



# 利用贝叶斯线性回归结合混合 Copula 函数分析 风电功率的相关性

苏晨博<sup>1</sup>, 刘崇茹<sup>1</sup>, 徐诗甜<sup>1</sup>, 岳昊<sup>2</sup>

(1. 华北电力大学 电气与电子工程学院, 北京 102206; 2. 国网冀北经济技术研究院, 北京 100038)

**摘要:** 分析风电场之间的出力相关性有助于合理规划风机的功率输送以及调度优化, 从而提高传输线路的利用率。以冀北地区风场为例, 首先分析了该区域风速的分布特性, 然后利用贝叶斯线性回归算法建立混合 Copula 函数模型, 拟合得到 4 个机群风速序列的联合分布, 计算出风电场之间的出力相关性, 并与其他相关性函数建模进行对比研究。研究结果表明, 基于贝叶斯线性回归的混合 Copula 函数模型能够提高参数估计的精确性, 从而使得计算出的相关性更为准确, 并且由其拟合得到的出力概率分布与实际风场出力的概率分布较为一致。

**关键词:** 贝叶斯线性回归; 相关性; 混合 Copula 函数; 风电功率; 出力概率

**DOI:** 10.11930/j.issn.1004-9649.201907121

## 0 引言

风能作为可再生能源的重要组成部分, 装机量日益剧增。风电出力具有间歇性和随机性, 会加剧电网功率波动, 影响电力系统稳定运行。降低弃风率, 更合理的建设和有效利用外送通道, 是当前风力发电发展的重要研究方向, 有助于更加充分合理地利用资源, 节约成本, 提高收益<sup>[1]</sup>。分析风场之间出力相关性<sup>[2-3]</sup>特征有助于实现优化调度和降低弃风率。距离相近、地理条件相近的风场, 风速分布具有较为明显的相关特性<sup>[4-5]</sup>; 距离较远、地理条件差别较大的风场, 风速分布相关系数较低。风速的分布特性与威布尔分布相似, 区域间风速相关性呈非线性, 以线性相关系数为基础的方法很难准确捕捉这种非线性相关性, 因此需要寻求一种更为合适的数学模型。

Copula 函数可以通过变量的边缘分布函数拟合变量的联合分布函数, 因此可以用来计算风电功率相关性, 有学者在这方面已经开展了一些研究。文献 [6-7] 将 Copula 函数运用于风电出力相关性研究。文献 [8] 验证了 Copula 函数产生的相关性概率序列能够有效应用于风电场建模。

文献 [9] 利用正态 Copula 函数对风电功率之间相关性进行分析, 并利用沿海地区的 4 个的风场进行验证, 正态 Copula 函数适用于对称性的数据结构, 对于非对称性的数据不能较好的表征。文献 [10] 利用 T-Copula 函数刻画了风电场之间的出力相关性, 通过潮流计算验证考虑相关性的建模对制定调度计划是必要的。单一的 T-Copula 函数不能呈现数据的尾部特性, 计算出的相关性存在一定偏差。文献 [11] 进一步考虑风速的非对称性分布, 利用 Gumbel-Copula 函数对风电相关性进行建模。但是 Gumbel-Copula 函数只能描述上尾相关性而无法准确刻画下尾相关性。文献 [12] 提出采用 T-Copula 和 Gumbel-Copula 函数构造混合 Copula 函数的概念, 但对模型的构造及验证未进行详细的描述。文献 [13] 利用最小二乘法构建混合 Copula 函数模型计算风电场之间的出力相关性, 最小二乘法选择的混合 Copula 参数可能存在过拟合问题<sup>[14]</sup>。

本文在文献 [13-15] 的基础上, 结合贝叶斯线性回归算法建立混合 Copula 函数模型对多个不同特征风场进行相关性计算; 分析相关性对风电场联合出力的影响。该方法的优点是: 能够将数据的先验信息与迭代相结合, 来计算后验分布, 从而在风场维度增加的情况下依然能够准确计算不同风场的联合概率密度函数。最后, 采用冀北地

收稿日期: 2019-07-15; 修回日期: 2020-03-23。



区风场风速的实际历史数据对模型效果进行了验证。

## 1 Copula 函数基本理论及相关性分

### 1.1 Copula 函数的基本概念

概率统计中，求解多个变量的相关性问一般先要求出变量间的联合分布，但在实际应用中很难得到变量之间的联合分布函数，并且求解联合分布函数时，要求各边缘分布函数与联合分布函数的分布类型一致，这个条件加大了求解难度。

文献 [16] 提出可以将一个  $N$  维联合分布函数分解为  $N$  个边缘分布函数和一个 Copula 函数，这个 Copula 函数描述了变量之间的相关性。1999 年 Nelsen 给出了 Copula 函数的严格定义，Copula 函数是把随机向量  $X_1, X_2, \dots, X_N$  的联合分布函数  $F(x_1, x_2, \dots, x_N)$  与各自的边缘分布函数  $F(x_1), F(x_2), \dots, F(x_N)$  相连接的连接函数，如式 (1) 所示，且当  $F_1(x), F_2(x), \dots, F_N(x)$  连续时，Copula 函数唯一确定。

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_N(x_N)) \quad (1)$$

### 1.2 Copula 函数相关性分析

Copula 函数用来度量连续随机变量之间的相关性，设有连续采样的二维随机变量  $(X, Y)$ ，它们对应的边缘分布函数为  $F(x)$  和  $G(y)$ ，与其对应的联合函数为  $C(u, v)$ 。 $\tau$ 、 $\rho$ 、 $\lambda$  分别为随机变量的秩相关系数、Spearman 相关系数以及尾部系数，它们之间的关系为

$$\tau = \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (2)$$

$$\rho = 12 \int_0^1 \int_0^1 uv dC(u, v) - 3 \quad (3)$$

$$\begin{aligned} \lambda^{up} &= \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, v)}{1 - u} \\ \lambda^{lo} &= \lim_{u \rightarrow 0^+} \frac{C(u, v)}{u} \end{aligned} \quad (4)$$

式中： $U = F(x) \sim U(0, 1)$ ； $V = G(y) \sim U(0, 1)$ ；

$$\hat{C}(1 - u, 1 - v) = P(U > u, V > v) = 1 - u - v + C(u, v)$$

### 1.3 混合 Copula 函数

Copula 函数包括正态 Copula 函数，T-copula 函数以及阿基米德 Copula 函数。其中阿基米德 Copula 函数主要包含 Gumbel-Copula，Clayton-Copula，Frank-Copula 3 种函数。文献 [16] 研究表明混合 Copula 函数比某个特定的 Copula 函数更能

准确地反映非线性变量之间的相关性与非对称结构。通过构建不同的 Copula 函数的概率密度可知，T-Copula 函数只能反映变量之间的对称相关性。Clayton-Copula 函数具有高度敏感的下尾特性。Gumbel-Copula 函数与 Clayton-Copula 函数正好相反，能够较好地跟踪上尾特性。由此得出，单一的 Copula 函数只能针对随机变量的某种特性，不能完全表征复杂随机变量之间的相关性。

混合 Copula 函数是利用多种 Copula 函数乘以权重系数进行叠加，因此能够更加全面的表征变量的分布特性。其数学表达式为

$$C(u_1, u_2, \dots, u_n) = \sum_{i=1}^n \omega_i C_i(u_1, u_2, \dots, u_n; \theta_i) \quad (5)$$

式中： $C(u_1, u_2, \dots, u_n)$  为已知的 Copula 函数； $\theta_i$  为 Copula 函数的相关参数； $\omega_i$  为权重系数，且  $\omega_1 + \omega_2 + \dots + \omega_n = 1$ ， $0 \leq \omega_i \leq 1$ 。可以通过改变权重系数  $\omega_i$  使得混合 Copula 函数能够更好地表征变量之间的对称性与非对称性结构

## 2 风电功率相关性建模

### 2.1 风速的分布特性

风电场出力跟风速的三次方成正比，一般情况下，风速服从双参数威布尔分布，如图 1 所示。其概率密度表达式为

$$f(v_w, c_w, k_w) = \frac{k_w}{c_w} v_w^{k_w-1} e^{-\left(\frac{v_w}{c_w}\right)^{k_w}} \quad (6)$$

式中： $v_w$  表示风速序列； $k_w$  为韦伯分布的形状参数； $c_w$  为尺度参数。

对式 (6) 进行积分可以得到风速的累计分布函数为

$$W(v_w) = 1 - e^{-\left(\frac{v_w}{c_w}\right)^{k_w}} \quad (7)$$

风电出力相关性受到地貌、气象、海拔高度等因素的影响并且具有一定的气候惯性。一般情况下，若地形相仿时，风速很大概率具有高相关性；若地形差异较大，则该片区内风速相关性可能较小。

线性相关系数并不能很好地表示风电场之间的出力相关性。因此，对于分析这类呈非线性分布的相关性问题，有必要构建混合 Copula 函数模型来表征不同条件下风电场间出力的相关性。

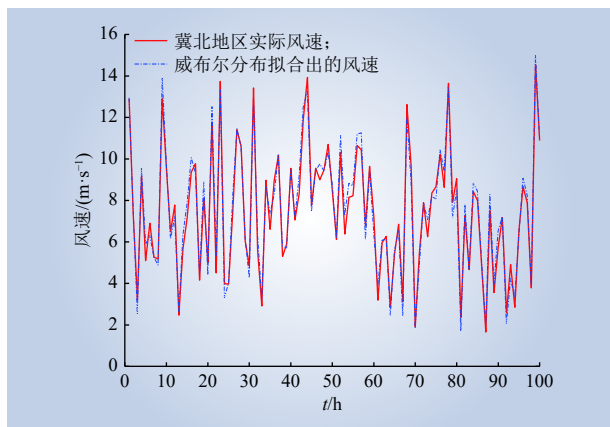


图 1 威布尔分布与实际风速分布对比  
Fig. 1 Comparison between weibull distribution and actual wind speed

### 2.2 混合 Copula 函数参数估计

本文根据风电场的多类型特征，采用 T-Copula 函数、Clayton-Copula 函数和 Gumbel-Copula 函数建立混合 Copula 函数模型为

$$C_M(v_1, v_2) = \omega_1 C_T(v_1, v_2; \theta_1) + \omega_2 C_G(v_1, v_2; \theta_2) + \omega_3 C_c(v_1, v_2; \theta_3) \quad (8)$$

式中： $v_1$ 和 $v_2$ 分别代表 2 个风场的风速序列； $\theta_1$ 为 T-Copula 函数表达式中的相关参数； $\theta_2$ 为 Gumbel-Copula 函数中的相关参数； $\theta_3$ 为 Clayton-Copula 函数中的相关参数； $\omega$ 为权重系数。

由于所有种类的 Copula 函数拟合得到的联合分布都是属于 0-1 均匀分布，故得到的混合 Copula 函数也属于 0-1 均匀分布，并且有 $\omega_1 + \omega_2 + \omega_3 = 1$ ，权重系数 $\omega$ 的大小反映了风速联合分布特性。如 $\omega_1$ 反映了风速联合分布的对称相关性，而 $\omega_2$ 和 $\omega_3$ 分别反映了风速联合分布函数的上尾特性和下尾特性。

参数估计的方法主要有极大似然估计，最大期望值 (EM) [17-19] 及贝叶斯线性回归。考虑数据处理的精确性并防止过拟合现象，本文采用贝叶斯线性回归对 $\omega$ 的值进行估算。贝叶斯线性回归中心思想是利用样本进行最大后验估计，最大后验估计与最大似然估计的显著区别就是附加了一项由先验分布决定的惩罚项，先验的好处主要在于代数上的方便性，可以直接给出后验分布的封闭形式，并且可以在一定程度上防止过拟合，充分合理的利用样本，其数学模型为

$$L(\omega|T) = \prod_{i=1}^n p(t_i|x_i, \omega, \beta^{-1}I) p(\omega|\alpha) \quad (9)$$

式中： $L(\omega|T)$ 为在样本 $T(t_1, t_2, \dots, t_n)$ 下参数 $\omega$ 的最大置信区间； $p(t_i|x_i, \omega, \beta^{-1}I)$ 为样本 $(x, t)$ 的输出 $t$ 关于输入 $x$ 、权重参数 $\omega$ 以及样本分布方差 $\beta$ 的概率密度函数； $p(\omega|\alpha)$ 为参数 $\omega$ 关于噪声 $\alpha$ 的先验分布。

贝叶斯线性回归函数为

$$\ln(L(\omega|T)) = -\frac{\beta}{2} \sum_{i=1}^n [t_i - y(x_i, \omega)]^2 - \frac{\alpha}{2} \xi + c \quad (10)$$

式中： $y(x_i, \omega)$ 为关于输入 $x$ 的函数，在本文特指式 (8) 中的 $C_M(v_1, v_2)$ ； $c$ 为常量，与其他参数无关；且有 $\xi = \omega_1^2 + \omega_2^2 + \dots + \omega_n^2$ 。

可以看出，贝叶斯线性回归在一般的误差函数的基础上添加了由样本数据选择的正则化项，通过调整 $\alpha$ 值，使得模型能够在有限大小的数据集上进行训练，这样就能够在一定程度上限制模型的复杂程度，从而能够抑制过拟合的现象。噪声 $\alpha$ 采用迭代的方法求解， $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$ 的误差函数为

$$E(\omega) = E(m_N) + \frac{1}{2}(\omega - m_N)^T A (\omega - m_N) \quad (11)$$

其中

$$A = \alpha I + \beta \phi^T(x_i) \phi(x_i) \quad (12)$$

$$m_N = \beta A^{-1} \phi^T(x_i) t$$

式中： $\phi(x_i)$ 为基函数； $m_N$ 为众数；矩阵 $A$ 的特征值为 $\alpha + \lambda_i$ ； $\ln|A|$ 关于 $\alpha$ 的导数为

$$\frac{d}{d\alpha} \ln|A| = \frac{d}{d\alpha} \ln \prod_i (\lambda_i + \alpha) = \sum_i \frac{1}{\lambda_i + \alpha} \quad (13)$$

关于 $\alpha$ 的驻点满足

$$0 = \frac{M}{2\alpha} - \frac{1}{2} m_N^T m_N - \frac{1}{2} \sum_{i=1}^M \frac{1}{\lambda_i + \alpha} \quad (14)$$

两侧同时乘以 $2\alpha$ ，整理可得

$$\alpha m_N^T m_N = M - \alpha \sum_{i=1}^M \frac{1}{\lambda_i + \alpha} = \gamma \quad (15)$$

可以得到 $\alpha$ 满足

$$\alpha = \frac{\gamma}{m_N^T m_N} \quad (16)$$

这里 $\alpha$ 是一个隐式解，不仅因为 $\gamma$ 与相关，还因为众数 $m_N$ 也与 $\alpha$ 的选择有关，所以用迭代算法求解，首先选一个 $\alpha$ 的初始值，使用这个初



始值找到  $m_N^{(1)}$ ，利用式 (15) 计算得到  $\gamma^{(1)}$ ，然后代入式 (16) 中重新计算  $\alpha$ ，直到收敛。

为了进一步计算混合 Copula 函数中的未知参数，引入经验 Copula 函数用以拟合实际风速的联合分布。假设两风场风速序列为  $(v_1, v_2)$ ，设变量  $(x_i, y_i)$  取自样本  $(v_1, v_2)$ ，它们的边缘分布函数分别为  $F_n(x_i)$  和  $G_n(y_i)$ ，定义样本经验 Copula 为

$$\hat{C}_n(v_1, v_2) = \frac{1}{n} \sum_{i=1}^n I_{[F_n(x_i) \leq v_1]} I_{[G_n(y_i) \leq v_2]} \quad (17)$$

式中： $I_{[\cdot]}$  为示性函数；当  $F_n(x_i) \leq v_1$  时  $I=1$ ；当  $F_n(x_i) > v_1$  时  $I=0$ 。将式 (8)、(11) 代入式 (10) 得到

$$L(\omega; \theta) = -\frac{\beta}{2} \sum_{i=1}^n \left\{ \hat{C}_n(v_1, v_2) - [\omega_1 C_T(v_1, v_2; \theta_1) + \omega_2 C_G(v_1, v_2; \theta_2) + \omega_3 C_C(v_1, v_2; \theta_3)] \right\}^2 - \frac{\alpha}{2} \tau \quad (18)$$

式中： $\tau = \omega_1^2 + \omega_2^2 + \omega_3^2$ ； $\theta$  是 Copula 函数的参数，与模型参数  $\omega$  的分布相互独立，故对式 (18) 左右两边求关于参数  $\theta$ 、 $\omega$  的偏导，即

$$\begin{cases} \frac{\partial L(\omega; \theta)}{\partial \theta_1} = 0 & \frac{\partial L(\omega; \theta)}{\partial \theta_2} = 0 & \frac{\partial L(\omega; \theta)}{\partial \theta_3} = 0 \\ \frac{\partial L(\omega; \theta)}{\partial \omega_1} = 0 & \frac{\partial L(\omega; \theta)}{\partial \omega_2} = 0 & \end{cases} \quad (19)$$

### 2.3 风电出力模型

若风速已知，则风电场出力分段函数为

$$P_w = \begin{cases} 0, & v < v_{wi}, v > v_{wo} \\ \frac{v^3 - v_{wi}^3}{v_r^3 - v_{wi}^3} P_r, & v_{wi} < v < v_r \\ P_r, & v_r \leq v < v_{wo} \end{cases} \quad (20)$$

式中： $v_{wi}$  表示切入风速； $v_{wo}$  表示切出风速； $v_r$  代表额定风速； $P_r$  为风电机组的额定功率。本文计算时参考实际风电场，设切入风速为 3 m/s，切出风速为 24 m/s，额定风速为 13 m/s。

### 2.4 不同风场相关性计算

结合 2.2 混合 Copula 函数参数估计，得出风电出力相关性计算过程。

(1) 采用非参数核密度估计方法计算风速序列，假设  $X_1, X_2, X_3, \dots, X_N$  是取自风速序列  $\{v_n\}$  的样本数据，任意风速  $v$  的核密度估计定义表达式为

$$\hat{f}_h(v) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{v - X_i}{h}\right) \quad (21)$$

式中： $K(\cdot)$  为核函数； $h$  为窗宽。为了保证  $\hat{f}_h(v)$  作为核密度的合理性，要求满足

$$K(x) \geq 0, \int_{-\infty}^{+\infty} k(x) dx = 1 \quad (22)$$

(2) 得到风速序列的核密度估计后，建立混合 Copula 函数表达式并利用贝叶斯线性回归对其中的参数进行估计，计算出风速联合分布的概率密度，进一步求得满足联合分布特性的风速序列。

(3) 将风速序列代入式 (14) 求出风电出力序列，得到不同风场之间的出力 Spearman 相关系数，并进一步求得各个风电场的出力概率分布状态，针对计算结果进行对比分析。

## 3 张北地区风电机群相关性分析

选取张北地区 4 个风电机群并将它们划分为 2 组进行相关性研究，具体流程如下。

(1) 读取近 2 年来冀北地区的历史风速序列及出力数据。

(2) 利用核密度估计各风场的风速边缘分布，将其与实际历史风速拟合得到的经验分布进行对比，如图 2 所示。

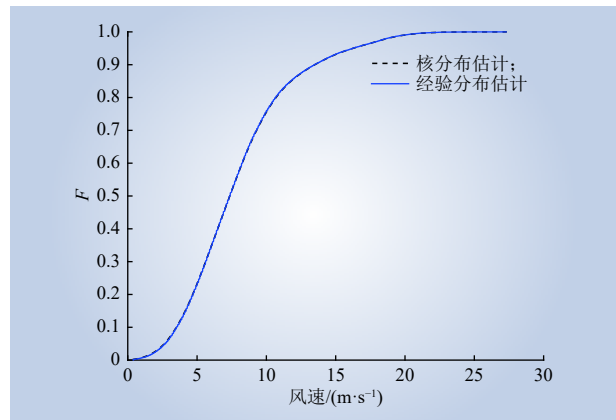


图 2 风场 1 风速核分布估计与经验分布对比  
Fig. 2 Comparison between wind speed kernel distribution and empirical distribution in 1st wind field.

从图 2 中可以看出，核分布估计出的风速边缘分布与经验分布基本一致，说明核分布估计能够有效表示风场风速的边缘分布。

(3) 将步骤 2 中得到的边缘分布代入式 (11)



得到 2 组经验联合分布 $(\eta_{12}, \eta_{34})$ ，再由式 (12) ~ (13) 计算得到混合 Copula 函数 $C_M(v_1, v_2)$ 、 $C_M(v_3, v_4)$ ，然后作出联合概率密度图与实际风速的联合概率密度直方图做对比 (如图 3~4 所示)。

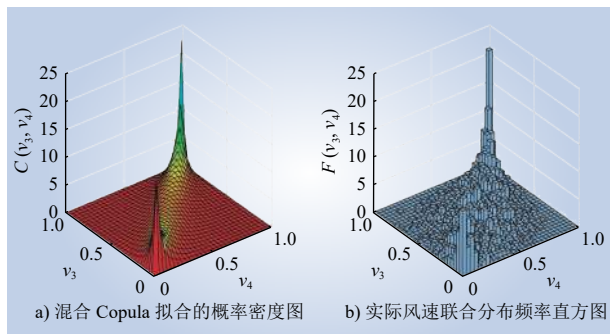


图 3 风电场 1、2 的联合分布特性  
Fig. 3 Joint distribution characteristics of wind farm 1 and 2

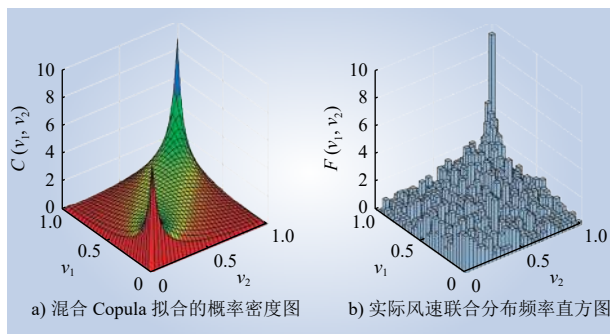


图 4 风电场 3、4 的联合分布特性  
Fig. 4 Joint distribution characteristics of wind farm 3 and 4

从图 3~4 中可以看出，针对不同地区的风场，所提算法估计得到的联合概率密度函数与历史风速的联合概率密度直方图趋势基本一致。因此可以说明所提算法的有效性。

为了更加直观的表现算法的优越性，利用单一 T-Copula 函数，由最小二乘法产生的混合 Copula 函数和贝叶斯线性回归得到的混合 Copula 函数分别进行相关性计算，结果如表 1 和表 2 所示。从表中可以看出，针对不同地区的风速序列，所提算法估计得到的联合分布函数与经验分布的欧式距离均为最小，在一定程度上可以说明算法的优越性。表中  $d$  表示计算得到的 Copula 函数与经验 Copula 函数的欧式距离，表达式为

$$d = \sum_{i=1}^n |C_M^{(i)}(v_1, v_2) - \hat{C}_n^{(i)}(v_1, v_2)| \quad (23)$$

表 1 不同算法拟合风电场 1、2 风速联合分布特性  
Table 1 Different algorithms fit wind speeds' joint distribution between wind fields 1 and 2.

计算方法	Copula函数	$\omega$	$\theta$	$d$
$C_B(v_1, v_2)$	T-Copula	0.63	0.876	0.6891
	G-Copula	0.25	2.845	
	C-Copula	0.12	1.592	
$C_{WLS}(v_1, v_2)$	T-Copula	0.68	0.864	0.9057
	G-Copula	0.23	2.652	
	C-Copula	0.09	1.672	
$C_T(v_1, v_2)$	—	—	0.871	2.6937

表 2 不同算法拟合风电场 3、4 风速联合分布特性  
Table 2 Different algorithms fit wind speeds' joint distribution between wind fields 3 and 4.

计算方法	Copula函数	$\omega$	$\theta$	$d$
$C_B(v_1, v_2)$	T-Copula	0.53	0.472	1.2756
	G-Copula	0.29	1.773	
	C-Copula	0.18	1.031	
$C_{WLS}(v_1, v_2)$	T-Copula	0.55	0.475	1.4459
	G-Copula	0.30	1.814	
	C-Copula	0.15	1.098	
$C_T(v_1, v_2)$	—	—	0.469	3.7762

注：表中 $C_B(v_1, v_2)$ 表示用贝叶斯线性回归拟合出的混合 Copula 函数； $C_{WLS}(v_1, v_2)$ 表示用最小二乘法拟合出的混合 Copula 函数； $C_T(v_1, v_2)$ 表示 T-Copula 函数； $\omega$  表示权重系数； $\theta$  表示 Copula 函数中的相关参数。

仅仅通过欧式距离可能无法全面表现该算法的优越性，所以利用 K-S 检验研究不同计算方法得到的混合 Copula 函数与经验 Copula 函数的标准差与分布差异情况。K-S 检测是通过求布朗运动上确界得到的随机变量的分布。它的累积分布函数可以写为

$$\Pr(K \leq x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 x^2} \quad (24)$$

K-S 检测结果如表 3 所示，选取的显著水平为 5%，其中  $p$  表示渐进程度，样本尺寸越大  $p$  值越精确。 $k$  表示检测指标， $k$  值越小，说明进行对比的累积分布函数之间的差异越小。可以看到，利用贝叶斯线性回归选择的混合 Copula 函数对不同风场之间出力的联合分布函数进行拟合时，K-S 检测值最小，说明该方法计算出的联合分布函数更贴合实际数据。



表 3 不同算法的K-S检测结果  
Table 3 K-S Detection Results of Different Algorithms

计算方法	$p$	$k$
$C_B(v_1, v_2)$	0.0112	0.156
$C_{WLS}(v_1, v_2)$	0.0156	0.187
$C_T(v_1, v_2)$	0.0654	0.372

通过欧式距离以及 K-S 检测的结果对比发现，不同地理位置的风电机群，风速联合分布特性差异明显，针对这种情况利用贝叶斯线性回归方法选择出的混合 Copula 函数拟合出的联合分布与实际风速联合分布的欧式距离最小，说明该方法比单一 Copula 函数及最小二乘法选择出的混合 Copula 函数更能反映实际风速的联合分布特性，所以用贝叶斯线性回归缔造出的混合 Copula 函数计算得到的风电功率的相关性更加准确。

(4) 由上一步得到 2 组风电机群风速的联合分布函数  $C_M(v_1, v_2)$ ,  $C_M(v_3, v_4)$  可以生成 4 组风速分布序列  $\eta(\eta_{v1}, \eta_{v2}, \eta_{v3}, \eta_{v4})$ ，利用威布尔逆分布式 (18) 可以得到 4 组风速序列  $V(V_1, V_2, V_3, V_4)$ ，序列包含全年 (8760 小时) 的风速。

$$v_w = W^{-1}(u) = c_w^{k_w} \sqrt{\ln \frac{1}{1-u}} \quad (25)$$

每组风电机群的形状参数  $k_w$  和尺度参数  $c_w$  如表 4 所示。将获得的风速序列带入式 (14) 可以求得全年功率分布序列为  $P(P_1, P_2, P_3, P_4)$ ，并计算出它们的出力 Spearman 相关系数  $\rho_{12} = 0.873$ 、 $\rho_{34} = 0.461$ 。假设每个风电机群的最大出力为额定值 1 (标么值)，将  $P_1, P_2$  叠加得到风电机群 1 和 2 的联合出力分布，然后计算其功率概率分布如图 5 中虚线所示，同理利用  $P_3, P_4$  叠加可以得到风电机群 3 和 4 的功率概率分布如图 6 中虚线所示。

将拟合的功率概率分布与风场实际出力的概率分布曲线 (图 5、图 6 中的实线) 作对比发现，

表 4 Weibull 分布参数估计值  
Table 4 Estimated value of Weibull distribution parameters

风电机群编号	尺度参数 $c_w$	形状参数 $k_w$
1	8.9828	2.1142
2	8.8971	2.1092
3	7.4522	1.2036
4	7.1601	1.7866

2 种曲线的贴合度很高，形状较为一致，说明利用贝叶斯线性回归方法得到的混合 Copula 函数不仅能够很好地表征风速之间的相关性，同时也可以较大程度地还原实际风场之间的出力相关性和他们的联合出力的概率分布。

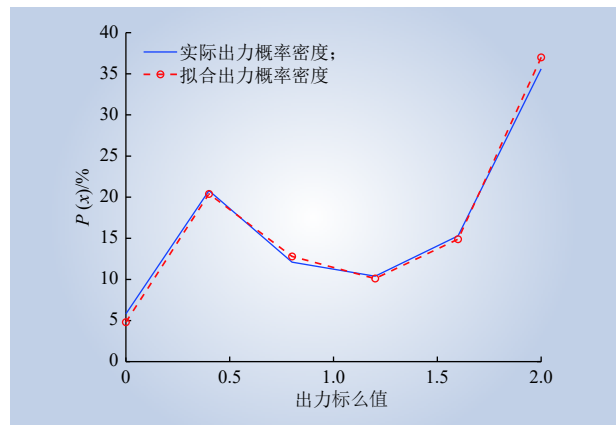


图 5 风场 1, 2 联合出力概率分布状态  
Fig. 5 Probability distribution of combined output between two wind fields 1 and 2.

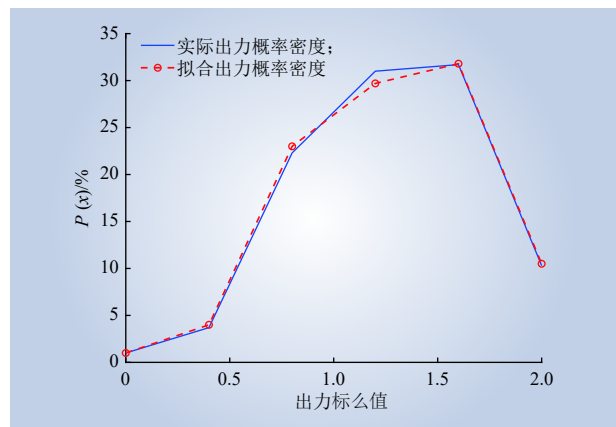


图 6 风场 3, 4 联合出力概率分布状态  
Fig. 6 Probability distribution of combined output between two wind fields 3 and 4.

除此之外，研究结果阐释了相关性对风电场联合出力的影响。风电机群 1 和 2 之间的相关系数为  $\rho_{12} = 0.873$ ，风速相关性较强，说明风速在下一时刻的增减情况大概率相似，并且由图 5 可以看出风场 1 和 2 的联合出力概率分布集中在 0.5 p.u. 以下和 1.6 p.u. 以上，两部分的和为 62.2%，说明风场 1 和 2 同时满发或者不发电的概率较高。进一步研究风电机群 3 和 4，它们的相关系数为  $\rho_{34} = 0.461$ ，相关性较弱，说明下一时刻风速变化的相依性较



低,变化不同的概率较大。从图 6 可以看出联合出力概率分布集中在 0.8 p.u.和 1.6 p.u.之间,占比 76.4%,理论较为相符。因此考虑风场之间的出力相关性是很有必要的,它能在一定程度上反映风场之间联合出力的概率分布情况。

## 4 结论

(1) 基于贝叶斯线性回归选择的混合 Copula 函数能够充分的利用样本数据,并且高度还原实际风场之间的风速联合分布,与最小二乘法 and 单一 T-Copula 函数相比,由该方法得到的相关系数更贴合实际。

(2) 用该方法得到的联合分布函数生成的风电场出力概率分布曲线与实际风电场出力概率分布进行对比,结果表明,基于贝叶斯线性回归的混合 Copula 函数拟合出的出力概率分布模型与实际结果比较一致,能够较为准确地反映实际风场间的联合出力状态。

(3) 风电场之间相关系数对它们的联合出力概率密度有较大影响,在一定距离内,若风场之间的相关性较强,风速在下一时刻的增减情况大概率相似,表征为风电场之间同时满发或者停发的概率比较大;当相关性较小时,风电场之间出力联合出力特性比较理想。

## 参考文献:

- [1] XIE Z Q, JI T Y, LI M S, *et al.* Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function[J]. *IEEE Transactions on Power Systems*, 2018, 33(2): 2239–2247.
- [2] PAPAETHYMIU G, SCHAVEMAKER P H, VAN DER SLUIS L, *et al.* Integration of stochastic generation in power systems[J]. *International Journal of Electrical Power & Energy Systems*, 2006, 28(9): 655–667.
- [3] 吕颖,鲁广明,谢昶,等.考虑大规模集中接入风电功率波动相关性在线概率安全评估[J]. *电网技术*, 2018, 42(4): 1140–1148.  
LÜ Ying, LU Guangming, XIE Chang, *et al.* Online probabilistic security assessment considering the power fluctuation correlation of large-scale centralized access wind power[J]. *Power System Technology*, 2018, 42(4): 1140–1148.
- [4] 杨洪明,王爽,易德鑫,等.考虑多风电场出力相关性的电力系统随机优化调度[J]. *电力自动化设备*, 2013, 33(1): 114–120.  
YANG Hongming, WANG Shuang, YI Dexin, *et al.* Stochastic optimal dispatch of power system considering multi-wind power correlation[J]. *Electric Power Automation Equipment*, 2013, 33(1): 114–120.
- [5] 杨柳青,林舜江,刘明波,等.考虑风电接入的大型电力系统多目标动态优化调度[J]. *电工技术学报*, 2014, 29(10): 286–295.  
YANG Liuqing, LIN Shunjiang, LIU Mingbo, *et al.* Multi-objective dynamic optimal dispatch for large-scale power systems considering wind power penetration[J]. *Transactions of China Electrotechnical Society*, 2014, 29(10): 286–295.
- [6] PAPAETHYMIU G, KUROWICKA D. Using copulas for modeling stochastic dependence in power system uncertainty analysis[J]. *IEEE Transactions on Power Systems*, 2009, 24(1): 40–49.
- [7] WANG Y, GAN D H, SUN M Y, *et al.* Probabilistic individual load forecasting using pinball loss guided LSTM[J]. *Applied Energy*, 2019, 235: 10–20.
- [8] STEPHEN B, GALLOWAY S J, MCMILLAN D, *et al.* A copula model of wind turbine performance[J]. *IEEE Transactions on Power Systems*, 2011, 26(2): 965–966.
- [9] 张宁,康重庆.风电出力分析中的相依概率性序列运算[J]. *清华大学学报(自然科学版)*, 2012, 52(5): 704–709.  
ZHANG Ning, KANG Chongqing. Dependent probabilistic sequence operations for wind power output analyses[J]. *Journal of Tsinghua University (Science and Technology)*, 2012, 52(5): 704–709.
- [10] 蔡菲,严正,赵静波,等.基于 Copula 理论的风电场间风速及输出功率相依结构建模[J]. *电力系统自动化*, 2013, 37(17): 9–16.  
CAI Fei, YAN Zheng, ZHAO Jingbo, *et al.* Dependence structure models for wind speed and wind power among different wind farms based on copula theory[J]. *Automation of Electric Power Systems*, 2013, 37(17): 9–16.
- [11] 谢敏,熊靖,刘明波,等.基于 Copula 的多风电场出力相关性建模及其在电网经济调度中的应用[J]. *电网技术*, 2016, 40(4): 1100–1106.  
XIE Min, XIONG Jing, LIU Mingbo, *et al.* Modeling of multi wind farm output correlation based on copula and its application in power system economic dispatch[J]. *Power System Technology*, 2016, 40(4): 1100–1106.
- [12] VALIZADEH HAGHI H, TAVAKOLI BINA M, GOLKAR M A, *et al.* Using Copulas for analysis of large datasets in renewable



- distributed generation: PV and wind power integration in Iran[J]. *Renewable Energy*, 2010, 35(9): 1991–2000.
- [13] 潘雄, 王莉莉, 徐玉琴, 等. 基于混合 Copula 函数的风电场出力建模方法 [J]. *电力系统自动化*, 2014, 38(14): 17–22.  
PAN Xiong, WANG Lili, XU Yuqin, *et al.* A wind farm power modeling method based on mixed copula[J]. *Automation of Electric Power Systems*, 2014, 38(14): 17–22.
- [14] 季峰, 蔡兴国, 王俊. 基于混合 Copula 函数的风电功率相关性分析 [J]. *电力系统自动化*, 2014, 38(2): 1–5, 32.  
JI Feng, CAI Xingguo, WANG Jun. Wind power correlation analysis based on hybrid copula[J]. *Automation of Electric Power Systems*, 2014, 38(2): 1–5, 32.
- [15] TUTZ G, BINDER H. Boosting ridge regression[J]. *Computational Statistics & Data Analysis*, 2007, 51(12): 6044–6059.
- [16] SKLAR A. Fonctions de repartition à n dimensions et leurs marges[J]. *Publication de l'Institut de Statistique l'Universite Paris*, 1959, 8: 229–231.
- [17] 徐玉琴, 陈坤, 李俊卿, 等. Copula 函数与核估计理论相结合分析风电场出力相关性的一种新方法 [J]. *电工技术学报*, 2016, 31(13): 92–100.  
XU Yuqin, CHEN Kun, LI Junqing, *et al.* A new method analyzing output correlation of multi-wind farms based on combination of Copula function and kernel estimation theory[J]. *Transactions of China Electrotechnical Society*, 2016, 31(13): 92–100.
- [18] 王玲玲, 王昕, 郑益慧, 等. 计及多个风电机组出力相关性的配电网无功优化 [J]. *电网技术*, 2017, 41(11): 3463–3469.  
WANG Lingling, WANG Xin, ZHENG Yihui, *et al.* Reactive power optimization of distribution network considering output correlation of multiple wind turbines[J]. *Power System Technology*, 2017, 41(11): 3463–3469.
- [19] GARDES L, GIRARD S. Nonparametric estimation of the conditional tail Copula[J]. *Journal of Multivariate Analysis*, 2015, 137: 1–16.

#### 作者简介:

苏晨博 (1991—), 男, 博士, 从事风电场规划, 风电机组接入系统的稳定性分析研究, E-mail: 373607885@qq.com;

刘崇茹 (1977—), 女, 通信作者, 教授, 从事交直流混合系统分析与仿真、运行与控制研究, E-mail: chongru.liu@ncepu.edu.cn.

(责任编辑 吴恒天)

## Mix Copula Function Based Wind Power Correlation Analysis: A Bayesian Linear Regression Approach

SU Chenbo<sup>1</sup>, LIU Chongru<sup>1</sup>, XU Shitian<sup>1</sup>, YUE Hao<sup>2</sup>

(1. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China; 2. State Grid Jibei Electric Economic Research Institute, Beijing 100038, China)

**Abstract:** Analyzing the output correlation among wind farms is beneficial to reasonably plan the power transmission and scheduling optimization, so as to improve the utilization rate of transmission lines. Taking wind field in northern Hebei region as an example, this paper analyzes the characteristics of wind power first, and then improves a mix Copula method to model the relationship among the wind power correlation structure. What's more, we use the Bayesian linear regression method to establish a mixed Copula function model to calculate the correlation of wind speed sequences from different wind farm groups. In this way, we can fit the joint distribution function between them and analyze the impact of correlation on the joint output of wind farms. In addition, this method has been verified to be effective and accurate, it was also be compared with other correlation function modeling. The results show that the mixed Copula function model based on Bayesian linear regression can well calculate the correlation of wind power output, from which the output probability distribution obtained can get more accurate fitting results.

**Keywords:** Bayesian linear regression; correlation; mixed Copula function; wind power; output probability