

# 基于STL与自适应残差多分支融合网络的短期电力负荷预测

麻少辉, 邓醉茶\*, 李谋磊, 梁明睿

兰州交通大学数理学院, 甘肃 兰州

收稿日期: 2026年8月7日; 录用日期: 2026年9月1日; 发布日期: 2026年9月14日

## 摘要

新型电力系统建设中分布式光伏、储能装置的大规模接入及用户侧柔性负荷的多元增长, 使得区域用电时序有高度非线性、强时变非平稳的特征, 短期负荷预测质量直接关联电网运行调度、备用及市场交易的经济效率。目前主流深度学习方法尚不能同时处理长周期负荷模式识别和由短期气候扰动所导致的负荷异常, 所用多尺度特征提取手段往往有失均衡。针对工程实际需要, 文中将时序分解方法同异构深度网络加以结合, 提出STL-APRFNet类短期电力负荷预测方案。所用鲁棒STL时序分解法对原负荷序列做加性拆分, 把长期趋势、日内循环、气象相关残差划为各自独立的分量, 由此缓解初始时序非平稳的困难; 按三类分量不同性质分别设计特征提取通道: 以多层感知机分析负荷慢变的长期基线, 利用PatchTST的切块注意机制考察周或日级长程联系, 另以一维卷积作跨模态设计, 同时考虑气象变量及残差序列, 以局部滑动方式将气象急变与负荷瞬态结合于同一特征空间。多通道所得特征在融合时应用可调缩放因子, 使STL分解所提示的先验物理知识与网络拟合结果达到平衡, 不因深度训练而忽略负荷变化基本规律。依据GEFCom2014公开电力数据集完成基准对照试验, STL-APRFNet的均方根误差、平均绝对误差是12.72 MW、8.88 MW, 较典型LSTM基准模型分别减少25.2%、27.8%; 模型平均百分比误差仅5.754%,  $R^2$  (决定系数)可取到0.937。从实测角度考察, 文中方法所建模型的误差控制效果更优、长期预测稳定度高, 对日前调度具有明确帮助意义, 是很有潜力的工程解决方案。

## 关键词

短期电力负荷预测, STL时序分解, 跨模态特征融合, 切块注意力Transformer, 自适应残差权重校正

## Short-Term Power Load Forecasting Based on STL and Adaptive Residual Multi-Branch Fusion Network

Shaohui Ma, Zuicha Deng\*, Moulei Li, Mingrui Liang

School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou Gansu

\*通讯作者。

文章引用: 麻少辉, 邓醉茶, 李谋磊, 梁明睿. 基于 STL 与自适应残差多分支融合网络的短期电力负荷预测[J]. 应用数学进展, 2026, 15(9): 88-100. DOI: 10.12677/aam.2026.159376

## Abstract

The integration of distributed photovoltaics, energy storage and diversified user-side flexible loads makes power load time series highly nonlinear, time-varying and non-stationary in new power systems. Reliable short-term load forecasting is crucial for efficient grid scheduling, reserve arrangement and market trading. Existing deep learning methods fail to jointly model long-term load patterns and meteorology-caused load anomalies caused by short-term climate disturbances, with limited multi-scale feature extraction balance. In response to the practical needs of engineering, this paper proposes an STL-APRFNet short-term load forecasting model combining STL time series decomposition and heterogeneous deep networks. The robust STL method decomposes raw load data into long-term trend, intraday periodicity and meteorological residual components to mitigate time series non-stationarity. Targeted feature extraction branches are designed based on the different properties of three types of components: a multi-layer perception captures long-term load baselines, PatchTST acquires long-range temporal correlations. Moreover, 1D convolution is deployed for cross-modal modeling to jointly incorporate meteorological variables and residual sequences, merging abrupt meteorological changes and load transients into a unified feature space through local sliding operations. Adjustable scaling factors are utilized in multi-channel feature fusion to balance physical prior knowledge from STL decomposition and network fitting outcomes, so that deep training will not overlook the fundamental laws of load variation. Validated on the GEFCom2014 dataset, the proposed model achieves RMSE of 12.72 MW and MAE of 8.88 MW, reducing errors by 25.2% and 27.8% compared with LSTM. It obtains a MAPE of 5.754% and  $R^2$  of 0.937. Validated from practical measurements, the model constructed by the proposed method delivers superior error control performance and high stability for long-term prediction. It provides clear support for day-ahead scheduling and represents a promising engineering solution.

## Keywords

Short-Term Power Load Forecasting, STL Time Series Decomposition, Cross-Modal Feature Fusion, Patch Attention Transformer, Adaptive Residual Weight Correction

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

对短期负荷作准确预测是保障电网安全运行的基本前提，调度所依据预测安排发电单元启停及旋转备用的规模，若预测误差太大将导致冗余机组运行或新能源不能充分利用这类经济损失，极端气候高温或寒潮亦可能造成局部缺电现象。按双碳思路推进能源转型以来，工商业可控负荷和家庭光伏已较普遍地并入配电网，温度、湿度乃至降水都对用电有即时影响，负荷随时间演化的规律同时包含日、周两类周期及随机项，构成复杂非平稳时序事实，给预测模型的建立带来明显困难[1]。

国内外已有众多学者就短期负荷预测做过较完整研究。有关传统统计的模型基本采用自回归积分滑动平均思路，按时间序列平稳及线性关联来设计建模方案，不能很好对应现在所遇的复杂用电波动；像支持向量回归或随机森林这类机器学习法对基本非线性有一定适应力，但要依靠人来挑特征、定输入，若把多维气象因素都带进来，就会有特征重叠现象，所建模型的泛化能力反而大大削弱[2]。

深度学习因具有自动提取特征的能力而被当作时序预测主流方法。循环神经网络一类的模型以门控

形式管理历史时序所涉信息，但用长序列做输入时梯度可能衰减，早期负荷状态无从得知；标准 Transformer 所依据的全局自注意规则适合于周期性分析，却不利于识别局部短暂气象因素导致的负荷跃变，亦不能很好抵抗高维气象噪声的干扰。“分解预测”类一体化架构基本按同构思路将趋势、周期及残差同时处理，没有考虑各频段对应时序的特征特点，未把时序分解用于真正有效的降维分析。

针对已有模型所遇的各种困难，这里提出以“物理先验时序分解 - 分频段定向特征提取 - 动态自适应加权融合”为思路的混合结构 STL-APRFNet。有关的创新点有三项：

- (1) 采用训练、验证、测试集独立切分后再分解的隔离式 STL 滚动分解方案，降低负荷时序非平稳性，针对趋势、周期、残差三类分量分别搭建专用异构网络，实现分频段精准特征挖掘；
- (2) 设计气象 - 负荷残差联合一维卷积支路，摒弃简单通道拼接操作，利用卷积局部感受野实现气象扰动与瞬时负荷波动的深度时空对齐；
- (3) 在多支路特征融合阶段引入可学习缩放权重，自动平衡负荷变化物理机理与数据驱动拟合结果，在降低预测误差的同时提升模型训练稳定性。

2. 理论基础与模型架构

单一网络结构难以同步完成长期趋势锁定、跨周期规律提取、高频噪声过滤三类任务，为此本文设计 STL-APRFNet 三层级核心架构，整体分为 STL 鲁棒时序分解层、多分支并行特征提取层、自适应残差校正融合层，完整模型结构如图 1 所示。

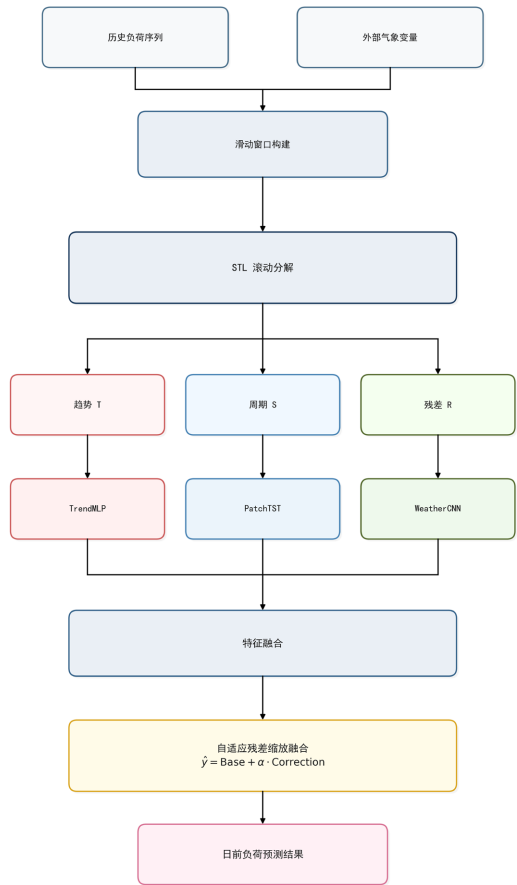


Figure 1. Overall architecture of the STL-APRFNet model  
图 1. STL-APRFNet 整体模型架构图

负荷时序中包含极端气温、设备检修、节假日等突变样本,传统经验模态分解算法处理此类数据易产生端点失真问题。本文选用基于局部加权回归的 STL 时序分解算法[3],该算法对时序异常值具备强鲁棒性,可将原始负荷时序  $Y$  完成加性拆解,得到三类具备清晰物理含义的独立分量:趋势分量  $T$ 、周期分量  $S$ 、残差分量  $R$ 。

$$Y_t = T_t + S_t + R_t \quad (1)$$

趋势分量是有关区域用电的长期、平缓基本变化项;周期分量以日或周为单位给出重复性节律所对应的用电波动;而残差部分在扣除确定性趋势及周期之后,基本由极端天气和测量误差所导致的短时负荷扰动来填充。

时序预测建模存在典型的数据泄露风险,若采用全量时序统一分解会让测试集提前获取未来时段的负荷特征,造成实验结果虚高失真。为规避该缺陷,本文先按照时序先后完成训练、验证、测试样本的划分,三组子集分别独立完成 STL 分解运算,全程隔离测试集与未来时序信息,以此确保各组对照实验的评价结果具备客观有效性。

本文使用小时级的电力负荷数据开展预测。用户用电行为普遍存在显著的日周期特性,因此将 STL 分解的周期参数设定为 24,以 24 小时为周期提取序列的季节分量。考虑到真实负荷会受极端天气、设备故障等因素产生突变,研究选用鲁棒 STL 算法,削弱异常样本对分解效果带来的干扰。周期参数依据电力负荷本身的周期特点确定:24 h 周期既可以刻画日内峰谷演变,也能够规避短周期带来季节特征缺失、长周期抹平局部波动的缺陷。

## 2.1. 多分支并行特征提取

按 STL 所述方法分解得到的三类分量就其频域或时域而论是不同的,故 STL-APRFNet 对三种特征做平行、异构的处理加以适应。将所用的历史时间步数定为  $L$ ,预测结果也以  $L$  为步长。

### (1) 宏观趋势支路

趋势分量曲线是平滑型、低速变化的时序现象,用复杂网络来拟合它易出现过拟合。针对该类问题现将层归一化用于深度多层感知机设计,建立整体上静态的映射形式,按少参数原则分析远期负荷的演变规律,因而可使计算本身更节省资源。

$$\hat{Y}_T = W_{T2} \left( \text{Dropout} \left( \text{GELU} \left( \text{LayerNorm} \left( W_{T1} X_T + b_{T1} \right) \right) \right) \right) + b_{T2} \quad (2)$$

### (2) 长程周期支路

周期分量有高度的跨周期依赖性,要完整利用多日或多周已知负荷的联系应采用本支路所提的 PatchTST 切块方法[4]。对一维序列按滑动窗口来分块得到若干小图块,把原序列作适当缩短,再以高维形式将图块送入多头自注意 Transformer 的编码部分[5],依靠全局注意力机制提取跨日、跨周的深层时序关联信息,最终输出周期分量对应的预测结果。

$$N = \left\lfloor \frac{L_x - P}{S_{\text{stride}}} \right\rfloor + 1 \quad (3)$$

### (3) 气象 - 残差跨模态卷积支路

残差分量的高频震荡大多是由外部多维气象因素引起的,文中把一维负荷残差及  $d$  个气象协变量按通道方式合并为联合输入矩阵。以一维卷积方法作局部信息处理,利用时序上滑动的窗口对多模态特征做交叉分析;所用卷积核取 3 或 5 两种大小,专门考虑短期负荷局部变化,把气象影响同残差的阶跃式波动按时间 - 空间关系加以对应、辨识。

$$Z_1 = \text{GELU}(\text{Conv1D}_{k=3}(M_{RW})) \quad (4)$$

$$Z_2 = \text{GELU}(\text{Conv1D}_{k=5}(Z_1)) \quad (5)$$

$$\hat{Y}_R = W_R(\text{Flatten}(Z_2)) + b_R \quad (6)$$

## 2.2. 自适应残差缩放融合机制

三条异构支路完成特征提取后, 本文采用两阶段自适应融合策略整合多源特征信息, 先将三路支路输出特征拼接为联合特征矩阵。

$$U = [\hat{Y}_T; \hat{Y}_S; \hat{Y}_R] \quad (7)$$

(1) 动态基准重构: 依托 Softmax 门控网络, 根据单时刻样本的数据特征动态计算趋势、周期、残差三条物理支路的置信权重:

$$[w_t, w_s, w_r] = \text{Softmax}(W_{w2}(\text{GELU}(W_{w1}U + b_{w1})) + b_{w2}) \quad (8)$$

结合权重加权生成贴合负荷物理变化规律的基准预测值:

$$\text{Base} = w_t \hat{Y}_T + w_s \hat{Y}_S + w_r \hat{Y}_R \quad (9)$$

(2) 自适应非线性校正: 所称的线性加权方法对高维气象及负荷所涉的非线性关联不能作完整模拟, 故考虑用多层感知机来求取非线性残差修正值, 在最后模型中以可学习的标量  $\alpha$  来调节输出。

$$\hat{Y} = \text{Base} + \alpha \cdot \text{Correction} \quad (10)$$

融合方法在模型刚开始训练时即利用 STL 对物理基准作出分解, 按梯度反向思路来确定各项缩放权重, 把由先验引导的物理形式及数据拟合两者调和至最合适的平衡点。

## 3. 实验设置与评价指标

### 3.1. 数据预处理与样本构建

本次实验选用电力时序预测领域权威公开数据集 GEFCom2014 的 L1-train.csv 子集[6]。数据集包含高分辨率实际负荷时序, 同步采集 25 维同步微气象协变量, 可充分验证气象 - 负荷跨模态建模效果。

不同特征量纲、数值区间差异会干扰神经网络训练收敛速度, 因此特征输入网络前, 采用 Min-Max 缩放算法将全部连续特征归一化至[0, 1]区间。为杜绝测试集引入未来统计信息, 归一化所需最大值、最小值仅基于训练集数据统计计算, 统一复用至验证集、测试集完成变换, 其变换公式为:

$$\tilde{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (11)$$

式中:  $X$  为原始特征序列,  $X_{\min}$  与  $X_{\max}$  分别对应特征在训练集上的极小值与极大值。

时序预测任务必须遵循时间先后顺序划分样本, 按照 7:1:2 比例切分训练集、验证集、测试集; 以滑动窗口方法构造输入 - 输出的对应关系, 选取最近 7 天即 168 小时历史负荷与气象数据作为输入, 实现未来 24 小时日前负荷曲线预测。

为消除实验变量带来的干扰, 所用网络及全部结构参数都已按表 1 列出, 各比较模型在数据分割和训练上是同样处理的, 因而能对彼此公平的性能对照。输入窗口 168 小时, 覆盖一周用电周期; 预测长度 24 小时, 适配日前负荷预测需求。训练启用早停策略, 减轻过拟合, 提升泛化性能。

**Table 1.** Network training and architecture parameter settings  
**表 1.** 网络训练与结构参数配置

| 参数名称     | 参数设置          | 参数名称             | 参数设置  |
|----------|---------------|------------------|-------|
| 输入历史长度   | 168 h         | 预测长度             | 24 h  |
| 数据划分比例   | 7:1:2         | 批次大小             | 128   |
| 优化器      | Adam          | 初始学习率            | 0.001 |
| 权重衰减     | 1.00E-05      | 最大训练轮数           | 100   |
| 早停策略     | Patience = 10 | 损失函数             | MSE   |
| STL 周期参数 | 24            | Patch 长度         | 16    |
| Patch 步长 | 8             | Transformer 编码层数 | 2     |
| 多头注意力数量  | 4             | 隐藏维度             | 64    |

模型训练采用均方误差 MSE 损失函数。该函数对大误差项惩罚力度更大，可有效控制大幅度预测偏差，提升负荷波动区间的预测质量。对于未来实际场景中峰谷负荷样本分布不均的问题，后续将探索结合 MAE 损失或峰值加权方式的复合损失函数。

### 3.2. 评价指标

要对所用模型的预测误差及拟合状况作较完整的量化分析，文中将采用四个行业标准的评价指标，即均方根误差 RMSE、平均绝对误差 MAE、平均绝对百分比误差 MAPE、决定系数  $R^2$ 。其数学定义如下：

RMSE：对大幅度预测偏差更加敏感，用于衡量负荷实际值与预测值的离散度程度。

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_{\text{real}}(i) - y_{\text{pre}}(i))^2} \quad (12)$$

MAE：直观反映单样本预测误差平均绝对大小。

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |y_{\text{real}}(i) - y_{\text{pre}}(i)| \quad (13)$$

MAPE：消除负荷数值量级影响，方便不同数据集横向对比误差水平。

$$\text{MAPE} = \frac{100\%}{M} \sum_{i=1}^M \left| \frac{y_{\text{real}}(i) - y_{\text{pre}}(i)}{y_{\text{real}}(i)} \right| \quad (14)$$

$R^2$ ：代表模型对原始负荷数据变异特征的解释能力，数值越趋近 1，模型拟合效果越优异。

$$R^2 = 1 - \frac{\sum_{i=1}^M (y_{\text{real}}(i) - y_{\text{pre}}(i))^2}{\sum_{i=1}^M (y_{\text{real}}(i) - \bar{y})^2} \quad (15)$$

各项指标计算公式沿用行业标准定义，其中  $M$  为测试集样本总量， $y_{\text{real}}(i)$ 、 $y_{\text{pre}}(i)$  分别为真实负荷、模型预测负荷， $\bar{y}$  为测试集真实负荷均值。

### 3.3. 对比模型与消融实验设计

为论证 STL 所述的 APRFNet 模型及其各基本模块作较完整的检验，文中将安排 5 种主要基准模型、2 组消融对照模型，共 7 项对比方案：

LSTM 长短期记忆网络：传统循环时序建模代表，用于横向对比长时序依赖信息的留存效果[7]。

GRU 门控循环单元: LSTM 的轻量化改进结构, 精简内部门控更新逻辑, 整体参数量实现缩减[8]。

TCN 时域卷积网络: 依托因果空洞卷积搭建时序提取框架, 用以检验纯卷积结构对长周期负荷特征的捕获能力[9]。

基础 Transformer: 搭载绝对位置编码的原生自注意力网络, 可直观体现单一注意力架构难以捕捉短时负荷突变的固有缺陷。

PatchTST-CF: 业内通用分块时序预测基准, 用于区分单一切块注意力结构与本文多分支异构融合框架的性能差异。

消融模型 1 不含 STL 分解: 把前端有关的时序分解予以剔除, 按原样使用负荷及气象资料, 考察 STL 对非平稳序列的帮助。

消融模型 2 不含自适应残差融合: 把 STL 的分解层予以保留, 不设异构分支或自适应缩放融合, 只利用基本线性方法将三路结果相加, 考察有关支路及修正项的精度贡献。

## 4. 实例分析与结果讨论

### 4.1. 实验运行环境

文中所列各种对比及消融测试都是在 Windows 平台下做的, 以 PyCharm 作为开发工具来设计和实现 PyTorch 模型; 就大规模时序矩阵计算、网络收敛速度而言, 所有训练过程都启用 GPU 类硬件加速。有关数据分割、归一化或构造的步骤均完全按 3.1 节给出的预处理方法加以统一处理, 不因数据处理不同而影响实验结论的正确性。

### 4.2. 序列特征分解

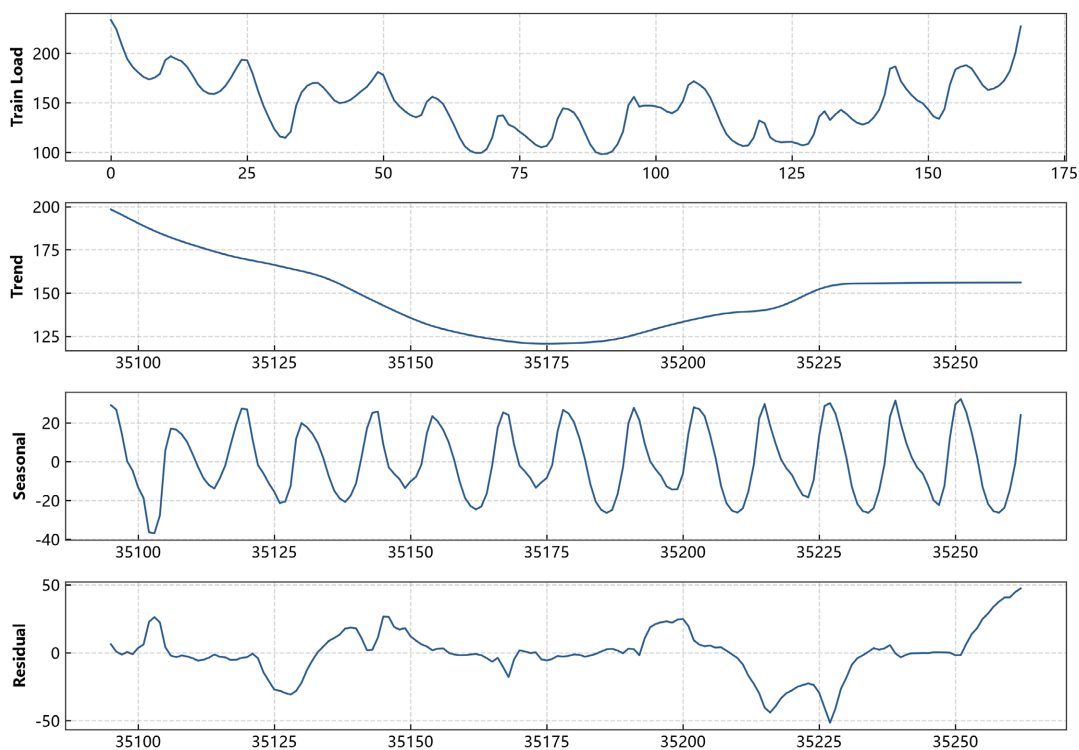


Figure 2. STL feature decomposition plot of GEFCom2014 load series

图 2. GEFCom2014 负荷序列 STL 特征分解图

电网负荷所涉的时序有多种尺度的周期、高度非平稳及剧烈波动现象，若用原数据直接作网络建模会加大特征分析难度。文中采用 STL 对时序做初步分解处理，将所研究的负荷序列划为趋势、周期与残差三个各自独立的部分，有关 GEFCom2014 数据的分解结果以图 2 形式给出。

按所分三类来考察各分量的频域特征时，低频走向、中频循环及高频残差都已分开，原序列所含非线性或不稳现象因而大体被解除，对各分支路来说可得到较准确的特征输入样本。

4.3. 预测结果分析

将 STL-APRFNet 与 5 组基准模型、2 组消融变体开展对照测试，全部模型四项误差指标统计结果如表 2 所示，MAE 指标直观对比柱状图见图 3。

Table 2. Comparison results of evaluation indicators for each model  
表 2. 各模型的评价指标对比结果

| Model        | MAE   | RMSE  | MAPE   | $R^2$ | NMAE   |
|--------------|-------|-------|--------|-------|--------|
| LSTM         | 12.30 | 17.00 | 7.912% | 0.888 | 0.0782 |
| GRU          | 11.97 | 16.63 | 7.752% | 0.893 | 0.0762 |
| TCN          | 14.04 | 18.72 | 9.109% | 0.864 | 0.0893 |
| Transformer  | 11.80 | 16.25 | 7.665% | 0.898 | 0.0751 |
| PatchTST-CF  | 12.79 | 17.54 | 8.217% | 0.881 | 0.0814 |
| Without STL  | 11.95 | 16.46 | 7.761% | 0.895 | 0.0760 |
| Without APRF | 10.33 | 14.39 | 6.753% | 0.920 | 0.0657 |
| STL-APRFNet  | 8.88  | 12.72 | 5.754% | 0.937 | 0.0565 |

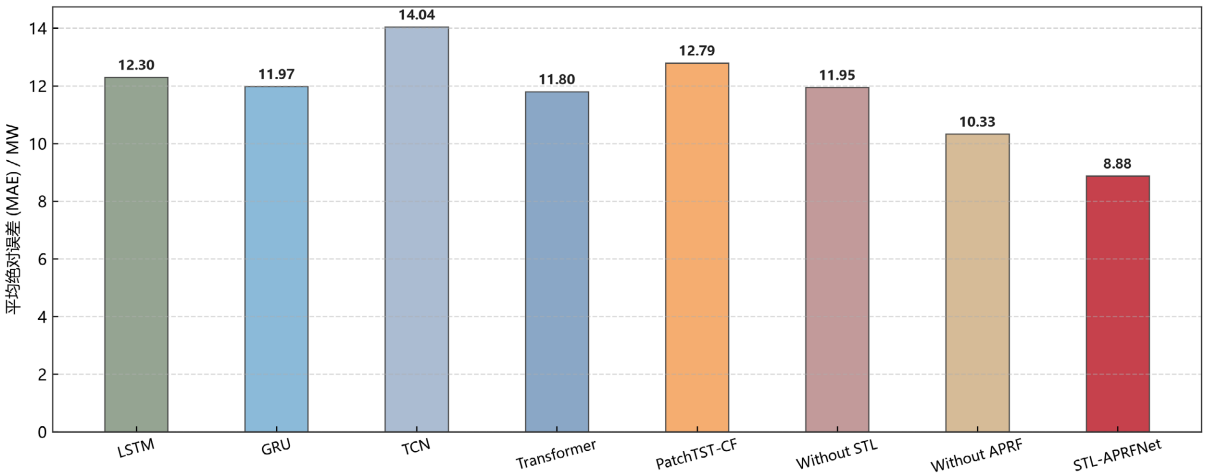


Figure 3. Bar chart for comparison of mean absolute error (MAE) metrics of each model  
图 3. 各模型平均绝对误差 MAE 指标对比柱状图

表 2 是全部模型定量误差数据，图 3 以柱状图直观呈现各组 MAE 数值，结合图表分析。综合全部评价指标能够看出，以总体误差为标准考察所提模型及各对比方案，则本文模型占优。TCN 单靠卷积分析局部、短期用电现象，对跨周长的周期性无把握，整体误差最大，RMSE 是 18.72 MW；LSTM 是传统循环类基准方法，不能很好记住长序列信息，预测有滞后现象，MAE 到达 12.30 MW。STL-APRFNet 模型 RMSE 只需 12.72 MW，比 LSTM 少 25.2%，与 TCN 相比降 32.1%；MAE 的控制值是 8.88 MW，较 LSTM

缩减 27.8%；MAPE 所示比例仅 5.754%，即高低负荷情形下误差都基本可控；拟合质量  $R^2$  为 0.937，说明该模型能解释 93.7% 有关负荷的时序规律。

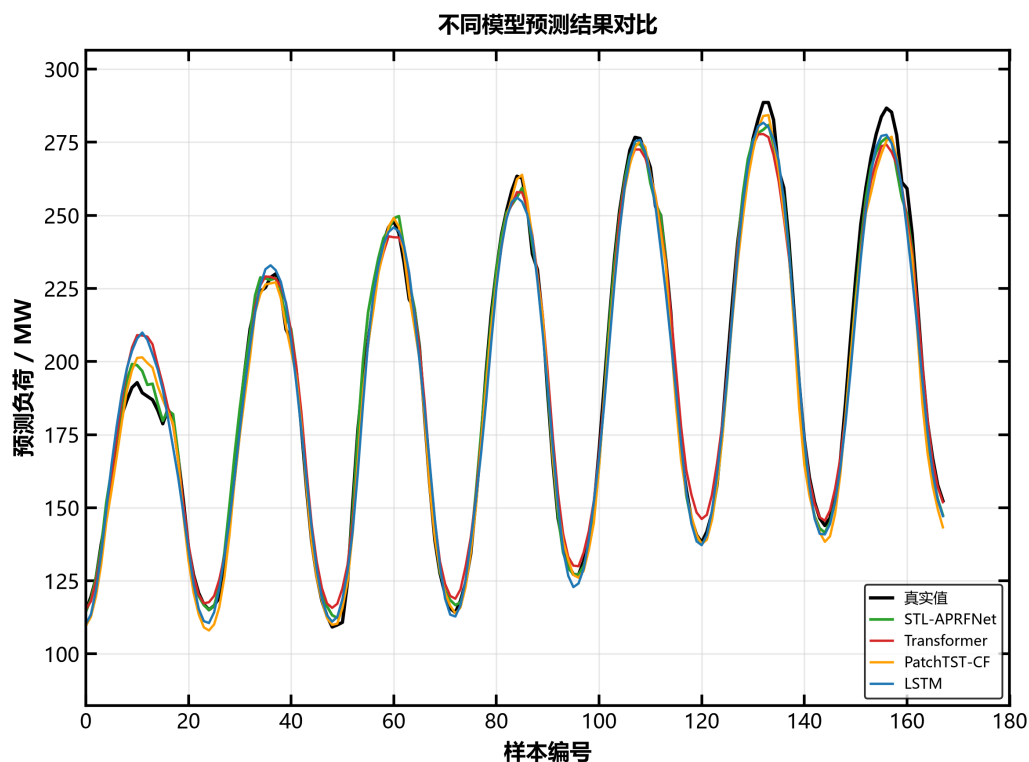


Figure 4. Comparison of fitting curves of prediction results for each model

图 4. 各模型日前预测结果对比拟合曲线

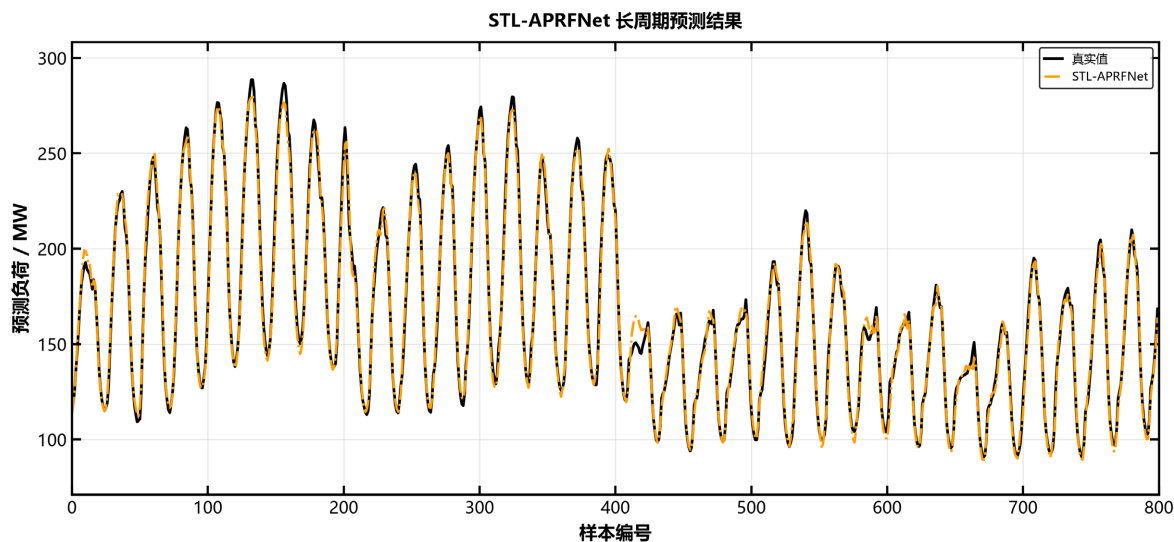


Figure 5. Panoramic tracking plot of long-cycle continuous prediction by the STL-APRFNet model

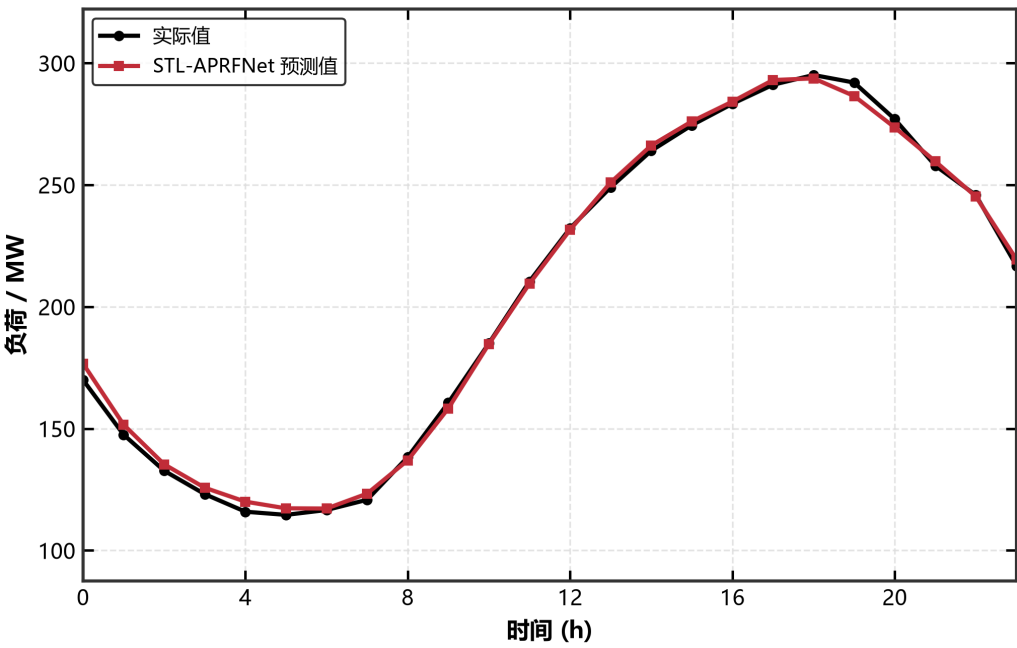
图 5. STL-APRFNet 模型长周期连续预测全景追踪图

按日前预测来考察各模型所拟合的曲线如图 4，就早晚用电高峰及深夜负荷低谷这类极值时段而论，LSTM 和标准 Transformer 的预测曲线有较大偏移、超调情况；STL-APRFNet 采用多分支分频处理及自

动权重调节方法，对实际负荷把握较准，所作极值点预测的精确度是很大的进步。从长周期连续预测角度分析如图 5 可知：模型长期追踪时间序列不混乱，所发误差不会随时间不断积累。

为进一步验证模型面对复杂负荷波动的预测性能，本文选取测试集内日内峰谷特征显著的样本开展案例分析。如图 6，该负荷序列周期性特征突出，凌晨负荷偏低，随用电需求抬升逐步上涨，傍晚抵达峰值。

由预测结果可见，STL-APRFNet 拟合曲线与真实负荷趋势契合度高，可有效捕捉负荷抬升、峰值维持与回落过程。即便负荷剧烈波动，模型预测误差依旧较小。这说明 STL 分解可实现多尺度特征解耦，配合多分支网络挖掘趋势、周期与短时扰动信息，提升复杂情况下预测稳定性。



**Figure 6.** Prediction results of STL-APRFNet under complex load fluctuation scenarios  
**图 6.** 复杂负荷波动场景下 STL-APRFNet 预测结果

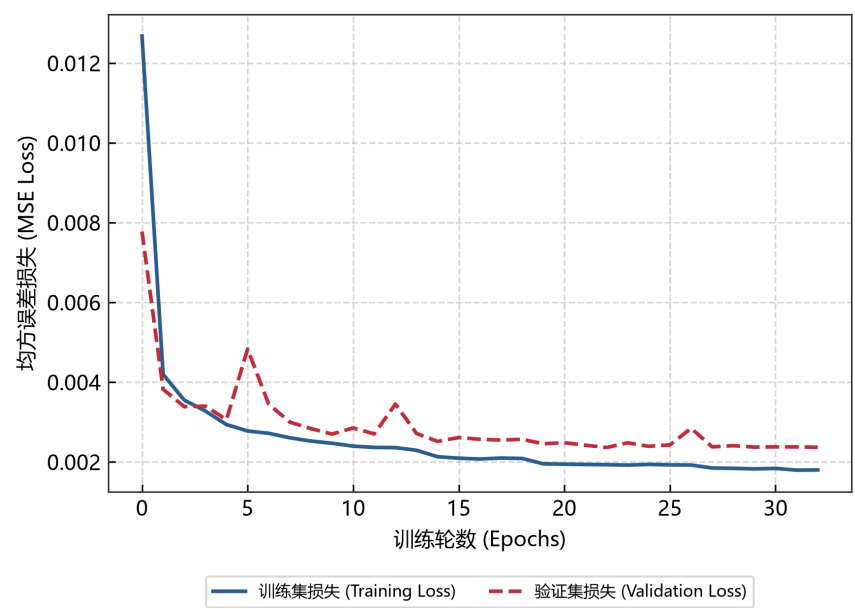
#### 4.4. 消融实验与收敛性分析

采用控制变量消融实验量化各核心模块对整体预测性能的贡献度：

- (1) 移除 STL 分解层后，模型 MAE 上升至 11.95 MW，精度衰减幅度显著，证明 STL 时序分解能够有效降低原始序列非平稳性，大幅降低网络特征提取难度；
- (2) 把自适应残差缩放的融合方法拿掉，所用模型 MAE 就能达到 10.33 MW，即使不废除时序分解步骤，没有按频段来设计提取路径、没有对权重作动态修正，仍不能很好利用多频信息的配合效应。

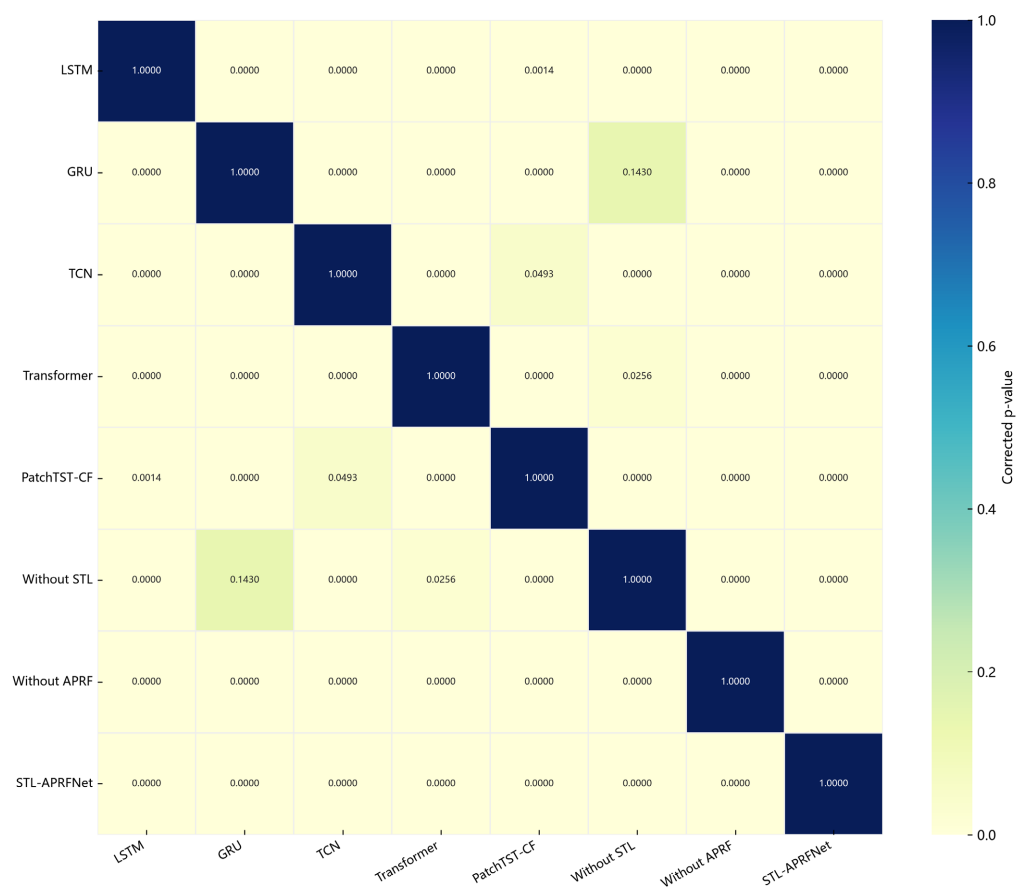
图 7 为模型迭代全过程损失变化曲线，观察训练后期走势能够发现，STL-APRFNet 训练集、验证集的 MSE 损失同步平缓走低，两条曲线数值贴近、无大幅分离分叉的情况，充分说明该网络训练过程平稳，未出现显著过拟合缺陷

本文采用 Bonferroni 校正后的 DM 检验方法对多步预测误差开展统计学显著性分析，对应检验热力图如图 8 所示。从量化检验结果能够看出，STL-APRFNet 与其余所有对比模型之间的预测误差显著性 p 值均小于 0.001 [10]，该统计学结论证明本文模型与各类基线、消融方案的预测精度存在本质性差距，从统计层面进一步说明 STL-APRFNet 的预测性能具备显著优势。



**Figure 7.** MSE loss convergence curves of training and validation sets for the STL-APRFNet model

**图 7.** STL-APRFNet 训练集与验证集均方误差损失(MSE)收敛曲线



**Figure 8.** Heatmap of DM significance test for multi-step prediction based on Bonferroni correction

**图 8.** 基于 Bonferroni 校正的多步预测 DM 显著性检验热力图

4.5. 模型复杂度与运行效率分析

为评估模型工程落地可行性，在统一 GPU 加速环境下统计所有对比模型参数量、单轮完整训练耗时，统计结果如表 3 所示。

Table 3. Comparison of computational complexity and training time consumption of different models  
表 3. 不同模型的计算复杂度与训练耗时对比

| 模型           | 参数量     | 训练耗时     |
|--------------|---------|----------|
| LSTM         | 58,392  | 21.11 s  |
| GRU          | 44,184  | 24.54 s  |
| TCN          | 68,712  | 13.73 s  |
| Transformer  | 822,104 | 652.39 s |
| PatchTST-CF  | 98,803  | 309.87 s |
| without -STL | 205,842 | 48.50 s  |
| without-APRF | 112,968 | 42.97 s  |
| STL-APRFNet  | 215,574 | 86.35 s  |

传统 LSTM、GRU、TCN 轻量化模型参数量少、单轮训练速度更快，但缺少多尺度时序建模能力，预测误差偏高；标准 Transformer、PatchTST-CF 依靠全局注意力建模，理论特征拟合能力更强，但全局自注意力计算复杂度极高，Transformer 参数量达到 822,104，单轮训练耗时超 10 分钟，大规模电网时序场景下计算成本过高。

文中所提的 STL-APRFNet 是对多分支异构提取结构及自适应融合作改进的模型，所用参数及训练时间略高于轻量化方法，但总的训练时间仍属分钟级，GPU 硬件上计算负担是完全可以接受的。它利用小幅度的计算资源增加来达到预测精度的明显改善，不使误差与复杂度过分悬殊，因而很适合实际电力调度中各类工程情况的运用。

5. 结论与展望

本文讨论的是短期电力负荷预测中遇到的时序强非线性、多尺度周期关联及气象性随机干扰诸项困难，提出用 STL 分解时序与自适应残差多分支方法来处理这些情况，建立 STL-APRFNet 预测模型。有关分析以 STL 算法为基础将负荷各尺度特征分别划出，就每类频段时间序列做特征支路设计，按动态方式把多源信息含残差加以融合，使所建模型能更准确地对应实际复杂的用电演化过程。

利用 GEFCom2014 公布的完整电力数据做对照试验分析，STL-APRFNet 模型 RMSE = 12.72 MW、MAE = 8.88 MW、MAPE = 5.754%、 $R^2 = 0.937$ ，所列各项指标都比 LSTM、GRU、TCN、Transformer、PatchTST-CF 和两种消融变体要好。从实验本身看，所用模型能较完整地识别负荷长期走向、周期变化及气候影响诸项因素，对调度、对交易都是可靠的数值资源，因而所提架构既有效又具实际工程意义。

文中所提模型就预测性能来说是较成功的，但仍有两项工程应用上的不足：第一是模型对完整、无缺项的多维气象输入有很强依赖性，若所用气象资料不全或发生中断，短期预测的可靠性就会减弱，此时宜考虑以缺失值补全方法改善整体鲁棒性；第二是现在模型按离线方式固定训练，对于电网负荷随季节而变、用电状况发生漂移的情形不能自动调节网络状态，所涉动态适应性还有改进的可能。

对所涉三个方向多源异构数据的结合、在线增量式学习、动态自适应预测要作后续的拓展分析，同时把有关分布式新能源、用户弹性用电及当前电网状况的各种信息一并加以考虑；就极端气候或高波动

负荷现象做模型改进, 努力使所建模型有较好的泛化性、较低的误判率, 由此给出针对新型电力系统的较完整、较现实的数据驱动型短期负荷预测思路与方法框架。

## 参考文献

- [1] Hong, T. and Fan, S. (2016) Probabilistic Electric Load Forecasting: A Tutorial Review. *International Journal of Forecasting*, **32**, 914-938. <https://doi.org/10.1016/j.ijforecast.2015.11.011>
- [2] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015) Time Series Analysis: Forecasting and Control. 5th Edition, Wiley.
- [3] Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, **6**, 3-73.
- [4] Nie, Y., Nguyen, N.H., Sinthong, P. and Kalagnanam, J. (2023) A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. 2023 *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*, Kigali, 1-5 May 2023.
- [5] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. arXiv: 1706.03762.
- [6] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. and Hyndman, R.J. (2016) Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, **32**, 896-913. <https://doi.org/10.1016/j.ijforecast.2016.02.001>
- [7] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/d14-1179>
- [9] Bai, S., Kolter, J.Z. and Koltun, V. (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv: 1803.01271.
- [10] Diebold, F.X. and Mariano, R.S. (1995) Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, **13**, 253-263. <https://doi.org/10.1080/07350015.1995.10524599>