

基于Copula联合概率密度模型和BP神经网络的风电场选址研究

赵经阳, 马海婧*

内蒙古工业大学理学院, 内蒙古 呼和浩特

收稿日期: 2026年4月27日; 录用日期: 2026年5月19日; 发布日期: 2026年5月28日

摘要

为提升风能资源评估准确性并为风电场选址提供科学依据, 本研究综合联合概率模型与神经网络方法, 选取内蒙古自治区通辽、阿尔山、阿巴嘎旗、东胜及鄂托克旗五个气象站点2020~2024年的气象观测数据, 依据不同下垫面特征选取对应风切变指数将地面风速与空气密度统一外推至50米轮毂高度后分析各站点风向、风速及空气密度的统计特征, 发现风速与空气密度分别服从广义极值分布与混合正态分布且通过Frank Copula函数构建二者联合概率模型以捕捉相依结构, 在此基础上构建以风矢量分量、气压、空气密度和海拔为输入的BP神经网络模型并应用于新站点锡林浩特的预测验证, 结果显示模型具备良好泛化能力与预测精度, 所构建的“概率模型-神经网络预测”融合框架可系统评估区域风能资源, 为风电场科学选址提供可靠理论基础与技术支持。

关键词

风切变指数, 联合概率分布, 风功率密度, BP神经网络, 风电场选址

A Study on Wind Farm Siting Based on Copula Joint Probability Density Models and Backpropagation Neural Networks

Jingyang Zhao, Haijing Ma*

College of Science, Inner Mongolia University of Technology, Hohhot Inner Mongolia

Received: April 27, 2026; accepted: May 19, 2026; published: May 28, 2026

*通讯作者。

文章引用: 赵经阳, 马海婧. 基于 Copula 联合概率密度模型和 BP 神经网络的风电场选址研究[J]. 统计学与应用, 2026, 15(5): 226-244. DOI: 10.12677/sa.2026.155121

Abstract

To enhance the accuracy of wind energy resource assessment and provide scientific basis for wind farm siting, this study integrates probabilistic models with neural network methods. Meteorological observation data are from 2020 to 2024 at five stations: Tongliao, Aarshan, Abaga Banner, Dongsheng, and Etuoke Banner in Inner Mongolia Autonomous Region. Based on different land surface characteristics, corresponding wind shear indices were selected to uniformly extrapolate surface wind speed and air density to a hub height of 50 metres. The statistical characteristics of wind direction, wind speed, and air density at each site were then analysed. It was found that wind speed and air density respectively followed a generalized extreme value distribution and a mixed normal distribution. A joint probability model was constructed using the Frank Copula function to capture their interdependent structure. Building upon this, a BP neural network model was developed with wind vector components, barometric pressure, air density, and altitude as inputs. This model was applied to predict and validate data for the new site in Xilinhot. Results demonstrate the model's robust generalisation capability and predictive accuracy. The developed "probabilistic model-neural network prediction" integrated framework enables systematic assessment of regional wind energy resources, providing a reliable theoretical foundation and technical support for the scientific siting of wind farms.

Keywords

Wind Shear Index, Joint Probability Distribution, Wind Power Density, BP Neural Network, Wind Farm Siting

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

内蒙古自治区是我国风能资源最富集的区域之一, 广袤的土地与独特的气候条件为风电产业发展提供天然优势, 《“十四五”现代能源体系规划》[1]明确提出加快内蒙古等可再生能源基地建设, 风电场选址作为风电开发的前置性关键环节, 其研究方案的合理性直接影响项目经济收益与生态可持续性。当前风电场选址研究中, 站点与评价指标的选择是核心技术问题, 需同时满足“区域覆盖全面”与“参数表征精准”的双重要求。区域风电场风能资源评估结果将直接影响区域风电场的前期建设发展规划[2]-[4]。

空气的水平运动称为风, 不仅有数值的大小, 还具有方向。地面风向常采用 16 方位制描述, 相邻方位间夹角为 22.5° , 风玫瑰图是专业的风向风速统计与可视化工具[5]。在早期风速概率分布研究中, 威布尔分布、高斯分布及瑞利分布等单一模型被广泛用于拟合风速数据[6]。然而, 实际风速分布常呈现双峰或多峰形态, 此类复杂结构难以被单一分布准确刻画[7]。为此, 学者陆续提出各类混合分布模型[8]。该类模型通过将两个或多个单一分布函数进行加权组合, 能够有效描述多峰风速分布特征, 显著提升了拟合精度[9] [10]。由于风速数据来自于不同的站点, 应选取不同的风切变指数, 风切变指数不仅受地表粗糙度影响, 在不同下垫面的风切变指数不同[11] [12], 还受热力学因素影响, 同一地区在不同时间段内风切变指数也不同。Jung 和 Schindler [13]比较了 Gamma、Weibull、Burr、Dagum、Generalized Extreme Value 和 Wakeby 分布的拟合性能, 最终采用五参数 Wakeby 分布对 140 米涡轮机高度的空气密度进行建模, 进而分析了德国全境的年平均风能产量。该研究实现了空气密度对高海拔山区与低海拔沿海地区风能产出

影响的定量刻画, 结果凸显了在风能资源评估中纳入空气密度因素的必要性。实际站点的密度概率分布往往呈现多峰的形态, 需考虑混合分布能够更好的拟合。

在风力发电机选址前, 需对目标区域进行准确的风能评估以保障经济效益。然而, 传统评估方法通常过度依赖风速单一变量, 未能充分纳入风速与风向、风速与空气密度之间的相关性, 存在一定局限性。Copula 函数能够有效刻画上述变量间的内在关联, 通过建立多变量联合概率分布, 显著提升风能评估的准确性[14] [15]。为风力发电机的精确选址提供重要依据。文献[16]提出了一种基于风速的截断正态-Weibull 混合分布和风向的 von Mises 分布构建的联合分布模型。文献[17]结合风速和风向联合概率密度评估风向对机组载荷的影响。

在短期风电功率预测中, 文献[18]提出了卷积注意力机制和长短时记忆网络相结合的风功率预测模型, 为优化机组策略提供了指导。BP 神经网络及其改进模型得到了广泛研究。Wang 等[19]采用基础 BP 神经网络进行预测; Zhang 等[20]则引入 LCASO-BP 神经网络, 并通过与 GA-BP 和 PSO-BP 模型的对比, 验证了其更好的预测性能。此外, Wang [21]将 BP 神经网络与 ERA5 再分析资料相结合, 对我国北方某风电场进行功率预测, 为观测数据稀缺或质量不高的区域提供了可行的预测参考。

总之, 现有研究虽已关注空气密度与风速对风能产量的影响, 但普遍缺乏能够精准表征二者相关性的数学统计模型, 难以充分反映风资源的耦合特征。本研究以内蒙古自治区多气象站点数据为基础, 首先结合不同下垫面的风切变指数差异对原始风速数据进行修正, 提升风速数据的局地适配性; 随后分别采用广义极值分布拟合风速的边缘分布、混合正态分布拟合空气密度的边缘分布, 并通过 Copula 函数构建二者的联合分布模型, 实现对风速-空气密度耦合关系的量化表征; 在此基础上, 将新候选站点的气象数据输入训练完成的 BP 神经网络模型, 借助模型的非线性拟合能力完成风资源潜力的快速评估, 最终为区域风电场的宏观选址提供兼具精准性与高效性的科学参考依据。

2. 数据来源与处理及研究方法

2.1. 数据概括

本研究选取内蒙古自治区的通辽、阿尔山、阿巴嘎旗、东胜及鄂托克旗五个气象站点, 站点经纬度见表 1, 2020 年 1 月 1 日至 2024 年 12 月 31 日的地面观测数据作为研究样本, 数据来源于中国地面气象观测资料, 包括 10 m 高度风速、2 m 高度气温及海平面气压的 3 小时观测记录。日平均风速由每日 8 次观测值平均计算获得, 空气密度则根据气压与气温数据间接计算得出。为贴近风电场运行实际, 所有风速与空气密度数据均根据内蒙古地区一般陆上风力发电机组的设计参数, 统一换算至轮毂高度 50 米处。

Table 1. Geographic information of meteorological observation stations in inner Mongolia autonomous region

表 1. 内蒙古自治区观测气象站点地理信息

站点	经度(°E)	纬度(°N)	海拔(m)
通辽	122.16	43.36	178.7
阿尔山	119.56	47.10	997.2
阿巴嘎旗	114.57	44.01	1126.1
东胜	109.59	39.50	1461.9
鄂托克旗	107.58	39.05	1381.4

站点的选取充分考虑内蒙古风能资源“西强东弱、北丰南欠”的分布特征及地形气候影响: 位于锡林郭勒盟北部的阿巴嘎旗代表北部风能富集带; 鄂尔多斯市中部的东胜和西部的鄂托克旗反映西部荒漠草原区的中等风能资源及沙尘影响特征; 东部的通辽对应农牧交错带的风能开发条件; 东北部的阿尔山

则代表林区复杂地形下的风能特征。五个站点横跨东、中、西三大区域, 覆盖森林草原、典型草原与荒漠草原等多种地貌, 其科学布局与长期观测历史确保了数据代表性, 能够系统揭示不同区域风能资源的差异规律, 从而为各类地貌下的风电场选址提供具有普适性的参考依据。

2.2. 数据处理

为尽可能降低误差, 将空气密度与风速数据集外推至统一的轮毂高度。特定轮毂高度 H_{hub} 对应的空气密度 $\rho_{H_{hub}}$, 可借理想气体定律[22]计算得到, 其表达式为:

$$\rho_{H_{hub}} = \frac{P_{H_{hub}}}{R \cdot T_{H_{hub}}} \quad (1)$$

式中, $\rho_{H_{hub}}$ 、 $T_{H_{hub}}$ 分别为轮毂高度处的气压与气温; R 为干燥空气的大气气体常数, 取值为 $287.058 \text{ J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$ 。

轮毂高度处的气压与气温, 通过气压定律、气温定律外推获得, 公式如下:

$$P_{H_{hub}} = P_{H_0} \left(1 - \frac{0.0065 \cdot \Delta h}{T_{H_0}} \right)^{5.2561} \quad (2)$$

$$T_{H_{hub}} = T_{H_0} - 0.0065 \cdot \Delta h \quad (3)$$

上述公式中, P_{H_0} 、 T_{H_0} 代表基准高度处的气压与温度; Δh 为轮毂高度 H_{hub} 与基准高度 H_0 之间的垂直高度差。

轮毂高度 H_{hub} 下的风速 v_{hub} 通过幂律公式[23]计算得到, 数学形式如下:

$$v_{hub} = v_{10} \left(\frac{H_{hub}}{10} \right)^\alpha \quad (4)$$

其中, v_{10} 是距地面 10 米处的风速; α 是风切变指数, 本研究参照《风电场风能资源评估方法》(GB/T 18710-2002) [24], 结合各站点下垫面类型与区域气候特征选取风切变指数, 基于幂律公式将 10 米高度实测风速外推至 50 米轮毂高度。通辽站位于农牧交错带, 地表粗糙度中等, 风切变指数取 0.16; 阿尔山站地处东北林区, 地形复杂且地表粗糙度较高, 风切变指数取 0.25; 阿巴嘎旗站为典型草原, 地势开阔且地表粗糙度极低, 风切变指数取 0.14; 东胜站处于荒漠草原与城镇交错区, 地表粗糙度中等偏低, 风切变指数取 0.15; 鄂托克旗站位于西部荒漠草原, 植被稀疏、地势开阔且大气层结稳定, 风切变指数取 0.13。所选风切变指数可合理反映不同下垫面条件下风速的垂直分布规律。

计算 50 m 高度处的风矢量分量可为后续 BP 神经网络模型输入特征构建提供数据基础, 风矢量具体分解公式如下:

$$U = -v \times \sin \left(\theta \times \frac{\pi}{180} \right) \quad (5)$$

$$V = -v \times \cos \left(\theta \times \frac{\pi}{180} \right) \quad (6)$$

在二维直角坐标系下, 风矢量可分解为以正东为正方向的東西向 U 分量和以正北为正方向的南北向 V 分量, 设 v 为风速、 θ 为风向, 需先将风向从角度制转换为弧度制再通过三角函数计算分量, 公式中负号源于气象学风向为风的来向而分量定义为风的去向, 需通过符号转换实现方向对应。

2.3. 研究方法

(1) 数据获取: 获取气象站的相对海平面气压、海面温度、2 米气温及 10 米风速数据。

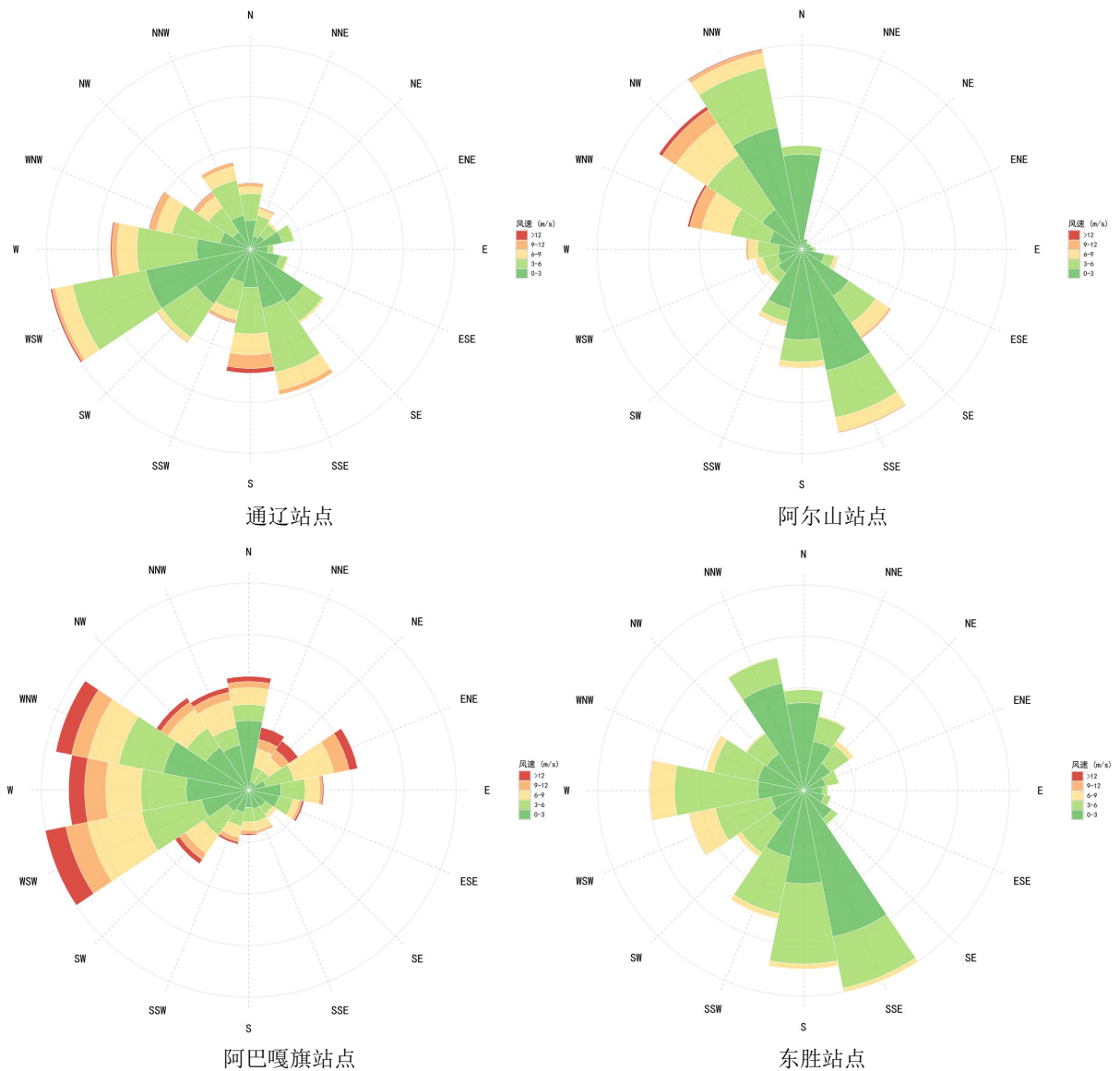
(2) 轮毂高度数据转换: 针对目标风电场区域, 结合其下垫面类型, 采用幂律公式将 10 米风速换算为 50 米轮毂高度处的风速; 同时利用理想气体定律, 结合换算得到的 50 米气压、气温数据, 计算 50 米轮毂高度处的空气密度。

(3) 联合分布构建: 分别对 50 米高度的风速与空气密度进行单变量分布拟合, 通过拟合优度指标筛选出各自的最优理论分布; 在此基础上采用 Copula 函数刻画两变量间的相依结构: 先基于其边缘经验分布通过概率积分变换将原始变量转换为[0, 1]区间的均匀分布变量, 进而构建二元联合分布模型。

(4) 风电场选址预测: 将风矢量 U 分量、风矢量 V 分量、气压、空气密度、海拔这五个特征代入训练完成的 BP 神经网络模型, 预测目标区域的风功率密度, 结合国标风功率密度分级标准, 并综合考量实际地形、地质及环境等约束条件, 完成风电场选址的可行性评估。

3. 站点风资源特征及概率分布模型对比研究

3.1. 风向特征分析



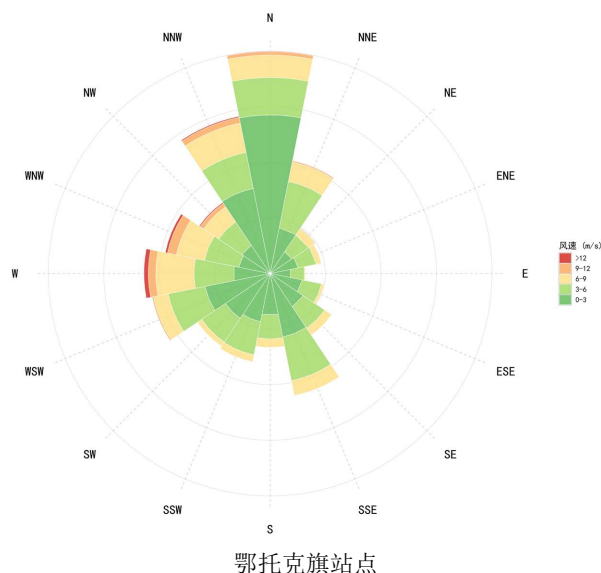
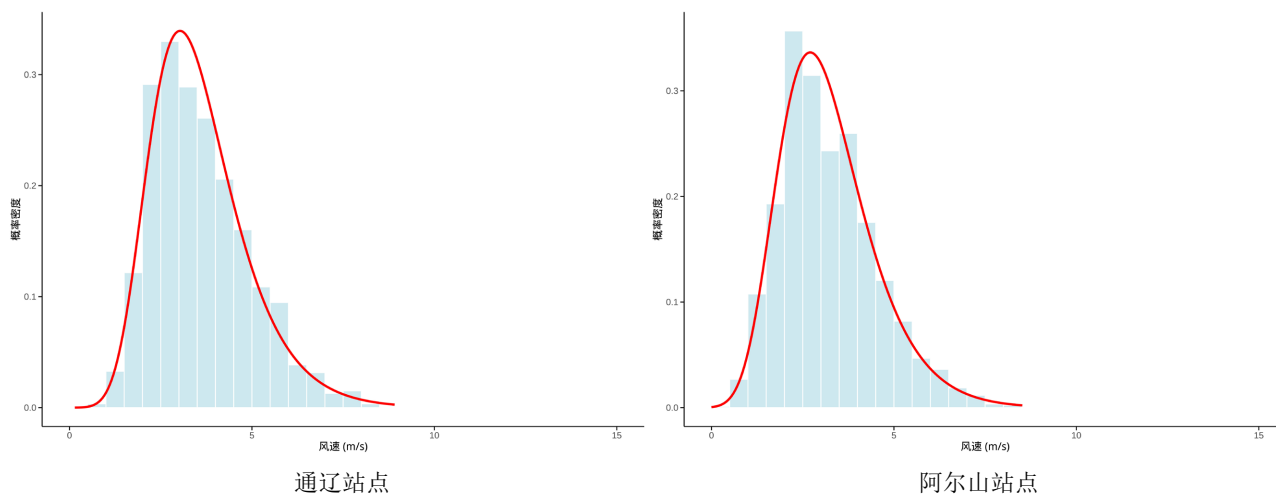


Figure 1. Wind rose diagrams of each station
图 1. 各站点风向玫瑰图

对内蒙古五个站点风向数据的统计分析绘制出风向玫瑰图(图 1), 各站点主导风向分布类型及频率特征存在明显差异, 通辽站为偏西风主导型, 第一、第二、第三主导风向分别为 WSW (14.5%)、SSE (10.5%)、W (9.9%), 前三风向合计频率 34.9%, 阿尔山站呈南北双向型特征, 主导风向依次为 NNW (15.5%)、SSE (14.2%)、NW (13.0%), 前三风向合计频率 42.7%, 阿巴嘎旗站为偏西风集中型, 主导风向为 WSW (13.4%)、WNW (12.7%)、W (11.6%), 合计频率 37.7%, 东胜站为偏南风主导型, 主导风向为 SSE (13.3%)、S (11.6%)、W (10.0%), 合计频率 34.9%, 鄂托克旗站为偏北风主导型, 主导风向为 N (14.0%)、NNW (10.1%)、W (8.0%), 合计频率 32.1%, 总体来看阿尔山站风向集中度最高、鄂托克旗站最低, 偏西风在多个站点占主导地位但各站点受地理位置和地形影响呈现不同风向分布特征, 这些差异对风电场选址具有直接指导意义, 可优先在主导风向稳定且集中度高的区域布局风机并结合各站点风向特征优化风机排列朝向, 以提升风能捕获效率、降低机组偏航损耗, 为内蒙古不同区域风电场的差异化选址与布局提供实践依据。

3.2. 广义极值分布在多站点风速拟合中的拟合研究



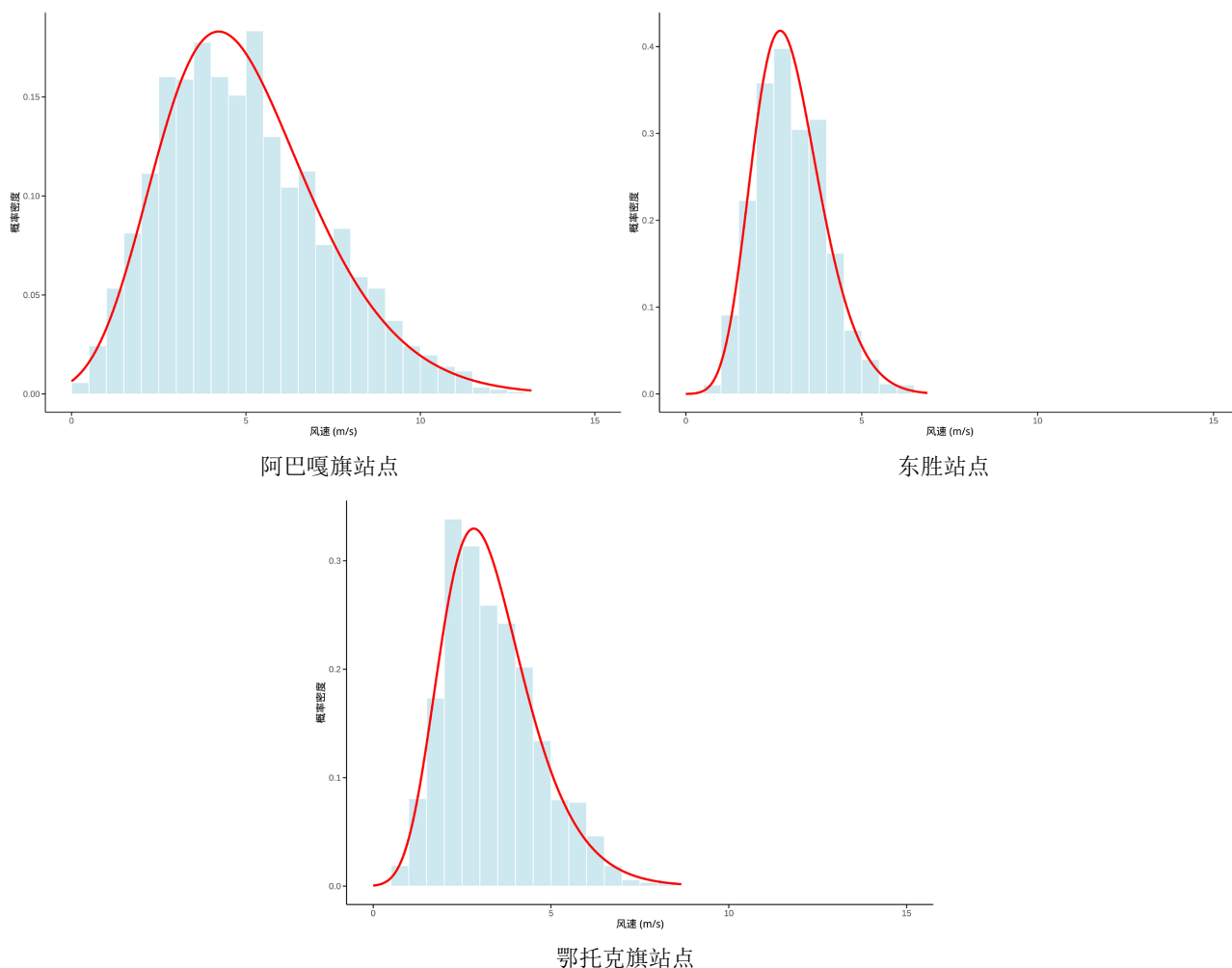


Figure 2. Fit of generalized extreme value distribution to stationary wind speed data
图 2. 广义极值分布对站点风速数据拟合图

Table 2. Statistical data of wind speed
表 2. 风速的统计数据

站点	均值(m/s)	标准差	偏度	峰度
阿巴嘎旗	5.2310	4.0741	1.0228	0.9368
阿尔山	3.4449	2.7203	1.0741	0.9876
东胜	2.9937	1.6423	0.6673	0.7369
鄂托克旗	3.4542	2.4562	1.0771	1.4418
通辽	3.8230	2.3631	1.4557	3.3102

如图 2 所示, 广义极值分布对上述五个站点的风速数据进行拟合时展现出显著综合优势, 尽管各站点风速直方图显示分布形态存在差异(如表 2 所示, 东胜风速均值仅 2.9937 m/s、集中于较低区间, 通辽、阿尔山均值分别为 3.8230 m/s、3.4449 m/s 且较高风速区间频率可观), 但 GEV 在全部站点均表现出稳健拟合性能, 从拟合图看 GEV 曲线能够较好捕捉各站点风速分布的主体与右尾较高风速区间形态、吻合良好。在统计指标方面, 由表 3 可知 GEV 的 AIC、KS 与 R² 等关键指标在五个站点中始终处于优或次优水

平, 未出现明显波动, 体现出良好跨地域适应性, 理论上 GEV 专门刻画极值分布形态, 其灵活形状参数能有效拟合风速分布的尾部特征, 更适用于风能评估中的高风速事件分析与极端风速风险建模, 此外 GEV 模型结构简洁、参数物理意义明确, 便于多站点统一比较、区域化分析以及后续联合概率建模。因此选用广义极值分布作为统一风速模型, 既能从统计上保证拟合稳健性与一致性, 也符合风速数据在能源与气象工程中的极值分析需求, 兼顾了统计解释力、理论适应性与工程实用性, 适合用于支持风电资源评估与风电场选取的后续研究。

Table 3. Fit indices of the generalized extreme value distribution model

表 3. 广义极值分布模型的拟合优度指标值

地区	$-\ln L$	AIC	BIC	KS	R^2	RMSE	MAE
通辽	-2810	5627	5643	0.036099	0.997683	0.013899	0.011371
阿尔山	-2803	5613	5630	0.038892	0.997549	0.014273	0.010335
阿巴嘎旗	-3831	7669	7686	0.026617	0.998124	0.012480	0.010756
东胜	-2371	4748	4764	0.035992	0.997525	0.014341	0.011095
鄂托克旗	-2780	5566	5582	0.031869	0.997853	0.013358	0.010606

3.3. 空气密度的时空变化特征及混合正态分布拟合研究

3.3.1. 空气密度的时空变化特征

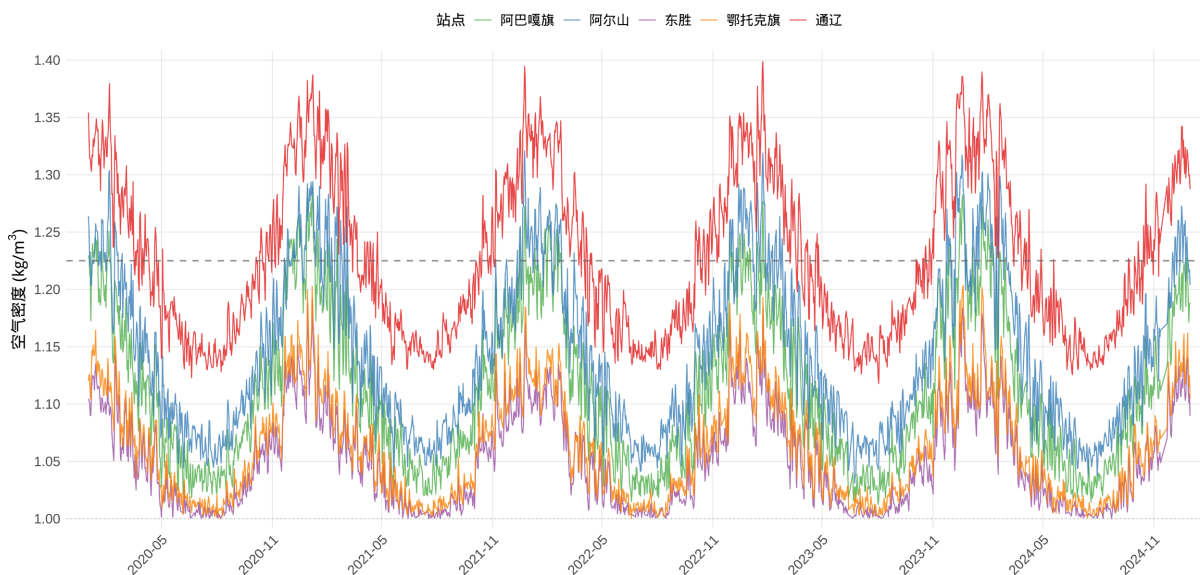


Figure 3. Variation of air density at a height of 50 meters with time scale

图 3. 五十米高空处的空气密度随时间尺度的变化

根据上述五个站点 2020~2024 年空气密度时序曲线图可以看出(见图 3), 各站点空气密度均呈现规律且同步的季节性周期波动, 表现为冬季达峰值、夏季降至谷值, 春季逐月递减至夏季维持在稳定低区间、秋季则逐月回升, 尽管整体变化模式一致, 但各站点空气密度水平存在明显差异, 这主要反映其地理环境与区域气候条件的影响, 且研究时段内序列无显著长期变化趋势, 表明区域空气密度在年际尺度上保持稳定, 这一季节性规律与空间差异特征对评估不同区域风能潜力、优化风电机组选型与布局具有关键意义。

3.3.2. 空气密度的混合正态分布拟合研究

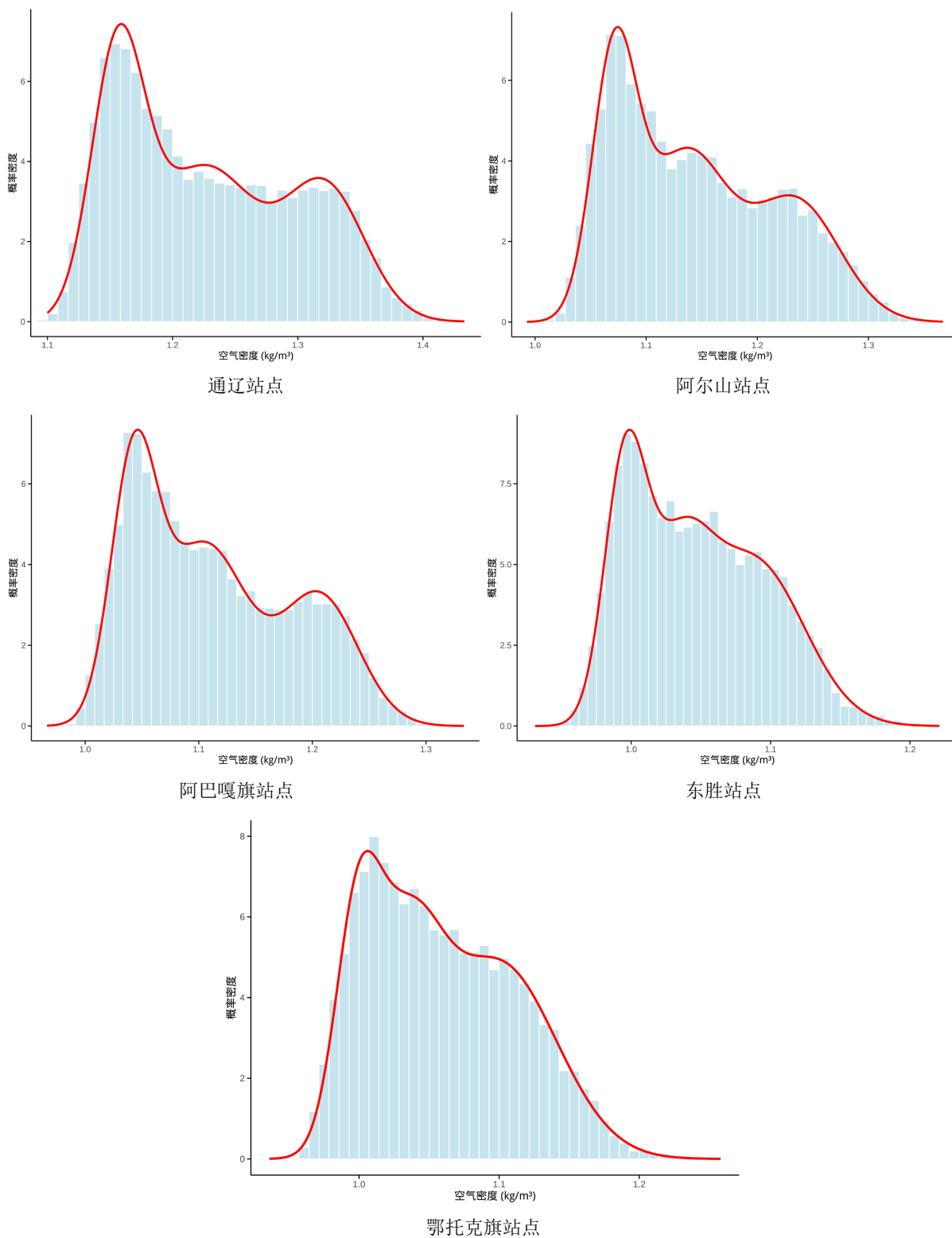


Figure 4. Fitting graph of mixed normal distribution for site air density data
图 4. 混合正态分布对站点空气密度数据拟合图

Table 4. Statistical data on air density**表 4.** 空气密度的统计数据

地区	均值(kg/m ³)	标准差	偏度	峰度
阿巴嘎旗	1.1132	0.0703	0.4496	-0.8966
阿尔山	1.1444	0.0724	0.4564	-0.8563
东胜	1.0473	0.0487	0.3997	-0.6902
鄂托克旗	1.0585	0.0532	0.4170	-0.6439
通辽	1.2273	0.0719	0.3803	-1.0335

Table 5. Goodness-of-fit index values of the mixed normal distribution model**表 5.** 混合正态分布模型的拟合优度指标值

地区	-ln L	AIC	BIC	KS	R ²	RMSE	MAE
通辽	-19088	-38161	-38100	0.0103	0.9718	0.000098	0.000073
阿尔山	-18739	-37463	-37403	0.0112	0.9799	0.000086	0.000061
阿巴嘎旗	-19227	-38439	-38378	0.0058	0.9848	0.000076	0.000057
东胜	-23751	-47487	-47426	0.0060	0.9946	0.000050	0.000039
鄂托克旗	-22289	-44563	-44502	0.0081	0.9945	0.000052	0.000036

混合正态分布对上述五个站点的数据拟合展现出全面且显著的建模优势(见图 4), 由拟合图可知各站点空气密度分布均呈非对称、局部多峰或厚尾特征, 传统单峰分布难以充分捕捉此类复杂形态, 而混合正态分布凭借多成分组合的灵活性, 在 KS 检验、信息准则及误差指标上均一致优于对比模型, 其拟合曲线与经验分布在峰值区域、尾部走向上均高度吻合; 结合表 4 空气密度统计数据与表 5 拟合优度指标, 采用混合正态分布作为上述地区空气密度概率建模的统一优选模型, 该选择既具备统计上的显著优越性, 也符合空气密度受温度、湿度、海拔等多因素交织影响形成的复合分布特点, 可为后续风速 - 空气密度联合概率建模、风功率密度估算提供可靠的边缘分布基础。

4. 构建风速和空气密度的二元分布

在完成风速与空气密度各自边缘分布的参数估计与检验后, 为进一步刻画二者在实际风能评估中的耦合关系, 需构建其联合概率分布模型。本章节将分别给出风速与空气密度的边缘分布函数, 并基于 Copula 理论建立二者的二元联合分布, 以描述两变量间的统计相依结构。具体建模过程如下:

4.1. 风速的边缘概率密度函数

风速随机变量服从广义极值分布, 其累积分布函数为:

$$F(v; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{v - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (7)$$

定义域满足: $1 + \xi(v - \mu)/\sigma > 0$, 位置参数 $\mu \in R$, 反映风速的基准水平, 尺度参数 $\sigma > 0$, 反映风速的离散程度, 形状参数 $\xi \in R$ 是核心参数, 决定分布的尾部类型。

对应的概率密度函数通过分布函数求导可得:

$$f(v; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{v - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{v - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \quad (8)$$

形状参数 ξ 提供了风速极端值的统计推断基础。

4.2. 空气密度的边缘概率密度函数

空气密度随机变量服从混合正态分布, 其累计分布函数是各个分量分布函数的加权和:

$$F(\rho; \theta) = \sum_{k=1}^K \pi_k \cdot \Phi(\rho; \mu_k, \sigma_k^2) \quad (9)$$

其中 $\Phi(x; \mu_k, \sigma_k^2)$ 是第 k 个正态分布的累积分布函数。

$$\Phi(\rho; \mu_k, \sigma_k^2) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\rho - \mu_k}{\sigma_k \sqrt{2}} \right) \right] \quad (10)$$

这里的 $\operatorname{erf}(\cdot)$ 是误差函数。

概率密度函数为:

$$f(\rho; \theta) = \sum_{k=1}^K \pi_k \cdot \phi(\rho; \mu_k, \sigma_k^2) \quad (11)$$

其中 $\sum_{k=1}^K \pi_k = 1$, $\pi_k \geq 0$, $\forall k$ 。

每个分量都是一个正态分布:

$$\phi(\rho; \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left(-\frac{(\rho - \mu_k)^2}{2\sigma_k^2} \right) \quad (12)$$

4.3. 风速空气密度联合概率分布建模

在完成风速与空气密度的边缘分布构建后, 采用 Copula 函数刻画二者相依结构, 其核心步骤为通过概率积分变换将原始变量转化为边缘分布服从均匀分布的新变量以构建联合分布, 首先对两个原始变量分别进行概率积分变换, 将其转化为服从 $[0, 1]$ 均匀分布的随机变量, 这是 Copula 理论的基础。

风速变量服从广义极值分布, 其累积分布函数为 $F_V(v; \mu_v, \sigma_v, \xi_v)$, 定义变换后的均匀变量为:

$$U = F_V(V) \sim U[0, 1] \quad (13)$$

空气密度变量服从混合正态分布, 其累积分布函数为 $G_\rho(\rho; \theta_\rho)$, 其中 $\theta_\rho = \{(\omega_k, \mu_k, \sigma_k^2)\}_{k=1}^k$ 为混合模型的全部参数。定义变换后的均匀变量为:

$$W = G_\rho(\rho) \sim U[0, 1] \quad (14)$$

本研究选用 Frank Copula 来描述 U 与 W 之间的统计相依关系。其二元累积分布函数定义为:

$$C(u, w; \theta) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta w} - 1)}{e^{-\theta} - 1} \right] \quad (15)$$

相应的 Copula 概率密度函数通过对分布函数求混合偏导得到:

$$c(u, w; \theta) = \frac{-\theta(e^{-\theta} - 1)e^{-\theta(u+w)}}{\left[(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta w} - 1) \right]^2} \quad (16)$$

结合边缘分布与 Copula 函数, 风速 V 与空气密度 ρ 的联合累积分布函数可表示为:

$$H(v, \rho) = C(F_V(v), G_\rho(\rho); \theta) \quad (17)$$

对应的联合概率密度函数则为:

$$h(v, \rho) = f_v(v; \mu_v, \sigma_v, \xi_v) \cdot g_\rho(\rho; \theta_\rho) \cdot c(F_v(v), G_\rho(\rho); \theta) \quad (18)$$

其中, $f_v(\cdot)$, $g_\rho(\cdot)$ 分别为风速广义极值分布与空气密度混合正态分布的概率密度函数。

4.4. 风速空气密度联合概率分布对比分析

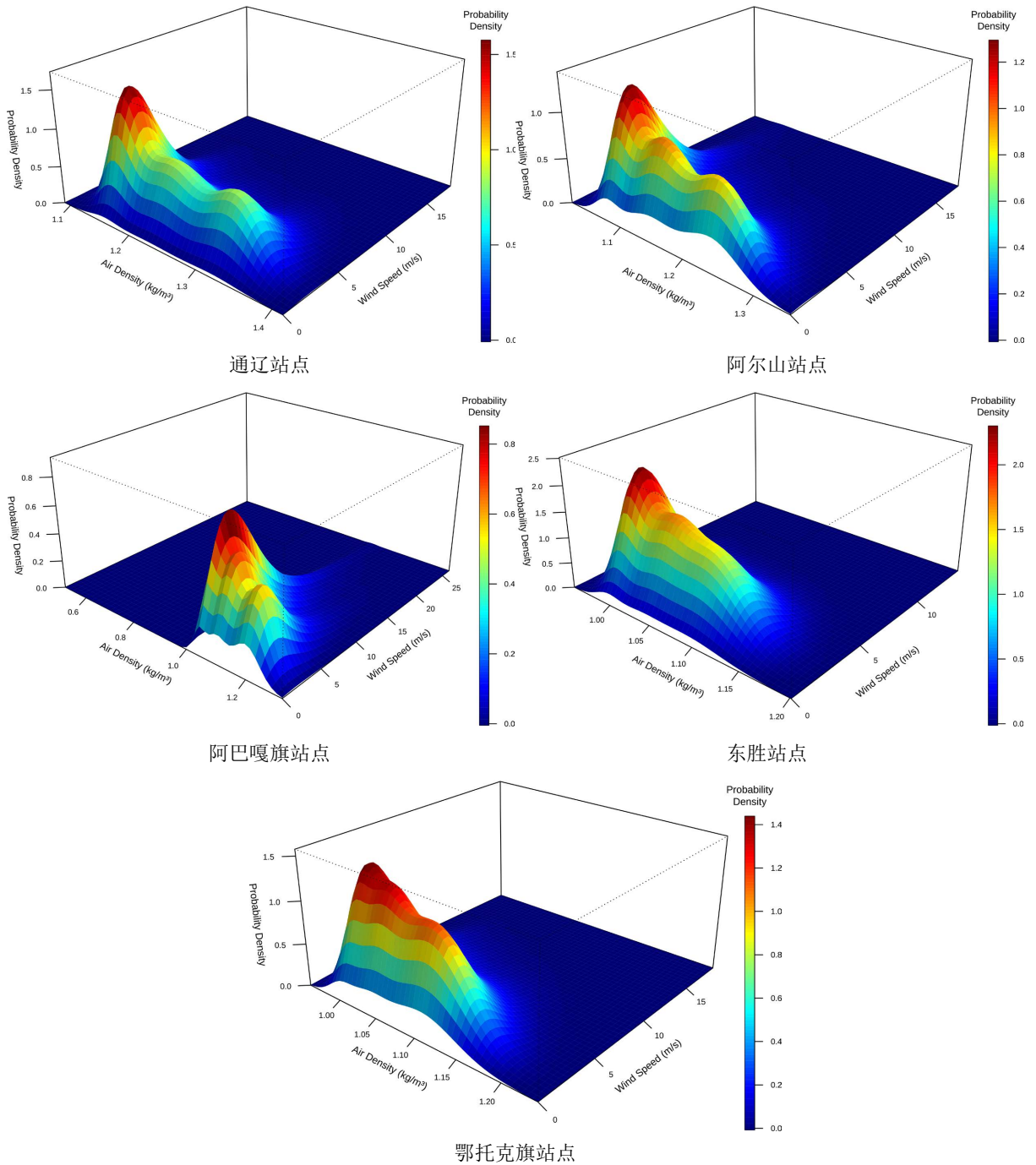


Figure 5. Joint distribution of wind speed and air density at each station and its fitting to empirical data

图 5. 各站点风速和空气密度的联合分布对经验数据的拟合

Table 6. Copula parameter values and model evaluation indicators
表 6. Copula 参数值和模型评价指标

地区	参数	AIC	RMSE	IA
通辽	-0.4521	-82.4773	0.0230	0.9971
阿尔山	-2.0988	-1831.2670	0.0307	0.9930
阿巴嘎旗	-1.5889	-1110.7070	0.0210	0.9970
东胜	-0.6008	-136.2314	0.0556	0.9841
鄂托克旗	-1.3678	-740.4542	0.0399	0.9907

根据对五个站点风速与空气密度联合分布的 Copula 拟合分析(见图 5), 尽管各站点最优连接函数存在差异, 但综合拟合优度与理论适配性, 如表 6 所示, 推荐采用 Frank Copula 作为统一的连接函数。主要依据在于: 首先 Frank Copula 在多数站点的 RMSE 与 IA 指标上表现最优, 且在其余站点与最优模型的差距微小, 展现出稳健的跨站点拟合能力, 其次其理论特性与数据特征高度匹配所有站点两变量均呈弱至中度负相关, Frank Copula 的参数范围可灵活刻画负相关, 且具有对称性与无尾部相依的特点, 恰好符合风速与空气密度的特征, 最后各站点 Frank Copula 的参数估计均为负值且变异较小, 保证了模型解释的一致性与稳定性。因此, 选用 Frank Copula 能够在兼顾多站点一致性与模型解释力的基础上, 实现对风速 - 空气密度相依结构的可靠描述。

5. 风功率密度

为了精准评估区域风能资源的开发价值、支撑风电厂的科学选址与效益预测, 合理的风能指标是实践中的核心依据。空气密度与风速是风电厂选址评估的关键要素, 探究两者的联合概率密度函数, 能更精准地刻画区域实际风力资源的分布特征, 为风电机组选型与布局提供针对性参考。风功率密度是风电厂选址的核心评估指标, 具体公式如下:

$$\text{WPD} = \frac{1}{2} \rho_{hub} v_{hub}^3 \quad (19)$$

基于 ρ 和 V 的联合概率密度函数 $f_{\rho V}(\rho_{hub}, v_{hub})$ 并结合上述公式, WPD 如下:

$$\text{WPD} = \frac{1}{2} \int_0^{\rho_{max}} \int_{\rho_{min}}^{\rho_{max}} \rho_{hub} v_{hub}^3 f_{\rho V}(\rho_{hub}, v_{hub}) dv_{hub} d\rho_{hub} \quad (20)$$

ρ_{min} 和 ρ_{max} 分别是所评估位置的全年或季节性数据的空气密度最大值和最小值。由公式(20)计算的各站点 50 m 高度年平均风功率密度见表 7。

Table 7. Annual average wind power density at 50 m height for each station
表 7. 各站点 50 m 高度年平均风功率密度

地区	风功率密度(W/m ²)
通辽	84.1290
阿尔山	83.0531
阿巴嘎旗	273.9724
东胜	55.0047
鄂托克旗	65.2262

根据国标 GB/T 18710-2002 [24], 风电场风资源评估方法在 50 m 高度下的风功率密度和年平均风速进行了具体规定, 如表 8 所示。

Table 8. Wind power density grade table

表 8. 风功率密度等级表

风功率密度等级	50 米高度风功率密度(W/m ²)	50 米高度年平均风速(m/s)	风力适宜度
1	<200	5.6	不适宜
2	200~300	6.4	不适宜
3	300~400	7.0	适宜
4	400~500	7.5	适宜
5	500~600	8.0	很适宜
6	600~800	8.8	很适宜
7	800~2000	11.9	很适宜

结合国标 GB/T 18710-2002 中 50 m 高度的风功率密度分级标准来看: 阿巴嘎旗年风功率密度达 273.9724 W/m², 处于 50 m 高度“200~300 W/m²”区间, 虽未达到国标“适宜”开发的等级 3, 但已接近临界值, 通辽、阿尔山、东胜、鄂托克旗处于“<200 W/m²”区间, 处于等级 1 区间。

风功率密度是风电场建设的核心决策指标: 它决定风能理论可开发量, 是评估发电量、投资回报的基础; 同时影响风机选型、机组布局, 直接关系项目经济可行性 - 风功率密度越高, 发电效益越优、投资回收期越短。而计算结果与国标分级的差距, 实则反映了区域实际资源与通用标准的适配性差异: 阿巴嘎旗地处内蒙古高原, 地形开阔、下垫面粗糙度低, 空气流动性强, 实际大风持续时间长, 虽年平均风功率密度未达国标“适宜”级, 但长期稳定的风况仍具备开发价值, 通辽、阿尔山受局地气候特征影响, 季节风资源较丰富, 可通过“分季节调度适配机型”提升利用效率; 东胜、鄂托克旗虽风功率密度偏低, 但结合低风速专用风机、优化机组布局等技术手段, 也可实现风能的有效开发。

因此, 阿巴嘎旗可优先规划布局风电场, 通辽、阿尔山可结合季节风特征针对性开发, 东胜、鄂托克旗则适配低风速机型推进项目落地, 实现区域风能资源的差异化利用。

6. 基于 BP 神经网络分析风电场选址模型的构建

BP 神经网络模型克服了传统统计模型对变量线性关系假设的依赖, 可自动学习输入特征与风功率密度间的非线性关联。对于多站点的风能资源数据, BP 神经网络能够从风矢量 U 分量、风矢量 V 分量、气压、空气密度、海拔等多维度输入中, 拟合出变量间的复杂相依性, 从而更精准地捕捉风功率密度的动态变化规律。结合内蒙古 5 个训练站点的历史气象数据集, 采用 BP 神经网络搭建预测模型, 实现对新站点风能资源的快速评估。

6.1. BP 神经网络

BP (back propagation) 神经网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家提出的概念, 是一种由数据信号正向传播以及数据误差信号反向传播而形成多层前反馈的神经网络, 有较强的适应学习能力[25]。可提取特征与目标的复杂关联, 不断优化特征映射关系, 最终实现对目标变量的精准预测。

6.2. 激活函数、损失函数

在人工神经网络的结构设计中, 激活函数和损失函数是除神经元、网络层级之外的关键组成部分,

对模型性能具有重要影响。在神经网络的每个神经元中, 激活函数通过对输入信号进行非线性变换, 将处理结果传递至下一层神经元, 为网络引入非线性映射能力, 从而使其能够学习和表示复杂的函数关系。本文隐藏层采用 sigmoid 激活函数, 具体数学表达式为:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (21)$$

监督学习中损失函数用于度量模型预测结果与实际观测值的差异程度且为模型优化的目标函数, 本文针对风功率密度预测这一回归问题采用均方误差作为损失函数, 通过计算预测值与实际值的误差平方和评估模型拟合效果, 计算公式如下:

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (22)$$

其中, y_i 为真实风功率密度值, \hat{y}_i 为模型预测值, n 为样本数量。在训练过程中通过最小化该损失函数模型参数向减小预测误差的方向调整。

6.3. 网络模型的构建

为实现新站点风功率密度的精准预测, 本文搭建单隐藏层 BP 神经网络模型, 通过层级特征映射拟合输入变量与风功率密度的非线性关系, 模型包含输入层、隐藏层、输出层三个层级, 输入层接收风矢量 U 分量、风矢量 V 分量、气压、空气密度、海拔 5 个特征变量, 隐藏层设 10 个神经元并采用 Sigmoid 激活函数对输入特征进行非线性转换以提取复杂关联特征, 输出层设 1 个神经元并采用 Linout 线性激活函数直接输出连续风功率密度预测值以适配回归任务数值输出需求; 模型训练阶段采用均方误差作为损失函数、梯度下降法作为优化器, 设置学习率 0.01、总迭代次数 5000 次, 通过“前向传播计算预测值、反向传播更新权重”的迭代过程最小化预测误差, 最终实现模型收敛与优化。

6.4. 模型训练与结果分析

模型在训练集与测试集的损失变化趋势如图 6、图 7 所示, BP 神经网络模型经 5000 次迭代训练后, 训练集与测试集的损失函数均呈快速下降趋势, 初始阶段损失值从约 4500 骤降至 100 以下, 后续逐渐收敛并稳定在极小区间内, 表明模型已实现有效学习且达到稳定状态。

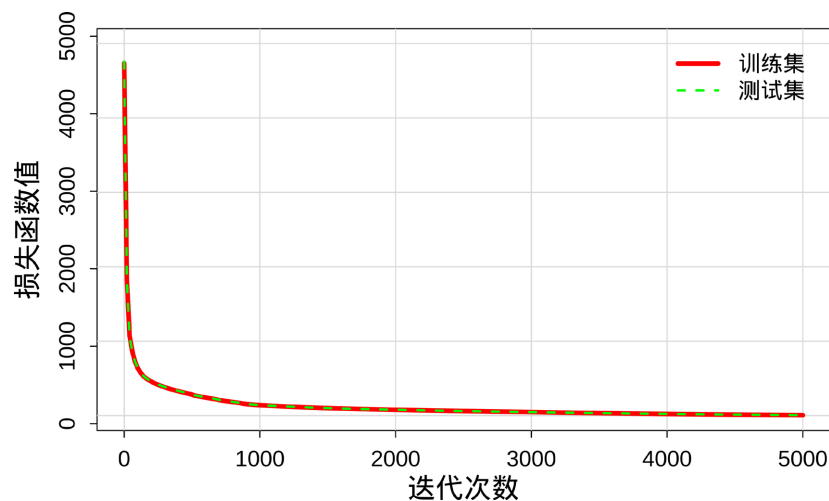


Figure 6. Trend chart of loss function

图 6. 损失函数变化趋势图

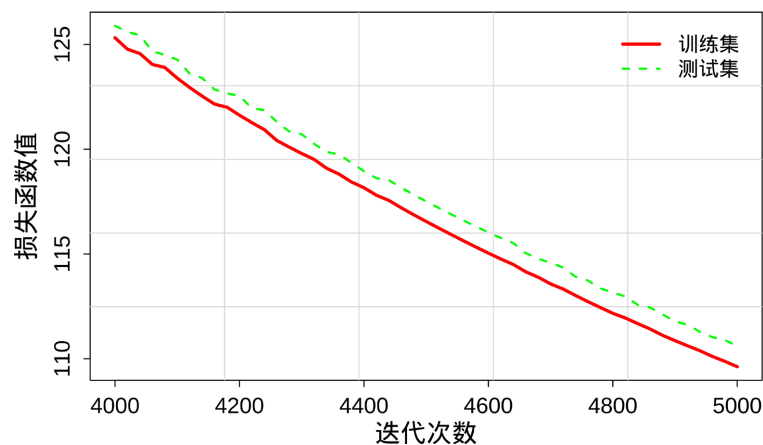


Figure 7. Local magnification of the trend of the loss function

图 7. 损失函数变化趋势局部放大图

从训练后期的损失函数局部变化可见, 训练集与测试集的损失曲线同步平缓下降且差距极小, 未出现训练集损失持续降低、测试集损失上升的过拟合现象, 进一步验证了模型的泛化能力。

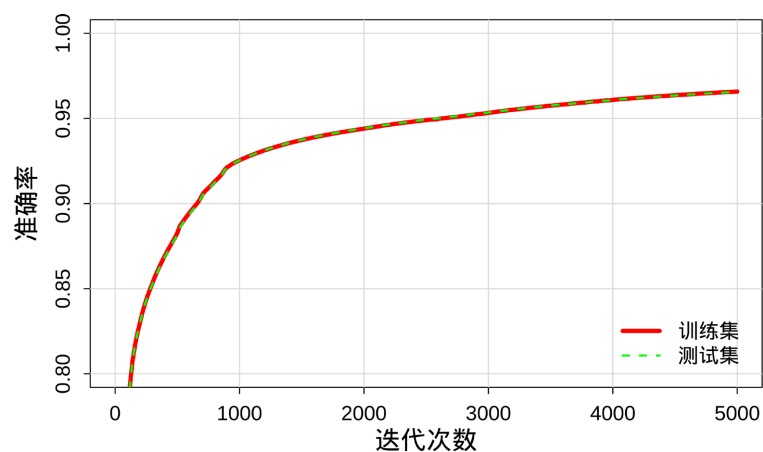


Figure 8. Trend chart of accuracy rate

图 8. 准确率变化趋势图

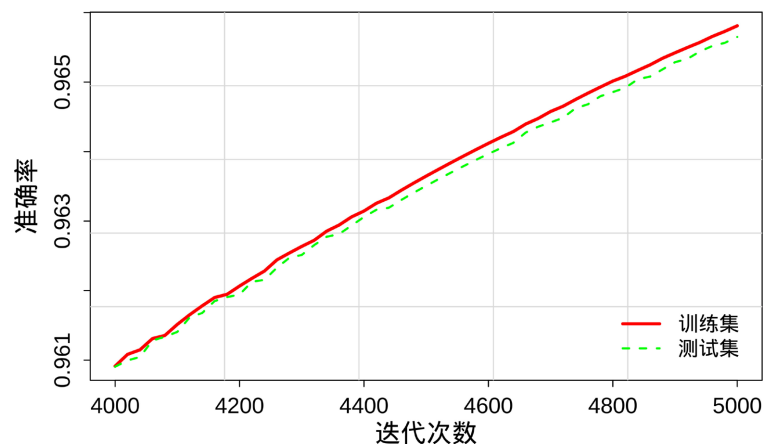


Figure 9. Partial enlarged view of the accuracy change trend

图 9. 准确率变化趋势局部放大图

为评估模型的泛化能力, 在模型训练完成后使用独立的测试集进行验证。测试集上的准确率变化趋势与训练集高度一致: 经过 5000 次迭代后, 测试集准确率持续上升并最终收敛至 0.96 以上, 与训练集准确率基本吻合, 且全程未出现明显偏离(图 8)。从训练后期的局部放大图(图 9)可见, 两条准确率曲线始终紧密贴合, 差距维持在极小范围内。以上结果表明, 模型在训练过程中未出现过拟合, 具有良好的泛化能力, 能够适应新数据的预测任务。

7. 实例仿真分析

7.1. 选址概括分析

综合地理分布、海拔特征、气候相似性及数据质量, 推荐锡林浩特站作为预测验证站点。该站地处训练站点地理中心, 海拔约 1100 米, 与训练站点平均海拔接近, 可降低模型外推误差; 其温带大陆性草原气候与训练站点风资源、气象要素特征高度相似, 利于检验模型适应性; 同时作为国家基准气象站, 该站观测数据长期连续、质量可靠, 为验证提供坚实支撑。选择该站验证, 既能评估模型空间性能, 也可作为当地风能资源评估提供参考, 兼具方法验证与实际应用价值。

7.2. 基于 BP 神经网络模型的锡林浩特站点风功率密度预测分析

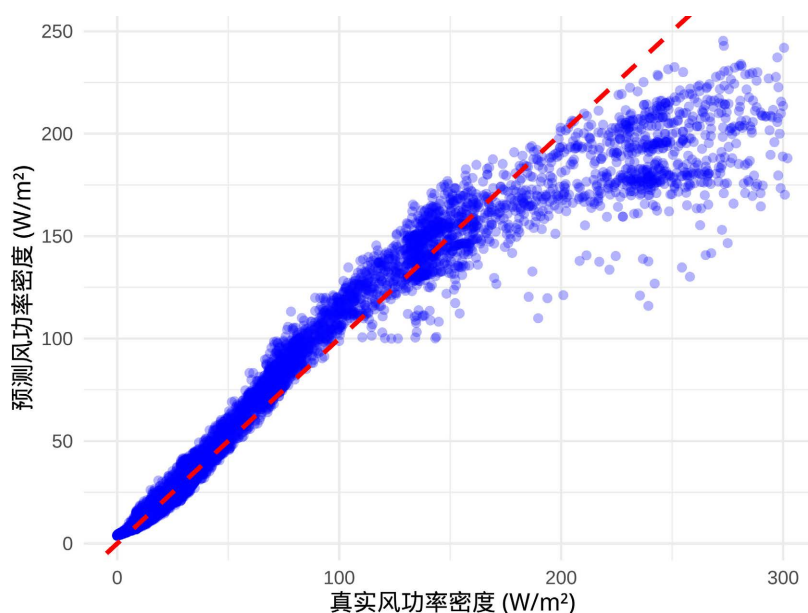


Figure 10. Comparison chart of predicted values and actual values at Xilinhot station
图 10. 锡林浩特站点预测值与真实值对比图

将锡林浩特 2020~2024 年 10050 组有效气象数据输入训练好的 BP 神经网络模型进行风功率密度预测。结果显示(见图 10), 该模型预测的锡林浩特站风功率密度均值为 49.75 W/m^2 , 范围 $3.84\sim 245.3 \text{ W/m}^2$, 符合该区域风能资源的实际分布特征。

从预测性能来看, 模型在锡林浩特站的 $R^2 = 0.9402$, $RMSE = 16.4 \text{ W/m}^2$, 结合图中“预测值 - 真实值”散点图可见, 数据点紧密围绕拟合线分布, 仅在高风功率密度区域存在少量随机离散点, 说明模型对锡林浩特站的风功率密度预测精度较高。结合实际风电开发标准, 该模型预测结果显示锡林浩特站多数时段的风功率密度满足开发需求, 同时锡林浩特当前已建成多家风电场, 与本模型的预测结论一致, 可见搭建的 BP 神经网络模型在新站点风能资源评估任务中具备良好的实用价值。

8. 结论

本研究以内蒙古多站点气象数据为基础, 根据不同下垫面类型选取对应风切变指数, 将风速、空气密度外推至统一轮毂高度, 并系统分析风向、风速及空气密度的统计特征, 进而采用 Copula 函数构建风速与空气密度的联合概率模型, 以量化二者的相依结构, 最后采用 BP 神经网络模型对锡林浩特站点风功率密度进行预测。主要研究结论如下:

(1) 在风能评估模型中, 风速与空气密度分别选取了广义极值分布和混合正态分布作为其理论边缘概率分布。研究验证了广义极值分布能较好地拟合风速的极值分布特征, 尤其在高风速区域具有显著优势; 混合正态分布则能有效刻画空气密度多峰、非对称的复杂分布形态, 为后续联合建模奠定了稳健的边缘分布基础。

(2) 为捕捉风速与空气密度之间的相依结构, 选用 Frank Copula 函数建立二者的二元联合概率分布。该函数在各项拟合优度指标上表现稳定, 能够有效刻画两者之间的相关关系, 从而构建出更为符合实际情况的联合概率模型。

(3) 采用 BP 神经网络构建了风功率密度预测模型, 模型以风矢量 U 分量、风矢量 V 分量、气压、空气密度、海拔为输入, 实现了对锡林浩特站点风功率密度的预测, 预测结果与实际值吻合良好($R^2 = 0.9402$, $RMSE = 16.4 \text{ W/m}^2$), 证明模型具有较强的泛化能力与预测精度。该模型可为缺乏长期观测数据或地形复杂的区域提供一种高效、可靠的风能资源评估工具, 为风电场前期选址决策提供直接的量化依据, 但在实际工程应用中需结合地形、地貌、电网接入等多因素进行综合论证。

基金项目

(1) 一流学科科研专项项目资助(编号: YLXKZX-NGD-027); (2) 内蒙古自治区自然科学基金项目(编号: 2025QN01027)。

参考文献

- [1] 《“十四五”现代能源体系规划》发布[J]. 风能, 2022(4): 8.
- [2] 黄林宏, 宋丽莉, 周荣卫, 等. 大型风电基地风电波动特征分析[J]. 中国电机工程学报, 2017, 37(6): 1599-1609.
- [3] Zhang, J., Chowdhury, S., Messac, A. and Hodge, B. (2014) A Hybrid Measure-Correlate-Predict Method for Long-Term Wind Condition Assessment. *Energy Conversion and Management*, **87**, 697-710. <https://doi.org/10.1016/j.enconman.2014.07.057>
- [4] Omrani, H., Drobinski, P., Arsouze, T., Bastin, S., Lebeaupin-Brossier, C. and Mailler, S. (2017) Spatial and Temporal Variability of Wind Energy Resource and Production over the North Western Mediterranean Sea: Sensitivity to Air-Sea Interactions. *Renewable Energy*, **101**, 680-689. <https://doi.org/10.1016/j.renene.2016.09.028>
- [5] 周积强, 魏巍, 姚肖萌, 等. 基于 Python 的三维风场风玫瑰图绘制方法研究与应用[J]. 气象水文海洋仪器, 2022, 39(3): 83-86.
- [6] 黄武枫. 风电场风速概率分布及其拟合模型研究[D]: [硕士学位论文]. 南宁: 广西大学, 2021.
- [7] Wais, P. (2017) A Review of Weibull Functions in Wind Sector. *Renewable and Sustainable Energy Reviews*, **70**, 1099-1107. <https://doi.org/10.1016/j.rser.2016.12.014>
- [8] 金国骅, 胡文忠. 风速频率分布混合模型的研究[J]. 太阳能学报, 1994, 15(4): 353-357.
- [9] Xiang, C., Chen, A. Li, Q. and Ma, R. (2021) Research on the Probability Model of Basic Wind Speed Estimation in China. *Wind Structure*, **32**, 587-596.
- [10] 杨茂, 董骏城. 基于混合分布模型的风电功率波动特性研究[J]. 中国电机工程学报, 2016, 36(S1): 69-78.
- [11] 胡亚男, 徐丽娜, 赵迦琪, 韩丽娜. 基于不同风切变指数算法推算风机轮毂高度风速[J]. 内蒙古气象, 2023(6): 14-18.
- [12] 龚玺, 朱蓉, 李泽椿. 我国不同下垫面的近地层风切变指数研究[J]. 气象, 2018, 44(9): 1160-1168.

-
- [13] Jung, C. and Schindler, D. (2019) The Role of Air Density in Wind Energy Assessment—A Case Study from Germany. *Energy*, **171**, 385-392. <https://doi.org/10.1016/j.energy.2019.01.041>
- [14] Han, Q. and Chu, F. (2021) Directional Wind Energy Assessment of China Based on Nonparametric Copula Models. *Renewable Energy*, **164**, 1334-1349. <https://doi.org/10.1016/j.renene.2020.10.149>
- [15] 王必强, 陈正洪, 孙朋杰, 等. 湖北省风电场大风速段风资源特征分析[J]. 可再生能源, 2022, 40(6): 792-797.
- [16] Carta, J.A., Ramírez, P. and Bueno, C. (2008) A Joint Probability Density Function of Wind Speed and Direction for Wind Energy Analysis. *Energy Conversion and Management*, **49**, 1309-1320. <https://doi.org/10.1016/j.enconman.2008.01.010>
- [17] 张广隶. 考虑风速风向联合概率分布的风电塔筒风致疲劳寿命评估[D]: [硕士学位论文]. 兰州: 兰州理工大学, 2020.
- [18] 张哲, 王勃. 基于 CBAM-LSTM 的风电集群功率短期预测方法[J]. 东北电力大学学报, 2024, 44(1): 1-8.
- [19] Wang, C., Zhao, Q. and Tian, R. (2023) Short-Term Wind Power Prediction Based on a Hybrid Markov-Based PSO-BP Neural Network. *Energies*, **16**, Article 4282. <https://doi.org/10.3390/en16114282>
- [20] Zhang, Y., Li, P., Li, H., Zu, W. and Zhang, H. (2023) Short-Term Power Prediction of Wind Power Generation System Based on Logistic Chaos Atom Search Optimization BP Neural Network. *International Transactions on Electrical Energy Systems*, **2023**, Article ID: 6328119. <https://doi.org/10.1155/2023/6328119>
- [21] 王婷婷, 李斯胜, 于伟, 能锋田, 李星南, 杨佳琳, 熊亮. 基于 BP 神经网络结合 ERA5 数据的风电功率预测[J]. 储能科学与技术, 2025, 14(1): 183-189.
- [22] Mandal, G., Kumar, A., Sharma, D.C. and Kumar, H. (2013) Comparative Analysis of Different Air Density Equations. *MAPAN*, **28**, 51-62. <https://doi.org/10.1007/s12647-013-0044-2>
- [23] Yang, X., Jiang, X., Liang, S., Qin, Y., Ye, F., Ye, B., *et al.* (2024) Spatiotemporal Variation of Power Law Exponent on the Use of Wind Energy. *Applied Energy*, **356**, Article 122441. <https://doi.org/10.1016/j.apenergy.2023.122441>
- [24] 中华人民共和国国家质量监督检验检疫总局. 风电场风能资源评估方法: GB/T 18710-2002 [S]. 北京: 中国标准出版社, 2004.
- [25] 黄凤新, 刘寿东, 祝赢, 等. 基于滚动极值处理的 BP 神经网络方法的 WRF 模式预报风速订正[J]. 科学技术与工程, 2013, 13(7): 1768-1772.