

# 基于证据理论的风速不确定性建模

郭小璇<sup>1,2</sup>, 龚仁喜<sup>1</sup>, 鲍海波<sup>3</sup>

(1. 广西大学 电气工程学院, 广西 南宁 530004; 2. 广西电网电力科学研究院, 广西 南宁 530023;

3. 广西电网南宁供电局, 广西 南宁 530031)

**摘要:** 风速决定了风电场的输出功率, 风速模型是研究含风电系统运行与规划的重要基础。提出一种基于证据理论的风速不确定建模方法。采用证据理论中基本可信度分配的概念描述风速; 提出依据实测历史数据确定基本可信度分配的焦元和信任函数的实现方法, 并设计等概率区间和等取值区间 2 种建模策略。以某风电场实测风速为例对基于所建模型和基于概率分布及区间分布的模型进行仿真比较, 结果表明所提模型能确定风速的似然累积概率分布和信任累积概率分布, 能更有效地描述和处理风速不确定性信息。

**关键词:** 风速; 风力发电; 证据理论; 基本可信度分配; 似然累积概率分布; 信任累积概率分布

**中图分类号:** TM 614

**文献标识码:** A

**DOI:** 10.16081/j.issn.1006-6047.2019.01.012

## 0 引言

发展风电是优化电力系统电源侧能源结构、减小碳排放量的重要手段之一, 但是风电的功率输出由风速决定而具有不确定性。为了分析电力系统不确定运行与规划, 需要建立更为合理的风速不确定性模型。

现有文献通常将风电场风速建模为随机变量<sup>[1-5]</sup>、模糊变量<sup>[6-7]</sup>或区间变量<sup>[8-10]</sup>等。文献[1-5]中假设风电场风速为服从 Weibull 分布、Rayleigh 分布、高斯分布、Beta 分布等的随机变量, 并通过极大似然估计等方法逼近获得随机分布的参数, 从而将风速建模为明确的随机变量。但是由于季节、时段、天气等因素不同对实际风速影响很大, 风速的随机特征具有较强的复杂性, 获得准确的风速概率分布函数及其参数通常较为困难。文献[6-7]基于模糊集理论方法将风速描述为模糊变量, 采用隶属度函数来表达各时段风速的不确定性, 需经过一定处理, 才能用于电力生产实际。文献[8-10]只关注风速波动的最大和最小边界值, 将风速表达为区间变量, 建模较容易实现。但其所蕴含的信息较少, 对电力系统运行参数描述的深度有限。

针对以上 3 种不确定量建模方式的缺陷, 文献[11]同时考虑了日风速的随机性和模糊性, 将风速 Weibull 分布的形状参数和尺度参数定义为模糊隶属函数, 建立了风速的随机模糊不确定模型。文献[12]提出了一种双峰 Weibull 分布以处理不规则的风速概率分布, 但其模拟精度不高, 限制了其实际应用。文献[13]基于证据理论描述风电不确定性, 构建了考虑风电的概率区间潮流模型, 但没有给出风

速证据理论建模的实现方法。本文将证据理论应用于风速建模, 提出了一种新的风速不确定性模型。

本文首先阐述了基于证据理论描述不确定量的理论基础, 将证据理论中可信度分配的概念用于描述风速的不确定性, 并将风速可信度分配的焦元设计为概率型数据结构; 然后, 依据风电场实测风速的历史数据, 提出了确定可信度分配的焦元和信任函数的实现方法, 并设计了等概率区间和等取值区间 2 种建模策略; 最后, 以国内某风电场实测风速为例, 对基于所建模型和基于概率分布及区间分布的模型进行了仿真计算及比较, 结果表明所提模型能确定风速的似然累积概率分布和信任累积概率分布, 包含了风速实际概率分布的可能情况, 更有效地描述和处理了风速不确定性信息, 更符合人们的思维习惯。

## 1 证据理论原理

证据理论<sup>[14-16]</sup>是建立在辨识框架  $\Theta$  上的一种不确定理论, 幂集  $2^\Theta$  是该辨识框架内所有子集的集合, 证据理论是对幂集元素进行基本概率赋值。传统概率理论表达不确定时, 将概率值分配到单个事件上, 建立单点到  $[0, 1]$  的映射。类似于传统概率论中有限的样本空间, 证据理论建立的是从集合到  $[0, 1]$  的映射。定义质量函数  $m$  的一个从集合  $2^\Theta$  到  $[0, 1]$  的映射,  $A$  表示辨识框架  $\Theta$  的任一子集, 记作  $A \subseteq \Theta$ , 且满足条件:

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \subseteq \Theta} m(A) = 1 \end{cases} \quad (1)$$

则也称  $m(A)$  为  $A$  的基本信任分配 BPA (Basic Probability Assignment) 函数, 表示证据对事件  $A$  发生信任程度。对于子集  $A$ , 只要有  $m(A) > 0$ , 则称  $A$  为焦元。

收稿日期: 2017-12-08; 修回日期: 2018-11-12

基金项目: 国家自然科学基金资助项目 (61561007)

Project supported by the National Natural Science Foundation of China (61561007)

由于事件  $A$  的知识信息(统计数据等)缺乏,对  $A$  的信任程度并不能推导出对事件  $\bar{A}$  的信任程度,即  $m(A) + m(\bar{A}) \leq 1$ ,这与概率论中  $p(A) + p(\bar{A}) = 1$  的性质是不同的。根据这一特性,证据理论定义了不确定区间的概念,以一个概率区间而不是单一概率值来表达事件的不确定程度,如图 1 所示。

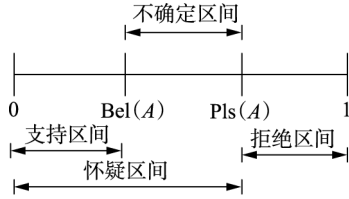


图 1 对事件  $A$  的不确定描述

Fig.1 Uncertainty description of event  $A$

对于命题  $A$ ,不确定区间的上、下限分别为似然函数  $\text{Pls}(A)$  和信任函数  $\text{Bel}(A)$ ,信任函数表示对  $A$  为真的信任程度,似然函数表示对  $A$  为非假的信任程度,二者分别定义如下:

$$\begin{cases} \text{Bel}(A) = \sum_{B \subseteq A} m(B) \\ \text{Pls}(A) = 1 - \text{Bel}(\bar{A}) = \sum_{B \cap A \neq \emptyset} m(B) \end{cases} \quad (2)$$

信任函数表示事件  $A$  发生的最小可能,而似然函数表示事件  $A$  发生的最大可能。

## 2 风速的基本可信度分配

### 2.1 风速的可信度分配

文献[15]按照证据的数学表现形式分类,将证据分为概率型、互补型、嵌套型和一般型 4 种类型。其中概率型结构也称为贝叶斯结构,就是将每一个基本元素赋予基本可信度分配。根据风速不确定特征,可采用概率型数据结构进行建模。风速可信度分配为:

$$v = \begin{cases} [v_1, v_2] & m = m_1 \\ [v_2, v_3] & m = m_2 \\ \vdots & \\ [v_k, v_{k+1}] & m = m_k \\ \vdots & \\ [v_{N_w}, v_{N_w+1}] & m = m_{N_w} \end{cases} \quad (3)$$

其中,风电场风速  $v$  的波动区间被分割为  $N_w$  个子区间,即  $v$  的可信度分配焦元,  $m_1, m_2, \dots, m_{N_w}$  为对应焦元的可信度。式(3)表达的含义是风速取值在焦元区间  $[v_k, v_{k+1}]$  内的可信度为  $m_k$ 。显然,风速可信度分配类似于经典概率论,每个焦元的可信度就类似于风速取值在该焦元区间的概率。

按照式(2),就可以由式(3)获得风速不同取值的信任函数和似然函数,进而获得风速的信任累积

概率分布和似然累积概率分布。以风速  $v \leq v_c$  的事件为例,  $v_c$  是介于  $v_k$  和  $v_{k+1}$  之间的任意值,则信任函数  $\text{Bel}(v \leq v_c)$  和似然函数  $\text{Pls}(v \leq v_c)$  为:

$$\begin{cases} \text{Bel}(v \leq v_c) = \sum_{i=1}^{k-1} m_i \\ \text{Pls}(v \leq v_c) = 1 - \sum_{i=k+1}^{N_w} m_i = \sum_{i=1}^k m_i \end{cases} \quad (4)$$

那么风速  $v \leq v_c$  发生的实际概率  $\text{Pr}(v \leq v_c)$  将介于  $\text{Bel}(v \leq v_c)$  和  $\text{Pls}(v \leq v_c)$  之间,即  $\text{Bel}(v \leq v_c) \leq \text{Pr}(v \leq v_c) \leq \text{Pls}(v \leq v_c)$ 。将  $v_c$  取值从  $v_1$  逐渐增加至  $v_{N_w+1}$ ,即可逐步构造出风速的信任累积概率分布和似然累积概率分布。

由式(3)可见,风速的可信度分配构造具有很大灵活性。不同科研人员将提出不同的数据组合,从而构造出不同的可信度分配模型,这些模型可能是互补的或对立的。传统概率理论难以处理该情况,而证据理论则可以按照一定的合成规则来综合考虑,将不同模型融合成一个一致性的可信度描述,提高模型的可用性。现有的合成规则有很多,具体可参考文献[16]进行选择。

### 2.2 2 种特殊建模策略

构建风速的可信度分配模型式(3),需要确定风速各个焦元区间的取值  $v_1, v_2, \dots, v_{N_w+1}$  和对应的可信度  $m_1, m_2, \dots, m_{N_w}$ 。按照证据理论的原理,不确定量建模可根据需要随意选择焦元区间和可信度。但为了方便实际应用,通常采用 2 种特殊的建模策略:等取值区间建模和等概率区间建模。

等取值区间建模策略就是定义每个焦元区间的半径相同来确定各个焦元对应可信度。风速可信度模型式(3)可转化为如下形式:

$$v = \begin{cases} [v_1, v_1+d] & m = m_1 \\ [v_1+d, v_1+2d] & m = m_2 \\ \vdots & \\ [v_1+(k-1)d, v_1+kd] & m = m_k \\ \vdots & \\ [v_1+(N_w-1)d, v_1+N_w d] & m = m_{N_w} \end{cases} \quad (5)$$

其中,  $d$  为各个区间宽度。显然,采用这种建模策略需要先确定参数  $v_1, v_2, \dots, v_{N_w+1}$ ,再确定参数  $m_1, m_2, \dots, m_{N_w}$ 。

图 2 是等取值区间建模得到的风速基本可信度分配示意图。图中每个矩形的宽度相同,高度和面积不同。可信度  $m_k$  的物理意义可认为是每个矩形的面积。图中  $m_{w_i,k}$  为第  $i$  个风电场风速第  $k$  个焦元区间的可信度。

等概率区间建模策略是定义每个焦元区间的可信度相同来确定各个焦元区间的边界。风速可信度

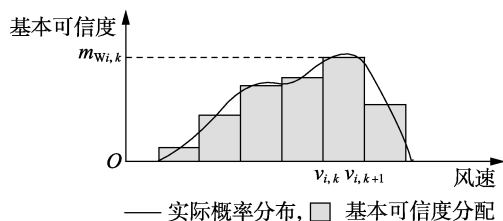


图 2 等取值区间可信度分配

Fig.2 Probability assignment of equal value interval  
模型式(3)可转化为如下形式:

$$v = \begin{cases} [v_1, v_2] & m = m_1 \\ [v_2, v_3] & m = m_2 \\ \vdots & \vdots \\ [v_k, v_{k+1}] & m = m_k \\ \vdots & \vdots \\ [v_{N_w}, v_{N_w+1}] & m = m_{N_w} \end{cases} \quad (6)$$

$$m_1 = m_2 = \dots = m_{N_w} = M \quad (7)$$

其中,各个焦元的可信度相等,均为  $M$ 。与等取值区间建模不同,这里需要先确定参数  $m_1, m_2, \dots, m_{N_w}$ ,再确定参数  $v_1, v_2, \dots, v_{N_w+1}$ 。

图 3 是等概率区间建模得到的风速基本可信度分配示意图,图中每个矩形的宽度和高度都不相等,而面积相等,即各焦元区间的可信度相等。

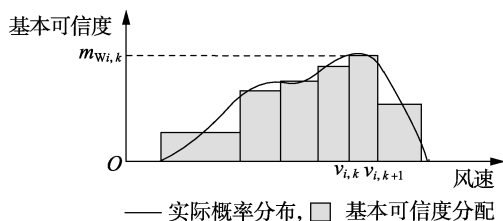


图 3 等概率区间可信度分配

Fig.3 Probability assignment of equal probability interval

### 3 风速建模的具体实现

构造式(3)的风速基本可信度描述,需要确定风速各焦元区间的取值  $v_1, v_2, \dots, v_{N_w+1}$  和对应可信度  $m_1, m_2, \dots, m_{N_w}$ 。基于已掌握的  $H$  个数据点的实测风速序列,风速建模的基本思路如下。

(1) 选择风速基本可信度分配的构建策略,可选择 2.2 节等取值区间或等概率区间原则。

(2) 求出实测风速序列的最大风速  $v_{\max}$  和最小风速  $v_{\min}$ 。

(3) 根据实测数据点数量  $H$  的大小和系统分析模型的需求,确定焦元区间的数量  $N_w$ 。按照等取值区间或等概率区间原则,确定风速可信度分配的焦元和可信度。

a. 若采用等取值区间策略,可确定  $v_1$  和  $d$  为:

$$v_1 = v_{\min} \quad (8)$$

$$d = \frac{v_{\max} - v_{\min}}{N_w} \quad (9)$$

将  $v_1$  和  $d$  代入式(5)即可确定  $v_2, \dots, v_{N_w+1}$ ,从而确定各焦元区间。然后可根据风速实测数据在各个焦元区间  $[v_k, v_{k+1}]$  ( $k=1, 2, \dots, N_w$ ) 的分布情况,确定各焦元区间的可信度分配。假设各个焦元区间内的实测风速点的数量分别为  $H_k$  ( $k=1, 2, \dots, N_w$ )。那么各焦元区间的可信度分配为:

$$m_k = \frac{H_k}{H} \quad k=1, 2, \dots, N_w \quad (10)$$

最终确定风速的可信度分配。

b. 若采用等概率区间策略,可确定  $v_1$  和  $M$  为:

$$v_1 = v_{\min} \quad (11)$$

$$M = \frac{1}{N_w} \quad (12)$$

那么每个焦元区间内的实测风速数据点数量相等,具体为:

$$H_k = MH = \frac{H}{N_w} \quad k=1, 2, \dots, N_w \quad (13)$$

然后采用递推方法逐步获得  $v_2, \dots, v_{N_w+1}$ 。将  $v_1, v_2, \dots, v_{N_w+1}$  和  $M$  代入式(6)、(7)即可确定风速的可信度分配。

(4) 当风速有不同数据来源时,可以考虑采用文献[16]的合成方法来融合以获得一个统一的风速基本可信度分配。

## 4 计算结果分析

### 4.1 数据概况

计算机为 IBM-PC 兼容机, CPU 主频为  $2 \times 2.19$  GHz, 内存为 3 GB。本文采用的数据是新疆地区的一座风能观测塔接近 1 a (从 2015 年 10 月 16 日到 2016 年 10 月 7 日) 的资料, 风速测量频率为每小时 1 次, 共 8 082 个风速测量数据。

### 4.2 风速的证据理论模型分析

分析风速的数据序列, 获得风速最大值  $v_{\max} = 18.48$  m/s 和最小值  $v_{\min} = 0.35$  m/s。选择  $N_w = 8$ , 即风速基本可信度分配的焦元数为 8, 采用本文所提的等取值区间原则和等概率区间原则建立风速基本可信度分配如表 1 所示。

图 4 是根据表 1 结果构建出的风速基本可信度分配图。图中实际概率密度分布是依据风速测量数据统计获得的实际分布情况, 拟合概率密度分布是按照文献[1]方法拟合出的 Weibull 分布。对比图 4 (a)、(b) 可知, 等取值区间原则获得的基本可信度分配每个焦元的横坐标区间宽度相同, 即风速各焦元的取值区间相等; 而等概率区间原则获得的基本



表 1 8 个焦元的风速基本可信度分配

Table 1 BPA of wind speed with eight focal elements

| 等取值区间原则                                   |           | 等概率区间原则                                   |           |
|-------------------------------------------|-----------|-------------------------------------------|-----------|
| 风速区间/<br>( $\text{m}\cdot\text{s}^{-1}$ ) | 可信度<br>分配 | 风速区间/<br>( $\text{m}\cdot\text{s}^{-1}$ ) | 可信度<br>分配 |
| [0.35, 2.62]                              | 0.196 5   | [0.35, 2.03]                              | 0.125 0   |
| [2.62, 4.88]                              | 0.311 4   | [2.03, 3.01]                              | 0.125 0   |
| [4.88, 7.15]                              | 0.263 5   | [3.01, 3.90]                              | 0.125 0   |
| [7.15, 9.42]                              | 0.137 3   | [3.90, 4.83]                              | 0.125 0   |
| [9.42, 11.68]                             | 0.058 0   | [4.83, 5.76]                              | 0.125 0   |
| [11.68, 13.95]                            | 0.022 8   | [5.76, 6.88]                              | 0.125 0   |
| [13.95, 16.21]                            | 0.008 4   | [6.88, 8.67]                              | 0.125 0   |
| [16.21, 18.48]                            | 0.002 0   | [8.67, 18.48]                             | 0.125 0   |

可信度分配每个焦元的矩形面积相同,即风速各焦元的可信度相等。

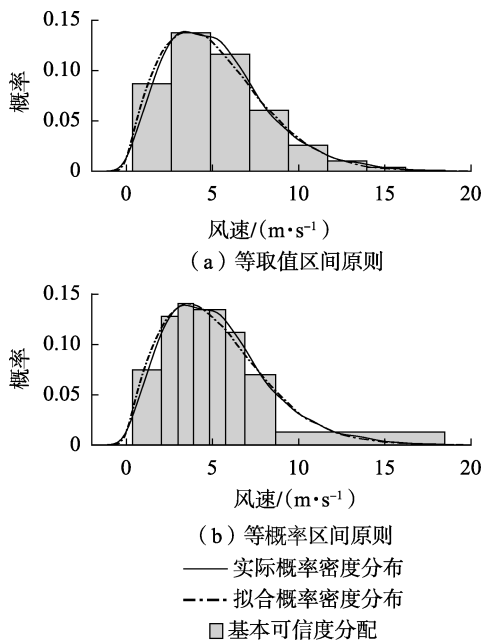


图 4 风速的可信度分配

Fig.4 Probability assignment of wind speed

图 5 和图 6 中给出了由表 1 结果合成的信任累积概率分布和似然累积概率分布。显然,根据信任和似然累积概率分布可以分别确定风速任意取值事

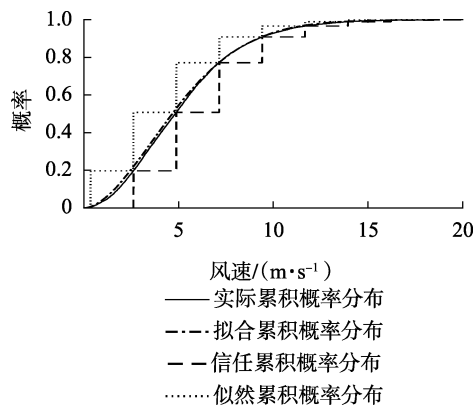


图 5 基于等取值区间策略的风速累积概率分布

Fig.5 Cumulative probability distribution of wind speed based on equal value interval strategy

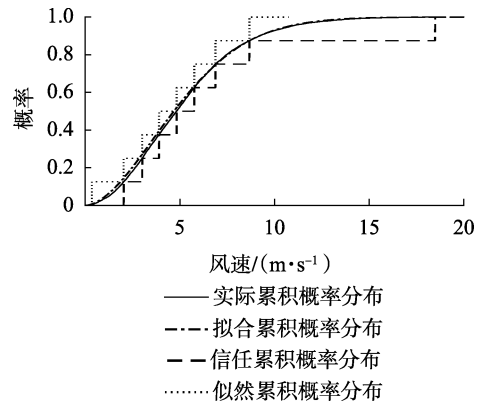


图 6 基于等概率区间策略的风速累积概率分布

Fig.6 Cumulative probability distribution of wind speed based on equal probability interval strategy

件对应的最小概率和最大概率,即风速取值事件发生的概率区间。

由图 5 和图 6 获得风速不同取值事件对应的概率区间如表 2 所示。以风速取值小于 4 m/s 为例,依据实测数据统计出实际概率为 0.390 2,依据文献 [1] 方法拟合的概率分布得到拟合概率是 0.408 8。虽然拟合结果得到的概率非常接近真实值,但仍然存在一定误差,并非正确值。而采用本文基本可信度模型得到概率区间为 [0.350 0, 0.507 9] 和 [0.375 0, 0.500 0], 实际概率 0.390 2 均在概率区间之内。显然,本文模型得到的结果是正确的。

表 2 风速取值的概率区间

Table 2 Probability interval of wind speed

| 风速取值/<br>( $\text{m}\cdot\text{s}^{-1}$ ) | 概率区间               |                    | 实际<br>概率 | 拟合<br>概率 |
|-------------------------------------------|--------------------|--------------------|----------|----------|
|                                           | 等取值区间原则            | 等概率区间原则            |          |          |
| < 2                                       | [0, 0.350 0]       | [0, 0.125 0]       | 0.121 3  | 0.140 5  |
| < 4                                       | [0.350 0, 0.507 9] | [0.375 0, 0.500 0] | 0.390 2  | 0.408 8  |
| < 6                                       | [0.507 9, 0.771 4] | [0.625 0, 0.750 0] | 0.655 2  | 0.659 2  |
| < 8                                       | [0.771 4, 0.908 7] | [0.750 0, 0.875 0] | 0.837 1  | 0.833 3  |
| < 10                                      | [0.908 7, 0.966 7] | [0.875 0, 1.000 0] | 0.928 5  | 0.931 7  |
| < 12                                      | [0.966 7, 0.989 5] | [0.875 0, 1.000 0] | 0.970 9  | 0.975 4  |
| < 14                                      | [0.989 5, 0.997 9] | [0.875 0, 1.000 0] | 0.990 1  | 0.992 1  |

#### 4.3 焦元数量对结果的影响

对于系统中第  $i$  个风电场,按照等取值区间建模原则,获得的 2、4、6 个焦元(即  $N_{wi}=2$ 、 $N_{wi}=4$ 、 $N_{wi}=6$ )风速可信度分配模型,构建不同基本可信度分配构建的信任和似然累积概率分布情况分别如图 7 和 8 所示。图中信任分布-2、信任分布-4、信任分布-6、信任分布-8 分别是 2、4、6、8 个焦元的信任累积概率分布,似然分布-2、似然分布-4、似然分布-6、似然分布-8 分别是 2、4、6、8 个焦元的似然累积概率分布。随着焦元数量的增多,似然和信任累积概率分布逐渐靠近。可以预见,当焦元数量接近无穷大时,似然累积概率分布和信任累积概率分布将重合为风速实际累积概率分布。

由图 7 和 8 可确定不同焦元个数得到的风速取

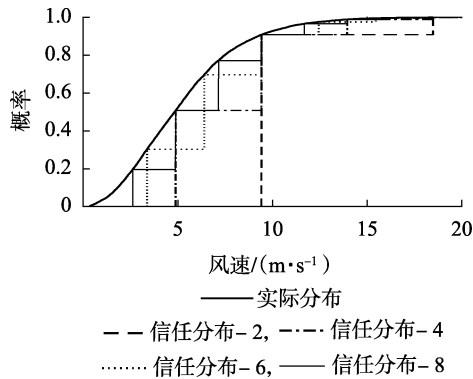


图 7 风速的信任累积概率分布

Fig.7 Cumulative belief probability distribution of wind speed

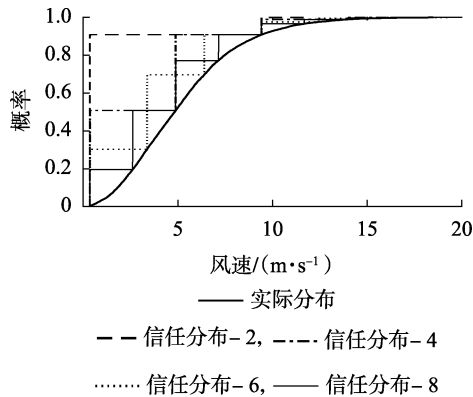


图 8 风速的似然累积概率分布

Fig.8 Cumulative plausibility probability distribution of wind speed

值概率区间,如表 3 所示。以风速小于 4 m/s 为例,当焦元数为 2 时,该事件发生的概率区间为 $[0, 0.908\ 7]$ ;当焦元数为 4 时,概率区间为 $[0, 0.507\ 9]$ ;当焦元数为 6 时,概率区间为 $[0.303\ 1, 0.697\ 3]$ ;当焦元数为 8 时,概率区间为 $[0.196\ 5, 0.507\ 9]$ 。随着焦元数的逐渐增多,风速小于 4 m/s 发生概率的区间半径逐渐减小。当焦元数量趋向于无穷大时,概率区间的半径减小为 0,概率区间退化为实际概率 0.390 2。

表 3 焦元数量对概率区间的影响

Table 3 Effect of focal element number on probability interval

| 焦元<br>数量 | 风速取值的概率区间              |                        |                        |
|----------|------------------------|------------------------|------------------------|
|          | <4 m/s                 | <8 m/s                 | <12 m/s                |
| 2        | $[0, 0.908\ 7]$        | $[0, 0.908\ 7]$        | $[0.908\ 7, 1.000\ 0]$ |
| 4        | $[0, 0.507\ 9]$        | $[0.507\ 9, 0.908\ 7]$ | $[0.908\ 7, 0.989\ 4]$ |
| 6        | $[0.303\ 1, 0.697\ 3]$ | $[0.697\ 3, 0.908\ 7]$ | $[0.908\ 7, 0.976\ 3]$ |
| 8        | $[0.196\ 5, 0.507\ 9]$ | $[0.771\ 4, 0.908\ 7]$ | $[0.966\ 7, 0.989\ 5]$ |
| 实际<br>概率 | 0.390 2                | 0.836 4                | 0.970 7                |

## 5 结论

本文将证据理论应用于风速建模,提出了一种新的风速不确定性模型。通过国内某风电场实测数据的建模结果表明:

a. 所提模型可以得到风速的似然累积概率分布和信任累积概率分布,确定风速任意取值发生概率的区间,即最大概率和最小概率;

b. 与风速概率模型相比,所提模型得到的似然累积概率分布和信任累积概率分布包含了风速实际概率分布的所有可能情况,更合理地描述风速不确定特性。

## 参考文献:

- [1] 龚伟俊,李为相,张广明. 基于威布尔分布的风速概率分布参数估计方法[J]. 可再生能源,2011,29(6):20-23.  
GONG Weijun, LI Weixiang, ZHANG Guangming. The estimation algorithm on the probabilistic distribution parameters of wind speed based on Weibull distribution[J]. Renewable Energy Resources, 2011, 29(6):20-23.
- [2] CARTA J A, RAMIREZ P, VELÁZQUEZ S. A review of wind speed probability distributions used in wind energy analysis; case studies in the Canary Islands[J]. Renewable and Sustainable Energy Reviews, 2009, 13(5):933-955.
- [3] JANGAMSHETTI S H, RAU V G. Siting of wind turbine generators[J]. IEEE Transactions on Energy Conversion, 2001, 16(1):8-13.
- [4] JANGAMSHETTI S H, RAU V G. Site matching of wind turbine generators; a case study[J]. IEEE Transactions on Energy Conversion, 1999, 14(4):1537-1543.
- [5] 凌丹,黄洪钟,张小玲,等. 混合威布尔分布参数估计的 L-M 算法[J]. 电子科技大学学报,2008,37(4):634-636.  
LING Dan, HUANG Hongzhong, ZHANG Xiaoling, et al. Parameter estimation for mixed Weibull distribution using L-M algorithm[J]. Journal of University of Electronic Science and Technology of China, 2008, 37(4):634-636.
- [6] 洪芦诚,石立宝,姚良忠,等. 计及风电场发电功率不确定性的电力系统模糊潮流[J]. 电工技术学报,2010,25(8):116-123.  
HONG Lucheng, SHI Libao, YAO Liangzhong, et al. Fuzzy modelling and solution of load flow incorporating uncertainties of wind farm generation[J]. Transactions of China Electrotechnical Society, 2010, 25(8):116-123.
- [7] VENKATESH B, YU PENG, GOOI H B, et al. Fuzzy MILP unit commitment incorporating wind generators[J]. IEEE Transactions on Power Systems, 2008, 23(4):1738-1746.
- [8] 刘兴杰,谢春雨. 基于贝塔分布的风电功率波动区间估计[J]. 电力自动化设备,2014,34(12):26-30.  
LIU Xingjie, XIE Chunyu. Wind power fluctuation interval estimation based on beta distribution[J]. Electric Power Automation Equipment, 2014, 34(12):26-30.
- [9] PINSON P, KARINIOTAKIS G. Conditional prediction intervals of wind power generation[J]. IEEE Transactions on Power Systems, 2010, 25(4):1845-1856.
- [10] 范磊,卫志农,李慧杰,等. 基于变分模态分解和蝙蝠算法-相关向量机的短期风速区间预测[J]. 电力自动化设备,2017,37(1):93-100.  
FAN Lei, WEI Zhinong, LI Huijie, et al. Short-term wind speed interval prediction based on VMD and BA-RVM algorithm[J]. Electric Power Automation Equipment, 2017, 37(1):93-100.
- [11] 马瑞,张强,吴瑕,等. 日风速随机模糊不确定模型[J]. 中国电机工程学报,2015,35(24):6351-6358.  
MA Rui, ZHANG Qiang, WU Xia, et al. Random fuzzy uncertain model for daily wind speed[J]. Proceedings of the CSEE, 2015, 35(24):6351-6358.

- [12] TIAN Pauchang. Estimation of wind energy potential using different probability density functions [J]. Applied Energy, 2011, 88 (5): 1848-1856.
- [13] 鲍海波, 韦化, 郭小璇, 等. 考虑风电不确定性的概率区间潮流模型与算法[J]. 中国电机工程学报, 2017, 37 (19): 5633-5642. BAO Haibo, WEI Hua, GUO Xiaoxuan, et al. The model and algorithm of probabilistic interval power flow considering wind power uncertainty [J]. Proceedings of the CSEE, 2017, 37 (19): 5633-5642.
- [14] 吴京秋, 孙奇, 杨伟, 等. 基于 D-S 证据理论的短期负荷预测模型融合[J]. 电力自动化设备, 2009, 29 (4): 66-70. WU Jingqiu, SUN Qi, YANG Wei, et al. Short-term load forecast model fusion based on D-S evidence theory [J]. Electric Power Automation Equipment, 2009, 29 (4): 66-70.
- [15] SHAFER G A. Mathematical theory of evidence [M]. Princeton, USA: Princeton University Press, 1976.

- [16] YAGER R R, KACPRZYK J, FEDRIZZI M. Advances in the Dempster-Shafer theory of evidence [M]. New York, USA: John Wiley & Sons, 1994.

#### 作者简介:



郭小璇

郭小璇 (1986—), 女, 辽宁鞍山人, 工程师, 博士, 主要研究方向为电力系统运行与分析 (E-mail: guo\_xiaoxuan@live.cn);

龚仁喜 (1962—), 男, 广西桂林人, 教授, 博士研究生导师, 博士, 通信作者, 主要研究方向为电力电子及新能源发电技术 (E-mail: rxgong@gxu.edu.cn);

鲍海波 (1988—), 男, 安徽合肥人, 高级工程师, 博士, 主要研究方向为电力系统最优运行与规划。

## Modeling of wind speed uncertainty based on evidence theory

GUO Xiaoxuan<sup>1,2</sup>, GONG Renxi<sup>1</sup>, BAO Haibo<sup>3</sup>

(1. School of Electrical Engineering, Guangxi University, Nanning 530004, China;

2. Electric Power Research Institute of Guangxi Power Grid Corporation, Nanning 530023, China;

3. Nanning Power Supply Bureau of Guangxi Power Grid Corporation, Nanning 530031, China)

**Abstract:** Wind speed determines the output power of a wind farm, and its model is an important basis for researching the operation and planning of power system with wind power. A modeling method of wind speed uncertainty based on the evidence theory is proposed. The concept of BPA (Basic Probability Assignment) in the evidence theory is used to describe wind speed. The implementation method of determining the focal points and trust functions of BPA based on the measured historical data is proposed, and two modeling methods of equal probability interval and equal value interval are designed. Taking the measured wind speed of a wind farm as an example, the simulative results of the proposed model are compared with those of the probability distribution and interval distribution model, and the results show that the proposed model can determine the cumulative plausibility probability distribution and the cumulative belief probability distribution of wind speed, and can describe and deal with the uncertainty information of wind speed more effectively.

**Key words:** wind speed; wind power; evidence theory; basic probability assignment; cumulative plausibility probability distribution; cumulative belief probability distribution

(上接第 77 页 continued from page 77)

## Application of XGBoost algorithm in prediction of wind motor main bearing fault

WANG Guilan, ZHAO Hongshan, MI Zengqiang

(School of Electrical & Electronic Engineering, North China Electric Power University, Baoding 071003, China)

**Abstract:** To improve efficiency and accuracy of the processing of scale wind turbine SCADA data, XGBoost algorithm is used to predict the main bearing fault of the wind turbine instead of traditional machine learning algorithm. Firstly, the feature analysis is carried out on the SCADA data of main bearing, the relationship between the feature and fault is excavated and discovered, and the importance of each feature is evaluated. Then the XGBoost algorithm is used to construct the main bearing fault prediction model, and the model is evaluated. Finally, the fault prediction model is trained and tested according to the measured data collected by SCADA system, and the main parameters of the XGBoost model are adjusted to improve the prediction accuracy. Compared with the diagnosis results of classical GBDT (Gradient Boosting Decision Tree) algorithm, the results show that XGBoost is superior to GBDT algorithm in terms of efficiency and accuracy on wind turbine main bearing fault prediction, and it is an effective tool to deal with large scale data sets of SCADA.

**Key words:** XGBoost; SCADA; main bearing; fault prediction; big data