The 2016 World Congress on
Advances in Civil, Environmental, and Materials Research (ACEM16)
Jeju Island, Korea, August 28-September 1, 2016

Copula Model for Multi-Dimensional Extreme Wind Speed Analysis

*Jingcheng Wang[1)], Yong Quan[2)], Ming Gu[3)], Peng Huang[4)] and Xuanyi Zhou[5)]
[1), 2), 3), 4), 5)] *State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, Shanghai 200092, China.*
[1)] *wangjingchengxx@163.com*

ABSTRACT

Some previous researchers built the multi-dimensional extreme wind speed probability distribution functionswith the copula theory. Their methods are based on Gaussian-copula model or fully-nested Gumbel-copula model for 16 dimensions.These copula models were checked with the wind speed data from meteorological station and a parametric bootstrap test based on the empirical copulain present paper. The fully-nested Gumbel-copula is rejected in some cases when its *2*nd deepest nested variates were checked by this bootstrap test.The reason about this phenomenon was discussed and then t-copula model was proposed.

1. INTRODUCTION

The importance of considering the effect of wind directionality on probabilistical estimationof wind load effects of structures has beenwell recognized(Zhang and Chen,2015). Some previous researchers (Simiuet al.,1985; Kanda and Itoi,2001; Itoi and Kanda,2002; Zhang and Chen,2015,2016) analyzed the correlation of extreme wind speed between different sectors with copula theory, in which the joint distribution model of multi-directional extreme wind speed estimation problem is regarded as the probability distribution of multi-dimensional random process.The problem then can be decomposed into two aspects base on Sklar theorem(Sklar，1959). First aspect is the estimation problem of one-dimensional marginal cumulative distribution functionin each sector. Second aspect is the estimate problem of copula model to consider the correlation of extreme wind speeds between different sectors. The problem of second aspect was discussed in present paper.

In research on the correlation of extreme wind speeds between different sectors, the copula models used by previous researchers are two-dimensional Gumbel-copula(Simiuet al.,1985), Partially Nested Gumbel-copula(Kanda and Itoi,2001), Fully Nested Gumbel-copula(Itoi and Kanda,2002) and Gaussian-copula(Zhang and Chen,2015). Nikoloulopouloset et al.(2009) indicated that the t-copulas are generally

---

[1)] Graduate Student
[2), 3), 4), 5)] Professor

superior to the Gaussian-copulas in the context of modeling multivariate financial return data. t-copula model perform well in multi-dimensional data analysis, therefore t-copula model was proposed to use to consider the correlation of extreme wind speeds between different sectors in present paper.

Because different copula model has different property, whether the copula model, which is fitted with observed data, can reflect the property of the observed data should be checked. The checked result also can be used to judged which model is fitted better with observed data. In the research area of probabilistic wind load, Gaussian-copula and Gumbel-copula was expanded to 16-dimensional by Zhang and Chen(2016) for current 16 wind directions system. Therefore Gaussian-copula, Gumbel-copula and t-copula were chosen to execute the hypothesis checking.

2. PARAMETIC BOOTSTRAP TEST BASED ON THE EMPIRICAL COPULA

2.1 fitting method of copula models

The recursion formula of the cumulative distribution function of the M-dimensional Fully Nested Gumbel-Copula is flowings:

$$C_{Gumbel}^{m}(p_1,p_2,...,p_m)=\begin{cases} e^{(-((-\ln(p_1))^{\alpha}+(-\ln(p_2))^{\alpha})^{1/\alpha})} & for\ \ m=2 \\ C_{Gumbel}^{2}(\mathrm{C}_{Gumbel}^{m-1}(p_1,p_2,...,p_{m-1}),p_m) & for\ 3\le m\le M \end{cases} \quad (1)$$

where M is the number of the wind directions, which is set as 16 generally. p_i is the one-dimensional marginal cumulative distribution function in each sector.

The cumulative distribution function of the M-dimensional Gaussian-Copula is:

$$C_{Gaussian}(p_1,p_2,...,p_M)=G_{[\rho]}(\Phi^{-1}(p_1),\Phi^{-1}(p_2),...,\Phi^{-1}(p_M)) \quad (2)$$

where Φ^{-1} is the inverse function of the one-dimensional standard normal distribution, $G_{[\rho]}$ is M-dimensional normal distribution with mean 0 and covariance matrix $[\rho]$.

The cumulative distribution function of the M-dimensional t-Copula is:

$$C_t(\mathrm{p}_1,\mathrm{p}_2,\cdots,\mathrm{p}_M)=\mathrm{t}_{[\rho],k}(\mathrm{t}_k^{-1}(p_1),\mathrm{t}_k^{-1}(p_2),\cdots,\mathrm{t}_k^{-1}(p_M)) \quad (3)$$

where t_k^{-1} is the inverse function of the one-dimensional standard t distribution, $\mathrm{t}_{[\rho],k}$ is M-dimensional t distribution with the non-centrality parameter 0, correlation matrix $[\rho]$ and k degrees of freedom.

Although the formula of copula model is complex, many matical softwares (such as MATLAB) have the built-in functions to estimate the parameters of each copula model conveniently.

There are several method to fitting copula model(王丽芳, 2012), maximum likelihood method(MLE), the method of inference functions for margins(IFM)(Xu, 1996), pseudo maximum likelihood estimator(PMLE)（Kim et al.,2007). The research result of Kim et al.(2007) indicated that PMLE is better than IFM and MLE in most cases. Therefore PMLE was proposed to estimate the parameters of each copula model, i.e.firstly, the no-exceedanced probabilities $[p]$ are estimated by observed extreme wind speeds $[V]$ based on empirical distribution in each wind direction, and then the parameters of copula function are estimated by maximum likelihood method.

2.2introduction of empirical copula distribution(Genest et al., 2009)

Marginal cumulative distribution probability values $[p]$ are estimated by observed extreme wind speeds $[V]$ in each wind direction, where p_{ji} is the no-exceedanced probability of the j^{th} observed extreme wind speed in the i^{th} wind direction. M is set as the number of wind directions and N is the number of observed sample points, the cumulative distribution function of empirical copula at any one point $\vec{p}=(p_1,...,p_M)$ is:

$$C_0(\vec{p})=\frac{1}{N}\sum_{j=1}^{N}I(p_{j1}\le p_1,...,p_{jM}\le p_M) \qquad (4)$$

Where $p_1,...,p_M\in[0,1]$, $\mathrm{I}(\bullet)$ are indicative functions, i.e.if $p_{j1}\le p_1,...,p_{jM}\le p_M$, $\mathrm{I}(\bullet)=1$; else $\mathrm{I}(\bullet)=0$.

2.3step of the test

Genest et al. (2009)summarized the testing methods of copula models completely and compared these methods with each other. The parametric bootstrap test based on the empirical copula has less intricate form and is more easy to understand. It also has good effect in application, so it was proposed to test t-copula, Gaussian-copula and Fully Nested Gumbel-copulain in present paper. Cramer–von Mises statistics S_n was selected as inspected value.

$[V]_{N\times M}$ are the observed n-day maxima of wind speeds in all wind directions, where M is the number of the wind directions, N is the number of observed sample points, the step of the parametric bootstrap test based on the empirical copula is flowings (Genest et al., 2009; Genest and Rémillard, 2008):

1） Let $i=0$ and then the no-exceedanced probabilities $[p_i]_{N\times M}$ of all extreme wind speeds are estimated with $[V]_{N\times M}$.

2） Calculate the empirical joint cumulative distribution function values $C_{0,i}(\vec{p})$ with $[p_i]_{N\times M}$ based on the cumulative distribution function of empirical copula(Eq. 4).

3） Fit copula model (t-copula model, Gaussian-copula model or Fully Nested Gumbel-copula model) with $[p_i]_{N\times M}$.

4） Calculate Cramer–von Mises statistics $S_{n,i}=\sum_{j=1}^{N}(C_{0,i}(\vec{\mathrm{p}}_j)-C_{\theta,i}(\vec{\mathrm{p}}_j))^2$. Where $\vec{p}_j$ is the j^{th} row of $[p_i]_{N\times M}$.

5） Generate N random M-dimensional number using function $C_{\theta,i}(\vec{p})$.Let $i=i+1$ and name the generated random numbers $[p_i]_{N\times M}$.The number of generated samples N should be equal to the number of observed sample points. Return to step 2) and repeat this circulationfrom step 2) to 5)K(K usually is several hundreds or one thousand) times and record $S_{n,i}$ for every step.

6） Calculate the P-value approximately: $P_{value}\approx\frac{1}{K}\sum_{i=1}^{K}\mathrm{I}(\mathrm{S}_{n,i}>\mathrm{S}_{n,0})$.

7） If $P_{value} \le P_0$, reject the Original hypothesis, where P_0 is significance level, which is set as 0.05 usually.

When the correlation between variates is weak, the probability of rejecting wrong Original hypothesis is very low by the bootstrap test(Genest et al.，2009). In the t-copula and Gaussian-copula model, it can be proved that the parameter ρ_{ij} in the ith row- jth column of coefficient matrix $[\rho]$ is equal to the parameter ρ'_{ij} of the $ith-jth$ two-dimensional marginal cumulative distribution function. Therefore, when the t-copula and Gaussian-copula model are tested, just all of the two-dimensional marginal cumulative distributions of the adjacent wind direction should be checked (the correlation of wind speed is stronger). When the Fully Nested Gumbel-copula model is checked, the model should be checked according to the nested sequence.

3. CONCLUSIONS

The directional Surface Hourly observation data of wind speed used in this study was recorded by fixed-weather-station 034820 at Marham, United Kingdomfrom gotten from the database of National Oceanic and Atmospheric Administration(NOAA). The wind speed data from January 1st, 1973 to July 31st, 2015 are selected as observed samples to test copula models. The unit of wind speed is m/s and the resolution of wind speed is 0.1. The resolution of wind direction is 1° and these directional wind speeds are partitioned into 16 directional sectors. The monthly maximum of wind speed data for each month at each directional sector(as $[V]_{N\times M}$) were used to fit the copula model(Zhang and Chen, 2015,2016).

The Gumbel-Copula model is nested in the order of dominance direction(Zhang and Chen, 2016; Itoi and Kanda, 2002). When significance level P_0 is set as 0.05，in present test, the Fully Nested Gumbel-Copula model was rejected at the *2*nd deepest of the nested structure.

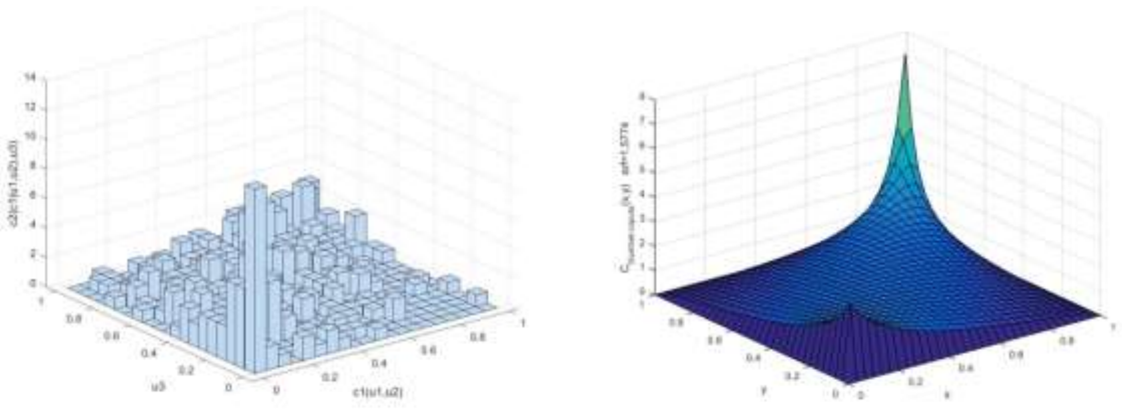

Fig. 1Frequency histogram of the *2*nddeepest nested variates(left) and Fitted probability densityof the *2*nddeepest nested variates(right)

In order to discuss the reason of this phenomenon, Fig. 1(left) shows the frequency histogram of variables in the second deepest of the Fully Nested Gumbel-Copula model.It can be treated as the probability density graphic of empirical copula distribution. Fig. 1(right) shows the probability density graphic of the Gumbel-copula which is fitted by variables in the second deepest of the nested structure. Obviously, the trend of the two figures is different, Fig. 1(left) has obvious lower tail, but in Fig. 1(right), the upper tail of the probability density graphic of the fitted Gumbel-Copula is very tall, it is much higher than the lower tail. The Fig. 1(left) and Fig. 1(right) cannot match well, it is induced that the Fully Nested Gumbel-Copulamodel is rejected at the *2*nd deepest of the nested structure.

Set the significance level $P_0 = 0.05$. When the t-copula and Gaussian-copula model were tested, all of the two-dimensional marginal cumulative distributions of the adjacent wind directions passed the parametric bootstrap test based on the empirical copula. But the experiment result of Genest et al. (2009) showed that the efficacy of all the test methods are limited. When the number of sample points is small, the test cannot reject wrong original hypothesis or accept right original hypothesis strictly. In the experiment of Genest et al. (2009), 150 samples were generated to test, the K in step 6) is 1000, the parametric bootstrap test based on the empirical copula was repeated 10000 times. The probability of reject the right original hypothesis(when the samples is generated from the same distribution of the original hypothesis) is no greater than 5% when the t-copula and Gaussian-copula were tested.

Because the experiment of Genest et al. (2009) is time-consuming, a simple experiment was made to test the model more strictly by compared with the experiment result of Genest et al. (2009). The step of the simple experiment is similar to Genest et al. (2009), but the times of repetition is reduced to 100 times, the K in step 6) is reduced to 100 in each time. If the probability of $P_{value} \le P_0$ is larger (greater than 5%), the original hypothesisis rejected.

Table1 when the original hypothesis is t-copula or Gaussian-copula,
the probability of reject the original hypothesis in the simple experiment.

风向	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
风向	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1
t-copula	0.04	0.02	0.04	0.05	0.05	0.02	0.01	0.03	0.03	0.02	0.02	0.02	0.05	0.03	0.03	0.02
Gaussian-copula	0.07	0.07	0.06	0.03	0.05	0.04	0.06	0.03	0.02	0.04	0.01	0.04	0.01	0.04	0.03	0.08

When the original hypothesis is t-copula, the probability of reject the original hypothesis is no greater than 5% in every adjacent wind directions, then the original hypothesisis accepted. When the original hypothesis is Gaussian-copula, the probability of rejecting the original hypothesis is greater than 5% in some adjacent wind directions. In the results of simple experiment, t-copula is much harder to be rejected and fitted better with the observed wind speed data than Gaussian-copula.

REFERENCES

Coles, S. G., & Walshaw, D. (1994), “Directional modelling of extreme wind speeds.” Applied Statistics, 139-157.
Genest, C., Rémillard, B., & Beaudoin, D. (2009), “Goodness-of-fit tests for copulas: A review and a power study”. Insurance: Mathematics and economics, 44(2), 199-213.
Itoi, T., & Kanda, J. (2002), “Comparison of correlated Gumbel probability models for directional maximum wind speeds.” Journal of Wind Engineering and Industrial Aerodynamics, 90(12), 1631-1644.
Kanda, J., & Itoi, T. (2001), “Correlated Gumbel probability model for directional wind speeds.” Structural Safety and Reliability: ICOSSAR'01, 2001.
Kim, G., Silvapulle, M. J., & Silvapulle, P. (2007), “Comparison of semiparametric and parametric methods for estimating copulas.” Computational Statistics & Data Analysis, 51(6), 2836-2850.
Nelsen,R.(2006), An Introduction to Copulas,2nded.,Springer,NewYork.
Nikoloulopoulos, A. K., Joe, H., & Li, H. (2009), “Extreme value properties of multivariate t copulas.” Extremes, 12(2), 129-148.
Simiu, E., Hendrickson, E. M., Nolan, W. A., Olkin, I., & Spiegelman, C. H. (1985), “Multivariate distributions of directional wind speeds.” Journal of Structural Engineering, 111(4), 939-943.
Sklar, M. (1959), Fonctions de répartition à n dimensions et leurs marges. Université Paris 8.
Xu, J. J. (1996), Statistical modelling and inference for multivariate and longitudinal discrete response data (Doctoral dissertation, University of British Columbia).
Zhang, X., & Chen, X. (2015), “Assessing probabilistic wind load effects via a multivariate extreme wind speed model: A unified framework to consider directionality and uncertainty.” Journal of Wind Engineering and Industrial Aerodynamics, 147, 30-42.
Zhang, X., & Chen, X. (2016), “Influence of dependence of directional extreme wind speeds on wind load effects with various mean recurrence intervals.” Journal of Wind Engineering and Industrial Aerodynamics, 148, 45-56.
王丽芳. (2012), Copula 分布估计算法. 机械工业出版社