

# 复合极值分布在台风风暴潮增水重现期计算中的应用研究

付翔, 梁森栋, 李涛, 吴少华

(国家海洋环境预报中心, 北京 100081)

**摘要:** 考虑小样本或无规律极值样本的情况下, 引入描述随机事件发生次数的离散型分布, 以三沙站为例, 研究了 Poisson-Gumbel 复合极值分布在台风风暴潮重现期计算中的应用效果。结果表明: 复合极值分布计算结果对样本数据较为敏感, 标准差小的样本数据其复合分布的低频值显著偏低; 对具有一定离散度的短时间序列数据, 复合分布的计算结果稳定, 且对低频值的估算也更准; 对于较长时间序列的数据, 阈值取值偏低则高频次发生值更准, 阈值取值偏高则低频次发生值更准。

**关键词:** 复合极值分布; 重现期; 台风风暴潮; 阈值

**中图分类号:** P731.23 **文献标识码:** A **文章编号:** 1003-0239(2022)04-0001-08

## 1 引言

风暴潮灾害是我国最主要的海洋灾害, 特别是台风风暴潮。自1989年有统计数据以来, 风暴潮(含近岸浪)灾害造成的直接经济损失占全部海洋灾害直接经济损失的90%以上, 严重的台风风暴潮灾害几乎每年都会在不同地区发生<sup>[1]</sup>。因此, 在涉及海洋工程建设、海岸环境损害评价、保险的灾害风险计算以及气候变化影响评估时, 需考虑风暴潮的致灾危险性。风暴潮致灾因子的危险性体现在风暴潮的强度、频率以及叠加浪、涌和天文潮后引起的高潮位上, 极值发生频率( $P$ )是常用的量化指标之一<sup>[2-3]</sup>。

极值发生频率和重现期( $T = 1/P$ )传递的是极端事件最大可能的信息。它有两层含义: 一是对于某频率 $p$ 的极值, 下次出现该值的平均等待时间为 $Ta$ ; 二是在 $Ta$ 里出现该极值的次数平均是1次<sup>[4]</sup>。对于台风风暴潮, 由于台风影响具有随机性, 某一区域并不是每年都遭受显著的台风风暴潮, 但在某些年却可能遭受一次以上的台风风暴潮灾害, 例如2014年连续遭受台风“威马逊”和台风“海鸥”袭击的雷州半岛, 以及2015年台风“苏迪罗”和台风“杜

鹏”连续登陆的福建中部。因此, 如果使用年极值法来计算风暴潮重现期, 每年只取一个极值, 有可能会取到无台风或弱台风影响年的小值而舍去了频发年的次大值, 使得计算数据失真, 估算的重现期极值出现较大偏差。考虑随机事件发生次数的复合极值分布可以有效地解决这一问题。

引入一个描述无规律事件发生次数的离散型概率分布——Poisson分布, 使其与一个连续型的极值分布复合, 组成的复合极值分布不仅能拟合每年实测变量随机的数据序列, 对于资料年限较短的情况也能可靠应用<sup>[5]</sup>。复合极值分布提出后, 不少学者在台风波高、极值风速和风暴潮的计算预测中进行了研究和应用<sup>[6-10]</sup>。目前 Poisson-Gumbel 复合极值分布已被推荐用于台风波高的设计值计算<sup>[11-12]</sup>。本文将该分布法用于台风风暴潮的重现期计算, 分析其优缺点及应用效果。

## 2 计算方法和资料

### 2.1 极值 I 型频率曲线分布

港口与海岸工程设计中, 常采用极值 I 型分布

收稿日期: 2021-08-09; 修回日期: 2021-09-13。

基金项目: 国家重点研发计划(2017YFA0604903, 2017YFA0604901)。

作者简介: 付翔(1980-), 女, 高级工程师, 硕士, 主要从事风暴潮预警研究工作。E-mail: fx@nmefc.cn

律推求指定频率或重现期的风暴潮和极端高低潮位值<sup>[11-12]</sup>。它是广义极值分布(Fisher-Tippett)的一种,最早由Gumbel<sup>[13]</sup>用于水文统计。其概率分布函数为双指数形式,表示为:

$$G(x) = e^{-e^{-(x-\mu)/\alpha}} \quad (1)$$

式中, $\mu$ 为位置参数,即分布的众数; $\alpha$ 为尺度参数,代表分布的离散性。这两个参数可用多种方法进行估计<sup>[14-15]</sup>。

## 2.2 离散型 Poisson 分布

连续的极值 I 型分布律仅能用于规律的极值样本,如年极值或月极值等。而对于某一地区,每年台风风暴潮的发生次数和强度都是随机的,它们构成的是离散型分布。描述稀有事件发生次数的 Poisson 分布非常适合用以描述台风风暴潮的发生概率。假定某地一定强度以上的台风风暴潮的发生频次  $n$  符合 Poisson 分布,其概率质量函数表示为:

$$P_k = e^{-\lambda} \frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \dots, n) \quad (2)$$

式中, $\lambda$ 为年平均台风风暴潮发生次数。

## 2.3 Poisson-Gumbel 复合极值分布

根据复合极值分布式<sup>[5-6]</sup>,给定设计频率  $p'$  的极值分布为:

$$F(x) = \sum_{k=0}^{\infty} P_k [G(x)]^k = p \quad (3)$$

式中, $p = 1 - p'$ 。定义  $T = \frac{1}{p'}$ ,称为重现期。将式(1)、(2)代入(3),则某  $T$  年一遇的台风风暴潮值  $x_p$  可表示为:

$$x_p = \mu + \alpha \left( -\ln \left( -\ln \left( 1 + \frac{1}{\lambda} \ln p \right) \right) \right) \quad (4)$$

具体推算见文献[6]。

## 2.4 过阈法

过阈法(Peak Over Threshold, POT)是极值理论下确定计算样本序列的一种方法,即先确定某一观测值下限(阈值),凡超过此值的均列入统计序列,而后再进行极值分布拟合。阈值的确定十分关键,阈值过高会使可利用的样本数过少,从而导致分布函数估计的参数方差偏高;阈值过低又会使分布的渐进性得不到满足,产生有偏估计。阈值的选取有

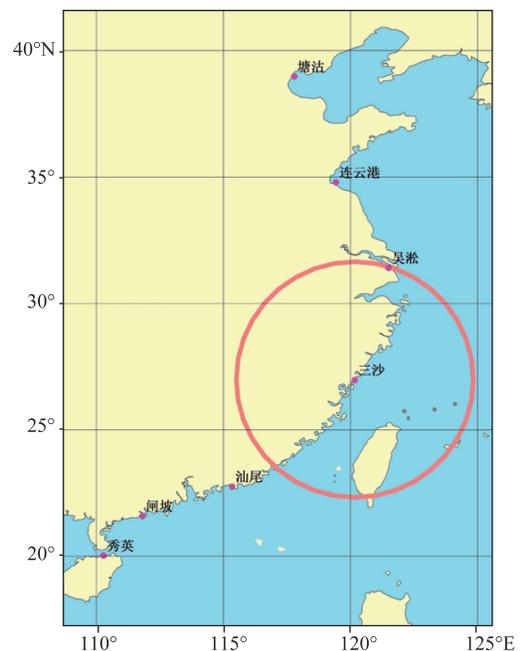
多种方法,如平均剩余寿命图法(Mean Residual Life Plot)、平均超越期望图法(Mean Excess Plot)、Hill图(Hill plot)等。算法包括峰度法和渐进均方误差法等。

## 2.5 资料来源

台风信息采用中国气象局的《台风年鉴》(1980—1988年)和《热带气旋年鉴》(1989—2016年),以及热带气旋资料中心最佳路径数据集<sup>[16]</sup>(网址: <http://tcdata.typhoon.org.cn>)中的资料(2017—2019年)。风暴潮数据来自国家海洋环境预报中心,在“908”专项台风风暴潮灾害历史数据集基础上进行延长和补充,增水值由台风影响期间测站实测逐时水位值去除天文潮得到,实测水位值根据国家标准<sup>[17]</sup>进行观测,数据经过严格的质量控制,排除了由于台站迁址、仪器变更和地面沉降等产生的影响。

## 3 计算个例

三沙验潮站位于福建省东北部(见图1),几乎每年都会受到台风影响。以其为例,利用 Poisson-Gumbel 复合极值分布计算其多年一遇台风风暴潮。



注:图中圆圈为距三沙站 500 km 半径圆

图1 部分代表验潮站分布图

Fig.1 Location of some typical water gauge

第一步:确定阈值。采用平均剩余寿命图法<sup>[17]</sup>和百分位法确定阈值范围,挑选1980—2019年间进入距三沙站约500 km范围以内的台风,共计171个,统计每次过程的最大风暴潮值,剔除其中的减水过程后建立经验频率序列。根据平均剩余寿命图(见图2),考虑到近似线性和置信区间范围<sup>[18]</sup>,阈值可在20~80 cm之间选取。取其中值,同时也是增水极值累积频率的中位数50 cm为统计阈值(见图3),使其发生次数符合Poisson分布。

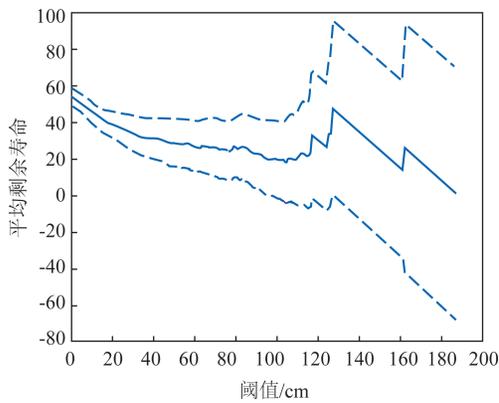


图2 三沙站台风风暴潮增水极值序列平均剩余寿命图 (虚线为95%置信区间)

Fig.2 Mean residual life of typhoon storm surge extreme value in Sansha station (The dotted line is the 95% confidence interval)

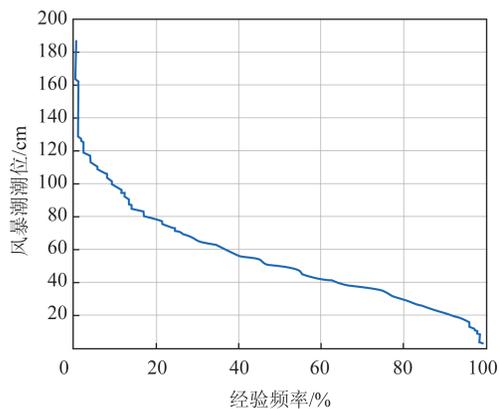


图3 三沙站台风风暴潮增水极值累积频率曲线

Fig.3 Accumulative frequency curve of typhoon storm surge extreme value in Sansha station

第二步:Poisson分布检验。统计三沙站超过阈值的台风风暴潮过程,由式(2)计算其发生次数的

理论频率分布(见图4),采用卡方分布  $\chi^2 = \sum_{i=0}^k \frac{(f_i - nP_i)^2}{nP_i}$  ( $k = 6$ )进行检验。取显著水平0.05,查自由度为5 ( $k - 1$ )的卡方分布临界表  $\chi_{0.05}^2 = 11.07$ ,由  $\chi^2 < \chi_{0.05}^2$ 可知三沙站最大增水50 cm以上的台风风暴潮发生频数符合Poisson分布。

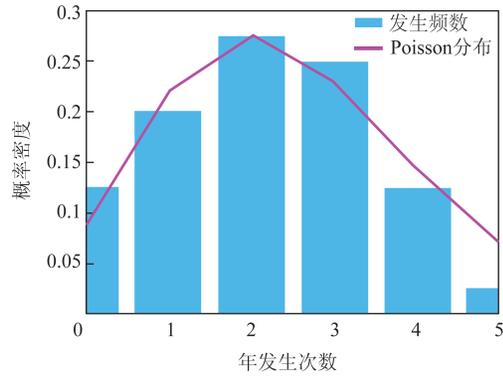


图4 三沙站台风风暴潮发生频次及概率密度分布

Fig.4 Distribution of probability density and typhoon storm surge occurrence frequencies in Sansha station

如果发生频数拟合不能通过检验,则应该适当提高阈值大小减少样本量,再进行表1计算,直至通过检验。

表1 台风风暴潮发生次数拟合检验

Tab.1 Fitting test of the number of typhoon storm surge events

发生次数	分组出现次数 $f_i$	Poisson 分布概率 $P_i$	$\chi^2$
0	5	0.082 1	0.896 667
1	8	0.205 2	0.005 271
2	11	0.256 5	0.053 372
3	10	0.213 8	0.245 171
4	5	0.133 6	0.022 144
5	1	0.066 8	1.046 251
共计	40	—	2.268 877

第三步:Gumbel分布检验。P-P图是变量的累计比例与指定分布的累积比例之间的关系图,可直观地检验样本数据是否符合指定概率分布。当数据符合指定分布时,P-P图中各点近似呈一条直线,

即代表样本数据的点基本在代表理论分布的对角线上。根据P-P图,三沙站过阈台风风暴潮极值序列符合 Gumbel 分布(见图 5),也可采用 Kolmogorov-Smirnov 检验进行数据序列的 Gumbel 分布检验。

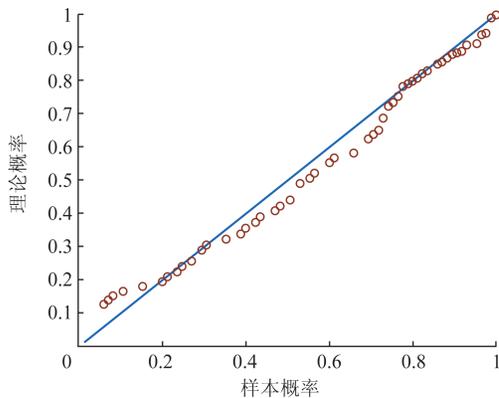


图5 Gumbel分布概率检验图

Fig.5 Gumbel distribution probability plot

第四步:复合极值分布拟合。根据式(4)拟合三沙站过阈风暴潮增水极值序列,利用有限样本分布函数代替总体分布函数的 Gumbel 法<sup>[13]</sup>估算参数  $\mu$  和  $\alpha$ ,然后计算不同重现期的台风风暴潮极端增水值。

## 4 比较分析

### 4.1 与年极值序列计算结果比较

建立三沙站 1980—2019 年的台风风暴潮年极值增水序列,利用 Gumbel 分布拟合计算三沙站台风风暴潮多年一遇重现期增水值,与复合极值分布拟合的结果进行比较(见表 2)。结果显示,对于 10 年一遇以下增水,Gumbel 计算的增水值比复合分布拟合的结果小,而对于 10 年一遇以上增水,Gumbel 计算值比复合分布拟合的结果大,且随着重现期增

大,差比也增大。在中高频次发生范围内,两种分布的计算值比较接近,但在高频次和低频次发生范围内的计算差异较大。究其原因,采用年极值序列拟合计算时,由于无台风影响年低值的存在会导致 Gumbel 曲线的斜率增大,使得高频发生值更小,低频发生值更大;而过阈法序列不仅包含了在相同年份发生的高值,还剔除了弱过程,避免了增水小值带来的较大观测误差和潮汐预报误差,但由于相对低值样本增多,导致低频次发生的风暴潮值计算偏小。

根据三沙站台风风暴潮原始统计序列,40 a 间增水大于 79 cm 的过程为 33 次,约为 1.2 年一遇;增水大于 90 cm 以上的过程为 23 次,约为 1.7 年一遇;增水大于 95 cm 的过程为 20 次,正为 2 年一遇。可见 Poisson-Gumbel 复合极值分布的高频次发生值的结果更为合理。另外,由图 6 也可看出,对于过阈原始序列的拟合,复合极值分布对极大值的拟合结果明显好于 Gumbel 分布。

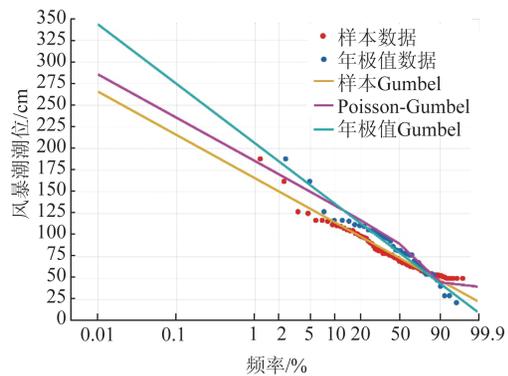


图6 分布拟合比较

Fig.6 Comparison between curves of different distribution

### 4.2 阈值敏感性试验

由于接近阈值的小值数量的多少对拟合估算结果影响较大,因此需要调整阈值大小,比较不同阈值下样本序列的复合分布拟合结果。取最大增水阈值 60 cm、65 cm、70 cm 和最高 75 cm,分别截

表 2 不同分布拟合的重现期增水值比较(单位:cm)

Tab.2 Comparison of surge return values between different distribution (unit: cm)

重现期/a	2	5	10	20	50	100	500	1 000
年极值 Gumbel 分布	79.8	113.6	135.9	157.4	185.1	205.9	254.0	274.7
Poisson-Gumbel 分布	90.1	117.2	134.0	149.9	170.2	185.4	220.5	235.6

取台风风暴潮过程样本序列进行复合分布计算。所有序列的发生频数分布均通过0.05显著性检验的Poisson分布。结果显示(见表3),阈值选取增加,所有重现期的计算值都会升高,在阈值为70 cm时达到最高值,后又开始下降。文献资料显示<sup>[19]</sup>,三沙站1949年以来台风风暴潮极值为190~200 cm(6614号台风风暴潮),与阈值为70 cm和75 cm的计算结果相近。由此可见,虽然高频次发生值的估计偏高,但对于低频范围的高年遇值,高阈值序列的拟合效果更好。

### 4.3 分段计算结果比较

根据相关标准<sup>[11-12]</sup>,在进行极端水位值确定时,应用不少于连续20 a的潮位实测资料进行计算。将三沙站40 a的资料序列分为前20 a和后20 a,以阈值70 cm为例(其余结果略),分别对两种方法的拟合计算进行比较,结果见表4。我们发现,两种方法两个时间段的差值均自高频向低频增大,复合分布

结果仅在高频处差值较小,10年一遇以上低频结果两个时间段的差值的复合分布远高于Gumbel年极值分布。其原因在于三沙站这两个时间段的台风风暴潮极值序列差别较大,前20 a极值偏小且分布均匀,以阈值70 cm为例,前20 a的极值样本仅为18个,均值为89.8 cm,标准差为16.8 cm;而后20 a的样本为26个,极值均值为99.7 cm,标准差为27.3 cm。样本数据的较大差异导致了复合极值分布对低频发生值的估算差别显著,前20 a数据的计算值严重偏低。

在包含有40 a间最大增水的原始过阈样本和年极值样本中,任意挑选两组20 a的数据序列进行两种分布拟合计算,由于观测年代较短,复合分布应取低阈值以增加样本数量,这里取50 cm。由两种分布的两组结果比较可见(见表5),在观测年代较短且样本较少的情况下,复合分布的拟合结果比单一的年极值Gumbel分布结果更为稳定,两组数据结果的差值更小,低频值也更为合理。

表3 拟合不同阈值样本的Poisson-Gumbel复合分布重现期增水值比较(单位:cm)

Tab.3 Comparison of surge return values of Poisson-Gumbel distributions fitting samples with different threshold (unit: cm)

阈值/cm	重现期/a							
	2	5	10	20	50	100	500	1 000
50	90.1	117.2	134.0	149.9	170.2	185.4	220.5	235.6
60	93.1	120.5	137.2	152.8	172.9	187.9	222.4	237.2
65	97.8	125.1	141.8	157.5	177.6	192.5	227.0	241.8
70	102.3	129.6	146.3	162.0	182.0	197.0	231.5	246.3
75	98.0	127.0	144.0	159.9	180.2	195.2	230.0	244.9

表4 连续分段20 a样本数据重现期增水值比较(单位:cm)

Tab.4 Comparison of surge return values between different distribution of consecutive 20 a subsection data (unit: cm)

分布类型	年份和偏差	重现期/a							
		2	5	10	20	50	100	500	1 000
年极值Gumbel分布	1980—1999年	70.2	103.2	125.0	145.9	173.0	193.3	240.3	260.4
	2000—2019年	90.3	126.9	151.0	174.2	204.3	226.8	278.8	301.1
	偏差	20.1	23.7	26	28.3	31.3	33.5	38.5	40.7
Poisson-Gumbel分布	1980—1999年	95.1	115.4	127.8	139.4	154.3	165.3	190.9	201.9
	2000—2019年	98.3	132.8	153.1	171.9	196	213.9	255.2	273
	偏差	3.2	17.4	25.3	32.5	41.7	48.6	64.3	71.1

表5 随机分段20 a样本数据重现期增水值比较(单位:cm)

Tab.5 Comparison of surge return values between different distribution of random 20 a subsection data (unit: cm)

分布类型	年份和偏差	重现期/a							
		2	5	10	20	50	100	500	1 000
年极值Gumbel分布	1995—2014年	84.5	121.7	146.3	170.0	200.5	223.4	276.4	299.1
	2000—2019年	90.3	126.9	151.0	174.2	204.3	226.8	278.8	301.1
	偏差	5.8	5.2	4.7	4.2	3.8	3.4	2.4	2
Poisson-Gumbel分布	1995—2014年	93.1	124.7	144.3	162.9	186.6	204.4	245.3	262.9
	2000—2019年	98.3	128.9	148.1	166.3	189.7	207.1	247.4	264.7
	偏差	5.2	4.2	3.8	3.4	3.1	2.7	2.1	1.8

## 5 沿海代表站台风风暴潮重现期分布

根据第2部分的计算步骤,选取部分代表验潮站(见图1),利用Poisson-Gumbel复合分布拟合1980—2019年40 a的台风风暴潮增水极值序列,计算出各站50年一遇和100年一遇的台风风暴潮增

水值,阈值选取均在累积分布中位数附近并保证序列通过分布检验。由表6可见,塘沽、吴淞、闸坡和秀英站可能遭受的台风风暴潮强度较强,不仅增水值大,100年一遇和50年一遇增水的差值也大,发生频次低的风暴潮强度更强,说明台风在这些地方可能引起的极端风暴潮更高,危险性也更大。

表6 各站台风风暴潮重现期值(单位:cm)

Tab.6 Return values of typhoon storm surge in several typical stations (unit: cm)

重现期/a	塘沽	连云港	吴淞	三沙	汕尾	闸坡	秀英
50	237.2	185.8	232.6	180.2	177.7	230.8	240.9
100	263.8	203.2	253.8	195.2	193.1	254.0	266.1

## 6 小结

考虑随机事件发生频率的Poisson-Gumbel复合极值分布是解决观测资料年限较短或是出现无年极值(无台风影响)情况下,风暴潮极值重现期计算的有效方法。复合极值分布的计算结果对样本数据较为敏感,特别是标准偏差较小的样本数据,低频次发生值的计算结果显著偏低。阈值的选取对结果影响较大,对于较长时间序列数据,复合分布的阈值取值偏低则高频次发生值更准,阈值取值偏高则低频次发生值更准;而对于短时间序列数据,复合分布的计算结果更稳定,且对低频发生值的估算也更准,此时由于样本数量较少,应取偏小阈值以增加样本数量。对几个典型潮位站的重现期计算表明,塘沽、吴淞、闸坡和秀英站台风引起的可能

极端风暴潮较强,台风风暴潮致灾危险性较高。

复合分布在风暴潮极值计算中的特征可延伸至潮位极值计算。但由于其对数据样本的敏感性,在工程应用中,仍建议以差比法延长资料年限后再行计算。若考虑无年极值或一年多极值的情况,则需根据对高频值或低频值的需求,确定具体阈值。

### 参考文献:

- [1] 自然资源部/国家海洋局. 中国海洋灾害公报[EB/OL]. 1989-2019 [2020-07-10]. <http://www.nmdis.org.cn/hygb/zghyzhgb/>.  
Ministry of Natural Resources/State Oceanic Administration[EB/OL]. Bulletin of China Marine Disaster. 1989-2019[2020-07-10]. <http://www.nmdis.org.cn/hygb/zghyzhgb/>.
- [2] 冯爱青,高江波,吴绍洪,等. 气候变化背景下中国风暴潮灾害风险及适应对策研究进展[J]. 地理科学进展, 2016, 35(11): 1411-1419.

- FENG A Q, GAO J B, WU S H, et al. A review of storm surge disaster risk research and adaptation in China under climate change [J]. *Progress in Geography*, 2016, 35(11): 1411-1419.
- [3] 张月霞, 王辉. 台风风暴潮灾害风险评估研究综述[J]. *海洋预报*, 2016, 33(2): 81-88.
- ZHANG Y X, WANG H. Review of risk assessment of typhoon storm surge disaster[J]. *Marine Forecasts*, 2016, 33(2): 81-88.
- [4] 万仕全, 周国华, 潘柱, 等. 南京过去100年极端日降水量模拟研究[J]. *气象学报*, 2010, 68(6): 790-799.
- WAN S Q, ZHOU G H, PAN Z, et al. A simulative study of extreme daily rainfall in Nanjing for the past 100 years[J]. *Acta Meteorologica Sinica*, 2010, 68(6): 790-799.
- [5] 刘德辅, 马逢时. 极值分布理论在计算波高多年分布中的应用[J]. *应用数学学报*, 1976(1): 23-37.
- LIU D F, MA F S. Application of extreme value distribution theory in calculating multi-year distribution of wave heights[J]. *Acta Mathematicae Applicatae Sinica*, 1976(1): 23-37.
- [6] 马逢时, 刘德辅. 海洋工程建筑中设计波高推算的新方法[J]. *科学通报*, 1979, 24(1): 33-37.
- MA F S, LIU D F. A new method for designing wave height calculation in offshore engineering construction[J]. *Chinese Science Bulletin*, 1979, 24(1): 33-37.
- [7] LIU T F, MA F S. Prediction of extreme wave heights and wind velocities[J]. *Journal of the Waterway, Port, Coastal and Ocean Division*, 1980, 106(4): 469-479.
- [8] 董胜, 李奉利, 孙瑞文. 风暴增水随机分析的过阈法及其统计计算模式[J]. *青岛海洋大学学报(自然科学版)*, 2000, 30(3): 542-548.
- DONG S, LI F L, SUN R W. On the peak over threshold method in the statistical analysis of storm surge elevation and its probability model[J]. *Journal of Ocean University of Qingdao*, 2000, 30(3): 542-548.
- [9] 董胜, 宋艳, 杨志. 考虑季节变化的风暴增水随机统计分析[J]. *海洋工程*, 2003, 21(3): 68-72.
- DONG S, SONG Y, YANG Z. Statistical analysis of storm surge elevation considering seasonal variation[J]. *The Ocean Engineering*, 2003, 21(3): 68-72.
- [10] 张容焱, 张秀芝, 蔡连娃. 沿海风工程设计风速中泊松-耿贝尔法的应用[J]. *应用气象学报*, 2010, 21(2): 237-242.
- ZHANG R Y, ZHANG X Z, CAI L W. Application of Poisson-Gumbel distribution to wind speed calculation for the southeast coastland of China[J]. *Journal of Applied Meteorological Science*, 2010, 21(2): 237-242.
- [11] 中华人民共和国交通运输部. JTS 145-2-2013 海港水文规范[S]. 北京: 人民交通出版社, 2013.
- Ministry of Communications of the People's Republic of China. JTS 145-2-2013 Code of hydrology for sea harbour[S]. Beijing: People's Communications Press, 2013.
- [12] 中华人民共和国交通运输部. JTS 145-2015 港口与航道水文规范[S]. 北京: 人民交通出版社, 2015.
- Ministry of Communications of the People's Republic of China. JTS 145-2015 Code of hydrology for harbour and waterway[S]. Beijing: People's Communications Press, 2015.
- [13] GUMBEL E J. *Statistics of extremes*[M]. New York: Columbia University Press, 1958.
- [14] 罗纯, 王筑娟. Gumbel分布参数估计及在水位资料分析中应用[J]. *应用概率统计*, 2005, 21(2): 169-175.
- LUO C, WANG Z J. The estimates of the parameters of Gumbel distribution and their application to the analysis of the water level data[J]. *Chinese Journal of Applied Probability and Statistics*, 2005, 21(2): 169-175.
- [15] 董双林. Gumbel分布的参数估计方法的统计分析[J]. *水利学报*, 1989(11): 35-42.
- DONG S L. Statistical analysis of parameter estimation methods for Gumbel distribution[J]. *Journal of Hydraulic Engineering*, 1989(11): 35-42.
- [16] YING M, ZHANG W, YU H, et al. An overview of the China Meteorological Administration tropical cyclone database[J]. *Journal of Atmospheric and Oceanic Technology*, 2014, 31(2): 287-301.
- [17] 国家市场监督管理总局, 国家标准化管理委员会. GB/T 14914.2-2019 海洋观测规范 第2部分: 海滨观测[S]. 北京: 中国标准出版社, 2019.
- State Administration of Market Supervision, State Standardization Administration. GB/T 14914.2-2019 The Specification for marine observation-Part 2: Offshore observation[S]. Beijing: China Standards Press, 2019.
- [18] COLES S. *An introduction to statistical modeling of extreme values*[M]. London: Springer, 2001.
- [19] 于福江, 董剑希, 叶琳, 等. 中国风暴潮灾害史料集: 1949-2009[M]. 北京: 海洋出版社, 2015.
- YU F J, DONG J X, YE L, et al. *Collection of storm surge disasters historical data in China 1949-2009*[M]. Beijing: China Ocean Press, 2015.

## Application study of compound extreme value distribution in calculating the return period of typhoon storm surge

FU Xiang, LIANG Sendong, LI Tao, WU Shaohua

*(National Marine Environmental Forecasting Center, Beijing 100081, China)*

**Abstract:** Taking Sansha station as an example, this paper studies the application effect of Poisson-Gumbel compound extreme value distribution in calculating the return period of typhoon storm surge, which considers the conditions of small samples or irregular extreme value samples and introduces the discrete distribution describing the occurrence of random events. The results show that the calculation results of compound extreme value distribution is relatively sensitive to samples, and the low frequency value of the compound distribution of the sample data with small standard deviation is significantly lower. For the short time series observation data with a certain degree of dispersion, the calculation result of the compound distribution is stable, and the estimation of the low-frequency value is also more accurate. For longer time series data, the high frequency occurrence value is more accurate if the threshold value is a lower, and the low frequency occurrence value is more accurate if the threshold value is a higher.

**Key words:** compound extreme value distribution; return period; typhoon storm surge; threshold