第 52 卷 第 7 期：3243-3258 高电压技术 Vol.52, No.7: 3243-3258
2026 年 7 月 31 日 High Voltage Engineering July 31, 2026
DOI: 10.13336/j.1003-6520.hve.20251121

融合多源特征的 Dual-MTSA 超短期风电功率预测

陈世轲，张　维，李玉洪，杨　楠，王小玲，毕贵红
（昆明理工大学电力工程学院，昆明 650500）

摘　要：随着大规模风电并网比例不断提升，其输出的高度非线性与多尺度振荡特征，以及数据缺失与异常值干扰，给超短期功率预测带来严峻挑战。为此该文提出了一种基于数据清洗填补、混合分解与深度学习融合的超短期风电功率预测模型。首先采用三次埃尔米特插值填补缺失值，结合最小协方差行列式与马氏距离检测异常点，再通过随机森林回归修复异常点，确保输入数据的一致性；然后利用最大互信息系数(maximal information coefficient, MIC)和 Spearman 相关性分析筛选关键气象因素构建环境特征矩阵，并通过跳变模态分解(jump plus AM-FM mode decomposition, JMD)与变分模态分解(variational mode decomposition, VMD)提取多尺度功率模态分量，充分剖析信号的频谱结构并构造功率特征矩阵，接着将两个特征矩阵输入 Dual-MTSA 深度学习模型。模型将 TKAN 模块融入 Mamba 模块的选择性状态空间后，利用 Mamba 模块的线性时间复杂度优势高效捕捉跨时间点与跨变量的长程依赖，同时借助 B 样条核与门控融合机制精细刻画高频振荡与低频趋势；利用自注意力模块(self-attention, SA)对上述多源、多尺度特征进行全局加权重标定，并在全连接映射层生成最终的 15 min~3 h 超短期功率预测结果。实证结果表明，该方法相较对比模型有效降低预测误差，显著提升预测精度与鲁棒性。
关键词：风电功率预测；超短期预测；数据填补与异常修复；双分支输入；深度学习；Mamba

Dual-MTSA Ultra-short-term Wind Power Forecasting Integrating Multi-source Features

CHEN Shike, ZHANG Wei, LI Yuhong, YANG Nan, WANG Xiaoling, BI Guihong
(Faculty of Electric Power Engineering , Kunming University of Science and Technology, Kunming 650500, China)

Abstract: With the ever-increasing grid integration of large-scale wind power, its pronounced nonlinearity, multi-scale oscillations, and issues of missing or anomalous data pose serious challenges to ultra-short-term power forecasting. To address these issues, this paper presents an ultra-short-term wind power forecasting framework that integrates data cleansing and imputation, hybrid signal decomposition, and deep learning. First, missing values are recovered via cubic Hermite interpolation; anomalies are detected using the minimum covariance determinant and Mahalanobis distance and subsequently corrected with a random forest regressor to ensure data consistency. Then, the maximal information coefficient (MIC) and Spearman's correlation analysis are employed to select key meteorological variables and construct an environmental feature matrix, meanwhile, the jump-plus AM-FM mode decomposition (JMD) and variational mode decomposition (VMD) are used to extract multi-scale power components to thoroughly dissect the signals' spectral structure and form a power feature matrix. These two feature matrices are then fed into a Dual-MTSA deep learning model. Within its encoder, TKAN is embedded into Mamba's selective state-space framework, and the Mamba's linear-time complexity is used to efficiently capture long-range dependencies across both time steps and variables. Moreover, a B-spline kernel with gated fusion is utilized to finely represent high-frequency oscillations and low-frequency trends. A self-attention module is then used to globally recalibrate these multi-source, multi-scale features, and a final fully connected mapping layer generates ultra-short-term forecasts spanning 15 minutes to 3 hours. Empirical results demonstrate that the proposed method can be used to effectively reduce the prediction errors and significantly improves the prediction accuracy and robustness when compared with the benchmark models.
Key words: wind power prediction; ultra-short-term prediction; data imputation & anomaly correction; two-branch input; deep learning; Mamba

0　引言

在“双碳”战略目标驱动下，我国电力行业的低碳化进程加速推进[1]。作为清洁能源与低碳发展的核心构成要素，风电技术渗透率呈现显著增长趋势[2]，其规模化并网所引发的功率波动特性显著增加电力系统运行不确定性[3-4]。因此高精度风电功率预测，是提升电力系统调度科学性、保障新能源电

基金资助项目：国家重点研发计划(2022YFB2703500)。
Project supported by National Key R&D Program of China (2022YFB2703500).

网稳定运行的关键支撑[5]。

风电功率预测研究历经多阶段发展，研究者普遍使用物理建模、统计学习与深度学习等方法[6]。物理建模方法求解高维非线性方程推导轮毂高度风场参数，继而通过机组功率特性曲线实现机械能到电能转换预测[7]。统计方法如马尔可夫链模型[8]，差分整合移动平均自回归模型[9]。这两种方法虽具结构简捷性优势，然而显著的数据物理驱动依赖性导致其跨场景泛化性能和外推预测能力存在显著约束。随着人工智能技术的快速发展，深度学习方法凭借其非线性拟合优势在新能源预测领域备受关注[10]。文献[11]基于 Holland 气压场模型和 Batts 梯度风模型构建 Physical Net 模型，利用嵌入层动态插值台风预报信息以对齐时间尺度，将 Physical Net 输出与改进的含时序模式注意力机制的长短期记忆网络(temporal pattern attention long short-term memory, TPA-LSTM)网络结合，通过注意力机制优化时序特征权重分配，实现风速多步预测。然而台风期间极端天气下传感器损坏或通信中断导致的数据缺失问题，可能影响模型输入质量。文献[12]提出一种基于低风速功率修正和损失函数改进的超短期风电功率预测模型，低风速修正模块利用风速预测值和当前功率动态调整输出，缓解了低风速段功率拟合偏差，改进的损失函数引入余弦相似度与序列形状约束，强化了预测值与真实值在时序趋势上的关联性。然而模型核心结构以单步预测为目标设计，对多步预测的误差累积和时序依赖性衰减缺乏有效抑制机制。

信号分解通过将复杂时序数据拆解为多尺度模态分量，能显著降低输入序列的非平稳性与噪声干扰[13]。文献[14]采用 DBSCAN 算法清洗异常数据，并通过随机森林填补缺失值，降低异常数据对预测的干扰，模型融合长短时记忆网络(long short-term memory, LSTM)的短期时序建模能力与 Transformer 的全局注意力机制，结合余弦退火热重启策略优化学习率，实现了对风电功率波动特征的高效捕捉。然而风电功率数据具有显著的非平稳性，直接使用原始时序数据可能导致高频噪声与低频趋势的耦合干扰。文献[15]提出一种基于时间卷积网络(temporal convolutional network, TCN)与 Informer 的双通道风电功率预测模型，通过变分模态分解(variational mode decomposition, VMD)抑制非平稳性噪声，结合 TCN 的因果卷积与 Informer 的概率稀疏化注意力机制，显著提升了超短期预测精度。然而该方法仅采用单一分解策略，对风电功率中多尺度耦合特征解析能力不足。

针对上述问题，本文构建了一种集成数据清洗填补、混合分解与深度学习的超短期风电功率预测框架。首先采用分段三次埃尔米特插值(piecewise cubic Hermite interpolating polynomial, PCHIP)对原始功率序列缺失值进行填补，并通过最小协方差行列式(minimum covariance determinant, MCD)与马氏距离协同检测异常点，继而利用随机森林(random forest, RF)回归对异常值实施物理一致性的修复，以确保输入数据的完整性与可靠性。随后利用最大互信息系数(maximal information coefficient, MIC)对风速、风向等气象因子与功率输出的关联度进行定量评估，挑选出与功率高度相关的关键变量构建环境变量特征矩阵；并采用跳变模态分解(jump plus AM-FM mode decomposition, JMD)与 VMD 相结合对功率序列进行混合分解，以获得在不同频段上互补的模态成分并生成多尺度功率特征矩阵。随后功率特征矩阵与环境特征矩阵输入 Dual-MTSA 模型，Mamba 模块通过选择性状态空间机制，以线性时间复杂度高效捕捉序列的长程依赖与关键动态，其隐状态更新方式既能保持全局时序一致性，又能突出突发扰动的特征响应。TKAN 模块借助 B 样条核函数对多阶特征进行平滑插值，同时依托门控机制实现非线性信息的逐步记忆与融合，既能精准刻画高频振荡，又能稳健提炼低频趋势。最后自注意力模块(self-attention, SA)对 TKAN 模块两路融合后的输出进行全局重标定，通过注意力权重自适应地聚焦于更具预测价值的时空特征，并在全连接映射层生成最终的超短期风电功率预测值。实验结果表明本方法在保证数据物理一致性与特征丰富性的基础上，显著提升了多步短期预测的鲁棒性与精度。

1 数据预处理

本文采用 2021 年中国国家电网主办的可再生能源发电量预测竞赛数据[16]。装机容量为 99 MW，将 2020 年 3 月 1 日至 2020 年 5 月 31 日的春季实际风电采集数据作为实验样本数据，样本间隔时间为 15 min。

1.1 数据缺失值清洗与异常值填补

风电功率数据在实际采集过程中，常因传感器故障或传输中断而出现缺失值或异常值[17]。若直接删除或采用简单填补策略，往往会破坏数据原有分

布特性，进而影响后续模型的预测精度。针对这一问题，本文提出了一种基于功率物理连续性特征的三阶段混合插补与异常处理框架，用于高效地清洗与修复风电功率数据。首先针对缺失值采用 PCHIP 进行初步填补。与传统插值方法相比，PCHIP 通过对插值曲线导数施加严格约束，可在保持数据局部单调性的同时，有效避免过冲或振荡现象，生成符合物理规律的功率值。随后利用 MCD 估计数据协方差矩阵，结合马氏距离对潜在异常值进行识别。MCD 方法通过筛选高密度数据子集，能够在协方差估计中抵御异常值的干扰，避免出现掩蔽效应或淹没效应。鉴于直接删除异常值会导致数据连续性中断，不利于模型训练，本文采用 RF 回归对识别出的异常值进行补估。RF 模型通过集成多棵决策树，能够刻画风速、温度与功率之间的非线性关系，为异常点生成合理的功率预测值，从而保证数据的连续性与物理一致性。

1.2　环境因素相关性分析

在风电功率预测的多维输入场景中，合理筛选输入特征对于提升模型的泛化能力和计算效率具有决定性作用。虽然风速、风向、气温、气压、湿度等气象因素与风电出力存在潜在关联，但并非所有变量都能为预测模型带来增益，部分气象因子对功率变化的解释力较弱，甚至可能引入噪声，降低模型性能。将所有观测指标纳入训练，不仅会令模型结构臃肿，还容易引入冗余或噪声，从而导致过拟合并拖慢训练速度。为此本文采用 MIC 与 Spearman 对风电功率与气象因素之间的信息相关性进行线性与非线性评估。两个指标经过归一化步骤保证了该指标的取值范围限定在[0,1]之内，因此当两变量存在高度依赖关系时，其值趋近于 1；而若二者相互独立或关联较弱，则该值则接近于 0。经 MIC 和 Spearman 分析后的特征两两相关性热力图如附录图 A1 和 A2 所示。表 1 为功率与不同气象因素间的相关性指标。

1.3　归一化

鉴于风电功率及其相关气象因子的数值量纲相差悬殊，若不加处理便直接投入模型训练，容易引入尺度偏差，进而削弱算法的收敛性与预测精度。为此本文针对风电功率以及气象因素特征，统一采用 min-max 归一化策略，将所有变量的取值线性映射至[0,1]区间。

1.4　混合分解

风电功率序列通常呈现显著的非线性与非平稳特征，其振荡幅度和周期结构在多重时间尺度上交织分布，且易受气象条件变化、地理环境及机组运行状态等多重因素的共同作用。单一分解方法往往难以同时应对上述复杂性与多样性，从而限制了其在长短期预报中的精度表现。

表 1　风电功率与各气象因素的相关系数

Table 1　Correlation coefficients between wind power output and various meteorological factors

气象因素	特征编号	MIC	Spearman
10 m 高处风速	1	0.797	0.847
10 m 高处风向	2	0.286	0.322
30 m 高处风速	3	0.844	0.911
30 m 高处风向	4	0.329	0.355
50 m 高处风速	5	0.861	0.927
50 m 高处风向	6	0.346	0.359
风机轮毂处风速	7	0.867	0.931
风机轮毂处风向	8	0.350	0.352
温度	9	0.068	0.098
大气压强	10	0.165	0.189
湿度	11	0.317	0.298

JMD 是一种基于变分优化框架的信号分解方法，可从非平稳信号中同步提取调幅-调频(amplitude and frequency-modulated，AM-FM)窄带振荡成分与宽带跳变成分，通过设计联合优化目标函数，利用交替方向乘子法(alternating direction method of multipliers, ADMM)迭代求解，克服传统方法在跳跃-振荡混合信号中的局限性[18]。VMD 是一种基于变分优化的信号分解方法，通过构造带约束的变分模型，将信号分解为多个具有稀疏频带的模态分量[19]。本文将 JMD 分解和 VMD 分解相结合，充分挖掘原始功率数据的内在结构与频谱特征，为后续的预测模型提供更具针对性的输入分量。

1.5　双分支输入矩阵设置

在构造双分支输入矩阵时将风电功率序列经过 JMD 和 VMD 两种方法分解后，提取出子序列拼接构成功率特征矩阵 $\boldsymbol{F}_{\mathrm{p}}$。用同样的方法对经 MIC 和 Spearman 筛选后的气象变量序列进行横向合并，形成环境变量特征矩阵 $\boldsymbol{F}_{\mathrm{v}}$。针对不同预测步长情况设定 $s\in\{1,6,12\}$ 与固定输入窗口长度为 36，采用滑动窗口技术对矩阵 $\boldsymbol{F}_{\mathrm{p}}$ 与 $\boldsymbol{F}_{\mathrm{v}}$ 进行划分。图 1 中以 $s=12$ 为例，示意了风电功率与气象变量两条支路分别切分并汇合为模型输入的全过程；其余 $s=1$ 和 $s=6$ 的情况下，仅需调整滑动步幅，其余划分与拼接流程保持一致。双输入分支的构造方式，能同时捕获功率序列与环境变量的时序特征，灵活适配不同预测

跨度，为后续模型训练提供了多尺度、多模态的输入支持。

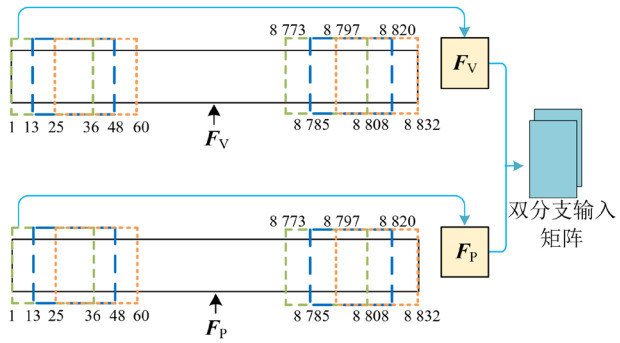


图 1 构造双分支输入矩阵

Fig.1 Constructing double branch input matrix

2 双分支输入的深度学习模型

2.1 Mamba 模型

Mamba 是一种基于选择性状态空间模型的新型序列建模框架[20]，设计核心在于将传统状态空间理论与动态参数化机制相结合，以高效建模长序列的全局依赖关系[21]。模型结构如图 2 所示，Mamba 通过引入数据驱动的选择性机制，将状态转移矩阵 $\boldsymbol{A}$、输入投影矩阵 $\boldsymbol{B}$ 及时间步长 Δ 动态调整为输入序列的函数，实现对关键上下文信息的自适应聚焦；同时借助零阶保持离散化方法，将连续状态空间方程转化为递归形式，并通过并行扫描算法优化计算流程，将复杂度降低至线性级别。为进一步提升硬件效率，Mamba 采用结构化全局卷积核替代传统自注意力机制，利用隐状态序列的全局卷积视角显式捕获长程依赖，同时结合硬件感知设计加速训练与推理。

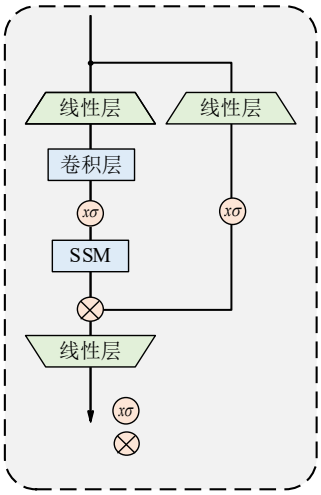


图 2 Mamba 模型结构

Fig.2 Mamba model structure

2.2 TKAN 模型

TKAN 是一种创新的神经网络框架，设计灵感源于 Kolmogorov-Arnold Networks 定理 (KAN)与门控循环机制的有机结合[22]。TKAN 模型结构如图 3 所示，TKAN 通过循环 Kolmogorov-Arnold Networks 定理(recurrent Kolmogorov-Arnold networks，RKAN)可学习的 B 样条核函数对时序数据进行非线性映射，利用函数组合的方式近似复杂动力学规律；同时借鉴 LSTM 中的遗忘门、输入门与输出门结构，对记忆单元执行精细的信息保留与更新控制，以防止梯度消失或爆炸。TKAN 采用多层核激活模块的级联融合策略，将各层输出拼接并统一输入至后续计算单元，从而在不同时间尺度上并行建模数据的周期性、趋势性及突变特征。该架构在提升长序列预测精度与稳定性的同时，为处理风电功率等高波动、高噪声信号提供了有效而灵活的手段。

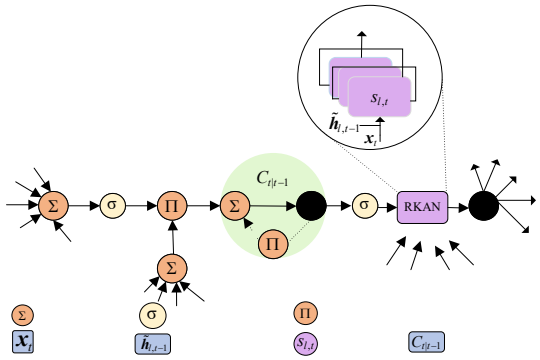


图 3 TKAN 模型结构

Fig.3 TKAN model structure

2.3 Dual-MTSA 深度学习模型

为克服传统自注意力机制在超短期风电功率预测中对局部连贯性与全局动态捕捉能力的不足，并全面挖掘功率序列与多维气象变量在不同时间尺度上的耦合特征，本文构建了 Dual-MTSA 双分支预测模型(见图 4)。两条分支分别以风电功率矩阵和环境特征矩阵为输入，分支序列先后经过 Mamba 与 TKAN 级联处理，其中 Mamba 模块以线性时间复杂度有针对性地保留或遗忘状态，从而同时捕获长程依赖与突发扰动；紧接其后的 TKAN 模块则利用 B 样条核函数对多阶特征进行平滑插值，构建高频与低频子空间，并通过门控机制在子空间间选择性地融合与记忆，精细提炼波动与趋势信号。当两条分支完成上述特征抽取后，其输出向量沿时间维度拼接，融合了功率与环境的多层次表达；拼接后

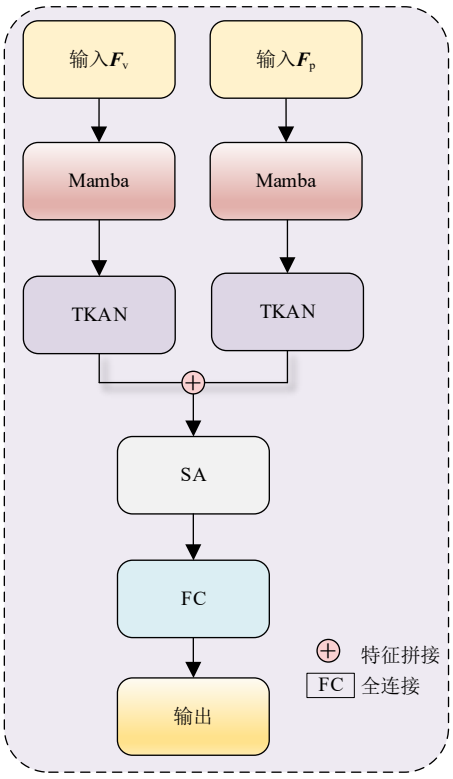


图 4　深度学习模型结构

Fig.4　Deep learning model architecture

的特征再输入 SA 模块，通过注意力权重对融合特征进行全局重标定，自适应地突出最具预测价值的时空模式；最终重标定后的特征向量经全连接层映射，生成超短期风电功率预测结果。本设计通过底层状态空间建模、多尺度耦合插值和全局注意力校准的 3 级结构，有效兼顾了局部连贯性和全局动态性，显著提升了多步超短期预测的鲁棒性与精度。将先前构建的功率特征矩阵 $\boldsymbol{F}_{\mathrm{p}}$ 和环境变量特征矩阵 $\boldsymbol{F}_{\mathrm{v}}$ 分别输入至 Mamba 模块。以下给出功率分支的计算流程，环境变量分支的处理过程与此相同：

$$\begin{cases}\boldsymbol{h}'(t)=\boldsymbol{A}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{h}(t)+\boldsymbol{B}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{F}_{\mathrm{p}}\\ \boldsymbol{y}(t)=\boldsymbol{C}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{h}(t)\end{cases}\tag{1}$$

$$\begin{cases}\boldsymbol{h}_t=\overline{\boldsymbol{A}}_t\boldsymbol{h}_{t-1}+\overline{\boldsymbol{B}}_t\boldsymbol{F}_{\mathrm{p}}\\ \overline{\boldsymbol{A}}_t=\mathrm{e}^{\boldsymbol{\Delta}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{A}(\boldsymbol{F}_{\mathrm{p}})}\\ \overline{\boldsymbol{B}}_t=(\boldsymbol{\Delta}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{A}(\boldsymbol{F}_{\mathrm{p}}))^{-1}(\mathrm{e}^{\boldsymbol{\Delta}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{A}(\boldsymbol{F}_{\mathrm{p}})}-\boldsymbol{I})\boldsymbol{B}(\boldsymbol{F}_{\mathrm{p}})\\ \boldsymbol{f}_{\mathrm{pma}}=\boldsymbol{C}(\boldsymbol{F}_{\mathrm{p}})\boldsymbol{h}_t\end{cases}\tag{2}$$

式中：$\boldsymbol{h}(t)$ 为隐藏状态矩阵；$\boldsymbol{h}'(t)$ 为隐藏状态矩阵求一阶导；$\boldsymbol{A}(\boldsymbol{F}_{\mathrm{p}})$ 为动态状态转移矩阵；$\boldsymbol{B}(\boldsymbol{F}_{\mathrm{p}})$ 为动态输入投影矩阵；$\boldsymbol{C}(\boldsymbol{F}_{\mathrm{p}})$ 为动态输出投影矩阵；$\boldsymbol{y}(t)$ 为隐藏状态的线性投影；$\boldsymbol{\Delta}(\boldsymbol{F}_{\mathrm{p}})$ 为动态时间步长；$\boldsymbol{h}_t$ 为第 t 时刻的隐藏状态；$\overline{\boldsymbol{A}}_n$、$\overline{\boldsymbol{B}}_n$ 分别为离散化后的动态参数；$\boldsymbol{I}$ 为单位矩阵；$\boldsymbol{f}_{\mathrm{pma}}$ 为功率分支输出，同理 $\boldsymbol{f}_{\mathrm{vma}}$ 为环境分支输出。

随后将 Mamba 模块输出的隐状态向量输入 TKAN 模块，以利用 B 样条函数与门控机制进一步丰富模型的多尺度时序表达能力。以下给出功率分支在 TKAN 中的具体计算流程，环境变量分支的处理过程与此相同：

$$\phi_l(\boldsymbol{f}_{\mathrm{pma}})=\sum_{m=1}^{n_{\mathrm{base}}}(c_{l,j,n,m}\cdot B_m^{(k)}(\boldsymbol{f}_{\mathrm{pma}}))\tag{3}$$

$$\begin{cases}\boldsymbol{i}_t=\sigma\left(\boldsymbol{W}_{\mathrm{i}}\cdot\phi_{\mathrm{in}}(\boldsymbol{f}_{\mathrm{pma}})+\boldsymbol{U}_{\mathrm{i}}\cdot\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{i}}\right)\\ \boldsymbol{f}_t=\sigma\left(\boldsymbol{W}_{\mathrm{f}}\cdot\phi_{\mathrm{forget}}(\boldsymbol{f}_{\mathrm{pma}})+\boldsymbol{U}_{\mathrm{f}}\cdot\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{f}}\right)\\ \tilde{\boldsymbol{c}}_t=\tanh\left(\boldsymbol{W}_{\mathrm{c}}\cdot\phi_{\mathrm{cell}}(\boldsymbol{f}_{\mathrm{pma}})+\boldsymbol{U}_{\mathrm{c}}\cdot\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{c}}\right)\end{cases}\tag{4}$$

$$\begin{cases}\boldsymbol{o}_t=\sigma\left(\boldsymbol{W}_{\mathrm{o}}\cdot\phi_{\mathrm{out}}(\boldsymbol{f}_{\mathrm{pma}})+\boldsymbol{U}_{\mathrm{o}}\cdot\boldsymbol{h}_{t-1}+\boldsymbol{b}_{\mathrm{o}}\right)\\ \boldsymbol{c}_t=\boldsymbol{f}_{\mathrm{pma}}\odot\boldsymbol{c}_{t-1}+\boldsymbol{i}_t\odot\tilde{\boldsymbol{c}}_t\\ \boldsymbol{h}_t=\boldsymbol{o}_t\odot\tanh(\boldsymbol{c}_t)\end{cases}\tag{5}$$

$$\begin{cases}\boldsymbol{w}_l=\mathrm{Softmax}\left(\boldsymbol{W}_{\mathrm{g}}\cdot\boldsymbol{h}_t+\boldsymbol{b}_{\mathrm{g}}\right)\\ \boldsymbol{g}=\sum_{l=1}^{L}\boldsymbol{w}_l\odot\phi_l(\boldsymbol{f}_{\mathrm{pma}})\\ \boldsymbol{f}_{\mathrm{pt}}=\mathrm{LayerNorm}\left(\boldsymbol{h}_t+\mathrm{Agg}(\boldsymbol{h}_t)+\mathrm{Proj}(\boldsymbol{g})\right)\end{cases}\tag{6}$$

式中：$\phi_l(\cdot)$ 为第 l 个 KAN 子层的 B 样条激活函数；n_{base} 为基函数数量；k 为 B 样条阶数；$c_{l,j,n,m}$ 为可学习系数，其中 j 为第 j 个输入神经元、n 为第 n 个输出神经元；$B_m^{(k)}$ 为 B 样条基函数；σ 为 Sigmoid 激活函数；$\boldsymbol{c}_t$ 为 t 时刻记忆单元状态；$\tilde{\boldsymbol{c}}_t$ 为 t 时刻候选记忆单元状态向量；$\phi_{\mathrm{in}}(\cdot)$、$\phi_{\mathrm{forget}}(\cdot)$、$\phi_{\mathrm{cell}}(\cdot)$、$\phi_{\mathrm{out}}(\cdot)$ 分别为输入门、遗忘门、细胞状态、输出门的非线性变换函数；$\boldsymbol{i}_t$、$\boldsymbol{f}_t$、$\boldsymbol{o}_t$ 分别为输入门、遗忘门、记忆单元状态；$\boldsymbol{W}_{\mathrm{i}}$、$\boldsymbol{U}_{\mathrm{i}}$、$\boldsymbol{b}_{\mathrm{i}}$ 为输入门的权重矩阵与偏置；$\boldsymbol{W}_{\mathrm{f}}$、$\boldsymbol{U}_{\mathrm{f}}$、$\boldsymbol{b}_{\mathrm{f}}$ 为遗忘门的权重矩阵与偏置；$\boldsymbol{W}_{\mathrm{c}}$、$\boldsymbol{U}_{\mathrm{c}}$、$\boldsymbol{b}_{\mathrm{c}}$ 为记忆单元的权重矩阵与偏置；$\boldsymbol{W}_{\mathrm{o}}$、$\boldsymbol{b}_{\mathrm{o}}$ 为输出门的权重矩阵与偏置；$\boldsymbol{W}_{\mathrm{g}}$ 为门控权重矩阵；$\boldsymbol{b}_{\mathrm{g}}$ 为门控偏置项；$\boldsymbol{w}_l$ 为第 l 个 KAN 子层的门控权重向量；L 为 KAN 总子层数；$\boldsymbol{g}$ 为全局门控融合特征；$\odot$ 为逐元素乘积；Agg 为全局特征整合；Proj 为全局特征整合；Softmax 为归一化激活函数；LayerNorm 为层归一化；$\boldsymbol{f}_{\mathrm{pt}}$ 为功率分支输出，同理 $\boldsymbol{f}_{\mathrm{vt}}$ 为环境分支输出。

完成两条分支的多尺度特征提取后，将它们在时间维度上拼接融合，再送入自注意力模块，通过可学习的注意力权重对融合后的多变量特征进行全局重标定，其具体计算过程见式(8)。

$$\boldsymbol{f}_{\mathrm{T}}=\boldsymbol{f}_{\mathrm{pt}}\oplus\boldsymbol{f}_{\mathrm{vt}}\tag{7}$$

$$\begin{cases}(\boldsymbol{Q}_i,\boldsymbol{K}_i,\boldsymbol{V}_i)=\Sigma(\boldsymbol{f}_{\mathrm{T}}\boldsymbol{W}_{\mathrm{q},i},\boldsymbol{f}_{\mathrm{T}}\boldsymbol{W}_{\mathrm{k},i},\boldsymbol{f}_{\mathrm{T}}\boldsymbol{W}_{\mathrm{v},i})\\ \boldsymbol{N}_{\mathrm{head}_i}=\sum\mathrm{Softmax}(\dfrac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}})\boldsymbol{V}_i,\quad i=0,1,2,3\\ \boldsymbol{f}_{\mathrm{SA}}=[N_{\mathrm{head}_1},\cdots,N_{\mathrm{head}_4}]\boldsymbol{W}_0\end{cases}\tag{8}$$

式中：$\boldsymbol{f}_{\mathrm{T}}$ 为两条支路特征融合后的特征向量；⊕ 为对特征进行拼接；$\boldsymbol{W}_{\mathrm{q},i}$、$\boldsymbol{W}_{\mathrm{k},i}$、$\boldsymbol{W}_{\mathrm{v},i}$ 分别为自注意力机制第 i 组对应的查询、键、值权重矩阵；$\boldsymbol{Q}_i$、$\boldsymbol{K}_i$、$\boldsymbol{V}_i$ 分别为第 i 组对应的查询、键、值矩阵；d_{k} 为归一化参数；$\boldsymbol{N}_{\mathrm{head}_i}$ 为第 i 组注意力权重；$\boldsymbol{W}_0$ 为初始化矩阵；$\boldsymbol{f}_{\mathrm{SA}}$ 为注意力权重融合后的特征输出。最后经过全连接层处理，以生成最终的功率预测值 $\boldsymbol{Y}$。

2.4 模型评价指标

为全面量化所提模型在超短期风电功率预测任务中的表现，本文选取了 4 种广泛应用于回归评价的统计度量指标：平均绝对误差(e_{MAE})、平均绝对百分比误差(e_{MAPE})以及均方根误差(e_{RMSE})和决定系数(R^2)。

3 实验流程及框架

为深入挖掘风电功率序列在超短期多步预测中的内在规律并提升模型的泛化鲁棒性，本文设计了系统化的实验流程，涵盖从数据预处理到模型训练与性能评估的完整流程。具体的预测实验流程如图 5 所示。在数据预处理阶段首先采用 PCHIP 对原始功率序列中的缺失样本进行初步填补；紧接着利用 MCD 与马氏距离联合识别异常点，并以 RF 回归对异常值进行修复，以保证数据连续性与物理一致性。随后基于 MIC 与 Spearman 定量评估风速与风向等气象因素对功率输出的影响，通过阈值敏感性分析实验选择最佳阈值，详见附录 B。最终选取

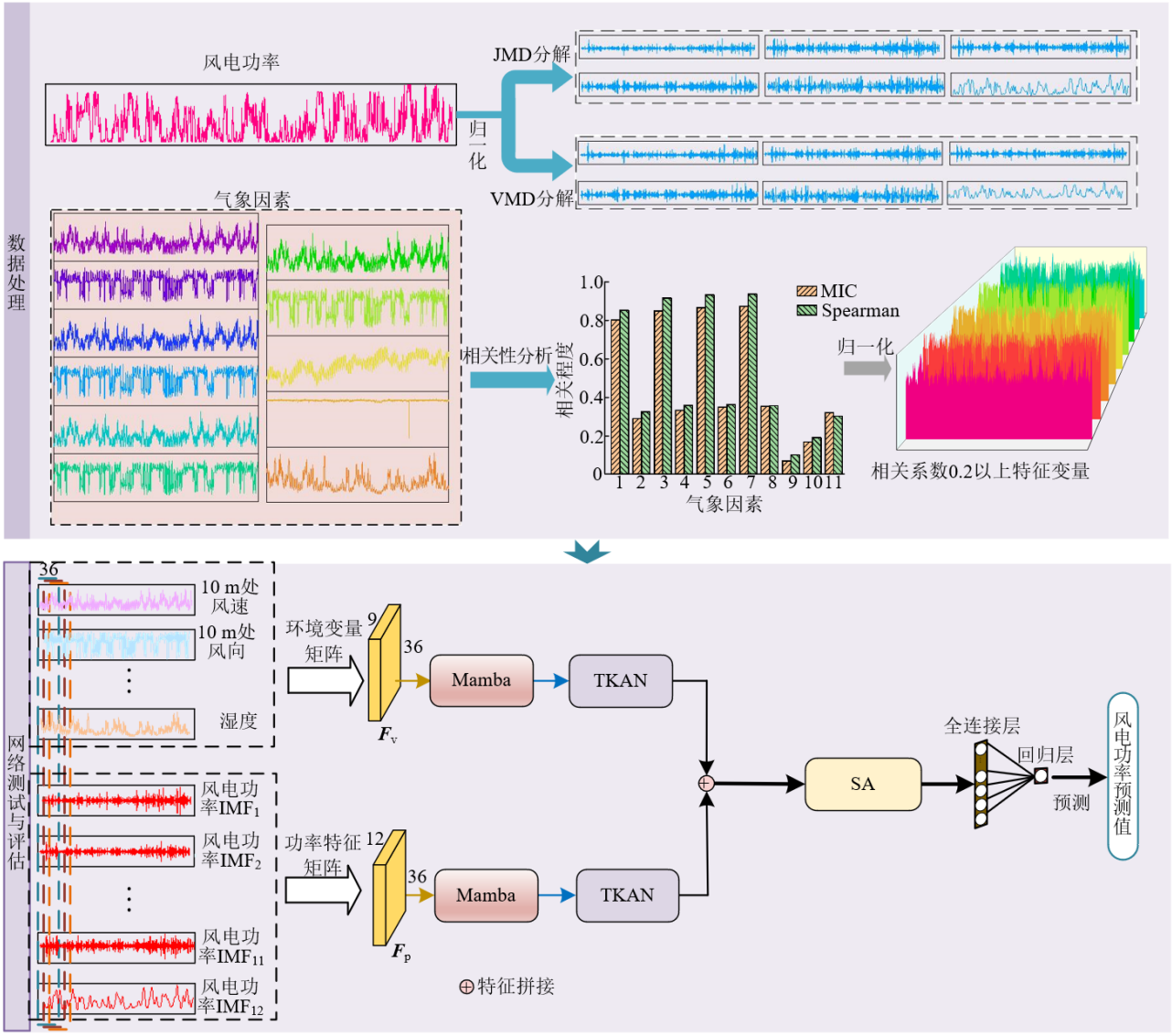


图 5 预测实验流程框架图

Fig.5 Framework of the forecasting experiment pipeline

10 m、30 m、50 m 处的风速与风向，以及机舱处风速、风向、湿度共 9 项高相关特征构建环境变量特征矩阵；为消除量纲差异引起的偏倚，所有环境变量与功率数据均经 min–max 归一化至[0,1]区间内。针对数据波动的多尺度特性，先后应用 JMD 与 VMD 对归一化功率序列进行分解，各自产生 6 条本征模函数(intrinsic mode function，IMF)子序列，最终将这 12 条 IMF 子序列拼接成功率特征矩阵，为模型功率分支提供丰富的时频信息，风电功率分解结果如附录图 A3 所示。将功率与环境变量特征矩阵分别按步长 $s\in\{1,6,12\}$ 与固定窗口长度为 36 切分窗口滑动时，依次提取连续 36 个数据点的特征作为输入，输出则对应未来 s 个时间步的功率值。

在模型训练与评估阶段，将得到功率特征矩阵与环境变量特征矩阵，分别输入两条 Mamba 模型，每条支路通过选择性状态空间机制动态更新隐状态，紧接着由 TKAN 模块基于 B 样条核插值与门控融合提炼多尺度时序特征。完成单分支特征抽取后，将两路输出沿时间维度拼接，送入自注意力层进行全局加权重标定，从而在所有时刻与变量维度上自适应提升关键信息的表征强度，最后再通过多层全连接网络映射为超短期功率预测值。在训练过程中，以均方误差 MSE 为损失函数。综合 e_{MAE}、e_{MAPE}、e_{RMSE} 和 R^2 误差指标，验证所提方法在超短期风电功率预测任务中的鲁棒性与精度优势。

4　实验结果分析

本文设计了 4 个实验系统评估所提模型的有效性与泛化能力。不同信号分解实验验证数据预处理效果；消融实验逐步去除核心组件以识别其对整体性能的贡献；不同季节对比实验突出模型在不同气象条件与功率波动模式下的泛化性；模型对比实验通过误差指标证明本文方法的优势；实验选取连续 92 天的风电功率与气象观测数据，按照 8:1:1 的比例划分为训练集、验证集和测试集，并对每组实验进行 10 次取平均值以降低偶然性，确保结果的可靠性。每组预测实验将误差表现最优的方法在表格中以加粗形式予以标注。模型超参数的最优配置方案通过对测试集预测性能的综合评估确定，详见附录 C。

4.1　分解对比实验

为了验证所提取的混合分解策略在预测性能方面较单一分解方法的优越性，首先分别采用 JMD 和 VMD 对原始序列进行预处理，并将得到的各模态分量送入双分支 Dual-MTSA 网络进行预测。为考察该混合方案的有效性，引入集合经验模态分解(ensemble empirical mode decomposition, EEMD)作为第 3 方对照，与单一分解和混合分解相结合，展开对比实验。在 1 步、6 步及 12 步这 3 种预测步长下，运用多维度误差评估指标对各方法的结果见表 2，图 6 为分解对比实验预测曲线，分析不同分解策略在预测中的稳定性和性能差异。

由表 2 可见，本文所提混合分解策略在 1 步、6 步及 12 步预测任务中，均显著优于单一分解方法及对比混合方案。1 步预测中，本文方法 e_{MAE} 仅为 0.704，相较单一最优模型 JMD 降低 34.1%，较 JMD＋EEMD 降低 19.2%；e_{MAPE} 为 10.665，较 JMD 减少 20.2%，较 JMD＋EEMD 下降 27.8%；e_{RMSE} 由 JMD 的 1.642 和 JMD＋EEMD 的 1.286 降至 1.106；R^2 提升至 0.998，高于 JMD＋EEMD 的 0.997。6 步预测中，e_{MAE} 为 1.871 较 JMD 降低 28.0%，较 JMD＋EEMD 降低 27.3%；e_{MAPE}、e_{RMSE} 相比 12 步时分别下降 16.4%与 27.2%，R^2 达 0.987。12 步预测时，e_{MAE} 仅为 2.537，较 JMD 和 JMD＋EEMD 分别降低 25.2%与 27.2%；e_{MAPE} 由单一最优 JMD 的 41.161 和 JMD＋EEMD 的 39.758 降至 31.252；e_{RMSE} 由 JMD 的 5.111 和 JMD+EEMD 的 4.844 降至 3.896；R^2 为 0.975。上述结果证明，JMD 对局部波动的精准刻画与 VMD 对非平稳信号的自适应分离在混合框架下实现了互补协同，有效抑制了模态混叠与误差累积，显著提升了多步风电功率预测的准确性与鲁棒性。

表 2　分解对比实验预测误差对比

Table 2　Comparison of prediction errors for decomposition comparison experiments

模型	e_{MAE}/MW			e_{MAPE}/%			e_{RMSE}/MW			R^2		
	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步
EEMD	2.314	3.963	5.471	32.821	54.366	78.999	3.217	5.646	7.668	0.982	0.947	0.903
VMD	1.564	2.747	3.547	20.419	29.926	47.000	2.362	4.068	5.093	0.990	0.972	0.957
JMD	1.069	2.599	3.393	13.363	33.546	41.161	1.642	3.861	5.111	0.995	0.975	0.957
JMD+EEMD	0.871	2.577	3.483	14.784	27.900	39.758	1.286	3.390	4.844	0.997	0.975	0.961
本文方法	**0.704**	**1.871**	**2.537**	**10.665**	**26.112**	**31.252**	**1.106**	**2.836**	**3.896**	**0.998**	**0.987**	**0.975**

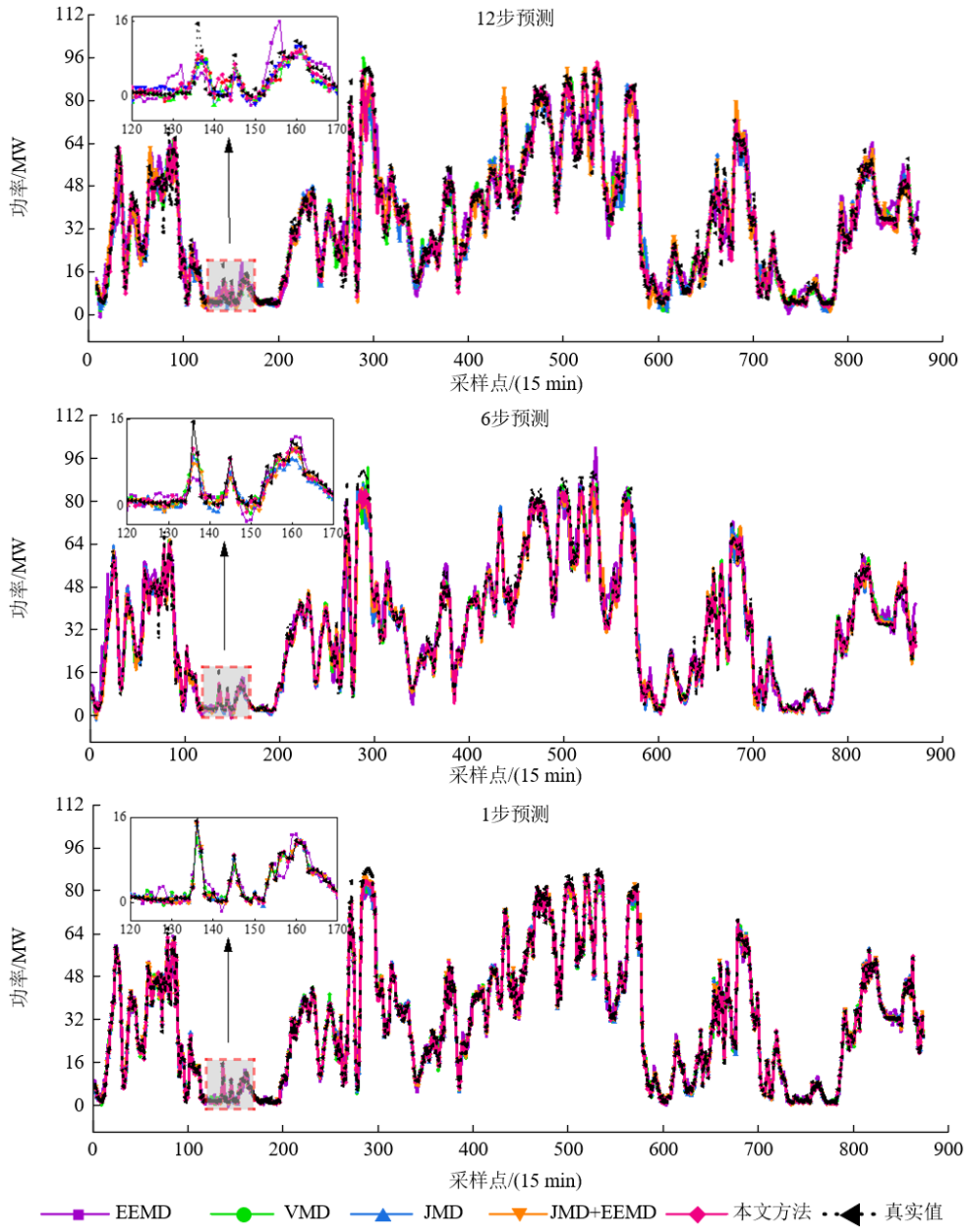


图 6 分解对比实验预测曲线

Fig.6 Prediction curves for decomposition comparison experiments

4.2 模型消融实验

为深入检验本文模型合理性与各模块对预测性能的贡献以及各模块的参数规模，本文设计了系统化的消融实验，逐步引入关键模块并剖析其内在关联。首先以仅包含双分支 Mamba 模块 Dual-M 为基准模型，验证 Mamba 在特征提炼方面的独立作用。并且为了进一步探讨 TKAN 的多尺度时序特征提取能力，设计仅包含双分支 TKAN 模块得到 Dual-T 模型。随后再将 TKAN 模块集成于 Dual-M 之后，构建 Dual-MT，验证 TKAN 在多频信号捕获与门控融合层面对现有模型性能的增益。最后通过全量整合 Mamba、TKAN 与 SA 模块，形成完整的 Dual-MTSA 模型，从底层特征提炼到全局依赖建模，全面验证各模块间的互补性与协同提升效果。并且引入模型参数规模(Params)、浮点运算量(FLOPs)及步长为 12 情况下程序完整运行时间，展示了收益与开销对比。表 3 为消融实验预测误差与各模块参数量对比。

由表 3 可知，所提 Dual-MTSA 模型在 1、6、12 步长预测任务中均显著优于消融对比模型，验证了模块协同作用的有效性。单独采用 Mamba 模块时 Params、FLOPs 和运行时间都是最少的，计算开

表 3　模型消融实验预测误差与参数对比

Table 3　Model ablation experiment prediction error vs. parameters

模型	e_{MAE}/MW			e_{MAPE}/%			e_{RMSE}/MW			R^2			Params/10^6	FLOPs/10^9	运行时间/min
	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步			
Dual-M	1.753	2.428	2.985	18.091	29.667	44.042	2.371	3.507	4.707	0.990	0.979	0.963	0.047	0.385	3.125
Dual-T	0.846	2.573	3.250	15.492	29.871	45.910	1.112	3.781	4.979	0.997	0.976	0.959	0.111	0.952	7.301
Dual-MT	1.005	2.030	2.839	16.779	27.604	40.169	1.246	3.051	4.188	0.997	0.984	0.971	0.122	1.180	8.011
本文方法	**0.704**	**1.871**	**2.537**	**10.665**	**26.112**	**31.252**	**1.106**	**2.836**	**3.896**	**0.998**	**0.987**	**0.975**	0.127	0.858	6.256

销低，预测误差 e_{MAE} 在 1.753 ~2.985 之间波动，但在多步预测和多尺度特征建模方面能力有限。相比之下单独采用 TKAN 在一步预测上取得最低误差 e_{MAE} 为 0.846，优于 Mamba 的 1.753，表达能力上带来提升。但 Params、FLOPs 和运行时间在单一模块中是最大的，表明 TKAN 对多尺度局部特征提炼的收益是以较大算力和时间代价换来的，并且随着预测步长的增加，TKAN 的误差增长更为显著。Dual-MT 通过融合两者优势，相较于仅使用 TKAN 模块时 Params 小幅增加，从 0.111 增加至 0.122 增幅仅 9.9%，FLOPs 增幅为 24.0%，运行时间增幅为 9.7%，表明级联带来的计算复杂度主要源自模块间的时序变换与中间表示的额外前向计算，而非单纯的参数堆积。并且 Dual-MT 使得各步长预测误差均低于任一单模块模型，尤其在 12 步预测中提升尤为明显，e_{MAPE} 为 40.169 比单一使用 Mamba 和 TKAN 模型分别降低了 8.8%和 12.5%；e_{RMSE} 为 4.188 比单一使用 Mamba 和 TKAN 模型分别降低了 11.0%和 17.3%，说明级联结构能以可接受的额外开销换取跨时间尺度的建模收益。引入 SA 后 Dual-MTSA 模型相比 Dual-MT 模型 Params 从 0.122 增至 0.127 仅微增 4.1%，却实现了 FLOPs 从 1.180 降至 0.858 降幅为 27.3%，运行时间从 8.011 分钟降至 6.256 分钟降幅为 21.9%，表明模块间的协同与实现细节优化能够在几乎不增加参数量的前提下改善整体推理效率。并且在 12 步长预测任务中 e_{MAPE} 仅为 31.252，相较于 Dual-M、Dual-T 和 Dual-MT 分别降低了 29.0%、31.9%、22.2%。综合来看 Dual-MTSA 在保证多步预测精度和鲁棒性的同时，实现了计算效率与性能的最佳折中，彰显了全局重标定与多源特征融合的互补优势，从而在超短期风电功率预测任务中达成更高的鲁棒性与精度。

4.3　不同季节与极端工况及不同缺失率对比实验

为检验本文 Dual-MTSA 模型在不同季节场景、极端工况及不同缺失率下的泛化能力，本文选取 2019 年 12 月 1 日至 2020 年 11 月 30 日之间数据，分为冬、春、夏、秋 4 个季节，其中 2020 年 3 月 1 日至 2020 年 5 月 31 日的春季数据为主实验，其中选取未填补缺失值以及填补一半缺失值数据作为对比实验。剩余冬、夏、秋 3 个季节的数据集进行对比实验，并在全年数据中选取最具代表性的极高风速、风速突变情况进行实验，各不同情境数据在相同模型相同超参数配置和训练流程下各自独立测试，保证结果的稳定性。通过横向比较不同情境下预测的性能表现，全面评估模型在不同气候和时间特征下的适应性与可靠性。实验误差结果见表 4，不同季节拟合度见图 7。

由表 4 可见，本文方法在春、夏、秋、冬 4 季均能实现稳定且高精度的多步风电功率预测，春季风场波动适中，1 步至 12 步预测 e_{MAE} 仅为

表 4　不同季节和工况实验预测误差对比

Table 4　Comparison of experimental prediction errors in different seasons and operating conditions

不同季节和工况	e_{MAE}/MW			e_{MAPE}/%			e_{RMSE}/MW			R^2		
	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步
冬季	0.565	1.838	2.339	23.817	63.174	77.938	0.850	2.716	3.490	0.998	0.986	0.977
春季	0.704	1.871	2.537	10.665	26.112	31.252	1.106	2.836	3.896	0.998	0.987	0.975
夏季	0.641	1.368	2.578	7.539	20.164	26.912	0.921	1.997	4.041	0.998	0.992	0.967
秋季	0.426	0.476	0.998	39.140	42.929	87.835	0.591	0.690-	1.494	0.996	0.995	0.977
极高风速	0.404	1.394	1.899	10.025	23.274	34.571	0.533	1.934	2.595	0.997	0.960	0.939
风速突变	0.896	2.735	4.387	32.996	46.893	65.979	1.139	3.633	5.921	0.996	0.959	0.919
未填补缺失值	2.976	3.997	4.623	33.543	48.315	85.212	3.978	5.618	6.059	0.961	0.942	0.921
填补一半缺失值	1.256	2.688	3.391	17.730	45.233	46.953	1.949	3.914	4.973	0.993	0.975	0.959

0.704~2.537，e_{MAPE} 维持在 10.665~31.252，R^2 高达 0.975~0.998，体现了模型对中频波动的精确捕捉。夏季风速持续且功率平稳，尽管 12 步预测 e_{MAE} 略升至 2.578，e_{MAPE} 为 26.912，但 R^2 仍保持 0.967 以上，显示出对长序列平稳态的鲁棒性。秋季模型在 1 步预测时 e_{MAE} 误差最低为 0.426，但 12 步预测时 e_{MAPE} 为 87.835 有所增大，说明在极端波动期需更多短期重构；冬季低温下的断续强风令 12 步 e_{RMSE} 达 3.490，但 R^2 依然维持在 0.977 附近，验证了对复杂非平稳工况的适应能力。模型在极高风速条件下表现出较好的短期预测能力，1 步 e_{MAE} 仅为 0.404 误差集中且稳定，尽管 6 步和 12 步时略有上升，但整体仍维持在可接受范围内，表明模型能够较好地捕捉高风速下的趋势变化。风速突变情况长步预测误差相对增加，12 步 e_{MAE} 和 e_{RMSE} 分别为 4.387 和 5.921 但即便在如此复杂的工况下，模型依然保持了较强的拟合能力和趋势捕捉能力，显示出相对于传统方法的稳健性优势。未填补缺失值时预测误差明显偏高，1 步 e_{MAE} 达到 2.976，12 步 e_{RMSE} 高达 6.059，R^2 下降至 0.921，表明缺失数据会削弱模型的预测能力。而填补一半缺失值后误差指标明显改善，1 步至 12 步 e_{MAE} 介于 1.256 至 3.391 之间，R^2 显著提升至 0.959~0.993，显示合理的数据填补策略能够有效缓解缺失值对模型的负面影响。由图 7 散点图看 4 个季节，预测值与真实值分布高度吻合，借助 JMD 与 VMD 的互补分解与双分支深度学习模型的协同作用，模型在多样化季节特征中均展现出兼顾短期精度与长期稳定的优秀性能。

4.4 模型对比实验

为系统评估所提双分支深度学习模型在风电功率预测任务中的性能，本文首先对数据进行预处理，除了与常见的双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)、卷积神经网络(convolutional neural network, CNN)、

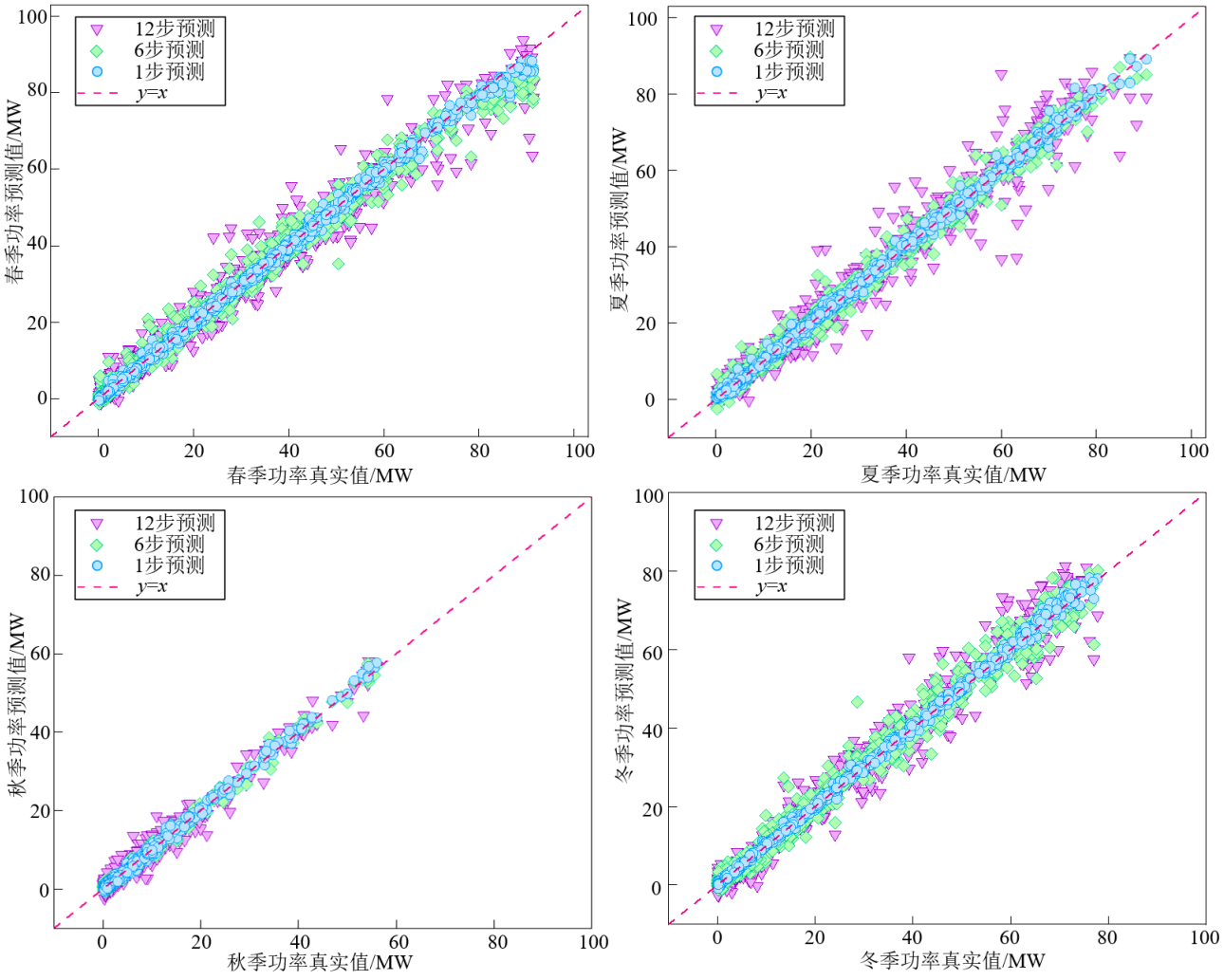


图 7 不同季节拟合对比

Fig.7 Comparison of fits for different seasons

TCN 等常见轻量级方法进行对比，还与近年来国内外高水平文献中提出的组合模型进行对比，分别是文献[23]的 VMD-LSTM-ARMA 模型、文献[24]的 AMLSTM-FCN 模型和文献[25]的 MTPNet 模型；该部分实验所有模型使用了相同的数据预处理、输入窗口、预测步长、训练轮次、优化器以及训练集、验证集、测试集划分，每种模型均进行 10 次独立试验。表 5 为各模型在关键误差指标上的差异。表 A1 为本文模型与对比模型在不同步长下的 e_{MAE} 误差指标配对样本 t 检验。图 8 为不同误差在 3 种步长下的带箱体的小提琴图，验证本文方法在不同步长下的预测性能。

表 5　不同模型的预测误差对比

Table 5　Comparison of prediction errors of different models

模型	模型编号	e_{MAE}/MW			e_{MAPE}/%			e_{RMSE}/MW			R^2		
		1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步	1 步	6 步	12 步
CNN	A	1.499	4.787	7.586	45.069	80.040	122.546	2.087	6.542	10.271	0.982	0.929	0.826
BiLSTM	B	1.206	3.903	6.903	10.573	34.330	70.352	2.262	5.621	9.848	0.981	0.947	0.840
TCN	C	1.382	3.306	5.995	18.841	38.536	90.356	2.141	4.708	9.342	0.992	0.963	0.865
VMD-LSTM-ARMA	D	2.393	3.966	6.576	35.328	65.500	119.844	3.362	5.508	9.762	0.981	0.950	0.843
AMLSTM-FCN	E	1.198	3.883	5.301	15.072	52.217	67.823	1.743	5.433	7.586	0.994	0.951	0.905
MTPNet	F	**0.671**	2.283	3.538	**9.492**	29.811	47.842	**0.967**	3.405	4.999	**0.998**	0.980	0.958
本文方法	G	0.704	**1.871**	**2.537**	10.665	**26.112**	**31.252**	1.106	**2.836**	**3.896**	**0.998**	**0.987**	**0.975**

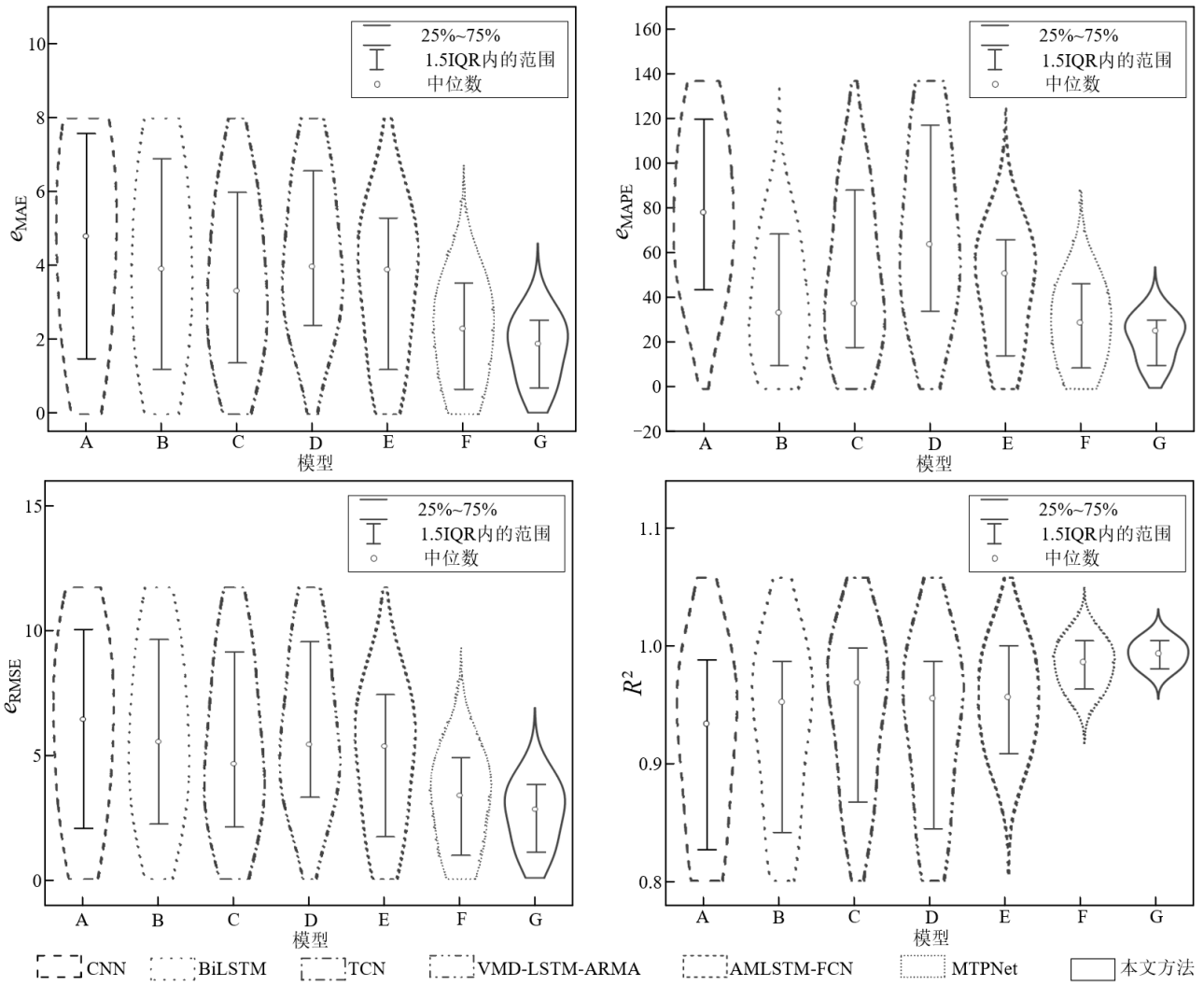


图 8　不同误差类型带箱体的小提琴图对比

Fig.8　Comparison of violin diagrams with different error types and case attachments

由表 5 可见，本文方法在多项误差指标分布上均展现出优异特征。e_{MAE} 与 e_{MAPE} 分布的小提琴箱线图中，本文方法及 MTPNet 均呈现出明显的集中性与低值化特征，小提琴形态狭窄，表明误差取值的离散性较小，预测结果更为稳定。其中本文方法的中位数整体处于更低区间，进一步体现了在平均误差水平上的优势。相比之下，CNN、BiLSTM、TCN、VMD-LSTM-ARMA 和 AMLSTM-FCN 等模型的误差分布则更为宽泛，离散性显著偏大，且中位数普遍偏高，说明其在不同预测步长下的稳定性与整体误差控制方面存在不足。e_{RMSE} 分布可以看出本文方法表现出高度的集中性与低值化优势，其分布窄而集中，中位数位于较低水平，表明在误差的控制上具有良好的稳定性和可靠性；而其他模型均呈现出分布宽泛、离散性强以及中位数偏高的特征，反映出其在误差稳定性与整体水平上的不足。图 8 为各模型重复实验下的预测性能分布小提琴图，内嵌箱线的短横线对应 25%~75%分位区间，覆盖半数实验结果；工字形竖线表示 1.5IQR 内的正常波动范围，超出该范围可判定为离群误差；中间圆点为多次实验结果的中位数；外围轮廓为核密度曲线，反映结果在不同取值的分布密集程度。本文方法的 R^2 分布小提琴图呈现出狭窄且集中于高值区间的特征，中位数明显高于对比模型，显示出较强的拟合效果。其他模型的 R^2 分布则相对宽泛且整体偏低，说明其不同步长下预测效果不及本文方法。由表 A1 e_{MAE} 误差指标配对样本 t 检验可知，本文方法在 12 步长和 6 步长下与其他模型显著不同，显著优于现有方法，仅在 1 步长下与 MTPNet 差异不显著，说明在极短期条件下，方法性能接近，预测难度较低，表明本文在不同步长下超短期风电功率预测中的有效性和鲁棒性。

5 结论与展望

提出基于数据清洗填补、双分支输入的 Dual-MTSA 风电功率预测方法，并通过系列消融实验与同类模型对比验证，得出以下结论：

1）采用 JMD 与 VMD 组合分解对原始功率数据进行预处理，成功提取了多层次模态分量，显著缓解了原始序列的非线性和间歇性特征，使后续深度模型输入更加平滑稳定；

2）在双分支 Mamba 后引入 TKAN 模块，使模型对关键时刻信息的响应更加灵敏；SA 注意力机制进一步通过自适应加权协调多通道特征，从而在短、中、长时域内均保持预测精度；

3）与 VMD-LSTM-ARMA、AMLSTM-FCN 和 MTPNet 组合预测模型在同一数据集和相同评价指标下进行对比，分别计算本文方法相对 3 类对比模型的误差降幅后取均值，本文方法在 1、6、12 步预测中分别实现 e_{MAE}、e_{MAPE}、e_{RMSE} 平均降低 28%以上，e_{MAPE} 最低达 10.665，R^2 稳定不低于 0.97，进一步证明了各子模块的互补的协同增益效应；

4）未来可引入地理位置和设备运行状态等多源异构信息进行多模态融合；并在更多分布式风电场场景下开展大规模验证，以检验方法在分布式场景下和复杂地形下的推广性。

附录见本刊网络版(http://hve.epri.sgcc.com.cn)。

参考文献 References

[1] 廖扬，蔡帜，邵佳扬，等. 多区互联综合能源系统分散协调低碳经济调度[J]. 高电压技术，2023，49(1)：138-146.
LIAO Yang, CAI Zhi, SHAO Jiayang, et al. Decentralized coordinated low-carbon economic dispatch method for multi-regional interconnected integrated energy system[J]. High Voltage Engineering, 2023, 49(1): 138-146.

[2] 王永生，李海龙，关世杰，等. 基于变换域分析和 XGBoost 算法的超短期风电功率预测模型[J]. 高电压技术，2024，50(9)：3860-3870.
WANG Yongsheng, LI Hailong, GUAN Shijie, et al. Ultra-short-term wind power prediction model based on transform domain analysis and XGBoost algorithm[J]. High Voltage Engineering, 2024, 50(9): 3860-3870.

[3] 吕云龙，胡琴，熊俊杰，等. 基于时空注意力卷积模型的超短期风电功率预测[J]. 电网技术，2024，48(5)：2064-2073.
LÜ Yunlong, HU Qin, XIONG Junjie, et al. Ultra-short-term wind power prediction based on spatiotemporal attention convolution model[J]. Power System Technology, 2024, 48(5): 2064-2073.

[4] 刘雅婷，杨明，于一潇，等. 基于多场景敏感气象因子优选及小样本学习与扩充的转折性天气日前风电功率预测[J]. 高电压技术，2023，49(7)：2972-2982.
LIU Yating, YANG Ming, YU Yixiao, et al. Transitional-weather-considered day-ahead wind power forecasting based on multi-scene sensitive meteorological factor optimization and few-shot learning[J]. High Voltage Engineering, 2023, 49(7): 2972-2982.

[5] 郑颖颖，李鑫，陈延旭，等. 基于 Stacking 多模型融合的极端天气短期风电功率预测方法[J]. 高电压技术，2024，50(9)：3871-3882.
ZHENG Yingying, LI Xin, CHEN Yanxu, et al. Short-term wind power forecasting method in extreme weather based on Stacking multi-model fusion[J]. High Voltage Engineering, 2024, 50(9): 3871-3882.

[6] 钟吴君，李培强，涂春鸣. 基于尾流关联的动态超图风电功率超短期预测方法[J]. 中国电机工程学报，2025，45(12)：4693-4706.
ZHONG Wujun, LI Peiqiang, TU Chunming. Dynamic hypergraph wind power ultra short term prediction method based on wake correlation[J]. Proceedings of the CSEE, 2025, 45(12): 4693-4706.

[7] 杨茂，王达，王小海，等. 基于数据物理混合驱动的超短期风电功

率预测模型[J]. 高电压技术，2024，50(11)：5132-5141.
YANG Mao, WANG Da, WANG Xiaohai, et al. Ultra-short term wind power prediction method based on data physics hybrid driven model[J]. High Voltage Engineering, 2024, 50(11): 5132-5141.

[8] 周玮，钟佳成，孙辉，等. 基于隐马尔可夫模型的日内风电功率预测误差区间滚动估计[J]. 电力系统自动化，2018，42(21)：90-95，184.
ZHOU Wei, ZHONG Jiacheng, SUN Hui, et al. An interval rolling estimation method for daily wind power forecast errors based on hidden Markov Model[J]. Automation of Electric Power Systems, 2018, 42(21): 90-95, 184.

[9] LIU X L, LIN Z, FENG Z M. Short-term offshore wind speed forecast by seasonal ARIMA-a comparison against GRU and LSTM[J]. Energy, 2021, 227: 120492.

[10] 吉兴全，赵国航，叶平峰，等. 基于 QMD-HBiGRU 的短期光伏功率预测方法[J]. 高电压技术，2024，50(9)：3850-3859.
JI Xingquan, ZHAO Guohang, YE Pingfeng, et al. Short-term PV forecasting method based on the QMD-HBiGRU model[J]. High Voltage Engineering, 2024, 50(9): 3850-3859.

[11] 何锦华，刘洋，朱誉，等. 融合物理信息的 TPA-LSTM 网络应用于台风条件下海上风电场风速多步预测[J]. 电网技术，2023，47(10)：4152-4160.
HE Jinhua, LIU Yang, ZHU Yu, et al. Physics-informed TPA-LSTM network applied to multi-step wind speed prediction of offshore wind farms under typhoon conditions[J]. Power System Technology, 2023, 47(10): 4152-4160.

[12] 臧海祥，赵勇凯，张越，等. 基于低风速功率修正和损失函数改进的超短期风电功率预测[J]. 电力系统自动化，2024，48(7)：248-257.
ZANG Haixiang, ZHAO Yongkai, ZHANG Yue, et al. Ultra-short-term wind power prediction based on power correction under low wind speed and improved loss function[J]. Automation of Electric Power Systems, 2024, 48(7): 248-257.

[13] BOMMIDI B S, TEEPARTHI K, KOSANA V. Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function[J]. Energy, 2023, 265: 126383.

[14] 潘鹏程，张帅，刘晖，等. 基于 CAWR-LSTM-TRF 模型的超短期风电功率预测[J]. 太阳能学报，2026，47(3)：452-461.
PAN Pengcheng, ZHANG Shuai, LIU Hui, et al. Ultra-short-term wind power forecasting based on CAWR-LSTM-TRF mode[J]. Acta Energiae Solaris Sinica, 2026, 47(3): 452-461.

[15] LI Q, REN X Y, ZHANG F, et al. A novel ultra-short-term wind power forecasting method based on TCN and Informer models[J]. Computers and Electrical Engineering, 2024, 120: 109632.

[16] CHEN Y B, XU J J. Solar and wind power data from the Chinese State Grid Renewable Energy Generation Forecasting Competition[J]. Scientific Data, 2022, 9(1): 577.

[17] 李丹，唐建，缪书唯，等. 考虑时序特征缺失值动态插补的超短期风电功率预测[J]. 中国电机工程学报，2025，45(17)：6790-6803.
LI Dan, TANG Jian, MIAO Shuwei, et al. Ultra-short-term wind power prediction considering dynamic interpolation of missing values of time series features[J]. Proceedings of the CSEE, 2025, 45(17): 6790-6803.

[18] NAZARI M, KORSHØJ A R, REHMAN N U. Jump plus AM-FM mode decomposition[J]. IEEE Transactions on Signal Processing, 2025, 73: 1081-1093.

[19] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.

[20] GU A, DAO T. Mamba: Linear-time sequence modeling with selective state spaces[Z]. arXiv preprint arXiv: 2312.00752, 2024.

[21] 潘苗绒，王燚. 基于 ConvNeXt-Mamba 的双编码器图像伪造检测[J]. 计算机工程与应用，2026，62(5)：336-345.
PAN Miaorong, WANG Yi. Double encoder image forgery detection based on ConvNeXt and Mamba[J]. Computer Engineering and Applications, 2026, 62(5): 336-345.

[22] GENET R, INZIRILLO H. TKAN: Temporal kolmogorov-arnold networks[Z]. arXiv preprint arXiv: 2405.07344, 2025.

[23] LI C, CAO D S, ZHAO Z T, et al. Forecast for wind power at ultra-short-term based on a composite model[J]. Energy Reports, 2024, 12: 4076-4082.

[24] 张昊立，张菁，倪建辉，等. 引入注意力机制的 LSTM-FCN 海上风电功率预测[J]. 太阳能学报，2024，45(6)：444-450.
ZHANG Haoli, ZHANG Jing, NI Jianhui, et al. LSTM-FCN offshore wind power forecasting with introduction of attention mechanism[J]. Acta Energiae Solaris Sinica, 2024, 45(6): 444-450.

[25] 徐武，范鑫豪，沈智方，等. 多尺度特征提取的 Transformer 短期风电功率预测[J]. 太阳能学报，2025，46(2)：640-648.
XU Wu, FAN Xinhao, SHEN Zhifang, et al. Short-term wind power prediction using Transformer with multi-scale feature extraction[J]. Acta Energiae Solaris Sinica, 2025, 46(2): 640-648.



陈世轲
2001—，男，硕士生
主要从事人工智能在电力系统中的应用研究
E-mail: Chenshk001:@163.com

CHEN Shike



毕贵红(通信作者)
1968—，男，博士，教授，硕导
主要从事新能源系统建模、优化与智能数据分析方面的研究工作
E-mail: KM_BGH@163.com

BI Guihong
Ph.D., Professor
Corresponding author

收稿日期　2025-07-28　在线出版日期　2025-12-26　　编辑　陈　蔓

附录 A

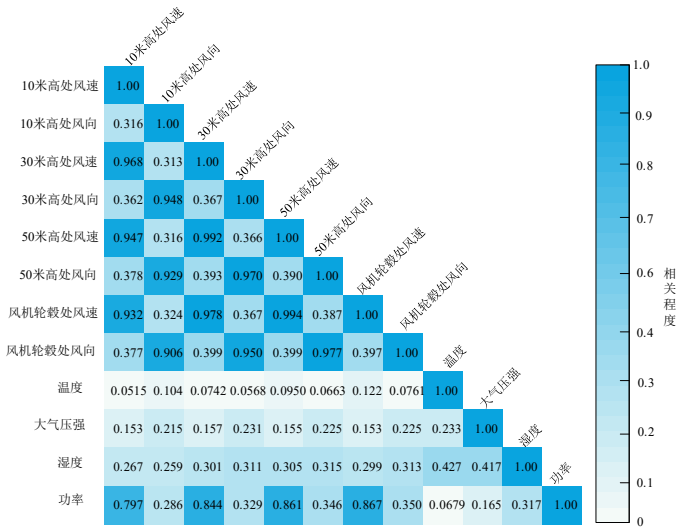


图 A1　MIC 分析后的特征两两相关性热力图

Fig.A1　Heatmap of pairwise correlations among features following MIC analysis

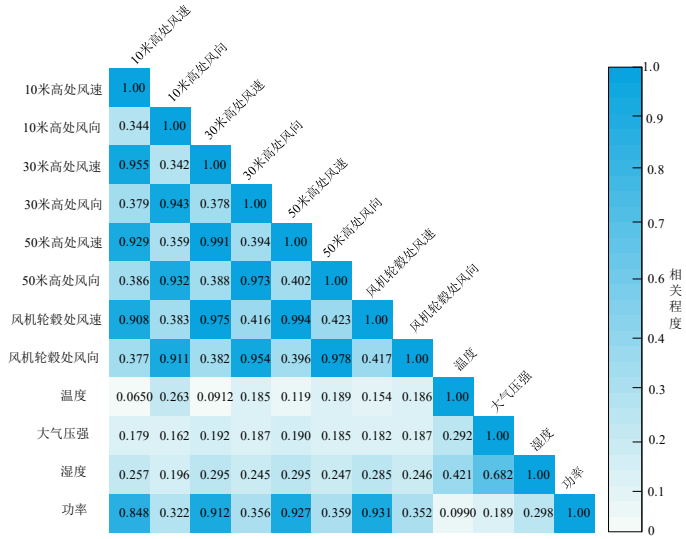


图 A2　Spearman 分析后的特征两两相关性热力图

Fig.A2　Heatmap of pairwise correlations among features following Spearman analysis

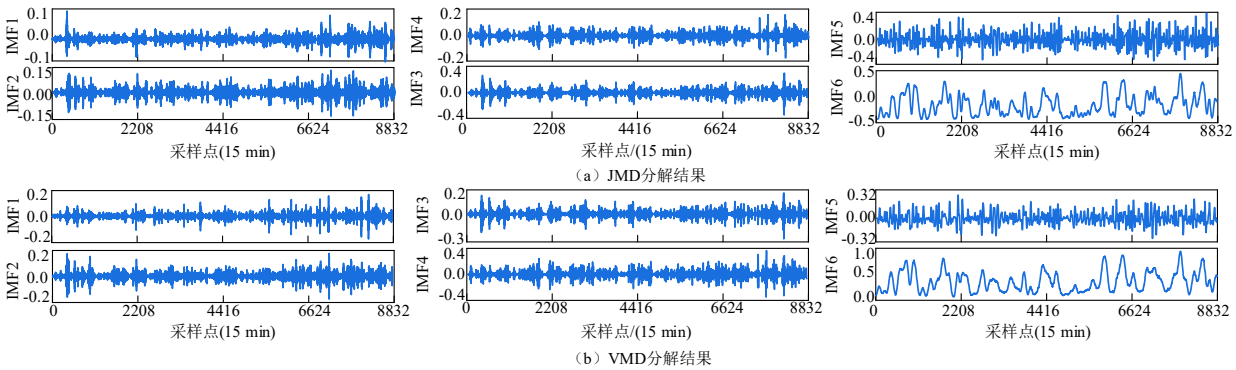


图 A3　风电功率分解结果

Fig.A3　Wind power decomposition results

表 A1　本文模型与对比模型 e_{MAE} 误差在不同步长的配对样本 t 检验结果

Table A1　Results of paired sample t-tests comparing e_{MAE} errors between the proposed model and the baseline model across different stride lengths

模型	P 值		
	1 步	6 步	12 步
CNN	<0.000 1	<0.000 1	<0.000 1
BiLSTM	<0.000 1	<0.000 1	<0.000 1
TCN	<0.000 1	<0.000 1	<0.000 1
VMD-LSTM-ARMA	<0.000 1	<0.000 1	<0.000 1
AMLSTM-FCN	<0.000 1	<0.000 1	<0.000 1
MTPNet	0.6500 9	<0.000 1	<0.000 1

附录 B

基于 Mamba 模块在步长为 12 的多步预测中，开展阈值敏感性实验以确定输入气象变量的最优筛选阈值。实验通过将候选特征按若干阈值区间进行划分并选取相应子集输入模型，比较不同阈值下的预测性能变化，以评估阈值对模型泛化能力的影响。预测性能使用若干常用误差指标进行量化，并以这些误差随阈值变化的趋势来判定最佳阈值区间。不同阈值区间的预测误差对比见表 B1，阈值选取与风电功率预测平均误差分析见图 B1。

表 B1　不同阈值区间的预测误差对比

Table B1　Comparison of prediction errors across different threshold intervals

阈值区间	e_{MAE}/MW	e_{RMSE}/MW	e_{MAE}/MW	e_{RMSE}/MW
	MIC		Spearman	
0~1	3.789	5.391	3.801	5.403
0.2~1	3.089	4.245	3.102	4.249
0.4~1	3.506	4.989	3.511	5.016
0.6~1	3.506	4.989	3.511	5.016
0.8~1	3.518	5.017	3.520	5.115

由表 B1 与图 B1 可见，当阈值区间设定为 0~1 时模型预测误差整体较大；当阈值范围收缩至 0.4~1、0.6~1 和 0.8~1 时，预测性能有所改善误差处于中等水平；而在阈值区间为 0.2~1 时，模型表现最佳预测误差达到最低值。由此可确定最优阈值选择区间为 0.2~1，以保证模型在风电功率预测任务中的精度与稳定性。

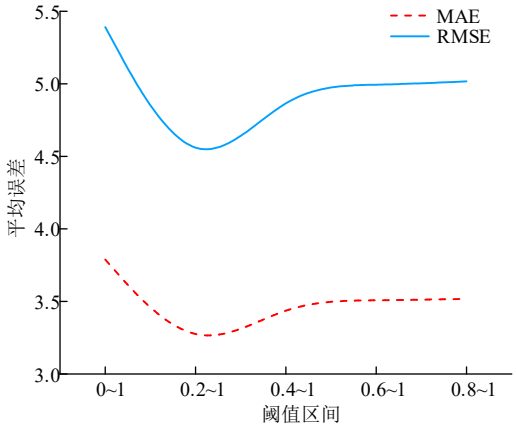


图 B1　阈值选取与风电功率预测平均误差分析

Fig.B1　Threshold selection and average error analysis in wind power forecasting

附录 C

实验平台基于 CPU(Intel Core i5-13400F 2.5 GHz)、RAM(32 GB)、GPU(RTX 3060 12G)等硬件条件下进行。深度学习模型的搭建在 PyCharm2025.1 环境下进行，训练过程中选用 Adam 优化算法以提升收敛效率与稳定性。模型超参数的最优配置由 Dual-MTSA 在步长为 12 的多步预测任务中基于测试集预测性能的综合评估确定。所有实验均采用经过统一预处理的春季数据作为输入，确保结果的可比性与可靠性，其具体配置与实验结果详见表 C1。

表 C1　不同超参数实验误差对比

Table C1　Error comparison across different hyperparameter configurations

编号	训练轮次	批次大小	学习率	运行时间/min	e_{MAE}/MW	e_{MAPE}/%	e_{RMSE}/MW	R^2
					Dual-MTSA			
1	1 000	64	0.001	17.527	2.777	34.382	4.117	0.972
2	1 000	128	0.001	10.166	2.733	38.398	4.157	0.971
3	1 000	256	0.001	4.171	3.110	38.977	4.465	0.967
4	**1 500**	**256**	**0.001**	**6.256**	**2.537**	**31.252**	**3.896**	**0.975**
5	1 500	512	0.001	4.571	3.059	41.480	4.332	0.969
6	1 500	256	0.000 1	7.012	4.493	58.901	6.244	0.935

由表 C1 共 6 组实验表明，学习率设为 0.001 时模型能稳定收敛且效果最好，在相同学习率批次大小为 256 下，将训练轮次从 1 000 提升到 1 500 使 e_{MAE} 从 3.110 降至 2.537，e_{RMSE} 从 4.465 降至 3.896，R^2 提升至 0.975，同时运行时间保持在可接受范围，因此可认定为性能上的最优配置。对比不同批次大小可见，增大批次大小有助于缩短单次训

练的运行时间，但过大的批次大小会带来泛化性能下降;而过小的批次大小虽在部分指标上接近最优，但显著增加运行时间，在工程化部署中并不经济。学习率过小反映为收敛变慢且性能退化，误差显著上升。