

张倩影, 钟少波, 黄全义. 基于模糊 C 均值的广东省登陆热带气旋路径分类研究[J]. 灾害学, 2015, 30(4): 187–193. [Zhang Qianying, Zhong Shaobo and Huang Quanyi. Classification of tropical cyclone tracks striking Guangdong from 1949–2013 based on fuzzy c-means[J]. Journal of Catastrophology, 2015, 30(4): 187–193.]

## 基于模糊 C 均值的广东省登陆热带 气旋路径分类研究<sup>\*</sup>

张倩影, 钟少波, 黄全义

(清华大学 工程物理系/公共安全研究院, 北京 100084)

**摘要:** 利用中国台风网(www.typhoon.gov.cn)的“CMA-STI 热带气旋最佳路径数据集”获取 1949–2013 年间登陆广东省的热带气旋的登陆位置、强度、路径长度等主要的特征参数, 利用模糊 C 聚类(FCM)法建立了登陆广东省热带气旋路径分类方法, 并对分类得到的 6 个路径类型热带气旋在空间分布特征、强度和生命史以及活动时间等方面进行了对比分析。结果表明各类热带气旋特征量有明显的差异。从 65 年的变化趋势看, 广东省热带气旋存在着西多东少的频次规律, 而强度分布却是东强西弱。

**关键词:** 热带气旋; 路径插值; 路径聚类; 模糊 C 均值; 空间特征; 广东

**中图分类号:** P4; X43 **文献标志码:** A **文章编号:** 1000–811X(2015)04–0187–07

doi: 10.3969/j.issn.1000–811X.2015.04.033

热带气旋是世界上最灾难性的自然灾害之一, 统计数据显示, 在所有气象灾害中, 热带气旋造成的保险损失为最高<sup>[1]</sup>。对于位于热带气旋易发区域来说, 热带气旋(TC)是最具破坏性的自然灾害之一。由于其登陆过程中总是伴随强劲的阵风, 暴雨和风暴潮, 每年热带气旋的登陆都会给沿海和内陆区域造成巨大的损失。研究热带气旋的形成及其演化过程具有很高的科学价值, 社会价值和经济价值<sup>[2]</sup>。热带气旋登陆我国的沿海区域显示了典型的季节性、年际和年代际变化, 合理地根据历史数据分析热带气旋登陆的时间特征, 可以有效地提前做好准备, 为减缓灾害造成的损失提供科学依据<sup>[3]</sup>。

热带气旋的登陆通常发生在中国东南部的沿海地区, 本文的主要研究区域是我国的沿海大省——广东省。广东省是一个财富高度集中的沿海地区, 并且每年都在因热带气旋的登陆遭受巨大的损失<sup>[4]</sup>。随着广东省人口数量的增长, 由热带气旋登陆造成的财产损失、经济损失以及由台风造成的人口伤亡也在逐渐增加。例如, 在 2013 年, 台风 Usagi(NO.1319)从广东省汕尾市登陆, 估计共造成 177.6 亿元的经济损失, 造成至少 29 人死亡。这样惊人的例子更加强调对热带气旋的形成时间和空间移动特征研究的迫切需要。由于气候变化影响了气象系统的动态平衡, 热带气旋时序变化特征以及热带气旋的空间上的路径变化

及登陆频率和强度的变化也受到了一定程度的影响<sup>[4–5]</sup>。认识热带气旋的时空特征可以有效提高风险评估的准确性, 增强预防和减缓灾害造成的损失的能力。

为了能够准确地预测热带气旋登陆的可能性, 并且提前减缓其带来的经济损失, 研究热带气旋登陆路径特征是很有必要的。之前的一些学者<sup>[1, 6–8]</sup>表明, 可以有效地说明不同热带气旋路径特征的方法是根据定义好的类别, 将相应路径进行归类。针对不同区域, 一些探索性的研究将热带气旋路径归类到有限的形式。Hodanish 和 Gray<sup>[9]</sup>针对西北太平洋的热带气旋, 根据其行进过程中的折弯程度, 将其分为四种不同的形式: 大幅向后弯曲, 逐渐向后弯曲, 左转, 直线。Harr 和 Elsberry<sup>[10]</sup>基于异常大气环流与季风槽和副热带高压脊的活动, 将西北太平洋的热带气旋登陆路径划分为三类: 直线, 向南折回和向北折回。Lander<sup>[11]</sup>也同样将中国南海登陆的热带气旋分为了四大类: 直线移动, 向后折回移动, 向北移动和停留在南海。

数值聚类方法已经成为了热带气旋路径分类主要选择。聚类的主要优点是其客观性, 因为它最大可能性地排除了分析人的主观性。Elsner<sup>[12]</sup>表明通过利用路径中最大强度的坐标以及其消逝坐标, K-means 聚类方法可以有效地对热带气旋路径将进行分类。这个方法也被应用到太平洋飓风路

<sup>\*</sup> 收稿日期: 2015–04–03 修改日期: 2015–05–27

基金项目: 国家自然科学基金项目(91224004)

作者简介: 张倩影(1990–), 女, 河南周口人, 硕士, 主要从事气象灾害研究. E-mail: skrvvtg@163.com

径分类和北太平洋温带气旋路径的分类。Camargo 等人<sup>[13]</sup>指出由于 k-means 聚类方法要求矢量数据必须具有相同的长度,所以利用 k-means 聚类方法并不能够覆盖路径上包含的所有点。由于不同热带气旋的路径是由点组成的, k-means 的这种处理数据的局限性会使路径的一些重要信息缺失,造成分类结果的偏差。为了克服这种局限性, Camargo 提出一种新的聚类方法—混合回归模型,并且利用该方法对不同特征的热带气旋进行了分类及特征分析,在此基础上分析了其与大尺度环境变化的一些关系。Nakamura 等人<sup>[14]</sup>建议基于热带气旋路径的第一和第二质量矩阵,也就是其形状和长度的接近性进行分类。通过将质量矩阵应用于 k-means 的方法,他们针对起源于太平洋的热带气旋路径得到了可靠的分类结果。这个研究也解决了不同长度的矢量数据聚类的问题。

在本研究中,我们将利用另一种方法——模糊 C 均值聚类方法(FCM),针对登陆广东省的热带气旋路径进行聚类。大量的热带气旋路径在地图上看起来非常的杂乱无章,以至于无法确定可以划分的模式类型。这种类型的数据的模糊属性,使其非常适合利用模糊聚类方法。k-means 聚类或者分层聚类属于硬聚类结果,也就是说每个数据被特定地分配到一个类中。相反的,在模糊聚类方法中,数据并不是直接被分配到一个类别中,而是允许保留数据的模糊性。这种方法中,一个数据最初以不同的隶属度属于所有的类别,隶属度从 0 变化到 1(0 代表一个数据完全不属于一个类别,而 1 表示该数据完全包含在这个类别)<sup>[15]</sup>。隶属度代表一个数据属于某个特定类别的概率大小,综合热带气旋路径的模糊属性来看,利用模糊聚类方法对一组热带气旋路径进行分析,可以得到更为合理的聚类结果。

利用模糊 C 均值聚类方法,我们试图的从 1949–2013 年间登陆广东省的热带气旋路径找到的聚类中心,并且这些被归为一类的热带气旋路径有着相似的路径形状和邻近的地理路径。在分类的基础上,分别就路径空间分布特征,强度和生命史的特征以及时间分布特征进行了分析。

## 1 资料和方法

### 1.1 资料介绍

(1) 本文采用的台风路径数据来自中国台风网(www.typhoon.gov.cn)的“CMA-STI 热带气旋最佳路径数据集”。该数据集中包括 1949–2012 年西太平洋和南海(赤道以北, 180°E 以西)海域生成的热带气旋每 6 h 的位置和强度, 包括台风中心点位置, 中心气压, 近中心最大风速, 热带气旋活动天数等。

#### (2) 热带气旋基础资料

1949–2013 年登陆广东省的热带气旋名称、登陆时间、强度及登陆地点的相关资料, 来源于

中国气象局发布的《台风年鉴》和《热带气旋年鉴》。热带气旋登陆期间广东省各站点的风速、雨量等逐日资料, 来源于中国气象科学数据共享服务网。

### (3) 地理信息数据

广东省地理区划信息数据主要是来自国家基础地理信息系统全国 1:400 万数据库。图 1 中展示的 1949–2013 年间登陆广东省的热带气旋的路径图。

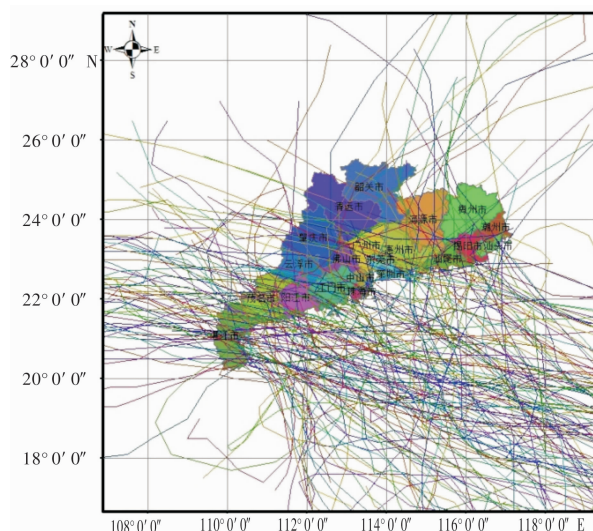


图 1 1949–2013 年间登陆广东省的热带气旋路径

## 1.2 方法简介

### 1.2.1 FCM 聚类方法

模糊聚类分析就是把模糊数学的概念引入聚类分析中, 用来研究“物以类聚”的一种多元统计分析方法, 即用数学方法把原来样品之间的模糊关系定量地确定关系, 从而客观地进行分类, 因此模糊聚类分析的理论基础即是模糊理论。模糊理论是建立在模糊集合基础之上的, 是描述和处理人类语言中所特有的模糊信息的理论, 其主要概念包括模糊集合、隶属度函数、模糊算子、模糊运算和模糊关系等。

本文使用的是模糊 C-均值聚类, 在众多模糊聚类算法中, 模糊 C-均值算法应用最为广泛且较成功, 它通过优化目标函数得到每个样本点对所有类中心的隶属度, 从而决定样本点的类属以达到自动样本数据进行分类的目的。假设样本集合为热带气旋路径及  $X = \{x_1, x_2, \dots, x_n\}$ , 根据模糊 C-均值聚类算法将其划分成 C 个模糊组, 并求每组的聚类中心  $C_j (j=1, 2, \dots, C)$ , 使目标函数达到最小。

目标函数定义为  $J$ ,  $J$  的算法为:

$$J = \sum_{i=1}^C \sum_{k=1}^K (\mu_{ik})^m \|x_k - c_i\|^2. \quad (1)$$

其中

$$\mu_{ik} = \left[ \sum_{j=1}^C \left( \frac{x_k - c_j^2}{x_k - c_i^2} \right)^{2/(m-1)} \right]^{-1}, \quad (2)$$

并且

$$c_i = \frac{\sum_{k=1}^K (\mu_{ik})^m x_k}{\sum_{k=1}^K (\mu_{ik})^m}. \quad (3)$$

式中:  $\mu_{ik}$  表示第  $k$  条热带气旋路径属于第  $i$  个聚类中心的隶属度,  $m$  表示大于 1 的模糊系数,  $x_k$  表示第  $k$  条热带气旋路径,  $c_i$  表示的是第  $i$  个聚类中心,  $C$  表示聚类中心的数目,  $K$  表示热带气旋路径集。任意向量范数  $\| \cdot \|$  表示热带气旋路径与聚类中心的距离。

在本文研究中, 我们使用欧几里得范数来计算, 其在模糊 C-均值聚类算法使用最为广泛。为了优化 C-均值目标函数  $J$ , 我们给其定义了两个限制条件:  $\mu_{ik} \geq 0$  和  $\sum_{i=1}^C \mu_{ik} = 1$ 。模糊系数  $m$  表示每条路径属于不同聚类中心的重叠度, 也就是说, 如果  $m$  的值较小时, 越靠近聚类中心的数据, 所被赋予的权重值也会更大, 同理可得, 越是远离聚类中心的数据, 所被赋予的权重值只会更小。 $m$  系数值越是接近 1, 对于远离某个聚类中心的路径数据所得到的隶属度值  $\mu_{ik}$  会收敛于 0, 而对于靠近某个聚类中心的路径数据而言, 所得到的隶属度值  $\mu_{ik}$  会收敛于 1。在本文中, 我们将模糊系数  $m$  设定为模糊 C-均值算法最常用的值 2。

为了使得目标函数最小化, 采用如何的迭代的过程:

对于给定的数据集  $x$ , 选择聚类的数量为  $1 < C < K$ , 加权指数  $m > 1$ , 迭代停止阈值  $\varepsilon > 0$ , 划分矩阵为:

$$U = \begin{bmatrix} \mu_{1,1} & \mu_{2,1} & \cdots & \mu_{1,C} \\ \mu_{1,2} & \mu_{2,2} & \cdots & \mu_{2,C} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{K,C} & \mu_{2,C} & \cdots & \mu_{K,C} \end{bmatrix}, \quad (4)$$

其中初始的划分矩阵  $U^{(0)}$  是随机生成的。然后重复下列步骤直到  $\|J^{(l)} - J^{(l-1)}\| \leq \varepsilon$ , 其中  $l = 1, 2, \dots$ 。

第一步: 计算聚类中心:

$$c_j^{(l)} = \frac{\sum_{k=1}^K (\mu_{ik}^{(l-1)})^m}{\sum_{k=1}^K (\mu_{ik}^{(l)})^m}, \quad 1 \leq i \leq C. \quad (5)$$

第二步: 更新划分矩阵:

$$\mu_{ik}^{(l)} = \left[ \sum_{j=1}^C \left( \frac{x_k - c_j^{(l-1)}}{x_k - c_j^{(l)}} \right)^{2/(m-1)} \right]^{-1}. \quad (6)$$

整个计算过程就是反复修改聚类中心和分类矩阵的过程, 因此常称此类方法为动态聚类或者逐步聚类法。

### 1.2.2 TC 路径数据插值

与  $k$ -均值聚类算法类似, 模糊 C-均值聚类同样也要求样本数据有着相同的长度, 为了能够处理不同长度的路径, 之前利用  $k$ -均值和质量矩阵的一些研究多是在热带气旋路径上选取一些特定的点, 一般是最大强度或者消逝点。与这些方法不同的是, 利用插值的方法可以保证在聚类过程中保证热带气旋路径的完整性。为了实现这种目的, 在插值的过程中, 将每条热带气旋路径插值成相等长度的  $M$  段, 也就是  $M+1$  个数据点。

利用每个热带气旋路径原始最佳路径集中每 6 h

的位置信息, 可以得到整条路径的长度,  $d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$ , 其中  $i = 1, \dots, N-1$ ,  $(x_i, y_i)$  表示热带气旋路径的第  $i$  个经纬度点,  $N$  表示的是每 6 h 热带气旋路径观测点的数量。插值部分的长度可以用  $ed = 1/M \sum_{i=1}^{N-1} d_i$  表示,  $M$  表示插值的段数。经过插值后的热带气旋路径的坐标可以用下列公式计算得到:

$$\begin{cases} \hat{x}_j = x_1, \hat{y}_j = y_1 & \text{for } j = 1; \\ \hat{y}_j = y_N, \hat{y}_j = y_N & \text{for } j = M+1. \end{cases} \quad (7)$$

$$\begin{cases} \hat{x}_j = x_i + \frac{(x_{i+1} - x_i)}{d_i} [(j-1)ed - \sum_{i=1}^{l-1} d_i], & \text{for } j = 2, \dots, M \\ \hat{y}_j = y_i + \frac{(y_{i+1} - y_i)}{d_i} [(j-1)ed - \sum_{i=1}^{l-1} d_i], & \text{for } j = 2, \dots, M. \end{cases} \quad (8)$$

式中:  $l$  代表的是满足的  $\sum_{i=1}^{l-1} d_i \leq (j-1) \times ed < \sum_{i=1}^l d_i$ , for  $j = 2, \dots, M$ 。整个插值的过程可以根据最佳路径集中每 6 h 的路径坐标, 使得每条路径获得相等数量的新的位置坐标, 并且在插值的过程中保留了原始路径的形状、长度和地理路径信息。在本文的研究中, 我们选取  $M = 20$  作为插值确定新的数据点。经过插值的过程, 对于较短的路径, 在原始观测点中间得到了较密集的新的数据点, 而对于长路径, 中间一些连续的位置点就在被略过了。

为了对样本数据进行模糊聚类分析, 插值后的热带气旋观测点的经纬度坐标点用下列的列向量来表示:

$$x_k = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{(M+1)}, \hat{y}_1, \hat{y}_2, \dots, \hat{y}_{(M+1)}]^T \quad (9)$$

for  $k = 1, \dots, K$ .

式中:  $x_k$  表示的是第  $k$  条热带气旋路径的列向量,  $\hat{x}$  和  $\hat{y}$  分别表示的是经过插值后的热带气旋路径的经纬度坐标。在这里  $k$  是我们所得到的所有的气旋路径的数量, 也就是 180。这个列向量作为样本数据用在式(5)中。然后在优化隶属度函数和聚类中心的过程中, 使得  $C$  均值目标函数达到最小值。

### 1.2.3 最优聚类数的确定

对于任何一种聚类分析来讲, 确定最佳聚类中心数目都是最关键的步骤。FCM 聚类的结果极大地依赖于聚类中心的数目, 因此在进行聚类算法之前必须优先确定最佳聚类数目。在本文中, 聚类中心数目有四个参数来确定: 划分系数( $PC$ ), 分类指数( $SC$ ), 离散指数( $S$ )和  $Dunn$  指数( $DI$ )。

划分系数( $PC$ )是由 Bezdek<sup>[16]</sup> 定义用来检测聚类中心的重叠性。 $PC$  最大的缺陷在于其的计算仅仅依赖于隶属度, 因此, 它缺少与样本数据的地理属性的直接联系。

$$PC(c) = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^N (\mu_{ij})^2, \quad (10)$$

式中:  $\mu_{ij}$  表示数据  $j$  在聚类中心  $i$  的隶属度, 当  $PC$  值最大时, 聚类中心数目达到最优值。

分类指数( $SC$ )是由 Bensaid<sup>[17]</sup> 提出的, 表示聚类的紧凑性和分散性和的比例, 是指每个聚类的模

糊基数归一化的参数和。它在验证聚类紧凑性的同时也考虑到分散性,其中紧凑性是由样本数据与聚类中心的平均距离得到,而分散性是指一个聚类中心到所有聚类中心的距离的和:

$$SC(c) = \frac{\sum_{i=1}^c \sum_{k=1}^K \mu_{ik}^m \|x_j - c_i\|^2}{\sum_{k=1}^K \mu_{ik} \sum_{j=1}^c \|c_j - c_i\|^2} \quad (11)$$

与分类指数( $SC$ )相反的是,由 Xie 和 Beni<sup>[18]</sup>提出的离散指数( $S$ )利用最短距离来进行分类来保证划分的有效性:

$$S(c) = \frac{\sum_{i=1}^c \sum_{k=1}^K (\mu_{ik})^2 \|x_j - c_i\|^2}{K \min_{i,j} \|c_j - c_i\|^2} \quad (12)$$

Dunn 指数( $DI$ )最初由 Dunn<sup>[19]</sup>提出用于验证聚类结果的紧凑性和分散性。这个指数表示的是两个样本到各自聚类中心的最短距离与最长聚类的比值,在本文中,它主要用于每个样本数据分配到指定聚类中心后:

$$DI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{k \in c, k \neq i} d(x, y)}{\min_{x_k \in c} \{ \min_{x, y \in c} d(x, y) \}} \right\} \right\} \quad (13)$$

与  $PC$  不同的是,其他的指数( $SC$ ,  $S$ ,  $DI$ )都是用来检测聚类的紧凑度和分散度,所以,其值越小得到的聚类结果越好。但是必须注意的是,单纯满足某一指数并不能够得到最佳的聚类结果,应该需要同时满足四个指数才可以得到最佳的聚类中心。图 2 展示了以聚类中心数目为函数的四个参数的值。从图 2 中我们可以看出,  $PC$  指数随着聚类中心的增加而逐渐递减,表明最佳的聚类中心数目应该是 2。  $SC$  和  $S$  指数最小值使聚类中心数目为 13,另一方面,  $DI$  指数在聚类中心数目为 6 时达到最小值。综合四个指数,我们可以得到  $PC$  和  $S$  指数在聚类中

心为 6 时变化最大,而  $SC$  指数也在聚类中心为 6 时发生了突变,所以我们将聚类中心的数目确定为 6。

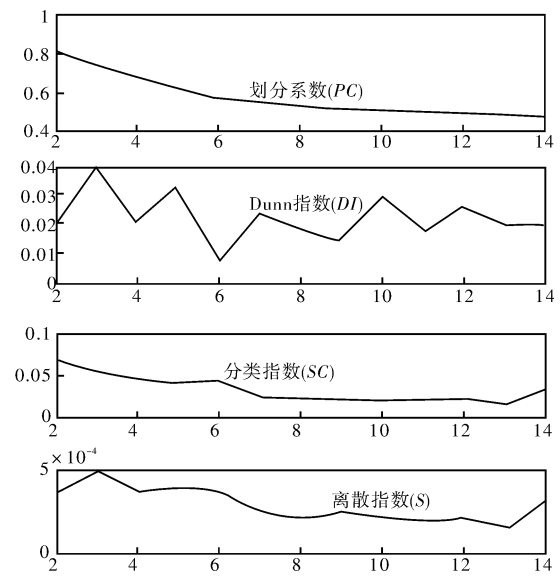


图 2 四个有效性参数与聚类个数的关系图

## 2 聚类结果分析

### 2.1 空间分布特征

图 3 展示了 1949 – 2013 从广东省登陆的热带气旋路径的模糊聚类结果,为了对比,图 3 中也显示了所有热带气旋路径。从图中我们可以得到 6 个聚类中心在空间路径上有着明显的区别,但是根据路径的起源地及地理空间走向大致可以分为三类:西行型,转折型和北上型。其中 B 类路径属于西行型,该类路径登陆的热带气旋主要来自西北太平洋,热

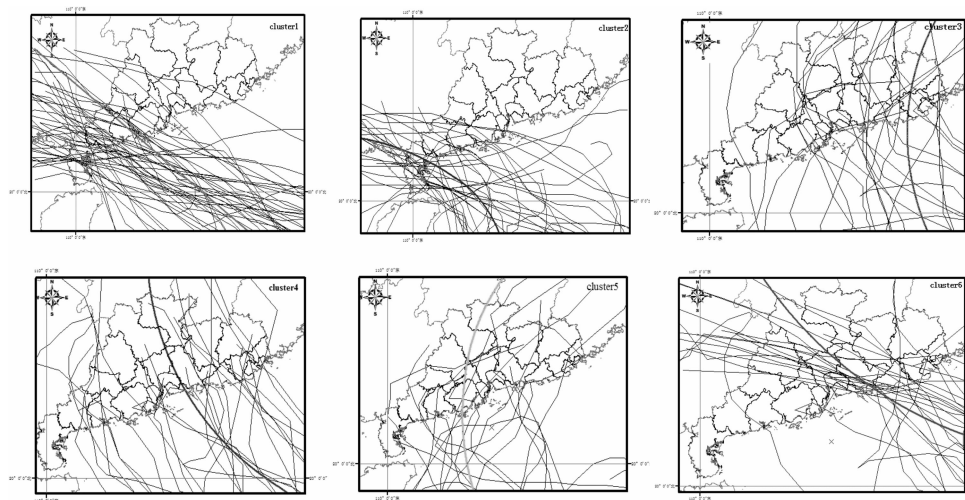


图 3 登陆广东省的热带气旋 6 类聚类结果

表 1 180 条热带气旋路径聚类结果

类别	A	B	C	D	E	F	全部
个数	50	31	28	30	13	28	180
比例/%	0.27027	0.178378	0.156756757	0.162162	0.07567	0.156756	1

表 2 热带气旋路径空间聚类分布特征

参数	A	B	C	D	E	F
经度/°	113.048	110.638	114.746	114.383	112.756	115.927
纬度/°	22.164	21.245	22.806	22.600	21.914	22.820
最低中心气压/hPa	966.2	979.0	970.5	971.2	977.9	950.3
最大风速/m/s	39.3	30.3	35.3	36.6	29.6	45.3
平均生命史/(6h)	27.2	24.2	31.9	25.4	24.2	30.1

带气旋形成后向西北偏西方向移动, 进入南海后继续西北行, 进入广西后逐渐消逝。而 A、D 和 F 类属于转折型路径, 此类路径登陆的台风主要也是来自西北太平洋。热带气旋形成后向偏西方向移动, 但是由于“副高”位置东撤或冷空气及其他天气系统广东而影响, 路径发生转向, 程抛物线型转向东北移动。C 和 E 类路径属于北上型, 该类路径热带气旋多是形成与南海中、北部海面。但是在形成后借助热带气旋内力和副高西元偏北气流引导, 使其向北移动, 分别在福建省和湖南省逐渐消逝。

表 1 和表 2 分别对热带气空间路径聚类结果以及分布特征进行一个比较。热带气旋的登陆点很大程度低依赖其路径, 表 3 展示了 A-F 类热带气旋登陆地分布情况。广东省海岸线较长, 在对登陆点的研究中, 按广东省的地理环境将此海岸线分为了三个沿岸: 粤东沿岸, 珠江沿岸, 粤西沿岸。登陆广东省的热带气旋基本集中在 4 个地区, 自西向东为雷州半岛、阳江-台山、珠江口两侧以及汕尾-惠来。根据表 3 可以得出, 不同类别的热带气旋登陆点的分布情况: 热带气旋的总体情况大致为西多东少, 对于 A 类路径来说, 94% 的登陆点在粤西沿岸和珠江沿岸, 几乎没有从粤东沿岸登陆的情况发生; 而同样为转折型的路径 E、F 类路径来说, 其登陆点的情况较一致, 多是选择在粤东沿岸登陆。B 类路径作为西行型路径的代表, 多是选择珠江口岸登陆, 而对于 C 和 E 类路径来说, 粤西沿岸登陆数约等于珠江三角洲和粤东沿岸之和, 而且从粤东沿岸登陆数最少。

表 3 不同类别热带气旋的登陆分布情况

类别	A	B	C	D	E	F	全部
粤东沿岸	3	1	9	15	1	14	43
珠江沿岸	34	26	6	7	6	5	84
粤西沿岸	13	6	13	6	6	9	53

## 2.2 强度分布特征

热带气旋登陆的情况呈现的是一种西多东少的分布, 然而针对强度来说, 总体呈现的是一种东强西弱的分布趋势, 这个与广东省所处的地理纬度、大气环流系统的平均位置有密切的关系。图 4 以盒须图的形式展示了各类 TC 的登陆过程最大风速的分布状况。在各类路径中, 平均最大风速最大值为转折型路径 A 和 F 类。按照不同台风等级分类来看, 各类路径的强度分布情况如图 5 所示。从总体强度分布来说, 强热带风暴以上强度的热带气旋登陆次数的概率是基本近似的, 而热带气压和热带

风暴的登陆分别在 13% 和 28% 左右。A、C 和 D 类路径的强度分布情况与总体情况基本保持一致, 相反的是, E 和 F 类路径在强度的分布上表现出一定的异常, E 类为北上路径, 多是在粤西沿岸登陆, 并且多是以强热带风暴和台风的强度级别登陆, 强热带风暴强度的热带气旋数占此类路径总数的 85% 左右。F 类为转折路径, 主要是从粤东沿岸进行登陆, 此类路径总占最大比例的是强台风 (50%), 而

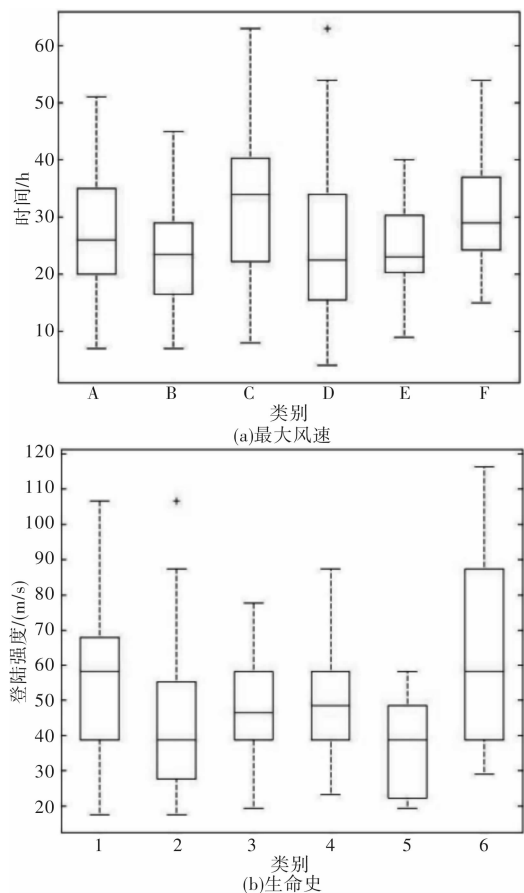


图 4 广东省热带气旋最大风速和生命史箱图 (1949-2013)

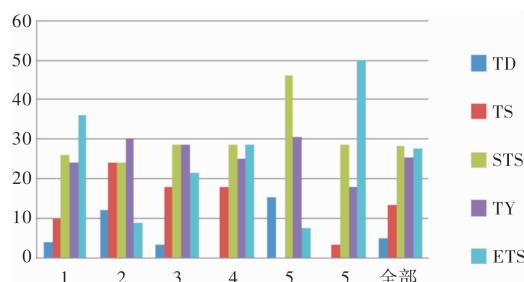


图 5 不同级别热带气旋在各类别的分布特征

且强台风多是从粤东沿岸登陆,几乎没有从珠江和粤西沿岸登陆,其次台风占有 29%,这类强度中,只有两个是从珠江口岸登陆,其余均是从粤东沿岸登陆。西行型的 B 类路径中,热带风暴、强热带风暴及台风三类强度分布情况平均在 9 个,超强台风出现的概率较低,仅有的三个全部都是从粤东沿岸登陆,然后进入广西后逐渐消逝。热带气旋强度与路径的长度也有着一定的关系,也就是与其生命史有着一定的关系。各类热带气旋的生命史分布见图 4(b),图中方框的上下界分别为各类路径 75% 和 25% 分位值,方框内短横线为中位值,方框外短横线为各类上、下边界值。生命史变化幅度从 1 d 到 60 d。其中 C 和 F 类热带气旋的平均生命史最长,其次是 A、D 和 E 类,西型路径的 B 类生命史最短。热带气旋的平均最大风速和生命史的相关系数达 0.70,在海表温度高和垂直风切变小的条件下,一般生命史越长,气旋发展的可能性就越大。

### 2.3 活动时间

不同类别登陆广东省的热带气旋路径有着不同的季节性变化,但总体而言,随着季节的推移,不同类别的热带气旋路径也在发生相应的变化。在 4 月份之前由于“副高”位置偏南、偏东,鲜有热带气旋登陆广东。

图 6 展示了不同类别的 TC 登陆广东省月份的变化,各类热带气旋有明显的季节变化。由图 6 可知,各种路径的热带气旋整体集中在 7-10 月,也就是说夏秋季节。但是针对不同类别,热带气旋登陆的频次峰值也各有不同:(1)A 类、D 类和 F 类:7 月份;(2)B 类:8 月份;(3)C 类:六月份和 8 月份;(4)E 类:9 月份。热带气旋路径频数的这种季节变化与副热带高压季节性的北移和南退密切相关,并且与海平面的温度的变化有着不可分割的关联。

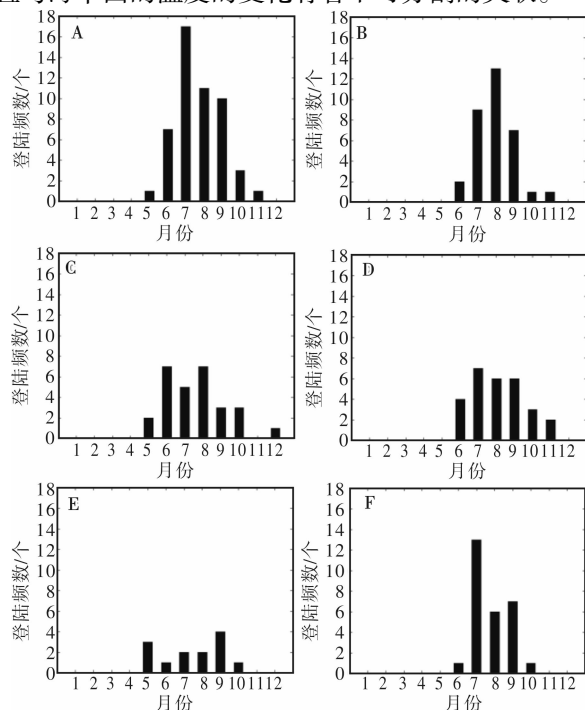


图 6 各类热带气旋的每月平均分布数

针对不同的月份而言,台风的路径也有所不同:5 月时,台风路径多是以 E 类走向为主,此类热带气旋多是在珠江口岸登陆,且源地多为南海;6 月,多是以 A 类和 C 类为主,此类路径分别属于转折型和北上型,期间登陆热带气旋共计 22 个,其中有 8 个来自南海,占 42%;7 月份,登陆的路径以 A 类为主,其次为 F 类,这两类都属于转折型路径,北上路径在七月份时明显减少;8 月份,多是以 A 类和 B 类,其中以西行型路径为主,其次是转折型路径;9 月份, A 类路径最多, C 类和 F 类其次;10-12 月份时,以 A 和 D 类转折型路径为主,北上路径明显减少,其中在 11-12 月份期间,基本没有北上 C 和 E 类路径的发生。

## 3 结论

用 FCM 聚类方法,对广东省热带气旋路径进行了分类,并分析了不同类别的空间分布特征、强度和时间分布特征。

(1)利用插值的方法,将每条登陆广东省的热带气旋的路径都划分为 20 段,在此过程中极大地保证了路径空间完整性。基于 1949-2013 年的热带气旋路径数据,利用 PC、SC、S 和 DI 四个参数,定量地确定了使用 FCM 聚类算法时最优的聚类数目为 6。将热带气旋路径分为了 6 类:三类转折型路径、1 类西行型路径和 2 类北上型路径。

(2)在对热带气旋路径空间特征进行分析时,将广东省的海岸线分为粤东沿岸、珠江沿岸、粤西沿岸。总体来说,热带气旋的登陆情况是西多东少,结合聚类结果,转折型路径 A 多是选择粤西沿岸登陆,而 D 和 F 类多是选择粤东沿岸登陆,西行型的典型代表类 B 路径,多是选择珠江口岸登陆, C 和 E 类作为北上型路径,多是选择从粤西沿岸和珠江沿岸进行登陆,然后分别进入湖南省和福建省。

(3)在对热带性强度和生命史分析时,可以得出,最大风速与生命史具有明显的正相关型,这两个特征量的系数在 0.70 左右。其中 A、C 和 F 类的强度和生命史居前三位,而西行型 B 类和北上型 E 类在强度和生命史上均居于后两位

(4)就热带气旋活动时间而言,5-11 月份是热带气旋出现的集中期,但是不同类别在登陆时间上也有着一定的差异性。5 月份是北上型路径 E 类的高发期,其他类别出现的频率相对较少。8 月份以西行型路径为主,而 9-12 月是转折型路径的登陆高峰期,

广东省热带气旋灾害发生的频率高并且影响范围较广,利用 FCM 方法对路径进行聚类分析,并且根据聚类结果对路径进行空间分析,分析路径的空间格局及其高发时段的时空耦合规律,进而对广东省因地制宜、因时制宜地采取针对性的适应对策提供一定的参考依据,可有效减轻热带气旋灾害损失。

## 参考文献:

- [1] Zuki Z M, Lupo A R. Interannual variability of tropical cyclone activity in the southern South China Sea[J]. *Journal of Geophysical Research*. 2008, 113(D6): 10–19.
- [2] 曹祥村. 2005 年登陆我国热带气旋特征分析[J]. *应用气象学报*, 2007, 18(3): 412–416.
- [3] 王毅, 石汉青, 黄思训. 西北太平洋热带气旋的时空分布特征分析[J]. *自然灾害学报*, 2009, 19(6): 166–174.
- [4] 林继生, 罗金铃. 登陆广东的热带气旋灾害评估和预测模式[J]. *自然灾害学报*, 1995, 3(1): 92–97.
- [5] 周俊华, 史培军, 陈学文. 1949–1999 年西北太平洋热带气旋活动时空分异研究[J]. *自然灾害学报*, 2002, 10(3): 44–49.
- [6] Camargo S J, Robertson A W, Gaffney S J, et al. Cluster Analysis of Typhoon Tracks. Part II: Large-Scale Circulation and ENSO[J]. *Journal of Climate*. 2007, 20(14): 3654–3676.
- [7] Hsu K C, Li S T. Clustering spatial-temporal precipitation data using wavelet transform and self-organizing map neural network[J]. *Advances in Water Resources*, 2010, 33(2): 190–200. *Advances in Water Resources*, 2010, 33(2): 190–200.
- [8] 唐丽丽, 胡德勇, 李小娟. 1951–2006 年西北太平洋热带气旋活动时空特征[J]. *自然灾害学报*, 2012, 21(1): 31–38.
- [9] Ho C, Baik J, Kim J, et al. Interdecadal Changes in Summertime Typhoon Tracks[J]. *Journal of Climate*. 2004, 17(9): 1767–1776.
- [10] Harr P A, Elsberry R L. Large-scale circulation variability over the tropical western North Pacific. Part I: Spatial patterns and tropical cyclone characteristics[J]. *Monthly weather review*, 1995, 123(5): 1225–1246.
- [11] Lander M A. Specific tropical cyclone track types and unusual tropical cyclone motions associated with a reverse-oriented monsoon trough in the western north pacific[J]. *Weather and Forecasting*. 1996, 11(2): 170–186.
- [12] Elsner J B. Tracking hurricanes[J]. *Bulletin of the American Meteorological Society*. 2003, 84(3): 353–356.
- [13] Camargo S J, Robertson A W, Gaffney S J, et al. Cluster analysis of typhoon tracks. Part I: General Properties[Z]. 2007; 20, 3635–3653.
- [14] Nakamura J, Lall U, Kushnir Y, et al. Classifying north atlantic tropical cyclone tracks by mass moments[J]. *Journal of Climate*. 2009, 22(20): 5481–5494.
- [15] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版社, 2004: 212.
- [16] Bezdek J C. Pattern Recognition with Fuzzy Objective Function Algorithms[M]. Kluwer Academic Publishers, 1981: 256.
- [17] Amine M, Bensaid L O H J. Validity-guided (re) clustering with applications to image segmentation[J]. *IEEE Transactions on Fuzzy Systems*. 1996, 4(2): 112–123.
- [18] Xie X L, Beni G A. A validity measure for fuzzy clustering[J]. *IEEE Transactions on pattern Analysis and Machine Intelligence*. 1991, 13(8): 841–847.
- [19] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters[J]. *Cybernetics and Systems*. 1973, 3(3): 32–57.

## Classification of Tropical Cyclone Tracks Striking Guangdong from 1949–2013 based on Fuzzy c-means

Zhang Qianying, Zhong Shaobo and Huang Quanyi  
(Tsinghua University, Department of Engineering Physics/Institute of Public Safety Research, Beijing 100084, China)

**Abstract:** Based on the best-track dataset from the [www.typhoon.gov.cn](http://www.typhoon.gov.cn) between 1949–2013, a method is established for objective classification of tropical cyclone (TC) tracks striking Guangdong province by Fuzzy c-means (FCM) that takes into account such characteristic parameters as the position, intensity, path length, etc. Six clusters from classification are analyzed for comparison in terms of the TC's genesis and landfall position, intensity and lifespan, the temporal characteristics was analyzed. These parameters are quite different among the clusters. From the trends in the 65 years, the frequency of the western is more than the eastern, while the intensity is opposite.

**Key words:** tropical cyclone; interpolation; cluster analysis; FCM; spatial characteristic; Guangdong

## 《灾害学》杂志征订启事

2016 年《灾害学》杂志仍为季刊, 每季初月 20 日出版, 70 克铜版纸印刷, 大 16 开, 220 个页码, 彩色插页, 国内统一刊号: CN 61–1097/P。

《灾害学》杂志 2016 年定价每期 50 元, 全年 200 元(含邮费)。订阅方式有:

1. 到全国各地邮局订阅, 邮发代号: 52–47。
2. 通过邮局直接汇款至编辑部订阅, 即: 陕西省西安市碑林区边家村水文巷 4 号《灾害学》编辑部, 邮编: 710068。
3. 通过全国非邮发报刊联合征订服务部征订。地址: 天津市大寺泉集北里别墅 17 号; 代号: 9875; 邮编: 300385。