

报名序号：004010

赛题题目:A 题: 光伏电站发电功率日前预测问题

A 题: 光伏电站发电功率日前预测问题

摘 要

随着国家“碳达峰、碳中和”战略的深入推进，以光伏为代表的新能源成为构建新型电力系统的核心支柱。光伏发电的间歇性与波动性对电网功率平衡及频率调节提出了严峻挑战，提升其发电功率预测精度是保障电网安全运行的关键技术问题。本题围绕光伏电站发电功率日前预测展开研究，结合历史功率数据与数值天气预报（NWP）信息，构建多场景预测模型，分析地理、气象及季节因素对预测精度的影响，为电力调度部门优化资源配置提供理论支撑^[1]。

针对问题一：首先基于光伏电站地理位置信息构建太阳辐照与光伏转换模型，计算其理论可发功率；其次，解析光伏发电的长周期与短周期特性，发现光伏发电长周期具有显著的季节性规律，短周期与辐照强度有强相关性；然后，引入均方根误差和平均绝对偏差量化实际与理论功率的偏差，表明夏季辐照稳定时精度最高（ $E_{rmse} < 5\%$ ），冬季因散射光及低辐照影响偏差显著（ $E_{rmse} > 15\%$ ），为后续预测模型优化提供关键数据特征。

针对问题二：基于第一题中的分析，为了针对光伏发电功率预测中传统线性模型对光伏出力复杂时序波动适应不足的问题，本题提出了 LSTM 与 SVR 融合预测模型。实验结果表明，本题所用模型使均方根误差较线性回归降低 73.7%，相关系数提升至 0.8947，同时合格率为 93.30%，有效验证了该混合模型对光伏出力动态特征的精准刻画能力。

针对问题三：由于第二题中仅仅之利用了历史功率进行光伏发电预测，对 NWP 信息的利用并不充分。所以本题提出一种融合空间特征与时序依赖的双通道 CNN-BiLSTM-AM 预测模型。实验表明，该模型在 2/5/8/11 月测试集的均方根误差（ E_{rmse} ）较 LSTM-SVR 降低 3.1%，相关系数提升至 0.9068，说明 NWP 信息可以有效的提升光伏预测精度。

为进一步解决季节交替导致的时序混淆问题，引入最大互信息系数（MIC）筛选总云量、地表辐照和温度三大关键气象因子，通过加权 K-means 聚类划分春/夏/秋/冬四类典型场景。场景化建模后，模型的预测合格率（ C_R ）提升至 94.3%，系统性偏差（ E_{mc} ）趋近于零，进一步验证了划场景后再进行预测可以步提升其精度。

针对问题四：由于大尺度数的 NWP 信息已无法满足 MW 级光伏电站精细化功率预测需求。所以本题提出基于回声状态网络（ESN）的 NWP 空间降尺度方法：通过 ESN 特有的稀疏储层结构，融合高分辨率地理信息与历史观测数据，成功构建了多尺度气象场的动态映射关系。实证结果表明，通过解析辐照度和云层突变特征，最终将功率预测均方根误差（ E_{rmse} ）从 0.0591 降至 0.0474。

关键词：光伏预测、双向长短期记忆网络、卷积神经网络、注意力机制、地理条件

0.问题重述

0.1 问题背景

2020 年 9 月，我国明确提出“碳达峰、碳中和”战略目标，要求加快构建以新能源为主体的新型电力系统，推动可再生能源规模化替代化石能源。截至 2023 年底，全国光伏装机容量已突破 6 亿千瓦，连续八年位居全球首位，但其出力受太阳辐照、气象条件等因素影响显著，呈现强间歇性与波动性。典型光伏电站的日内功率波动幅度可达装机容量的 80%以上，在阴雨天气下更会出现断崖式下跌，这种特性导致光伏发电功率预测不确定性高，加剧了电力系统供需平衡压力。据国家电网统计，光伏预测误差每增加 1 个百分点，将导致系统备用容量需求增加约 150 万千瓦，直接影响电网经济运行效率。

在此背景下，光伏发电功率日前预测作为电网调度的核心输入，其精度直接决定调度计划的合理性与经济性。传统预测方法多依赖单一气象要素或简单统计模型，难以应对复杂天气演变过程。值得关注的是，国家能源局在《新型电力系统发展蓝皮书》中明确要求，到 2025 年光伏功率预测准确率应提升至 90%以上。因此，融合气象信息与历史数据，构建高精度光伏发电预测模型，不仅是提升电网消纳能力、保障运行安全的技术基础，更是推动“双碳目标”下源网荷储协同优化的必然要求。

0.2 问题提出

为了更好的对光伏电站的出力进行预测，首先便是分析该光伏电站的理论发电功率以及其长/短周期的发电特性。接着，依次分析不同因素对光伏出力预测所带来的影响。例如，总云量、辐照强度、温度、风速等。并结合现阶段的强化学习，神经网络等技术，采用不同的预测方法进行预测评估。最后，研究如何在有限的气象数据下，能够更精确的对光伏电站进行出力预测，从而最终解决电网对风电的调度难题。

本次 4 个问题可以从以下 5 角度进行分析：

- 1.结合光伏电站地理位置，根据太阳辐照计算理论进行光伏电站长短周期的发电特性分析。
- 2.基于历史功率，利用深度学习的知识，对光伏发电进行预测。
- 3.研究融合 NWP 信息后对光伏发电会带来什么样的影响。
- 4.进一步划分不同场景，并进行预测模型的更新，对比不同模型光伏发电预测的精度。
- 5.研究降低传统气象预报空间的尺度的方法，使其能够适应当今 MW 级发电站精细化预测的要求。

0.3 数据来源

表 1 光伏电站的历史发电功率和 NWP 数据规格及要求

数据规格	参数值
数据来源(公开数据集链接)	https://www.datafountain.cn/datasets/7021
光伏电站装机容量	<u>1.5</u> MW
发电功率和 NWP 数据时间分辨率	采样点/15min
发电功率和 NWP 数据起始-截至时间(一年)	2022.01.03- 2023.01.02
NWP 属性	地表气压(Pa)、相对湿度(%）、总云量(%)、10 米风速(m/s)、10 米风向(°)、2 米气温(K)、辐照强度(J/m²)、降水量(m)、100 米风速(100m/s)

具体数据见附件。

问题 1：基于历史功率的光伏电站发电特性分析

1.1 解题思路

该题需要基于光伏电站经纬度、海拔及时区计算太阳角度，结合地理位置信息与太阳辐照计算理论，建立太阳辐照与光伏转换的模型，用于计算得到光伏电站的理论可发功率。

1.2 数据处理

针对光伏气象大数据集的预处理工作主要分为两个核心部分：单位标准化与数据分类。通过统一单位体系和按时间维度划分数据集，提升数据的可解释性并适配后续模型的输入需求。

1.2.1 单位标准化

为确保数据量纲一致，需对以下数据进行单位转换：

温度：开尔文（ K ） \rightarrow 摄氏度（ $^{\circ}C$ ），温度（ $^{\circ}C$ ）=温度（ K ）-273.15

气压：帕斯卡（ Pa ） \rightarrow 百帕（ hPa ），气压（ hPa ）=气压（ Pa ）/100

降水：米（ m ） \rightarrow 毫米（ mm ），降水（ mm ）=降水（ m ） $\times 100$

辐照强度：焦耳每平方米（ J/m^2 ）（表示累计辐射能量） \rightarrow 瓦特每平方米（ W/m^2 ）

（表示瞬时辐照强度），辐照强度（ W/m^2 ）= $\frac{\text{辐照强度}(J/m^2)}{\text{时间间隔}(s)}$ ，本题所用光伏气象大数据

集时间间隔为 15 分钟即 900 秒。

1.2.2 数据分类

基于大数据集时间列，将数据按日、月、季节划分，以适配不同时间尺度的分析需求。

按日分类：分析日内光伏发电的周期性波动。

按月分类：捕捉不同月份气象变化对光伏发电效率的影响（如雨季与旱季的辐照差异）。

按季节分类：以北半球为例，春季为 3-5 月，夏季为 6-8 月，秋季为 9-11 月，冬季为 12-次年 2 月。用以研究季节性气象模式（如冬季低辐照强度、夏季高云量）对光伏发电的影响。

1.3 模型建立

1.3.1 太阳辐照模型

在计算光伏板所接收到的光照强度之前，需要引入表示太阳与地球的相对位置的太阳赤纬角 δ 、时角 H 、太阳高度角 β 、太阳方位角 ϕ_s ，计算方法为：

$$\delta = 23.45 \sin\left[\frac{360}{365}(n - 81)\right] \quad (1)$$

$$H = \frac{15^{\circ}}{h} t_m \quad (2)$$

$$\beta = \arcsin(\cos L \cos \delta \cos H + \sin L \sin \delta) \quad (3)$$

$$\phi_s = \arcsin\left(\frac{\cos \delta \sin H}{\cos \beta}\right) \quad (4)$$

其中， n 是日期，以 1 年为 365 天计； t_m 为当前时刻与正午时刻所相距的小时数； L

为当地纬度； h 为单位小时数。

为了计算昼长时间内的光伏板可以接受到的光照强度，首先需要计算太阳光直射地表的光照强度^[2]。

$$I_B = I_{SC} \exp\left(-\frac{k}{\sin \beta}\right) \quad (5)$$

$$I_{SC} = 1370\left[1 + 0.034\left(\frac{360n}{365}\right)\right] \quad (6)$$

$$k = 0.174 + 0.035 \sin\left[\frac{360}{365}(n - 100)\right] \quad (7)$$

其中， I_B 是太阳光直射地表的光照强度； I_{SC} 是大气层上接收的总光照强度； k 是光学深度，由经验公式得到。

在晴天条件下，光伏板可以接受到的光照强度 I_C 包括直射光照强度 I_{BC} ，散射光照强度 I_{DC} ，反射光照强度 I_{RC} ，具体计算方法如下所示：

$$I_C = I_{BC} + I_{DC} + I_{RC} \quad (8)$$

$$I_{BC} = I_B \cos \alpha \quad (9)$$

$$I_{DC} = CI_B \left(\frac{1 + \cos \xi}{2}\right) \quad (10)$$

$$I_{RC} = \rho I_B (\sin \beta + C) \left(\frac{1 - \cos \xi}{2}\right) \quad (11)$$

$$\alpha = \arccos[\cos \beta \cos(\phi_s - \phi_c) \sin \xi + \sin \beta \cos \xi] \quad (12)$$

$$C = 0.095 + 0.04 \sin\left[\frac{360}{365}(n - 100)\right] \quad (13)$$

其中， α 是太阳光入射角，即光伏板表面法线与入射光线间的夹角； C 是散射因数； ξ 是光伏板倾斜角，即光伏板与地平面的夹角； ρ 是地面反射率； ϕ_c 是光伏板方位角。为了便于理解，附录图表示出了这些角度。

1.3.2 光伏转换模型

光伏电池板的额定功率一般是在标准测试条件（STC）下确定的，即在 1000 W/m^2 的辐照度和 25°C 电池温度条件下。对于光伏发电量影响最大的是辐射量，其次就是温度。光伏发电的效率随着温度的增加而降低，而电池组件表面的风加速散热，故与温度的影响相反，即随着风速的增加光伏发电的效率会增加。因此，有必要根据当地的具体气象条件，通过具体的温度模型来模拟光伏电池的温度。

本题采用桑迪亚阵列性能模型（Sandia Array Performance Model, SAPM）来计算光伏电池的温度 T_{cell} ，其公式为：

$$T_{cell} = T_m + \frac{G_c}{1000} \Delta T \quad (14)$$

其中， ΔT 为在 1000 W/m^2 的条件下，光伏板的后表面模块温度与电池温度之间的温差； T_m 为后表面模块温度，其公式为：

$$T_m = G_c \exp(a + b \times WS) + T_a \quad (15)$$

其中， T_a 为环境温度； WS 为 10 m 高度的风速； a 和 b 为系数，主要由光伏模块的结构和安装配置来确定。 ΔT 、 a 和 b 的值一般由光伏厂商来提供。在本题中， ΔT 的值取 3， a 的值取 -3.56 ， b 的值取 -0.0045 。

在获得倾斜面总辐射量和电池温度后，便可计算光伏发电的直流功率。本题采用 NREL 所开发的 PVwatts 模型，其公式为：

$$P_{solar} = \frac{P_{STC} G_c}{1000W / m^2} [1 + \gamma(T_{cell} - T_{ref})] \quad (16)$$

其中, P_{STC} 为在 STC 条件下光伏板的额定功率; γ 是太阳能电池的温度系数, 一般取值约为 $-0.4\% / ^\circ C$; T_{ref} 为设定温度即 $25^\circ C$ 。

1.4 结果分析

1.4.1 长周期（季节性变化）特性分析

针对光伏发电长周期（季节性变化）的特性分析, 本题选取 2022 年光伏数据进行功率特性分析, 计算得月理论发电功率总量如图 1.1 所示。

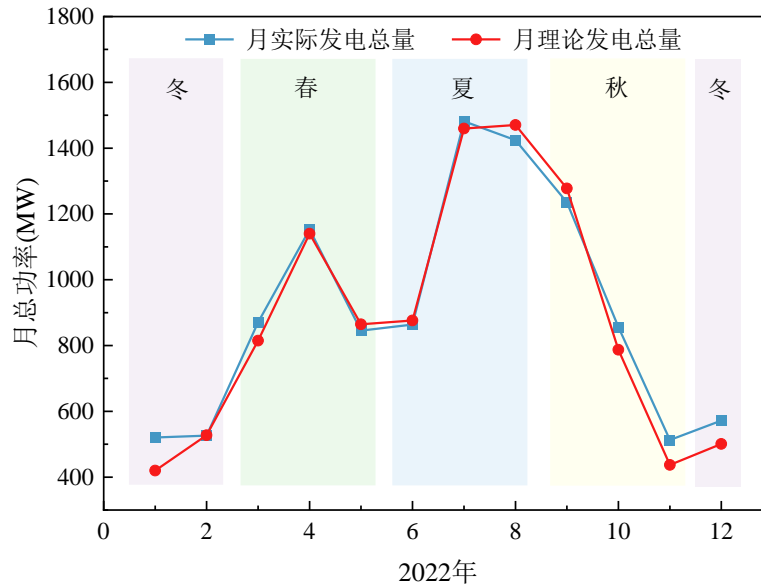


图 1.1 2022 年光伏月发电功率总量

由图可知, 光伏电站的发电特性呈显著的季节性规律, 夏季为发电高峰期, 冬季为低谷期。春季 (3-5 月) 光伏电站发电量随日照时间增加而上升; 夏季 (6-8 月) 发电量因日照时间长、辐照强度高而达到峰值; 秋季 (9-11 月) 发电量逐步下降; 冬季 (12 月到次年 2 月) 因日照时间最短且太阳高度角最低, 发电量达到谷底。

1.4.2 短周期（日内波动）特性分析

针对光伏发电短周期（日内波动）的特性分析, 本题选取 2 月、5 月、8 月与 11 月的最后一天进行功率特性分析, 计算得日内理论发电功率如图 1.2 所示。

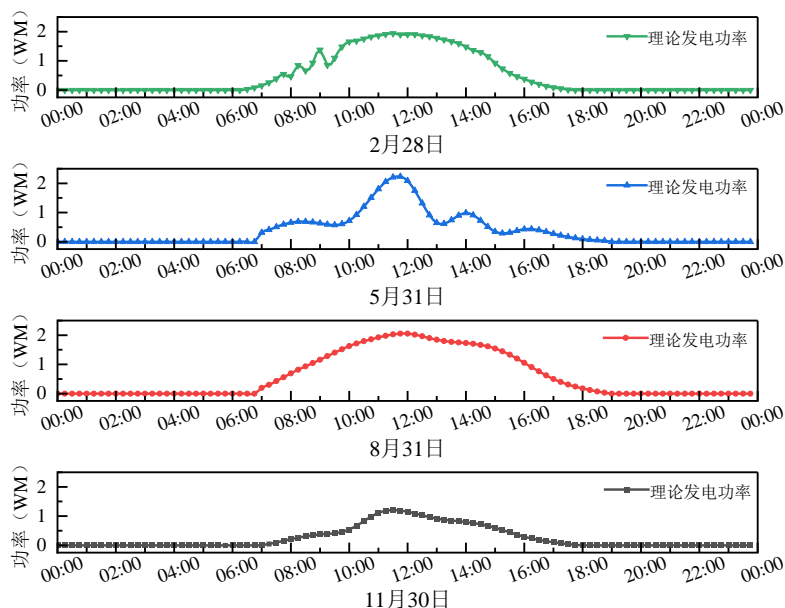


图 1.2 2022 年典型日光伏发电功率

通过观察分析可以得出日内发电功率的典型规律。早晨至正午，发电功率随太阳高度角升高、辐照强度增强而逐步上升，通常在正午前后达到峰值。此时太阳直射角度最佳，光伏板接收的辐照强度最高；正午后至傍晚，发电功率随太阳高度角下降逐渐降低，但午后云量变化或局部气象波动可能导致短时功率波动；夜间，发电功率趋近于零，因无太阳辐照。

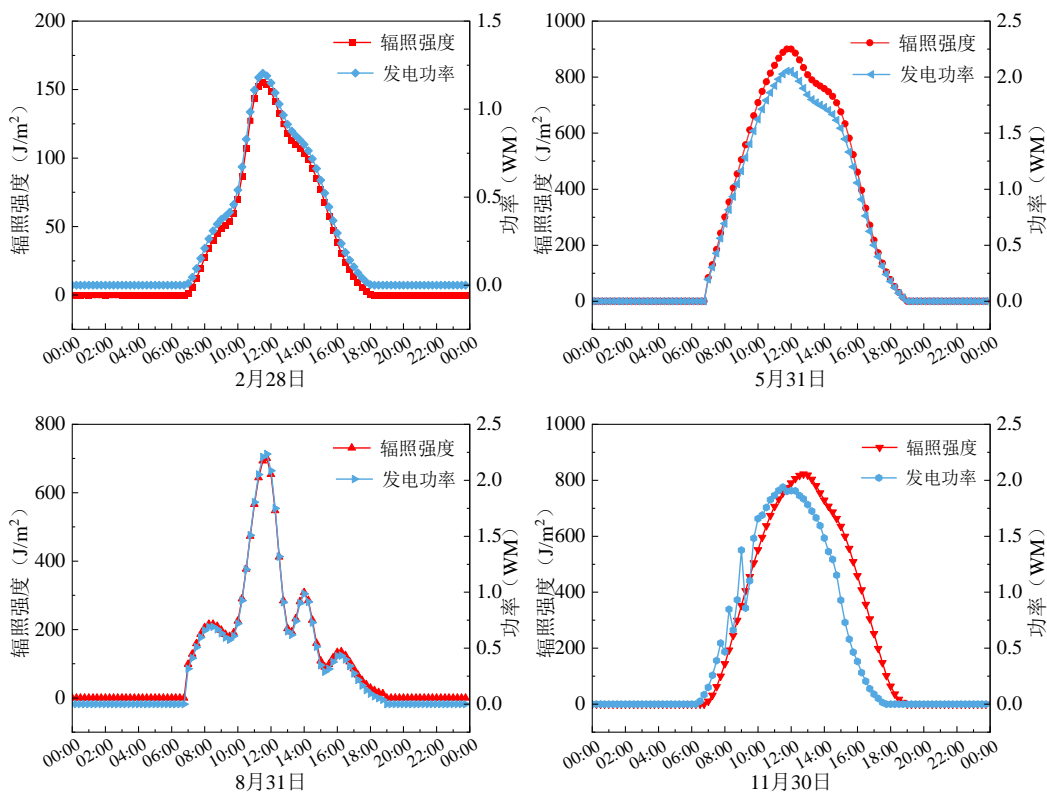


图 1.3 光伏发电功率与辐照强度关系

由图 1.3 可知辐照强度与光伏发电功率具有强相关性，发电功率曲线与辐照强度曲

线高度同步，表明辐照强度是日内发电量的主导因素。辐照强度的瞬时波动（如正午后的短时下降）直接导致发电功率的同步下降，反映了光伏发电对光照条件的敏感性。

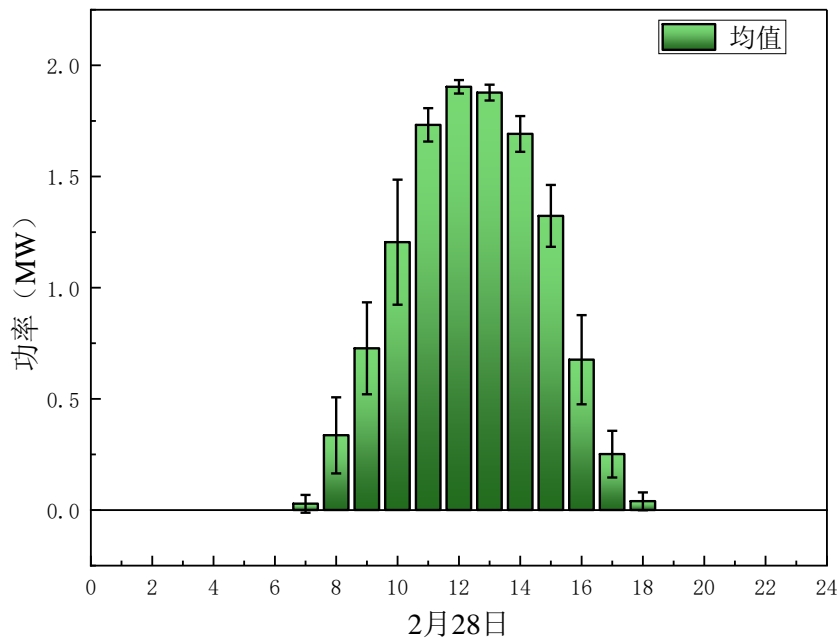


图 1.4 光伏发电每小时平均功率及波动范围

对光伏电站日内发电波动强度进行分析，如图 1.4 所示。高波动时段出现在早晨（7：00-9：00）、午后（13：00-15：00）及傍晚（16：00-18：00），低波动时段出现在正午（11：00-13：00）及夜间。早晨太阳高度角快速变化导致辐照强度波动显著，功率波动幅度较大；午后受局部气象（如云量突变、阵风散热）影响，功率可能出现短时剧烈波动；傍晚太阳高度角迅速下降叠加散射光影响，波动幅度再次增大。正午辐照强度稳定，功率波动幅度最小；夜间无发电活动，波动幅度为零。

1.4.3 偏差分析

引入均方根误差与平均绝对偏差，对光伏电站长/短周期发电功率进行偏差特性分析。

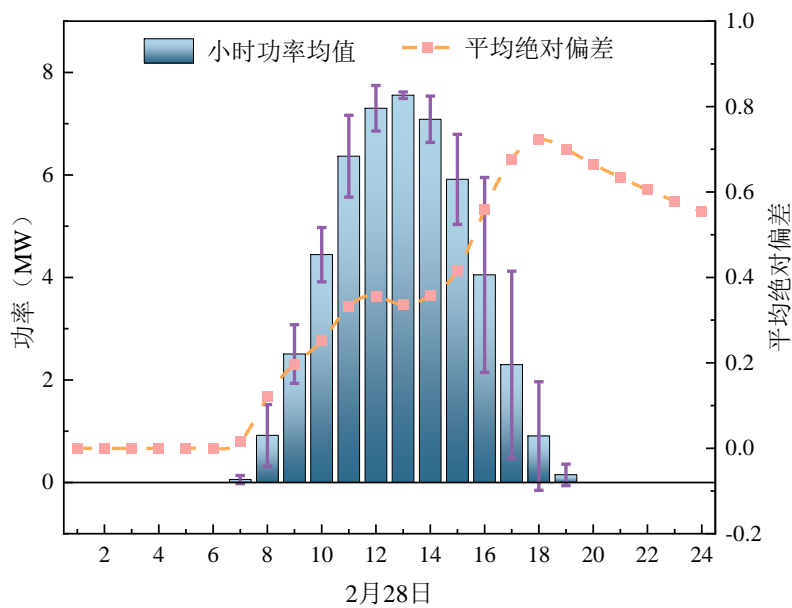


图 1.5 典型日光伏实际功率与理论可发功率均方根误差与平均绝对偏差

由图 1.5 分析可知，2 月 28 日，正午时段发电功率均值最高，但 RMSE 和 MAE 相对较低，表明在辐照强度最大的时段计算较准确。而日出和日落时段（如早晨 6-8 点、傍晚 16-18 点），功率均值较低，但 RMSE 与 MAE 显著升高，可能与太阳高度角快速变化、云层遮挡或局部气象波动（如晨雾、晚霞反射）未被计算模型充分捕捉有关。

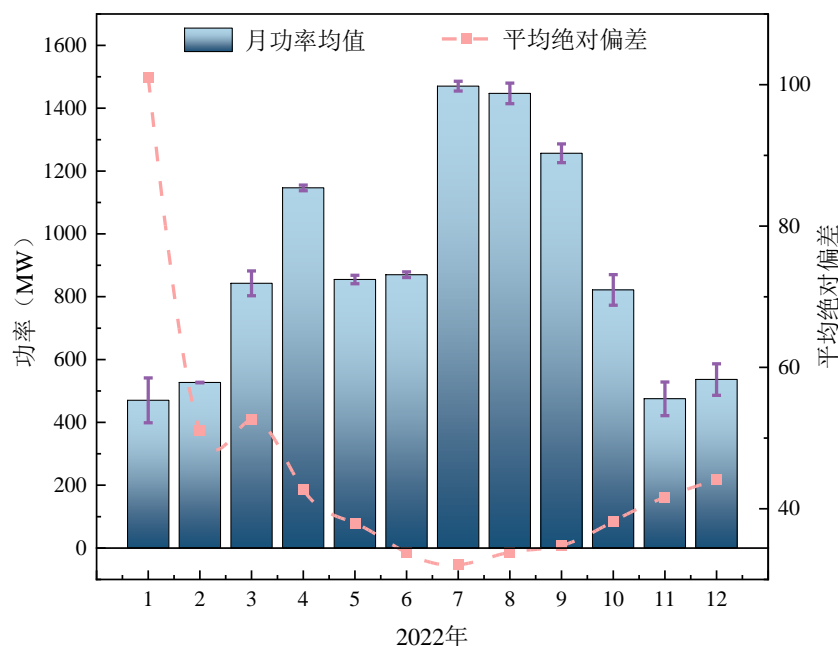


图 1.6 2022 年光伏月实际功率与理论可发功率均方根误差与平均绝对偏差

由图 1.6 分析可知，夏季（6-8 月）：RMSE 与 MAE 均处于全年最低水平，可能与计算模型对高温、高辐照条件的模拟较精确有关。冬季（12-2 月）：RMSE 与 MAE 显著升高，可能因模型低估了低辐照条件下散射光的影响。过渡季节（春、秋季）：偏差呈波动状态，可能与计算模型对多云、降水等复杂气象条件的适应性不足有关。

问题 2：建立融入 NWP 信息的光伏电站日前发电功率预测模型

2.1 解题思路

问题要求建立基于历史功率的光伏电站日前发电功率预测模型，进行发电功率预测，故提出了一种基于长短期记忆网络(Long Short-Term Memory,LSTM)和支持向量回归(Support Vector Regression,SVR)混合模型的光伏功率预测方法。通过对 SVR 模型引入松弛变量，对 LSTM 模型中较大的预测误差进行修正，从而达到更好的预测准确度。

2.2 LSTM-SVR 混合预测模型

2.2.1 LSTM 神经网络模型

LSTM 不同于传统的前馈神经网络，其在卷积神经网络的结构上改进了隐藏层的内容，不仅添加了外部自循环模块，同时也加入了细胞模块并包含了输入门 input 为 f_t 、遗忘门 forget 为 i_t 、输出门 output 为 o_t 三类不同的门及储存单元。输入门所代表的是输入给控制记忆单元有用的信息数据，输出门的含义是实输出信息数据，遗忘门的含义在于遗忘控制记忆单元的无用信息数据。

在 t 时刻，遗忘门将本单位的 x_t 和上一单元的输出 h_{t-1} 作为输入，通过激活函数得到 f_t ，并将 f_t 传递给上一时刻细胞状态 C_{t-1} ， f_t 计算公式为：

$$f_t = \sigma(W_f[c_{t-1}, h_{t-1}, x_t] + b_f) \quad (17)$$

其中， σ 为 SIGMOID 函数； W_f 为遗忘门与输入 x_t 和中间输出 h_{t-1} 乘积的矩阵权重； b_f 为遗忘门的偏移量。

而输入门将本单元的输入 x_t 和上一单元的输出 h_{t-1} 作为输入，通过激活函数得到输入门的输出 i_t ，同时产生 T 时刻细胞候选状态 c_t ：

$$c_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (18)$$

其中，TANH 为候选状态单元的激活函数； W_c 为候选状态单元与输入 x_t 和中间输出 h_{t-1} 乘积的矩阵权重； b_c 为候选状态单元的偏移量。更新当前时刻的细胞状态。而当前时刻的 c_t ，可由 f_t 和 i_t 分别与上一时刻 c_{t-1} 和当前时刻 c_t 组合得到：

$$c_t = f_t c_{t-1} + i_t c_t \quad (19)$$

而细胞单元中输出门结果和可记忆信息的隐藏层状态计算如下：首先将本单元的输入 x_t 和上一单元的输出 h_{t-1} 作为输出门的输入，然后通过激活函数得到输出门的结果 o_t ，将更新后的细胞状态值使用 TANH 函数激活，再与 o_t 相乘得到最终的输出结果，其中 o_t 和 h_t 为

$$\begin{aligned} o_t &= \sigma(W_o[c_t, h_{t-1}, x_t] + b_o) \\ h_t &= o_t \tanh(c_t) \end{aligned} \quad (20)$$

其中， W_o 为遗忘门与输入 x_t 和中间输出 h_{t-1} 乘积的矩阵权重； b_o 为输出门的偏移量； x_t 为输入参数； h_t 为输出参数； o_t 为输出门的输出参数； c_t 为 T 时刻神经元的状态。

2.2.2 支持向量机回归

SVR 线性回归模型为：

$$f(x) = \omega\phi(x) + b \quad (21)$$

首先, SVR 的模型预测结果允许存在一定的误差, 若光伏功率的预测值在 $f(x)$ 两侧的两倍误差内, 就可以当作预测结果是正确的。为了构建的 SVR 结构风险最小化, 可以引入松弛变量和间隙距离, 将其转化为二次规划凸问题:

$$\min \frac{1}{2} w^2 + C \sum_{i=1}^m \xi_i - \xi_i^* \quad s.t. \begin{cases} f(x) - y_i \leq \varepsilon + \xi_i \\ y_i - f(x) \leq \varepsilon + \xi_i^* \end{cases} i = 1, 2, \dots, m, \xi_i, \xi_i^* \geq 0 \quad (22)$$

同时利用对偶原理, 引入拉格朗日乘子 α_i 和 α_i^* , 具体数学表达为:

$$\begin{aligned} \max_{\alpha, \alpha^*} & \sum_{i=1}^m y_i (\alpha_i^* - \alpha_i) - (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m y_i (\alpha_i^* - \alpha_i) (\alpha_i^* + \alpha_i) x_i^T x_j \\ s.t. & \sum_{i=1}^m (\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha_i^*, \alpha_i \leq C \end{aligned} \quad (23)$$

而对上述二次凸规划问题进行求解, 即可获得预测后的光伏出力数据。

2.2.3 LSTM-SVR 混合模型

在 LSTM-SVR 混合模型中, LSTM 网络拥有识别和存储这一特性, 而 RNN 网络的循环构造在挖掘模型深层特征中具有显著的优势。同时 SVR 模型因其对于非线性函数强大的映射功能, 对于细节分量数列的预测、SVR 参数集合的搜寻及择优选取有着很好的效果。

光伏数据属于时间序列数据, 具有极大的不确定性, 采用单一算法模型不能准确把握光伏变化规律, 实现光伏精准预测。为充分发挥组合预测模型优秀的预测性能, 引入 SVR 算法对 LSTM 进行误差修正, 利用 LSTM-SVR 组合模型实现充电站光伏预测。利用 SVR 模型引入松弛变量法, 对 LSTM 每个时刻的误差 e_t 进行修正, 有利于优化光伏预测准确度。具体求解流程如图 2.3。

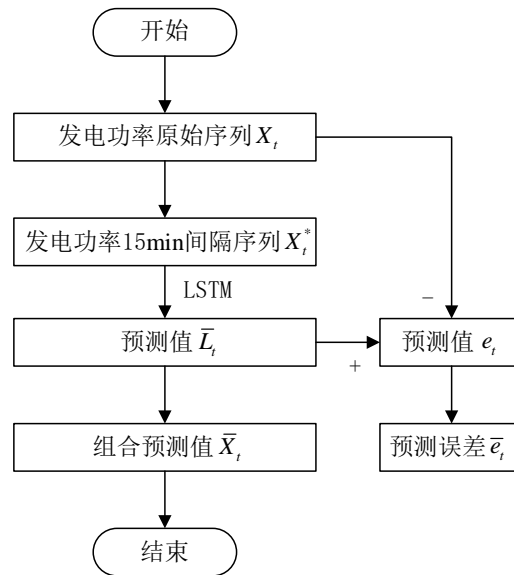


图 2.3 LSTM-SVR 混合模型求解流程

步骤一: 清洗数据。对数据进行预处理, 消除数据中的错误或异常因素。

步骤二: 将光伏原始数据归一化处理得到 X_t^* , 利用训练集对 LSTM 模型进行训练, 利用训练好的 LSTM 模型预测输出一组预测值 \bar{L}_t 。

步骤三：计算 LSTM 模型在 t 时刻的预测误差，即 $e_t = x_t^* - \bar{L}_t$ ；

步骤四：采用 SVR 对 LSTM 模型在 t 时刻的预测 e_t 误差进行回归预测，并通过交叉验证和网格搜索寻找最优的核函数参数 g 和惩罚因子 c ，并得到预测值。

步骤五：合并 LSTM 模型预测值 \bar{L}_t 与修正预测误差 \bar{e}_t ，得到基于 LSTM-SVR 组合算法模型的光伏预测结果 \bar{X}_t ，即 $\bar{X}_t = \bar{L}_t + \bar{e}_t$ 。

2.3 结果分析

为了体现出本题采用的 LSTM-SVR 模型具有可行性，利用传统线性回归的方式进行对比实验。采用题目要求的 2、5、8、11 月的最后一周为测试集，其他为训练集，其中图 2.4 为传统线性回归的预测结果，图 2.5 为本题所用 LSTM-SVR 混合模型预测结果。同时按照附录 1 中的误差计算方式，对比两种不同预测方法，具体结果如表 2.1。

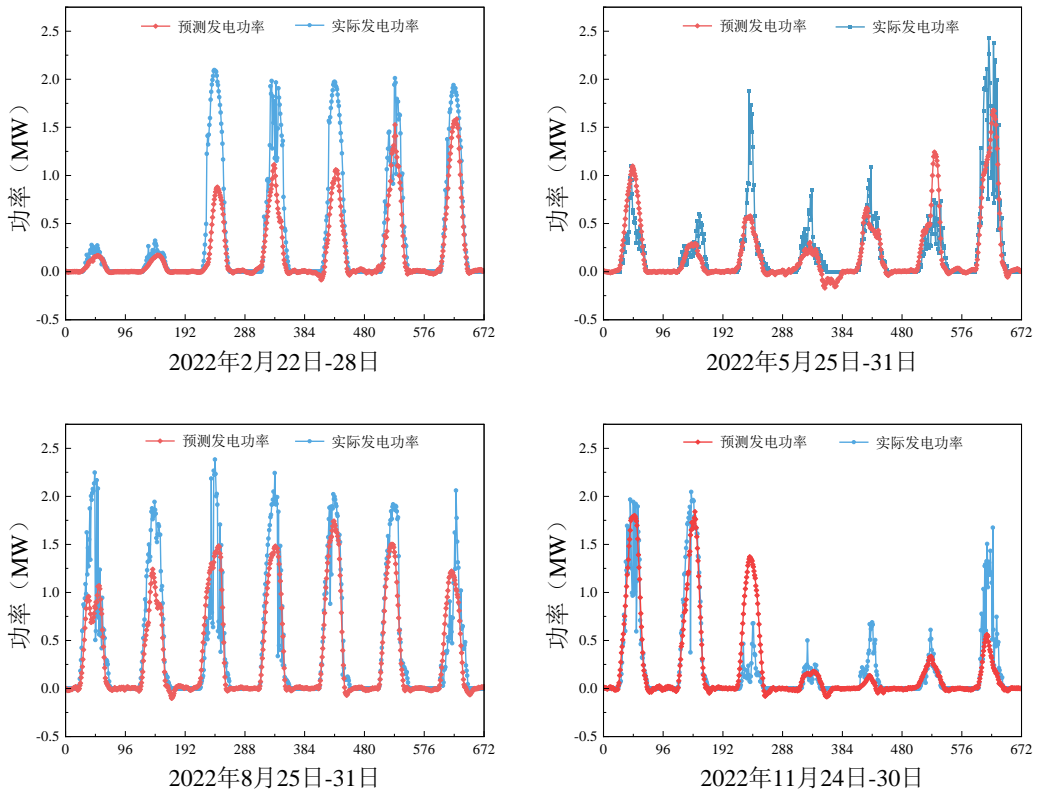


图 2.4 传统线性预测结果

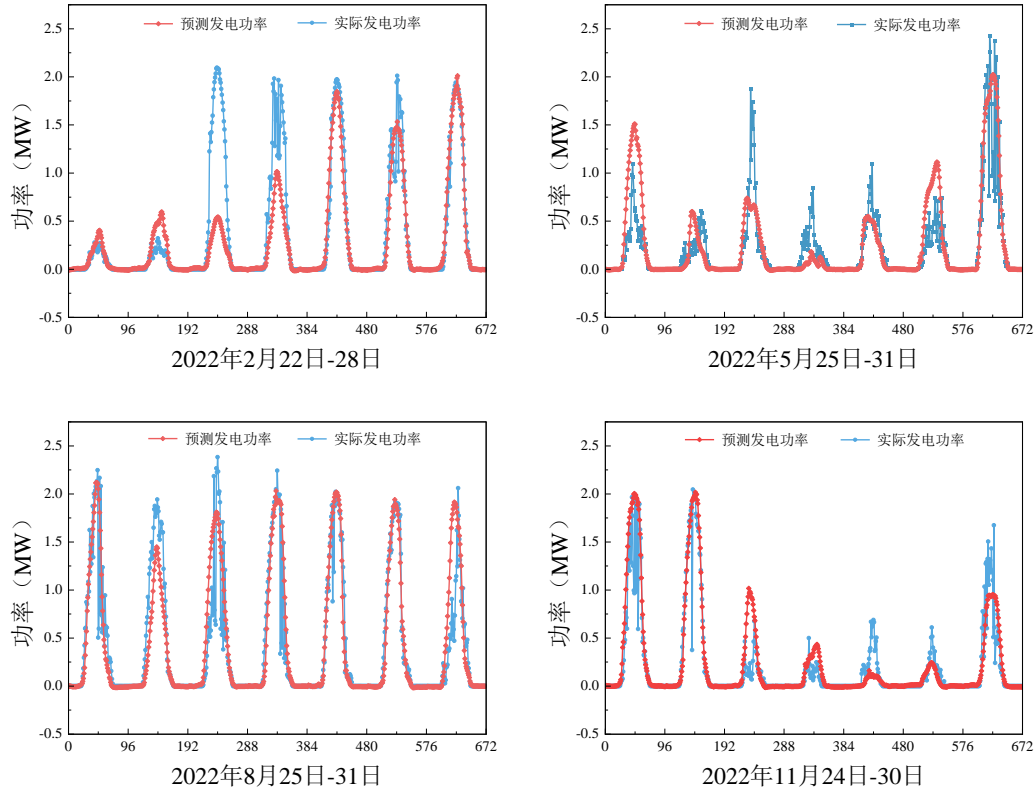


图 2.5 LSTM-SVR 混合模型预测结果

表 2.1 不同预测方法的结果对比

预测方法	E_{rmse}	E_{mac}	E_{me}	r	C_R	Q_R
传统线性	0.2647	0.1642	0.0204	0.7412	73.53%	74.54%
LSTM-SVR	0.0696	0.0278	0.0028	0.8947	93.04%	93.30%

从图中可以看出，传统的线性回归的预测方法的准确率是小于本题所用的 LSTM-SVR 混合预测模型的。而从表 2.1 中可以得到：传统线性模型在光伏出力预测中表现较差，其均方根误差（ $E_{rmse}=0.2647$ ）和平均绝对误差（ $E_{mac}=0.1642$ ）显著高于 LSTM-SVR 方法（ $E_{rmse}=0.0696$ ， $E_{mac}=0.0278$ ）。这表明传统模型对复杂的光伏功率波动（如天气变化、云层遮挡）的适应性不足，导致预测偏差较大。

相比之下，LSTM-SVR 方法通过结合长短期记忆网络（LSTM）对时序特征的捕捉能力与支持向量回归（SVR）的非线性拟合优势，大幅提升了预测精度： E_{rmse} 和 E_{mac} 分别降低 73.7% 和 83.1%，表明预测值更接近实际功率。而 E_{me} 从 0.0204 降至 0.0028，系统性偏差趋近于零，避免了传统模型的整体高估问题。 r 从 0.7412 提升至 0.8947，验证了预测趋势与实际变化的高度一致性； C_R 和准确率 Q_R 均从约 74% 提升至 93% 以上，说明 93% 以上的预测结果满足误差阈值要求。

从而说明 LSTM-SVR 模型能有效学习光伏出力的非线性时序特征（如昼夜周期、天气突变），而传统线性模型仅依赖静态关系，无法捕捉动态波动规律。

问题 3：建立融入 NWP 信息的光伏电站日前发电功率预测模型

3.1 问题 3(3.1)

3.1.1 解题思路

本题要求将 NWP 信息融入进光伏电站的日前发电功率预测模型中，通过建立 CNN-BiLSTM-AM 模型，采用双通道结构：CNN 处理 NWP 栅格数据(如云量分布)，BiLSTM-AM 同步分析历史功率与站点级气象标量数据，通过注意力机制动态分配辐照强度、风速、总云量等突变时段的特征权重。

3.1.2 模型建立

3.1.2.1 卷积神经网络(Convolutional Neural Networks, CNN)

为了克服传统神经网络存在的参数爆炸问题，CNN 可以根据局部连接、权重共享和下采样技术构造网络结构，完整架构如下图所示：

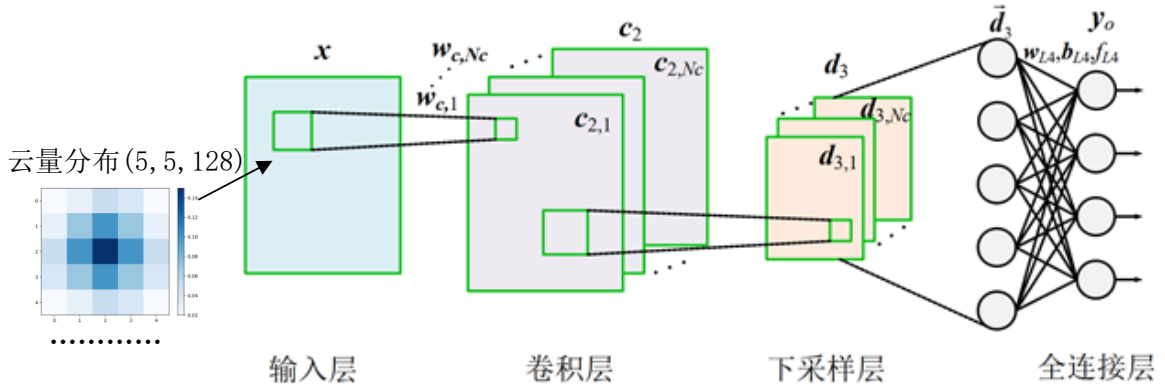


图 3.1 卷积神经网络结构

其中，输入层为二维的输入数据 x 卷积层包含 s 个卷积核，卷积核的整体以 w_c 表示，第 k 个卷积核为 $w_{c,k}$ 。卷积层神经元数据以 c_2 表示，卷积层中包含了 Nc 个二维特征平面，第 k 个卷积层的二维平面数据为 $c_{2,k}$ 。下采样层神经元数据以 d_3 表示，下采样层包含了 Nc 个二维下采样特征平面，第 k 个下采样特征平面 $d_{3,k}$ 。对 d_3 进行向量化排布后获得 \vec{d}_3 。下采样层之后是全连接层，同时也是输出层。全连接层的权重、偏置、激活函数和输出分别为 w_{L4} 、 b_{L4} 、 $f_{L4}()$ 和 y_o 。CNN 中各级连接关系如下：

$$c_{2,k} = f_C(x, w_{c,k}, b_{c,k}), \quad k=1, \dots, Nc \quad (24)$$

$$d_{3,i} = f_D(c_{2,k}), \quad k=1, \dots, Nc \quad (25)$$

$$d_3 = [d_{3,1}, d_{3,2}, \dots, d_{3,Nc}] \quad (26)$$

$$\vec{d}_3 = \text{reshape}(d_3) \quad (27)$$

$$y_o = f_{L4}(w_{L4}\vec{d}_3 + b_{L4}) \quad (28)$$

将 NWP 中的云量、辐照强度等空间分布参数，按照经纬度插值生成 5×5 的网络矩阵，中心点对应的为光伏电站位置。而处理后的网络矩阵作为输入层进入到卷积神经网络中来到卷积层，这里的目的是通过局部连接与权重共享提取空间特征(如云团聚集区、辐照强度梯度变化)。而下采样层是为了压缩特征维度，保留了显著空间模式。最后到全连接层中将空间特征映射为高维向量，供后续跨模态融合做准备。

3.1.2.2 双向长短期记忆网络注意力机制架构(BiLSTM-AM)

为了克服单个 LSTM 单元只能捕捉先前状态信息而无法捕捉未来状态信息的局限，本文引入了双向长短期记忆网络(, BiLSTM)模型。BiLSTM 的网络结构是基于 LSTM，其核心特点是包含两个 LSTM 层，分别从正向和反向两个方向处理输入序列，如图 2 所示。正向的 LSTM 层处理输入序列，输出隐状态序列;反向的 LSTM 层从序列末尾开始处理输入序列，输出反向的隐状态序列。最后，将正向和反向的隐状态拼接在一起，形成 BiLSTM 模型的输出。BiLSTM 模型的优点是能很好地处理双向上下文信息，从而更好地捕捉序列中的长期依赖关系。此外，BiLSTM 能够综合利用输入序列的正向和反向信息，从而更全面地捕捉序列的特征。这种能力有助于提高模型的准确性，使其更好地处理复杂的序列数据。

在 BiLSTM 网络结构图中， x_t 为 t 刻的输入数据， $\vec{h}_t = (\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n)$ 为 t 时刻正向 LSTM 隐含层的输出， $\bar{h}_t = (\bar{h}_1, \bar{h}_2, \dots, \bar{h}_n)$ 为 t 时刻反向 LSTM 隐含层的输出， $y_t = (y_1, y_2, \dots, y_n)$ 为 t 时刻 BiLSTM 网络的输出,最终输出向量是正向和反向信息流的综合效应 $y_t = f(\vec{h}_t, \bar{h}_t)$ ，其数学模型如所示：

$$\begin{cases} \vec{h}_t = \sigma(w_{hx}x_t + w_{hh}h_t + b_h) \\ \bar{h}_t = \sigma(w_{hx}w_t + w_{hh}h_t + b_h) \\ y_t = w_{yh}\vec{h}_t + w_{y\bar{h}}\bar{h}_t + b_y \end{cases} \quad (29)$$

而注意力机制(Attention Mechanism, AM)是一种一种资源分配机制，可以通过对输入特征赋予不同的权重，使包含重要信息的特征不会随着步长的增加而消失，突出更加重要信息的影响，使模型更易学习序列中长距离相互依赖的关系，具体实现流程如图 3.3 所示。

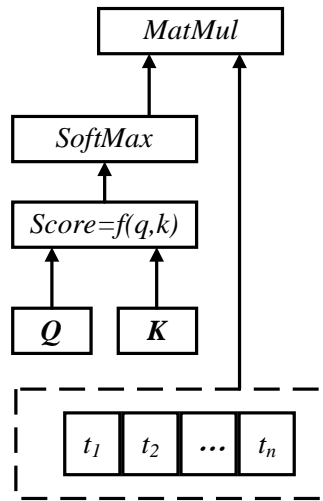


图 3.3 注意力机制

其中， t_1 为 t 时刻的第 1 个特征输入； t_n 为 t 时刻的第 n 个特征输入； Q 为整个输入序列形成的 Query 向量； K 为整个输入序列形成的特征 Key 向量；Score 为计算的注意力得分； f 为计算的函数；SoftMax 为计算后的注意力权重得分；MatMul 为权重得分与输入序列本身点乘后的输出结果。

通过建立 BiLSTM-AM 通道对历史发电功率曲线与站点级气象标量数据(如历史辐照度、气温、风速等)进行时间序列建模，用来挖掘潜在的时序依赖关系^[3]。同时模型引

入注意力机制，在时间维度上动态调整关键气象因素(如辐照强度、风速、总云量)在突变时段的特征权重，从而增强模型对突发天气变化的响应能力与判别能力，实现对复杂气象条件下光伏出力的更加精准预测。

3.1.3 结果分析

本题利用 NWP 信息作为先验知识融入到 CNN-BiLSTM-AM 模型中，进行发电功率预测。同样的采用题目要求的 2、5、8、11 月的最后一周为测试集，其他为训练集，其中图 3.5 为融入 NWP 信息的 CNN-BiLSTM-AM 模型的预测结果，同时按照附录 1 中的误差计算方式，对比不同预测方法，具体结果如表 3.1。

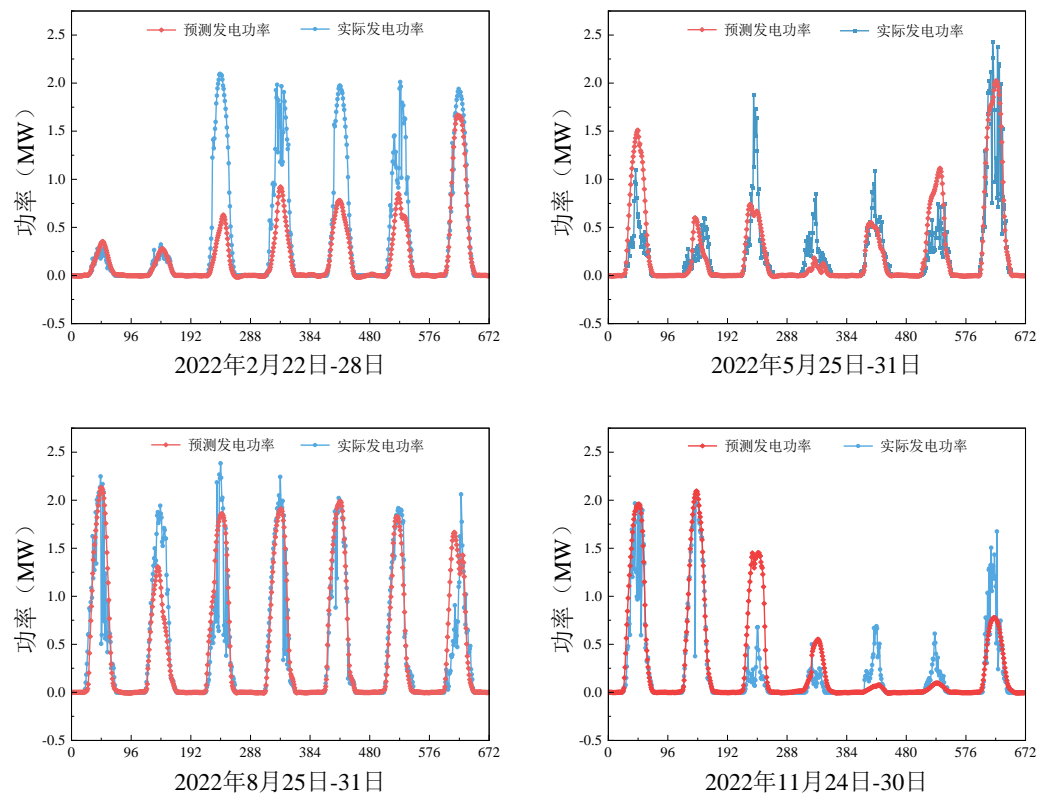


图 3.5 融入 NWP 信息的 CNN-BiLSTM-AM 预测模型

表 3.1 不同预测方法的结果对比

预测方法	E_{rmse}	E_{mae}	E_{me}	r	C_R	Q_R
传统线性	0.2647	0.1642	0.0204	0.7412	73.53%	74.54%
LSTM-SVR	0.0696	0.0278	0.0028	0.8947	93.04%	93.30%
CNN-BiLSTM-AM	0.0675	0.0237	0.0015	0.9068	93.25%	93.50%

本题所用的 CNN-BiLSTM-AM 双通道模型通过融合数值天气预报（NWP）栅格数据与站点级时序气象数据，显著提升了光伏功率预测精度，其综合性能优于传统线性模型和 LSTM-SVR 方法。具体优势如下：CNN 对 NWP 栅格数据（如云量分布）的局部空间特征建模能力，有效捕捉云层移动对光伏功率的区域性影响，使 E_{rmse} 从 LSTM-SVR 的 0.0696 降至 0.0675， E_{mae} 从 0.0278 降至 0.0237。注意力机制（AM）强化了辐照强度突变、风速骤变等关键时段的特征权重，减少极端天气下的预测偏差，平均误差（ E_{me} ）

趋近于零 (0.0015)，系统性偏差几乎消除。而 BiLSTM-AM 对历史功率与站点气象标量数据 (温度、湿度等) 的长短期依赖建模，结合 CNN 的栅格数据特征，使相关系数 (r) 从 LSTM-SVR 的 0.8947 提升至 0.9068，预测趋势与实际出力波动更同步。

3.2 问题 3(3.2)

3.2.1 解题思路

通过上一题的分析可知，当 NWP 信息融入到光伏发电预测模型中时，可以显著提升日前预测的精度。然而，由于季节变换和昼夜交替的影响，一年内存存在大量气象条件相似但时序不同的情形。如图 3.6 所示，不同的气象信息会导致不同的输出功率，而相同的气象信息光伏电站的输出功率接近。基于此，本题提出一种结合最大互信息系数 (Maximum Information Coefficient, MIC) 的场景划分方法，通过识别与光伏出力高度相关的气象特征对历史数据进行分类建模，从而进一步提升预测模型的泛化能力与适应性。

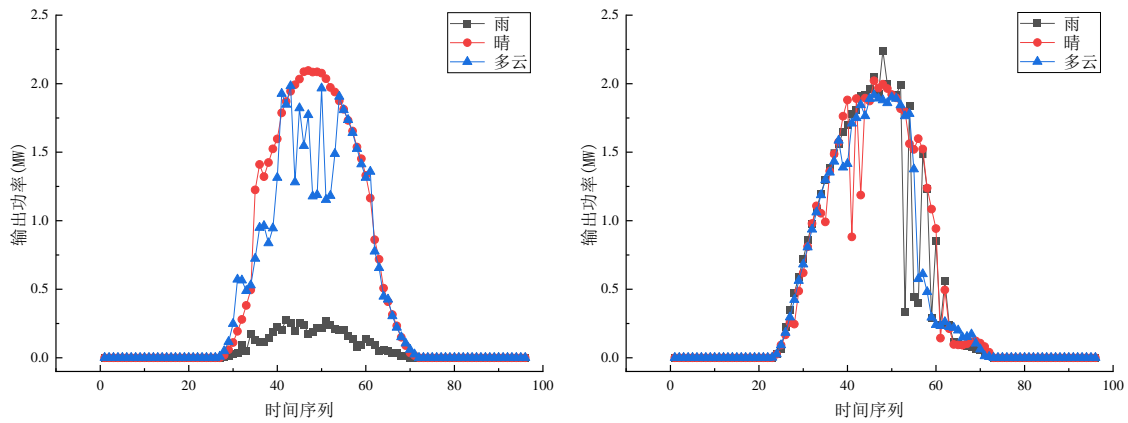


图 3.6 不同/相同天气类型下光伏功率关系曲线图

3.2.2 场景划分与建模方法

为充分利用 NWP 信息与光伏出力之间的内在关联性，设计了如下三步建模流程：

(1) 特征相关性分析

首先，对历史 NWP 数据中的气象信息与实际光伏出力数据进行逐一配对，计算 MIC 值，具体计算公式如下：

$$MIC = \max_{XY < B} \frac{I(X;Y)}{\ln(\min(X,Y))}, \quad B=f(data_size)=n^{0.6} \quad (30)$$

其中， B 为网格的分辨率， $\ln(\min(X,Y))$ 为变量 X 和变量 Y 最小取值的自然对数。而变量 X 和变量 Y 分别为每次配对的气象信息与光伏实际出力。而选择 MIC 的原因是其可以捕获包括功能性和非功能性的广泛关联，具有普遍性和公平性。普遍性意味着在足够的样本容量下，MIC 可以获取到广泛的相关关系，而不是仅仅局限于特定的功能类型 (如线性的、指数的或周期的)，甚至包括所有的函数关系；公平性则意味着 MIC 对于不同类型的包含同等噪声的关系可以给出相似的分。

根据 MIC 值的排序，选取相关性前 3 的气象特征作为场景划分依据。

(2) 多特征联合场景划分

在选定了前 3 个高 MIC 值气象特征 (记为 F_1, F_2, F_3) 后，要对特征标准化与加权处理。首先进行 Z-score 标准化，消除量纲差异：

$$F'_i = \frac{F_i - \mu_i}{\sigma_i} \quad (31)$$

其中， μ_i 和 σ_i 分别为特征 F_i 的均值和标准差。紧接着是动态权重分配，根据 MIC 值赋予特征权重，强化高相关性特征的主导作用。具体计算公式为：

$$\omega_i = \frac{MIC(F_i, P)}{\sum_{j=1}^3 MIC(F_j, P)} \quad (32)$$

其中， P 表示光伏处理。加权后的特征向量为 $F = [\omega_1 F_1', \omega_2 F_2', \omega_3 F_3']$ 。

通过 K-means 算法对标准化加权后的特征向量 F 进行聚类从而得出最后的聚类数 k ，这里注意在初始化聚类是要优先选择高密度区域样本作为初始中心，避免陷入局部最优的情况^[4]。

最后计算每个场景内样本的 MIC 值方差，验证同类气象特征的处理关联稳定性，具体计算公式如下：

$$Var_{MIC} = \frac{1}{N_c} \sum_{i=1}^{N_c} (MIC(F_i, P) - \overline{MIC}_c)^2 \quad (33)$$

其中， N_c 为场景 c 的样本数， \overline{MIC}_c 为该场景的平均 MIC 值。若 Var_{MIC} 的值越小，则其同类气象特征之间的关联越稳定。

(3) 建立不同场景的 CNN-BiLSTM-AM 模型

对每个不同的场景，构建独立的 CNN-BiLSTM-AM 模型，整体框架如 3.1.2 所示，在此不过多赘述。

3.2.3 结果分析

本题根据上述所提的基于 MIC 信息的场景划分方式，分别划分出春、夏、秋、冬四个季节中的典型周，同样利用一周的时间作为测试集，其他样本为训练集。其中图 3.5 为融入 NWP 信息的 CNN-BiLSTM-AM 模型的预测结果，同时按照附录 1 中的误差计算方式，对比不同预测方法，具体结果如表 3.2。

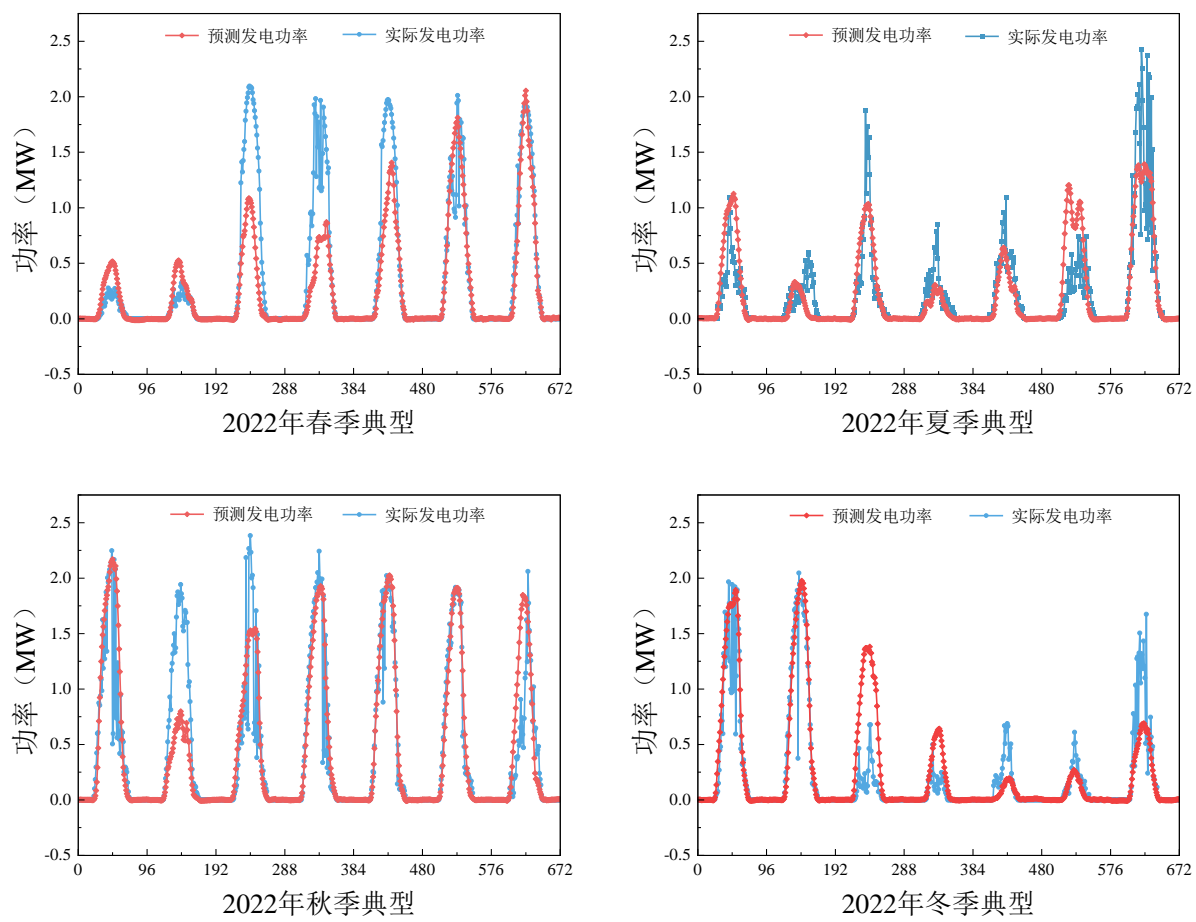


图 3.7 划分场景后的 CNN-BiLSTM-AM 模型预测结果

表 3.2 不同预测方法的结果对比

预测方法	E_{rmse}	E_{mac}	E_{me}	r	C_R	Q_R
传统线性	0.2647	0.1642	0.0204	0.7412	73.53%	74.54%
LSTM-SVR	0.0696	0.0278	0.0028	0.8947	93.04%	93.30%
CNN-BiLSTM-AM	0.0675	0.0237	0.0015	0.9068	93.25%	93.50%
CNN-BiLSTM-AM(S)	0.0591	0.0254	-0.0010	0.9114	94.09%	94.30%

本题通过引入 MIC 对场景进行划分，对历史气象数据进行分类建模，筛选出高 MIC 特征作为场景划分依据，具体数值见表 3.2。可以看出总云量、地表辐照和温度对光伏出力的影响较大，所以选择这 3 个气象特征作为划分依据。同时根据第一题所提出的分析，将一年的气象数据分为春、夏、秋、冬四个典型周进行训练与测试。

表 3.2 不同气象特征的 MIC 数值

气象特征	总云量	地表辐照	温度	相对湿度	风速	风向	气压
MIC 值	0.85	0.92	0.73	0.6	0.40	0.38	0.32

这样的场景划分使得 CNN-BiLSTM-AM(S)模型在光伏功率预测中实现了精度与泛化能力的双重提升。首先，误差显著降低： E_{rmse} 从 0.0675 降至 0.0591，降幅达 12.4%，表明 MIC 场景划分有效区分了气象条件相似但时序不同的样本，减少模型因时序混淆导致的预测偏差。 Q_R （合格率）提升至 94.30%，验证了模型对复杂气象组合（如昼夜交替、季节变换）的适应性更强。系统性偏差也得到了优化： E_{me} 从 0.0015 变为-0.0010，绝对值进一步减小，表明 MIC 划分后子模型的误差分布更均衡，系统性偏差趋近于零，部分场景下轻微低估实际功率（负值），但整体预测更稳定。同时相关性得到了提升：相关系数（ r ）从 0.9068 升至 0.9114，说明预测曲线与实际功率波动的同步性增强，尤其在气象突变时段（如云层快速移动）的响应更精准。

与无场景划分下的 CNN-BiLSTM-AM 模型进行对比， E_{rmse} 降低 12.4%，MIC 划分后，模型对相似气象但不同季节/时段的样本分类训练，减少时序混淆误差。 Q_R 提升 0.8%（93.50%→94.30%）：更多预测结果满足 25%误差阈值，尤其在季节过渡期表现更稳健。

与 LSTM-SVR 混合模型进行对比： E_{mac} 降低 8.9%（0.0278→0.0254）：MIC 筛选的高相关特征被强化建模，减少平均绝对误差。 C_R （准确率）提升 1.05%（93.04%→94.09%）：场景划分后，模型对“易错场景”（如低辐照多云天气）的预测更精准。

问题 4：探讨 NWP 空间降尺度能否提高光伏电站发电功率预测精度

4.1 解题思路

由于传统的大尺度气象预报信息难以适用于如今的 MW 级的光伏电站预测信息要求，所以需要精细化气象预报空间尺度的分辨率^[5]。回声状态网络(Echo State Network, ESN)作为融合动态系统理论与时序数据驱动的典型方法，具备对时空序列的天然建模能力和高效训练优势，可有效捕捉气象场中的时空非线性动力学特征。具体而言，ESN 通过随机生成的稀疏储层(Reservoir)隐式记忆历史状态，仅需训练输出层权重即可实现复杂时空映射，这种特性使其能够同时处理高维气象变量的空间关联性与时间依赖性。因此，本题利用 ESN 建立低分辨率 NWP 数据、地理信息与高分辨率气象观测数据之间的动态关系，通过储层状态传播提取多尺度气象场的时空演化规律，从而实现高分辨率空间降尺度。结合空间降尺度后的预测结果，可进一步验证 ESN 在捕捉局部气象突变(如云层快速移动)和地形效应上的可行性。

4.2 数据处理

具体建模步骤：

(1)选取气象变量

本次选取的气象变量包括地表气压、地表相对湿度、总云量、地表 10 米风速、10 米风向、地表 2 米气温、水平面总辐照度、降水量、地表 100 米风速作为主要因素来映射光伏电站的出力变化。并且将气象变量扩展为时空序列，输入维度为[时间步长×低分辨率网格×气象变量数]。

(2)数据集划分

1.时间维度划分：

按照第一题中所分析的季节特性进行划分数据集，确保训练/测试集负荷不同气候模式，其中春夏秋数据占比 70%作为训练集，而剩下的 30%由冬季数据来组成，用于测试冷启动与极端天气适应性。

2.空间维度划分：

将光伏电站所在区域划分为 5×5 个子区域，随即保留一个子区域作为测试集，用于验证模型在未见过空间上的泛化能力。

3.数据预处理：

对所有选取的气象变量进行时空归一化，具体如下式：

$$X_{norm} = \frac{X - \mu_{i,j}}{\sigma_{i,j}} \quad (34)$$

其中， μ 和 σ 按照每个空间网格在不同季节单据计算。

(3)模型建立

ESN 不同于传统循环神经网络，其设置了一种储存池结构作为隐藏层来输入特征进行非线性的表示，具备了不容易陷入局部最优和收敛慢的优点。

ESN 主要为3个部分：输入层、储存池和输出层，储存池内部主要是一个所谓的动态记忆库结构，该结构主要以大量的神经元相互链接而成，是 ESN 模型的核心。储存池的权值矩阵 w 中的每个值代表着内部各个神经元的连接权重，其在 ESN 模型初始化时自然生成，数值大概在 0.05 左右。输入层的作用则是与储存池内的动态记忆库以权重 w_{in} 进行连接，该权值也是 ESN 模型初始化时自然呈现，且维持不变。最后，通过输出层与储存池之间的权重 w_{out} 待输入信息到模型学习后输出相应维度的结果。

如下式所示，分别表示在时刻 n 下的输入层、输出层以及储存池内部各个神经元的状态表示。

$$\begin{aligned} U(n) &= [u(1) \ u(2) \ u(3) \ \dots \ u(n-1) \ u(n)]^T \\ X(n) &= [x(1) \ x(2) \ x(3) \ \dots \ x(n-1) \ x(n)]^T \\ Y(n) &= [y(1) \ y(2) \ y(3) \ \dots \ y(n-1) \ y(n)]^T \end{aligned} \quad (35)$$

在第 $n+1$ 时刻，输出层和储存池状态会按照下式进行迭代更新：

$$\begin{aligned} X(n+1) &= f(WX(n) + W_{in}U(n) + W_{back}Y(n)) \\ \tilde{Y}(n+1) &= f_{out}(W_{out}X(n-1)) + W_{bias}^{out} \end{aligned} \quad (36)$$

其中， W_{in} 、 W 、 W_{out} 含义如上所述， W_{back} 表示输出层的反馈连接权重， W_{bias}^{out} 表示偏置权重， f 和 f_{out} 表示储存池和输出层的激活函数。以上参数对 ESN 模型的训练和学习具有重要作用，其中参数 W_{in} 、 W 、 W_{back} 均为随机产生，但一旦生成就永远不变，模型训练过程中只有参数 W_{out} 会变化。

输入层接收低分辨率的 NWP 信息数据，其中维度为气象变量数。而储层结构由 1000 个神经元组成的稀疏网络加上设置好的谱半径，从而保证 ESN 的记忆能力。

4.3 结果分析

本题根据上述所提基于 ESN 的 NWP 空间尺度降维，将降尺度空间后的 NWP 信息作为新的气象信息放入数据集中。同样的，选取 2、5、8、11 月的最后一周作为测试集，其他样本为训练集。其中图 4.2 为融入 NWP 空间降尺度后的高分辨率信息的 CNN-BiLSTM-AM 模型的预测结果，同时按照附录 1 中的误差计算方式，对比不同预测方法，具体结果如表 4.1。

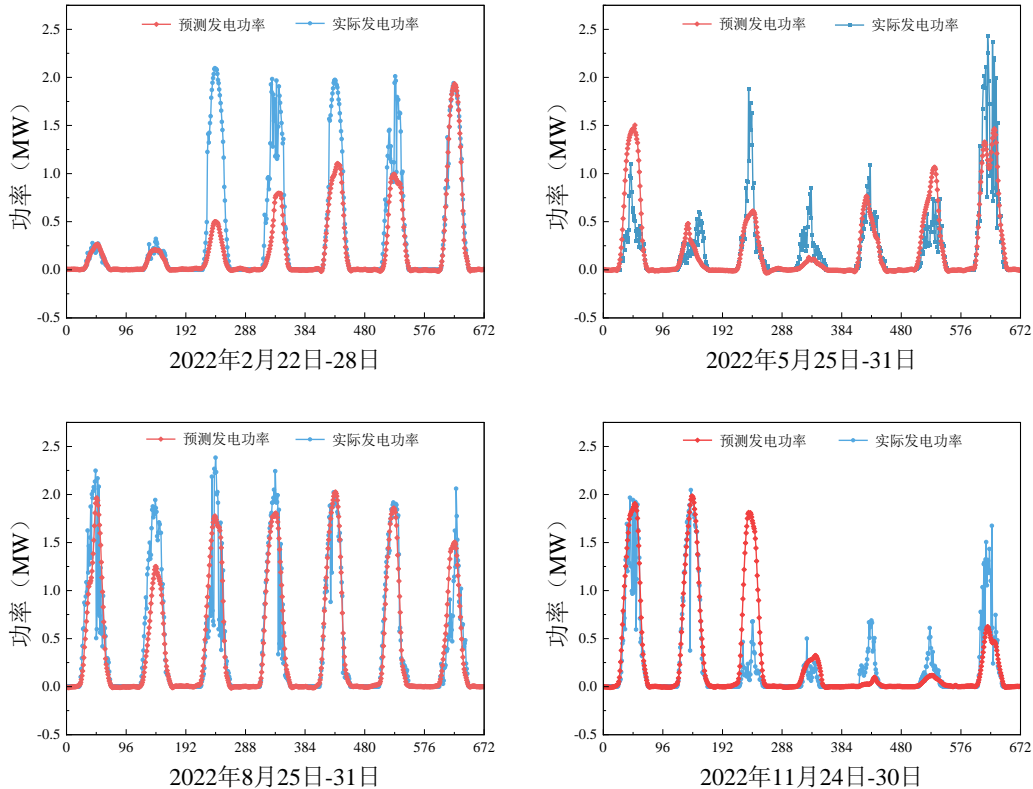


图 4.2 融入 NWP 空间降尺度信息的 CNN-BiLSTM-AM 模型预测结果

表 4.1 不同预测方法的结果对比

预测方法	E_{rmse}	E_{mae}	E_{me}	r	C_R	Q_R
传统线性	0.2647	0.1642	0.0204	0.7412	73.53%	74.54%
LSTM-SVR	0.0696	0.0278	0.0028	0.8947	93.04%	93.30%
CNN-BiLSTM-AM	0.0675	0.0237	0.0015	0.9068	93.25%	93.50%
CNN-BiLSTM-AM(S)	0.0591	0.0254	-0.0010	0.9114	94.09%	94.30%
CNN-BiLSTM-AM(ESN)	0.0474	0.0231	0.0005	0.9172	95.26%	95.45%

通过引入回声状态网络(ESN)对气象数据进行空间降尺度,CNN-BiLSTM-AM(ESN)模型显著提升了光伏功率预测的时空分辨率与局部突变捕捉能力,其综合性能超越所有基准模型。首先是误差全面最低, E_{rmse} 降至 0.0474, 较 CNN-BiLSTM-AM(S)进一步降低 19.8%, 表明 ESN 通过高分辨率气象场建模,可以有效捕捉云层移动、地形遮挡等局部突变的影响。 E_{mae} 降至 0.0231, 验证了降尺度后的气象数据对光伏出力空间异质性的精准刻画。而 E_{me} 为 0.0005, 绝对值接近零且符号为正, 表明模型在消除季节/昼夜偏差的同时,未引入新的系统性误差,预测结果高度均衡。而相关系数(r)提升至 0.9172, 反映预测曲线与实际功率的时空同步性更强,尤其在云层快速移动时段的拟合度优化显著。 Q_R (合格率)达 95.45%, 创历史最优,说明模型对光伏出力预测鲁棒性极佳。

5 结论

本次题目围绕光伏电站发电功率日前预测的关键技术问题，首先计算了基于地理位置信息下，结合太阳辐照计算了该光伏电站的理论可发功率。并研究了其长/短周期的波动特性。同时结合理论发电与实际发电，分析了该光伏电站的发电功率特性。

基于上述的分析，利用深度学习知识，建立了 LSTM-SVR 的混合预测模型，并基于历史功率进行发电预测，结果表明，本模型的预测均方根误差较传统线性模型降低了 73.7%。接着将 NWP 信息融入到预测模型中，并结合深度学习理论，将预测模型更新为 CNN-BiLSTM-AM。该模型利用卷积神经网络提取 NWP 信息中的空间特征，例如云量梯度分布。同时结合历史功率序列与气象标量数据进行双向时序建模，并利用注意力机制动态增强辐照突变时段的特征权重。最终结果将相关系数提升至 0.9068，较单一时序模型预测合格率提升了 0.95%。同时创新性的利用 MIC 信息进行场景的划分与建模，同样提升了预测精度。

最后，利用 ESN 驱动的 NWP 空间降尺度技术，通过稀疏储层结构捕捉气象场的时空演化规律，将预测误差进一步降低 19.8% ($E_{rmse}=0.0474$)，验证了高分辨率气象建模对局部突变特征的精准刻画能力。但是，本模型存在着推理时间较长，难以满足电网调度快速要求。

6 参考文献

- [1] 魏泓屹, 卓振宇, 张宁, 等. 中国电力系统碳达峰·碳中和转型路径优化与影响因素分析[J]. 电力系统自动化, 2022, 46(19): 1-12.
- [2] 吴沂洋, 刘继春, 伍峻杭. 考虑不同区域气象与地理条件差异的广域光伏电站规划[J]. 电网技术(9): 3653-3662.
- [3] 张姗, 冬雷, 纪德洋, 等. 基于 NWP 相似性分析的超短期光伏发电功率预测[J]. 太阳能学报(4): 142-147.
- [4] 叶林, 裴铭, 路朋, 等. 基于天气分型的短期光伏功率组合预测方法[J]. 电力系统自动化(1): 44-54.
- [5] 吴钰姝, 申建建, 程春田. 融合高精度气象数据的水风光系统长短期嵌套消纳调度方法[J]. 电网技术, 2025.