

基于平均条件超越率和插值法的风速极值预测

滕寰宇, 叶仁传, 田阿利*, 赵元帅

(江苏科技大学 船舶与海洋工程学院, 江苏 镇江 212003)

摘 要:船舶和海洋结构物长期服役在恶劣的海洋环境中,受到风、浪、流等各种随机载荷甚至极端载荷的作用,而极端载荷将造成不可逆的结构损伤,严重影响海上作业安全。确定船舶与海洋结构物所在海域的极端天气情况,包括重现期内的最大风速或最大浪高等成为结构设计的关键问题之一。基于平均条件超越率(Average Conditional Exceedance Rate, ACER)方法的基本原理,利用蒙特卡罗法产生多组随机风速载荷样本,建立基于极值分布规律的插值计算方法,利用平均超越率和极值的关系,对由蒙特卡罗法产生的模拟随机风速数据进行了重现期为 25, 50 和 100 a 的极值预测。本方法适用于包括非独立同分布在内的任何分布类型数据的极值预测,且通过对比验证了方法的预测精度。

关键词:随机载荷;插值法;ACER;极值预测

中图分类号:P425

文献标识码:A

文章编号:1002-3682(2018)04-0029-06

doi:10.3969/j.issn.1002-3682.2018.04.004

极端天气引发的风载荷和波浪载荷,对使用中的船舶和各种海洋结构物有着十分巨大的破坏作用^[1]。因此,在船舶和各种海洋结构物运营海域,未来数年里可能发生的极端天气预测,特别是风速的精确预测,成为船舶与海洋结构物设计需要考虑的重要因素,开展相关研究具有重要的意义。

国内外学者对海洋极值环境的预报,开展了大量的研究工作。其中刘甜甜等根据极值理论选取了 P-III 分布和 Gumbel 分布求矩适线法,对浙江近海热带气旋极值风速进行了重现期的计算^[2]。郑崇伟等利用 Gumbel 曲线法和 ERA-interim 海浪再分析资料对海上丝绸之路的各个海域每年特定时间里的极值风速和极值浪高做出了预测^[3],为海上风力发电和海浪发电提供科学指导,对解决能源危机和环境危机以及海上丝绸之路的建设有着十分重要的意义^[4-5]。Naess 等提出了连续随机过程中,超越给定阈值的概率的条件近似方程,并在此基础上提出了平均条件超越率(ACER)的概念^[6-7]。以平均条件超越率的概念为基础,Naess 等还发现当随机数据是独立同分布的时候,将 Gumbel 方程 $P = \exp[-\exp(-a(x-b))]$ 对数化,得到 $\ln P = -\exp[-a(x-b)]$ 并加入 2 个新参数 c 和 q ,得到 $\ln P = -q \cdot \exp[-a(x-b)^c]$ 作为平均条件超越率(ACER)的优化方程,新方程可以较好的拟合平均条件超越率尾部数据点的分布,由此得到一种新的极值预测方法,即平均条件超越率(ACER)方法,并用此方法成功预测了挪威附近一些海域 100 a 的极值风速^[8]。Wei 等成功利用 ACER 方法随机船冰载荷的极值进行了预测^[9]。当随机数据并没有明显的分布特征时,ACER 尾部数据分布会和另一个更为复杂的方程近似^[8],在数据分布更接近独立同分布时,此方程的优化效果不及改进后的 Gumbel 方程。要使用此方法得到更为精确的结果,就必须确定随机数据的分布形式。

本研究试图在未知随机数据详细的极值分布情况下的,获得一种对各类数据均通用的有效极值预测方法,基于三次样条插值理论和平均衰减率,采用蒙特卡罗法模拟的随机风速数据,通过在 ACER 图形上进行

收稿日期:2018-03-13

作者简介:滕寰宇(1993-),男,硕士,主要从事船舶与海洋结构物设计与制造方面研究。E-mail:790798463@qq.com

* 通讯作者:田阿利(1980-),女,副教授,博士,主要从事船舶与海洋结构物可靠性、力学性能等方面研究。Email:tianali@just.edu.cn

(李 燕 编辑)

插值计算,结合平均衰减率对尾部数据的优化进行了改进,建立一种通用的 ACER 方程尾部数据的优化方法,该方法具有对数据极值分布限制少,运算简单,结果可靠等优势。通过与 Naess 等^[8]中的 2 种优化方程结果进行对比,验证了本文方法的可行性和预测精度。

1 平均条件超越率

假设已知的随机过程 $X(t)$, 在时间 $(0, T)$ 里, 被观测到有限的 N 次, 观测时间依次为 $t_1, \dots, t_N$, 每个时间点观测到的值记为 $X_1, \dots, X_N$, 该值可为某时刻的点值, 也可为某记值周期内的平均值。 $X_j (j = 1, \dots, N)$ 令为 $X(t_j)$ 的函数, 即 $X_j = X(t_j)$, 则每一个 X_j 都是一个随机变量。现在要确定极值 $\hat{X}_N = \max\{X_j\}$ ($J = 1, \dots, N$) 的分布方程 $F(\eta) = P(\hat{X}_N \leq \eta)$ 以及每个额定值 η 所对应的概率值^[6,10-11]。

$$\begin{cases} F(\eta) \approx F_k(\eta) = \exp\left[-\sum_{j=k}^N \alpha_{kj}(\eta)\right] & k = 1, 2, \dots \\ \epsilon_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^N \alpha_{kj}(\eta) & k = 1, 2, \dots \end{cases}, \quad (1)$$

式中, $\epsilon_k(\eta)$ 为 k 阶平均条件超越率在一定重现期里的极值, $\sum_{j=k}^N \alpha_{kj}(\eta)$ 为前 $k-1$ 个没有超过阈值时第 k 个超过阈值的有效数学期望。

T 年重现期由下式确定^[12]:

$$\epsilon_k(\eta^{T_{yr}}) = -\frac{n_y}{(N-k+1)} \ln\left(1 - \frac{1}{T}\right), \quad (2)$$

式中, n_y 为观测的年数, T 为重现期, 单位是 a; N 为待处理的数据总数。

95% 置信区间表达式为^[8]

$$CI^\pm(\eta) = \epsilon_k(\eta) \cdot \left(1 \pm \frac{1.96}{\sqrt{(N-k+1) \cdot \epsilon_k(\eta)}}\right). \quad (3)$$

2 ACER 模拟随机风速极值预测

2.1 模拟数据极值理论计算

若每年有 100 个观测值, 则 20 a 的随机风速共有 2 000 个观测值。假设潜在的随机过程 $X(t)$ 是平稳的、平均值为 0、标准差为 1 的高斯分布, 均值为 0 的超越率 $\nu^+(0)$, 且 $\nu^+(0)T = 10^3$, $T = 1$ a, 这是典型的风速过程^[13]。随机过程 $X(t)$ 的年极值的分布方程可以用式(4)表示^[10], 式(4)为

$$F^{1_{yr}}(\eta) = \exp[-\nu^+(\eta)t] = \exp\left[-\nu^+(\eta) \exp\left(-\frac{\eta^2}{2}\right)\right] = \exp\left[-10^3 \exp\left(-\frac{\eta^2}{2}\right)\right], \quad (4)$$

式中, $t = 1$ a; $\nu^+(\eta)$ 是每年的平均超越率; η 是风速; 重现期为 T_{yr} 年的预测值 $\eta^{T_{yr}}$ 可以用方程 $F^{1_{yr}}(\eta^{T_{yr}}) = 1 - 1/T$ 的解表示。因此, 重现期为 100 a 的极值为 $\eta^{100_{yr}} = 4.80$ m/s。同理, 重现期为 50 和 25 a 的极值分别为 $\eta^{50_{yr}} = 4.65$ m/s, $\eta^{25_{yr}} = 4.49$ m/s。

2.2 基于插值法的 ACER

根据式(2)预估 25, 50, 100 a 风速极值的重现期水平分别为 4×10^{-4} , 2×10^{-4} 和 1×10^{-4} 。图 1 为考虑每一个超越阈值, 即 $k = 1$ 时, 其中一组模拟数据的 ACER 图形, 图中纵坐标为超越概率, 横坐标为极值风速预测值, 图中曲线的每个点代表阈值确定时, 超越阈值的概率。

由于不同的概率分布模型尾部特性不同,所用拟合概率模型不同,引起的长期极值预测也存在一定差异。从图 1 中可以发现,基于 ACER 法的尾部超越率表现为非连续线段较为离散,其主要原因是较大的风速值出现的次数较少,当阈值在 2 个相邻的较大值之间变化时,造成在 ACER 曲线上为离散的线段,如图 1 尾部数据。由于线段所表示的少数较大的值,与之前的极值变化规律有较大不同,为得到更精确的结果,需要将该部分线段去除。

插值法是一种利用范围十分广泛的方法^[14],在极端气候预测领域也有应用。周北平等^[15]利用包括插值法在内的方法,对长江下游流域 1960—2012 年间的极端气温和降水的特征进行了分析。三次样条插值法既能克服高次插值的缺陷又能保证插值曲线具有一定的光滑性,具体做法是在拼接处一阶和二阶导数连续。本文使用第二类边界条件^[16],给定端点处的二阶导数值: $S'(x_1)=y'_1, S''(x_1)=y''_1$ 。在首尾的特殊情况为 $y''_1=y''_N=0$ 时得到的样条被称为自然三次样条,同时自然样条可能是最光滑的插值样条^[14,17]。

很明显,插值计算过程涵盖了所有的概率点,为顺利完成插值计算,将基于 ACER 法的尾部非连续线段进行处理,对相同的小概率仅保留其中一点数据。由于样条插值的特性,只能较为准确地画出 2 个端点之间的曲线,而由于重现期水平线又要低于 ACER 图形的最低点,所以为了能够将样条曲线与相应重现期的直线相交,需要将样条曲线在插值区间外按照阈值每增加一个单位,平均超越率衰减 μ 的方式将曲线进行延伸。其中,引进的平均超越衰减率 μ 计算式为

$$\mu = \frac{\sum_{i=2}^n [\epsilon_k(\eta_i)/\epsilon_k(\eta_{i-1})]}{n}, \quad (5)$$

式中, n 表示在图 1 中有 n 组阈值和平均超越率,平均超越衰减率至少应精确到千分位。

图 2 所示为经过三次样条插值处理后的 ACER 图形,是图 1 舍弃尾部不够连续的数据后,经过插值计算得到曲线,其中横坐标表示阈值,每个点的纵坐标则表示当阈值为此值时,数据中超过此阈值的概率。由于所用数据中 0~1.5 的值较少,表现在 ACER 图上就是 0~1.5 段离 1.0 很近,随着阈值 η 逐渐增大,平均超越率 $\epsilon_k(\eta) = ACER_k(\eta)$ 逐渐减小,而整段曲线的平均衰减率 $\mu = 0.9958$ 。

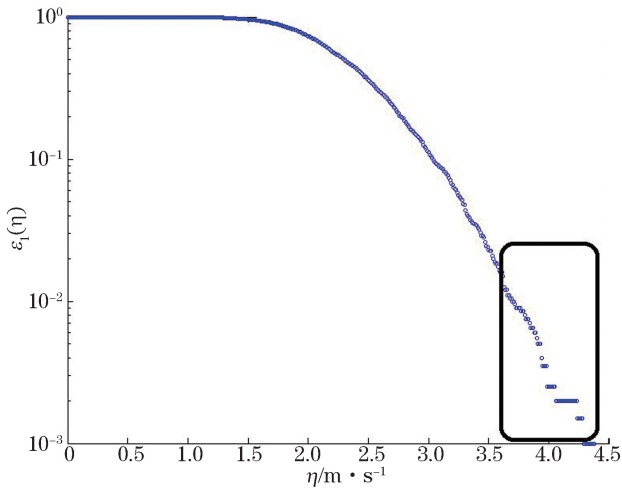


图 1 ACER 图形
Fig.1 ACER graph

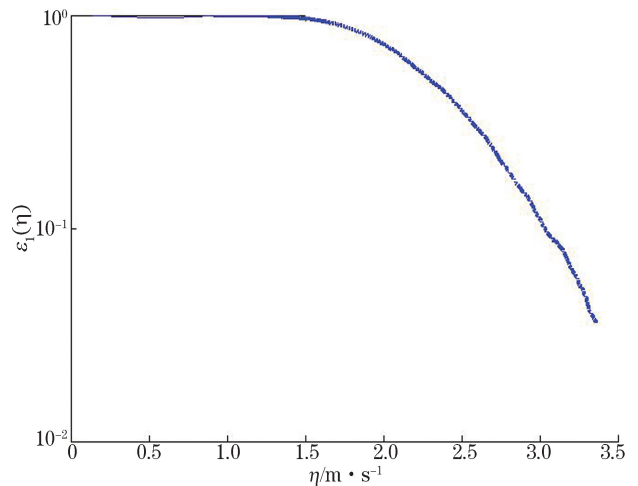


图 2 经过插值处理后的 ACER 图形
Fig.2 The ACER graph after the interpolation

3 预测结果

3.1 单组模拟数据预测结果

假设模拟风速的随机过程为均值为 0,标准差为 1 的标准高斯分布,观测并记录随机风速数据,每年观测 100 次,即每 3.65 d 观测一次,一共有 2 000 个观测数据。由于每 2 次观测间隔 3.65 d,可以认为每个数据都相互独立,所以 $k=1$,即超越阈值的每一个值都代入计算。在保证数据独立性的情况下可以充分利用每一个数据,提高精度。

图 3 是用平均衰减率 $\mu=0.995\ 8$ 将 ACER 图形的衰减规律向下传递,使图形延伸并分别与 25 a 重现期对应的纵坐标为 4×10^{-4} 的水平线,50 a 重现期对应的纵坐标为 2×10^{-4} 水平线以及 100 a 重现期对应的纵坐标为 1×10^{-4} 的水平线相交,交点对应的横坐标值即为预测值。根据图 3,25 a 重现期的极值风速为 4.459 m/s,50 a 重现期的极值风速为 4.619 m/s,100 a 重现期的极值风速为 4.790 m/s。

3.2 多组模拟数据预测结果

采用本文提出的方法,对 100 组模拟随机风速的观测值进行重现期分别为 25,50 和 100 a 的预测结果计算,并与模拟数据的理论计算结果进行对比。平均衰减率根据各组数据情况确定,由于单组数据量较少,容易出现较大的偏差,故采用多组数据结果求平均值的方法降低误差。如表 1 所示,当重现期 $T=25\text{ a}$ 时,本文预测极值风速平均为 4.464 m/s,与理论结果误差为 0.6%;当 $T=50\text{ a}$ 时,本文预测极值平均为 4.631 m/s,与理论结果误差为 0.4%;当 $T=100\text{ a}$ 时,本文预测极值平均为 4.798 m/s,与理论结果误差为 0.04%。通过对比分析结果显示,本文提出的基于三次样条插值法的 ACER 方法误差较小,可以准确预测出相应重现期的极值风速。

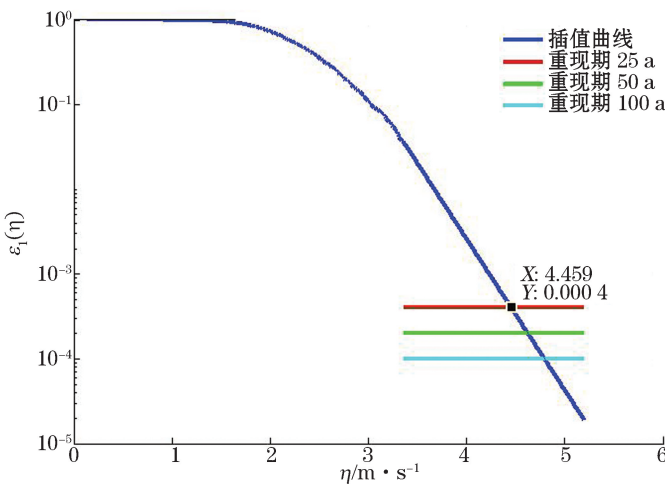


图 3 平均衰减率 $\mu=0.995\ 8$ 时的预计结果

Fig.3 Results of prediction under the average attenuation rate $\mu=0.995\ 8$

表 1 重现期分别为 25,50 和 100 a 的的预估结果 (η)和其 95%置信区间 (CI)

Table 1 The results predicted for the reappearing periods of 25, 50 and 100 a, respectively and CI=95%

序 号	25 a 重现期		50 a 重现期		100 a 重现期	
	$\eta/\text{m}\cdot\text{s}^{-1}$	CI	$\eta/\text{m}\cdot\text{s}^{-1}$	CI	$\eta/\text{m}\cdot\text{s}^{-1}$	CI
1	4.485	[4.392,4.578]	4.622	[4.528,4.715]	4.767	[4.671,4.863]
2	4.494	[4.401,4.587]	4.688	[4.593,4.781]	4.862	[4.765,4.959]
3	4.475	[4.382,4.568]	4.641	[4.547,4.734]	4.806	[4.710,4.902]
4	4.456	[4.363,4.549]	4.617	[4.523,4.710]	4.785	[4.689,4.881]
5	4.452	[4.360,4.544]	4.619	[4.525,4.712]	4.786	[4.690,4.882]
6	4.421	[4.329,4.513]	4.587	[4.493,4.679]	4.749	[4.653,4.845]
7	4.483	[4.390,4.576]	4.658	[4.563,4.751]	4.836	[4.740,4.932]

续表						
序 号	25 a 重现期		50 a 重现期		100 a 重现期	
	$\eta/\text{m}\cdot\text{s}^{-1}$	CI	$\eta/\text{m}\cdot\text{s}^{-1}$	CI	$\eta/\text{m}\cdot\text{s}^{-1}$	CI
8	4.468	[4.375,4.561]	4.645	[4.551,4.738]	4.822	[4.726,4.918]
9	4.448	[4.356,4.540]	4.614	[4.520,4.706]	4.780	[4.684,4.876]
10	4.459	[4.366,4.552]	4.619	[4.525,4.711]	4.790	[4.702,4.894]
均值	4.464	[4.372,4.557]	4.631	[4.537,4.724]	4.798	[4.703,4.895]
理论值	4.49	—	4.65	—	4.80	—
误差	0.6%	—	0.4%	—	0.04%	—

注:“—”处表示无数据

4 结 语

针对 ACER 概念开展的极值预测方法存在对数据分布类型较为严格的问题,基于三次样条插值法和平均衰减率的概念,建立了新的极值预测方法,此方法几乎对任何类型的随机数据都能有比较好的计算结果,对数据分布类型要求很低。可以得出结论:

1) 针对数据较少和数据为非独立同分布的情况,提出了将三次样条插值法和平均条件超越率(ACER)相结合的方法,用于预测一定重现期里的风速等随机数据。该方法将对数据的要求降到最低,同时计算难度相对较小。

2) 利用三次样条插值与平均条件超越率(ACER)相结合的方法,进行重现期分别为 25,50 和 100 a 的极值预测,结果与理论结果吻合较好,证明了该方法的精确性和稳定性。

参考文献(References):

[1] KARPA O. Development of bivariate extreme value distributions for applications in marine technology[D]. Norway, Trondheim: Norwegian University of Science and Technology, 2015: 1-3.

[2] LIU T T, WANG Y H, PEI J F, et al. Extreme wind speed and return period of cyclones passing through Zhejiang coastal area[J]. Advances in Marine Science, 2017, 35 (1): 107-116. 刘甜甜, 汪一航, 裴军峰, 等. 浙江近海热带气旋极值风速统计及重现期分析[J]. 海洋科学进展, 2017, 35(1): 107-116.

[3] ZHENG C W, GAO Z S, ZHANG Y, et al. Marine environment characteristics of maritime silk road in the 21st century: Extreme wind speed and extreme wave height[J]. Ocean Development and Management, 2015, 32 (11): 5-9. 郑崇伟, 高占胜, 张雨, 等. 经略 21 世纪海上丝路之海洋环境特征: 极值风速和极值波高[J]. 海洋开发与管理, 2015, 32 (11): 5-9.

[4] ZHENG C W, LI C Y. Development of the islands and reefs in the South China Sea: wind power and wave power generation[J]. Periodical of Ocean University of China, 2015, 45(9): 1-6. 郑崇伟, 李崇银. 中国南海岛礁建设: 重点岛礁的风候、波候特征分析[J]. 中国海洋大学学报(自然科学版), 2015, 46(9): 1-6.

[5] ZHENG C W, LI C Y. Overview of site selection difficulties foe marine new energy power plant and suggestions: wave energy case study [J]. Journal of Harbin Engineering University, 2018, 39(2): 200-206. 郑崇伟, 李崇银. 关于海洋新能源选址的难点及对策建议——以波浪能为例[J]. 哈尔滨工程大学学报, 2018, 39(2): 200-206.

[6] NAESS A, GAIDAI O. Monte Carlo methods for estimating the extreme response of dynamical systems[J]. Journal of Engineering Mechanics, 2008, 134(8): 628-636.

[7] NAESS A, GAIDAI O, HAVER S. Efficient estimation of extreme response of drag-dominated off shore structures by Monte Carlo simulation[J]. Ocean Engineering, 2007, 34(16): 2188-2197.

[8] NAESS A, GAIDAI O, KARPA O. Estimation of extreme values by the average conditional exceedance rate method[J]. Journal of Probability and Statistics, 2013(3): 207-221.

- [9] WEI C, LEIRA B J, NAESS A. Probabilistic methods for estimation of the extreme value statistics of ship ice loads[J]. *Cold Regions Science and Technology*, 2018, 146: 87-97.
- [10] NAESS A, MOAN T. *Stochastic Dynamics of Marine Structures*[M]. New York: Cambridge University Press, 2012: 349-353.
- [11] NAESS A. Technical notes: On the long-term statistics of extremes[J]. *Applied Ocean Research*, 1984, 6(4): 227-228.
- [12] NAESS A, GAIDAI O, BATSEVYCH O. Prediction of extreme response statistics of narrow-band random vibrations[J]. *Journal of Engineering Mechanics*, 2010, 136(3): 290-298.
- [13] NAESS A, CLAUSEN P. Combination of the peaks-over-threshold and bootstrapping methods for extreme value prediction[J]. *Structural Safety*, 2001, 23(4): 315-330.
- [14] KINCAID D, CHENEY W. *Numerical Analysis*[M]. Beijing: China Machine Press, 2005: 245-284. KINCAID D, CHENEY W. 数值分析[M]. 王国荣, 俞耀明, 徐兆亮, 译. 北京: 机械工业出版社, 2005: 245-284.
- [15] ZHOU B P, XUE H X, GOU S, et al. Characteristics of extreme climate in downstream catchment of Yangtze River during 1960-2012[J]. *Water Power*, 2017, 43(9): 26-30. 周北平, 薛华星, 苟尚, 等. 1960年~2012年长江下游流域气候变化特征分析[J]. *水力发电*, 2017, 43(9): 26-30.
- [16] LIAN D Z, WU W C, YOU D Y, et al. A new calculation method of cubic spline interpolation function value[J]. *Journal of Longyan University*, 2017, 35(5): 4-7. 连德忠, 吴文城, 游德有, 等. 三次样条插值的新算法[J]. *龙岩学院学报*, 2017, 35(5): 4-7.
- [17] ZHANG G C, ZHANG L. *Practical Numerical Analysis*(The second edition)[M]. 2nd ed. Sichuan: Sichuan University Press, 2001: 128-132. 张广澄, 张雷. 实用数值分析[M]. 2版. 四川: 四川大学出版社, 2001: 128-132.

Prediction of Wind Speed Extremum Based on Average Conditional Exceedance Rate and Interpolation Method

TENG Huan-yu, YE Ren-chuan, TIAN A-li, ZHAO Yuan-shuai

(School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology,
Zhenjiang 212003, China)

Abstract: As servicing long in a harsh marine environment, the watercrafts and marine structures are subjected to the actions of various random loads such as winds, waves and currents and even extreme loads. The extreme loads will cause irreversible structural damage and seriously affect the safety of maritime operations. Therefore, it is one of the keys of structural design to determine the extreme weather conditions including the extreme wind speed and extreme wave height in the sea area where ships and marine structures are located. According to the fundamental principle of average conditional exceedance rate (ACER), an interpolation calculation method is established for extreme value distribution by using multi-groups of random wind load samples obtained by means of Monte Carlo method. Based on the relationship between the average exceedance rate and the extreme value, the prediction of extreme values with reappearing periods of 25a, 50a and 100a is carried out for the simulated random wind speed data produced by using the Monte Carlo method. The method presented in the paper can be applied to the extreme value prediction of the data of any distribution types including the non-independent and identically distributed data, and its prediction accuracy has been verified through comparison.

Key words: random load; interpolation method; average conditional exceedance rate; extreme value prediction

Received: March 13, 2018