doi:10.11676/qxxb2021.062

基于雷达组合反射率拼图和深度学习的中尺度对流系统识别、追踪与分类方法*

南刚强[1,2]　陈明轩[2]　秦　睿[2]　韩　雷[1,2]　曹伟华[2]
NAN Gangqiang[1,2]　CHEN Mingxuan[2]　QIN Rui[2]　HAN Lei[1,2]　CAO Weihua[2]

1. 中国海洋大学, 青岛, 266100
2. 北京城市气象研究院, 北京, 100089
1. *Ocean University of China*，*Qingdao* 266100，*China*
2. *Institute of Urban Meteorology*，*CMA*，*Beijing* 100089，*China*

2021-03-29 收稿, 2021-08-23 改回.

南刚强, 陈明轩, 秦睿, 韩雷, 曹伟华. 2021. 基于雷达组合反射率拼图和深度学习的中尺度对流系统识别、追踪与分类方法. 气象学报, 79(6): 1002-1021

Nan Gangqiang, Chen Mingxuan, Qin Rui, Han Lei, Cao Weihua. 2021. Identification, tracking and classification method of mesoscale convective system based on radar composite reflectivity mosaic and deep learning. ***Acta Meteorologica Sinica*****, 79(6):1002-1021**

Abstract　Mesoscale convective system (MCS) is the main cause of lots of convective weather, which can lead to severe meteorological and hydrological disasters such as thunderstorms, tornadoes and flash floods. Accurate identification and tracking of MCS and the realization of MCS classification based on the tracking trajectory as well as understanding the MCS features are of great importance for the analysis and forecast of catastrophic weather. Based on the radar composite reflectivity mosaic data in the Beijing-Tianjin-Hebei region from 2010 to 2019, the support vector machines (SVM), the random forest (RF), the Extreme Gradient Boosting (XGBoost) and the deep neural network (DNN) are used to develop an automatic recognition algorithm for MCSs in Beijing-Tianjin-Hebei region. Secondly, the tracking and matching of identified MCS slices are completed according to spatiotemporal overlap tracking, and a tracking database of MCS is established, which includes MCS intensity and spatial and temporal information. Finally, on the basis of distinction between linear convection and non-linear convection and starting from three conceptual models and structural characteristics of thee classical quasi-linear MCSs, i.e., the trailing, leading, and parallel stratiform precipition, an algorithm for quasi-linear MCS classification is established based on the area ratios of stratiform and intense convection on both sides of the approximate major axis of MCS and its movement direction, which is calculated according to tracking trajectory. The recognition of MCS is subsumed under binary classification. Taking POD, FAR, CSI and ACC as evaluation indexes, the DNN model is better than the SVM, RF and XGBoost models in MCS recognition after comprehensive comparison. Spatiotemporal overlap tracking is used to track MCS slices identified by the DNN model. The analyses of two tracking examples suggest that the algorithm used in this research has achieved good tracking results, which further demonstates the accuracy and advantage of deep learning in identifying MCS. The accurate realization of MCS classification including TS, LS and PS provides a technical idea for the life cycle prediction of quasi-linear MCS and objective prediction of disasterous weather, especially the intensity, location and duration of short-term heavy precipitation by combining the movement direction of MCS at a single radar snashot with the distributions of stratiform and intense convection in MCS slices.

Key words　Deep learning，Mesoscale convective system，Identifying，Tracking，Classification

---

* 资助课题: 国家自然科学基金项目(41575050 和 41801022)、北京市科技计划课题(Z171100004417008)。

作者简介: 南刚强, 主要从事机器学习算法应用研究。E-mail: gqnan_ouc@163.com

通信作者: 陈明轩, 主要从事短时临近天气预报研究。E-mail: mxchen@ium.cn

摘　要　中尺度对流系统（Mesoscale Convective System, MCS）是很多对流性天气的主要致灾体，可导致严重的气象和水文灾害，如雷暴大风、冰雹、龙卷风和山洪。对MCS进行准确的识别和追踪，并根据追踪轨迹及获得的MCS特征实现MCS的分类，对灾害天气的分析和预报有重要意义。基于京津冀地区2010—2019年的雷达组合反射率拼图资料，分别使用支持向量机（SVM）、随机森林（RF）、极度梯度提升决策树（XGBoost）和深度神经网络（DNN）4种机器学习方法，研发了京津冀地区MCS的自动识别算法。使用时、空重叠追踪法对识别的MCS进行追踪匹配，得到包含强度、时间和空间信息的MCS追踪数据资料。在区分线状对流系统和非线状对流系统的基础上，进一步从经典的尾随层云（Trailing Stratiform, TS）、前导层云（Leading Stratiform, LS）和平行层云（Parallel Stratiform, PS）三类准线性MCS的概念模型和结构特征出发，根据追踪轨迹计算MCS的运动方向和MCS近似长轴两侧层状云和强对流云的面积占比，建立准线性MCS的分类算法。MCS的识别属于二分分类问题，以命中率（POD）、虚警率（FAR）、临界成功指数（CSI）和准确率（ACC）为评价指标，综合对比各项指标发现DNN模型较SVM、RF和XGBoost模型对MCS的识别效果更好。使用时、空重叠追踪法对DNN模型识别的MCS进行追踪，结合对两个追踪实例的分析，发现本研究所用的算法取得了很好的追踪结果，也进一步说明了深度学习方法识别MCS的准确性和优势。根据追踪轨迹计算某时刻MCS的运动方向，结合识别的层状云和强对流云的分布位置，准确实现了TS、LS和PS型准线性MCS的分类，为准线性MCS的生命史预测及其致灾天气特别是短时强降水的强度、位置和持续时间的客观预报提供了一种技术思路。

关键词　深度学习，中尺度对流系统，识别，追踪，分类

中图法分类号　P458.2

1　引　言

中尺度对流系统（Mesoscale Convective System, MCS）是具有旺盛对流性运动的天气系统，其水平尺度大约为10—2000 km，生命期在3 h以上。Schumacher等（2006）研究了美国地区1999—2003年的极端降水事件，发现所有事件中有66%和暖季事件中有74%与MCS有关，并且美国北部几乎所有的极端降雨事件都是由MCS引起的。Schumacher等（2020）研究表明MCS会产生很大比例的暖季降雨，且在气候变暖的情况下，MCS的频率和强度也可能会增大。中国国家气候中心分析结果显示，于1954、1969、1980、1991、1996、1998、1999、2003和2007年发生的特大暴雨洪涝都与MCS存在直接的关联，这些灾害给国民经济和人民生命财产安全造成了重大损失（王晓芳等，2011）。自2012年以来，华北中东部暴雨事件频发（雷蕾等，2020），对社会造成了巨大损失，并且这些暴雨特别是短时强降水的形成均与MCS存在直接关系。因此，做好MCS及其致灾天气的预报、预警，对人们了解暴雨、龙卷风和山洪等气象灾害的发展及演化有很大的帮助。

资料的选择对MCS的研究有着至关重要的影响。从中尺度天气的角度判断，MCS的尺度范围相对较大，且空间变化较广，形态较为复杂，因此近几十年来，气象学家通常使用较大范围的卫星或雷达组网数据进行MCS的监测、识别、追踪和预报（Houze, 2018）。

基于雷达探测资料的常用识别MCS的方法有2类。一类是基于雷达拼图资料的TITAN算法（Thunderstorm Identification, Tracking, Analysis and Nowcasting）（Dixon, et al, 1993）。TITAN属于对流风暴三维特征自动识别、跟踪、分析算法的典型代表，后续经过了多次改进和完善，并在多个临近预报系统中得到应用（Mueller, et al, 2003；韩雷等，2007；Han, et al, 2009；陈明轩等，2006，2010）。另一类是基于雷达拼图资料开发的SCIT算法（Storm Cell Identification and Tracking Algorithm）（Johnson, et al, 1998），并借助Davis等（2006a，2006b）开发的模式评估工具（MODE, Method for Objective-based Diagnostic Evaluation）进行识别。但TITAN和SCIT均属于风暴“质心”识别和追踪算法，对尺度较小的超级单体风暴或孤立的风暴单体的识别效果更好，而对于结构和形态较为复杂的MCS的识别有时不够准确。人们为了能够利用SCIT准确地识别MCS，对SCIT算法进行了一定改进，将SCIT算法中识别的位置比较接近的风暴单体组成MCS，以便对MCS进行跟踪和预报。随着机器学习算法的广泛应用，人们开始借助人工智能来实现MCS的自动识别，Haberlie等（2018a）使用随机森林、梯度提升和极度梯度提升3种分类算法实现了美国MCS的自动识别。

MCS的移动轨迹追踪通常也使用TITAN算法或改进的SCIT算法实现，但是这类风暴“质心”算法也存在与上述识别MCS类似的追踪缺陷。另一种常见的MCS移动轨迹追踪方法是基于雷达回波

的交叉相关追踪(Tracking Radar Echoes by Cross-correlation, TREC)(Rinehart, et al, 1978),该方法同样适用于基于卫星观测资料的MCS追踪。杨吉等(2015)利用TREC和面积重叠算法实现了新的MCS追踪预报方法。最近,曹伟华等(2019)将TITAN算法和TREC算法进行融合,发挥不同识别追踪算法的优势,以提升强对流系统的识别和临近预报水平。但是,TREC算法最大的问题是交叉相关矩阵的计算设置与对流系统回波的尺度密切相关,使得不同尺度对流系统的追踪效果和精度差异较大。对于MCS的追踪,还有Skok等(2009)提出的时间空间目标建立法,但是,该方法有一个很大的弊端,对多个对象的合并(分裂)将导致一个单一的、过度扩展的风暴带。作为一种替代方法,可以使用Lakshmanan等(2009)提出的时、空重叠追踪法,该方法将时、空对象构建过程仅应用于两个相邻时次雷达图像在空间上重叠的风暴。

准线性MCS包含一条对流线,也就是一个连续或接近连续的对流回波链,该回波链共享一个几乎共同的前缘,并以近似串联的方式移动,包括其按照一个接近直线或中等弯曲的弧线方式排列(Parker, et al, 2000)。准线性MCS(如飑线)的分类是研究MCS的一个重要课题,尤其对短时强降水和暴雨特征的研究有重要意义。Parker等(2000)使用2 km分辨率的美国雷达组合反射率因子数据,研究了MCS的主要组织形态,根据对流线和层状云的相对位置将准线性MCS分为尾随层云(Trailing Stratiform, TS)、前导层云(Leading Stratiform, LS)和平行层云(Parallel Stratiform, PS)3类,并研究了每种类型的基本特征,形成了经典的线状MCS分类概念模型。Wang等(2014)借鉴上述工作,利用2010年6—7月长江流域的雷达拼图和观测资料,分析了长江中下游地区梅雨季MCS的类型和特征。Ashley等(2019)使用图像分类和机器学习方法对22 a的美国地区雷达拼图数据进行分割、分类和准线性对流系统(Quasi-Linear Convective Systems, QLCS)追踪,该研究更进一步地说明了自动风暴形态分类的实用性,减少了研究人员手动形态学分类的耗时和时空限制。Jergensen等(2020)使用机器学习并基于雷达探测数据和邻近探空资料,将雷暴有效地分为3类:超级单体、QLCS和无组织对流。

MCS的自动识别、跟踪和分类本身就是一个复杂的工作,涉及到很多核心技术与算法。鉴于此,文中结合机器学习算法来实现MCS的自动识别,将MCS的识别转化为从特定MCS切片中抽取到的样本的预测问题。并且,基于追踪得到的运动轨迹和准线性MCS中TS、LS和PS三种类型的组织结构,提出了新的分类算法,也就是根据MCS运动方向与层状云和强对流云区域在识别的MCS切片中的分布特征,实现对准线性MCS的分类。

文中首先通过分割雷达拼图数据和抽取MCS切片中的特征将MCS的识别转换为二分类问题,并使用机器学习算法训练数据集得到最优分类器进而实现MCS的自动识别。再对机器学习模型识别的MCS进行追踪,得到包含MCS信息的数据集和追踪轨迹。最后根据轨迹矢量与MCS切片拟合椭圆短轴的夹角以及拟合椭圆长轴两侧的层状云和强对流云面积之比,建立准线性MCS的分类算法。

2 模 型

2.1 深度学习简介

深度学习是机器学习的一个重要分支,它能自动地从输入数据中抽取更加复杂的特征,使网络模型的权重学习变得更加简单有效。早期的深度学习受到了神经学的启发,使得深度学习可以胜任很多人工智能的任务,到如今,深度学习已经从最初的图像识别领域扩大到了机器学习的各个领域。

文中使用深度学习中的深度神经网络(Deep Neural Networks, DNN)进行MCS的特征识别,并将其训练所得模型的预测结果与传统的机器学习算法做对比。由于用到的其他3种普通机器学习分类算法(支持向量机(SVM)、随机森林(RF)、极度梯度提升决策树(XGBoost))都是基于开源的Scikit-Learn库(Pedregosa, et al, 2011)实现的,在此不予介绍,读者可参考相关文献。下文将主要介绍DNN模型的实现。

2.2 深度神经网络(DNN)模型

2.2.1 网络结构

文中使用的DNN模型(Bengio, 2009)是一个4层的全连接神经网络结构,包含2个不同节点的隐藏层,第1层为输入层,节点数为MCS样本的特征数量(共14个,后面会详细介绍这些特征的定义);第4层为输出层,含有2个节点,分别对应预测

结果 MCS(标记为 1)和 non-MCS(非 MCS, 标记为 0)。

DNN 模型的主要参数见表 1。表中的 GradientDescent 即梯度下降法, 是一种常用的优化器; Relu 是激活函数, 表达式见式(1), Relu 函数在正区间内的斜率为常数, 避免了模型训练过程中梯度消失的情况, 并且在梯度下降过程中使得模型能够快速收敛。

$$\mathrm{Relu}(x)=\max(0,x) \tag{1}$$

表 1 DNN 模型主要参数

Table 1 Main parameters of the DNN model

	DNN
输入层节点	14
输出层节点	2
激活函数	Relu
优化器	GradientDescent

2.2.2 学习率和损失函数设置

在训练神经网络时, 需要设置学习率来控制网络参数更新速度, 学习率决定了网络参数每次更新的幅度。学习率太小, 会导致模型收敛过于缓慢, 进而增加训练的时间成本, 有时甚至导致模型出现“无学习能力”的情况; 学习率太大, 使得模型无法靠近或达到最优解, 最终导致模型无法收敛。为了解决此问题, 使用指数衰减法来控制学习率的变化, 使模型趋于最优解。

$$\mathrm{lr}=\mathrm{lr\_base}\times\alpha^{\frac{\mathrm{train\_step}}{\mathrm{decay\_step}}} \tag{2}$$

式中, lr 是学习率; lr_base 是初始学习率; α是小于 1 的衰减率, 在本试验中取 0.99; decay_step 是常数, 表示衰减速度; train_step 是训练轮次。

损失函数是模型优化的对象, 通过最小化损失函数使模型达到收敛状态, 减少模型预测值的误差。本试验解决的是二分类问题, 所以用交叉熵作为该模型的损失函数。交叉熵用来刻画两个概率分布的距离, 对于两个特定的概率分布 p 和 q, 交叉熵的计算方法为

$$H(p,q)=-\sum_{x}p(x)\ln(q(x)) \tag{3}$$

在本试验中, p 表示样本的标签, q 表示网络输出结果的概率分布。

根据本研究的需要, 为了将神经网络的输出结果转化为概率分布, 用 Softmax 回归作为网络输出层的额外处理层。假设原始网络的输出为 $y_i(i=1,2,\cdots,n)$, 则经过 Softmax 回归处理后的结果为

$$S_i=\mathrm{softmax}(y_i)=\mathrm{e}^{y_i}\Big/\sum_{i=1}^{n}\mathrm{e}^{y_i}\qquad i=1,2,\cdots,n \tag{4}$$

2.2.3 过拟合问题

在神经网络的训练过程中, 模型经常会出现过拟合的情况, 也就是模型在训练集上的拟合效果很好, 但在测试集上的预测值和真实值差异却很大。为了解决训练得到的模型出现过拟合问题, 通常会在损失函数中引入正则化。正则化就是在损失函数中加入刻画模型复杂度的指标来限制权重的大小, 进而减小训练数据中的随机噪声对模型拟合的影响。常用的有 L1 正则化和 L2 正则化

$$R_{\mathrm{L1}}(w)=\sum_{i}\left|w_i\right| \tag{5}$$

$$R_{\mathrm{L2}}(w)=\sum_{i}\left|w_i^2\right| \tag{6}$$

式中, w 表示网络的权重, 模型的参数复杂度由网络的所有权重系数(w)决定。L1 正则化更趋向于产生一个稀疏模型, 而 L2 正则化可以更好地防止模型过拟合, 故本试验使用 L2 正则化。假设模型的损失函数为 $L(\theta)$, 正则化系数为 λ, 则引入 L2 正则化后的优化函数如下

$$\mathrm{Loss}=L(\theta)+\lambda R_{\mathrm{L2}}(w) \tag{7}$$

此时, 在优化模型时会直接优化 Loss 函数, 而不是损失函数 $L(\theta)$。需要特别说明的是, 本试验为了增加 DNN 模型在测试集上的健壮性(即模型稳定高效且性能优越), 引入了滑动平均模型。在采用梯度下降法训练神经网络时, 使用滑动平均模型在很多应用中都可以一定程度上提高最终模型在测试数据上的性能。简单来说, 就是数据每次训练得到的模型都受到之前模型的影响, 进而影响后面模型的训练, 这个影响随着训练次数的增加而减小, 这样可以让模型的训练更加趋于稳定。

3 试验设计

3.1 试验数据及预处理

文中所用的雷达拼图数据的格点分辨率为 1 km×1 km, 覆盖整个京津冀地区, 区域大小为 800 km×800 km, 时间间隔为 6 min。该数据具有高

时、空分辨率特征，并且覆盖范围较广，非常适合于京津冀地区 MCS 的识别与追踪。该雷达拼图数据是北京自动临近预报系统(BJ-ANC)的产品(陈明轩等, 2010)，BJ-ANC 系统在形成上述雷达拼图资料过程中对京津冀地区每部雷达基数据均进行了较为严格的质量控制，包括地物杂波、超折射回波、0℃层亮带回波的自动识别和剔除(陈明轩等, 2010)，这里不再赘述。

京津冀地区原始雷达拼图数据的投影坐标为非等间隔经纬度投影，为了方便后面试验的进行，需要对每个网格的经度和纬度等间隔化。经处理后每个网格在地理上的大小近似为 1 km^2，数据的经纬度范围(36.21°—43.40°N, 112.03°—120.90°E)。这样处理只是细微地改变了每个网格点的经纬度，每个网格的值依旧保持不变。因为夏季是京津冀地区 MCS 的高发季节，并且要得到足够多的样本来训练模型，故选择 2010—2019 年中 5—9 月的数据进行试验，其中 2010 和 2014 年缺失 5 月的数据。

3.2 数据分割及 MCS 切片提取规则

为了用机器学习模型识别雷达拼图中的 MCS，首先需要分割雷达拼图数据得到候选 MCS 切片，进而抽取样本特征。这里的 MCS 切片，是指通过搜索满足特定阈值大小和强度标准的雷达回波图像中的相连通像素组，而组合得到的雷达探测强对流区域，用该 MCS 切片表示单个时刻 MCS 的空间强度和形态特征。在本研究中，参考 Parker 等(2000)的工作(简称 PJ00 标准)，PJ00 标准将 MCS 定义为一个至少持续 3 h 且包含连续或半连续深湿对流的降水区域，该降水区域的长轴不小于 100 km。根据 PJ00 标准，分割雷达拼图数据中 MCS 切片的阈值如表 2 所示，其中对流区域搜索半径和层状云区域搜索半径并不是唯一的，对流区域搜索半径的常用取值有 6、12、24 和 96 km，而层状云区域搜索半径的常用取值有 48、96 和 102 km。根据 Haberlie 等(2018b)关于美国中纬度地区 MCS 追踪的研究，对流区域搜索半径取 24 km、层状云区域搜索半径取 96 km 时，追踪效果最好。所以本试验这两个指标也分别设为 24 和 96 km 进行雷达拼图数据的分割及 MCS 的追踪。

以图 1 所示原始雷达拼图数据为例，分割过程可以总结为以下 3 个步骤：(1)确定至少包含一个强对流回波(≥50 dBz)像素的对流回波(≥40 dBz)区域，并将面积大于 40 km^2 的对流区域选定，如图 2a 中黑色实线标记的区域；(2)如果选定的对流区域的距离在指定半径 24 km 内，则将它们连接，若连接后区域的最佳拟合椭圆的主轴长度(即 MCS 核长度)至少为 100 km，则将其视为候选 MCS 核，如图 2b 黑色实线区域；(3)将指定半径 96 km 内的层状云回波(≥20 dBz)区域与其各自的候选 MCS 核相关联，并用黑色轮廓线勾勒出最终的候选 MCS 切片，如图 2c 所示。

表 2 用于分割雷达拼图中 MCS 的指标阈值

Table 2 Various thresholds used to segment MCS in radar mosaic data

指标名称	阈值
层状云(dBz)	20
对流(dBz)	40
强对流(dBz)	50
对流区域面积(km^2)	40
MCS核长度(km)	100
对流区域搜索半径(km)	24
层状云区域搜索半径(km)	96

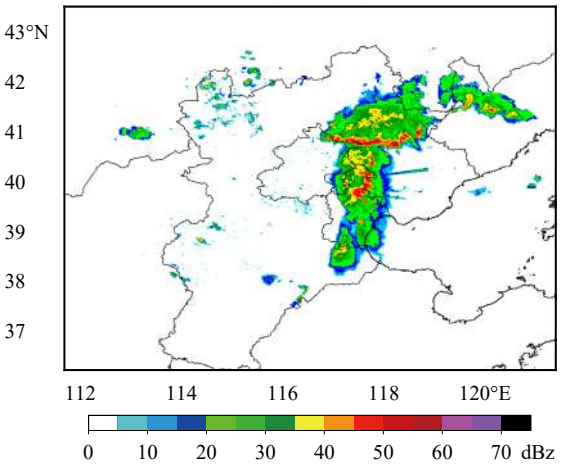


图 1 原始雷达拼图数据

(2014 年 6 月 17 日 11 时 59 分 36 秒(世界时, 下同))

Fig. 1 Original radar mosaic data

(11: 59: 36 UTC 17 June 2014)

3.3 MCS 特征化及识别

为了实现文中的 MCS 分类目标，必须将 MCS 切片信息具体特征化从而得到训练样本。每个 MCS 特征的选择是参考先前的相关研究而确定的(Haberlie, et al, 2018a)，并使用 Scikit-Image(van der Walt, et al, 2014)中的图像处理函数来完成特征值计算。共选取 14 个 MCS 特征，可以简单将其

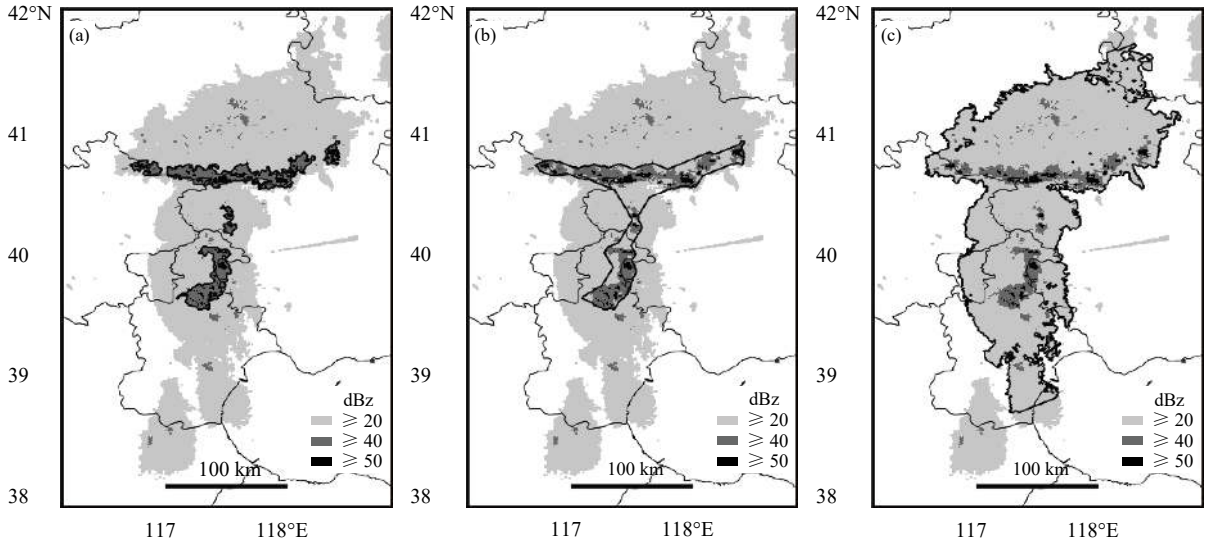


图 2　使用雷达拼图数据（2014 年 6 月 17 日 11 时 59 分 36 秒）演示候选 MCS 切片的分割过程
（a. 包含强对流单元且面积大于 40 km^2 的对流区域；b. 连接指定半径 24 km 内的对流区域，将主轴长度超过 100 km 的连接区域确认为 MCS 核；c.关联 MCS 核指定半径 96 km 内的层云区域得到候选 MCS 切片）

Fig. 2　Demonstration of segmentation steps for candidate MCS slices using radar mosaic data (11: 59: 36 UTC 17 June 2014)
(a. convection areas greater than 40 km^2 with intense convection; b. connected convection area within a specified radius (24 km), and the connected area is considered to be the MCS core if its major axis length is at least 100 km; c. candidate MCS slice is identified by connecting the strtatiform pixels that are within the specified radius (96 km) of MCS core)

分为面积特征、比值特征、几何特征和统计特征，具体参见表 3。

由于每个网格的面积是 1 km^2，因而面积特征大小即为满足阈值的网格数。14 个 MCS 特征的计算都比较简单，含义也很明确，此处对较复杂的几何特征做一些简单说明。几何特征主要涉及到 MCS 拟合椭圆和凸包两大形态，对应的相关特征就是拟合椭圆的长轴、短轴和离心率以及凸包区域的面积。凸包（图 3a）是将不规则图形的最外层点连接起来而得到的凸多边形，即该不规则图形的最小外接凸多边形。拟合椭圆是指与不规则图形区域具有相同标准二阶中心矩的椭圆（图 3b），即最佳拟合椭圆。离心率是该椭圆的焦距与长轴之比，用来衡量椭圆的扁平程度，取值范围为（0, 1），离心率越大椭圆越扁平。

抽取完每个候选 MCS 切片的 14 个特征后，为每个样本主观分配 MCS 和 non-MCS 标签，将其制作成含有大量样本的数据集，并将数据集按照年份划分为训练集和测试集，具体见表 4。数据集的划分遵循以下 2 个原则：（1）训练集和测试集的比例要适当，既要保证足够多的样本来训练模型，也要有充足的测试集来评估模型的性能，通常按照 7∶3 的比例划分训练集和测试集；（2）要保证训练集中正、负样本的平衡性。训练集用来训练分类器得到最优的机器学习模型，而测试集则用作独立数据来评估模型的分类性能，根据最优模型来识别候选 MCS 切片是否为真实的 MCS。如前所述，文中用 4 种常见的机器学习算法作为试验的分类器，分别是 RF、SVM、XGBoost 和 DNN，前 3 种算法都是基于 Scikit-Learn 库实现，属于传统机器学习算法，对解决二分类问题有很好效果。DNN 模型是基于 Tensorflow 框架搭建的全连接层神经网络，该模型的可调控参数较多，优化器和损失函数的选择较为灵活，并且可以调用 GPU 加速模型的训练速度，都极大提高了模型的潜力和应用空间。

3.4　MCS 追踪

根据 PJ00 标准，从对流系统的结构规模来看，由对流单体或者对流簇形成的 MCS 及其伴随的中尺度环流必须持续足够长的时间。鉴于此准则，对雷达拼图中的 MCS 进行追踪，必须满足如下条件：（1）尺度和强度要求的分块必须在时间序列上进行时、空关联；（2）该关联必须至少持续 3 h 以上。追踪的目的是在时间和空间上关联机器学习模型识别出的 MCS 切片，以生成包含强度、空间和时间信

表 3 MCS 样本特征列表

Table 3 Sample features of MCS

	特征	定义
面积特征	面积	切片总面积(km²)
	强对流面积	超过强对流阈值的像素所覆盖面积(km²)
	对流面积	超过对流阈值的像素所覆盖面积(km²)
比值特征	面积-凸包面积比值	总面积与凸包面积之比
	强对流-层状云比值	强对流面积与总面积之比
	强对流-对流比值	强对流面积与对流面积之比
	对流-层状云比值	对流面积与总面积之比
几何特征	椭圆长轴长度	切片最佳拟合椭圆的长轴长度(km)
	椭圆短轴长度	切片最佳拟合椭圆的短轴长度(km)
	椭圆离心率	切片最佳拟合椭圆的离心率
	凸包面积	切片的外接多边形所覆盖的面积(km²)
统计特征	方差	切片区域像素值的方差
	平均值	切片区域像素值的均值(dBz)
	最大值	切片区域像素值的最大值(dBz)

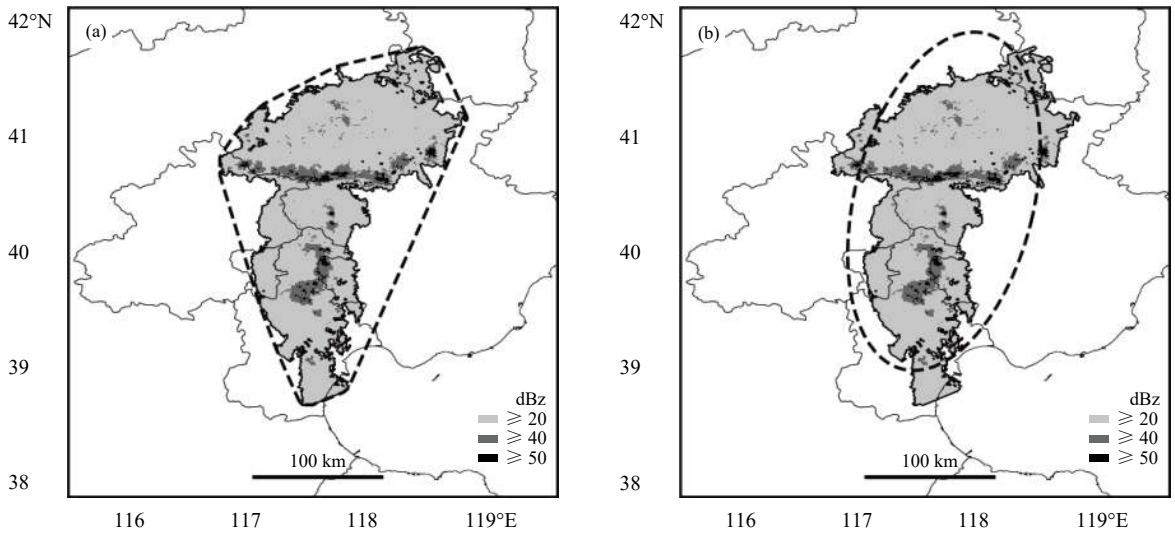


图 3 MCS 切片的凸包(a)和拟合椭圆(b)示意

Fig. 3 Convex hull (a) and fitting ellipse (b) of MCS slice

息的 MCS 条带数据集,并根据追踪轨迹实现准线性 MCS 中 TS、LS 和 PS 三种模型的特征分类。

本试验使用时空重叠追踪法(Lakshmanan, et al, 2009)进行 MCS 追踪,该方法对两个相邻时次雷达拼图在空间上相重叠的风暴进行匹配。对于 2018 和 2019 年 5—9 月的所有时间间隔为 6 min 的测试集雷达数据,根据 DNN 模型识别 MCS 的评估结果确定分类阈值为 0.5,依此阈值来选择当前时刻和下一时刻的 MCS 切片。匹配过程中将建立一个二维矩阵,“矩阵行”表示在现有追踪轨迹内的一个当前时刻 MCS 切片,“矩阵列”表示下一时刻未经匹配的 MCS 切片。分别计算前、后 2 个时刻重叠的 MCS 切片的相似度,根据最小相似度进行匹配并确定追踪的 MCS 回波轨迹。此处的相似度是指经过最大值归一化后的两个长度为 14 的样本特征之间的欧几里德距离。对于下一个时刻未匹配的 MCS 切片,则将其视为新追踪轨迹的起始,并为其分配新的 MCS 序号用于后续的追踪匹配。

如图 4 所示,分别计算 MCS 切片 N 与 S_1、S_2 的欧几里德距离,当前时刻切片 N 与下一时刻切片

表 4 不同类别和年份的训练集和测试集样本数
Table 4 Training and testing counts by classification and year

	年份	MCS样本数	non-MCS样本数
训练集	2010	741	635
	2011	1606	903
	2012	1367	1018
	2013	2278	1070
	2014	965	1042
	2015	516	323
	2016	2145	1848
	2017	1445	1644
	总计	11063	8483
测试集	2018	967	1737
	2019	1765	1211
	总计	2732	2948

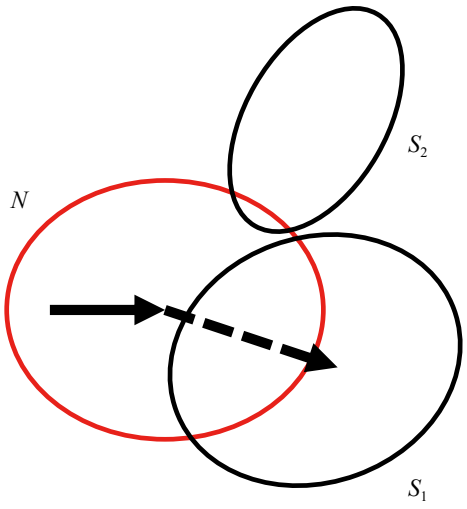


图 4 追踪过程示意（N 为当前时刻的 MCS 切片，S_1 和 S_2 为下一时刻的 2 个 MCS 切片）

Fig. 4 Tracking process (N is a MCS slice at the current moment, S_1 and S_2 are the two MCS slices at the next moment)

S_1 更相似，所以追踪轨迹指向 S_1（图中虚线箭头所指方向）。切片 S_2 则被标记为新的 MCS 并用于后面的追踪，依此类推。显然，对于前后 2 个时刻只有一个重叠的切片，则该算法就类似于简单的重叠匹配；如果存在多个重叠切片，则选择最为相似的切片与现有的追踪轨迹相关联。

3.5 准线性 MCS 分类

根据准线性 MCS 的定义，首先用主观判断法从各 MCS 切片的雷达回波图中选择满足定义的准线性 MCS；再根据追踪得到的 MCS 轨迹矢量，计算 MCS 正方向与轨迹矢量的夹角以及层状云和强对流云在拟合椭圆长轴两侧的占比，从而建立准线性 MCS 的分类算法。

（1）MCS 正方向定义

定义沿 x 轴的正方向为基准，根据 MCS 切片的最佳拟合椭圆长轴的斜率 k 来确定椭圆短轴的正方向。若 $k\geqslant 0$，则以右下侧短轴为正方向；若 $k<0$，则以右上侧短轴为正方向，如图 5 所示。

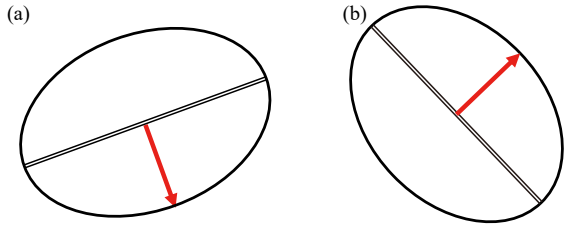


图 5 MCS 正方向的定义
（a. $k\geqslant 0$, b. $k<0$; 红色箭头为短轴的正方向）

Fig. 5 Definition of the positive direction of MCS
(a. $k\geqslant 0$, b. $k<0$; red arrow is the positive direction of the minor axis)

（2）MCS 分类特征计算

根据前述 TS、LS 和 PS 三种类型 MCS 的气象学特征，在此定义 3 个特征来实现 3 类 MCS 的分类，分别为短轴正方向与轨迹矢量的夹角（θ）、长轴两侧层状云区域面积比值（R_S）和长轴两侧强对流区域面积比值（R_I）。R_S 和 R_I 是正方向一侧的面积与负方向一侧的面积之比。轨迹矢量是当前 MCS 到下一时刻 MCS的运动方向，在数学上，夹角的取值范围[0, 180°]，此处为了区分正负方向的角度，当 $\theta>90°$ 时，将其转换为 $\theta-180°$。此时，夹角（θ）的取值范围[−90°, 90°]，其中[0, 90°]表示 MCS 沿短轴正方向运动，[−90°, 0]表示 MCS 沿短轴负方向运动。根据定义的上述特征对 TS、LS 和 PS 型 MCS 进行分类，如表 5 所示（表格中的 thre 是分类阈值，根据 R_I 的计算结果及分类正确率，本试验 thre 的取值为 10）。

表 5 TS、LS 和 PS 型 MCS 的分类规则
Table 5 MCS classification rules for TS, LS and PS

	$R_I\geqslant$thre	$R_I\leqslant$1/thre	$R_S\approx 1$ 且 1/thre$<R_I<$thre
$\theta(0, 90°)$	TS	LS	PS
$\theta(-90°, 0)$	LS	TS	

4 试验结果

4.1 检验方法

文中试验属于有监督机器学习中的分类问题，所以用基于“观测”与“预测”按类别分类后列出频率表进行统计，通常将该表称为混淆矩阵(Zheng, 2015)，如表 6 所示。表中 TP 表示实际样本为 MCS、模型预测也为 MCS; FP 表示实际样本为 non-MCS、但模型将其预测为 MCS; FN 表示实际样本为 MCS、但模型将其预测为 non-MCS; TN 表示实际为 non-MCS、模型预测也为 non-MCS。也就是说，TP 和 TN 都是分类正确的度量值，而 FP 和 FN 都是分类错误的度量值。

表 6 预测和实际标签的混淆矩阵

Table 6 Confusion matrix for predictions and actual labels

		实际	
		MCS	non-MCS
预测	MCS	TP	FP
	non-MCS	FN	TN

根据混淆矩阵的统计结果，计算命中率(probability of detection, POD)、虚警率(false alarm ratio, FAR)、临界成功指数(critical success index, CSI)和准确率(accutacy, ACC)对结果进行综合评估。各评分标准的计算公式如下

$$\mathrm{CSI}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{8}$$

$$\mathrm{FAR}=\frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} \tag{9}$$

$$\mathrm{POD}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{10}$$

$$\mathrm{ACC}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}} \tag{11}$$

4.2 MCS 识别结果分析

使用训练好的 SVM、RF、XGBoost 和 DNN 四个模型分别对测试集样本进行 MCS 识别，得到各个模型的混淆矩阵，如表 7 所示。可以发现在测试集上，XGBoost 模型对应的 TP 值最大，SVM 模型对应的 TP 值最小，且二者相差较大，说明 XGBoost 模型对 MCS 类的识别效果最好，达到 91.22%，而 SVM 模型对 MCS 类的识别效果最差，仅为 88.10%。对于这一点，在 FN 上也得以很好的体现，在测试集的 2732 个 MCS 类样本中，SVM 模型将其中 325 个样本预测为 non-MCS，而 XGBoost 模型对应的该值为 240。对于 non-MCS 类样本的预测，DNN 模型取得了最高的准确率，对测试集中 non-MCS 类的分类正确率达到了 90.16%，SVM 模型仅次之。

表 7 SVM、RF、XGBoost 和 DNN 模型在测试集上的混淆矩阵

Table 7 Confusion matrix of the SVM, RF, XGBoost and DNN models on testing set

	TP	FN	FP	TN
SVM	2407	325	303	2645
RF	2470	262	408	2540
XGBoost	2492	240	409	2539
DNN	2428	304	290	2658

混淆矩阵仅仅展示了模型预测效果的频率，为了更全面地对比这 4 个模型的分类性能，根据混淆矩阵计算它们各自的 CSI、POD、FAR 和 ACC，如表 8 所示。DNN 模型的 CSI 值最高，达到 0.8034，这充分说明了 DNN 模型整体上对 MCS类识别的性能优于其他模型，再结合 ACC，更体现出 DNN 模型的优良性能。POD 值反映了模型对正样本 MCS 类的识别率，XGBoost 模型的 POD 值最高，达到 0.9112，与前面对混淆矩阵的分析是极度吻合。而 FAR 值的大小反映了模型将负样本 non-MCS 类别识别为 MCS 类所占的比重，DNN 模型的 FAR 值最小，说明其对 non-MCS 有很高的识别率。

表 8 SVM、RF、XGBoost 和 DNN 模型在测试集上的评分

Table 8 Scores of the SVM, RF, XGBoost and DNN models on testing set

	CSI	POD	FAR	ACC
SVM	0.7931	0.8810	0.1118	0.8894
RF	0.7866	0.9041	0.1418	0.8820
XGBoost	0.7934	0.9122	0.1410	0.8857
DNN	0.8034	0.8887	0.1067	0.8953

综合来看，DNN 模型对 MCS 的识别性能优于其他 3 种机器学习模型，但该模型也存在一定缺点：对 MCS 类的识别正确率次于 XGBoost 和 RF 模型。考虑到后面的 MCS 轨迹追踪，若模型将 non-MCS 类预测为 MCS 类的次数较多，则会导致轨迹追踪出现一些属于非 MCS 的部分，对追踪结果正确性的影响会比较大；若模型将个别时刻雷达

拼图中的MCS识别为non-MCS，中断的追踪路径可以重新再匹配进行连接，对整体的轨迹追踪不会有太大影响。因此，后面将选择使用DNN模型识别的MCS切片信息进行追踪，进而生成MCS条带数据。

4.3 MCS追踪结果分析

本节主要选取2个具体的MCS个例来分析追踪结果，分别发生在2019年5月17日09时24分—15时和2019年7月13日13时42分—22时54分。追踪结果的分析以下面原则为切入点：(1)若未匹配的追踪结果不连续，则重点分析断点处的雷达拼图是否为MCS；(2)若未匹配的追踪结果是连续的，则重点分析其轨迹起始处的雷达拼图是否为MCS。据此，对MCS生命期内的追踪结果进行主观分析。

(1)2019年5月17日MCS个例

图6显示了2019年5月17日的MCS发展演变过程，组成该MCS每个时刻的MCS切片样本由DNN模型识别，并且将分类阈值设置为0.5。当模型对样本的预测值不小于0.5时，将该样本对应的候选MCS切片进行追踪合并。该MCS始于09时24分，此时对流云团基本处于北京北部，并一路向南移动，至13时06分结束，持续近4 h，主要影响北京、廊坊和天津等地。

该时段的MCS轨迹是不连续的(最下面有两条断开的轨迹)。查看实际雷达拼图数据发现，13时06—56分的雷达拼图数据缺失，但13时56分—14时30分的雷达数据正常，原始数据如图7所示，分割后的MCS切片如图8所示，并且DNN模型将其识别为MCS，生成的追踪数据也对该时段的MCS进行了关联。

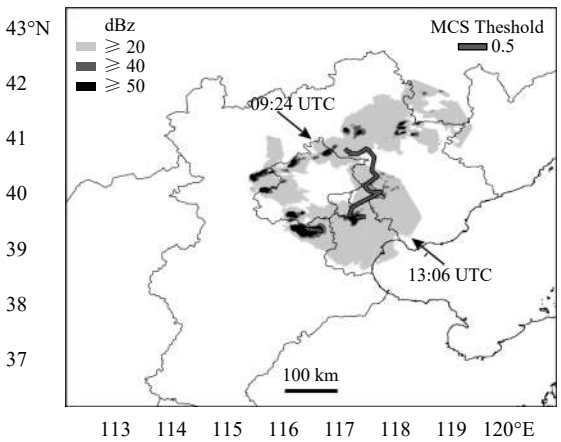


图6 2019年5月17日09时18分—15时MCS追踪轨迹

Fig. 6 Tracking path of MCS during 09: 18—15: 00 UTC 17 May 2019

试验结果表明，如果深度学习模型预测候选MCS样本的值未达到0.5，则会造成MCS的不连续，同时，某时段雷达拼图数据的缺失也会导致MCS的轨迹追踪中断，在这两种情况下时、空匹配过程将无法创建连续的MCS条带。尽管使用较高概率阈值的目的是减少non-MCS事件的错误识别，但实际情况表明，此方法也可能会删除或截断合理的MCS区域。由于匹配过程仅检查当前时刻和下一个6 min时刻的MCS切片匹配，因此，如果模型对某一个雷达拼图中的MCS切片的预测值未超过分类阈值，则追踪结束。

解决该问题的一种方法是重新分析追踪数据库来连接以前未连接的轨迹，也就是尝试将包含至少2个切片的条带末端(持续时间为12 min)连接到具有至少2个切片的条带开始端。要找到合适的匹配项，规定必须满足以下条件：(1)匹配的候选MCS条带的开始时间距上一个MCS条带的结束时间不超过60 min；(2)匹配的候选MCS条带的第一个切片与前一个条带的最后一个切片必须重叠或者相距100 km之内。图9是一个经过匹配的追踪轨迹，此时MCS的起止时间分别为09时24分和14时30分，很明显该MCS条带较未匹配前在结尾处有延伸(图9红色虚线标注区域)，整个轨迹是连续的(与图6对比)。

(2)2019年7月13日MCS个例

图10显示了2019年7月13日的一个MCS过程，雷达观测该MCS大约始于13时42分，并一路向东南方向移动，途径北京、天津及河北东部，并经渤海湾进入山东省境内，至22时54分逐渐减弱消退，持续超过9 h。

对DNN模型识别的MCS切片进行重新分析匹配，追踪轨迹如图11所示。显然，该MCS的轨迹较未匹配前有所延长(红色虚线标注区域)，延长区域主要分布在河北省北部，并靠近北京市北部。这是由于DNN模型将某时刻MCS分类为non-MCS导致的中断，匹配后对其重新建立了连接。

对上述MCS个例轨迹追踪中18时42分—

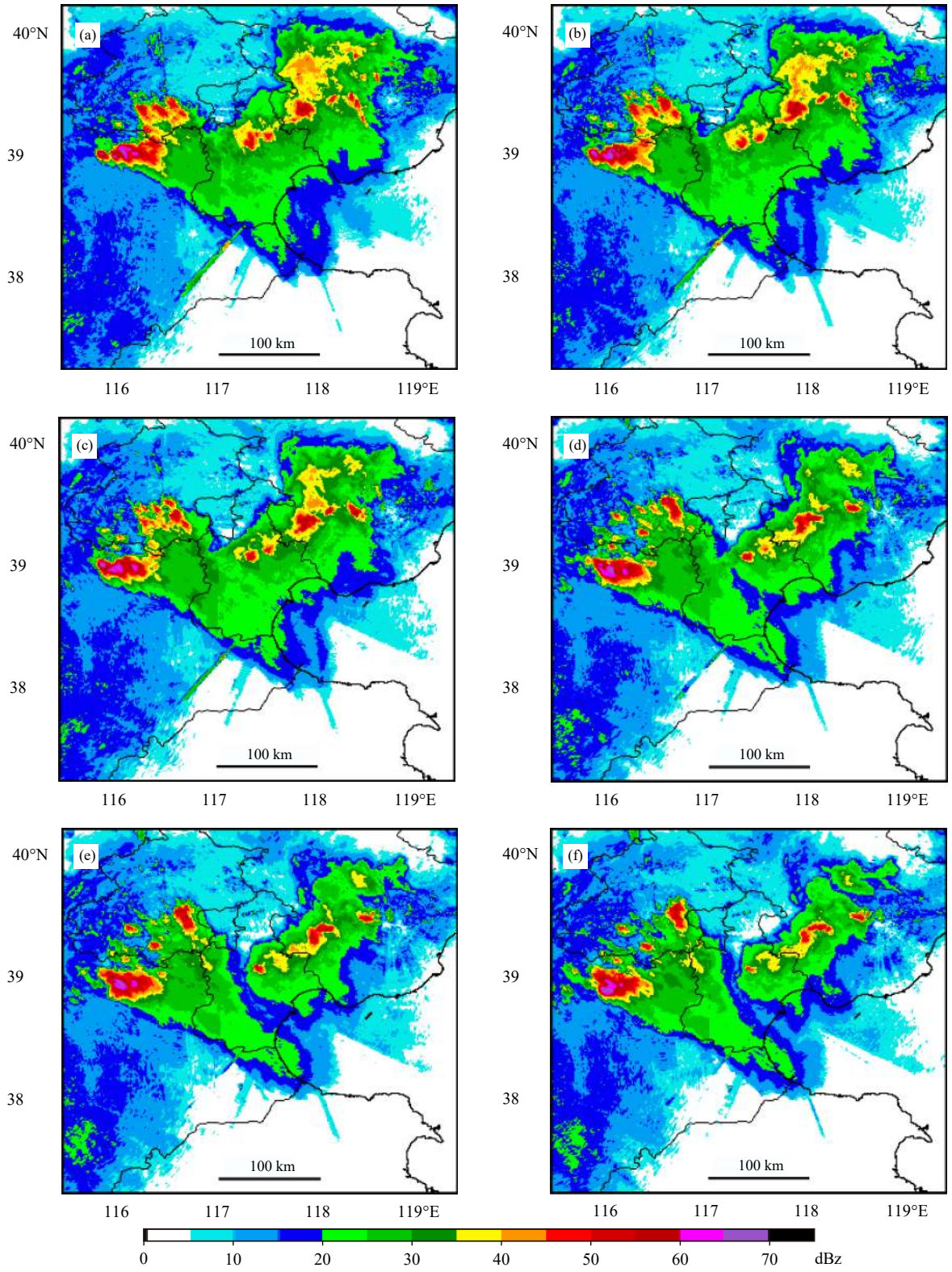


图 7 2019 年 5 月 17 日 13 时 56 分—14 时 30 分原始雷达拼图数据

（a—f, 时间间隔: 6 min）

Fig. 7 Original radar mosaic data at 13: 56—14: 30 UTC 17 May 2019

（a—f, interval: 6 min）

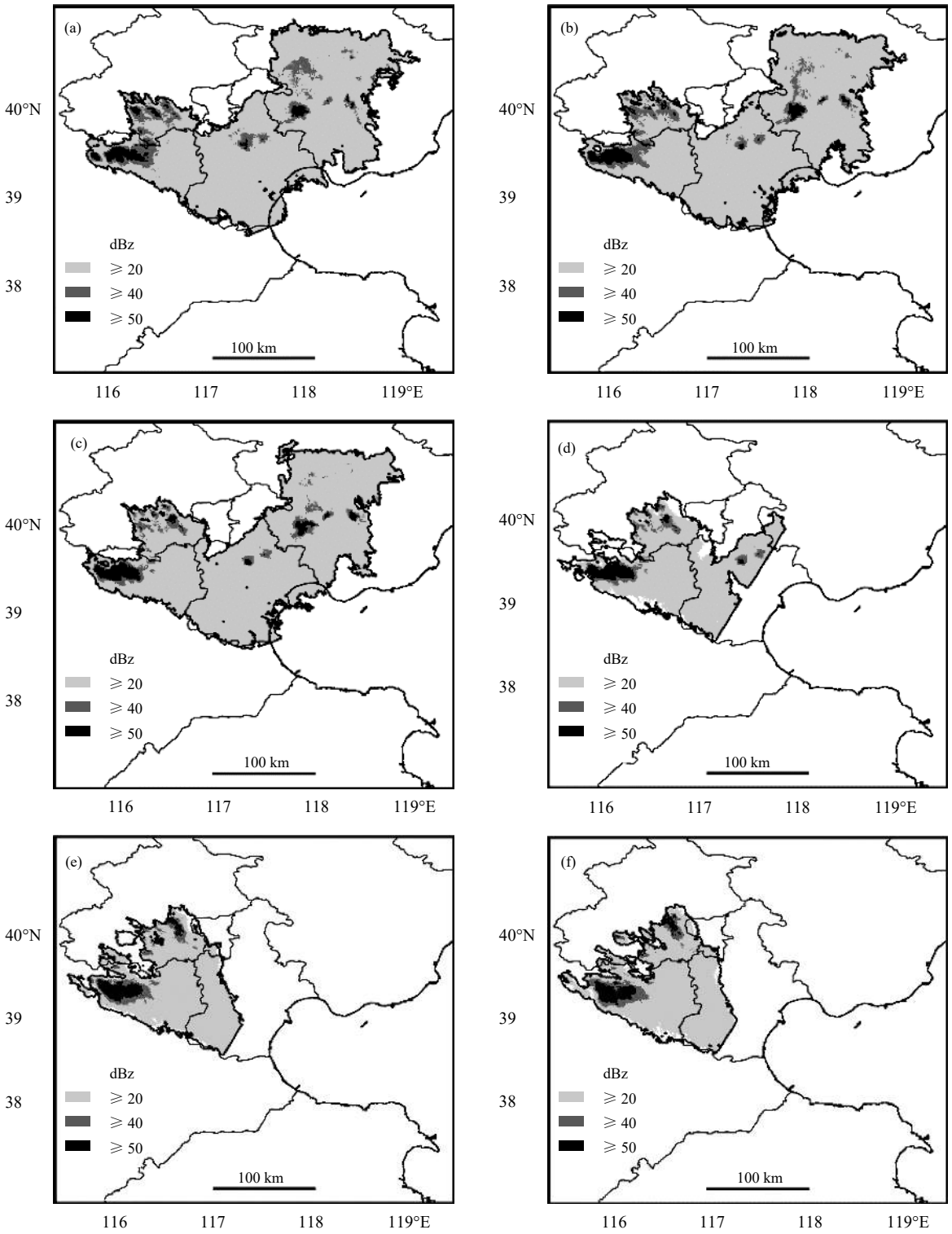


图 8　2019 年 5 月 17 日 13 时 56 分—14 时 30 分的 MCS 切片

（a—f, 间隔：6 min）

Fig. 8　Display of MCS slices during 13: 56—14: 30 UTC 17 May 2019

（a—f, interval: 6 min）

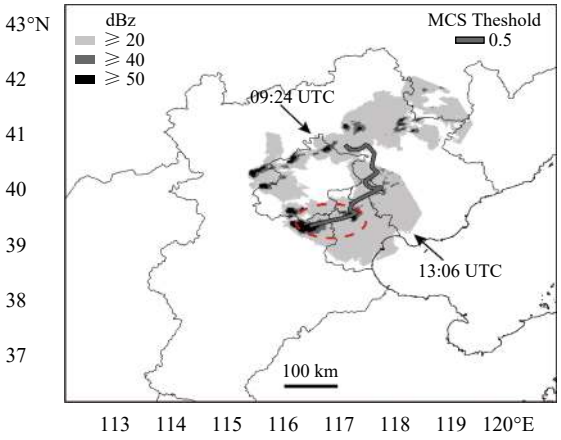


图 9 2019 年 5 月 17 日 09 时 18 分—15 时 MCS 追踪路径（已匹配）

Fig. 9 Tracking path of MCS during 09: 18—15: 00 UTC 17 May 2019 (rematched)

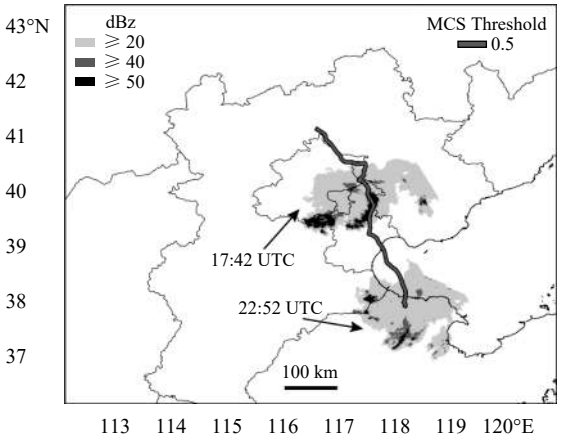


图 10 2019 年 7 月 13 日 13 时 42 分—22 时 54 分 MCS 追踪路径

Fig. 10 Tracking path of MCS during 13: 42—22: 54 UTC 13 July 2019

19 时 11 分的雷达数据（图 12）和其所对应的 MCS 切片（图 13）进行分析发现，雷达拼图分割时通常会得到一个候选 MCS 切片，但对于雷达回波结构和形态较为复杂的区域性对流天气过程，可能会出现 2 个（图 13b—e，分割得到 2 个候选 MCS 切片）、有时甚至更多个候选切片。当子图中出现多个 MCS 切片时，表示在该区域的同一时段出现了多个 MCS，进行追踪时会得到 2 条不同的轨迹路径。本试验的追踪结果只有 1 条，是因为发生在山东省北部的 MCS 切片虽然满足 MCS 的客观定义，但 DNN 模型将其识别为 non-MCS，与雷达观测实际分析完全一致，图 13b—e 右下角的 MCS 切片回

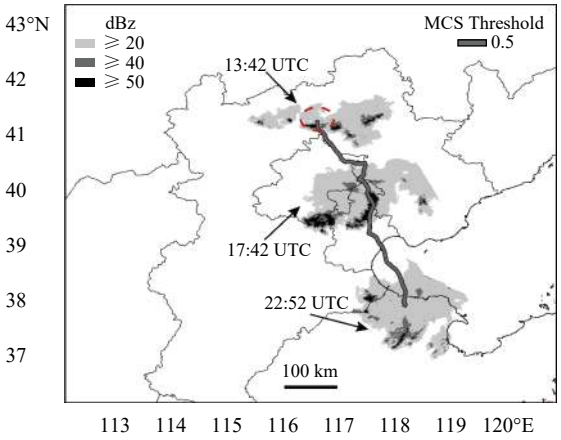


图 11 2019 年 7 月 13 日 13 时 42 分—22 时 54 分 MCS 追踪路径（已匹配）

Fig. 11 Tracking path of MCS during 13: 42—22: 54 UTC 13 July 2019 (rematched)

波特征只持续了 24 min 左右，无法形成真正的 MCS。

4.4 准线性 MCS 分类结果分析

根据 3.5 节的分类算法，对 2018 和 2019 年 5—9 月测试集数据的准线性 MCS 进行分类，可分为 TS、LS 和 PS 三类（表 9）。统计结果显示，京津冀地区 TS 型在这 3 类准线性 MCS 中占据主体（71% 左右）。Parker 等（2000）的研究也表明，美国中纬度地区的准线性 MCS 以 TS 型为主。

为了分析试验结果，此处选择了 3 个时段的 R_S、R_I 和 θ 的计算值，分别与 LS、TS 和 PS 这 3 类准线性 MCS 对应，如表 10 所示。

（1）LS 型：2019 年 5 月 17 日 12 时 41 分—13 时 05 分的 MCS 切片属于 LS 型。根据表 5 的分类算法，LS 的类别由 R_I 和 θ 决定。表 10 显示该 MCS 个例的 R_I 值均小于 0.1，且夹角 θ 值为正，与表 5 定义一致；结合 MCS 切片（图 14，2019 年 5 月 17 日 12 时 41、47、53 分和 13 时 05 分 4 个时刻的 MCS 切片），4 个 MCS 切片整体向南移动，根据其对流和强对流区域的分布，判定为 LS 型。

（2）TS 型：2019 年 7 月 13 日 14 时 17 分—15 时 59 分的 MCS 切片属于 TS 型。表 10 显示该 MCS 个例的 R_I 值均大于 10，且夹角 θ 值为正，与表 5 对 TS 型的定义一致；结合 MCS 切片（图 15，2019 年 7 月 13 日 14 时 17、47 分、15 时 17 和 47 分 4 个时刻 MCS 切片），4 个 MCS 切片整体向南移动，根据其对流和强对流区域的分布，判定为 TS 型。

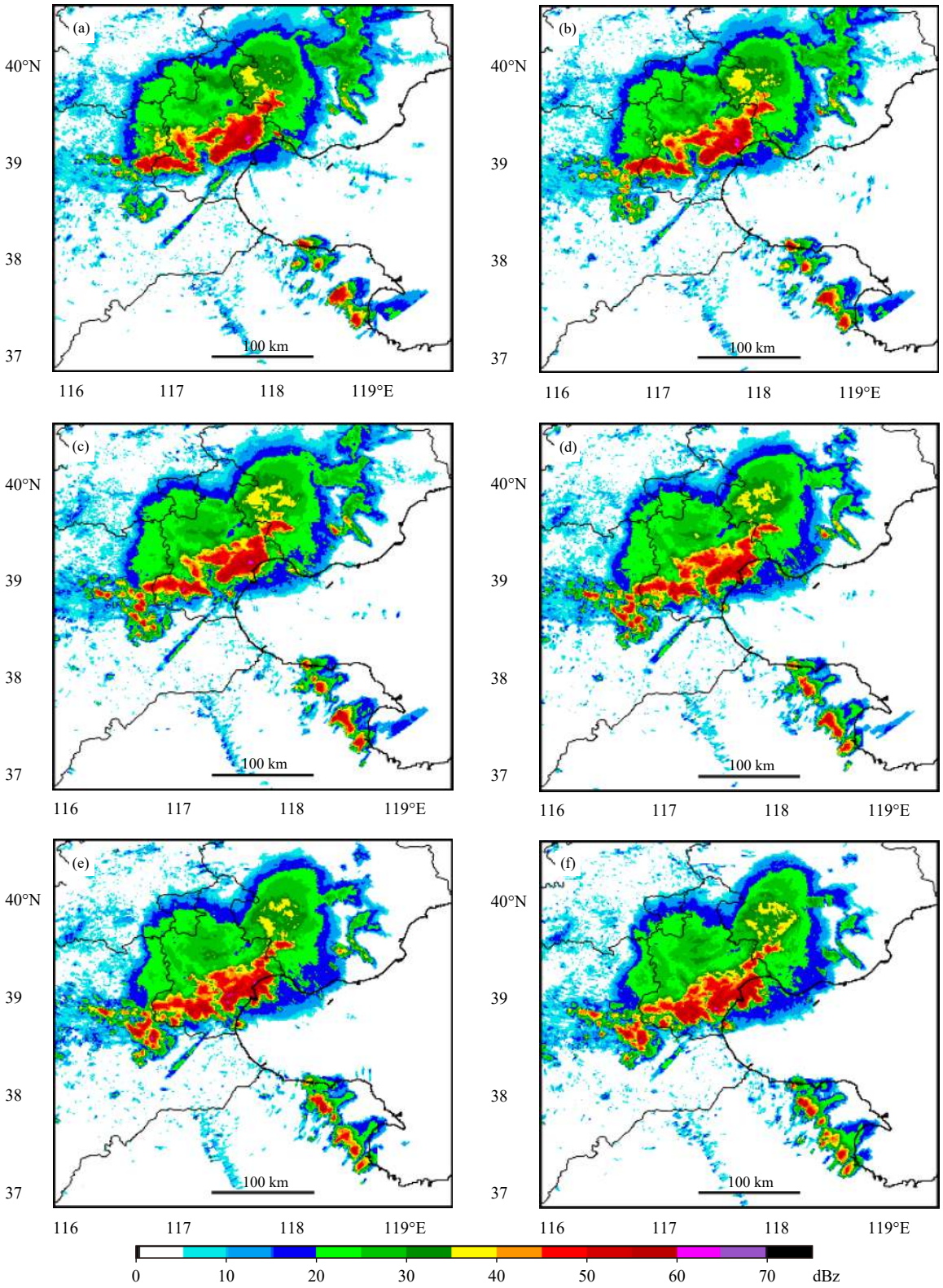


图 12　2019 年 7 月 13 日 18 时 41 分—19 时 11 分（a—f, 间隔: 6 min）的原始雷达拼图数据

Fig. 12　Original radar mosaic data during 18: 41—19: 11 UTC 13 July 2019 (a–f, interval: 6 min)

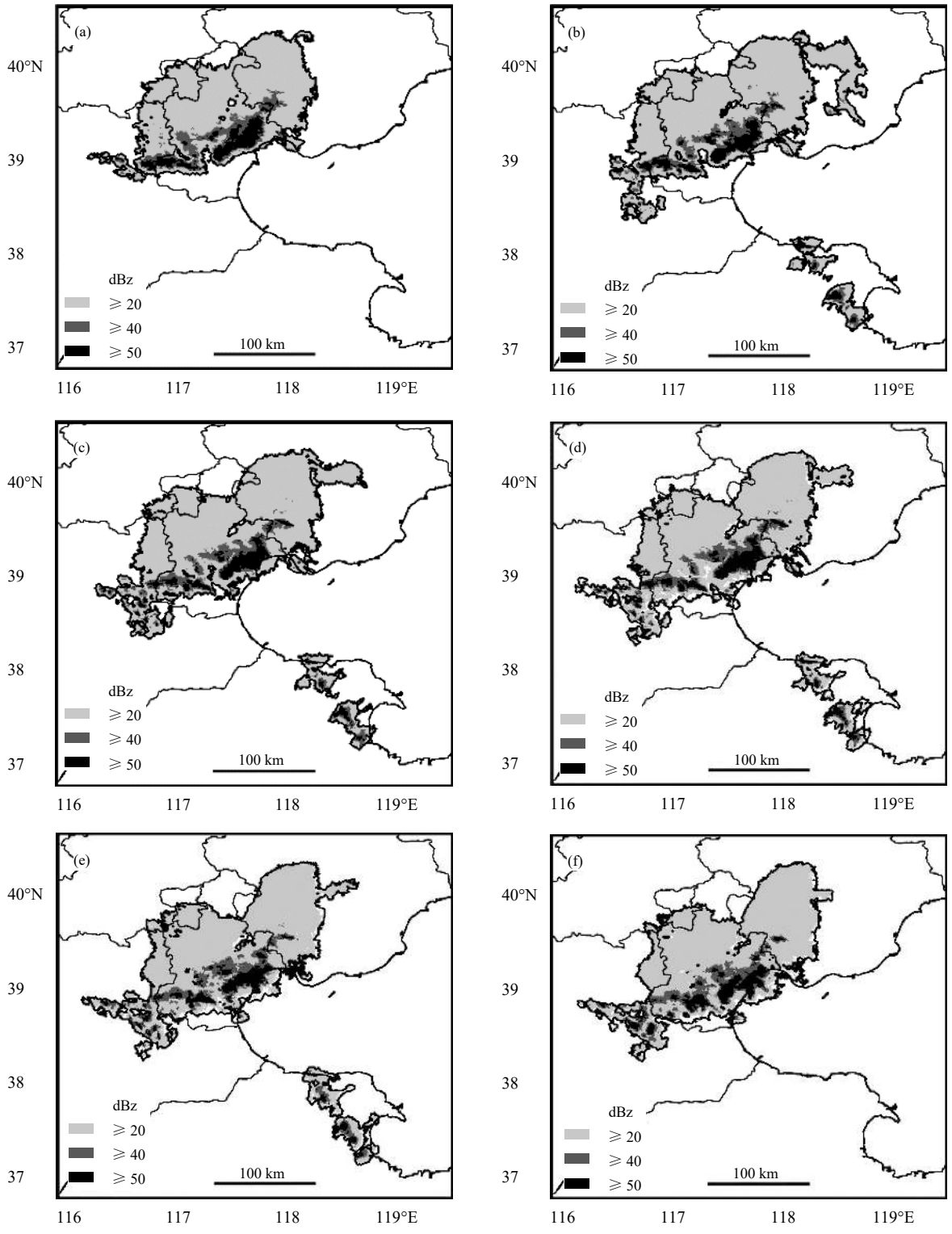


图 13　2019 年 7 月 13 日 18 时 41 分—19 时 11 分（a—f, 间隔: 6 min）的 MCS 切片展示
（b—e 子图中有 2 个 MCS 切片）

Fig. 13　Display of MCS slices during 18: 41—19: 11 UTC 13 July 2019 (a—f, interval: 6 min)
(there are two MCS slices in the b—e panels)

表 9 2018 和 2019 年 MCS 切片中 TS、LS 和 PS 型的个数统计

Table 9 Numbers of TS, LS and PS in MCS slices in 2018 and 2019

年份	2018	2019	总计
PS	41	57	98
LS	26	63	89
TS	112	356	468

（3）PS 型：2019 年 7 月 25 日 05 时 47 分—07 时 05 分的 MCS 切片属于 PS 型。根据表 5 的分类算法，LS 型由 R_S 和 R_I 决定。表 10 中该 MCS 个例的 R_S 值均接近 1，且 R_I 值在[0.1, 10]；结合 MCS 切片（图 16，2019 年 7 月 25 日 05 时 47 分、06 时 11、41 分和 07 时 05 分 4 个时刻的 MCS 切片），发现与对流线相关的大部分层状云降水区域平行于该对流线，符合 PS 型特征。

综合以上分析发现，表 5 提出的 TS、LS 和 PS 分类算法取得了良好结果，证明该分类算法的合理性与可行性，为准线性 MCS 的自动客观分类提供了一种新的方法，可在强对流天气特别是强降水时、空特征的预报中得到应用。

表 10 分类出的 LS、TS 和 PS 型准线性 MCS 所对应的 R_S、R_I 和 θ 的计算值

（比值的分母为 0 时用−9999.000 表示计算值；此处只选择了 3 个时间段）

Table 10 Calculated values of R_S, R_I and θ, which correspond to the classified LS, TS and PS of Quasi-linear MCSs

（−9999.000 is used to represent their values when the denominator of R_S and R_I is 0, only three time periods are selected here）

日期 时间(UTC)	R_S	R_I	θ	MCS类型	日期 时间(UTC)	R_S	R_I	θ	MCS类型
20190517 12: 41	1.161	0.002	66.626	LS	20190713 15: 23	0.711	880.000	32.700	TS
20190517 12: 47	1.217	0.016	18.913	LS	20190713 15: 47	0.661	53.773	3.644	TS
20190517 12: 53	1.223	0.002	13.530	LS	20190713 15: 59	0.703	−9999.000	7.494	TS
20190517 12: 59	1.195	0.017	27.797	LS	20190725 05: 47	0.979	1.614	64.397	PS
20190517 13: 05	1.195	0.023	52.510	LS	20190725 05: 59	0.963	1.749	−73.618	PS
20190713 14: 17	0.669	81.250	84.425	TS	20190725 06: 11	0.853	2.461	39.279	PS
20190713 14: 29	0.625	322.500	50.068	TS	20190725 06: 29	0.860	5.632	33.337	PS
20190713 14: 47	0.659	344.500	15.620	TS	20190725 06: 41	0.834	3.686	12.589	PS
20190713 14: 59	0.701	43.824	3.385	TS	20190725 07: 05	0.876	1.827	−49.447	PS
20190713 15: 11	0.679	−9999.000	9.676	TS					

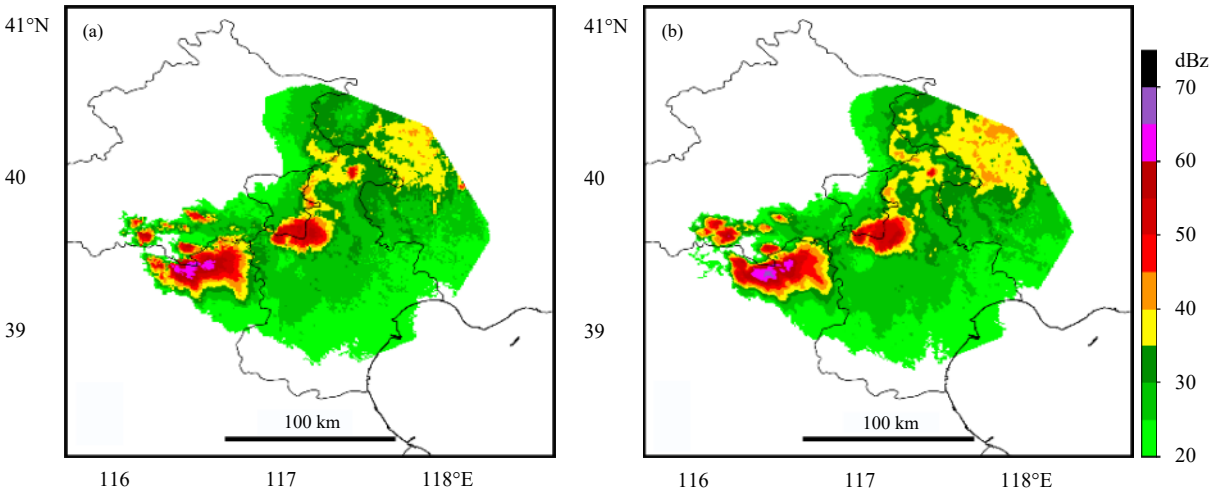
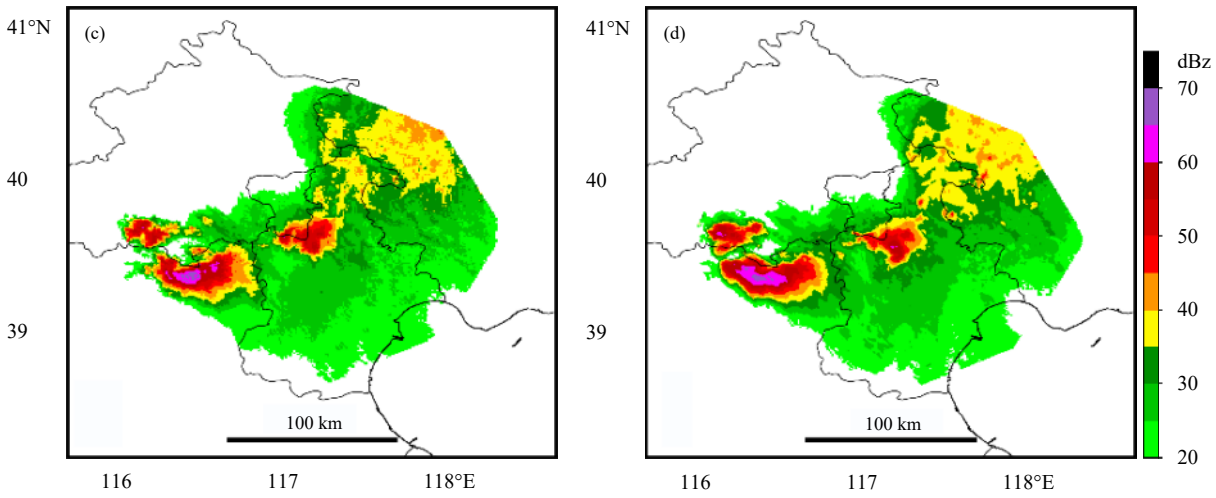


图 14 2019 年 5 月 17 日的 LS 型 MCS 雷达回波

（a. 12 时 41 分，b. 12 时 47 分，c. 12 时 53 分，d. 13 时 05 分）

Fig. 14 Classified LS MCS radar reflectivity on 17 May 2019

（a. 12: 41 UTC, b. 12: 47 UTC, c. 12: 53 UTC, d. 13: 05 UTC）



续图 14

Fig. 14 Continued

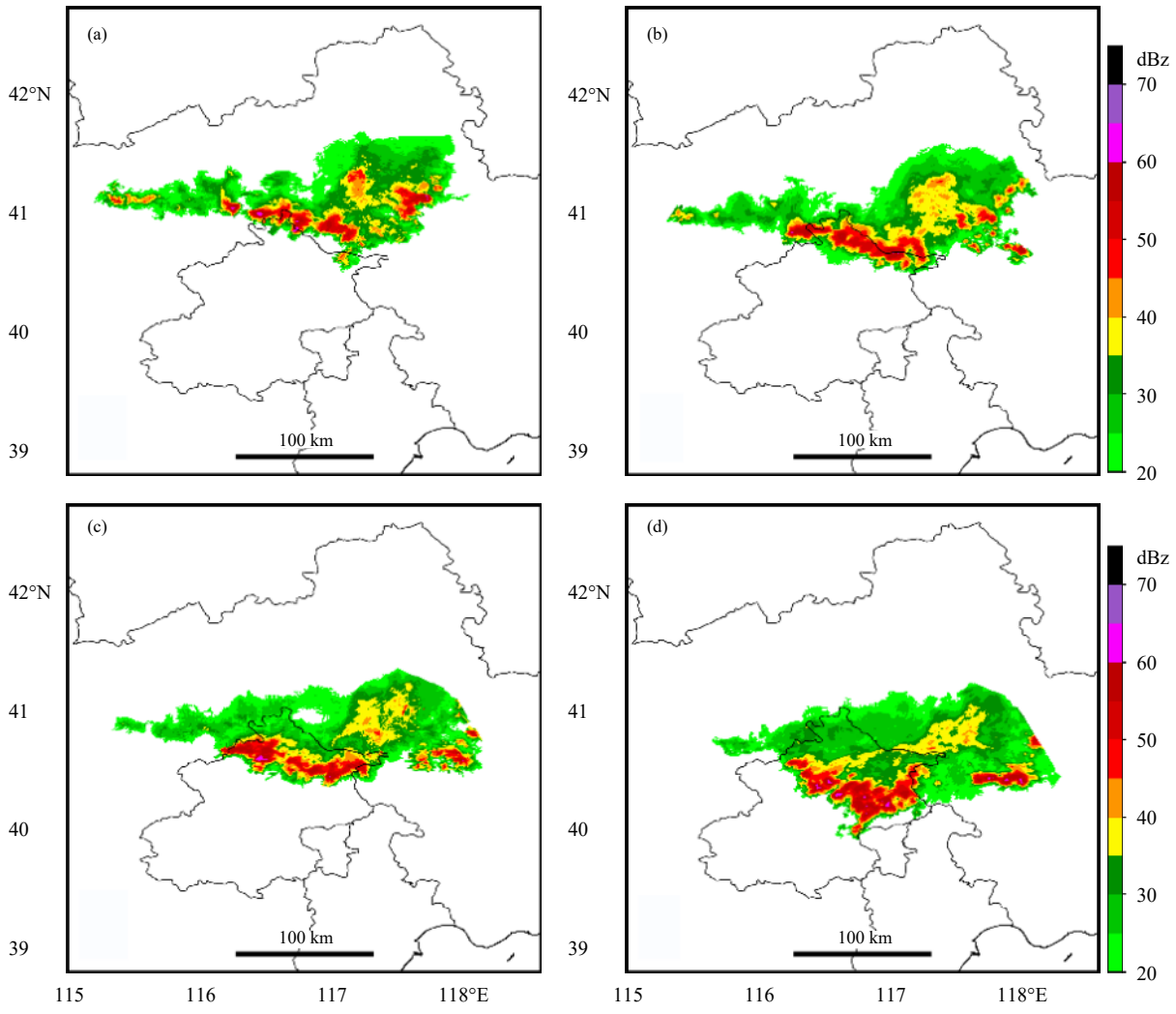


图 15 2019 年 7 月 13 日 TS 型 MCS 雷达回波(a. 14 时 17 分, b. 14 时 47 分, c. 15 时 17 分, d. 15 时 47 分)

Fig. 15 Classified TS MCS radar reflectivity on 13 July 2019 (a. 14: 17 UTC, b. 14: 47 UTC, c. 15: 17 UTC, d. 15: 47 UTC)

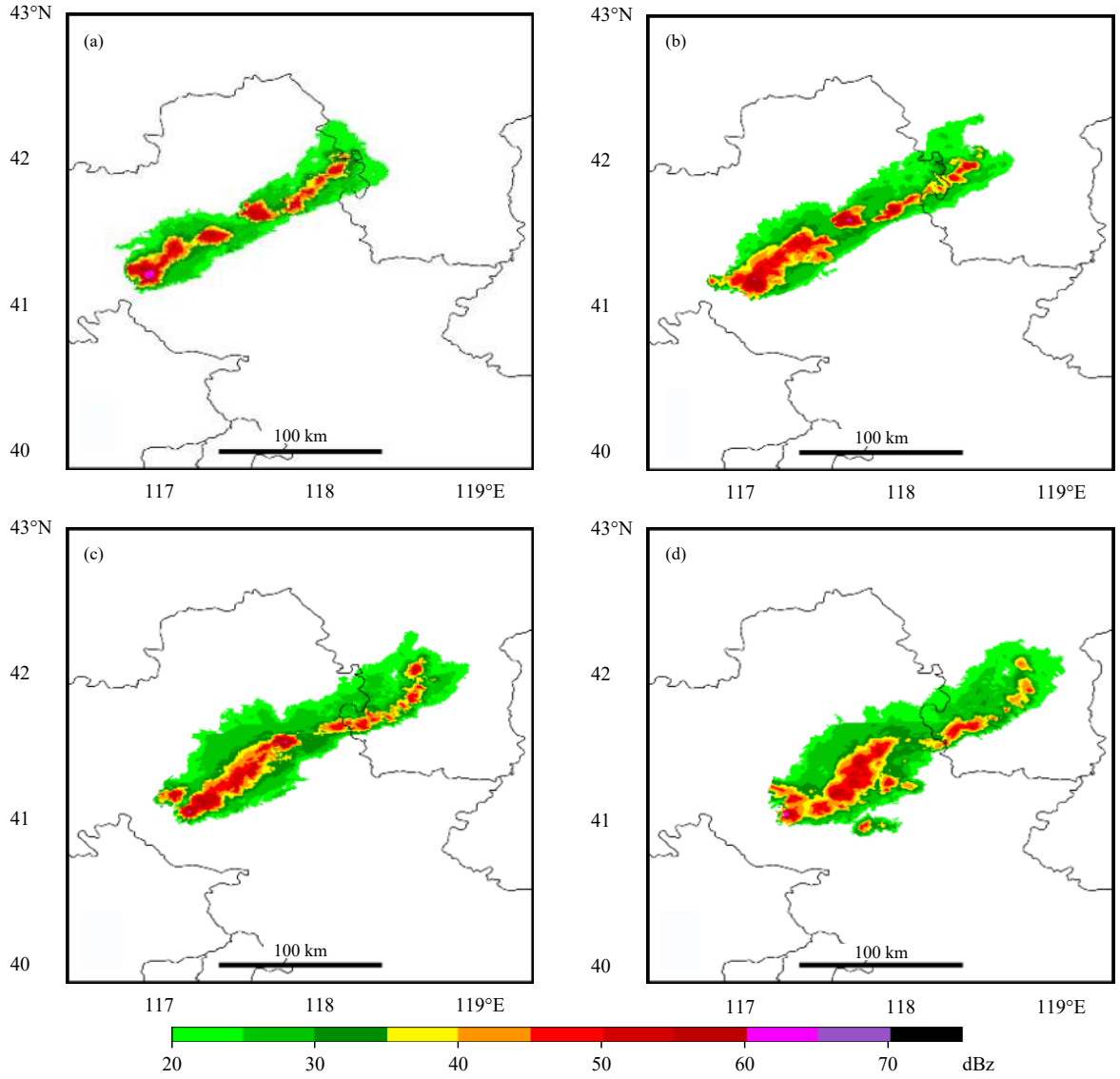


图 16　2019 年 7 月 25 日 PS 型 MCS 雷达回波（a. 05 时 47 分，b. 06 时 11 分，c. 06 时 41 分，d .07 时 05 分）

Fig. 16　Classified PS MCS radar reflectivity on 25 July 2019（a. 05: 47 UTC, b. 06: 11 UTC, c. 06: 41 UTC, d. 07: 05 UTC）

5　结论与讨论

选取 2010—2019 年共 10 a 夏季的京津冀地区雷达拼图数据，基于机器学习开展了 MCS 的自动识别、追踪及分类试验研究。（1）对原始雷达拼图数据进行预处理以保证试验数据的有效性，根据 PJ00 标准按照特定的分割参数对原始雷达数据进行分割得到候选 MCS 切片，并从每个切片中抽取 14 个 MCS 特征值构建 MCS 特征识别数据集。（2）使用深度学习方法建立了一个二分类 DNN 模型，将预测结果与其他 3 种传统机器学习算法（RF、SVM 和 XGBoost）的结果进行对比。试验结果表明，DNN 模型识别 MCS 的性能优于其他 3 种算法，能够有效判别 MCS 和 non-MCS。并且，DNN 模型将 non-MCS 识别为 MCS 的频率是最低的，有利于后续的 MCS 追踪。（3）将 DNN 模型识别的 MCS 切片用于 MCS 追踪，使用改进的时空重叠追踪法完成 2018 和 2019 年京津冀地区的 MCS 追踪，得到包含强度、空间和时间信息的 MCS 条带数据集。（4）根据追踪得到的 MCS 轨迹矢量计算 MCS 切片的运动方向，并求得 MCS 切片拟合椭圆长轴两侧的层状云和强对流云区域的面积占比，实现了 TS、LS 和 PS 三类准线性 MCS 的自动分类，对提升 MCS

致灾天气的预报、预警具有重要意义。

MCS 回波结构复杂,对其进行有效识别在气象领域是一件较为复杂的工作。文中使用深度学习算法建立了自动识别 MCS 的方法,对 MCS 的研究具有重要意义。本研究还存在一些不足,如用搜索半径 96 km 来限定 MCS 切片的层状云区域,在以后工作中还需要继续改进;对 MCS 分块进行人工特征抽取,没有发挥卷积神经网络(Convolutional Neural Networks, CNN)自动抽取图像特征的优势;并且,对准线性 MCS 的分类也是基于人工抽取特征再进行映射而实现。因此,在未来的研究中,可以从以下两方面做深入探索:(1)CNN 可以自动从输入数据中抽取到复杂的内在纹理特征,能够更加精确地捕捉到 MCS 分块中各个强度区域之间的空间联系,进行更高效地识别 MCS。可以考虑使用 CNN 模型实现 MCS 切片的自动识别,但首先得解决 CNN 网络如何训练大小不同的 MCS 切片数据,或者解决如何将 MCS 切片数据的大小进行统一处理。(2)利用深度学习实现准线性 MCS 或者准线性对流系统(QLCS)中的 TS、LS 和 PS 型的特征分类(Parker, et al, 2000)或实现 MCS 中强降水特征的分类识别(Schumacher, et al, 2005, 2020)。

致　谢:文中使用的机器学习算法源自 Scikit-Learn 开源库(代码地址:https://github.com/scikit-learn/scikit-learn.git)以及 Google 公司的 TensorFlow 平台(https://github.com/tensorflow/tensorflow.git),谨此致谢。

参考文献

曹伟华, 陈明轩, 高峰等. 2019. 雷暴区域追踪矢量与雷暴单体追踪矢量融合临近预报研究. 气象学报, 77(6): 1015-1027. Cao W H, Chen M X, Gao F, et al. 2019. A vector blending study based on object-based tracking vectors and cross correlation tracking vectors. Acta Meteor Sinica, 77(6): 1015-1027 (in Chinese)

陈明轩, 俞小鼎, 谭晓光等. 2006. 北京 2004 年"7.10"突发性对流强降水的雷达回波特征分析. 应用气象学报, 17(3): 333-345. Chen M X, Yu X D, Tan X G, et al. 2006. Radar echoes characteristics of the sudden convective rainstorm over Beijing area on July 10, 2004. J Appl Meteor Sci, 17(3): 333-345 (in Chinese)

陈明轩, 高峰, 孔荣等. 2010. 自动临近预报系统及其在北京奥运期间的应用. 应用气象学报, 21(4): 395-404. Chen M X, Gao F, Kong R, et al. 2010. Introduction of auto-nowcasting system for convective storm and its performance in Beijing olympics meteorological service. J Appl Meteor Sci, 21(4): 395-404 (in Chinese)

韩雷, 郑永光, 王洪庆等. 2007. 基于数学形态学的三维风暴体自动识别方法研究. 气象学报, 65(5): 805-814. Han L, Zheng Y G, Wang H Q, et al. 2007. 3D storm automatic identification based on mathematical morphology. Acta Meteor Sinica, 65(5): 805-814 (in Chinese)

雷蕾, 邢楠, 周璇等. 2020. 2018 年北京"7.16"暖区特大暴雨特征及形成机制研究. 气象学报, 78(1): 1-17. Lei L, Xing N, Zhou X, et al. 2020. A study on the warm-sector torrential rainfall during 15—16 July 2018 in Beijing area. Acta Meteor Sinica, 78(1): 1-17 (in Chinese)

王晓芳, 崔春光. 2011. 暴雨中尺度对流系统研究的若干进展. 暴雨灾害, 30(2): 97-106. Wang X F, Cui C G. 2011. A number of advances of the research on heavy rain mesoscale convective systems. Torrential Rain Disaster, 30(2): 97-106 (in Chinese)

杨吉, 郑媛媛, 夏文梅等. 2015. 雷达拼图资料上中尺度对流系统的跟踪与预报. 气象, 41(6): 738-744. Yang J, Zheng Y Y, Xia W M, et al. 2015. Mesoscale convective systems (MCSs) tracking and nowcasting based on radar mosaic data. Meteor Mon, 41(6): 738-744 (in Chinese)

Ashley W S, Haberlie A M, Strohm J. 2019. A climatology of quasi-linear convective systems and their hazards in the United States. Wea Forecast, 34(6): 1605-1631

Bengio Y. 2009. Learning deep architectures for AI. Found Trends Mach Learn, 2(1): 1-127

Davis C, Brown B, Bullock R. 2006a. Object-based verification of precipitation forecasts. Part Ⅰ: Methodology and application to mesoscale rain areas. Mon Wea Rev, 134(7): 1772-1784

Davis C, Brown B, Bullock R. 2006b. Object-based verification of precipitation forecasts. Part Ⅱ: Methodology and application to convective rain systems. Mon Wea Rev, 134(7): 1785-1795

Dixon M, Wiener G. 1993. TITAN: Thunderstorm identification, tracking, analysis, and nowcasting: A radar-based methodology. J Atmos Ocean Technol, 10(6): 785-797

Haberlie A M, Ashley W S. 2018a. A method for identifying midlatitude mesoscale convective systems in radar mosaics. Part Ⅰ: Segmentation and classification. J Appl Meteor Climatol, 57(7): 1575-1598

Haberlie A M, Ashley W S. 2018b. A method for identifying midlatitude mesoscale convective systems in radar mosaics. Part Ⅱ: Tracking. J Appl Meteor Climatol, 57(7): 1599-1621

Han L, Fu S X, Zhao L F, et al. 2009. 3D convective storm identification, tracking, and forecasting: An enhanced TITAN algorithm. J Atmos Ocean Technol, 26(4): 719-732

Houze R A Jr. 2018. 100 years of research on mesoscale convective systems. Meteor Monogr, 59(1): 17.1-17.54

Jergensen G E, McGovern A, Lagerquist R, et al. 2020. Classifying convective storms using machine learning. Wea Forecast, 35(2): 537-559

Johnson J T, MacKeen P L, Witt A, et al. 1998. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. Wea Forecast, 13(2): 263-276

Lakshmanan V, Hondl K, Rabin R. 2009. An efficient, general-purpose

technique for identifying storm cells in geospatial images. J Atmos Ocean Technol, 26(3): 523-537

Mueller C, Saxen T, Roberts R, et al. 2003. NCAR auto-nowcast system. Wea Forecast, 18(4): 545-561

Parker M D, Johnson R H. 2000. Organizational modes of midlatitude mesoscale convective systems. Mon Wea Rev, 128(10): 3413-3436

Pedregosa F, Varoquaux G, Gramfort A, et al. 2011. Scikit-learn: Machine learning in python. J Mach Learn Res, 12: 2825-2830

Rinehart R E, Garvey E T. 1978. Three-dimensional storm motion detection by conventional weather radar. Nature, 273(5660): 287-289

Schumacher R S, Johnson R H. 2005. Organization and environmental properties of extreme-rain-producing mesoscale convective systems. Mon Wea Rev, 133(4): 961-976

Schumacher R S, Johnson R H. 2006. Characteristics of U. S. extreme rain events during 1999–2003. Wea Forecast, 21(1): 69-85

Schumacher R S, Rasmussen K L. 2020. The formation, character and changing nature of mesoscale convective systems. Nat Rev Earth Environ, 1(6): 300-314

Skok G, Tribbia J, Rakovec J, et al. 2009. Object-based analysis of satellite-derived precipitation systems over the low-and midlatitude Pacific Ocean. Mon Wea Rev, 137(10): 3196-3218

van der Walt S, Schönberger J L, Nunez-Iglesias J, et al. 2014. Scikit-image: Image processing in python. Peer J, 2: e453

Wang X F, Cui C G, Cui W J, et al. 2014. Modes of mesoscale convective system organization during Meiyu season over the Yangtze River basin. Acta Meteor Res, 28(1): 111-126

Zheng A. 2015. Evaluating Machine Learning Models: A Beginner's Guide to Key Concepts and Pitfalls. O'Reilly Media, Inc: 8-9