Yang, Y. Group Pooling for Deep Tourism Demand Forecasting. *Ann. Tour. Res.* **2020**, *82*, 102899. [\[CrossRef\]](#)
2. Li, X.; Law, R.; Xie, G.; Wang, S. Review of Tourism Forecasting Research with Internet Data. *Tour. Manag.* **2021**, *83*, 104245. [\[CrossRef\]](#)
3. Zhang, C.; Li, M.; Sun, S.; Tang, L.; Wang, S. Decomposition Methods for Tourism Demand Forecasting: A Comparative Study. *J. Travel Res.* **2021**, *61*, 1682–1699. [\[CrossRef\]](#)
4. Xie, G.; Qian, Y.; Wang, S. A Decomposition-Ensemble Approach for Tourism Forecasting. *Ann. Tour. Res.* **2020**, *81*, 102891. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Goh, C.; Law, R. Modeling and Forecasting Tourism Demand for Arrivals with Stochastic Nonstationary Seasonality and Intervention. *Tour. Manag.* **2002**, *23*, 499–510. [\[CrossRef\]](#)
6. Song, H.; Qiu, R.T.R.; Park, J. A Review of Research on Tourism Demand Forecasting: Launching the Annals of Tourism Research Curated Collection on Tourism Demand Forecasting. *Ann. Tour. Res.* **2019**, *75*, 338–362. [\[CrossRef\]](#)
7. Zhang, Y.; Li, G.; Muskat, B.; Law, R. Tourism Demand Forecasting: A Decomposed Deep Learning Approach. *J. Travel Res.* **2020**, *60*, 981–997. [\[CrossRef\]](#)
8. Chen, C.-F.; Lai, M.-C.; Yeh, C.-C. Forecasting Tourism Demand Based on Empirical Mode Decomposition and Neural Network. *Knowl.-Based Syst.* **2012**, *26*, 281–287. [\[CrossRef\]](#)
9. Xie, G.; Xie, R.; Li, X.; Wang, S. A Secondary Decomposition-Ensemble Approach for Interval Forecasting of Tourism Demand. *Syst. Eng.—Theory Pract.* **2025**, *45*, 2784–2797. [\[CrossRef\]](#)
10. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
11. Ahajjam, M.A.; Bonilla Licea, D.; Ghogho, M.; Kobbane, A. Experimental Investigation of Variational Mode Decomposition and Deep Learning for Short-Term Multi-Horizon Residential Electric Load Forecasting. *Appl. Energy* **2022**, *326*, 119963. [\[CrossRef\]](#)
12. Niu, H.; Xu, K. A Hybrid Model Combining Variational Mode Decomposition and an Attention-GRU Network for Stock Price Index Forecasting. *Math. Biosci. Eng.* **2020**, *17*, 7151–7166. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Yu, L.; Wang, Z.; Tang, L. A Decomposition-Ensemble Model with Data-Characteristic-Driven Reconstruction for Crude Oil Price Forecasting. *Appl. Energy* **2015**, *156*, 251–267. [\[CrossRef\]](#)
14. Law, R.; Li, G.; Fong, D.K.C.; Han, X. Tourism Demand Forecasting: A Deep Learning Approach. *Ann. Tour. Res.* **2019**, *75*, 410–423. [\[CrossRef\]](#)
15. Nguyen, D.T.; Li, Y.; Peng, C.L.; Cho, M.-Y.; Nguyen, T.-P. Monthly Tourism Demand Forecasting with COVID-19 Impact-Based Hybrid Convolution Neural Network and Gate Recurrent Unit. *Int. J. Tour. Res.* **2024**, *26*, e2812. [\[CrossRef\]](#)
16. Bi, J.-W.; Han, T.-Y.; Yao, Y.; Yang, T. Tourism Demand Forecasting under Conceptual Drift during COVID-19: An Ensemble Deep Learning Model. *Curr. Issues Tour.* **2024**, *27*, 4084–4103. [\[CrossRef\]](#)
17. Santoro, D.; Ciano, T.; Ferrara, M. A Comparison between Machine and Deep Learning Models on High Stationarity Data. *Sci. Rep.* **2024**, *14*, 19409. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Song, H.; Li, G. Tourism Demand Modelling and Forecasting—A Review of Recent Research. *Tour. Manag.* **2008**, *29*, 203–220. [\[CrossRef\]](#)
19. Li, X.; Pan, B.; Law, R.; Huang, X. Forecasting Tourism Demand with Composite Search Index. *Tour. Manag.* **2017**, *59*, 57–66. [\[CrossRef\]](#)

20. Li, X.; Zhang, Y.; Zhu, M.; Wang, S. Forecasting tourism demand with a novel robust decomposition and ensemble framework. *Expert Syst. Appl.* **2024**, *236*, 121388. [[CrossRef](#)]
21. Qin, L.; Li, W. A Combination Approach Based on Seasonal Adjustment Method and Echo State Network for Energy Consumption Forecasting in USA. *Energy Effic.* **2020**, *13*, 1505–1524. [[CrossRef](#)]
22. Li, X.; Law, R. Forecasting Tourism Demand with Decomposed Search Cycles. *J. Travel Res.* **2020**, *59*, 52–68. [[CrossRef](#)]
23. Bosupeng, M. Forecasting Tourism Demand: The Hamilton Filter. *Ann. Tour. Res.* **2019**, *79*, 102823. [[CrossRef](#)]
24. Hassani, H.; Webster, A.; Silva, E.S.; Heravi, S. Forecasting U.S. Tourist Arrivals Using Optimal Singular Spectrum Analysis. *Tour. Manag.* **2015**, *46*, 322–335. [[CrossRef](#)]
25. Silva, E.S.; Hassani, H.; Heravi, S.; Huang, X. Forecasting Tourism Demand with Denoised Neural Networks. *Ann. Tour. Res.* **2019**, *74*, 134–154. [[CrossRef](#)]
26. Kummong, R.; Supratid, S. Thailand Tourism Forecasting Based on a Hybrid of Discrete Wavelet Decomposition and NARX Neural Network. *Ind. Manag. Data Syst.* **2016**, *116*, 1242–1258. [[CrossRef](#)]
27. Hassani, H.; Silva, E.S.; Antonakakis, N.; Filis, G.; Gupta, R. Forecasting Accuracy Evaluation of Tourist Arrivals. *Ann. Tour. Res.* **2017**, *63*, 112–127. [[CrossRef](#)]
28. Chen, J.; Li, G.; Wu, D.; Shen, S. Forecasting Seasonal Tourism Demand Using a Multiseries Structural Time Series Method. *J. Travel Res.* **2017**, *58*, 92–103. [[CrossRef](#)]
29. Yi, S.; Chen, X.; Tang, C. Time Series Transformer for Tourism Demand Forecasting. *Sci. Rep.* **2025**, *15*, 29565. [[CrossRef](#)] [[PubMed](#)]
30. Luo, Q.; Cai, S.; Lv, N.; Fu, X. Daily Forecasting of Tourism Demand: An ST-LSTM Model with Social Network Service Co-Occurrence Similarity. *Inf. Manag.* **2025**, *62*, 104056. [[CrossRef](#)]
31. Sun, S.; Wei, Y.; Tsui, K.-L.; Wang, S. Forecasting Tourist Arrivals with Machine Learning and Internet Search Index. *Tour. Manag.* **2019**, *70*, 1–10. [[CrossRef](#)]
32. Kulshrestha, A.; Krishnaswamy, V.; Sharma, M. Bayesian BILSTM Approach for Tourism Demand Forecasting. *Ann. Tour. Res.* **2020**, *83*, 102925. [[CrossRef](#)]
33. Bi, J.-W.; Li, H.; Fan, Z.-P. Tourism Demand Forecasting with Time Series Imaging: A Deep Learning Model. *Ann. Tour. Res.* **2021**, *90*, 103255. [[CrossRef](#)]
34. Li, X.; Xu, Y.; Law, R.; Wang, S. Enhancing Tourism Demand Forecasting with a Transformer-Based Framework. *Ann. Tour. Res.* **2024**, *107*, 103791. [[CrossRef](#)]
35. Hu, M.; Li, M.; Chen, Y.; Liu, H. Tourism Forecasting by Mixed-Frequency Machine Learning. *Tour. Manag.* **2025**, *106*, 105004. [[CrossRef](#)]
36. Li, H.; Hu, M.; Li, G. Forecasting tourism demand with multisource big data. *Ann. Tour. Res.* **2020**, *83*, 102912. [[CrossRef](#)]
37. Hu, M.; Dong, N.; Hu, F. Tourism demand forecasting using short video information. *Ann. Tour. Res.* **2024**, *109*, 103838. [[CrossRef](#)]
38. Park, E.; Park, J.; Hu, M. Tourism demand forecasting with online news data mining. *Ann. Tour. Res.* **2021**, *90*, 103273. [[CrossRef](#)]
39. Wen, L.; Liu, C.; Song, H.; Liu, H. Forecasting Tourism Demand with an Improved Mixed Data Sampling Model. *J. Travel Res.* **2021**, *60*, 336–353. [[CrossRef](#)]
40. Cao, W.; Hu, M.; Wei, S.; Yang, L. Denoising search query improves tourism forecasting. *Ann. Tour. Res.* **2025**, *115*, 104049. [[CrossRef](#)]
41. Kohns, D.; Bhattacharjee, A. Nowcasting growth using Google Trends data: A Bayesian Structural Time Series model. *Int. J. Forecast.* **2023**, *39*, 1384–1412. [[CrossRef](#)]
42. Li, C.; Zheng, W.; Ge, P. Tourism demand forecasting with spatiotemporal features. *Ann. Tour. Res.* **2022**, *94*, 103384. [[CrossRef](#)]
43. Jiao, X.; Li, G.; Chen, J.L. Forecasting international tourism demand: A local spatiotemporal model. *Ann. Tour. Res.* **2020**, *83*, 102937. [[CrossRef](#)]
44. Jiao, X.; Chen, J.L.; Li, G. Forecasting tourism demand: Developing a general nesting spatiotemporal model. *Ann. Tour. Res.* **2021**, *90*, 103277. [[CrossRef](#)]
45. Zheng, W.; Huang, L.; Lin, Z. Multi-attraction, hourly tourism demand forecasting. *Ann. Tour. Res.* **2021**, *90*, 103271. [[CrossRef](#)]
46. Gao, H.; Li, H.; Zhang, C.J. Time and feature varying tourism demand forecasting. *Ann. Tour. Res.* **2025**, *112*, 103959. [[CrossRef](#)]
47. Vatsa, P. Seasonality and cycles in tourism demand-redux. *Ann. Tour. Res.* **2021**, *90*, 103105. [[CrossRef](#)]
48. Montero-Manso, P.; Athanasopoulos, G.; Hyndman, R.J.; Talagala, T.S. FFORMA: Feature-based forecast model averaging. *Int. J. Forecast.* **2020**, *36*, 86–92. [[CrossRef](#)]
49. Talagala, T.S.; Li, F.; Kang, Y. FFORMPP: Feature-based forecast model performance prediction. *Int. J. Forecast.* **2022**, *38*, 920–943. [[CrossRef](#)]
50. Lim, B.; Arik, S.O.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [[CrossRef](#)]
51. Li, L.; Kang, Y.; Li, F. Bayesian forecast combination using time-varying features. *Int. J. Forecast.* **2023**, *39*, 1287–1302. [[CrossRef](#)]
52. Song, H.; Li, G.; Cai, Y. Tourism forecasting competition in the time of COVID-19: An assessment of ex ante forecasts. *Ann. Tour. Res.* **2022**, *96*, 103445. [[CrossRef](#)] [[PubMed](#)]

53. Liu, A.; Lin, V.S.; Li, G.; Song, H. Ex Ante Tourism Forecasting Assessment. *J. Travel Res.* **2022**, *61*, 64–75. [[CrossRef](#)]
54. Zhang, H.; Song, H.; Wen, L.; Liu, C. Forecasting tourism recovery amid COVID-19. *Ann. Tour. Res.* **2021**, *87*, 103149. [[CrossRef](#)] [[PubMed](#)]
55. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. M5 accuracy competition: Results, findings, and conclusions. *Int. J. Forecast.* **2022**, *38*, 1346–1364. [[CrossRef](#)]
56. Hasson, H.; Maddix, D.C.; Wang, B.; Gupta, G.; Park, Y. Theoretical Guarantees of Learning Ensembling Strategies with Applications to Time Series Forecasting. *Proc. Mach. Learn. Res.* **2023**, *202*, 12616–12632.
57. Yuan, L. Mine Gas Time-Series Data Prediction and Fluctuation Monitoring Method Based on Decomposition-Enhanced Cross-Graph Forecasting and Anomaly Finding. *Sensors* **2025**, *25*, 7014. [[CrossRef](#)] [[PubMed](#)]
58. Li, F.; Li, R.; Tian, L.; Chen, L.; Liu, J. Data-Driven Time-Frequency Analysis Method Based on Variational Mode Decomposition and Its Application to Gear Fault Diagnosis in Variable Working Conditions. *Mech. Syst. Signal Process.* **2019**, *116*, 462–479. [[CrossRef](#)]
59. Wang, X.; Hyndman, R.J.; Li, F.; Kang, Y. Forecast combinations: An over-50-year review. *Int. J. Forecast.* **2023**, *39*, 1518–1547. [[CrossRef](#)]
60. Timmermann, A. Forecast combinations. In *Handbook of Economic Forecasting*; Elliott, G., Granger, C.W.J., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; Volume 1, pp. 135–196. [[CrossRef](#)]
61. Cerqueira, V.; Torgo, L.; Soares, C. A case study comparing machine learning with statistical methods for time series forecasting: Size matters. *J. Intell. Inf. Syst.* **2022**, *59*, 415–433. [[CrossRef](#)]
62. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [[CrossRef](#)]
63. Samieiyan, B.; Awasthi, A. A Custom Transformer-Based Framework for Joint Traffic Flow and Speed Prediction in Autonomous Driving Contexts. *Future Transp.* **2026**, *6*, 15. [[CrossRef](#)]
64. Gligorić, Z.; Gligorić, M.; Halilović, D.; Beljić, Č.; Urošević, K. Hybrid Stochastic-Grey Model to Forecast the Behavior of Metal Price in the Mining Industry. *Sustainability* **2020**, *12*, 6533. [[CrossRef](#)]

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