ARTICLES

Chinese Science Bulletin 2006 Vol. 51 No. 23 2926—2930

DOI: 10.1007/s11434-006-2186-x

Theory of multivariate compound extreme value distribution and its application to extreme sea state prediction

LIU Defu[1], WANG Liping[2] & PANG Liang[3]

1. Disaster Prevention Research Institute, Ocean University of China, Qingdao 266003, China;
2. Department of Mathematics, Ocean University of China, Qingdao 266071, China;
3. College of Engineering, Ocean University of China, Qingdao 266071, China

Correspondence should be addressed to Liu Defu (email:liu@ouc.edu.cn)

Received October 18, 2005; accepted February 21, 2006

Abstract In this paper, a new type of distribution, multivariate compound extreme value distribution (MCEVD), is introduced by compounding a discrete distribution with a multivariate continuous distribution of extreme sea events. In its engineering application the number over certain threshold level per year is fitting to Poisson distribution and the corresponding extreme sea events are fitting to Nested Logistic distribution, then the Poisson-Nested logistic trivariate compound extreme value distribution (PNLTCED) is proposed to predict extreme wave heights, periods and wind speeds in Yellow Sea. The new model gives more stable and reasonable predicted results.

Keywords: multivariate compound extreme value distribution, Nested-logistic model, extreme sea state, threshold value.

Compound extreme value distribution (CEVD) was derived in 1980 as Poisson-Gumbel CEVD for typhoon-induced extreme events in South China Sea[1] and Poisson-Weibull model for hurricane characteristics along Atlantic coasts and the Gulf of Mexico in 1982[2]. After their publication, the CEVD has aroused great interests of scientists and engineers[3−7]. According to the statistics for about forty coastal structures, the CEVD has been successfully used to predict design wave height. Especially, the catastrophe in New Orleans induced by Hurricane Katrina in 2005 shows that the design basis of New Orleans protection structures is 'Standard Project Hurricane' proposed by NOAA[8], which only corresponded to 38 years return period of hurricane central pressure and Hurricane Katrina corresponded to 60 years return period predicted by Poisson-Weibull model[2].

Bivariate compound extreme value distribution was derived in different forms and applied to some engineering areas, such as platform deck clearance design, disaster prevention design criteria for estuarine city Shanghai and coastal city Qingdao, etc.[9−13].

In this paper, the Poisson-nested logistic trivariate compound extreme value distribution (PNLTCED) is derived by compounding the discrete distribution of the number of data sampling over certain threshold level per year (Poisson distribution) into the multivariate continuous distribution (Nested logistic trivariate distribution). The difference of this model from our previous study lies in the data sampling of discrete distribution. In this paper, the number of data sampling over threshold level is taken account in the CEVD instead of typhoon occurrence frequency. Comparison between the predicted results based on the long term observed data and short term data shows its stability in prediction.

1 Fundamental theory of the multivariate compound extreme value distribution

Let N be a random variable (the number of storms in a given year), with their corresponding probability $P\{N=k\}=p_k$, $k=1,2,\cdots$; let $(\xi_{11},...,\xi_{n1})$, $(\xi_{12},...,\xi_{n2})$... be an independent sequence identically distributed random vectors (the observed extreme sea environments in the sense defined above within the successive storms) with common density $g(\cdot)$. Then we are interested in the distribution of

$$(X_1,\ldots,X_n)=(\xi_{1i},\ldots,\xi_{ni}),$$

where ξ_{1i} is the maximum value of ξ_{1j}, $1\leqslant j\leqslant N$, $N=1,2,...$

It represents the maximum annual value of the principal variable, together with the simultaneously occurring values of the concomitant variables. There is a reasonable approximation in definition of $(X_1,\ldots,X_n)$, the case of N=0 should be neglected, because no extreme value of interest can occur outside the storm when N=0. A more detailed discussion of the model correction in case of $p(N=0)$ can be found in ref. [1].

When multivariate continuous cumulative distribution is $G(x_1,\ldots,x_n)$, we can derive the MCEVD as

$$F(x_1,\ldots,x_n)=\sum_{i=1}^{\infty}p_i\cdot i\cdot\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1}G_1^{i-1}(u)\cdot g(u_1,\ldots,u_n)\,du_1\ldots du_n, \qquad (1)$$

where $G_1(u_1)$ is marginal distribution of $G(x_1,\ldots,x_n)$, and $g(u_1,\ldots u_n)$ is density function.

This can be proved as follows:

$$\begin{aligned}F(x_1,\ldots,x_n)&=P(X_1<x_1,\ldots,X_n<x_n)\\&=P(\bigcup_{i=1}^{\infty}\{X_1<x_1,\ldots,X_n<x_n\}\cap\{N=i\})\\&=\sum_{i=1}^{\infty}p_iP(X_1<x_1,\ldots,X_n<x_n\mid N=i)\\&=\sum_{i=1}^{\infty}p_iP(\bigcup_{k=1}^{i}\{X_1<x_1,\ldots,X_n<x_n\}\\&\quad\cap\{\underset{1\leqslant j\leqslant i}{Max}\,\xi_{1j}=\xi_{1k}\}\mid N=i)\\&=\sum_{i=1}^{\infty}p_i.i.P(\{X_1<x_1,\ldots,X_n<x_n\}\\&\quad\cap\{\underset{1\leqslant j\leqslant i}{Max}\,\xi_{1j}=\xi_{11}\}\mid N=i)\\&=\sum_{i=1}^{\infty}p_i.i\cdot P(\xi_{11}<x_1,\ldots,\xi_{n1}<x_n,\xi_{11}>\xi_{1j},\\&\quad j=2,3,\cdots i\mid N=i)\\&=\sum_{i=1}^{\infty}p_i\cdot i\cdot\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_1}G_1^{i-1}(u)\\&\quad\cdot g(u_1,\ldots,u_n)\,du_1\ldots du_n.\end{aligned}$$

Therefore, eq. (1) is proved.

As mentioned above, the frequency of typhoon (hurricane, winter storm) occurrence can be fitted to Poisson distribution

$$P_i=\frac{e^{-\lambda}\lambda^i}{i!}.$$

Then Poisson-Nested logistic trivariate compound extreme distribution can be obtained from formula (1):

$$F_0(x_1,x_2,x_3)=e^{-\lambda}(1+\lambda\int_{-\infty}^{x_3}\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}e^{\lambda\cdot G_1(u_1)}\cdot g(u_1,u_2,u_3)du_1\,du_2du_3. \qquad (2)$$

Nested-Logistic trivariate distribution is expressed as

$$G(x_1,x_2,x_3)=\exp\left[-\left\{\left[\left(1+\xi_1\frac{x_1-\mu_1}{\sigma_1}\right)^{-1/(\alpha\beta\xi_1)}+\left(1+\xi_2\frac{x_2-\mu_2}{\sigma_2}\right)^{-1/(\alpha\beta\xi_2)}\right]^{\beta}+\left(1+\xi_3\frac{x_3-\mu_3}{\sigma_3}\right)^{-1/(\alpha\xi_3)}\right\}^{\alpha}\right], \qquad (3)$$

in which ξ_j, μ_j, σ_j are the shape, location and scale parameters of marginal distributions $G(x_j)$ to x_j (j = 1, 2, 3), respectively. And dependent parameters α, β can be obtained through moment estimation[14]

$$\hat{\alpha}=\frac{\sqrt{1-r_{13}}+\sqrt{1-r_{23}}}{2},$$

$$\hat{\beta}=\frac{\sqrt{1-r_{12}}}{\hat{\alpha}},$$

where $r_{i,j}$ is correlation coefficient, $I<j$, i,j = 1, 2, 3.

Trivariate layer structure (α, outside layer; β, inside layer) shows that the correlation between x_1 and x_2 is stronger than those among x_1, x_3 and x_2, x_3.

As shown above, the PNLTCED can be obtained through estimation of parameters of marginal distributions and their dependent parameters.

2 Application of the PNLTCED to predictions of extreme sea states at Yellow Sea

2.1 Data sampling, marginal distribution and parameter estimation

The observed wind speed, wave height and period data at a certain oceanologic station at Yellow Sea in a total of 26 years from 1963 to 1988 are classified into two groups: Af data (1963—1988) and Bf data (1973—1988). By checking residual life plot[15−18] it is shown that the smoother part of wind data in residual life plot gives a threshold level of 21 m/s. Then the data series of wind speed, wave height and period dominated by wind can be obtained. The probability plot, return level plot, quantile plot and density plot are shown in Fig. 1. The confidence probability is 95% for all data. So all data can be used for marginal distribution analysis and correlation analysis for groups Af and Bf (Table 1). The linear correlation analysis gives corresponding correlation coefficients and correlation parameters $\hat{\alpha}$, $\hat{\beta}$ for inside layer and outside layer (Table 2).

Table 2 shows that there is a stronger correlation between wave height and wind speed than those between others, so the wave height (x_1) and wind speed (x_2) can be taken as inside layer variables, and the wave period (x_3) as outside layer variable.

Instead of the BCEVD, in this paper the daily extreme sea state data over the certain threshold value are used as the sample of discrete random variable. Checking *K*-*S* with the significance level of 0.05, we see that the data fit to Poisson distribution very well (Table 3).

2.2 *The physical characteristics of the PNLTCED model*

The tridimensional perspective projection clearly shows some characteristics of the PNLTCED model (Figs. 2 and 3). The new model considers the most disadvantageous combination of three kinds of environment loads. The joint return periods of different combinations of wind speed and wave period with some return periods are listed in Tables 4 and 5. It can be seen that the joint return periods calculated using short term data (Bf) are close to those calculated using long term data (Af). And the results are stable and reason-

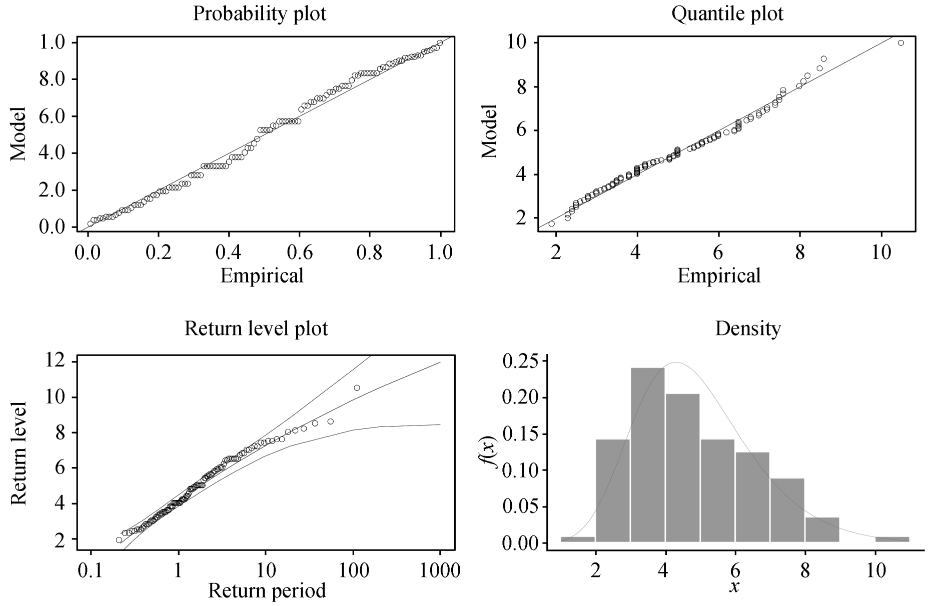

Fig. 1. Distribution diagnostic testing of wave height (group Af).

Table 1 Parameters of marginal distribution (groups Af and Bf)

Variables		Wave height		Wind speed		Wave period	
		Af	Bf	Af	Bf	Af	Bf
Parameters	location parameter μ	4.17	4.07	22.0	22.1	5.99	6.27
	standard deviation of μ	0.16	0.23	0.18	0.20	0.14	0.18
	scale parameter σ	1.48	1.59	1.28	1.32	1.39	1.35
	standard deviation of σ	0.12	0.17	0.16	0.17	0.09	0.12
	shape parameter ξ	–0.08	–0.06	0.38	0.29	–0.25	–0.35
	standard deviation of ξ	0.09	0.12	0.18	0.16	0.05	0.06

Table 2 Linear correlation coefficient and dependent parameter

Data group	x_1 vs. x_2 r_{12}	x_1 vs. x_3 r_{13}	x_2 vs. x_3 r_{23}	$\hat{\alpha}$	$\hat{\beta}$
Af	0.84	0.66	0.80	0.51	0.77
Bf	0.82	0.62	0.78	0.54	0.79

Table 3 The parameters of over threshold data for groups Af and Bf

Over threshold numbers per year	0	1	2	3	4	5	6	7	8	9	10	11	Total data numbers/total years	Mean value λ
Years (Af)	1	1	4	4	5	4	3	2	1	0	0	1	112/26	4.308
Years (Bf)	1	1	2	1	3	4	2	3	0	0	0	0	66/16	4.125

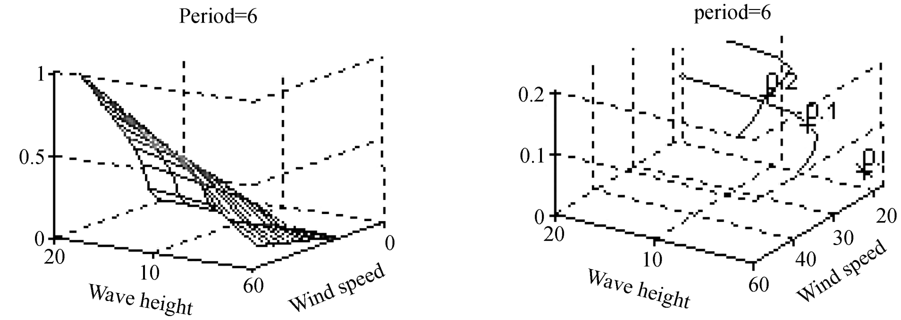

Fig. 2. Joint probability distribution function and the contour line.

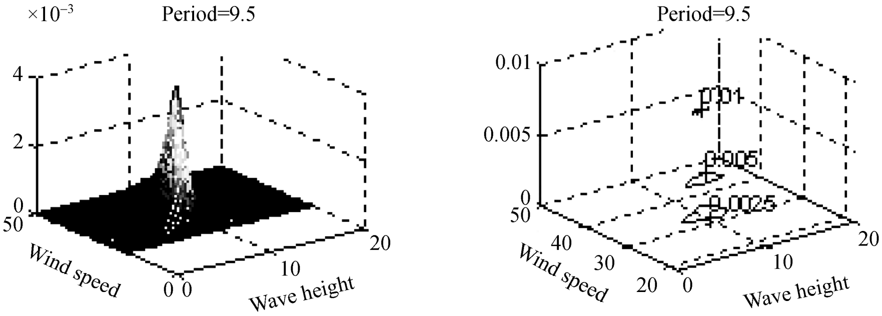

Fig. 3. Joint probability density function and the contour line.

Table 4 The joint return period of 100-year wave height and corresponding wind speed, wave period with different return periods

Wind speed Af \ Joint return period Af \ Wave period Af	100	200	500	Wind speed Af \ Joint return period Af \ Wave period Af	100	200	500
100	189	286	588	100	196	294	588
200	278	357	625	200	286	370	625
500	667	714	833	500	667	714	909

Table 5 The joint return period of 500yrs wave height and corresponding wind speed, wave period with different return periods

Wind speed Af \ Joint return period Af \ Wave period Af	100	200	500	Wind speed Af \ Joint return period Af \ Wave period Af	100	200	500
100	385	526	833	100	370	526	845
200	455	556	909	200	435	556	909
500	625	769	1000	500	625	769	1031

able. So the PNLTCED model can be used to predict reasonably the return level of a certain return period in engineering environments when the data period is not too long.

2.3 Comparison of predicted results by different methods

Comparison of predicted results by different methods is shown in Tables 6 and 7.

(i) Using 100-year return period wind speed combined with 100-year wave height and 100-year wave period as the extreme load combination, their joint return periods are 189 years (group Af) and 196 years (group Bf), calculated using the PNLCED. This combination will always yield conservative predictions and in some cases may be too conservative[19].

(ii) Using 100-year return period wave height with the associated wind speed and associated wave period as the traditional definition of the extreme load combination, (here the meaning of ‘associated’ is the statistically expected value of wind speed and wave period coexisting with the 100 year return period wave height), their joint return periods are 152 years (Af) and 161 years (Bf), as calculated by the PNLTCED.

Table 6 Comparison of the calculated results using different methods (Af)

Method	Joint return period	Load combination		
		wave height (m)	wind speed (m/s)	wave period (s)
1	189	11.25	51.88	10.33
2	152	11.25	48.45	10.23
3	100	10.60	44.17	10.10

Table 7 Comparison of the calculated results using different methods (Bf)

Method	Joint return period	Load combination		
		wave height (m)	wind speed (m/s)	wave period (s)
1	196	12.03	43.28	9.64
2	161	12.03	39.45	9.45
3	100	11.25	38.61	9.52

(iii) Using the PNLTCED model to predict the load combination of wave height, wind speed and wave period with 100-year joint return period, the reducing rates of design wave height, wind speed and wave period using data Af are 5.8%, 14.9% and 2.2% respectively, while the reducing rates using data Bf are 6.5%,10.8% and 1.3%.

3 Conclusions

The PNLCED model takes account of the randomicity of the marine data sampling or the extreme wind frequency, and its dissymmetrical form shows the layered structure of the variables with different correlations (α-outside layer, β-inside layer). So the new model gives the full probability information of the extreme sea state factors.

The new theory is a kind of probability model which considers both the probability characteristics of inducement and the multivariate joint probability of its induced abnormal variables with some correlations. The new model can be applied in many kinds of engineering design, the probability prediction of natural disasters and finance risks.

The study of wave climate is important for marine engineering, nautical engineering and ocean voyage[20]. The use of multivariate compound extreme value distribution in the joint probability analysis of wind speed, wave height and wave period affords a new way to study wave climate.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 50379051).

References

1 Liu T F, Ma F S. Prediction of extreme wave heights and wind velocities. Journal of the Waterway Port Coastal and Ocean division, ASCE, 1980, 106, No.WW4, 469－479
2 Liu T F. Long term distribution of hurricane characteristics. Proc. Offshore Technology Conference OTC 4325, 1982, 305－313
3 Langlev R M, Ei-Shaarawi A H. On the calculation of extreme wave heights: A review. Ocean Eng, 1986, 13(1): 93－118
4 Kirby W H, Moss M E. Summary of flood-frequency analysis in the United States. J Hydrology, 1987, 96: 5－14
5 Ochi M K. Stochastic analysis and probabilistic prediction of random seas. Adv Hydrosci, 1982, 13: 5－14
6 Naffa M G, Fanos A M, Elganiny M Y. Characteristics of waves off the Mediterranean Coast of Egypt. J.Coastal Res, 1991, 7: 38－49
7 Quek S T, Cheong H F. Prediction of extreme gusts accounting for seasonal effects. J Struct Safety, 1992, 11(2): 121－129
8 United States Government Accountability Office. History of the Lake Pontchartrain and Vincinity Hurricane Protection Project, Nov.9, 2005, GAO-06-244T
9 Liu D F, Wen S Q, Wang L P. Poison-gumbel mixed compound distribution and its application. Chin Sci Bull, 2002, 47(22): 1901－1906
10 Liu D F, Li H J. Prediction of extreme significant wave height from daily maxima. China Ocean Eng, 2001, 15(1): 97－106
11 Liu D F, Ma L, Jing K. Risk analysis of the disaster prevention design criteria for an estuarine city—Shanghai. In: Proceedings Offshore Mechanics & Arctic Engineering, Vancouver, Canada, OMAE 2004－51505
12 Liu D F, Song Y. Poisson-logistic compound bivariate extreme distribution and its application for designing of platform deck clearance. In: Proceedings Offshore Mechanics & Arctic Engineering, Cancum, Mexico, OMAE 2003-37395
13 Liu D F, Dong S, et al. System analysis of disaster prevention design criteria for coastal and estuarine cities. China Ocean Eng, 2000, 14(1): 69
14 Shi Daoji, Zhou Shengsheng. Moment estimation for multivariate extreme value distribution in a nested logistic model. Ann Inst Statist Math, 1999, 51(2): 253－264
15 Coles S G, Tawn J A. Modelling extreme multivariate events. J R Statist Soc B, 1991, 53(2): 377－392
16 Smith R L. Extreme value analysis of environmental time series: An example based on ozone data (with discussion). Statist Sci, 1989, l.4: 367－393
17 Davison A C, Smith R L. Models for exceedances over high thresholds (with discussion). J R Statist Soc B, 1990, 52: 393－442
18 Pickands J. Statistical inference using extreme order statistics. Ann Statist, 1975, l.3: 119－131
19 American Petroleum Institute. Planning, Designing and Constructing Fixed Offshore Platform－Load and Resistance Factor Design. RP2A-LRFD (ISO 13819-2), 1995(E)
20 Simoes O. Wave climate modeling south of Ria de Janeiro in Brazil. J Continental Shelf Res, 2002, 22: 2021