

风电场输出功率异常数据识别与重构方法研究

朱倩雯¹, 叶林¹, 赵永宁¹, 郎燕生², 宋旭日²

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 中国电力科学研究院, 北京 100192)

摘要: 电力大数据是电力发展的重要资源, 数据来源于电力生产和电能使用的各个环节。风电运行数据是电力大数据的重要组成部分, 随着风电穿透功率的增大, 风电数据的采集、处理、分析对风电场运行、控制与并网研究有重要意义。然而, 从风电场收集到的大量数据中通常包含异常数据点, 这样的异常点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。分析了风电场历史运行数据中存在的异常数据的主要来源, 并针对该实际问题, 采用基于四分位算法的数学模型对异常数据进行识别。在数据缺失的情况下, 以可用历史数据为基础, 采用基于临近风电场出力模式性的方法和多点三次样条插值方法重构出完整的时间序列。算例分析给出了两种重构方法的重构效果以及各自的适应性, 结果表明采用所提出的方法能够有效识别、剔除异常数据并重构缺失数据, 对不同风电场有较强的通用性, 具有一定的工程实用价值。

关键词: 风电场; 风电运行数据; 电力大数据; 异常数据; 重构

Methods for elimination and reconstruction of abnormal power data in wind farms

ZHU Qianwen¹, YE Lin¹, ZHAO Yongning¹, LANG Yansheng², SONG Xuri²

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China;
2. China Electric Power Research Institute, Beijing 100192, China)

Abstract: Electric power big data is an important resource for electric power development and comes from the procedures of electricity production and energy utilization. Wind power operating data is the major part of electric power big data. With the dramatic increase of wind power penetration, it is of great significance for wind farm operation, control and integration research by collection, processing and analysis of real historical operating data from wind farms. However, amounts of data collected from wind farms usually contain abnormal data, which have adverse impact on the study of fluctuation characteristics of wind power, wind power prediction, etc. The main source of abnormal data existed in wind farm historical operation data is analyzed and a model for eliminating abnormal data based on quartile method is presented. In the cases of missing data, methods based on patterns of similarity between neighboring wind farms outputs and multi-point cubic spline are used on the basis of historical data to reconstruct the discontinuous time series respectively. The case study indicates that the presented models are efficient for eliminating abnormal data and reconstructing missing data, which can be applied in practical engineering.

This work is supported by National Natural Science Foundation of China (No. 51477174 and No. 51077126).

Key words: wind farm; wind power operating data; big data; abnormal data; reconstruction

中图分类号: TM619 文献标识码: A 文章编号: 1674-3415(2015)03-0038-08

0 引言

随着电力大数据时代的到来, 电力大数据在应用过程中存在着对行业内外能源数据、天气数据等多类型数据的大量关联分析需求, 而这些都直接导

致了电力数据类型的增加, 从而极大地增加了电力大数据的复杂度。电力行业对大数据的需求, 其迫切性将大大超越其他基础能源行业。电力大数据将贯穿未来电力工业生产及管理各个环节, 起到独特而巨大的作用。在电力生产环节, 风光储等新能源的大量接入, 打破了传统相对“静态”的电力生产, 使得电力生产的计量和管理变得日趋复杂^[1]。

基金项目: 国家自然科学基金项目 (51477174, 51077126)

风电运行数据是电力大数据的重要组成部分。当前, 风力发电是可再生能源开发利用中技术最成熟、最具开发规模和商业化发展前景的发电形式。随着风力发电技术的不断发展, 由于风电功率所呈现的随机性、波动性和间歇性, 风电大规模并网对电力系统的影响也越来越明显, 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。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^[2-4]、风电功率对电网影响评估及控制策略^[5-6]、风电功率曲线的计算、测试以及修正^[7]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^[8], 可以用于评估风机和风电场的性能和运行状况、预测风机和风电场的年发电量等; 对于预测服务商, 历史实测风速和风电功率数据作为预测系统的基础输入数据, 对预测精度有显著影响。

然而, 从风电场收集到的大量数据中通常包含异常数据点, 这样的异常点表现出数据缺失, 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^[9]: 风机叶片受污垢和冰等环境因素的影响形成异常点体现在风电场功率曲线上; 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。

因此, 有必要对风电历史运行数据中的异常数据点进行识别剔除和还原重构。文献[9]建立了基于逻辑函数的风电功率曲线非线性模型, 并通过进化策略算法对模型参数进行求解。这个模型通过残值法和控制图对异常值进行识别、剔除。但是模型的建立需要大量的正常数据样本, 且进化策略算法对大样本数据的多参数寻优过程速度较慢, 具有较大局限性。文献[10]在数据预处理阶段通过简单的数据范围检查和一致性检验对异常数据进行剔除, 在功率曲线建模过程中忽略缺失数据, 然后根据互相关函数去除风速和功率时间序列之间的延时特征, 从而得到一个静态的(无延时)功率曲线模型。文献[11]采用基于小波模极大值的方法对风速数据中的奇异点进行辨识, 该方法将阈值判断与李氏指数相结合, 判断出风速信号中的奇异性及发生时刻, 剔除其中的奇异数据, 并采用 ARMA 模型对奇异点进行修正。但该方法仅针对时间序列的奇异点辨识, 并没有考虑到风速-功率的对应关系, 在奇异点修正过程中采用单一的 ARMA 模型修正效果并不

理想。文献[12]针对弃风情况下风电场历史运行数据存在的异常数据的实际问题, 自定义了在有功功率时间序列中异常数据所应遵循的时序变化特性, 并根据这些特性识别异常数据。但实际上由于风电所固有的较强波动性和随机性, 很难从时序变化特性上完全区分正常数据和异常数据, 尤其是在弃风现象不明显时, 其识别过程具有很大的局限性。文献[13]通过简单的数据范围检查和一致性检验对异常数据进行剔除, 并不适用于处理大量分散型异常数据。

本文分析了风电场历史运行数据中存在的异常数据的主要来源, 并针对该实际问题, 采用基于四分位算法的数学模型对异常数据进行识别; 并在此基础上通过实际算例对比分析了采用基于临近风电场出力模式性和多点三次样条插值方法对风电场输出功率异常数据进行重构的效果, 从而为研究风电功率波动特性、风电功率预测等风电相关问题提供有效的数据集。

1 异常数据的识别与剔除

1.1 异常数据分析

文献[14]对德国和丹麦风电场的研究表明, 导致风机故障的主要原因是电气控制或系统部件(电网或电气系统, 偏航系统以及机械或变桨控制系统)而不是机械组件, 比如变速箱。风电功率的异常数据主要是由上述故障和风机特性的不稳定造成的。大量的异常数据并不容易识别, 这些异常数据的存在严重影响了研究建模和风资源评估的结果。检验建模所用数据质量对于建模是十分重要的, 风电功率数据集中存在大量异常数据, 因此对该类问题的自动识别与剔除方法的研究是很有必要的。

考虑到难以从单一的风电功率时序变化特性中自识别分散型异常数据, 本文借助风速-功率散点图分析, 体现在 $v-p$ 散点图上存在大量分散型异常数据点和堆积型异常数据簇。异常数据的检验识别首先是要剔除风速-功率数据组中的连续数据以及越限数据, 这些异常数据通常是由表计读数错误引起的。但是仅对这部分数据进行剔除并不足以功率曲线建模、风电功率预测提供真实可靠的数据组。风电出力数据存在很多异常值, 这是因为风速的随机性和波动性导致了两个参量之间相关关系的时变性。图 1 是国内某风电场 70 m 测风塔风速与实测功率散点箱线图, 该风电场的切入风速 $v_{ci}=4$ m/s, 切出风速 $v_{co}=20$ m/s, 额定风速 $v_r=13$ m/s, 额定功率 $P_r=49.5$ MW。由图中显而易见的是风电功率数据集中存在大量异常数据, 表现在箱线图中,

数据整体分布不集中, 离群值数量大且分布不规律, 对风电功率预测、风电功率曲线的计算、测试以及修正等方面研究带来负面影响。就风电出力而言, 如果功率值是每 15 min 风机出力的平均值, 必须保证这台风机在这 15 min 内正常工作, 否则, 这些数据并不能代表风机的正常运行状态。

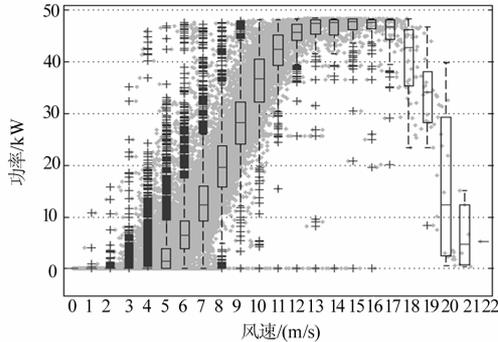


图 1 风电场风速-风电功率散点箱线图

Fig. 1 Wind speed-power scatter box-plot of a wind farm

1.2 识别与剔除异常数据的原理与过程

基于正态分布的 3σ 法则或 Z 分数方法是以假定数据服从正态分布为前提的, 但实际数据往往并不严格服从正态分布。它们判断异常值的标准是样本数据的均值和标准差, 而均值和标准差的耐抗性极小, 异常值本身会对它们产生较大影响, 这样产生的异常值个数不会多于总数 0.7%。显然, 将这种方法应用于非正态分布数据的异常值识别, 其有效性是有限的。箱线图的绘制依靠实际数据, 不需要事先假定数据服从特定的分布形式, 没有对数据作任何限制性要求, 它真实直观地表现数据形状的本来面貌; 另一方面, 箱线图判断异常值的标准以四分位距为基础, 四分位数具有一定的耐抗性, 多达 25% 的数据可以变得任意远而不会很大地扰动四分位数, 因此箱线图识别异常值的结果比较客观, 具有一定的优越性。

通过对体现在 $v-p$ 散点图上的大量分散型异常数据点和堆积型异常数据簇的分析, 根据异常数据在 $v-p$ 散点图中的随机分布特性, 本文采用基于四分位原理^[15-16]的数学模型对风电场异常数据进行识别与剔除。其异常数据识别和剔除过程如下:

(1) 当风速小于切入风速 v_{ci} 时, 风机的理论出力恒等于 0。因此在 $[0, v_{ci}]$ 区间内删除掉负功率值点, 保留风电功率的非负值点, 有利于后续对其他异常数据的处理工作。

(2) 当风速大于切入风速 v_{ci} 且小于切出风速 v_{co} 时, 该段数据体现在 $v-p$ 散点图中测风塔风速对应的风电功率波动幅度较大且分散, 大多数异常数据

点出现在这个风速区间内。在 $[v_{ci}, v_{co}]$ 区间内, 小于等于 0 的风电功率值应视为异常数据, 即将散点图中最底部由于弃风限电造成的堆积型异常数据簇剔除。此外, 还将大于额定值的功率数据及其对应的风速删除掉。

(3) 采用横向四分位法将横向分布的分散性异常数据点剔除。

从散点图中可以看出, 在理论功率曲线的包络线外存在许多分散性异常数据点以及堆积型异常数据簇, 如果直接采用散点图中的原始数据进行功率曲线建模、风电功率预测等研究, 这些异常数据的存在会严重影响研究分析的结果。因此可采用四分位法将这些异常数据剔除掉。首先, 对每个功率区间内(例如以 0.5 MW 为间隔单位将 $[0, 50 \text{ MW}]$ 的风电功率划分为相等的 100 个功率区间)对应的风速数据集合采用四分位法, 剔除掉位于风速数据集合内限 $[F_1, F_u]$ 之外的数据点。风速数据集合内限 $[F_1, F_u]$ 可以通过式(1)进行计算。

$$[F_1, F_u] = [Q_1 - 1.5IQR, Q_3 + 1.5IQR] \quad (1)$$

对于一个升序排列的样本 $X=(x_1, x_2, \dots, x_n)$, 式中, Q_1 和 Q_3 分别是样本序列的第一四分位数和第三四分位数; IQR 为四分位距(InterQuartile Range)。

(4) 采用纵向四分位法将数据左上方的分散性异常数据点剔除。

对每个风速区间内(如以 0.5 m/s 为间隔单位将 $[0, 20 \text{ m/s}]$ 的风电功率划分为相等的 40 个风速区间)对应的功率数据集合采用四分位法, 剔除掉功率数据集合大于上限 F_u 的数据点。最终保留的数据即为该风速区间内所期望的正常数据。

2 缺失数据点重构方法

上述基于四分位算法的异常数据识别与剔除算法虽然对异常数据的剔除效果较好, 但是被剔除的数据中难免存在被误删的正常数据, 这在很大程度上对数据的充裕度以及功率时间序列的完整性造成了破坏, 不利于对数据的再利用。因此, 有必要对剔除异常数据后的风电功率时间序列进行合理有效的重构, 以保证在异常数据被有效剔除的前提下, 能够较好地保持原始数据集的完整性以及对序列再利用的有效性。

将 $v-p$ 散点图中通过四分位算法识别与剔除后得到的所期望的正常风电出力数据绘制在风电出力时间序列图中, 找到时序图中的缺失数据, 通过以下方法对缺失数据点进行重构。

2.1 基于风电场出力相关性重构方法

2.1.1 风电出力模式性

电力系统除了在负荷侧表现出运行的模式性外,在电源侧也是有模式可循的。即使是随机性较强的风光等间歇性电源,模式性往往也是比较明显的^[7]。计 N 个地理位置临近的风电场同一时刻的输出功率分别为 $P_i(t)(i=1,2,\dots,N)$, MW, 各个风电场的输出功率占总功率的比值分别为 $X_i(t)(i=1,2,\dots,N)$, 其关系为

$$X_i(t) = \frac{P_i(t)}{\sum_{i=1}^N P_i(t)} \quad (2)$$

以我国距离约 10 km 的某两个临近风电场 A、B 为例,图 2 给出了风电场 A、B 一天内的输出功率以及输出功率与总功率比值的时间序列曲线,时间间隔为 15 min。对比分析可以看出,A、B 两个风电场的运行呈现出很明显的模式性:虽然在一天之内两个风电场的输出功率波动较大,呈现出较强的随机性和时变性,但是由于地理位置临近,气象条件相似,风速、风向变化特性相近,因此两个风电场的输出功率变化趋势以及所占总功率的比值相对稳定。虽然由于两个风电场昼、夜风电功率分布特性、风电场中机组维护和检修计划安排以及弃风率的差异等导致在不同时段其模式性受到一定程度的破坏,但风电场 A、B 输出功率占总功率的比值分别基本维持在 0.6 和 0.4 左右,相较于风电场出力波动幅度小,比值较稳定。

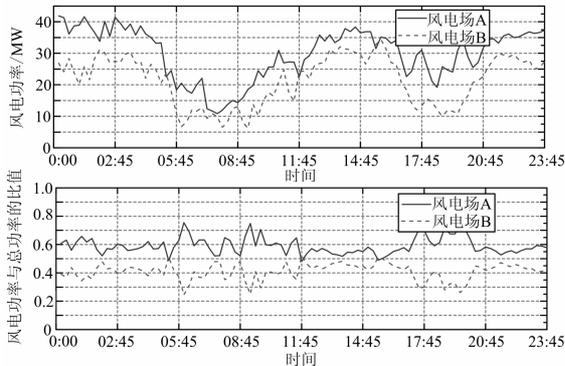


图 2 两个临近风电场输出功率与比值时间序列图

Fig. 2 Time series plots of wind power and its ratios of 2 adjacent wind farms

表 1 给出了 A、B 两个风电场的输出功率以及输出功率与总功率比值在一天内的波动率,波动率计算表达式为

$$F = \frac{P_{\max} - P_{\min}}{P_{\min}} \quad (3)$$

式中: F 表示波动率; P_{\max} 表示一天内时间序列的最大值, MW; P_{\min} 表示一天内时间序列的最小值, MW。

由表 1 可知,由于地理位置邻近的风电场的运行存在一定程度的模式性,模式性的存在使得新定义的功率比值时间序列比原功率时间序列更具平稳性,因此也更具预测性。故可以通过对单个风电场所占总功率的比值进行预测,由已知风电场的输出功率求得待重构风电场的输出功率。

表 1 两个风电场输出功率及其比值的日内波动率

风电场 A		风电场 B	
P_a	X_a	P_b	X_b
2.87	0.56	4.55	1.10

2.1.2 基于风电场出力模式性的重构方法

假设缺失数据所在时段为 $[i+1, i+N]$, 待重构风电场 A 以及与其地理位置临近的风电场 B 在 t 时刻的出力分别为 $P_a(t)$ 、 $P_b(t)$ 。取两风电场在 $[i-M, i]$ 时段内的风电功率历史运行数据,通过式(2)计算待重构风电场 A 在已知风电历史出力时段 $[i-M, i]$ 内的风电功率占总功率的比值,得到历史比值序列 $X_a(t)(t=i-M, \dots, i)$ 。采用 Box-Jenkins 建模方法,利用 AIC 准则(Akaike Information Criterion)^[18]对比值序列 $X_a(t)$ 定阶,关于 ARMA(p, q)模型,其 AIC 评价指标定义为

$$I_{\text{AIC}}(p, q) = n \ln \hat{\sigma}_\varepsilon^2(p, q) + 2(p + q) \quad (4)$$

式中: $\ln \hat{\sigma}_\varepsilon^2(p, q)$ 是拟合 ARMA(p, q)模型时的残差的方差; n 为样本数据长度。

并在此基础上建立自回归滑动平均模型 ARMA(p, q)对风电场 A 在缺失数据段 $[i+1, i+N]$ 的风电功率占总功率的比值 $X_a(t)$ 进行预测,即^[19]

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (5)$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

式中: X_t 为时间序列,这里是比值序列 $X_a(t)$; p 和 q 是模型的自回归阶数和滑动平均阶数; ϕ_i 和 θ_j 为

自回归系数和滑动平均系数; $\sum_{i=1}^p \phi_i X_{t-i}$ 为自回归项

AR,即为过去观测值的线性组合; $\sum_{j=0}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$ 为白噪声序列的滑动平均项 MA; N 代表正态分布。

则风电场 A 在缺失数据段 $[i+1, i+N]$ 的出力可以表示为

$$P_a(t) = \frac{X_a(t)}{1 - X_a(t)} P_b(t) \quad (6)$$

风电场 A 缺失数据段 $[i+1, i+N]$ 的功率 $P_a(t)$ 进而可通过临近风电场 B 的功率 $P_b(t)$ 由式(6)进行重构。

为证明该方法能有效对缺值数据进行重构,更好地适应风电功率的时变特性,在功率时间序列中挑选一段在短时间内大范围波动的采样数据点模拟缺值数据段,分别采用普通 ARMA 模型和基于风电场之间出力模式性模型对缺值数据进行重构。功率时间序列分辨率为 15 min, 所得结果的逐点绝对误差值如图 3 所示。

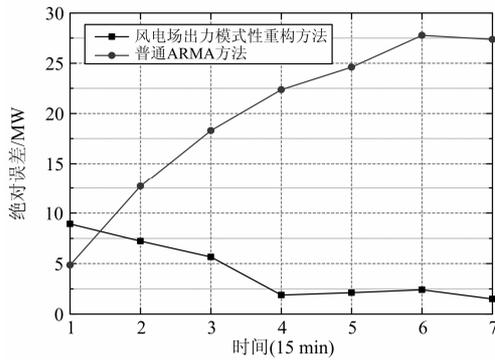


图 3 两种形式的 ARMA 重构误差

Fig. 3 Reconstruction errors of ARMA based on two different methods

由图 3 可知,采用基于风电场之间出力模式性模型对缺值进行重构,通过充分利用临近风电场实测数据以及风电场之间的模式性,能更好地适应风电功率的时变特性。

2.2 基于多点三次样条插值重构方法

三次样条(Cubic Spline)插值方法是数值分析理论中的经典内插方法,是一种能克服高次多项式插值的缺陷的分段插值方法^[20-21]。

设函数 $S(x) \in C^2[a, b]$, 在 $[a, b]$ 上有一个划分

$$\Delta: a = x_0 < x_1 < \dots < x_n = b \quad (7)$$

且在每个子区间 $I_j = [x_{j-1}, x_j]$ ($j=1, 2, \dots, n$)上都是三次多项式, 则称 $S(x)$ 是关于划分 Δ 的一个三次样条函数。若三次样条函数 $S(x)$ 在节点 x_j 上给定函数值 $y_j = f(x_j)$ ($j=0, 1, \dots, n$), 并成立

$$S(x_j) = y_j, j = 0, 1, \dots, n \quad (8)$$

则称 $S(x)$ 为三次样条插值函数。

本文通过三弯矩插值法构造样条插值函数 $S(x)$, 令 $M_j = S''(x_j)$ ($j = 0, 1, \dots, n$), 由于 $S(x)$ 在 $[x_j, x_{j+1}]$ 上是二阶光滑的分段三次多项式, $S''(x)$ 是分段线性连续函数。

$$S''(x) = \frac{x_{j+1} - x}{h_{j+1}} M_j + \frac{x - x_j}{h_{j+1}} M_{j+1}, x \in [x_j, x_{j+1}] \quad (9)$$

式中: M_j 为待定参数, $h_{j+1} = x_{j+1} - x_j$ ($j=0, 1, \dots, n-1$)。对式(9)两端在 $[x_j, x_{j+1}]$ 上积分两次, 因此 $S(x)$ 在 $[x_j, x_{j+1}]$ 上有表达式

$$S(x) = \frac{(x_{j+1} - x)^3}{6h_{j+1}} M_j + \frac{(x - x_j)^3}{6h_{j+1}^3} M_{j+1} + \frac{x - x_j}{h_{j+1}} [f_{j+1} - f_j - \frac{h_{j+1}^2}{6} (M_{j+1} - M_j)] + f_j - \frac{h_{j+1}^2}{6} M_j, x \in [x_j, x_{j+1}] \quad (10)$$

式中, $f_j = S(x_j)$ 。

在自然边界条件下, 得线性方程组

$$\begin{bmatrix} 2 & \lambda_1 & & & & & \\ \mu_2 & 2 & \lambda_2 & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \mu_{n-2} & 2 & \lambda_{n-2} & \\ & & & & \mu_{n-1} & 2 & \\ & & & & & & \lambda_{n-1} \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n-2} \\ d_{n-1} \end{bmatrix} \quad (11)$$

式中: $\lambda_j = \frac{h_{j+1}}{h_j + h_{j+1}}$; $\mu_j = \frac{h_j}{h_j + h_{j+1}}$; $d_j = 6f[x_{j-1}, x_j \cdot x_{j+1}]$ ($j=1, 2, \dots, n-1$)。

对缺失数据点插值的效果并不仅仅取决于缺值区间的端点数据, 还取决于间隔与其缺值间隔等距的若干数据点。采用该方法不会使时间序列中的周期信号受损, 计算出的插值点不会造成较大的失真。采用的插值区间满足各个选定点之间间隔和插值点之间间隔相等, 同时就保证了取样间隔与奈奎斯特频率之间的关系^[22]。而采用其他插值方法, 该条件很难满足。通过试验选取对待插值点影响较大的前后各 M 个点, 远离待插值点前后各 M 个点以外的数据对插值计算的影响微乎其微。假设缺失数据所在区间为 $[i+1, i+N]$, 待重构风电场在第 t 点的出力为 $P(t)$ 。多点三次样条插值^[22]重构方法是在三次样条插值的基础上每次在连续缺失 N 个数据点的数据段每次进行一个数据点的插值, 每次插值计算都采用三次样条插值方法。比如在总数为 1 000 个点的风电功率时间序列中连续缺失 7 个数据点(即 $N=7$), 其缺值区间为 $[501, 507]$, 则多点三次样条插值重构方法首先对 $P(504)$ 点进行插值计算。第一步计算时采用数据间隔为 3 点的 $P(500), P(496), \dots, P(504-4 \times M); P(508), P(511), \dots, P(504+4 \times M)$ 以上 $2 \times M$ 个点进行插值计算。每次通过多点三次样条插值重构计算参数所用到的数据间隔为连续缺失数据区间的一半, 继续进行以上过程直到完成

缺失数据段的插值重构。

3 算例分析

为验证所采用的异常数据识别、缺失数据重构方法的有效性, 对比不同方法的适应性, 以国内某风电场的历史运行数据为研究对象, 数据采样间隔为 15 min。

3.1 异常数据识别与剔除策略检验

本文以国内某风电场 2013 年整年的风电历史运行数据为研究对象, 通过采用上文提出的模型对异常数据进行识别、剔除, 整年包含约 30 000 个风速-风电功率数据对, 图 1 为该风电场的原始 $v-p$ 散点箱线图, 风电场异常数据经过基于四分位原理算法剔除的结果如图 4 所示。可见, 通过该模型对原始风速-功率数据的自识别, 分散型异常数据点等高群值都已被有效剔除。该模型从风速-功率数据集中识别并剔除了 22.67% 的数据。

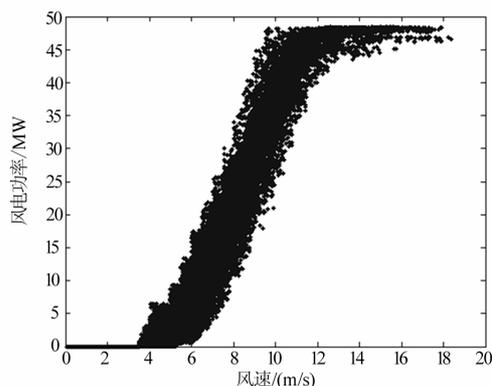


图 4 模型识别后风电场风速-功率散点图

Fig. 4 Wind farm scatter diagram after model elimination

该模型对此类风电场数据集中的异常数据提供了合理有效的识别方法, 对不同风电场有较强通用性, 并且能够为异常数据分布特性未知的其他数据集中的异常数据识别提供有效参考。

3.2 缺失数据重构方法检验

图 5 为装机容量为 49.5 MW 的风电场 A 在 2013 年 11 月 28 日的实际功率时序曲线, 为了验证本文提出的重构方法对风电场缺失数据的重构能力, 在算例分析中分别删除短时间内大范围波动的采样点数据(2013 年 11 月 28 日 7 个采样点数据), 模拟数据缺失情况以检验重构方法有效性。分别采用基于风电场之间出力模式性和基于多点三次样条插值重构方法, 与常规的数学方法(多项式拟合、递推式非邻均值补全法和普通三次样条插值方法)进行比较。

3.2.1 基于风电场出力模式性的重构方法

通过与待重构风电场 A 地理位置临近的风电

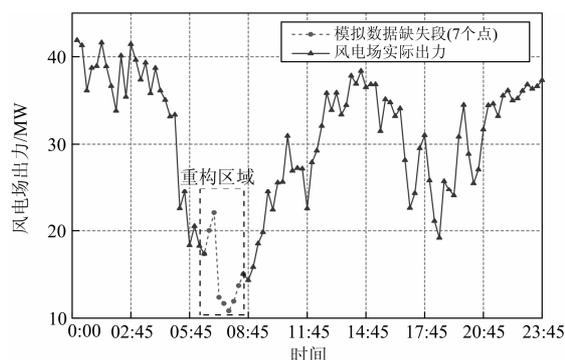


图 5 对缺失 7 点数据的模拟

Fig. 5 Simulation of missing 7 data

场 B 的实际出力对风电场 A 的缺失数据段进行重构。基于 ARMA 模型对待重构风电场 A 的风电出力占总功率(风电场 A 和风电场 B 功率之和)的比值时间序列建模, 对缺值数据段比值序列样本选取 ARMA(2,1)模型进行预测。最后, 通过预测得到的缺失数据所在时段的比值与临近风电场 B 的功率对待重构风电场 A 缺失数据段的功率通过式(6)进行重构。得到对风电场输出功率时间序列缺值数据段重构结果以及逐点的绝对误差值如图 6 所示。

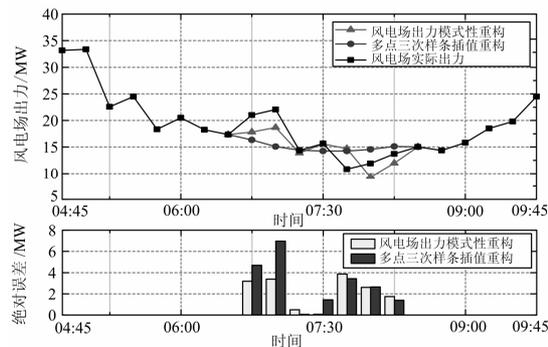


图 6 缺值 7 点重构结果对比

Fig. 6 Reconstruction effect comparison of missing 7 data

3.2.2 基于多点三次样条插值重构方法

由于远离待插值点前后各 M 个点以外的数据对插值计算的影响较小, 因此通过试验选取对待插值点影响较大的前后各 M 个点对缺失数据进行重构。当 M 达到一定值时, 通过增加影响插值的数据个数(M 值)对待插值数据段进行重构计算对插值结果的误差影响很小(1×10^{-6})。因此, 对缺值的数据样本选择前后各 5 点进行多点三次样条重构计算。得到对风电场输出功率时间序列缺失的数据段重构结果以及逐点的绝对误差值如图 6 所示。

本文在重构结果对比分析中, 选择多项式拟合法、递推式非邻均值补全法和普通的三次样条插值方法对缺失数据进行重构, 与上述两种方法的误差对比结果如表 2 所示, 其中 MAE(Mean Absolute

Error)是平均绝对误差。

算例表明,对风电功率时间序列缺失数据进行重构,基于临近风电场出力模式性对缺失数据的重构效果最好,但是当临近风电场数据缺失、异常、不可用时,可采用多点三次样条插值重构方法对缺失数据进行重构,亦可得到较理想的重构效果。采用以上两种重构方法对缺失数据段进行重构的结果均优于采用多项式拟合、递推式非邻均值补全法和普通三次样条插值方法重构得到的结果。值得注意的是,由于出力模式性模型对相邻风电场可用数据的依赖性较高,因此基于临近风电场出力模式性的重构方法对时间尺度更长的数据缺失以及在待重构风电场功率与总功率比值突变情况下的重构有一定限制。此外,基于多点三次样条插值重构方法对缺失数据中的突变点的拟合效果不如基于临近风电场出力模式性对缺失数据的重构。

表2 重构结果对比

Table 2 Effect comparison of different reconstruction methods

方法	最大绝对误差	最小绝对误差	平均绝对误差
风电场出力模式性	3.865	0.072	2.195
多点三次样条插值重构	6.954	0.059	2.933
多项式拟合	8.114	0.205	3.122
递推式非邻均值补全法	5.293	0.525	3.185
普通三次样条	6.176	1.569	3.715

4 结论

从风电场收集到的大量数据中通常包含异常运行数据点,这样的异常点对风电功率波动特性、风电功率预测、风电功率对电网影响评估及控制策略、风电功率曲线的计算、测试以及修正等方面研究带来负面影响。本文从 $v-p$ 散点图的角度对异常数据进行识别剔除,并在此基础上利用风电场出力模式性和改进的多点三次样条方法对缺失数据进行重构。主要结论如下。

(1) 由于风电所固有的较强波动性和随机性,很难从时序变化特征上完全区分正常数据和异常数据。这些异常数据主要来源于数据采集、测量、传输的各个环节中发生的故障和干扰以及维修或弃风限电造成的风电场停机。

(2) 本文从 $v-p$ 散点图的角度出发采用的基于四分位算法的数学模型对于不同类型、不同数量异常数据的剔除效果显著,且有较强的通用性和较高的工程实用价值。

(3) 针对少量数据缺失和连续缺失较多数据的

情况,分别采用基于临近风电场出力模式性和多点三次样条插值方法对异常缺失数据进行重构,并通过算例分析证明两种方法对缺失数据进行重构可信用度较高,改善了数据的完整性,有利于数据的再利用。

参考文献

- [1] 中国电机工程学会信息化专委会. 中国电力大数据发展白皮书[R]. 2013.
- [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.
YANG Mao, XIONG Hao, YAN Gangui, et al. Real-time prediction of wind power based on data mining and fuzzy clustering[J]. Power System Protection and Control, 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.
XIA Dong, WU Junyong, HE Dian, et al. A novel combined model for wind power forecasting based on maximum entropy principle[J]. Transactions of China Electrotechnical Society, 2011, 26(S1): 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.
WANG He, HU Zhijian, CHEN Zhen, et al. A hybrid model for wind power forecasting based on ensemble empirical mode decomposition and wavelet neural networks[J]. Transactions of China Electrotechnical Society, 2013, 28(9): 137-144.
- [5] 于芑,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.
YU Peng, ZHAO Yu, ZHOU Wei, et al. Research on the method based on hybrid energy storage system for balancing fluctuant wind power[J]. Power System Protection and Control, 2011, 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.
HU Guowei, BIE Zhaohong, WANG Xifan. Optimal dispatch in wind integrated system considering operation reliability[J]. Transactions of China Electrotechnical Society, 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.
SHEN Xinhe, PAN Donghao, TANG Jiguang. Analysis and modification method for wind turbine power curve[J].

- Applied Energy Technology, 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.
MA Ping, LIU Changhua. Validation on the power curve of wind power unit[J]. Renewable Energy Resources, 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] SCHLECHTINGEN M, FERREIRA S I. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection[J]. Mechanical Systems and Signal Processing, 2011, 25(5): 1849-1875.
- [11] 李丽, 叶林. 风速数据奇异点辨识研究[J]. 电力系统保护与控制, 2011, 39(21): 92-97.
LI Li, YE Lin. Identification of singular points in wind speed data[J]. Power System Protection and Control, 2011, 39(21): 92-97.
- [12] 张东英, 李伟花, 刘燕华, 等. 风电场有功功率异常运行数据重构方法[J]. 电力系统自动化, 2013, 37(4): 1-6.
ZHANG Dongying, LI Weihua, LIU Yanhua, et al. Reconstruction method of active power historical operating data for wind farm[J]. Automation of Electric Power Systems, 2013, 37(4): 1-6.
- [13] 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.
- [14] TAVNER P J, XIANG J, SPINATO F. Reliability analysis for wind turbines[J]. Wind Energy, 2007, 10(1): 1-18.
- [15] HYNDMAN R J, FAN Y. Sample quantiles in statistical packages[J]. The American Statistician, 1996, 50(4): 361-365.
- [16] 李国柱. 统计学[M]. 北京: 科学出版社, 2004.
- [17] 段炼, 于继来, 徐泰山, 等. 基于运行模式的注入功率联合概率分布研究[EB/OL]. [2013-05-31]. <http://www.paper.edu.cn/releasepaper/content/201305-491>.
- [18] 徐国祥. 统计预测和决策[M]. 上海: 上海财经大学出版社, 2008.
- [19] 陈刚, 段晓, 张继红, 等. 基于 ARMA 模型的低频振荡模式在线辨识技术研究[J]. 电网技术, 2010, 34(11): 48-54.
CHEN Gang, DUAN Xiao, ZHANG Jihong, et al. A new approach for online identification of low frequency oscillation modes based on auto-regressive moving-average model[J]. Power System Technology, 2010, 34(11): 48-54.
- [20] 李庆扬, 王能超, 易大义. 数值分析[M]. 北京: 清华大学出版社, 2001.
- [21] 吴勃英. 数值分析原理[M]. 北京: 科学出版社, 2003: 135-144.
- [22] 武艳强, 黄立人. 时间序列处理的新插值方法[J]. 大地测量与地球动力学, 2005, 24(4): 43-47.
WU Yanqiang, HUANG Liren. A new interpolation method in time series analyzing[J]. Journal of Geodesy and Geodynamics, 2005, 24(4): 43-47.

收稿日期: 2014-05-04

作者简介:

朱倩雯(1991-), 女, 硕士研究生, 从事电力系统运行与控制、新能源发电技术方面的研究;

叶林(1968-), 男, 教授, 博士生导师, 主要研究方向为电力系统自动化、新能源发电与超导电力应用; E-mail: yl@cau.edu.cn

赵永宁(1990-), 男, 博士研究生, 从事电力系统运行与控制、新能源并网技术方面的研究。