JOINT PROBABILITY ANALYSIS OF SIGNIFICANT WAVE HEIGHTS AND EXTREME WATER LEVELS AT POINT REYES STATION

[1]XINGHUA ZHU, [2]SUDONG XU, [3]KE YANG

[1,2,3]Department of Harbor, Waterway and Coastal Engineering,
School of Transportation, Southeast University, Nanjing 210096, China.
E-mail: [2]sudongxu@seu.edu.cn

Abstract - To ensure the safety, practicality and rationality of engineering design in the coastal areas, it is important to consider various environmental factors and their extreme values. Water level and wave height directly produce an effect on engineering design in several dominant environmental factors and they are correlative, so their correlation are seriously considered. The objective of this study is analyzing joint probability of extreme water levels and significant wave heights at Point Reyes Station. For this purpose, two joint probability distribution models, that are Gumbel □model and Gumbel Logistic model, are used to fit to joint probability distribution of significant wave heights and extreme water levels. Fitting results of two models are both effective though Gumbel Logistic model is slightly better. Two established models are applied to analyze the classic joint occurrences of water level and wave height, that are the combination of 10-year return period water level and 100-year return period wave height, the combination of 50-year return period water level and 50-year return period wave height, the combination of 100-year return period water level and 10-year return period wave height. It shows that joint return period is different from the return period of single variable. The return period contour lines of various combinations show that the growth rate of joint return period is increased with constant growth of single variable's return period.

Keywords - Point Reyes, California; Extreme Water Levels; Significant Wave Heights; Joint Probability; Return Period.

I. INTRODUCTION

Project sites of coastal engineering are located in the transitional zone of land and ocean. Coastal engineering is affected by some environmental factors from land and various marine environmental factors, such as typhoon, storm surge, etc. These factors are not independent. They are interrelated and interact on each other. In order to ensure safety, practicality and reasonable cost, we consider joint probability of several dominant environmental factors. Since the single variable analysis can not reflect the effects of complicated environmental factors, people began to focus on multiple variables analysis[3]. So far, several methods and models were proposed to analyze joint probability of multiple variables. Each of them has advantages and disadvantages, what make them suitable for different situations. There are complete probability method, stochastic simulation method, normal transforms method, multivariate extreme value theory, maximum entropy model, copula function, etc[3].

Water levels and wave heights are considered very important in various environmental factors. They directly produce an effect on engineering design. So we use correlation methods to analyze joint probability of water levels and wave heights in this paper. At present, the engineering design usually use extreme value data sets. In order to avoid the deterministic and periodicity of extreme value data sets, we use annual extreme value data sets[2]. In this paper, extreme water levels are annual maximum water levels (AMWL), and significant wave heights are annual significant wave heights (ASWH). Traditional analysis method is to superpose the extreme values of specific return periods[1]. This method is too conservative that its predictions are unreasonable[1]. Besides, the probability of two specific return periods' appearing simultaneously is small[3]. As a result, it is necessary to analyze joint probability of extreme water levels and significant wave heights.

II. DATA SETS AT POINT REYES STATION

In order to analyze joint probability of ASWH and AMWL, it is desired to have a long-term data set covering a period of at least 20 years. Point Reyes Station, the water level and wave height measurement station, is located in western Marin County, California. Point Reyes Station is located at 38°04′09″N, 122°48′25″W, just south and east of the southern end of Tomales Bay (**Fig.1.**).



Fig.1. Studied site：Point Reyes，California

In this study, wave heights data set and water levels data set in Point Reyes Station in a same period are collected. 20-year observed wave height data set for the period of 1995-2014 is obtained from The Coastal Data Information Program (CDIP). CDIP is the website of Integrative Oceanography Division (http://cdip.ucsd.edu). It provides the latest coastal conditions, wave models, forecasts, archived wind, wave and temperature data in history. Wave height data is recorded once half an hour or an hour at CDIP. 20-year observed water level data set for the period of 1995-2014 was obtained from The National Oceanographic and Atmospheric Administration (NOAA) of the United States. NOAA has maintained a national network of water level monitoring stations distributed in regional scale that has been operating for several decades (http://tidesonline.nos.noaa.gov/geographic.html). Hourly water levels over a period of several decades have been processed and are available for online download from the NOAA website[11]. Based on initial collected data, 20-year ASWH and 20-year AMWL are gained by calculation and selection. The results are presented in **Fig.2.** and **3.**.

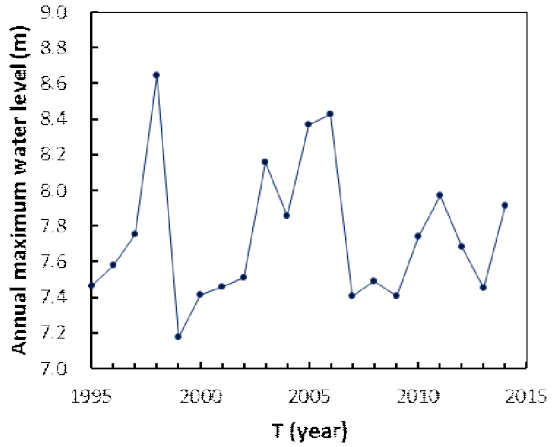


Fig.2. 20-Year data set of AMWL in Point Reyes Station (1995–2014).

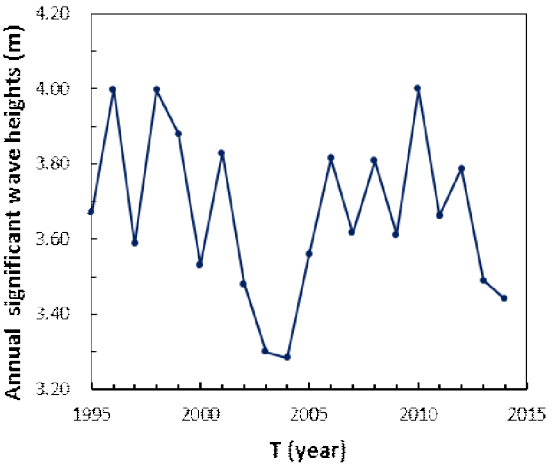


Fig.3. 20-Year data set of ASWH in Point Reyes Station (1995–2014).

III. JOINT PROBABILITY DISTRIBUTION MODEL

The studied objects in this paper are ASWH and AMWL. We found that the distributions of ASWH and AMWL are nonlinear correlated. Extreme value analysis for two variables is based on that for one variable[1]. The marginal distribution function of each variable and the correlation model describing the correlation of each variable are established. Then, the correlation of two variables is linked by the function and model.

In recent years, two-dimensional joint probability distributions are often used in the marine environment and hydrological analysis. Here are some main kinds: Gumbel distribution, normal distribution, Pearson type □ distribution, Weibull distribution, Gamma distribution, equivalent maximum entropy distribution, etc[3]. According to the hydrological factors in this paper and experience in this research field, we use two-dimensional Gumbel distribution to fit to them. Gumbel □model and Gumbel Logistic model are employed.

3.1. One-dimensional Gumbel distribution

In general, the theoretical frequency distribution of climate and hydrological elements always use Gumbel distribution[3], the function of its cumulative probability expression is:

$$F(x) = \mathrm{P}\{x_0 < x\} = \exp[-\exp(-\frac{x-\mu}{\sigma})], -\infty < x < +\infty \quad (1)$$

Where μ is location parameter, σ is scale parameter.

3.2. Gumbel □distribution model

There are any two values named x_1 and x_2. $F(\eta \geq x_1, Hs \geq x_2)$ is the probability of extreme water level higher than the x_1 and significant wave height higher than the x_2 occurring simultaneously. It is called joint probability of significant wave heights and extreme water levels[2].

$$F(\eta \geq x_1, Hs \geq x_2) = \exp[-(y_1^{1/\alpha} + y_2^{1/\alpha})^{\alpha}] \quad (2)$$

$$y_i = -\ln\left\{1 - \exp\left[-\exp\left(-\frac{x_i - \mu_i}{\sigma_i}\right)\right]\right\}, i = 1,2 \quad (3)$$

Where μ_i, σ_i are the marginal distribution parameters. They are determined by the least square method using the statistical characteristic value of the sample. μ_i is the location parameter of one- dimensional Gumbel distribution and σ_i is the scale parameter of one-dimensional Gumbel distribution. α is correlation coefficient linking the two single variables. Its value ranges from 0 to 1. The value of α is determined by the least square method and graphing method[2].

3.3. Gumbel Logistic model

There are any two values named x_1 and x_2. $P(\eta < x_1, Hs < x_2)$ is the probability of extreme water level lower than the x_1 and significant wave height lower than the x_2 occurring simultaneously. It

is called joint probability of significant wave heights and extreme water levels[1].

$$P(\eta < x_1, Hs < x_2) = \exp[-(y_1^{1/\alpha} + y_2^{1/\alpha})^{\alpha}] \quad (4)$$

$$y_i = -\ln(F(x_i)), i = 1, 2 \quad (5)$$

$$F(x_i) = \exp\left[-\exp\left(-\frac{x_i - \mu_i}{\sigma_i}\right)\right], i = 1, 2 \quad (6)$$

Where $F(x_1) = P(\eta < x_1)$ is cumulative probability of marginal distribution about AMWL. $F(x_2) = P(H_s < x_2)$ is cumulative probability of marginal distribution about ASWH. The meaning of μi, σi and α are the same as above.

Joint transcendental probability of Gumbel Logistic model is following equations[1]:

$$F(\eta \geq x_1, H_s \geq x_2) = 1 + P(\eta < x_1, H_s < x_2) - F(x_1) - F(x_2) \quad (7)$$

IV. MODEL-FITTING ANALYSIS

4.1. Parameters evaluation

The marginal distributions of AMWL or ASWH conform one-dimensional Gumbel distribution. The location parameter and the scale parameter are determined by the least square method. The results are shown in **Table 1**. The parameter α shows correlation between AMWL and ASWH. Its value affects the goodness of fit. The values of α determined by the least square method and graphing method[1][2] are 0.5388 for Gumbel model and 0.6345 for Gumbel Logistic model (**Table 1**).

Table1: Parameters of marginal distribution and joint probability model

Model		Gumbel II	Gumbel Logistic
AMWL	μ1	7.5652	7.5652
	σ1	0.3088	0.3088
ASWH	μ2	3.5697	3.5697
	σ2	0.1698	0.1698
	α	0.5388	0.6345

4.2. Model fitting analysis

Supposing there are n years' data set and one significant wave height is regarded as the standard. To find out all the years of which ASWH and AMWL are both higher than the standard. The m-group data obtained in whole sample combinations is the final result. Joint distribution of wave height and water level for the specific year is cumulative frequency (m/n). The cumulative frequency of all sample combinations can be calculated according to this method[2].

The comparison for joint probability of model simulations and actual is presented in **Fig.4.** and **5.**. The figures show that the two-dimensional flat points are distributed on both sides of the ideal straight line uniformly and intensively[2]. It shows that fitting results are good. Both models are available for describing the joint probability of AMWL and ASWH in Point Reyes Station.

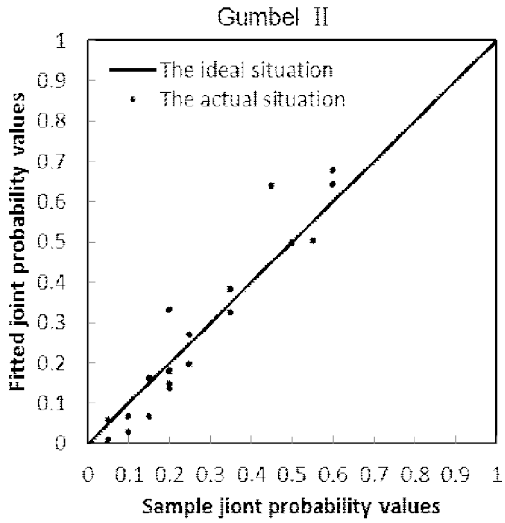


Fig.4. The comparison of Gumbel II model simulation and sample value.

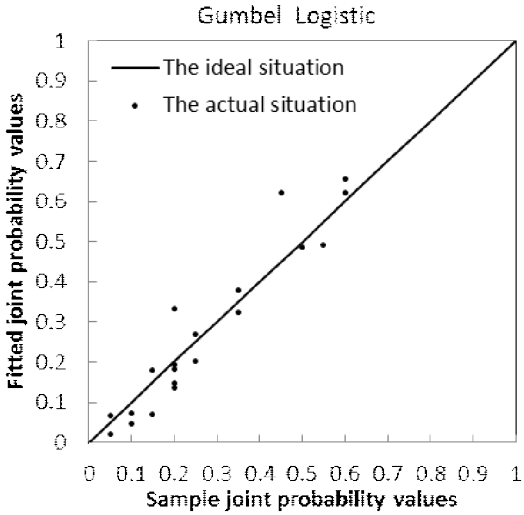


Fig.5. The comparison of Gumbel Logistic model simulation and sample value.

4.3. Combination of joint probability analysis

In the field of ocean engineering, the commonly used design method is probability analysis of each environmental factor, and one probability is selected as a design standard. This method is not reasonable enough because joint probability of environmental factors is far less than the probability of independent events[2]. Gumbel□model and Gumbel Logistic model are used to explain the fact.

There is not exact equation to connect joint return period and joint probability. In order to analyze joint return period expediently, equation $T = 1/F(\eta \geq x_1, h_s \geq x_2)$ is used to estimate joint return period T[2]. **Table 2** lists the classic joint occurrences: The combination of 10-year return period water level and 100-year wave height, the

combination of 50-year return period water level and 50-year wave height, the combination of 100-year return period water level and 10-year wave height.

Table 2 Different joint probabilities and return periods of classic joint occurrences.

Joint events	Gumbel II	
	Joint probability	Return period/year
$\eta_{10}+H_{s100}$	0.01286	78
$\eta_{50}+H_{s50}$	0.0055	182
$\eta_{100}+H_{s10}$	0.00545	183
Joint events	Gumbel Logistic	
	Joint probability	Return period/year
$\eta_{10}+H_{s100}$	0.01154	87
$\eta_{50}+H_{s50}$	0.00633	158
$\eta_{100}+H_{s10}$	0.00535	187

Table 2 shows that the joint return period of 10-year return period water level and 100-year return period wave height is less than 100 years, so it is dangerous if 100-year joint return period is regarded as the standard. The joint return period of 100-year return period water level and 10-year return period wave height is more than 100 years, so it is too conservative if 100-year joint return period is regarded as the standard. The joint return period of 50-year return period water level and 50-year return period wave height is much more than 50 years, so it is wasteful if 50-year joint return period is regarded as the standard. Above all, extreme water levels should be considered as the dominant variable in the studied site when researching the joint probability of extreme water levels and significant wave heights.

4.4. The most likely probability event study for different return period

Using the established Gumbel□model and Gumbel Logistic model to calculate the joint probability and return period of variable combinations of extreme wave level and significant water height. The contour lines of joint return periods are drawn on the map (**Fig.6.** and **7.**).

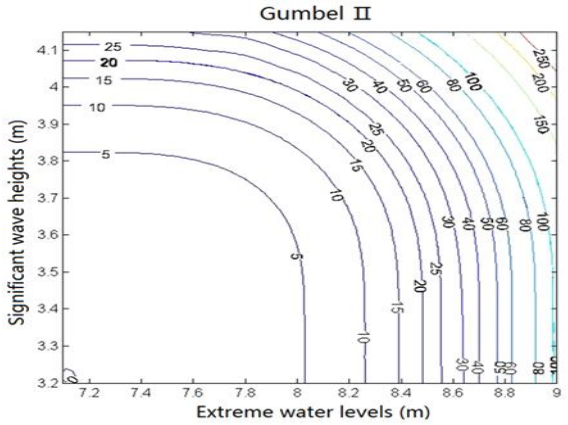


Fig.6. The return periods contour lines of various combinations determined by Gumbel II model.

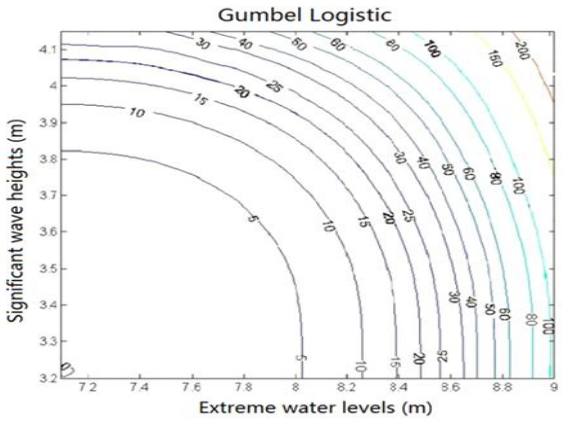


Fig.7. The return periods contour lines of various combinations determined by Gumbel Logistic model.

The greater extreme water level represent the greater independent return period of it, as well as significant wave height. **Fig.2.** and **3.** show that contour lines are more and more intensive toward the upper right corner when the extreme water level is above 8 m or significant wave height is above 3.8 m. With the increase of extreme water levels or significant wave heights, the growth rate of joint return period is increased. And return period of single variable is different from the joint occurrences.

V. MODEL COMPARISON AND DISCUSSION

Fitting results of Gumbel □model and Gumbel Logistic model are both good. They are showed in **Fig.4.** and **5.**. There isn't much difference though the fitting result of Gumbel Logistic model is slightly better. Gumbel □model is more sensitive to single variable than Gumbel Logistic model and the difference is more obvious with the return period of single variable increasing. In this paper, extreme water levels have a greater influence on joint return period and prediction in studied site. Therefore, direct linear superposition for single variable's return period is inappropriate. Joint probability and joint return period are taken into account on engineering design.

CONCLUSIONS

Joint return period about extreme water level and significant wave height is different from return period of single variable. Joint return period is much longer than return period of single variable when return periods of two variables are the same. Such as the joint return period of 50-year return period water level and 50-year return period wave height is closed to 180~200 years, 3~4 times as much as 50 years. If the growth rate of single variable's return period is constant, the growth rate of joint return period is increased. Fitting results of Gumbel II model and Gumbel Logistic model are both effective though Gumbel Logistic model is slightly better. Gumbel II model is more sensitive to single variable than Gumbel Logistic model, and the difference is more

obvious with the growth of single variable's return period. There is a dominant variable in the joint events. Extreme water level is the dominant variable of joint probability of extreme water levels and significant wave heights in this studied site. The prediction of joint probability and return period determined by Gumbel II model or Gumbel Logistic model are available. The results are reasonable for the engineering design.

ACKNOWLEDGMENTS

This study was supported by The Natural Science Foundation for Young Scientists of Jiangsu Province, China (BK2012341); and National Natural Science Foundation of China (Research grant #51209040).

REFERENCES

[1] D. Xie, Y. Chen, C Zhang, J. Li, "Joint probability analysis of extreme waves and surges in Jiangsu and nearby offshore area", The Ocean Engineering , vol.32, no.4, pp.64-71, 2014. (in Chinese)

[2] S. Li, W. Sha, Y. Qi, "Wave and storm surge of the Wulei Station at the Beibu Bay", Marine Science Bulletin, vol.25, no.4, pp.23-28, 2006. (in Chinese)

[3] S. Tao, "Study on multivariate maximum entropy models and their application in coastal and ocean engineering", Qing Dao: Ocean University of China, 2013. (in Chinese)

[4] M. Masina, A. Lamberti, R..Archetti, "Coastal flooding : A copula based approach for estimating the joint probability of water levels and waves", Coastal Engineering, vol.97,pp.37-52, 2015.

[5] J.E. Heffernan, and J.A. Tawn, "A conditional approach for multivariate extreme values", Journal of Royal Statistical Society, vol.66,pp.497-546, 2004.

[6] Peter J. Hawkes, "Joint probability analysis for estimation of extremes", Journal of Hydraulic Research, vol.46, pp.246-256, 2008.

[7] Jules J. Beersma, T. Adri. Buish, "Joint probability precipitation and discharge deficits in the Netherlands", Water Resources research, vol.40, no.12, W12308, 2004.

[8] D.A. Gaffney, G.L. Williams, "Joint probability of superelevated water levels and wave heights at Duck, North Carolina". Proceedings of the 2nd International Symposium on Ocean Wave Measurement and Analysis, pp. 905-917, 1994.

[9] Hugo C. Winter, Jonathan A. Tawn, "Modeling heatwaves in central France: a case-study in extremal dependence", Journal of Royal Statistical Society, Applied Statistics, Series C. pp.1-21, 2015.

[10] C.W. Li, Y. Song, "Correlation of extreme waves and water levels using a third-generation wave model and a 3D flow model", Ocean engineering, vol.33, pp.635-653, 2006.

[11] W. Huang, S. Xu, S. Nnaji, "Evaluation of GEV model for frequency analysis of annual maximum water levels in the coast of United States", Ocean Engineering, vol.35, pp.1132-1147, 2008.

★★★