

# 风电机组异常数据智能识别方法

■文 | 滕繁 李思亮 申云 郭鹏程

风脉（武汉）可再生能源技术有限责任公司

**摘要：**从风电场收集到的大量数据中通常包含异常数据，这样的异常数据给风电功率波动特性、风电功率预测等方面研究带来负面影响，影响着预测的精度。本文分析了风机 SCADA 系统的风速功率曲线中异常数据的来源以及产生机理，并采用三种不同的算法分三个步骤对异常数据进行自动识别，从而为后续研究提供有效的数据集。规模化测试结果表明，该系列算法识别异常数据的准确率在 94.6%，初步符合服务现场风电场数据分析的要求。

**关键词：**风电机组 功率曲线 SCADA 异常数据 智能识别

## 1. 前言

当前，风力发电是可再生能源开发利用中技术最成熟、最具开发规模 and 商业化发展前景的发电形式，研究风电对系统的影响并在此基础上提出相关技术解决措施已成为当前的研究热点和重要课题<sup>[1]</sup>。风电运行数据是电力大数据的重要组成部分。风电历史运行数据是研究风电功率波动特性、风电功率预测<sup>[2-4]</sup>、风电功率对电网影响评估及控制策略<sup>[5-6]</sup>、风电功率曲线的计算、测试以及修正<sup>[7]</sup>的基础，因此，对风电场历史运行数据的记录就尤为重要。

对于风力发电运营管理方，通过实测风速和风电功率得到的功率曲线是衡量风电机组和风电场经济技术水平的最佳标尺<sup>[8]</sup>，可以用于评估风机和风电场的性能和运行状况。然而，从风电场收集到的大量数据

中通常包含异常数据点，这样的异常数据表现出数据缺失，越限，功率波动梯度过大等特征。通过对风电场的运行经验和统计结果的分析，异常数据点的主要来源有<sup>[9]</sup>：风机叶片受污垢和附冰；维修或弃风限电造成的风电场停机；传感器故障等控制系统问题；由表计测量误差导致的异常观测值等。

李闯等人<sup>[10]</sup>采用功率—风速 Bin 功率曲线拟合数据滤波方法，去除中高风速段中限电严重的异常数据，该滤波方法利用了功率 Bin 的数据分布特性和功率曲线的出力特性。赵永宁等人<sup>[11]</sup>针对弃风造成的数据问题，提出一种基于四分位原理和聚类算法的数学模型，采用该模型可有效识别并剔除弃风异常数据。以上两方法都只是去除了部分原因的数据，未能将功率曲线中不同类型的异常数据一并有效识别。KUSIAK

A 等人<sup>[12]</sup>涉及对类似异常数据的处理，但所采用数学模型为学习算法，需要经过大量的正常数据进行训练，实用性较差；MANGALOVA E 等人<sup>[13]</sup>采用参数化公式表示的条件准则来筛选并剔除散点图中的异常数据，该方法模型简单，但对不同的散点图需要设置不同的参数，通用性不强。

前人研究大部分是以功率曲线的图形出发，采用各种方法从图形曲线上删除功率曲线以外的异常数据，对于功率曲线异常数据的机理分析并不是十分深入，采用的方法但是在功率曲线内部也可能存在部分异常数据，如果不从异常数据的来源分析，会使得这一部分异常数据隐含在正常数据当中，另外，功率曲线边缘的部分数据也有可能是正常数据，通过曲线图形来删除也可能将此正常数据删除，因此，本

文分析了风机 SCADA 系统的风速功率曲线中异常数据的来源以及产生机理, 并采用三种不同的算法分三个步骤对各种类型的异常数据进行自动识别。

## 2. 异常数据分类

风机 SCADA 实测风机功率数据中, 往往含有大量不符合风机正常出力特性的异常运行点, 如在较大的实测风速下, 风机功率低于标准值很多甚至为 0。其原因是多方面的。如风机组出现故障、维护停机、机组限出力运行<sup>[14]</sup>、风速传感器失灵等。我们将异常数据主要标记为三类: 底端有风停机异常数据, 中部限功率异常数据, 其余分散异常数据。

以某风场某台双馈机组为例, 该风机额定风速 12 m/s, 额定功率为 1 500 kW。从风电场 SCADA 系统中导出 2014 年 1-6 月 10 min 内的平均风速和功率数据, 绘制风速 (WS) - 功率 (WP) 曲线散点图, 如图 1 所示。

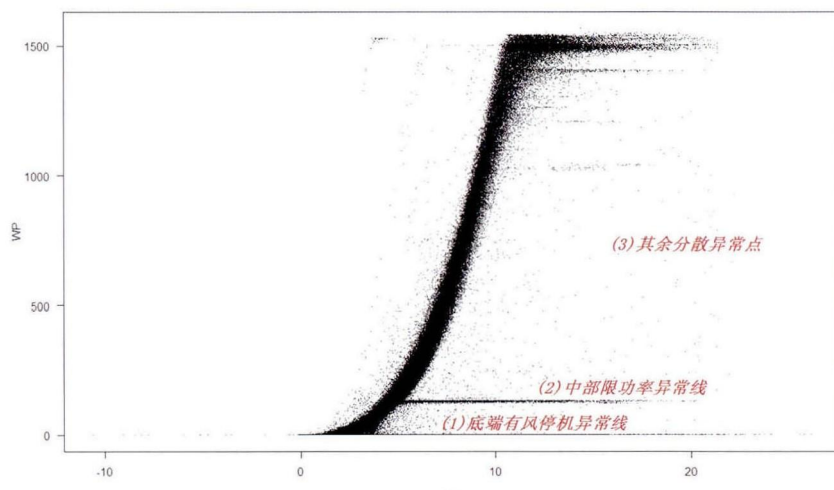


图1: 风力发电机组风速-功率散点

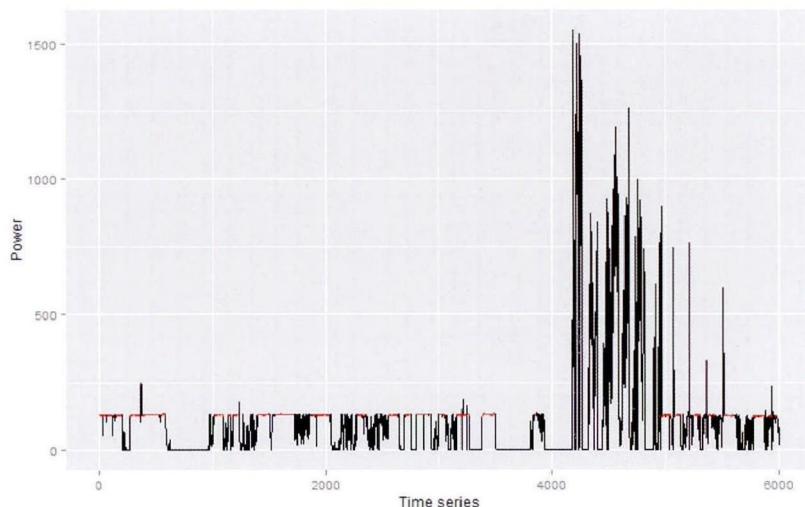


图2: WP-WS随时间变化率

## 3. 智能识别步骤

### 3.1 底端有风停机异常数据

底端异常数据产生的原因是有风停机, 限功率等。底部异常数据常见于单条或多条密集的异常数据线, 异常数据线一般在功率为零的附近。根据工程经验, 如果发现某一类风电机组的底端异常数据都为零值附近, 可以采用“直接设功率

上限值法”, 以零值为上限值, 将零值以下的所有点全部标记出来; 在没有工程经验指导的情况下, 本文设计了一种基于百分位数极差的自动寻功率上限值的算法, 可自动寻找到有风停机异常数据的功率上限值, 将功率上限值以下的所有数据全部标记, 简称“自动寻功率上限值法”。该算法认为底端异常功率点为功率分布密度较大, 且低于该点功率分位数距也很大。简要过

程如下:

(1) 限定一个较小的功率上限值  $P_h$ , 建立该功率以下的功率分布  $F_p$ 。

(2) 寻找功率分布密度最高点, 并考察该点以下的功率分位数距是否超过限值  $D$ 。如果超过则将该功率点定为搜索下限  $P_l$ 。如果没有超过则终止。

(3) 在上限值  $P_h$  和下限  $P_l$  之间重复步骤 2, 直到终止。

### 3.2 中部限功率异常数据

风速 (WS) - 功率 (WP) 曲线在限功率后, 会存在中部的异常数据, 当限功率点密集的时候会形成多条线。限功率时, WS 仍然会随时间 T 波动变化, 而 WP 不会随时间 T 变化, 限定在某一个值, 限功率是随着时间 T 连续发生的现象, 如图 2 所示。

根据以上原理, 设计了一种基于连续时间识别功率变化率的算法, 核心是找到满足设定条件的连续不发生变化的功率片段, 且该功率片段随风速变化不发生明显改变。其基本思想如下:

(1) 将整个功率划分成多个连续的功率区间, 区间的宽度不宜过高, 一般为 6kW。

(2) 在每个区间内寻找时间序列上连续功率片段, 这些点为潜在可能发生限电的功率点, 即为  $P_i$ 。

(3) 对上述 2 中每个连续的功率片段, 拟合对应时间序列上的功率, 判定对应的功率风速数据的拟合曲线斜率是否接近于 0。如果是, 则说明其发生了限电。

4. 重复步骤 2, 遍历所有划分的功率区间, 标记出符合条件的数据。

### 3.3 其余分散异常数据

完成上述 2 个步骤后, 最后在风速 (WS) - 功率 (WP) 曲线左右两侧还存在许多异常的散点, 可能是由于风电机组满发或者亏发、机组故障、风速传感器失灵等原因造成的, 我们设计了一种基于 Stepwise 的预测值置信区间的算法进行识别, 主要思想是:

(1) 采取拟合样条曲线, 设定预测值的置信区间, 区间大小通过 t 统计量确定, 标记区间曲线以外的点为异常工况。

(2) 采用正常工况点重新拟合样条曲线并更新置信区间, 同时对之前标记为异常工况的点重新判定。

(3) 重复步骤 1, 2 直到没有新的点标记为异常, 或者没有已标记为异常的重新判定为正常的情况出现。

## 4. 智能识别结果

### 4.1 底端有风停机异常数据

经过 3.1 算法处理后, 可以标记出底部所有异常数据, 如图 3 红色标记点所示。

### 4.2 中部限功率异常数据

经过 3.2 节算法处理后, 可以标记出异常数据 “limited”, 如图 4 所

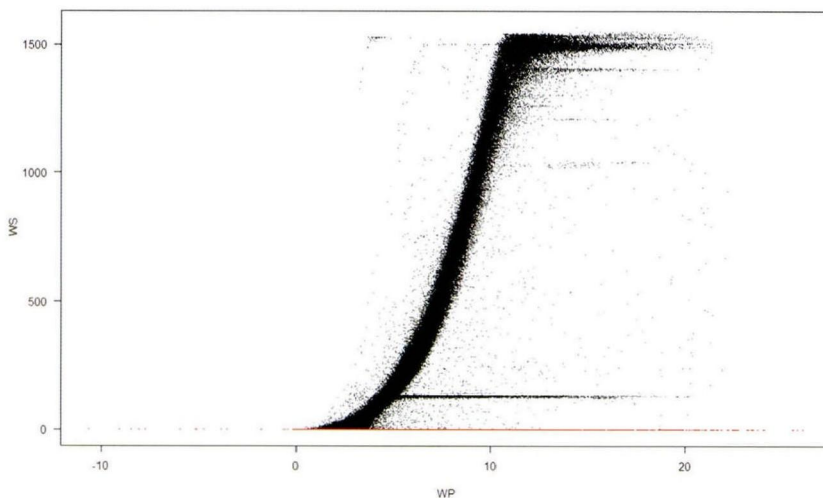


图3: 底端有风停机数据标记

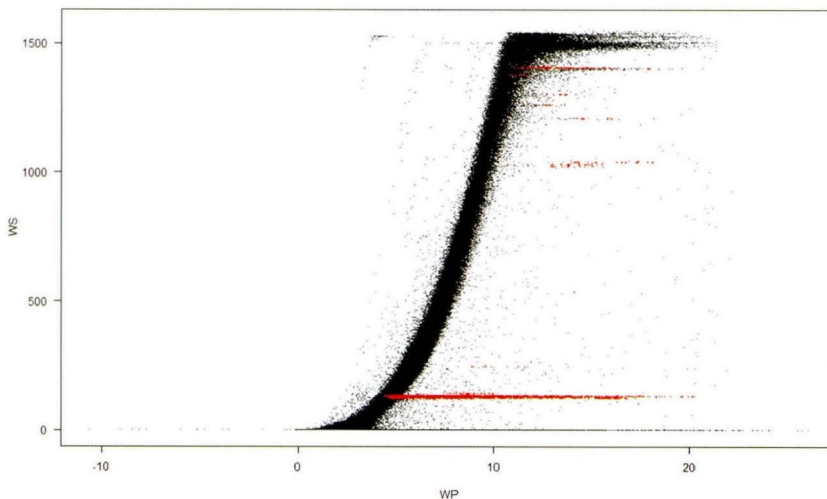


图4: 风机SCADA中部限功率异常数据标记

示。从图中可以看出，部分红色的标记点隐含在了功率曲线内部，其实按照有风停机的原理出发，这一部分隐含在正常数据中的数据为属于有风停机数据，因此通过该算法可以将其标识出来。

#### 4.3 其余分散异常数据

图 5 是通过 3.3 中的算法标记了红点的部分为其余分散异常数据。

通过以上三个步骤处理后，基本可以将原始数据中的异常数据识别出来，最终呈现结果如图 6 所示。

#### 4.4 规模测试结果

采用某风电场 33 台风机 SCADA 数据对数据预处理模块的上述 3 个算法进行整体测试。数据时间跨度为 2010 年 1 月-2014 年 4 月，10 分钟数据记录频率，被随机分割成每 3 个月数据为一组，一共 319 组测试数据。测试的软硬件平台如下：

CPU: Intel(R) Core(TM) i7-4790 CPU @ 3.6GHz 3.60GHz

Memory: ARM 8.00GB

System: Windows 7 Service Pack 1 (SP1) x64

HD: ATA AT1000DM003-1ER1 SCSI Disk Device

开发平台: VS2010, R (programming language)

通过对算法参数的优化调整，测试结果对比风电机组记录的故障数据，综合正确识别率为 94%。性能结果如下：

(1) 算法的识别率：3 个算法能够有效识别所有异常数据，综合识别率较高，配置一套合理的算法

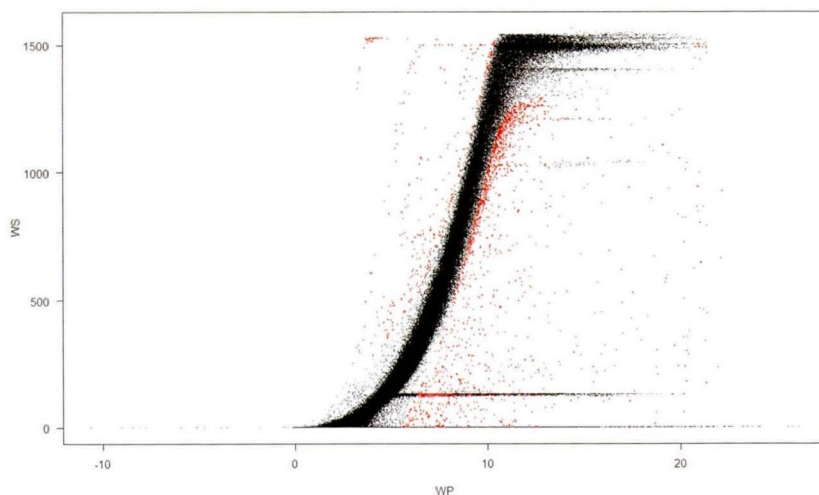


图5：其余分散异常数据标记

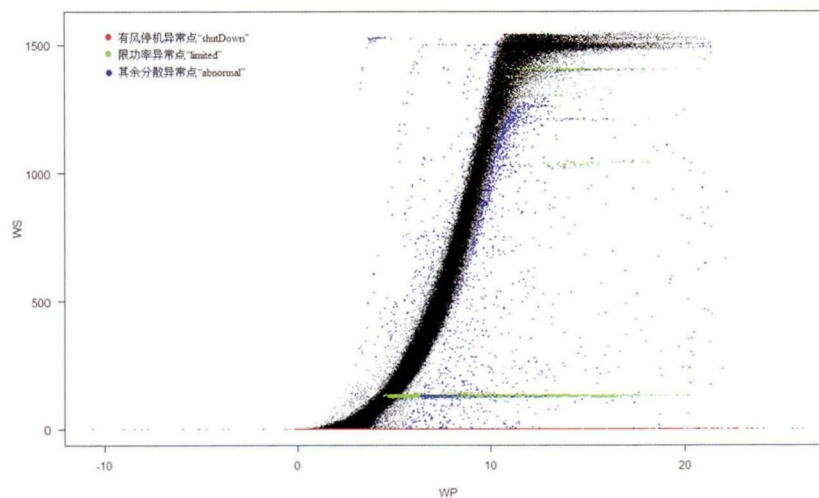


图6：所有异常数据分类标记

参数，基本可以达到 95% 以上的识别率，该参数基本可以用于一个风场的全部数据，显示了较好的智能处理效果。

(2) 算法的稳健性：算法在测试过程中没有报错，稳健性较好。

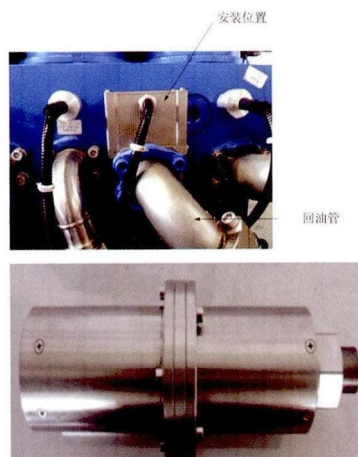
(3) 算法的耗时：在对 3 个月的数据进行异常识别的过程中，算法整体耗时基本在 20s 左右，耗时在合理接受范围内。

## 5. 结论与应用展望

本文从风速 (WS) - 功率 (WP) 曲线散点图的异常数据原理出发，通过分析了风机 SCADA 系统的风速功率曲线中异常数据的来源以及产生机理，采用三次分步智能识别算法，依次将不同类型异常数据自动标记出来，从而为后续研究风电功率波动特性、风电功率预测等风电

(下接第 70 页)





静电传感器安装与实物

该传感器有三种不同内径的规格（1.5in, 1in, 0.5in）可供选用。

②颗粒成份分析技术，是指对润滑油进行定期取样，采用 MOA 型多元素油料分析光谱仪对油样进行分析，可以获得油液中磨损金属的成分和含量及颗粒、水分状况，从而有针对性采取措施。

以上是对风电后市场齿轮箱污染控制的粗浅认识，希望与业界共同探讨，共同为此市场的健康发展而努力！ **WIND**



德国KLOTZ（克劳斯）颗粒计数系统分析系统

收稿日期：2015 年 8 月 15 日

作者简介：刘克树，硕士学历，高级工程师。现从事风电产品运营管理工作。

（上接第 63 页）

相关问题提供有效的数据集。

本识别方法适用于无故障数据或者故障数据表和历史信息表无法映射时对风电机组是否故障做出初步判断，同时该系列算法对计算资源要求不高，有较好的应用潜力。 **WIND**

## 参考文献：

- [1] 朱倩雯，叶林，赵永宁等，风电场输出功率异常数据识别与重构方法研究，电力系统保护与控制，2015，Vol.43 No.3，38-45
- [2] 杨茂，熊昊，严干贵，等，基于数据挖掘和模糊聚类的风电功率实时预测研究[J]，电力系统保护与控制，2013，41(1): 1-6.
- [3] 夏冬，吴俊勇，贺电，等，一种新型的风电功率预测综合模型[J]，电工技术学报，2011，26(增刊 1): 262-266.
- [4] 王贺，胡志坚，陈珍，等，基于集合经验模态分解和小波神经网络的短期风功率组合预测[J]，电工技术学报，2013，28(9): 137-144.
- [5] 于凡，赵瑜，周玮，等，基

于混合储能系统的平抑风电波动功率方法的研究[J]，电力系统保护与控制，2012,39(24): 35-40.

[6] 胡国伟，别朝红，王锡凡，考虑运行可靠性的含风电电力系统优化调度[J]，电工技术学报，2013，28(5):58-65.

[7] 申新贺，潘东浩，唐继光，大型风电机组功率曲线的分析与修正[J]，应用能源技术，2009(8): 28-32.

[8] 马平，刘昌华，风力发电机组功率曲线的验证[J]，可再生能源，2008，26(6): 82-84.

[9] KUSIAK A, ZHENG H, SONG Z. Models for monitoring wind farm power[J]. Renewable Energy, 2009, 34(3):583-590.

[10] 李闯，申焮等，基于统计阈值的风机功率曲线拟合数据滤波方法[J]，电力行业信息化年会论文集，2013: 360-362.

[11] 赵永宁，叶林等，风电场弃风异常数据簇的特征及处理方法[J]，电力系统自动化，2014，38(21): 39-46

[12] KUSIAK A, ZHANG H, SONG Z, Models for monitoring wind farm

power[J], Renewable Energy, 2009, 34(3): 583-590

[13] MANGALOVA E, AGAFONOV E. Wind power forecasting using the k-nearest neighbors algorithm[J], International Journal of Forecasting, 2013, 29(3): 402-406

[14] 范高峰，赵海翔，戴慧珠，大规模风电对电力系统的影响和应对策略，电网与清洁能源，2008，24(7): 44-48

收稿日期：2014 年 8 月 6 日

## 作者简介：

李思亮（1985-），通信作者，男，硕士研究生，中级工程师，主要从事风电技术咨询中的算法开发和数据分析工作。

滕繁（1986-），女，硕士研究生毕业，主要研究方向有风电机组故障预警系统，风机备件智能管理系统等。

申云（1979-），男，本科，信息系统项目管理师，软件设计师，主要研究方向：大型软件设计、系统架构、软件开发项目管理。