

徐碧裕,徐加民,于东海,等.基于DBSCAN聚类方法的强回波定位追踪[J].气象研究与应用,2020,41(3):61-64.

Xu Biyu, Xu Jiamin, Yu Donghai, et al. Strong echo location and tracking based on DBSCAN clustering method[J]. Journal of Meteorological Research and Application, 2020, 41(3): 61-64.

基于DBSCAN聚类方法的强回波定位追踪

徐碧裕¹, 徐加民², 于东海², 胡丽华¹, 叶朗明^{*1}

(1.广东省江门市气象局, 广东 江门 529000; 2.江门市新会区气象局, 广东 新会 529000)

摘要:运用PHP网络编程语言和DBSCAN聚类分析方法,提出了一种新的解析应用该资料的强回波定位追踪方法,针对江门蓬江X波段相控阵雷达组合反射率图片资料,统计2020年6月持续104次强回波定位追踪情况以验证此方法的准确率。结果显示,在DBSCAN聚类参数Eps=10,minPts=2设置下,强回波聚类追踪效果较优,各项结果准确率均值达到85.35%,能满足实际业务所需。

关键词:DBSCAN聚类;强天气;雷达监控;自动报警

中图分类号:P429

文献标识码:A

doi: 10.19849/j.cnki.CN45-1356/P.2020.3.11

OSID:



引言

华南汛期时间长,强对流天气如短时强降水、雷电、大风、冰雹等多发,多数强天气过程生命史短,但突发性强、破坏力大,难以及时有效监测,往往给人们生命财产造成严重损失。因此,提高短时临近天气监测能力,实现强回波定位追踪,是防灾减灾和开展精细化专业气象服务的必由之路。近年来,许多学者^[1-5]运用或改进DBSCAN聚类方法,分析闪电定位资料,实现了对雷电的有效监测预警。但与此同时,Carey^[6]等研究发现一些非常强的风暴并没有产生大量的正地闪;有些学者^[7-8]对强风暴的地闪活动特征进行分析发现,一些强风暴产生非常低的地闪频数,而并非通常认为的对流越强,雷暴云中的地闪活动也越强。因此,利用闪电定位资料常常会忽略对这类强天气的监测定位。采用快速扫描的相控阵雷达组合反射率图片资料,能填补对没有闪电发生的强回波的监控;相控阵雷达资料1~2min内完成一次体扫,同样能增加警报提前量^[9-10];TITAN算法需要雷达体扫资料进行雷暴追踪^[11-12],但使用的雷达

体扫资料成本较高,相比之下雷达图片产品更易于获得,且具有轻量级的优势。因此,基于雷达图片产品采用DBSCAN聚类方法的强回波定位追踪,不仅能节省计算机消耗,也能为短时临近强天气监测提供更可靠的保障。

1 资料与方法

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)聚类算法:是由M.Ester,H.P.Kriegel等人提出的一种基于密度的空间聚类算法^[3]。通过用户给定的密度阈值,将空间对象中具有一定密度的区域划分为类。密度阈值由Eps和minPts两个参数组成,其中Eps为邻域半径,minPts为半径邻域内应该包含的对象个数^[4]。

皮克定理:是指一个计算点阵中顶点在格点上的多边形面积公式,该公式可以表示为 $2S=2a+b-2$,其中a表示多边形内部的点数,b表示多边形边界上的点数,S表示多边形的面积。

$$S=a+\frac{b}{2}-1$$

收稿日期:2020-06-20

基金项目:广东省气象局科学技术研究项目(GRMC2019Z04)、江门市气象局科技项目(201901)

作者简介:徐碧裕(1992—),男,工程师,主要研究暴雨及微信开发。E-mail:bixuejianke@yeah.net

通讯作者:叶朗明(1982—),男,高级工程师,主要研究暖区暴雨。E-mail:ye_langming@126.com

2 坐标网格及颜色识别

2.1 坐标转换

用图像识别的方法解析蓬江雷达产品,需要将监控服务点的地理坐标(经、纬度值),转换为 PHP 网络编程中的屏幕坐标(网页坐标)。根据蓬江雷达所处纬度 22.628°N、雷达监测范围 22.245~23.006°N,112.648~113.468°E。产品的图片像素横、纵范围均为 0~700px,计算得单个像素的经度值和纬度值如下:

$$\text{ScaleX} = (113.468 - 112.648) / 700 = 0.00117 (\text{单位: } ^\circ \text{E/px})$$

$$\text{ScaleY} = (23.006 - 22.250) / 700 = 0.00108 (\text{单位: } ^\circ \text{N/px})$$

最终,得到任意经纬度(Lon, Lat)转换为屏幕坐标(X, Y)的公式如下:

$$X = (\text{Lon} - 112.648) / 0.00117$$

$$Y = (23.006 - \text{Lat}) / 0.00108$$

2.2 监控网格化及颜色识别

蓬江雷达监测覆盖范围为直径 84km 圆形区域,其生成的组合反射率图片分辨率为 700*700 像素,每个像素的长度约为 0.12km。本文以每隔 5 像素距离设置一个格点,构造 61*61 的监控网格,网格分辨率为 5*5 像素(即 0.6*0.6km);设置监控范围为 36*36km(300*300 像素)(图 1a 红框)。为采集网格点上组合反射率图片颜色,需要将监控网格坐标,转换为屏幕坐标。图 1b 网格中心坐标(30,30)对应图 1a 中的屏幕坐标是(350px,350px),图 1b 中红框 4 个角的网格坐标顺时针依次为 (0,0)、(60,0)、(60,60)、(0,60),对应图 1a 屏幕坐标为(200px,200px)、(500px,200px)、(500px,500px)、(200px,500px)。根据转换的屏幕坐标,用 for 循环函数,按列

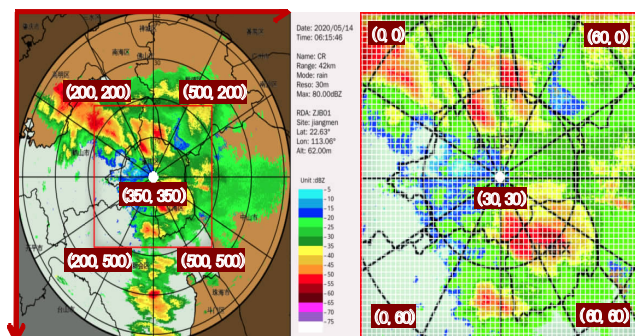


图 1 (a)蓬江相控阵雷达组合反射率图(红框:网格化监控区域,对应为屏幕坐标,单位:px);(b)以雷达/服务点为中心的 61*61 点网格化监控区域,对应格点坐标

采集所有网格点的 RGB 颜色值。其核心代码如下:

```
$gridrgb[$row][$col]=getrgb($im,$X-150+$row*5,$Y-150+$col*5)
```

上式中,\$row、\$col 为网格坐标,变化范围均为 0 到 60,共 61*61 个格点;\$im 为图片地址;\$X、\$Y 为服务点对应的屏幕坐标,设为雷达中心,即\$X=350px、\$Y=350px。

对于计算机来说,图像是一个由数字组成的巨大的三维数组。通常有 3 个颜色通道,分别是红、绿和蓝(Red,Green,Blue,简称 RGB),每个通道的数字都是处于范围 0 至 255 之间的整型,其中 0 表示黑色,255 表示白色^[13]。在 getrgb(\$im,\$x,\$y)内,PHP 语言图像处理函数 ImageColorAt(\$im,\$x,\$y)用于获取雷达组合反射率图片\$中位置(\$x,\$y)的 RGB 颜色值(图 2a);根据获取的 RGB 颜色值,用 rgbtype(\$r,\$g,\$b)函数 Switch 语句,找出对应的 dBz 回波强度值(图 2b)。通过以上颜色识别流程,即可得到 61*61 个网格点的回波强度值。

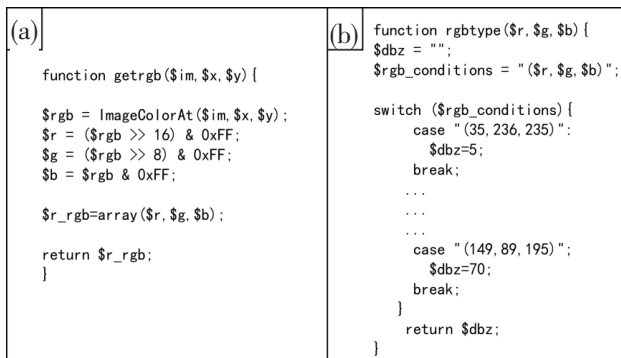


图 2 (a)获取 RGB 颜色值函数;(b)查找 dBz 回波强度值函数

3 强回波聚类定位追踪关键技术

3.1 多簇回波聚类分析

在监控区域内出现多个强回波中心时,为了确定每个强回波簇的质心位置,采用 DBSCAN 聚类分析方法,并以此计算强回波簇的运动方向以及移动速度、外推影响时间等等。以 2020 年 5 月 14 日 06 时 15 分 46 秒的一次强天气过程为例(图 1a、b),获取网格化监控区域 61*61 个格点的 RGB 颜色值,并根据 RGB 值查找对应回波强度 dBz 值,将雷达组合反射率图片中所有回波强度值大于 50dBz 的点对应网格坐标(x,y)筛选出来,进行聚类分析。

调整并设置 DBSCAN 聚类参数 Eps=10, minPts=2(即设置样本的邻域距离阈值为 10 个格点

距离;设置样本个数最小阈值为 2 个格点)。结果可以将监控区域内大于 50dBz 以上的强回波分成 4 个类簇(图 3),计算每个强回波簇的质心坐标(即该类簇下所有网格点坐标平均值,图 3 红点),并根据每个强回波簇质心坐标与服务点中心(30,30)(图 3 棕色菱形点)距离,筛选出最近距离的回波簇(图 3 紫色点),计算相关属性值。

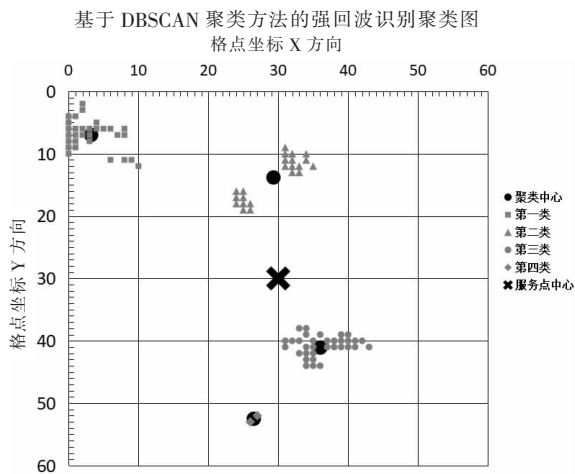


图 3 基于 DBSCAN 聚类方法的强回波识别聚类图
(范围对应图 1b;横、纵坐标为格点坐标)

3.2 强回波簇属性值计算

强回波簇运动参数计算,需要至少 2 个时次以上的组合反射率图。同一强回波簇 T1 时刻质心点①(a,b)移动到 T2 时刻质心点②(x,y),认为代表了整块回波的移动传播。采用文献^[14]中 8 个方位角的划分方法后,得出以下计算公式:

$$(1) \text{ 当 } x > a \text{ 时, } A = \arccos\left(\frac{b-y}{\sqrt{(x-a)^2+(y-b)^2}}\right),$$

如图 4a 中,从点①(1.4,2.4)移动到②(4.2,2)时,其角度为:

$$A = \arccos\left(\frac{2.4-2}{\sqrt{(4.2-1.4)^2+(2-2.4)^2}}\right) = 81.869^\circ,$$

向偏东方向移动(见图 4b)。

$$(2) \text{ 当 } x < a \text{ 时, } B = 360 - \arccos\left(\frac{b-y}{\sqrt{(x-a)^2+(y-b)^2}}\right),$$

如图 4a 中,从点②(4.2,2)移动到③(1.8,4.4)时,其角度为:

$$B = 360 - \arccos\left(\frac{2-4.4}{\sqrt{(1.8-4.2)^2+(4.4-2)^2}}\right) = 255^\circ,$$

向西南方向移动。

(3) 当 $x=a$ 时,如果 $y>b$,则角度为 180° ,向偏南方向移动;如果 $y=b$,则无移动;如果 $y<b$,则角度为

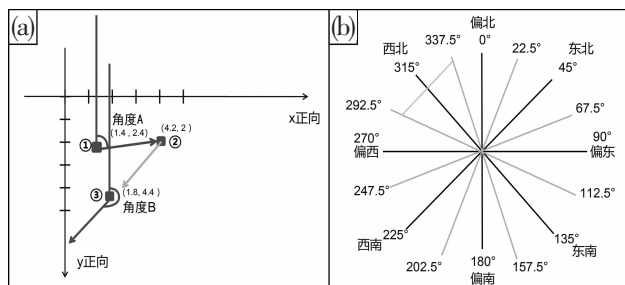


图 4 (a)聚类回波簇运动方向角度值计算
(b)八大方位角划分示意图

0° ,向偏北方向移动。

在雷达图片采样单位时间间隔 ΔT (93s)内,对于单一聚类簇的强回波其质心①(a,b)移动到②(x,y)移动速度计算如下:

$$\text{移动距离 } S = \sqrt{(x-a)^2+(y-b)^2}$$

$$\text{速度 } V = \frac{\sqrt{(x-a)^2+(y-b)^2}}{\Delta T}$$

文中,1 个网格距离等于 0.6km。由皮克定理可以算出当前时刻最近距离强回波簇(大于 50dBz 的回波)的所有格点围成的强回波中心面积。

4 应用成效

当以服务点为中心的 18km 方形监控区域范围内出现 50dB 以上强回波时,经过 DBSCAN 聚类分析后,可以准确定位到强回波所处的实际位置,并计算其动态参数,在微信企业号中弹出报警通知消息。如 2020 年 5 月 14 日 06 时 15 分 46 秒强回波的移动速度为 $27\text{km}\cdot\text{h}^{-1}$,向偏北方向移动。且最近距离的强回波簇中心位于服务点东南方向 7.6km,中心面积为 5.76km^2 ,预计 3min 后会影响到服务点。该信息基于快速扫描的雷达资料,能实现监控加密,并提供短时临近的强回波定位追踪信息。统计 2020 年 6 月持续 104 次的基于 DBSCAN 聚类方法的强回波定位追踪结果,与 S 波段相近时间强雷达回波实况进行对比分析后显示:50dBz 及以上强回波中心追踪准确率如下(以下括号为误差范围):距离($\leq 1\text{km}$)82.69%,方位角($\leq 45^\circ$)87.5%,中心强度($\leq 5\text{dBz}$)97.12%,移速($\leq 10\text{km/h}$)87.36%,移向($\leq 45^\circ$)75.86%,最近距离的强回波簇中心距离($\leq 1\text{km}$)90.80%、方位角($\leq 45^\circ$)90.80%。强回波簇中心面积误差主要出现在簇边界区域,与强回波中心形态密切相关,中心面积准确率($\leq 10\%$)83.91%,外推影响时间准确率(≤ 5 分钟)72.09%。综上,聚类分析效果与聚类参数密切相关,在参数 $\text{Eps}=10, \text{minPts}=2$ 设

置下,强回波聚类追踪效果较优,强回波追踪的各项结果准确率均值达到 85.35%,能满足实际业务所需。

5 结论与讨论

借助高时空分辨率的蓬江相控阵雷达图像资料,通过 DBSCAN 聚类分析方法,可以实现对监控区域内多个强回波的定位追踪,识别出强回波中心强度,确定其方向、距离、面积大小;进一步外推运算,可以得到最近距离的强回波移动方向、影响时间等要素。强回波定位追踪方法已经应用在“江门市气象公共服务中心”微信企业号服务中,对专业气象用户防御灾害性强天气过程以及开展精细化专业气象服务具有重要参考意义。

在实际应用过程中,设置不同的 DBSCAN 聚类参数 Eps 、 $minPts$,对强回波聚类追踪的准确率影响较大。因此,如何根据监控区域大小,筛选出最高准确率对应的聚类参数,仍然需要做大量统计工作。基于 DBSCAN 聚类算法的强回波定位追踪,仍有许多问题需要进一步优化解决,包括聚类参数的自适应选择、影响服务点回波簇的筛选等等。

参考文献:

- [1] 梁丽,雷勇,张帅弛,等.基于 DBSCAN 与网格搜索的雷电定位算法[J].应用气象学报,2019,30(3):267-278.
- [2] 侯荣涛,朱斌,冯民学,等.基于 DBSCAN 聚类算法的闪电临近预报模型[J].计算机应用,2012,32(3):847-851.
- [3] 高攀,田浩,李健,等.基于改进 DBScan 算法的雷暴挖掘与研究[J].高压电器,2019,55(4):169-177.
- [4] 李涛,张帅弛,张灿,等.基于自适应密度聚类分析的雷电定位算法[J].科学技术与工程,2018,18(36):53-59.
- [5] 赵琴,郝笑,钱嘉星,等.广州雷电监测预警服务平台的设计与实现[J].广东气象,2020,42(1):71-74.
- [6] Carey L D, Rutledge S A. Electrical and multiparameter radar observations of a severe hailstorm[J]. J Geophys Res, 1998, 103(D12): 13979-14000.
- [7] Lang T J, Rutledge S A, Dye J E, et al. Anomalous low negative cloud-to-ground lightning flash rates in intense convective storms observed during STERAO-A [J]. Mon Wea Rev, 2000, 128(1): 160-173.
- [8] Soula S, Seral Y, Feral L, et al. Cloud-to-ground lightning activity in hail-bearing storms[J]. J Geophys Res, 2004, 109, D02101.
- [9] 杨金红,高玉春,程明虎.相控阵天气雷达与多普勒雷达的探测精度与探测能力对比研究[J].气象,2008(5):44-51,129.
- [10] 张志强.相控阵天气雷达强度探测与评估方法研究[D].中国气象科学研究院,2011.
- [11] Dixon M, Wiener G. TITAN: Thunderstorm Identification, Tracking, Analysis and Nowcasting—A Radar-based Methodology [J]. Journal of Atmospheric and Oceanic Technology, 1993, 10(10): 785-797.
- [12] 周毓荃,潘留杰,张亚萍. TITAN 系统的移植开发及个例应用[J].大气科学学报,2009,32(6):752-764.
- [13] 魏溪含.深度学习与图像识别原理与实践[M].北京:机械工业出版社,2019.
- [14] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.地面气象观测规范 风向和风速:GB/T 35227-2017[S].北京:中国标准出版社,2017.

Strong echo location and tracking based on DBSCAN clustering method

Xu Biyu¹, Xu Jiamin², Yu Donghai², Hu Lihua¹, Ye Langming^{*1}

(1. Jiangmen Meteorological Bureau, Jiangmen Guangdong 529000;

2. Meteorological Bureau of Xinhui District, Jiangmen Guangdong 529000)

Abstract: Based on the high spatial-temporal resolution image data of Pengjiang phased array radar, using PHP network programming language and DBSCAN clustering analysis method, a new strong echo location and tracking method was proposed for the combination reflectivity image data of X-band phased array radar in Pengjiang, Jiangmen. In order to verify the accuracy of this method, 104 times of strong echo tracking in June 2020 were counted. The results show that when the DBSCAN clustering parameters $EPS=10$, $minpts=2$, the strong echo clustering tracking effect is better, and the average accuracy of strong echo tracking results is 85.35%, which can meet the needs of actual business.

Key words: DBSCAN clustering; strong weather; radar monitoring; automatic alarm