applied sciences

MDPI

Article

A Hybrid LSTM Framework for Short-Term Regional Wind Speed Forecasting Based on PCA and SSA-Optimized VMD

Huachen Li, Zhengzheng Ma *, Liang Chen, Qinglin Zhu, Xiang Dong, Bin Xu, Yuanming Li and Mantong Zhang

China Research Institute of Radiowave Propagation, Qingdao 266107, China; 15891047332@163.com (H.L.); 13505210936@163.com (Y.L.); 13628066758@163.com (M.Z.)
* Correspondence: m_z_z_@126.com

Abstract

Accurate regional wind speed forecasting is critical yet challenging due to inherent spatiotemporal correlations and data non-stationarity. This paper proposes a hybrid framework combining Principal Component Analysis (PCA), Variational Mode Decomposition (VMD), and Long Short-Term Memory (LSTM) networks. First, PCA extracts dominant spatial features from a regional wind field (9×9 grid), retaining 99.5% of the information to reduce redundancy. Next, an adaptive VMD strategy, optimized by the Sparrow Search Algorithm (SSA), decomposes these components to mitigate temporal non-stationarity. High-correlation sub-signals are then fed into the LSTM predictor. Experimental results demonstrate that the framework achieves an average coefficient of determination (R^2) of approximately 0.41 in the first forecasting step. Crucially, it significantly mitigates error accumulation in multi-step forecasting, maintaining a stable R^2 of 0.39 in the third step. Conversely, complex spatiotemporal models like ConvLSTM achieve high initial accuracy but suffer severe degradation (R^2 dropping from 0.70 to 0.24) alongside significantly higher computational overhead. The proposed strategy effectively prevents overfitting to high-frequency noise, ensuring a computationally efficient and robust solution for multi-step regional wind forecasting.

Keywords: regional wind speed forecasting; multi-step forecasting; spatiotemporal feature extraction; Variational Mode Decomposition (VMD); Principal Component Analysis (PCA); Sparrow Search Algorithm (SSA); Long Short-Term Memory (LSTM)

1. Introduction

With the increasing environmental pressure from fossil fuels and the proposal of global carbon reduction goals, the global energy system is undergoing a profound transition [1]. Among renewable energy sources, wind energy has emerged as a pivotal force driving this transition, owing to its abundance, renewability, and technological maturity [2].

However, the inherent intermittency and volatility of wind energy pose severe challenges to grid stability and energy dispatch. As a critical component of wind power exploitation, accurate wind field forecasting provides essential support for grid dispatching, market trading, and energy storage optimization. Precise forecasting of wind speed and power helps reduce wind curtailment rates, minimize equipment wear, and enhance the penetration of wind power in power systems. Particularly at the regional scale, wind field forecasting not only reflects the spatial distribution of wind resources but also provides a scientific basis for cross-regional grid scheduling. Therefore, achieving high-precision regional wind forecasting is key to addressing these challenges.

Check for updates

Academic Editors: Nir Krakauer and Alessandro Lo Schiavo

Received: 20 March 2026
Revised: 11 April 2026
Accepted: 23 April 2026
Published: 26 April 2026

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

The methodologies for wind speed forecasting have evolved from physical models to statistical models, and recently to deep learning approaches. Early research relied primarily on physical models, such as Numerical Weather Prediction (NWP) [3,4], which simulate atmospheric dynamics. While effective at large scales, these methods suffer from high computational costs and sensitivity to initial conditions. Subsequently, statistical models (e.g., ARIMA [5,6], Kalman Filtering [7], SVM [8,9]) were developed to capture temporal patterns from historical data. Despite their computational efficiency, they often struggle with the non-linearity of complex weather conditions. In recent years, deep learning methods [10,11]—such as CNNs [12–14], LSTMs [15–17], and GNNs [18–20]—have become a research hotspot. These models demonstrate significant advantages in extracting multi-dimensional features and capturing spatiotemporal dependencies.

Despite these advancements, applying deep learning directly to raw wind data remains problematic. First, raw wind signals contain significant high-frequency random noise and non-stationary components. Direct prediction often leads neural networks to overfit this noise, resulting in poor generalization and robustness. Second, traditional point-wise forecasting methods ignore the strong spatial correlations within the wind field and incur prohibitive computational costs when applied to regional-scale grids, lacking the capability for efficient parallel prediction.

Effectively balancing spatiotemporal feature extraction with computational efficiency remains a non-trivial challenge, even with the rapid proliferation of deep learning architectures in this field. Many existing hybrid models still depend on empirically chosen parameters for signal decomposition, which constrains their adaptability to the highly non-stationary characteristics of regional wind fields across diverse geographical scales.

The physical rationale for adopting a 'decomposition-first' strategy is deeply rooted in atmospheric dynamics. Classic meteorological studies on the atmospheric kinetic energy spectrum, notably the seminal work by Van der Hoven (1957) [21], have established that wind energy is not uniformly distributed across all scales. Instead, it concentrates in distinct frequency bands, primarily the synoptic scale and the micro-scale. Crucially, a distinct spectral gap exists between these two peaks. This physical phenomenon implies that regional wind signals can be decoupled into deterministic macro-trends and random micro-fluctuations. By employing signal decomposition techniques like VMD, we mathematically operationalize this physical separation, preventing the deep learning models from blindly fitting the chaotic turbulence within the spectral gap.

Specifically, direct modeling of a 9×9 regional grid (81 variables) introduces significant spatial redundancy and 'dimensionality curse', which not only increases the risk of overfitting but also incurs prohibitive computational costs. While some decomposition-based models exist, they often ignore the critical step of spatial feature compression or fail to adaptively optimize the decomposition process for varying spatial components. This leaves a research gap in developing a framework that is both spatially compact and temporally adaptive.

To bridge these gaps, this study proposes an integrated PCA–VMD–SSA–LSTM framework. Unlike traditional end-to-end models, we adopt a 'decomposition-first' strategy grounded in the Van der Hoven atmospheric kinetic energy spectrum. The main contributions of this manuscript are explicitly summarized as follows:

1. Spatial-Temporal Decoupling: We introduce PCA to compress the 9×9 regional wind field into a low-dimensional subspace (51 PCs), effectively eliminating spatial redundancy while retaining 99.5% of the physical information.
2. Adaptive Signal Refinement: An SSA-optimized VMD strategy is developed to adaptively decompose each principal component, ensuring that deterministic trends are separated from stochastic noise without the bias of manual parameter tuning.

3. Robust Multi-step Performance: Through comprehensive evaluations, we demonstrate that the proposed framework achieves exceptional stable multi-step capability (maintaining $R^2 \approx 0.39$ in step 3), significantly outperforming complex spatiotemporal baselines like ConvLSTM in long-term stability and computational efficiency.

2. Materials and Methods

2.1. Data Description and Preprocessing

The wind field data utilized in this study were obtained from the ERA5 reanalysis dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF, Reading, UK). ERA5 provides hourly estimates of a large number of atmospheric, land, and oceanic climate variables and is widely recognized for its high spatial and temporal resolution. For the purpose of regional wind speed forecasting, a specific geo-graphical area was selected, spanning from 35° N to 37° N latitude and 119° W to 121° W longitude. The data covers the period from 1 January 2025 to 30 June 2025, with an hourly temporal resolution. Given the ERA5 spatial resolution of 0.25° × 0.25°, the selected region corresponds to a grid of 9 × 9 spatial points, resulting in a total of 81 observation stations. This grid structure serves as the basis for the regional spatiotemporal analysis.

The raw data includes both the U-component (zonal) and V-component (meridional) of the wind. To obtain the scalar wind speed (v), the following conversion was applied:

$$v = \sqrt{U^2 + V^2} \tag{1}$$

where U and V represent the horizontal wind components. In this study, a continuous subset of 1000 time steps was extracted from the dataset to verify the effectiveness of the proposed model. The final dataset is represented as a 3D tensor with dimensions ($T \times H \times W$), where T = 1000, H = 9, and W = 9.

To ensure the numerical stability of the deep learning model while strictly preserving the physical characteristics and multiscale energy distribution of the wind field, the following preprocessing steps were implemented:

Unlike standard Z-score standardization (which normalizes each feature dimension independently to zero mean and unit variance) or Min-Max normalization (which relies on specific training set statistics and may cause "out-of-bounds" issues), this study employs a fixed global scaling strategy. Considering the physical limits of wind speed in the target region, a constant scaling factor of 50.0 was applied. The normalized wind speed x'_t is calculated as

$$x'_t = \frac{x_t}{50.0} \tag{2}$$

The rationale for avoiding independent Z-score standardization lies in the core design of the proposed "decomposition-first" framework. Both PCA and VMD decompose the original wind field into multiscale sub-signals (i.e., PCs and IMFs) with explicit physical meanings and intrinsic energy hierarchies. Typically, low-frequency trend components carry large amplitudes, whereas high-frequency stochastic modes possess minimal variance. Applying standard Z-score normalization independently to each decomposed channel would artificially equalize their variances, thereby severely amplifying the high-frequency turbulence and distorting the relative energy contributions. By utilizing a fixed global scaling factor, the proposed method ensures that all input values fall within a range conducive to neural network training (typically [0, 1] for wind speeds under 50 m/s), while inherently preserving the relative amplitude ratios across all multiscale components. This preservation is crucial for the downstream LSTM network to effectively distinguish the main deterministic trends from local stochastic fluctuations.

Sliding Window Sequence Generation: To construct the supervised learning samples for the LSTM network, a sliding window approach was applied. Let L_{in} be the input sequence length and L_{out} be the prediction horizon. The dataset was transformed into pairs of inputs $\boldsymbol{X} \in \mathbb{R}^{B \times L_{in} \times N}$ and targets $\boldsymbol{Y} \in \mathbb{R}^{B \times L_{out} \times N}$, where N = 81 (flattened spatial features from PCA components in the methodology) and B represents the batch size of the neural network input.

Input Sequence Length (L_{in}): Set to 72, corresponding to the historical wind information of the past 3 days (72 h).

Prediction Horizon (L_{out}): Set to 3, aiming to predict the wind speed for the next 3 h.

Dataset Partitioning: To ensure rigorous model evaluation and prevent data leakage, the dataset partitioning was strictly performed after the sliding window sequence generation. The initial 1000 continuous hours of raw data yielded a total of 926 supervised sequence samples. These samples were sequentially divided into three distinct subsets: a training set (648 samples, approx. 70%), a validation set (92 samples, approx. 10%), and a testing set (186 samples, approx. 20%). The training set was utilized for model training and parameter optimization (SSA-VMD). The validation set served for hyperparameter tuning and early stopping monitoring to prevent overfitting. The testing set was strictly reserved as unseen data for the final performance evaluation.

2.2. Theoretical Background

2.2.1. Principal Component Analysis (PCA)

PCA is a statistical procedure used to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCA, implemented via scikit-learn version 1.7.2, NumFOCUS, Austin, TX, USA) [22]. Given the wind speed matrix $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{T} \times \boldsymbol{N}}$, PCA aims to find an orthogonal projection matrix $\boldsymbol{W}$ to maximize the variance. The core operation involves the eigendecomposition of the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \frac{\mathbf{1}}{\boldsymbol{T} - \mathbf{1}} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{X} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{\boldsymbol{T}} \tag{3}$$

where $\boldsymbol{\Lambda}$ represents the eigenvalues. In this study, PCA is employed to extract the dominant spatial features of the regional wind field, reducing the input dimension from 81 to a lower-dimensional subspace while retaining 99.5% of the variance.

2.2.2. Variational Mode Decomposition (VMD)

Unlike EMD, VMD is a non-recursive signal decomposition method that decomposes the input signal $f(t)$ into K discrete sub-signals (modes) u_k, each with a specific center frequency ω_k (VMD, implemented via the vmdpy package version 0.1, Montana State University, Bozeman, MT, USA) [23]. The decomposition is formulated as a constrained variational problem to minimize the sum of the estimated bandwidths of each mode:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} ||\partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} ||_2^2 \right\} \tag{4}$$

$$\text{s.t.} \sum_{k=1}^{K} u_k = f \tag{5}$$

This method is crucial for handling the non-stationarity of wind speed by separating high-frequency noise from the main trend.

2.2.3. Sparrow Search Algorithm (SSA)

The SSA is a novel swarm intelligence optimization algorithm inspired by the foraging and anti-predation behavior of sparrows (SSA, implemented using Python version 3.10.9, Python Software Foundation, Wilmington, DE, USA) [24]. It demonstrates high convergence speed and stability. In this study, the SSA is utilized to adaptively optimize the penalty factor α and the number of modes K for VMD.

The objective function (fitness function) for the optimization is the Envelope Entropy (E_p), which measures the sparsity of the signal. A lower entropy indicates less noise and more feature-rich decomposition. The entropy is defined as

$$E_p = -\sum_{i=1}^{N} p_i \log p_i,\ p_i = \frac{a(i)}{\sum_{j=1}^{N} a(j)} \tag{6}$$

where $a(i)$ is the envelope signal of the IMFs obtained from the Hilbert transform.

2.2.4. Long Short-Term Memory (LSTM)

LSTM is a special type of Recurrent Neural Network (RNN) designed to learn long-term dependencies (LSTM, implemented via PyTorch version 2.5.1, Meta Platforms, Inc., Menlo Park, CA, USA) [25]. It introduces a memory cell and three gates (forget gate, input gate, and output gate) to control the flow of information. The core transition equations are as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_t + b_i) \tag{7}$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_t + b_f\right) \tag{8}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_t + b_o) \tag{9}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \tag{10}$$

$$h_t = o_t \odot \tanh(c_t) \tag{11}$$

In this framework, LSTM serves as the predictor for the screened subsequences extracted by the hybrid decomposition module.

2.3. The Proposed Forecasting Framework

The overall framework of the proposed regional wind speed forecasting model is illustrated in Figure 1. The workflow is composed of three sequential phases: spatial feature extraction, adaptive temporal decomposition, and nonlinear prediction. The detailed implementation procedure is described as follows:

Phase 1: Spatial Dimensionality Reduction via PCA

Given the raw regional wind field data represented as a matrix $\boldsymbol{X} \in \mathbb{R}^{T\times N}$ (where $N = 81$ grid points), direct modeling would incur high computational costs and suffer from spatial redundancy.

To address this, Principal Component Analysis (PCA) is first applied to $\boldsymbol{X}$. By calculating the cumulative contribution rate (CCR), the top k principal components (PCs) that satisfy $CCR \geq 99.5\%$ are retained. This step effectively compresses the high-dimensional spatial information into a lower-dimensional subspace $\boldsymbol{P} \in \mathbb{R}^{T\times k}$ (in this study, $k = 51$), decoupling the spatial correlations while preserving the dominant wind field characteristics.

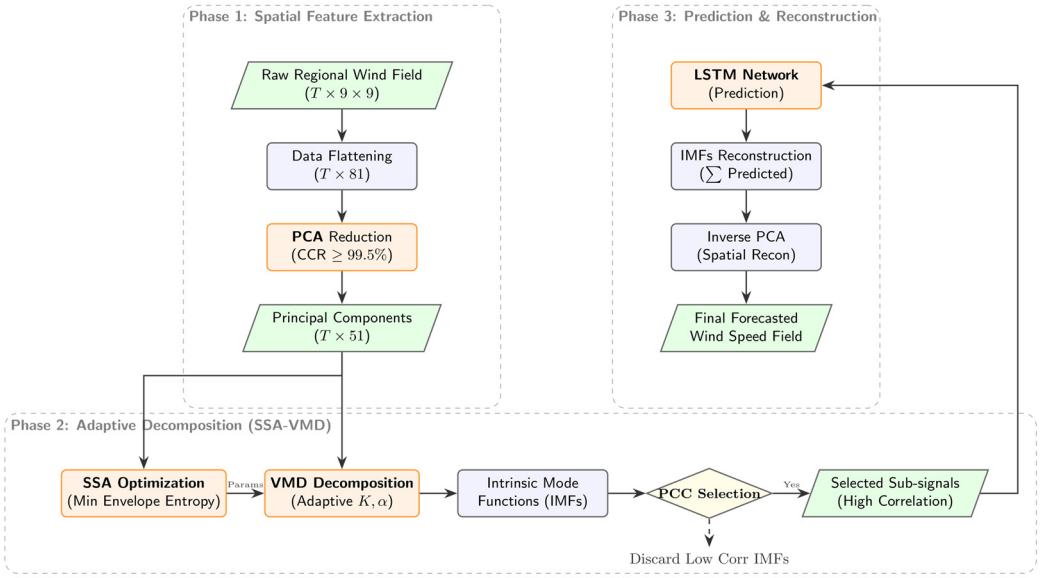


Figure 1. Flowchart of the proposed PCA-VMD-SSA-LSTM forecasting framework.

Phase 2: Adaptive Temporal Decomposition (SSA-VMD)

The extracted PCs still contain non-stationary fluctuations and noise. To enhance predictability, an adaptive decomposition strategy is employed for each principal component $\boldsymbol{P_i}$ $(i = 1, \ldots, k)$:

Parameter Optimization: The Sparrow Search Algorithm (SSA) is utilized to automatically search for the optimal combination of the mode number K and penalty factor α for VMD. The objective function is defined as the Minimum Envelope Entropy (as described in Equation (6)) of the decomposed modes. To ensure reproducibility, the random seed for the optimization algorithm was set to 42.

Signal Decomposition: With the optimized parameters $[K_{opt}, \alpha_{opt}]$, VMD decomposes the $\boldsymbol{P_i}$ into a set of Intrinsic Mode Functions (IMFs).

Feature Selection: To filter out high-frequency noise and redundant modes, the Pearson Correlation Coefficient (PCC) is calculated between each IMF and the original $\boldsymbol{P_i}$. High-correlation sub-signals are validated and retained via a Pearson Correlation Coefficient (PCC) threshold strategy, ensuring comprehensive feature input for the prediction model (108 input features in this study).

Phase 3: Sequence Prediction and Reconstruction

In the final phase, the selected high-quality IMFs are normalized using the fixed global scaling strategy and fed into the LSTM network.

Training: The LSTM model is trained to predict the future values of these IMFs based on historical sequences.

Reconstruction: The final regional wind speed is obtained through a two-step reconstruction process:

$$\widehat{\boldsymbol{P}}_{\boldsymbol{i}} = \sum \text{Predicted IMFs} \tag{12}$$

$$\widehat{\boldsymbol{X}} = \widehat{\boldsymbol{P}} \cdot \boldsymbol{W}^{\boldsymbol{T}} + \mu \tag{13}$$

where $\widehat{\boldsymbol{X}}$ is the forecasted wind field, $\boldsymbol{W}$ is the PCA loading matrix, and μ is the mean vector of the original data.

The parameters of all algorithms and neural networks used in this paper are summarized in Table 1.

Table 1. Model configuration.

Module	Parameter	Value
Input Data	Time Resolution	1 h
	Grid Size	9 × 9 (81 points)
	Train/Val/Test Split	648/92/186 samples (7:1:2)
PCA	Cumulative Contribution Rate	≥99.5%
	Reduced Dimensions	51
SSA-VMD	Population Size (SSA)	20
	Max Iterations (SSA)	30
	Objective Function	Minimum Envelope Entropy
LSTM	Input Size	108
	Hidden Layers	2
	Hidden Units	256, 256
	Optimizer	Adam (lr = 0.001)
	Batch Size	32
	Epochs	300

All deep learning algorithms and forecasting experiments were implemented using Python (version 3.10.9, Python Software Foundation, Wilmington, DE, USA) and the PyTorch framework (version 2.5.1, Meta Platforms, Inc., Menlo Park, CA, USA). The computations were accelerated on a workstation equipped with an NVIDIA GeForce RTX 4060 Laptop GPU (NVIDIA Corporation, Santa Clara, CA, USA).

2.4. Evaluation Metrics

To strictly evaluate the prediction performance of the proposed model and benchmarks, three standard statistical metrics are adopted: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).

Assuming y_i represents the observed wind speed, $\widehat{y_i}$ denotes the predicted value, $\overline{y}$ is the mean of observed values, and N is the number of samples, the definitions are as follows:

Root Mean Square Error (RMSE): RMSE measures the standard deviation of the prediction errors. It gives higher weight to large errors, making it sensitive to outliers. A lower RMSE indicates better stability.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2} \tag{14}$$

Mean Absolute Error (MAE): MAE calculates the average magnitude of errors in a set of predictions, without considering their direction. It provides a straightforward assessment of the prediction accuracy.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y_i}| \tag{15}$$

Coefficient of Determination (R^2): R^2 evaluates the goodness of fit, representing the proportion of the variance in the dependent variable that is predictable from the independent variables. A R^2 score closer to 1 indicates a perfect fit, while a score near 0 (or negative) implies poor predictive capability.

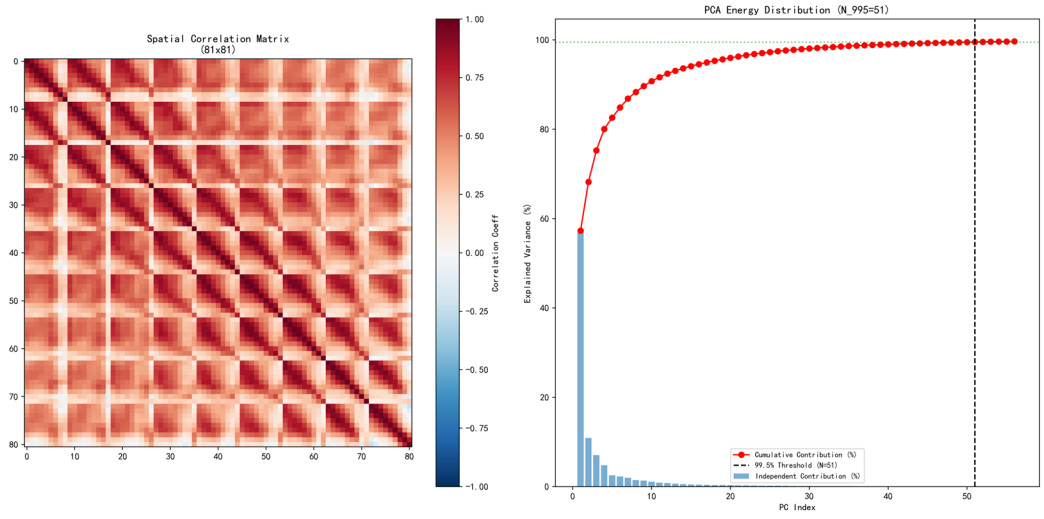
$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_i)^2}{\sum_{i=1}^{N}(\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2} \tag{16}$$

Statistical Significance Testing: To rigorously determine whether the proposed framework exhibits significant improvements over the baseline models, formal statistical analyses were conducted on the prediction results. Specifically, the Wilcoxon signed-rank test and the Diebold-Mariano (DM) test were employed to evaluate the statistical significance of the forecasting errors between the proposed model and the baselines. A p-value of less than 0.05 was considered statistically significant.

3. Results

3.1. Visualization of Decomposition

To investigate the spatial dependencies within the regional wind field, the Pearson correlation coefficient matrix of the 81 grid points was calculated, as shown in Figure 2. The visualization reveals a strong positive correlation (indicated by dark-red blocks) among spatially adjacent grid points. This phenomenon is consistent with the physical continuity of fluid dynamics, where wind speed variations at a specific location are closely related to its neighbors. However, the correlation coefficient gradually decreases as the spatial distance increases (transitioning from red to light orange). The widespread high-correlation areas in the matrix indicate significant spatial redundancy in the raw 9×9 grid data. Directly using all 81 points as input features for the neural network would not only introduce collinearity issues but also significantly increase the computational burden.



(**a**) (**b**)

Figure 2. Spatial feature analysis of the regional wind field. (**a**) Spatial correlation matrix heatmap showing high redundancy among adjacent grid points; (**b**) PCA variance explanation chart, where the first 51 principal components preserve over 99.5% of the cumulative information.

Following spatial dimensionality reduction, the extracted Principal Components (PCs) exhibit varying degrees of complexity. Low-order PCs typically dominate the main trend, while high-order PCs contain intricate spatiotemporal details and noise. To address this heterogeneity, the SSA-VMD module was applied to adaptively determine the optimal number of modes (K) and penalty factor (α) for each component.

The optimization results reveal significant diversity in the decomposition strategies. As summarized in Table 2, simpler components like PC1, PC3, and PC4 were decomposed into fewer modes ($K = 3$), indicating that separating the main trend from the residual is

sufficient for these signals. In contrast, complex components like PC14 ($K = 12$) and PC24 ($K = 10$) required a finer granularity to effectively disentangle the mixed frequencies. This adaptive behavior validates the necessity of the proposed strategy, preventing the model from under-decomposing complex signals or over-decomposing simple ones.

Table 2. SSA-VMD parameter optimization results. The search bounds for the mode number (K) and penalty factor (α) were set to [3, 12] and [100, 3000], respectively. The SSA population size was 20 with 30 iterations.

(PC1–PC20)			(PC21–PC40)			(PC41–PC51)		
Component	K	α	Component	K	α	Component	K	α
PC1	3	2093	PC21	8	100	PC41	9	100
PC2	5	100	PC22	8	100	PC42	8	100
PC3	3	100	PC23	3	906	PC43	6	100
PC4	3	348	PC24	10	100	PC44	8	3000
PC5	5	1508	PC25	8	100	PC45	7	100
PC6	5	3000	PC26	3	740	PC46	9	100
PC7	3	661	PC27	8	100	PC47	6	100
PC8	3	100	PC28	8	100	PC48	3	2041
PC9	3	631	PC29	9	100	PC49	4	100
PC10	5	100	PC30	6	106	PC50	4	100
PC11	3	2977	PC31	6	100	PC51	12	100
PC12	4	136	PC32	4	148			
PC13	4	100	PC33	12	100			
PC14	12	100	PC34	3	118			
PC15	4	2336	PC35	9	100			
PC16	3	100	PC36	3	100			
PC17	3	412	PC37	6	100			
PC18	5	100	PC38	4	100			
PC19	3	188	PC39	9	100			
PC20	8	100	PC40	5	100			

To demonstrate the optimization process, the 2nd Principal Component (PC2) is selected as a representative case study. Figure 3 illustrates the SSA convergence curve for PC2. The algorithm initially settled at a local optimum with $K = 10$ (Fitness ≈ 6.758). However, at the 6th iteration, the sparrow search mechanism successfully escaped this local trap, identifying a superior configuration with $K = 5$ and a refined $\alpha = 100$ (Fitness ≈ 6.7536). The curve stabilizes rapidly thereafter, demonstrating the algorithm's strong global search capability.

Figure 4 presents the decomposition results of PC2 using these optimal parameters. In the time domain (left), the non-stationary PC2 signal is decomposed into five distinct IMFs. The corresponding Power Spectral Density (PSD) plots (right) show that the spectral bands of the five IMFs are well-separated with minimal overlap. This confirms that the optimized VMD effectively extracts the intrinsic features of the wind signal without mode mixing, providing high-quality inputs for the subsequent LSTM prediction.

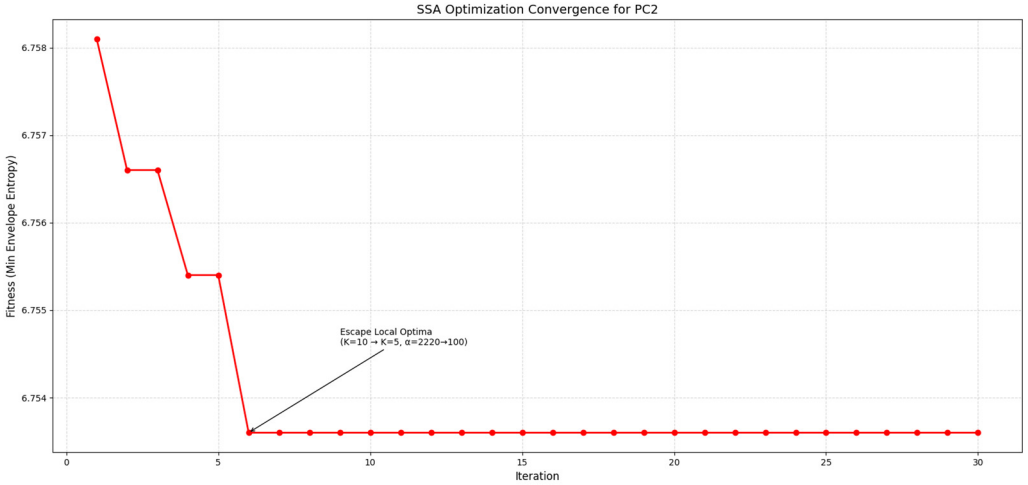


Figure 3. SSA optimization convergence curve for PC2. Sparrow search mechanism successfully escaped the local trap.

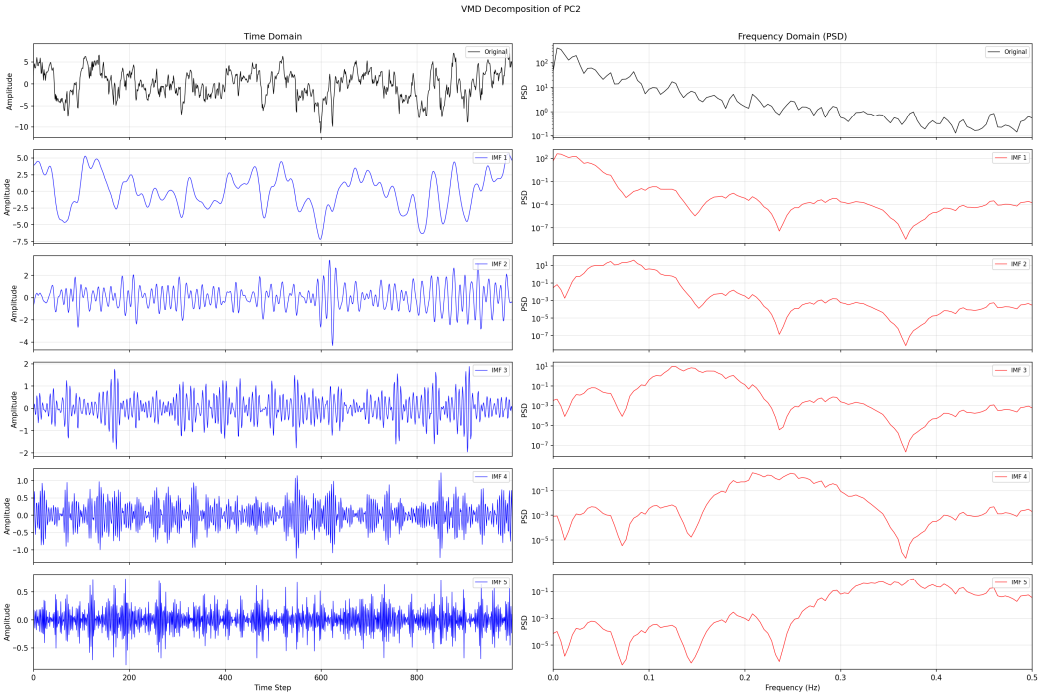


Figure 4. Results of PC2 using these optimal parameters.

3.2. Spatiotemporal Prediction Performance

To quantitatively assess the overall forecasting performance of the proposed framework, Table 3 summarizes the average evaluation metrics (RMSE, R^2, MAE, and MAPE) across the entire 9×9 grid for multiple forecasting horizons (Step 1 to Step 3, corresponding to 1 to 3 h ahead). This comprehensive comparison establishes the baseline trajectories of each model over time. As indicated in the table, the baseline models either fail to capture the wind dynamics across all steps (e.g., Raw LSTM and Random Forest) or suffer from rapid performance degradation over successive prediction horizons (e.g., Raw ConvLSTM). In contrast, the proposed PCA-VMD-SSA-LSTM framework exhibits a highly stable predictive capability, maintaining an R^2 of 0.3902 and a MAPE of 41.30% even at the third forecasting step. This multi-step robustness highlights the advantage of the decomposition-based strategy. To further dissect these overall metrics, the following sections provide a detailed spatial and temporal analysis of the model performances.

Table 3. Quantitative comparison of the average forecasting performance across the regional grid for multiple prediction steps (1 to 3 steps).

Model	RMSE	R^2	MAE	MAPE
		Step 1		
Proposed Model	0.9127	0.4112	0.7159	40.49%
Raw LSTM	1.0039	0.3250	0.7606	40.27%
Raw ConvLSTM	0.6485	0.6980	0.4911	27.00%
Random Forest	1.0987	0.2872	0.8089	40.40%
Ridge Regression	0.9987	0.2957	0.7727	41.35%
		Step 2		
Proposed Model	0.9254	0.3904	0.7253	40.87%
Raw LSTM	1.0819	0.2601	0.8133	41.34%
Raw ConvLSTM	0.9041	0.4647	0.6771	34.84%
Random Forest	1.1547	0.2296	0.8487	41.36%
Ridge Regression	1.1511	0.1128	0.8870	45.52%
		Step 3		
Proposed Model	0.9273	0.3902	0.7271	41.30%
Raw LSTM	1.1748	0.1632	0.8750	42.40%
Raw ConvLSTM	1.1117	0.2377	0.8215	40.06%
Random Forest	1.2135	0.1642	0.8914	42.28%
Ridge Regression	1.2556	−0.0204	0.9653	48.21%

To visually evaluate the forecasting capability across the entire 9×9 regional grid, the spatial distributions of the Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) scores for the first forecasting step (Step 1) are presented in Figures 5 and 6. The heatmap visualizations reveal significant spatial heterogeneity and varying model sensitivities to the highly noisy wind field dataset.

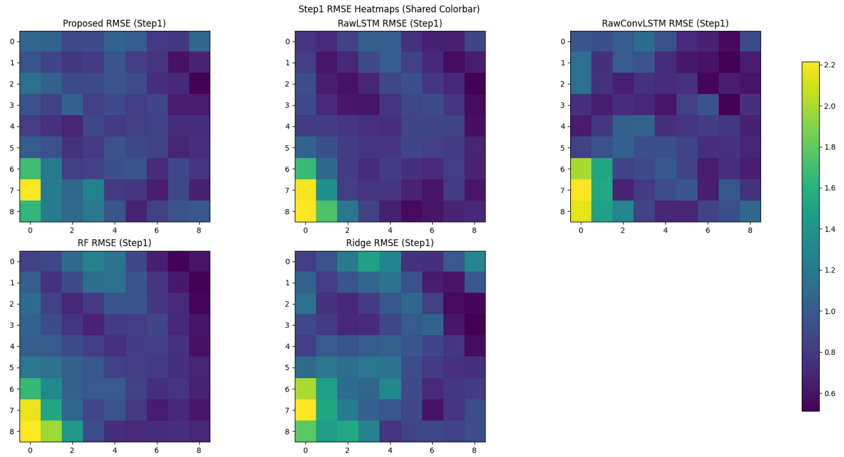


Figure 5. Spatial distribution of the Root Mean Square Error (RMSE) for the proposed model and four baselines at the first forecasting step (Step 1).

As shown in the figures, applying basic deep learning or ensemble methods directly to raw temporal data yielded sub-optimal performance. The Raw LSTM and Random Forest models struggled to achieve high accuracy, resulting in average R^2 scores of 0.3250 and 0.2872, respectively. While they captured the general trend of the wind dynamics—avoiding complete predictive failure—their limited accuracy is fundamentally attributed to their vulnerability to temporal non-stationarity, which leads them to partially overfit the high-frequency stochastic noise inherent in the raw wind data. Similarly, the linear Ridge Regression model achieved an average R^2 of only 0.2957. In contrast, the proposed PCA-VMD-SSA-LSTM framework achieved a significantly higher R^2 of 0.4112 in Step 1,

proving that the signal decomposition effectively extracts predictable deterministic trends and breaks through the performance bottleneck of conventional models.

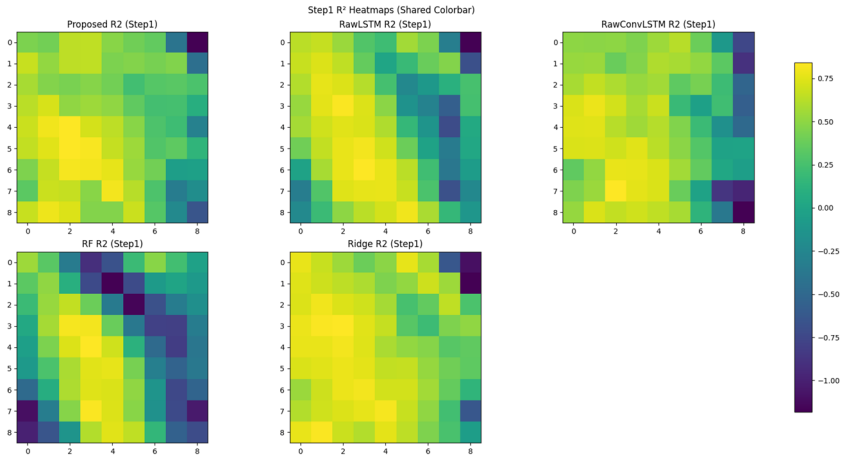


Figure 6. Spatial distribution of the Coefficient of Determination (R^2) scores for the proposed model and four baselines at the first forecasting step (Step 1).

The RawConvLSTM model exhibited the highest accuracy for the ultra-short-term Step 1 prediction, achieving an average R^2 of 0.6980 and the lowest RMSE of 0.6485 This high initial performance occurs because its complex spatiotemporal architecture explicitly captures localized, high-frequency spatial correlations across the grid. However, the proposed PCA-VMD-SSA-LSTM framework also demonstrated highly competitive robustness, achieving an average R^2 of 0.4112 and an RMSE of 0.9127. Spatially, the proposed model exhibited excellent stability across the vast majority of the central and western grid regions (indicated by the widespread bright yellow and green areas in the R^2 heatmap). Although a localized performance drop was observed at the top-right boundary corner, this boundary effect is a common phenomenon in regional topological modeling.

It is crucial to highlight the underlying trade-off revealed here. While RawConvLSTM outperforms the proposed model in Step 1 by assimilating raw, high-frequency spatial details, this behavior acts as a double-edged sword. By preserving and reacting to chaotic turbulence, ConvLSTM is highly vulnerable to error accumulation over time. The proposed framework, by design, sacrifices a fraction of this initial high-frequency sensitivity to ensure global stability. As will be quantitatively analyzed in the subsequent multi-step forecasting section, the denoising strategy grants the proposed model exceptional error-mitigating capability and long-term robustness, ultimately outperforming ConvLSTM in multi-horizon tasks.

Statistical Significance Analysis: To rigorously compare the predictive performance, the Wilcoxon signed-rank test and the Diebold-Mariano (DM) test were conducted on the testing set (using MSE as the loss function, $\alpha = 0.05$). The statistical results reveal that the proposed framework significantly outperforms the traditional RawLSTM model (Wilcoxon $p < 0.001$, DM test $p < 0.01$). When compared to the RawConvLSTM, the statistical difference across the averaged 3-step horizon is not strictly significant. This aligns with our expectations: ConvLSTM possesses an absolute advantage in capturing high-frequency spatial turbulence in the ultra-short term (Step 1), thereby masking its subsequent severe degradation in the averaged metrics. However, as the prediction horizon extends, the proposed model demonstrates statistically superior stability, which is further analyzed in the multi-step performance section.

3.3. Time-Series Analysis at Representative Grid

To provide a deeper physical interpretation of the models' dynamic forecasting behaviors, a detailed time-series analysis was conducted at representative grid points. First, Grid (8,0) was selected to demonstrate the fundamental denoising capability of the proposed framework in a wind regime characterized by distinct macro-trends coupled with continuous micro-fluctuations.

Figure 7 illustrates the 186 h predictive trajectories at Grid (8,0) for the first forecasting step (Step 1). As observed in the true wind speed sequence, the raw signal is highly non-stationary, filled with dense, high-frequency stochastic jitters. The proposed PCA-VMD-SSA-LSTM framework exhibits a remarkable smoothing effect. By leveraging the adaptive VMD preprocessing, it successfully filters out the chaotic turbulence and extracts the deterministic physical backbone of the wind field. Consequently, its predictive trajectory tightly traces the macro-level peaks and valleys without being distracted by localized noise.

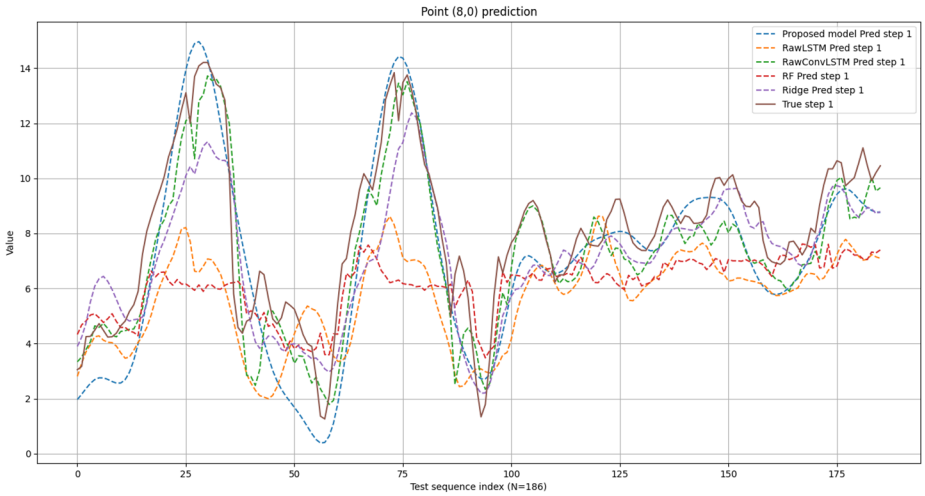


Figure 7. Time-series forecasting results at Grid (8,0) over a 186-step test sequence for Step 1.

Conversely, baseline models trained directly on the raw temporal data display highly unstable behaviors. As shown in Figure, models such as Raw LSTM and Random Forest (RF) exhibit severe oscillatory predictive curves. By aggressively attempting to fit the unpredictable high-frequency noise inherent in the raw signal, leading to significant phase lags and structural distortion. This visual evidence perfectly explains their poor quantitative performance across the grid, reaffirming that deep learning models are highly vulnerable to overfitting when exposed to raw, low-signal-to-noise-ratio meteorological data.

To further explore the models' behaviors under extreme volatile conditions, a second representative node, Grid (4,4), was analyzed. Located at the geographic center of the target region, this grid experiences drastic wind speed surges and deep troughs within short intervals, making it a rigorous testbed for evaluating the trade-off between peak capturing and noise filtering.

As depicted in Figure 8, the true wind speed at Grid (4,4) contains severe, sudden spikes. Here, a deliberate trade-off in the proposed PCA-VMD-SSA-LSTM framework becomes apparent: amplitude damping. Because the VMD module inherently filters out ultra-high-frequency, instantaneous spikes as stochastic noise, the proposed model slightly underestimates the absolute maximum peaks. In contrast, the RawConvLSTM model successfully captures these extreme instantaneous spikes in the ultra-short-term (Step 1). By utilizing its spatial convolutional kernels directly on the raw grid, ConvLSTM accurately assimilates the localized, high-frequency spatial turbulence into its single-step prediction.

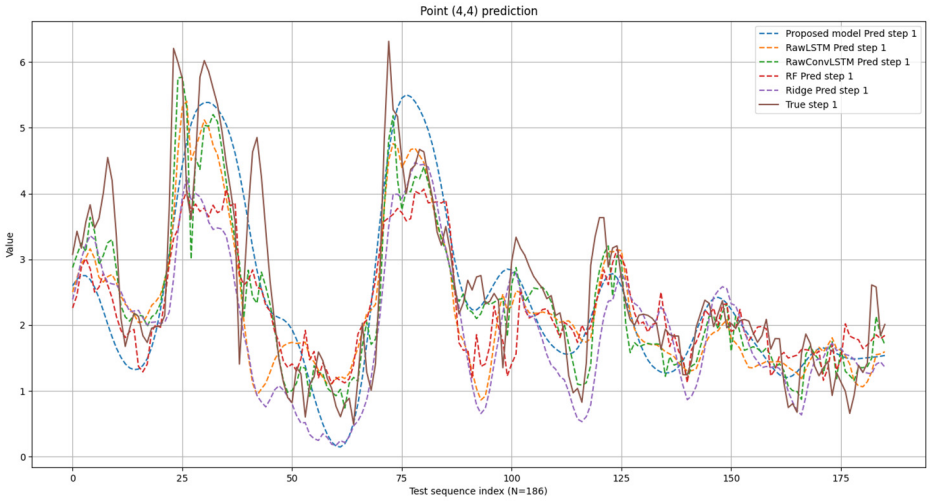


Figure 8. Time-series forecasting comparison at the highly volatile central node (4,4).

However, this initial advantage of the RawConvLSTM acts as a double-edged sword. While fitting raw high-frequency features allows for precise peak matching in Step 1, it simultaneously implies that the model has internalized chaotic, unpredictable turbulence as a learnable pattern. As will be quantitatively revealed in the subsequent multi-step analysis, this extreme sensitivity to transient noise ultimately becomes the primary catalyst for severe error accumulation over longer forecasting horizons. The proposed model, despite the slight peak underestimation, intentionally maintains a smoother trajectory to ensure structural stability.

3.4. Multi-Step Forecasting Performance

To evaluate the models' robustness over longer prediction horizons, multi-step forecasting (Step 1 to Step 3) was conducted. As previously summarized in Table 3, the regional average metrics indicate a distinct divergence in multi-step performance among the evaluated models. To visually illustrate this degradation, Figure 9 plots the R^2 decay trajectories for both the regional mean and the central node, Grid (4,4).

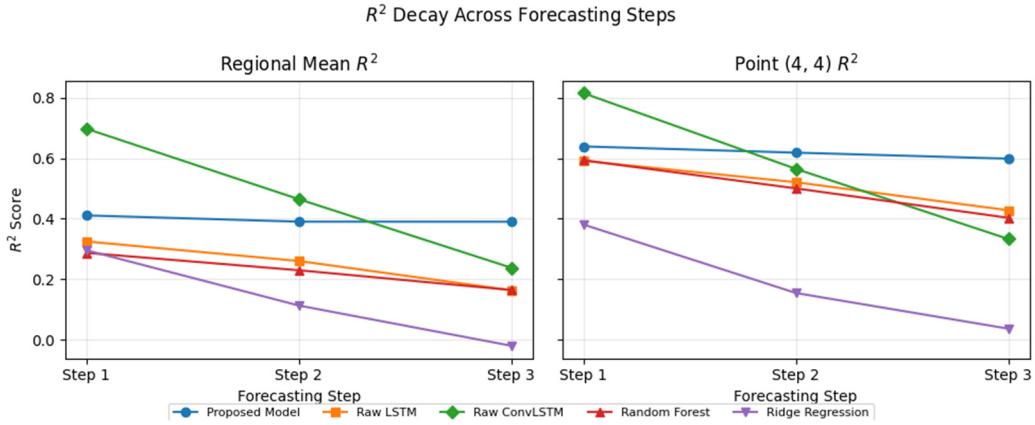


Figure 9. The R^2 score decay trajectories across forecasting steps for the regional mean and the central Grid (4,4). The proposed model shows minimal performance decay, whereas RawConvLSTM shows severe performance degradation.

As shown in Figure 9, the performance degradation over the prediction horizons is remarkably clear. The RawConvLSTM model, despite its superior initial accuracy (R^2 = 0.8190 at Grid (4,4) in Step 1), experiences a drastic performance drop, plummeting to 0.5645 in Step 2 and further to 0.3370 in Step 3. A continuous, albeit gentler, degra-

dation pattern is also observed in the Ridge Regression model. In sharp contrast, the proposed PCA-VMD-SSA-LSTM framework maintains a highly consistent predictive capability, yielding a near-horizontal trajectory with R^2 scores of 0.6412, 0.6188, and 0.5995 across the three steps.

This error-mitigating characteristic is not an isolated phenomenon but is spatially universal. Table 4 details the multi-step R^2 scores for the Proposed model, Ridge Regression, and ConvLSTM at four randomly selected grids across the region. At all sampled locations, the model directly processing raw spatial data (ConvLSTM) suffers from continuous and steep performance degradation, whereas the proposed framework robustly sustains its predictive accuracy. The underlying mechanisms driving this stark contrast in multi-step error propagation will be thoroughly analyzed in Section 4.

Table 4. Multi-step R^2 scores at four randomly selected grid points, demonstrating the universal error-mitigatingcapability of the proposed framework.

	Proposed	Ridge	ConvLSTM	Proposed	Ridge	ConvLSTM
		Point 1			Point 2	
Step 1	0.4696	0.2827	0.7997	0.6498	0.1119	0.7428
Step 2	0.4929	−0.0169	0.5822	0.5775	−0.0657	0.5835
Step 3	0.4144	−0.2587	0.3548	0.5466	−0.1847	0.3365
		Point 3			Point 4	
Step 1	0.2166	0.2762	0.2898	0.5062	0.4938	0.8435
Step 2	0.1582	0.2785	0.0441	0.5223	0.2712	0.5899
Step 3	0.0899	0.3162	−0.1132	0.4768	0.1566	0.3731

3.5. Ablation Study

To explicitly clarify the added value of individual components in the proposed framework, an ablation study was conducted. Two variant models were constructed: PCA-LSTM (removing the VMD and SSA modules to evaluate the necessity of temporal decomposition) and VMD-LSTM (using fixed, empirical VMD parameters without SSA optimization to assess the value of adaptive search). The regional average metrics are compared in Table 5.

Table 5. Performance comparison between ablation models.

Model	RMSE	R^2	MAE	MAPE
		Step 1		
Proposed Model	0.9127	0.4112	0.7159	40.49%
Ablation PCA-LSTM	1.0801	0.2765	0.8154	38.91%
Ablation VMD-NoSSA	0.9001	0.3991	0.6889	40.57%
		Step 2		
Proposed Model	0.9254	0.3904	0.7253	40.87%
Ablation PCA-LSTM	1.1437	0.2073	0.8594	40.42%
Ablation VMD-NoSSA	0.8870	0.4098	0.6815	40.13%
		Step 3		
Proposed Model	0.9273	0.3902	0.7271	41.30%
Ablation PCA-LSTM	1.2122	0.1312	0.9038	41.58%
Ablation VMD-NoSSA	0.8982	0.3890	0.6884	40.81%

As shown in Table 5, removing the VMD module (PCA-LSTM) leads to a rapid collapse in multi-step forecasting (R^2 dropping to 0.13 in Step 3), proving that VMD is the core driving force behind the model's stable multi-step capability. Furthermore, while the unoptimized VMD(No SSA) achieves comparable stability to the proposed model, it relies heavily on manual parameter tuning, which lacks generalizability. The introduction of the

SSA automatically finds the optimal decomposition parameters (K and α) for diverse spatial components, ensuring high fidelity and robust forecasting without human intervention.

3.6. Computational Efficiency Analysis

In real-world regional grid dispatching, predictive accuracy must be balanced against computational cost. Table 6 summarizes the training time, prediction latency (per sample), and peak GPU memory usage of the implemented models.

Table 6. Comparison of computational costs between different models.

Model/Stage	Train (s)	Predict (s)	GPU Peak (MB)
PCA	0.016	—	—
SSA-VMD	73.567	—	—
PCC	0.020	—	—
Proposed LSTM	20.130	0.070	207.4
Raw LSTM	13.785	0.022	205.4
Raw ConvLSTM	174.690	0.226	1496.4
RF	46.777	0.015	—
Ridge	0.058	0.005	—
Ablation PCA-LSTM	12.933	0.022	202.1
Ablation VMD	85.994	0.025	229.6

Although the proposed framework incorporates an SSA-VMD preprocessing phase (taking approx. 73.5 s for parameter search), its total computational footprint remains extremely light. Most notably, compared to the end-to-end spatiotemporal RawConvLSTM model, the proposed framework reduces the peak GPU memory consumption by nearly 86% (from 1496.4 MB to 207.4 MB) and accelerates the inference latency by more than 3 times (0.070 s vs. 0.226 s). This demonstrates that the "decomposition-first" strategy effectively circumvents the massive parameter overhead inherent in 3D convolutions, offering a highly efficient and deployable solution for edge-computing environments in wind farms.

4. Discussion

The primary objective of this study was to develop a robust, stable multi-step regional wind speed forecasting framework capable of operating in highly stochastic and noisy meteorological environments. While current literature heavily favors increasingly complex end-to-end deep learning architectures, the empirical results of this study—particularly the multi-step forecasting trajectories—reveal a critical limitation of directly training end-to-end ConvLSTM models on raw wind fields in this setting. By explicitly decoupling spatial redundancy and filtering temporal non-stationarity, the proposed PCA-VMD-SSA-LSTM framework demonstrates a more reliable pathway for practical grid dispatching. The underlying mechanisms driving these performance divergences are discussed below.

4.1. The Double-Edged Sword of Complex Spatiotemporal Models

A significant finding of this study is the stark contrast in the multi-step forecasting behaviors between the RawConvLSTM and the proposed decomposition-based framework. In the ultra-short-term forecasting horizon (Step 1), RawConvLSTM achieved the highest accuracy. This superior initial performance can be attributed to its spatial convolutional operations, which directly assimilate localized, high-frequency spatial turbulence across the 9×9 grid into its prediction state. As validated by the statistical significance analysis in Section 3.2, ConvLSTM possesses a localized advantage in the ultra-short term (Step 1).

However, the time-series analysis (Figure 8) and multi-step decay evaluations (Figure 9) reveal that this high-frequency sensitivity acts as a double-edged sword. Wind speed data inherently comprises a low-frequency deterministic backbone superimposed with high-frequency aleatoric uncertainty (stochastic turbulence). By training directly on the raw wind field, ConvLSTM is forced to internalize both components. Consequently, it effectively overfits the stochastic noise, misinterpreting chaotic turbulence as a predictable, learnable pattern.

In an autoregressive multi-step forecasting scenario, this overfitting triggers a severe error accumulation. From an autoregressive perspective, each forecasting step implicitly builds upon the previous noisy estimates, so high-frequency errors are recursively propagated and amplified across steps. Because the high-frequency turbulence is fundamentally chaotic, the predicted noise rapidly diverges from reality, leading to cascading structural distortions, severe phase lags, and false alarm fluctuations. This mechanistic flaw fundamentally explains why the RawConvLSTM's performance plummeted to an R^2 of 0.2377 by the third forecasting step, rendering it unsuitable for reliable long-term grid scheduling.

4.2. Denoising vs. Fidelity: The Trade-Off in Signal Decomposition

To overcome the inherent vulnerability of complex spatiotemporal models, the proposed framework adopts a "decomposition-first, prediction-later" paradigm. The exceptional multi-step robustness demonstrated in Figure 9 is fundamentally driven by the SSA-VMD module. By adaptively decomposing the principal components and strategically discarding low-correlation, high-frequency IMFs, the framework effectively functions as an intelligent low-pass filter. This mechanism isolates the deterministic physical backbone of the wind field, allowing the subsequent LSTM network to focus exclusively on learning stable, low-frequency atmospheric dynamics rather than chasing unpredictable turbulence.

However, this signal decomposition strategy inherently introduces a crucial trade-off between denoising and signal fidelity. As visually evidenced in the time-series analysis at the highly volatile Grid (4,4) (Figure 8), the proposed model exhibits a slight "amplitude damping" effect. Extreme, instantaneous wind speed spikes are often the result of sudden, high-frequency localized gusts superimposing on the main trend. Because the VMD algorithm identifies and filters these transient extremes as stochastic noise, the reconstructed input signal underestimates the absolute peak values in the ultra-short-term forecasting (Step 1).

Despite this localized loss of extreme peak fidelity, this deliberate design choice yields overwhelming benefits for multi-step forecasting. By sacrificing the network's sensitivity to chaotic transient spikes, the proposed framework becomes fundamentally immune to the cascading error propagation described in Section 4.1. The LSTM net-work, fed with clean and deterministic features, avoids internalizing false alarm fluctuations, thereby sustaining a minimal decay rate across consecutive forecasting steps (Table 4).

From a practical engineering perspective, this trade-off is highly advantageous. For wind farm operators and power grid dispatchers, the primary goal of regional forecasting is not to perfectly predict a 10 min random gust, which cannot be practically utilized by large-scale turbines. Instead, the objective is to reliably foresee the macroscopic wind power trend over the next several hours to ensure safe energy scheduling and reserve allocation. The proposed framework, by prioritizing long-term temporal stability and structural robustness over transient extreme fitting, aligns perfectly with these real-world operational requirements.

4.3. The Critical Impact of Normalization on Multiscale Energy Hierarchy

A crucial, yet often overlooked, aspect of developing hybrid decomposition-prediction frameworks is the choice of data normalization. To investigate this, an ablation study was conducted comparing the proposed fixed global scaling strategy ($x'_t = x_t/50.0$) against conventional independent Z-score standardization (zero mean, unit variance). The empirical results revealed a striking phenomenon: applying Z-score standardization caused the proposed framework to completely collapse (with the regional average R^2 plummeting to -0.3514 in Step 1). Conversely, the RawConvLSTM model benefited from Z-score normalization but suffered performance degradation when using fixed scaling.

This stark contrast is deeply rooted in the physical and mathematical structures of the input data. The RawConvLSTM directly processes the raw wind field, which is a homogeneous physical quantity (wind speed). For such raw spatial matrices, independent Z-score standardization effectively stabilizes neural network gradients without destroying any underlying structure.

However, the input to the proposed LSTM predictor is fundamentally different; it consists of multiscale sub-signals (PCs and IMFs) generated by PCA and VMD. These decomposition algorithms intrinsically distribute the wind field's energy across different frequency bands. Typically, the low-frequency modes (representing the main physical trend) carry the vast majority of the signal's energy and possess large variances. In contrast, the high-frequency modes (representing stochastic turbulence) contain minimal energy and exhibit extremely small variances. This uneven variance distribution forms a critical "multiscale energy hierarchy."

When independent Z-score standardization is applied to these decomposed channels, it forcibly equalizes their variances to 1. Mathematically, this acts as a massive artificial amplifier for the high-frequency noise and a dampener for the low-frequency main trend. The neural network is subsequently overwhelmed by the artificially amplified turbulence, negating the denoising efforts of the SSA-VMD module.

The fixed global scaling strategy employed in this study elegantly resolves this issue. By dividing all decomposed components by a single, physically meaningful constant, the normalization bounds the data within a range suitable for LSTM training (typically [0, 1]) while strictly preserving the relative amplitude ratios and the intrinsic energy hierarchy among the IMFs. This finding highlights a critical methodological guideline: for any "decomposition-first" deep learning framework, the normalization strategy must be physically aware and strictly preserve the multiscale energy distribution of the sub-signals.

5. Conclusions

This study proposes a hybrid deep learning framework (PCA–VMD–SSA–LSTM) for short-term regional wind speed forecasting, specifically designed to mitigate spatiotemporal redundancy and strong stochastic noise. Based on comprehensive multi-step evaluations on a (9×9) regional grid, the main findings can be summarized as follows:

1. Robustness in Multi-Step Forecasting: Although complex spatiotemporal networks such as RawConvLSTM achieve very competitive accuracy in ultra-short-term, single-step prediction, the experiments in this study show that their performance deteriorates rapidly as the forecasting horizon extends, mainly due to their tendency to fit high-frequency turbulent components in the raw wind field. In contrast, the proposed decomposition-first framework exhibits a much slower decay of (R^2) across consecutive steps, indicating more stable multi-step forecasting behavior under noisy meteorological conditions. Furthermore, as demonstrated by the computational efficiency analysis, the proposed framework achieves this stability while reducing peak

GPU memory consumption by nearly 86% and accelerating inference by over 3 times compared to ConvLSTM, making it highly viable for edge-computing deployments.

2. A practical denoising trade-off: The SSA-optimized VMD module effectively acts as a data-driven low-pass filter. While this inevitably introduces slight amplitude damping for some instantaneous extreme spikes, it successfully extracts a smoother, physically interpretable backbone of the wind field and suppresses random fluctuations. From the perspective of grid operation, this trade-off—sacrificing a small amount of peak-fitting ability in exchange for improved overall stability and reduced false alarms—is favorable for reliable short-term dispatching.
3. Importance of preserving the multiscale energy hierarchy: The ablation study on normalization strategies reveals that conventional channel-wise standardization (e.g., independent Z-score) can be harmful for decomposition-based models, as it artificially equalizes the variances of low-frequency and high-frequency sub-signals, thereby amplifying noise and weakening the large-scale trend information. A simple global scaling strategy, by contrast, keeps all inputs within a numerically appropriate range for LSTM training while preserving the relative amplitude and energy hierarchy across PCs and IMFs. This suggests that, for decomposition-first deep learning frameworks, normalization should be designed to be consistent with the underlying multiscale structure rather than applied in a purely algorithmic manner.

Overall, the proposed PCA–VMD–SSA–LSTM framework provides a reliable and stable multi-step forecasting solution in the investigated case study and better satisfies the robustness and stability requirements of real-world wind power grid dispatching than the benchmark schemes considered in this work.

Author Contributions: Conceptualization, H.L. and Z.M.; methodology, H.L. and Z.M.; software, H.L., Y.L. and M.Z.; validation, Z.M., Q.Z., X.D., B.X. and L.C.; formal analysis, Z.M. and Y.L.; data curation, H.L. and M.Z.; writing—original draft preparation, H.L.; writing—review and editing, H.L., M.Z. and L.C.; visualization, H.L. and Z.M.; supervision, Z.M. and B.X.; project administration, H.L.; funding acquisition, L.C. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Aerospace Advanced Research Fund (Grant No. A240202570) and the Stable Support Project for National Basic Research Institutes (Grant No. A240204150) and The APC was funded by China Research Institute of Radiowave Propagation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw regional wind field data (ERA5 reanalysis dataset) used in this study are publicly available from the European Centre for Medium-Range Weather Forecasts (ECMWF) through the Climate Data Store at https://cds.climate.copernicus.eu/datasets/derived-era5-land-daily-statistics?tab=overview (accessed on 22 April 2026). The processed datasets and generated results during the current study are available from the corresponding author upon reasonable request. The custom computer code developed for the PCA-VMD-SSA-LSTM forecasting framework, along with all parameters used to run the software analyses, are openly available in the GitHub repository at [https://github.com/WST123-hello/PCA-VMD-SSA-LSTM-Wind-Forecasting] (accessed on 22 April 2026).

Acknowledgments: During the preparation of this manuscript, the authors used Gemini(version 3 Pro, Google, Mountain View, CA, USA) for the purposes of language polishing and structural refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication. No other individuals were included in the acknowledgments who require consent.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Autoregressive Integrated Moving Average
CCR	Cumulative Contribution Rate
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
GNN	Graph Neural Network
IMF	Intrinsic Mode Function
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NWP	Numerical Weather Prediction
PC	Principal Component
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SSA	Sparrow Search Algorithm
SVM	Support Vector Machine
VMD	Variational Mode Decomposition

References

1. Höök, M.; Li, J.; Johansson, K.; Snowden, S. Growth Rates of Global Energy Systems and Future Outlooks. *Nat. Resour. Res.* **2012**, *21*, 23–41. [CrossRef]
2. Ahmad, H.; Yaqub, M.; Lee, S.H. Global trends in carbon neutrality: A scientometric review on energy transition challenges, practices, policies, and opportunities. *Environ. Dev. Sustain.* 2025, *in press*. [CrossRef]
3. Wang, S.; Liu, H.; Yu, G. Short-term wind power combination forecasting method based on wind speed correction of numerical weather prediction. *Front. Energy Res.* **2024**, *12*, 1391692. [CrossRef]
4. Yang, M.; Jiang, Y.; Che, J.; Han, Z.; Lv, Q. Short-term forecasting of wind power based on error traceability and numerical weather prediction wind speed correction. *Electronics* **2024**, *13*, 1559. [CrossRef]
5. Elsaraiti, M.; Merabet, A. A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed. *Energies* **2021**, *14*, 6782. [CrossRef]
6. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1970.
7. Wang, H. Extreme learning Kalman filter for short-term wind speed prediction. *Front. Energy Res.* **2022**, *10*, 1047381. [CrossRef]
8. Szostek, K.; Mazur, D.; Drałus, G.; Kusznier, J. Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production. *Energies* **2024**, *17*, 4803. [CrossRef]
9. Yamin, M.; Giyats, A.F. Support Vector Regression Approach for Wind Forecasting. *Int. J. Adv. Sci. Comput. Eng.* **2022**, *4*, 95–101. [CrossRef]
10. Janakiraman, V.; P, C. Wind Speed Forecasting Using Deep Learning. In Proceedings of the 2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC), Coimbatore, India, 28–29 June 2024; pp. 250–254. [CrossRef]
11. Durap, A. Explainable deep learning techniques for wind speed forecasting in coastal areas: Integrating model configuration, regularization, early stopping, and SHAP analysis. *Neural Comput. Appl.* **2025**, *37*, 21219–21257. [CrossRef]
12. Mugware, F.W.; Sigauke, C.; Ravele, T. Evaluating wind speed forecasting models: A comparative study of CNN, DAN2, random forest and XGBOOST in diverse South African weather conditions. *Forecasting* **2024**, *6*, 672–699. [CrossRef]
13. Wang, H.K.; Song, K.; Cheng, Y. A hybrid forecasting model based on CNN and Informer for short-term wind power. *Front. Energy Res.* **2022**, *9*, 788320. [CrossRef]
14. Liu, T.; Huang, Z.; Tian, L.; Zhu, Y.; Wang, H.; Feng, S. Enhancing Wind Turbine Power Forecast via Convolutional Neural Network. *Electronics* **2021**, *10*, 261. [CrossRef]

15. Demirtop, A.; Sevli, O. Wind speed prediction using LSTM and ARIMA time series analysis models: A case study of Gelibolu. *Turk. J. Eng.* **2024**, *8*, 524–536. [CrossRef]
16. Juyal, V.D.; Kakran, S. CNN-LSTM-based wind forecasting for a residential energy management system. *Neural Comput. Appl.* **2025**, *37*, 27117–27135. [CrossRef]
17. Beu, C.M.L.; Landulfo, E. Machine-learning-based estimate of the wind speed over complex terrain using the long short-term memory (LSTM) recurrent neural network. *Wind Energy Sci.* **2024**, *9*, 1431–1450. [CrossRef]
18. Daenens, S.; Verstraeten, T.; Daems, P.-J.; Nowé, A.; Helsen, J. Spatio-temporal graph neural networks for power prediction in offshore wind farms using SCADA data. *Wind Energy Sci.* **2025**, *10*, 1137–1152. [CrossRef]
19. Mo, S.; Chen, X.; Wang, Z.; Peng, Y.; Wang, B.; Su, Y. Short-Term Wind Power Forecasting Based on Spatio-Temporal Adaptive Graph Convolutional Recurrent Network. *Energies* **2026**, *19*, 92. [CrossRef]
20. Li, H. Short-Term Wind Power Prediction via Spatial Temporal Analysis and Deep Residual Networks. *Front. Energy Res.* **2022**, *10*, 920407. [CrossRef]
21. Van der Hoven, I. Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. *J. Atmos. Sci.* **1957**, *1*, 160–164. [CrossRef]
22. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 498–520. [CrossRef]
23. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
24. Xue, J.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control Eng.* **2020**, *8*, 22–34. [CrossRef]
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [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.