

郭雪星, 瞿建华, 叶凌梦, 等. 基于朴素贝叶斯的 FY-4A/AGRI 云检测方法. 应用气象学报, 2023, 34(3): 282-294.

DOI: 10.11898/1001-7313.20230303

# 基于朴素贝叶斯的 FY-4A/AGRI 云检测方法

郭雪星 瞿建华\* 叶凌梦 韩旻 史墨杰

(北京华云星地通科技有限公司, 北京 100081)

## 摘要

光学遥感云检测是定量遥感和遥感应用的基础, 尝试将朴素贝叶斯的机器学习方法应用于风云四号气象卫星 A 星(FY-4A)搭载的先进静止轨道辐射成像仪(AGRI)红外通道数据云检测。因辐射物理方法的云检测采用可见光通道导致存在日夜不连续现象, 仅选取 FY-4A/AGRI 载荷 7 个红外通道的光谱数据, 构建 10 种特征分类, 利用正交偏振云-气溶胶激光雷达(CALIOP)与 FY-4A/AGRI 时空匹配数据, 对不同地表类型和不同季节的数据集进行分类训练和验证。与 CALIOP 数据交叉验证显示除积雪上空云识别准确率约为 81%, 深海、浅水、陆地和荒漠上空的云识别准确率均高于 92%, 误判率基本低于 10%, 总体云识别精度达到 90%; 与 2021 年 10 月和 2022 年 1, 4, 7 月 MODIS 2 级云检测产品比对, 深海、浅水云识别准确率均在 88% 以上, 误判率分别低于 3% 和 10%, 夏季云识别效果最佳, 总体云识别准确率达 90%。云检测结果不仅得到云、可能云、可能晴空和晴空 4 种分类结果, 还得到每种特征和综合特征云检测分类器的不确定性概率值, 这为云和地表相关检测产品提供重要参考。

**关键词:** FY-4A/AGRI; 朴素贝叶斯; 云检测

## 引言

云在地球大气系统的辐射收支平衡中的作用由其光学性质决定<sup>[1]</sup>。国际卫星云气候计划(International Satellite Cloud Climatology Project, ISCCP)提供的全球云量数据显示地球表面 66% 以上区域经常被云覆盖<sup>[2-5]</sup>。

天气现象均与云的生消和变化密切相关。气象卫星对地球表面和云层持续进行大范围观察, 卫星云图所蕴含的气象信息非常丰富, 综合展示了大气热力和动力过程演变<sup>[6-13]</sup>。卫星云区产品是当前数值预报同化的热点和难点, 尤其是云检测精度对数值天气预报准确率具有较大影响<sup>[14]</sup>。

目前气象卫星光学载荷云检测大多基于可见光反射和红外辐射特性, 通过统计方法或经验统计, 设定静态或动态阈值开展云检测<sup>[15-20]</sup>, 阈值受空间、时

间和下垫面等因素的影响较大。物理阈值法包括 ISCCP 法、APOLLO 法、CLAVE 法和 CO<sub>2</sub> 薄片法<sup>[21]</sup>。随着机器学习或深度学习技术的发展与成熟, 国内外对其在云检测中的应用进行了广泛而深入的研究<sup>[22-29]</sup>。基于机器学习或深度学习的方法可以获取云或晴空与不同光谱或不同图像的隐藏关系, 可在一定程度上避免阈值误差引起的云检测偏差, 其中朴素贝叶斯机器学习方法已应用于美国和欧洲 AVHRR(advanced very high resolution radiometer)、VIIRS(visible infrared imaging radiometer)、ABI(advanced baseline imager)和 SEVIRI(spinning enhanced visible and infrared imager)等多个气象卫星有效载荷的云检测产品<sup>[30-35]</sup>。国内亦有将朴素贝叶斯方法应用于 AGRI(advanced geostationary orbital radioimager)和 MERSI(medium resolution spectral imager)等多个风云气象卫星有效载荷的云检测算法研究<sup>[21]</sup>。

2022-12-23 收到, 2023-03-23 收到再改稿。

资助项目: 国家自然科学基金项目(42201399)

\* 通信作者, 邮箱: qujh@cma.gov.cn

无论是阈值法、机器学习还是深度学习等云检测方法均区分白天、夜间模式,白天选取可见光和红外通道数据,夜间仅选取红外通道数据。可见光通道可提高云的识别率,尤其是卷云的识别率,但受到太阳高度角度的影响,仅可在一定范围内使用,导致云检测结果存在日夜不连续现象,特别是在晨昏时段。因此,本文提出基于机器学习的FY-4A/AGRI云检测方法。该方法选择高时间分辨率静止气象卫星FY-4A/AGRI全部红外通道数据,以正交偏振云-气溶胶激光雷达(cloud-aerosol lidar with orthogonal polarization, CALIOP)数据为确定云或晴空的有效训练数据,采用朴素贝叶斯方法对不同地表类型和不同季节的数据集进行分类训练和判别,构建云/晴空特征分类器,实现FY-4A/AGRI全天时一致、连续的精准云识别。

## 1 数据

本文主要使用4类数据:第1类是FY-4A/AGRI红外通道数据,根据不同红外通道亮温定义云检测分类器;第2类是正交偏振云-气溶胶激光雷达云层产品数据,其提供的云探测信息是被动遥感中云探测的重要检验数据,本文以2级(L2)1 km云层产品提供的云检测信息为参考,进行模型训练和精度验证;第3类是中分辨率成像光谱仪(moderate resolution imaging spectroradiometer, MODIS)云检测产品;第4类是基于多源数据构建动态全年12个月不同地表类型文件。

### 1.1 FY-4A/AGRI数据

FY-4A是我国的静止轨道气象卫星,其上搭载的先进静止轨道辐射成像仪有14个通道,波长为 $0.45\sim 13.8\ \mu\text{m}$ ,空间分辨率为 $0.5\sim 4\ \text{km}$ ,不仅可以观测云、水汽、植被和地表,还具备捕捉气溶胶和冰雪的能力。本研究选用全部红外通道,包括中波红外(通道8)、水汽通道(通道9、通道10和通道11)和长波红外通道(通道12、通道13和通道14)。

### 1.2 CALIOP数据

CALIPSO(Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation)卫星发射于2006年4月28日,是NASA和CNES的太阳轨道地球侦察卫星。CALIOP是该卫星的主要载荷之一,可测量

532 nm反向散射的两个正交偏振分量和1064 nm反向散射的总强度。每个CALIOP激光雷达足迹直径约为100 m,足迹之间(中心到中心)的物理间隔为335 m,平均3个CALIOP足迹可以获得1 km云产品。CALIOP 2级产品提供全球气溶胶和云观测数据,分辨率为333 m,1 km和5 km<sup>[36]</sup>。本文使用CALIOP 2级1 km云层产品,该产品提供1 km水平分辨率下的云探测信息,包括垂直方向的云分布层数、分层云顶高度等信息。云层产品包括每个扫描像元的云层计数(最大层数为10),若云层计数大于0(云顶高度非填充值),该扫描像元被定义为云像元;若云层计数等于0(云顶高度为填充值),该扫描像元被定义为晴空像元,且其云顶高度标记为0<sup>[36]</sup>。本文使用2020年12月及2021年1—11月共12个月的数据进行时空匹配生成对应数据集,每月随机选取80%的数据作为训练数据集,用于构建云检测分类器,选取2021年1,4,7,10月剩余20%的数据作为测试数据集,用于精度验证。

### 1.3 MODIS云检测C6版本数据

为了进一步评估本文算法精度,采用MODIS云检测产品C6版本MYD35进行评估。MODIS是EOS(Earth Observation System)卫星搭载的光学遥感仪器,有36个光谱波段,光谱范围宽为 $0.4$ (可见光) $\sim 14.4\ \mu\text{m}$ (热红外)。空间分辨率为250,500 m和1000 m,扫描宽度为2330 km。C6版本云检测算法使用多个通道组合和阈值计算置信度以区分云或晴空,云检测产品中每个像元被标记为晴空、可能晴空、可能云或云。本文MODIS云检测分类中可能晴空划归为晴空,可能云划归为云。选取2021年10月及2022年1,4,7月共4个月数据进行精度验证。

### 1.4 地表类型

将地球表面分为5种类型:1为深海,2为浅水,3为陆地,4为积雪,5为荒漠。结合MODIS地理定位文件、SEEBOR表面发射率数据库中 $3.75\ \mu\text{m}$ 表面发射率文件<sup>[37]</sup>、EOSMLT卫星AVHRR提供的近实时逐日全球合成5 km冰雪产品文件和FY-4A气象卫星导航文件,获得FY-4A全圆盘全年12个月动态地表类型文件数据。本文以1,4,7,10月作为典型月份进行展示。图1为各季节典型月份FY-4A全圆盘5种地表类型分布。

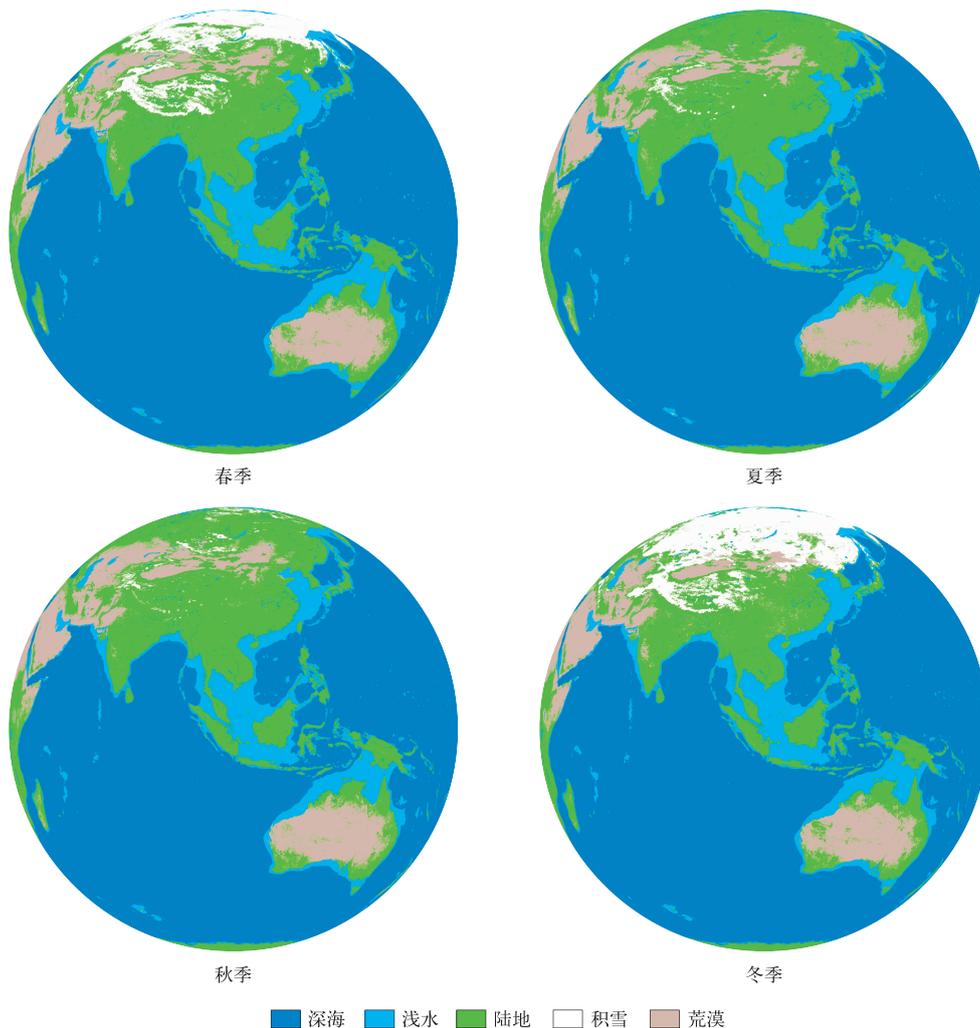


图1 各季节典型月份FY-4A全圆盘地表覆盖

Fig.1 FY-4A global surface coverage classification in typical month for seasons

## 2 方法

基于机器学习的FY-4A/AGRI云检测方法(advanced bayesian cloud mask,简称ABCM),包括分类训练和应用两部分。分类训练主要是对2020年12月、2021年1—11月每月80%的FY-4A/AGRI全圆盘、中国区数据和CALIPSO/CALIOP 2级云层产品进行时空匹配,构建训练数据集,利用不同月份的动态地表类型文件,对训练数据集进行地表类型归类,基于贝叶斯原理计算各云检测分类器的贝叶斯模型参数,按季节分类训练生成不同季节的云检测分类器,其中2020年12月—次年2月为2020年冬季,2021年3—5月、6—8月、9—11月为2021年春、夏、秋季。根据季节将对应的动态地表

类型文件和云检测分类器应用于AGRI云检测,设置分类阈值,得到云检测结果和云属概率。

### 2.1 AGRI与CALIOP数据时空匹配

FY-4A/AGRI和CALIPSO/CALIOP的运行轨道不同,观测方式、观测时间和空间分辨率均存在差异。FY-4A/AGRI星下点4 km空间分辨率的1个像元对应CALIPSO/CALIOP 2级1 km云层产品的多个像元。为了获取更准确的云检测信息,提高分类器的云识别能力,结合云瞬息万变的特点,采用阈值控制和云覆盖率相结合的方式进行时空匹配:以FY-4A/AGRI观测像元为基准,过滤出观测时间差在5 min内的CALIOP像元,选取球面距离在4 km范围内的CALIOP像元;针对匹配到的多个CALIOP像元进行云覆盖率计算,将CALIOP有云像元的有云概率设定为1,晴空或其他情况的有云概率

设定为0,计算与1个FY-4A/AGRI像元空间匹配的多个CALIPSO像元的平均云覆盖率,当CALIPSO平均云覆盖率大于50%时,认定为云像元。

## 2.2 云检测分类器

### 2.2.1 贝叶斯原理

按照条件概率理论,气象卫星光学载荷的云检测问题可以描述为已知像元的若干特征 $F$ ,判断其属于云的条件概率 $P(C_{\text{cloud}}|F)$ ,

$$P(C_{\text{cloud}}|F) = \frac{P(C_{\text{cloud}})P(F|C_{\text{cloud}})}{P(F)} \quad (1)$$

式(1)中, $P(C_{\text{cloud}})$ 代表在不考虑特征时像元属于云的先验概率, $P(F)$ 是特征 $F$ 的像元概率, $P(F|C_{\text{cloud}})$ 为已知属于云的像元中特征 $F$ 的条件概率。特征 $F$ 可以由依据卫星成像仪和其他先验知识计算得到的能够反映像元云属特性的属性值。在贝叶斯理论中, $P(C_{\text{cloud}}|F)$ 被称为后验概率。

由于云判识问题的复杂和观测的局限性,单一特征不足以反映云属类别的差异,需要组合多种特征 $F_i, i=1, \dots, N$ 。在经典的贝叶斯条件概率(式(1))中, $P(F|C_{\text{cloud}})$ 反映多维特征联合的条件概率,需要计算 $N$ 组联合分布 $P(\{F_1, F_2, \dots, F_n\}|C_{\text{cloud}})$ 。充足的样本具有良好的代表性,但伴随巨大的计算量以及特征可能的发散性。为了兼顾训练样本的代表性和计算量,引入朴素贝叶斯原理,在假设各特征分布相互独立的条件下,基于大数定律和中心极限定理对特征进行分析,特征分布实际趋向于正态分布或区间截断的正态分布。因此,对各类特征引入正态分布假设,即假设总体云属像元在某类特征上的分布符合正态分布

$$I(F_i^{(j)}, C_{\text{yes}}) \sim N(\mu_i, \sigma_i) \quad (2)$$

式(2)中, $\mu_i$ 为云属像元在第 $i$ 类特征上的总体期望, $\sigma_i$ 为云属像元在第 $i$ 类特征上的总体标准差,经由样本标准差无偏估计得到。

另一方面, $P(F)$ 可分解为

$$P(F) = P(C_{\text{cloud}}) \prod_{i=1}^N P(F_i|C_{\text{cloud}}) + P(C_{\text{clear}}) \prod_{i=1}^N P(F_i|C_{\text{clear}}) \quad (3)$$

式(3)中, $P(C_{\text{clear}})$ 为不考虑特征时像元属于晴空的先验概率, $P(F|C_{\text{clear}})$ 为已知属于晴空的像元特征 $F$ 的条件概率。

### 2.2.2 分类器定义

针对FY-4A/AGRI红外波段,设计以下10种

云相关特征<sup>[30-34]</sup>:① $3 \times 3$ 像素矩阵的 $10.7 \mu\text{m}$ 通道亮温标准差,该特征不适用于海岸线或高山像素<sup>[31]</sup>;② $3 \times 3$ 像素矩阵 $10.7 \mu\text{m}$ 通道亮温的最大值与目标像元亮温之差,可检测云边缘和其他小规模云特征<sup>[31]</sup>,当 $3 \times 3$ 像素矩阵内的最大亮温大于300 K、目标像元为海岸线、冷地表或雪时,该特征不可用;③ $10.7 \mu\text{m}$ 通道亮温,可检测厚的高云,但当目标像元为冷地表或雪时,该特征不可用<sup>[31]</sup>。应用时根据月份和海拔高度的变化对该特征进行适当的亮温补偿<sup>[36]</sup>,补偿原则为12月一次年2月海拔高度每升高100 m补偿0.2 K,3—5月和9—11月补偿0.5 K,6—8月补偿0.7 K;④目标像素 $10.7 \mu\text{m}$ 和 $8.5 \mu\text{m}$ 通道的亮温差。因两个通道对大气水汽的吸收不同,导致亮温存在差异<sup>[38]</sup>。一般该特征在云像元处为负值,在晴空像元处为正值,且仅适用于液态水体下垫面<sup>[31]</sup>;⑤目标像素 $3.75 \mu\text{m}$ 和 $10.7 \mu\text{m}$ 通道的亮温差,夜间该特征为正值时,可用于检测部分云或薄云<sup>[31]</sup>;⑥ $3.75 \mu\text{m}$ 通道的伪发射率<sup>[31]</sup>。对于半透明云和冷云,该特征很大,但当目标像元 $10.7 \mu\text{m}$ 通道亮温大于310 K时,该特征不可用;⑦目标像素 $10.7 \mu\text{m}$ 和 $12.0 \mu\text{m}$ 通道的亮温差,该特征为正值时,可用于检测半透明云<sup>[31]</sup>;⑧目标像素 $10.7 \mu\text{m}$ 和 $6.25 \mu\text{m}$ 通道的亮温差,当 $10.7 \mu\text{m}$ 通道的观测亮温大于220 K时,该特征可用于检测薄云<sup>[31]</sup>;⑨目标像素 $10.7 \mu\text{m}$ 和 $7.1 \mu\text{m}$ 通道的亮温差,可识别夜间陆地上空的高层和中层云;⑩目标像素 $10.7 \mu\text{m}$ 和 $13.5 \mu\text{m}$ 通道的亮温差,可识别高云。

### 2.2.3 分类器构建

为了降低卫星观测特征随时间变化对云检测结果产生的影响,提高云检测分类器的云识别能力,基于动态地表类型文件对训练数据集进行地表类型归类,同时兼顾训练样本量,将训练数据集分别作为冬、春、夏、秋的训练样本,最终生成不同季节的云检测分类器。

采用概率密度函数(probability density function, PDF)对5种地表类型分别统计,生成各分类器云像元的PDF分布图。图2是夏季亮温补偿后 $10.7 \mu\text{m}$ 通道的亮温对应5种地表类型的像元特征和PDF分布图。由图2可见,深海的训练样本量最多,其次是陆地、浅水和荒漠,积雪的样本量最少,荒漠上有云和晴空像元的特征区别明显。

如上所述,利用云检测分类器计算各像元的云属概率  $P(F)$ ,设定云检测分类规则如下:若  $P(F) \geq 0.85$ ,将像元划分为云;若  $0.68 \leq P(F) < 0.85$ ,将

像元划分为可能云;若  $0.4 < P(F) < 0.68$ ,将像元划分为可能晴空;若  $P(F) \leq 0.4$ ,将像元划分为晴空。

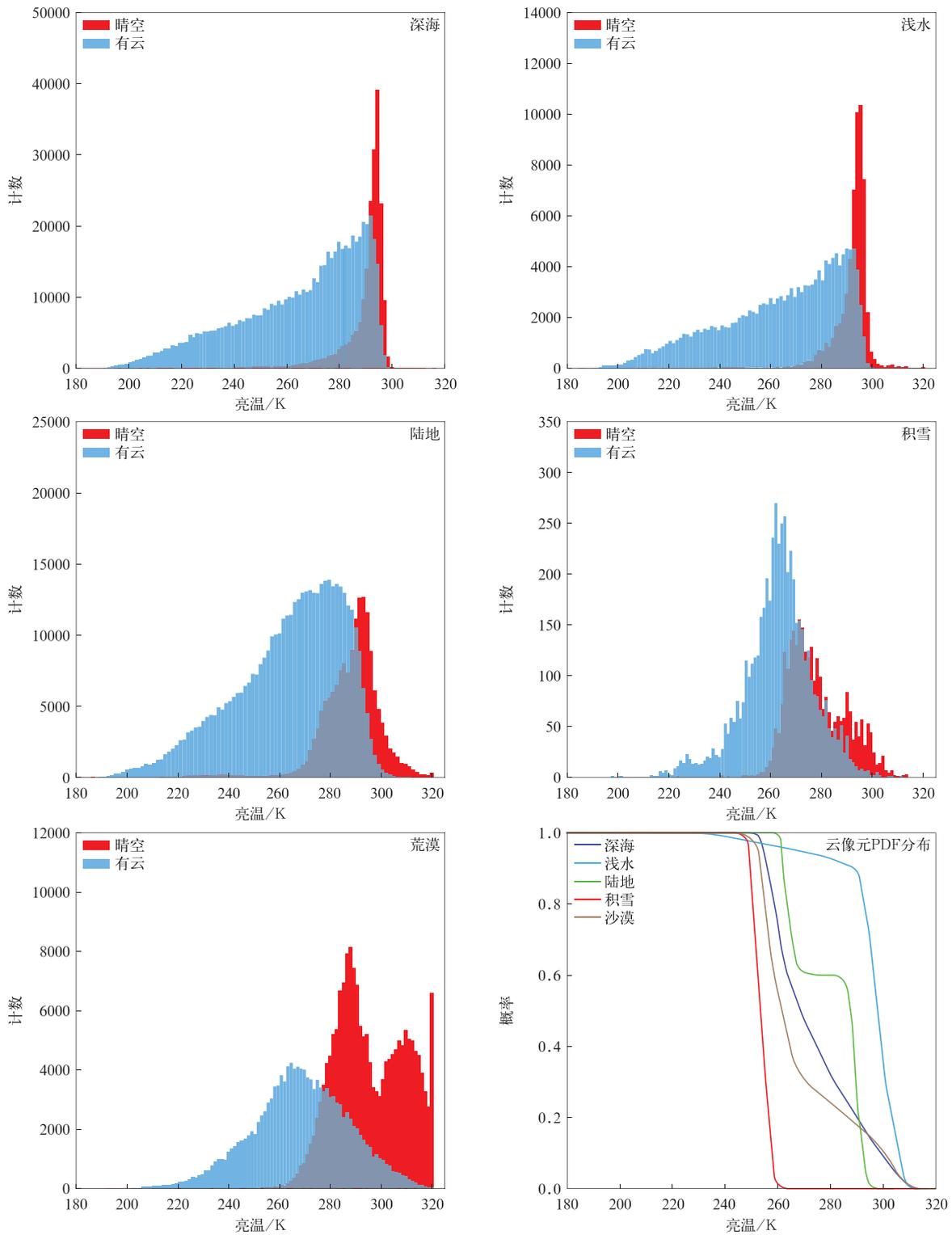


图 2 夏季 10.7  $\mu\text{m}$  通道亮温特征

Fig. 2 Brightness temperature features for 10.7  $\mu\text{m}$  channel in summer

### 3 云检测验证

为验证不同季节云检测分类器精度,定量评估 ABCM 云检测结果,以 CALIOP 2 级 1 km 云层产品和 MODIS 2 级 1 km 云产品 MOD35\_L2 作为检测源,与本文 ABCM 云检测结果做交叉验证。采用 5 种评估参数评价云检测结果:①云识别准确率,ABCM 云检测与检验源均为云的像元数量与 ABCM 云检测判识为云的像元数量的比值;②云识别误判率,ABCM 云检测判识为晴空而检验源判识为云的像元数量与检验源为云的像元数量的比值;③晴空识别准确率,ABCM 云检测和检验源均为晴空像元数量与 ABCM 云检测判识为空空的像元数量比值;④晴空误判率,ABCM 云检测判识为云而检验源为晴空的像元数量与检验源为晴空的像元数量的比值;⑤总体云识别准确率,ABCM 云检测和检验源判识结果相同的像元数量与全部像元数量的比值。以 FY-4A/AGRI 太阳高度角是否大于  $75^\circ$  为原则,分白天、夜间、全天和 5 种地表类型分别进行统计,同时与中国气象局国家卫星气象中心(National Satellite Meteorological Center, NSMC)存档系统 FY-4A/AGRI 云检测结果比对。本文 ABCM 云检测的可能云归为云像元,可能晴空归为晴空。

### 3.1 与 CALIOP 交叉验证

选取 2021 年 1,4,7,10 月的 CALIOP 测试数据集为冬、春、夏、秋季验证月份,与本文 ABCM 云检测结果做交叉验证。表 1~表 4 是 2021 年 1,4,7,10 月 FY-4A/AGRI 云检测与 CALIOP 比对的参数评估结果。由表 1~表 4 可见,4 个季节中深海、浅水、陆地和荒漠上空的云识别准确率均在 93% 以上,平均云识别准确率高于 92%,云识别误判率低于 10%,平均云识别误判率低于 8%。积雪上空云识别准确率约为 80%,云误判率约为 20%,且冬季大于 20%。对日夜判识结果,ABCM 云检测方法在不使用可见光波段的条件下,白天也具有较好的云判识效果,除积雪外,其他云识别准确率均大于 90%,且白天与夜间的云识别准确率及云识别误判率基本一致,云识别准确率相差小于 1%。整体上,ABCM 云检测方法在不同季节云识别准确率均大于 90%,夏季的识别效果最好,云识别准确率最高为 91.8%,其次是秋季,为 90.9%,冬季的效果相对较差。与云判识准确率相比,晴空判识准确性略有降低,基本为 80%~90%,误判率约为 15%,而且积雪上空的晴空判识效果与其他地表类型上空的晴空判识精度无显著差异。以上表明 ABCM 云检测方法对云判识更为有利。

表 1 2021 年 1 月 FY-A/AGRI 与 CALIOP 比对

Table 1 Evaluated parameters of FY-4A/AGRI with reference of CALIOP in Jan 2021

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别 误判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	21572	92.8	3.1	87.5	25.5	91.7
	夜间	19024	93.4	9.6	86.3	9.6	90.4
	全天	40596	93.0	5.8	86.7	15.8	91.1
浅水	白天	18425	93.1	7.8	87.5	11.2	90.9
	夜间	22101	93.5	5.2	82.9	20.8	91.1
	全天	40526	93.3	6.2	85.6	15.3	91.0
陆地	白天	24158	93.3	8.7	85.5	11.4	90.5
	夜间	24007	93.1	8.4	86.9	10.9	90.6
	全天	48165	93.2	8.5	86.2	11.1	90.5
荒漠	白天	23880	93.2	8.9	82.8	13.5	89.6
	夜间	19541	93.2	8.3	81.7	15.1	89.6
	全天	43421	93.2	8.6	82.3	14.2	89.6
积雪	白天	4037	80.5	21.7	80.6	17.4	80.6
	夜间	4498	81.1	22.4	81.5	15.5	81.3
	全天	8535	80.8	22.1	81.1	16.4	80.9

表 2 2021 年 4 月 FY-4A/AGRI 与 CALIOP 比对

Table 2 Evaluated parameters of FY-4A/AGRI with reference of CALIOP in Apr 2021

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别误 判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	22136	93.6	3.9	89.1	17.1	92.4
	夜间	25761	93.0	5.2	84.2	20.4	90.9
	全天	47897	93.3	4.6	86.5	18.8	91.6
浅水	白天	20095	93.9	6.2	84.4	15.4	91.2
	夜间	19940	93.0	5.7	85.4	17.4	90.9
	全天	40035	93.5	5.9	84.9	16.4	91.1
陆地	白天	18896	93.5	6.0	87.0	14.1	91.5
	夜间	15410	93.2	7.8	87.8	10.7	91.1
	全天	34306	93.4	6.7	87.4	12.4	91.3
荒漠	白天	23317	93.1	9.2	82.8	13.2	89.4
	夜间	21424	93.2	8.6	83.6	13.3	89.8
	全天	44741	93.1	8.9	83.2	13.3	89.6
积雪	白天	4019	81.4	20.9	82.7	15.3	82.1
	夜间	4022	81.2	18.9	82.2	17.8	81.7
	全天	8041	81.3	19.9	82.5	16.5	81.9

表 3 2021 年 7 月 FY-4A/AGRI 与 CALIOP 比对

Table 3 Evaluated parameters of FY-4A/AGRI with reference of CALIOP in Jul 2021

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别 误判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	25200	94.7	4.3	86.1	16.5	92.7
	夜间	23527	95.2	4.8	86.3	13.5	92.9
	全天	48727	95.0	4.6	86.2	15.0	92.8
浅水	白天	28502	94.9	4.6	86.1	16.5	92.7
	夜间	21297	94.3	4.8	84.0	18.7	91.9
	全天	49799	94.6	4.7	85.2	16.8	92.4
陆地	白天	19857	93.9	5.8	86.6	14.0	91.7
	夜间	20576	93.5	4.4	90.0	14.3	92.4
	全天	40433	93.7	5.1	88.3	14.2	92.1
荒漠	白天	20837	93.2	6.4	87.0	13.8	91.1
	夜间	21477	93.7	4.5	90.4	13.1	92.7
	全天	42314	93.5	5.5	88.7	13.4	91.9
积雪	白天	4594	82.3	17.9	82.9	16.9	82.6
	夜间	4502	81.4	20.1	80.9	17.6	81.1
	全天	9096	81.8	19.0	81.9	17.2	81.9

表 4 2021 年 10 月 FY-4A/AGRI 与 CALIOP 比对

Table 4 Evaluated parameters of FY-4A/AGRI with reference of CALIOP in Oct 2021

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别 误判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识 别准确率/%
深海	白天	23669	94.0	5.5	87.2	13.7	92.0
	夜间	21668	93.4	4.6	87.6	17.4	91.8
	全天	45337	93.7	5.1	87.4	15.4	91.9
浅水	白天	22142	94.1	6.5	85.9	13.0	91.5
	夜间	19195	94.1	5.1	84.7	17.6	91.7
	全天	41337	94.1	5.8	85.4	14.9	91.6
陆地	白天	19861	93.5	5.7	86.8	14.9	91.5
	夜间	20131	93.4	4.6	87.6	17.1	91.9
	全天	39992	93.5	5.2	87.1	16.0	91.7

续表 4

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别 误判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识 别准确率/%
荒漠	白天	21569	93.6	7.2	85.0	13.5	90.8
	夜间	23935	92.7	10.5	85.1	10.5	89.5
	全天	45504	93.2	8.8	85.1	11.8	90.1
积雪	白天	4695	81.9	17.2	83.1	17.8	82.5
	夜间	4088	82.6	19.8	81.1	16.6	81.8
	全天	8783	82.2	18.4	82.1	17.2	82.2

3.2 与 MODIS 交叉验证

选取 2021 年 10 月和 2022 年 1, 4, 7 月的 MOD35\_L2 数据作为秋、冬、春、夏季验证月份,与本文 ABCM 云检测结果进行交叉验证。表 5~表 8 是 2021 年 10 月和 2022 年 1, 4, 7 月 FY-4A/AGRI 云检测与 MODIS 比对的参数评估结果。由表 5~表 8 可见,从地表类型看,本文方法在深海、浅水上云判识效果最好,云识别准确率均在 88%以上。秋季深海云识别准确率最高可达 97.5%,夏季浅水云识别准确率最高可达 92.0%,云识别准确率均在 86%以上。4 个季节深海的云识别误判率均低于 3%,浅水云识别误判率为 5%~10%。相比而言,积雪上空的云判识效果较差,平均云识别准确率约

为 85%,云识别误判率为 10%~20%。从日夜标识看,白天、夜间的总体云识别准确率分别在 84%, 82%以上,同一地表类型的云识别准确率和误判率基本一致。整体上,本文方法在 4 个季节的云识别准确率均在 86%以上,夏季识别效果最好,云识别准确率达到 90%,除积雪外,其他地表类型的云识别误判率均在 8%以下,晴空判识准确性基本高于 84%;秋季的云识别准确率为 88%,较夏季的晴空判识准确性偏低。冬春两季的识别效果相对较差,云识别准确率约为 86%。以上结果表明:本研究的云检测算法对于不同地表类型和不同季节的云判识效果无显著差异。

表 5 2021 年 10 月 FY-4A/AGRI 与 MODIS 比对  
Table 5 Evaluated parameters of FY-4A/AGRI with reference of MODIS in Oct 2021

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别误 判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	144057	97.5	1.8	84.1	20.1	96.1
	夜间	97999	96.8	1.7	82.4	28.5	95.6
	全天	242056	97.2	1.8	83.4	23.9	95.9
浅水	白天	16167	91.5	5.5	88.9	16.8	90.6
	夜间	10584	86.3	5.6	89.0	24.9	87.2
	全天	26751	88.9	5.5	88.9	20.2	89.3
陆地	白天	220165	88.7	5.6	83.8	29.1	87.5
	夜间	149421	85.0	8.4	81.2	30.7	83.9
	全天	369586	87.3	6.7	82.6	29.8	86.1
荒漠	白天	11598	88.2	8.8	82.3	22.9	86.2
	夜间	8604	90.6	9.1	70.6	30.1	85.9
	全天	20202	89.3	9.0	78.2	25.4	86.1
积雪	白天	157873	87.1	9.8	80.4	24.9	84.9
	夜间	139159	83.3	6.6	82.3	37.7	83.1
	全天	297032	85.2	8.3	81.2	30.7	84.1

表 6 2022 年 1 月 FY-4A/AGRI 与 MODIS 比对  
Table 6 Evaluated parameters of FY-4A/AGRI with reference of MODIS in Jan 2022

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别误 判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	503820	90.8	0.8	89.7	57.4	90.8
	夜间	333403	88.0	1.6	87.1	55.5	87.9
	全天	837223	89.7	1.1	88.5	56.5	89.6

续表 6

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别误 判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
浅水	白天	33987	90.3	12.4	84.3	12.5	87.6
	夜间	35916	87.4	16.0	81.2	14.9	84.5
	全天	69903	88.8	14.2	82.7	13.7	86.0
陆地	白天	174214	89.1	11.3	72.8	26.3	84.3
	夜间	179412	83.9	13.8	75.0	28.6	80.8
	全天	353626	86.7	12.5	74.0	27.6	82.5
荒漠	白天	102027	88.9	6.7	75.8	35.9	86.1
	夜间	95487	83.7	7.0	69.4	53.5	81.3
	全天	197514	86.3	6.8	73.0	44.5	83.8
积雪	白天	58203	86.4	15.3	75.5	22.0	82.2
	夜间	49709	79.8	24.3	78.4	17.9	79.0
	全天	107912	83.8	18.9	77.1	19.8	80.7

表 7 2022 年 4 月 FY-4A/AGRI 与 MODIS 比对

Table 7 Evaluated parameters of FY-4A/AGRI with reference of MODIS in Apr 2022

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别 误判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	301247	91.6	2.6	89.9	27.8	91.3
	夜间	395136	89.1	3.3	94.7	16.6	91.1
	全天	696383	90.3	3.0	93.3	20.1	91.2
浅水	白天	37351	89.3	10.4	87.1	13.2	88.3
	夜间	33876	89.1	15.9	81.7	12.7	85.6
	全天	71227	89.2	13.0	84.6	13.0	87.0
陆地	白天	557695	89.1	17.0	80.3	12.8	84.8
	夜间	760995	88.2	21.4	80.1	11.0	83.7
	全天	1318690	88.6	19.5	80.2	11.7	84.2
荒漠	白天	112893	86.1	20.1	92.3	5.1	90.7
	夜间	95256	84.9	18.6	92.8	5.8	90.5
	全天	208149	85.5	19.4	92.6	5.4	90.6
积雪	白天	424781	86.4	7.3	80.1	32.5	84.9
	夜间	361103	85.1	10.6	78.8	28.4	83.1
	全天	785884	85.7	8.8	79.8	30.5	84.1

表 8 2022 年 7 月 FY-4A/AGRI 与 MODIS 比对

Table 8 Evaluated parameters of FY-4A/AGRI with reference of MODIS in Jul 2022

地表类型	日夜标识	像元数量	云识别 准确率/%	云识别误 判率/%	晴空识别 准确率/%	晴空误 判率/%	总体云识别 准确率/%
深海	白天	569426	92.9	2.3	91.9	22.4	92.7
	夜间	532286	91.4	3.0	92.0	20.7	91.6
	全天	1101712	92.2	2.6	92.0	21.5	92.1
浅水	白天	52739	92.0	3.7	85.6	26.8	90.7
	夜间	51601	92.7	4.6	84.5	22.9	90.9
	全天	104340	92.3	4.2	85.0	24.8	90.8
陆地	白天	445400	87.4	6.7	92.4	14.0	89.7
	夜间	473079	85.7	7.6	93.6	12.2	89.8
	全天	918479	86.6	7.1	93.1	13.0	89.8
荒漠	白天	33485	87.3	1.3	97.8	19.4	90.9
	夜间	29325	84.2	2.5	96.5	20.6	89.1
	全天	62810	85.9	1.8	97.2	20.0	90.0
积雪	白天	46343	90.3	9.9	81.5	18.1	87.2
	夜间	35481	91.6	13.4	76.8	15.2	86.0
	全天	81824	90.8	11.4	79.4	16.9	86.7

### 3.3 与 NSMC 云检测验证

图 3 为 2022 年 9 月 15 日 10:00(世界时,下同) FY-4A/AGRI 云检测结果及云图。由图 3 可见,观测域内有部分区域无可见光观测数据,对比彩色合成图,可发现明显的晨昏分界线。在 ABCM 云检测结果图像上,晨昏过渡区云检测结果连续,几乎没有晨昏交界线,与红外通道图的云分布基本一致,比

NSMC 云检测表现更好。整体上,ABCM 云检测与 NSMC 云检测的云分布大致相同,ABCM 云检测判识的晴空像元比业务云检测多,可能晴空像元较少。对比同时次 10.7  $\mu\text{m}$  红外通道图,在中亚、印度和印度洋区域 ABCM 云检测表现更好。夜间部分 NSMC 云检测算法在日本附近出现云过判,ABCM 云检测更为准确。

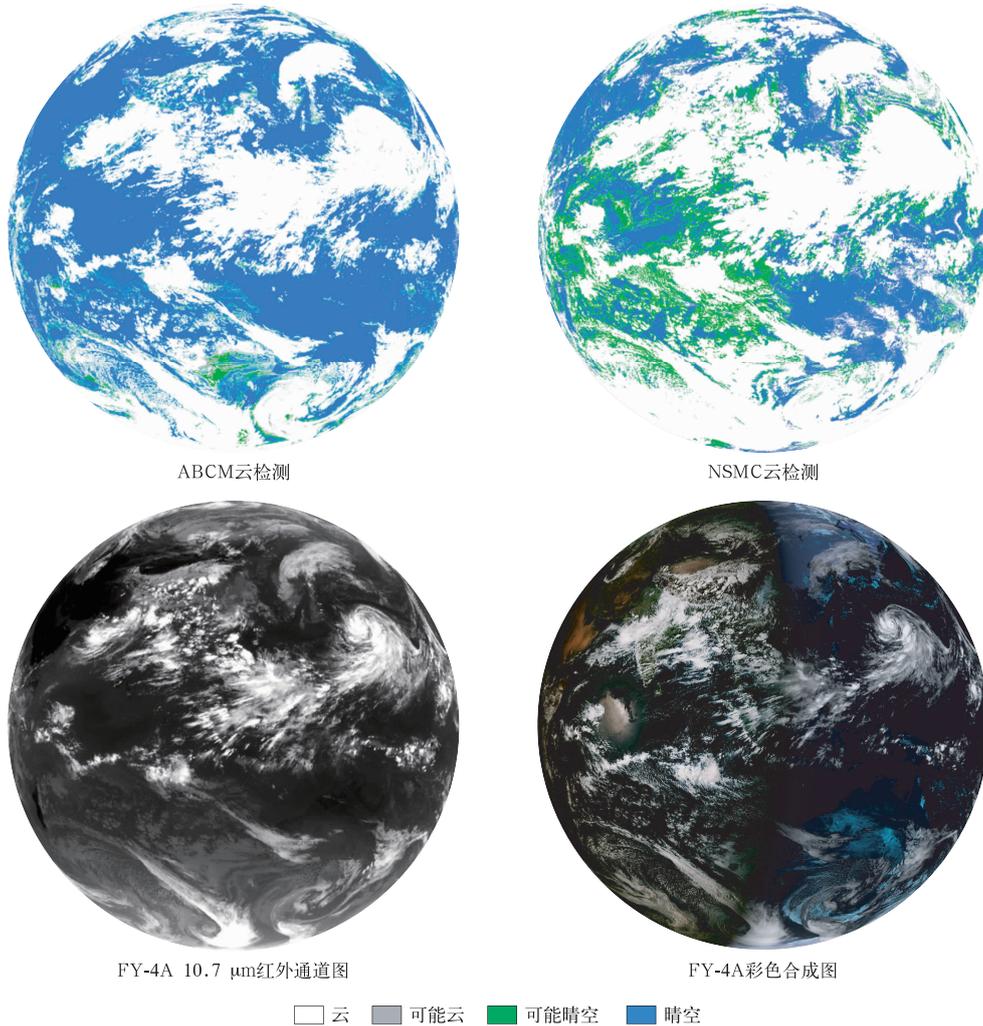


图 3 2022 年 9 月 15 日 10:00 云检测结果及云图

Fig. 3 Cloud detection results and cloud images at 1000 UTC 15 Sep 2022

## 4 结论与讨论

本文使用 FY-4A/AGRI 红外通道数据,结合 CALIPSO/CALIOP 2 级 1 km 云层产品,在时空匹配的基础上,构建基于朴素贝叶斯原理的云检测分类器,并采用 4 分类方法,提出基于机器学习的 FY-

4A/AGRI 云检测方法,并与 CALIOP 和 MODIS 云检测数据开展交叉验证,得到主要结论如下:

- 1) 基于 FY-4A/AGRI 红外通道构建多种云检测分类器,避免太阳辐射的直接影响,在不选用可见光波段的条件下,白天、夜间均具有良好的云判识效果,不仅可保证日夜云检测精度,也可保证日夜云检测的连续性。

2) ABCM 云检测方法在夏秋两季的云识别效果较好,其次是春季,冬季的云识别效果较差,且对低云存在漏判现象。

3) 训练数据集的样本主要集中在深海,其次是陆地、浅水和荒漠,积雪地表类型的样本量最少。因此对深海、浅水的云识别准确率普遍较高,有植被覆盖的陆地和荒漠次之,积雪类型的判识不稳定,相对较差。

ABCM 云检测方法不仅可得到云、可能云、可能晴空和晴空 4 种云检测分类结果,还可获得每个云检测分类器的不确定性概率值,可为云和地表相关检测产品提供重要参考。该方法不依赖外部相关数据,降低云检测算法的复杂程度,提高算法执行效率,可应用 FY-4B/AGRI, FY-3D/MERSI 和 FY-3E/MERSI 载荷,尤其 FY-3E/MERSI 载荷仅有红外通道。

本文主要基于 FY-4A/AGRI 红外通道构建云检测分类器,今后将考虑加入其他波段、辅助数据(如海温)或辐射传输模式以改善分类器精度,提高云识别能力。同时,结合其他云观测数据,通过统计分析得到各分类器用于判识云或晴空的可靠阈值。

### 参考文献

- [1] 林晔. 大气探测学教程. 北京:气象出版社,1995.  
Lin Y. Course of Atmospheric Exploration. Beijing: China Meteorological Press, 1995.
- [2] 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. 应用气象学报, 2007, 18(2): 158-164.  
Liu J, Zhang W J, Zhu Y J, et al. Case study on cloud properties of heavy rainfall based upon satellite data. *J Appl Meteor*, 2007, 18(2): 158-164.
- [3] Zhang Y C, Rossow W B, Lacis A A, et al. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J Geophys Res*, 2004, 109 (D19): D19105.
- [4] Rossow W B, Walker A W, Beusichel D E, et al. International Satellite Cloud Climatology Project (ISCCP) Documentation of New Cloud Datasets. WMO TD-No. 737, 1996: 1-115.
- [5] Rossow W B, Walker A W, Gander L C. Comparison of ISCCP and other cloud amounts. *J Climate*, 1993, 6: 2394-2418.
- [6] 肖海霞,张峰,王亚强,等. 基于生成对抗网络和卫星数据的云图临近预报. 应用气象学报, 2023, 34(2): 220-233.  
Xiao H X, Zhang F, Wang Y Q, et al. Nowcasting of cloud images based on generative adversarial network and satellite data. *J Appl Meteor Sci*, 2023, 34(2): 220-233.
- [7] 高洋,蔡森,曹治强,等. “21·7”河南暴雨环境场及云的宏观特征. 应用气象学报, 2022, 33(6): 682-695.  
Gao Y, Cai M, Cao Z Q, et al. Environmental conditions and cloud macro and micro features of “21·7” extreme heavy rainfall in Henan Province. *J Appl Meteor Sci*, 2022, 33(6): 682-695.
- [8] Rossow W B, Schiffer R A. ISCCP cloud data products. *Bull Amer Meteor Soc*, 1991, 72(1): 2-20.
- [9] 马瑞阳,郑栋,姚雯,等. 雷暴云特征数据集及我国雷暴活动特征. 应用气象学报, 2021, 32(3): 358-369.  
Ma R Y, Zheng D, Yao W, et al. Thunderstorm feature dataset and characteristics of thunderstorm activities in China. *J Appl Meteor Sci*, 2021, 32(3): 358-369.
- [10] 翁理国,张旭,夏曼,等. 关于卫星气象云图准确检测的仿真研究. 计算机仿真, 2019, 36(1): 429-436.  
Weng L G, Zhang X, Xia M, et al. Simulation study on accurate detection of satellite meteorological cloud images. *Computer Simulation*, 2019, 36(1): 429-436.
- [11] 任素玲,牛宁,覃丹宇,等. 2021年2月北美极端低温暴雪的卫星遥感监测. 应用气象学报, 2022, 33(6): 696-710.  
Ren S L, Niu N, Qin D Y, et al. Extreme cold and snowstorm event in North America in February 2021 based on satellite data. *J Appl Meteor Sci*, 2022, 33(6): 696-710.
- [12] 任素玲,方翔,卢乃锰,等. 基于气象卫星的青藏高原低涡识别. 应用气象学报, 2019, 30(3): 345-359.  
Ren S L, Fang X, Lu N M, et al. Recognition method of the Tibetan Plateau vortex based on meteorological satellite data. *J Appl Meteor Sci*, 2019, 30(3): 345-359.
- [13] 杨军. 气象卫星及其应用. 北京:气象出版社, 2012.  
Yang J. Meteorological Satellites and Their Applications. Beijing: China Meteorological Press, 2012.
- [14] 任强,董佩明,薛纪善. 台风数值预报中受云影响微波卫星资料的同化试验. 应用气象学报, 2009, 20(2): 137-146.  
Ren Q, Dong P M, Xue J S. The use of microwave satellite data affected by cloud in numerical forecast of typhoon. *J Appl Meteor Sci*, 2009, 20(2): 137-146.
- [15] Seze G, Rossow W B. Time-cumulated visible and infrared radiance histograms used as descriptors of surface and cloud variations. *Int J Remote Sens*, 1991, 12(5): 877-920.
- [16] 刘希,许健民,杜秉玉. 用双通道动态阈值对GMS-5图像进行自动云检测. 应用气象学报, 2005, 16(4): 434-444.  
Liu X, Xu J M, Du B Y. A bi-channel dynamic threshold algorithm used in automatically identifying clouds on GMS-5 imagery. *J Appl Meteor*, 2005, 16(4): 434-444.
- [17] Wylie D, Jackson D L, Menzel W P, et al. Trends in global cloud cover in two decades of HIRS observations. *J Climate*, 2005, 18(15): 3021-3031.
- [18] Rossow W B, Garder L C. Cloud detection using satellite measurements of infrared and visible radiances for ISCCP. *J Climate*, 1993, 6(12): 2341-2369.
- [19] 刘健. 中国区域云特性分析及在FY-2云检测中的应用. 应用气象学报, 2009, 20(6): 673-681.

- Liu J. Cloud properties analysis and its application in FY-2 cloud detection. *J Appl Meteor*, 2009, 20(6): 673-681.
- [20] 武艳, 银燕, 师春香, 等. 基于动态阈值法的 NOAA 系列卫星云检测结果检验. *高原气象*, 2012, 31(3): 745-751.
- Wu Y, Yin Y, Shi C X, et al. Validation of NOAA/AVHRR cloud detections by an automated dynamic threshold cloud-masking algorithm. *Plateau Meteor*, 2012, 31(3): 745-751.
- [21] 鄢俊洁, 郭雪星, 瞿建华, 等. 基于朴素贝叶斯方法的 FY-4A/AGRI 云检测模型. *自然资源遥感*, 2022, 34(3): 33-42.
- Yan J J, Guo X X, Qu J H, et al. FY-4A/AGRI cloud detection model based on naive Bayesian algorithm. *Sensing for Natural Resources*, 2022, 34(3): 33-42.
- [22] Liu C, Yang S, Di D, et al. A Machine learning-based cloud detection algorithm for the Himawari-8 spectral image. *Adv Atmos Sci*, 2022, 39: 1994-2007.
- [23] Amato U, Antoniadis A, Cuomo V, et al. Statistical cloud detection from SEVIRI multispectral images. *Remote Sens Environ*, 2008, 112(3): 750-766.
- [24] Gomez-Chova L, Camps-Valls G, Amoros-Lopez J, et al. New Cloud Detection Algorithm for Multispectral and Hyperspectral Images; Application to ENVISAT/MERIS and PROBA/CHRIS Sensors//IEEE International Conference on Geoscience and Remote Sensing Symposium (GRSS), 2007: 2757-2760.
- [25] Hughes M, Daniel H. Automated detection of cloud and cloud shadow in single-date landsat imagery using neural networks and spatial post-processing. *Remote Sensing*, 2014, 6(6): 4907-4926.
- [26] 瞿建华, 鄢俊洁, 薛娟, 等. 基于深度学习的 FY3D/MERSI 和 EOS/MODIS 云检测模型研究. *气象与环境学报*, 2019, 35(3): 87-93.
- Qu J H, Yan J J, Xue J, et al. Research on FY3D/MERSI and EOS/MODIS cloud detection models based on deep learning. *J Meteor Environ*, 2019, 35(3): 87-93.
- [27] Francis A, Sidiropoulos P, Muller J P. CloudFCN: Accurate and robust cloud detection for satellite imagery with deep learning. *Remote Sensing*, 2019, 11(19): 2312.
- [28] 康晓光, 孙龙祥. 基于人工神经网络的云自动检测算法. *解放军理工大学学报(自然科学版)*, 2005, 6(5): 506-510.
- Kang X G, Sun L X. Cloud automatic detection algorithm based on artificial neural network. *Journal of PLA University of Science and Technology(Nat Sci Ed)*, 2005, 6(5): 506-510.
- [29] 靳泽群, 张玲, 刘神聪, 等. 基于 BP 神经网络的云检测和云相态识别. *光学与光电技术*, 2016, 14(5): 74-77.
- Jin Z Q, Zhang L, Liu S C, et al. Cloud detection and cloud phase recognition based on BP neural network. *Optical & Optical Technology*, 2016, 14(5): 74-77.
- [30] Merchant C J, Harris A R, Maturi E, et al. Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quart J Roy Meteor Soc*, 2010, 131(611): 2735-2755.
- [31] Heidinger A K, Evan A T, Foster M J, et al. A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *J Appl Meteor Climatol*, 2012, 51(6): 1129-1144.
- [32] Frey R A, Ackerman S A, Liu Y H, et al. Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for collection 5. *J Atmos Ocean Technol*, 2008, 25(7): 1057-1072.
- [33] Yang J, Zhang Z, Wei C, et al. Introducing the new generation of Chinese geostationary weather satellites Fengyun-4. *Bull Amer Meteor Soc*, 2017, 98(8): 1637-1658.
- [34] Wang X, Min M, Wang F, et al. Intercomparisons of cloud mask products among Fengyun-4A, Himawari-8, and MODIS. *IEEE Trans Geosci Remote Sens*, 2019, 57(11): 8827-8839.
- [35] Zhuge X Y, Zou X, Wang Y. A fast cloud detection algorithm applicable to monitoring and nowcasting of daytime cloud systems. *IEEE Trans Geosci Remote Sens*, 2017, 55(11): 6111-6119.
- [36] 王曦, 刘健, 杨冰韵. 基于 FY-3D/MERSI-II 红外波段夏季北极云检测模型研究. *红外与毫米波学报*, 2022, 41(2): 483-492.
- Wang X, Liu J, Yang B Y. Research on summer Arctic cloud detection model based on FY-3D/MERSI-II infrared band. *J Infrared Millim Waves*, 2022, 41(2): 483-492.
- [37] Seemann S W, Borbas E E, Knuteson R O, et al. Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multi-spectral satellite radiance measurements. *J Appl Meteor Climatol*, 2007, 47(1): 108-123.
- [38] 马晓静, 邓志辉, 陈梅花, 等. 中国大陆卫星红外亮温与海拔高度关系的研究. *地震地质*, 2008, 30(2): 562-572.
- Ma X J, Deng Z H, Chen M H, et al. Study on the relationship between the satellite infrared brightness temperature and ground elevation in the mainland of China. *Seismology and Geology*, 2008, 30(2): 562-572.

## FY-4A/AGRI Cloud Detection Method Based on Naive Bayesian Algorithm

Guo Xuexing Qu Jianhua Ye Lingmeng Han Min Shi Mojie

(Beijing Huayun Shinetek Science and Technology Co., Ltd., Beijing 100081)

### Abstract

Optical remote sensing cloud detection is the foundation for subsequent quantitative remote sensing and applications. A cloud detection method based on naive Bayesian algorithm is studied and applied to the advanced geostationary orbital radioimager (AGRI) on Fengyun-4A satellite. Cloud detection method considering radiation physics of visible light channels is discontinuous between day and night. To avoid the direct impact of solar radiation, the spectral data of 7 infrared channels loaded by AGRI are analyzed to construct 10 cloud detection feature classifiers. Using cloud polarized lidar with orthogonal polarization (CALIOP) data as the true value of cloud detection, and using its spatiotemporal matching data with AGRI, classification training and validation are conducted for datasets of different surface types and different seasons. The cloud detection results and CALIOP data cross-verification show that the cloud recognition accuracy over snow is about 81%, the cloud recognition accuracy rate over the deep sea, shallow water, land and desert is higher than 92%, the false positive rate is basically less than 10%, and the overall cloud recognition accuracy reaches 90%. Compared with MODIS level 2 cloud detection products in October of 2021 and January, April and July of 2022, the recognition accuracy rate of deep-sea and shallow water clouds is above 88%, and the false positive rate is lower than 3% and 10%, respectively. The overall cloud recognition accuracy rate in four seasons is more than 86%, of which the summer cloud recognition effect is the best, and the overall cloud recognition accuracy rate is as high as 90%. The recognition effects of the method are good during both day and night, ensuring not only the accuracy of day and night cloud detection, but also the continuity of cloud detection in the morning and evening transition zone. Due to the use of dynamic surface type files and sufficient training sample sizes for deep and shallow waters, the overall cloud recognition accuracy of the method is relatively ideal in four seasons, with the best performance in summer and autumn. The cloud recognition accuracy of deep and shallow water is generally high, but there are still omissions and misjudgments. The method can output classification results of cloud including probable cloud, probable clear sky, and clear sky, and it also outputs the uncertainty probability value of each feature and a comprehensive feature cloud detection classifier, which can provide important reference for cloud and surface related detection products.

**Key words:** FY-4A/AGRI; naive Bayesian algorithm; cloud detection