

基于气象分类的光伏发电预测

吴林河^{1,2}, 万建飞^{1*}, 杨 杭²

¹长安大学理学院, 陕西 西安

²中创新能网络科技有限公司, 陕西 西安

收稿日期: 2026年6月17日; 录用日期: 2026年7月5日; 发布日期: 2026年7月16日

摘 要

随着节能降碳行动方案的推出, 光伏发电产业得到空前的发展, 其发电量的不确定性对电网的稳定性带来重大冲击, 所以, 对光伏发电的精准预测是确保电网稳定运行的重要环节。通过对光伏发电的主要影响因素的分析, 将气象等非定量且强非线性的因素进行分类, 对各分类情况分别进行建模, 得到分类预测模型。充分利用我们数据完整的优势, 确定分类预测模型的参数, 对测试集数据的比较分析表明, 分类预测模型稳定性好, 鲁棒性强, 预测精度高。

关键词

光伏发电, 发电预测, 分类预测

Meteorological Classification-Based Photovoltaic Power Generation Forecasting

Linhe Wu^{1,2}, Jianfei Wan^{1*}, Hang Yang²

¹School of Sciences, Chang'an University, Xi'an Shaanxi

²Wealthypower Network Technology Co., Ltd., Xi'an Shaanxi

Received: June 17, 2026; accepted: July 5, 2026; published: July 16, 2026

Abstract

With the introduction of energy conservation and carbon reduction action plans, the photovoltaic (PV) power generation industry has experienced unprecedented growth. However, the uncertainty of PV power output poses significant challenges to the stability of the power grid. Therefore, accurate forecasting of PV power generation is crucial for ensuring reliable grid operation. By analyzing the main influencing factors of PV power generation, non-quantitative and strongly nonlinear factors,

*通讯作者。

文章引用: 吴林河, 万建飞, 杨杭. 基于气象分类的光伏发电预测[J]. 统计学与应用, 2026, 15(7): 119-126.

DOI: 10.12677/sa.2026.157154

such as weather conditions, are classified. A separate prediction model is developed for each category, resulting in a classification-based forecasting model. Leveraging the advantage of complete data, the parameters of the classification-based prediction model are determined. Comparative analysis of the test set data demonstrates that the proposed model exhibits good stability, strong robustness, and high prediction accuracy.

Keywords

Photovoltaic Power Generation, Power Generation Forecasting, Classification-Based Prediction

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

新能源发电受环境及气象等因素的影响,特别是气象因素,其随机性严重影响发电量,造成新能源发电量的波动,这对整个电网的稳定、安全造成不利的影响,所以,迫切需要对未来的天气状况进行精准预测,进而预测出未来的发电量,电网可以根据预测值以及负荷需求通过调整火力或其他发电能力对冲新能源发电的波动。此外,新能源的发展,也推动了电力市场的变革,将导致虚拟电厂的发展,而新能源发电量的精准预测正是虚拟电厂业务的基本依据。所以,基于天气预测的新能源发电预测,是当今以及未来的电力行业运行管理的必要且基本的需求。

新能源电厂的分布广泛,有很多分布在农村或山区,现有的气象预测的空间粒度太大(通常是行政区划单位),很难精准预测新能源厂址的天气情况,从而难以预测发电量,所以需要研发基于多源数据的天气预测模型,鉴于其重要性,该问题的研究也受到国内外的重视,并成为近年来的热点研究问题。

光伏发电预测涉及的环节很多,包括数据清洗算法、天气预测、辐照预测等[1]-[7]。在发电量预测方面, Saida El Bakali 等(2025)通过优化变分模态分解(VMD)参数来提高特征提取的精度[8]。Woo Gyun Shin 等(2025)从单二极管模型(SDM)出发,引入辐照度与温度修正系数,推导出阵列级功率推断方程,从而得到高精度的回归模型,同时具备良好的泛化性能与可解释性[9]。Zhang 等(2024)利用支持向量机(SVM)筛选特征后,采用高斯混合模型(GMM)将数据样本按气象与功率特征分类,构建多样化数据特征库,提高模型学习效率[10]。随着机器学习与深度学习技术的不断发展,研究者们开始探索多算法融合的混合预测框架,以充分发挥不同模型的互补优势, Zhu 等(2025)设计了一种融合长短期记忆网络(LSTM)与极端梯度提升(XGBoost)的混合预测模型,通过两种算法的协同作用提升预测精度[11]。龙小慧等(2024)基于相似日聚类,利用改进的模态分解(ICEEMDAN)对序列进行多尺度分解与重构,高频部分采用 CNN-BiGRU-Attention 网络预测,低频与趋势部分则由 LSSVR 处理,实现对光伏功率的精细化预测[12]。此外, M. Alghassab (2025)将粒子滤波深度嵌入循环神经网络(RNN)的端到端训练过程中,通过在特征空间中进行粒子采样,实现对平滑且稳健的状态转移函数的逼近,从而显著提升了模型应对输入数据高变异性的鲁棒性[13]。臧鹏发等(2025)构建了一个多特征提取的混合模型框架,将数据驱动与物理机理模型深度融合,并系统区分不同时期的天气状况,实现预测精度的显著提升[14]。对于近年来兴起的 Transformer 模型,任惠等(2026)提出的 DGAT-Trans 组合算法,分别预测光伏功率的周期与趋势,并加以重构,总体解决了区域光伏处理数据质量增强与超短期预测问题[15]。

影响发电量的因素很多,各文献一定程度上基于其具备的数据特征展开研究,我们是基于光伏运维

和电力交易的相关业务及数据积累展开研究,提出了分类建模,分时段回归的预测模型,并运用到业务系统平台上。

2. 光伏发电影响因素的分析

以某光伏电站实际数据为例,对发电量与各因素的关系进行了分析(部分数据如表 1)。

Table 1. Measured radiation, power generation and environmental data on a certain day

表 1. 某天实测辐射量、发电量以及环境数据

小时	温度	湿度	光照强度	辐射量	站 1 发电量
6~7	10.445	96.255	135.5	4.583333	0.086667
7~8	11.26667	94.95083	1893	73.25	1.334167
8~9	18.85545	77.67727	8845.727	357.1818	7.116667
9~10	25.17417	50.79667	11586.08	464.9167	17.29
10~11	26.46273	46.90182	10321	405.4545	30.32333
11~12	24.53636	53.43909	9002	346.2727	30.22167

分析发现辐射量和发电量呈现强相关性(如图 1 所示)。

图 1 是 9 月下旬到 10 月中下旬的数据,雨天较多,发电量与辐照量的不一致程度较高,可能的原因是: (1) 温度因素, (2) 清洁度因素, (3) 散射比例因素等。

为此,在预测模型中引入温度、湿度因素,而透明度、清洁度等数据采集成本较高的指标不纳入模型(会一定程度影响预测精度)。

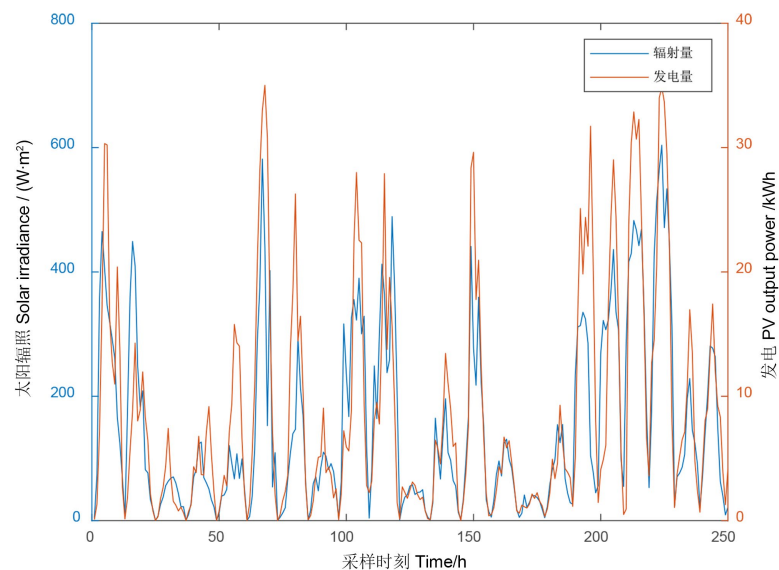


Figure 1. Comparison of radiation and power generation

图 1. 辐射量与发电量比较

3. 预测模型的构建

至于气象情况,比如阴天、晴天、多云等定性指标,对发电量的影响极大,又无法精准量化,并且不同类型的天气下,发电量与其他因素的关系呈现较强的不一致性,所以针对不同的天气类型,分别建立

不同的预测模型。

3.1. 晴天预测模型

针对理想条件下晴天，由于发电量主要受直射辐照决定，于是其与高度角是强相关(如图2所示)。

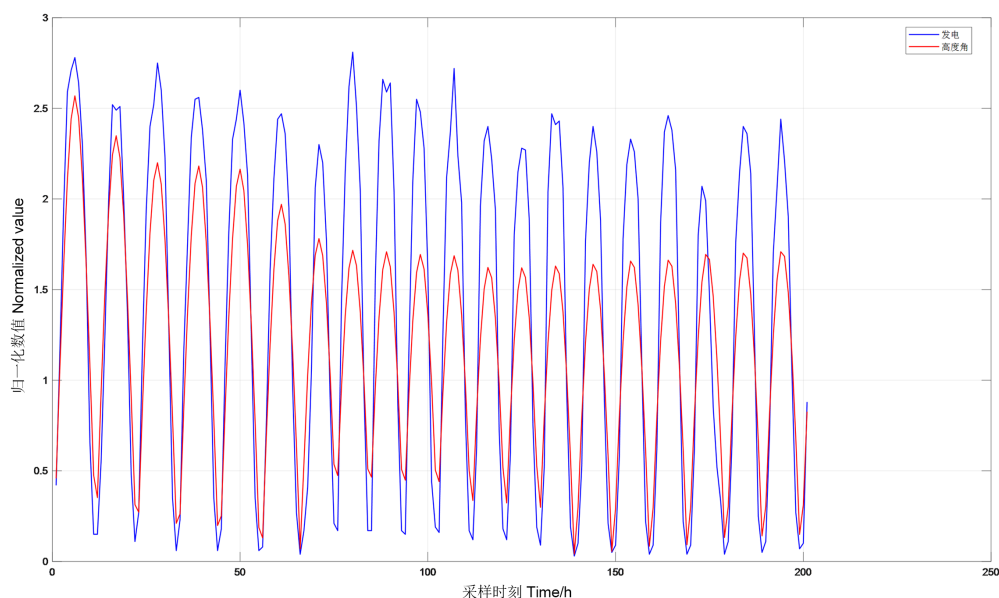


Figure 2. Relationship of altitude angle and power generation

图2. 高度角与发电量关系

图中红色曲线的峰点为当日高度角的最大值点，该时间点的发电量也达到峰值，说明发电量与高度角呈现强相关性。

于是建立预测模型如下：

设发电量用 y 表示， x_1, x_2 分别表示温度和湿度， h 表示高度角。

$$y = c_0 + c_1 x_1 + c_2 x_2 + c_3 h + c_4 h^2 + c_5 h^3$$

考虑到高度角大的时候，发电量相对更大，为了提高预测精度，希望在发电量大时预测精度更高，所以采用加权最小二乘法(以高度角作为权系数)，对模型系数进行求解。

设晴天的学习样本有 n 个，即 $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,n})^T$, $\mathbf{x}_2 = (x_{2,1}, x_{2,2}, \dots, x_{2,n})^T$, $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ ，并记

$$A = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{h}, \mathbf{h}^2, \mathbf{h}^3), \text{ 权矩阵 } D = \begin{pmatrix} h_1 & & & & \\ & h_2 & & & \\ & & \dots & & \\ & & & & h_n \end{pmatrix}, Y = (y_1, y_2, \dots, y_n)^T.$$

回归方程的加权最小二乘解为 $C = (c_0, c_1, c_2, c_3, c_4, c_5)^T = (A^T D A)^{-1} A^T D Y$ 。

为了提高预测精度，不同高度角范围(早晚、早上、正午、下午)的时间，由于环境条件差异大，散射辐照比例有差别，发电量与各因素的关系有所不同，所以，对不同的时段，利用同一个模型分别求解回归系数，得到各自的预测模型。

对实际发电数据的观察，发现多云天气时，出力特征与晴天的情况比较接近(如图3所示)，每日高度角的最大值点对应着发电量的峰值点，说明其发电特征与晴天的情况类似。所以，对多云天气发电量的

预测也采用晴天的模型。

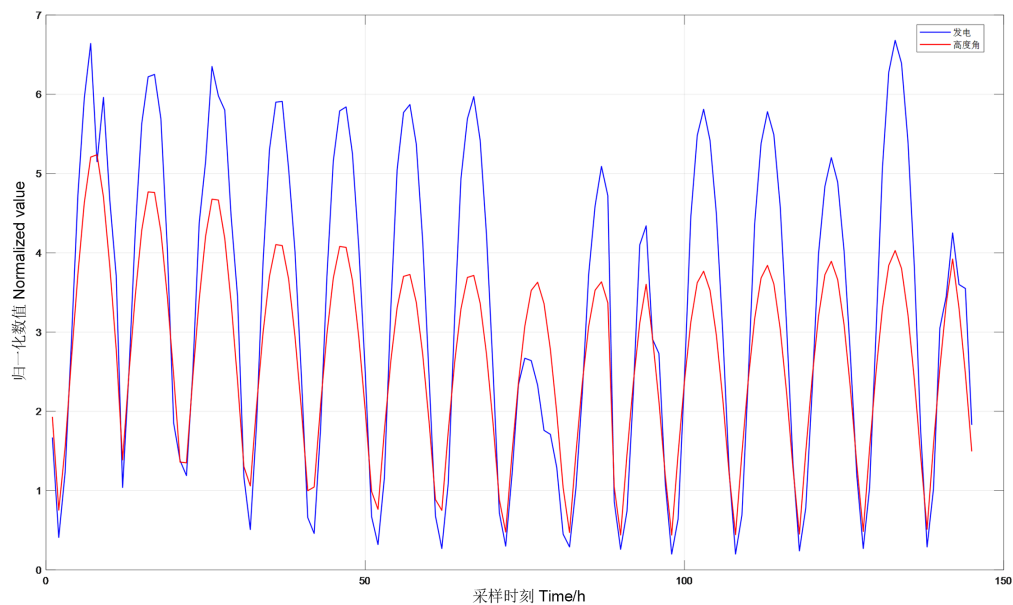


Figure 3. Comparison of altitude angle and power generation in cloudy conditions
图 3. 多云天气下高度角与发电量对比

3.2. 其他天气情况的预测回归模型

晴天的辐照以直射为主，其他天气情况的辐照以散射等为主，但影响的方式有所不同，所以，对此采用同一类模型，但相关的参数是通过不同天气类型的样本分别进行估计。以阴天为例，模型的建立及参数的确定如下：

$$y = c_0 + c_1x_1 + c_2x_2 + c_3h$$

设阴天的学习样本有 n 个，即 $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \dots, x_{1,n})^T$, $\mathbf{x}_2 = (x_{2,1}, x_{2,2}, \dots, x_{2,n})^T$, $\mathbf{h} = (h_1, h_2, \dots, h_n)^T$ ，并记 $A = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{h})$, $Y = (y_1, y_2, \dots, y_n)^T$ 。

回归方程的加权最小二乘解为 $C = (c_0, c_1, c_2, c_3)^T = (A^T A)^{-1} A^T Y$ 。

至于某些小众的天气类型(如大雪、雾霾等)历史样本相对较少，如果要取足够多的样本，时间跨度很长(可能要跨多个年份)，设备情况可能已发生一定的变化，不利于精准预测，所以，对样本量不足的情况，在预测时，不去建立回归模型，直接利用高度角进行估计，认为发电量与高度角强相关，即两组样本数据标准化后基本相同，从而假设 $\frac{\hat{y} - \bar{y}}{\sqrt{D_y}} = \frac{h - \bar{h}}{\sqrt{D_h}}$ ，得到估计值 $\hat{y} = \bar{y} + \frac{h - \bar{h}}{\sqrt{D_h}} \sqrt{D_y}$ 。

4. 异常数据的识别及学习样本的清理

原始数据包括温度、湿度、天气及发电量等采集的数据(高度角是基于日期、时间通过计算得到，是准确的)，这些数据均有可能是非正常情况下的值，或者不一致，比如发电量，有可能是维修(或其他随机因素)造成的非正常发电水平，比如天气，它代表的是一个区域的整体情况，电厂所在的局部区域有时可能整体情况有差异，在进行预测时，由于天气非真实的电厂天气，异常结果会有明显的差别，或者，预报的天气本来就有误，而这些问题无法通过模型类弥补，并且在预测的时候，无法知道是不是有异常，

但当预测日的实际发电量出来后,才能检验当天是否有问题。由于这天的数据将预测未来发电量的样本,避免这些非正常样本对未来预测结果的负面影响,需要对这些数据进行标记,不作为未来预测的样本。

为了检测当前的发电量是否正常,首先利用回归方程 $\hat{\sigma} = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-k)}$ $y = g(x) + \varepsilon$, 其中 $\varepsilon \sim N(\hat{y}, \hat{\sigma}^2)$ (y 为实际发电量, $\hat{y}$ 为理论发电量, x 为影响因素值), 可得其标准差为, 其中 n 为样本数, k 为回归方程中常数项以外的项数, 给定信度水平为 α , 则 $\frac{|\hat{y} - y|}{\hat{\sigma}} \geq U_{1-\alpha/2}$ 视为异常。

一旦异常值被标注, 在对未来一天进行预测时, 就可以从该天之前的正常样本中找出所需的样本。

5. 预测精度指标

预测一定存在误差, 设预测日实际值为 y_i , 预测值为 $\hat{y}_i$, $i=1, 2, \dots, n$, 其绝对误差为 $e_i = |y_i - \hat{y}_i|$, 相对误差为 $e_{ri} = \frac{|y_i - \hat{y}_i|}{y_i}$, 当 y_i 很小时, 尽管绝对误差较小, 但相对误差可能很大, 而这个时候, 关注的主要是绝对误差, 所以, 相对误差不能很好地体现精度描述的需求。提出两个精度指标: (1) 平均相对误差; (2) 整体相对误差。

平均相对误差: $\bar{e}_r = \frac{1}{n} \sum_{i=1}^n e_{ri}$, 这个指标一定程度削弱了“小发电量导致大相对误差的问题”, 所以

采用加权平均, 取实际发电量作为权重, 即 $\bar{e}_r = \frac{1}{n} \sum_{i=1}^n e_{ri} y_i / \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n e_i / \sum_{i=1}^n y_i$, 也就是整体相对误差。

整体相对误差: $Ge_r = \sum_{i=1}^n e_i / \sum_{i=1}^n y_i$, 反映了单位总发电量的误差, 比较有效地克服了小发电量相对误差大的问题, 也能很好地体现实际的精度水平。

6. 数值实验

通过上述预测方法, 对 3 个电站分别进行预测分析, 采集一年的发电数据和气象数据, 采集频率是 5 分钟一次, 训练数据集与测试数据集比例为 8:2, 具体结果如图 4~6。

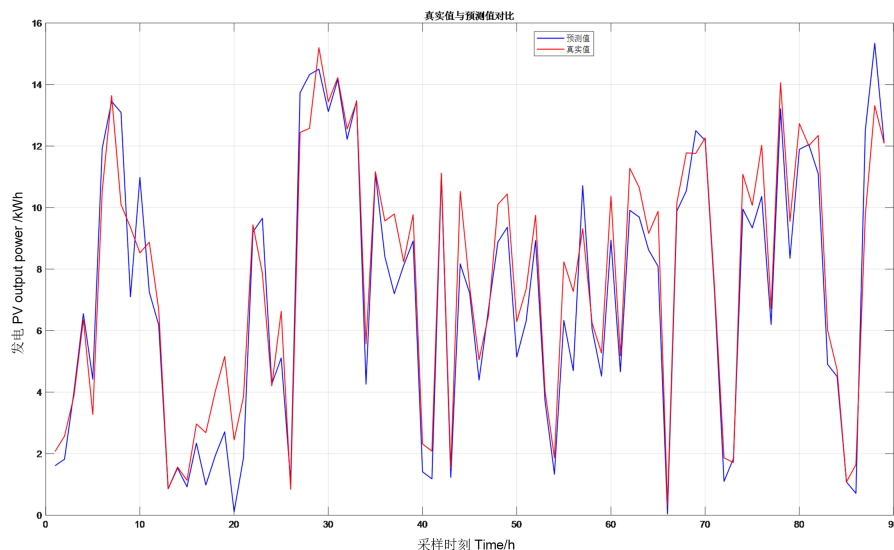


Figure 4. Comparison of predicted and actual values for power station 1

图 4. 电站一预测值与实际值对比

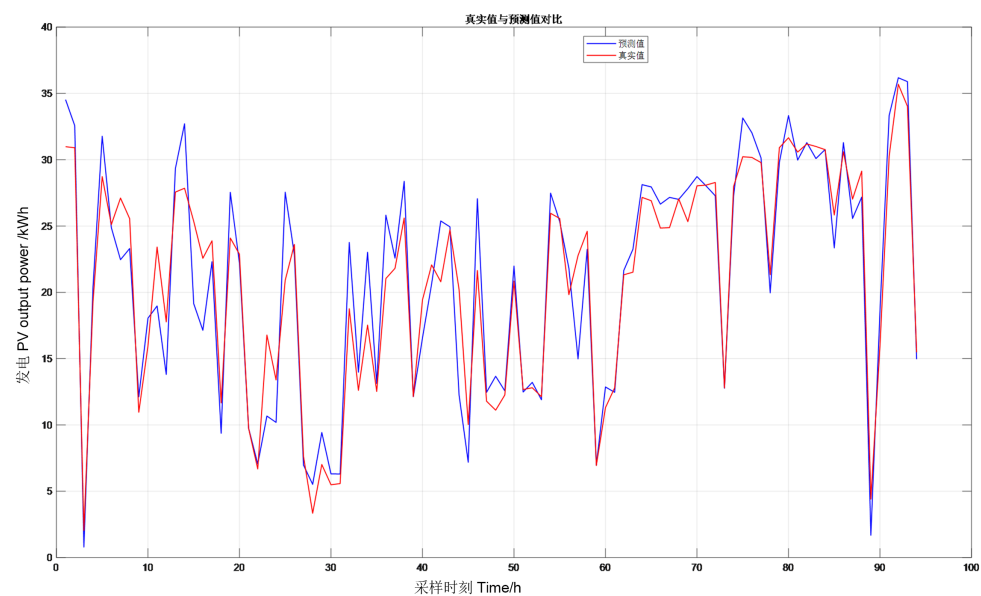


Figure 5. Comparison of predicted and actual values for power station 2
图 5. 电站二预测值与真实值对比

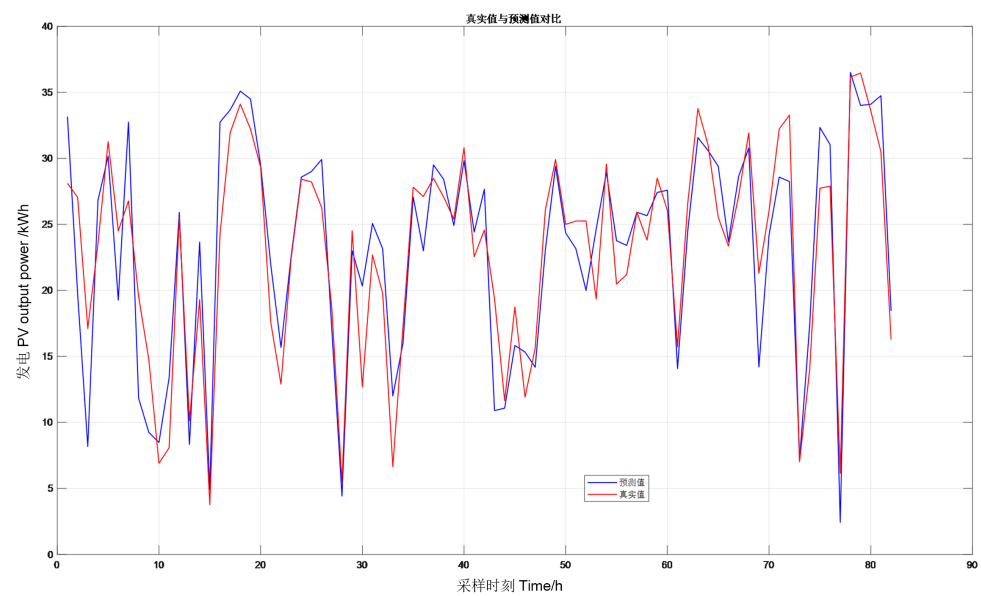


Figure 6. Comparison of predicted and actual values for power station 3
图 6. 电站三真实值与预测值对比

预测误差如下表 2 所示。

Table 2. Result data of standard test system
表 2. 标准试验系统结果数据

电站编号	整体相对误差	平均相对误差
电站 1	0.0815	0.1075
电站 2	0.0707	0.0901
电站 3	0.0757	0.0886

7. 结论

发电量预测涉及的因素很多, 有的因素难以量化, 有的因素采集成本高, 有的因素预测困难, 有的因素随机性大, 这一切均不同程度影响了发电量预测的精细度和精准度。论文主要采用了机理分析及统计方法, 由于数据采集困难的问题, 无法将一些重要的影响因素(云层厚度、空气透明度等)引入预测模型, 在精度上产生一定的负面影响。此外, 平台获取的气象数据与电站的真实气象未必完全一致, 这也影响到预测精度, 有待后续的深入研究。

基金项目

陕西省自然科学基金基础研究计划(2025JC-YBMS-070)项目。论文全部数据由中创新能网络科技有限公司提供。

参考文献

- [1] Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., *et al.* (2021) Can Deep Learning Beat Numerical Weather Prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **379**, Article 20200097. <https://doi.org/10.1098/rsta.2020.0097>
- [2] Taillardat, M., Mestre, O., Zamo, M. and Naveau, P. (2016) Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics. *Monthly Weather Review*, **144**, 2375-2393. <https://doi.org/10.1175/mwr-d-15-0260.1>
- [3] Rasp, S., Pritchard, M.S. and Gentine, P. (2018) Deep Learning to Represent Subgrid Processes in Climate Models. *Proceedings of the National Academy of Sciences*, **115**, 9684-9689. <https://doi.org/10.1073/pnas.1810286115>
- [4] Ham, Y.G., Kim, J.H. and Luo, J.J. (2019) Deep Learning for Multi-Year ENSO Forecasts. *Nature*, **573**, 568-572. <https://doi.org/10.1038/s41586-019-1559-7>
- [5] 王娜, 霍鹏, 韩毅, 等. 基于深度学习的气象预报模型研究综述[J]. 中国电机工程学报, 2023, 43(20): 7929-7940.
- [6] 张童彦, 廖清芬, 唐飞, 等. 基于气象资源插值与迁移学习的广域分布式光伏功率预测方法[J]. 中国电机工程学报, 2023, 43(20): 7929-7940.
- [7] Sezen, M., Shabani, M. and Demir, I. (2024) Spatio-Temporal Graph Neural Networks for Multi-Site Photovoltaic Power Forecasting. *Applied Energy*, **362**, Article 123000.
- [8] El Bakali, S., Ouadi, H. and Gheouany, S. (2025) Wind Turbine and PV Power Prediction Using a Deterministic Data-Driven Model with Variational Mode Decomposition Preprocessing. *Transactions of the Institute of Measurement and Control*, **47**, 1413-1428. <https://doi.org/10.1177/01423312241274044>
- [9] Shin, W.G., Lee, J.S., Ju, Y.C., Hwang, H.M. and Ko, S.W. (2025) Data Preprocessing and Machine Learning Method Based on Ameliorated Mathematical Models for Inferring the Power Generation of Photovoltaic System. *Energy Conversion and Management*, **333**, Article 119793. <https://doi.org/10.1016/j.enconman.2025.119793>
- [10] Zhang, X., Deng, Q., Jia, M., Dai, X., Gao, X. and Zhou, H. (2024) Integrating a Hybrid Data Processing Strategy into an Optimized Light Gradient Boosting Machine for Photovoltaic Power Forecasting. *Neural Computing and Applications*, **37**, 1633-1650. <https://doi.org/10.1007/s00521-024-10691-4>
- [11] Zhu, C., Tu, Y., Wei, Q., Zang, Y., Zhou, P., Yang, L., *et al.* (2025) Short-Term Power Prediction of Photovoltaic Power Station Based on LSTM-XGBoost Model. *Solar Energy*, **300**, Article 113819. <https://doi.org/10.1016/j.solener.2025.113819>
- [12] 龙小慧, 秦际赞, 张青雷, 等. 基于相似日聚类及模态分解的短期光伏发电功率组合预测研究[J]. 电网技术, 2024, 48(7): 2948-2957.
- [13] Alghassab, M. (2025) Advancing Photovoltaic Power Forecast Accuracy: A Hybrid Particle Filter and Recurrent Neural Network Model for Data-Scarce Scenarios toward Thermal Adaptation. *Case Studies in Thermal Engineering*, **75**, Article 107247. <https://doi.org/10.1016/j.csite.2025.107247>
- [14] 臧鹏飞, 王珂琪, 张中伟, 等. 基于多特征提取的超短期光伏发电功率预测研究[J]. 激光与光电子学进展, 2025, 62(5): 333-343.
- [15] 任惠, 于光发, 强涵悦, 等. 基于 DGAT-Transformer 组合算法的区域光伏出力数据质量增强与超短期功率预测方法[J]. 太阳能学报, 2026, 47(5): 639-649.