

基于精细化气象预报数据的分布式光伏 日发电量预测研究

——以南浔地区为例

喻 樾¹, 翟晓瑶¹, 王丹丹², 吴 越^{1*}, 王国祯¹, 许肖璐¹, 王建东¹

¹湖州市南浔区气象局, 浙江 湖州

²湖州市气象局, 浙江 湖州

收稿日期: 2026年7月1日; 录用日期: 2026年8月3日; 发布日期: 2026年8月20日

摘 要

为提高分布式光伏日发电量预测精度, 本文以浙江省南浔地区多个分布式光伏电站为研究对象, 收集电站发电数据以及降水、温度、相对湿度、风速、云量、能见度和晴空辐射等逐小时气象数据, 构建了基于XGBoost的分布式光伏日发电量预测模型, 并分析了不同气象要素的重要性及时变特征, 以及模型预测误差与气象要素之间的关系。结果表明, 云量和能见度是影响分布式光伏日发电量预测的关键气象因素, 其中云量在白天各时段均保持较高的重要性, 能见度的重要性主要集中在中午时段, 不同气象要素对光伏发电具有明显的时间依赖特征。所建立的XGBoost模型能够有效挖掘多源气象要素与日发电量之间的非线性关系, 在验证集上的决定系数达到0.98, 平均绝对百分比误差为13%, 预测准确率达到87%, 预测结果与实测值具有较高的一致性。进一步的误差分析表明, 模型预测误差未随云量、能见度、温度及晴空辐射等单一气象要素变化而表现出明显的系统性规律, 说明模型能够综合利用多种气象信息实现稳定预测, 具有较好的泛化能力和鲁棒性。研究结果可为区域分布式光伏发电预测、电网调度及新能源消纳提供技术参考。

关键词

分布式光伏, 日发电量预测, XGBoost, 气象要素, 时变特征

Daily Power Generation Prediction of Distributed Photovoltaic Systems Using High-Resolution Weather Forecast Data

—A Case Study in Nanxun District

Yue Yu¹, Xiaoyao Zhai¹, Dandan Wang², Yue Wu^{1*}, Guozhen Wang¹, Xiaolu Xu¹,
Jiandong Wang¹

*通讯作者。

文章引用: 喻樾, 翟晓瑶, 王丹丹, 吴越, 王国祯, 许肖璐, 王建东. 基于精细化气象预报数据的分布式光伏日发电量预测研究[J]. 可持续发展, 2026, 16(8): 187-198. DOI: 10.12677/sd.2026.168280

¹Huzhou Nanxun Meteorological Bureau, Huzhou Zhejiang²Huzhou Meteorological Bureau, Huzhou Zhejiang

Received: July 1, 2026; accepted: August 3, 2026; published: August 20, 2026

Abstract

To improve the accuracy of daily power generation forecasting for distributed photovoltaic (PV) systems, multiple distributed PV power stations in Nanxun District, Zhejiang Province were selected as the study area. Historical power generation data together with hourly meteorological data, including precipitation, air temperature, relative humidity, wind speed, cloud cover, visibility, and clear-sky solar radiation, were collected to develop a daily PV power generation forecasting model based on XGBoost. The importance and temporal characteristics of different meteorological variables, as well as the relationship between prediction errors and meteorological factors, were further investigated. The results indicate that cloud cover and visibility are the most influential meteorological variables affecting daily PV power generation forecasting. Cloud cover maintains consistently high importance throughout the daytime, whereas the importance of visibility is mainly concentrated around midday, demonstrating a clear temporal dependence of meteorological factors on PV power generation. The proposed XGBoost model effectively captures the nonlinear relationships between multiple meteorological variables and daily power generation, achieving a coefficient of determination (R^2) of 0.98, a mean absolute percentage error (MAPE) of 13%, and a prediction accuracy of 87% on the validation dataset. The predicted values show high consistency with the measured observations. Furthermore, error analysis reveals no significant systematic relationship between prediction errors and individual meteorological variables, including cloud cover, visibility, temperature, and clear-sky solar radiation, indicating that the model can comprehensively utilize multiple meteorological inputs to achieve stable predictions with strong generalization ability and robustness. The findings provide a useful technical reference for regional distributed PV power forecasting, power grid dispatching, and renewable energy integration.

Keywords

Distributed Photovoltaic, Daily Power Generation Forecasting, XGBoost, Meteorological Factors, Temporal Characteristics

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“双碳”目标的深入推进和新能源装机规模的快速增长，分布式光伏发电凭借建设灵活、就地消纳、投资成本低等优势，已成为中国新能源发展的重要组成部分[1]。近年来，分布式光伏装机容量持续增加，在提高可再生能源利用率、优化能源结构等方面发挥了重要作用。然而，光伏发电受太阳辐射、云量、降水、温度等气象条件影响显著，其发电量具有较强的随机性和波动性[2]-[4]。准确预测光伏发电量不仅有助于提升电网调度效率、促进新能源消纳，而且能够为电力交易、运行计划制定以及分布式能源管理提供重要依据，因此具有重要的理论意义和工程应用价值。

目前，光伏发电预测主要包括短时功率预测和发电量预测两类。其中，短时功率预测由于能够为实

时调度和电网安全运行提供支撑，一直是国内外研究的重点，相关研究已取得了较为丰富的成果[5]-[8]。近年来，机器学习和深度学习方法在光伏预测领域得到了广泛应用，随机森林、支持向量机、人工神经网络、长短期记忆网络(LSTM)以及极端梯度提升(XGBoost)等模型均表现出了较好的预测性能[9]-[11]。其中，XGBoost 具有较强的非线性建模能力、较高的计算效率以及良好的泛化性能，在处理多维气象特征与光伏发电之间复杂关系方面具有明显优势。

相比于短时功率预测，针对日发电量预测的研究相对较少。日发电量预测更加关注一天内累计发电能力，可直接服务于日前发电计划制定、电网运行安排和新能源消纳分析，其预测目标与评价方式均不同于短时功率预测。由于日发电量受到全天气象条件的综合影响，各时段气象要素对最终发电量的贡献存在明显差异，简单依赖单一时刻或少量气象特征难以充分反映全天气象变化对发电量的影响[12] [13]。因此，如何充分利用逐小时气象信息，提高日发电量预测精度，并揭示不同气象要素对预测结果的贡献规律，是当前值得进一步研究的问题。深入分析不同气象要素在一天内各时段对模型预测的影响，不仅有助于提高模型的可解释性，而且能够为气象特征选择和模型优化提供依据，对提升分布式光伏发电预测的可靠性具有重要意义[14]。

基于此，本文以浙江省南浔地区多个分布式光伏电站为研究对象，收集电站历史发电数据以及降水、温度、相对湿度、风速、云量、能见度和晴空辐射等逐小时气象数据，构建基于 XGBoost 的分布式光伏日发电量预测模型。在此基础上，分析不同气象要素的重要性及时变特征，并进一步探讨模型预测误差与主要气象要素之间的关系，评估模型在不同气象条件下的预测性能和稳定性，为区域分布式光伏日发电量预测及电网调度提供参考。

2. 数据与方法

2.1. 分布式光伏发电数据

本文采用南浔地区分布式光伏电站实际运行数据作为研究对象，数据来源于锦浪云监控平台。选取 2025 年 1 月至 2025 年 10 月期间南浔地区 41 座分布式光伏电站的发电数据，其中电站装机容量范围为 23~1534 kWp，涵盖不同规模的分布式光伏系统，具有较好的区域代表性。原始数据为电站日发电量数据，时间分辨率为 1 d。为保证数据质量，对原始数据进行了完整性检查和异常值筛查，剔除了因设备故障、通信异常等原因导致的缺测和异常记录，并将各电站日发电量数据与对应的装机容量进行关联，形成电站级发电数据集。

2.2. 气象数据

考虑到分布式光伏发电主要受气象条件影响，为构建分布式光伏日发电量预测模型，本文采用浙江省智能网格气象预报融合数据作为气象预报数据来源。该数据由浙江省气象部门提供，融合多源观测资料、数值预报产品及预报员订正结果生成。数据空间分辨率为 10 km，时间分辨率为 1 h，能够提供未来 30 天的滚动更新预报产品。结合分布式光伏发电特性，本文选取气温、相对湿度、总云量、风速、降水、能见度和 1 h 累计降水等气象要素，提取预测日 24 h 逐小时预报数据作为模型输入特征，并与对应日期的区域分布式光伏日发电量进行时空匹配，用于后续发电量预测模型的训练与验证。

2.3. 晴空辐射模型

太阳辐射是影响光伏发电量的关键因素，而实际到达地表的太阳辐射受太阳位置、地理位置及大气条件等多种因素共同影响。为表征理论晴空条件下太阳辐射的变化规律，本文采用 Ineichen 晴空辐射模型计算研究区域的晴空全球水平辐照度(Global Horizontal Irradiance, GHI)。Ineichen 模型综合考虑太阳天

顶角、大气质量、海拔及大气浑浊度等因素, 具有计算效率高、精度较好的特点, 已广泛应用于太阳能资源评估及光伏发电预测等研究。晴空条件下的全球水平辐照度可表示为:

$$GHI_{cs} = DNI_{cs} \cos \theta_z + DHI_{cs}$$

式中, GHI_{cs} 为晴空全球水平辐照度, DNI_{cs} 为晴空直射法向辐照度, DHI_{cs} 为晴空散射水平辐照度, θ_z 为太阳天顶角。本文基于研究区域的经纬度、海拔及预测时刻信息, 采用 Ineichen 晴空辐射模型计算预测日 24 h 逐小时晴空水平辐照度, 并将计算得到的晴空辐射作为模型输入特征之一, 与气温、相对湿度、总云量、风速、能见度及 1 h 累计降水等气象要素共同构建分布式光伏日发电量预测模型。

2.4. XGBoost 模型

XGBoost (Extreme Gradient Boosting) 是一种基于梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 的集成学习算法, 由多个弱学习器按照加法模型进行迭代训练, 通过不断拟合前一轮模型的残差, 逐步提高模型的预测精度。由于 XGBoost 能够有效处理非线性关系、变量间复杂交互作用及高维特征数据, 同时具有较强的泛化能力, 因此被广泛应用于光伏发电预测、风电功率预测及负荷预测等领域。XGBoost 采用加法模型构建预测函数, 其第 t 次迭代后的预测结果可表示为:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$

其中, $\hat{y}_i^{(t)}$ 为第 i 个样本经过第 t 次迭代后的预测值, x_i 为输入特征, $f_t(x_i)$ 为第 t 棵决策树输出的修正量。模型训练过程中, 以最小化目标函数为优化目标, 其目标函数定义为:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

式中, n 为训练样本数; l 为损失函数, 用于衡量预测值与真实值之间的误差, y_i 为第 i 个样本真实值, $\Omega(f_k)$ 为第 k 棵决策树的正则化项, 用于控制模型复杂度, 防止模型过拟合, 其表达式为:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

其中, T 为决策树叶节点数量, w_j 为第 j 个叶节点对应的权重, γ 和 λ 为正则化系数。通过引入正则化项, XGBoost 在提高模型泛化能力的同时, 有效降低了模型过拟合风险。模型训练过程中, XGBoost 利用目标函数的一阶梯度和二阶梯度信息, 不断寻找最优分裂点, 并生成新的决策树对上一轮预测误差进行修正。经过多轮迭代后, 各决策树共同组成最终预测模型, 能够有效学习输入特征与目标变量之间复杂的非线性关系。本文以南浔地区分布式光伏日发电量为预测目标, 将预测日前 24 h 逐小时气温、相对湿度、总云量、风速、能见度、1 h 累计降水以及晴空辐射等特征作为模型输入, 利用 XGBoost 建立分布式光伏日发电量预测模型, 通过学习气象要素与光伏发电量之间的非线性映射关系, 实现基于精细化气象数据的分布式光伏日发电量预测。

3. 南浔地区分布式光伏日发电量特征

3.1. 分布式光伏电站基本特征

以浙江省南浔地区分布式光伏电站为研究对象, 本文收集了区域内多个分布式光伏电站的连续发电数据。样本涵盖不同装机规模的电站, 能够较好地反映南浔地区分布式光伏系统的运行特征。图 1 给出了研究区域内已收集分布式光伏电站的装机容量分布, 覆盖了从百千瓦级到兆瓦级等不同规模。其中, 中小容量电站数量较多, 同时也包含若干兆瓦级分布式光伏电站, 具有较好的代表性。

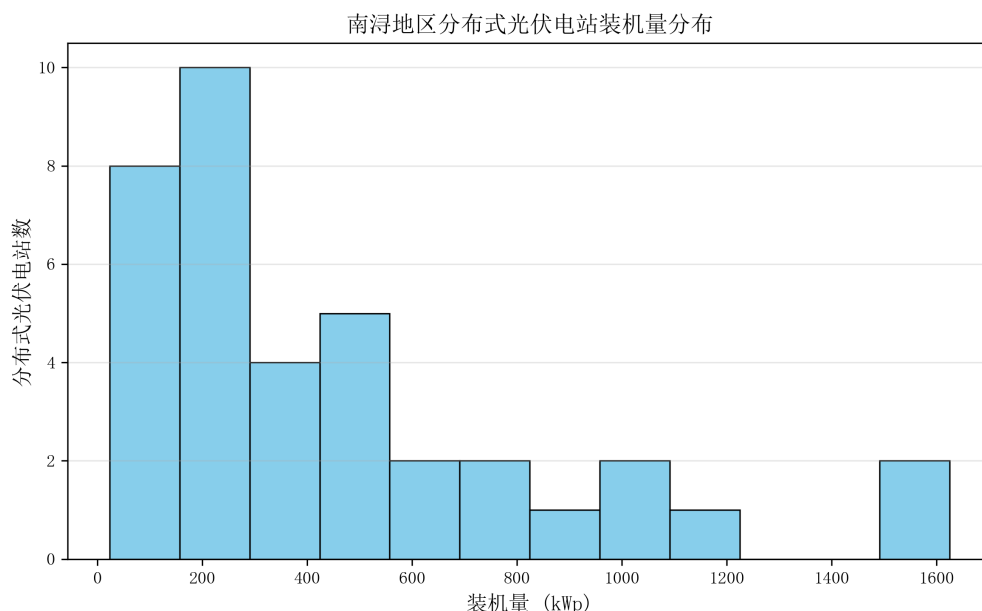


Figure 1. Distribution of installed capacity of sampled distributed photovoltaic power plants in Nanxun Region
图 1. 南浔地区样本分布式光伏电站装机容量统计分布

由于不同装机规模电站的绝对发电量差异较大，若直接比较发电量，难以客观反映不同电站的运行性能。因此，采用单位装机容量发电量作为评价指标，即将日发电量统一换算为每 100 kWp 装机容量对应的发电量(kWh/100 kWp)，以消除装机容量差异带来的影响，提高不同电站之间的可比性。为进一步分析不同规模电站的发电特征，本文选取了 4 座具有代表性的分布式光伏电站，其装机容量分别约为 160 kWp、550 kWp、1047 kWp 和 1625 kWp。图 2 展示了这 4 座典型电站单位装机容量日发电量和周发电量的时间变化。从图中可以看出，各电站均表现出较为一致的季节变化规律：夏季受太阳辐射资源丰富、日照时间较长等因素影响，单位装机容量发电量整体较高；冬季则受太阳高度角降低、日照时长缩短及天气条件影响，发电量明显下降。虽然不同规模电站在个别时段存在一定差异，但总体变化趋势基本一致，说明区域内光伏发电主要受太阳辐射和气象条件控制，而装机规模对归一化后的发电特性影响较小。这也表明采用单位装机容量发电量作为统一评价指标具有合理性，为后续发电特性分析和模型构建奠定了基础。

3.2. 分布式光伏气象要素分析

为分析不同气象要素对分布式光伏发电的影响程度，采用 XGBoost 模型对输入特征进行了重要性分析。模型输入包括装机容量以及逐小时气象要素，涵盖降水、温度、相对湿度、风速、云量、能见度和晴空辐射等变量。图 3 给出了各输入特征的总体重要性排序。从整体结果来看，云量和能见度的重要性最高，是影响分布式光伏发电预测精度的主要因素。这是由于云层变化直接影响太阳辐射的透过率，而能见度能够在一定程度上反映大气透明度及气溶胶等因素对太阳辐射传播的影响，因此两者与光伏出力具有较强的相关性。温度和晴空辐射的重要性次之，其中温度影响光伏组件的转换效率，晴空辐射则反映了理想天气条件下的太阳辐照水平，对发电功率具有较好的表征能力。装机容量的重要性位于中等水平，说明不同电站的规模差异对发电能力具有一定影响，但其贡献低于主要气象因素。相比之下，1 h 降水、相对湿度和风速的重要性相对较低，表明这些因素对分布式光伏发电的直接影响较弱，更多表现为对太阳辐射或组件运行状态的间接影响。

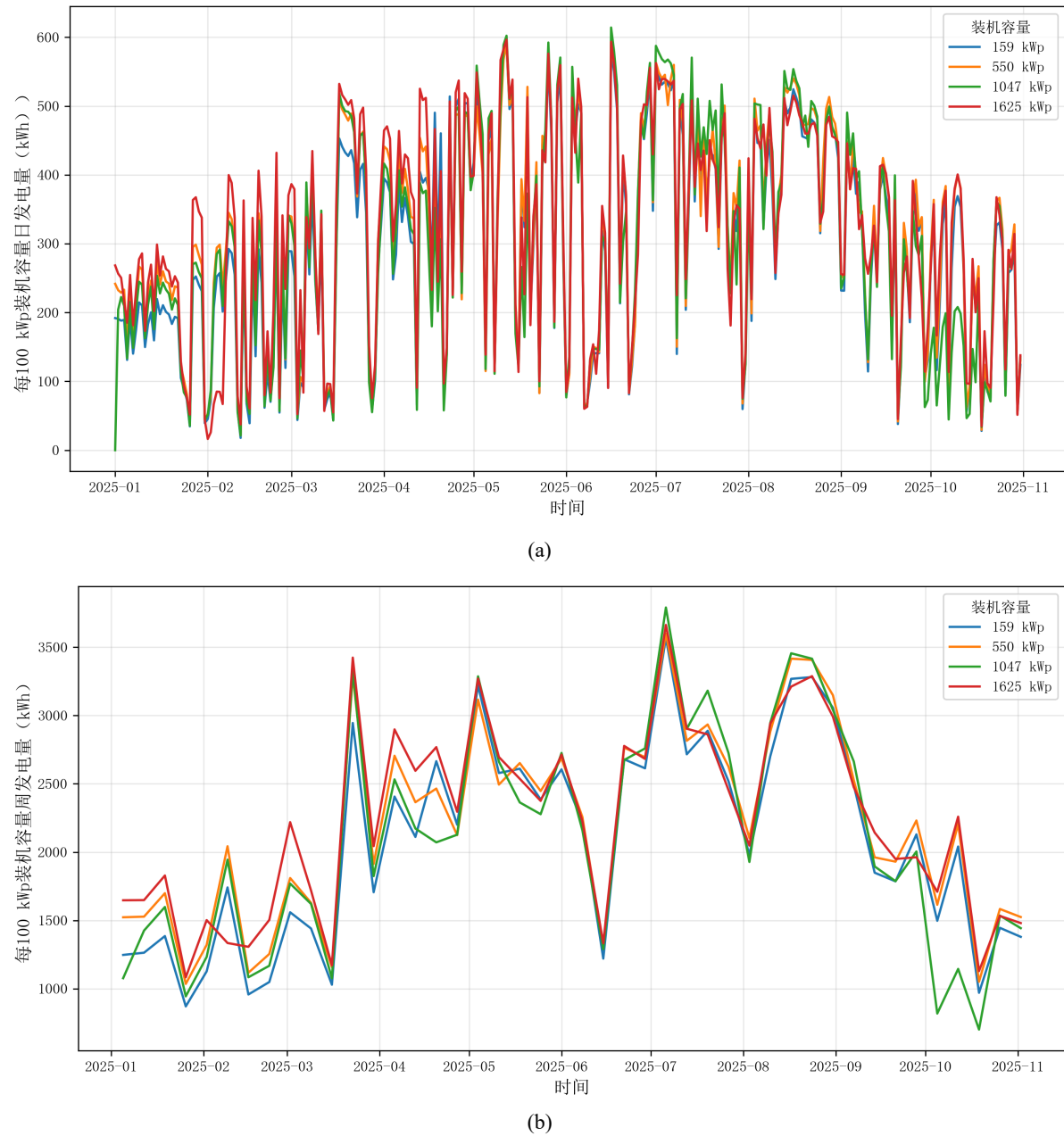


Figure 2. Temporal variations of daily (a) and weekly (b) electricity generation per unit installed capacity for typical power plants in Nanxun Region
图 2. 南浔地区典型电站单位装机容量日发电量(a)和周发电量(b)时间变化

为了进一步分析不同气象要素在一天内各时段对光伏发电的影响差异,对 4~20 时各小时输入特征的重要性进行了统计分析,结果如图 4 所示。总体来看,各气象要素的重要性具有明显的时变特征。其中,云量的重要性在 7~17 时期间均保持较高水平,并在 12 时左右达到峰值,表明白天尤其是太阳辐射最强时段,云层变化对光伏发电功率的影响最为显著。这是由于云量直接影响太阳辐射的透过率,其变化会导致光伏组件接收到的辐照度发生明显波动,从而对发电功率产生直接影响。相比之下,能见度的重要性主要集中在 12~14 时,其余时段的重要性相对较低。能见度能够反映大气透明度以及气溶胶、水汽等因素对太阳辐射传播的影响,在太阳高度角较高、太阳辐射较强的中午时段,其对光伏发电的影响更加

明显，因此表现出较高的特征重要性。而在早晚时段，由于太阳辐射本身较弱，能见度变化对发电功率的影响相对有限。其余气象要素的重要性整体低于云量和能见度，虽然在部分时段存在一定波动，但总体贡献较小。上述结果表明，不同气象要素对分布式光伏发电的影响具有显著的时间依赖性，其中云量是影响全天发电过程的关键因素，而能见度主要在中午时段发挥重要作用。

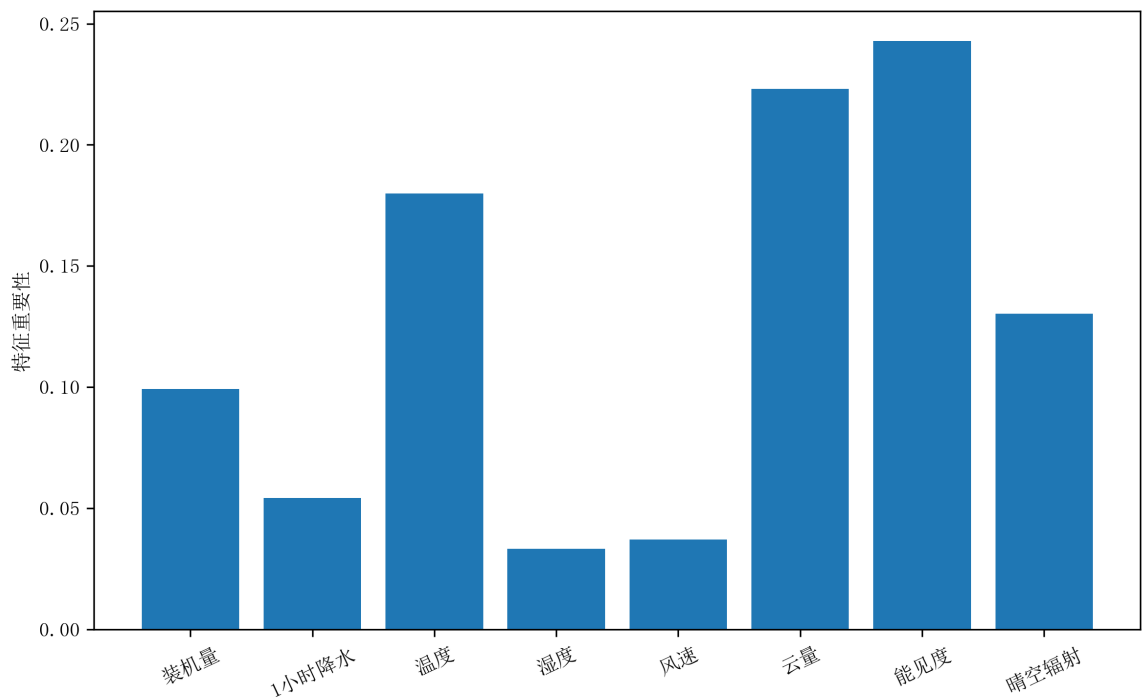


Figure 3. Feature importance ranking of input variables in the XGBoost model
图 3. XGBoost 模型输入特征重要性排序

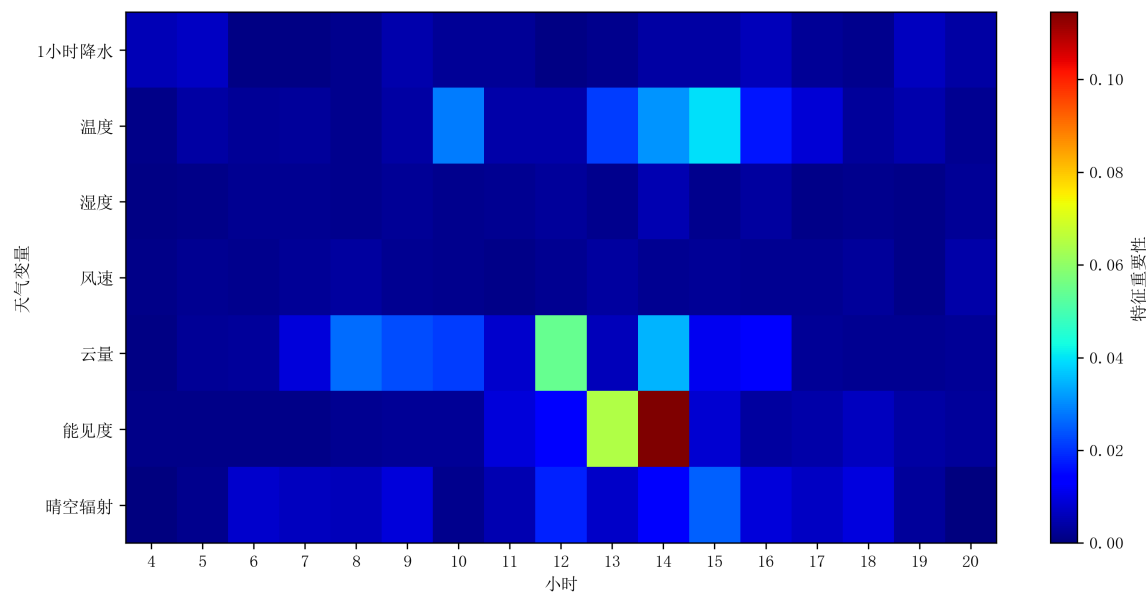


Figure 4. Feature importance ranking of input variables for each time period from 04:00 to 20:00
图 4. 4~20 时各时段输入特征重要性排序

4. 南浔地区分布式光伏日发电量预测

4.1. 分布式光伏日发电量预测模型训练结果

为了验证所建立模型对分布式光伏日发电量的预测性能,采用 XGBoost 模型对研究样本进行训练。模型输入包括装机容量以及 4~20 时逐小时气象要素,涵盖降水、温度、相对湿度、风速、云量、能见度 and 晴空辐射等变量,输出为对应日期的日发电量。实验中,将全部样本按照 8:2 的比例随机划分为训练集和验证集,其中 80% 的样本用于模型训练,20% 的样本用于模型验证。为了评价模型的预测性能,本文采用决定系数(R^2)、平均绝对百分比误差(MAPE)以及预测准确率(Accuracy)作为评价指标。其中, R^2 用于衡量模型对日发电量变化规律的拟合能力,MAPE 用于反映预测值与真实值之间的相对误差,预测准确率定义为:

$$\text{Accuracy} = 1 - \text{MAPE}$$

该指标能够直观反映模型预测结果与真实值之间的接近程度,预测准确率越高,说明模型预测性能越好。

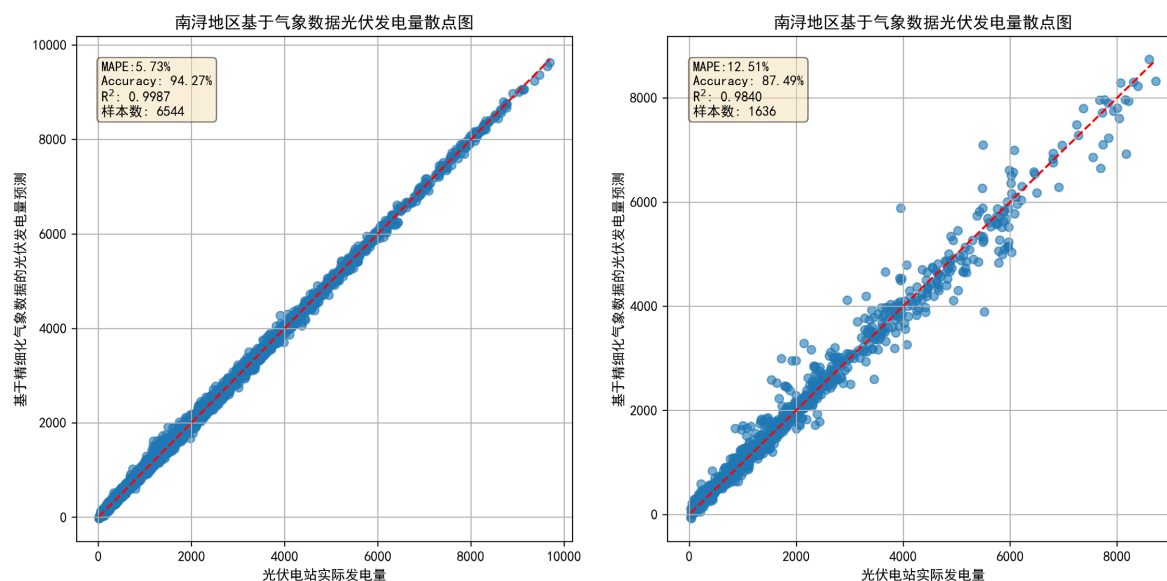


Figure 5. Scatter plots of predicted versus observed values of distributed photovoltaic daily electricity generation for the training (left) and validation (right) sets using the XGBoost model

图 5. XGBoost 模型训练集(左图)和验证集(右图)分布式光伏日发电量预测值与真实值散点图

模型训练结果如图 5 所示。训练集的决定系数 R^2 达到 0.99, 验证集 R^2 达到 0.98, 表明 XGBoost 模型能够较好地学习分布式光伏日发电量与各输入特征之间的非线性关系,并具有良好的泛化能力。训练集和验证集的 MAPE 分别为 5%和 13%,对应的预测准确率分别为 95%和 87%。相比训练集,验证集预测误差有所增加,但总体仍保持较高的预测精度,说明所建立的模型能够较准确地反映分布式光伏日发电量的变化规律。图 5 给出了模型预测值与真实值的对比结果。从图中可以看出,模型预测值与真实值整体具有较高的一致性,大部分样本均分布在 1:1 参考线附近,仅在高日发电量样本处存在一定的预测偏差。这主要是因为高日发电量通常对应于装机容量较大的分布式光伏电站,而研究样本中大装机容量电站数量相对较少,导致高发电量样本占比较低,训练数据存在一定的样本分布不均衡。受此影响,模型对高发电量样本特征的学习不够充分,因此在该区间的预测误差相对较大。尽管如此,由于高发电量

样本所占比例较小,模型在整体样本上的预测性能仍保持较高水平,验证集 R^2 达到 0.98,预测准确率达到 87%,表明所建立的 XGBoost 模型具有较好的预测性能,可为分布式光伏日发电量预测提供可靠的技术支撑。

4.2. 分布式光伏日发电量预测误差分析

为进一步分析模型在不同样本条件下的预测稳定性与适用性,本节对模型预测误差进行深入分析(见图 6)。从误差分布结果可以看出,在日发电量较小的样本中,部分数据的绝对百分比误差相对较大。这主要是由于低发电量通常对应光照条件较差或阴雨天气,此时发电功率本身较低,导致相同的绝对误差在相对意义上被放大,从而使得绝对百分比误差偏高。进一步分析装机容量与预测误差的关系可以发现,装机容量与日发电量之间存在明显的正相关关系,即装机容量越大,其对应的日发电量通常也越高。在装机容量较小的样本中,由于发电水平整体偏低且波动较大,误差较大的样本点数量相对更多。然而,从整体趋势来看,绝对百分比误差与装机容量之间并未表现出明显的单调相关关系,说明模型并未对不同装机规模的样本产生系统性偏差。仅在装机容量较大的样本中,由于对应的日发电量较高,其绝对误差的数值相对更大,但从绝对百分比误差角度来看并未显著恶化。综合来看,模型在不同发电水平及不同装机容量条件下均具有较稳定的预测性能,仅在低发电量样本中由于相对误差放大效应以及样本分布差异,绝对百分比误差表现出一定波动,但整体误差分布较为合理,模型具有较好的鲁棒性。

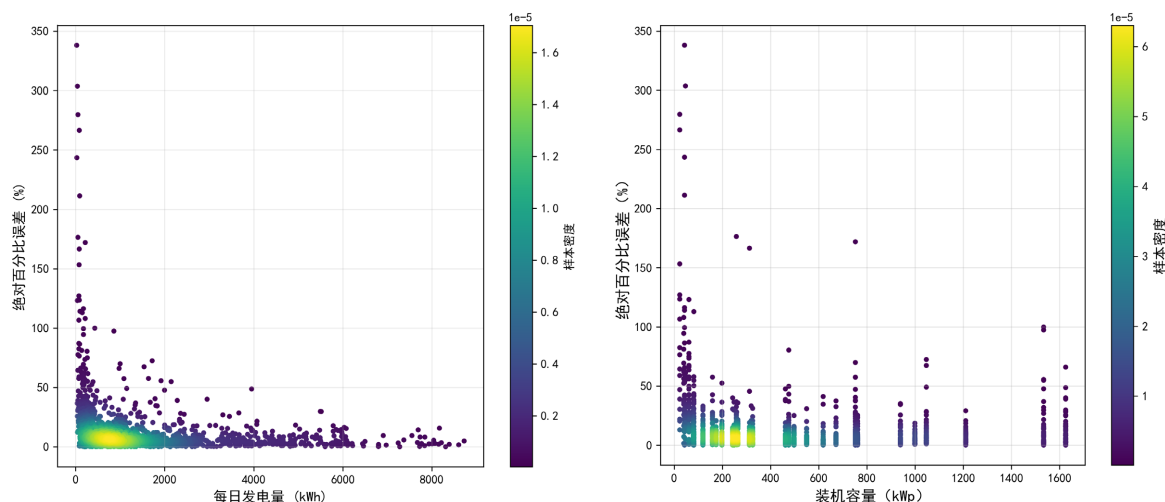


Figure 6. Relationships between the absolute percentage error of distributed photovoltaic daily electricity generation prediction and daily electricity generation (left) and installed capacity (right)

图 6. 分布式光伏日发电量预测绝对百分比误差与日发电量(左图)及装机容量(右图)关系

从气象要素影响分析可以发现(见图 7),平均云量与能见度与预测绝对百分比误差之间并未表现出明显的单调依赖关系。尽管在特征重要性分析中,云量和能见度均具有较高的权重,但在误差分布图中可以看出,不同云量及能见度条件下的绝对百分比误差值整体分布较为分散,并未呈现出随云量增加或能见度降低而系统性增大的趋势。具体而言,大部分样本的绝对百分比误差均集中在 20%以内,在不同云量与能见度区间内均保持较为稳定的分布。这表明模型虽然在特征选择阶段对云量和能见度赋予了较高的重要性,但在实际预测误差层面,这两类气象要素并未对误差产生显著的单向影响。这一结果说明,模型能够较好地综合利用多气象要素信息进行日发电量预测,其误差主要受多因素共同作用影响,而非由单一气象变量主导。与上文特征重要性分析结果相比,虽然云量与能见度在模型输入中占据重要地位,

但其影响主要体现在整体预测精度提升，而非直接决定误差大小规律。

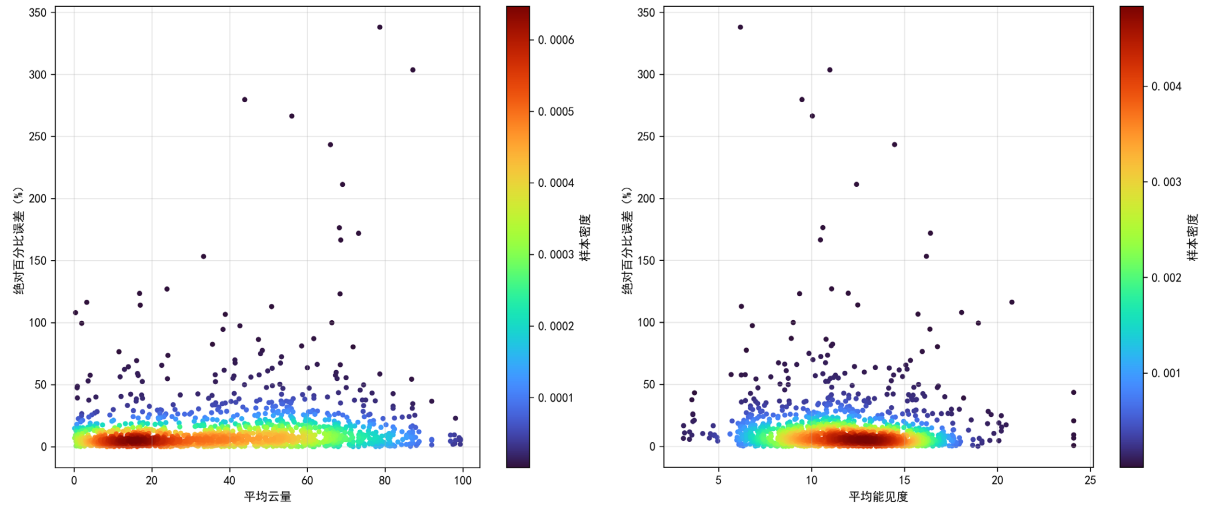


Figure 7. Relationships between the absolute percentage error of distributed photovoltaic daily electricity generation prediction and average cloud cover (left) and average visibility (right)

图 7. 分布式光伏日发电量预测绝对百分比误差与平均云量(左图)及平均能见度(右图)关系

进一步分析平均温度和平均晴空辐射与预测绝对百分比误差之间的关系(见图 8)可以发现，两者与预测误差同样未表现出明显的单调依赖关系。虽然温度和晴空辐射在模型特征重要性分析中具有一定贡献，但其重要性低于云量和能见度。从误差分布结果来看，不同温度及晴空辐射条件下的绝对百分比误差整体分布较为均匀，未出现误差随温度升高或晴空辐射增强而显著变化的趋势。同时，大部分样本的绝对百分比误差仍集中在 20%以内，说明模型在不同温度和晴空辐射条件下均能够保持较稳定的预测性能。这表明，温度和晴空辐射虽然能够为模型提供有效的辅助信息，但对预测误差的影响并非独立存在，而是与云量、能见度及其他气象要素共同作用于日发电量预测过程。

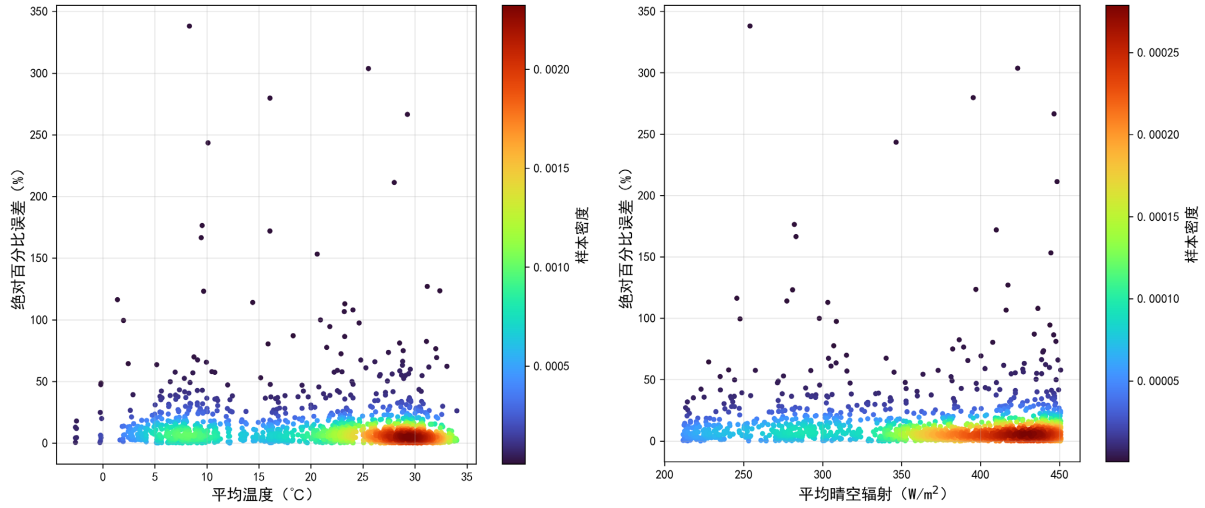


Figure 8. Relationships between the absolute percentage error of distributed photovoltaic daily electricity generation prediction and average temperature (left) and average clear-sky irradiance (right)

图 8. 分布式光伏日发电量预测绝对百分比误差与平均温度(左图)及平均晴空辐射(右图)关系

5. 结论

本文以浙江省南浔地区分布式光伏电站为研究对象, 基于不同装机规模电站的发电数据及逐小时气象数据, 构建了分布式光伏日发电量预测模型, 并对影响发电预测的主要气象因素及模型误差特征进行了分析。研究表明, 采用单位装机容量发电量能够有效消除不同规模电站之间的容量差异, 不同装机规模电站在归一化后均表现出一致的季节变化规律, 说明区域分布式光伏发电主要受太阳辐射资源和气象条件控制, 装机规模对归一化后的发电特性影响相对有限, 为不同电站发电特性的统一分析和模型构建提供了基础。

气象要素分析结果表明, 不同气象因素对分布式光伏发电的影响存在明显差异。云量和能见度是影响发电预测的关键气象要素, 其中云量在白天各时段均具有较高的的重要性, 而能见度主要在中午时段发挥较大作用, 反映出太阳辐射变化及大气透明度是影响区域分布式光伏发电的重要因素。相比之下, 温度、晴空辐射、降水、相对湿度和风速等气象要素虽然能够为模型提供有效信息, 但其影响程度相对较低。

基于 XGBoost 构建的日发电量预测模型能够充分挖掘装机容量与多种气象要素之间的非线性关系, 预测结果与实际发电量保持较高的一致性, 在训练集和验证集上均表现出良好的拟合能力和泛化性能。误差分析进一步表明, 模型在不同装机容量及不同气象条件下均具有较好的稳定性, 预测误差未随云量、能见度、温度及晴空辐射等单一气象要素变化而表现出明显的系统性规律, 说明模型能够综合利用多种气象信息进行预测, 其误差主要受多因素共同作用影响。未来可结合 SHAP 等模型解释方法, 对极端天气过程和高发电量场景下的预测误差来源进行深入分析。此外需要指出的是, 本文模型基于浙江南浔地区分布式光伏运行数据构建, 虽然能够较好地反映该区域光伏发电特征, 但不同气候区域在气象条件和光伏运行特征方面存在差异, 模型的跨区域泛化能力仍需进一步验证。

基金项目

湖州市公益性应用研究项目“基于气象大数据的湖州地区光伏发电效率预报服务研究(2024GZ73)”。

参考文献

- [1] “十四五”能源领域科技创新规划[J]. 中国科技奖励, 2022(9): 23-42.
- [2] 栗峰, 丁杰, 周才期, 等. 新型电力系统下分布式光伏规模化并网运行关键技术探讨[J]. 电网技术, 2024, 48(1): 184-199.
- [3] 宋祥斌. 温度对屋顶分布式光伏系统发电效率的影响分析[J]. 能源研究与信息, 2018, 34(3): 136-140.
- [4] 许国泽. 考虑多因素影响的光伏发电功率智能预测研究[J]. 自动化技术与应用, 2023, 42(10): 162-166.
- [5] 蔡源, 吴浩, 唐丹. 光伏发电功率预测方法综述[J]. 四川电力技术, 2024, 47(2): 25-31.
- [6] Cantillo-Luna, S., Moreno-Chuquen, R., Celeita, D. and Anders, G. (2023) Deep and Machine Learning Models to Forecast Photovoltaic Power Generation. *Energies*, **16**, Article No. 4097. <https://doi.org/10.3390/en16104097>
- [7] Mohamad Radzi, P.N.L., Akhter, M.N., Mekhilef, S. and Mohamed Shah, N. (2023) Review on the Application of Photovoltaic Forecasting Using Machine Learning for Very Short- to Long-Term Forecasting. *Sustainability*, **15**, Article 2942. <https://doi.org/10.3390/su15042942>
- [8] Scott, C., Ahsan, M. and Albarbar, A. (2023) Machine Learning for Forecasting a Photovoltaic (PV) Generation System. *Energy*, **278**, Article ID: 127807. <https://doi.org/10.1016/j.energy.2023.127807>
- [9] 栗然, 李广敏. 基于支持向量机回归的光伏发电出力预测[J]. 中国电力, 2008(2): 74-78.
- [10] 谈玲, 康瑞星, 夏景明, 等. 融合多源异构气象数据的光伏功率预测模型[J]. 电子与信息学报, 2024, 46(2): 503-517.
- [11] 余光正, 陆柳, 汤波, 等. 基于云图特征提取的改进混合神经网络超短期光伏功率预测方法[J]. 中国电机工程学报, 2021, 41(20): 6989-7003.

- [12] 孙景博, 王阳, 杨晓帆, 等. 中国风光资源气候风险时空变化特征分析[J]. 中国电力, 2023, 56(5): 1-10.
- [13] 闫全全, 李芬, 王丽娟, 等. 不同季节光伏发电量与气象影响因子的诊断分析[J]. 上海电力学院学报, 2016, 32(3): 261-266.
- [14] 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.