DOI: 10.19666/j.rlfd.202511051

基于融合全局残差 Mamba 的短期电价预测方法研究

谢云明[1]，许鹏辉[2]，吴 涛[3]，李 杰[1]，姜 鹏[1]，黄 涛[1]，
黄晓明[1]，陈 伟[1]，王 慧[3]，赖 菲[3]
（1. 中国华能集团有限公司山东分公司，山东 济南 250000；
2. 华能济南黄台发电有限公司，山东 济南 250000；
3. 西安热工研究院有限公司，陕西 西安 710054）

［摘 要］【目的】针对传统时间序列预测模型难以同时捕捉跨尺度非线性波动及其长程依赖特征，导致电力现货市场短期电价预测精度不高的问题，提出一种基于融合全局残差 Mamba 的短期电价预测方法。【方法】该方法首先通过移动平均滤波器将电价序列解耦为趋势项和残差项，并嵌入高维特征空间；随后设计包含变量相关性支路与特征交互支路的并行 Mamba 编码器，分别从跨变量耦合和特征交互 2 个维度提取电价序列的局部−全局动态特征；对局部−全局动态特征进行拼接融合，将拼接后的特征引入全局残差学习机制，实现原始异构信息的加性融合，以缓解梯度退化并增强跨尺度特征捕捉能力；最后采用贝叶斯优化算法对模型关键超参数进行寻优。【结果】基于澳大利亚能源市场运营商的实测数据进行 24 h 电价预测验证，结果表明，所提模型的均方根误差相较于基准 Mamba 模型与长短时记忆神经网络 LSTM 模型分别降低了 29.22%和 35.33%，验证了所提方法在提升短期电价预测精度方面的有效性。【结论】消融实验进一步验证了序列分解模块的核心作用以及变量相关性支路在双分支架构中的关键地位，同时全局残差连接的引入显著提升了模型训练的稳定性与特征表达能力。

［关 键 词］Mamba 模型；残差学习；电力现货市场；电价预测；贝叶斯优化

［引用本文格式］谢云明. 许鹏辉，吴涛，等. 基于融合全局残差 Mamba 的短期电价预测方法研究[J]. 热力发电，2026，55(5)：169-177. XIE Yunming, XU Penghui, WU Tao, et al. Study on a short-term electricity price forecasting method based on fused global residual Mamba[J]. Thermal Power Generation, 2026, 55(5): 169-177.

Study on a short-term electricity price forecasting method based on fused global residual Mamba

XIE Yunming[1], XU Penghui[2], WU Tao[3], LI Jie[1], JIANG Peng[1], HUANG Tao[1],
HUANG Xiaoming[1], CHEN Wei[1], WANG Hui[3], LAI Fei[3]
(1. Shandong Branch, China Huaneng Group Co., Ltd., Jinan 250000, China;
2. Huaneng Jinan Huangtai Power Generation Co., Ltd., Jinan 250000, China;
3. Xi'an Thermal Power Research Institute Co., Ltd., Xi'an 710054, China)

Abstract: [Objective] Traditional time-series forecasting methods often struggle to simultaneously capture cross-scale nonlinear fluctuations and long-range temporal dependencies, which leads to limited accuracy in short-term electricity price prediction for spot markets, especially when prices exhibit spikes, volatility clustering, and pronounced non-stationarity. [Methods] To address these challenges, this study proposes a short-term electricity

收 稿 日 期：2025-11-15 修 回 日 期：2026-01-20 接 受 日 期：2026-01-22
基 金 项 目：中国华能集团有限公司总部科技项目（HNKJ23-HF62）
Supported by: Science and Technology Project of China Huaneng Group Co., Ltd. (HNKJ23-HF62)
第一作者简介：谢云明（1976），男，高级工程师，主要研究方向为火力发电生产运行优化，xymwyyx@163.com。
通信作者简介：吴涛（1993），男，硕士，工程师，主要研究方向为电力生产信息化智能化，wutaox@tpri.com.cn。

price forecasting framework based on a fused global-residual Mamba model that combines series decomposition, dual-branch selective state-space encoding, and global residual learning to strengthen representation power and improve training stability. First, a moving average filter is applied to the normalized electricity price sequence to decouple it into a trend component and a residual component, separating relatively stable low-frequency movements from high-frequency stochastic variations. The decomposed sequences are concatenated along the temporal dimension and mapped through a high-dimensional embedding layer to obtain a richer latent representation capable of characterizing complex market dynamics. To better reflect the multi-factor formation mechanism of spot prices, the model also incorporates exogenous variables, such as regional load and weather-related information (e.g., temperature and meteorological conditions). Building on these inputs, a parallel dual-branch Mamba encoder is designed to extract local-to-global dynamic features from complementary perspectives. The variable-correlation branch focuses on learning time-varying interdependencies between electricity prices and exogenous drivers, explicitly modeling cross-variable coupling and market co-movements. In parallel, the feature-interaction branch targets nonlinear transformations and interactions within the embedded feature space; by permuting tensor dimensions so that selective scanning operates along the embedding dimension rather than only along time, it uncovers abstract interaction patterns that conventional temporal scanning may overlook. To integrate heterogeneous information from both branches, their outputs are concatenated and passed to a global residual learning module, which performs additive fusion between the fused representations and the original embedded input. This global residual pathway provides a stable channel for information flow, alleviates gradient degradation in deeper state-space architectures, and enhances the model's ability to capture multi-scale patterns by preserving original signals while enriching them with learned cross-variable and cross-feature dynamics. For robust performance and reduced manual effort, Bayesian optimization is applied under a time-series cross-validation (TS-CV) setting to tune key hyperparameters, while training further adopts learning-rate scheduling and early stopping to improve efficiency and stability. [Result]Experiments on real operational data from the Australian Energy Market Operator (AEMO) for 24-hour-ahead forecasting demonstrate clear performance gains: the proposed Residual Mamba reduces RMSE by 29.22% relative to the baseline Mamba model and by 35.33% relative to LSTM, confirming superior accuracy and robustness. [Conclusion]Ablation results further highlight the essential role of the series decomposition module, the importance of the variable-correlation branch in the dual-branch design, and the effectiveness of global residual connections in stabilizing training and improving feature expression for highly volatile spot-market price series.

Key words: Mamba model; residual learning; electricity spot market; electricity price forecasting; Bayesian optimization

随着全球能源结构变革与电力市场自由化，电价作为核心经济信号的重要性日益凸出。其形成机制受多种复杂因素影响，呈现高度非线性、非平稳性和波动性特征[1-2]。对于电力现货市场，短期电价精准预测是市场参与者制定竞价策略、管理风险的关键依据[3]，也是保障市场稳定运行、提升资源配置效率的必要前提[4]。电价序列的多尺度非线性导致传统统计模型（如自回归（auto regression，AR）模型、自回归整合移动平均（auto-regressive integrated moving average，ARIMA）模型等）对短期电价预测精度较低[5]。

目前研究通常采用统计学习和深度学习进行短期电价预测。深度学习模型由于其处理非线性特征的优势成为电价预测的主流。勾玄等利用长短时记忆（long short-term memory，LSTM）神经网络捕捉电价序列数据的波动性，验证了其对电价预测的有效性[6]。郭鑫炜等结合特征选择与深度学习方法，通过增加市场耦合因素显著提升电价预测精度[7]。龚丹丹等融合变分模态分解（VMD）、改进郊狼算法（ICOA）与双向长短时记忆神经网络（BiLSTM）提出混合模型，精度优于单一 LSTM 模型[8]。但 LSTM 神经网络的循环结构在处理超长序列时存在梯度衰减的问题，因此其无法有效解决长距离依赖问题。

Transformer 能有效捕捉电价的时序演变模式，有利于解决长距离依赖问题，其预测精度明显优于传统 LSTM 神经网络方法[9]。然而，Transformer 的自注意力机制在处理超长序列时面临计算资源挑战，导致其应用范围受限[10-11]。

状态空间模型（state space model，SSM）及 Mamba 模型为降低序列建模线性复杂度提供了新思路[12-13]。Mamba 通过结构化状态空间和选择性扫描机制高效处理长序列，降低计算复杂度[14-15]。林琦淇等基于选择性状态空间构建时间序列预测模型，实现了对正向和逆向时序信息同时建模[16]。张鼎开等设计了双向 Mamba 状态空间模型，结合多

尺度特征提取与双向状态空间建模，进一步扩展了状态空间建模的应用范围[17]。然而，Mamba 模型在深度网络中可能面临梯度消失问题，限制其在深层架构中的特征提取能力[18-19]。因此，Mamba 模型在电价预测领域仍然面临由于深层网络梯度退化而导致的预测精度较低的问题。

为了解决上述问题，本研究提出基于残差 Mamba 模型的短期电价预测模型，以解决电价时间序列中复杂动态特性建模与长序列依赖捕捉的难题。残差连接构建深层网络结构，促进梯度传播，缓解退化问题，实现电价序列数据深层特征学习，增强学习稳定性和捕捉复杂动态（尖峰、季节性、长期依赖），提升预测精度。

1 研究方法

1.1 残差 Mamba 模型整体架构

图 1 为本文构建的电价预测模型。模型由序列分解与嵌入层、双分支并行 Mamba 编码层、全局残差模块与全局残差模块以及输出映射层 4 个核心部分构成，并引入基于贝叶斯优化的超参数自动寻优模块，以确定模型最优配置。模型基于残差学习思想，将全局残差连接嵌入 Mamba 模型中。实现电价数据的局部−全局特征的加性融合，构成残差 Mamba 主干结构。该设计在不增加复杂度的前提下，为双分支选择性扫描与全局残差模块提供了稳定梯度通道，实现了对跨尺度动态的深层捕捉。

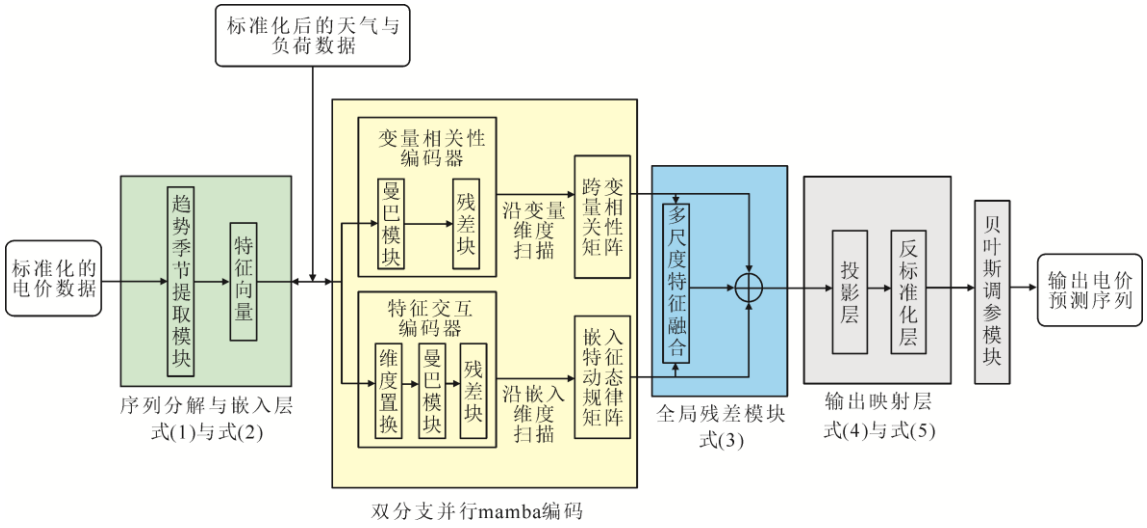


图 1 残差 Mamba 模型架构

Fig.1 Architecture of the residual Mamba model

序列分解与嵌入层基于移动平均滤波方法将原始电价序列分解为趋势项和残差项，并通过嵌入操作将趋势项与残差项融合映射至高维特征空间。双分支并行 Mamba 编码层依托变量相关性支路与特征交互支路的选择性扫描机制，从变量相关性和特征交互编码器 2 条路径同步提取电价序列的局部与全局动态特征。全局残差模块将 2 条编码支路的输出与原始嵌入向量拼接构成残差通路，实现异构信息的加性整合，增强特征表达力并缓解梯度退化。输出映射层通过线性投影将融合特征张量解码为预测时间步的标准化预测序列，并经维度置换输出为规范的时序格式。为获得具有实际物理意义的最终电价，对标准化预测值执行反标准化操作，即利用预处理阶段基于训练数据计算的均值与标准差，将模型输出还原至原始电价量纲。

1.2 序列分解与嵌入层

电价序列分解旨在将标准化后的电价输入序列 $X_{\text{in}}\in\mathbf{R}^{B\times L\times V}$（$B$ 为批量大小，L 为序列长度，V 为特征维度）分解为趋势项和残差项。该过程通过移动平均滤波器（moving average filter，MAF）实现：

$$\boldsymbol{X}_{\text{trend}}=\text{MovingAvg}(\boldsymbol{X}_{\text{in}}) \tag{1}$$

$$\boldsymbol{X}_{\text{res}}=\boldsymbol{X}_{\text{in}}-\boldsymbol{X}_{\text{trend}} \tag{2}$$

分解后，在序列长度维度 L 上进行拼接，形成一个新的中间张量 $\boldsymbol{X}_{\text{concat}}\in\mathbf{R}^{B\times 2L\times V}$，此拼接后的序列随后通过一个高维线性嵌入层，将其映射至一个预设的更高维度（n_{embed}）的特征空间，生成嵌入向量 $\boldsymbol{X}_{\text{emb}}\in\mathbf{R}^{B\times 2L\times n_{\text{embed}}}$。这一过程旨在增强模型对输入序列复杂表示的容量。接着针对变量相关性分支，将该嵌入向量直接供给输入，针对特征交互分支，

将嵌入向量进行维度置换重组，使输入张量的形状变为[B, n_{embed}, $2L$]，因此，模型将不再沿时间或变量维度扫描，而是沿嵌入特征空间的不同维度进行扫描，从而提取高维抽象特征之间的交互动力学特征。

1.3 双分支并行 Mamba 编码层

嵌入向量 $\boldsymbol{X}_{\text{emb}}$ 与外生变量（天气数据、负荷数据）并行输入到 2 个结构特异化 Mamba 编码分支。第 1 分支为变量相关性分支，旨在捕捉不同输入电价与外生变量（负荷、天气数据）之间的动态互相关性。外生变量与电价序列在输入维度 V 上进行拼接，形成统一的输入张量 $\boldsymbol{X} \in \mathbf{R}^{B\times s_{\text{eqlen}}}$。随后被投影至维度为 n_{embed} 的嵌入空间。Mamba 模块的扫描操作沿变量维度 V 进行，其内部采用大小为 $k_{\text{ernel,size}}$ 的卷积核生成选择性扫描参数，并利用维度为 d_{state} 的状态空间矩阵进行序列建模，同时引入残差块，通过一个快捷路径与经过状态空间变换的输出进行逐元素相加。最终输出为跨变量依赖表征矩阵 $\boldsymbol{O}_v \in \mathbf{R}^{(B\times s_{\text{eqlen}}\times n_{\text{embed}})}$，以捕获市场耦合与因果传导等跨变量依赖关系。

第 2 分支为特征交互分支，用于学习高维嵌入特征内部的深度非线性变换与交互。输入为与特征交互分支相同的输入特征向量，经维度置换后得到 $\boldsymbol{X} \in \mathbf{R}^{(B\times V\times s_{\text{eqlen}}\times n_{\text{embed}})}$，使得 Mamba 扫描沿高维嵌入特征维度（$n_{\text{embed}}$）进行。该分支采用与变量相关性分支相同的结构化状态空间配置，包括 $k_{\text{ernel,size}}$、d_{state} 和 dropout 等共享参数，以捕捉价格尖峰、波动集聚等非线性动力学行为，输出为特征交互表示矩阵 $\boldsymbol{O}_{\text{f}} \in \mathbf{R}^{(B\times V\times s_{\text{eqlen}}\times n_{\text{embed}})}$。

2 个分支的输出融合后，经包含 fc_{dropout} 的全连接层映射为长度为 pred_{len} 的预测结果。整个模型的训练过程依赖于对学习率 $\text{learning}_{\text{rate}}$、批次大小 $\text{batch}_{\text{size}}$ 及训练轮数 $\text{train}_{\text{epochs}}$ 等超参数的系统优化。

1.4 全局残差模块

全局残差模块负责对双分支并行编码器所提取的异构特征进行整合，并引入全局残差连接以优化信息流。该设计不仅通过特征拼接保留了 2 个分支的完整信息，增强了表达能力，而且利用加性残差机制确保了原始信息的无衰减传递，缓解了深度网络训练过程中的梯度消失问题[20]。

该模块所拼接的路径共包含 3 条：1）由双分支并行 Mamba 模块融合的保留了跨变量依赖与深度交互信息的统一特征矩阵($\boldsymbol{X}_{\text{temporal}}$+$\boldsymbol{X}_{\text{variate}}$)；2）变量相关性编码器输出的跨变量相关性矩阵 $\boldsymbol{X}_{\text{temporal}}$ 直接前传，构成恒等映射路径；3）特征交互编码器输出的嵌入特征动态规律矩阵 $\boldsymbol{X}_{\text{variate}}$ 直接前传，构成恒等映射路径。最终，经过融合的特征矩阵与原始嵌入向量通过逐元素相加进行合并。其残差拼接的融合公式为：

$$\boldsymbol{X}_{\text{fused}} = \text{Concat}(\boldsymbol{X}_{\text{temporal}}, \boldsymbol{X}_{\text{variate}}, \boldsymbol{X}_{\text{temporal}} + \boldsymbol{X}_{\text{variate}}) \quad (3)$$

1.5 输出映射层

输出映射层将由全局残差模块产生的高维特征张量 $\boldsymbol{X}_{\text{fused}}$ 通过全连接层实现解码并投影到预设的未来预测时间步长 T_{pred} 上，其数学表达式为：

$$\boldsymbol{Y}_{\text{proj}} = \text{GELU}(\boldsymbol{X}_{\text{fused}}\boldsymbol{W}_{\text{map}}^{\text{T}} + \boldsymbol{b}_{\text{map}}) \quad (4)$$

式中：$\boldsymbol{X}_{\text{fused}} \in \mathbf{R}^{B\times V\times 4D}$ 为上一层融合后的特征张量，其中 D 为模型的核心嵌入维度；$\boldsymbol{W}_{\text{map}} \in \mathbf{R}^{T\text{pred}\times 4D}$ 为该线性层的权重矩阵；$\boldsymbol{b}_{\text{map}}$ 为该线性层的偏置向量；GELU(·)为高斯误差线性单元激活函数；$\boldsymbol{Y}_{\text{proj}} \in \mathbf{R}^{B\times V\times T\text{pred}}$，为经过线性投影和激活函数处理后的输出张量。

在执行映射后，为了使张量格式符合时间序列处理的通用约定，模型会进行一次反标准化操作：

$$\hat{y}_{\text{actual}} = \hat{y}_{\text{proj}} \times \sigma + \mu \quad (5)$$

式中：y_{actual} 为模型输出的经过反标准化处理的最终电价；σ 为在数据预处理阶段，基于训练集数据计算得到的该特征的标准差；μ 为在数据预处理阶段，基于训练集数据计算得到的该特征的均值。

1.6 贝叶斯调参模块

为在超参数空间寻找最优模型配置，本研究采用贝叶斯优化方法[21]，通过构建概率代理模型拟合目标函数，利用采集函数指导下一次采样点[22]，目标是找到能最小化交叉验证损失的超参数组合 $\boldsymbol{X}^*$。

$$\boldsymbol{X}^* = \arg\min_{x\in X} f(x) \quad (6)$$

式中：$\boldsymbol{X}$ 为一个包含所有待优化超参数（表 1）的向量；x 为预定义的超参数搜索空间；$f(x)$为目标函数，即使用超参数组合 x 训练模型并在时间序列交叉验证（TS-CV）得到平均验证损失，该平均验证损失计算公式为：

$$f(x) = \frac{1}{K}\sum_{k=1}^{K} \ell_k(\boldsymbol{X}) \quad (7)$$

式中：K 为时间序列交叉验证的总折数；$\ell_k(\boldsymbol{X})$ 为采用超参数组合 x 训练的模型在第 k 个验证折上计算得到的损失值。

表 1 贝叶斯模型优化关键参数指标及寻优范围
Tab.1 Key parameters and optimization ranges for Bayesian model optimization

参数名称	参数类型	寻优范围	说明
learning_rate	连续型	$[10^{-5}, 10^{-2}]$，对数分布	优化器的学习率，采用对数均匀分布采样，在不同数量级上探索
batch_size	离散型	{256, 512, 1 024}	训练过程中每批次包含样本数量，影响梯度下降稳定和收敛速度
seq_len	离散型	{120, 240, 360}	输入序列的长度，决定了模型回看历史数据的时间窗口大小
pred_len	离散型	{24, 48, 72}	模型的预测序列长度，即需要向前预测的时间步数量
n_embed	离散型	{128, 256, 512}	嵌入层的维度，决定了输入特征被映射到的向量空间大小
d_state	离散型	{128, 256, 512}	Mamba 模型的状态空间维度，是影响模型容量和复杂度的核心参数
dropout	连续型	[0.05, 0.50]	应用于模型内部的 Dropout 比率，用于防止过拟合
fc_dropout	连续型	[0.1, 0.5]	应用于全连接层的 Dropout 比率，提供额外的正则化
train_epochs	整型	[1, 20]	模型的训练轮数，决定了模型在训练集上迭代学习的总次数
kernel_size	离散型	{5, 7, 9, 11}	模型中可能存在的卷积层的卷积核大小
revin	离散型	{0, 1}	是否启用可逆实例归一化，1 为启用，0 为禁用

2 实验分析

2.1 数据与预处理

为验证所提模型的有效性，本文选取澳大利亚电力市场（Australian energy market operator，AEMO）某地区的电负荷与电价数据，时间范围为 2025 年 1 月到 2025 年 3 月，采样间隔为 5 min，共 16 993 条数据。同时，外生变量采用相同时间尺度下的区域气温变化与区域天气情况，其分布如图 2 所示。

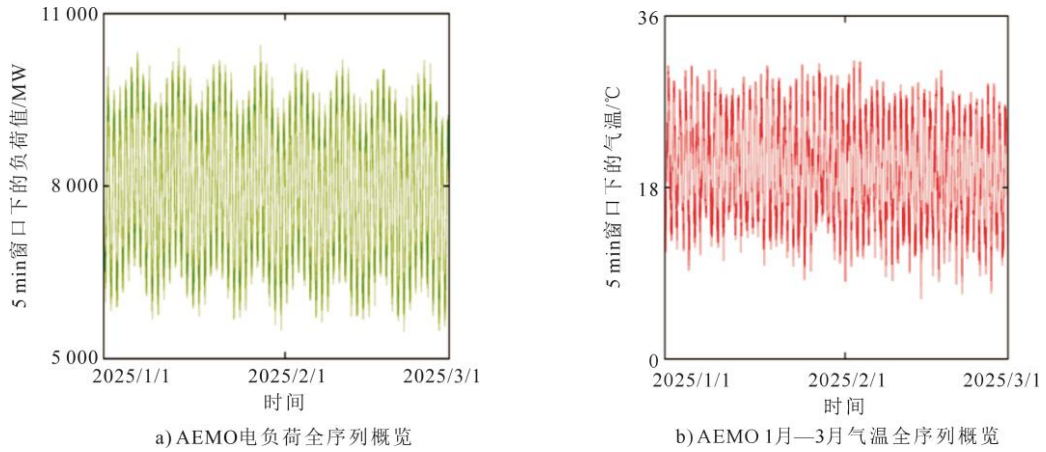


图 2 外生变量全序列概览
Fig.2 Overview diagram of the full sequence of exogenous variables

2.1.1 原始数据及分析

图 3 为 AEMO 参考电价（reference price，RRP）全序列概览。从图 3 中 RRP 时间序列可见，电价数据具有高度非平稳性和波动性。序列通常围绕基准值波动，但周期性出现极端尖峰，包括正负值。这种非线性行为源于电力市场供需失衡、竞价策略和输电阻塞等复杂因素。图 4 为参考电价全序列数据分布直方图。由图 4 可见，电价分布呈现多峰形态，主峰集中在 80 单位附近，同时存在多个次级峰值。此外，分布呈现右偏和长尾特性，意味着高电价事件虽然发生频率较低，但其数值幅度极大。因此，RRP 序列是一个集高度非线性、非平稳性、多重季节性、长记忆性与短期强相关性于一体的复杂时间序列。

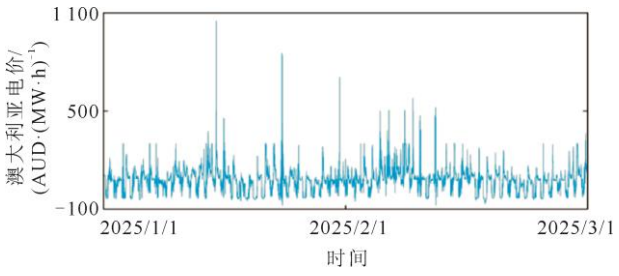


图 3 AEMO 参考电价全序列概览
Fig.3 Overview of the full series of AEMO reference electricity prices

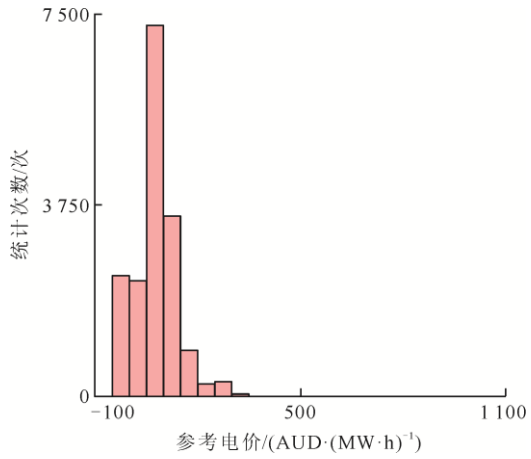


图 4 AEMO 参考电价全序列数据分布直方图
Fig.4 Histogram of the full series data distribution of AEMO reference electricity prices

2.1.2 数据预处理及划分

本文以历史 RRP 序列为输入，未来 RRP 为预测目标，1 h 为步长预测未来 24 步 RRP 序列。数据标准化采用全局标准化模块，消除量纲影响并稳定训练过程，统计量在 TS-CV 每一折中基于训练数据动态计算，保证验证有效性。

首先对现有数据集采用 3:1 的比例来划分开发集与测试集，测试集在模型训练过程中完全隔离，用以最后的评估。其中，开发集内采用时间序列交叉验证（TS-CV）框架，实现扩展窗口 *K*-fold 机制：其中 K 为 8 折，训练数据从序列开始到第 i 切分点，取$[t_0, t_k]$为训练集，取$(t_k, t_{k+1}]$作为验证集，验证数据取自后续窗口。起始点滑动，确保多历史时期检验。标准化仅用训练数据拟合以防泄露。所有折损失均值作为超参数优化目标。

2.2 评价指标

为了实现全面评估模型的预测性能，本研究评估指标采用平均绝对误差δ_{MAE}、平均绝对百分比误差 δ_{MAPE}、均方根误差δ_{RMSE}，其计算公式为：

$$\delta_{\mathrm{MAE}}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{8}$$

式中：n 为样本总数；y_i 为第 i 个样本的真实值；$\hat{y}_i$ 为第 i 个样本的预测值。

$$\delta_{\mathrm{MAPE}}=\frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i-\hat{y}_i\right|}{\left|y_i\right|} \tag{9}$$

$$\delta_{\mathrm{RMSE}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{10}$$

3 实验结果与讨论

3.1 超参寻优结果

为加速收敛并提高稳定性，学习率调度采用了 OneCycleLR 策略。在每一折交叉验证中，均启用提前停止机制。为了优化模型预测性能，在训练阶段，初始随机值设定为 50 个，共进行 200 轮超参寻优，优化过程如图 5 所示。贝叶斯优化在有限评估次数内从超参数空间定位最优解区域并收敛于稳定性能点，找到使验证集最优的超参数组合。

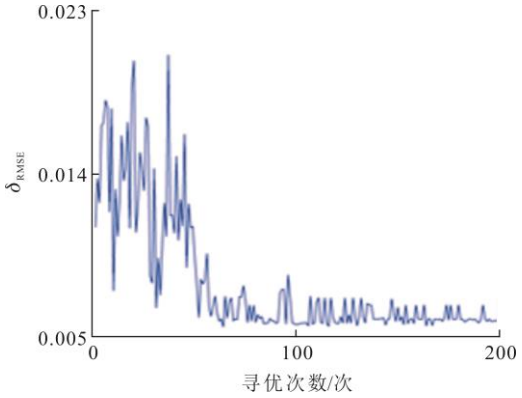


图 5 贝叶斯超参寻优收敛
Fig.5 Bayesian hyperparameter optimization convergence graph

通过 200 轮的超参寻优，最终确定最优超参值如表 2 所示。将最优超参数组合代入本文所构建残差 Mamba 模型并迭代 60 次后，得到经过归一化处理后的性能指标。其中，均方根误差为 0.059 1，平均绝对误差为 0.037 4，平均绝对百分比误差为 12.93%。

表 2 贝叶斯模型优化关键参数指标最终值
Tab.2 Final values of the key parameters for Bayesian model optimization

参数名称	最优值	参数名称	最优值
learning_rate	0.001 1	dropout	0.315 9
batch_size	512	fc_dropout	0.110 8
seq_len	360	train_epochs	13
pred_len	24	kernel_size	9
n_embed	128	revin	0
d_state	256		

在最佳超参数下，残差 Mamba 模型经过归一化处理后，均方根误差极低表明预测鲜有极端偏差，平均绝对误差低证实数值准确，平均绝对百分比误差低表示预测值与实际值之间的平均相对误差占实际值的比例非常低。

图 6 为预测值真实值散点拟合。由图 6 可见，散点大部分围绕完美预测线分布，表明模型成功学习了 RRP 变化的基本规律，预测值与真实值之间存在强正相关性。图 7 为平均绝对误差损失整体分布。该分布呈现出右偏态，其特征展示出高频小误差，表明模型在大部分情况下的预测准确性较高。

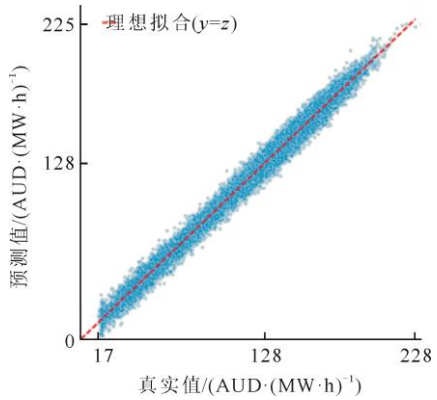


图 6 预测值真实值散点拟合
Fig.6 Scatter fitting of the predicted values and actual values

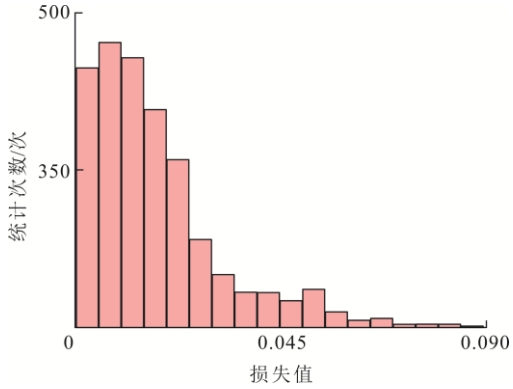


图 7 平均绝对误差损失整体分布
Fig.7 Overall distribution of MAE loss

图 8 为电价预测效果，图 8a)为 1 天的电价预测走势，图 8b)为局部预测电价走势。由图 8 可见，预测曲线可以成功追踪实际数据的主要趋势。在电价的跳变中，模型预测表现出良好的相位对齐和方向一致性，表明模型可学习到该时间序列内在的确定性动态模式。

3.2 消融实验

为了定量评估所提残差 Mamba 模型中各核心组件的独立贡献，设计并进行了系统的消融实验。实验在相同的 AEMO 数据集上进行，采用时间序列交叉验证，评估指标包括均方误差与平均绝对误差。实验设置了 4 种模型变体进行对比，表 3 为消融实验结果。

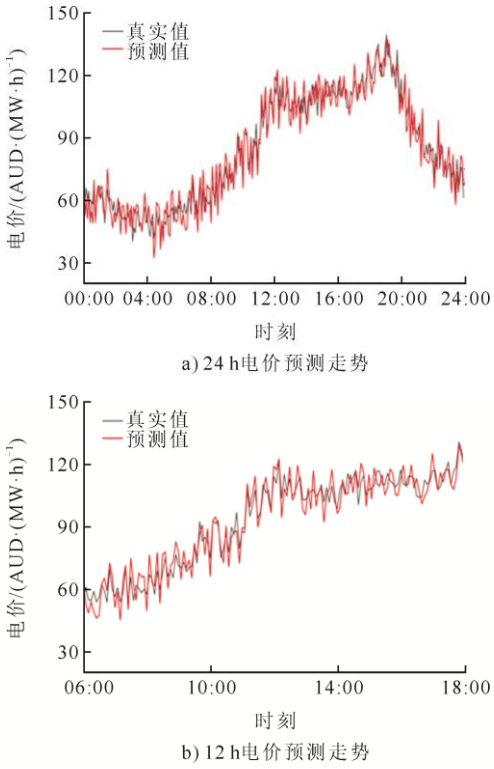


图 8 电价预测效果
Fig.8 Trends of actual and predicted values of electricity price

表 3 消融实验结果对比
Tab.3 Comparison of ablation experiment results

模型	δ_{RMSE}	δ_{MAE}
残差 Mamba 模型	0.059 1	0.037 4
无残差连接模型	0.081 2	0.076 2
特征交互分支模型	0.071 1	0.066 9
无序列分解模型	0.066 0	0.070 5

由表 3 可见，残差 Mamba 模型取得最优性能，均方根误差为 0.059 1，平均绝对误差为 0.037 4。移除全局残差连接后，模型性能下降最显著，均方根误差升至 0.081 2，平均绝对误差升至 0.076 2，全局残差连接的移除引发了梯度传播不稳定，削弱了模型对电价尖峰和波动集聚等极端事件的响应能力。将双分支编码器简化为仅保留特征交互编码器的单一分支，均方根误差上升至 0.071 1，这表明捕捉外生变量与电价间的跨维度耦合关系对于预测精度具有决定性影响。双分支并行编码通过从变量维度和特征维度进行选择性扫描，能够更全面地建模电价形成的复杂动力学机制。移除序列分解预处理模块，均方根误差升至 0.066 0，因其破坏了电价序列中确定性趋势与随机波动的有效解耦，显著增加了模型对混合信号的拟合难度。残差连接、双分支架构与序列分解模块共同构成高效预测系统，缺失

组件会导致捕捉电价非线性动态能力下降，消融实验证明了各核心组件的必要性与协同作用。

3.3 对比实验

为了进一步评估提出的残差 Mamba 模型在时间序列预测任务中的相对优势，本文将其与当前领域内具有代表性的基准模型进行了性能比较（包括 LSTM、itransformer、Dlinear）。表 4 为基于 RRP 的模型预测性能对比。由表 4 可见，原始 Mamba 模型由于其扫描方向仅沿时间维度展开，难以捕获外生变量与电价之间的交互依赖关系，在性能上表现中等，平均绝对误差为 0.062 3，均方根误差为 0.083 5，表明预测误差的离散程度较大；平均绝对百分比误差为 24.46%，显示相对误差占比偏高。相比之下，残差 Mamba 模型实现了大幅提升，平均绝对误差降至 0.037 4，同时均方根误差大幅下降至 0.059 1，平均绝对百分比误差也降低至 12.93%，显著改善了预测的准确性和稳定性。同时，本文所提出的残差 Mamba 模型与 LSTM、itransformer 与 Dlinear 模型相比较，其在平均绝对误差、均方根误差与平均绝对百分比误差上均实现了提升。其中，所提模型的均方根误差相比较原始 Mamba 模型与 LSTM 模型分别降低了 29.22%与 35.33%，验证了残差学习与双分支并行机制在提升电价序列预测性能方面的有效性。

表 4 基于 RRP 的模型预测性能对比
Tab.4 Comparison of model prediction performance based on RRP

对比模型	δ_{MAE}	δ_{RMSE}	δ_{MAPE}/%
LSTM	0.085 7	0.091 4	30.11
残差 Mamba	0.037 4	0.059 1	12.93
Mamba	0.062 3	0.083 5	24.46
itransformer	0.039 2	0.073 5	19.61
Dlinear	0.066 3	0.089 2	28.18

4 结论与展望

本文提出了融合全局残差学习机制与双分支并行编码器的残差 Mamba 模型。该模型创新性地将全局残差学习机制引入 Mamba 状态空间架构以缓解梯度退化问题，同时利用双分支并行编码器分别沿变量维与嵌入特征维进行选择性扫描，以捕捉跨变量耦合与深层交互特征。此外，研究还结合时间序列交叉验证与贝叶斯超参数优化方法，确保了模型评估鲁棒性与参数搜索效率，主要结果如下。

1）在电价预测准确率方面，残差 Mamba 模型在 AEMO 数据集上表现出色：平均绝对误差从 LSTM 模型的 0.085 7 降至 0.037 4，显著收窄了预测值与实际值的平均偏差；同时相对误差指标（平均绝对百分比误差）从 LSTM 模型的 30.11%降至 12.93%，体现了其在电价预测任务中更高的准确性与稳定性。

2）相比其他模型，残差 Mamba 模型实现了显著的性能提升。与 itransformer 模型相比，平均绝对误差从 0.039 2 降至 0.037 4、均方根误差从 0.073 5 降至 0.059 1、平均绝对百分比误差从 19.61%降至 12.93%，在误差离散度和相对误差上表现更优；与 Dlinear 模型相比，平均绝对误差从 0.066 3 降至 0.037 4，均方根误差从 0.089 2 降至 0.059 1，平均绝对百分比误差从 28.18%降至 12.93%。

本文解决了现有 SSM 在电价预测中梯度消失与跨维信息缺失的两大痛点。残差 Mamba 模型的结构创新（全局残差+双分支）与策略创新（Bayesian auto-tuning）共同构成完整的技术闭环。尽管本文提出的残差 Mamba 模型在电价预测的准确性和稳定性方面整体表现更为突出。但是实验也具有一定的局限性，未来工作将进一步开展输入扰动与参数不确定性条件下的系统敏感性分析，以完善模型的应用鲁棒性评估。

[参 考 文 献]

[1] POGGI A, DI PERSIO L, EHRHARDT M, et al. Electricity price forecasting via statistical and deep learning approaches: the german case[J]. Applied Math, 2023, 3: 316-342.

[2] 王来增, 冯璐, 张新敬, 等. 考虑分时电价的风光储系统技术经济性研究[J]. 热力发电, 2024, 53(10): 66-71.
WANG Laizeng, FENG Lu, ZHANG Xinjing, et al. Techno-economic study of wind-photovoltaic-storage systems considering time-of-use electricity pricing[J]. Thermal Power Generation, 2024, 53(10): 66-71.

[3] LAGO J, DE RIDDER F, SCHUTTER B D, et al. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms[J]. Applied Energy, 2018, 221: 386-405.

[4] 李斌, 王雨萌, 张庆来, 等. “光火储”一体化发电系统研究[J]. 热力发电, 2022, 51(2): 56-64.
LI Bin, WANG Yumeng, ZHANG Qinglai, et al. Study on an integrated “PV-thermal-storage” power generation system[J]. Thermal Power Generation, 2022, 51(2): 56-64.

[5] FU Z, WANG F, WANG R Y, et al. Electricity price forecasting based on improved ARIMA model[C]. 2024 6th International Academic Exchange Conference on Science and Technology Innovation (IAECST). IEEE, 2024: 1023-1027.

[6] 勾玄, 肖先勇. 基于经验模式分解与 LSTM 神经网络的短期电价预测模型[J]. 西安理工大学学报, 2020,

36(1): 129-134.
GOU Xuan, XIAO Xianyong. A short-term electricity price forecasting model based on empirical mode decomposition and LSTM neural network[J]. Journal of Xi'an University of Technology, 2020, 36(1): 129-134.
[7] 郭鑫炜, 赵耀. 基于相似波动聚类与 SA-LightGBM 的日前电价预测[J]. 上海电力大学学报, 2024, 40(6): 587-594.
GUO Xinwei, ZHAO Yao. Day-ahead electricity price forecasting based on similar-volatility clustering and SA-LightGBM[J]. Journal of Shanghai University of Electric Power, 2024, 40(6): 587-594.
[8] 龚丹丹. 基于 VMD-ICOA-BiLSTM 混合模型的日前电价预测[J]. 电气技术, 2023, 24(11): 28-34.
GONG Dandan. Day-ahead electricity price forecasting based on a VMD-ICOA-BiLSTM hybrid model[J]. Electrical Engineering, 2023, 24(11): 28-34.
[9] 孙欣, 王思敏, 谢敬东, 等. 考虑多维影响因素的改进 Transformer-PSO 短期电价预测方法[J]. 上海交通大学学报, 2024, 58(9): 1420-1431.
SUN Xin, WANG Simin, XIE Jingdong, et al. A short-term electricity price forecasting method using an improved transformer-PSO considering multi- dimensional influencing factors[J]. Journal of Shanghai Jiao Tong University, 2024, 58(9): 1420-1431.
[10] 许越, 李强, 崔晖. 基于 MIC-EEMD-改进 Informer 的含高比例清洁能源与储能的电力市场短期电价多步预测[J]. 电网技术, 2024, 58(9): 1420-1431.
XU Yue, LI Qiang, CUI Hui. Short-term multi-step electricity price forecasting for power markets with high shares of clean energy and energy storage based on MIC-EEMD and an improved informer[J]. Power System Technology, 2024, 58(9): 1420-1431.
[11] 王腾飞, 杨力, 孙龙, 等. 基于 C-Transformer 的光伏负荷预测方法[J]. 微电子学与计算机, 2025, 42(6): 67-74.
WANG Tengfei, YANG Li, SUN Long, et al. A photovoltaic load forecasting method based on C-Transformer[J]. Microelectronics & Computer, 2025, 42(6): 67-74.
[12] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[C]//First Conference on Language Modeling. 2024.
[13] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces[J]. arXiv preprint. arXiv:2111.00396.
[14] BERGSTRA J, BENGIO Y. Random search for hyper-parameter optimization[J]. Journal of Machine Learning Research, 2012, 13(1): 281-305.
[15] SMITH J T H, WARRINGTON A, LINDERMAN S W, et al. Simplified state space layers for sequence modeling[J]. arXiv preprint. arXiv: 2208.04933, 2022.
[16] 林琦淇. 基于选择性状态空间的时间序列预测模型[J]. 建模与仿真, 2025, 14(5): 293-303.
LIN Qiqi. Time series prediction model based on selective state space[J]. Modeling and Simulation, 2025, 14(5): 293-303.
[17] 张鼎开, 王一喆, 王鹏飞, 等. 基于双重时空掩码与双向 Mamba 状态空间的高速公路交通异常事件检测模型[J]. 交通运输工程学报, 2025, 25(4): 311-327.
ZHANG Dingkai, WANG Yizhe, WANG Pengfei, et al. Highway traffic incident detection based on dual spatio-temporal masking and a bidirectional Mamba state-space model[J]. Journal of Traffic and Transportation Engineering, 2025, 25(4): 311-327.
[18] 刘京华, 魏祥麟, 范建华, 等. 基于时序深度残差收缩网络的混叠信号调制识别方法[J]. 电信科学, 2024, 40(10): 27-38.
LIU Jinghua, WEI Xianglin, FAN Jianhua, et al. A modulation recognition method for aliased signals based on a temporal deep residual shrinkage network[J]. Telecommunications Science, 2024, 40(10): 27-38.
[19] 王盈, 蔡述庭, 熊晓明. 基于主动学习和视觉状态空间模型的热点检测器[J]. 广东工业大学学报, 2024, 41(6): 45-51.
WANG Ying, CAI Shuting, XIONG Xiaoming. A hotspot detector based on active learning and a visual state-space model[J]. Journal of Guangdong University of Technology, 2024, 41(6): 45-51.
[20] WU J, CHEN X Y, ZHANG H, et al. Hyperparameter optimization for machine learning models based on Bayesian optimization[J]. Journal of Electronic Science and Technology, 2019, 17(1): 26-40.
[21] LIU Y, QIN G, SHI Z, et al. Sundial: a family of highly capable time series foundation models[J]. arXiv preprint. arXiv: 2502.00816, 2025.
[22] WU H, HU T, LIU Y, et al. Timesnet: temporal 2d-variation modeling for general time series analysis[J]. arXiv preprint. arXiv: 2210.02186, 2022.

（责任编辑　杜亚勤）