

曹诗嘉, 侯静惟, 方伟华, 等. 基于站点实测风速的中国台风大风重现期估计[J]. 灾害学, 2019, 34(1): 229–234. [CAO Shijia, HOU Jingwei, FANG Weihua, et al. Estimation on the frequency of typhoon wind speed over China based on historical wind observations from weather stations[J]. Journal of Catastrophology, 2019, 34(1): 229–234. doi: 10.3969/j.issn.1000-811X.2019.01.041.]

基于站点实测风速的中国台风大风重现期估计^{*}

曹诗嘉^{1,2}, 侯静惟^{1,2}, 方伟华^{1,2}, 林伟^{1,2}

(1. 北京师范大学地理科学学部环境演变与自然灾害教育部重点实验室, 北京 100875;

2. 北京师范大学地理科学学部减灾与应急管理研究院, 北京 100875)

摘要: 首先基于 1951–2014 年中国地面气象站点历史风速观测数据和 CMA-STI 热带气旋最佳路径数据集, 提取了站点历史台风过程影响风速; 其次, 利用极大似然法估计 Gumbel 分布、Frechet 分布以及 Weibull 3 种分布模型参数, 优选出最适模型进行站点极值风速年最大值风速拟合, 计算不同重现期风速值并利用 bootstrap 方法对各重现期计算结果的不确定性进行量化, 得到典型重现期下中国台风大风风速的 90% 置信区间; 最后, 根据包括克里金法、反距离权重法、自然邻域法、最近邻法在内的 4 种空间插值方法对于台风大风的插值效果得到中国东部沿海 1km 网格不同重现期台风风速分布及其 90% 置信区间。结果表明: 东南沿海地区是台风极值大风影响最严重的区域, 典型重现期台风风速估计的不确定性较大, 不确定性的高值区域也主要分布在东南沿海地区。

关键词: 台风大风; 极值分布; 危险性评估; 不确定性量化; 空间插值

中图分类号: X43; P44; P49 **文献标志码:** A **文章编号:** 1000-811X(2019)01-0229-06

doi: 10.3969/j.issn.1000-811X.2019.01.041

中国是遭受台风灾害最严重的国家之一, 常因台风造成严重的人员伤亡和巨大的经济损失。例如, 2006 年第 8 号台风“桑美”在浙江省苍南县登陆, 引起了狂风巨浪, 共造成 483 人死亡, 千余条船沉没, 直接经济损失达 196.5 亿元人民币^[1], 2014 年 10 号台风“威马逊”登陆风速达到 17 级, 造成海南省 216 个乡镇(街道)受灾, 受灾人口 325.8 万人, 直接经济损失 108.28 亿元^[2]。近年来我国沿海台风易发区经济快速发展, 对于台风致灾因子的暴露程度不断增加; 而沿海桥梁、铁路、石油石化以及核电站等重大工程的选址及设计需要对局地风速进行评估。因此, 面向台风灾害风险管理及工程设计需求, 采取有效手段对中国台风大风危险性进行评估具有重要意义。

目前学界对于台风大风危险性评估方法主要有三类, 其中第一类一般通过极值分布模型对历史台风风速序列进行模拟从而计算单点的风速重现期; 第二类通常针对某个区域, 通过统计该区域历史台风参数概率分布, 并结合风场模型和极值分布模型计算该小区域的风速重现期^[3]; 第三

类一般采用全路径模拟的方法生成大样本台风随机事件, 然后结合风场模型和极值分布模型计算台风影响区域的风速重现期^[3]。与第二类和第三类方法相比, 第一类方法在历史风速观测时间序列较长时, 可充分利用历史样本信息得到较为准确的风速重现期计算结果, 其缺点在于当历史数据不足时, 重现期计算结果存在一定不确定性。国外利用第一种方法即历史观测数据进行台风大风危险性评估采用的数据类型主要包括台风近中心历史风速观测数据及气象站点历史风速观测数据两类。一些学者利用台风近中心历史风速观测数据采用韦伯分布(Weibull Distribution)、广义帕累托分布(General Pareto Distribution, GPD)等极值分布函数对美国区域最大风速进行拟合, 并计算了一定空间尺度的年发生超越概率^[4,5], 此类方法可以满足台风综合危险性评估需要, 但其不足之处在于台风近中心历史风速数据无法反映整个区域的空间差异性, 因此不适宜用于大区域范围台风灾害危险性评估; 另外一些学者采用气象站点历史风速观测数据, 同样利用极值模型对强风危

* 收稿日期: 2018-06-09 修回日期: 2018-08-14

基金项目: 国家重点研发计划课题“多灾种重大自然灾害情景构建与动态模拟技术”(2018YFC1508803); 国家重点研发计划重点专项第 3 子课题“海岸带和沿海地区综合风险评估关键技术研究”(2017YFA0604903)

第一作者简介: 曹诗嘉(1990-), 女, 辽宁丹东人, 博士研究生, 主要从事台风灾害损失评估研究。

E-mail: shijia.cao@mail.bnu.edu.cn

通讯作者: 方伟华(1973-), 男, 湖北嘉鱼人, 教授, 主要从事自然灾害风险评估研究。E-mail: weihua.fang@bnu.edu.cn

险性进行评估^[3], 此类方法可以较好地反应风速分布的空间差异性, 缺点在于气象站点历史风速观测数据往往存在数据量不足的问题。

中国大风危险性研究多是针对小区域或台站进行的不同分布模型对比研究或大风重现期推算实证研究^[6-9], 如综合岛屿站观测、台风记录、船舶报告以及数值模式计算风速, 组成大风年最大值序列, 利用矩估计法拟合大风序列, 通过检验选定最适风速序列及分布模型, 绘制中国近海 50 年一遇和 100 年一遇大风极值等值线^[10], 此类研究的缺点在于面向区域较小, 且缺少对评估结果不确定性的讨论。利用气象站点历史风速观测数据计算重现期, 数据不足、数据一致性问题、测量误差是结果不确定性的 3 个最主要来源。为了提高结果可靠性, 需要对不确定性进行量化, 目前国内已有研究采用广义极值分布模型, 利用极大似然法及渐进分布理论推导出不同年遇水平设计风速和一定置信度下的置信区间, 并通过气象站点年最大风速资料进行实证研究^[11], 该研究对单个站点重现期计算结果的不确定性进行了刻画, 对于工程的选址设计具有一定意义, 而为了反映整个中国东部沿海台风大风危险性分布特征及不确定性大小, 还需对站点计算结果进行插值, 从而得到整个空间连续的重现期评估结果。

本文的研究目标为利用气象站点历史风速观测资料, 评估不同重现期下中国东部台风风速空间分布并给出一定置信度下的风速置信区间。具体过程为, 首先, 基于中国地面气象站点历史风速观测数据, 提取每个站点历史台风过程影响风速的年最大值; 其次, 针对每个站点利用极大似然法估计 3 种经典极值分布模型的参数估计值, 分别选择最适模型进行站点台风年最大值风速拟合; 然后, 利用 bootstrap 重采样方法, 得到 90% 置信度下的风速置信区间; 最后, 从 4 种空间插值方法中选择最优者插值得到不同重现期及 90% 置信度下的中国台风大风空间分布并对中国台风大风危险性区域分布及不确定性特征进行分析。

1 数据

1.1 地面气象站点观测数据及研究区

本文采用的风速资料来自中国 756 个一般气象观测站 1951–2014 年日值风速观测值, 指标为日极大风速(3 s 瞬时风速的日最大值)。由于中国东部沿海地区是受台风影响的主要地区, 因此选取中国东部、中部、南部 13 个省、直辖市、自治区作为本文的研究区, 考虑到研究区边缘插值风速的准确性及连续性, 故保留中国东部、中部、南部 21 个省、直辖市、自治区的 359 个一般气象观测站的气象观测数据, 359 个一般气象观测站空间分布如图 1 所示。

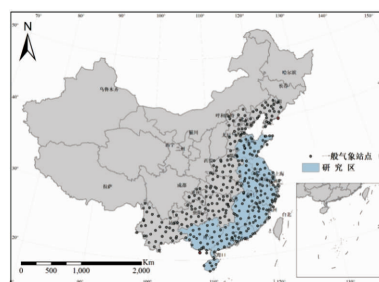


图 1 研究区地面一般气象观测站点分布图

1.2 台风路径数据

本文采用中国气象局上海台风研究所(CMA-STI)整编的西北太平洋热带气旋最佳路径数据^[12], 包括 1949–2014 年共 2182 场台风, 其中登陆中国台风共计 611 场。数据具体指标包括: 国际热带气旋编号、中国热带气旋编号、英文名称、每个台风路径点的年、月、日、时、等级、经度、纬度、近中心最低气压以及最大持续风速等, 记录时间间隔为 6 h。

2 方法

2.1 风速提取

台风大风重现期计算的前提是提取每个气象站点受台风影响风速, 其提取方法简述如下: ①基于站点经纬度信息, 提取一定空间范围内历史台风的路径点; ②获得历史台风对该站点影响的起止时间, 并提取该时间范围对应的站点观测风速, 作为历史台风对该站点的影响风速。这种方法的关键在于台风影响范围的界定, 由于台风是由比较均匀的热带海洋气团发展起来, 因此台风气压场、风场分布具有一定的对称性, 可近似将台风看作圆对称的涡旋, 其半径变化范围小到上百公里, 大到上千公里^[13]。可基于台风中心位置, 经验地设定台风影响半径从而划定影响范围。若半径设定太小, 则无法获取台风外围大风信息, 可能导致获取的格点风速样本不足; 若半径设定过大, 虽可获得完整的台风大风序列, 但可能将其他天气系统引起的局地大风误认为台风大风^[14], 在本研究中, 经验性地将台风影响范围设定为距台风中心 500 km。

利用上述气象站点台风影响风速提取方法, 对 359 个气象站点的历史台风影响风速进行提取, 对部分历史样本小于 15 个的站点予以剔除, 最终得到风速数据提取结果。图 2 为站点历史台风过程影响风速最大值统计图, 从图中可以看出, 台风风速样本数超过 15 个的站点共有 275 个, 且空间分布相对较广泛, 其最大值出现在浙江大陈岛站, 为 59.5 m/s, 站点台风风速高值区主要分布在东南沿海地区, 部分内陆地区也出现了风速高值, 如山东泰山站及安徽黄山站, 体现了地形对于风

速的影响作用。

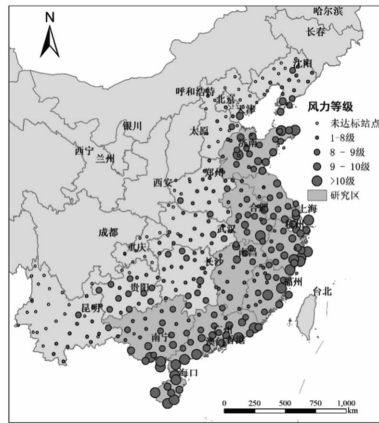


图2 359个一般气象观测站历史台风风速过程最大值

2.2 风速空间插值

空间插值方法主要包括点插值和面插值两种,已知某点数据推求空间区域内任一点数值应采用点插值方法进行插值^[15],点插值方法又包括了克里金法、多项式回归法、最近邻法、反距离权重法等。为确定最适用中国东部台风大风风速插值的方法,以2.1中得到的站点历史台风风速最大值作为检验样本,选用克里金法、反距离权重法、自然邻域法以及最近邻法4种点插值方法对站点风速进行空间插值,然后通过交叉验证法确定最适插值方法。各插值模型主要参数设置如下:最大搜索半径设为500 km,搜索范围内的最小样本数设为15,其中反距离权重法采用固定搜索半径,幂参数设为3,克里金插值法的半变异函数采用指数模型。

具体计算方法为对站点*i*,利用其他站点的记录,选用一种方法计算该站点的插值风速,并将其与该站点的实测值进行比较,循环计算所有站点的均方根误差(*RMSE*)以及决定系数(*R*²),以所有站点*RMSE*及*R*²的均值作为插值效果比较的判定标准,均方根误差及决定系数的公式分别如式(1)

及式(2)所示,其中,*n*为插值站点样本量,*V*_{obs}为站点风速实测值,*V*_{int}为站点风速插值结果, \bar{V}_{int} 为站点风速插值结果的均值。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_{obs} - V_{int})^2}; \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (V_{obs} - V_{int})^2}{\sum_{i=1}^n (V_{int} - \bar{V}_{int})^2} \quad (2)$$

利用达标站点台风过程风速的最大值的统计结果作为检验样本,比较4种空间插值方法模拟结果的交叉检验误差,如表1所示。可以看出,对于台风风速的插值,克里金插值法效果最好,站点平均*RMSE*为6.45,*R*²达到了0.41。

表1 4种站点风速空间插值方法交叉验证比较

插值方法	风速/(m/s)	
	均方根误差(<i>RMSE</i>)	决定系数(<i>R</i> ²)
克里金法	6.45	0.41
反距离权重法	6.59	0.24
自然邻域法	12.64	0.31
最近邻法	7.01	0.34

2.3 极值大风重现期估计及不确定性量化

气候统计学中,通常采用经典极值理论对气象要素极值进行拟合^[16],经典极值理论包括Gumbel模型,Fréchet模型以及Weibull模型3种,其累积概率分布函数分别如式(3)~式(5)所示,其中μ为位置参数,σ为尺度参数,α为形状参数,*x*为连续型随机变量,超过定值*x*则表示极端事件发生。根据历史样本年最大值分布计算极端事件的重现期(Return Period, RP)是极值统计最重要的应用之一,由式(6)可知每一个重现期对应一个极值分位数,表示极端事件的极值变量的数值大小。同时,对于给定重现期,极值分位数越大说明超越概率越小,则极端事件发生的可能性也就越小。

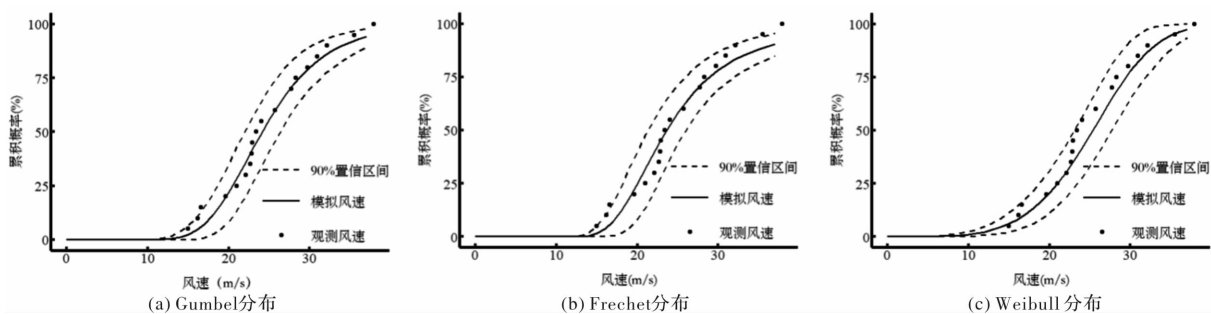


图3 (No. 59758)海口站3种极值函数拟合下的台风年极值风速累积概率及90%置信区间

表2 (No. 59758)海口站不同极值分布拟合参数及拟合效果检验比较

风速指标	极值模型	μ	σ	α	<i>D</i> 统计量	<i>P</i> 值	<i>RMSE</i>
极大风速	Gumbel	20.9313	7.3076	/	0.0946	0.6940	0.0353
	Fréchet	0	19.6026	2.7552	0.1375	0.2457	0.0552
	Weibull	/	27.9487	3.2481	0.0861	0.7951	0.0395

对于每一个气象站点,利用年极值抽样方法,选择该站点历史台风影响风速的极大值作为极值拟合的样本,然后从3种极值分布模型中选择最适者进行极值拟合,得到模拟风速的期望值(即位于所有风速估计值中50%分位数的风速值)并利用 bootstrap 方法得到90%置信度下的风速区间。

$$\text{Gumbel 分布: } F(x) = \exp\{-\exp[-(\frac{x-\mu}{\sigma})]\}; \quad (3)$$

$$\text{Frechet 分布: } F(x) = \exp\{-(\frac{x-\mu}{\sigma})^{-\alpha}\}, \mu=0, x>\mu; \quad (4)$$

$$\text{Weibull 分布: } F(x) = 1 - \exp\{-(\frac{x}{\sigma})^\alpha\}, x \geq 0; \quad (5)$$

$$RP = \frac{1}{1 - F(x)}. \quad (6)$$

为定量刻画不同极值模型的拟合效果,①采用 Kolmogorov-Smirnov (KS) 检验方法评价拟合结果与样本概率分布的一致性。KS 检验根据样本得到经验分布与假设分布函数之间的偏差,描述两个独立统计样本相似程度^[17-18]。统计量 D 及其对应的 P 值可用于表征模拟分布与样本分布的相似性, P 值越高,则相似性越大;②针对每个站点,计算每个模型模拟 CDF 与经验 CDF 的均方根误差 RMSE,用于协助判断拟合效果。RMSE 越小模拟平均误差越小,若出现不同模型的 KS 检验结果相

同的情况,则选择 RMSE 最小者作为最适模型。

以海口站点为例,利用上述站点大风重现期及不确定性估计方法,得到3种极值模型模拟下的台风年极值风速累积概率分布90%置信区间与观测值的对比图。由图3可知海口站点采用 Weibull 分布函数的拟合效果较好,观测值均位于模拟风速的90%置信区间内,特别是对于样本累积概率分布尾部的拟合效果较好。Gumbel 函数及 Frechet 函数的拟合效果较差,样本累积概率分布尾部超出了模拟风速的90%置信区间。

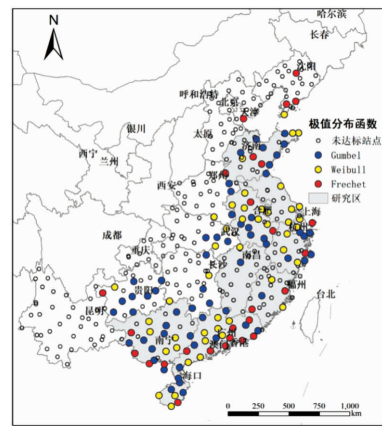


图4 台风极大风速数据达标站点及其最适极值分布类型

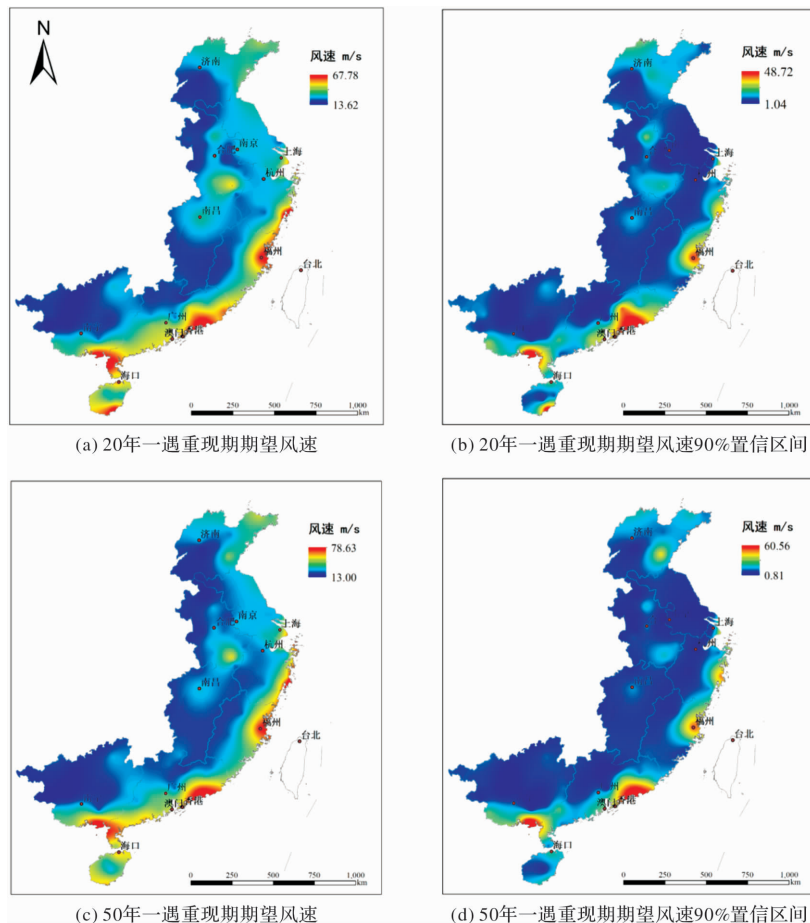


图5 不同重现期及90%置信度的中国台风极大风速空间分布

表2给出了利用三种极值函数对海口站台风风速年最大值进行拟合得到的模型参数^[19],以及对应的KS检验及RMSE结果。可以看出,对于海口站年极值风速拟合,采用Weibull分布的D统计量最小、P统计量最大、RMSE结果与Gumbel模型结果相差较小,因此对海口站点台风年极大风速适合采用Weibull函数进行拟合。其他站点最适极值模型的确定方法与海口站点相同。

3 结果

对359个气象站点,利用2.3中的站点大风重现期估计方法,提取台风影响期间站点风速的年最大值作为极值拟合的样本,利用Gumbel、Frechet以及Weibull函数进行拟合。计算过程中,为了减小因数据量不足导致的参数估计的不确定性及满足极值分布函数对拟合样本数量的要求,对极值拟合样本数少于15个的站点予以剔除,最终得到台风极大风速数据达标站点及其对应的最适极值分布(如图4所示)。可以看出:①359个站点中共计133个站点满足拟合样本数量需求,最适极值分布确定为Gumbel、Frechet以及Weibull的站点个数分别为60个,27个以及46个,即对于不同站点的风速极值,应采用不同极值分布函数进行拟合;②359个站点中位于研究区外符合样本数量要求的站点较少,这是因为这些区域的站点受台风影响频次较少,影响时长较短;③359个站点中位于研究区内江西省、福建省的站点满足拟合样本数量要求的数量偏少,江西18个站点中仅3个站点满足要求,福建省22个站点中仅5个站点满足要求,造成这一结果的主要原因是尽管此区域台风影响频次较多,影响时间较长,但未达标站点历史台风观测数据较少。

利用各达标站点最适极值分布模型,分别计算各站点20年一遇及50年一遇的台风风速期望值及90%置信区间;然后,利用克里金法插值得到中国1 km网格分辨率的20年一遇及50年一遇重现期风速期望值及90%置信区间,计算结果如图5所示。从图5可以看出:①台风极值大风高值区主要分布在我国东部地区,尤其是东南沿海地区,该区域是我国历史上受台风灾害影响最严重的地区,发生台风大风极端事件的可能性也最大;②部分内陆地区也出现了风速高值区,主要原因是由于历史台风影响的观测风速样本不足;③典型重现期台风风速估计的不确定性较大,20年一遇的台风风速90%置信区间约49 m/s,50年一遇的台风风速90%置信区间近61 m/s,置信区间的高值区主要分布在中国东南沿海,置信区间整体由东南沿海向内陆递减。

4 结论与讨论

本文基于中国地面气象站点历史风速观测数据,提取了每个站点历史台风过程影响风速,在此基础上优选出最适极值分布以及空间插值模型,计算了不同重现期下,中国东部地区空间上连续分布的台风风速及90%置信度下的风速区间,并对其区域分布特征及不确定性进行了分析。主要结论如下:

(1)采用交叉检验的方法比较了4种空间插值方法对于站点极大风速的插值总误差,结果表明,对于站点台风极大风速插值,克里金法插值效果最优。

(2)由于历史样本数量不足、数据一致性问题及测量误差等原因,基于站点历史风速观测数据的台风大风重现期估计结果不确定性较大,为了保证评估结果的可靠性,需要对其进行不确定性量化。

(3)结合极值分布模型以及空间插值方法评估中国台风大风危险性,相较于传统小区域或台站尺度的评估有一定改进,但仍存在一些局限性,包括:①气象站点历史风速观测样本不足可能影响极值风速评估精度及区域插值效果;②本文经验地将台风影响范围确定为500公里,主观性较强,今后的研究可通过敏感性分析方法对台风影响最适范围进行评估^[14];③由于台风登陆后下垫面情况比较复杂,局地地形以及地表粗糙度的变化对于近地表风速影响较大,仅依靠空间插值的方法难以精确刻画该影响机制。后续的台风大风危险性研究可基于大样本台风路径事件集,结合台风风场模型得到空间上连续分布并且样本充足的台风风速序列^[7-9],在此基础上进行不同重现期极值风速评估将更具可靠性^[19]。

参考文献:

- [1] 于玉斌. 中国近海热带气旋强度突变研究[M]. 北京: 气象出版社, 2012.
- [2] 况昌勋. 救灾资金, 省政府再拨6亿[N]. 海南日报, 2014-07-22(001).
- [3] 方伟华, 林伟. 面向灾害风险评估的台风风场模型研究综述[J]. 地理科学进展. 2013, 32(6): 852-867.
- [4] Jagger T, Elsner J B, Niu X. A Dynamic probability model of hurricane winds in coastal counties of the united states[J]. Journal of Applied Meteorology. 2001, 40(5): 853-863.
- [5] Elsner J B, Jagger T H, Liu K B. Comparison of hurricane return levels using historical and geological records[J]. Journal of Applied Meteorology & Climatology. 2008, 47(2): 368-374.
- [6] 郑运霞, 应明, 李永平. 影响广东省中部沿海热带气旋大风气候特征和重现期分析[J]. 海洋与湖沼. 2014, 45(1): 157-166.
- [7] 史军, 徐家良, 谈建国, 等. 上海地区不同重现期的风速估

- 算研究[J]. 地理科学. 2015, 35(9): 1191–1197.
- [8] 马玉峰, 宋进华, 高春香. 内蒙古东部电网最大风速及其重现期极值分布特征[J]. 气象科技. 2015, 43(1): 162–167.
- [9] 阎俊岳, 黄爱芬. 中国近海大风极值计算方法研究[J]. 气象学报. 1996, 54(2): 233–239.
- [10] 阎俊岳, 黄爱芬. 中国近海大风极值计算方法研究[J]. 气象学报. 1996, 54(2): 233–239.
- [11] 刘聪, 秦伟良, 江志红. 基于广义极值分布的设计基本风速及其置信限计算[J]. 东南大学学报(自然科学版). 2006, 36(2): 331–334.
- [12] Ying M, Zhang W, Yu H, et al. An Overview of the China Meteorological Administration Tropical Cyclone Database[J]. Journal of Atmospheric & Oceanic Technology. 2014, 31(2): 287–301.
- [13] 陈联寿. 西太平洋台风概论[M]. 北京: 科学出版社, 1979.
- [14] 任福民, Byron, Gleason. 一种识别热带气旋降水的数值方法[J]. 热带气象学报. 2001, 17(3): 308–313.
- [15] 朱求安, 张万昌, 余钧辉. 基于 GIS 的空间插值方法研究[J]. 江西师范大学学报: 自然科学版. 2004, 28(2): 183–188.
- [16] 陈兴旺. 广义极值分布理论在重现期计算的应用[J]. 气象与减灾研究. 2008, 31(4): 52–54.
- [17] 孟庆珍, 杜健. 成都地面风速年极值的 4 种分布函数拟合结果的比较[J]. 成都信息工程学院学报. 2001, 16(2): 97–104.
- [18] 侯澍旻, 李友荣, 刘光临. 一种基于 KS 检验的时间序列非线性检验方法[J]. 电子与信息学报. 2007, 29(4): 808–810.
- [19] 赵阿兴, 马宗晋. 自然灾害损失评估指标体系的研究[J]. 自然灾害学报. 1993(3): 1–7.

Estimation on the Frequency of Typhoon Wind Speed over China Based on Historical Wind Observations from Weather Stations

CAO Shijia^{1,2}, HOU Jingwei^{1,2}, FANG Weihua^{1,2} and LIN Wei^{1,2}

(1. Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science, the Beijing Normal University, Beijing 100875, China; 2. State Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China)

Abstract: China has been suffering great loss from typhoon wind damage over the past few decades, and typhoon wind hazard assessment plays an essential role in typhoon risk management. Based on historical ground wind observation data and the CMA-STI best track dataset, historical wind speeds induced by typhoon at each weather station of China are obtained. Kriging, Inverse Distance Weighted, Natural Neighbor and Nearest Neighbor methods were compared according to cross validation results of typhoon wind interpolation. By using maximum likelihood method, the optimal distribution model was selected from Gumbel distribution, Frechet distribution and Weibull distribution for each station, whose annual maximum historical typhoon wind speed was extracted, then the spatial distribution of maximum typhoon wind speeds of expected value and 90% confidence intervals at typical return periods were derived with the application of the optimal spatial interpolation method. The findings indicate that: the maximum typhoon wind speed and corresponding uncertainty decreases from the southeast to the northwest gradually at a certain return period, the southeast coastal region seems to be the most affected area.

Key words: typhoon wind speed; extreme value distribution; risk assessment; uncertainty quantification; spatial interpolation