

暴雨云团的多尺度识别方法及其在临近预报中的应用

王改利^{1, 2} 刘黎平¹

1 中国气象科学研究院灾害天气国家重点实验室, 北京 100081

2 南京信息工程大学, 南京 210044

摘 要 将一种暴雨云团的多尺度识别方法——层级聚类法, 应用于 β 中尺度对流系统识别及临近预报中。该方法的基本思路是: 将笛卡尔坐标下的雷达反射率因子进行聚类, 得到比较详细的较小尺度的暴雨云团, 然后设定阈值, 将云团之间差异小于阈值的进行合并, 可以得到较大尺度的云团, 逐步放宽合并阈值, 可得到更大尺度的云团。选取广州雷达 2005 年 3 月的飑线过程和温州 2005 年 9 月的台风过程对这种方法的识别结果进行了详细说明, 结果表明: 该方法能够识别不同系统的 β 中尺度对流云团, 并能识别出其中反射率较强的 γ 中尺度云团, 识别结果合理。采用这种方法识别不同尺度的暴雨云团, 有利于跟踪、预报造成中国暴雨主要原因的 β 中尺度系统, 也可兼顾 β 中尺度系统中的 γ 中尺度对流单体。根据预报时效的不同, 可以选择不同的云团识别尺度。

关键词 多尺度识别 层级聚类 跟踪 预报

文章编号 1006-9895(2007)03-0400-10

中图分类号 P412

文献标识码 A

A Multiscale Identifying Algorithm for Heavy Rainfall and Application in Nowcasting

WANG Gai-Li^{1, 2} and LIU Li-Ping¹

1 *State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081*

2 *Nanjing University of Information Science & Technology, Nanjing 210044*

Abstract Heavy rainfall in the Meiyu front and typhoon precipitation are mainly caused by meso- β Mesoscale Convective Systems (MCS). In order to identify the convective systems with different spatial scales, a multi-scale cloud cluster identification algorithm called hierarchical K -Means clustering method is developed in the paper. The algorithm is clustering radar reflectivity data in Cartesian coordinate using K -Means cluster to classify all grids according to a criterion, and the detailed and smaller clusters are identified in the first step. The coarser cloud clusters are formed by merging the clusters with differences less than the threshold. The algorithm and its application in nowcasting are described in detail for the squall line heavy rainfall and the typhoon observed by Guangzhou and Wenzhou radars. The main conclusions are gotten as follows: (1) The clustering algorithms, which are widely used in the field of market analysis and medicinal practice, are successfully used in meso- β -scale and meso- γ -scale convective cells identification, and the results are reasonable for the two cases. The multiscale algorithm is helpful for identifying, tracking and forecasting meso- β -scale and meso- γ -scale systems. (2) The convective systems with different scales can be tracked and extrapolated with different forecast time. The further study to improve the forecast and the evolvement of storm with the algorithm still need to be done.

收稿日期 2006-03-09 收到, 2006-09-22 收到修改稿

资助项目 国家自然科学基金资助项目 40375008, 国家重点基础研究发展规划项目 2004CB418305

作者简介 王改利, 女, 1973 年出生, 工程师, 主要从事利用雷达资料进行暴雨的临近预报工作。E-mail: wgl3111@cams.cma.gov.cn

Key words multi-scale, hierarchical clustering, tracking, forecasting

1 引言

新一代多普勒天气雷达可以获得高时空分辨率的降水强度的信息, 是监测暴雨, 并进行临近预报的重要手段之一。美国建立的临近预报系统的重要资料来源就是多普勒天气雷达。目前, 我国也正在建立新一代多普勒天气雷达观测网, 其探测覆盖面、探测能力和探测资料将会迅速增加。因此, 可以利用多普勒天气雷达对暴雨进行实时监测, 分析暴雨的一些重要特征, 对其进行预警和临近预报, 以减少气象灾害的发生。

利用雷达资料进行的临近预报, 主要是对雷达回波的跟踪及外推。在国外, 最普遍的方法是质心跟踪法, 质心跟踪法对雷暴进行识别并计算其特征, 通过对相邻时刻的雷暴进行匹配, 进行雷暴的跟踪^[1, 2]。以质心跟踪为基础发展起来的算法有雷暴的识别、跟踪、分析和临近预报 (TITAN)、风暴单体识别和跟踪算法 (SCIT), 另一种技术是利用神经网络将反射率因子模拟成神经元来预报将来时刻反射率的位置^[3]。第三种方法是利用矩形网格在搜索半径内寻找最大相关系数的方法来跟踪雷达回波 (TREC)^[4, 5]。Browning 提出了预报此类问题的另外一种方法——物理概念模型法 (CM)^[6], 定性地描述了一些气象现象的结构、形成机制和生命周期的相互联系。此外, Lakshmanan 等^[7]提出了一种雷暴的多尺度识别方法, 该方法采用聚类分析对雷暴进行识别, 再用层级聚类的方法对识别的雷暴进行合并得到不同尺度的雷暴, 从而实现雷暴的多尺度识别。以这些风暴跟踪算法为基础发展起来的临近预报系统有美国国家大气研究中心 (NCAR) 的 Auto-Nowcast System (ANC)^[8], 该系统综合了雷达、卫星、探空、闪电等多种传感器的资料进行风暴的临近预报, 对风暴的跟踪主要采用 TITAN、TREC 算法; 英国的 Generating Advanced Nowcasts for Development in Operational Land-surface Flood Forecasts System (GANDOLF) 主要采用神经网络法和物理概念模型法进行云团的分类和外推^[9]; 香港天文台发展的“小涡旋”临近预报系统的主要部分是采用 TREC 方法跟踪雨带的移动, 并结合地面雨量站预测未来数小时的雨量分布。美国强

风暴实验室 (NSLL) 的 Warning Decision Support System-Integrated Information (WDSS-II) 采用了 SCIT、冰雹监测算法、中尺度气旋检测算法、雷暴多尺度识别等算法来监测、预警强对流天气。

国内各个省市气象局都非常重视强对流天气的临近预报, 并做了一些工作, 如新疆、山东的冰雹短时预报分析^[10], 江西的飊线雷达回波特征分析^[11]。上海市气象局还以中尺度数值预报模式预报产品为基础, 人机交互方式为工作平台, 综合使用了卫星、雷达、自动站及高空地面观测等资料, 结合预报员经验建立了上海地区强对流天气短时预报系统^[12]。肖艳姣等^[13]利用质心法对风暴进行实时地自动识别、跟踪、结构分析和临近预报。陈家慧等^[14]讨论了利用 BP 模型进行临近预报的方法, 说明了人工神经网络方法用于临近预报的可行性。盛春岩等^[19]讨论了多普勒雷达资料对中尺度短时预报的影响, 认为同时利用雷达反射率因子和径向风能够改善中尺度数值模式短时定量降水预报。

近年来气象学家的研究表明: 我国的梅雨锋暴雨、台风暴雨主要由 β 中尺度对流系统引发。如贝耐芳等^[15]研究了 1998 年长江流域特大暴雨发生的中尺度特征, 认为 β 中尺度系统是造成暴雨的主要原因; 张小玲等^[16]指出 β 中尺度系统暴雨是我国最常见的一类暴雨, 由其引发的降水量, 占降水总量的 60%~70%; 李献州^[18]对广州台风特大暴雨的研究认为: 台风特大暴雨主要是由其云系中的中尺度强降水系统造成。因此, 研究 β 中尺度系统暴雨的雨带自动识别方法, 以及对雨带的跟踪和未来 0~2 小时的降雨量分布作出定量的预报, 减少气象灾害所造成的损失, 是一项很有意义的工作。

现在发展的质心跟踪方法主要适用于 γ 中尺度的较小风暴单体, 对于上百公里的 β 中尺度云团效果不是很理想。TREC 方法将整幅雷达图像分成若干个矩形次网格, 对云团的划分物理意义不够明确。神经网络方法需要对神经元进行大量的资料训练。本文采用雷暴的多尺度识别方法, 对我国的暴雨云团进行了多尺度识别, 并针对我国对流系统的特征, 对 Lakshmanan^[7]提出的风暴多尺度识别方法中的参数进行了调整, 并在云团合并时既考虑了云团特征值之间的差异, 也考虑了云团质心之间的

距离。这种多尺度识别方法既可以识别 β 中尺度系统的云团,也可以识别 β 中尺度系统中的 γ 中尺度风暴单体,比较适合我国的强对流天气过程,并以我国的飑线过程和台风过程为例对该识别方法进行了验证。该方法的优点是用于临近预报时可根据预报时效的长短选择云团的识别尺度。

2 基于聚类算法的云团的多尺度识别方法及跟踪、预报方法介绍

2.1 聚类方法介绍

聚类分析是一种新兴的多元的统计方法,简单地说,聚类就是将一个对象的集合分割成几个类,每个类内的对象之间是相似的,但与其他类的对象是不相似的。聚类分析在市场研究领域、医学实践中已经得到广泛应用,将其用于雷达图像的识别,其优点是可以实现雷达回波的多尺度分割。

聚类有多种方法,本研究选择基于 K -Means 算法的层级聚类法, K -Means 算法的基本思想是选择合适的 K 值(要生成类的数目),按照相似性判断准则,将那些具有相似性的对象聚集起来,同时将性质不同的对象分开,相似性可以通过欧几里德距离来衡量。该算法优点是算法简单直观,但是必须事先给出合适的 K 值。 K -Means 聚类算法的关键是描述数据特征的向量和判断准则。

2.1.1 特征向量

聚类算法应用于雷达图像的分析中,聚类的目的就是将那些具有相似性的格点聚成一类,从而实现雷达图像暴雨云团的自动识别。我们可以把雷达图像上的每个格点看作一个多维向量,其中每个分量表示格点的一个数据特征。在本文中采用的向量的分量如下:

- (1) 单点的反射率因子 Z_{xy} ,
- (2) 区域的反射率平均值 \bar{Z}_{xy}

$$\bar{Z}_{xy} = \frac{\sum_{i \in N_{xy}} Z_i}{N_{xy}}, \quad (1)$$

- (3) 与相邻格点的标准差

$$s_{xy} = \sqrt{\frac{\sum_{i \in N_{xy}} (Z_i - \bar{Z}_{xy})^2}{N_{xy}}}, \quad (2)$$

其中, N_{xy} 为格点 (x, y) 相邻区域的格点数目,本研究选取的区域为 5×5 的网格,因此 $N_{xy} = 25$ 。

2.1.2 判断准则

类是一个元素的集合,类内元素的相似性要高,由于我们聚类的对象为图像数据,因此我们要同时考虑图像像素在区域上的连续性,即尽可能使得每一个像素与其周围相邻的像素属于同一类^[17]。因此判断每一像素属于哪一类的判断准则应综合以上两个方面的因素,第一种因素是像素点的特征向量与该像素所属类的均值向量之间的欧几里德距离 $d_m(k)$,

$$d_m(k) = \|\boldsymbol{\mu}_k - \mathbf{T}_{xy}\|. \quad (3)$$

这里, $\boldsymbol{\mu}_k$ 为第 k 类的均值向量, \mathbf{T}_{xy} 为格点的特征向量。第二种因素是像素与其相邻像素之间的连续性测量 $d_c(k)$,其测量该像素点与其周围相邻像素点所属类的差异性。

$$d_c(k) = \sum_{\tilde{ij} \in N_{xy}} (1 - \delta(s_{\tilde{ij}} - k)), \quad (4)$$

其中, s_{ij} 为格点 (i, j) 所属类的序号, $\delta(s_{ij} - k)$ 为单位脉冲函数, N_{xy} 为相邻的格点数,聚类的结果应使每个格点所属的类都综合了上述两种因素。格点 (x, y) 的类的序号 k 应使得价值函数 $E(k)$ 最小。 $E(k)$ 定义如下:

$$E(k) = \lambda d_m(k) + (1 - \lambda) d_c(k), \quad (5)$$

这里 λ 为加权系数,本研究中为 0.7。归纳起来,雷达数据的聚类包括以下步骤:

(1) 选择样本的 k 个类的初始划分,计算这些类的质心。为了节约计算机时间,我们只是简单地将数据的测量范围进行了等分。

(2) 把每个格点分配其特征向量与类的均值向量的欧几里德距离最小的一个类。

(3) 在每个格点上计算以下内容:

(a) 将格点周围相邻格点的类的序号作为候选;

(b) 计算连续性测量 $d_c(k)$;

(c) 第 k 个类的均值向量与该格点特征向量的欧几里德距离 $d_m(k)$;

(d) 将 $E(k)$ 最小的类分配给该格点。

(4) 如果任一格点的类发生了变化,则需更新每个类的均值向量,迭代第 3 步直到每个格点所属的类不再发生变化。

经过以上 K -Means 聚类步骤,雷达 Constant Altitude Plan Position Indicator (CAPPI) 的每个格点被分配到某个类中,并用类的序号进行标志,

通过区域增长的方法得到相互独立的云团。

2.2 云团的层级聚类

层级聚类就是根据一些规则将较小的云团合并成更大的云团,直到满足一些预先设定的条件。具体做法是:计算每两个云团之间的距离(包括质心距离和云团的均值向量之间的欧几里德距离),设定阈值,使得有一半左右的云团对处于该阈值以下,合并这些云团,直到所有云团两两之间的距离都大于阈值。合并后的云团尺度要大一些,包括了合并前的一个或多个云团,从这种意义上说,这两种尺度的云团之间具有层级关系,如果我们放宽阈值,则可以得到更大尺度的云团。

2.3 云团的跟踪

一旦雷达图像上的云团被识别出来,通过最大相关系数法跟踪雷达回波(TREC),求出每个云团的移动方向和速度。这与传统的TREC方法有所不同,传统的TREC方法将雷达图像分成若干个矩形区域,对每个矩形区域求最大相关系数,来实现回波的跟踪,但这种矩形区域的划分是随意的。本文中用到的TREC方法,以云团为区域划分,对每个云团求最大相关系数,来实现云团的跟踪。

通过连续两个时次的雷达CAPPI资料计算云团的移动。具体做法是:将第一时刻识别出的每一块云团作为一个区域,在搜索半径内向任一方向移动一定的距离,然后计算此云团区域与第二时刻相同大小的区域之间的交叉相关系数 R ,对于不同的移动位置,会得到不同的相关系数值,直到找到极大值 R_{\max} 为止,具有最大相关系数的移动就是云团的TREC矢量,用作云团的外推预报,计算公式如下:

$$R = \frac{\sum Z_1(i)Z_2(i) - n^{-1} \sum Z_1(i) \sum Z_2(i)}{\left[\left(\sum Z_1^2(i) - n\bar{Z}_1^2 \right) \left(\sum Z_2^2(i) - n\bar{Z}_2^2 \right) \right]^{1/2}}, \quad (6)$$

式中, Z_1 、 Z_2 分别为 T 时刻和 $T+\Delta t$ 时刻云团的反射率, n 为云团的格点数。那么,通过式(6)就可以求出间隔 Δt 时间的两个云团的相关系数,重复这个过程,直到找到最大的相关系数,此时,从 T 时刻云团的中心位置指向 $T+\Delta t$ 时刻云团中心位置的矢量即为TREC矢量。据此外推云团未来时刻的位置。

搜索半径根据最大期望速度计算:

$$R_r = v \Delta t, \quad (7)$$

式中, v 是风暴移动的最大期望速度(这里采用120 km/h), Δt 为时间间隔。

2.4 预报及评价

云团的预报是以运动估算、反射率的增长和减弱信息以及当前数据为基础,进行外推。对于每一个云团来说,通过计算云团均值的变化可以得到云团反射率增长消亡的信息,以此预报云团反射率的趋势变化。通过2.3节所描述的云团跟踪方法得到TREC矢量,预报云团未来时刻的位置。

根据预报时效的不同,我们可以进行不同尺度的云团识别。小于30分钟的预报,我们选择小尺度的云团识别,对于大于30分钟的预报,选择较大尺度的云团识别。

为了对预报结果进行定量的分析。将预报的数据和预报时刻雷达实际观测的数据逐个格点进行对比,如果实测的格点数据和预报的格点数据都大于阈值,则认为该格点是成功,如果实测的格点数据大于阈值而预报的格点数据小于阈值,则该格点是漏报,如果实测的格点数据小于阈值而预报的格点数据大于阈值,则该格点是虚假警报。探测概率(P_{OD})、虚假警报比(F_{AR})、临界成功指数(C_{SI})按以下定义计算:

$$P_{OD} = \frac{n_s}{n_s + n_f}, \quad (8)$$

$$F_{AR} = \frac{n_a}{n_s + n_a}, \quad (9)$$

$$C_{SI} = \frac{n_s}{n_s + n_f + n_a}, \quad (10)$$

其中, n_s 、 n_f 、 n_a 分别为预报成功的格点数、漏报的格点数和虚假警报的格点数。

3 参数敏感性试验及云团识别结果验证

3.1 参数的敏感性试验

变化参数值,以研究不同参数条件下聚类方法的效果和存在的问题。

3.1.1 K 值的影响

我们采用 K -Means聚类将雷达图像分成 K 个类,然后采用层级聚类技术进行云团的合并。在我们的聚类过程中,采用 $K=4$,这里讨论一下改变 K 值的影响。

K 值的大小确定了最初图像划分的等级,改变 K 值将改变最详细划分的云团的数目。改变 K 值

的影响如图 1 (见文后彩图) 所示, 随着 K 值的增加, 雷达图像类所包括的格点范围减小, 云团数目增多。因此, 如果想得到较小尺度云团的详细信息, 我们就可以选取大一点的 K 值。 K 值变化对云团最详细的划分影响较大, 但由于我们采用了层级技术, 通过迭代的过程对云团进行合并, 因此, K 值的改变不会对较大尺度的云团数量产生直接的影响。台风的结构是强对流系统中比较复杂的, 但从图 1 可以看出, 在 $K > 3$ 时, 这种聚类方法对台风中三个主要强回波带的识别, 轮廓清晰, 结构合理, 随着 K 值的增大, 可以识别出强回波带中反射率更强的核心云团。

3.1.2 加权系数 λ 的影响

在公式

$$E(k) = \lambda d_m(k) + (1 - \lambda) d_c(k)$$

中, 权重 λ 是价值函数的欧几里德距离 $d_m(k)$ 相对于连续性测量 $d_c(k)$ 的权重系数。在我们的研究中, 选取 $\lambda = 0.7$, 这里讨论改变 λ 的影响。

欧几里德距离描述的是格点的特征向量与所属类的均值向量之间的数据差异, 连续性测量描述的是格点与其周围格点属于同一类的平滑度。

假设 $\lambda = 0.0$, 那么我们忽略数据的差异性, 而只考虑连续性的影响。换句话说, 不管相邻格点的特征向量与类的均值之间差异多大, 都把它们归到同一类中去。我们将得到平滑的区域边界, 但失去了图像的数据特征。假设 $\lambda = 1.0$, 则没有考虑连续性测量, 将格点归到欧几里德距离最小的类中, 得到的是一个比较杂乱的图像分割。我们在 $0 \sim 1$ 之间选取 λ 的值进行云团的自动识别, 发现 $0.9 > \lambda > 0.5$ 时, 识别的云团结果有差异, 但差异不是很大, 特别是对流云团的识别, 几乎没有差别。图 2 为 $K = 4$, 改变 λ 时的聚类结果, 从图 2 可以看出, 在 0.6 、 0.7 、 0.8 之间, 聚类结果差异很小, λ 可以随便取值。本文研究过程中, 我们选取 $\lambda = 0.7$ 。

3.1.3 特征向量的影响

聚类算法的一个关键问题就是特征向量的选取, 特征向量是和格点联系在一起的。很难一致同意哪组特征向量是最佳的, 或者有标准的特征向量。前面描述的特征向量只是包括了格点数据的一些特征, 在实践中可以根据实际情况减少或增加特征向量的维数, 以增加聚类的准确性。在我们的参数试验中, 也增加了特征向量的维数, 比如增加了

格点反射率与相邻格点的对比度、变化率等, 但对对流云团的识别效果影响并不是很大 (图略)。考虑到对云团的识别要求和计算机时间, 我们选取 2.1.1 节中描述的三个因子作为格点的特征向量。

3.2 云团识别结果验证

层级聚类法进行云团的多尺度识别, 采用 C Builder 语言编制成软件, 计算机 (主频 2.8 GHz) 运行时间为 30 秒左右。下面, 我们选取了 2005 年 3 月 22 日广州雷达的一次飊线过程和 2005 年 9 月温州雷达的卡努台风过程来具体说明层级聚类方法对云团的识别。

文中研究的雷达资料均为笛卡儿坐标下的 CAPPI 资料, 高度为 2 km, 以雷达为中心, 半径 200 km。图 3 为层级聚类方法对广州 2005 年 3 月 22 日的一次飊线过程的云团识别结果, 图 3 (a)、图 3 (b)、图 3 (c) 分别为不同时刻雷达观测资料 and 云团的多尺度识别结果, 自左向右分别为雷达观测图像, 层级聚类方法识别的云团系统, 云团系统包含的较强的对流单体。从图 3 可以看出, 在雷达观测区域内, 云团被识别为两个系统 (深蓝色和浅蓝色区域), 深蓝色的云团系统包含了一个带状强对流云团, 浅蓝色的云团系统包含了若干个小的强对流单体。从不同时刻的识别结果来看, 不管是对云团系统的识别, 还是对较小的强对流单体的识别, 在时间上都是比较稳定的。对照雷达观测图像可以看出, 在连续的几个时次, 聚类方法对不同云团系统的划分是正确的, 雷达图像上有两个云团系统, 用深蓝色和浅蓝色表示, 可以看出深蓝色云团与浅蓝色云团不属于同一个系统, 我们也可以从它们面积大小及反射率均值的变化得到进一步的解释。图 4 (见文后彩图) 为连续 15 个时刻 (时间从 2005 年 3 月 22 日 01:56~03:20) 两个云团的面积和反射率变化情况, 从图中可以看出: 深蓝色云团在面积上随时间是先增后减的, 也就是说, 深蓝色云团的面积在 3 月 22 日 02:39 之前是增加的, 然后又开始递减, 云团的平均反射率随时间递减, 在 30~35 dBZ 之间。而浅蓝色云团的面积随时间是一直递增的, 而且增长速度比较快, 云团的平均反射率随时间变化不大, 在 25~27 dBZ 之间。视觉的直观判断和数据分析说明层级聚类方法对云团系统的识别是正确、稳定的。对云团系统的识别是层级聚类方法不同于其他方法之处。

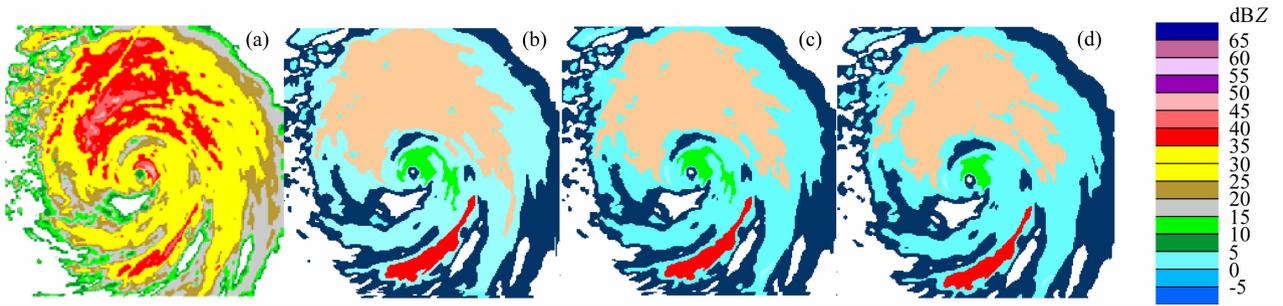


图 2 改变 λ 值对小尺度云团识别的影响：(a) 雷达观测图像；(b) $\lambda=0.6$ ；(c) $\lambda=0.7$ ；(d) $\lambda=0.8$

Fig. 2 The effect of varying λ on small-scale cloud cluster identification: (a) Original radar image; (b) $\lambda = 0.6$; (c) $\lambda = 0.7$; (d) $\lambda = 0.8$

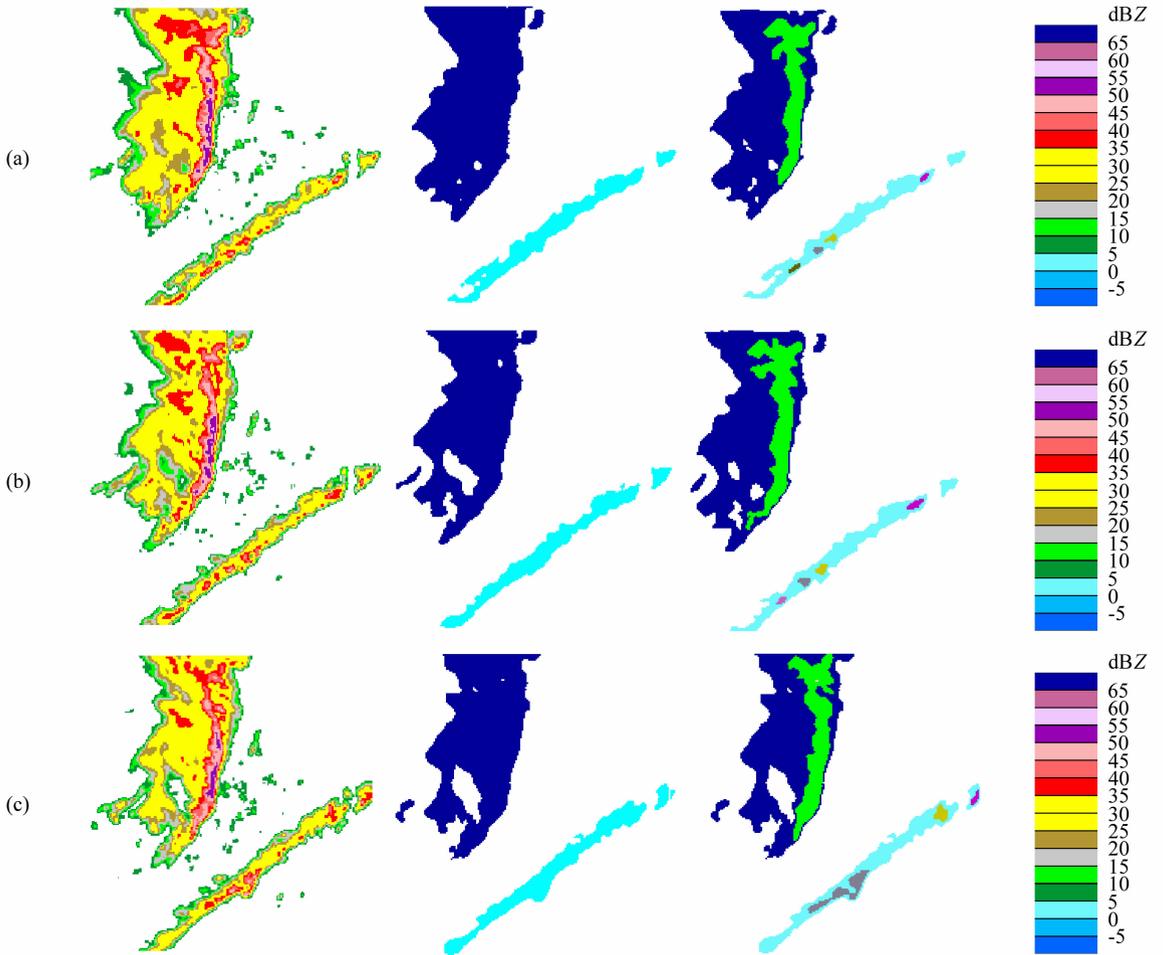


图 3 广州 2005 年 3 月 22 日不同时刻的观测图像 (左)、云团系统识别 (中) 以及云团系统中强对流单体识别 (右)：(a) 02:51；(b) 03:03；(c) 03:15

Fig. 3 Observation image (left), cloud clusters identification (middle) and convective cells identification (right) at different time on 22 Mar 2005 in Guangzhou: (a) 0251 UTC; (b) 0303 UTC; (c) 0315 UTC

层级聚类方法的另一优点是可以对云团进行不同尺度的识别，在识别不同云团系统的同时，可以对云团系统中的强对流单体进行识别，从图 3 中右侧的图像可以看出，在研究的时间段内，深蓝色云

团包含了一个强对流雨带，从连续的几个时刻来看，对强对流雨带的识别也是稳定的。浅蓝色云团中包含了几个强对流单体，从连续的几个时刻可以看出，强对流单体在发展的同时，产生了合并，并

在东北方向新生了一个单体, 对照雷达观测图, 我们也可以看到同样的变化, 这说明聚类方法对强对流单体的识别也是正确的、稳定的。

以上分析说明了聚类方法可以实现云团的多尺度识别, 这样, 在对云团预报的时候, 就可以根据不同的预报时效, 选择不同的云团识别尺度, 这点是不同于其他云团识别方法的。

在对该方法的研究中, 我们还选取了一些暴雨个例进行研究。下面以台风暴雨为例进行分析。图 5 (见文后彩图) 为温州 2005 年 9 月 11 日的卡努台风, 图中给出了 01:59 分到 02:25 分的雷达观测图像和云团识别结果。

图 5a~c 分别为不同时刻雷达观测资料和云团识别结果。从图 5 中三个时刻的识别结果可以看出, 最大面积的强回波比较稳定, 面积变化不大, 较小面积的云团增长、消亡变化比较大, 在 02:25 的图像上新增了紫色和黄色两个云团, 而枣红色的云团在慢慢消亡, 对照雷达观测图, 可以看到类似的变化。

4 预报结果及评价

利用聚类方法对云团识别的一个重要目的就是 对云团未来时刻的位置和强度作出预报。本研究选取广州 2005 年 3 月 22 日的飑线过程进行云团的识别、跟踪及临近预报。图 6 (见文后彩图) 为 2005 年 3 月 22 日 03:09 的雷达观测图像、云团不同尺度识别和 15 分钟、30 分钟、60 分钟的预报图像及对应时刻的雷达观测图像。小尺度云团的识别、跟踪, 用作 15 分钟的预报, 云团系统的识别、跟踪, 用作 30 分钟、60 分钟的预报。

通过预报图像和雷达实测图像的比较, 可以得到预报结果的定性描述, 从图 6 可以看出, 云团 30 分钟以内的预报还是比较准确的, 预报图像和观测图像比较相似, 但随着预报时间的增加, 预报准确率在降低, 这是由于随着预报时间的增加, 云团在不断地发展变化, 60 分钟后云团在逐渐消亡, 而线性外推并不能预报这种变化。在今后的工作中, 应进行云团发展演变规律和预报方法的研究, 以提高预报准确率。

P_{OD} 、 F_{AR} 、 C_{SI} 可以定量地评价预报结果, 经聚类方法识别云团后的预报结果的 P_{OD} 、 F_{AR} 、 C_{SI} 随预报时效的变化情况如表 1 所示。可以看出, 预报

表 1 层级聚类法预报结果的定量评价结果

Table 1 Quantitative estimate results of the hierarchical clustering method

预报时效 Forecast time	P_{OD}	F_{AR}	C_{SI}
15 min	0.88	0.31	0.65
30 min	0.80	0.38	0.48
60 min	0.72	0.50	0.36

结果的好坏与预报时效密切相关, C_{SI} 和 P_{OD} 随预报时效的增加而减少。 F_{AR} 随预报时效的增加而增加。

5 总结和讨论

通过上述讨论和资料分析结果, 可以得到以下几点结论:

(1) 聚类方法虽然在市场研究领域、医学实践中得到了广泛应用, 但在雷达图像的分割上是一种新的尝试, 通过对几个暴雨过程的雷达资料的试验研究, 我们认为基于 K -Means 的层级聚类方法能够实现 对暴雨云团的多尺度识别, 能够识别不同系统的云团, 这比较适合引发我国主要降水的 β 中尺度对流系统, 同时也可以识别出 β 中尺度对流系统中较强的对流单体, 这是该方法的优点。通过对雷达连续时次图像的识别结果分析可以认为: 云团识别结果是稳定、合理的。这种识别云团的方法对参数的变化不是很敏感, 有利于将该方法适用于不同的天气情况。

(2) 在暴雨的临近预报中, 可以根据预报时效的不同, 选择不同的尺度来识别云团, 这是我们进行云团多尺度识别的目的所在。

(3) 云团的跟踪过程中虽然能够得到云团反射率的增长或减弱信息, 但这种信息不是云团的发展演变规律, 并不能预报云团的发展、消亡。为了提高临近预报的准确性, 减小线性外推带来的预报误差, 我们将继续利用其他暴雨过程的雷达资料, 进行云团的多尺度识别及云团的演变规律研究。

参考文献 (References)

[1] Michael D, Gerry W. TITAN: Thunderstorm identification, tracking, analysis, and nowcasting radar-based methodology. *J. Atmos. Oceanic Technol.*, 1993, 1 (6): 785~796

[2] Johnson J T, Pamela M, Witt A, et al. The storm cell iden-

- tification and tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, 1998, **13**: 263~276
- [3] French M, Krajewski W, Cuykendall R. Rainfall forecasting in space and time using a neural network. *J. Hydrol.*, 1992, **137**: 1~31
- [4] Rinehart R, Garvey E. Three-dimensional storm motion detection by conventional weather radar. *Nature*, 1978, **273**: 287~289
- [5] Tuttle J D, Foote G B. Determination of the boundary layer airflow from a single Doppler radar. *J. Atmos. Oceanic Technol.*, 1999, **7**: 218~232
- [6] Browning K A. Conceptual models of precipitation systems. *Wea. Forecasting*, 1986, **1**: 23~41.
- [7] Lakshmanan V, Rabin R, DeBrunner V. Mutiscale storm identification and forecast. *J. Atmos. Res.*, 2003, **4**: 367~380
- [8] Mueller C, Saxen T, Roberts R, et al. NCAR auto-nowcast system. *Wea. Forecasting*, 2003, **18**: 545~561
- [9] Pierce C E, Rbert E, Seed A W, et al. The nowcasting of precipitation during Sydney 2000: An appraisal of the QPF algorithms. *Wea. Forecasting*, 2004, **19**: 7~21
- [10] 杨传风. 强降雹回波系统与地面中尺度系统的分析. *山东气象*, 1994, **14** (4): 13~16
Yang Chuanfeng. The analysis of severe hail echo system and MCS on surface. *Shandong Meteorology* (in Chinese), 1994, **14** (4): 13~16
- [11] 应冬梅, 郭艳. 江西省飑线的雷达回波特征分析. *气象*, 2001, **27** (3): 42~45
Ying Dongmei, Guo Yan. The analysis of the radar echo characteristics of squall line in Jiangxi Province. *Meteorology* (in Chinese), 2001, **27** (3): 42~45
- [12] 杜秉玉, 官莉, 姚祖庆, 等. 上海地区强对流天气短时预报系统. *南京气象学院学报*, 2000, **23** (2): 242~250
Du Bingyu, Guan Li, Yao Zuqing, et al. Nowcasting system of severe convective weather in Shanghai. *Journal of Nanjing Institute of Meteorology* (in Chinese), 2000, **23** (2): 242~250
- [13] 肖艳姣, 汤达章, 李中华, 等. 风暴的自动识别、跟踪与预报. *南京气象学院学报*, 1998, **21** (2): 223~229
Xiao Yanjiao, Tang Dazhang, Li Zhonghua, et al. Storm automatic identification, tracking and forecasting. *Journal of Nanjing Institute of Meteorology* (in Chinese), 1998, **21** (2): 223~229
- [14] 陈家慧, 张培昌. 用天气雷达回波资料作临近预报的 BP 网络方法. *南京气象学院学报*, 2000, **23** (2): 284~287
Chen Jiahui, Zhang Peichang. The back propagation network method for nowcasting for using weather radar data. *Journal of Nanjing Institute of Meteorology* (in Chinese), 2000, **23** (2): 284~287
- [15] 贝耐芳, 赵思雄. 1998 年“二度梅”期间突发强暴雨系统的中尺度分析. *大气科学*, 2002, **26** (4): 526~540
Bei Naifang, Zhao Sixiong. Mesoscale analysis of severe local heavy rainfall during the second stage of the 1998 Meiyu season. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2002, **26** (4): 526~540
- [16] 张小玲, 陶诗言, 张顺利. 梅雨锋上的三类暴雨. *大气科学*, 2004, **28** (2): 187~205
Zhang Xiaoling, Tao Shiyang, Zhang Shunli. Three types of heavy rainstorms associated with the Meiyu front. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2004, **28** (2): 187~205
- [17] Lakshmanan V, Rabin V R, DeBrunner V. Identifying and tracking storms in satellite images. Second Artificial Intelligence Conference, American Meteorological Society, Long Beach, CA, 2000, 90~95
- [18] 李献州, 贺忠. 广州台风特大暴雨预报. *气象*, 1995, **21** (8): 17~22
Li Xianzhou, He Zhong. Forecast of typhoon heavy rainfall in Guangzhou. *Meteorology* (in Chinese), 1995, **21** (8): 17~22
- [19] 盛春岩, 蒲一芬, 高守亭. 多普勒天气雷达资料对中尺度模式短时预报的影响. *大气科学*, 2006, **30** (1): 93~107
Sheng Chunyan, Pu Yifen, Gao Shouting. Effect of Doppler radar data on nowcasting output of mesoscale model. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2006, **30** (1): 93~107

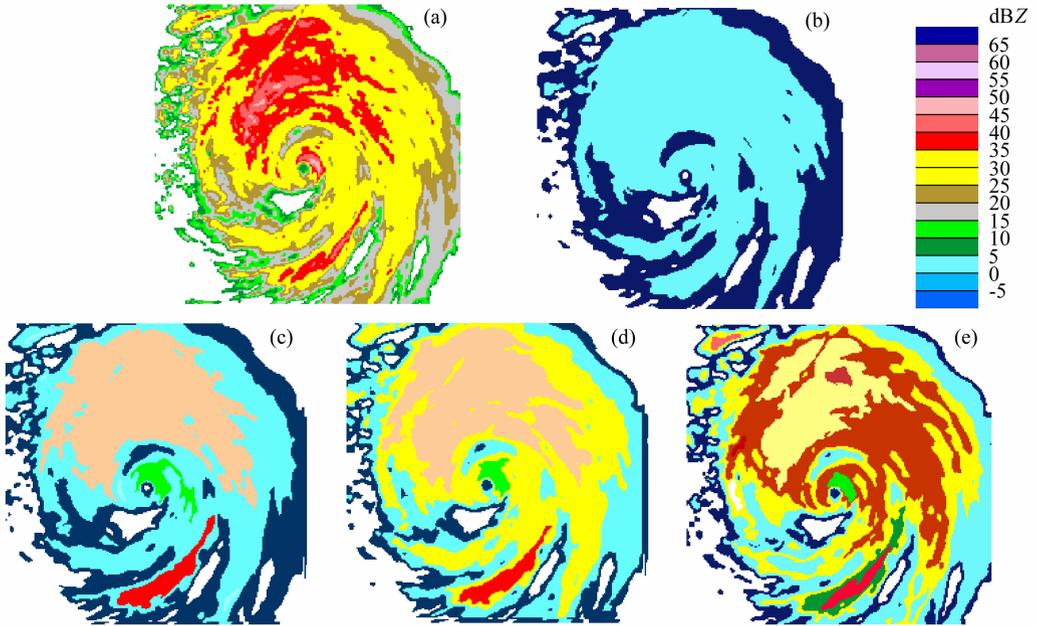


图 1 不同 K 值对聚类结果的影响: (a) 2005 年 9 月 11 日 02: 25 (国际协调时, 下同) 温州雷达观测图像; (b) $K=3$; (c) $K=4$; (d) $K=5$; (e) $K=6$

Fig. 1 The effect of different K on clustering results: (a) Original radar image at 0225 UTC Sep 2005 from Wenzhou; segmentation results at (b) $K = 3$, (c) $K=4$, (d) $K=5$, (e) $K=6$

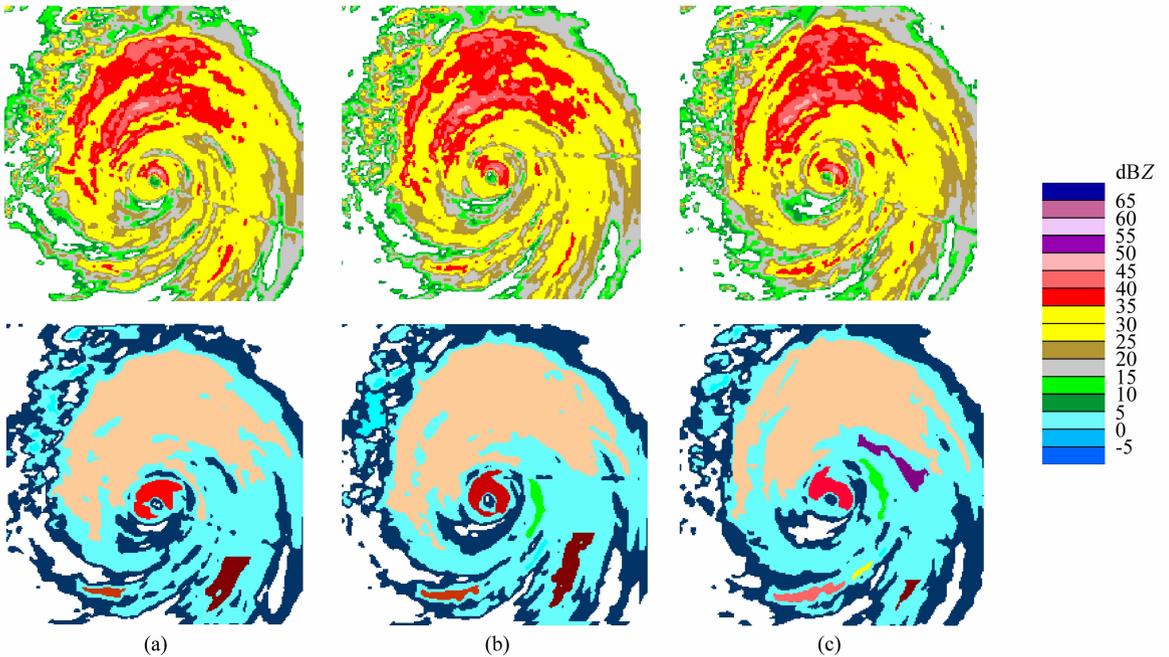


图 5 温州卡努台风 2005 年 9 月 11 日不同时刻的观测图像(上)、台风强螺旋雨带的识别(下): (a) 01:59; (b) 02:13; (c) 02:35

Fig. 5 Radar images (top) and severe spiral rain band identification (bottom) for typhoon Khanun at Wenzhou on 11 Sep 2005: (a) 0159 UTC; (b) 0213 UTC; (c) 0235 UTC

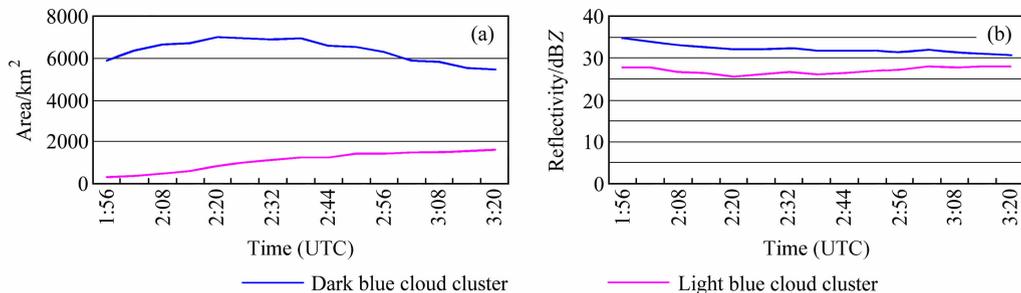


图 4 云团的面积 (a) 和平均反射率因子 (b) 变化曲线
 Fig. 4 The variations of the area (a) and average reflectivity (b) of cloud cluster

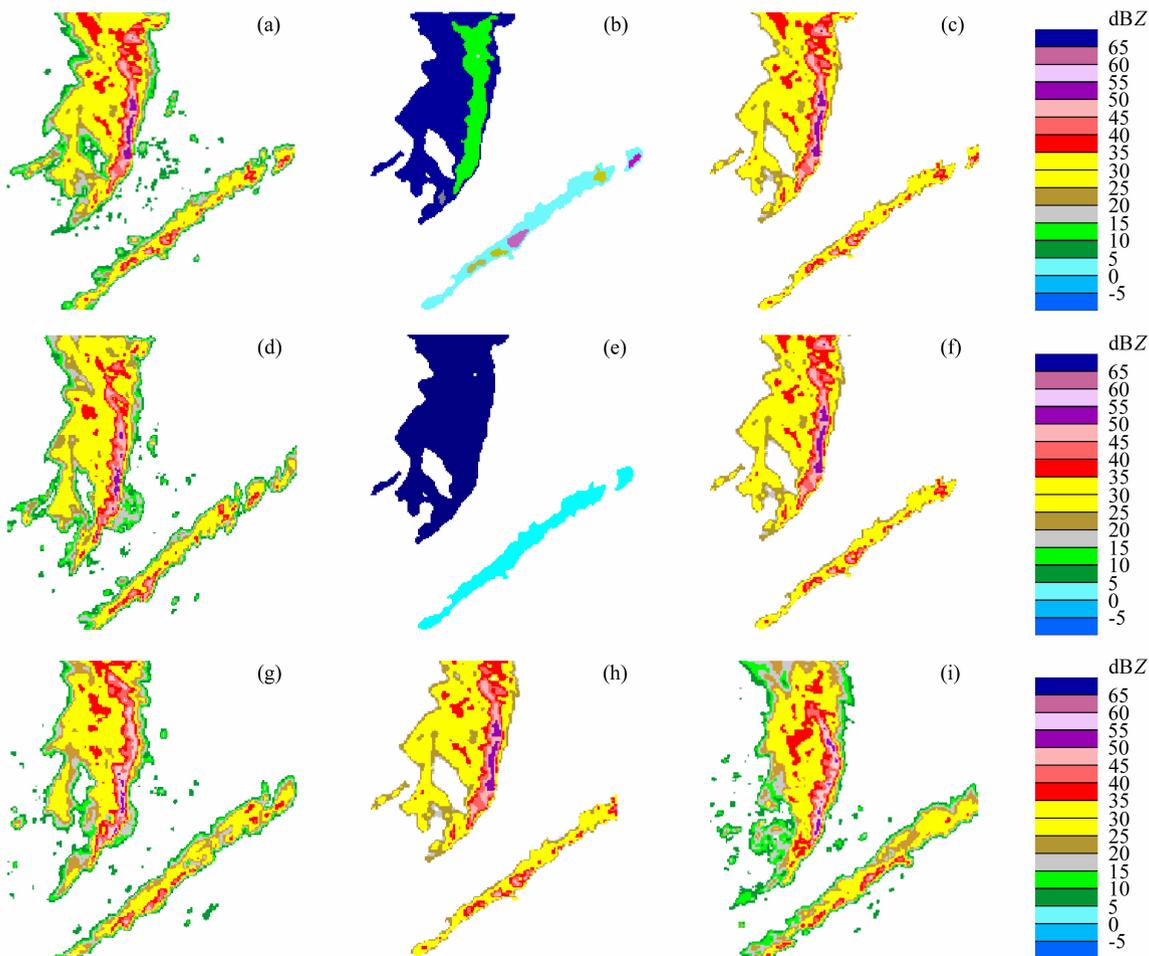


图 6 2005 年 3 月 22 日广州雷达反射率因子聚类及预报: (a) 03: 09 雷达观测图像; (b) 小尺度云团识别; (c) 预报 15 分钟图像; (d) 15 分钟后的雷达观测图像; (e) 云团系统识别; (f) 预报 30 分钟图像; (g) 30 分钟后的雷达观测图像; (h) 预报 60 分钟图像; (i) 60 分钟后的雷达观测图像

Fig. 6 Reflectivity clustering results and nowcasts on 22 Mar 2005 in Guangzhou: (a) Observation image of radar at 0309 UTC in Guangzhou; (b) small-scale cloud cluster identification; (c) image of 15-min forecast; (d) observation image at 0324 UTC; (e) cloud cluster identification; (f) image of 30-min forecast; (g) observation image at 0339 UTC; (h) image of 60-min forecast; (i) observation image at 0409 UTC