

# 计及风向信息的风电功率异常数据识别研究

杨 茂<sup>1</sup>, 杨春霖<sup>1</sup>, 杨琼琼<sup>1</sup>, 苏 欣<sup>2</sup>

(1. 现代电力系统仿真控制与绿色电能新技术吉林省重点实验室(东北电力大学), 吉林 132012;

2. 东北电力大学理学院, 吉林 132012)

**摘 要:** 针对目前风电功率异常数据识别的算法复杂繁琐且识别效果不理想, 算法通用性差, 对复杂多变的风电功率数据难以有效甄别等问题, 依据实测风速功率数据中异常数据来源的特征提出一种不同风向上的不同风速区间内异常数据的识别方法。该算法基于不同风向上不同风速区间的组内最优方差来识别其相应的异常数据, 最后经归类整理后识别出机组全部的异常数据。研究结果表明该算法可有效识别风电异常数据。

**关键词:** 异常数据; 风向; 数据识别; 组内最优方差

**中图分类号:** TM71

**文献标识码:** A

## 0 引 言

近年来随着风电市场的不断扩大及大数据分析技术的日趋成熟, 风电运行数据的价值得到极大重视, 风电运行数据的研究已成为该领域研究的热门课题<sup>[1-3]</sup>。

风电运行数据中风速和功率数据研究价值最高。通过实测的某台机组的风速和功率数据可得到该机组的风速功率曲线, 通过对该曲线的分析可为机组的设计、选址、评估发电能力、估算损失电量等一系列技术问题提供重要的参考。然而机组的实测风速功率曲线分布错乱无章并存在大量异常数据, 这些异常数据主要来自于 2 个方面: 1) 在数据采集和测量等环节出现故障或干扰, 导致数据的异常、缺失; 2) 因风电接纳能力有限而造成的弃风限电所形成的数据异常。使得实测风速功率曲线不能正确反映机组的实际运行状态<sup>[4-9]</sup>。故对由数据采集和监控(supervisory control and data acquisition, SCADA)系统采集到的原始风速和功率进行异常数据识别显得十分必要。文献[10]在数据预处理阶段借助数据范围的检查及一致性检验识别异常数据; 文献[11]通过参数化公式表示的条件准则对风速功率曲线中的异常数据进行筛选和去除; 文献[12]提出一种对异常数据处理的组合筛

选模型, 采用两次四分位法对分散型异常数据去除, 再对堆积型异常数据使用聚类法去除, 并采用二次聚类的思想来解决  $k$ -means 聚类中  $k$  的取值问题。文献[13]根据历史数据分析异常数据产生的原因, 针对这些原因提出一种基于四分位法的异常数据识别法, 算法还可实现采用临近出力模式下的数据对严重缺失数据进行补齐并重构出完整的时间序列。

上述文献中存在数据识别不彻底、识别过程繁琐、识别条件苛刻和通用性不强等问题。基于此, 本文以风速功率曲线为研究基础, 提出一种计及风向的组内最优方差算法, 该算法识别异常数据的原理过程简单, 无需使用历史数据集进行训练和学习, 具有较强的通用性, 可对异常数据分布类型不同机组进行有效识别。

## 1 风速功率曲线与风玫瑰图

### 1.1 风速功率曲线

空气流动所携带的动能称之为风能。风电机组发电的基本原理是将由叶片捕获的风能转换为机械能, 再将机械能转换为电能。假设当风速为  $V$  时, 风力机叶片单位时间内捕获的风能  $E$  理论值为:

收稿日期: 2017-05-15

基金项目: 国家重点研发计划(2018YFB0904200)

通信作者: 杨 茂(1982—), 男, 博士、教授, 主要从事风力发电技术方面的研究。yangmao820@163.com

$$E = 0.5\rho_0 AV^3 \quad (1)$$

式中,  $\rho_0$  ——空气密度,  $\text{kg/m}^3$ ;  $A$  ——叶片扫掠面积,  $\text{m}^2$ 。

在式(1)的风能条件下, 根据贝兹理论可获得的最大功率  $P_{\max}$  为:

$$P_{\max} = \frac{8\rho_0 AV^3}{27} \quad (2)$$

由式(1)、式(2)可得风电机组的最大理论风能利用系数  $C_p$  [4] 为:

$$C_p = \frac{P_{\max}}{E} = \frac{2P_{\max}}{\rho_0 AV^3} \approx 0.593 \quad (3)$$

对于实际风电机组, 由于风电机组叶片结构不满足理想条件, 且传动系统和发电机等能量转换环节都存在损失, 因此实际风能利用系数低于理论风能利用系数, 即  $C_p \leq 0.593$ 。

风电机组实际能得到的有用功率输出可表示为:

$$P = 0.5\rho_0 AV^3 C_p \quad (4)$$

风电机组的风速功率曲线包括由厂家提供的标准风速功率曲线和实测风速功率曲线。根据 IEC 61400-12 标准可获得机组的标准风速功率曲线, 该曲线表征的是风速与功率数据 10 min 平均值之间的对应函数关系[15]。机组的实测风速功率曲线是根据实测的风速和功率数据绘制的散点图, 简称风速功率曲线。在理想条件下, 风速功率曲线如图 1 所示。

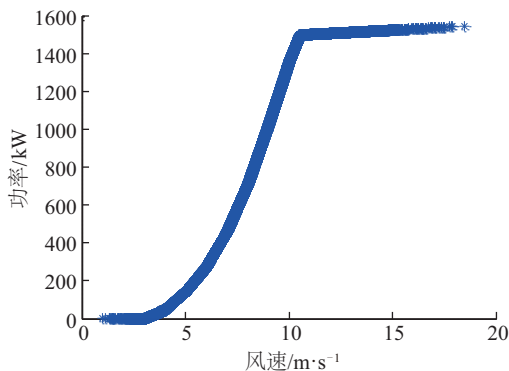


图1 理想状态下标准风速功率曲线

Fig. 1 Ideal wind power curve

但实际的风速功率曲线不可能完全与理想状态下的曲线保持一致, 而是会在一个较宽的范围内左右上下波动, 这是由于机组常运行在风速和风向不断变化的情况下且机组的实际运行环境与理想

状态下的设计环境存在很大差异造成的。故对实际运行状况下风电数据的研究意义重大。

## 1.2 风玫瑰图

风玫瑰图是根据某一地区多年平均统计的各个风向和风速的百分数值, 按一定比例绘制而成。将  $360^\circ$  方位按照每  $22.5^\circ$  一格划分成 16 格, 将实时采集的各个风向统计到 16 个方向上。玫瑰图上所表示的风向, 是指从外面吹向地区中心的方向[15]。风玫瑰图中必须包含风速、风向、风向频率 3 个要素的定量值。若要计算风向频率, 首先需要将一段时期内的风向按照 16 个方向进行统计, 各个方向的风向频率和平均风速(除静风以外)定义为:

$$g_N = f_N / \left( \sum_{N=1}^{15} f_N + C \right) \quad (5)$$

$$\bar{v}_N = \sum_{i=1}^{f_N} v_{Ni} / f_N \quad (6)$$

$$\bar{v}_C = \sum_{i=1}^C v_{Ci} / C \quad (7)$$

式中,  $g_N$  ——  $N$  方向上的风向频率,  $N=1, 2, \dots, 16$ ;  $\bar{v}_N$  ——  $N$  方向上的平均风速;  $f_N$  ——在所统计时间段内,  $N$  方向风观测到的次数;  $v_{Ni}$  ——在所统计时间段内在  $n$  方向上观测到的第  $i$  次风的风速;  $\bar{v}_C$  ——在所统计时间段内静风平均风速;  $C$  ——所统计时间段内观测到的静风次数;  $v_{Ci}$  ——在所统计时间段内, 所观测到的第  $i$  次静风的风速。

## 2 含有风向的最优组内方差算法

含有风向信息的最优组内方差算法的原理是通过不同风向不同风速段内的滑动方差来识别机组发电性能偏高或偏低的数据集, 分别作为上部异常数据和下部异常数据, 其余数据为正常数据。首先通过绘制风玫瑰图来统计所要研究风力机在不同风向范围内所捕获的风能的大小(通过风速大小来体现); 然后对每个风向范围内的风速进行统计, 得出该风向范围内的风速波动范围; 其次, 在该风速波动范围内再将波动范围细化, 在细化出的风速小区间中将不同风速对应的功率进行排序, 将该区间内排序后的功率序列从第一个功率值开始滑动求取方差, 最后将所求方差与阈值对比, 从而确定异常数据的位置。

在已知风速、功率数据样本的集合  $Q$ , 可设

$W=\{W_1, W_2, \dots, W_n\}$  为目标集合,  $W_n$  为正常数据集,  $W_1$  为输出功率偏低的数据集,  $W_2$  为输出功率偏高的数据集, 且满足  $W_1 \cap W_2 \cap \dots \cap W_n = \emptyset$ ,  $W_1 \cap W_2 \cap \dots \cap W_n = E$ 。

该算法就是对  $Q$  进行异常数据识别, 其主要步骤如下:

步骤 1: 对含有风向、风速和功率的数据序列  $Q$  进行处理, 即通过绘制风玫瑰图统计出风能的分布范围, 按照每  $22.5^\circ$  为一个风向范围进行划分, 则共有  $D=\{D(1), D(2), \dots, D(16)\}$ , 共 16 个扇区。统计出所要研究风电机组运行数据的分布扇区  $D(i)$  ( $i=1, 2, \dots, 16$ )。

步骤 2: 设风电运行数据分布于  $D(i)$  扇区, 将该扇区再按每  $10^\circ$  一个小区间细化, 设细化后某一区间为  $[a, b]$ , 假设该区间内有  $E(i)=\{(s_1, p_1), (s_2, p_2), \dots, (s_k, p_k)\}$  共  $k$  对风速功率数据,  $s_i$  代表风速,  $p_i$  代表相应风速对应的功率 ( $i=1, 2, \dots, k$ )。

步骤 3: 在由步骤 2 得到的区间  $[a, b]$  中, 对风速区间进行细化, 即统计风向区间  $[a, b]$  内的风速大小的分布, 然后确定风速的合理范围, 假设为  $[s_x, s_y]$ , 其中  $s_x, s_y$  分别代表该范围内的某风速值。以  $T=0.5 \text{ m/s}$  为间隔时间再将该区间划分为  $n$  个小区间, 划分后的第 1 个区间为  $[s_{x1}, s_{y1}]$ 。将  $[s_{x1}, s_{y1}]$  区间内的功率值进行降序排列, 得到新的风电功率运行数据  $E'(i)=\{(s'_1, p'_1), (s'_2, p'_2), \dots, (s'_n, p'_n)\}$ ,

满足  $p'_i < p'_{i+1}$ 。

步骤 4: 计算最优组内方差。

$$V_i = \frac{1}{\lambda} \sum_{i=1}^{\lambda} (\bar{p}' - p'_i)^2 \quad (8)$$

式中,  $\lambda$  —— 向前滚动的次数,  $\lambda=1, 2, \dots, n$ ;  $\bar{p}'$  —— 从新风电功率数据  $p'_1$  到  $p'_\lambda$  的平均值;  $p'_i$  —— 第  $i$  个点的功率,  $i=1, 2, \dots, \lambda$ 。

步骤 5: 异常数据识别, 即将步骤 2 中得到的方差序列的方差值与最大阈值  $R$  及最小阈值  $r$  进行比较, 找出大于  $R$  或小于  $r$  的方差所对应的位置  $x_{\min}, x_{\max}$ , 大于  $x_{\max}$  的坐标点对应的风速、功率数据为下异常数据, 小于  $x_{\min}$  的坐标点对应的风速、功率数据为上异常数据, 其余数据为正常数据。由此便识别出了风速区间  $[s_{x1}, s_{y1}]$  内的全部异常数据。

步骤 6: 重复步骤 3, 识别出在风向  $[m_1, n_1]$  下全部风速范围  $[s_x, s_y]$  中的所有异常数据,  $m_1$  和  $n_1$  代表风向值, 且  $m_1 < n_1$ 。

步骤 7: 对除风向  $[m_1, n_1]$  外的所有风向范围, 重复步骤 1~步骤 4, 得出该机组的全部异常数据。

计及风向的组内最优方差算法的实现流程图如图 2 所示。在考虑风向的组内最优方差算法中各个方向各个风速区间的最大阈值与最小阈值是不同的, 需根据各区间的具体情况来确定阈值以确保算法实现效果最佳。

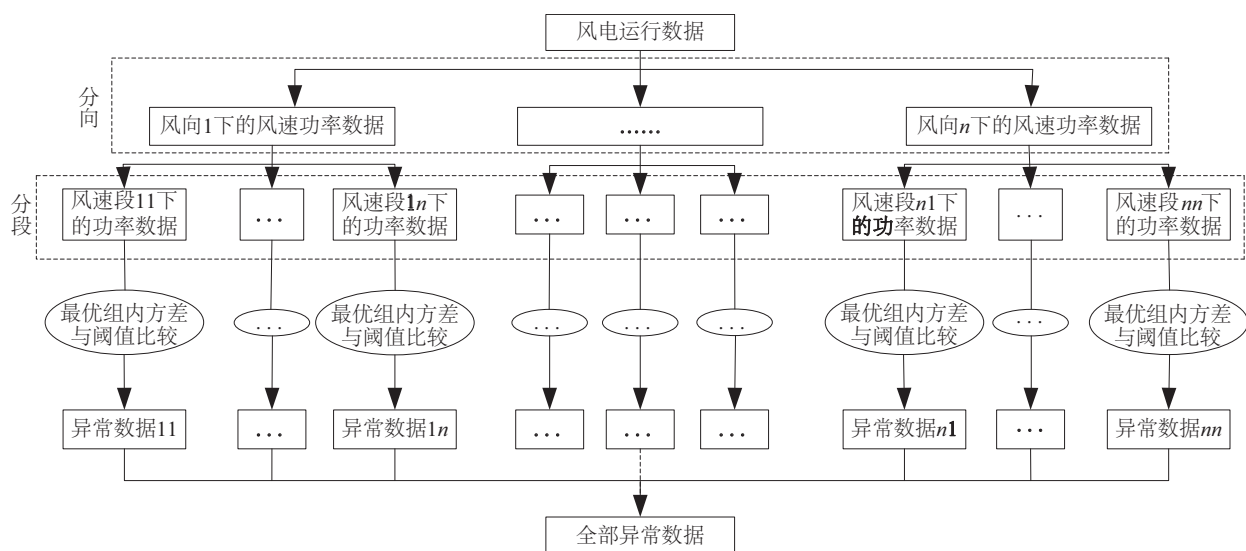


图2 计及风向的组内最优方差计算流程图

Fig. 2 Flow chart of most optimal variance algorithm include wind direction

### 3 算例分析

以中国东北地区某风电场 1 号和 18 号风力机运行数据为例,对计及风向的组内最优方差算法进行验证,机组的装机容量为 1500 kW,采样间隔为 15 min,基本运行数据包括风向、风速和功率。

1 号机组和 18 号机组的风速功率曲线如图 3 和图 4 所示。由图 3 可知,1 号机组的风速功率散点图更接近于理想状态想的风速功率散点图,异常数据点较少,风速功率数据集中一个较小的区域内。

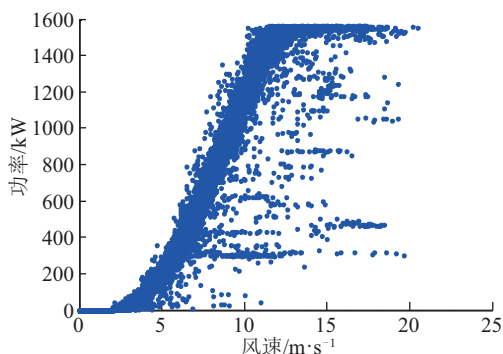


图3 1号风力机风速功率散点图

Fig. 3 Wind power scatter diagram of No.1 wind turbine

由图 4 可知,18 号机组的风速功率散点比较杂乱,风速功率数据在一个较大的范围内波动,且存在很多与理想状态的风速功率散点分布规律极其相似的点,这部分点没有分布在风力机生产厂商提供的标准曲线附近,而是略微滞后,这可能是由于出厂风力机不合格或者在运行中风力机受损所致,这些点也称为异常点。

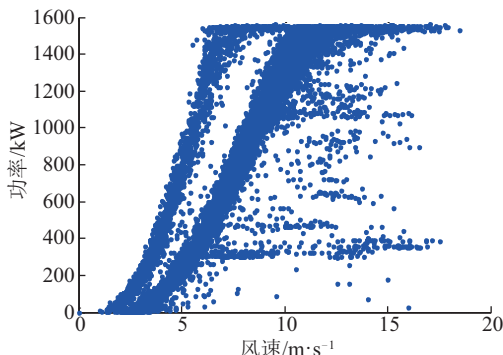


图4 18号风力机风速功率散点图

Fig. 4 Wind power scatter diagram of No.18 wind turbine

受文章篇幅所限,下文仅以 18 号风力机 2015 年

年 03 月 01 日至 2015 年 07 月 01 日这一时间段,采样间隔为 15 min 的风电运行数据共 11520 个点为例进行研究说明,图 5 为此时间段运行数据的风玫瑰图,由图可知风向主要分布在第 12 扇区和第 13 扇区之间(按顺时针方向依次为第 1 扇区,第 2 扇区, ..., 第 16 扇区),图 5 中  $W_s$  表示风速。

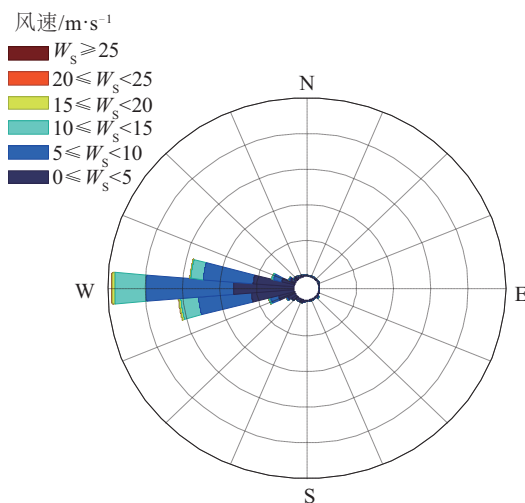


图5 18号风力机风玫瑰图

Fig. 5 Wind rose chart of No.18 wind turbine

经统计得到风向的分布概率密度如图 6 所示,可知风向的主要分布范围为  $[150^\circ, 200^\circ]$ ,与风玫瑰图所得范围相符,其累计概率分布如图 7 所示,可知风向范围在  $[150.0^\circ, 200.3^\circ]$  之间的点占总数的 98.32%,即有 11326 个。故为了缩小研究范围,降低计算的复杂度,取 18 号风力机的风向在  $[150^\circ, 200^\circ]$  之间的风电运行数据作为研究对象。在  $[150^\circ, 200^\circ]$  范围内,为进一步细化研究区间,对风向按每  $10^\circ$  进行一次划分,得到 5 个细化区间。每

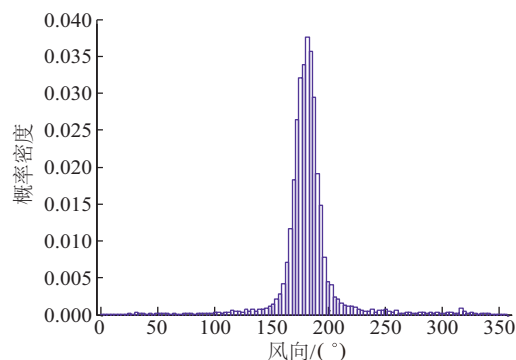
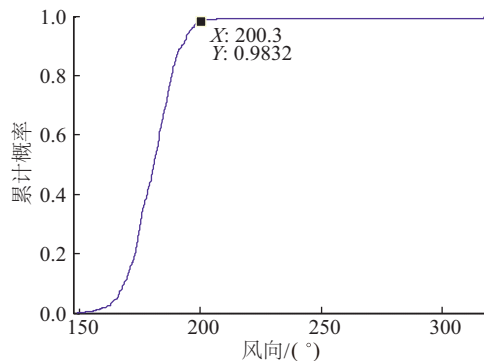


图6 18号风力机风向分布概率密度图

Fig. 6 Probability density distribution chart of wind direction of No.18 wind turbine



图7 风向 $[150^{\circ}, 200^{\circ}]$ 之间风向分布概率密度图Fig. 7 Wind direction distribution probability distribution between wind direction  $[150^{\circ}, 200^{\circ}]$ 

一个区间的内的风电运行数据量如表 1 所示,  $[150^{\circ}, 160^{\circ}]$  之间共有 476 个, 占总数据量的 4.25%。 $[160^{\circ}, 170^{\circ}]$  之间风电运行数据个数为 1425, 占总数的 12.58%。

表 1 各风向区间内风电数据的数量

Table 1 Data volume of each wind direction

区间	总数/个	风电运行数据/个	占总数百分比/%
$[150^{\circ}, 160^{\circ}]$		476	4.20
$[160^{\circ}, 170^{\circ}]$		1425	12.58
$[170^{\circ}, 180^{\circ}]$	11326	3676	32.46
$[180^{\circ}, 190^{\circ}]$		4209	37.16
$[190^{\circ}, 200^{\circ}]$		1540	13.60

$[170^{\circ}, 180^{\circ}]$  之间共有 3676 个, 达到总数的 32.46%。 $[180^{\circ}, 190^{\circ}]$  之间的数据量最多共有 4209 个, 占总数的 37.16%。 $[190^{\circ}, 200^{\circ}]$  之间共有 1540 个, 占总数的 13.60%。仅取风向在  $[170^{\circ}, 180^{\circ}]$  范围内的风电运行数据来说明计及风向的组内最优方差算法的具体实现步骤。经统计, 风向在  $[170^{\circ}, 180^{\circ}]$  范围内的风速累计概率分布如图 8 所示, 可知风速全部分布在  $[0, 20]$  m/s 的范围之内, 当风速为 18 m/s 时, 累计概率为 0.9988, 即意味着  $[0, 18]$  m/s 之间的风速占比为 99.88%, 该区间外的风速占比为 0.12%。所以取风向在  $[170^{\circ}, 180^{\circ}]$  范围内风速在  $[0, 18]$  m/s 之间的数据作为研究对象。

再将风速区间  $[0, 18]$  以 0.5 m/s 为间隔划分为 36 各区间, 以第 15 个区间, 即风速为  $[7.0, 7.5]$  m/s 之间的风速和功率数据为例。将在该区间内的功率值(该区间内共有 191 对风电功率运行数据)按从

大到小的顺序排列, 将其带入式(6)中可求出相应的最优方差, 求出的最优方差曲线如图 9 所示。图 9 中最优方差曲线出现了 2 个显著变化的点(斜率变化最大的点), 将这 2 个点对应的方差定义为阈值  $R$  与  $r$ , 将其余方差与该阈值进行比较, 找到大于  $R$  或小于  $r$  的方差所对应的风速功率数据。于是将阈值大于  $R$  的方差所对应的风速功率数据定义为上异常数据, 将阈值小于  $r$  的方差所对应的风速功率数据定义下部异常数据, 其余为正常数据。

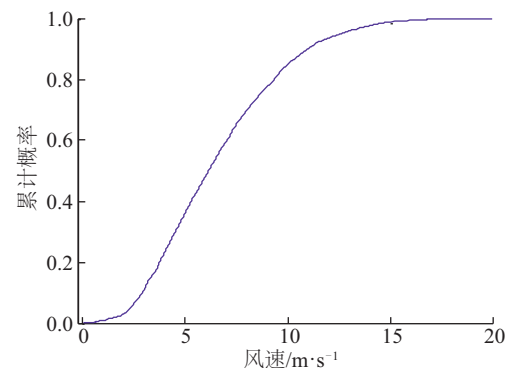
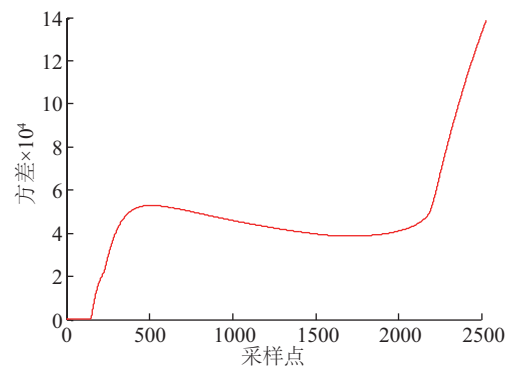
图8  $[170^{\circ}, 180^{\circ}]$  风向风速累计概率图Fig. 8  $[170^{\circ}, 180^{\circ}]$  wind direction wind speed cumulative probability map

图9 组内最优方差曲线

Fig. 9 Optimal interclass variance curve

对风向在  $[170^{\circ}, 180^{\circ}]$  下风速在为  $[7.0, 7.5]$  m/s 之间异常数据的识别结果如图 10 所示。

对风向在  $[170^{\circ}, 180^{\circ}]$  下的另外 35 个区间的风速和功率数据做同样处理, 根据上文展示的步骤, 将得到每个小区间数据中的正常和异常数据, 再通过整理后得到 18 号风力机风向在  $[170^{\circ}, 180^{\circ}]$  下功率曲线中的正常数据集和上下异常数据集。异常数据识别后的风速功率散点如图 11 所示。

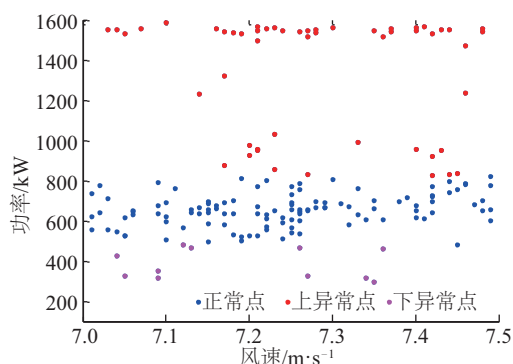
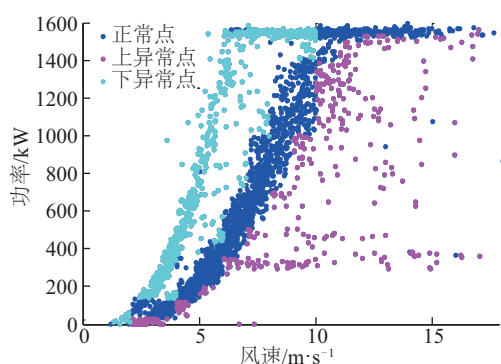


图 10 异常数据分离结果

Fig. 10 Abnormal data separation results

图 11  $[170^\circ, 180^\circ]$  风向下异常数据识别结果Fig. 11  $[170^\circ, 180^\circ]$  wind anomaly data recognition results

最后,分别将其余风向范围内的风速功率数据同样按 $[170^\circ, 180^\circ]$ 之间数据的处理方式进行处理,经归类整理后可得到 18 号风力机的全部异常数据和正常数据,如图 12 所示,天蓝色为上异常数据,紫色为下异常数据,蓝色散点代表正常数据。对 18 号风力机识别完毕之后发现,左侧的天蓝色异常数据存在极强的规律性且规律与正常数据相

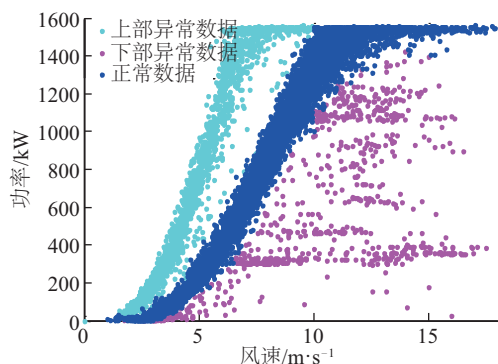


图 12 18 号风力机全部异常数据识别结果

Fig. 12 No. 18 wind turbine all abnormal data recognition results

似。这部分数据之所以会被识别为异常数据是因为其并没有完全遵循生产厂商给定的标准所致,这写数据的存在说明该风力机在出厂时就存在一定质量问题,亦或是在运行过程中出现了一定的故障,此时应联系生产厂商协调解决或者及时检修以免故障加剧。

1 号风力机全部异常数据的识别结果如图 13 所示。

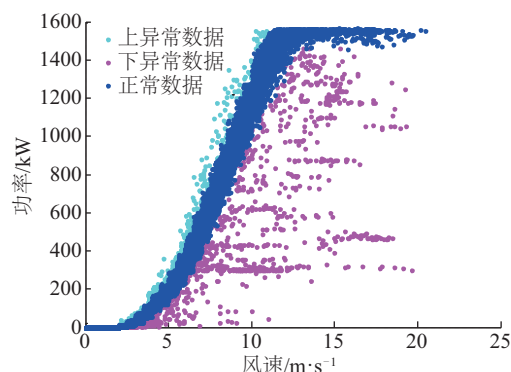


图 13 1 号风力机全部异常数据识别结果

Fig. 13 No. 1 wind turbine all abnormal data recognition results

对 18 号风力机的异常数据进行剔除后的风速功率散点图如图 14 所示,其中红线表示该机组的标准风速功率曲线。对 18 号风力机异常数据识别前相对于标准风速功率曲线的平均绝对误差为 193.39 kW,对异常数据识别并进行处理后,其相对于标准风速功率曲线的平均绝对误差减小为 93.72 kW,减小了将近 99.67 kW,表明异常数据剔除后风速功率散点与标准风速功率曲线更加接近,说明剔除效果良好,体现了算法的有效性。

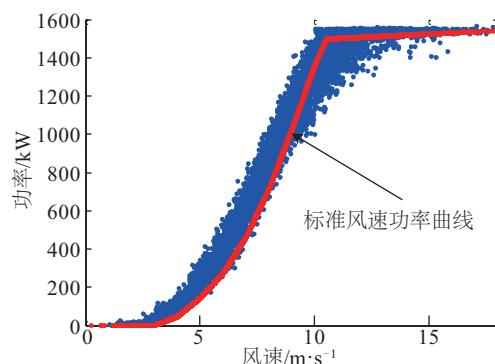


图 14 18 号风力机异常数据处理后与标准风速功率曲线

Fig. 14 No. 18 wind turbine abnormal data processing and standard wind power curve

对 1 号风力机的异常数据进行剔除后的风速功率散点图如图 15 所示,其中红线表示该机组的标准风速功率曲线。

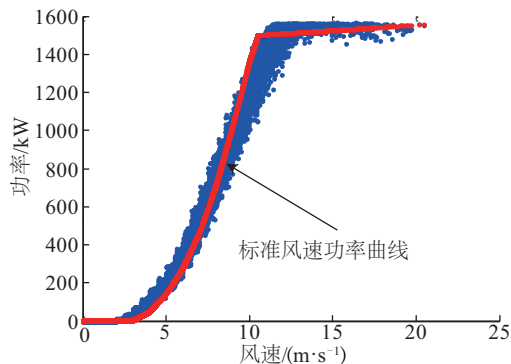


图 15 1 号风力机异常数据处理后与标准风速功率曲线

Fig. 15 Abnormal data processing and standard wind power curve of No. 1 wind turbine

如表 2 所示,对 1 号风力机异常数据识别前相对于标准风速功率曲线的平均绝对误差为 94.47 kW,对异常数据识别并进行处理后,其相对于标准风速功率曲线的平均绝对误差降低为 53.06 kW,减小了 41.42 kW。所以,对 1 号机组的识别后的剔除效果同样说明了计及风向的最优方差算法的有效性。

表 2 对异常处理前后的平均绝对误差

Table 2 Absolute error before and after exception handling

风力机	处理前平均 绝对误差/kW	处理后平均 绝对误差/kW
1 号	94.47	53.06
18 号	193.39	93.72

## 4 结 论

由于各种不可校正因素的存在,实测风电功率数据中必然存在种类繁杂的异常数据,与以往基于复杂数学算法的异常数据识别法相比,计及风向信息的组内最优方差法意义明确,原理简单,计算复杂度小,解决了因技术局限性造成的异常数据识别算法通用性不强等问题,且研究结果表明,对识别出的 1 号和 18 号风力机的异常数据进行处理后,其平均绝对误差较处理之前分别减小了 41.41 和 89.67 kW,说明了算法的有效性。同时该算法也可为风电机组制造商和风力发电企业在运行过程中功率曲线考核评估时提供了一种客观、科学的数据

处理方法。

## [参考文献]

- [1] 刘 波,贺志佳,金 昊. 风力发电现状与发展趋势[J]. 东北电力大学学报, 2016, 36(2): 7—13.
- [1] Liu Bo, He Zhijia, Jin Hao. Current situation and development trend of wind power[J]. Journal of Northeast Electric Power University, 2016, 36(2): 7—13.
- [2] 张丽英,叶廷路,辛耀中,等. 大规模风电接入电网的相关问题及措施[J]. 中国电机工程学报, 2010, 30(25): 1—9.
- [2] Zhang Liying, Ye Tinglu, Xin Yaozhong, et al. Problems and measures of large scale wind power access to power grid[J]. Proceedings of the CSEE, 2010, 30(25): 1—9.
- [3] 杨 茂,张 强. 风电功率超短期预测误差的非参数估计分布研究[J]. 东北电力大学学报, 2018, 38(1): 15—20.
- [3] Yang Mao, Zhang Qiang. Nonparametric estimation distribution of ultra-short-term prediction errors of wind power[J]. Journal of Northeast Electric Power University, 2018, 38(1): 15—20.
- [4] 刘红柳,杨 茂. 基于小波变换和支持向量机的风电功率爬坡事件识别与预测[J]. 东北电力大学学报, 2016, 36(6): 30—35.
- [4] Liu Hongliu, Yang Mao. Identification and prediction of wind power ramp events based on wavelet transform and support vector machines[J]. Journal of Northeast Dianli University, 2016, 36(6): 30—35.
- [5] 杨 茂,齐 玥. 基于相空间重构的风电功率波动性分析及其对预测误差影响[J]. 中国电机工程学报, 2015, 35(24): 6304—6314.
- [5] Yang Mao, Qi Yue. Analysis of wind power fluctuation based on phase space reconstruction and its influence on prediction error[J]. Proceedings of the CSEE, 2015, 35(24): 6304—6314.
- [6] 杨 茂,杨春霖. 考虑风向的风电功率预测不确定性研究[J]. 东北电力大学学报, 2018, 38(5): 9—15.
- [6] Yang Mao, Yang Chunlin. Uncertainty study on wind power prediction considering wind direction[J]. Journal of Northeast Electric Power University, 2018, 38(5): 9—15.
- [7] 龚莺飞,鲁宗相,乔 颖,等. 基于 Copula 理论的光伏功率高比例异常数据机器识别算法[J]. 电力系统

- 自动化, 2016, 40(9): 16—22, 55.
- [7] Gong Yingfei, Lu Zongxiang, Qiao Ying, et al. Copula theory based machine identification algorithm of high proportion of outliers in photovoltaic power data[J]. Automation of Electric Power Systems, 2016, 40(9): 16—22, 55.
- [8] 郎斌斌, 穆 钢, 严干贵, 等. 联网风电机组风速-功率特性曲线的研究[J]. 电网技术, 2008, 32(12): 70—74.
- [8] Lang Binbin, Mu Gang, Yan Gangui, et al. Research on wind speed vs output power characteristic curve of wind power generator interconnected with power grid[J]. Power System Technology, 2008, 32(12): 70—74.
- [9] 杨 茂, 杨春霖. 基于模糊粒计算的风电功率实时预测研究[J]. 东北电力大学学报, 2017, 37(5): 1—7.
- [9] Yang Mao, Yang Chunlin. Research on real-time prediction of wind power based on fuzzy granular computing[J]. Journal of Northeast Electric Power University, 2017, 37(5): 1—7.
- [10] Kusiak A, Zheng Haiyang, Zhe Song. On-line monitoring of power curves [J]. Renewable Energy, 2009, 34(6): 1487—1493.
- [11] Schlechtingen M, Santos I F, Achiche S. Using data-mining approaches for wind turbine power curve monitoring: A comparative study[J]. IEEE Transactions on Sustainable Energy, 2013, 4(3): 671—679.
- [12] Mangalova E, Agafonov E. Wind power forecasting using the  $k$ -nearest neighbors algorithm[J]. International Journal of Forecasting, 2014, 30(2): 402—406.
- [13] 朱倩雯, 叶 林, 赵永宁. 风电场输出功率异常数据识别与重构方法研究[J]. 电力系统保护与控制, 2015, 43(3): 38—45.
- [13] Zhu Qianwen, Lin Lin, Zhao Yongning. Study on the identification and reconstruction of abnormal power output in wind farm [J]. Power System Protection And Control, 2015, 43(3): 38—45.
- [14] 娄建楼, 胥 佳, 陆 恒, 等. 基于功率曲线的风电机组数据清洗算法[J]. 电力系统自动化, 2016, 40(10): 116—121.
- [13] Lou Jianlou, Xu Jia, Lu Heng, et al. Wind turbine data-cleaning algorithm based on power curve [J]. Automation of Electric Power Systems, 2016, 40(10): 116—121.
- [15] 曹 潇. 风玫瑰图在电网气象信息系统中的应用研究[J]. 风能, 2010, (10): 44—45.
- [15] Cao Xiao. Research on the application of Wind rose chart in the meteorological information system of power grid[J]. Wind Energy, 2010, (10): 44—45.

## STUDY ON DATA RECOGNITION OF WIND POWER ABNORMALITY CONSIDERING WIND DIRECTION INFORMATION

Yang Mao<sup>1</sup>, Yang Chunlin<sup>1</sup>, Yang Qiongqiong<sup>1</sup>, Su Xin<sup>2</sup>

(1. Key Laboratory of Modern Power System Simulation and Green Energy & Renewable Energy, Ministry of Education, Northeast Electric Power University, Jilin 132012, China; 2. School of Science, Northeastern Electric Power University, Jilin 132012, China)

**Abstract:** In view of the complexity of the algorithm and the poor recognition effect, as well as the poor generality of the algorithm, it is difficult to effectively identify the complex wind power data. According to the characteristics of the abnormal data source in the measured wind speed and power data, a method for identifying the abnormal data based on the different wind speeds is proposed. The algorithm is based on the optimal variance of the wind speed in different wind directions to identify the corresponding abnormal data. Finally, all the abnormal data of the unit are identified after sorting. The results show that the algorithm can effectively identify the abnormal wind data.

**Keywords:** abnormal data; wind direction; data identification; optimal interclass variance