

引用格式: Wang Sheng, Yang Xiaofeng, Ma Wentao, *et al.* Tropical Cyclone Genesis Prediction based on Space-borne Microwave Brightness Temperature Observations in the Northwest Pacific Ocean[J]. Remote Sensing Technology and Application, 2021, 36(1): 198-207. [汪胜, 杨晓峰, 马文韬, 等. 基于星载微波亮温数据的西北太平洋热带气旋生成预测方法[J]. 遥感技术与应用, 2021, 36(1): 198-207.]

doi:10.11873/j.issn.1004-0323.2021.1.0198

基于星载微波亮温数据的西北太平洋热带气旋生成预测方法

汪 胜^{1,2}, 杨晓峰^{1,3}, 马文韬^{1,3}, 相坤生^{1,4}, 胡 蝶²

(1. 中国科学院空天信息创新研究院, 遥感科学国家重点实验室, 北京 100101;

2. 中国科学院大学, 北京 100049;

3. 三亚中科遥感研究所, 海南省地球观测重点实验室, 海南 三亚 572029;

4. 北京航天宏图信息技术股份有限公司, 北京 100195)

摘要:从热带云团中识别有可能生成热带气旋的云团,对台风灾害的早期预警至关重要。基于星载微波辐射计亮温数据和决策树方法,开展了西北太平洋热带气旋生成过程识别研究,构建了热带气旋生成预测模型。样本数据来自2005~2009年间每年5~10月台风季获取的SSM/I微波辐射计观测数据,共包含326个热带气旋样本和2 112个未生成气旋的热带扰动样本。基于该数据集计算得到了35个亮温特征参数,包括25个统计量参数和10个描述像元空间分布的景观格局参数。然后,将数据集分为训练集(2007~2009年)和测试集(2005~2006年),在训练集上模型的总体精度为85.27%,热带气旋正确预报率为90.91%,错误预报率为20%;上述3项指标在测试集上的结果中分别为79.47%、80%和20.79%;表明利用微波亮温数据构建的热带气旋生成模型能有效预测热带气旋的发展。

关键词:热带气旋生成;SSM/I亮温;决策树方法;西北太平洋

中图分类号:P444;TP79 **文献标志码:**A **文章编号:**1004-0323(2021)01-0198-10

1 引言

热带气旋(Tropical Cyclone, TC)是发生在热带洋面上的强烈气旋性涡旋,又称台风(西北太平洋)、飓风(大西洋和东北太平洋)等。作为地球上最具破坏性的自然现象之一,由其引起的风暴潮不仅使沿海地区面临严重威胁,而且对海上交通运输、军事活动、海洋工程和渔业都带来严重影响^[1]。相比于其他大洋,西北太平洋的热带扰动发生得最为活跃,平均每年会生成数以百计的热带扰动,但能发展成热带气旋的仅有20个左右,准确的热带气

旋全生命周期观测和预测对防灾减灾至关重要^[2-4]。

针对热带气旋的预报问题,一种方法是使用数值天气预报模型(Numerical Weather Prediction Models, NWP)。例如, Halperin等^[5]对比了5个NWP模型针对2004年至2011年间在北大西洋上发生的热带气旋的预报能力,结果表明其条件正确预报率为26%~44%。然而,大多数的NWP模型难以判断一个热带扰动最终是发展成热带气旋还是不会继续发展并逐渐消亡^[6]。另一种方法是使用统计模型,这类方法是从大量的、不完全的、有噪声的随机数据中提取出与热带扰动和热带气旋发展

收稿日期:2019-12-01;修订日期:2020-12-23

基金项目:国家自然科学基金面上项目(41871268),国家重点研发计划项目(2017YFB0502800)。

作者简介:汪 胜(1994—),男,湖北武汉人,硕士研究生,主要从事海洋遥感方面的研究。E-mail:wangsheng@radi.ac.cn

通讯作者:杨晓峰(1982—),男,四川成都人,研究员,主要从事海洋微波遥感研究。E-mail: yangxf@radi.ac.cn

过程密切相关的有用信息,然后利用数据挖掘算法分析所获取的信息以构建热带气旋识别的统计模型。数据挖掘算法主要包括关联规则、分类、聚类等分析方法。热带气旋的预测本质上是将热带扰动分为发展(Developing, DEV)和非发展(Non-Developing, NON_DEV)的分类问题^[7]。Schumacher^[8]等综合利用美国国家环境预测中心(National Center for Environmental Prediction, NCEP)、美国国家大气研究中心(National Center for Atmospheric Research, NCAR)再分析获得的大尺度环境数据^[9]和多个地球同步卫星的观测数据,采用线性判别分析(Linear Discriminant Analysis, LDA)的方法估测了热带气旋在各种大洋海盆的 $5^{\circ}\times 5^{\circ}$ 子区域内的热带气旋生成情况。类似的,Hennon等^[10]基于NCEP-NCAR再分析数据获取的8个大尺度预报因子也使用LDA方法进行热带气旋的生成预测。此外,Fan等^[11]基于NCEP-NCAR再分析数据提取的大尺度环境因子进行多元线性回归分析,进而对大西洋的热带气旋生成进行了预测。

上述基于线性模型的方法在使用时需要假设预测因子和预测结果之间的线性关系,因而存在较大的局限性^[12-13]。由于机器学习的方法无需线性假设的前提^[14],使得机器学习方法在热带气旋的探测^[15-17]和其他遥感研究^[18-22]中得到了广泛应用。通常机器学习算法包括决策树(Decision Trees, DT)、随机森林(Random Forest, RF)、支持向量机(Support Vector Machine, SVM)以及近年来兴起的深度学习(Deep Learning, DL)等。其中,决策树是根据类似树的决策规则递归地将数据划分为子集,而随机森林则可以视为分类树的集成^[23]。支持向量机则是通过超平面对数据集进行划分,达到分类的目的^[24]。相比于随机森林和支持向量机算法,决策树分类算法具有简单和分类规则可见的优势,因此在热带气旋发生的预测中运用最为广泛。

针对热带气旋生成预测问题,学者们基于机器学习方法提出了一系列预测模型。Zhang等^[15]从粗分辨率的全球预报系统中计算得到一系列大尺度的环境指数,并基于决策树方法对上述预测因子构建了分类规则,使得热带气旋的总体预测精度为81.72%,而在气旋发生前24 h提前预测的精度则达到了84.6%。Park等^[16]使用WindSat卫星遥感风场数据提取了预测因子构建了决策树规则,其总体分类精度达到81.05%,正确预报率为88.35%,且

93.55%的热带气旋能在发生前13.5 h内准确预测。在Park等^[16]的研究基础上, Kim等^[25]使用相同的数据集对比评估了线性判别分析、决策树、随机森林以及支持向量机4种方法在预测热带气旋发生中的表现。研究结果表明,3种机器学习的方法的正确预报率(94%~96%)均高于线性判别分析方法(77%),但其错误预报率(21%~28%)也高于线性判别分析方法(13%)。此外,机器学习方法能提前26~30 h预测热带气旋的发生,而线性判别方法仅能提前5~9 h成功预测。此外,Matsuoka等^[26]基于计算模拟得到的外向长波辐射(OLR)数据和卷积神经网络(CNN)方法提出了一种用于识别热带气旋及其前兆的学习模型。实验结果表明该模型在西北太平洋正确检测热带气旋及其前兆的精度为79.9%~89.1%,错误警报率为32.8%~53.4%。Zhang等^[27]开发了一种机器学习框架来预测中尺度对流系统(MCS)是否会演变成热带气旋。结果表明,机器学习分类器AdaBoost在气旋生成前6 h可较为准确地预测。当预测的前置时间延长到12、24和48 h时,该分类器在北大西洋和西北太平洋仍具有较好的表现。Su等^[28]基于卫星遥感获取的热带气旋内部结构信息,使用机器学习方法,有效提高了对热带气旋快速增强过程的预报精度,并降低了误报率。

上述研究多是使用再分析获得环境数据或遥感反演得到的产品资料来预测热带气旋的生成,相比直接获取的卫星原始辐射亮温观测数据(Brightness Temperature, TB),这些数据的质量会受到中间处理过程的影响,且由于存在中间过程,导致处理时效性较低。因此,本研究将直接使用卫星观测的微波亮度温度数据,采用决策树方法,构建热带气旋生成预测模型。

2 数据说明

2.1 SSM/I 亮温数据

美国国防气象卫星计划(Defense Meteorological Satellite Program, DMSP)于1999年12月12日发射了DMSP F15卫星,其采用高度约为850 km,倾角98.7度的近极地轨道。卫星上搭载了特种微波成像仪SSM/I,该辐射计以45°的天底角做圆锥形扫描,因此当它扫描地球时,观察区域和倾斜路径几乎保持不变,能有效避免跨轨迹横扫描仪器中的临边效应^[29-30]。SSM/I提供7个窗口通道,用来检测和反演大气和地表参数,如降水量、海冰覆盖

率、海表面风速、柱积分水汽含量、地表温度、发射率以及土壤水分等^[31-33]。SSM/I辐射计7个频段的主要技术指标详见表1。

表 1 SSM/I 通道特性

Table 1 Channel characteristic of SSM/I

| 频段/GHz | 极化方式 | NEDT/K | 采样间隔/km |
|--------|-------|--------|---------|
| 19.35 | H / V | 0.8 | 25 |
| 22.235 | V | 0.8 | 25 |
| 37.0 | H / V | 0.6 | 25 |
| 85.5 | H / V | 1.1 | 12.5 |

为尽量与热带云团数据的覆盖周期相匹配,综合参考现有^[16,25]研究中使用数据的时间周期后,本研究使用的SSM/I微波辐射计数据的时间范围为2005~2009年期间每年5~10月,研究范围为西北太平洋区域($0^{\circ}\sim 30^{\circ}\text{N}$, $150^{\circ}\text{E}\sim 180^{\circ}$)。对于刚生成的热带气旋,由于对流发展较弱,冰粒子较少,此时85 GHz通道上亮温对其不敏感,因此本文使用5个低频段通道上获取的数据作为实验样本,这些频段分别是19 GHz(H)、19 GHz(V)、22 GHz(V)、37 GHz(H)和37 GHz(V)。

2.2 最佳路径数据

研究采用的西北太平洋的热带气旋最佳路径数据由美国台风联合预警中心(Joint Typhoon Warning Center, JTWC)提供。该数据集收集了热带气旋随着时间发展的经度、纬度、风速等信息,其时间分辨率为6 h。使用该数据集获取热带气旋的位置信息,进而提取出热带气旋发生的样本数据。为保证最佳路径数据与亮温影像更为准确地匹配,需将最佳路径数据的时间分辨率插值为1 h。

2.3 热带云团数据

研究采用到的西北太平洋的热带云团资料来自1982~2009年的全球热带云团数据集^[34](Global Tropical Cloud Cluster Dataset v01r01),该数据集通过分析红外卫星资料云顶亮温来确定和跟踪热带云团的移动和发展情况,其时间分辨率为3 h。利用该数据集提供的位置信息提取得到未发展成热带气旋的样本数据。为保证云团路径数据与亮温影像匹配更加准确,需将热带云团路径数据的时间分辨率插值为1 h。

3 方 法

研究技术流程如图1所示。首先通过星载亮温数据和描述热带气旋或扰动位置的路径数据与热

带云团数据提取 $16^{\circ}\times 16^{\circ}$ 范围内的辐射计影像,然后在各个频段上提取特征参数,进而得到本研究所需的样本数据集,然后基于决策树C4.5算法和训练集数据构建热带气旋生成预测模型,最后利用测试集数据对模型进行验证。

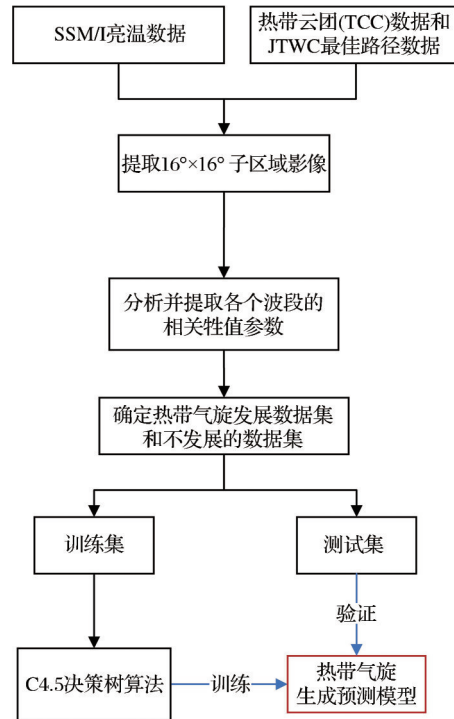


图 1 热带气旋生成预报建模流程图

Fig.1 The flowchart of tropical cyclone prediction

3.1 数据预处理

在获取热带气旋的样本时,参考Park等^[16,23]的做法,将热带气旋的最大风速第一次达到13 m/s (25 knot)的时刻定义为气旋发生时刻,为了保证有足够的样本并尽可能不引入成熟期热带气旋的干扰,仅从气旋发生后72 h以内来获得发展类样本。此外,还需对样本进行如图2所示的质量控制:以热带扰动中心为圆心, 4° 范围为半径确定为核心区域,只有该区域内的非空值像元比例大于60%时,该样本才被划分为有效数据。根据以上规则从2005~2009年的辐射计影像中共提取出326个发展成热带气旋的正样本,2 112个未发展为热带气旋的负样本,其详细信息见表2。然后,将所有数据按时间分为训练集和测试集,其中2007~2009年的数据划分为训练集,2005~2006年数据划分为测试集。由于发展类和非发展类的样本在数量上存在明显差距,因此将对训练集中非发展类样本采用随机提取的方式进行欠采样,使得两类样本数量相当。最终训练集中包含187个发展类样本,200个非发展类

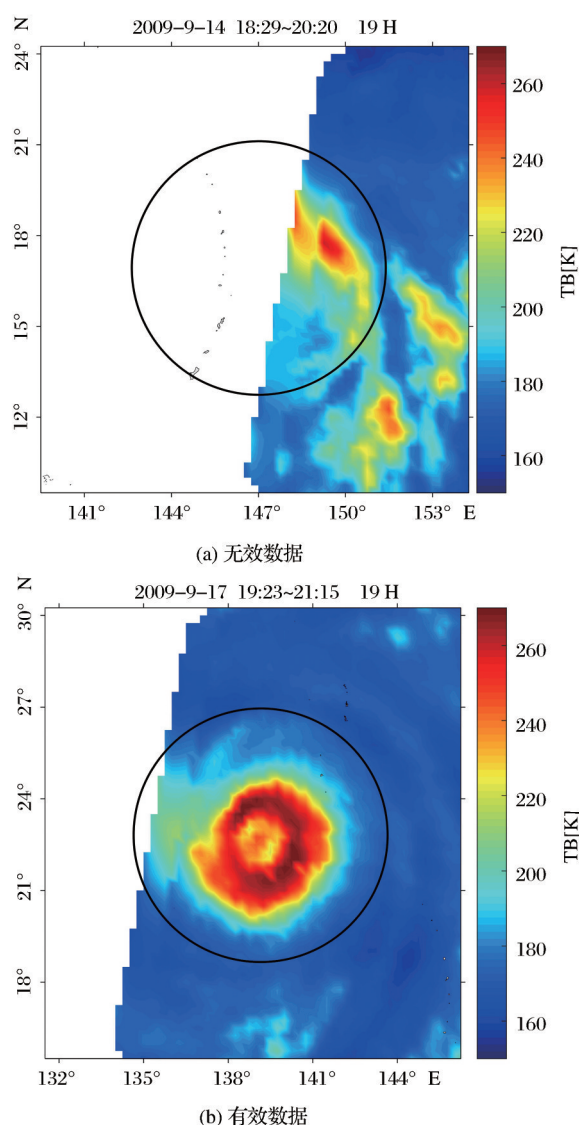


图 2 无效数据与有效数据

Fig.2 The invalid data and valid data

样本。对于测试集,两类样本数量没有条件约束,均采用随机采样的方式分别从发展类和非发展类中提取 50 个和 101 个样本作为测试集数据。

3.2 特征参数选择

研究采用的特征参数主要有两类,分别是亮温统计量和描述高亮温像元空间分布的景观格局参

表 2 训练集(2007~2009 年)和测试集(2005~2006 年)
所包含的发展类和非发展类扰动数量

Table 2 The Sample size of developing and non-developing tropical disturbances in training dataset (2007~2009) and test dataset (2005~2006)

| 数据集名称 | 发展类 | 非发展类 | 总计 |
|-------------------|-----|-------|-------|
| 训练集(2007~2009 年) | 187 | 1 233 | 1 420 |
| 测试集(2005~2006 年) | 139 | 879 | 1 018 |
| 总数据集(2005~2009 年) | 326 | 2 112 | 2 438 |

量。由于未发展成气旋的热带扰动与热带气旋在亮温统计量之间的差异性较小,难以直接将二者进行准确区分。考虑到热带气旋往往具有一定的螺旋特性,其高亮温像元会集中分布在气旋中心附近;而未发生气旋的热带扰动往往不具备这种螺旋特性,其高亮温像元分布较为分散^[35]。以 2009 年台风“彩云”和同时期的一个非气旋热带扰动为例,图 3 为二者在各个通道上的亮温图。对比图中任一通道上的两幅子图,其亮温的大小并没有明显的差别。为此,以 19 H 波段为例,统计了两组影像核心区域(中心 4° 范围)的均值、最大值、最小值以及它们之间的差值。从图 4 中可以看出,这两组样本在 19 H 波段下的各个统计值的区别很小。进一步,考虑到图 3 中非发展类的扰动的高亮温值分布较为分散,而台风“彩云”的高亮温像元主要集中在中心区域,因此引入描述像元分布特征的景观格局参量来弥补亮温统计量在区分能力上的不足。在景观格局参量的选择过程中,参考 Park 等^[16]基于风速和降雨率的集中度指数以及相似邻接比参数来分析热带气旋发生的研究,选择高亮温值像元的集中度指数(Clumpiness Index, CI)和相似邻接比(Percentage of Like Adjacencies, PLADJ)作为特征参数。集中度指数的取值范围从-1~1,数值从小到大表示越来越集中,相似邻接比的取值范围为 0~100,值越大表示相同类别之间连接处斑块更加集中。以上两个景观格局参数可使用 FRAGSTATS 软件直接计算,软件的下载地址为(<http://www.umassee.edu/landeco/research/fragstats/fragstats.html>)。表 3 给出了研究使用的特征参数的定义及其缩写。在计算 CI 和 PLADJ 之前,需要将亮温影像分为高亮温区和低亮温区,以总数据集中两类样本在各个通道的亮温均值为参考,确定了各个通道的亮温分割阈值(表 4)。将亮温大于该阈值的像元的标签值设置为 1,小于该阈值则设置为 0,利用 FRAGSTATS 软件计算得到集中度指数和相似邻接比。最终,共确定了 7 个特征参数,在每个通道上计算以上参数,共获得了 35 个特征参数。

为分析各个通道的数据样本的分布情况,以 19 GHz 水平极化通道为例进行了简要分析,图 5 为正、负两类样本数据分布的箱线图。结果表明所选择的 7 个特征参数均具有较好的区分度。从图 5 中最上层子图可以看出,发展类样本与非发展类热带扰动相比,其最大值 MAX、均值 MEAN 和最小值

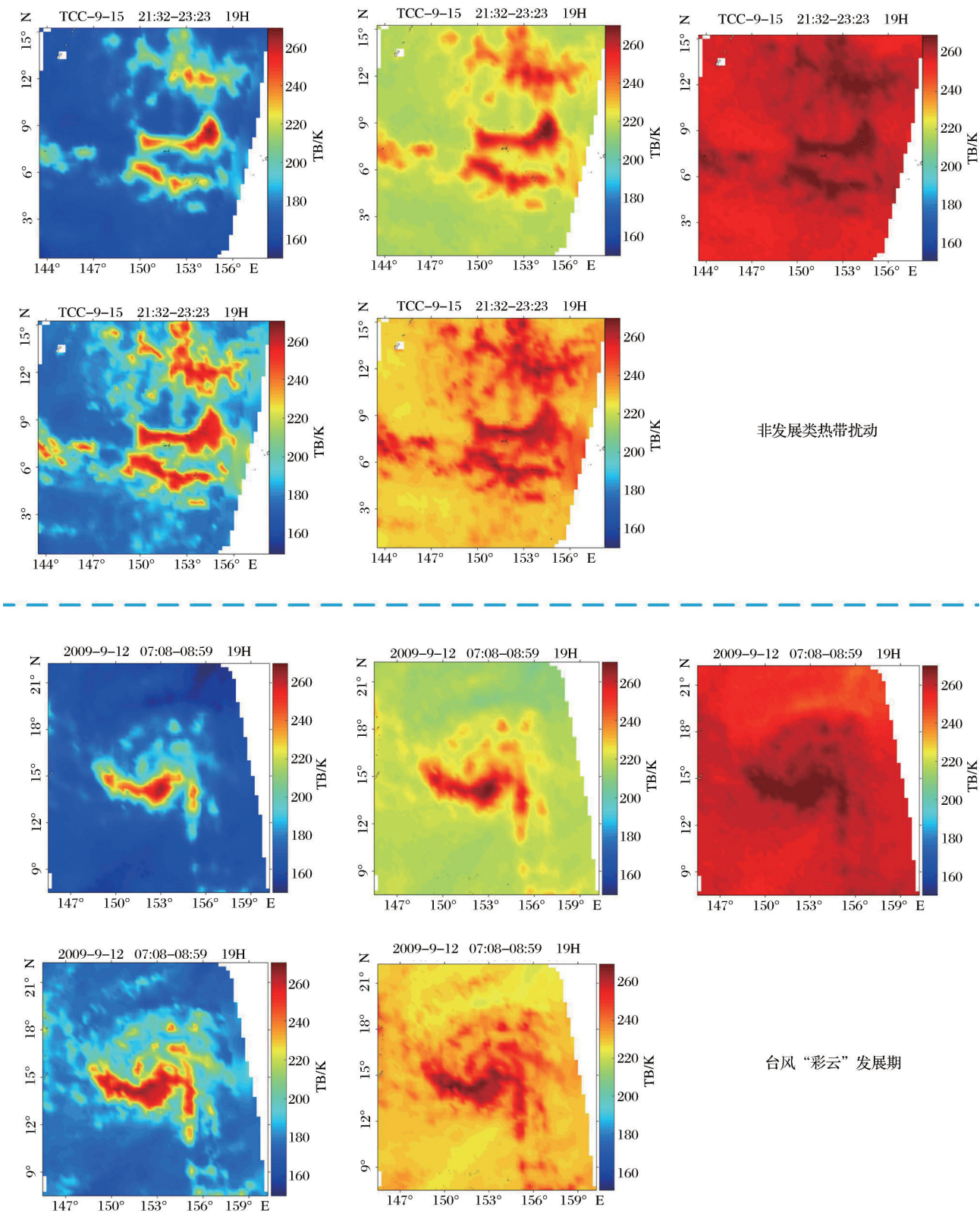


图 3 非发展类热带扰动与发展类热带气旋各个通道亮温图

Fig.3 Bright temperature maps of non-developing tropical disturbance and developing tropical cyclone for each channel

MIN均较大。对于发展类样本,其最大值分布较非发展类更为集中,均值和最小值的分布范围更广。另外,从中间子图可以看出,无论是最大值与均值的

差值还是均值与最小值的差值,发展类样本结果均高于非发展类,且分布更为集中。最后,集中度指数CI和相似邻接比PLADJ的数据分布情况也具有较

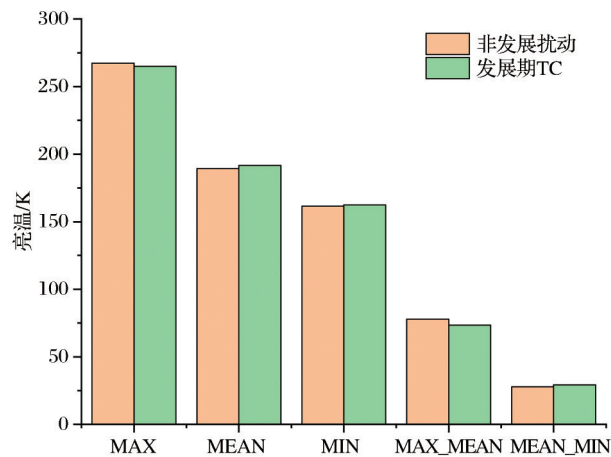


图 4 非发展类扰动与热带气旋的统计量直方图

Fig.4 The histogram of statistics of non-developing disturbance and tropical cyclone

表 3 参数标识的缩写及含义

Table 3 The abbreviations and meanings of parameters

| 参数缩写 | 含义 |
|--------------------|--|
| MAX | 核心区域 ¹ 亮温最大值 |
| MIN | 核心区域 ¹ 亮温最小值 |
| MEAN | 核心区域 ¹ 亮温平均值 |
| MAX_MEAN | 核心区域 ¹ 亮温最大值减去平均值 |
| MEAN_MIN | 核心区域 ¹ 亮温平均值减去最小值 |
| CI ² | 集中度指数 (Clumpiness Index) |
| PLADJ ³ | 相似邻接比 (Percentage of Like Adjacencies) |

注:¹以扰动中心为圆心 4°范围内的区域,即图 2 所示的黑色圆形区域;

CI²表示集中度指数,其计算公式为 $G_i = \frac{g_{ii}}{\sum_{k=1}^m g_{ik}}$;

$$CI = \begin{cases} \frac{G_i - P_i}{1 - P_i} & G_i \geq P_i \\ \frac{G_i - P_i}{1 - P_i} & G_i < P_i; P_i \geq 0.5 \\ \frac{P_i - G_i}{-P_i} & G_i < P_i; P_i < 0.5 \end{cases} \quad \text{其中 } g_{ii} \text{ 表示 } i \text{ 类像元之间的邻接数}$$

(双重计数法), g_{ik} 表示 i 类和 k 类像元之间的邻接数 (双重计数法), P_i 表示 i 类像元占总像元的比例:

$$^3 \text{ PLADJ 为相似邻接比,其计算公式为 } PLADJ = \left(\frac{\sum_{i=1}^m g_{ii}}{\sum_{i=1}^m \sum_{k=1}^m g_{ik}} \right) 100$$

表 4 亮温分割阈值

Table 4 The segmentation thresholds of brightness temperature

| 频段/GHz | 阈值/K |
|--------|------|
| 19 H | 190 |
| 19 V | 230 |
| 22 V | 260 |
| 37 H | 200 |
| 37 V | 240 |

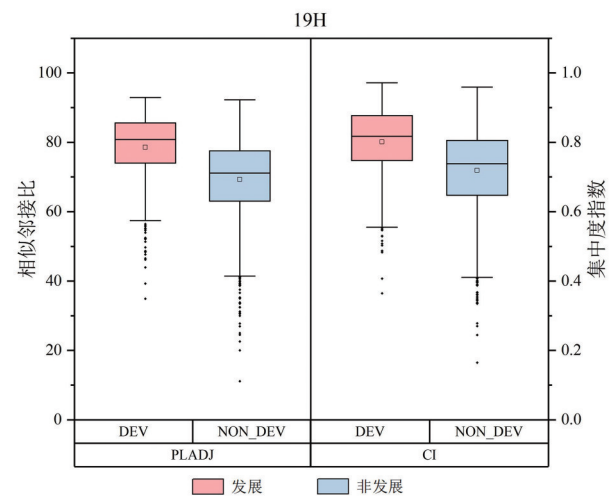
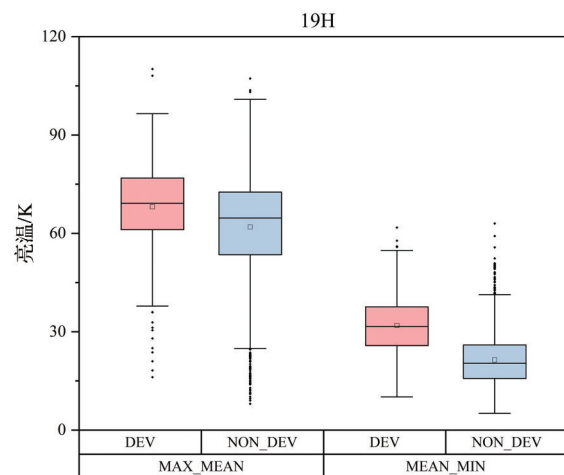
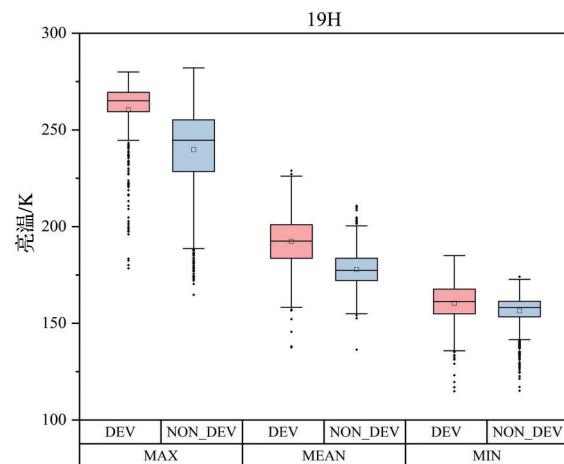


图 5 19 GHz 水平极化下各个参数数值分布箱线图

Fig.5 The box plot of distribution of parameter value in 19 GHz channel (H-polarized)

好的区分度,发展类样本的 CI 和 PLADJ 取值偏大且分布的范围更为集中。

因此,当热带气旋发生时,不仅其亮温数值可能较高,其高亮温像元的空间分布也相对集中,反之当高亮温像元的空间分布较为分散时,该热带扰动很可能不会发展成热带气旋。

3.3 C4.5 决策树算法

C4.5 决策树算法通过属性信息增益的大小来选择每个节点的分类属性,根据能够带来最大增益的属性拆分样本,递归拆分直到满足停止条件^[7]。本文使用的 C4.5 算法可利用 Weka 软件实现,该软件是集成了机器学习算法的数据挖掘并开源软件(<http://www.cs.waikato.ac.nz/ml/weka/index.html>)。

在决策树算法中,最小叶子节点的尺寸决定了规则的复杂程度。当最小叶子节点较小时,决策树较复杂,但更大的树更可能是训练样本过拟合的结果,通常会导致对未来事件的预测准确性不佳。另一方面,较小的树又无法从训练样本中获取足够的信息。因此,如何确定最佳的最小叶片尺寸是构建决策树算法的关键。最常见的策略是将训练样本中传达信息最少的节点删除^[15],通常情况下选择总样本数量的 5% 作为最小叶子节点的尺寸,但在实际研究中会根据不同的预测因子以及预测对象发生改变。图 6 为本研究决策树模型的总体精度和正

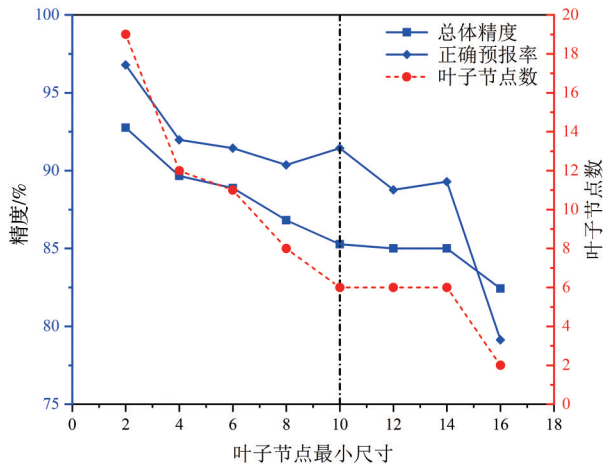


图 6 叶子节点数目及模型精度折线图

Fig.6 Line diagram of the number of leaf nodes and the model accuracy

确预报率对叶子节点最小尺寸变化的折线图。从图中可以看出,当叶子节点最小尺寸较小时,模型总体精度虽然较高,但是叶子节点数也相应较大,即决策树规则较为繁琐;而当叶子节点最小尺寸大于 14 时,模型精度过低。当叶子节点最小尺寸为 10 时,既能保证较高的模型精度,也能使得模型的规则较为简洁。因此,为保证决策树模型能从样本中获得足够多的信息,尽可能避免数据过拟合并且保证模型的简洁性,将最小叶子节点的尺寸设置为 10 (图 6 黑色虚线所示),其余参数均采用 C4.5 算法的默认值。

4 热带气旋生成预测模型

图 7 给出了利用训练集数据获得的决策树模型,其叶子节点共计 6 个,采用的 5 个特征参数分别为 19H_MAX、37H_AVG、19_MEAN_MIN、22V_PLADJ 和 37V_MAX。该模型在训练集上测试得到的分类混淆矩阵见表 5。结果表明训练集得到的总体精度

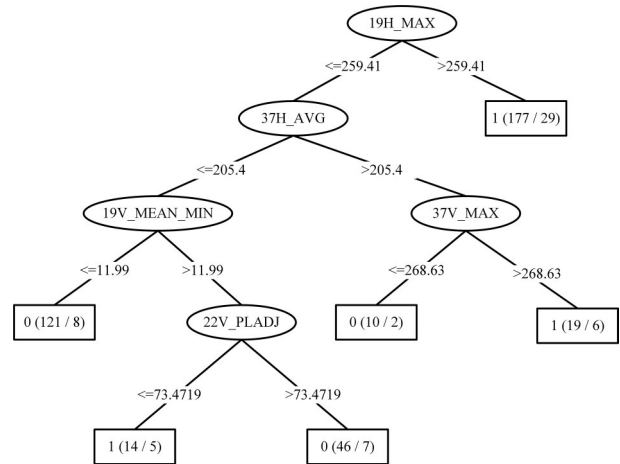


图 7 C4.5 算法建立的决策树

Fig. 7 The decision tree built by C4.5 Algorithm

表 5 决策树模型训练结果

Table 5 The training results of the Decision Tree Model

| 训练集 | | 观测 | | 分类 总和 |
|--------------------|-----|---------|-----|----------|
| | | 发展 | 非发展 | |
| 分类 | 发展 | 170 | 40 | 210 |
| | 非发展 | 17 | 160 | 177 |
| 观测数据总和 | | 187 | 200 | 387 |
| 总体精度 ¹ | | 85.27 % | | |
| 正确预报率 ² | | 90.91 % | | |
| 错误预报率 ³ | | 20.00 % | | |

注:¹ 总体精度 = $\frac{\text{正确预测的发展类和非发展类数目之和}}{\text{发展和非发展类总数目}}$

² 正确预报率 = $\frac{\text{正确预测的发展类气旋数目}}{\text{观测到的发展类气旋总数目}}$

³ 错误预报率 = $\frac{\text{错误预测的发展类气旋数目}}{\text{观测到的非发展类扰动总数目}}$

为 85.27%,对于发展为热带气旋的正确预报率为 90.91%,而错误预报率为 20.00%。

表 6 为本文建立的 6 条分类规则,其分类精度为 64.28%~93.39%。其中,规则 1、3、4 和规则 6 的分类精度均在 80% 以上,且规则 4 的精度达到了 93.39%。使用这 4 个高精度的规则可区分占总样本数量 91.47%(354/387) 的样本。相比而言,规则 2 和规则 5 的分类精度则均低于 70%,但这些规则的分类样本仅占总样本数的 8.53%(33/387)。以上

表 6 决策树规则说明以及各规则精度

Table 6 The Descriptions and accuracies of the rules derived from the decision tree

| 规则 | 规则描述 | 结果 | 精度 |
|----|---|-----|--------|
| 1 | 如果 19H_MAX>259.41 | 发展 | 83.62% |
| 2 | 如果 19H_MAX≤259.41, 37H_AVG>205.4 且 37V_MAX>268.63 | 发展 | 68.42% |
| 3 | 如果 19H_MAX≤259.41, 37H_AVG>205.4 且 37V_MAX≤268.63 | 非发展 | 80.00% |
| 4 | 如果 19H_MAX≤259.41, 37H_AVG≤205.4 且 19V_MEAN_MIN≤11.99 | 非发展 | 93.39% |
| 5 | 如果 19H_MAX≤259.41, 37H_AVG≤205.4, 19V_MEAN_MIN>11.99 且 22V_PLADJ≤73.471 9 | 发展 | 64.28% |
| 6 | 如果 19H_MAX≤259.41, 37H_AVG≤205.4, 19V_MEAN_MIN>11.99 且 22V_PLADJ>73.471 9 | 非发展 | 84.78% |

述图 3 中的两组样本为例,非发展热带扰动和台风“彩云”的 19H_MAX 分别为 265.08 和 267.27,按照规则 1,它们均被划分为发展类热带气旋,出现了一个错分现象。事实上,由于规则 1 不涉及集中度指数 CI 或者相似邻接比 PLADJ 相关的特征参数,因此这两个案例无法正确区分。规则 1 的错分数目占总的错分数目的 50.88%(29/57),这也是本文决策树模型的主要误差来源。假如按照图 6 将叶子节点最小尺寸设置为 2 时,两个示例云团的 37V_PLADJ 分别为 76.757 8 和 81.359 6,而决策规则关于 37V_PLADJ 的阈值为 78.819,因此可以将这两个案例进行准确区分。但是此时的决策树过于复杂,规则达到了 19 个,不利于实际应用。因此,为保证决策模型的简洁性并保证较高精度,最终只选择了 6 条规则构建决策模型。

表 7 为本文决策树模型在验证集上的精度测试结果:其总体观测精度为 79.47%,正确预报率为 80.00%,错误预报率为 20.79%。与训练精度相比,总体预测精度和正确预报率均有所下降,原因可能是数据量不足和不同年份的数据之间存在一定特征差异。即便如此,本文模型的正确预报率达到了 80%,说明模型较为稳定。此外,无论是训练集还是验证集,错误预报率均为 20% 左右,这与其他学者的研究结果基本一致^[15-16,23]。

表 7 热带气旋生成预测模型验证结果

Table 7 The validating results of the tropical cyclone genesis model based on decision tree

| 测试集 | 观测 | | 分类 总和 |
|--------|-----|--------|----------|
| | 发展 | 不发展 | |
| 分类 | 发展 | 40 | 21 |
| | 不发展 | 10 | 80 |
| 观测数据总和 | | 50 | 101 |
| 总体精度 | | 79.47% | |
| 正确预报率 | | 80.00% | |
| 错误预报率 | | 20.79% | |

5 结 论

本文基于 2005~2009 年间 SSM/I 微波辐射计 5 个通道(19 H、19 V、22 V、37 H、37 V)的亮温观测数据和决策树方法开展了西北太平洋热带气旋生成预测研究。首先,从辐射计影像中提取描述亮温数值特征的 5 个统计量参数和描述高亮温像元空间分布特征的 2 个景观格局参数作为模型输入量;然后,使用 C4.5 决策树算法和训练集样本构建了西北太平洋热带气旋生成预测模型;最后,在测试集样本上验证了模型精度,结果表明新模型的总体精度达到 80% 以上。本文模型的主要创新如下:

(1)与使用反演获取的风场数据或再分析得到的环境数据作为输入的方法相比,本文直接使用亮温数据作为模型输入能有效避免中间处理过程造成误差并提高了处理效率,为快速预报热带气旋发生提供了可能。

(2)在描述高亮温像元的空间分布特征时,本文引入了集中度指数(CI)和相似邻接比(PLADJ)两个景观格局参数,有效提高了模型精度。

但是,本文构建的基于微波亮温的预测模型与使用风场数据作为输入的模型在准确性上还存在一定差距,误报率也较高^[16]。后续研究将考虑使用深度学习算法来替代决策树算法进一步提高热带气旋生成预测的效果。

参考文献(References):

- [1] Zou Juhong, Lin Mingsen, Zou Bin, *et al.* Automated Cyclone Detection Using HY-2 Satellite Data[J]. *Haiyang Xuebao*, 2015, 37(1): 73-79.[邹巨洪,林明森,邹斌,等. HY-2 卫星散射计热带气旋自动识别算法[J]. *海洋学报*, 2015, 37(1): 73-79.]
- [2] Zhou Xuan, Yang Xiaofeng, Li Ziwei, *et al.* Estimation of Tropical Cyclone Parameters and Wind Fields from SAR Images[J]. *Science China: Earth Sciences*, 2014, 44(2): 355-366.[周旋,杨晓峰,李紫薇,等. 基于星载 SAR 数据的台风参数估计及风场构建[J]. *中国科学: 地球科学*, 2014, 44

- (2): 355-366.]
- [3] Li Xiaofeng, Zhang Biao, Yang Xiaofeng. Remote Sensing of Sea Surface Wind and Wave from Spaceborne Synthetic Aperture Radar[J]. Journal of Radars, 2020, 9(3): 425-443.[李 晓峰, 张彪, 杨晓峰. 星载合成孔径雷达遥感海洋风场波浪场[J]. 雷达学报, 2020, 9(3): 425-443].
- [4] Wang S, Yang X F, Li H Y, *et al.* An Improved Asymmetric Hurricane Parametric Model based on Cross-Polarization SAR Observations[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14(1): 1411-1422. doi:10.1109/JSTARS.2020.3043246.]
- [5] Halperin D J, Fuelberg H E, Hart R E, *et al.* An Evaluation of Tropical Cyclone Genesis Forecasts from Global Numerical Models[J]. Weather and Forecasting, 2013, 28(6): 1423-45. doi: 10.1175/waf-d-13-00008.1.
- [6] Park M S, Elsberry R L. Latent Heating and Cooling Rates in Developing and Nondeveloping Tropical Disturbances during TCS-08: TRMM PR Versus ELDORA Retrievals[J]. Journal of the Atmospheric Sciences, 2013, 70(1): 15-35. doi: 10.1175/jas-d-12-083.1.
- [7] Huang Chao. Research of Developing Versus Nondeveloping Tropical Cloud Clusters Classification and Key Factors Mining based on Decision Tree Algorithm[D]. Nanjing: Nanjing University of Information Science and Technology, 2016.[黄超. 基于决策树方法的热带云团发展分类及关键因子挖掘研究[D]. 南京: 南京信息工程大学, 2016.]
- [8] Schumacher A B, Demaria M, Knaff J A. Objective Estimation of the 24 h Probability of Tropical Cyclone Formation[J]. Weather and Forecasting, 2009, 24(2): 456-71. doi: 10.1175/2008waf200710.9.1.
- [9] Kalnay E, Kanamitsu M, Kistler R, *et al.* The NCEP/NCAR 40-year Reanalysis Project[J]. Bulletin of the American Meteorological Society, 1996, 77(3): 437-71. doi: 10.1175/1520-0477(1996)077<0437:Tn yrp>2.0.Co;2.
- [10] Hennon C C, Hobgood J S. Forecasting Tropical Cyclogenesis over the Atlantic Basin Using Large-scale Data[J]. Monthly Weather Review, 2003, 131(12): 2927-40. doi: 10.1175/1520-0493(2003)131<2927:F tcota>2.0.Co;2.
- [11] Fan K. A Prediction Model for Atlantic Named Storm Frequency Using a Year-by-Year Increment Approach [J]. Weather and Forecasting, 2010, 25(6): 1842-51. doi: 10.1175/2010waf2222406.1.
- [12] Hennon C C, Marzban C, Hobgood J S. Improving Tropical Cyclogenesis Statistical Model Forecasts through the Application of a Neural Network Classifier[J]. Weather and Forecasting, 2005, 20(6): 1073-83. doi: 10.1175/waf890.1.
- [13] Xiang K S, Yang X F, Zhang M, *et al.* Objective Estimation of Tropical Cyclone Intensity from Active and Passive Microwave Remote Sensing Observations in the Northwestern Pacific Ocean[J]. Remote Sensing, 2019, 11: 627. doi:10.3390/rs11060627.
- [14] Li X F, Liu B, Zheng G, *et al.* Deep Learning-based Information Mining from Ocean Remote Sensing Imagery[J]. National Science Review, 2020, 7(10): 1584-1605. doi: 10.1093/nsr/nwaa047.
- [15] Zhang W, Fu B, Peng M S, *et al.* Discriminating Developing Versus Nondeveloping Tropical Disturbances in the Western North Pacific through Decision Tree Analysis [J]. Weather and Forecasting, 2015, 30(2): 446-54. doi: 10.1175/waf-d-14-00023.1.
- [16] Park M S, Kim M, Lee M I, *et al.* Detection of Tropical Cyclone Genesis via Quantitative Satellite Ocean Surface Wind Pattern and Intensity Analyses Using Decision Trees[J]. Remote Sensing of Environment, 2016, 183: 205-214. doi: 10.1016/j.rse.2016.06.006.
- [17] Jin S H, Li X F, Yang X F, *et al.* Identification of Tropical Cyclone Centers in SAR Imagery based on Template Matching and Particle Swarm Optimization Algorithms [J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(1): 598-608. doi: 10.1109/TGRS.2018.2863259.
- [18] Han H, Lee S, Im J, *et al.* Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite based on Machine Learning Approaches[J]. Remote Sensing, 2015, 7(7): 9184-204. doi: 10.3390/rs70709184.
- [19] Mercer A, Grimes A. Atlantic Tropical Cyclone Rapid Intensification Probabilistic Forecasts from an Ensemble of Machine Learning Methods[M]. Dagli C H. Complex Adaptive Systems Conference with Theme: Engineering Cyber Physical Systems, Cas.2017:333-40. doi:10.1016/j.procs.2017.09.036.
- [20] Liu B, Li X F, Zheng G. Coastal Inundation Mapping from Bitemporal and Dual-Polarization SAR Imagery based on Deep Convolutional Neural Networks[J]. Journal of Geophysical Research: Oceans, 2019, 124: 9101-9113. doi: 10.1029/2019JC015577.
- [21] Zheng G, Li X F, Zhang R H, *et al.* Purely Satellite Data-driven Deep Learning Forecast of Complicated Tropical Instability Waves[J]. 2020, 6(29): eaba1482. doi:10.1126/sciadv.aba1482.
- [22] Wang W S, Yang X F, Li X F, *et al.* A Fully Polarimetric SAR Imagery Classification Scheme for Mud and Sand Flats in Intertidal Zones[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(3): 1734-1742. doi: 10.1109/TGRS.2016.2631632.
- [23] Defries R S, Chan J C W. Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data[J]. Remote Sensing of Environment, 2000, 74(3): 503-15. doi: 10.1016/S0034-4257(00)00142-5.
- [24] Mountrakis G, Im J, Ogole C. Support Vector Machines in Remote Sensing: A Review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 247-59. doi: 10.1016/j.isprsjprs.2010.11.001.
- [25] Kim M, Park M S, Im J, *et al.* Machine Learning Approaches for Detecting Tropical Cyclone Formation Using Satellite Data [J]. Remote Sensing, 2019, 11(10). doi: 10.3390/rs11101195.
- [26] Matsuoka D, Nakano M, Sugiyama D, *et al.* Deep Learning Approach for Detecting Tropical Cyclones and Their Precursors in the Simulation by a Cloud-resolving Global Nonhydrostatic Atmospheric Model[J]. Progress in Earth and Planetary

- Science, 2018, 5. doi: 10.1186/s40645-018-0245-y.
- [27] Zhang T, Lin W Y, Lin Y L, *et al.* Prediction of Tropical Cyclone Genesis from Mesoscale Convective Systems Using Machine Learning[J]. Weather and Forecasting, 2019, 34(4): 1035-49. doi: 10.1175/waf-d-18-0201.1.
- [28] Su H, Wu L T, Jiang J H, *et al.* Applying Satellite Observations of Tropical Cyclone Internal Structures to Rapid Intensification Forecast with Machine Learning[J]. Geophysical Research Letters, 2020, 47(17). doi: 10.1029/2020gl089102.
- [29] Tan Yongqiang, Fei Jianfang, Zhang Shaohong, *et al.* Using SVM Network to Adjust the Limb Effect of AMSU-A[J]. Journal of Tropical Meteorology, 2010, 26(2): 187-193.[谭永强, 费建芳, 张少洪, 等. 支持向量机在 AMSU-A 临边调整中的应用[J]. 热带气象学报, 2010, 26(2): 187-193.]
- [30] Goldberg M D, Crosby D S, Zhou L H. The Limb Adjustment of AMSU-A Observations: Methodology and Validation[J]. Journal of Applied Meteorology, 2001, 40(1): 70-83. doi:10.1175/1520-0450(2001)040<0070:Tlaaaa>2.0.Co;2.
- [31] Weng F Z, Grody N C. Retrieval of Cloud Liquid Water Using the Special Sensor Microwave Image (SSM/I)[J]. Journal of Geophysical Research-Atmospheres, 1994, 99(D12): 25535-51. doi: 10.1029/94jd02304.
- [32] Ferraro R R, Weng F Z, Grody N C, *et al.* An Eight-year (1987~1994) Time Series of Rainfall, Clouds, Water Vapor, Snow Cover, and Sea Ice Derived from SSM/I Measurements[J]. Bulletin of the American Meteorological Society, 1996, 77(5): 891-905. doi: 10.1175/1520-0477(1996)077<0891:Aeytso>2.0.Co; 2.
- [33] Ferraro R R. Special Sensor Microwave Imager Derived Global Rainfall Estimates for Climatological Applications[J]. Journal of Geophysical Research: Atmospheres, 1997, 102(D14): 16715-35. doi: 10.1029/97jd01210.
- [34] Hennon C C, Helms C N, Knapp K R, *et al.* An Objective Algorithm for Detecting and Tracking Tropical Cloud Clusters: Implications for Tropical Cyclogenesis Prediction[J]. Journal of Atmospheric and Oceanic Technology, 2011, 28(8): 1007-1018. doi: 10.1175/2010jtecha1522.1.
- [35] Geng Xiaoqing, Li Ziwei, Yang Xiaofeng. Tropical Cyclone Auto-recognition from Stationary Satellite Imagery[J]. Journal of Image and Graphics, 2014, 19(6): 964-970.[耿晓庆, 李紫薇, 杨晓峰. 静止卫星图像热带气旋云系自动识别[J]. 中国图象图形学报, 2014, 19(6): 964-970.]

Tropical Cyclone Genesis Prediction based on Spaceborne Microwave Brightness Temperature Observations in the Northwest Pacific Ocean

Wang Sheng^{1,2}, Yang Xiaofeng^{1,3}, Ma Wentao^{1,3}, Xiang Kunsheng^{1,4}, Hu Die²

(1. The Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute,

Chinese Academy of Sciences, Beijing 100101, China.;

2. University of Chinese Academy of Sciences, Beijing 100049, China.;

3. The Key Laboratory of Earth Observation of Hainan Province, Sanya 572029, China.;

4. Beijing Piesat Information Technology Co. Ltd, Beijing 100195, China)

Abstract: Forecasting and identifying the clouds that may develop into Tropical Cyclones (TCs) from tropical clouds are essential for early warning disasters. In this study, spaceborne microwave brightness temperature observations are used to build a new model to predict TC genesis in the Northwest Pacific. A sample dataset consists of 326 developing TC samples, and 2 112 non-developing tropical disturbance samples is established. They are collected by DMSP satellite during the typhoon season (May to October) from 2005 to 2009. Based on this dataset, 35 brightness temperature characteristic parameters were calculated including 25 statistical parameters and 10 landscape pattern parameters. Landscape pattern parameters are included to describe the spatial distribution of brightness temperature near TC center. The dataset is then divided into a training set (2007~2009) and a test set (2005~2006). For the training set, the model's overall accuracy is 85.27%, the hit rate and false alarm rate of tropical cyclones are 90.91% and 20%, respectively. For the test set, the above three parameters are 79.47%, 80%, and 20.79%, respectively. Therefore, the tropical cyclone generation model based on microwave brightness temperature data can effectively predict tropical cyclones' development.

Key words: Tropical cyclone genesis; SSM/I brightness temperature; Decision tree method; Northwest Pacific Ocean