

DOI: 10.11835/j.issn.2096-6717.2021.148



开放科学(资源服务)标识码 OSID:



基于 Vine Copula 函数的风浪要素联合概率分布模型

王望, 朱金, 康锐, 李永乐

(西南交通大学桥梁工程系, 成都 610031)

摘要:随着全球气候变暖的加剧, 极端气候现象发生的频率和强度均可能加大, 这对海岸和近海结构的安全不利。基于中国东海的连云港海洋观测站实测风浪数据和 Vine Copula 理论, 建立风浪要素中风速、波高、波浪周期、风向和波向五维随机变量之间的联合概率分布模型。采用极大似然法确定各风浪要素边缘分布模型参数, 通过 AIC 信息准则和均方根误差 RMSE 进行拟合优度评价, 由此建立风浪要素的边缘分布。采用带有基于残差的高斯似然函数的贝叶斯框架估计二维 Copula 函数的参数, 结合 AIC 信息准则进行拟合优度评价并确定最优 Copula 函数。绘制最优联合分布概率密度图, 与二维频率直方图进行对比以评价模型效果。采用 Vine Copula 函数建立多维联合概率模型并结合 AIC 值评价其拟合优度。研究表明: 建立的 Vine Copula 模型可以较好地刻画风速、波高、波浪周期、风向和波向五维随机变量之间的联合概率分布。

关键词:风浪联合概率分布模型; 风浪荷载; 参数估计; 拟合优度检验

中图分类号:U441.2 **文献标志码:**A **文章编号:**2096-6717(2023)04-0083-11

Joint probability distribution model of wind and wave with Vine Copula function

WANG Wang, ZHU Jin, KANG Rui, LI Yongle

(Department of Bridge Engineering, Southwest Jiaotong University, Chengdu 610031, P. R. China)

Abstract: With the intensification of global climate warming, the probabilities and load intensities of extreme weather phenomenon are gradually increasing, which could threaten the safety of coastal and offshore infrastructures. The present study presents a joint probability distribution model of wind speed, wave height, wave period, wind direction and wave direction with Vine Copula function based on monitoring data from Lianyungang Ocean Station in the East China Sea. Firstly, the marginal probability distributions of wind and wave data are determined, in which the AIC criteria and RMSE index are employed to select the optimal probability distribution model and the maximum likelihood method is used to determine the model parameters.

收稿日期:2021-05-06

基金项目:国家自然科学基金(51908472); 四川省科学技术厅科技计划(2020YJ0080); 中国博士后科学基金(2019M663554、2019TQ0271)

作者简介:王望(1994-), 男, 主要从事大跨桥梁风浪耦合动力学研究, E-mail: wangwang@my.swjtu.edu.cn.

朱金(通信作者), 男, 助理研究员, 博士(后), E-mail: zhujin@home.swjtu.edu.cn.

Received: 2021-05-06

Foundation items: National Natural Science Foundation of China (No. 51908472); Project of Science and Technology Department of Sichuan Province (2020YJ0080); China Postdoctoral Science Foundation (No. 2019M663554, 2019TQ0271)

Author brief: WANG Wang (1994-), main research interest: coupled wind-wave-bridge dynamics of long-span bridges, E-mail: wangwang@my.swjtu.edu.cn.

ZHU Jin (corresponding author), associate research fellow, postdoctor, E-mail: zhujin@home.swjtu.edu.cn.

Subsequently, the optimal two-dimensional Copula function for wind and wave data is determined via the AIC criteria, and the model parameters are fitted with a Bayesian framework with a residual-based Gaussian likelihood function. To illustrate the goodness of fit, the binary frequency histogram of the original wind and wave data is compared with the proposed two-dimensional Copula function. Finally, the multi-dimensional joint probability distribution model of wind and wave data is established with the Vine Copula function based on the AIC criteria. The results show that the proposed Vine Copula model is able to describe the joint probability distribution between the wind speed, wave height, wave period, wind direction and wave direction.

Keywords: joint probability distribution model of wind and wave; wind and wave load; parameter estimation; goodness of fit test

海洋环境恶劣,往往存在大风和巨浪的组合作用,同时,跨海大桥具有基础结构尺寸大、主跨轻柔、阻尼小、刚度小等特点,导致跨海大桥对风浪等海洋环境荷载非常敏感,极端风浪荷载成为威胁跨海大桥安全的主要因素之一。目前,跨海大桥多灾害作用的研究中通常研究单一灾害(如风、波浪、地震、冲刷等)对桥梁结构的作用,未能很好地考虑各灾害之间的相关性。特别是风浪要素,其包含了众多影响跨海大桥动力特性的参数,如风速、风向、波高、波向、波浪周期等,各参数是随时间变化的随机变量,且相互之间具有不同的尾部相关结构(即非线性)。因此,构造风浪要素的联合分布函数是研究海洋环境中风浪耦合效应及进一步探究风浪各要素对跨海大桥动力响应特征影响规律的基本前提。

Copula 函数将多维联合分布分解为相应的边缘分布和 Copula 函数之积,可以灵活地表示两两随机变量之间的相关结构。近年来,Copula 函数在土木工程领域逐渐得到关注。Li 等^[1]基于 C-Vine Copula 理论,研究了阿拉斯加南部海岸风浪相关性,并由建立的模型推算了特定重现期的长期荷载。Zhang 等^[2]采用椭圆 Copula 和三维 Plackett Copula 建立了风速、风向和温度的三维联合分布。Bai 等^[3]采用混合二维 Copula 来描述风浪环境变量的相关性并与采用单一 Copula 的情况进行对比,研究表明,混合 Copula 能更好地描述多维变量之间复杂的相关关系。Wang 等^[4]基于 Copula 理论,提出了一种利用气象资料研究大跨桥梁风温联合作用的方法并应用于常泰长江大桥。Zhang 等^[5]基于非对称 Copula 函数,模拟了海洋环境要素的相关关系,侧重于捕捉环境要素的非对称相关关系。Li 等^[6]在浮式风力发电机组的概率疲劳评估研究中,基于 Copula 函数建立了多维概率模型,用来定义风浪环境参数相关性。Vanem^[7]对特征波高和波浪周期的同时进行了多维分布研究,结果表明,使用非对称 Copula 函数能较好地模拟非对称的相关关系。Yang 等^[8]基于三维 Copula 函数建立了风速、风暴潮和暴雨的联合概率分布,提出了一种有

效的 PSO 算法来估计边缘分布和联合分布的参数,并与极大似然法和对数似然法进行对比验证。Wang 等^[9]通过建立挪威 Sulafjord 的短期近海数据和桥址区数据之间的定量关系,推算了桥址区长期风浪数据,从而建立了风浪联合分布。Zhang 等^[10]通过基于 Copula 的多维概率模型,研究了考虑多种环境因素的近海结构长期荷载,由此建立了多种常见海况参数之间的联合概率模型。Xu 等^[11]将 Copula 函数应用于海洋结构可靠度的分析,描述了波高、波峰和平均风速 3 个近海环境参数之间的相关性,研究表明,该方法可用于考虑长期疲劳荷载和极端响应的海洋结构可靠度分析。陈子桑等^[12]采用非对称 Archimedean Copula 函数分析极端波况下波高、风速和周期的三变量联合概率分布,计算了“或”重现期、“且”重现期和二次重现期,探讨了规范设计值的安全性并给出了建议。Copula 理论不仅在海洋环境要素相关性研究方面有广泛的应用,在结构可靠度等其他领域也有重要的应用。陈建兵等^[13]基于 Copula 理论,建立了混凝土参数之间的相关性模型,由生成的样本计算了混凝土本构全曲线。樊学平等^[14]基于 R-Vine Copula 理论对大跨桥梁主梁检测点失效概率的相关性进行了研究,建立了检测点的相关性模型。刘月飞等^[15]基于桥梁检测点的极值应力数据,建立了描述检测变量非线性相关的 Vine Copula 模型。宋帅等^[16]将混合 Copula 方法应用于桥梁系统地震易损性分析中,准确描述了构件地震需求的非对称相关关系,简化了联合分布模型的建立过程。

综上所述,以上基于 Copula 函数的海洋环境参数联合分布的研究极大促进了跨海桥梁的建设,但是目前相关研究多针对二维及三维的海洋环境参数,这对准确模拟复杂多变的海洋环境来说是不够的。如前所述,海洋风浪要素中的风速、风向、波高、波向、波浪周期对于跨海大桥的动力响应均有重要影响,然而,目前针对风浪要素多维联合分布的研究还鲜有报道。笔者在单一 Copula 函数的基础上,基于 Vine Copula 理论建立了海洋风浪要素中风速、风向、波高、波向、

波浪周期五维变量之间的联合分布模型,从而准确刻画了风浪要素之间的相关关系。首先,建立风浪各个要素的边缘概率模型,采用均方根误差(RMSE)进行拟合优度评价;在得到风浪各要素边缘分布的基础上,基于Copula理论,建立风浪要素两两之间的二维联合概率分布模型,通过AIC信息准则和均方根误差RMSE进行拟合优度评价,并考察风浪要素两两之间的相关性;基于Vine Copula理论,采用C-Vine结构构建了风浪要素中风速、风向、波高、波向、波浪周期五维变量之间的联合概率分布模型。通过AIC准则对模型进行拟合优度评价。

1 边缘分布

1.1 数据说明

采用位于中国东海的连云港海洋观测站2016—2020年波浪和风场观测数据,数据由中国国家科技资源共享服务平台——国家海洋科学数据中心(<http://mds.nmdis.org.cn/>)提供数据支撑。选用的风浪要素包括10 m高度处最小平均风速、特征波高、波浪周期、风向和波向,测量频率为每小时测量一次,站点的经纬度为34°47'0"N 119°26'0"E,最大风速为22 m,达到强风等级,最大波高为2.4 m。风浪数据信息如表1所示。需要说明的是,笔者在进行数据处理时发现,海洋站的观测数据中有很少一部分数据存在缺失的情况,即有个别或多个要素的数据没有观测到。针对这种情况,将缺失的样本数据予以剔除,尽可能多地保留其余数据样本,最后得到的样本总量为29 363个。

表1 风浪要素数据信息

Table 1 Information of wind and wave data

风浪要素	说明
风速	10 min 平均风速,单位 m/s
波高	特征波高,单位 m
波周期	单位 s
风向	单位度(°),正北为0°,度数沿顺时针增加
波向	单位度(°),正北为0°,度数沿顺时针增加

1.2 边缘分布

首先,需要建立风浪要素的边缘分布模型。研究中发现,风速、波高、波周期样本具有单峰分布的特征(图1(a)~(c)),采用常见的单峰分布模型进行拟合,包括Weibull、广义极值分布(GEV)、含有尺度参数和位置参数的 t 分布等。风向和波向具有多峰分布的特征(图1(d)、(e)),采用混合模型进行拟合,包括混合Gaussian分布、混合Gamma分布、混合Weibull分布。

1) Weibull分布

$$f(x, \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \quad (1)$$

式中: λ 为尺度参数; k 为形状参数。

2) 广义极值分布(GEV)

$$F_X(x) = P(X < x) =$$

$$\exp - \left\{ - \left[1 - \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{1/\xi} \right\}, \xi \neq 0 \quad (2)$$

式中: ξ, β, μ 为分布函数参数; ξ 为形态参数; σ 为尺度参数; μ 为位置参数。

3) 含有尺度参数和位置参数的 t 分布

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left[\frac{\nu + \left(\frac{x-\mu}{\sigma}\right)^2}{\nu} \right]^{-\left(\frac{\nu+1}{2}\right)} \quad (3)$$

式中: ν, σ, μ 均为分布函数参数,其中: ν 为形态参数; σ 为尺度参数; μ 为位置参数。

4) 混合Gaussian模型

$$f(\theta) = \sum_{i=1}^3 w_i \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{(\theta - \mu_i)^2}{\sigma_i^2} \right\} \quad (4)$$

式中: $\sum_{i=1}^3 w_i = 1; \mu_i, \sigma_i$ 分别为相应部分Gaussian分布的均值和方差。

5) 混合Gamma模型

$$f(\theta) = \sum_{i=1}^c w_i \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \theta^{\alpha_i-1} e^{-\beta_i\theta}, \theta > 0 \quad (5)$$

式中: $\sum_{i=1}^c w_i = 1; \alpha_i, \beta_i$ 分别为相应部分Gamma分布的形状参数和尺度参数。

当采用上述概率分布模型对风浪要素进行拟合时,概率分布模型的参数估计采用极大似然法。另外,为了评价不同概率分布模型的拟合效果,采用AIC、BIC和RMSE对概率分布模型进行拟合优度评价,并据此选取最优的概率分布模型。AIC、BIC和RMSE的计算式为

$$AIC = -2 \sum_{i=1}^n \ln f(x_i) + 2k \quad (6)$$

式中: x_i 为样本值; n 为样本数量; $f(x_i)$ 为备选边缘分布函数的密度函数; k 为备选边缘分布函数中分布参数的数量。

$$BIC = -2 \sum_{i=1}^n \ln f(x_i) + k \ln n \quad (7)$$

式中: x_i 为样本值; n 为样本数量; $f(x_i)$ 为备选边缘分布函数的密度函数; k 为备选边缘分布函数中分布参数的数量。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [P_c(i) - P_0(i)]^2} \quad (8)$$

式中： n 为样本数量； P_c 为多维 Copula 联合分布理论频率值。RMSE 值越小，拟合的效果越好。

通过观察风向概率直方图(图 1(d))可以看出，风向的概率分布有 3 个较为明显的峰值，因此，采用的混合概率模型中应包含 3 个单峰分布，故采用混合维度为 3 的混合模型(3 个单峰分布函数混合)来拟合。与风向类似，波向频率直方图(图 1(e))也有多个峰值，分别采取二维和三维混合模型对波向的

概率分布进行拟合，通过对比 RMSE 发现，三维混合模型的概率拟合效果较好。表 2 和表 3 给出了风速、波高、波浪周期、风向、波向 5 个风浪要素的最优边缘概率分布类型和相应参数。此外，图 1 还给出了这 5 个风浪要素样本的直方图及最优边缘概率分布曲线。由图 1 可以看出，选取的最优边缘概率分布曲线对 5 个风浪要素样本的拟合效果较好。拟合优度评价结果表明：风速、波高、波周期的最优拟合分布分别为 Weibull 分布、广义极值分布(GEV)、含有尺度参数和位置参数的 t 分布；而风向和波向的最优拟合分布均为混合 Gaussian 分布。

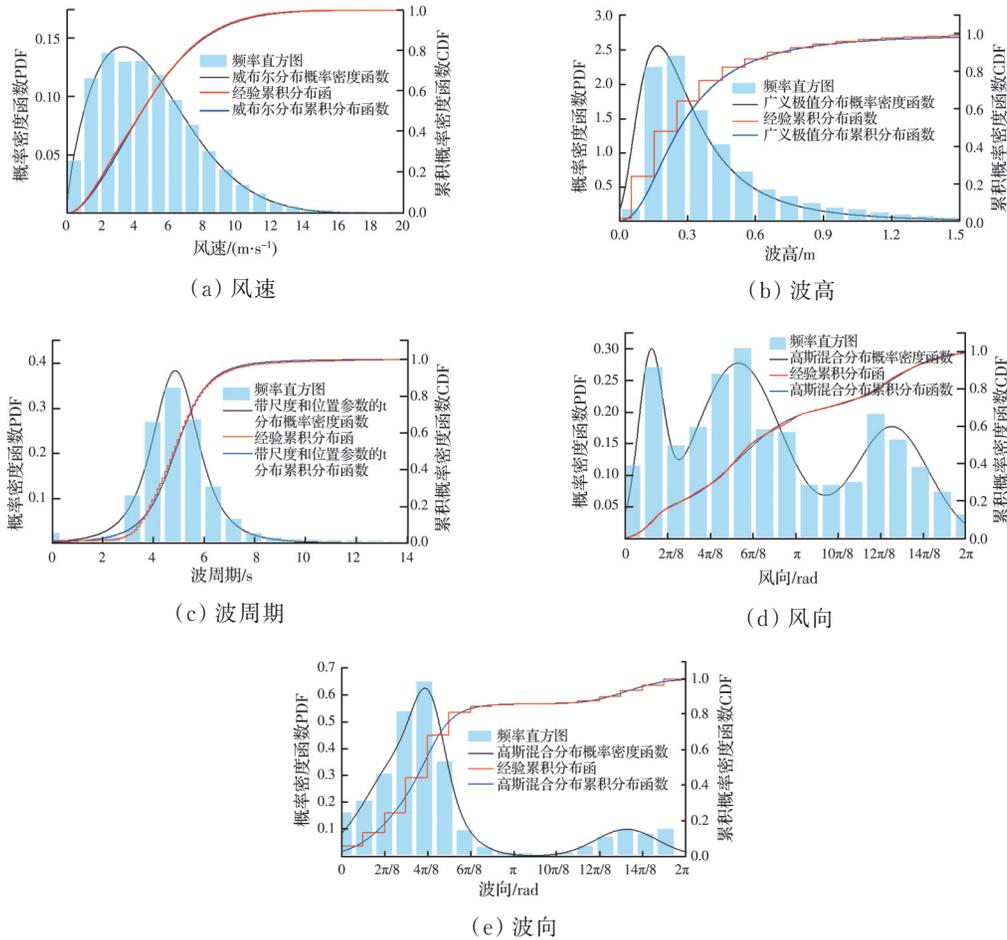


图 1 风浪要素概率分布直方图及最优边缘概率分布曲线

Fig. 1 Histogram of wind and wave data and the optimal marginal probability distribution curve

表 2 风速、波高、波周期最优边缘概率分布

Table 2 The optimal marginal probability distribution of wind speed, wave height and wave period

环境要素	分布	BIC	AIC	尺度参数	形态参数	位置参数	RMSE
风速 V_w	Weibull	140 950	140 930	5.496 2	1.728 8		0.006 3
波高 H_s	Generalized Extreme Value	-7 914	-7 939	0.150 4	0.313 2	0.205 8	0.052 7
波周期 T_p	七分布	101 130	101 110	0.959 7	3.090 1	4.863 0	0.015 3

表 3 风向、波向边缘分布

Table 3 The marginal distribution of wind direction and wave direction

环境要素	函数类型	拟合参数									RMSE		
风向 θ_w	混合 Gaussian	参数	w_1	w_2	w_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	0.006 3	
		取值	0.149 7	0.552 3	0.298 0	0.461 0	2.080 3	4.923 6	0.225 5	0.793 8	0.672 0		
	混合 Gamma	参数	w_1	w_2	w_3	λ_1	λ_2	λ_3	k_1	k_2	k_3		0.007 3
		取值	0.381 1	0.346 2	0.272 6	0.212	0.827	0.075 9	11.428 3	1.388 9	65.938 1		
波向 θ_s	混合 Gaussian	参数	w_1	w_2	w_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	0.046 5	
		取值	0.646 6	0.212 8	0.140 7	1.173 2	1.590 6	5.202 6	0.677 6	0.283 1	0.564 1		
	混合 Weibull	参数	w_1	w_2	w_3	λ_1	λ_2	λ_3	k_1	k_2	k_3		0.048 5
		取值	0.799 2	0.142 4	0.058 4	1.533 5	5.433 6	0.000 1	2.748 1	10.909 7	63.643 4		

2 二维联合分布

2.1 二维 Copula 理论

根据 Sklar 定理^[17],多维联合分布和其边缘分布可以写为

$$F(x_1 \cdots x_n) = C(F_1(x_1) \cdots F_n(x_n); \theta) \quad (9)$$

式中: $F_i(x_i)$ 为随机变量 x_i 的边缘概率分布函数; $C(\cdot)$ 为 Copula 函数; θ 为 Copula 函数的参数。

将海洋环境变量定义为一个 n 维随机变量 $X = (x_1, x_2 \cdots x_i \cdots x_n)$, 基于 Copula 理论, 联合概率密度可以表示为

$$f(x_1, x_2 \cdots x_i \cdots x_n) = c(F_1(x_1), F_2(x_2) \cdots F_i(x_i) \cdots F_n(x_n)) \prod_{i=1}^n f(x_i) \quad (10)$$

式中: $F_i(x_i)$, $f(x_i)$ 分别为随机变量 x_i 的边缘概率分布的分布函数和概率密度函数; c 为 Copula 密度函数。

由式(10)可得二维联合分布概率密度函数公式

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f(x_1)f(x_2) \quad (11)$$

式中: c 为二维 Copula 函数的密度函数。

由式(11)可知,为了研究风浪要素之间的二维联合概率分布,首先应建立风浪要素两两变量之间的 Copula 函数。选取风速-波高、波高-波周期、风速-风向、波高-波向 4 个随机变量对来研究并建立二维 Copula 函数。目前,研究风浪要素联合分布的大多数文献均将这些变量对作为研究对象。其次,这些变量对中的两个变量之间往往存在较大的相关性,而这些相关性在进行结构(如海上风机、跨海桥梁等)的动力响应分析时又至关重要。样本数据经过式(12)可以变换为范围 0~1 的标准分布,样本的标准分布忽略了样本数值大小的差异,只保留数据的相对大小关系,以便于更清晰地研究数据之间的相关性。图 2 是风速-波高经验 Copula 频率直方图,可以看出,风速和波高具有较强的尾部相关性。

$$U = \frac{n - R + 0.5}{n} \quad (12)$$

式中: U 为将样本变换为范围为 0~1 的标准分布后的随机变量; n 为样本个数; R 为样本点在所有样本中的排序。

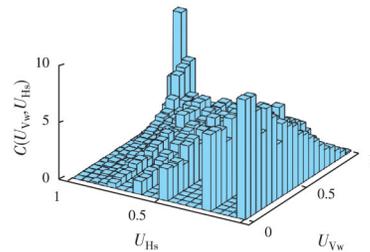


图 2 风速-波高标准分布二维频率直方图

Fig. 2 Binary frequency histogram of standard distribution of wind and wave

常用的二维 Copula 函数类型有: Gaussian、T、Clayton、Gumbel、Frank, 如表 4 所示。

表 4 5 种典型的 Copula 函数

Table 4 Five typical Copula functions

Copula 函数	函数公式	ρ 范围
Gaussian	$\Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2) \rho)$	$[-1, 1]$
T	$t_{\rho,n}(t_n^{-1}(u_1), t_n^{-1}(u_2) \rho)$	$[-1, 1]$
Clayton	$(u_1^{-\rho} + u_2^{-\rho} - 1)^{-1/\rho}$	$(0, \infty)$
Gumbel	$\exp(-((-\ln u_1)^\rho + (-\ln u_2)^\rho)^{1/\rho})$	$[1, \infty)$
Frank	$-\frac{1}{\rho} \ln \left(1 + \frac{(e^{-\rho u_1} - 1)(e^{-\rho u_2} - 1)}{e^{-\rho} - 1} \right)$	$(-\infty, \infty) \setminus \{0\}$

2.2 拟合优度评价

要进行拟合优度评价,首先要计算经验 Copula,经验 Copula 可以通过式(13)计算^[18]。

$$\hat{C}_E(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n I(u_{1i} \leq u_1, u_{2i} \leq u_2) \quad (13)$$

式中: n 为样本的大小,对每一个 $1 \leq i \leq n$, 满足 $u_{1i} \leq u_1, u_{2i} \leq u_2$ 时, $I(u_{1i} \leq u_1, u_{2i} \leq u_2) = 1$ 。

对应三维的情况,经验 Copula 可通过式(14)计算。

$$\hat{C}_E(u_1, u_2, u_3) = \frac{1}{n} \sum_{i=1}^n I(u_{1i} \leq u_1, u_{2i} \leq u_2, u_{3i} \leq u_3) \quad (14)$$

式中: n 为研究样本的大小, 对每一个 $1 \leq i \leq n$, 满足 $u_{1i} \leq u_1, u_{2i} \leq u_2, u_{3i} \leq u_3$ 时, $I(u_{1i} \leq u_1, u_{2i} \leq u_2, u_{3i} \leq u_3) = 1$ 。

为了对选取的 Copula 函数进行拟合优度评价, 选用均方根误差 (RMSE) 作为评价标准来评价模型的优劣。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left| \hat{C}_E[(u_1)_i, (u_2)_i] - C[(u_1)_i, (u_2)_i] \right|^2} \quad (15)$$

式中: \hat{C}_E 为 Copula 经验值; C 为建立的模型计算出的理论值。

S_n 越小, 说明拟合的 Copula 模型与经验 Copula 越接近, 拟合效果越好。

建立二维联合分布可以对多种 Copula 函数分别进行参数估计, 采用贝叶斯框架和基于残差的高斯似然函数进行参数估计^[19]。

贝叶斯分析已在多个领域应用于参数估计。当获得新信息时, 贝叶斯理论更新假设的先验概率, 将所有建模的不确定因素归因于参数, 通过式 (16) 估计模型参数的后验分布。

$$p(\theta | \tilde{Y}) = \frac{p(\theta) p(\tilde{Y} | \theta)}{p(\tilde{Y})} \quad (16)$$

式中: $p(\theta)$ 和 $p(\theta | \tilde{Y})$ 分别表示参数的先验和后验分布; $p(\tilde{Y} | \theta)$ 可代表似然函数; $p(\tilde{Y}) = \int_{\theta} p(\tilde{Y} | \theta) d\theta$ 为证据, 证据在每次建模中都是常数。

在缺乏参数先验分布的有效信息时, 可以采用均匀先验, 假设残差不相关, 同方差、均值为零的高斯分布, 那么似然函数就可以通过式 (17) 表示。

$$L(\theta | \tilde{Y}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} [\tilde{y}_i - y_i(\theta)]^2\right\} \quad (17)$$

式中: $\tilde{\sigma}$ 为测量误差的标准偏差估计值; \tilde{y}_i 为观测变量的联合概率; $y_i(\theta)$ 为 Copula 预测概率。为了简洁和数值稳定, 式 (17) 可以采用对数转换成为

$$l(\theta | \tilde{Y}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \tilde{\sigma}^2 - \frac{1}{2\tilde{\sigma}^2} \sum_{i=1}^n [\tilde{y}_i - y_i(\theta)]^2 \quad (18)$$

通过 AIC、BIC 及 RMSE 对不同类型的 Copula 模型进行拟合优度评价, 拟合优度评价结果见表 5, 参数估计值见表 6。风速-波高的最优联合分布为

$\theta = 0.8461$ 的 Gaussian Copula; 波高-波浪周期的最优联合分布为 $\theta = 0.4653$ 的 Gaussian Copula; 风速-风向的最优联合分布为 $\theta = -1.1457$ 的 Frank Copula; 波高-波向的最优联合分布为 $\theta = 0.5410$ 的 Gaussian Copula。图 3 为 $\theta = 0.8461$ 的 Gaussian Copula 函数图像。

表 5 二维 Copula 拟合优度评价

Table 5 The performance evaluation of 2-dimensional Copula distribution

随机变量	Copula 类型	AIC	BIC	ML	RMSE
风速-波高	Gaussian	-191 776	-191 768	95 889	6.54
	Clayton	-187 416	-187 408	93 709	7.05
	Frank	-190 322	-190 314	95 162	6.71
	Gumbel	-190 648	-190 640	95 325	6.67
	Joe	-187 587	-187 579	93 795	7.02
波高-周期	Plackett	-189 264	-189 256	94 633	6.83
	Gaussian	-197 802	-197 794	98 902	5.90
	Clayton	-193 411	-193 402	96 706	6.36
	Frank	-197 103	-197 095	98 552	5.97
	Gumbel	-196 943	-196 934	98 472	5.99
风速-风向	Joe	-195 444	-195 435	97 722	6.15
	Plackett	-196 352	-196 343	98 176	6.05
	Gaussian	-271 710	-271 701	135 855	1.68
	Clayton	-225 937	-225 929	112 969	3.66
	Frank	-274 369	-274 361	137 185	1.60
波高-波向	Gumbel	-225 919	-225 911	112 960	3.66
	Joe	-225 954	-225 946	112 978	3.66
	Plackett	-274 571	-274 562	137 286	1.60
	Gaussian	-180 147	-180 139	90 074	7.97
	Clayton	-176 902	-176 894	88 452	8.43
波高-波向	Frank	-178 757	-178 749	89 379	8.16
	Gumbel	-179 820	-179 812	89 910	8.02
	Joe	-179 358	-179 350	89 680	8.08
	Plackett	-178 097	-178 089	89 049	8.26

表 6 最优二维 Copula 分布及拟合参数

Table 6 The optimal 2-dimensional Copula distribution and fitting parameters

联合分布	Copula 类型	θ
风速-波高	Gaussian	0.8461
波高-周期	Gaussian	0.4653
风速-风向	Frank	-1.1457
波高-波向	Gaussian	0.5410

由图 2 和图 3 可以看出, Gaussian Copula 能较好地模拟风速-波高的相关性。此外, 样本数据的二维频率直方图可以表示联合分布的经验图像, 与

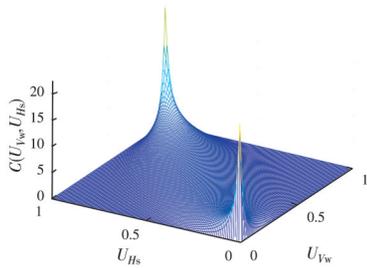


图 3 Gaussian Copula ($\theta=0.8461$) 概率密度图

Fig. 3 Probability density of Gaussian Copula ($\theta=0.8461$)

图 4 不同的是,图 3 只表示两随机变量之间的相关关系,略去了两随机变量之间数值大小的影响,通过图 3 能更清晰地了解二者的相关关系。图 4 则表示出两随机变量之间联合分布的图像,保留了样本数值大小的信息。通过二维频率直方图和联合分布的概率密

度图可以直观比较本文选取的二维 Copula 模型的拟合效果,如图 4 所示。由图 4 可以看出,对于风速-波高、波高-周期、风速-波向,其二维联合概率密度图与经验图像的峰值和形状都比较接近,说明相关性模拟较为合理,较好地考虑了风浪要素之间的相关性。同时,从图 4 也可以看到波高-波向的联合分布与经验直方图的峰值大小存在一定差异,这是由于该样本波高-波向之间的相关性比较复杂,采用单一的 Copula 函数模拟它们之间相关性的效果有限,在后续研究中,可以采用混合 Copula 函数来进行模拟,并优化拟合效果。由于不同海域的风浪要素之间的相关关系可能会有较大的差异,本研究侧重于探究建立风浪联合分布的方法和思路,因此未对波高-波向的二维联合概率模型作进一步优化。

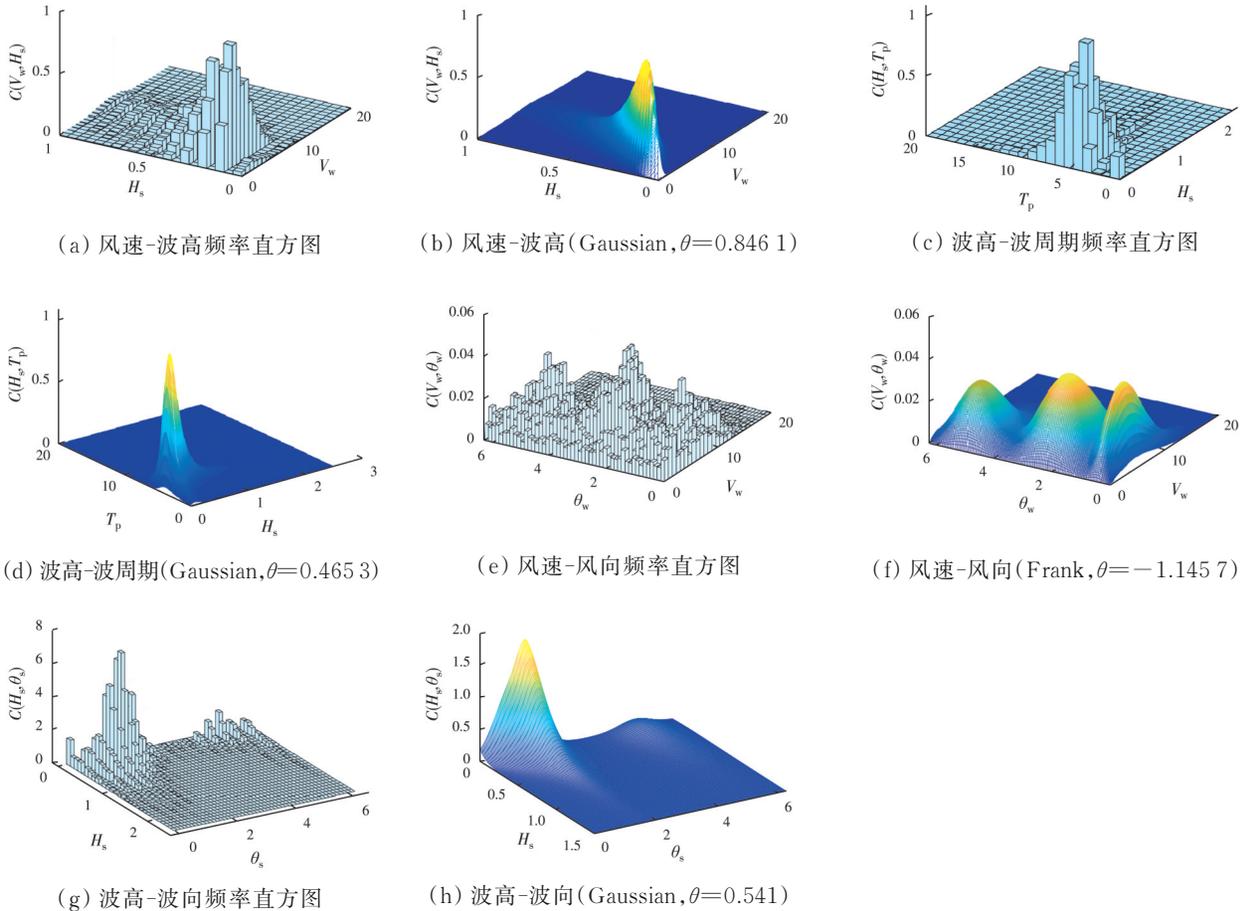


图 4 风浪要素两两变量之间的频率直方图及最优二维 Copula 函数

Fig. 4 Binary frequency histogram of wind and wave data and the corresponding optimal 2-dimensional Copula distribution

3 多维联合分布

可以采用多维 Copula 函数建立多维随机变量的联合分布,但可供选用的多维 Copula 函数类型有限,且灵活性较弱。Joe^[20]提出了采用 Vine Copula 来构建多维随机变量的联合分布的方法, Vine

Copula 具有结构多样灵活的优点。

3.1 Pair-Copula 理论

Bedford 等^[21]提出了通过 Pair-Copula 构建多维随机变量联合分布概率模型的方法,基于条件概率可以将多维随机变量分解成一系列的 Pair-Copula 结构,分解时采用的逻辑结构不同,可以构造出不

同的模型。Pair-Copula 结构推动了 Copula 理论在多维随机变量应用中的发展。

根据条件概率,多维联合分布概率密度函数可以表示为

$$f(x_1 \cdots x_n) = f_n(x_n) \cdot f(x_{n-1}|x_n) \cdot f(x_{n-2}|x_{n-1}, x_n) \cdots f(x_1|x_2 \cdots x_n) \quad (19)$$

由式(19)可以得到二维随机变量的联合概率密度函数为

$$f(x_a, x_j) = c_{aj}(F_a(x_a), F_j(x_j))f(x_a)f(x_j) \quad (20)$$

式中: $a, j=1, 2 \cdots n$, 且 $a \neq j$; $c_{aj}(F_a(x_a), F_j(x_j))$ 为 x_a 和 x_j 的二维 Copula 密度函数。

由式(20)可以推导 x_j 已知的情况下 x_a 的条件概率密度

$$f(x_a|x_j) = c_{aj}(F_a(x_a), F_j(x_j))f(x_a) \quad (21)$$

由式(21)可得在 n 维随机变量 u 已知的条件下,任意随机变量 x 的条件密度函数

$$f(x|u) = c_{xu|u-a}(F(x|u-a), F(u_a|u-a)) \cdot f(x|u-a) \quad (22)$$

式中: u_a 是 n 维随机变量 u 中的一个分量; u_{-a} 是 n 维随机变量 u 中去掉 u_a 之后的 $n-1$ 维分量。

以三维联合分布为例说明分解过程,根据式(19),三维的联合分布函数可写为

$$f(x_1, x_2, x_3) = f_1(x_3) \cdot f(x_2|x_3) \cdot f(x_1|x_2, x_3) \quad (23)$$

由式(22)可以得到

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot f(x_1|x_2) \quad (24)$$

将 $f(x_1|x_2) = c_{12}\{F_1(x_1), F_2(x_2)\}f_1(x_1)$ 带入式(24)可得

$$f(x_1|x_2, x_3) = c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{12}\{F(x_1), F(x_2)\} \cdot f_1(x_1) \quad (25)$$

将 $f(x_2|x_3) = c_{23}\{F_2(x_2), F_3(x_3)\}f_2(x_2)$ 和式(25)代入式(23)即可将多维联合分布分解为 Pair-Copula 结构和边缘分布的概率密度函数的乘积

$$f(x_1, x_2, x_3) = f_3(x_3) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2) \cdot c_{13|2}\{F(x_1|x_2), F(x_3|x_2)\} \cdot c_{12}\{F_1(x_1), F_2(x_1)\} \cdot f_1(x_1) \quad (26)$$

3.2 Vine 结构

高维 Copula 函数可以通过 R-Vine 结构来建立。基于两两随机变量之间的相依组合,结合条件概率可以建立多维 Copula 函数。R-Vine 中有两类特殊的 Vine: C-Vine 和 D-Vine, 这两类 Vine 有各自的逻辑结构,用于建立高维变量之间的联合分布。

C-Vine 结构的特点是每层树都有一个主节点,主节点连接其他所有节点。为方便阐述构造原理,图 5 给出了四维的 C-Vine 结构图,该 C-Vine 的结构有 3 棵树,每棵树有一个主节点,主节点和其他节点 (node) 相连形成边 (edge), 边在下一棵树中作为主

节点,以此类推,直到连接所有树的节点。对任意维 n 维 C-Vine, 有 $n-1$ 棵树, 有 $n(n-1)/2$ 条边, 每条边对应一个 Pair-Copula 结构。

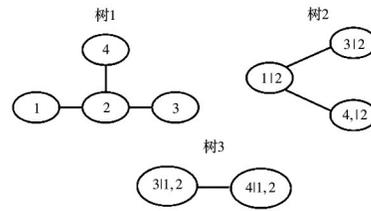


图 5 四维 C-Vine 分解结构
Fig. 5 Decomposition structure for four-dimensional C-Vine model

D-Vine 结构的特点是每层树的节点依次相连,呈直线状。图 6 给出了四维 D-Vine 的结构,该 D-Vine 结构有 3 棵树,每棵树中节点一次连接,形成边,边在下一棵树中成为节点,用同样的方式类推,直至连接所有节点。对任意维 n 维 D-Vine, 同样有 $n-1$ 棵树, 有 $n(n-1)/2$ 条边, 每条边对应一个 Pair-Copula 结构。

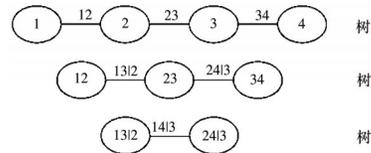


图 6 四维 D-Vine 分解结构
Fig. 6 Decomposition structure for four-dimensional D-Vine model

C-Vine 的多变量分解结构的确定需要先确定根节点和其他节点的顺序, C-Vine 的根节点通常选取与其他变量相关性最强的节点, 可以采用 Aas 等^[22]与 Dißmann 等^[23]提出的序贯估计法来确定。

1) 从第 1 层树开始, 计算所有随机变量两两组合的经验 Kendall 相关系数 τ , 如表 7 所示, 选取和其他随机变量的相关系数 τ 之和最大的随机变量作为

表 7 经验 Kendall's τ 矩阵及 Kendall's τ 之和

参数	风速	波高	波周期	风向	波向	τ_{sum}
风速	1	0.415 3	0.081 4	0.147 6	0.064 2	1.708 5
波高	0.415 3	1	0.037 5	0.274 5	0.162 5	1.889 8
波周期	0.081 4	0.037 5	1	0.008 6	0.001 6	1.129 1
风向	0.147 6	0.274 5	0.008 6	1	0.166 9	1.597 6
波向	0.064 2	0.162 5	0.001 6	0.166 9	1	1.973 0

根节点, 这样就确定了第 1 层树的结构。

2) 选择第 1 层树中二维随机变量的 Copula 函数种类并使用极大似然法估计参数 θ 。

3)结合第2)步确定的 Copula 函数及参数 θ , 通过式(25)将原随机变量转换为条件随机变量, 即树 2 的随机变量。

$$F(x|v) = \frac{\partial C_{xv|v_{-j}}(F(x|v_{-j}), F(v_j|v_{-j}))}{\partial F(v_j|v_{-j})} \quad (27)$$

式中: v_{-j} 为 v 中除去 v_j 的 $n-1$ 维向量, $F(x|v)$ 为条件分布函数, $C_{xv|v_{-j}}$ 为连接 $F(x|v_{-j})$ 与 $F(v_j|v_{-j})$ 的二维 Copula 函数。

4)使用确定树 1 结构的方法确定剩余的所有树。

采用 C-Vine Copula 构建多维联合分布, 图 7 给出了最优 C-Vine 的构造, 1~5 分别表示风速、波高、周期、风向、波向, 五维 C-Vine Copula 有 4 层树, 在第 1 层树中, 随机变量 2(波高)与其余 4 个变量分别通过二维 Copula 连接, 形成 4 条边; 在第 2 层树中, 这 4 条边作为节点, 选取一个主节点, 再一次通过二维 Copula 进行连接, 以此类推, 直至连接所有节点。

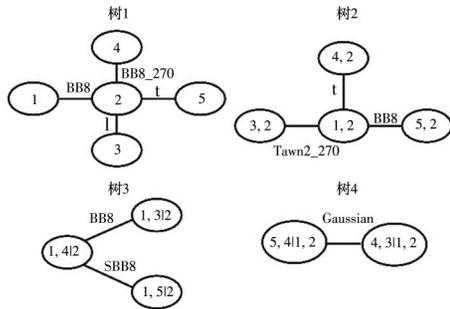


图 7 维 C-Vine Copula 结构

Fig. 7 The optimal five-dimensional C-Vine model

由式(10)和图 7 可得, 五维随机变量 $(x_1, x_2, x_3, x_4, x_5)$ 的 C-Vine 联合概率密度函数如式(28)所示。

$$f(x_1, x_2, x_3, x_4, x_5) = f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4)f_5(x_5)c_{12}(F_1(x_1), F_2(x_2)) \times c_{23}(F_2(x_2), F_3(x_3)) \times c_{24}(F_2(x_2), F_4(x_4))c_{25}(F_2(x_2), F_5(x_5)) \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) \times c_{14|2}(F_{1|2}(x_1|x_2), F_{4|2}(x_4|x_2)) \times c_{15|2}(F_{1|2}(x_1|x_2), F_{5|2}(x_5|x_2)) \cdot c_{34|12}(F_{3|12}(x_3|x_1,$$

$$x_2), F_{4|12}(x_4|x_1, x_2)) \times c_{45|12}(F_{4|12}(x_4|x_1, x_2), F_{5|12}(x_5|x_1, x_2)) \times c_{35|124}(F_{3|124}(x_3|x_1, x_2, x_4), F_{5|124}(x_5|x_1, x_2, x_4)) \quad (28)$$

式中: $F_i(\cdot)$ 为每个随机变量的累积分布函数; $F_{j|\cdot}(\cdot|\cdot)$ 为条件分布函数; $c_{\cdot}(\cdot, \cdot)$ 为 Copula 密度函数; $c_{j|\cdot}(\cdot|\cdot)$ 为 Copula 密度函数。

可以通过比较 AIC 值进行 C-Vine Copula 模型的拟合优度评价, AIC 的值越小说明拟合效果越好。一般来说, C-Vine 和 D-Vine 均可用于构建数据的联合分布, 具体采用哪一种模型应该取决于数据本身, 通常应通过试算来最终确定较优的模型。分别采用 C-Vine 和 D-Vine 来构建风浪联合分布模型, 并对比两种模型的 AIC 值, 发现 C-Vine 能更好地模拟该海洋站点的风浪联合分布模型。因此, 最终采用 C-Vine 来构建风浪的联合分布模型。此外, 采用序贯估计法得到风浪联合概率分布的 C-Vine 模型。通过序贯估计法得出的最优根节点顺序为 2、1、4、5、3(1~5 分别表示风速、波高、周期、风向、波向), AIC 值为 -27 508, 如表 8 所示。采用遍历所有根节点顺序的方法(即考虑了所有可能的根节点顺序), 对由序贯估计法得出的 C-Vine 风浪联合概率分布模型进行拟合优度评价。五维随机变量的根节点顺序共有 $A_5^3 = 60$ 种, 针对每一种根节点顺序, 首先采用 AIC 准则建立相应的 C-Vine 模型, 并通过 AIC 值的大小对 60 个 C-Vine 模型进行排序, 结果如表 9 所示。限于篇幅, 仅给出了前 20 个较优的 C-Vine 模型及相应的根节点顺序。通过比较发现, 通过序贯估计法建立的 C-Vine 模型的 AIC 值在 60 种情况中排第 3, 并且其 AIC 值与前面 2 个模型的 AIC 值非常接近(相差 0.31% 以内)。因此, 序贯估计法能高效且准确的找出较优的 C-Vine 联合概率模型, 该模型能很好地描述风浪要素间的相关关系。随着变量维度的提升, 根节点的组合会迅速增加, 采用遍历所有根节点顺序的方法会极大地增加计算负

表 8 最优 C-Vine 参数及 AIC 值

Table 8 The optimal C-Vine parameters and AIC values

树	连接边	Copula	参数 1	参数 2	AIC	总 AIC
T1	2, 1	BB8	2.07	0.99	-13 473	-27 508
	2, 3	t	0.05	3.54	-2 143	
	2, 4	rotated BB8 Copula (270 degrees)	-2.21	-0.83	-5 116	
	2, 5	t	-0.23	3.6	-2 570	
T2	1, 3 2	rotated Tawn type 2 Copula (270 degrees)	-1.45	0.37	-2 024	
	1, 4 2	t	0.04	9.06	-348	
	1, 5 2	BB8	1.18	0.92	-376	
T3	4, 5 2, 1	rotated BB8 Copula (180 degrees; "survival BB8")	1.48	0.85	-996	
	4, 3 2, 1	BB8	1.25	0.88	-436	
T4	3, 5 2, 1, 4	Gaussian	0.03		-22	

担,由此也体现出序贯估计法的优越性。

表 9 不同 C-Vine 模型的根节点顺序和 AIC 值

Table 9 The root node order and AIC value of different C-Vine models

根节点顺序	AIC	根节点顺序	AIC
2,1,5,4,3	-27 593.8	2,3,4,5,1	-27 338.8
2,1,3,5,4	-27 555.6	2,5,4,3,1	-27 244.4
2,1,4,5,3	-27 508.1	3,2,1,5,4	-26 133.6
2,5,1,4,3	-27 454.9	3,2,5,4,1	-26 128.8
2,4,1,3,5	-27 424.2	3,2,4,5,1	-26 080.5
2,4,3,5,1	-27 409.7	5,2,1,3,4	-26 048.4
2,3,1,5,4	-27 385.6	5,2,3,4,1	-26 029.7
2,5,3,4,1	-27 384.4	1,2,4,3,5	-26 007.7
2,3,5,4,1	-27 375.2	1,2,5,4,3	-25 977.3
2,4,5,3,1	-27 368.1	1,2,3,5,4	-25 945.7

通过建立的 C-Vine 模型仿真了风速、波高、波浪周期、风向和波向五维随机变量之间的累积概率密度(CDF)。为了更直观地观察拟合效果,以风速、波高、风向为例,结合式(14),图 8 给出了其三联合分布的累积概率密度,黑色网格是原始数据的经验 Copula 图像,彩色图像是由 C-Vine 模拟产生的图像。由图 8 可知,建立的 C-Vine 模型能较好地模拟风速、波高、风向的相关关系。

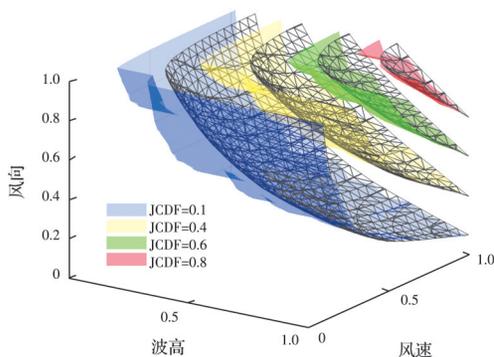


图 8 三维累积概率密度图

Fig. 8 Diagram of three-dimensional cumulative probability density

4 结论

基于 Vine Copula 函数研究了我国东海连云港海洋观测站的风浪要素之间的联合概率分布,得出以下结论:

1) 风速、波高、波周期的概率分布为单峰分布,最优拟合分布分别为 Weibull 分布、广义极值分布、 t 分布;而风向和波向的概率分布为多峰分布,最优拟合分布均为混合 Gaussian 分布。

2) 风浪要素中两两随机变量之间的联合概率分布研究表明,风速-波高、波高-波周期、风速-风向、波高-波向 4 个二维变量对的最优二维联合概率分布分别为 Gaussian Copula、Gaussian Copula、Frank Copula 和 Gaussian Copula。

3) 建立的 C-Vine 模型可以较好地刻画风速、波高、波浪周期、风向和波向五维随机变量之间的联合概率分布。

4) 采用 Vine Copula 函数建立了东海连云港海洋观测站风浪要素之间的联合概率分布模型,对于中国其他海域海洋观测站风浪要素之间的联合分布规律还有待进一步研究。但该研究方法和思路可为中国其他海域海洋观测站风浪要素之间的联合分布研究提供借鉴。

致谢:感谢中国国家科技资源共享服务平台

——国家海洋科学数据中心(<http://mds.nmdis.org.cn/>)提供数据支撑。

参考文献

- [1] LI X, ZHANG W. Long-term assessment of a floating offshore wind turbine under environmental conditions with multivariate dependence structures [J]. *Renewable Energy*, 2020, 147: 764-775.
- [2] ZHANG W M, WANG Z W, LIU Z. Joint distribution of wind speed, wind direction, and air temperature actions on long-span bridges derived via trivariate metaelliptical and plackett copulas [J]. *Journal of Bridge Engineering*, 2020, 25(9): 04020069.
- [3] BAI X Y, JIANG H, LI C, et al. Joint probability distribution of coastal winds and waves using a logtransformed kernel density estimation and mixed copula approach [J]. *Ocean Engineering*, 2020, 216: 107937.
- [4] WANG Z W, ZHANG W M, TIAN G M, et al. Joint values determination of wind and temperature actions on long-span bridges: Copula-based analysis using longterm meteorological data [J]. *Engineering Structures*, 2020, 219: 110866.
- [5] ZHANG Y, KIM C W, BEER M, et al. Modeling multivariate ocean data using asymmetric copulas [J]. *Coastal Engineering*, 2018, 135: 91-111.
- [6] LI X, ZHANG W. Probabilistic fatigue evaluation of floating wind turbine using combination of surrogate model and copula model [C]//AIAA Scitech 2019 Forum. San Diego, California. Reston, Virginia: AIAA, 2019.
- [7] VANEM E. Joint statistical models for significant wave height and wave period in a changing climate [J]. *Marine*

- Structures, 2016, 49: 180-205.
- [8] YANG X, QIAN J. Joint occurrence probability analysis of typhoon-induced storm surges and rainstorms using trivariate Archimedean copulas [J]. Ocean Engineering, 2019, 171: 533-539.
- [9] WANG J G, LI L, JAKOBSEN J B, et al. Metocean conditions in a Norwegian fjord in relation to the floating bridge design [J]. Journal of Offshore Mechanics and Arctic Engineering, 2019, 141(2): 021604.
- [10] ZHANG Y, BEER M, QUEK S T. Long-term performance assessment and design of offshore structures [J]. Computers & Structures, 2015, 154: 101-115.
- [11] XU X, LAI F, LI G J, et al. A novel vibration suppression device for floating offshore wind generator [C]//Proceedings of ASME 2018 12th International Conference on Energy Sustainability Collocated With the ASME 2018 Power Conference and the ASME 2018 Nuclear Forum. Lake Buena Vista, Florida, USA: 2018.
- [12] 陈子燊, 路剑飞, 于吉涛. 基于非对称 Archimedean Copula 的三变量风浪重现水平分析[J]. 海洋通报, 2017, 36(6): 631-637.
CHEN Z S, LU J F, YU J T. Analysis on return levels of trivariate stormy waves based on asymmetric Archimedean Copula function [J]. Marine Science Bulletin, 2017, 36(6): 631-637. (in Chinese)
- [13] 陈建兵, 陶金聚, 任晓丹, 等. 基于 Copula 理论的混凝土受压本构全曲线参数相关性研究[J]. 土木工程学报, 2020, 53(7): 52-63.
CHEN J B, TAO J J, REN X D, et al. Research on the dependence between parameters in complete stress-strain constitutive curve of concrete based on copula theory [J]. China Civil Engineering Journal, 2020, 53(7): 52-63. (in Chinese)
- [14] 樊学平, 杨光红, 肖青凯, 等. 大跨桥梁主梁失效概率分析的最优 R-Vine Copula [J]. 吉林大学学报(工学版), 2021, 51(4): 1296-1305.
FAN X P, YANG G H, XIAO Q K, et al. Optimal R-Vine Copula information fusion for failure probability analysis of long-span bridge girder [J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(4): 1296-1305. (in Chinese)
- [15] 刘月飞, 樊学平. 失效非线性相关的桥梁截面可靠性 Vine-Copula 数据融合[J]. 同济大学学报(自然科学版), 2019, 47(3): 315-321.
LIU Y F, FAN X P. Data fusion about Vine-Copula for bridge section reliability considering nonlinear correlation of failure modes [J]. Journal of Tongji University (Natural Science), 2019, 47(3): 315-321. (in Chinese)
- [16] 宋帅, 钱永久, 吴刚. 基于多元 Copula 函数的桥梁体系地震易损性分析方法研究[J]. 振动与冲击, 2017, 36(9): 122-129, 208.
SONG S, QIAN Y J, WU G. Seismic fragility analysis of a bridge system based on multivariate Copula function [J]. Journal of Vibration and Shock, 2017, 36(9): 122-129, 208. (in Chinese)
- [17] SKLAR A. Fonctions de repartition a n dimensions et leurs marges [J]. Publications de l'Institut de statistique de l'Université de Paris, 1959, 8: 229-231.
- [18] CARNICERO J A, AUSÍN M C, WIPER M P. Non-parametric copulas for circular-linear and circular-circular data: An application to wind directions [J]. Stochastic Environmental Research and Risk Assessment, 2013, 27(8): 1991-2002.
- [19] SADEGH M, RAGNO E, AGHAKOUCHAK A. Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework [J]. Water Resources Research, 2017, 53(6): 5166-5183.
- [20] JOE H. Families of m -variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters [M]//Institute of Mathematical Statistics Lecture Notes-Monograph Series. Hayward, CA: Institute of Mathematical Statistics, 1996: 120-141.
- [21] BEDFORD T, COOKE R M. Probability density decomposition for conditionally dependent random variables modeled by vines [J]. Annals of Mathematics and Artificial Intelligence, 2001, 32(1/2/3/4): 245-268.
- [22] AAS K, CZADO C, FRIGESSI A, et al. Pair-copula constructions of multiple dependence [J]. Insurance: Mathematics and Economics, 2009, 44(2): 182-198.
- [23] DIBMANN J, BRECHMANN E C, CZADO C, et al. Selecting and estimating regular vine copulae and application to financial returns [J]. Computational Statistics & Data Analysis, 2013, 59: 52-69.

(编辑 胡玲)