

蒙金凤,龙俊宇,梁虹,等.基于K-均值聚类法和EOF分析的贵港市台风暴雨特征分析[J].气象研究与应用,2024,45(3):44-49.  
Meng Jin Feng, Long Jun Yu, Liang Hong, et al. Analysis the Characteristics of typhoon rainstorm in Guigang based on K-Means Clustering Method and EOF Analysis[J]. Journal of Meteorological Research and Application, 2024, 45(3): 44-49.

## 基于K-均值聚类法和EOF分析的贵港市台风暴雨特征分析

蒙金凤, 龙俊宇, 梁虹\*, 梁毅进

(贵港市气象局, 广西 贵港 537100)

**摘要:**利用实况观测资料、中国气象局台风最佳路径数据集,采用K-均值聚类法,结合经验正交函数(EOF)分析方法,分析研究2010—2023年间影响贵港市的65个台风造成的暴雨特征。结果表明:(1)有39个台风造成暴雨,每年平均有2.8次台风暴雨,7、8月是台风暴雨的集中期,港南区南部、覃塘区中西部及桂平北部山区台风暴雨多且高发。(2)根据K-均值聚类分析,将65个台风分为A、B、C类。A、B类台风源地为西北太平洋,C类台风多来自南海。A、B类台风以西北行为主,C类台风路径复杂且在北部湾活跃。B类台风强度最强,C类台风强度最弱。(3)根据贵港市三类台风暴雨第一特征向量空间分布情况分析,A、B类台风暴雨年降雨量表现为全市普遍偏多或者普遍偏少分布型式,C类台风暴雨的年降雨量变化趋势主要为桂平东北部、南部和平南西北部偏多,其余地区偏少,或者桂平东北部、南部和平南西北部偏少,其余地区偏多。

**关键词:**台风暴雨;K-均值聚类法;EOF分析;时空变化

**中图分类号:**P457

**文献标识码:**A

**doi:** 10.19849/j.cnki.CN45-1356/P.2024.3.06

台风是最强的暴雨天气系统之一,我国有记录的暴雨极值都是由台风导致<sup>[1]</sup>。广西位于华南地区,南临北部湾,是我国台风暴雨(日降雨量 $\geq 50$  mm)的高发地区之一。据统计,平均每年约有5个台风影响广西,常给广西带来强风、暴雨、风暴潮等灾害<sup>[2]</sup>。贵港市地处广西东南部,是台风灾害影响较严重的地区,对人们生产生活有不利影响。因此,开展台风暴雨研究具有重要科学意义,也是趋利避害的现实需要。

目前,广西台风暴雨研究,单个台风暴雨过程分析丰富且较为成熟<sup>[3-5]</sup>,对于气候态的台风暴雨特征研究相对较少。与以往采取主观分类分析不同台风路径下的风雨影响<sup>[6-8]</sup>不同,近年来一些学者尝试使用K-均值聚类法对进行西北太平洋台风路径客观分类<sup>[9-10]</sup>,赋予每个时次的风速一定权重,通过台风每个时次的经度、纬度信息计算各个台风路径的质心和方差,作为路径形状、长度和位置的特征量,更大程度地考虑台风在较强时的路径特征。

EOF分析和聚类分析一样,都是气象预报和分析不可缺少多元统计工具,在我国气象业务预报和科研工作中起着重要的作用<sup>[11]</sup>。EOF分析方法是在一组变量中找出其方差和协方差矩阵的特征量,把多个变量通过降维的方法转化为少数几个综合变量的统计分析方法,能有效地分析某个地区气象要素的时空分布类型,在气象研究上应用比较常用<sup>[12-14]</sup>。本研究在前人研究基础上,通过K-均值聚类分析方法,对2010—2023年影响贵港的台风路径进行聚类分析,最后通过EOF分析研究不同聚类结果下贵港市台风暴雨的空间分布特征,以为台风降雨的预报提供参考依据。

### 1 资料与方法

#### 1.1 研究资料

所使用的台风数据为上海台风研究所整编的CMA-STI台风最佳路径数据资料集。其提供1949—2016年西北太平洋(含南海,赤道以北,东经

收稿日期:2023-05-16

基金项目:广西气象科研计划项目(桂气科2023QN16、桂气科2023M27、桂气科2023M28)

第一作者:蒙金凤,助理工程师,主要从事天气预报研究。E-mail:17866636243@163.com

\*通讯作者:梁虹,工程师,主要从事天气预报与预警研究。E-mail:37145228@qq.com

180以西)海域台风每6 h的位置和强度;2017年起,对于登陆我国的台风,在其登陆前24 h时段内,最佳路径时间频次加密为逐3 h一次;2018年起,对于登陆我国的台风,在其登陆前24 h及在我国陆地活动期间,最佳路径时间频次加密为逐3 h一次。

采用的雨量数据为2010—2023年贵港市3个国家气象站及95个具有三年以上雨量观测资料的区域气象站的20—20时日雨量数据。

## 1.2 研究方法

### 1.2.1 影响贵港的台风暴雨

为了方便获取更多研究样本,将影响广西的台风<sup>[15]</sup>记为影响贵港市的台风。本研究将2010—2023年贵港市国家气象观测站及自动站日降雨量(20—20时)大于等于50 mm的过程识别出来,根据台风进入影响区的起止时间,在此期间发生的暴雨过程即为影响贵港市的台风暴雨过程。

### 1.2.2 研究方法

(1)K-均值聚类法将数据聚成K类,使得每个样本点都属于且仅属于其中一个类,并且希望不同类的样本点之间的距离尽可能的大,而同一类中的样本点之间的距离尽可能小。本研究将影响贵港的台风路径坐标简化成质心描述<sup>[9]</sup>,以台风每个定位时次的风速为权重,用经纬度坐标计算各个台风路径的质心,利用K-均值聚类法,对各个台风进行分类。应用Calinski-Harabasz (CH) 指标评估聚类效果,该指标基于类间和类内的离差矩阵的迹,CH值越高表明聚类质量越好,类别内部紧凑且类别间区分明显。

(2)为了客观描述贵港市各类台风路径下的台风暴雨年降雨量的空间分布特点,根据聚类分析后

的台风类别划分,计算各类台风的台风暴雨年降雨量,并使用经验正交函数分析方法(empirical orthogonal function, 缩写为EOF),分析各类台风暴雨的降水分布情况。EOF基本原理是将气象要素的场分解为只与时间或空间有关的两部分<sup>[16]</sup>,与空间有关的部分由正交函数组成,称为特征向量,计算特征向量的方差贡献率,方差贡献率越高说明其对应的空间模态越重要,对总方差的贡献越大;与时间有关的部分表示各正交函数随时间的变化,称为时间系数。最后,使用North等<sup>[17]</sup>提出的研究对特征值误差范围进行显著性检验,确定实际资料分析中得到的空间模态是否具有物理意义。

## 2 结果与分析

### 2.1 近十年贵港市台风暴雨时空分布特征及其规律

自2010—2023年间,共计有65次台风事件影响了贵港市,其中直接引发暴雨过程的台风数量为39个。由图1a可知,贵港市台风暴雨呈南部多、中部少的空间分布特征。贵港市台风暴雨中心位于覃塘区西部,港南区南部,桂平市的北部山区,这些地区的年平均台风暴雨雨量达到100 mm以上,其中最大为152.1 mm,出现在桂平北部。

进一步分析台风暴雨的年均发生天数空间分布(见图1b),可以发现台风暴雨的高发区域主要集中在贵港市的港北区西部、港南区南部、覃塘区的西北部以及桂平市的南北部山区。桂平市北部地区台风暴雨的发生频率最高,年均达1.4 d。此外,从2010—2023年台风暴雨年平均天数空间分布图(图1b)来看,台风暴雨高频区主要位于贵港市港北区西部,港南区南部,覃塘区西、北部及桂平市南北

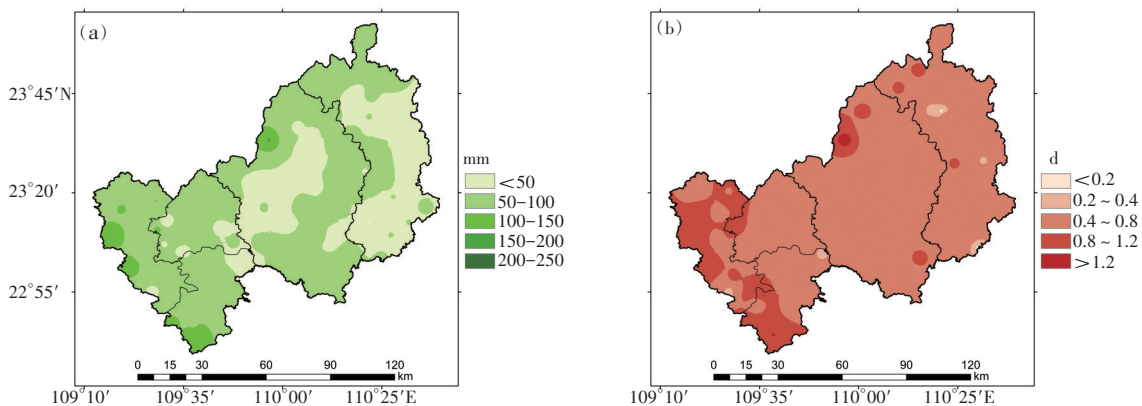


图1 2010—2023年贵港市台风暴雨年平均雨量和年平均天数空间分布

(a)暴雨年平均雨量;(b)暴雨年平均天数

部山区。台风暴雨最频繁的区域同样出现在桂平北部,每年达到1.4 d。由此可见,港南区南部、覃塘区西部及桂平北部山区是台风暴雨中心,同时也是台风暴雨的高频区。

从时间变化特征来看,2010—2023年贵港市台风暴雨频数变化趋势不明显,平均每年有2.8次由台风直接造成的暴雨过程,其中2013年出现最多,达到6次,2022年、2023年最少,没有台风暴雨

的影响(图2a)。分析台风暴雨频次的月变化趋势(图2b),2010—2023年,贵港市台风暴雨频次月际变化呈单峰型分布。台风暴雨过程主要出现在7—8月,共有23次,占总数的59.0%;其中7月发生频次最多,达15次,占总数的38.5%,8月次之,出现8次,占20.5%。贵港市台风暴雨平均持续天数最多的月份是8月和10月,均为1.6 d,其次是9月,为1.4 d。

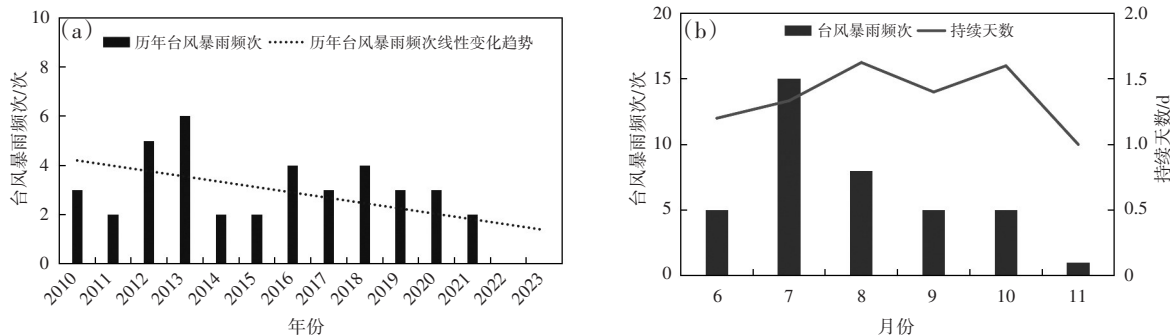


图2 2010—2023年贵港市台风暴雨的频次与月平均持续天数(折线)

(a)年际变化;(b)月际变化(填色)

## 2.2 贵港市台风暴雨分型及变化规律

### 2.2.1 台风分类及特征

对2010—2023年期间影响贵港市的65个台风进行聚类分析,根据CH指标结果,考虑选定3个聚类中心。

A类台风共有21个,根据质心的分布位置来看,主要位于 $18^{\circ} \sim 25^{\circ} \text{N}$ ,  $118^{\circ} \sim 135^{\circ} \text{E}$ 的海域内,即在南海东部和菲律宾北部海域内。此类台风大多源于菲律宾以东的热带洋面上;台风最强的强度在强热带风暴到强台风之间;台风路径整体来看以西北行为主;登陆点主要位于广东雷州半岛以东到福建沿海一带,且登陆后移动的西向分量增大,一般向西行以热带风暴进入桂东,进而影响贵港。

B类台风共有13个,其质心主要位于 $10^{\circ} \sim 18^{\circ} \text{N}$ ,  $112^{\circ} \sim 135^{\circ} \text{E}$ 的海域内,即在南海中部和菲律宾中北部一带。从源地来看,多在菲律宾以东, $5^{\circ} \sim 15^{\circ} \text{N}$ 的洋面上生成,路径同样以西北行为主。整体来看,B类台风的强度为三类中最强,大部分最大强度能达到强台风等级。登陆点或再次登陆点通常在桂南沿海到雷州半岛一带,进入广西的台风强度减弱明显,以强热带风暴到热带风暴为主,登陆后台风移动的北向分量偏大。

C类台风共有31个,是三类台风中对广西地区造成影响最为频繁的一类。C类台风的质心分布相对偏西,位置主要位于 $18^{\circ} \sim 22^{\circ} \text{N}$ ,  $105^{\circ} \sim 115^{\circ} \text{E}$ 的海域内,即在北部湾到南海西北部海域内。此类台风的源地通常在南海中北部及菲律宾东部沿海附近,移动路径较A、B复杂,以西行到西北行路径为主,部分台风出现连续转向的复杂路径。大部分C类台风西行进入北部湾后,中心移动范围在广西海岸线以南,最终登陆越南后逐渐消散;少数C类台风在广东登陆后从桂东南进入广西内陆,其登陆点主要在广东湛江到澳门一带。C类台风相对A、B类台风的整体强度偏弱,其峰值强度多为强热带风暴级,仅有部分达到台风级别。

### 2.2.2 台风暴雨的分型特征

#### (1) A类台风暴雨

表1给出A类台风暴雨EOF分解后前6个模式的特征值及方差结果。可以看到6个特征根误差范围不重叠,彼此相互独立,但第一模态具有较大的特征值,第一特征向量反映了气候平均场的分布情况,且对应的累积方差贡献率为69.81%。根据North等<sup>[17]</sup>提出的研究对特征值误差范围进行显著性检验时,显著性检验是在相对较大的特征值中进



行的,较小的特征值误差一般被认为是由噪声引起的,不具有显著性。因此,对第一模态的特征向量进行分析,研究2010—2023年A类台风影响下贵港台风暴雨的空间分布类型。

图3为2010—2023年贵港市A类台风暴雨做EOF展开的第一模态特征向量,可见第一模态的特征向量均为正值,高值区域主要位于港北区北部、桂平市南部,整体来看,特征向量大小由南北山区向中部盆地递减,表明2010—2023年贵港市A类台风暴雨年降雨量变化趋势整体具有一致性,即A类台风影响下,全市台风暴雨降雨量呈现出普遍偏多或者普遍偏少的特征<sup>[18]</sup>。从表1可知第一特征值的方差贡献率占比69.81%,即第一模态可以表征贵港市A类台风暴雨年降雨量空间分布的主要类型。

(2)B类台风暴雨

B类台风暴雨的EOF分解结果如表2所示。虽然B类台风暴雨的7个空间模态的特征根的误差范

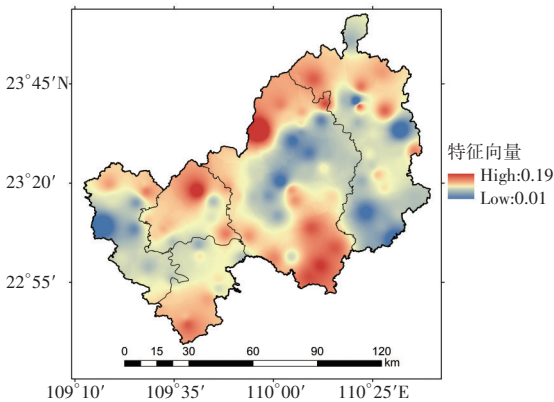


图3 A类台风贵港市台风暴雨第一模态特征向量空间分布

围都不重叠,通过显著性检验,但第一模态的特征值远高于第二模态,累积贡献率占81.71%,因此同样采用第一模态的特征根解释贵港市近十年B类台风暴雨的分布类型。

表1 A类台风暴雨 EOF 分解结果					
模态	特征值	方差贡献率/%	累计方差贡献率/%	特征根误差范围	
				下限	上限
1	111 440.22	69.81	69.81	95 438.34	127 442.10
2	23 716.44	14.86	84.67	20 310.96	27 121.92
3	12 727.94	7.97	92.64	10 900.32	14 555.57
4	8 653.00	5.42	98.06	7 410.50	9 895.50
5	2 002.57	1.25	99.31	1 715.02	2 290.12
6	1 096.33	0.69	100.00	938.91	1 253.76

由图4可知B类台风暴雨第一模态的特征向量均为正值,高值区域主要位于覃塘区中西部、港南区南部及平南北部,这一特征向量代表2010—2023年B类台风影响下,全市台风暴雨年降雨量同样呈

表2 B类台风暴雨 EOF 分解结果					
模态	特征值	方差贡献率/%	累计方差贡献率/%	特征根误差范围	
				下限	上限
1	305 316.98	81.87	81.87	260 543.13	350 090.84
2	41 825.45	11.22	93.09	35 691.87	47 959.03
3	15 831.16	4.25	97.34	13 509.57	18 152.75
4	4 343.67	1.16	98.50	3 706.69	4 980.66
5	3 408.97	0.91	99.42	2 909.05	3 908.88
6	1 832.78	0.49	99.91	1 564.01	2 101.56
7	347.23	0.09	100.00	296.31	398.15

现出普遍偏多或者普遍偏少的特征。

(3)C类台风暴雨

表3为C类台风暴雨的EOF分解结果,给出前4个特征向量方差贡献率及特征根误差范围,由表

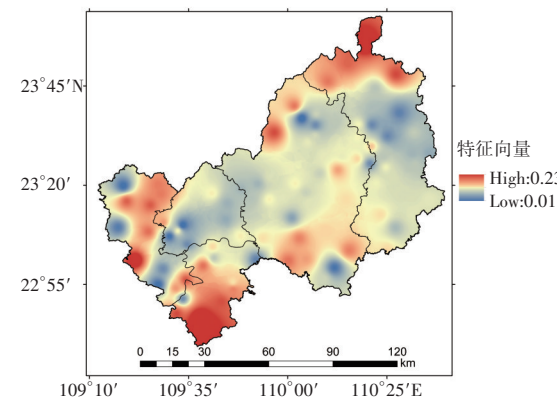


图4 B类台风下贵港市台风暴雨第一模态特征向量空间分布

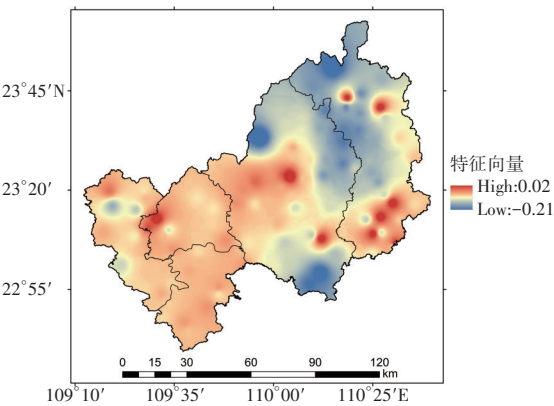


图5 C类台风下贵港市台风暴雨第一模态特征向量空间分布

可知,第3个特征根的误差范围与第4个有重叠,前3个特征根的误差范围不重叠,但由于第一模态的特征值较大,方差贡献率达到72.55%,根据前面分析,第一特征向量反映了气候平均场的分布情况,因此对第一模态进行空间分布分析。

表3 C类台风暴雨EOF分解结果

模态	特征值	方差贡献率/%	累计方差贡献率/%	特征根误差范围	
				下限	上限
1	168 633.27	72.55	72.55	145 685.19	191 581.35
2	35 838.69	15.42	87.97	30 961.66	40 715.72
3	8 554.86	3.68	91.65	7 390.69	9 719.03
4	7 489.14	3.22	94.87	6 470.00	8 508.28

C类台风暴雨的第一模态特征向量在桂平东北部、南部及平南西北部为负值,其余地区为正值,也就是说C类台风暴雨的年降雨量变化趋势主要为桂平东北部、南部和平南西北部偏多,其余地区偏少,或者桂平东北部、南部和平南西北部偏少,其余地区偏多。此外,由于第一模态的正值中心位于港北区西北部山区、桂平城区附近及平南南部山区,负值中心位于桂平东北部山区及南部山区、平南北部山区,说明这些地区为这种分布型式的中心。

### 3 结论与讨论

本研究采用K-均值聚类分析与EOF分析相结合的方法,系统评估2010—2023年间影响广西贵港市的65个台风事件,揭示台风暴雨的时空分布特征与台风活动的内在关联。与以往研究不同,本研究对研究台风个例的质心使用K-均值聚类分析进行分类,不仅能够更全面的描述台风路径、强度和速度的变化的动态演变特征,而且运用K-均值聚类分析,能够更客观分类,减少人为判断的主观性,提供更为科学的台风活动的统计分析特征。得出以下主要结论:

(1)台风暴雨的时空分布显示港南区南部、覃塘区中西部及桂平北部山区是台风暴雨中心,同时也是台风暴雨的高频区,暴雨中心的分布可能与我市的南北山区、中部平原的地形及台风移动方向、路径有关,在台风暴雨预报中应关注这些区域。从时间特征来看,贵港市台风暴雨过程平均每年有2.8次,最多时达到6次;从月变化趋势,7月和8月是台风暴雨的集中期。

(2)通过K-均值聚类分析将台风分为A、B、C类;A类台风21个、B类13个、C类31个。从台风源地来看,A、B类多个台风源地为西北太平洋,不同的是A类偏北,B类偏东;C类台风源地通常在南海中北部及菲律宾东部沿海附近。从移动路径和登陆点来看,A类台风主要为西北行,在广东雷州半岛以东到福建沿海一带登陆;B类台风同样以西北行为主,在桂南沿海到雷州半岛一带进入广西;C类台风移动路径复杂,大部分西行到西北行,且主要在北部湾活跃。从台风的强度来看,B类最强,C类最弱。

(3)根据三类台风暴雨第一特征向量空间分布情况分析,A、B类台风暴雨呈全局性一致型,表现为贵港市台风暴雨年降雨量普遍偏多或者普遍偏少;但二者分布型式的中心位置不同,A类台风暴雨主要分布型式的中心在港北区北部、桂平市南北山区,B类则是覃塘区中西部、港南区南部及平南北部。C类台风暴雨的年降雨量变化趋势主要为桂平东北部、南部和平南西北部偏多,其余地区偏少,或者桂平东北部、南部和平南西北部偏少,其余地区偏多,这种分布形势的中心在桂平东北部山区及南部山区、平南北部山区。

本研究在台风暴雨的时空特征分析方面仍存在局限性,主要包括数据样本量的限制和模型假设的简化,如仅使用2010—2023年的数据可能不足以完全说明气候特征。未来的研究可以考虑延长分析时间跨度,纳入更多历史数据,以提高统计显著性。

#### 参考文献:

- [1] 任福民,杨慧.1949年以来我国台风暴雨及其预报研究回顾与展望[J].暴雨灾害,2019,38(5):526-540.
- [2] 肖志祥,姚才,赵金彪,等.广西热带气旋预报业务和研究进展[J].气象研究与应用,2020,41(4):20-27.
- [3] 韦美闹,赵金彪,陈绍河,等.南海台风“韦帕”生成环境条件及移动转向成因分析[J].气象研究与应用,2021,42(4):20-25.
- [4] 肖志祥,黎金玲,罗小莉.1415号台风“海鸥”广西极端暴雨预报偏差分析及地形订正[J].气象研究与应用,2021,42(4):14-19.
- [5] 赵飞,包文雯,张雪波,等.台风“浪卡”(2016)暴雨成因及数值预报模式偏差分析[J].气象研究与应用,2021,42(3):83-87.
- [6] 孟莹,卢娟,缪启龙.影响台湾岛海域的西太平洋台风

- 特征分析[J]. 热带气象学报, 2005, 21(3): 315–322.
- [7] 蒙焰臻, 陈见, 韩慎友, 等. 广西台风残涡暴雨发生特征分析[J]. 气象研究与应用, 2017, 38(1): 20–25.
- [8] 钟利华, 李仲怡, 李勇, 等. 西江流域台风暴雨面雨量分布特征及天气概念模型[J]. 气象研究与应用, 2017, 38(3): 13–22.
- [9] 郑颖青, 余锦华, 吴启树, 等. K-均值聚类法用于西北太平洋热带气旋路径分类[J]. 热带气象学报, 2013, 29(4): 607–615.
- [10] 彭跃华, 易大江, 王挺, 等. 西北太平洋台风路径的聚类分析[J]. 海洋预报, 2019, 36(5): 63–69.
- [11] 谢炯光, 曾琼, 纪忠萍. 中国近30年来气象统计预报进展[J]. 气象科技, 2003, 31(2): 67–83.
- [12] 庞铁舒, 祝从文, 刘凯. 中国夏季降水异常EOF模态的时间稳定性分析[J]. 大气科学, 2014, 38(6): 1137–1146.
- [13] 王栋, 梁忠民, 王靖, 等. 基于主成分聚类分析的云南省干旱自然分区[J]. 南水北调与水利科技, 2017, 15(2): 15–21.
- [14] 段晓梅, 王英. 基于EOF的内蒙古地区降水时空变化特征分析[J]. 河南科学, 2020, 38(7): 1125–1130.
- [15] DB45/T 2154—2020, 影响和登陆广西的热带气旋[S]. 南宁: 广西壮族自治区市场监督管理局, 2020.
- [16] 邓爱军, 陶诗言, 陈烈庭. 我国汛期降水的EOF分析[J]. 大气科学, 1989, 13(3): 289–290.
- [17] NORTH G R, BELL T L, CAHALAN R F, et al. Sampling errors in the estimation of empirical orthogonal function[J]. Monthly Weather Review, 1982, 110(7): 699–706.
- [18] 魏凤英. 现代气候统计诊断与预测技术[M]. 北京: 气象出版社, 2022: 107–114.

## Analysis the Characteristics of typhoon rainstorm in Guigang based on K-Means Clustering Method and EOF Analysis

MENG Jinfeng, LONG Junyu, LIANG Hong\*, LIANG Yijin  
(Guigang Meteorological Bureau, Guangxi Guigang 537100, China)

**Abstract:** Using the observation data and the best track dataset of typhoons from the China Meteorological Administration to analyze the characteristics of rainstorms caused by 65 typhoons affecting Guigang from 2010 to 2023 using the K-means clustering method combined with the Empirical Orthogonal Function (EOF) analysis method. The results indicate that: (1) there are 39 typhoons caused rainstorms, with an average of 2.8 typhoon-induced rainstorms per year, and the concentrated period for these rainstorms is in July and August. The southern part of Gangnan District, the central and western part of Qin Tang District, and the northern mountainous area of Guiping are the areas with more frequent and severe typhoon rainstorms. (2) Based on the K-means clustering analysis, the 65 typhoons were classified into three categories: A, B, and C. Typhoons of categories A and B originated from the northwest Pacific, while category C typhoons mostly came from the South China Sea. Typhoons of categories A and B were primarily characterized by northwestward movement, whereas category C typhoons had complex paths and were active in the Beibu Gulf. Category B typhoons were the strongest, while category C typhoons were the weakest. (3) The analysis of the spatial distribution of the first principal component vectors of the three types of typhoon rainstorms in Guigang showed that the annual rainfall of categories A and B typhoon rainstorms was generally more or less distributed throughout the city, while the annual rainfall variation trend of category C typhoon rainstorms was mainly characterized by more rainfall in the northeastern part of Guiping, the southern part, and the northwestern part of Pingnan, with less rainfall in the remaining areas, or less rainfall in the northeastern part of Guiping, the southern part, and the northwestern part of Pingnan, with more rainfall in the remaining areas.

**Key words:** typhoon rainstorm; K-means clustering method; EOF analysis; spatial and temporal variations