Advances in Applied Mathematics 应用数学进展, 2023, 12(5), 2410-2419
Published Online May 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.125243

Hans汉斯

基于D-Vine Copula构建内蒙古四邻近地区风速相依模型

刘文博，彭秀云*，郑　洋

内蒙古工业大学理学院，内蒙古　呼和浩特

收稿日期：2023年4月28日；录用日期：2023年5月21日；发布日期：2023年5月30日

摘　要

采用D-Vine Copula方法测度内蒙古四近邻地区最大风速的关联性，该方法将多元联合分布通过Pair-Copula分解成边缘密度和二元Copula函数的乘积形式。根据Kendall秩相关系数选择最优的D-Vine Copula结构，使用两阶段法求解模型参数。首先构建边缘密度，然后使用逐树估计和联合估计方法估计二元Copula参数并选择最优分布。通过比较赤池信息量(AIC)发现，相比于逐树估计方法，联合估计参数的结果拟合效果更优。模拟发现四近邻地区间风速间存在不同形式的关联性，D-Vine Copula方法能够灵活的测度这种高维随机变量的关联性差异。

关键词

D-Vine Copula，Pair-Copula结构，风速，相关性

Modeling Dependency Structure of Wind Speed by D-Vine Copula in Four Neighboring Areas of Inner Mongolia

Wenbo Liu, Xiuyun Peng*, Yang Zheng

School of Science, Inner Mongolia University of Technology, Hohhot Inner Mongolia

Received: Apr. 28th, 2023; accepted: May 21st, 2023; published: May 30th, 2023

Abstract

The D-Vine Copula method is used to measure the correlation of maximum wind speed in four

*通讯作者。

文章引用: 刘文博, 彭秀云, 郑洋. 基于 D-Vine Copula 构建内蒙古四邻近地区风速相依模型[J]. 应用数学进展, 2023, 12(5): 2410-2419. DOI: 10.12677/aam.2023.125243

neighboring areas of Inner Mongolia. Based on this method, the multivariate joint distribution is decomposed into a product of the marginal densities and the bivariate Copula functions in terms of the Pair-Copula technique. The optimal D-Vine Copula structure is selected according to Kendall rank correlation coefficient, and two-stage strategy is used to solve the model parameters. The marginal density is fitted first, and then the Pair-Copula function is simulated by the tree by tree estimation and global joint density estimation methods. By Akaike Information Criterion (AIC), it can be seen that the fitting effect of the global joint density estimation is better than that of tree by tree. It is found that there are different forms of correlations between wind speeds in neighboring areas, and the D-Vine Copula method can flexibly measure the correlation of such high-dimensional random variables.

Keywords

D-Vine Copula, Pair-Copula Construction, Wind Speed, Correlation

1. 引言

风速概率分布对风能潜力的评估十分重要[1]。对于地理位置相距较近的临近地区，在相同时间基本处于同一风速带，其风速间具有一定的相关性，研究这种相依关系对合理利用风能以及建设高效输出的风电场都有重大意义[2]。

有很多经典的测度二元随机变量相关性的方法，如皮尔逊相关系数、斯皮尔曼相关系数、Kendall 秩相关系数、联合密度等，但这些方法延拓到高于二维问题时面临困难。Skla [3]提出了使用 Copula 函数从概率角度测度多元变量相关性的方法，广泛应用于经济和金融领域。近来也应用于风能方面。Schindler 等[4]使用高斯 Copula 函数估计德国地面以上 100 米的风能输出。Haghi 等[5]介绍了基于 Copula 函数构建电力系统随机变量之间联合概率计算方法，并利用正态 Copula 函数描述离岸风电场与近岸风电场出力间的相关性。Stephen [6]将 Copula 函数应用到风电不确定性分析中，采用 Copula 函数模拟单个风电场数据。张宁等人[7]应用 Copula 理论对江苏某 4 个风电场出力分布之间的相依结构进行拟合，验证了相依概率性序列运算在风电场建模与计算上的有效性。

单纯的采用多元 Copula 函数测度随机变量的相关性具有明显的局限性，比如会忽略两两组合间尾部相关性的差异。为了克服这类局限性，Joe [8]引入了 Vine Copula 结构分解多元联合密度函数。随后 Bedford 和 Cooke [9]作了进一步的研究，引入了 Pair-Copula 函数，将多元联合概率密度函数通过 Vine 结构分解为二元条件概率密度函数的乘法，构建了 R-vine 结构。Cooke [10]等人详细介绍了 Vine 这种新的相依随机变量的图形模型，以及 D-Vine 和 C-Vine 这两个特殊的 R-Vine 结构。Barthel [11]认为 Vine 结构可以灵活地捕捉多变量的关联性，并展示了选择合适的 Vine Copula 模型的方法。

Vine Copula 的灵活特性，为多维变量的相关性研究提供了新方法。例如，在金融领域，胡月等人[12]构建 Vine Copula 模型研究东盟国家汇率市场的整体相关性。Diszmann 等人[13]探讨了 R-Vine Copula 中 Copula 函数的选择和参数估计方法，并成功地将该方法应用于一个 16 维金融数据集，不仅探讨了参数的顺序估计方法，还提供了整体的极大似然估计方法。

在地理和气象方面，Guilherme [14]等使用 R-Vine Copula 构建了水力发电厂水流的时间和空间的依

赖关系模型，证明了该模型可以通过减少参数数量达到降低模型复杂性的效果，并将其应用于一组巴西水电站月径流数据集。为了对不同径流量级和不同预见期下河流短期径流预报的不确定性进行定量评估，刘源[15]等人引入 R-Vine Copula 构建预测模型。结果表明，通过这种模型计算的各统计量与实测数据相差较小，可以有效降低预报的不确定性。曾文颖[16]等人基于 R-Vine Copula 函数构建了河南四个地区极端降水的多因子联合概率分布模型，探讨了 R-Vine Copula 模型可以较好地保持原序列 Spearman 和 Kendall 秩相关系数。

在风速方面，Cai 等[17]提出用斜正态混合模型和 D-Vine Copulas 构造风速联合分布研究风电容量信用评价问题。Goh 等[18]结合主成分和 R-Vine 方法，研究了风速的统计特性。

目前，关于 Vine Copula 的理论研究比较完善，在金融和气象等领域应用比较广泛。但在拟合风速数据方面研究较少，特别是对邻近地区间最大风速这种比较极端的气候现象的相关性分析缺少研究。

D-Vine Copula 是应用广泛的一类 R-Vine Copula，其特点是结构简洁规范灵活。本文利用 D-Vine Copula，对位于内蒙古四个近邻地区(呼和浩特市区、东胜、集宁和四子王旗)气象站采集的日最大风速分布以及相关性进行探究，文章安排如下。第 2 节给出 Vine Copula 的相关知识。第 3 节构建四个近邻地区日最大风速分布并探讨相关性。第 4 节是全文的总结。

2. Vine Copula 相关知识

2.1. 基于 Copula 函数的相关性测度

对于二维随机变量 X 和 Y，设边缘分布函数为 $u=F(x)$ 和 $v=G(y)$，对应的边缘密度函数为 $f(x)$ 和 $g(y)$。用定义在空间 $[0,1]^2$ 上的二元 Copula 函数 $C(u,v;\theta)$ 表示 X 和 Y 的联合分布函数 $H(x,y)$，公式为：

$$H(x,y)=C(F(x),G(y);\theta)$$

其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_k),k\geq 1$ 为 Copula 参数，用于度量变量的相关性。X 和 Y 的联合密度函数 $h(x,y)$ 为：

$$h(x,y)=\frac{\partial^2 C(F(x),G(y);\theta)}{\partial x\partial y}=c(F(x),G(y);\theta)\cdot f(x)\cdot g(y) \tag{1}$$

这里 $c(F(x),G(y);\theta)=\dfrac{\partial^2 C(F(x),G(y);\theta)}{\partial F(x)\partial G(y)}$ 是 Copula 密度函数。

Kendall τ 相关系数和 $C(u,v;\theta)$ 的关系为：

$$\tau=4\int_0^1\int_0^1 C(u,v;\theta)\mathrm{d}C(u,v;\theta)-1$$

推广到多维情形，设 n 维随机变量 $X_1,X_2,\cdots,X_n,n\geq 2$，对应边缘分布函数为 $F_i(x_i)$，边缘密度函数为 $f_i(x_i)$，$i=1,2,\cdots,n$，则多元联合分布函数和密度函数用多元 Copula 表示为：

$$H(x_1,x_2,\cdots,x_n)=C(F_1(x_1),F_2(x_2),\cdots,F_n(x_n);\theta) \tag{2}$$

$$h(x_1,x_2,\cdots,x_n)=c(F_1(x_1),F_2(x_2),\cdots,F_n(x_n);\theta)\cdot f_1(x_1)\cdot f_2(x_2)\cdots f_n(x_n) \tag{3}$$

这里 $c(F_1(x_1),F_2(x_2),\cdots,F_n(x_n);\theta)=\dfrac{\partial^n C(F_1(x_1),F_2(x_2),\cdots,F_n(x_n);\theta)}{\partial F_1(x_1)\partial F_2(x_2)\cdots\partial F_n(x_n)}$ 为 Copula 密度函数。

后续为了符号简洁，记 $F_i(x_i)=F(x_i)$，$f_i(x_i)=f(x_i)$，$i=1,2,\cdots,n$，即省略第 i 个分布函数和密度函数表达式的下标，直接用变量 x_i 的下标 i 表示对应的分布与密度函数。

2.2. Pair-Copula 结构

利用多元 Copula 函数，直接引入参数建立高维相依结构时，难以捕捉到不同变量间的复杂多变的关系，具有一定的局限性。因此在讨论多变量相依结构问题时，Joe 提出了构造 Pair-Copula 结构的方法[8]，Bedrord 和 Cooke 进一步的发展，提出了 Pair-Copula 结构建立多元概率模型[9]。这种方法将多元概率密度函数分解为两两一组的条件二元 Pair-Copula 函数乘以边缘分布函数的形式。X 的 Pair-Copula 表示为：

$$h\left(x|v;\theta\right)=c_{x,v_j|v_{-j}}\left(F\left(x|v_{-j}\right),F\left(v_j|v_{-j}\right);\theta\right)\cdot f\left(x|v_{-j}\right)$$

其中 v 是一个 n 维向量，v_j 是 v 的一个任意分量，v_{-j} 则表示由向量 v 中不包含 v_j 的部分组成的向量。Pair-Copula 结构中的条件边缘分布 $F\left(x|v\right)$ 则由以下公式得到：

$$F\left(x|v\right)=\frac{\partial C_{x,v_j|v_{-j}}\left(F\left(x|v_{-j}\right),F\left(v_j|v_{-j}\right);\theta\right)}{\partial F\left(v_j|v_{-j}\right)}$$

2.3. Vine Copula 结构

对于高维联合分布，Pair-Copula 分解有很多种方法，逻辑结构也不唯一。本文采用 Bedford 和 Cooke 引入的 R-Vine Copula 这种图形结构，将 n 元概率分布函数通过 Vine 结构分解为二元 Pair-Copula 函数的乘积，每对 Pair-Copula 的选择是独立的，可以用 $n-1$ 棵树表示[10]。R-Vine Copula 的两种特殊结构：C-Vine Copula 和 D-Vine Copula 应用最为广泛。n 维随机变量的 C-Vine Copula 的密度函数为：

$$h\left(x_1,x_2,\cdots,x_n\right)=\prod_{k=1}^{n}f_k\left(x_k\right)\prod_{i=1}^{n-1}\prod_{j=1}^{n-i}c_{i,i+j|1:(i-1)}\left(F\left(x_i|x_1,\cdots,x_{i-1}\right),F\left(x_{i+j}|x_1,\cdots,x_{i-1}\right)\middle|\theta_{i,i+j|1:(i-1)}\right)\tag{4}$$

C-Vine Copula 结构的特点是每棵树需有一个根节点(变量)与其于各节点相连，这限制了其应用，因为实践中相关的多维变量往往没有主变量。D-Vine Copula 结构灵活性强，可以有效克服这个弱点。n 维随机变量的 D-Vine Copula 的密度函数表示为：

$$h\left(x_1,x_2,\cdots,x_n\right)=\prod_{k=1}^{n}f_k\left(x_k\right)\prod_{i=1}^{n-1}\prod_{j=1}^{n-i}c_{j,j+i|(j+1):(j+i-1)}\left(F\left(x_j|x_{j+1},\cdots,x_{j+i-1}\right),F\left(x_{j+i}|x_{j+1},\cdots,x_{j+i-1}\right)\middle|\theta_{j,j+i|(j+1):(j+i-1)}\right)\tag{5}$$

3. 建立四地区 D-Vine Copula 风速相依模型

本文分析的数据为位于内蒙古西部的呼浩特市、东胜、集宁和四子王旗四个近邻地区气象站 2016~2017 年两年间日最高风速(单位：m/s)。数据来源为：中国气象数据网(https://data.cma.cn/)。Vine Copula 的选择由 R 软件的 RVineCopula 程序包实现。

模型的构建采用 IFM (即两步极大似然法)计算模型参数的极大似然估计(MLE)。具体来说，第一步是计算各个变量边缘分布参数的 MLE，第二步是通过参数估计得到的边缘分布密度函数计算关于 Vine Copula 参数的 MLE，同时在候选模型中基于赤迟信息量准则 AIC 选择最优 D-Vine Copula 分布。

3.1. 边缘分布

根据数据，绘制出四个气象站最大风速数据的核密度和柱状图，如图 1 所示。从图 1 形状看，拟合四地最大风速的边缘分布，应选择具有正偏态的分布。因此，采用两参数的威布尔分布、伽马分布和对数正态分布这三种常见的右偏分布来拟合四地最大风速的边缘分布，密度函数见表 1。分别对四组数据拟合威布尔分布、伽马分布和对数正态分布，参数的 MLE 结果如表 2 所示。

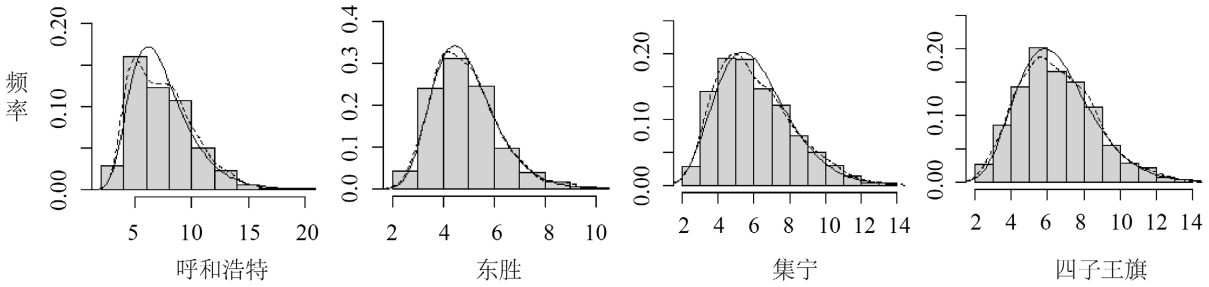

Figure 1. Histogram and fitting density curves (Axis: wind speed (m/s); Sold line: fitting density, Dotted line: kernel density)
图 1. 数据柱状图和拟合最优密度曲线(横轴：风速(m/s)；实线：拟合最优密度，虚线：核密度)

Table 1. Densities of the Weibull, the Gamma and the Log-normal distributions
表 1. 威布尔分布、伽马分布和对数正态分布密度函数

分布	密度	参数
威布尔分布	$f(x)=\alpha\beta(\beta x)^{\alpha-1}\exp\{-(\beta x)^{\alpha}\}, x>0$	$\alpha>0,\beta>0$
伽马分布	$f(x)=\frac{\lambda^{r}}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, x>0$	$\lambda>0,r>0$
对数正态	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\ln x-\mu)^{2}}{2\sigma^{2}}\right\}, x>0$	$\mu\in R,\sigma>0$

Table 2. Parameter MLEs for the wind speed marginal distributions and the corresponding RMESs
表 2. 风速边缘分布参数的 MLE 及对应的 RMSE

分布	站点	参数一	参数二	RMSE
威布尔分布	呼和浩特	2.9249	8.3237	0.4865
	东胜	3.8728	5.3794	0.3164
	集宁	3.0446	6.8166	0.3462
	四子王旗	3.3507	7.2556	0.2560
伽马分布	呼和浩特	8.1143	1.0936	0.2679
	东胜	15.4058	3.1521	0.1350
	集宁	**8.5729**	**1.4080**	**0.1471**
	四子王旗	**9.7088**	**1.4903**	**0.1186**
对数正态	**呼和浩特**	**1.9413**	**0.3545**	**0.2618**
	东胜	**1.5539**	**0.2552**	**0.0742**
	集宁	1.7469	0.34674	0.1926
	四子王旗	1.8217	0.32935	0.2871

对于这三种分布的拟合结果，通过均方根误差(RMSE)，选择出最优的边际分布，其值越小，拟合效果越好，表达式为：

$$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{n}\left(x_{obs}-x_{pre}\right)^{2}\Big/n}$$

其中，x_{obs} 表示观测值，x_{pre} 表示拟合值，RMSE 的计算结果列于表 2 最后一列。

从表 2 看，RMSE 越小则分布的拟合效果越好，所以选择对数正态分布拟合呼和浩特和东胜两地风速的边缘分布，而集宁和四子王旗两地的风速选择伽马分布拟合为最佳，由此得出的四地风速数据最优边缘分布拟合以及参数估计见表 2 (斜黑体)。最优拟合密度曲线添加在图 1 中。由图 1 可知，柱状图、核密度图和拟合密度图相匹配，充分说明选择的密度对数据拟合效果很好。

3.2. 变量顺序的选择

拟合 Vine-Copula 模型首先要确定合适的 Vine 结构。C-Vine 需要确定每棵树的根节点，而 D-Vine 则需要确定第一棵树的变量顺序。本文使用 Kendall τ 值作为选择变量顺序的标准。对四维风速序列，计算得 Kendall τ 矩阵为：

$$\tau = \begin{bmatrix} 1.0000 & 0.2197 & 0.2901 & 0.2073 \\ 0.2197 & 1.0000 & 0.2970 & 0.3698 \\ 0.2901 & 0.2970 & 1.0000 & 0.4430 \\ 0.2073 & 0.3698 & 0.4430 & 1.0000 \end{bmatrix}$$

由 τ 值可以看出，第一列中第三行的元素值最大，第二列和第三列中均为第四行的元素值最大，即呼和浩特和集宁、东胜和四子王旗、集宁和四子王旗相关性最强。显然，四地的风速数据中，并没有和另外三个变量均有强相关性的主变量，因此无法选择根节点建立 C-Vine 结构，故选择 D-Vine。由此得到第一棵树的结构，进而确定整体 D-Vine 结构如图 2 所示。在图 2 中，呼和浩特、东胜、集宁和四子王旗对应站点依次编码为 1，2，3 和 4，目的是为了 Pair-Copula 结构清晰简洁。后续表 4 和图 3，以及变量下标也是如此。

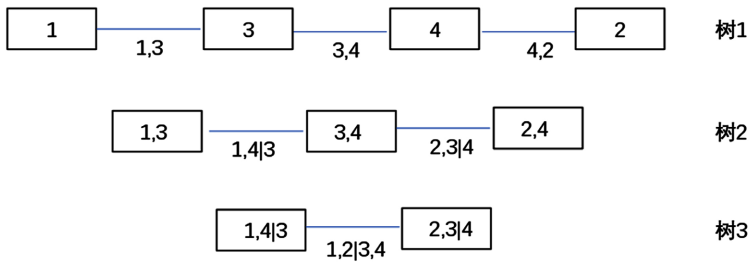

Figure 2. D-Vine Copula structure plot
图 2. D-Vine Copula 结构图

3.3. Copula 族的选择和参数估计

确定 D-Vine 结构后，进一步对两两之间的二元 Pair-Copula 进行选择，并对其参数进行估计。

3.3.1. Copula 族的选择

二元 Copula 主要分为椭圆分布族和阿基米德分布族。椭圆分布族中常用的主要是 Normal Copula 和 t-Copula 两种，用于描述对称的尾部相关性。阿基米德分布族中 Clayton Copula，Gumbel Copula 和 Frank Copula 是常用的三种 Copula 函数，分别拟合上尾、下尾和对称相关性。这五种 Copula 函数代表了变量间常见的关联特性，因此本文选择这五种 Copula 函数作为候选的 Pair-Copula，分布函数以及对应的参数列于表 3。通过计算每种 Copula 拟合结果的 AIC 值选择最优的 Pair-Copula，AIC 值越小则说明拟合效果越好。

Table 3. Five bivariate Copula probability functions
表 3. 五种二元 Copula 函数分布

名称	Copula 分布函数	参数
Normal-Copula	$C(u,v;\rho)=\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right\}\mathrm{d}s\mathrm{d}t$	$\lvert\rho\rvert<1$
t-Copula	$C(u,v;\rho,k)=\int_{-\infty}^{t_k^{-1}(u)}\int_{-\infty}^{t_k^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{s^2-2\rho st+t^2}{k(1-\rho^2)}\right)^{-(k+2)/2}\mathrm{d}s\mathrm{d}t$	$\lvert\rho\rvert<1,k>0$
Clayton	$C(u,v;\theta)=\max\left\{\left(u^{-\theta}+v^{-\theta}-1\right)^{\frac{1}{\theta}},0\right\}$	$\theta\in[-1,+\infty)\setminus 0$
Gumbel	$C(u,v;\theta)=\exp\left\{-\left[(-\ln u)^{\theta}+(-\ln v)^{\theta}\right]^{\frac{1}{\theta}}\right\}$	$\theta\in[1,+\infty)$
Frank	$C(u,v;\theta)=-\frac{1}{\theta}\ln\left(1+\frac{\left(\mathrm{e}^{-\theta u}-1\right)\left(\mathrm{e}^{-\theta v}-1\right)}{\mathrm{e}^{-\theta}-1}\right)$	$\theta\in(-\infty,+\infty)\setminus 0$

3.3.2. Copula 参数估计

确定了 D-Vine 结构和选择的 Copula 族后，对每组 Pair-Copula，采用逐树和联合密度两种方法对参数做 MLE。设 $\left(x_1^{(1)},x_2^{(1)},\cdots,x_n^{(1)}\right),\left(x_1^{(2)},x_2^{(2)},\cdots,x_n^{(2)}\right),\cdots,\left(x_1^{(N)},x_2^{(N)},\cdots,x_n^{(N)}\right)$ 是维数为 n，容量为 N 的样本。首先，按顺序逐树对每组 Pair-Copula 参数计算 MLE。由公式(5)，第 m, $m=1,2,\cdots,n-1$ 棵树的似然函数为：

$$L_m(\theta)=\prod_{k=1}^{N}\prod_{i=m}^{n-1}\prod_{j=1}^{n-i}c_{i,j|(j+1):(j+i-1)}\left(F\left(x_j^{(k)}\middle|x_{j+1}^{(k)},\cdots,x_{j+i-1}^{(k)}\right),F\left(x_{j+i}^{(k)}\middle|x_{j+1}^{(k)},\cdots,x_{j+i-1}^{(k)}\right);\theta_{j,j+i|(j+1):(j+i-1)}\right)$$

由此计算出第 m 棵树候选 Pair-Copula 函数参数 θ 的 MLE，然后根据 AIC 准则选择第 m 棵树的最优 Pair-Copula 函数。本文中，$n=4$。表 4 中列出了逐树计算得到的最优 Pair-Copula 函数以及对应参数的 MLE。

Table 4. Pair-Copula functions and the parameter MLEs based on the tree by tree method and joint density method
表 4. Pair-Copula 选择及参数逐树估计和联合估计结果

树	Pair-Copula	Copula 类型	(逐树方法)参数	(联合方法)参数	Kendall τ
树 1	1, 3	Gumbel Copula	1.27	1.37	0.22
	2, 4	Frank Copula	3.90	2.59	0.38
	3, 4	Normal Copula	0.62	0.65	0.43
树 2	1, 4|3	Normal Copula	0.06	0.13	0.04
	2, 3|4	Clayton Copula	0.30	0.24	0.13
树 3	1, 2|3, 4	Clayton Copula	0.72	1.33	0.26

其次，由公式(5)中得出的 n 维随机变量的 D-Vine Copula 联合分布密度函数计算 D-Vine Copula 参数的 MLE，其似然函数表示为：

$$L(\theta)=\prod_{k=1}^{N}\prod_{i=1}^{n-1}\prod_{j=1}^{n-i}c_{j,j+i|(j+1):(j+i-1)}\left(F\left(x_j^{(k)}\middle|x_{j+1}^{(k)},\cdots,x_{j+i-1}^{(k)}\right),F\left(x_{j+i}^{(k)}\middle|x_{j+1}^{(k)},\cdots,x_{j+i-1}^{(k)}\right);\theta_{j,j+i|(j+1):(j+i-1)}\right)$$

联合密度方法计算 D-Vine Copula 的相关参数后，得出的 D-Vine Copula 的 Pair-Copula 选择类型没有变化，参数估计结果仍列在表 4。

计算得到逐树拟合与联合拟合的似然值分别为 324.87 和 352.39，对应的 AIC 值为−637.74 和−692.78。

3.4. 模型分析与讨论

由 3.1 节以及对应的表 2 可知，基于 RMSE 值，集宁和四子王旗的风速边缘分布选用伽马分布拟合，呼和浩特和东胜风速则用对数正态拟合。尽管威布尔分布常用于拟合风速[19] [20] [21]，但对本文讨论的日最大风速数据而言，并不是最优的选择。

根据秩相关性，建立了四地日最大风速 D-Vine Copula 相依模型，获得了 Pair-Copula 结构。表 4 可知，每对 Pair-Copula 对应不同的二元 Copula 函数。这充分说明，四地风速相依概率模型不适于用常规的联合概率(3)拟合，因为(3)难以捕捉到不同变量之间相异的关联性，缺乏灵活性。D-Vine Copula 则能够较好的捕捉到多个变量之间不同的关联性。

在所选择的二元 Copula 函数中，联合拟合和逐树拟合得到的 Pair-Copula 参数 MLE 有差异，但 Pair-Copula 选择一致，见表 4。从 AIC 值看，联合拟合效果优于逐树拟合，这是因为联合拟合是全局最优估计的结果。但是，D-Vine Copula 分解的待估计参数的个数随着变量的增加而成倍增加。例如，n 个维度的变量，增加一个维度得到 $n+1$ 个变量，即使每个 Pair-Copula 函数均采用单参数模型，联合拟合参数估计方程也增加 n 个，而逐树拟合在每棵树上参数只增加一个。众所周知，MLE 往往依赖于初值。因此，如果建立联合密度计算参数 MLE，建议逐树拟合得到的 MLE 作为其初值，以降低初值选择的困难。

依据表 4 构建了关于四个站点间日最大风速的四维 D-Vine Copula 风速相依模型，Pair-Copula 参数采用联合估计的结果，联合密度函数的 Pair-Copula 的分解表达式为：

$$\begin{aligned}f(x_1,x_2,x_3,x_4)=&c_G\left(F_1(x_1),F_3(x_3);1.37\right)\cdot c_F\left(F_2(x_2),F_4(x_4);2.59\right)\cdot c_N\left(F_3(x_3),F_4(x_4);0.65\right)\\&\cdot c_N\left(F(x_1|x_3),F(x_4|x_3);0.13\right)\cdot c_C\left(F(x_2|x_4),F(x_3|x_4);0.24\right)\\&\cdot c_C\left(F(x_1|x_3,x_4),F(x_2|x_3,x_4);1.33\right)\cdot f_{LN}(x_1)\cdot f_{LN}(x_2)\cdot f_{Ga}(x_3)\cdot f_{Ga}(x_4)\end{aligned}\tag{6}$$

其中，$c_N(\cdot),c_C(\cdot),c_G(\cdot),c_F(\cdot)$ 分别表示 Normal Copula，Clayton Copula，Gumbel Copula，Frank Copula 的密度函数，该函数由表 3 分布函数对 u,v 求混合二阶偏导即可得到；$f_{LN}(\cdot)$ 和 $f_{Ga}(\cdot)$ 分别表示对数正态密度和伽马密度以及对应参数估计见表 1 和表 2。由此，我们完整构建了测度四个站点间最大风速相关性的 D-Vine Copula 相依模型，轮廓图如图 3。

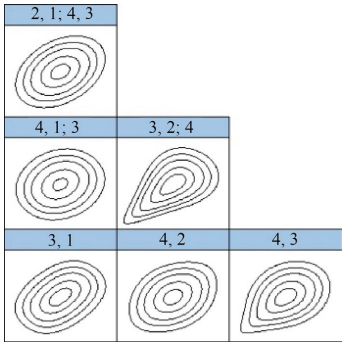

Figure 3. D-Vine contour plot
图 3. D-Vine 轮廓图

由图 3 发现，两两近邻的最大风速具有较为明显的上尾和下尾相关性。四子王旗和集宁，以及在已知四子王旗最大风速的条件下，东胜和集宁的两两下尾相关性较强一些，其他地区相关性呈对称状态。如前面所讲，针对高维相关的随机变量概率分布问题，如果用多元 Copula 函数，无法反应不同地区风速的相关性差异，而 D-Vine Copula 方法可以灵活清晰的反应多元变量之间的不同的关联关系，如对称相关、非对称相关和尾部相关。

4. 结论

本文用 D-Vine Copula 方法建立了内蒙古近邻四地日最高风速的多元联合分布。该方法是将联合分布分解为边缘分布和 Pair-Copula 函数乘积的形式。使用两阶段法得到了边缘分布和 Pair-Copula 函数的最优选择和参数的极大似然估计。其中，Pair-Copula 参数的极大似然估计采用了逐树估计和联合估计两种方法，拟合结果表明联合估计方法优于逐树估计。模拟表明风速间存在关联性，多元联合分布的 D-Vine Copula 分解方法灵活清晰的反应了不同地区的关联性差异。

参考文献

[1] Wu, J., Wang, J. and Chi, D. (2013) Wind Energy Potential Assessment for the Site of Inner Mongolia in China. *Renewable and Sustainable Energy Reviews*, **21**, 215-228. https://doi.org/10.1016/j.rser.2012.12.060

[2] 吴巍, 汪可友, 韩蓓, 等. 基于 Pair Copula 的随机潮流三点估计法[J]. 电工技术学报, 2015(9): 121-128.

[3] Sklar, A. (1959) Fonctions de Répartitionà n Dimensions et Leurs Marges. *Annales de l'ISUP*, **8**, 229-231.

[4] Schindler, D. and Jung, C. (2018) Copula-Based Estimation of Directional Wind Energy Yield: A Case Study from Germany. *Energy Conversion & Management*, **169**, 359-370. https://doi.org/10.1016/j.enconman.2018.05.071

[5] Haghi, H.V., Bina, M.T., Golkar, M.A., *et al*. (2010) Using Copulas for Analysis of Large Datasets in Renewable, Distributed Generation: PV and Wind Power Integration in Iran. *Renewable Energy*, **35**, 1991-2000. https://doi.org/10.1016/j.renene.2010.01.031

[6] Stephen, B., Galloway, S.J., McMillan, D., *et al*. (2011) A Copula Model of Wind Turbine Performance. *IEEE Transactions on Power Systems*, **26**, 965-966. https://doi.org/10.1109/TPWRS.2010.2073550

[7] 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, 2012, 52(5): 704-709.

[8] Joe, H. (1996) Families of m-Variate Distributions with Given Margins and m(m−1)/2 Bivariate Dependence Parameters. In: Rüschendorf, L., Schweizer, B. and Taylor, M.D., Eds., *Distributions with Fixed Marginals and Related Topics*, Lecture Notes—Monograph Series, Vol. 28, Institute of Mathematical Statistics, Beachwood, 120-141. https://doi.org/10.1214/lnms/1215452614

[9] Bedford, T. and Cooke, R. (2001) Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines. *Annals of Mathematics and Artificial Intelligence*, **32**, 245-268.

[10] Cooke, B. (2002) Vines: A New Graphical Model for Dependent Random Variables. *Annals of Statistics*, **30**, 1031-1068. https://doi.org/10.1214/aos/1031689016

[11] Barthel, N., Geerdens, C., Killiches, M., *et al*. (2018) Vine Copula Based Likelihood Estimation of Dependence Patterns in Multivariate Event Time Data. *Computational Statistics & Data Analysis*, **117**, 109-127. https://doi.org/10.1016/j.csda.2017.07.010

[12] 胡月, 雷柳荣, 王甜甜, 等. 东盟国家货币兑人民币汇率相依性研究——基于 Vine Copula 模型[J]. 数学的实践与认识, 2021, 51(18): 89-101.

[13] Diímann, J., Brechmann, E.C., Czado, C., *et al*. (2013) Selecting and Estimating Regular Vine Copulae and Application to Financial Returns. *Computational Statistics & Data Analysis*, **59**, 52-69. https://doi.org/10.1016/j.csda.2012.08.010

[14] Pereira, G., Veiga, A., Erhardt, T., *et al*. (2016) Spatial R-Vine Copula for Streamflow Scenario Simulation. 2016 *Power Systems Computation Conference* (*PSCC*), Genoa, 20-24 June 2016, 540-546. https://doi.org/10.1109/PSCC.2016.7540939

[15] 刘源纪, 昌明, 张验科, 等. 基于 Vine Copula 的短期径流预报不确定性分析[J]. 水力发电学报, 2022, 41(7): 95-105.

[16] 曾文颖, 徐明庆, 宋松柏, 等. 基于 R-Vine Copula 函数的极端降水联合分布模型及风险识别[J]. 水资源保护, 2022, 38(6): 96-103.

[17] Cai, J.L., Xu, Q.S., Cao, M.J., *et al*. (2019) Capacity Credit Evaluation of Correlated Wind Resources Using Vine Copula and Improved Importance Sampling. *Applied Sciences*, **9**, Article No. 199. https://doi.org/10.3390/app9010199

[18] Goh, H., Peng, G.M., Zhang, D.D., *et al*. (2022) A New Wind Speed Scenario Generation Method Based on Principal Component and R-Vine Copula Theories. *Energies*, **15**, Article No. 2698. https://doi.org/10.3390/en15072698

[19] Wais, P. (2017) A Review of Weibull Function in Wind Sector. *Renewable & Sustainable Energy Reviews*, **70**, 1099-1107. https://doi.org/10.1016/j.rser.2016.12.014

[20] Patidar, H., Shende, V., Baredar, P., *et al*. (2022) Comparative of Optimal Weibull Parameters for Wind Power Predictions Using Numerical and Metaheuristic Optimization Methods for Different Indian Terrains. *Environmental Science and Pollution Research*, **30**, 30874-30891. https://doi.org/10.1007/s11356-022-24395-6

[21] Hussain, I., Haider, A., Ullah, Z., *et al*. (2023) Comparative Analysis of Eight Numerical Methods Using Weibull Distribution to Estimate Wind Power Density for Coastal Areas in Pakistan. *Energies*, **16**, Article No. 1515. https://doi.org/10.3390/en16031515