

光学学报

江淮梅雨降水云顶光谱信息与降水强度关系研究

范荣峰, 赵文川, 李欣艳, 陈绎冰, 李天依, 赵幽, 刘超, 杨元建*

南京信息工程大学大气物理学院, 江苏 南京 210044

摘要 云顶的光学和微观物理特征是基于卫星光谱定量反演降水强度的关键信息。基于 1998—2007 年江淮梅雨期内热带测雨卫星的光谱和雷达融合观测资料, 建立了基于降水云顶光谱信息反演降水强度的随机森林算法模型, 进一步探究了梅雨降水云顶微观物理特征与降水强度变化的关系。研究表明, 在随机森林模型反演降水强度的测试集中, 预测降水强度与观测降水强度的相关系数 R 为 0.67, 均方根误差为 4.06 mm/h, 测试集具有较高的降水预测精度。随机森林模型中的云水含量 (CWP, 即云水质量浓度) 在所有输入的云微物理变量重要性排序中位于前列。进一步分析表明: 当 CWP 小于 $1.0 \text{ kg} \cdot \text{m}^{-3}$ 时, 江淮梅雨期降水主要以小雨等级为主, 而当 CWP 大于 $1.5 \text{ kg} \cdot \text{m}^{-3}$ 时, 大雨和暴雨等级的降水概率明显增大; 降水的云粒子有效半径 (CER) 大小主要位于 $10 \mu\text{m}$ 以上, 且降水概率总体上随着 CER 的增大而单调递增; 各等级降水概率随着云光学厚度 (COT) 的增大而不断增大, 当 COT 大于 120 时, 各级降水的概率显著增强, 尤其是强降水。

关键词 大气光学; 卫星光谱; 反演降水; 随机森林; 云微物理变量

中图分类号 O436

文献标志码 A

doi: 10.3788/AOS202141.2201001

Relationship Between Cloud-Top Spectral Information and Precipitation Intensity of Meiyu Precipitation over the Yangtze-Huaihe Valley

Fan Rongfeng, Zhao Wenchuan, Li Xinyan, Chen Yibing, Li Tianyi,
Zhao You, Liu Chao, Yang Yuanjian*

*School of Atmospheric Physics, Nanjing University of Information Science & Technology,
Nanjing, Jiangsu 210044, China*

Abstract The optical and microphysical features of the cloud top are key information for quantitative precipitation retrieval based on satellite spectra. With the spectral and radar observation data of the Tropical Rainfall Measuring Mission satellite collected during the Meiyu periods of the Yangtze-Huaihe Valley from 1998 to 2007, this paper built a random forest algorithm model that used the cloud-top spectral information of precipitating clouds to retrieve precipitation intensity. It also investigated the relationship between cloud-top microphysical characteristics and the intensity variation of Meiyu precipitation. The results show that in the precipitation retrieval testing set of the random forest model, the correlation coefficient R between predicted precipitation intensity and observed precipitation intensity is 0.67, and the root-mean-square error is 4.06 mm/h. This means the random forest model has high precipitation prediction accuracy. In the model, the cloud water path (CWP) ranks high in the importance hierarchy of all input variables. Further analysis shows that when the CWP is less than $1.0 \text{ kg} \cdot \text{m}^{-3}$, precipitation at light rain level is dominant during the Meiyu period of the Yangtze-Huaihe Valley, while when the CWP is more than $1.5 \text{ kg} \cdot \text{m}^{-3}$, the probability of precipitation at heavy rain and torrential rain levels increases significantly. On the whole, the probability of precipitation at all levels increases monotonically with the increase of cloud effective radius (CER), which is mainly over $10 \mu\text{m}$. It also increases with the increase of cloud optical thickness (COT).

收稿日期: 2021-04-06; 修回日期: 2021-05-03; 录用日期: 2021-06-03

基金项目: 国家重点研发计划(2018YFC1507402)、国家自然科学基金(42061134009)、南京信息工程大学大学生创新创业训练计划项目(202010300057Y)

通信作者: *yyj1985@nuist.edu.cn

When COT is more than 120, it increases significantly, especially the probability of heavy precipitation.

Key words atmospheric optics; satellite spectrum; precipitation retrieval; random forest; cloud microphysical parameter

OCIS codes 010.0280; 000.5490; 010.1615

1 引言

随着卫星遥感技术的不断发展,可见光/红外光谱扫描仪已成为目前最为广泛的星载探测仪器,尤其是静止卫星上的可见光/红外仪器,他们可以提供大范围高时空分辨率的探测资料。利用卫星观测的可见光/红外信号进行降水云和非降水云的辨别,并在此基础上利用统计关系来估算地表降水率,一直都是卫星遥感降水研究中的重要课题^[1-4]。

国内外关于卫星云顶光谱参数与降水关系的研究已经取得很多重要进展。例如, Rosenfeld and Gutman^[5]研究了 NOAA 卫星反演的云粒子有效半径(CER)与降水的关系,提出有效半径大于 14 μm 是云中产生降水的阈值。此外,研究表明,不同的云微物理参数和地面降水强度密切相关且有显著性差异,例如云光学厚度、云顶高度、云液态含水量、云粒子有效半径与地面降水强度基本呈正相关关系^[2,6-9],云顶温度则与地面降水强度呈反相关关系^[2]。特别地,云光学厚度对地面降水概率和降水强度的影响最为显著^[8-9],稳定少变的大光学厚度云或云光学厚度显著增大均可带来强的地面降水^[7]。随着云光学厚度的增加,降水概率呈增加趋势。当云光学厚度大于 20 时,降水概率显著增大。云顶温度和云顶高度对降水强度的影响次之。当云顶温度小于 $-15\text{ }^{\circ}\text{C}$ 时,随着云顶温度的降低,降水强度和概率都随之增大;强降水过程中降水强度高值区内云相为冰云,冰云云顶高度在 8~17 km ^[10]。而云粒子有效半径的增大也将引起降水强度和降水概率的增大。可见,上述结果表明,利用卫星遥感云顶光谱参数信息的统计在反演降水方面具有重要的科学意义和应用价值。

然而,对于静止卫星等多光谱卫星,在降水监测方面面临的一个重要问题就是很难时空同步地获取光谱信号和降水信息。以往的研究多以地面降水观测资料来验证由卫星光谱信号估算的降水信息,特别地,由于地面降水监测站点在空间上稀疏,这对卫星光谱遥感反演降水带来了很大的不确定性和挑战,故不能准确地发展由卫星光谱估算降水的方法。1997 年底美日共同研制的热带测雨卫星(Tropical Rainfall Measuring Mission,简称 TRMM)发射升

空,利用其搭载的测雨雷达(Precipitation Radar,简称 PR)、微波成像仪(TRMM Microwave Imager,简称 TMI)和可见光/红外辐射计(Visible and Infrared Sensor,简称 VIRS)等仪器,为研究热带、副热带地区的降水云光谱信号特征及其与降水的关系提供了新的契机^[4,11-12]。由于 TRMM 上的 PR 和 VIRS 的测量具有类似的跨轨扫描模式,在观测同一目标时几乎可以立即进行同步,因此将这两者结合使用是可行的,能够用于精确地探索降水云的光谱特征。同时,由于 VIRS 传感器可以装载到对地静止卫星上且具有较高的采样频率,因此它将有效地提高和扩展使用可见/红外扫描仪测量识别云微物理特性的能力。例如,傅云飞等^[11]通过融合 PR 和 VIRS 10 年间的探测资料,利用 PR 对降水云的直接识别,分析了夏季热带和副热带地区对流、层状降水云光谱信号的气候分布特征。傅云飞等^[4]开展了利用 TRMM 的 PR 和 VIRS 的融合资料,基于双光谱反射率算法反演获得降水云参数和降水强度信息,并使其应用于气溶胶-云-降水相互作用以及降水云监测等方面的研究。Yang 等^[13]利用 TRMM 卫星 15 年的同步光谱和雷达观测资料,对降水云的光学和微物理特征进行了统计,并分析了与江淮流域夏季地面降水率的关系。可见,TRMM 上 PR 和 VIRS 的时空同步观测降水与云光谱信号的能力优势明显且应用广泛。

江淮梅雨是东亚夏季风从华南向北推进后形成的季风降水雨带^[14]。作为发生在我国东部的特征性天气气候现象,梅雨期降水是江淮流域汛期降水的重要组成部分^[15]。研究表明,江淮流域是我国夏季旱涝灾害最频发的地区^[16],特别是梅雨期引发的洪涝灾害,常给人们带来巨大经济损失,甚至导致严重的人员伤亡。因此,梅雨期降水(尤其暴雨)的监测与预报一直是广大气象、水文工作者和防汛调度人员十分关注的问题。可见,进行梅雨期降水监测与预报方面的研究具有重要的科学意义和社会价值。

鉴于此,本文将将以 1998—2007 年江淮流域梅雨期降水为研究对象,利用长达 10 年的 TRMM PR 和 VIRS 的综合探测资料,通过随机森林的机器学习训练,建立卫星光谱和云参数与降水的关系模型,

进行降水的反演和预测,并试图揭示江淮流域梅雨期降水云的多光谱信号时空分布特征及其与地表降水率之间的关系,旨在进一步提升江淮梅雨期降水的多光谱卫星监测技术能力,为该地区的防灾减灾提供科学依据。

2 研究数据与方法

2.1 江淮梅雨概况

江淮地区位于淮河以南,长江下游地区,地理坐标为 $110^{\circ}\text{E}\sim 112^{\circ}\text{E}$, $28^{\circ}\text{N}\sim 35^{\circ}\text{N}$,主要包括苏、豫南和皖中等地区,分布着众多的低山、丘陵和岗地,是我国东部地区南北和东西的过渡地带。每年 6 月中旬,由于西太平洋副热带高压由南向北推进,季风雨带到达江淮地区,江淮地区进入梅雨期。梅雨期内雨量充沛,相对湿度很大,日照时间短。其中,降水多为连续性降水,但也常伴有阵雨和雷雨,有时可达到暴雨程度。基于之前的研究结果,将江淮地区每年的梅雨期定义为符合广义梅雨准则降水特征的时间^[17],表 1 显示了 1998—2007 年间的梅雨期。

表 1 1998—2007 年江淮地区的梅雨期

Table 1 Meiyu periods over the Yangtze-Huaihe river basin from 1998 to 2007

Year	Meiyu period (dd/mm)
1998	23/06-04/07, 16/07-03/08
1999	08/06-17/06, 22/06-01/07, 06/07-18/07
2000	24/05-04/06, 20/06-30/06
2001	17/06-25/06
2002	19/06-28/06
2003	20/06-11/07, 17/07-22/07
2004	14/06-27/06, 10/07-20/07
2005	05/07-13/07
2006	22/06-11/07
2007	19/06-26/07

2.2 数据介绍

TRMM 卫星轨道为近赤道非太阳同步轨道,轨道高度为 350 km,倾角为 35° ,周期为 96 min。其覆盖范围以赤道为中心,在南北纬 38° 之间。由于 TRMM 卫星 35° 的轨道倾角远小于其他极轨卫星,这使得 TRMM 卫星具有较高的时间分辨率。TRMM 卫星每天可以覆盖全球 1~3 次,其搭载的传感器包括 PR、TMI、VIRS、云和地球辐射能量测量系统(CERES)和闪电成像感应器(LIS)。VIRS 是一个五通道成像分光辐射计,350 km 轨道宽度和 $\pm 45^{\circ}$ 的扫描角度决定了其 720 km 的扫描宽度(2001 年 8 月 7 日后约为 833 km)。星下点瞬时视

场为 2.11 km (2001 年 8 月 7 日后约为 2.4 km), 45° 扫描角度上的瞬时视场增加到 3.02 km。VIRS 有五个波段,一个可见光波段($0.6\ \mu\text{m}$)、一个近红外($1.6\ \mu\text{m}$)、一个中波红外($3.7\ \mu\text{m}$)和两个热红外($10.8\ \mu\text{m}$ 和 $12\ \mu\text{m}$)。

由于 PR 和 VIRS 在同一时刻的探测目标不同,且两者具有不同的分辨率,因此需要对 1B01 和 2A25 资料进行融合处理^[18-19],即将 2A25 水平分辨率内的 1B01 像素进行距离加权平均。在融合资料的基础上,利用 PR 对降水云进行识别。值得注意的是,PR 的探测灵敏度约为 16~18 dBz,对应的降水率约为 0.5 mm/h,即对毛毛雨等弱降水的识别能力有限。因此,本文将采用 0.5 mm/h 的 PR 地表降水率作为阈值来识别降水云。经过上述处理,能够获得 1998—2007 年梅雨期的轨道级降水云垂直廓线对应的三维降水率和云参数融合数据集。

2.3 研究方法

随机森林是 2001 年由 Leo Breiman 和 Culter Adele 开发的一种数据挖掘方法,是一种现代分类与回归的机器学习技术,同时也是一种组合式的自学习技术^[20]。随机森林的基本思想为:将 bagging 算法和 random subspace 算法结合,生产多棵互不相关的决策树,每棵决策树都能够独立地利用训练样本进行训练,训练完成后,对这些决策树采取投票选取机制,确定最终分类结果^[21-23]。其优越性体现在同等运算率下的高预测精度,以及相较于传统的统计方法,对非线性数据具有更好的拟合效果。此外,随机森林能够进行变量重要性分析,对比神经网络和支持向量机等其他方法,其在分析变量关系上存在优势^[23]。本研究主要利用随机森林的机器学习来训练和建立卫星光谱和云参数与降水的关系模型,并进行降水的反演和预测。随机森林模型构建步骤如下:

1) 确定本次研究衡量随机森林模型准确度的指标为相关系数(R)和均方根误差(RMSE, E_{RMS})。其中, R 越大模型估算效果越好, RMSE 越小模型估算效果越好。

2) 根据已有的研究^[17,23],将降水强度作为因变量,并对其进行自然对数化处理,将其他 16 个参数信息作为输入自变量,具体输入自变量的物理解释见表 2。

3) 采用十折交叉验证的方法。将建模所用数据集 D 划分为 10 个大小相似的互斥子集。每个子集 D_i 都尽可能保持数据分布的一致性,即从 D 中

通过分层采样得到。然后每次用 9 个子集的并集作为训练集,余下的子集作为测试集,这样就可获得 10 组训练/测试集,从而可进行 10 次训练和测试,最终返回的是这 10 个测试结果的均值。十折交叉验证的方法最大限度地利用了现有的样本数据,保证了每个样本都分别用作训练样本和测试样本,这在一定程度上避免了过度拟合的结果。同时,为了进一步验证所训练的模型,本研究从数据集中预留了未参与建模过程的 9121 轨道数据并将其作为全新的测试数据,用于评估此模型的泛化能力^[24]。

4) 随机模型的建立需要确定两个重要参数,分别为树节点的输入变量数 M_{try} 和决策树的棵数 N_{tree} 。其中, M_{try} 越大,随机森林算法的过拟合效应就越小, N_{tree} 越大,子模型间的差异性就越小。 M_{try}

的数值应小于输入的变量参数,一般情况下, M_{try} 的取值为 $1, k/2, \sqrt{k}$ 和 $\ln k + 1$ 等, k 为模型输入的变量个数。本文选取 $k/2$ 作为树节点的输入变量数,即 $M_{\text{try}} = 8$ 。决策树的棵数 N_{tree} 的取值在 $500 \sim 800$ 时为比较理想的分类,本文将其设置为 550。

5) 建模。本次研究通过 R 语言中的 Random Forest 数据包构建随机森林模型。该数据包的估算实质如下:随机产生 M_{try} 个变量用于节点上的二叉树,二叉树变量的选择仍然满足节点不纯度最小原则;应用 Bootstrap 自助法在原数据集中有放回地随机抽取 N_{tree} 个样本集,组成 N_{tree} 棵决策树,而对于未被抽取的样本,将其用于单棵决策树的预测;根据 N_{tree} 棵决策树组成的随机森林对待分类样本进行分类或预测,分类的原则是投票法,预测的原则是简单平均。

表 2 模型输入预测因子表

Table 2 Model input predictive factor table

Number	Variable	Physical meaning	Number	Variable	Physical meaning
1	RR / (mm · h ⁻¹)	Rain rate	10	SRF12	RF1 minus RF2
2	RT	Precipitation type	11	STB34	TB3 minus TB4
3	RF1	Reflectivity of 0.63 μm channel	12	STB45	TB4 minus TB5
4	RF2	Reflectivity of 1.6 μm channel	13	STB35	TB3 minus TB4
5	TB3 /K	Black body temperature of 3.7 μm channel	14	DRF12	RF1 divide RF2
6	TB4 /K	Black body temperature of 10.8 μm channel	15	DTB34	TB3 divide TB4
7	TB5 /K	Black body temperature of 12 μm channel	16	DTB45	TB4 divide TB5
8	CER /μm	Cloud effective radius	17	DTB35	TB3 divide TB5
9	CWP / (kg · m ⁻³)	Cloud water path			

为了方便统计和绘图,本研究根据 10 年梅雨期数据,对数据集作格点化处理,将地图空间分为 0.5° 网格,统计落在小网格内的数据个数,并计算得到其平均数,将其用作这个小网格的代表数据,最终生成空间分辨率为 0.5° 的格点化数据,以判断梅雨期的降水基本态(见图 1)。图 1(a)显示了梅雨期内的降水样本数,可以发现,降水样本数由北向南呈现递减趋势,高样本数(>2100)集中在江苏东北面的海面上。每个小网格内的数据样本数都在 500 以上,这说明根据格点化数据分析梅雨期降水基本气候态是可靠的。

3 结果与分析

3.1 地表降水强度和降水云顶光谱信息的气候态空间分布特征

降水信息(降水率、样本数等)的水平空间分布在一定程度上可以反映出降水云系的性质和状态。图 1(b)~(i)分别显示了梅雨期内地表降水率、光谱

反射率、亮温和云水含量(CER,COT,CWP)的 0.5° 网格平均分布。如图 1(b)所示,江淮大部分地区地表平均降水率为 $3 \sim 4$ mm/h,而在豫南、皖中以及苏东等地区最大可达到 6 mm/h 以上(将这三个地区统计为降水大值区)。在降水大值区内,图 1(d)~(f)恰巧为小值区,图 1(g)和图 1(i)为大值区,两者有着明显的对应关系。

$0.63 \mu\text{m}$ 通道是水汽非吸收通道,其大小主要由云的光学厚度和云滴粒子的大小决定,云光学厚度越大,云滴粒子越小,光谱反射率值就越大。因此,其光谱反射率值(RF1)与 COT 有着十分相似的空间水平分布[见图 1(c)和图 1(h)]。 $1.6 \mu\text{m}$ 通道是一个重要的水汽吸收通道[见图 1(d)],受云滴粒子大小的影响非常明显。由于液相和冰相的水汽辐射差异,该通道可以用来区分云的相态。大的冰云粒子会导致小的光谱反射率值(RF2),与此相对应的是,降水大值区内的 TB4 值均低于 230 K[见图 1(f)],且在图 1(g)中,云粒子有效半径均大于

20 μm 。3.7 μm 通道也是一个水汽吸收通道,其光谱反射值大小受云滴粒子大小和相态的影响,大的冰相云滴粒子会导致小的 TB3 值[见图 1(e)]。对比图 1(e)和图 1(f)可发现,江淮地区 TB3 的水平空间分布和 TB4 相似,并且在同一

区域 TB3 的值要比 TB4 高 25 K 左右。10.8 μm 通道为热红外窗区通道,受大气中的水汽和其他气体的影响极小,所以该通道的亮温值(TB4)最接近于云顶最真实的亮温。通常情况下,云顶高度越高, TB4 的值越小。

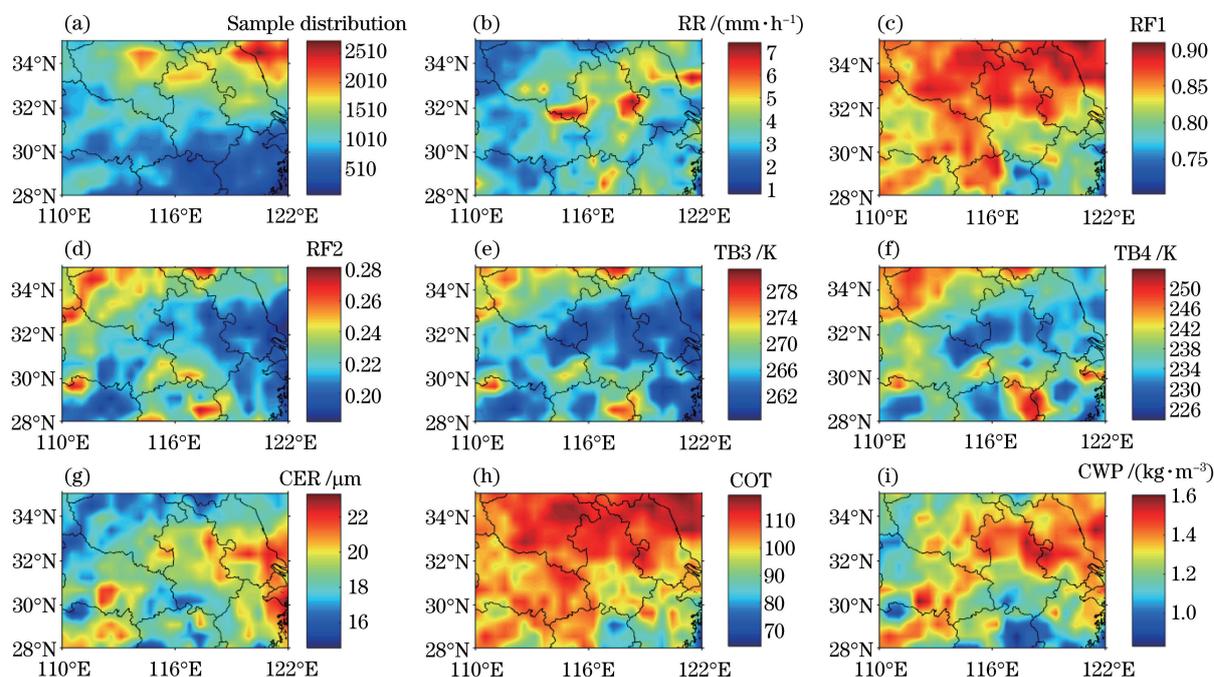


图 1 江淮地区梅雨期内降水样本数、地表降水率、云微物理和光谱参数的 0.5° 网格平均分布。(a)降水样本数;(b)地表降水率;(c) 0.63 μm 通道光谱反射率;(d) 1.6 μm 通道光谱反射率;(e) 3.7 μm 通道云顶亮温;(f) 10.8 μm 通道云顶亮温;(g)云粒子有效半径;(h)云光学厚度;(i)云水含量

Fig. 1 0.5° grid average distribution of precipitation samples, average surface rain rate, cloud microphysics and spectral parameters during Meiyu period over the Yangtze-Huaihe River Valley. (a) Precipitation sample number; (b) rain rate; (c) reflectivity of 0.63 μm channel; (d) reflectivity of 1.6 μm channel; (e) black body temperature of 3.7 μm channel; (f) black body temperature of 10.8 μm channel; (g) cloud effective radius; (h) cloud optical thickness; (i) cloud water path

云粒子有效半径(CER)代表了云滴粒子的大小,云水含量体现着云中含水量的多少。云粒子有效半径越大,云水含量越高,云产生的降水粒子越多、越大,地表降水率越大的可能性也就越高。综上所述,越大的 CER 与 CWP 和越低的 RF2、TB3 与 TB4 对应着越高的地表平均降水率,这说明梅雨期内的强降水主要由云顶高度高、云水含量多、云滴粒子大的降水云产生。因此,降水云顶光谱信息的气候态空间分布决定了地表降水强度的气候态分布。

3.2 基于云顶光谱信息和微物理参数预测降水强度的随机森林模型

图 2(a)显示了随机森林方法反演的训练集降水强度与实际观测降水强度的散点图,相关系数 R 为 0.94, RMSE 为 2.33 mm/h, 反演精度很好。图 2(b)显示了随机森林方法反演的测试集降水强

度和实际观测降水强度的散点图,相关系数 R 为 0.67, RMSE 为 4.06 mm/h, 反演精度较好,随机森林对于地表降水强度的反演有较好的适用性。综合图 2(a)和图 2(b)可发现,两图中样本大部分聚集在 1:1 线周围,有较高的拟合度。以 40 mm/h 为界限,降水强度低于 40 mm/h 的样本比高于 40 mm/h 的样本更贴近 1:1 线,这表明随机森林在降水强度较低时反演精度更好。同时,该模型降水强度小于 5 mm/h 的数据量密度最大,可以预期该模型在预测弱降水情况时将具有更高的准确性。

此外,可以发现本模型测试集的相关系数相较于训练集有较为显著的下降,其主要原因:降水是个非常复杂的物理过程,降水强度与很多的光谱参数及云微物理参数相关。此次研究的模型中输入的预测变量较少,且训练样本数据较少。尽管采用了十

折交叉验证方法,增加了样本数据量,但是总体上样本量仍然较少,且存在一定的样本数据过度利用的情况和一定的过拟合现象,这使得测试集的精度降低。此外,模型训练样本的因变量降水率

的大小主要集中在小于 5 mm/h 的范围,因变量样本的差异较小且集中,故对于大于 5 mm/h 的降水情况的预测能力较弱,由此导致测试集的精度较低。

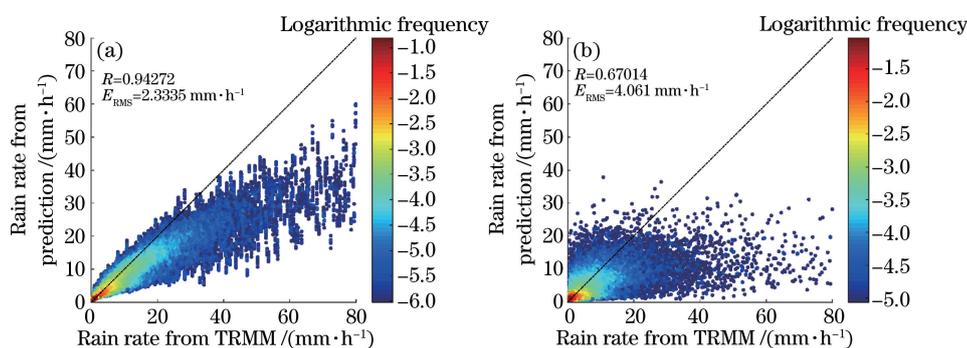


图 2 随机森林模型的训练集和测试集降水强度预测与实测的散点密度分布图。(a)训练集;(b)测试集

Fig. 2 Scatter of predicted rain rate of random forest model and observed rain rate for training set and testing set.

(a) Training set; (b) testing set

R 语言提供的重要性函数可直接对变量重要性进行分析,其主要评价指标为精度平均减少值 IncMSE 和节点不纯度平均减少值 IncNode Purity。IncMSE 指将该变量随机取值后随机森林模型估算误差相对于原来误差的升高幅度。IncMSE 值越大,表明该变量越重要。IncNode Purity 是指该变量对各个决策树节点的影响程度。IncNode Purity 值越大,表明该变量越重要^[23,25]。表 3 将降水强度随机森林模型的变量按重要性排序,从表 3 可以看出,在云顶光谱-微物理信息中,对反演精度影响较大的前 5 个输入预测变量,分别是 RT, TB3, STB45, DTB45 和 CWP。

RT 为降水类型,由 TRMM 卫星上搭载的 PR 降水雷达探测,并分为对流降水、层云降水和“其他”类型降水三类。若 PR 回波出现零度层亮带,则该降水廓线被定义为层云降水廓线;如果 PR 回波无亮带,但回波中出现超过 39 dBz 的信号,则该降水廓线被定义为对流降水廓线;非上述两种情形的降水廓线被定义为“其他”类型降水廓线。降水类型由零度层亮带、降水雷达回波强度等因素确定,零度层亮带内的粒子融化过程中的碰并加强与破碎程度减弱,会影响降水效率和地面降水强度^[26],雷达回波强度则是对降水强度的直接反映,且降水类型是对降水特性的分类,因此降水类型对模型的影响最大,占比高达 42.92%。

TB3 为 3.7 μm 通道云顶亮温,被广泛用于反演云粒子有效半径^[17],而云粒子有效半径是影响降水强度大小的重要因子,因此 TB3 也成为了仅次于

降水类型且影响降水强度预测的重要因子。

表 3 随机森林变量重要性

Table 3 Importance of forests random variables

Parameter	IncMSE /%	IncNode Purity
RT	42.92	33524.4
TB3	5.66	5531.0
STB45	4.98	4377.8
DTB45	4.86	4304.8
CWP	4.6	5686.3
STB35	4.12	2879.3
RF1	3.91	6631.2
DRF12	3.74	5356.4
STB34	3.74	2908.1
RF2	3.69	3697.9
CER	3.64	3108.0
DTB34	3.34	2857.7
DTB35	3.14	2921.6
SRF12	2.65	9617.4
TB4	2.61	6172.1
TB5	2.4	4856.2

STB45 为 10.8 μm 通道云顶亮温与 12.0 μm 通道云顶亮温的差值,是光学厚积云和光学薄卷云的良好指示器^[17],这使得其对降水强度的预测具有重要作用。

DTB45 为 10.8 μm 通道云顶亮温与 12.0 μm 通道云顶亮温的商,10.8 μm 波长的辐射代表物体的热温度,如果目标可以被认为是黑体,则 TB4 近似等于物体的表面温度。通常情况下,降水云的云顶越高, TB4 就越低。TB5 的性能与 TB4 相似^[17]。两者均与降水云云顶高度有关,因此两者之商对降水强度的预测影响较大。

CWP 为云水含量,体现着云中含水量的多少,云粒子有效半径越大,云水含量越高,云产生的降水粒子越多、越大,地表降水率越大的可能性也随之增加。因此,CWP 对于降水强度的预测具有重要作用。

3.3 随机森林模型的降水预测应用

图 3 给出了随机森林模型预测的降水强度、实测样本以及两者之差在 0.5° 网格平均分布。通过对比图 3(a)与图 3(b),图 3(d)与图 3(e),1998—2007 年江淮梅雨和 9121 轨道的降水强度的预测分布和实际观测分布均具有较好的一致性,图 3(c)和

图 3(f)中均显示降水强度的预测与实际观测之差总体较小,前者预测与实际观测之差的范围为 -4.0 mm/h 至 1.0 mm/h ,后者绝大多数的降水强度预测误差小于 2.5 mm/h ,这说明该模型对于降水强度的预测效果较好,且具有较好的适用性和推广价值。同时,由图 3(c)和图 3(f)可发现,在 1998—2007 年江淮梅雨和 9121 轨道的降水强度反演中,实际观测降水强度较低时,预测的降水强度大于实际观测的降水强度。反之,预测的降水强度小于实际观测的降水强度,相同的验证结果说明该模型的稳定性较好。

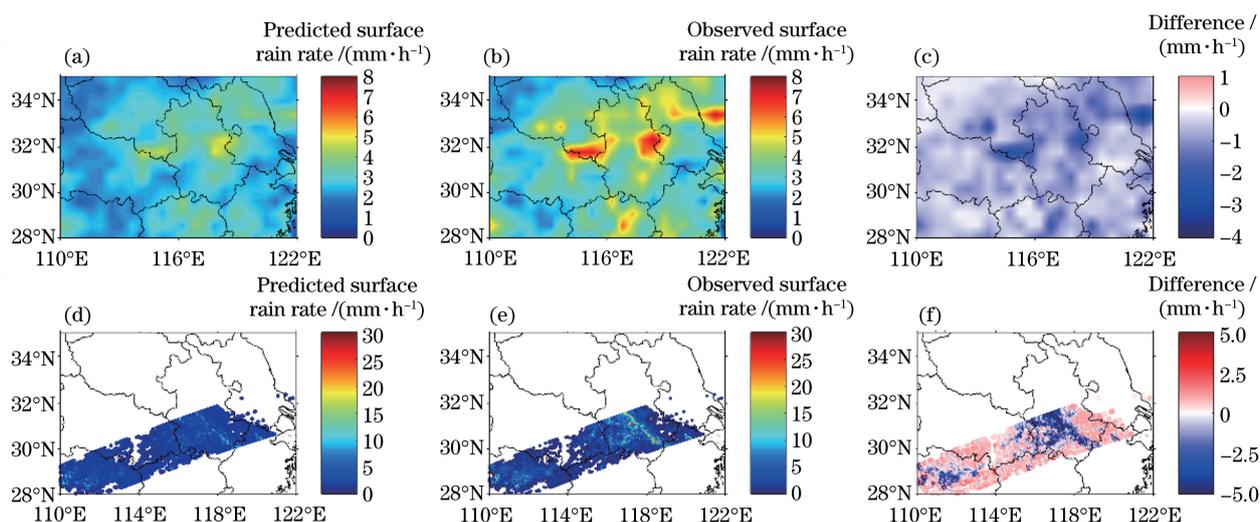


图 3 降水强度预测、实测和两者之差的 0.5° 网格平均分布。(a) 1998—2007 年长江淮河流域降水强度预测;(b) 1998—2007 年长江淮河流域降水强度实测;(c) 1998—2007 年长江淮河流域降水强度预测和实测两者之差;(d) 9121 轨道降水强度预测;(e) 9121 轨道降水强度实测;(f) 9121 轨道降水强度预测和实测两者之差

Fig. 3 0.5° grid average distribution of precipitation intensity prediction, observation and their difference. (a) Predicted surface rain rate over the Yangtze-Huaihe River Valley from 1998 to 2007; (b) observed surface rain rate over the Yangtze-Huaihe River Valley from 1998 to 2007; (c) difference of predicted and observed surface rain rates over the Yangtze-Huaihe River Valley from 1998 to 2007; (d) predicted surface rain rate in orbit 9121; (e) observed surface rain rate in orbit 9121; (f) difference of predicted and observed surface rain rates in orbit 9121

图 4(a)和图 4(b)显示了 1998—2007 年间降水和 9121 轨道云微物理参数和光谱参数预测的降水强度和观测到的降水强度的散点密度分布。如图 4(a)所示,1998—2007 年间降水云微物理参数和光谱参数通过该模型预测的降水强度与观测的降水强度在 40 mm/h 以下具有较好的一致性,降水强度在 40 mm/h 以上时预测准确度下降,由此验证了该模型在预测弱降水情况将具有更高准确性的猜想。经过 1998—2007 年间降水的降水强度预测验证,可得相关系数 R 为 0.76 ,这说明该模型的适用性较好,同时证明通过模型预测降水强度是可行的。如图 4(b)所示,9121 轨道云微物理参数和光谱参数

通过该模型预测的降水强度与观测的降水强度同样具有较好的一致性。经过 9121 轨道的降水强度预测验证,相关系数 R 为 0.61 ,再次验证了该模型的适用性和通过该模型预测降水强度的可行性。

3.4 降水云参数对地表降水强度的指示意义

云的结构特征参数与云的辐射特性、降水条件和降水效率等息息相关,为了系统地研究江淮梅雨期降水过程中云结构特征参数与降水强度的关系,基于 10 年的梅雨期降水数据,选取 4 个云结构特征参数(CWP, TB4, CER, COT),并对它们与降水强度分别进行了分档与统计。根据梅雨期降水率 RR 的大小,将降水雨强分为 5 级,即 $RR \in$

(0.5 mm/h, 2.5 mm/h] 时为小雨; $RR \in (2.5 \text{ mm/h}, 8 \text{ mm/h}]$ 时为中雨; $RR \in (8 \text{ mm/h}, 16 \text{ mm/h}]$ 时为大雨; $RR \in (16 \text{ mm/h}, 24 \text{ mm/h}]$ 时为暴雨; $RR \in (24 \text{ mm/h}, +\infty)$ 时为特大暴雨。并将大雨、暴雨和特大暴雨一起记为强降水。如表 4 所示,为了统计各类云参数在各档出现的概率,将各类云特征参数按照参数范围大小分成 5 档。值得注

意的是, TB4 的大小反映着云的云顶高度和相态。根据已有研究,将液相云定义为 TB4 大于 253 K, 冰相云则定义为 TB4 小于 233 K, 之间的则定义为混合云^[27], 综合 TB4 与云顶高度的关系, 可以将 1 档定义为高的冰相云, 2 档定义为较高的冰相云, 3 档定义为中等高度的混合云, 4 档定义为较低的液相云, 5 档定义为低的液相云。

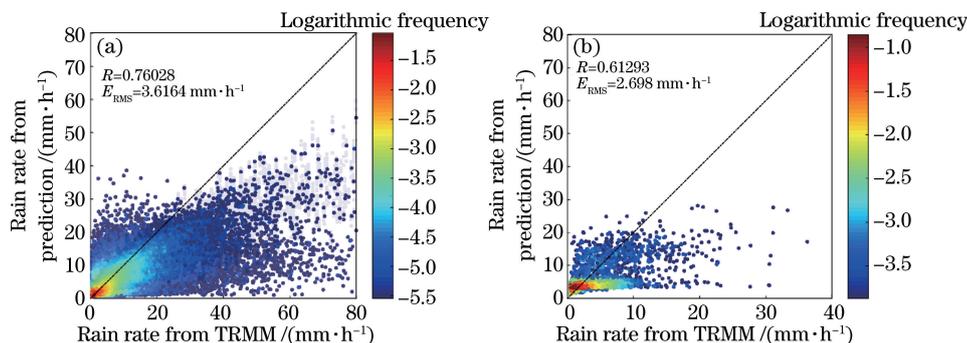


图 4 模型预测降水强度的散点密度分布图。(a)1998—2007 年江淮梅雨地区全部降水; (b) 9121 轨道降水
Fig. 4 Scatter of predicted and observed surface rain rate of proposed model. (a) All the rain rate during Mei-yu period over the Yangtze-Huaihe River Valley from 1998 to 2007; (b) all the rain rate in orbit 9121

表 4 云微物理参数和云顶温度统计分档表(各分档均只包括右边界值)

Table 4 Statistical classification table of cloud microphysical parameters and cloud top temperature (each segment contains only the right boundary value)

Classification	CWP / ($\text{kg} \cdot \text{m}^{-3}$)	TB4 / K	CER / μm	COT
1	(0, 0.5]	(0, 213]	(6, 10]	(0, 30]
2	(0.5, 1.0]	(213, 233]	(10, 14]	(30, 60]
3	(1.0, 1.5]	(233, 253]	(14, 18]	(60, 90]
4	(1.5, 2.0]	(253, 273)	(18, 22]	(90, 120]
5	(2.0, $+\infty$)	(273, $+\infty$)	(22, $+\infty$]	(120, $+\infty$]

不同云特征参数分档下各级降水强度的概率如图 5 所示。CWP 在 $0.5 \sim 1.0 \text{ kg} \cdot \text{m}^{-3}$ 范围中, 小雨的概率最大[图 5(a)], 而当 CWP 处于 $1.5 \sim 2.0 \text{ kg} \cdot \text{m}^{-3}$ 时, 其他 4 类雨的概率均增大, 这说明: 当 CWP 小于 $1 \text{ kg} \cdot \text{m}^{-3}$ 时, 江淮梅雨期降水主要以小雨为主, 而当 CWP 大于 $1.5 \text{ kg} \cdot \text{m}^{-3}$ 时, 大雨和暴雨的降水概率明显增大。由图 5(b) 可知, 当 TB4 处于 $233 \sim 253 \text{ K}$ 的范围时, 混合相的云下小雨的概率最大, 而当 TB4 的值越来越小, 梅雨期强降水的概率也越来越大, 中雨的峰值显示在 $213 \sim 233 \text{ K}$, 暴雨和特大暴雨的峰值显示在 $TB4 \leq 213 \text{ K}$ 的范围内。显而易见, 高度更高的冰相云能够产生强度更大的降水。对于 CER[图 5(c)], 各等级降水的 CER 大小主要位于 $10 \mu\text{m}$ 以上的范围, 降水概率总体上随着 CER 的增大而单调递增。由图 5(d)

可见, 对于云光学厚度 COT, 各级降水概率随着 COT 的增大而不断增大, 当 COT 不大于 120 时, 各分档内有较小概率的小雨和中雨, 强降水的概率并不大, 而当 COT 大于 120 时, 各级降水的概率显著增强, 尤其是强降水。

降水强度、云微物理参数与云顶高度的关系如图 6 所示。在图 6(a) 中, 梅雨期地表平均降水强度随着降水云顶高度降低而逐渐变小, 冰相云的平均降水强度要明显大于混合相云和液相云, 高度更高的冰相云的降水平均强度和强度范围要高于高度较低的冰相云, 液相云和混合相云的降水平均强度与强度范围则相差不大。随着云顶高度逐渐变高, 云的相态由液相到混合相再到冰相转换, 云滴粒子之间出现碰并现象, 云的 CER 值范围也出现了明显的增加[图 6(b)]。

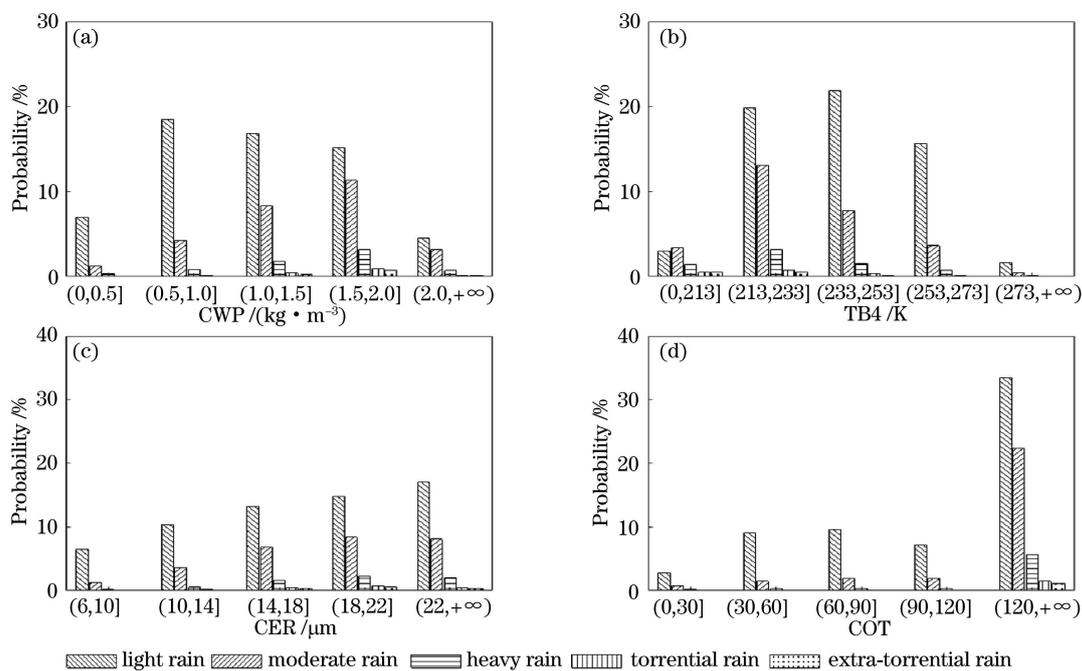


图 5 不同云特征参数分档下各级降水强度的概率分布。(a) CWP; (b) TB4; (c) CER; (d) COT

Fig. 5 Probability distribution of precipitation intensity at different levels (light rain, moderate rain, heavy rain, torrential rain, extra-torrential rain) under different cloud characteristic parameters. (a) CWP; (b) TB4; (c) CER; (d) COT

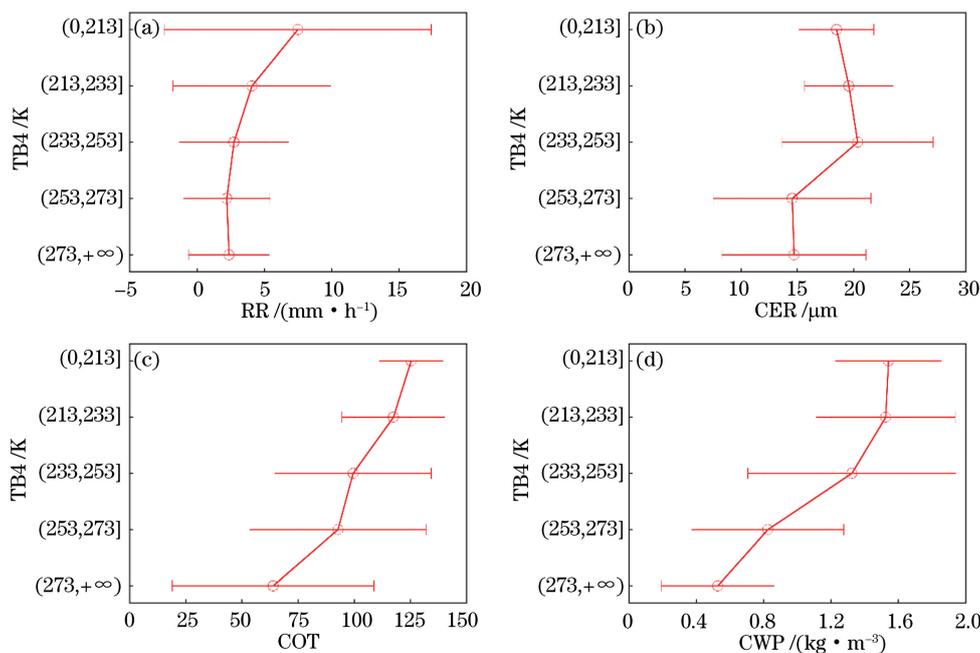


图 6 降水强度、云微物理参数与云顶高度的关系。(a) RR; (b) CER; (c) COT; (d) CWP

Fig. 6 Relationship between precipitation intensity, cloud characteristic parameters and cloud top height as cloud top height. (a) RR; (b) CER; (c) COT; (d) CWP

值得一提的是,在液相云中,平均 CER 出现明显的减小现象,明显低于高度更高的冰相云,其主要原因有:一方面是由于云发展不旺盛,气流较弱,云滴碰并较弱,云滴较小^[4,13];另一方面可能是由于气溶胶的间接效应作用,又因为江淮地区为气溶胶高

污染区域,故边界层中的气溶胶更容易进入较低的云层(4 km 以下)中,大量气溶胶更容易争食水汽,云滴长不大^[13]。图 6(c)和图 6(d)则显示了相似的图形,平均 COT 和平均 CWP 值都随着高度的降低而不断降低,唯一不同是,COT 值的范围会随着高

度降低而不断变大,然而 CWP 值得的范围却在混合相云中为最大。因此,在高度更高的冰相云中,云滴有效半径较大,非常大的云光学厚度和云水含量,产生的降水强度的数值范围更大。

4 结 论

基于 TRMM 卫星光谱和雷达的同步资料,采用随机森林方法对江淮地区梅雨期的降水进行反演,证明了随机森林方法在卫星反演降水中的优越性,并统计研究了梅雨期的云微物理和云顶光谱信息与降水强度的关系。结果表明:

随机森林模型反演的降水过程中,训练集中预测降水强度与观测的降水强度的相关系数 R 为 0.94,均方根误差为 2.30 mm/h;模型在十折交叉验证测试集中预测降水强度与观测的降水强度的相关系数为 0.67,均方根误差为 4.06 mm/h,具有较高的降水预测精度。利用模型对 10 年所有降水数据进行反演时,观测值和预测值的相关系数为 0.76,均方根误差为 3.60 mm/h;9121 轨道个例反演预测降水的相关系数为 0.61,均方根误差为 2.70 mm/h。降水强度的预测分布和实际观测分布有着较好的一致性。

江淮梅雨的地表平均降水强度随着降水云顶高度的降低而逐渐变小,中小雨一般由云水含量低、云滴半径小的液相云和混合云所产生,而强降水通常是由云水含量高、云滴半径大的冰相云产生。随机森林模型给出了输入预测因子的重要性排序,云微物理参数的 CWP 排在前列。进一步分析表明,当 CWP 小于 $1.0 \text{ kg} \cdot \text{m}^{-3}$ 时,江淮梅雨期降水主要以小雨为主,而当 CWP 大于 $1.5 \text{ kg} \cdot \text{m}^{-3}$ 时,大雨和暴雨的降水概率明显增大。对于 CER,各等级降水的 CER 大小主要位于 $10 \mu\text{m}$ 以上,降水概率总体上随着 CER 的增大而单调递增。对于云光学厚度 COT,各级降水概率随着 COT 的增大而不断增大,当 COT 不大于 120 时,各分档内有较小概率的小雨和中雨,强降水的概率并不大,而当 COT 大于 120 时,各级降水的概率显著增强,尤其是强降水。

总的来看,本研究利用 TRMM 卫星上同步的光谱和降水信息,通过机器学习方法建模和验证,得到光谱反演降水的机器学习算法框架,可利用该框架在不同卫星平台上进行光谱预测降水研究。此外,在以往的研究中,云顶信息包含了降水云与非降水云的混合光谱信息,本研究通过热带测雨雷达和光谱信息的融合,提取了同步观测的降水云云顶信

息与地表降水强度,给出了二者之间更为精准的对应关系。特别地,随机森林模型反演降水的预测因子重要性研究结果,将有助于深入理解降水云参数对地面降水强度变化的指示作用,对江淮梅雨期的暴雨监测、防灾减灾,以及人工影响降雨等有重要的科学支撑和应用价值。

致谢 感谢中国科学技术大学傅云飞教授提供热带测雨卫星的轨道级融合资料并对本文提出宝贵意见。

参 考 文 献

- [1] Wang C X, Yu F, Zhang C W. Retrieval research on rainfall intensity field at all time during Mei-Yu period by using MTSAT multi-spectral imagery[J]. Journal of Nanjing University (Natural Sciences), 2010, 46(3): 305-316.
王晨曦, 郁凡, 张成伟. 基于 MTSAT 多光谱卫星图像监测全天时我国华东地区的梅雨期降水[J]. 南京大学学报(自然科学版), 2010, 46(3): 305-316.
- [2] Zhou Y Q, Cai M, Ou J J, et al. Correlation between cloud characteristic parameters and precipitation[J]. Transactions of Atmospheric Sciences, 2011, 34(6): 641-652.
周毓荃, 蔡森, 欧建军, 等. 云特征参数与降水相关性的研究[J]. 大气科学学报, 2011, 34(6): 641-652.
- [3] Fu Y F, Pan X, Yang Y J, et al. Climatological characteristics of summer precipitation over East Asia measured by TRMM PR: a review[J]. Journal of Meteorological Research, 2017, 31(1): 142-159.
- [4] Fu Y F. Cloud parameters retrieved by the bispectral reflectance algorithm and associated applications[J]. Acta Meteorologica Sinica, 2014, 72(5): 1039-1053.
傅云飞. 利用卫星双光谱反射率算法反演的云参数及其应用[J]. 气象学报, 2014, 72(5): 1039-1053.
- [5] Rosenfeld D, Gutman G. Retrieving microphysical properties near the tops of potential rain clouds by multispectral analysis of AVHRR data [J]. Atmospheric Research, 1994, 34 (1/2/3/4): 259-283.
- [6] Zhang J, Zhang Q, Tian W S, et al. Remote sensing retrieval and analysis of optical character of cloud in Qilian mountains [J]. Journal of Glaciology and Geocryology, 2006, 28(5): 722-727.
张杰, 张强, 田文寿, 等. 祁连山区云光学特征的遥感反演与云水资源的分布特征分析[J]. 冰川冻土, 2006, 28(5): 722-727.
- [7] Liu J, Zhang W J, Zhu Y J, et al. Case study on

- cloud properties of heavy rainfall based upon satellite data[J]. *Journal of Applied Meteorological Science*, 2007, 18(2): 158-164.
- 刘健, 张文建, 朱元竞, 等. 中尺度强暴雨云团云特征的多种卫星资料综合分析[J]. *应用气象学报*, 2007, 18(2): 158-164.
- [8] Zhou Y Q, Chen Y Y, Li J, et al. Retrieval and preliminary test of cloud physical parameters from combination of FY-2C/D geostationary satellite data and other observation data [J]. *Meteorological Monthly*, 2008, 34(12): 27-35, 130-131.
- 周毓荃, 陈英英, 李娟, 等. 用FY-2C/D卫星等综合观测资料反演云物理特性产品及检验[J]. *气象*, 2008, 34(12): 27-35, 130-131.
- [9] Wang L, Zhou Y Q, Cai M, et al. Study on correlation between cloud characteristic parameters and precipitation in North China[J]. *Meteorological and Environmental Sciences*, 2019, 42(3): 9-16.
- 王磊, 周毓荃, 蔡森, 等. 华北云特征参数与降水相关性的研究[J]. *气象与环境科学*, 2019, 42(3): 9-16.
- [10] Zheng Q, Zheng Y F, Wang L W, et al. Macrophysical and microphysical properties of ice clouds during heavy rainfalls in Beijing—Tianjin—Hebei region in summer[J]. *Arid Land Geography*, 2019, 42(1): 67-76.
- 郑倩, 郑有飞, 王立稳, 等. 京津冀夏季强降水下冰云宏观特征[J]. *干旱区地理*, 2019, 42(1): 67-76.
- [11] Fu Y F, Liu P, Liu Q, et al. Climatological characteristics of VIRS channels for precipitating cloud in summer over the tropics and subtropics[J]. *Journal of Atmospheric and Environmental Optics*, 2011, 6(2): 129-140.
- 傅云飞, 刘鹏, 刘奇, 等. 夏季热带及副热带降水云可见光/红外信号气候分布特征[J]. *大气与环境光学学报*, 2011, 6(2): 129-140.
- [12] Yang Y J, Lu D R, Fu Y F, et al. Spectral characteristics of tropical anvils obtained by combining TRMM precipitation radar with visible and infrared scanner data [J]. *Pure and Applied Geophysics*, 2015, 172(6): 1717-1733.
- [13] Yang Y J, Wang H, Chen F J, et al. TRMM-based optical and microphysical features of precipitating clouds in summer over the Yangtze-Huaihe river valley, China [J]. *Pure and Applied Geophysics*, 2019, 176(1): 357-370.
- [14] Ding Y H, Liu J J, Sun Y, et al. A study of the synoptic-climatology of the Meiyu system in East Asia[J]. *Chinese Journal of Atmospheric Sciences*, 2007, 31(6): 1082-1101.
- 丁一汇, 柳俊杰, 孙颖, 等. 东亚梅雨系统的天气-气候学研究[J]. *大气科学*, 2007, 31(6): 1082-1101.
- [15] Liang P, He J H. Review for climate change of Meiyu over the Yangtze-Huaihe basins [J]. *Plateau Meteorology*, 2008, 27(S1): 8-15.
- 梁萍, 何金海. 江淮梅雨气候变化研究进展[J]. *高原气象*, 2008, 27(S1): 8-15.
- [16] Qian Y F, Wang Q Q, Huang D Q. Studies of floods and droughts in the Yangtze-Huaihe river basin [J]. *Chinese Journal of Atmospheric Sciences*, 2007, 31(6): 1279-1289.
- 钱永甫, 王谦谦, 黄丹青. 江淮流域的旱涝研究[J]. *大气科学*, 2007, 31(6): 1279-1289.
- [17] Yang Y J, Fu Y F, Chen F J, et al. Spectral characteristics of precipitating clouds during the Meiyu over the Yangtze-Huaihe River Valley from merged TRMM precipitation radar and visible/infrared scanner data [J]. *Proceedings of SPIE*, 2014, 9259: 92591K.
- [18] Fu Y F, Liu D, Wang Y, et al. Characteristics of precipitating and non-precipitating clouds in typhoon raman as viewed by TRMM combined measurements [J]. *Acta Meteorologica Sinica*, 2007, 65(3): 316-328.
- 傅云飞, 刘栋, 王雨, 等. 热带测雨卫星综合探测结果之“云娜”台风降水云与非降水云特征[J]. *气象学报*, 2007, 65(3): 316-328.
- [19] Fu Y F, Li H T, Zi Y. Case study of precipitation cloud structure viewed by TRMM satellite in a valley of the Tibetan Plateau [J]. *Plateau Meteorology*, 2007, 26(1): 98-106.
- 傅云飞, 李宏图, 自勇. TRMM卫星探测青藏高原谷地的降水云结构个例分析[J]. *高原气象*, 2007, 26(1): 98-106.
- [20] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [21] Cao C P, Wang H Q, Wang K, et al. Intelligent evaluation method of grottoes surface weathering based on multispectral imaging and random forest algorithm [J]. *Acta Optica Sinica*, 2020, 40(22): 2230001.
- 曹赤鹏, 王慧琴, 王可, 等. 基于多光谱成像和随机森林算法的石窟表面风化智能评估方法[J]. *光学学报*, 2020, 40(22): 2230001.
- [22] Tian M L, Ge X Y, Ding J L, et al. Coupled machine learning and unmanned aerial vehicle based hyperspectral data for soil moisture content estimation [J]. *Laser & Optoelectronics Progress*, 2020, 57(9): 093002.
- 田美玲, 葛翔宇, 丁建丽, 等. 耦合机器学习和机载高光谱数据的土壤含水量估算[J]. *激光与光电子学*

- 进展, 2020, 57(9): 093002.
- [23] Min M, Bai C, Guo J P, et al. Estimating summertime precipitation from himawari-8 and global forecast system based on machine learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(5): 2557-2570.
- [24] Yuan L S, Lou M Y, Liu Y Q, et al. Palm vein classification based on deep neural network and random forest [J]. Laser & Optoelectronics Progress, 2019, 56(10): 101010.
袁丽莎, 娄梦莹, 刘娅琴, 等. 结合深度神经网络和随机森林的手掌静脉分类[J]. 激光与光电子学进展, 2019, 56(10): 101010.
- [25] Dai P Q, Ding L X, Liu L J, et al. Tree species identification based on FCN using the visible images obtained from an unmanned aerial vehicle[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101001.
戴鹏钦, 丁丽霞, 刘丽娟, 等. 基于 FCN 的无人机可见光影像树种分类[J]. 激光与光电子学进展, 2020, 57(10): 101001.
- [26] Stewart R E, Marwitz J D, Pace J C, et al. Characteristics through the melting layer of stratiform clouds [J]. Journal of the Atmospheric Sciences, 1984, 41(22): 3227-3237.
- [27] Fu Y F, Qin F. Summer daytime precipitation in ice, mixed, and water phase as viewed by PR and VIRS in tropics and subtropics [J]. Proceedings of SPIE, 2014, 9259: 925906.