



一种基于 3σ 准则与 FCM 算法相结合的光伏电站直流侧故障定位方法

■ 于航¹ 刘阳¹ 王海政² 胡阳³ 朱红路^{2*}

(1. 龙源(北京)太阳能技术有限公司; 2. 华北电力大学可再生能源学院;

3. 华北电力大学控制与计算机工程学院)

摘要: 为了解决光伏电站故障组件定位困难的问题,提出了一种基于3倍标准差(3σ)准则与模糊C均值聚类(FCM)算法相结合的光伏电站直流侧故障定位方法,深入分析了光伏电站各组串电流数据之间的关联性与差异性,利用组串间的差异性结合 3σ 准则得到组串故障因子;然后利用故障因子的关联性,通过FCM算法得到其聚类中心,客观地找出故障因子的阈值;利用实际光伏电站的数据,实现了只借助组串电流数据定位光伏电站的直流侧故障。实验结果证明了该方法的可行性与准确性。

关键词: 光伏电站; 直流侧故障定位; 3σ 准则; FCM 算法

0 引言

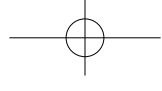
近年来,随着太阳能电池生产技术的提高和成本的降低,国内的光伏产业发展迅速,光伏电站装机容量不断增长。但是由于光伏组件长期在环境相对恶劣的户外运行,经常会出现各种故障问题,若不能及时排除故障,电站受其影响运行效率将会大幅降低^[1],因此,实时监控光伏阵列的运行状态,并及时发现故障、定位故障对光伏电站的安全运行十分重要。光伏发电功率密度低的特点导致光伏电站占地面积大、发电单元数量庞大且电站结构配置复杂,因此,光伏电站运行过程中产生的数据量庞大,如何从海量的运行数据中快速发现光伏电站运行过程中的故障,就成为光伏电站故障诊断中亟待解决的问题。

目前,光伏电站的故障检测受到了国内外学者的广泛关注^[2]。文献[3-4]提出,正常工作状态与故障状态的光伏组件之间存在温差,因此,可利用其红外图像的差异检测出故障组件。但该方法检测精度低、设备成本高、实时性差,难以运用到实际光伏系统中。通过在光伏阵列布置大量传感设备,以对传感设备的电信号进行分析来定位故障组件是一种有效的途径^[5]。但是此方法中光伏系统的安装过程复杂,对于大型光伏电站而言,所需要的传感器数量庞大,大幅增加了系统成本。文献[6]利用功率、电压和填充因子这3个指标对组件故障进行在线诊断。文献[7]也提出了以实测的直流侧电流、电压、功率值与模型得到的理论值之间的差值作为故障诊断依据。

收稿日期: 2018-04-16

基金项目: 国家重点研发计划项目(2017YFB0902100)

通信作者: 朱红路(1982—),男,博士、讲师,主要从事风能及太阳能利用技术的研究及应用。hongluzhu@ncepu.edu.cn



虽然该类方法可以通过偏差分析确定故障的发生,但是阈值设置不客观;同时电站性能退化等因素也会使阈值需要频繁校正。还有学者利用BP神经网络及粒子群算法等智能算法对光伏系统进行故障诊断^[8-9],但该类方法故障诊断模型需要故障样本训练,存在故障样本获取困难及模型泛化能力差的缺点。

针对光伏电站故障组串定位困难、故障诊断过程成本高、故障误判率高的问题,本文提出了一种基于 3σ 准则与FCM算法相结合的光伏电站直流侧故障诊断方法。首先,根据光伏电站组串电流信号的统计特性,基于 3σ 准则对组串的电

1 光伏电站运行数据特性分析

本文使用的数据样本采集于华东平原地区的某40 MW光伏发电系统。该光伏系统包括74台逆变器和553个汇流箱,每个汇流箱包含16个光伏组串,每个组串由16块光伏组件串联组成。由数据采集与监视控制(SCADA)系统对组串的电

流数据进行采集,但是组串运行数据受天气、逆变器最大功率跟踪效果及组串的实际性能等因素的影响波动剧烈,海量数据的并发及噪声的干扰给光伏电站性能分析和故障诊断带来极大难度。

对光伏电站组串、汇流箱、逆变器3个层级的运行数据进行比较,每个层级随机选取了3个组串在5天内的电流数据。图1a为同一汇流箱的3个不同组串

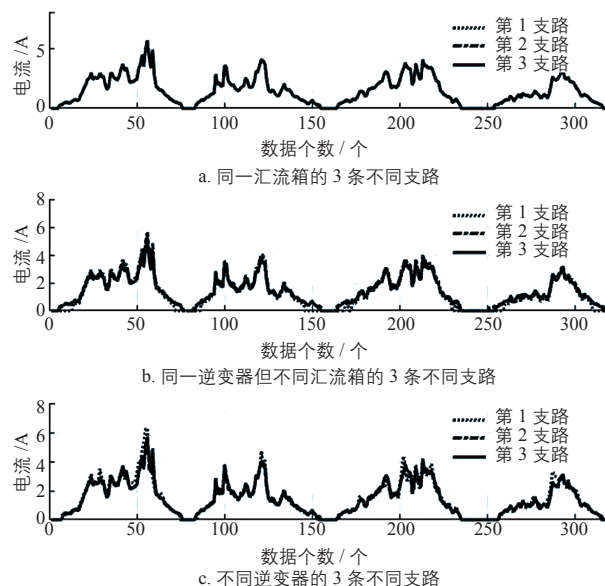


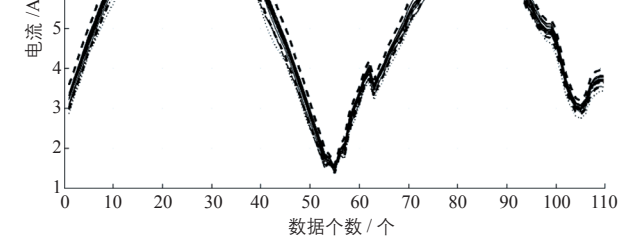
图1 电站不同层级的3条支路电流数据

相似度。此外,由于组串间的数据相似度较高,一些正常组串的电

流数据与故障组串的十分相近,无明显差异特征,故障辨识困难;即便出现了细微的差异也难以确定组串是否发生故障。

光伏电站不同组串之间的出力紧密耦合,外部因素导致的光伏出力的波动会带来基于偏差分析的故障诊断算法的误判^[7]。而这些误判多由短暂的局部阴影、天气波动等可恢复故障造成,但频繁的误判会掩盖组件的短路、开路、持续的阴影遮挡等不可恢复故障的判断。

为说明常规偏差分析带来的误判现象,从某一汇流箱中随机选取2天的组串电流数据(其中包含1个故障组串)进行分析,如图2所示。



备注: 每隔10 min 收集一次数据
图2 对同一汇流箱的不同组串进行故障辨识



2 天中出现了持续的组串电流过低, 表现为不可恢复的故障; 而误判组串只是在 2 个时间段出现了电流低于正常组串的情况, 除此之外的电流数据都保持在正常范围内, 属于可自行恢复的故障。

2 算法的提出

为实现光伏电站的故障快速定位并避免误判现象的发生, 本文首先基于 3σ 准则对组串的电

2.1 3σ 准则

假设数据样本符合正态或近似正态分布, 并且这组数据只含有随机误差, 对其进行计算后得到样本均值 \bar{X} 与标准偏差 σ , 进而得出概率区间 $(\bar{X}-3\sigma, \bar{X}+3\sigma)$, 超过这个区间的数据为异常数据^[10]。 3σ 准则可表示为:

$$3\sigma = 3\sqrt{\frac{\sum_{i=1}^n (\bar{X}_i - \bar{X})^2}{n}} \quad (1)$$

式中, X_i 为第 i 个数据样本; n 为数据样本个数。

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad (2)$$

$$X_i = \begin{cases} \text{正常}, & X_i \in (\bar{X}-3\sigma, \bar{X}+3\sigma) \\ \text{异常}, & X_i \notin (\bar{X}-3\sigma, \bar{X}+3\sigma) \end{cases} \quad (3)$$

当光伏电站中的组串处于正常运行状态时, 该组串与同一汇流箱的其他正常组串数据特征基本一致。此时, 组串每个时刻的运行数据基本都在 3σ 的概率区间内; 当组串发生故障时, 故障组串与同一汇流箱中的正常组串的运行状态相比会出现较大差异, 运行数据超出 3σ 的概率区间。因此, 可对同一汇流箱各组串的电

2.2 FCM 算法

FCM 算法作为无监督机器学习的主要技术之一, 是用模糊理论对重要数据分析和建模的方

法。它通过优化目标函数得到每个样本点对所有类中心的隶属度, 决定样本点的类属, 以达到自动对样本数据进行分类的目的, 能较为客观地反映数据的分布特征^[12]。

FCM 算法将 n 个向量 $x_i (i=1, 2, \dots, n)$ 分为 c 个模糊组, 求出每组的聚类中心, 使得非相似性指标的价值函数达到最小。FCM 算法采用模糊划分, 根据每个给定数据点值在 $[0, 1]$ 间的隶属度来确定其属于各个组的程度。与引入模糊划分相适应, 隶属矩阵 U 允许有取值在 $[0, 1]$ 间的元素。不过, 加上归一化规定, 一个数据集的隶属度 $u_{i,j}$ 的总和等于 1^[13]。

$$\sum_{j=1}^c u_{i,j} = 1 \quad (j=1, \dots, n) \quad (4)$$

式中, $u_{i,j} \in [0, 1]$ 。

FCM 算法的目标函数为:

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^n J_i = \sum_{i=1}^n \sum_{j=1}^c d_{i,j}^2 u_{i,j}^m \quad (5)$$

式中, J 是 FCM 的目标函数; c_i 为模糊组 I 的聚类中心; $d_{i,j} = \|c_i - X_j\|$ 为第 I 个聚类中心与第 j 个数据点间的欧几里德距离; m 为一个加权指数, $m \in [1, \infty)$ 。

构造如下新的目标函数, 可求得使式 (5) 达到最小值的必要条件:

$$J(U, c_1, \dots, c_c, \lambda_1, \dots, \lambda_n) = \sum_{i=1}^n J_i + \sum_{j=1}^n \lambda_j (\sum_{i=1}^c u_{i,j} - 1) \quad (6)$$

式中, λ_i 为式 (5) 中 n 个约束式的拉格朗日乘子。对所有输入参量求导, 使式 (5) 达到最小的必要条件为:

$$c_i = \frac{\sum_{j=1}^n X_j u_{i,j}^m}{\sum_{j=1}^n u_{i,j}^m} \quad (7)$$

$$u_{i,j} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{i,j}}{d_{k,j}} \right)^{\frac{2}{m-1}}} \quad (8)$$

式中, $d_{k,j} = \|c_k - X_j\|$ 为第 K 个聚类中心与第 j 个数据点间的欧几里德距离。

由上述两个必要条件可知, FCM 算法的运算过程是一个简单的迭代过程。

2.3 基于 3σ 准则和 FCM 算法的光伏故障定位方法

本文提出的故障诊断方法流程如图 3 所示。

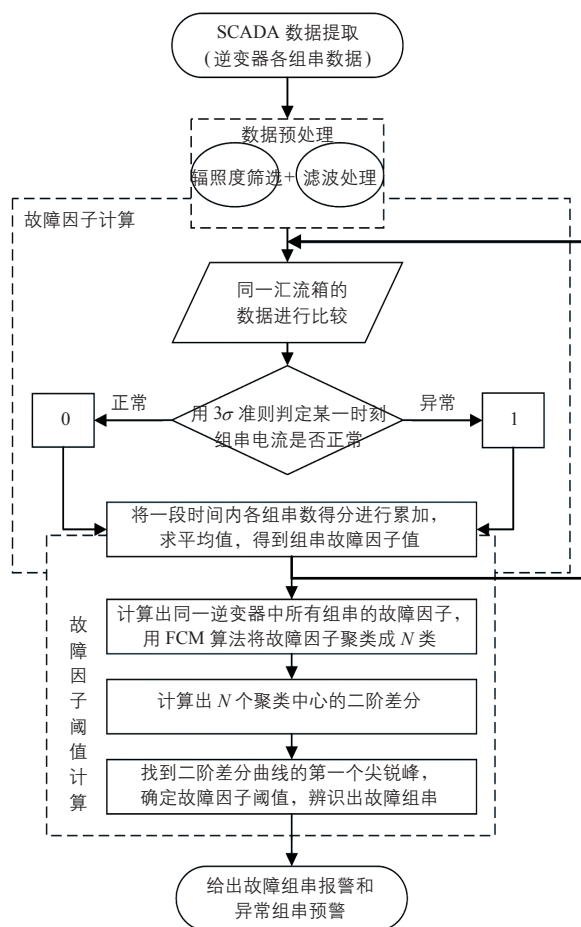
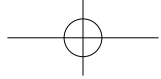


图 3 故障诊断流程图

2.3.1 统计特性分析及故障因子的提取

首先对采集到的组串电流数据进行筛选和平滑滤波等数据预处理。

基于 3σ 准则, 正常组串间的数据特征基本一致, 组串的运行数据落在 3σ 的概率区间内; 而故障组串的运行数据则落在 3σ 的概率区间外。

$$I_{i,j} = \begin{cases} \text{正常, } I_{i,j} \in (\bar{I}_j - 3\sigma, \bar{I}_j + 3\sigma) \\ \text{异常, } I_{i,j} \notin (\bar{I}_j - 3\sigma, \bar{I}_j + 3\sigma) \end{cases} \quad (9)$$

式中, $I_{i,j}$ 为组串 j 在 i 时刻的电流数据; \bar{I}_j 为 j 时刻各组串的电流平均值。

对各组串一段时间的电流数据进行数学统计, 并利用故障因子对组串的故障程度进行量化表征:

$$K_j = \frac{\sum_{i=1}^M f_i}{M} \quad (10)$$

式中, K_j 为第 j 个组串的故障因子; M 为该

组串电流数据的个数; f_i 为表征组串电流状态的量。

f_i 可表示为:

$$f_i = \begin{cases} 1, I_{i,j} \text{ 异常} \\ 0, I_{i,j} \text{ 正常} \end{cases} \quad (11)$$

2.3.2 聚类过程与阈值的设定

得到各组串的故障因子后, 可诊断出存在故障倾向的组串。为了确定故障组串的故障因子阈值, 本文考虑了同一逆变器范围内所有组串的故障因子。故障组串的故障因子之间存在一定的相似性, 而与正常组串之间存在差异性, 据此利用 FCM 算法将所有故障因子 (大于 100 个) 划分为 N 个群集。

对得到的 N 个聚类中心进行二阶差分计算, 进一步增强故障组串与正常组串的差异性, 这时得到的差分曲线的第一个峰对应的值即故障组串的故障因子阈值。故障因子高于阈值的组串被诊断为故障组串。

$$Result = \begin{cases} \text{故障, } K_j > \delta_1 \\ \text{预警, } \delta_2 \leq K_j \leq \delta_1 \\ \text{正常, } K_j < \delta_2 \end{cases} \quad (12)$$

式中, δ_1 为二阶差分曲线第一个尖锐峰处的故障因子; δ_2 为紧邻 δ_1 左侧的故障因子。

3 实际算例分析

利用实际的光伏电站数据对本文方法进行验证。此处选取了一个包含有 112 个组串的逆变器数据, 该逆变器共连接了 7 个汇流箱, 每个汇流箱包含 16 个组串。根据电站的故障维修记录显示, 该逆变器中的 1# 汇流箱的第 1 支路出现了组件破碎的故障, 并于 11 月 11 日进行了组件更换, 排除了故障。选取了 10 月 24-26 日的数据, 作为待辨识的逆变器数据样本, 并利用以上的数据筛选规则, 对电站各组串的电流数据进行了预处理。

根据第一步的故障辨识, 利用 3σ 准则对同一汇流箱的不同组串进行对比, 计算出同一汇流箱的各组串故障因子: 组串的故障因子值越大, 出现故障的可能性越大。图 4 为两个不同的逆

变器中的 32 个组串的故障因子在 3 天 (10 月 24-26 日) 内的值, 并辨识出 2 个潜在的故障组串。

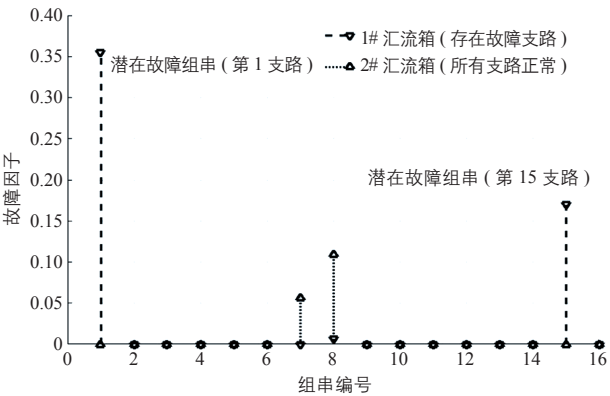


图 4 两个不同汇流箱的 32 个支路的故障因子

表 1 同一逆变器各汇流箱支路的故障因子

支路	1# 汇流箱	2# 汇流箱	3# 汇流箱	4# 汇流箱	5# 汇流箱	6# 汇流箱	7# 汇流箱
1	0.3546	0	0	0	0	0.1305	0
2	0	0	0	0	0	0.0284	0
3	0	0	0	0	0	0	0
4	0	0	0.0780	0	0	0	0
5	0	0	0	0	0	0	0.0355
6	0	0	0	0	0	0	0
7	0	0.0567	0	0	0	0	0
8	0.0071	0.0890	0	0	0	0	0.0284
9	0	0	0	0.0645	0.5667	0.07094	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0.0631
12	0	0	0.0142	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0.1702	0	0.0709	0	0	0	0.0426
16	0	0	0	0	0	0	0

表 2 故障因子二阶差分表

故障因子聚类中心	一阶差分	二阶差分
0.0001	—	—
0.0272	0.0271	—
0.0395	0.0123	-0.0148
0.0620	0.0225	0.0102
0.0729	0.0109	-0.0116
0.1089	0.0360	0.0251
0.1305	0.0216	-0.0144
0.1702	0.0397	0.0181
0.3546	0.1844	0.1447
0.5667	0.2121	0.0277

之后根据以上的方法计算出了整个逆变器共 112 条组串的故障因子值。其中, 5# 汇流箱第 9 支路故障因子的值最高, 为 0.5667, 其次为 1# 汇流箱第 1 支路; 此外, 故障因子较大的还有 1# 汇流箱第 15 支路。逆变器各支路故障因子值如表 1 所示。

根据第二层级的故障辨识, 将第一步所得的所有组串故障因子值进行聚类, 利用各聚类中心的二阶差分曲线的第一个尖锐峰的位置, 该位置对应的故障因子值的大小即为故障因子的阈值。表 2 为故障因子二阶差分表。

通过表 2 中的二阶差分结果可以发现, 正常情况下故障因子的二阶差分变化不大, 若出现突变点, 则该点所对应的值就是故障因子的阈值。

从图 5 可以看出, 故障因子聚类中心二阶差分的第一个尖锐峰对应的位置为第 9 个聚类中心, 该处的故障因子大小为 0.3546, 所以故障因子大于等于 0.3546 的组串为故障组串; 所以结合表 1 的各组串故障因子得出结论为, 1# 汇流箱的第 1 支路和 5# 汇流箱的第 9 支路为故障支路; 同时 1# 汇流箱的第 15 支路的故障因子与其他正常支路相比和阈值最为接近, 因此该支路为预警

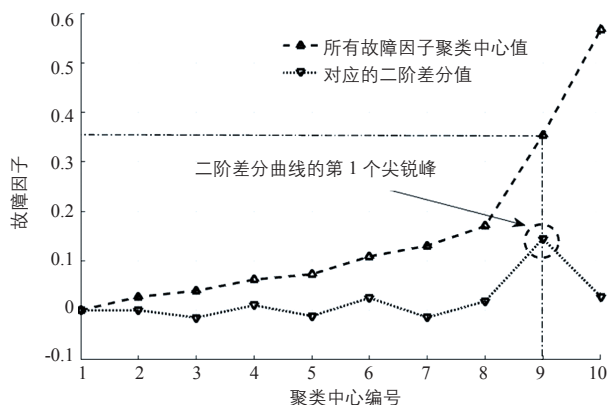
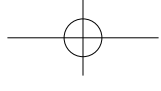


图5 对故障因子进行聚类并计算二阶差分

状态。

以上诊断结果与现实电站的实际故障状态一致。

4 结论

本文提出的基于 3σ 准则与FCM算法的直流侧故障辨识方法,通过对实际光伏电站的SCADA数据的全面分析,能准确地对光伏电站直流侧故障组串进行定位;同时基于统计分析方法,利用故障因子对电站实际运行数据的分布特征进行定量描述;并利用FCM算法对同一逆变器的组串故障因子进行聚类,通过对其进行差分处理客观地设定了故障因子阈值。使用该方法实现了对实际的40 MW光伏系统的故障组串进行评估,实验结果显示,该方法可以有效地检测出故障组串,并提供异常组串的预警报告;而且该方法原理简单、有效,不需要复杂的计算平台,便

于推广到其他光伏系统。由于数据的局限性,文中所提出的方法的准确性和普适性有待验证;其次对于光伏电站的故障辨识的故障类型诊断方法有待深入研究。

参考文献

- [1] 杨吉洲. 太阳能光伏发电技术应用中的相关问题探讨[J]. 水能经济, 2016, (1): 61.
- [2] 王元章, 李智华, 吴春华. 光伏系统故障诊断方法综述[J]. 电源技术, 2013, 37(9): 1700 - 1705.
- [3] 王培珍, 沈玉樑, 杨维翰. 太阳光伏阵列的温度与红外特性分析[J]. 太阳能学报, 2005, 26(1): 82 - 85.
- [4] 王培珍, 郑诗程. 基于红外图像的太阳能光伏阵列故障分析[J]. 太阳能学报, 2010, 31(2): 197 - 202.
- [5] 胡义华, 陈昊, 徐瑞东, 等. 基于最优传感器配置的光伏阵列故障诊断[J]. 中国电机工程学报, 2011, 31(33): 19 - 30.
- [6] 陈宇航, 闫腾飞, 谢添, 等. 基于功率损失和UI特性综合考虑的光伏组件故障诊断方法[J]. 电机与控制应用, 2016, 43(11): 92 - 97.
- [7] Harrou F, Sun Y, Taghezouit B, et al. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches[J]. Renewable Energy, 2017, (116): 23 - 27.
- [8] 王元章, 李智华, 吴春华, 等. 基于BP神经网络的光伏组件在线故障诊断[J]. 电网技术, 2013, 37(8): 2094 - 2100.
- [9] 尹春雨, 孙凤杰, 范杰清, 等. 基于粒子群优化小波神经网络的光伏组件故障诊断方法[J]. 计算机应用, 2017, 37(s1): 106 - 108.
- [10] 欧阳琳男. 粗大误差判断准则运用条件的相关分析[J]. 中国计量, 2017, (11): 106 - 107.
- [11] Liu Q, Zhao Y, Zhang Y, et al. Hierarchical Context-Aware Anomaly Diagnosis in Large-Scale PV Systems Using SCADA Data[A]. International Conference on Industrial Informatics[C]. Indin, 2017.
- [12] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48 - 61.
- [13] 孙晓霞, 刘晓霞, 谢倩茹. 模糊C-均值(FCM)聚类算法的实现[J]. 计算机应用与软件, 2008, 25(3): 48 - 50. 太阳能

(接第60页)

- [4] 汪宁渤, 马彦宏, 王建东. 大规模风电集中并网对电力系统安全稳定的影响[J]. 电力建设, 2011, 32(11): 77 - 80.
- [5] 姚伟, 程时杰, 文劲宇. 直流输电技术在海上风电场并网中的应用[J]. 中国电力, 2007, 40(10): 70 - 74.
- [6] 王锡凡, 卫晓辉, 宁联辉, 等. 海上风电并网与输送方案比较[J]. 中国电机工程学报, 2014, 34(31): 5459 - 5466.
- [7] 汤广福. 基于电压源换流器的高压直流输电技术[M]. 北京: 中国电力出版社, 2010.
- [8] Breuer W, Christl N. 大容量陆地和海上风电场的电网接入方案[J]. 中国电力, 2007, 40(3): 74 - 78.
- [9] 陈霞. 基于多端直流输电的风电并网技术研究[D]. 武汉: 华中科技大学, 2012.

- [10] Thomes Worzyk. Submarine Power Cables: Design, Installation, Repair, Environmental Aspects [M]. West Berlin, Germany: Springer Publishing Company, Incorporated, 2009.
- [11] 王志新, 王承民, 艾芊, 等. 近海风电场关键技术[J]. 华东电力, 2007, 35(2): 37 - 41.
- [12] 张建民, 谢书鸿. 海上风电场电力传输与海底电缆的选择[J]. 电气工程学报, 2011, (11): 33 - 35.
- [13] ABB. XLPE submarine cable systems - attachment to XLPE cable systems user's guide[EB/OL]. www.abb.com/cables, 2018-04-01.
- [14] ABB. It's time to connect - Technical description of HVDC light technology[R]. Sweden, 2005. 太阳能