

# 基于 Change-Point 的风电数据挖掘算法研究

胥 佳<sup>1</sup>, 李韶武<sup>1</sup>, 王桂松<sup>2</sup>, 刘瑞华<sup>1</sup>, 朱耀春<sup>1</sup>

(1. 龙源(北京)风电工程技术有限公司, 北京 100034; 2. 华北电力大学控制与计算机工程学院, 北京 102206)

**摘 要:** 风电机组运行过程中,一些故障导致设备状态发生改变,状态的改变发生在一个持续的时间序列中,找到变化点的时间对于故障回溯及根本原因分析具有重要价值。该文研究风电信号及状态时序变化的特点,引入统计学中的 Change-Point 算法,通过划分不同置信区间求取置信度方法解决奇异变点的不确定度问题。通过实验对算法进行验证,得出以下结论:Change-Point 算法能够有效挖掘到历史数据中的一维及二维模型数据的变化,并给出变点;Change-Point 算法思想是挖掘出数据本身的规律性,不受其他条件限制,因此可广泛应用于风电机组数据采集与监视控制(supervisory control and data acquisition, SCADA)系统变量数据挖掘中的问题回溯,快速定位 SCADA 数据状态变化点。

**关键词:** 数据挖掘; 数据采集与监视控制(SCADA)系统; 风电机组; Change-Point; 置信度

**中图分类号:** TK83

**文献标志码:** A

## 0 引 言

近年来,风能在世界能源结构中的地位越来越突出,装机量增大的同时也伴随着一系列问题。虽然风电机组在发电过程无原料成本,但目前中国的风电设备因长期运行在室外恶劣自然环境中,运行水平不佳,故障率比常规发电机组高、停机时间长,风电的度电维护成本仍维持在较高水平。如何快速找到风电机组存在的问题,定位机组劣势的时间点并及时开展优化预警以降低维护费用,成为风电场运营中迫切的需求。

风电场 SCADA 系统持续采集风电机组的上千个传感器的信号,数据中蕴藏着风力机运行状态的相关信息。目前大量专家和学者开展风电数据挖掘技术的研究,Andrew 开展基于数据的功率曲线监测技术研究<sup>[1]</sup>,郭鹏<sup>[2]</sup>、孙建平<sup>[3]</sup>、李思亮<sup>[4]</sup>等通过非线性状态评估模型开展齿轮箱、塔筒振动等故障监测研究,王欣、潘东浩等<sup>[5]</sup>采用统计概率分布分析方法开展风电机组偏航误差分析,赵永宁等<sup>[6]</sup>采用  $k$ -均值技术开展功率曲线异常数据识别工作以上状态监测方法改进复杂算法模型学习效率过低问题,可总体定位性能劣化或故障发生的时间段。在进行风电机组状态监测、预警试验

中,风电机组劣化或故障时间存在时间不确定度问题,当发现设备故障后,需要快速分析设备最早出现故障的时间点,用于关联这个时间点周围环境事件对设备异常产生的影响,分析设备异常根本原因,做出及时预防性维修。

本文主要通过研究风电信号及状态时序变化的特点,开展 Change-Point 算法构建,并将 Change-Point 算法引入风电数据挖掘应用;在功率曲线测试应用中,领域内有关开展对功率曲线劣化时间点定位分析的方法较少,通过 Change-Point 方法对功率曲线性能劣化时间点的关联分析找到功率曲线异常的根本原因,为功率曲线性能劣化分析提供新方法。

## 1 风电机组故障分类及其数据特性

风电机组故障通常来源于内部部件损坏或外部事件驱动,故障的发生会改变自身的运行状态及表现,而不同的物理机理变化也会体现在数据特性的变化上。本节介绍几种常见的风电设备故障及其数据特性。

第 1 类故障,部件的损坏是渐变的,其对应的数据变化也是渐进的。例如齿轮箱齿轮或轴承磨损会造成齿轮箱振幅增大、频率升高;变桨电机散热风扇

收稿日期: 2017-11-08

基金项目: 中国电机工程学会“青年人才托举工程”(J2B2017304)

通信作者: 王桂松(1992—),男,硕士研究生,主要从事风电机组大部件预警技术方面的研究。18810791715@163.com

轴承损坏,会导致变桨电机散热系统功能下降,变桨电机温度随着轴承损坏加剧,温度不断增高。图1为某2 MW风电机组叶片变桨电机温度时序图,采样数据为该机组10 min间隔叶片变桨电机温度数据。在变桨电机散热风扇产生故障一段时间内,变桨电机温度逐步升高。

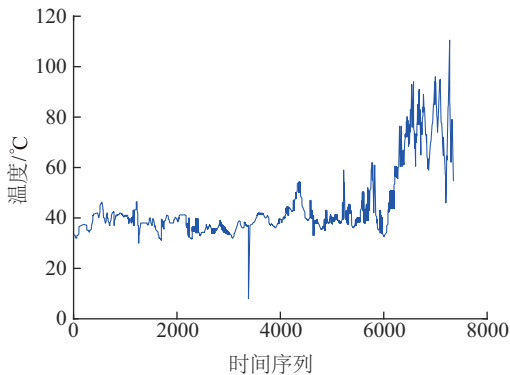


图1 某2 MW风电机组叶片变桨电机温度时序图

Fig. 1 Temperature sequence diagram of a 2 MW wind turbine blade variable propeller motor

第2类故障,部件的损坏或异常呈非趋势性渐变或突变,这类问题导致相应的传感器数据信号突变,变化后信号保持一个平稳状态。例如:传感器由于机舱振动造成了短时间内信号虚接,数据错误;现场工作时风电机组IP设置错误,误接入其他机组。图2为某1.5 MW风电机组齿轮箱油泵出口压力时序图。传感器虚接后信号产生突变,修正后信号恢复正常。

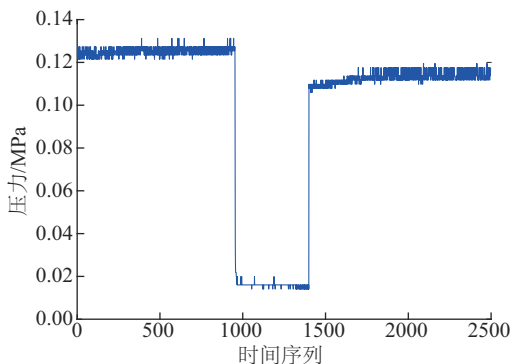


图2 某1.5 MW风电机传感器虚接的压力时序图

Fig. 2 Jumper pressure sequence diagram of a 1.5 MW wind turbine sensor

第3类故障,由于风电机组本身处在变化的自然环境中,当自然环境发生改变时设备参数随之变动,但风电机组本身的一些控制特性或出力特性是稳定的。通常由2个参数或多个参数构成的控制特性或出力特性模型也会由于外界环境发生变化,使观察到的信号不再

是一维的而是多维的。例如:功率曲线的变化,转速与扭矩的变化,转速与桨叶角度的变化。图3为某1.5 MW风电机组2015年5月功率曲线图。该风电机组在额定风速前出现功率特性劣化情况,而达到额定风速后功率较其他机组更高,发现是环境因素和控制策略改变综合作用的结果。

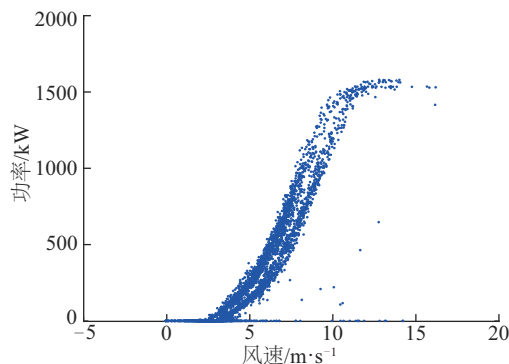


图3 某1.5 MW风电机组2015年5月功率曲线

Fig. 3 Power curve of a 1.5 MW wind turbine in May 2015

当风电机组发生类似上述故障时,时序信号统计特征都会出现连续的非偶然的变化,这个变化的时间点就称为变点。找到变点,再对变点时刻进行深度挖掘,往往可追溯到故障产生的根源。

## 2 Change-Point 算法

在分析历史数据状态时,特别是在数据集较大时,通过统计阈值计算的控制图法分析计算状态变化点时存在大量数据点超限的误报现象,而且无法准确定位到最开始的变点时刻,在传统控制图基础上构建的 Change-Point 方法在分析数据特性方面要明显优于单一控制图法,本文重点分析用 Change-Point 提取风电数据变点的应用。

Change-Point 算法利用一定的统计指标或统计方法对时间序列的状态进行观测,以准确估计出变点的位置。变点问题作为统计学中的一个重要课题,最早应用于工业质量控制,后延展到金融经济、计算机、气象学、流行病学等多个领域。变点问题分为连续和离散2种形式<sup>[7]</sup>,对于风电机组运行数据来说是时间上的连续序列,如若对样本所有观察点进行变点提取,应用连续形式;但抽样时往往只在一些指定的时刻(例如等距离时刻)去观察,样本是一种自然的离散形式,要根据具体问题具体分析;本文主要对连续时序数据进行处理,时序数据变点提取在数学上可描述为:假设存在一个数据

集,每个数据观测值相互独立,如果在某一时刻,模型中的某个或某些变量突然发生变化,即存在一个时间点,在该点之前,数据集符合一个分布,在该点之后,数据集符合另外一个分布,则该点为该数据集的变点。

经过分析总结,Change-Point 方法分析数据特性可总体归为 2 类状态:状态 1,存在一个时间点,在该点之前数据集符合一个分布,在该点之后突变为另一个分布,两者分布特性明显不一致,该状态对应第 1 节中的第 2 类故障。

从统计学角度状态 1 可描述为:

$$x_t = u(t) + e_t, \quad t = 1, 2, 3, \dots, n \quad (1)$$

式中,  $e_t$  ——随机变量,  $e_t$  的均值  $E(e_t) = 0$ ;  $u(t)$  ——  $x_t$  的期望值,假定格式为:

$$u(t) = \begin{cases} u_1, & t = 1, 2, \dots, k \\ u_2, & t = k + 1, \dots, n \end{cases} \quad (2)$$

式中,  $k$ 、 $u_1$ 、 $u_2$  ——未知常数,  $u_1 \neq u_2$ , 则  $k$  为变点。

状态 2,存在一个时间点,在该点前数据集符合一个分布,在该点之后,趋势性渐变为另一个分布,两者分布特性明显不一致,该状态对应第 1 节中的第 1 和第 3 类故障。

$$x_t = u(t) + e_t, \quad t = 1, 2, 3, \dots, n \quad (3)$$

$$u(t) = \begin{cases} u_1, & t = 1, 2, \dots, k \\ u_2(t), & t = k + 1, \dots, n \end{cases} \quad (4)$$

式中,  $u_2(t)$  ——起点为  $u_1$  的随时间变化的函数,当  $u_2(t)$  的一阶导数大于 0,则是状态 2 的形态,  $k$  为变点。

自 20 世纪 70 年代以来,许多统计学家投入到变点问题研究领域,估计和检测变点问题的方法也不断发展完善<sup>[8-9]</sup>,如最小二乘法、极大似然法、CUSUM 法(累积和法)、迭代累计平方和方法、Bayes 方法、局部比较法、小波分析法等。其中最小二乘法、极大似然法及 CUSUM 法应用最为广泛,CUSUM 法<sup>[10]</sup>通过对观测值与目标值之差的累积和来描点,方法简洁有效,在经济学等领域得到了广泛应用。一段时间内统计值高于全体均值,高于均值的数据量会持续累积,CUSUM 控制图呈现稳步增长趋势;相反,假设在一段时间内统计值低于全体均值,低于均值的数据量同样会持续累积,CUSUM 控制图呈现稳步下降趋势。某些数据特性在直观上不能发现显著变化,通过原始数据无法准确定位到变点时间域,而通过 CUSUM 控制图累加数据特性后可总体定位到变点位置,本文选用 CUSUM 法实现 Change-Point 算法。

CUSUM 的主要原理:假设有数据模型:

$$x_t = u(t) + \varepsilon_t, \quad t = 1, 2, 3, \dots, n \quad (5)$$

式中,  $u(t)$  ——一个非随机函数;  $\varepsilon_t$  ——线性过程,现考虑较简单的情形即  $u(t)$  只取 2 个值:

$$u(t) = \begin{cases} u_1, & t \leq k^* \\ u_2, & t > k^* \end{cases} \quad (6)$$

式(6)中  $u_1$ 、 $u_2$  和  $k^*$  未知。变点  $k^*$  的 CUSUM 估计定义为:

$$\hat{k}^* = \min \left\{ k: |U_k| = \max_{1 \leq j \leq n} |U_j| \right\} \quad (7)$$

其中

$$U_k = \left\{ \frac{k(n-k)}{n} \right\}^{1-\gamma} \left\{ \frac{1}{k} \sum_{i=1}^k X_i - \frac{1}{n-k} \sum_{i=k+1}^n X_i \right\}, \quad 0 \leq \gamma \leq 1 \quad (8)$$

使用 CUSUM 方法确定变点的总体位置,在搜寻变点位置时存在时间不确定度问题。为了解决不确定度问题,在实际应用中使用变异的 CUSUM 方法<sup>[11]</sup>,本文使用二分法进行分割搜索,通过 CUSUM 方法找到变点总体位置后进行二分法分割处理,通过计算置信度的大小定位变点发生区间,继续使用二分法分割,直到准确搜寻到变点时刻,算法具体流程如图 4 所示,其中置信度大小确定变点算法的敏感性,也就是信号波动的大小以及变点数量。

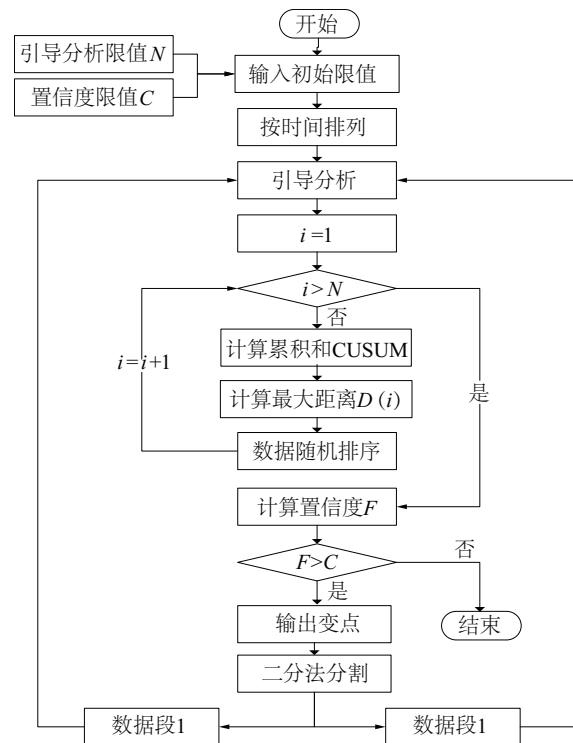


图 4 Change-Point 进行变点搜寻流程图

Fig. 4 Flow chart of Change-Point search

### 3 模型验证

在风电数据挖掘的应用中<sup>[12]</sup>,经常会遇到2个参数组合在一起表征设备的状态,例如风速与功率代表机组性能表现,转矩与转速表征机组的扭矩控制,转速和叶片桨距角表征机组的变桨控制等,这种模型的变点分析,可应用多维 Change-Point 技术,也可进行预处理,将二维数据转化成一维数据后进行分析。

案例1:选用2 MW 风电机组变桨电机温度信号作为测试信号,图5是该信号的时序图,SCADA 存储数据为10 min 平均值,从图5中可明显看到信号在序号约6000处有明显升高。在 Change-Point 算法中,将数据窗口设置精度为12,即找到变点的误差为2 h,置信度设置为0.95,通过计算找到变点1个,位置为5712,计算较准确。

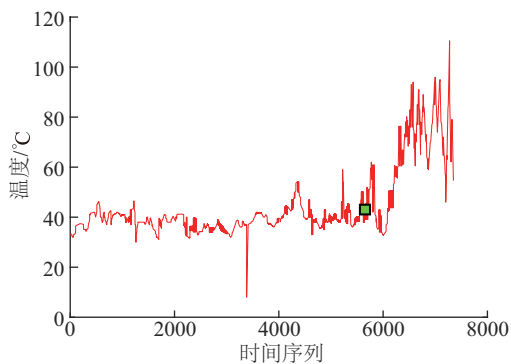


图5 一维数据 Change-Point 算法变点搜寻

Fig. 5 Change-Point search of one-dimensional data

案例2:某台1.5 MW 风电机组对比全场其他机组功率曲线偏低,但额定功率稍高,通过数据挖掘发现机组控制参数与其他机组不同,为了回溯问题产生原因,对数据进行长周期观察,发现机组功率曲线出现过明显变化,现使用 Change-Point 算法对变点进行搜寻(图6),

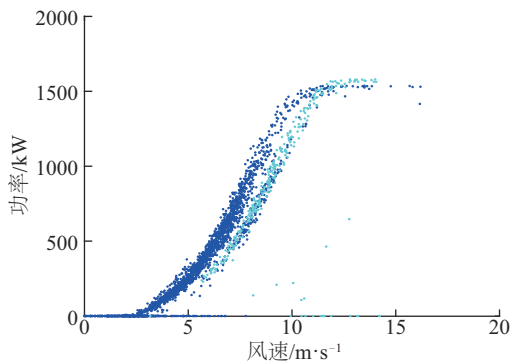


图6 功率曲线变点分析

Fig. 6 Power curve variation point analysis

将数据窗口设置精度为6,即找到变点的误差为1 h,置信度设置为0.9,通过计算找到变点1个,对应的时间为5月14日12:40。通过工单核对,发现5月14日有厂家工作人员对风电机组进行故障处理,在此过程中对机组更新错误的程序,由此导致机组性能下降,造成发电损失。

案例3:某风场风电机组某时间段内出现功率特性波动性下降问题,通过 Change-Point 方法绘制的机组功率曲线 PC 图呈现散点散乱现象,但又具有一定规律可寻,功率特性 PC 图如图7,呈现不同的功率曲线带。

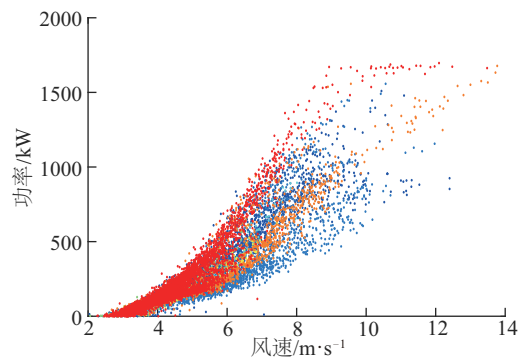


图7 某风场时间段内功率特性 PC 图

Fig. 7 Wind field characteristics period power PC figure

由图7可总体划分为不同功率曲线,全场各机组之间变化时间点较为统一,变化趋势相似;针对此问题对风场某台机组此时间段内10831条数据进行 Change-Point 变点搜寻,置信度设置为0.95,通过计算找到6个变点,变点位置分别在原始数据功率时序图、残差序列绝对值、CUSUM 累积和曲线中给出,如图8所示。

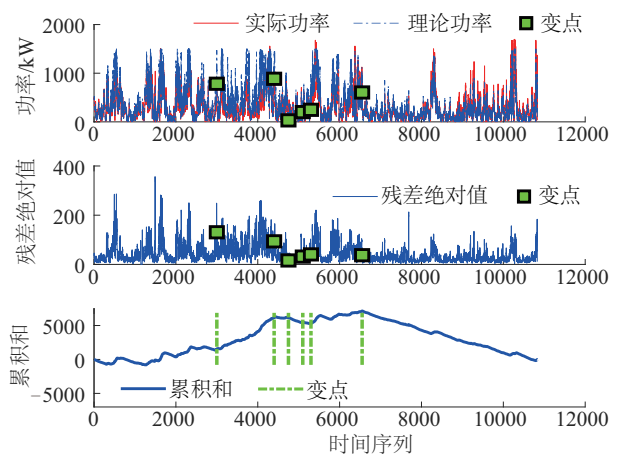


图8 某机组5月份 Change-Point 算法变点搜寻结果

Fig. 8 Wind turbine Change-Point search result in May

在变点时间发现功率特性相对于前一时间序列功



率特性有显著差异,变点时间和相对应的功率特性数据如表 1。

表 1 变点时间及其功率特性数据

Table 1 Variation point time and its power characteristics

变点时间	实际功率/kW	理论功率/kW	残差/kW
2014-05-01	339.67	577.92	-238.25
2014-05-12	337.33	538.80	-201.47
2014-05-15	301.91	322.00	-20.09
2014-05-20	68.72	66.00	2.72
2014-05-22	124.38	149.20	-24.82
2014-06-02	937.89	835.98	01.911

在追溯到时间点后针对该时间点回溯问题产生原因,发现大部分变点时间与当地天气变化时间点基本吻合,功率曲线表现与天气关系如表 2。

表 2 功率曲线表现与天气关系表

Table 2 Power curve performance vs. weather table

时间间隔	功率曲线表现	天气
2014-05-01~2014-05-09	严重下降	多云为主,无雨
2014-05-10~2014-05-11	略有提升	轻微降雨
2014-05-12~2014-05-13	略有提升	较大降雨
2014-05-15~2014-05-16	有提升	大雨
2014-05-20~2014-05-21	有提升	轻微降雨
2014-06-02~2014-06-03	明显提升	大雨(并加装叶片增功装置)

该风场功率特性与降雨有某种关联因素,在深层分析后得知,该风场风电机叶片气动性能较差,极易受灰尘污垢等影响,在多云少风天气随着叶片受灰尘、污渍的影响气动性能会逐渐变差;相反,在叶片受雨水冲刷后,叶片气动性能会发生明显改善。

以上 3 个实例验证了 Change-Point 算法的通用性强、时间定位准确的特点,Change-Point 方法在风电机组状态监测与预警工作中慢慢发展起来,为实现在线预警提供了新思路。

## 4 结 论

1)风电设备故障及变化发生时间点分析是风电数据挖掘必备的一项技术,如何找到相关的变点在工程应用中具有重要的研究价值。Change-Point 理论在统计学科中较为成熟,适用于风电领域对于故障回溯或变化时间点分析。本文通过对风电故障特征及数据特点的分析,将 Change-Point 技术与风电需求结合以求解决风

电设备故障及变化发生时间点查找问题,该技术已在大量实际工程业务中开展使用,文中挑选的 3 个实例证明了算法的有效性。在工程使用中同时需要注意风电机组故障停机或通讯中断后时序信号不连续及突变等数据的筛选清洗问题。

2)由于 Change-Point 技术通过大量数据统计实现异常数据识别,优势是适用性强、无需构建复杂模型,但缺点同样明显,就是该算法计算资源消耗高于一般算法,因此在预警技术这种实时分析要求较高的场景实用性较差。但随着大数据技术的发展,计算资源不再是应用瓶颈,因此近些年逐渐又引起学者和各应用行业的研究兴趣,很多行业用其开展状态监测与故障预警的应用。在风电行业如何应用,可作为后续研究的方向。

## [参考文献]

- [1] KUSIAK A, SONG Z. On-line monitoring of power curve [J]. Renewable energy, 2009, 34(6): 1487-1493.
- [2] 郭鹏, David Infield, 杨锡运. 风电机组齿轮箱温度趋势状态监测及分析方法[J]. 中国电机工程学报, 2011, 31(32): 129-136.  
GUO P, INFELD D, YANG X Y. Wind turbine gearbox temperature trend state monitoring and analysis methods[J]. Proceedings of the CSEE, 2011, 31(32): 129-136.
- [3] 孙建平, 朱雯, 翟永杰, 等. 基于 MSET 方法的风电机组齿轮箱预警仿真研究[J]. 系统仿真学报, 2013, 25(12): 3009-3014.  
SUN J P, ZHU W, ZHAI Y J, et al. simulation study on precautionary gearbox of wind turbine based on MSET method[J]. Journal of system simulation, 2013, 25(12): 3009-3014.
- [4] 李思亮, 郑禄佳, 申云, 等. 风电机组主要部件故障预警及其系统实现[J]. 风能产业, 2017(1): 68-74.  
LI S L, ZHENG Y J, SHEN Y, et al. Main components failure warning of wind turbine and its system realization [J]. Wind energy industry, 2017(1): 68-74.
- [5] 王欣, 吴根勇, 潘东浩, 等. 基于运行数据的风电机组偏航优化控制方法研究[J]. 可再生能源, 2016, 34(3): 413-420.  
WANG X, WU G Y, PAN D H, et al. Wind turbine yaw control optimization utilizing the running data[J]. Renewable energy resources, 2016, 34(3): 413-420.
- [6] 赵永宁, 叶林, 朱倩雯. 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.  
ZHAO Y N, YE L, ZHU Q W. Characteristics and

- processing method of abnormal data clusters caused by wind curtailments in wind farms[J]. Automation of electric power systems, 2014, 38(21): 39-46.
- [7] 陈希孺. 变点统计分析简介[M]. 北京: 人民教育出版社, 1979: 55-58.
- CHEN X R. Introduction to statistical analysis of change-point[M]. Beijing: People's Education Press, 1979: 55-58.
- [8] 王黎明. 变点统计分析的研究进展[J]. 统计研究, 2013(1): 50-51.
- WANG L M. Research progress of change-point statistical analysis[J]. Statistical research, 2013(1): 50-51.
- [9] PETTITT A N. A simple cumulative sum type statistic for the change-point problem with zero-one observations[J]. Biometrika, 1980, 67(1): 79-84.
- [10] HINKLEY D V. Inference about the change-point from cumulative sum tests[J]. Biometrika, 1971, 58(3): 509-523.
- [11] HINKLEY D, SCHECHTMAN E. Conditional bootstrap methods in the mean-shift model[J]. Biometrika, 1987, 74(1): 85-93.
- [12] LINDAHL S, HARMAN K, NGUYEN B. Analytical techniques for performance monitoring of modern wind turbines[C]//2012 EWEC, 2012: 16-19.

## WIND TURBINE DATA MINING ALGORITHM BASED ON CHANGE-POINT RESEARCH

Xu Jia<sup>1</sup>, Li Shaowu<sup>1</sup>, Wang Guisong<sup>2</sup>, Liu Ruihua<sup>1</sup>, Zhu Yaochun<sup>1</sup>

(1. Long Yuan(Beijing) Wind Power Engineering Technology Co., LTD., Beijing 100034, China;

2. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

**Abstract:** In the wind turbine operation, faults lead to turbine state changes, which occur in a continuous time series. To find out the change timing is of great value to fault backtracking and root cause analysis. We studied the characteristics of time series changes of wind power signals and states, and introduced a statistics algorithm Change-Point, which solved the uncertainty of singularities by calculating the confidence levels in different divided confidence intervals. We carried out experiments to verify the algorithm and concluded that the Change-Point algorithm can effectively mine the changes of the one-dimensional and two-dimensional model data in the historical data and illustrate the change points. The idea of Change-Point algorithm is to mine the regularity from the data, without limitation from other conditions. Therefore the algorithm can be widely applied to the fault backtracking in data mining of the system variables from wind turbines supervisory control and data acquisition (SCADA) system, and applied to rapidly positioning the SCADA data state change points.

**Keywords:** data mining; supervisory control and data acquisition (SCADA); wind turbine; Change-Point; confidence level