

DOI:10.12158/j.2096-3203.2020.05.017

基于 EKF 算法的分布式光伏发电异常数据排查技术

左松林¹, 陈伟¹, 付真斌¹, 赵骞¹, 江再玉², 郑昕昕³

(1. 国网安徽省电力有限公司电力科学研究院, 安徽 合肥 230009;

2. 国网信通产业集团北京中电普华信息技术有限公司, 北京 100085;

3. 合肥工业大学智能制造技术研究院, 安徽 合肥 230009)

摘要:分布式光伏发电运营过程中,设备故障、仪表测量误差、用户个人行为等干扰因素会导致运营系统采集异常数据,因此对异常数据进行排查,有助于保障光伏发电用户数据库的准确性和可靠性,提高存在运营问题的分布式光伏用户的识别度。文中针对分布式光伏系统的特殊性,提出一种光伏发电运营系统异常数据排查技术,结合温度、辐射量、纬度、时令等环境因素,采用中心复合设计方法,通过有限的的数据量建立较为精确的发电量数学模型,采用扩展卡尔曼滤波算法对采集到的发电量进行修正,从而排查和消除异常数据。该方法能预测发电数据,快速、可靠地排查异常数据,实验结果验证了该方法的有效性。

关键词:分布式光伏;发电量模型;环境因素;异常数据识别;扩展卡尔曼滤波

中图分类号:TM92

文献标志码:A

文章编号:2096-3203(2020)05-0120-06

0 引言

分布式光伏运营系统是为光伏发电用户提供服务和管理的核心系统,其对光伏发电系统进行运营监测和统计分析,实现对大量分布式光伏发电用户的高效管理,是分布式发电系统的重要组成部分^[1-3]。目前,以美国为首的发达国家已大量建设规模化分布式光伏发电系统,我国亦大力发展分布式光伏发电,因此需配套可靠、高效的分布式光伏运营系统。运营过程中,不可避免地会因各种干扰因素而产生异常数据,从而导致一系列问题。首先,作为光伏发电大数据平台的组成部分,采集到的异常数据会对用户数据库的准确性和可靠性产生干扰^[4-6];其次,部分用户个人行为导致的异常数据若不能被及时发现并排除,将会造成严重的经济损失,危害电网安全^[7-9]。因此,对光伏发电运营系统异常数据进行排查与消除,对维护分布式光伏发电系统的安全性、稳定性和经济性有着重要意义。

针对光伏发电异常数据的识别需要立足于待修复电站自身数据的统计分析,包括聚类算法、神经网络法、残差检测法等^[10-11]。文献[12]提出一种光伏电站输出功率修复办法,通过采集数据的相关性进行识别,提高了异常数据识别的可信度。文献[13]提出光伏功率高比例异常数据机器识别算法,建立了描述辐照度与光伏功率间相关关系的概

率功率曲线模型,能够适应高比例异常数据条件。这些方法均针对大型光伏电站,能够有效利用集中式大规模光伏电站的海量数据进行有效处理。相对于大型光伏电站,分布式光伏发电系统具有一定的特殊性,其具有各自独立的用户端,每个用户端都需要进行数据排查,而单个用户端数据量极其有限;此外,用户端行为具有一定的随机性,对各用户端数据检测的实时性要求高。因此,亟需开发分布式光伏运营系统异常数据排查技术。

文中提出基于扩展卡尔曼滤波(extended Kalman filter, EKF)算法的光伏发电运营系统异常数据排查技术,基于光伏系统环境因素和并网数据,通过中心复合设计(central composite designs, CCD)对光伏系统发电量进行建模,利用有限的的数据量建立较为精确的光伏发电量模型。在此基础上,采用EKF算法对系统采集到的发电量数据进行校核,检测偏差值,排查异常情况,从而判断用户行为,修正异常数据,保障分布式光伏运行系统的工作可靠性,结果表明该方法能有效检测并修复异常数据。

1 光伏发电量模型的建立

1.1 数据拟合方法

文中以光伏板斜面太阳辐射量为输出量,温度和发电量为输入量,采用CCD方法确定输入输出关系^[14],同时考虑纬度、光伏板安装倾角、时令等环境因素的影响,得到可直接测量的地表水平辐射量和光伏发电量的定量关系,从而建立光伏发电量模型。对于特定型号的光伏板,CCD设计输出和输入

收稿日期:2020-03-13;修回日期:2020-04-21

基金项目:国家电网有限公司科技项目(B3680117111400Z-R000000);安徽省科技重大专项(17030901061)

关系具有通用性,当其应用在分布式发电系统中时,光伏发电量的差异性由环境因素决定,从而提高模型对不同环境的适应性和准确性。CCD 输入变量边界条件如图 1 所示, T 为环境温度,其范围为 $[T_{\min}, T_{\max}]$; Q 为光伏板发电量,其范围为 $[Q_{\min}, Q_{\max}]$ 。图中 A, B, C, D 为边界点, E, F, M, N 为轴向点, O 为中心点。

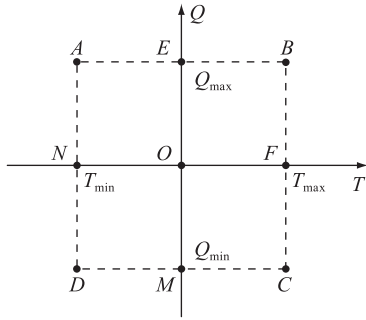


图 1 CCD 输入变量边界条件

Fig.1 Input variable boundary conditions of CCD

在试验设计中,采集边界点、轴向点和中心点对应的输入输出数据,将结果拟合表征光伏板斜面太阳辐射量 H_{bt} 和 Q, T 关系的二次曲面,并通过残差分布进行验证。当其符合六西格玛正态分布时,输入输出关系成立。根据某光伏电站测得数据,以春季为例,CCD 试验数据如表 1 所示。

表 1 CCD 试验数据

Table 1 Experimental data of CCD

点类型	$Q/(kW \cdot h)$	$T/^\circ C$	$H_{bt}/(kW \cdot m^2)$
边界点 A	1 114	-5.0	0.923 7
边界点 B	1 114	40.8	0.851 6
边界点 C	0	40.8	0
边界点 D	0	-5.0	0
轴向点 E	1 114	22.9	0.883 9
轴向点 F	557	40.8	0.708 3
轴向点 M	0	22.9	0
轴向点 N	557	-5.0	0.379 2
中心点 O	557	22.9	0.553 5
随机附加点	542	34.2	0.791 8
随机附加点	587	1.0	0.423 0
随机附加点	124	6.4	0.088 2

CCD 输入输出关系可表示为:

$$H_{bt} = C + a_1 Q + a_2 T + a_3 Q^2 + a_4 T^2 + a_5 QT \quad (1)$$

式(1)表征太阳辐射量 H_{bt} 为光伏板发电量 Q 和环境温度 T 的二次响应曲面,其中 C 为曲面与 H_{bt} 方向的交截值, $a_1 \sim a_5$ 为变量的系数,均为实数。根据表 1,可得响应残差分布如图 2 所示,可以看

出,其符合六西格玛正态分布,证明了响应曲面拟合的准确性。同理可以获得另外三季的光伏板斜面太阳辐射量 H_{bt} 和 Q, T 的关系。

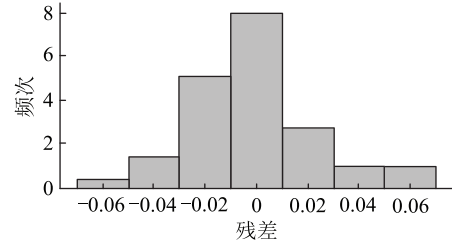


图 2 残差分布直方图

Fig.2 Residual distribution histogram

1.2 环境因素影响建模

对于分布式光伏发电用户,测量光伏板斜面太阳辐射量极大地增加了系统成本,可通过地表水平辐射量进行发电量建模,因此需要考虑实际安装环境的影响^[15-16]。图 3 为光伏板的安装环境,其中 H_b 为地表水平辐射量; φ 为光伏板安装地理纬度; β 为光伏阵列倾角; δ 为太阳赤纬角。

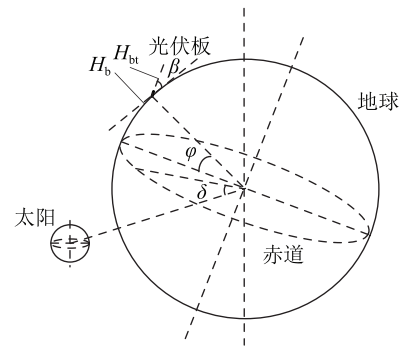


图 3 光伏板的安装环境

Fig.3 The installation environment of photovoltaic panels

H_{bt} 和 H_b 之间的关系可以表示为:

$$H_{bt} = H_b R_b = \frac{\cos(\varphi - \beta) \cos \delta \sin \omega'_s + K \omega'_s \sin(\varphi - \beta)}{\cos \varphi \cos \delta \sin \omega_s + K \omega_s \sin \varphi} \quad (2)$$

其中:

$$K = \frac{\pi}{180} \sin \delta \quad (3)$$

$$\begin{cases} \delta = 23.45 \sin\left(360 \times \frac{284 + n}{365}\right) \\ \omega_s = \arccos(-\tan \varphi \tan \delta) \\ \omega'_s = \min\{\omega_s, \arccos[-\tan(\varphi - \beta) \tan \delta]\} \end{cases} \quad (4)$$

式中: R_b 为斜面辐射量和水平辐射量之比; ω_s 为水平面日落角; ω'_s 为倾斜面日落角; n 为当日在一年中的序号。由此可得光伏发电量模型为:

$$H_b = \frac{C + a_1 Q + a_2 T + a_3 Q^2 + a_4 T^2 + a_5 QT}{R_b} \quad (5)$$

2 基于 EKF 算法的发电量预测

通过光伏逆变器工作瞬时有功功率 p , 可以计算得到用户端发电量, 即:

$$Q = \int p dt \quad (6)$$

p 受到干扰因素影响, 会导致异常数据, 采用 EKF 算法能够对异常数据进行排查与消除^[17-18], 根据式(6)可得光伏发电系统状态方程为:

$$x_{k+1} = x_k + p_k \Delta t + w_k \quad (7)$$

根据式(5), 得到系统观测方程为:

$$y_{k+1} = \frac{C}{R_b} + \frac{a_1}{R_b} x_k + \frac{a_3}{R_b} x_k^2 + \frac{a_5}{R_b} x_k T_k + \frac{a_2}{R_b} T_k + \frac{a_4}{R_b} T_k^2 + v_k \quad (8)$$

式中: w_k, v_k 分别为系统噪声和量测噪声; T_k 为该次迭代实时温度。设定 R_w 和 R_v 为两者的协方差矩阵, 系统先验估计均方误差为 P_k^- , 最优估计均方误差为 P_k^+ , 系统状态矩阵 A_k 和系统测量矩阵 C_k 分别定义为系统状态方程和观测方程对 x_k 的偏导数, 定义动态特性匹配系数 B 为状态方程对 p_k 的偏导数, 则 EKF 迭代过程如下:

(1) $k = 0$ 时, 系统状态量和估计误差初始值为:

$$\begin{cases} x_0 = Q(t_0) \\ P_0^+ = \text{var}(x_0) \end{cases} \quad (9)$$

(2) $k = 1, 2, \dots$ 时, 状态量时间更新, 误差协方差时间更新:

$$\begin{cases} \hat{x}_{k+1}^- = A_k \hat{x}_k^- + B P_k \\ P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + R_w \end{cases} \quad (10)$$

(3) 计算卡尔曼滤波增益:

$$L_k = P_k^- C_k^T [C_k P_k^- C_k^T + R_v]^{-1} \quad (11)$$

(4) 状态估计观测, 误差协方差观测更新:

$$\begin{cases} \hat{x}_k^+ = \hat{x}_k^- + L_k (y_{k+1} - C_k \hat{x}_k^-) \\ P_k^+ = (I_k - L_k C_k) P_k^- \end{cases} \quad (12)$$

图 4 为 EKF 预测流程, 确定初始值后, 进行状态量时间更新和误差协方差时间更新, 根据计算的卡尔曼滤波增益进行状态估计观测更新和误差协方差观测更新。通过算法循环迭代, 即可实现分布式光伏发电量的预测, 与用户端采集发电量进行对比, 即可进行异常数据的排查与消除。

通过计算得到发电量预测值后, 可以推导出发电功率预测值, 根据式(6), 发电量为瞬时功率的积分, 则发电功率为发电量对时间的微分。因此, 发电功率与发电量的关系在基于数字控制器的离散

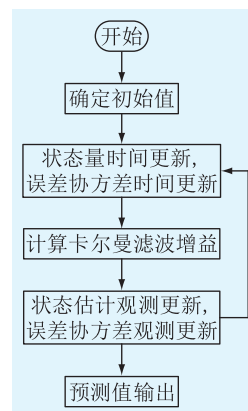


图 4 EKF 预测流程

Fig.4 EKF forecasting process

系统中可以表示为:

$$p = \Delta Q / \Delta t \quad (13)$$

式中: ΔQ 为相邻 2 次发电量的预测结果之差; Δt 为预测时间间隔。根据式(7), 每次迭代更新的 x_k 即为发电量预测结果, 故式(13)可以改写为:

$$p_k = (x_k - x_{k-1}) / \Delta t \quad (14)$$

当计算得到发电量预测结果 x_k 后, 通过式(14)即可求得发电功率的预测结果。

3 异常数据识别方法

图 5 为异常数据识别流程, 首先根据式(5)得到光伏发电量模型, 由此得到 EKF 算法的状态方程和观测方程, 通过 EKF 算法进行发电量预测。图中 ε 为发电量用户采集数据和 EKF 滤波结果的误差值, ε_{\max} 为误差值上限, 当 ε 未超过上限时, 输出行为一, 当 ε 超过上限时, 需要进一步判断。 T 为异常数据持续时间, T_{\max} 为异常数据持续时间上限。当 T 未超过上限时, 输出行为二; 当 T 超出上限时, 识别过程结束, 输出行为三。可以看出, 当 ε 或 T 超过上限时, 均认为数据异常。识别过程中, 需要针对不同情况做出相应行为, 图 5 中给出了 3 种行为, 具体如下: (1) 行为一。用户端数据正常, 向数据库上传用户端采集数据; (2) 行为二。用户端数据暂时异常, 向数据库上传 EKF 修正结果替换用户端采集数据; (3) 行为三。用户端数据长期异常, 需要排查用户异常行为。

根据图 5 即可实现光伏发电运营系统异常数据的排查与消除, 若异常数据发生后短时间内又恢复正常, 则仅需将 EKF 修正结果替代采集结果上传数据库, 从而消除异常数据; 若异常数据长时间发生, 则有可能是发电系统故障或用户主动行为, 需要对用户异常行为进行排查。

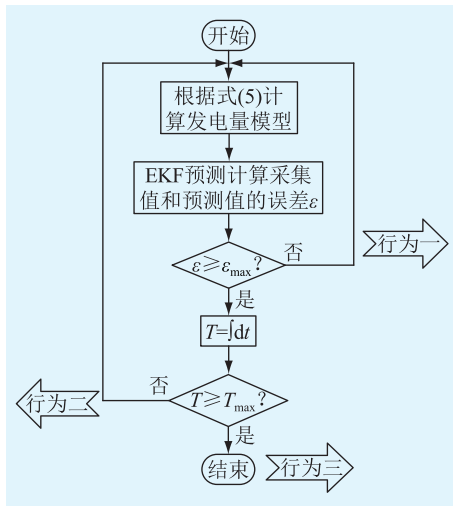


图5 异常数据识别流程

Fig.5 Abnormal data identification process

4 异常数据排查的实验验证

利用某地分布式光伏发电运营系统采集数据所提异常数据排查技术进行验证,选取该地 2018 年 12 月份共 31 d 数据,图 6 为发电量的采集数据和 EKF 预测结果对比。可以看出,采集值和预测值误差均在 5% 以内,证明了预测结果的准确性。

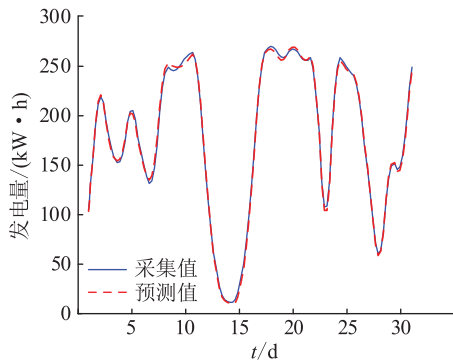


图6 未发生干预时发电量的采集和预测值对比

Fig.6 Comparison of collection and predicted generated energy before intervention

图 7 为对数据进行短期干预后发电量的采集值和预测值对比,虚线框内是干预后的情况,当干预消除后,预测值能够再次跟踪采集值,则干预发生时段的数据可由预测值替代。图 8 为对数据进行长期干预后发电量的采集值和预测值对比,虚线框内是干预后的情况,当长时间预测值无法跟踪实际采集值时,需要排查用户异常行为。

图 7 和图 8 的采集值分别是对图 6 中的实际数据进行了人为的短期和长期干预后,由光伏发电运营系统采集到的值,即 3 个图的实际值相同,但采集值不同。可见,当以天为单位进行预测时,长期预

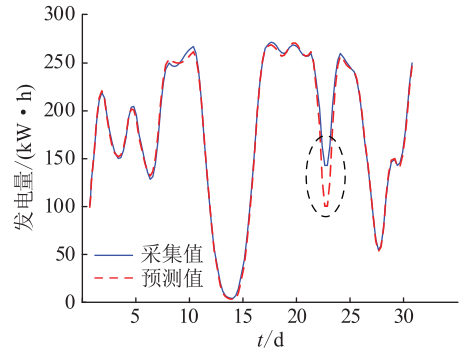


图7 短期干预时发电量的采集值和预测值对比

Fig.7 Comparison of collection and predicted generated energy with short-term intervention

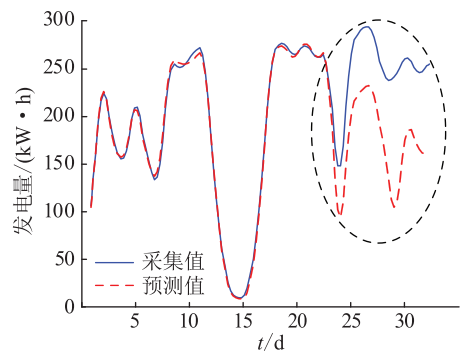
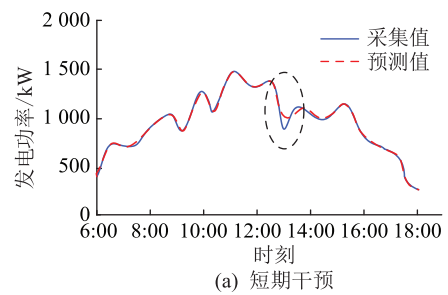


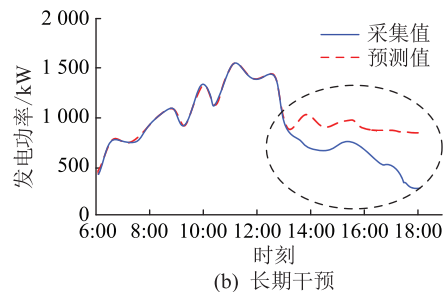
图8 长期干预时发电量的采集值和预测值对比

Fig.8 Comparison of collection and predicted generated energy with long-term intervention

测的数据异常判断较为滞后,上级电网的实用价值受到限制,故在此基础上缩小时间单位,将按天排查改为按小时排查,图 9 和图 10 分别为该地 2018 年 12 月 7 日(晴天)和 2018 年 12 月 15 日(阴天) 06:00—18:00 发电功率的采集值和 EKF 预测值。



(a) 短期干预



(b) 长期干预

图9 2018 年 12 月 7 日发电功率的采集值和预测值对比

Fig.9 Comparison of collection and predicted generated power on December 7, 2018

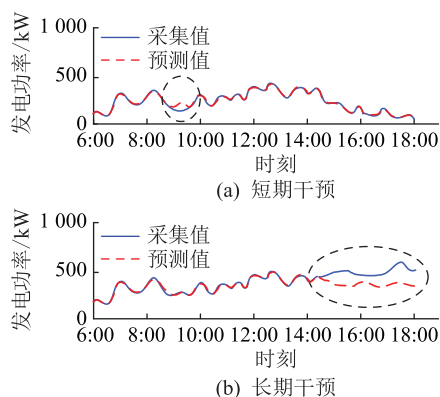


图 10 2018 年 12 月 15 日发电功率的采集值和预测值对比
Fig.10 Comparison of collection and predicted generated power on December 15, 2018

可以看出,缩小时间单位后,异常情况能够在较短的时间内被检测出来,且在不同的天气工况下均适用,验证了该方法的有效性。

5 结论

光伏发电运营系统运营过程中可能采集到异常数据,文中结合环境因素的影响对发电量进行建模,采用 EKF 算法进行发电量预估,从而实现对异常数据的排查,最后通过实验验证了该方法的有效性,为分布式光伏发电运营系统智能化、网联化发展发现提供了有力支持,具体创新点如下:

(1) 通过 CCD 试验设计给出了光伏板斜面太阳辐射量、温度和发电量的定量关系,有效利用有限的数据库建立较为精确的模型,尤其适用于多个独立客户端的分布式发电系统;

(2) 考虑了环境因素的影响,引入安装纬度、光伏阵列倾角、时令等变量,发电量模型的建立方法具有普适性,只需根据实际用户所处的环境参数进行调整,即可得到通用发电量模型;

(3) 以瞬时有功功率积分项为状态方程,发电量模型为观测方程,通过 EKF 算法实现了发电量的预测,并根据预测结果做出相应的行为以排查和消除异常数据,能够有效防止因干扰因素导致误判,检测可靠性高。

参考文献:

[1] VONMEIER A, STEWART E, MCEACHERN A, et al. Precision micro-synchrophasors for distribution systems: a summary of applications [J]. IEEE Transactions on Smart Grid, 2017, 8(6): 2926-2936.

[2] 任明珠, 管晟超, 谭佳, 等. 监测点优化配置下考虑谐波约束的多节点分布式光伏最大配置方案研究 [J]. 电力系统保护与控制, 2018, 46(8): 73-79.

REN Mingzhu, GUAN Shengchao, TAN Jia, et al. Configuration

analysis of multi-node distributed photovoltaic generation considering harmonic constraints under optimal monitoring program [J]. Power System Protection and Control, 2018, 46(8): 73-79.

[3] 宋森, 王克文. 主动配电网中考虑空间与时间相关性的分布式光伏并网规划 [J]. 电力系统保护与控制, 2019, 47(5): 106-114.

SONG Miao, WANG Kewen. Planning of distributed PV grid-access considering the space and time correlation in active distribution network [J]. Power System Protection and Control, 2019, 47(5): 106-114.

[4] PLATON R, MARTEL J, WOODRUFF N, et al. Online fault detection in PV systems [J]. IEEE Transactions on Sustainable Energy, 2015, 6(4): 1200-1207.

[5] 杨茂, 黄鑫. 基于光照过程特征分析的光伏功率异常数据识别算法 [J]. 电力系统自动化, 2019, 43(6): 64-74.

YANG Mao, HUANG Xin. Abnormal data identification algorithm for photovoltaic power based on characteristics analysis of illumination process [J]. Automation of Electric Power Systems, 2019, 43(6): 64-74.

[6] 杨茂, 孟玲建, 李大勇, 等. 基于类 3σ 准则的光伏功率异常数据识别 [J]. 可再生能源, 2018, 36(10): 25-30.

YANG Mao, MENG Lingjian, LI Dayong, et al. Identification of abnormal data of photovoltaic power based on class 3σ [J]. Renewable Energy Resources, 2018, 36(10): 25-30.

[7] 郑征, 刘刚, 张琳娟, 等. 基于 LSSVM 的光伏发电三层筛选窃电识别方法 [J]. 电力电子技术, 2017, 51(10): 30-32.

ZHENG Zheng, LIU Gang, ZHANG Linjuan, et al. Photovoltaic generation three-layer electricity stealing recognition method based on LSSVM [J]. Power Electronics, 2017, 51(10): 30-32.

[8] 沈才良. 光伏电站故障监测系统的设计与实现 [D]. 哈尔滨: 哈尔滨工业大学, 2015.

SHENG Cailiang. Design and implementation of photovoltaic power station fault monitoring system [D]. Harbin: Harbin Institute of Technology, 2015.

[9] 李明春. 光伏及无线技术在城市电缆沟安全监测中的应用探索 [J]. 低碳世界, 2016(28): 65-66.

LI Mingchun. Photovoltaic and wireless technology in cable trench safety monitoring, the application of exploration in the city [J]. Low Carbon World, 2016(28): 65-66.

[10] 吕学梅, 孙宗义, 曹张驰. 电池板温度和辐射量对光伏发电量影响的趋势面分析 [J]. 可再生能源, 2014, 32(7): 922-927.

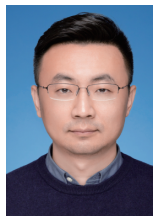
LYU Xuemei, SUN Zongyi, CAO Zhangchi. A trend surface analysis on effect of battery plate temperature and radiation levels on PV power output [J]. Renewable Energy Resources, 2014, 32(7): 922-927.

[11] ZHANG P, ZHANG L, WU T, et al. Detection and location of fouling on photovoltaic panels using a drone-mounted infrared thermography system [J]. Journal of Applied Remote Sensing, 2017, 11(1): 16-26.

[12] 于若英, 陈宁, 苗森, 等. 考虑天气和空间相关性的光伏电站输出功率修复方法 [J]. 电网技术, 2017, 41(7):

- 2229-2236.
YU Ruoying, CHEN Ning, MIAO Miao, et al. A repair method for PV power station output data considering weather and spatial correlations [J]. Power System Technology, 2017, 41(7): 2229-2236.
- [13] 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.
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.
- [14] 张志红, 何桢, 郭伟. 在响应曲面方法中三类中心复合设计的比较研究[J]. 沈阳航空工业学院学报, 2007, 24(1): 87-91.
ZHANG Zhihong, HE Zhen, GUO Wei. A comparative study of three central composite designs in response surface methodology[J]. Journal of Shenyang Institute of Aeronautical Engineering, 2007, 24(1): 87-91.
- [15] 王崇. 基于 PVsyst 的户用型光伏发电系统设计与仿真研究[D]. 沈阳: 沈阳农业大学, 2018.
WANG Chong. Design and simulation based on PVsyst household photovoltaic power generation system [D]. Shenyang: Shenyang Agricultural University, 2018.
- [16] 乔永力, 徐红伟, 孙坚. 家庭并网光伏发电系统仿真研究与分析[J]. 电源技术, 2019, 43(1): 133-135.
QIAO Yongli, XU Hongwei, SUN Jian. Simulation and analysis of home grid connected photovoltaic power generation system [J]. Chinese Journal of Power Sources, 2019, 43(1): 133-135.
- [17] KHUNTIA P, HAZRA R. An efficient deep reinforcement learning with extended Kalman filter for device-to-device communication underlying cellular network [J]. Transactions on Emerging Telecommunications Technologies, 2019, 30(9): 3671-3578.
- [18] POINTON H, MCLOUGHLIN B, MATTHEWS C, et al. Towards a model based sensor measurement variance input for extended Kalman filter state estimation [J]. Drones, 2019, 3(1): 19-24.

作者简介:



左松林

左松林(1981),男,学士,高级工程师,从事电力自动化相关工作(E-mail:zsl1988@163.com);

陈伟(1982),男,硕士,高级工程师,从事电力自动化相关工作;

付真斌(1985),男,博士,高级工程师,从事电力自动化相关工作。

Abnormal data inspection technology of photovoltaic power generation based on EKF algorithm

ZUO Songlin¹, CHEN Wei¹, FU Zhenbin¹, ZHAO Qian¹, JIANG Zaiyu², ZHENG Xinxin³

(1. State Grid Anhui Electric Power Co., Ltd. Research Institute, Hefei 230009, China; 2. Beijing China-Power Information Technology Co., Ltd., State Grid Information and Communication Industry Group, Beijing 100085, China; 3. Intelligent Manufacturing Technology Research Institute, Hefei University of Technology, Hefei 230009, China)

Abstract: During the operation of distributed photovoltaic power generation, abnormal data collected by the operating system would be caused by interference factors, such as equipment failure, meter measurement error, user personal behavior and so on. Therefore troubleshooting abnormal data helps to ensure the accuracy and reliability of the PV generation user database and identify distributed PV users with operational problems. For the above reasons, an abnormal data inspection technology for photovoltaic power generation operation system is proposed. Environmental factors such as temperature, radiation, latitude, and seasonality are taken into consideration. The mathematical model of power generation is established by the central composite design method. The accumulated power generation is corrected by the extended Kalman filter (EKF) algorithm. Thus abnormal data can be checked and eliminated. The proposed method can realize the prediction of power generation data. It can check abnormal data quickly and reliably. The principle of investigation and the specific implementation process are discussed. Finally the effectiveness of the proposed technique is proved by experimental results.

Keywords: distributed photovoltaic; power generation model; environmental factors; abnormal data identify; extended Kalman filter

(编辑 方晶)