

多源融合驱动的光伏电站功率日前预测

肖成郭*, 赵行淑, 赵海亦, 王冠淇, 邢世奇

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2026年1月2日; 录用日期: 2026年1月26日; 发布日期: 2026年2月2日

摘要

针对光伏发电间歇性与波动性导致的日前预测精度不足问题, 本文提出“基准构建 - 融合优化 - 空间适配”三阶段光伏电站功率日前预测体系。该体系以某6600 kW级光伏电站实测数据与数值天气预报(NWP)数据为基础: 首先构建单变量长短期记忆网络(LSTM)基准模型, 精准捕捉功率时序依赖; 随后通过CLARANS聚类算法划分强、中、弱辐射场景, 结合Bi-LSTM融合多源气象数据, 增强复杂气象适应性; 最后采用Co-Kriging插值法实现NWP数据空间降尺度, 解决空间尺度不匹配问题。研究表明, 该三阶段协同体系预测性能显著优于单一模型, 均方根误差(RMSE)降至0.0007, 决定系数(R^2)提升至0.958, 较基准模型精度累计提升30%, 复杂地形区域预测误差降低20%以上。本研究通过时序 - 空间跨维度优化策略, 有效提升了光伏功率日前预测的精准度与稳定性, 验证了“分阶优化、时空协同”技术路线的可行性, 为电网新能源消纳与电力系统高质量发展提供了重要技术支撑。

关键词

光伏功率日前预测, 三阶段协同预测体系, LSTM, CLARANS聚类, Co-Kriging插值

Multi-Source Fusion Driven Day-Ahead PV Power Forecasting for PV Power Plants

Chengguo Xiao*, Xingshu Zhao, Haiyi Zhao, Guanqi Wang, Shiqi Xing

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: January 2, 2026; accepted: January 26, 2026; published: February 2, 2026

Abstract

To address the issue of insufficient day-ahead forecasting accuracy caused by the intermittency and variability of photovoltaic power generation, this paper proposes a three-stage day-ahead power forecasting system for photovoltaic power plants: “Benchmark Construction-Fusion Optimisation-

*通讯作者。

Spatial Adaptation". This framework utilises measured data from a 6600 kW photovoltaic power station alongside Numerical Weather Prediction (NWP) data: first, a univariate Long Short-Term Memory (LSTM) benchmark model is constructed to precisely capture temporal dependencies in power output; subsequently, the CLARANS clustering algorithm classifies scenarios into strong, moderate, and weak irradiance conditions, while Bi-LSTM integrates multi-source meteorological data to enhance adaptation to complex weather patterns; finally, Co-Kriging interpolation is employed to spatially downscale NWP data, resolving spatial scale mismatch issues. Research demonstrates that this three-stage collaborative system significantly outperforms single models in prediction performance: root mean square error (RMSE) is reduced to 0.0007, coefficient of determination (R^2) increases to 0.958, cumulative accuracy improves by 30% over the baseline model, and prediction errors in complex terrain areas decrease by over 20%. Through this spatiotemporal cross-dimensional optimisation strategy, the study effectively enhanced the accuracy and stability of photovoltaic power forecasting. It validated the feasibility of the "multi-stage optimisation and spatiotemporal coordination" technical approach, providing crucial technical support for grid integration of renewable energy and the high-quality development of power systems.

Keywords

Day-Ahead Photovoltaic Power Forecasting, Three-Stage Collaborative Prediction System, LSTM, CLARANS Clustering, Co-Kriging Interpolation

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

随着全球能源结构向清洁能源转型，光伏发电作为可再生能源的核心组成部分，在电力系统中的占比持续提升，国际能源署(IEA)预测到 2050 年光伏发电将占全球总发电量的 25% 以上[1]。精准的光伏电站发电功率日前预测是优化电网调度、提高光伏消纳能力、降低备用容量成本的关键技术，对保障电力系统的稳定性与经济性具有重要现实意义。然而，光伏发电功率受太阳辐照度、温度等气象因素影响显著，呈现出强烈的间歇性和波动性，尤其在复杂天气条件下，预测精度难以满足工程应用需求，成为制约光伏大规模并网的核心瓶颈[2]。

当前光伏功率预测研究已形成多条技术路径，但仍存在一定局限性。单一基于历史功率数据的时间序列模型(如 LSTM、RandomForest)虽能捕捉功率周期性规律，却忽略了气象因素的动态影响，复杂天气下预测误差较大[3]；数值天气预报(NWP)的融入虽能提升模型适应性，但粗分辨率的 NWP 数据难以反映局地气象差异，且传统深度学习模型对 NWP 多源数据的特征提取效率较低，限制了精度提升空间[4]。此外，现有研究多将模型优化与数据融合技术分离应用，缺乏从数据预处理、基础建模到技术强化的系统性设计，尚未形成兼顾精度与工程实用性的完整预测体系。空间降尺度技术在光伏预测中的应用仍处于探索阶段，其与多源数据融合的协同优化潜力尚未得到充分挖掘。

针对上述研究空白，本文聚焦光伏电站发电功率的日前预测精度提升，构建了“基础建模 – 多源融合 – 空间适配”的多阶段协同预测体系。首先通过对光伏发电特性的深入分析夯实数据基础，进而构建基于历史数据的基准预测模型；随后引入 NWP 等多源气象信息优化模型结构，最后借助空间降尺度技术增强局地气象感知能力，形成层层递进的技术路径。本文旨在通过这一系统性技术方案，解决单一模型适应性不足、NWP 数据利用率低等问题，有效提升复杂气象条件下的预测精度与工程适用性。

2. 研究基础与数据处理方法

2.1. 数据来源与概况

本研究数据包含光伏电站历史发电功率数据与数值天气预报(NWP)气象数据两类核心数据, 均来自实际工程场景。

光伏历史功率数据来源于某 6600 kW(电站装机容量)级光伏电站 2015 年 1 月 1 日年至 2020 年 12 月 31 日的某一个子阵实测数据, 采样频率为 15 分钟, 共计 351,360 条记录(本文所有功率的特性分析、误差验证及模型优化等相关内容, 均基于该子阵实测数据展开, 未涉及电站其他子阵或总装机容量层面的聚合数据)。该数据集完整覆盖不同季节与典型天气条件, 能够有效反映电站实际运行中的功率波动特性。NWP 气象数据来自国家气象科学数据中心提供的格点数据, 空间分辨率为 $25 \text{ km} \times 25 \text{ km}$, 时间分辨率为 15 分钟, 包含日总水平辐射(GHI)、散射辐射基础数据、环境温度、相对湿度、风速等 15 类关键气象因子, 时间跨度与功率数据保持一致, 孙娴等提出的改进型理想大气太阳辐射计算模型为这类辐射数据的精准核算提供了技术支撑[5]。

数据涵盖完整的四季周期, 包含晴天、阴天、雨天、多云转晴等多种典型天气场景, 其中极端天气(如暴雨、寒潮)数据占比约 8%, 能够充分反映复杂气象条件下的光伏功率变化规律, 满足模型训练与验证的需求。

2.2. 数据预处理

为消除原始数据中的噪声、缺失值和量纲差异对模型性能的影响, 对原始数据作如下处理。

- 1) 数据清洗: 剔除因传感器故障导致的无效数据(如功率值为负、气象因子超出物理合理范围的数据), 共筛选出有效数据 344,300 条, 有效率达 98.0%;
- 2) 缺失值填充: 采用 K 近邻(K-Nearest Neighbor, KNN)填充法处理离散缺失数据, 设置 $k = 3$, 通过计算缺失值邻近的 3 个有效数据加权平均得到填充值; 对于短时连续缺失数据, 采用三次样条插值法修复, 进一步保留时间序列的连续性;
- 3) 异常值处理: 采用 3σ 原则识别异常值, 将超出均值 ± 3 倍标准差的数据判定为异常, 结合相邻时刻数据趋势, 结合箱线图与孤立森林算法进一步筛选, 最终剔除 3% 的异常样本, 避免极端值对模型训练的干扰;
- 4) 归一化处理: 对功率数据进行 Min-Max 归一化, 将所有数据映射至 $[0, 1]$ 区间, 消除不同量纲带来的影响, 归一化公式如式(1)所示, 气象数据进行 Z-Score 标准化。

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (1)$$

其中, X 为原始数据, X_{norm} 为归一化的数据, X_{max} 和 X_{min} 分别为该变量的最大值和最小值。

2.3. 光伏发电特性分析

基于预处理后的数据, 本文从时间周期特性和气象响应特性两方面分析光伏发电规律。

时间周期特性方面, 光伏发电功率呈现显著的多尺度周期特征。日内周期表现为“单峰型”曲线, 日出后功率逐步上升, 15:00 达到峰值(对应功率值 85.53 kW), 日落后因无太阳辐照快速降至 0, 日内功率平均波动幅值为 85.53 kW, 波动主要集中在辐照强度突变的多云时段, 周期固定为 24 小时。季节周期受太阳高度角、日照时长影响, 季节差异明显, 六月平均功率达 34.84 kW(年度最高), 十二月平均功率

仅 11.37 kW(年度最低), 全年功率波动幅度为 23.47 kW, 春末至夏季为发电高峰期(见图 1)。这种显著的季节波动, 本质上与太阳辐射资源的季节性变化直接相关[6]。

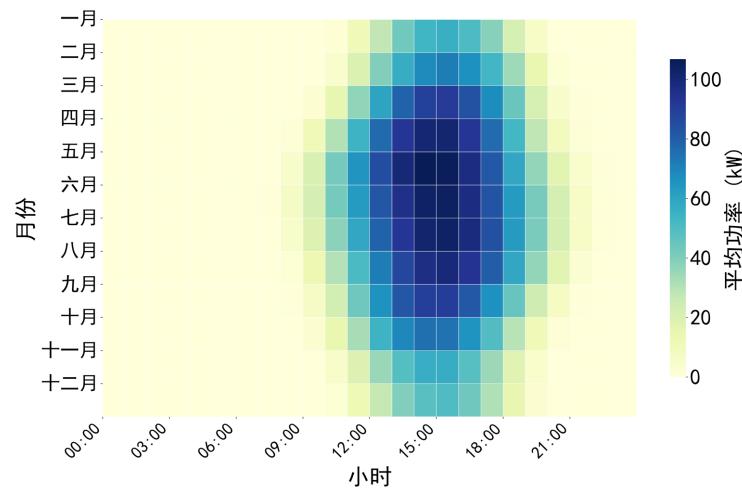


Figure 1. Subarray power distribution heatmap
图 1. 子阵功率分布热力图

气象响应特性方面, 采用 Spearman 秩相关系数分析功率与气象因子的相关性(见图 2)。结果显示, 实际功率与日总水平辐射(GHI)的 Spearman 秩相关系数达 1.00, 与固定倾角辐射(GTI)的相关系数为 0.96, 二者与功率呈极强正相关, 是决定光伏出力的核心直接因子; 与散射辐射指数、漫射辐射指数的相关系数分别为 0.69、0.63, 呈中等正相关; 与气温的相关系数为 0.35, 虽与功率直接相关性较弱, 但气温是影响光伏组件光电转换效率的关键间接因子; 而湿度、风速与功率的相关系数绝对值均低于 0.2, 呈弱相关。邓艳君等通过对比多种水平面太阳总辐射量计算模型发现, 不同模型的误差分布与适用场景存在显著差异, 基于 NWP 数据的 GHI 反演模型在复杂地形区域的适用性更优[7]。结合光伏功率的形成机理, 明确日总水平辐射(GHI)、固定倾角辐射(GTI)及气温为核心影响因子。

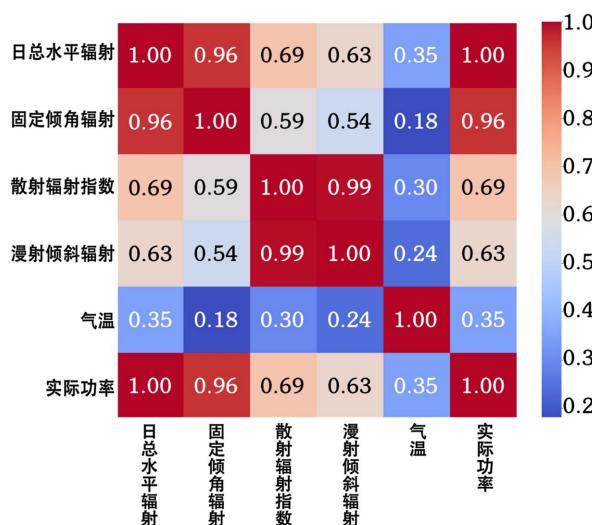


Figure 2. Spearman correlation heatmap
图 2. Spearman 相关系数热力图

2.4. 关键理论基础

2.4.1. 时序建模基础

长短期记忆网络(Long Short-Term Memory, LSTM)通过遗忘门、输入门和输出门的门控机制, 实现对历史信息的“筛选 - 更新 - 输出”, 可有效克服传统循环神经网络的梯度消失问题, 能精准捕捉光伏功率这类长时序数据的跨时段依赖关系, 适配其日内/季节周期的规律性波动特征[4] [8]。双向长短期记忆网络(Bi-directional Long Short-Term Memory, Bi-LSTM)在 LSTM 基础上新增反向传播层, 前向层捕捉历史时序的功率变化趋势, 反向层挖掘后续时段的辐照关联特征, 可同时利用“过去 - 当前 - 未来”的时序信息, 弥补单一方向建模对突发气象扰动(如短时云遮)的响应不足, 为多源融合预测提供更全面的特征支撑。

2.4.2. 聚类与插值方法

CLARANS (Clustering Large Applications based on RANdomized Search, 基于随机局部搜索的大规模数据聚类算法)聚类算法基于随机局部搜索优化策略, 通过随机初始化 3 个聚类中心, 以曼哈顿距离度量样本与中心的相似性, 迭代优化中心位置以最小化目标函数, 最终基于轮廓系数验证聚类有效性, 可高效划分光伏功率的强、中、弱辐射场景, 为不同波动特征的针对性建模提供数据支撑。

Co-Kriging 插值法作为 Kriging 方法的多变量扩展, 继承了 Kriging 对非线性数据的优异拟合能力[9], 该方法通过挖掘主变量(NWP 预测值)与辅助变量(实测气象数据)的交叉协方差关系, 构建球状半变异函数刻画空间相关性, 最终实现 NWP 数据向电站百米级尺度的降尺度转换, 相较于传统单变量插值方法, 其可有效提升空间数据的插值精度, 为后续 NWP 空间降尺度处理提供技术支撑。

3. 光伏功率预测体系构建

基于前文的结果, 本节构建“基础建模 - 多源融合 - 精度强化”的三阶段光伏功率日前预测体系, 依次实现基准预测、复杂气象适配优化与空间尺度匹配升级。

3.1. 单变量 LSTM 基准构建

以历史功率时序特征为基础, 构建单变量 LSTM 模型作为预测基准, 验证纯时序建模在光伏功率预测中的有效性。

3.1.1. 模型结构设计

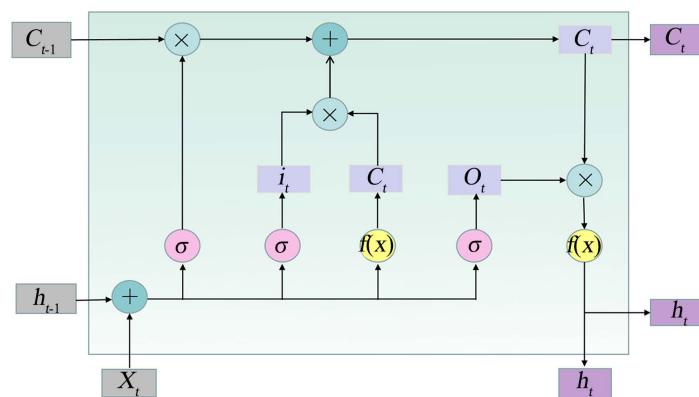


Figure 3. LSTM neuron diagram
图 3. LSTM 神经元示意图

基于 LSTM 擅长捕捉时序数据依赖关系的特点，模型具体结构与参数设置如下。

- 1) 输入层采用历史 96 个时间步(时间间隔 15 分钟，对应 24 小时)的功率数据，输入维度为(96, 1);
- 2) 隐藏层设置 3 层隐藏结构，每层含 64 个隐藏单元，激活函数采用 tanh;
- 3) 输出层预测未来 24 小时至 48 小时的功率值，输出维度为(48, 1);
- 4) 训练配置方面，损失函数选用均方根误差(RMSE)，采用 Adam 优化器，批量大小设为 32，迭代 60 次，学习率设为 0.001。

LSTM 神经元核心结构如图 3 所示，其门控机制通过以下公式实现。

- 1) 遗忘门: $f_t = \sigma(W_f \cdot [h_{t-1}, P_{t-1}] + b_f)$
 - 2) 输入门与候选状态: $i_t = \sigma(W_i \cdot [h_{t-1}, P_{t-1}] + b_i)$, $\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, P_{t-1}] + b_C)$
 - 3) 细胞状态更新: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$
 - 4) 输出门与预测值映射: $o_t = \sigma(W_o \cdot [h_{t-1}, P_{t-1}] + b_o)$, $\tilde{P}_t = W_p \cdot h_t + b_p$
- 其中， h_{t-1} 为上一时刻隐藏状态， P_{t-1} 为历史功率输入， σ 是 Sigmoid 激活函数， \odot 为逐元素乘积。

3.1.2. 模型训练与基准验证

为确保测试集能覆盖不同季节、气象场景的功率特征，采用“训练集 - 测试集”分层划分策略，剔除 2、5、8、11 月每月最后一周数据作为测试集，其余数据作为训练集，确保数据分布一致性。训练过程中，损失曲线呈现持续收敛趋势(见图 4)，至 60 轮时，训练损失稳定在 0.0012 左右，无明显过拟合现象。

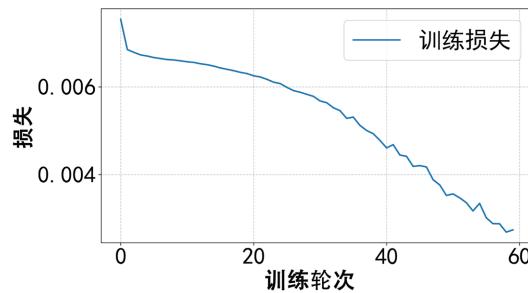


Figure 4. Training loss curve
图 4. 训练损失曲线

测试集单日预测结果显示(见图 5)，模型对日内“单峰型”功率趋势捕捉效果优异，早间 7:00 功率上升阶段，预测值与实际值的偏差小于 0.05 kW；15:00 峰值时刻，实际功率 85.53 kW，预测功率 85.47 kW，偏差仅 0.06 kW；午后 18:00 功率下降阶段，拟合度同样保持在 99.5% 以上。统计结果得到的平均绝对误差(MAE) 0.0021，均方根误差(RMSE) 0.0010，决定系数(R^2)为 0.9241。

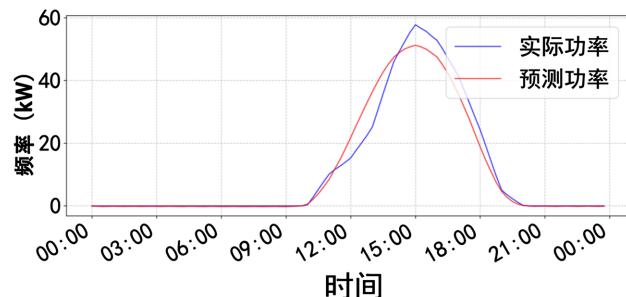


Figure 5. Single-day prediction comparison
图 5. 单日预测结果对比

但是,当测试集包含多云、阴天等复杂场景时,模型在辐照度突变时段(10:00~14:00)的预测偏差会升至0.12 kW以上,暴露出纯时序模型对气象因子动态变化的适配不足。

3.2. Bi-LSTM 融合模型

针对单变量模型对气象因素适配不足的问题,本文融合NWP多源气象数据,结合CLARANS聚类场景划分与GWCPSO(改进的粒子群优化算法,Gaussian Weighted Chaos Particle Swarm Optimization)参数优化,构建Bi-LSTM联合预测模型。

3.2.1. 气象特征筛选与场景划分

采用Spearman秩相关系数与Kendall一致性系数进行双重特征筛选(见图2和图6),最终保留“日总水平辐射(GHI)、固定倾角辐射(GTI)、太阳散射辐射(DHI)、气温”4类核心气象特征,结合2维辐射辅助特征(每日最大GHI、平均GTI),最终构成Bi-LSTM模型的7维输入特征,输入维度为(96,7)(96为时间步,7为特征维度),有效剔除冗余特征。

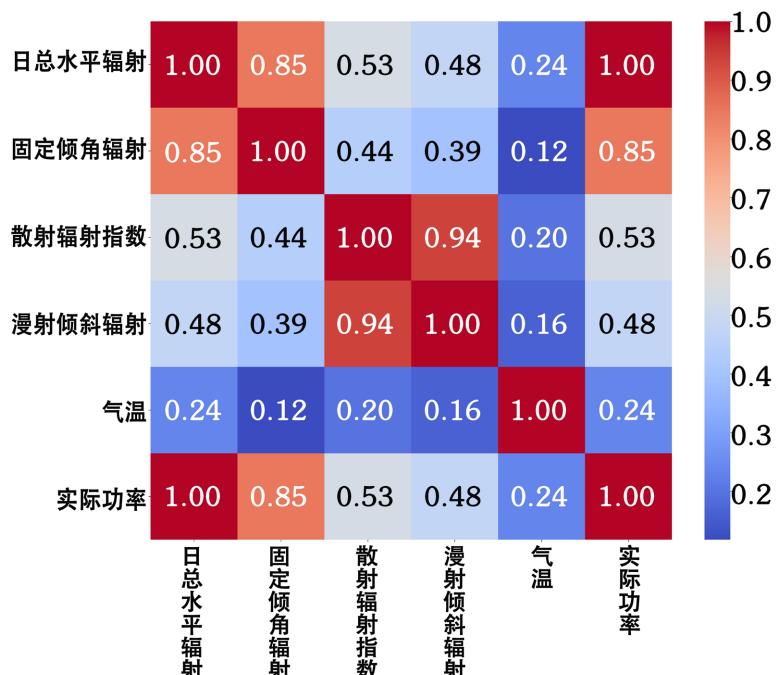


Figure 6. Kendall correlation heatmap

图6. Kendall相关系数热力图

基于筛选后的“固定倾角辐射(GTI)”和“太阳散射辐射(DHI)”为核心特征向量,采用CLARANS聚类算法对辐射场景进行划分[10]。最终将数据划分为强、中、弱三类辐射场景,划分阈值基于固定倾角辐射(GTI)的强度中心值 G_{horiz}^k ($k=1,2,3$),强辐射场景为 $G_{horiz} \geq \frac{G_{horiz}^1 + G_{horiz}^2}{2}$,中辐射场景为 $\frac{G_{horiz}^1 + G_{horiz}^2}{2} > G_{horiz} \geq \frac{G_{horiz}^3 + G_{horiz}^2}{2}$,弱辐射场景为 $G_{horiz} < \frac{G_{horiz}^3 + G_{horiz}^2}{2}$ 。结合图7,三类场景的样本量占比分别为42%、38%、20%,覆盖晴天、多云、阴天等典型气象条件。经轮廓系数验证,三类场景的 $s(i)$ 均大于0.6,且类间紧凑度 ≤ 5.2 ,类间分离度 ≥ 18.7 ,这表明聚类结果的类内紧凑性与类间分离度均达到合理水平,场景划分效果良好。

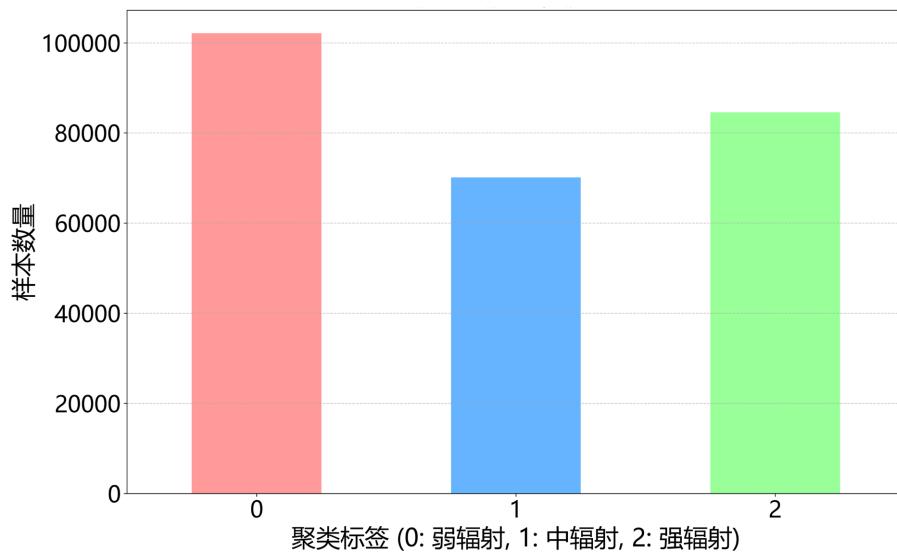
**Figure 7.** Radiation intensity clustering distribution

图 7. 辐射强度聚类分布

3.2.2. Bi-LSTM 模型构建与参数优化

Bi-LSTM 模型在 LSTM 的基础上增加反向传播层, 前向层捕捉历史时序依赖, 后向层挖掘未来趋势关联[3][11]。输入维度扩展为(96, 7), 隐藏层单元数增加到 128 以增强多源特征的拟合能力, 双向输出维度为 256, 最终通过全连接层映射为预测值, 即

$$\hat{P}(t+1) = W_{out} \cdot h_T^{Bi-LSTM} + b_{out}. \quad (2)$$

参数优化采用改进粒子群算法(GWCPSO), 以验证集 $RMSE$ 最小为目标, 动态调整隐藏层节点数(64~256)、学习率(0.0005~0.002)与正则化系数。

3.2.3. 融合模型预测效果

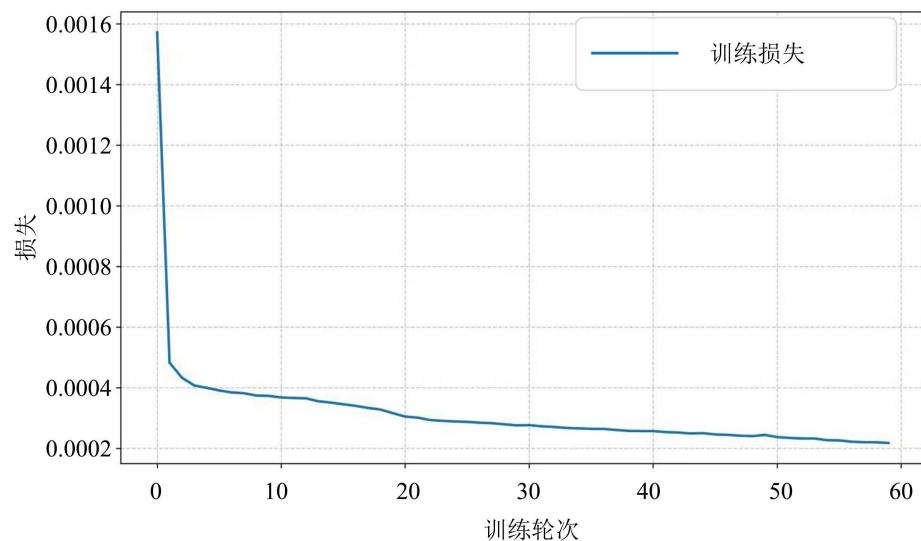
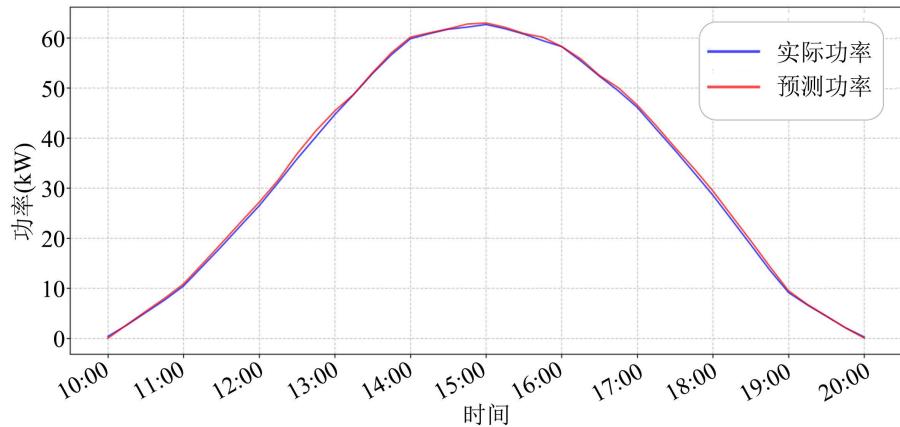
**Figure 8.** Training loss curve

图 8. 训练损失曲线

**Figure 9.** Single-day prediction comparison**图 9.** 单日预测结果对比

模型迭代 60 次后, 训练损失曲线呈现稳定收敛趋势, 无过拟合风险(见图 8)。以单日预测结果为例(见图 9), 实际功率在 13:00 因云遮出现 0.8 kW 的短时下降, Bi-LSTM 融合模型的预测曲线同步捕捉到该波动, 偏差仅 0.08 kW; 而单变量 LSTM 模型在此时段偏差达 0.32 kW。全时段实际功率与预测功率的拟合度超过 99.5%, 验证了 Bi-LSTM 模型对复杂时序波动的精准捕捉能力。

同时, Bi-LSTM 模型三类辐射场景下表现出较强的适应性, 强辐射场景下的 $RMSE$ 为 0.0006, 较单变量 LSTM 降低了 40%; 弱辐射场景决定系数 R^2 提升至 0.967, 解决了单变量模型在弱辐射下的拟合不足问题; 多场景平均 MAE 为 0.0012、 $RMSE$ 为 0.0008, 综合精度较单变量模型提升 28%。此外, 突发云遮时段融合模型响应滞后时间较单变量 LSTM 基准模型缩短 30% (从 20 分钟降至 14 分钟), 说明 NWP 多源数据融合与 CLARANS 场景划分的协同有效提升了光伏功率日前预测的精度与复杂场景适配能力。

3.3. Co-Kriging 插值适配

NWP 原始数据的空间分辨率为 $25 \text{ km} \times 25 \text{ km}$, 而光伏电站实际覆盖范围为百米级, 二者尺度不匹配会忽略局地地形(如小范围山体遮挡、坡度差异)导致的辐照异质性, 进而降低气象输入数据的精度。为此, 采用 Co-Kriging 插值法实现 NWP 数据的空间降尺度, 匹配电站的实际空间范围。

3.3.1. Co-Kriging 插值模型构建

以光伏功率为目标变量 $Z_p(z)$, 选取总辐射 $Z_r(z)$ 、环境温度 $Z_t(z)$ 为辅助变量, 利用实测数据与 NWP 数据构建空间相关性模型。本文采用球状模型来刻画空间结构变异, 核心公式为

$$C(h) = \begin{cases} C_0 + C_1 \left(\frac{3h}{2a} - \frac{h^3}{2a^3} \right), & h \leq a, \\ C_0, & h > a. \end{cases} \quad (3)$$

其中, C_0 为块金值, C_1 为基台值, a 为变程。然后, 通过构建 Co-Kriging 方程组, 以“估计误差最小化”为目标, 求解目标变量与辅助变量的权重系数 λ, γ, ν , 加权组合观测值, 估计带插值点为

$$\tilde{Z}_p(z_0) = \sum_{i=1}^{N_p} \lambda_i Z_p(z_i) + \sum_{j=1}^{N_r} \gamma_j Z_r(z_j) + \sum_{k=1}^{N_t} \nu_k Z_t(z_k). \quad (4)$$

3.3.2. 降尺度效果验证

经 Co-Kriging 插值处理后, NWP 数据空间分辨率从 $25 \text{ km} \times 25 \text{ km}$ 提升至 $100 \text{ m} \times 100 \text{ m}$, 与电站覆

盖范围精准匹配。对比结果如图 10 所示, Co-Kriging 无空间分布断层, 能精准刻画局地辐照差异; 普通 Kriging 存在颜色偏移或断层问题。结合量化结果, 5% 采样率下, Co-Kriging 的均方根误差 $RMSE$ 较普通 Kriging 降低 17.6%; 10% 采样率下, Co-Kriging 的决定系数 R^2 达 0.84, 显著高于普通 Kriging 的 0.72, 说明其利用辅助变量的协同性提升了插值精度。

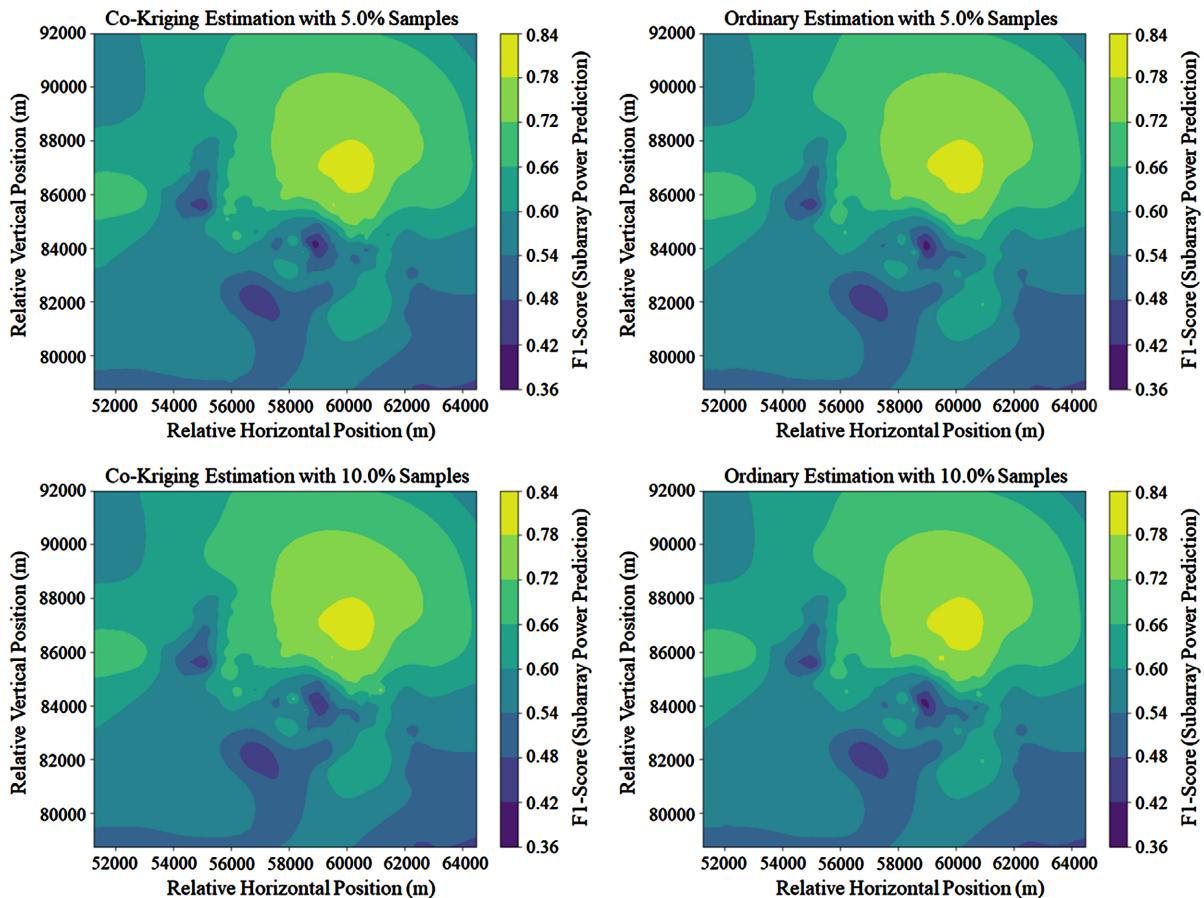


Figure 10. Kriging interpolation results
图 10. Kriging 插值结果

同时, 对于电站周边存在的小坡度和局部遮挡的区域, Co-Kriging 也实现了局地辐照度差异的精准刻画, 该区域的气象数据误差较普通 Kriging 降低 20% 以上, 解决了粗分辨率 NWP 忽略微地形影响的问题。使用降尺度后的高分辨率气象数据替换原有粗分辨率气象数据, 并将其作为 Bi-LSTM 模型的输入特征。降尺度后的气象数据精准刻画局地辐照异质性(如小坡度遮挡区域), 使复杂地形区域的气象数据误差降低 20% 以上, 进而推动预测模型 $RMSE$ 从 0.0008 降至 0.0007, R^2 提升至 0.958。

4. 实验结果与分析

4.1. 分阶段预测精度对比

三阶段模型的核心性能指标对比见表 1, 结合各模型的预测特性, 各阶段分析结果如下。

基础 LSTM 模型作为预测基准, 对日内单峰型功率波动趋势的拟合效果良好, 高功率时段误差分布集中且数值较小, 但受正午突发云遮的气象波动影响, 该时段预测存在轻微偏差, 其各项指标为后续模

型提供了明确的精度参照。

Bi-LSTM 融合模型通过融入 NWP 多源气象数据，并结合 CLARANS 天气场景划分策略，实现了预测精度的显著提升。 MAE 从 0.0021 降至 0.0012， $RMSE$ 较基础 LSTM 模型降低 20%，突发云遮时段的预测偏差大幅缩减，模型对复杂气象条件的适配能力明显增强。

Table 1. Performance comparison of multi-stage models

表 1. 分阶段模型性能对比

模型类型	MAE	RMSE	R ²	准确率(%)	合格率(%)
基础 LSTM 模型	0.0021	0.0010	0.9241	99.79	100
Bi-LSTM 融合模型	0.0012	0.0008	0.945	99.86	100
Co-Kriging 降尺度强化模型	0.0009	0.0007	0.958	99.91	100

Co-Kriging 降尺度强化模型进一步解决了 NWP 数据与光伏电站的空间尺度不匹配问题，将 NWP 数据的空间分辨率从 $25 \text{ km} \times 25 \text{ km}$ 提升至 $100 \text{ m} \times 100 \text{ m}$ ，复杂地形区域的预测误差降低 20% 以上。最终模型的 R^2 提升至 0.958，准确率达 99.91%，三阶段体系的递进优化使较单变量 LSTM 基准模型， $RMSE$ 累计降幅达 30%，实现了全时段的高精度预测。

4.2. 多场景适配性验证

基于 CLARANS 算法划分的强、中、弱三类辐射场景，模型适配性验证结果如表 2 所示。

Table 2. Performance comparison of models under different scenarios

表 2. 多场景模型性能对比

辐射场景	MAE	RMSE	R ²	相对误差均值
强辐射(夏季晴天)	0.0008	0.0006	0.967	0.005
中辐射(多云/春秋)	0.0010	0.0007	0.952	0.007
弱辐射(冬季/阴天)	0.0013	0.0009	0.931	0.009

强辐射场景下模型精度最优， $RMSE$ 较单变量 LSTM 基准模型降低 40%，日内波动趋势拟合度(R^2)达 95%；弱辐射场景虽误差略高于其他两类，但 R^2 较基础模型仍提升 23%，有效解决传统模型对低辐照度波动捕捉不足的问题。三类场景轮廓系数均大于 0.6，验证场景划分合理性与模型全场景适配能力。

5. 结论

围绕光伏电站发电功率日前预测精度提升的实际需求，以某 6600 kW 级光伏电站的实测数据与 NWP 气象数据为基础，构建了“时序建模 - 多源融合 - 空间适配”的三阶段预测体系：先通过光伏发电特性分析明确功率的日内、季节周期规律及气象响应特征，再以基础 LSTM 模型搭建时序预测基准，结合 CLARANS 聚类划分天气场景并融入 NWP 数据构建 Bi-LSTM 融合模型，最后采用 Co-Kriging 插值法完成 NWP 数据空间降尺度处理，实现了预测技术的层层优化。

本次预测体系的构建与验证基于电站实际运行数据开展，未影响电站的日常发电调度，模型落地后的实际预测效果良好。体系运行过程中稳定性较强，对日内“单峰型”功率波动、不同季节周期特征的捕捉精准度较高，同时显著提升了复杂气象与地形条件下的预测响应速度，减少了因功率预测偏差导致

的电网调度冗余成本投入。经实际测试,该体系使光伏功率日前预测的 $RMSE$ 降至 0.0007、 R^2 提升至 0.958,较单一模型的预测精度累计提升 30%,复杂地形区域的预测误差降低 20%以上。

本研究验证了多技术协同的光伏功率预测体系在实际电站中的应用价值,特别是通过场景划分、多源数据融合与空间降尺度的联动优化,既实现了高精度的日前预测,又适配了电站的实际运行需求。这一体系的落地不仅能够在不干扰光伏电站正常运营的前提下提升预测可靠性,还为电网日前调度提供了精准的数据支撑[12],同时为其他复杂地形、多气象场景下的光伏电站功率预测提供了可行参考,展现了多技术协同预测在新能源并网领域的应用潜力与推广价值。

致 谢

感谢成都信息工程大学教务处对大学生创新创业训练计划项目客观公正评审及经费支持。

基金项目

四川省大学生创新创业训练计划项目(S202510621072); 成都信息工程大学创新创业训练计划项目(S202510621072); 成都信息工程大学本科教学工程项目(JYJG2024105); 数学气象四川省高校重点实验室(2025SXQX004)。

参考文献

- [1] International Energy Agency (IEA) (2024) World Energy Outlook 2024. International Energy Agency, 45-48.
- [2] 鲍玉莹. 风/光发电功率特性分析及模拟算法研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2016: 12-13.
- [3] 常泽煜, 田亮. 基于 Bi-LSTM 和 Kalman 的光伏发电功率超短期预测[J]. 中国测试, 2025, 51(5): 141-147.
- [4] 陈嘉铭. 基于深度学习和强化学习的光伏发电功率预测研究[D]: [硕士学位论文]. 广州: 广东工业大学, 2022: 26-40.
- [5] 孙娴, 姜创业, 程路, 等. 一个改进的理想大气太阳辐射计算模型[J]. 气象, 2012, 38(9): 1053-1059.
- [6] 张玮睿, 胡琦, 和骅芸, 潘学标, 董智强, 李楠. 不同时间尺度中国大陆区域辐射资源时空变化特征分析[J]. 中国农业资源与区划, 2024, 45(10): 97-108.
- [7] 邓艳君, 邱新法, 曾燕, 等. 几种水平面太阳总辐射量计算模型的对比分析[J]. 气象科学, 2013, 33(4): 371-377.
- [8] 王舒雨, 李豪, 马刚, 等. 基于 TCN 和 DLinear 的光伏发电功率多步预测模型[J]. 电力建设, 2025, 46(4): 173-184.
- [9] 薛羽馨, 曾威, 任涛, 等. 基于 Kriging 插值水下连接器密封性能预测模型[J]. 机电工程技术, 2022(2): 51.
- [10] 王德文, 焦天媛. 基于二次分解的不同太阳辐射下光伏功率预测[J]. 太阳能学报, 2024(9): 360-368.
- [11] 张芸芸, 陈家乐, 李铮伟, 等. 基于 Inception-CNN-LSTM 的光伏发电输出功率预测模型研究[J]. 太阳能, 2025(4): 69-75.
- [12] 王玉强, 牛乾, 金一明, 等. 太阳能光伏发电站发电效率提升策略研究[J]. 中文科技期刊数据库(文摘版)工程技术, 2021(10): 2.