机器学习在西北太平洋热带气旋生成前期大尺度环流场分型

与识别中的应用

赵字慧^{1,2}陈光华¹王紫清^{1,2}方荻^{1,2,3} 1中国科学院大气物理研究所云降水物理与强风暴重点实验室,北京100029 2中国科学院大学,北京100049 3贵州省山地环境气候研究所,贵阳550002

摘要:基于 1979-2020 年 6-11 月的热带气旋最佳路径(IBTrACS)和欧洲中期天气预报中心的第五代再分析(ERA5)资料,本文根据以热带气旋(TC)生成位置为中心的 850hPa 水平风场特征,采用自组织映射网络(SOM)将西北太平洋 TC 生成前期的低层大尺度环流场分为 5 型:季风辐合型(MC)、季风涡旋型(MG)、强季风槽型(SMT)、弱季风槽型(WMT)及东风波型(EW)。MC 型 TC 生成于副热带高压南侧辐合带中,占比最高; MG、SMT 与 WMT 三型的 TC 生成受季风槽相关的气旋性切变或辐合区影响; EW 型 TC 由东风波增幅发展生成,占比最小。在对历史资料分型的基础上,为选取合适的机器学习方法用于 TC 环流型的自动识别,本文还对比分析了支持向量机(SVM)、k 近邻(KNN)及随机森林(RF)三种方法的识别效果,结果表明:SVM 的准确率达 0.965,对五类环流型识别的 召回率和精确率均达到 0.94 以上,对样本不均衡问题不敏感,并且对样本量的敏感性分析显示其在有限样本量下即可充分学习各型的环流场特征,识别效果明显优于 KNN 和 RF。 关键词:大尺度环流型;自组织映射网络;支持向量机;热带气旋

文章编号: 2022074B

doi:10.3878/j.issn.1006-9895.2208.22074

Application of Machine Learning in Clustering and

Discriminant Analysis of Large-scale Circulation Patterns

Favorable for Tropical Cyclogenesis over the Western North

Pacific

ZHAO Yuhui^{1,2}, CHEN Guanghua¹, WANG Ziqing^{1, 2}, FANG Di^{1, 2, 3}

1 Key Laboratory of Cloud-Precipitation Physics and Severe Storms, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

2 University of Chinese Academy of Science, Beijing 100049

收稿日期 2022-05-08; 网络预出版日期 作者简介 赵宇慧, 女, 1999年出生, 硕士研究生, 主要从事热带气旋研究。E-mail: zhaoyuhui21@mails.ucas.ac.cn

通讯作者 陈光华, E-mail: <u>cgh@mail.iap.ac.cn</u>

资助项目 国家重点研发计划项目 2017YFC1501901; 国家自然科学基金项目 42175073、41975071 **Funded by** National Key Research and Development Program of China (Grant 2017YFC1501901); National Natural Science Foundation of China (Grants 42175073, 41975071)

Abstract: Based on the IBTrACS dataset and ERA5 850-hPa winds from July to November in 1979-2020, the low-level large-scale circulations associated with tropical cyclogenesis over the western North Pacific are clustered into five patterns using self-organizing map (SOM). The five patterns are named Monsoon Confluence (MC), Monsoon Gyre (MG), Strong Monsoon Trough (SMT), Weak Monsoon Trough (WMT) and Easterly Wave (EW), respectively. Tropical cyclones (TCs) in the MC pattern form in the confluence zone south of the subtropical high, occupying the largest proportion. Cyclogeneses in the MG, SMT and WMT patterns are affected by the cyclonic wind shear or the confluence zone related to the monsoon trough. The EW pattern with the smallest number of cases features an easterly wave directly evolving into a TC. To select an optimal machine learning method for automatic pattern identification for a given TC circulation, comparison is carried out among three discriminant analysis models: support vector machine (SVM), k-nearest neighbors and random forest. The results show that SVM reveals the best accuracy of 0.965 and the least sensitivity to imbalanced data with recall rate and precision exceeding 0.94 for each circulation pattern. Meanwhile, the sensitivity to dataset size indicates that using SVM model, characteristic signals can be most effectively captured from relatively limited training data.

Key words: Large-scale circulation patterns; Self-organizing Map; Support Vector Machine; Tropical cyclone

1. 引言

西北太平洋是热带气旋(Tropical cyclone, TC)生成最为活跃的海域,生成于其中的 TC 大多在环境气流引导下移向东亚和东南亚沿岸国家,并带来大风、强降水、风暴潮、洪涝等 一系列强烈的天气现象和自然灾害。由于利于 TC 生成的环境因子及前期扰动与其所处的大 尺度环流背景密切相关,因此对与 TC 生成相关的大尺度环流场进行分类对预测 TC 生成进而 减小其危害有重要意义。

西北太平洋夏季海表温度通常都高于 26.5℃(Gray, 1968),能够为 TC 生成提供高温高 湿的大气环境,满足 TC 生成所必需的热力条件;除热力条件外,TC 生成还需基本的动力条 件,包括初始扰动、适宜的垂直风切变以及一定的地转偏向力(Gray, 1968; Briegel and Frank, 1997; Cheung, 2004)。西北太平洋夏秋季的季风环流能够为绝大多数 TC 生成创造有利的大尺 度环境条件,其中季风槽是主要的影响系统,研究表明西北太平洋 70%以上的 TC 在季风槽 中生成(Richie and Holland, 1999; Chen et al., 2004),它可以为 TC 生成提供低层辐合及气旋 性涡度、中下层充足的水汽、高层辐散及较小的垂直风切变等环境条件(冯涛等, 2013;黄荣 辉等, 2016)。Holland(1995)研究表明西南季风和盛行东风所产生的辐合区存在多尺度相互 作用,能够强化利于 TC 生成的条件,通过增幅传播到此的热带波动而增大 TC 生成的可能 性。因此,TC 生成与天气尺度扰动和大尺度环境场间的相互作用密切相关。Fu et al. (2006)曾将天气尺度扰动分为 TC 诱发的 Rossby 波能量频散、天气尺度波列及东风波三 类,且在所研究的 11 例天气尺度波列型 TC 中,有 4 例源自混合 Rossby 重力波转变为 TD 型 波动后产生的先兆扰动。

许多研究对影响西北太平洋 TC 生成的环流型进行了分类(Zehr, 1992; Briegel and Frank, 1997; Richie and Holland, 1999; 王蔚等, 2008; Feng et al., 2014)。Zehr (1992) 阐明了季风在 TC 生成中所发挥的重要作用,将与季风有关的低层大尺度环流分为东风型、季风-东风型、季风-弱西风型、季风-强西风型四类。Ritchie and Holland (1999)通过对 TC 生成前期散度、 涡度等气象要素的合成分析,将大尺度环流归结为季风切变(MS)、季风辐合(MC)、季风 涡旋(MG)、东风波、Rossby 能量频散五型,其中 MS 和 MC 有较强的气旋式切变和低层辐合,且伴有更为活跃的中尺度对流现象,生成的 TC 超过 70%。Feng et al. (2014)以 1991-2010 年间生成于西北太平洋台风季的 TC 为研究对象,认为有利于 TC 生成的大尺度环流型 有季风切变型、季风辐合型、季风倒槽型、季风涡旋型及东风型五类,前三类中生成的 TC 数量可达总数的 80%以上,且其 TC 生成的物理机制涉及正压能量转换过程。以上研究大都 是基于主观分类,而 Yoshida and Ishikawa (2013)针对 Ritchie and Holland (1999) 定义的五 类环流型提出过一套客观分类及识别标准,定义五类环流型对应的贡献分数公式,对于某一 TC,贡献分数最高者即为对该 TC 生成有主要贡献的大尺度环流型。

近年来,迅速发展的机器学习技术丰富了气象学的研究方法并已得到广泛应用(郑颖青等,2013; Bao and Wallace,2015;周璞等,2016; Neetu et al., 2020; Zhang et al., 2022)。Bao and Wallace (2015)采用自组织映射网络(Self-organizing map, SOM)对北半球冬季 500hPa 高度场进行分型,发现了西半球环流型这一新的大气低频模态,且相比层次聚类法,SOM 所得四种低频模态线性无关性更强,其中三型有助于解释北美及欧洲的三次历史极端气候事件。Jergensen et al. (2020)对比了多种机器学习方法对三类雷暴(超级单体、准线性对流系统及其他)的识别效果,结果显示梯度提升决策树和支持向量机表现最佳,准确率分别为0.77±0.02及0.70±0.02。Bhowmick et al. (2021)评估了随机森林(Radom Forest, RF)和决策树对西南太平洋热带低压和热带风暴的区分能力,其中 RF 以 91%的准确率体现了其作为集成方法的优势,并指出海盐气溶胶光学厚度、1000hPa 相对湿度和海表温度是区分二者的最重要变量。

本文主要包括两部分内容,首先采用 SOM 聚类方法对西北太平洋 TC 生成前期的低层大 尺度环流场进行分型,然后选用支持向量机(Support vector machine, SVM)、k 近邻(knearest neighbors, KNN)及 RF 三种在小样本识别问题上表现较好的有监督学习方法建立识别 模型(Cortes and Vapnik, 1995; Cotter and Polagye, 2020),客观评估不同机器学习方法对 TC 环流型识别的表现。所用数据及其预处理在第 2 节给出,模型的建立及评估方法于第 3 节说 明,第 4 节对 TC 生成前期不同环流型的特征进行分析,并给出三种识别模型的评估结果, 第 5 节为总结与讨论。

2. 数据

2.1 数据来源

本文将 1979-2020 年作为研究时间段,以为聚类及识别模型的建立提供充足的 TC 样本 规避过拟合问题,且仅考虑西北太平洋 TC 活动较为活跃的 6-11 月份。具体 TC 样本筛选自 联合台风警报中心最佳路径(International Best Track Archive for Climate Stewardship, IBTrACS)数据集(Knapp et al., 2010),资料的时间分辨率取 6h,主要应用的变量包括: TC 中心的经纬度及最大持续风速(MSW)。大尺度环境场数据来自欧洲中期天气预报中心的 ERA5 数据集(Hersbach et al., 2020),资料的时间分辨率取 6h,水平分辨率取 2.5°×2.5°;所 用变量包括 850hPa 的水平风(U、V风分量)、位势高度以及相对涡度,其中 U、V风分量用 于构造各 TC 样本的特征向量。

2.2 数据预处理

大尺度环流等外强迫在热带扰动发展至热带低压并热带低压初始发展过程中发挥着重要 作用,此后 TC 发展转入自主维持阶段,外强迫的作用相对弱化(Zehr, 1992);因此前者是 本研究关注时段,即 TC 生成前期。本文将 MSW 首次达到 12m/s 且至少维持 24h 不减弱的时 刻定义为 TC 生成时刻,此前 24h 为 TC 生成前期,其在 6h 时间分辨率下包含 0h,-6h,-12h,-18h,-24h 五个时刻。根据以上定义,从 IBTrACS 西北太平洋数据集中初步筛选得 TC 样本共计 985 例。

现有研究多以流场直观体现低层大尺度环流状况及类型,我们将其作为对环流场分型及 识别的客观依据。TC 样本的特征向量是由其生成位置(TC 生成时刻所在经纬度)周围 30°×30°区域内的 850hPa U、V 风分量串联而成的一维向量,在所取 2.5°×2.5°水平分辨率 下,其长度为 338。U、V 风场为 TC 生成前期的平均场,并由空间九点平滑滤去尺度较小的 扰动。为减小风速差异对结果的影响,采用如下公式对各特征向量进行归一化处理,将各维 数据映射到(-1,1)区间内:

$$X(i)' = \frac{X(i)}{||X(i)||} ,$$
 (1)

其中*X*(*i*)为样本*i*的特征向量, ||*X*(*i*)||为其欧几里得范数。此外, 区域的大小决定了 TC 样本的特征维度, 进而影响算法性能。前期多次试验表明, 30°×30°的区域可以充分体现 TC 生成前期的大尺度环流场特征,且能保证模型的高泛化性能。

3. 模型构建

3.1 SOM 聚类模型

SOM(Kohonen et al., 1982)包括输入层和竞争层两层结构,竞争层上各神经元都被赋 以一个权重向量,其长度与样本特征向量相同。作为一种无监督学习方法,SOM 根据权重向 量与输入样本间的相似性,判定对应于各输入样本的获胜神经元,并不同程度地调整获胜神 经元及其拓扑邻域内其他神经元的权重向量,使其移近对应的输入样本。以上竞争学习过程 的不断迭代优化最终使权重向量分布逼近输入样本分布,继而每一个神经元可代表一类样 本,且输入层中相似的两类样本所映射的神经元相近,即拓扑正确的映射。

3.1.1 聚类数目的选择

以 SOM 进行聚类分析,需要预先设定聚类数目。理想的聚类数目期望满足:类内样本 相似度尽可能高,而类间相似度尽可能低,且尽可能减小聚类结果对抽样变化的敏感性。

采用 Bao and Wallace (2015) 定义的方差比(Variance ratio, VR)来量化评估聚类结果的 独特性: VR 为"外部方差"与"总方差"之比,前者表示某类样本质心到总样本质心的平 方距离,后者表示该类中各样本到总样本质心平方距离的均值。这一指标可视为对某类样本 聚集程度的度量,其值越大表示样本聚集程度越高。计算不同聚类数目下各类的 VR 值并做 平均,结果如表 1 所示。可以看出,样本的平均 VR 值随聚类数目的增加而增加,但增幅逐 渐减小,若每个样本都自成一类,VR 即达到其极限值 100%。前期相对大的增幅符合环流分 型的目的(如表 1 示),当聚类数目超过某一限度,继续增加聚类数目是无意义的,并且会导 致模型过学习,其结果会对抽样变化非常敏感。

表 1 不同聚类数目下样本的平均方差比(VR) Table 1 Averaged variance ratio (VR) corresponding to different number of clusters

聚类数目	2	3	4	5	6
平均 VR 值	12.7%	21.2%	25.2%	28.3%	31.1%

通过计算不同聚类数目下各类样本质心间的空间相关性,并对各类的合成场进行天气学分析,发现当聚类数目由 2 依次增加到 5 的过程中,都会有一类具有独立特征的环流型被分离出来。如,当聚类数目取 4 (所得各类样本记为IV-1 至IV-4)和 5 (V-1 至V-5;与图 1 对应)时,各类样本质心间的空间相关值如表 2 所示,可知当聚类数目增加到 5 时,除V-4外,其它各类均与聚类数目为 4 时的各类有一一对应的高空间相关值(表 2 中粗体所示),天气学分析表明,V-4的出现代表弱季风槽及其切变背景对部分 TC 生成的作用得到分离及凸显,具体环流型命名及特征分析详见 4.1和 4.2节。当继续增加聚类数目至 6 时,所得各类(VI-1 至VI-6)均与V-1 至V-5 中的某类高度相关(表 3),并且存在两类 TC 样本(VI-3 和VI-5)同时与V-1 高度相关,这表明不再有新的环流型被分离出来,且开始出现天气学意义上的无效分型。综上,我们将与西北太平洋 TC 生成有关的低层大尺度环流型分为五类。

表 2 聚类数目为 4 和 5 时所得各类样本质心间的空间相关性。大于 0.9 的值以粗体表示。

Table 2 Spatial correlations between	patterns derived from 4 and 3	5 number of clusters.	Boldface denotes
	values higher than 0.9		

		values inglier th	dii 0.7.		
	IV-1	IV-2	IV-3	IV-4	
V-1	0.642	0.998	0.382	0.515	
V-2	0.595	0.478	0.078	0.986	
V-3	0.997	0.669	0.332	0.574	
V-4	0.550	0.484	0.749	0.751	
V-5	0.360	0.416	0.916	-0.051	

	Table :	S AS III Table 2 exc	ept for 5 and 6 m	inder of clusters	
	V-1	V-2	V-3	V-4	V-5
VI-1	0.488	0.998	0.581	0.591	-0.176
VI-2	0.410	0.685	0.515	0.981	0.384
VI-3	0.936	0.620	0.544	0.645	0.299
VI-4	0.566	0.585	0.984	0.544	0.318
VI-5	0.912	0.381	0.784	0.368	0.477
VI-6	0.299	-0.166	0.258	0.501	0.990

表 3 同表 2,但为聚类数目取 5 和 6 的结果 Table 3 As in Table 2 except for 5 and 6 number of cluster

3.1.2 聚类结果的评估

以可复现性参数(Reproducibility parameter, RP; Bao and Wallace, 2015)考察聚类结果的 稳健性,具体计算方法为:将输入样本随机分为三组,对每一组 TC 样本进行 SOM 聚类分 析,相应绘制合成流场(图略)并与五类环流型(图 1)相匹配,对各组间相匹配的合成流 场求取空间相关性,因此每一类环流型对应三个空间相关值,求其平均作为该环流型的 RP 值。结果表明,面对输入样本的变化,五类环流型都对应有较高(>0.85)的 RP 值(详见图 1),维持了稳定的聚类结果,这也可以作为确定五类典型环流型的佐证。

SOM 可以得到五类环流型及其所包含的 TC 样本,但由于 TC 环流存在复杂和多样性,因此部分 TC 样本的流场可能接近于其它型,而难以界定其确切归类,并且算法本身的局限性也可能导致某些 TC 归类"错误"。为给 3.2 节识别模型提供可靠的历史资料,基于轮廓系数(Rousseeuw, 1987)筛去分类模糊或分类"错误"的 TC 样本:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$
(2)

其中a(i)是样本 i 到它所属类中其他样本的平均欧氏距离, b(i)是此样本到与它相邻最近 类中各样本的平均欧氏距离, S(i)的取值范围为[-1,1]。相对其他样本而言, 若某一样本的 S(i) 值接近于 1, 说明它的内聚度和分离度都较优; 取值接近 0 表示样本位于边界处; 取值 为负则表示样本被分到别的类中, 即分类"错误"。

综合比较所有 TC 样本的轮廓系数值,将小于临界值 S₀(设为 0.005)的 95 例 TC 样本(占比 9.6%)移入"无效集"。经筛选得有效 TC 样本共计 890 例,以此保证各 TC 环流型类别的准确性,后续将不再考虑无效集中的样本。

3.2 识别模型

基于 SOM 提供的具有明确环流型的历史 TC 样本,本节将建立 SVM、KNN 及 RF 模型,并采用 Holdout 检验法客观评估三个模型对 TC 环流型的识别效果,将来业务可以应用本研究评估的最优模型对 TC 环流型进行自动识别。Holdout 检验法是一种分割样本集进而检验

模型效果的方法,它将样本集按一定比例随机划分为训练集和测试集,前者供以特征学习和 调参,后者用于评估模型性能(Rozas-Larraondo, 2014)。本研究中取 7:3 比例,即 70%的 TC 样本作为训练集,其余 30%为测试集,据此确定能以高可信度和高稳健性对 TC 环流型进 行自动识别的模型。

SVM(Cortes and Vapnik, 1995)最初是为解决二分类问题而提出。通过选择一个适当的 核函数,可以将输入向量映射到更高维的特征空间中,在此空间中寻求一个能以最大间隔将 两类样本完全或近似分隔开的线性决策面(最优超平面),在此基础上即可建立决策函数以 识别新样本类别(Burges, 1998; Neetu et al., 2020)。本文选用一对一算法(One-against-one, OAO; Knerr et al., 1990)将二分类 SVM扩展来解决多分类问题,在任意两类样本间寻找最优 超平面进而建立决策函数,所有决策函数的组合构成 OAO 多分类 SVM 模型,将待识别样本 的输入向量依次通过各决策函数,胜出次数最多的类别即为其最终归类。

在 KNN (Cover and Hart, 1967)中,待识别样本的归类直接取决于特征空间中与其相邻 最近的 k 个训练样本的类别,并最终归属于占比最高者或按距离加权后占比最高者。显然, 超参数 k 是决定模型泛化性能的关键因子,理论上训练集中噪音样本越多,关键参数 k 的最 优取值就越大 (Hall, 2010),且新样本识别的计算量会因 k 值及训练集样本量的不同而有显 著差异,具体参数的寻优见 3.2.1 节。

RF 是由多个决策树组成的集成学习方法(Breiman, 2001),不同决策树的生成源于对训练样本和分类特征的随机抽取,因此相互间的相关性较小。待识别样本的类别为各决策树输出类别中的占比最高者,因此一定程度上减小了单一决策树过拟合问题的影响进而有更高的模型准确率。

3.2.1 超参数寻优

为使模型在特定数据集上表现出较优的识别效果,需要对其涉及的关键超参数进行寻优。在 SVM 模型中,决策函数的构建主要涉及核函数和惩罚因子两个超参数的选择和优化,由于 SOM 所得任意两类样本在原始特征空间中的近似线性可分性,选用线性核作为核函数,同时也避免了因核函数设定而带来的附加参数,因此本研究只关注控制模型对训练集拟合程度的惩罚因子的取值。KNN 模型的构建基于 k 值的选择,且多次试验表明,不进行距离加权的 KNN 模型对 TC 环流型的识别准确率更高。单个决策树的建立可以人为干预的参数较多,但对多决策树集成的 RF 模型,我们仅考虑对效果影响最为显著的决策树个数的设定(Cotter and Polagye, 2020)。

以上各模型的关键超参数采用 10-折交叉验证进行寻优确定,具体而言:将训练集大致 均分为 10 组,循环取其一组作为验证集,另 9 组用于模型训练,共得 10 个训练模型及其相 应验证集的准确率,据其均值可以择优确定超参数取值。需要说明的是,由于 Holdout 法划 分的训练集是随机的,超参数的最优值会随训练集的不同而存在变动,但对模型效果的影响 几可忽略,因此各模型的超参数取值以表 4 为准。

表 4 三种识别模型的超参数及其取值	Ī
--------------------	---

Table 4	Hyperparameter	and its	value of	each model	

识别模型	超参数	参数值	
SVM	惩罚因子	2.4	
KNN	k	13	
RF	决策树个数	170	B

3.2.2 识别效果评估指标

本文选用准确率、召回率及精确率三个指标,基于测试集来量化评估模型的泛化性能。 准确率(Accuracy)指环流型被正确识别的TC样本数占测试集样本总数的比例,即:

$$Accuracy = \frac{\sum_{i} TP_{i}}{\sum_{i} N_{i}},$$
(3)

其中 TP_i 为类 i 中被正确识别的样本数, N_i 为该类的样本总数。准确率是对模型全局准确程度的评估,但由于各类环流型下生成的 TC 数量存在明显差异,样本的不均衡使得此指标不能全面代表模型对各型环流场的识别效果,因此我们借助召回率及精确率做出进一步的分析。

召回率(Recall rate)指某类中被正确识别的样本数占该类样本总数的比例,即:

$$R_i = \frac{TP_i}{N_i} , \qquad (4)$$

精确率(Precision)指被识别为某类环流型的样本中真正属于该类的比例,即:

$$P_i = \frac{TP_i}{TP_i + FP_i} , \qquad (5)$$

其中FPi为被错误识别为类i的样本数。

对 TC 样本的环流型集中识别时, 召回率和精确率分别衡量模型对各类 TC 识别的全面性和准确性。当对某一 TC 的环流型进行识别时, 精确率更能反映识别结果的可信程度。

在计算上述三个指标时,为减少样本划分带来的随机误差,我们采用 Holdout 法对所有 TC 样本进行 100 次随机划分,取其均值作为对模型性能的最终描述,也可借此评估不同模型 的稳健性。

3.2.3 样本量敏感性分析

样本量是影响模型性能的重要因素,我们期望所选模型通过较少的训练样本即可获得较优的学习效果,为此基于准确率定量考察三种模型对训练样本量的敏感性。为使准确率这一指标更具代表性,对前述 Holdout 划分做出如下调整以消除样本不平衡的影响:(1)以最少类的 TC 样本量为准(EW型,116例;详见图1),从其他四类中随机抽取同等量的样本组成新的 TC 样本集;(2)从各类中抽取 30%(35例)的样本组成测试集;(3)从各类剩余的 TC 样本中逐次抽取不同量的样本组成训练集用以模型建立及超参数寻优,各次取样量分别占各类样本总量的 10%(12例),20%(23例),30%(35例),40%(46例),50%(58例),60%(70例)及 70%(81例)。以上过程重复进行 100次,取其准确率均值作为评价特定训

练样本量下模型性能的指标。

4. 结果分析

4.1 TC 生成前期低层大尺度环流型

基于 850hPa 流场特征, SOM 将西北太平洋 TC 生成前期的低层大尺度环流场分为五种 典型模式(图 1): 季风辐合型(Monsoon Confluence, MC)、季风涡旋型(Monsoon Gyre, MG)、强季风槽型(Strong Monsoon Trough, SMT)、弱季风槽型(Weak Monsoon Trough, WMT)、东风波型(Easterly Wave, EW)。具体命名规则和演变特征将在 4.2 节介绍。在季风 气流影响下生成的 TC 占比达 87%, 其中 MC 型和 MG 型是影响 TC 生成的两个主要环流 型,生成于其中的 TC 比例分别为 27.5%(245 例)和 23.6%(210 例),其余 19.7%(175 例)和 16.2%(144 例)的 TC 生成在 SMT 和 WMT 两型中;此外 EW 型控制着约 13%(116 例)的 TC 生成活动。



图 1 五类低层大尺度环流型在 TC 生成前期的 850hPa 合成风场(流线)、合成位势高度场(填色,单位:位势米)。合成场以 TC 生成位置为中心。各型 RP 值及其包含的 TC 数量及比例于图中注明。(a) MC 型;(b) MG 型;(c) SMT 型;(d) WMT 型;(e) EW 型;(f) 全部 TC 的平均场。

Figure 1 Composite of 24h-averaged 850-hPa horizontal wind (streamlines) and geopotential height (shaded, unit: gpm) before TC genesis for (a) MC; (b) MG; (c) SMT; (d) WMT; (e) EW; (f) all TCs. The domain is centered at TC genesis position. TC number and percentage are shown in the title while the value of RP at the top right.

五类环流型在西北太平洋(0°-40°N,100°E-180°)的合成流场和位势高度场如图 2 所示,关注各环流型中季风槽与西太平洋副热带高压的位置、强度以及 TC 生成位置分布的差异。同时,大尺度系统的季节变化也令各环流型下 TC 生成较为活跃的月份有明显的差异(图 3)。

MC型(图2(a))是10-11月的主导环流型(图3),这一时期季风槽衰减为涡旋和辐合带(张翔等,2017),副热带高压西伸至亚洲大陆,TC生成在其南侧的赤道辐合带中,经向跨度较大,平均生成位置最偏南(12.8°N;图2(f)),且个别TC在5°N以南的低纬地区生成,这可能与冬季越赤道气流伴随的较大正涡度有关(王蔚等,2008;李艳等,2019)。MG型(图2(b))多发于西南季风和南半球越赤道气流盛期,即8-9月TC生成最为活跃(图3),副热带高压东撤,季风槽可东伸至150°E以东,它产生很强的气旋式切变使得TC集中生成其内部,中国南海北部和菲律宾群岛以东洋面为MG型TC的两个主要源地,TC平均生成位置是五类环流型中最偏西的(图2(f))。SMT型(图2(c))和WMT型(图2(d)) TC生成的季节变化基本一致,在8月达到峰值(图3),二者的合成场相仿,但SMT型中季风槽强于WMT型;而且,SMT型中TC多生成在季风槽东南侧,相比较而言,WMT型TC多生成在相对偏西偏北的更高纬地区,这种生成位置差异的产生原因将在4.2节进行讨论。除受越赤道气流及西南季风有关的辐合带或季风槽影响外,还有部分TC生成于副热带高压 西南侧的偏东信风中(图2(e)),尤其是当西南季风较弱,季风槽位置偏西,副热带高压较强时,因此EW型TC的平均生成位置偏东偏北(21.1°N,147.3°E;图2(f)),其TC生成在7月最为频繁(图3)。



图 2 五类低层大尺度环流型在西北太平洋的 850hPa 水平风场 (矢量,单位: 8 ms-1) 及位势高度场 (填色,

单位: 位势米)。环境场为在 TC 生成前期的平均合成场。彩色圆点表示各型 TC 的生成位置; (f) 图中相应颜色的 TC 符号表示其平均生成位置。(a) MC 型; (b) MG 型; (c) SMT 型; (d) WMT 型; (e) EW 型; (f) 全部 TC 的平均场。

Figure 2 Composite of 24h-averaged 850-hPa horizontal wind (vectors; unit: 8ms⁻¹) and geopotential height (shaded, unit: gpm) before TC genesis over the WNP for (a) MC; (b) MG; (c) SMT; (d) WMT; (e) EW; (f) all TCs. Colored dots denote the location of TCs genesis and TC symbols in (f) stand for the

mean locations.

80 60 TC number 40 20 0 Jun Jul Aug Oct Nov Sep EW MG MC MC SMT WMT

图 3 五类低层大尺度环流型中 TC 生成数量随月份的变化 Figure 3 Monthly variation of TC number in each pattern.

4.2 各低层大尺度环流型的演变特征

上节我们考察了各环流型下 TC 生成的一些统计特征,这一节将以 TC 生成位置为中心对 各型进行合成,由大尺度环境场在 TC 生成前 72h 内的演变讨论各型 TC 的成因。

MC型(图 4) TC 生成在赤道辐合带的向极一侧,合成图上没有明显的季风槽出现,使 其区别于下述 SMT 型。在副热带高压南侧,越赤道气流与东风带交汇形成强的辐合带,产 生纬向分布的正涡度区,TC 生成前 72h 时(图 4 (a)),相对涡度中心位于生成位置东侧, 此后随时间西移且不断加强致 TC 生成。





图 4 MC 型的 850hPa 水平风场(流线)、位势高度(填色,单位:位势米)和相对涡度(等值线,单位:10⁻⁶s⁻¹)的合成场在 TC 生成前 72h 内的演变情况。相对涡度等值线取值范围为 4-12,间隔为 2。(a) -72h; (b) -48h; (c) -24h; (d) 0h。

Figure 4 Evolution of composite 850-hPa horizontal wind (streamlines), geopotential height (shaded, unit: gpm) and relative vorticity (contours, unit: 10⁻⁶ s⁻¹) within 72h before TC genesis in the MC pattern.
Contours indicate values ranging from 4 to 12 with an interval of 2. (a) -72h; (b) -48h; (c) -24h; (d) 0h.

MG型(图 5) TC 生成于强季风槽内部,其南北两侧盛行的偏西风和偏东风产生了经向延伸的强切变,相应有强的带状正涡度分布。在 TC 生成前 48h 时(图 5 (b)),季风槽内部即发展出一个自组织的气旋性涡旋(季风涡旋),季风涡旋若进一步加强可转变生成尺度较大的 TC (图 5 (d))。许多研究曾对季风涡旋给出过不同的天气学定义(Lander, 1994; Feng et al., 2014;杨琼琼和吴立广, 2015)。如在 Feng et al. (2014)的定义中,季风涡旋被描述为直径大于 2500km 的近圆形孤立闭合涡旋,对流活动发生在东南部,西北太平洋 4.2%的 TC 生成与其有关。有别于此,本研究客观分型结果强调强季风槽为其内部 TC 生成提供了强气旋性切变背景,故此处"季风涡旋"的概念用于描述由此而自组织发展出的低压涡旋特征,不对涡旋大小及生命史等作进一步的约束。





图 5 同图 4, 但为 MG 型的演变情况。 Figure 5 As in Figure 4 except for the MG pattern.

Ŕ

在季风槽或季风涡旋东南侧,可建立季风气流与偏东信风的辐合,辐合区的湿对流有利 于季风维持和 TC 生成,若热带波动(如东风波或 Rossby 波能量频散波列)被其捕获,则可 作为触发机制导致 TC 生成(Holland, 1995; Ritchie and Holland, 1999)。图 6 所示 SMT 型的合 成场体现了东风波在上述过程中的作用:在 TC 生成前 72h 至 48h 内(图 6 (a) - (b)), 有东风波西传汇入辐合区,辐合区正涡度增强,此后西风向东发展与偏东风形成气旋式涡 旋,相应有独立于季风槽之外的低压中心形成(图 6 (c)),进而发展为 TC。此外,成熟涡 旋移动过程中可产生 Rossby 波能量频散,在其东南侧激发出气旋与反气旋交替排列的波列 (Carr and Elsberry, 1995),因此,西北象限若存在季风涡旋(图 6 (b)显示有涡旋中心)或 先前生成的 TC,可激发 Rossby 波能量频散波列向东南侧传播,则也可触发辐合区 TC 生成。

需要说明的是,与 Rossby 波能量频散有关的 TC 生成,与辐合或切变主导的环流型在流场上高度相似,可视为它们的特例(Ritchie and Holland, 1999; Feng et al., 2014),需依赖物理过程进行判断,因此在机器学习过程中隐匿于其他环流型中无法被单独归类。

MC型与 SMT 型皆以西南季风或越赤道气流与偏东信风交汇而成的辐合区为典型特征, 但结合辐合区与副热带高压的相对位置而言,前者关注副热带高压南侧大范围辐合带中的 TC 生成(图 4),而后者强调强季风槽的作用(图 6),这与两型的盛行期相符,同时也是二 者区别命名的依据。此外,考虑到环流特征的相似性,MC型也存在热带波动在辐合区积聚 能量以触发 TC 生成的情况。



图 6 同图 4,但为 SMT 型的演变情况。黑色线表征东风波位置。 Figure 6 As in Figure 4 except for the SMT pattern. The black lines indicate the easterly wave.

与 MG 型类似, WMT 型(图 7) TC 在季风槽中生成, 但季风槽相对 MG 和 SMT 两型 向西北太平洋西侧收缩(图 2 (d)), 强度偏弱, 气旋式涡度中心位于季风槽东端, 由西南季 风和副热带高压西南侧的东南气流间的切变产生。在 TC 生成前 72h 内, 季风槽东南侧的气 流辐合使西南季风不断加强 (Holland, 1995), 气旋式切变北抬发展, 季风槽中相对涡度不断 增大。TC 生成的差异使这一型 TC 的平均生成位置较 SMT 型偏西偏北。





Figure 7 As in Figure 4 except for the WMT pattern.

东风波和大尺度环境场间的相互作用有利于 TC 生成,除前述作为启动机制激发辐合区 TC 生成外,东风波扰动在适当环境条件下可直接增幅发展为 TC (Li et al., 2003; Fu et al., 2006),这样的 TC 生成在偏东信风中,基本不受季风环流的影响。如图 8 所示,在 TC 生成 前 72h,生成位置周围始终是偏东气流,从生成前 48h 开始,TC 初始扰动的振幅不断加大,相对涡度增强,且表现出明显的西传倾向。



图 8 同图 4, 但为 EW 型的演变情况。黑色线表征东风波位置。

Figure 8 As in Figure 4 except for the EW pattern. The black lines indicate the easterly wave.

4.3 识别模型效果检验

对五类环流型特征的分析明确了 SOM 客观分型的天气学意义,这是建立识别模型的前提。本节将对前述 SVM、KNN 和 RF 三种模型的识别性能进行客观评估,以确定最优的机器 学习方法用于 TC 环流型的自动识别。

首先关注三种模型的准确率(如 3.2.2 节定义)。图 9(左)显示三种模型都表现出了较好的全局准确度,其中 SVM 的准确率高达 0.965,模型偏差最小,其次 KNN 和 RF 也有 0.90 左右的准确率。此外,进一步对比各模型在训练集和测试集上的准确率差异,图 9(右)为其在 100 次随机分样下的统计分布,结果显示,平均而言 SVM 和 KNN 中这一差异低于 0.05,而在 RF 中则相对较高,这表明对于同样的 TC 样本集,RF 往往会面临过拟合问题,从而产生较高的模型方差。总体而言,SVM 模型的泛化能力相对最强且更稳健。



图 9 SVM、KNN 及 RF 的准确率(左)及模型对训练集和测试集识别准确率之差的统计分布(100次随机分样;右)。小提琴图中显示有中位数(白色实线)、四分位距(黑色箱子)及临界值(黑色竖线上下端)。 Figure 9 Accuracy of SVM, KNN, RF model, respectively (left). Distribution of the accuracy difference between training set and test set in 100 iterations (right). Each violin plot shows the median value (the white solid centerline in the box), the interquartile range (the black box) and the upper and lower adjacent value (the top and bottom of the black vertical line).

图 10 为三种识别模型对各型 TC 的召回率和精确率。SVM 对五类环流型的识别都同时 表现出了大于 0.94 的高召回率和高精确率,这也反映出各型 TC 样本量的差异对模型性能的 影响极小,模型对各环流型的识别皆值得信赖。相较之下,KNN 和 RF 对样本不均衡有较高 的敏感性,其在各型上的召回率和精确率也因此表现出了明显的制约关系,其中以 KNN 表 现最甚,例如它对各型召回率的最值之差可达 0.16 (*R_{MC}* = 0.99,*R_{EW}* = 0.83); SVM 和 KNN 的这一差异或可归因于 SVM 对未知样本的识别仅取决于各类中的少量训练样本(即支 持向量),而 KNN 则一定程度上取决于所有训练样本的分布。此外,图示各指标的四分位距 皆以 SVM 最小,这表明 SVM 模型具有相对强的稳健性(Cotter and Polagye, 2020)。



图 10 SVM、KNN 及 RF 对五类大尺度环流型的(a) 召回率及(b) 精确率。直线和填色区分别表征 100 次 随机分样的平均值和四分位距。

Figure 10 (a) Recall rates and (b) precisions of SVM, KNN and RF for five large-scale patterns. The colored line and shaded region indicate the mean value and the interquartile range for 100 iterations,

respectively.

图 11 为样本量敏感性分析(如 3.2.3 节所述)的结果,即三种模型的准确率随训练样本量的变化。三种模型的准确率皆随训练样本量的扩充呈现增长趋势,并且 SVM 在各样本量下都具有相对高的准确率,明显优于 KNN 和 RF。此外,当仅从各类中抽取 23 例(20%)样本供以模型学习时,SVM 即可有超过 0.90 的准确率,70 例时已对应有 0.95 左右的准确率,即说明 SVM 模型在有限样本量下就可充分学习各型的环流场特征,进而达到较理想的识别效果。



图 11 SVM, KNN 和 RF 的准确率随训练样本量的变化。横坐标为训练集中各类 TC 的样本量,纵坐标的准确 率为 100 次随机分样的平均值。

Figure 11 Accuracy of SVM, KNN and RF with varying volumes of training set. Note that the x-axis represents the sampling amount of each pattern and the accuracy shown in y-axis is the mean value for

100 iterations.

5. 总结与讨论

本研究基于 1979-2020 年 6-11 月西北太平洋 TC 的相关数据,并选取 TC 生成位置周围 30°×30°范围内的 850hPa 水平风场,采用自组织映射网络(SOM)对 TC 生成前 24h 平均的 低层大尺度环流场进行聚类分型,在此基础上分别采用支持向量机(SVM)、k 近邻 (KNN)及随机森林(RF)构建识别模型,通过对已有样本的机器学习训练,客观评估不同 模型对新样本环流型的识别效果,以期将来业务应用可以选择最优模型对 TC 环流型进行自 动识别。

SOM 将有利于 TC 生成的低层大尺度环流场分为季风辐合型(MC)、季风涡旋型(MG)、强季风槽型(SMT)、弱季风槽型(WMT)以及东风波型(EW)五种类型。聚类数目的选择权衡了分型的独特性、稳健性和天气学可解释性,前两者分别由方差比(VR)和可复现性参数(RP)来量化。此外借由轮廓系数优化聚类结果,以给识别模型提供分型明确的训练样本。

结果表明,约 87%的 TC 生成与季风环流有关,其中 MC 型(27.5%) TC 多生成于 9-11 月,此时季风槽西退,西太平洋副热带高压西伸,TC 生成在副热带高压南部的低压辐合带 中;MG 型(23.6%)、SMT 型(19.7%)与 WMT 型(16.2%) TC 生成的季节变化明显受到 季风槽的影响,当季风槽强度较强时,其内部强切变产生的低压涡旋(季风涡旋)可加强发 展为 MG 型 TC,季风槽或涡旋东南侧的辐合区有利于生成 SMT 型 TC, Rossby 波能量频散 及东风波等热带波动或可为其提供触发机制,而 WMT 型 TC 生成则与相对较弱的季风槽有 关。此外,还有 13%的 EW 型 TC 生成在副热带高压西南侧的偏东信风中,由东风波扰动增 幅发展而成。

对三种识别模型的对比分析表明, SVM 对 TC 环流型的识别准确率可达 0.965, 模型泛 化能力优于 KNN 和 RF。此外, 样本不均衡问题往往会降低模型对少数类的识别效果, 但 SVM 在各型上都有超过 0.94 的高召回率和高精确率, 对此问题不敏感且表现出了较高的稳 健性。对样本量的敏感性分析表明, SVM 在各训练样本量下都有高于 KNN 和 RF 的准确 率, 且仅依赖少量 TC 样本就可达到较好的学习效果。综上研究, SVM 可作为一种较理想的 客观方法对 TC 生成前期的低层大尺度环流型进行识别。

本文用机器学习方法对有利于 TC 生成的大尺度环流场进行分型与识别,并对比各型异 同给出合理的天气学分析。但尚未进一步考察各型的其它环境因子分布特征及其对 TC 生成 的影响。此外,天气尺度扰动在各型 TC 生成中的作用及其所占比例也有待进一步统计分 析。

参考文献:

- Bao M, Wallace J M. 2015. Cluster Analysis of Northern Hemisphere Wintertime 500-hPa Flow Regimes during 1920-2014 [J]. J. Atmos. Sci., 72(9): 3597-3608. doi: 10.1175/JAS-D-15-0001.1
- Bhowmick R, Trepanier J C, Haberlie A M. 2021. Southwest Pacific tropical cyclone development classification utilizing machine learning and synoptic composites [J]. International Journal of Climatology, 1-27. doi: 10.1002/joc.7457
- Breiman L. 2001. Random Forests [J]. Machine Learning, 45: 5-32. doi: 10.1023/A:1010933404324
- Briegel L M, Frank W M. 1997. Large-Scale Influences on Tropical Cyclogenesis in the Western North Pacific [J]. Mon. Wea. Rev., 125(7): 1397-1413. doi: 10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2
- Burges C. 1998. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 2(2): 121-167. doi: 10.1023/A:1009715923555
- Carr L E III, Elsberry R L. 1995. Monsoonal Interactions Leading to Sudden Tropical Cyclone Track Changes [J]. Mon. Wea. Rev., 123(2): 265-290. doi: 10.1175/1520-0493(1995)123<0265:MILTST>2.0.CO;2
- Chen T C, Wang S Y, Yen M C, et al. 2004. Role of the Monsoon Gyre in the Interannual Variation of Tropical Cyclone Formation over the Western North Pacific [J]. Wea. Forecasting, 19(4): 776-785. doi: 10.1175/1520-0434(2004)019<0776:ROTMGI>2.0.CO;2
- Cheung K K W. 2004. Large-Scale Environmental Parameters Associated with Tropical Cyclone Formations in the Western North Pacific [J]. J. Climate, 17(3): 466-484. doi: 10.1175/1520-0442(2004)017<0466:LEPAWT>2.0.CO;2
- Cortes C, Vapnik V. 1995. Support-Vector Networks [J]. Machine Learning, 20(3): 273-297. doi: 10.1007/BF00994018
- Cotter E, Polagye B. 2020. Automatic Classification of Biological Targets in a Tidal Channel Using a Multibeam Sonar [J]. J. Atmos. Oceanic Technol., 37(8): 1437-1455. doi: 10.1175/JTECH-D-19-0222.1

- Cover T M, Hart P E. 1967. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 13(1): 21-27. doi: 10.1109/TIT.1967.1053964
- 冯涛, 黄荣辉, 陈光华, 等. 2013. 近年来关于西北太平洋热带气旋和台风活动的气候学研究进展 [J]. 大气科学, 37(2): 364–382. doi:10.3878/j.issn.1006-9895.2012.12307. Feng Tao, Huang Ronghui, Chen Guanghua, et al. 2013. Progress in recent climatological research on tropical cyclone activity over the western North Pacific [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 37(2): 364–382. doi:10.3878/j.issn.1006-9895.2012.12307.
- Feng T, Chen G H, Huang R H, et al. 2014. Large-scale circulation patterns favourable to tropical cyclogenesis over the western North Pacific and associated barotropic energy conversions [J]. International Journal of Climatology, 34(1): 216-227. doi: 10.1002/joc.3680
- Fu B, Li T, Peng M S, et al. 2006. Analysis of Tropical Cyclogenesis in the Western North Pacific for 2000 and 2001 [J]. Wea. Forecasting, 22(4): 763-780. doi: 10.1175/WAF1013.1
- Gray W M. 1968. Global View of the Origin of Tropical Disturbances and Storms [J]. Mon. Wea. Rev., 96(10): 669-700. doi: 10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2
- Hall T J, Thessin R N, Bloy G J, et al. 2010. Analog Sky Condition Forecasting Based on a k-nn Algorithm [J]. Wea. Forecasting, 25(5): 1463-1478. doi: 10.1175/2010WAF2222372.1
- Hersbach H, Bell B, Berrisford P, et al. 2020. The ERA5 global reanalysis [J]. Quart. J. Roy. Meteor. Soc. doi: 10.1002/qj.3803
- Holland G J. 1995. Scale interaction in the Western Pacific Monsoon [J]. Meteor. Atmos. Phys., 56(1):57-79. doi: 10.1007/BF01022521
- 黄荣辉, 皇甫静亮, 武亮, 等. 2016. 关于西北太平洋季风槽年际和年代际变异及其对热带气旋 生成影响和机理的研究 [J]. 热带气象学报, 32(06):767-785. doi: 10.16032/j.issn.1004-4965.2016.06.001. Huang Ronghui, Huangpu Jingliang, Wu Liang, et al. 2016. Research on the Interannual and Interdecadal Variabilities of the Monsoon Trough and Their Impacts on Tropical Cyclone Genesis over the Western North Pacific [J]. Journal of Tropical Meteorology (in Chinese), 32(06):767-785. doi: 10.16032/j.issn.1004-4965.2016.06.001.
- Jergensen G E, Mcgovern A, Lagerquist R, et al. 2020. Classifying Convective Storms Using Machine Learning [J]. Wea. Forecasting, 35(2): 537-559. doi: 10.1175/WAF-D-19-0170.1
- Knapp K R, Kruk M C, Levinson D H, et al. 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data [J]. Bull. Amer. Meteor. Soc., 91(3): 363–376. doi: 10.1175/2009BAMS2755.1
- Knerr S, Personnaz L, Dreyfus G, et al. 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network [J]. Neurocomputing: Algorithms, Architectures and Applications, 41-50. doi: 10.1007/978-3-642-76153-9_5
- Kohonen T. 1982. Self-organized formation of topologically correct feature maps [J]. Biol. Cybern., 43(1): 59-69. doi: 10.1007/BF00337288
- Lander M A. 1994. Description of a Monsoon Gyre and Its Effects on the Tropical Cyclones in the Western North Pacific during August 1991 [J]. Wea. Forecasting, 9(4): 640-654. doi: 10.1175/1520-0434(1994)009<0640:DOAMGA>2.0.CO;2
- 李艳, 符彩芳, 金茹. 2019. 西北太平洋近赤道热带气旋生成的特征分析[J]. 大气科学学报, 42(05):695-704. doi: 10.13878/j.cnki.dqkxxb.20170427001. Li Yan, Fu Caifang, Jin Ru. 2019. Analysis of near-equatorial tropical cyclone genesis in the western North Pacific [J]. Journal of Nanjing Institute of Meteorology (in Chinese), 42(05):695-704. doi: 10.13878/j.cnki.dqkxxb.20170427001.

- Li T, Fu B, Ge X, et al. 2003. Satellite data analysis and numerical simulation of tropical cyclone formation [J]. Geophys. Res. Lett., 30(21): 1-4. doi: 10.1029/2003GL018556
- Neetu S, Lengaigne M, Vialard J, et al. 2020. Quantifying the Benefits of Nonlinear Methods for Global Statistical Hindcasts of Tropical Cyclones Intensity [J]. Wea. Forecasting, 35(3): 807-820. doi: 10.1175/WAF-D-19-0163.1
- Ritchie E A, Holland G J. 1999. Large-Scale Patterns Associated with Tropical Cyclogenesis in the Western Pacific [J]. Mon. Wea. Rev., 127(9): 2027-2043. doi: 10.1175/1520-0493(1999)127<2027:LSPAWT>2.0.CO;2
- Rousseeuw P J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis [J]. Journal of Computational and Applied Mathematics, 20: 53-65. doi: 10.1016/0377-0427(87)90125-7
- Rozas-Larraondo P, Inza I, Lozano J A. 2014. A Method for Wind Speed Forecasting in Airports Based on Nonparametric Regression [J]. Wea. Forecasting, 29(6): 1332-1342. doi: https://doi.org/10.1175/WAF-D-14-00006.1
- 王蔚, 朱伟军, 端义宏, 等. 2008. 大尺度背景下西北太平洋热带气旋的统计分析 [J]. 南京气象学 院学报, 31(02): 277-286. doi: 10.13878/j.cnki.dqkxxb.2008.02.016. Wang Wei, Zhu Weijun, Duan Yihong, et al. Statistical Analysis of Large-Scale Backgrounds Classification Associated with Tropical Cyclone Formations in the Western North Pacific [J]. Journal of Nanjing Institute of Meteorology, 31(02): 277-286. doi: 10.13878/j.cnki.dqkxxb.2008.02.016.
- 杨琼琼, 吴立广. 2015. 中国 24h 台风路径预报难点及其大尺度环流分析 [J]. 气象科学, 35(04):389-397. Yang Qiongqiong, Wu Liguang. 2015. doi: 10.3969/2014jms.0029. Analysis on difficult of 24 h typhoon track forecast in China and associated large-scale circulation [J]. Journal of the Meteorological Sciences, 35(04):389-397. doi: 10.3969/2014jms.0029.
- Yoshida R, Ishikawa H. 2013. Environmental Factors Contributing to Tropical Cyclone Genesis over the Western North Pacific [J]. Mon. Wea. Rev., 141(2): 451-467. doi: 10.1175/MWR-D-11-00309.1
- Zehr R M. 1992. Tropical Cyclogenesis in the Western North Pacific [R]. NOAA Tech. Rep. NESDIS 61.
- 张翔, 武亮, 皇甫静亮, 等. 2017. 西北太平洋季风槽的季节和年际变化特征及其与热带气旋生成大尺度环境因子的联系 [J]. 气候与环境研究, 22(04): 418-434. doi: 10.3878/j.issn.1006-9585.2016.16065. Zhang Xiang, Wu Liang, Huangpu Jingliang, et al. 2017. Seasonal and Interannual Variability of the Western North Pacific Monsoon Trough and Its Relationship to Large-Scale Environment Factors [J]. Climatic Environ. Res. (in Chinese), 22(04): 418-434. doi: 10.3878/j.issn.1006-9585.2016.16065.
- Zhang R, Liu Q, Hang R, et al. 2022. Predicting Tropical Cyclogenesis Using a Deep Learning Method From Gridded Satellite and ERA5 Reanalysis Data in the Western North Pacific Basin [J]. IEEE Trans. Geosci. Remote Sens., 60: 1-10. doi: 10.1109/TGRS.2021.3069217
- 郑颖青, 余锦华, 吴启树, 等. 2013. K-均值聚类法用于西北太平洋热带气旋路径分类 [J]. 热带气 象学报, 29(04): 607-615. Zheng Yingqing, Yu Jinhua, Wu Qishu, et al. 2013. K-means clustering method for classification of the Northwestern Pacific tropical cyclone tracks [J]. Journal of Tropical Meteorology (in Chinese), 29(04): 607-615.
- 周璞, 江志红. 2016. 自组织映射神经网络(SOM)降尺度方法对江淮流域逐日降水量的模拟评估 [J]. 气候与环境研究, 21(05):512-524. doi:10.3878/j.issn.1006-9585.2016.16097. Zhou Pu, Jiang Zhihong. 2016. Simulation and evaluation of statistical downscaling of regional daily

precipitation over Yangtze–Huaihe River basin based on self-organizing maps [J]. Climatic and Environmental Research (in Chinese), 21(5): 512–524, doi:10.3878/j.issn.1006-9585.2016.16097.