

# 风电场 SCADA 数据预处理方法及评价策略

戴巨川<sup>1</sup>, 曹俊伟<sup>1</sup>, 张 帆<sup>1</sup>, 刘德顺<sup>1</sup>, 沈祥兵<sup>2</sup>

(1. 湖南科技大学机电工程学院, 湘潭 411201; 2. 湘电风能有限公司, 湘潭 411000)

**摘 要:** 为了从 SCADA 数据中获得更明确的物理信息, 更好地判断风电机组运行状态, 分别将常规平均数法、最小二乘法及该文提出的非参数法(核密度-均值法)用于风电场 SCADA 数据预处理。建立风电场 SCADA 数据预处理算法的 3 个评价指标, 包括物理特性一致性、采样时间变化稳健性及采样频率变化稳健性。设计评价指标定量计算公式, 用以评价各种预处理算法效果, 计算结果表明非参数法(核密度-均值法)能够获得更好的预处理效果。最后, 基于核密度-均值法对全工况风电机组 SCADA 数据进行预处理, 分析风电机组运行特性, 包括风速与输出功率、轮毂转速的关系以及风能利用系数。

**关键词:** 风电机组; SCADA 数据; 数据预处理; 评价指标

**中图分类号:** TK83

**文献标识码:** A

## 0 引 言

随着风电机组单机容量的不断增加, 单机成本越来越高, 各类风电机组事故的发生对风电产业发展带来严重的负面影响。目前, 大型风电机组均配有数据采集与监视控制(supervisory control and data acquisition, SCADA)系统, 能够获取风电机组实时运行状态参数。在服役阶段基于 SCADA 数据分析风电机组运行状态, 实现实时预警与故障诊断是保证风电机组安全的重要手段。在学术界和工程界已有大量研究围绕 SCADA 数据展开, 如 Kusiak 等<sup>[1]</sup>通过对 3 台以上风电机组 SCADA 数据比较分析来识别有故障的风电机组; 文献[2~5]通过提取风电场 SCADA 数据对齿轮箱故障进行了识别研究; Schlechtingen 等<sup>[6]</sup>将自适应神经模糊推理系统用于风电场 SCADA 数据处理, 进行风电机组状态识别研究; 杨文献等<sup>[7]</sup>在对 SCADA 原始数据进行预处理的基础上, 构建状态判别的决策模型, 对风电机组叶片故障、发电机轴承故障等进行量化评估。在利用 SCADA 数据进行风电机组状态识别、预警与故障诊断时需要对数据进行预处理, 因为风电机组在服役过程中, 风电场风速风向是自然随机的, 还存在许多其他(如温度湿度等)自然随机因素和所接

电网、风电机组自身等设备随机因素, 这些随机性在整个能量转换过程中叠加和传递, 并且包含在 SCADA 数据之中。已有研究中, 有的采用平均值, 有的采用期望值来处理这些随机性, 但有关风电机组 SCADA 数据预处理的研究尚处于探索阶段, 对如何选择合理的预处理算法以及预处理算法的评价策略有待深入研究。

## 1 SCADA 原始数据基本格式

研究用风电机组为 2 MW 直驱型, 位于郴州鲁荷金风场, 其风轮直径 96 m, 切入风速 3 m/s, 额定风速 11 m/s, 轮毂转速 6~17 r/min。SCADA 系统能够获取风电机组实时运行状态参数, 基本参数包括风速、转速、电压、电流、功率、振动、温度等。SCADA 数据系统采用 1 s 采样的方式进行储存(采样频率为 1 Hz)。表 1 给出所使用的原始 SCADA 数据式样。

根据 SCADA 数据可绘制出风电机组不同状态参数之间相互关系, 图 1 所示为风速与风电机组轮毂转速之间的关系。从图 1 可看出, 直接由风电机组原始数据得到的状态关系图形较复杂, 难以直接进行定性分析, 从而获得明确的物理关系。

收稿日期: 2015-10-15

基金项目: 国家自然科学基金(51475160; 51205123); 湖南省教育厅科研优秀青年项目(15B084)

通信作者: 戴巨川(1979—), 男, 博士、副教授, 主要从事风电技术与装备方面的研究。daijuchuan@163.com

表1 SCADA采集的原始数据式样

Table 1 Original form of SCADA data

序号	时刻	风速/ $\text{m}\cdot\text{s}^{-1}$	轮毂转 速/ $\text{r}\cdot\text{min}^{-1}$	...	变频器发电 机侧功率/kW
1	14:20	6.1	12.78	...	694
2	14:20	6.5	12.78	...	677
3	14:20	6.0	12.80	...	674
4	14:20	6.4	12.81	...	676
⋮	⋮	⋮	⋮	⋮	⋮

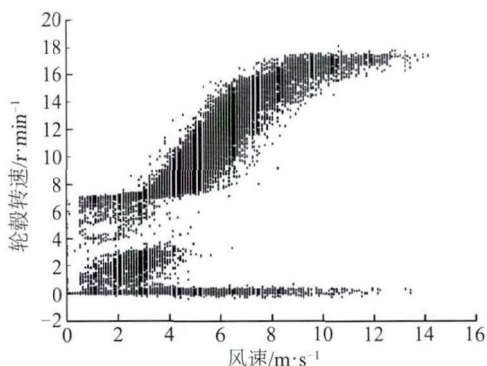


图1 基于原始数据的风电场风速-轮毂转速数据图

Fig. 1 Relationship between wind speed and rotor speed based on original SCADA data

## 2 SCADA数据预处理方法

为了从 SCADA 数据中获得更明确的物理信息,剔除相关干扰信息,应对 SCADA 数据进行预处理。数据预处理的方法很多,常用的有平均数法和最小二乘法,本文提出一种非参数法(核密度-均值法),并对这 3 种方法进行对比分析。

### 2.1 平均数法

平均数法是数据处理中常用的、简单易算的方法。但为了便于与后文中的最小二乘法、非参数法(核密度-均值法)进行比较分析,不宜直接对全工况数据进行处理,应考虑风速范围、桨距角状态等因素选择物理意义明确的阶段。下面以风电场 SCADA 数据中的风速、轮毂转速、发电机输出功率等参数为例介绍应用平均数法的具体实施方法。(步骤 1)~(4)为数据筛选,步骤 5)~(8)为数据处理):

1)选取出风电场原始 SCADA 数据中的风速、轮毂转速、发电机输出功率参数,分别用  $v$ 、 $r$ 、 $p$  表示,单位分别为  $\text{m/s}$ 、 $\text{r/min}$ 、 $\text{kW}$ 。

2)选取风速超过切入风速且低于额定风速下的

阶段进行分析,即风速在  $3\sim 10\text{ m/s}$  范围内以及与该风速范围相对应的轮毂转速和发电机输出功率参数集。

3)去除停机及变桨距阶段数据,即判断桨距角  $\beta > 0.5^\circ$ ,  $r \leq 0$  或  $p \leq 0$  是否成立,若成立则满足条件的相应各状态参数舍去。

4)将处理后的风速、轮毂转速、发电机输出功率分别用  $V$ 、 $R$ 、 $P$  表示。

经过筛选后的风速-轮毂转速如图 2 所示。

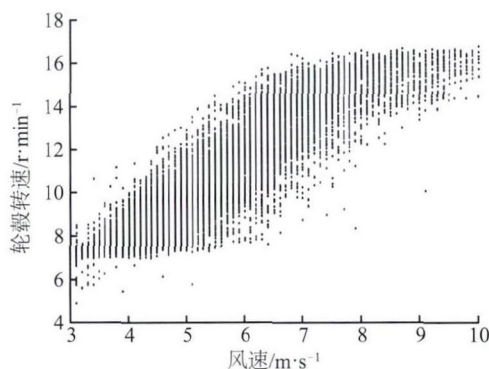


图2 数据筛选后的风速-轮毂转速数据图

Fig. 2 Relationship between wind speed and rotor speed after data screening

5)在  $V=3$  时,选取其风速下所对应的轮毂转速和发电机输出功率数据,并将其定义为  $V_3$ 、 $R_{i3}$ 、 $P_{i3}$  ( $i=1,2,\dots,N_3$ ),其中  $N_3$  为  $V=3$  时所包含的  $R_{i3}$ 、 $P_{i3}$  组数。

6)选取间隔为  $0.3\text{ m/s}$ ,将整段风速划分为若干个离散点,分别定义为  $V_{3+0.3n}$ ,  $R_{i(3+0.3n)}$ ,  $P_{i(3+0.3n)}$  ( $i=1,2,\dots,N_{3+0.3n}$ ,  $n=1,2,\dots$ ),图 3 为提取相应数据后对应的风速-轮毂转速数值图。

7)计算  $R_3 = \sum_{i=1}^{N_3} R_{i3}/N_3$  以及  $P_3 = \sum_{i=1}^{N_3} P_{i3}/N_3$ 。

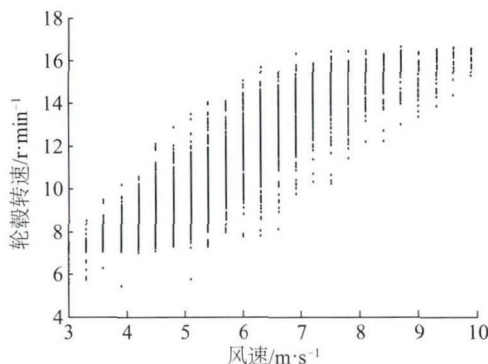


图3 风速离散化后风速-轮毂转速数据图

Fig. 3 Relationship between wind speed and rotor speed after data discretization

8) 利用步骤 7) 的计算方法, 得出  $R_{3+0.3n}$ ,  $P_{3+0.3n}$  ( $n=1,2,\dots$ )。

通过上述方法得出的风速-轮毂转速、发电机输出功率关系图如图 4 所示。

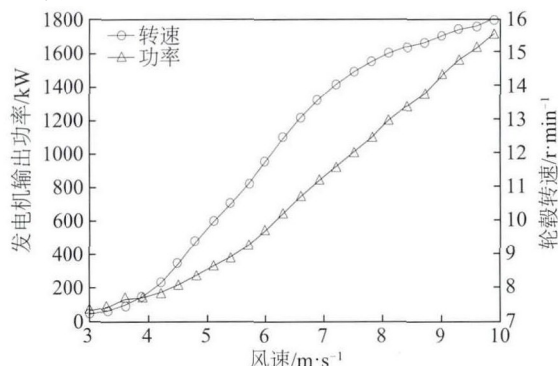


图4 风速-轮毂转速、发电机输出功率关系(平均数法)

Fig. 4 Relationships between wind speed and rotor speed, wind speed and generator power based on mean method

## 2.2 最小二乘法

最小二乘法通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法进行 SCADA 数据预处理的基本思路如下:

1) 提取在平均数法中筛选数据后得到的处理数据  $V$ 、 $R$ 、 $P$ 。

2) 在  $V=3$  时, 选取其风速下所对应的轮毂转速和发电机输出功率数据  $R_{i3}$ 、 $P_{i3}$  ( $i=1,2,\dots,N_3$ )。

3) 令  $R_{i3}'$  和  $P_{i3}'$  为估计值, 并符合要求:  $\min\left(\sum_{i=1}^{N_3}(R_{i3}-R_{i3}')^2\right)$ ,  $\min\left(\sum_{i=1}^{N_3}(P_{i3}-P_{i3}')^2\right)$ 。

4) 重复上述步骤, 求得  $R_{3+0.3n}'$ ,  $P_{3+0.3n}'$ 。

通过上述方法得出的风速-轮毂转速、发电机输出功率关系图如图 5 所示。

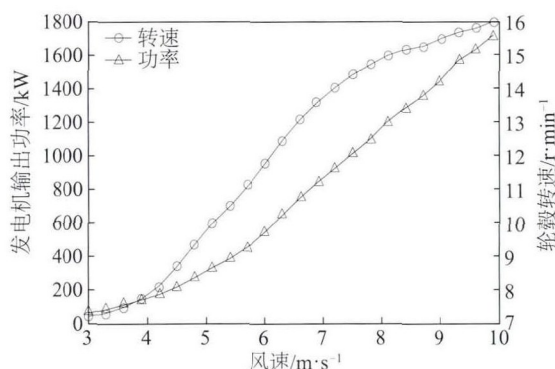


图5 风速-轮毂转速、发电机输出功率关系(最小二乘法)

Fig. 5 Relationships between wind speed and rotor speed, wind speed and generator power based on least square method

## 2.3 非参数法(核密度-均值法)

在上述数据预处理过程中, 未考虑数据非对称分布对处理结果的影响, 由此提出通过建立数据分布特征, 剔除小概率数据再求均值的方法, 即核密度-均值法。在前期经验不足条件下, 难以设定 SCADA 数据服从何种分布, 因此采用非参数核密度估计方法进行分析。核密度估计是在概率论中用来估计未知的密度函数, 无需有关数据分布的先验知识, 对数据分布不预设任何假定, 具有从数据样本自身出发获得数据分布特征的优点。核密度估计的一般定义为:

设  $K(\cdot)$  为核函数,  $h$  为窗宽,  $X_1, X_2, X_3, \dots, X_n$  是一元连续样本, 则在任意点  $x$  处的总体密度函数  $f(x)$  的核密度估计为:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (1)$$

式中,  $K(x) \geq 0$ ,  $\int_{-\infty}^{+\infty} K(x)dx = 1$ 。

数据处理基本思路为:

1) 提取在平均数法中筛选数据后得到的处理数据  $V$ 、 $R$ 、 $P$ 。

2) 令  $V=3$  时, 选取对应的轮毂转速和发电机输出功率数据, 分别为  $R_{i3}$ 、 $P_{i3}$  ( $i=1,2,\dots,N_3$ )。

3) 令  $V=3+0.3n$  ( $n=1,2,\dots,23$ ), 选取对应的轮毂转速和发电机输出功率数据, 分别为  $R_{i(3+0.3n)}$ 、 $P_{i(3+0.3n)}$  ( $i=1,2,\dots,N_{3+0.3n}$ )。

4) 将  $R_{i3}$  进行核密度处理, 依据密度分布, 对偏离较大的数据进行剔除。选取核密度曲线图中  $f(x)$  顶点数值的 10% 为临界值, 舍弃其下方的数值(图 6 为风速为 3 m/s 时的轮毂转速核密度曲线图)。

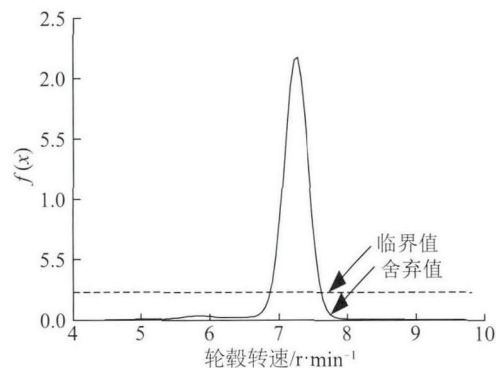


图6 风速为 3 m/s 时的轮毂转速核密度曲线

Fig. 6 Kernel density curve of rotor speed at 3 m/s wind speed



5) 令  $R_{i3}'' (i=1, 2, \dots, N_3')$  为剔除偏离较大数据后的轮毂转速数值,  $P_{i3}'' (i=1, 2, \dots, N_3')$  为其对应的发电机输出功率,  $N_3'$  为  $V=3$  时所包含的  $R_{i3}''$ 、 $P_{i3}''$  组数。

6) 计算  $R_3'' = \sum_{i=1}^{N_3'} R_{i3}'' / N_3'$  及  $P_3'' = \sum_{i=1}^{N_3'} P_{i3}'' / N_3'$ 。

7) 利用步骤 6) 的计算方法, 得出  $R_{3+0.3n}''$ 、 $P_{3+0.3n}'' (n=1, 2, \dots)$ 。

通过上述方法计算得出的风速-轮毂转速、发电机输出功率关系图如图 7 所示。

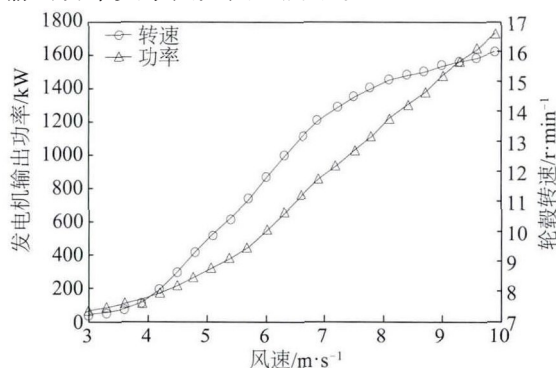


图 7 风速-轮毂转速、发电机输出功率关系  
(核密度-均值法)

Fig. 7 Relationships between wind speed and rotor speed, wind speed and generator power based on kernel density estimation-mean method

### 3 SCADA 数据预处理方法评价策略

为了评价不同数据预处理方法的性能, 需要构建合理的评价策略。本文提出的数据预处理评价有 3 个指标: 物理特性一致性、采样时间变化稳健性、采样频率变化稳健性。

#### 3.1 物理特性一致性

在制定风电机组控制策略时, 理想情况下可大致划分为 3 个区域, 风能利用系数  $C_p$  恒定区、转速恒定区和功率恒定区。 $C_p$  恒定区指在低于额定风速区域, 风电机组在给定的功率-转速曲线控制下, 使风电机组风能利用系数达到最大, 即保持  $C_p$  恒定。转速恒定区指保持  $C_p$  为最大值, 未达到额定功率但转速达到其极限值。功率恒定区指风速大于

额定风速, 需要通过变桨距限制功率为额定值的区域。

在高于额定风速时, 风电机组通过变桨距限制风能捕获, 风能利用系数随风速的增大而减小。因此, 更大的关注点在低于额定风速时如何调节转速以保持风能利用系数为最大。低于额定风速时, 功率与风轮利用系数之间的关系可表达为<sup>[8]</sup>:

$$P = T\omega = \frac{1}{2}C_p\rho\pi R^2v^3 = \frac{1}{2}\rho SC_p[(R/\lambda)\omega]^3 \quad (2)$$

式中,  $T$ ——转矩;  $\omega$ ——风轮转速;  $\rho$ ——空气密度;  $R$ ——风轮半径;  $v$ ——风速;  $S$ ——风轮扫略面积;  $\lambda$ ——叶尖速比。

在  $C_p$  恒定区, 有  $C_p = C_{p\max}$ ,  $\lambda = \lambda_{\text{opt}}$ , 则功率与转速三次方成正比。此时, 转矩与转速关系为:

$$T = \frac{1}{2}\rho SC_p(R/\lambda)^3\omega^2 \quad (3)$$

另一方面, 根据 SCADA 数据做出轮毂转速-发电机输出功率散点图, 如图 8 所示。不难看出, 当发电机输出功率介于 150~800 kW 之间时, 属于  $C_p$  恒定区, 即最大风能跟踪区域<sup>[9]</sup>。

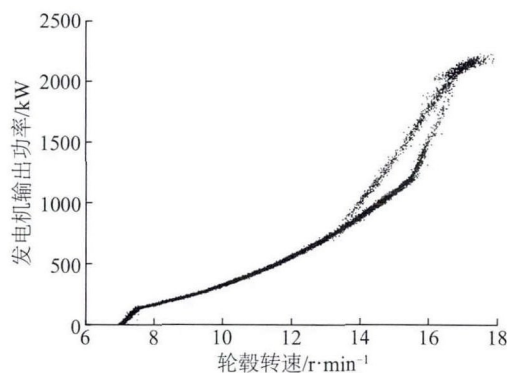


图 8 轮毂转速-发电机输出功率散点图

Fig. 8 Scatter plot of wind speed - generator power

具体到数据处理过程为: 1) 筛选出发电机输出功率介于 150~800 kW 之间 SCADA 数据; 2) 按照上述 3 种数据预处理方法分别来对筛选的数据进行预处理(表 2 为不同的数据预处理方法处理后的功率值); 3) 基于三次多项式  $P = f_n v^3 (n=1, 2, 3)$  来对计算出的散点进行最小二乘法曲线拟合。

根据上述步骤, 得到平均数法  $f_1 = 2.47$ , 最小二乘法  $f_2 = 2.48$ , 核密度-均值法  $f_3 = 2.41$ 。

表2 不同数据预处理方法处理后的功率(kW)

Table 2 Power values (kW) under different data preprocessing methods

处理方法	风速/ $\text{m}\cdot\text{s}^{-1}$															
	4.5	4.6	4.7	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
平均数法	236	249	265	280	299	316	333	348	365	385	404	427	442	466	483	505
最小二乘法	237	250	265	281	299	316	333	349	365	385	404	428	443	466	484	506
非参数法	231	244	260	276	295	310	327	343	361	382	403	429	446	470	488	511

为了获得一致性评价的定量指标,设计评价计算公式如下:

$$S_n = \sum_{i=4.5}^6 (P_i - f_n v^3)^2 \quad (i=4.5, 4.6, \dots, 6; n=1, 2, 3) \quad (4)$$

式中,  $P_i$ ——数据预处理后的功率值;  $v$ ——风速。

计算后,  $S_1 = 1353.90$ ,  $S_2 = 1450.20$ ,  $S_3 = 633.70$ 。因此在上述3种方法中,核密度-均值法是进行物理特性评价的最优方法。

### 3.2 采样时间变化稳健性

自然风是随机变化的,在不同采样时段具有不同特征。良好的数据预处理方法应对采样时间的变化具有较好的稳健性。

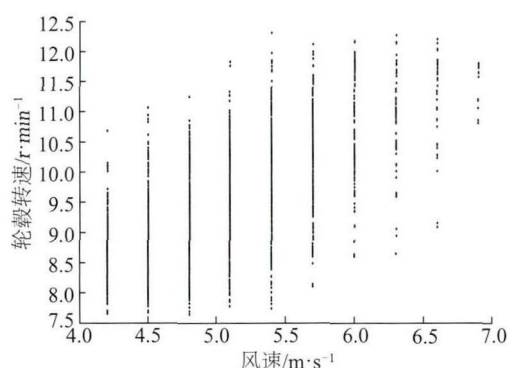
选取不同时段的两万组风况相近的数据进行分析,经过数据筛选,得到处理后的SCADA数据(图9为处理后的待预处理数据)。再将这些数据分别用平均数法、最小二乘法、核密度估计法进行进一步数据预处理。表3为用以上3种数据预处理方法处理后的转速、功率值。

为了获得采样时间变化稳健性的定量指标,设计评价计算公式如式(5):

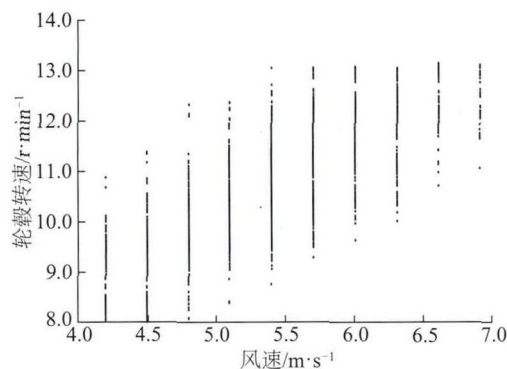
$$\begin{cases} S'_n = \sum_{i=4.2}^{6.9} (P_{1ni} - P_{2ni})^2, (i=4.2, 4.5, 4.8, \dots, 6.9; n=1, 3, 5) \\ S'_n = \sum_{i=4.2}^{6.9} (R_{1ni} - R_{2ni})^2, (i=4.2, 4.5, 4.8, \dots, 6.9; n=2, 4, 6) \end{cases} \quad (5)$$

式中,  $P_{1ni}$ 、 $P_{2ni}$ ——用3种数据预处理方法得到的前后取样时间段功率值;  $R_{1ni}$ 、 $R_{2ni}$ ——轮毂转速值;  $S'_n$ ——计算出的结果。

根据式(5),可计算得到,  $S'_1 = 1.06 \times 10^5$ ,  $S'_2 = 7.09$ ,  $S'_3 = 1.05 \times 10^5$ ,  $S'_4 = 7.29$ ,  $S'_5 = 9.88 \times 10^4$ ,  $S'_6 = 6.87$ ,  $S'_5$ 相较于 $S'_1$ 、 $S'_3$ 偏小,  $S'_6$ 相较于 $S'_2$ 、 $S'_4$ 偏小。可见核密度-均值法的采用时间变化稳健性更好。



a. 前段时间



b. 后段时间

图9 不同采样时间段的待预处理数据

Fig. 9 Data waiting for preprocessing under different sampling time period

### 3.3 采样频率变化稳健性

不同SCADA数据的采样频率可能不同,数据预处理方法应对采样频率的变化具有良好的稳健性。文中所用原始风电场SCADA数据的采样频率为1 Hz,将现有SCADA数据按照时间顺序,间隔5 s(0.2 Hz)选取数据,以模拟不同采样频率。两组数据分别运用上述3种方法进行数据预处理。处理后的轮毂转速、发电机输出功率部分值如表4所示。

表3 不同采样时间段数据处理后的转速、功率值

Table 3 Rotor speed and power values after data preprocessing under different sampling time periods

采样时段	风速/ $\text{m}\cdot\text{s}^{-1}$	轮毂转速/ $\text{r}\cdot\text{min}^{-1}$			发电机输出功率/ $\text{kW}$		
		平均数法	最小二乘法	核密度估计法	平均数法	最小二乘法	核密度估计法
前段 时间	4.2	8.61	8.61	8.58	208.84	209	206.12
	4.5	8.83	8.84	8.80	225.45	225	221.86
	4.8	9.19	9.19	9.18	255.88	256	255.04
	5.1	9.62	9.63	9.62	294.17	294	293.39
	5.4	10.00	10.01	9.99	329.98	330	327.62
	5.7	10.23	10.23	10.24	352.63	353	352.78
	6.0	10.62	10.60	10.66	392.64	393	397.25
	6.3	10.88	10.88	10.92	421.50	422	425.82
	6.6	11.23	11.25	11.32	462.52	466	471.13
	6.9	11.35	11.20	11.35	173.96	457	473.96
后段 时间	4.2	8.78	8.76	8.76	221.88	220	219.97
	4.5	9.38	9.38	9.34	270.52	271	266.42
	4.8	9.81	9.81	9.83	308.73	309	308.91
	5.1	10.25	10.24	10.21	352.13	352	347.40
	5.4	10.72	10.71	10.71	403.72	404	403.04
	5.7	11.26	11.25	11.27	483.60	469	469.08
	6.0	11.70	11.70	11.72	523.52	524	526.76
	6.3	11.96	11.95	11.96	557.63	558	558.34
	6.6	12.28	12.28	12.29	601.78	602	603.10
	6.9	12.36	12.33	12.38	612.15	610	614.78

表4 不同采样频率条件下轮毂转速、发电机输出功率部分值

Table 4 Rotor speed and power values after data preprocessing under different sample frequencies

采样频率	风速/ $\text{m}\cdot\text{s}^{-1}$	轮毂转速/ $\text{r}\cdot\text{min}^{-1}$			发电机输出功率/ $\text{kW}$		
		平均数法	最小二乘法	核密度估计法	平均数法	最小二乘法	核密度估计法
1 s	3.0	7.22	7.22	7.25	64.99	65	66.14
	3.9	7.75	7.74	7.66	142.70	143	136.16
	4.8	9.36	9.36	9.32	273.50	273	269.48
	5.7	11.08	11.09	11.07	455.81	456	449.84
	6.6	13.07	13.06	13.15	746.85	747	755.17
	7.5	14.41	14.41	14.49	1015.17	1015	1026.49
	8.4	15.14	15.15	15.22	1286.68	1287	1303.87
	9.3	15.67	15.66	15.67	1565.60	1566	1565.60
	9.9	15.99	15.99	16.02	1720.35	1720	1731.23
5 s	3.0	7.23	7.23	7.25	66.08	66	67.20
	3.9	7.75	7.76	7.65	143.25	143	135.08
	4.8	9.38	9.38	9.36	274.76	274	272.22
	5.7	11.05	11.06	11.05	451.21	451	447.61
	6.6	12.89	12.89	13.00	718.23	718	731.93
	7.5	14.44	14.44	14.48	1011.73	1011	1015.71
	8.4	15.14	15.14	15.22	1274.89	1273	1292.25
	9.3	15.66	15.65	15.63	1554.13	1563	1565.69
	9.9	16.02	16.02	16.06	1721.49	1707	1735.16

为了获得采样时间变化稳健性的定量指标,设计评价计算公式如式(6):

$$\begin{cases} S_n'' = \sum_{i=3}^{9.9} (P_{1ni} - P_{2ni})^2 \quad (i=3.0, 3.3, 3.9, \dots, 9.9; n=1, 3, 5) \\ S_n'' = \sum_{i=3}^{9.9} (R_{1ni} - R_{2ni})^2 \quad (i=3.0, 3.3, 3.9, \dots, 9.9; n=2, 4, 6) \end{cases} \quad (6)$$

式中,  $P_{1ni}$ 、 $P_{2ni}$ ——1 s、5 s 采样时用以上 3 种数据预处理方法得到的功率值;  $R_{1ni}$ 、 $R_{2ni}$ ——1 s、5 s 采样时用以上 3 种数据预处理方法得到的轮毂转速值;  $S_n''$ ——计算出的结果。

通过计算得到,  $S_1'' = 2.31 \times 10^3$ ,  $S_2'' = 0.05$ ,  $S_3'' = 2.86 \times 10^3$ ,  $S_4'' = 0.05$ ,  $S_5'' = 1.80 \times 10^3$ ,  $S_6'' = 0.04$ ,  $S_5''$  相较于  $S_1''$ 、 $S_3''$  偏小,  $S_6''$  相较于  $S_2''$ 、 $S_4''$  偏小。计算结果表明即核密度-均值法在采样频率的稳健性方面,表现得更好。

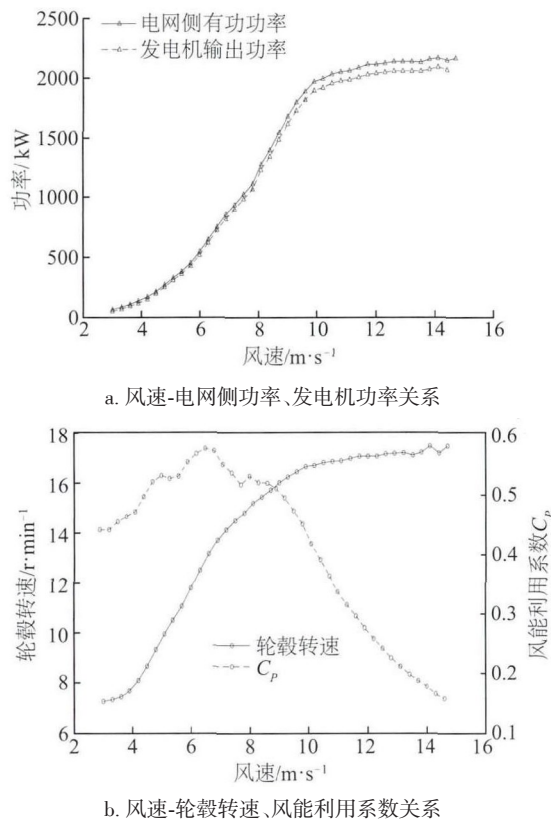
通过上述 3 种评价标准的分析可知,核密度-均值法是相较于平均数法和最小二乘法更好的数据预处理方法。

## 4 基于数据预处理的系统物理特性分析

基于上述分析结论,选择核密度-均值法对全工况下风电机组 SCADA 数据进行预处理,利用处理后的数据分析风电机组实际的现场运行特性。图 10 为基于数据预处理的风电机组现场运行特性图。经过数据预处理后,风电机组风速与输出功率关系转化为单值映射关系,如图 10a 所示。在额定风速以下,随着风速的增大输出功率快速增大;在超过额定风速以后,输出功率被控制在设定功率 2 MW 附近。对比风电机组风速-发电机输出功率关系和风速-变频器电网侧有功功率关系可知,发电机输出功率与变频器电网侧功率变化趋势一致,但发电机输出功率略大于变频器电网侧功率,这是因为经过全功率变流器后,功率有一定损失。

从图 10b 中风速与轮毂转速关系可看出,在额定风速以下,随着风速的增大轮毂转速上升,该阶段主要通过控制转速与风速之间的关系获得最佳叶尖速比,从而提高风能利用率。在额定风速以上,风轮转速基本保持在设定值附近,通过变桨距实现功率控制。从图 10b 中风电机组风能利用系数曲线可得如下结论:总的趋势是在额定风速以

下,风能利用系数较高;在额定风速以上,风能利用系数随风速的增大逐步降低。从计算结果来看,风能利用系数最大可达 0.55。需要说明的是,计算过程中采用的风速数据来自风机舱上的风速仪,由于叶片转动的影响,该检测数据比风轮前沿的风速低。因此,若以风轮前沿风速为计算基准,风能利用系数数值会略低。



a. 风速-电网侧功率、发电机功率关系

b. 风速-轮毂转速、风能利用系数关系

图 10 风电机组现场运行特性图

Fig. 10 The relationship with several parameters for wind turbine

## 5 结论

1) 分别将平均数法、最小二乘法和非参数法(核密度-均值法)用于风电场 SCADA 数据预处理,提出相应的处理步骤,获得了不同数据预处理结果;

2) 提出风电场 SCADA 数据预处理效果的评价指标,包括物理特性一致性、采样时间变化稳健性及采样频率变化稳健性;

3) 提出各种评价指标定量计算公式,基于这些计算公式可得到对应不同指标的评价值,计算结果表明:核密度-均值法是相较于平均数法和最小二乘



法更好的数据预处理方法;

4) 基于核密度-均值法对全工况风电机组 SCADA 数据进行预处理, 利用处理后的数据分析风速与输出功率、轮毂转速的关系, 并计算风能利用系数, 结果可用于实际风电场风电机组性能的评价。

### [参考文献]

- [1] Kusiak A, Li W. The prediction and diagnosis of wind turbine faults [J]. *Renewable Energy*, 2011, 36 (1): 16—23.
- [2] Feng Y, Qiu Y, Crabtree C J, et al. Use of SCADA and CMS signals for failure detection and diagnosis of a wind turbine gearbox [A]. *European Wind Energy Conference and Exhibition 2011, EWEC 2011 [C]*, Sheffield, 2011, 17—19.
- [3] Kim K, Parthasarathy G, Uluyol O, et al. Use of SCADA data for failure detection in wind turbines [A]. *ASME 2011 5th International Conference on Energy Sustainability. American Society of Mechanical Engineers [C]*, Washington, 2011, 2071—2079.
- [4] Zhai Yongjie, Wang Dongfeng, Zhang Junying, et al. Research on early fault diagnostic method of wind turbines [J]. *Telkomnika Indonesian Journal of Electrical Engineering*, 2013, 11(5): 2330—2341.
- [5] Zaher A, Mcarthur S, Infield D, et al. Online wind turbine fault detection through automated SCADA data analysis [J]. *Wind Energy*, 2009, 12(6): 574—593.
- [6] Schlechtingen M, Santos I F, Achiche S. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description [J]. *Applied Soft Computing*, 2013, 13(1): 259—270.
- [7] Yang W, Court R, Jiang J. Wind turbine condition monitoring by the approach of SCADA data analysis [J]. *Renewable Energy*, 2013, 53: 365—376.
- [8] Dai J C, Hu Y P, Liu D S, et al. Aerodynamic loads calculation and analysis for large scale wind turbine based on combining bem modified theory with dynamic stall model [J]. *Renewable Energy*, 2011, 36 (3): 1095—1104.
- [9] 姚 骏, 廖 勇, 瞿兴鸿. 直驱永磁同步风力发电机的最佳风能跟踪控制 [J]. *电网技术*, 2008, 32(10): 11—15.
- [9] Yao Jun, Liao Yong, Qu Xinghong. Optimal wind-energy tracking control of direct-driven permanent magnet synchronous generators for wind turbines [J]. *Power System Technology*, 2008, 32(10): 11—15.

## DATA PRE-PROCESSING METHOD AND ITS EVALUATION STRATEGY OF SCADA DATA FROM WIND FARM

Dai Juchuan<sup>1</sup>, Cao Junwei<sup>1</sup>, Zhang Fan<sup>1</sup>, Liu Deshun<sup>1</sup>, Shen Xiangbing<sup>2</sup>

(1. School of Electromechanical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;

2. XEMC Windpower Co., Ltd., Xiangtan 411000, China)

**Abstract:** In order to obtain more explicit physical information from the SCADA data and judge the operating state of wind turbines better, the mean value method, least square method and non-parametric method (kernel density-mean value method) are used to pre-process SCADA data in wind farm. Three evaluation indexes for pre-processing algorithm are presented, including: (1) the consistency of physical characteristics; (2) the robustness of the sampling time; (3) the robustness of the sampling frequency. The quantitative calculation formulas of the evaluation indexes are designed to evaluate the effect of various kinds of pre-processing algorithms. The results showed that the kernel density-mean value method can obtain better results. Finally, based on kernel density-mean method the SCADA data in full working condition are pre-processed, operating characteristics of wind turbines are analysed, including the relationship between the wind speed and output power, rotational speed of the wind rotor and the wind energy utilization coefficient.

**Keywords:** wind turbine; SCADA data; data pre-processing; evaluation index