

基于随机森林算法的 FY-4A 云底高度估计方法

谭仲辉¹, 马烁^{1*}, 韩丁², 高顶³, 严卫¹

(1. 国防科技大学气象海洋学院, 江苏 南京 210000;

2. 中国人民解放军 96901 部队, 北京 100000;

3. 中国人民解放军 61175 部队, 江苏 南京 210000)

摘要:基于 2017 年 8 月至 10 月 FY-4A 的云顶高度、云光学性质等上游产品和 A-Train 系列卫星星载毫米波雷达和激光雷达主动探测的云底高度资料, 利用随机森林算法建模, 提出了 FY-4A 对最上层云云底高度的估计算法, 并用 2017 年 11 月独立样本对算法进行了检验与评估. 结果表明, 该算法可以有效实现对最上层云云底高度的估计, 与星载主动探测结果相比, 平均绝对偏差为 1.29 km, 相关系数为 0.80. 对单层云的估计结果相对较好, 而多层云存在时云底高度的估计结果一般偏小.

关键词:卫星遥感; 云底高度; 随机森林; FY-4A

中图分类号:P4 **文献标识码:**A

Estimation of cloud base height for FY-4A satellite based on random forest algorithm

TAN Zhong-Hui¹, MA Shuo^{1*}, HAN Ding², GAO Ding¹, YAN Wei¹

(1. National University of Defense Technology, College of Meteorology and Oceanography, Nanjing 210000, China;

2. PLA 96901, Beijing 100000, China;

3. PLA 61175, Nanjing 210000, China)

Abstract: Based on upstream products of FY-4A and A-Train satellites data during August and October, 2017, an estimation algorithm of cloud base height for FY-4A has been presented utilizing Random Forest model. The algorithm is evaluated in the comparison with CloudSat and CALIPSO. The results show that cloud base height for top layer cloud can be generated by using upstream products of FY-4A. Compared with CloudSat and CALIPSO, the mean absolute error is less than 1km and the relationship coefficient is bigger than 0.8. The presence of multi-layer clouds may result in underestimate of cloud base height.

Key words: satellite remote sensing, cloud base height, Random Forest, FY-4A

PACS: 92. 60. Wc, 92. 60. Ox

引言

云是大气中水滴、过冷水滴、冰晶或其混合物组成的可见聚集体, 能显著影响地球的辐射传输和水汽循环. 云边界高度尤其是云底高度代表了云的垂直分布, 不仅在大气环流、辐射强迫、云结构的分析等研究领域起重要作用, 而且是数值天气预报模

式的基础参数. 因此, 准确获取云底高度信息对于定量描述天气系统辐射收支、研究气候变化以及保障飞行安全具有重要意义^[1].

随着气象卫星探测技术的发展, 卫星遥感逐渐成为业务上获取全球云参数的主要手段. 卫星遥感云底高度可分为主动遥感和被动遥感. 主动遥感如美国 A-Train 系列卫星中的 CloudSat 和 CALIPSO, 利

收稿日期: 2019-01-06, 修回日期: 2019-01-22

基金项目: 国家自然科学基金(41705007, 41575028)

Foundation items: Supported by the National Natural Science Foundation of China (41705007, 41575028)

作者简介 (Biography): 谭仲辉 (1995-), 男, 湖北孝感人, 硕士研究生, 主要研究领域为卫星遥感云物理参数. E-mail: tzh_go@126.com

* 通讯作者 (Corresponding author): E-mail: mashuo0601@163.com

Received date: 2019-01-06, revised date: 2019-01-22

用星载毫米波雷达和激光雷达,能获取全球云廓线信息,但都只能沿轨进行垂直探测,覆盖区域十分有限^[2];严卫等基于相对湿度阈值法,利用 COSMIC 掩星湿空气数据反演全球的云底高度^[3],并与探空仪、毫米波雷达反演结果进行了对比分析,发现掩星的云底高度反演结果除高层云和卷云外普遍偏小,且对低云的反演误差较大^[4]。

被动遥感器如 MODIS (Moderate Resolution Imaging Spectroradiometer)、VIIRS (Visible Infrared Imaging Radiometer Suite) 等多通道辐射成像仪,已经有很多方法对云顶高度 (CTH) 进行有效的反演^[5-7],但星载多通道辐射成像仪只能获取云顶的可见光通道反射率、红外通道亮温以及垂直空间的水汽总含量等信息,无法准确得到云底信息,因此对云底高度的反演仍然是一个难题^[8]。孙学金等基于光谱匹配,将 CloudSat/CALIPSO 的云底高探测结果外推至 MODIS 视场中,得到了三维云结构信息,但只能运用于主被动探测器在同一视场内的情况^[9];Hutchison 等基于 VIIRS 的被动遥感数据,假定各类云的云水含量为固定值,利用云光学厚度、云液态水路径等产品,通过统计分析确定了估计云几何厚度的经验公式,再由云顶高度减去云几何厚度得到云底高度,实现了 VIIRS 视场内的云底高度估计,但结果非常依赖于对云水含量的假定^[10]。但基于卫星资料很难准确识别云的类型,且即使是同一类云,其云水含量也会有显著差异,所以根据经验值确定的云水含量会给云底高度估计带来较大的不确定性^[8]。

Breiman 等建立的随机森林 (Random Forest, RF) 算法^[11]是一种用于分类和回归的集成学习方法,适合处理预测值与非线性输入间的复杂关系^[12]。随机森林算法已经在生态领域得到广泛应用^[13-14],现在也越来越多的用于卫星遥感领域^[15-16]。Mutanga 等将随机森林算法用于估算湿地植被的高密度生物量,结果优于传统的多元线性回归法^[17]。

我国第二代静止轨道气象卫星风云四号系列首颗科学实验卫星 FY-4A 于 2016 年 12 月 11 日发射成功^[18],并于 2018 年 5 月 1 日正式投入业务运行。FY-4A 搭载的可见光红外成像仪 AGRI (Advanced Geostationary Radiation Imager) 拥有 14 个光谱通道,15min 能进行一次全圆盘探测,通道信息如表 1 所示。由于静止轨道卫星的优势,FY-4A 能对约三分之一地球表面进行连续探测。基于 FY-4A 高时空分辨率的探测数据,中国气象局国家卫星气象中心开发

出云检测、云分类、云顶高度、云光学性质等多种业务产品,为云底高度的估计提供了有利条件。

表 1 FY-4A/AGRI 通道特性

Table 1 FY-4A/AGRI Channel Specifications

通道	波长 (μm)	空间分辨率 (km)	主要应用
1	0.45 - 0.49	1	气溶胶
2	0.55 - 0.75	0.5	云,雾
3	0.75 - 0.90	1	植被
4	1.36 - 1.39	2	卷云
5	1.58 - 1.64	2	云,雪
6	2.1 - 2.35	2	云,气溶胶
7	3.5 - 4.0 (high)	2	地表特征
8	3.5 - 4.0 (low)	4	
9	5.8 - 6.7	4	水汽
10	6.9 - 7.3	4	水汽
11	8.0 - 9.0	4	水汽,云
12	10.3 - 11.3	4	海表温度,云
13	11.5 - 12.5	4	海表温度,云
14	13.2 - 13.8	4	云

基于 FY-4A 的上游产品和 A-Train 卫星数据,利用随机森林算法建模,避免了对云水含量的假设,增加了对地理特征的考虑,提出新的云底高度估计算法,实现了基于 FY-4A 数据对最上层云的云底高度估计。第一章介绍了本文用到的数据,第二章分析了新算法的理论依据,并详细描述了基于随机森林的 FY-4A 云底高度估计算法,第三章对算法进行了应用和误差分析,同时选取了一个典型个例进行研究,第四章对全文做了总结。

1 数据

国家卫星气象中心自 2018 年 5 月 8 日起陆续公布 FY-4A 产品,基于 AGRI 探测数据提供的产品主要有两种,一种是经过质量检验、地理定位以及辐射定标后的 L1 产品,一种是基于 L1 数据和辅助数据开发的 L2 产品,表 2 列举了本文用到的 8 种产品,表中产品的空间分辨率均为 $4 \text{ km} \times 4 \text{ km}$ 。

表 2 FY-4A/AGRI 产品

Table 2 FY-4A/AGRI Productions used in the algorithm

名称	产品信息	主要物理量
CLM	云检测产品	0:有云;1:可能有云;2:可能晴空;3:晴空
CTH	云顶高度产品	云顶高度,km
CPD/ CPN	白天/夜间云光学 和微物理性质	云光学厚度,云有效粒子半径,云水(冰)路径
CLT	云类型产品	Water, supercool, mixed, ice, cirrus, overlap 六类云
CTT	云顶温度产品	云顶温度,K

CloudSat 数据处理中心 (Data Processing Center,

DPC)提供的标准数据产品中,一级产品为 CPR 的直接测量结果,二级产品是根据一级产品和其他卫星辅助数据联合处理后的结果,本文用到的 2B-GEOPROF 产品主要包含云顶高度、云底高度、MODIS 云量、CPR 掩膜以及海陆标志等信息。CALIPSO 卫星搭载的激光雷达 CALIOP 对薄卷云更加敏感,但对不透明云探测时衰减较大^[19]。本文用到的 CALIPSO 数据为 L2-1 km-Cloud-Layer 产品,包含 1 km 水平分辨率的云顶高度、云底高度以及云层结构等信息。需要指出的是,CloudSat 与 CALIPSO 原本在 A-Train 系列中组成编队飞行,两者的联合探测结果提供更加完整、准确的云廓线信息,但在 2011 年 CloudSat 机械故障之后不再与 CALIPSO 提供联合产品。

将 FY-4A 产品分别与 CloudSat 和 CALIPSO 数据进行精确的时空匹配,数据匹配时要求时间间隔不超过 5 min,空间间隔不超过 2 km。基于 2017 年 8 月至 10 月的数据,最终得到与 CloudSat 匹配的 159686 组训练样本,与 CALIPSO 匹配的 178542 组训练样本,同时,基于 11 月的数据分别得到 85986 组、96139 组测试样本。

2 算法描述

2.1 理论依据

随机森林以决策树为基学习器构建 Bagging 集成,并进一步在决策树的训练过程中引入了随机属性的选择,通过样本扰动和属性扰动增强了最终集成的泛化能力,相比传统的神经网络算法计算更快,抗噪声能力更强。

每颗决策树包含一个根节点、若干个内部节点和若干个叶节点,叶节点对应于决策结果,从根节点到每个叶节点的路径对应了一个判定测试序列。用于回归时,通过最小均方差法判定选取的特征值和特征值划分点。

Bagging 是一种并行式集成学习方法,基于自助采样法(Bootstrap Sampling),对采样集进行随机、放回的采样,得到多个不同的采样集。然后基于每个采样集训练一个决策树,再将这些决策树的结果进行结合,回归问题中通常使用简单平均法。

在云底高度研究中,云水含量与云顶高度、云光学厚度、云有效粒子半径、云水路径以及地理特征等信息有复杂且非线性的关系,通过随机森林建模,可以避免对云水含量的假设,同时可以引入经度、纬度包含的地理特征等更多影响因子,增强对云底高度

的估计能力。本文以 FY-4A 上游业务产品为输入特征,以 CloudSat 和 CALIPSO 卫星探测的云底高度为目标输出,对模型进行训练,然后建立基于 FY-4A 自身产品对最顶层云云底高度的估计算法。

2.2 算法框架

FY-4A 对最顶层云顶高度估计算法的原理图如图 1 所示,主要分为随机森林算法训练、云底高度估计算法两部分。

2.2.1 随机森林算法训练

对随机森林算法进行训练首先要建立包含输入特征和目标输出的训练集。本文基于 FY-4A 与 CloudSat 和 CALIPSO 在 2017 年 8 月至 10 月的匹配数据建立训练集。首先进行数据预处理:先进行云检测,然后基于 FY-4A 云分类产品,将有云样本分为 6 类;

针对水云(water)、过冷云(supercool)和混合云(mixed),由于毫米波雷达对非透明云探测能力较强,以云顶高度(CTH)、云光学厚度(COT)、云有效粒子半径(CER)、云液态水路径(LWP)以及样本点的经纬度作为输入特征,以 CloudSat 最顶层云的云底高度作为目标输出,对每种云分别建立样本集。针对冰云(ice)、卷云(cirrus)和重叠云(overlap),考虑到激光雷达对光学厚度较小的冰云敏感性更好,以 CALIPSO 最顶层云的云底高度作为目标输出。

建立训练集后,对 6 类云分别进行随机森林的训练,并将训练过的模型输入到云底高度估计算法中。

2.2.2 云底高度估计算法

如图 2 所示,FY-4A 仅依赖自身上游产品和训练过的随机森林算法即可得到对最上层云云底高度的估计。算法的主要步骤如下:

A. 输入上游产品。读取 FY-4A 的云检测、云分类、云顶高度以及云光学和微物理性质等产品;

B. 云检测。对数据进行云检测,只针对有云样本点进行云底高度估计;

C. 云分类。将云分为 6 类,对每类云提取相应的输入数据;

D. 输入随机森林算法。将每类云的输入数据代入随机森林算法,得到云底高度估计结果;

E. 质量控制。对估计的云底高度进行筛选,剔除无效值和异常值;

F. 输出。输出 FY-4A 对最上层云云底高度的估计结果。

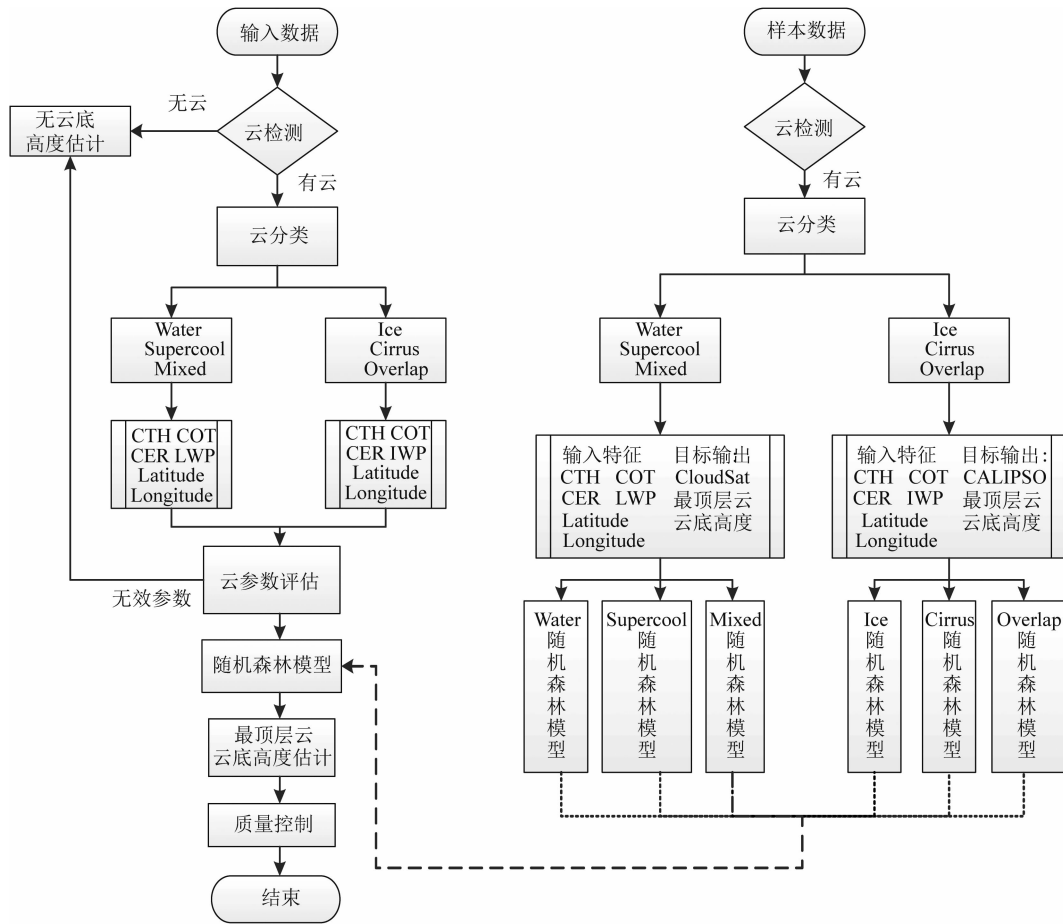


图 1 FY-4A 对最上层云云底高度估计原理图
Fig. 1 Diagram of estimation of top layer cloud CBH for FY-4A

3 算法应用与误差分析

为检验算法的有效性,利用 2017 年 8 月至 10 月的 FY-4A 与 A-Train 卫星数据,对随机森林算法进行了训练,然后将训练过的模型应用至 11 月的 FY-4A 上游业务产品,得到云底高度的估计结果,并与 CloudSat 和 CALIPSO 卫星探测的云底高度进行对比检验。

3.1 计算结果

图 2 所示为云底高度对比结果,横坐标表示 CloudSat 或 CALIPSO 探测的云底高度,纵坐标表示 FY-4A 的云底高度估计值。

基于 FY-4A 的云分类产品,将结果分为六类云进行分析。其中水云与过冷云的云底高度估计值与 CloudSat/CALIPSO 云底高度的一致性相对较好。而对混合云、冰云和卷云,当云底高度大于约 1 km 时,误差相对较小;当云底高度小于约 1 km 时,结果会出现了较大的误差。这些误差较大的样本主要为深

对流云,平均云几何厚度大于 7.5 km,而基于云光学厚度、云有效粒子半径、云水路径等信息无法准确估计深对流云的云底高度^[8]。另外,FY-4A 云分类算法对水云和过冷云的识别准确率相对较高,而对冰云、卷云和多云的识别准确率相对较低,因此可能有部分误差较大的样本源于对云类型的误判。

表 3 FY-4A 与 CloudSat、CALIPSO 云底高度对比结果
Table 3 Comparison of FY-4A、CloudSat and CALIPSO CBH

	Water	Supercool	Mixed	Ice	Cirrus	Overlap	All
平均云底高度(km)	1.25	2.15	3.14	4.83	7.44	7.28	4.19
平均云顶高度(km)	1.80	3.4	5.90	10.92	10.34	9.55	7.69
平均云层厚度(km)	0.66	1.55	2.98	6.4	3.16	2.12	3.14
平均绝对偏差(km)	0.35	0.64	1.08	1.72	1.56	1.33	1.29
STDE(km)	0.66	1.09	1.63	2.44	2.14	1.80	1.76
相关系数	0.84	0.83	0.81	0.77	0.80	0.79	0.80

表 3 列举了各类云的平均云底高度、平均云顶高度、平均云层厚度、平均绝对偏差(MAE),均方根误差(STDE)和相关系数(R)。表中可以看到,不同

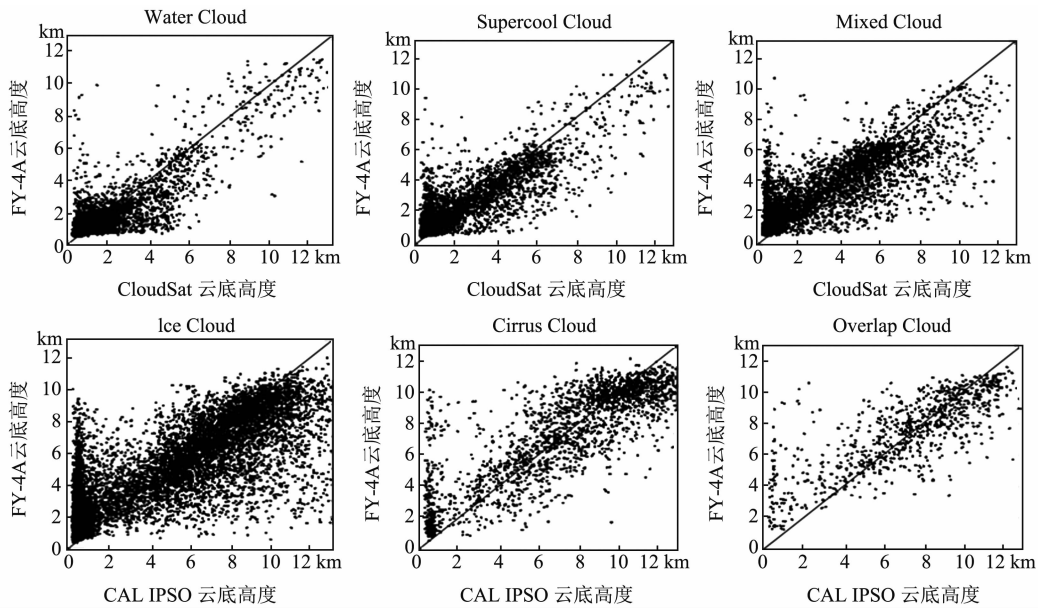


图 2 FY-4A 与 CloudSat、CALIPSO 云底高度对比散点图
 Fig. 2 Comparison of FY-4A、CloudSat & CALIPSO CBH

类型云的云底高度、云顶高度和云层厚度存在明显的差异,水云的云顶高度最小,云层厚度也最小,而冰云的平均云顶高度最大并且云层厚度最大.水云、过冷云和混合云的平均绝对偏差和均方根误差相对较小,相关系数都超过 0.8,说明其云底高度的误差相对冰云、卷云和重叠云较小.为深入分析误差来源,下面对多层云系统、云层厚度、云顶高度等因素进行研究.

3.1.1 多层云

根据 CloudSat 和 CALIPSO 卫星的云廓线,将结果分为单层云和多层云.图 3 为 6 类云的云底高度偏差的分布,实线代表单层云,虚线代表多层云.图中单层云的云底高度偏差都集中在 0 附近,表明单层云时云底高度偏差较小;多层云的云底高度偏差则比较分散且主要集中在负数区,表明多层云存在时云底高度估计的误差较大,并且 FY-4A 的云底高

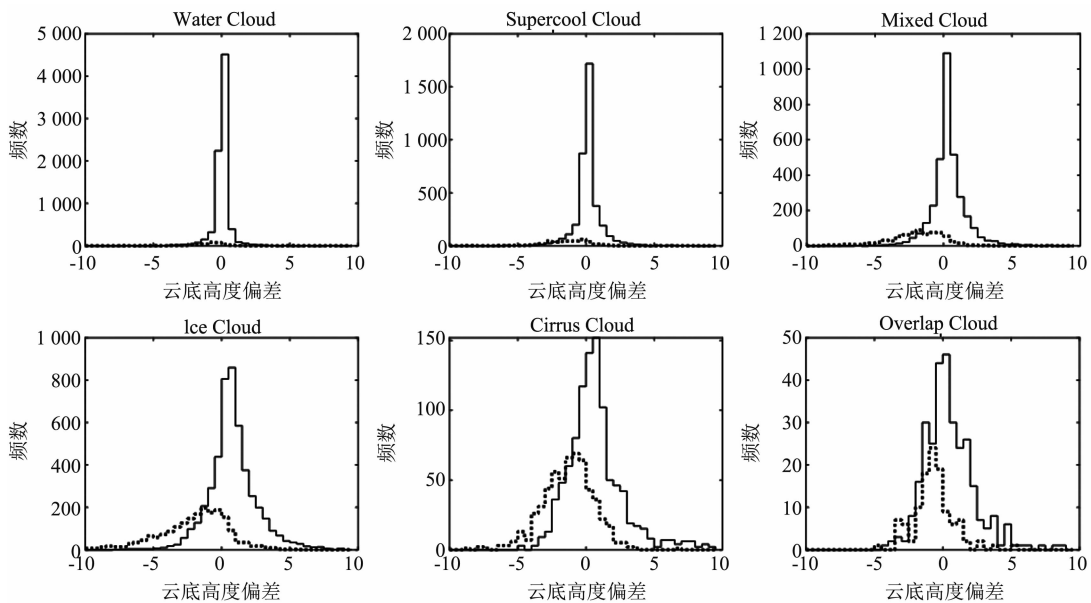


图 3 FY-4A 云底高度偏差在单层云和多层云条件下的频率分布
 Fig. 3 Distribution of FY-4A CBH bias for single-layer and multi-layer clouds

度一般小于 CloudSat 和 CALIPSO 卫星探测最上层云的云底高度. Holz 等研究发现基于多通道辐射计资料反演多层云云顶高度时,结果通常介于两层云之间,即反演结果通常小于最上层云云顶高度^[20]. 当多层云存在时,多通道辐射计对光学厚度、云有效粒子半径等光学参数的反演同样具有较大的不确定性^[21]. 因此,这些输入参数的误差是算法对多层云云底高度的估计误差较大的重要原因.

3.1.2 云顶高度

表 4 计算了水云情况下不同云顶高度误差对应的云层厚度的误差以及云底高度的误差. 从表中可以看出,云顶高度的误差越大,云层厚度的误差越大,云底高度的误差也更大. 同时,云底高度偏差小于 500 m 的比例也随云顶高度误差的增大而减小.

表 4 水云的云顶高度误差与云层厚度、云底高度误差的关系

Table 4 Relationship between CTH bias and cloud thickness & CBH errors

云顶高度误差率	云层厚度误差平均值(m)	云层厚度误差率	云底高度误差平均值(m)	云底高度误差率	云底高度偏差小于 500m 比例
10%	226.4	18%	240.7	26%	90%
20%	292.6	21%	389.3	31%	86%
30%	370.3	24%	558.2	37%	78%

表 5 对冰云计算了相应参数,结果表明冰云的云层厚度与云底高度的误差随云顶高度误差的变化趋势与水云相同. 此外,冰云的云层厚度误差和云底高度误差均大于 1 km,云底高度偏差小于 500 m 的比例也明显小于水云,表明 FY-4A 对冰云云底高度的估计误差相对较大.

表 5 冰云的云顶高度误差与云层厚度、云底高度误差的关系

Table 5 Relationship between CTH bias and cloud thickness & CBH errors

云顶高度误差率	云层厚度误差平均值(m)	云层厚度误差率	云底高度误差平均值(m)	云底高度误差率	云底高度偏差小于 500 m
10%	1106.4	19%	1190.7	24%	36%
20%	1223.5	22%	1480.5	30%	32%
30%	1434.7	26%	1670.2	38%	28%

3.1.3 云层厚度

图 4 所示为 FY-4A 云底高度偏差随云层厚度的变化趋势. 由图 4 可知,当云层厚度小于约 3 km 时, FY-4A 的云底高度偏差较小,表明对厚度较小的云, FY-4 的云底高度更接近 CloudSat 和 CALIPSO 探

测的云底高度;当云层厚度大于约 3 km 时,云底高度偏差逐渐增大,表明对厚度较大的云, FY-4A 云底高度误差较大. 另外,当云层厚度小于 3 km 时,云底高度偏差为负值,即 FY-4A 的云底高度低于 CloudSat 或 CALIPSO 探测的云底高度,而对厚度较大的云,云底高度偏差为正值,表明此时 FY-4A 的云底高度一般高于 CloudSat 或 CALIPSO 探测的云底高度.

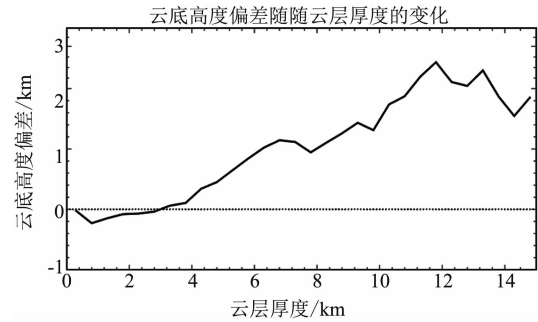


图 4 云底高度偏差随云层厚度的变化
Fig. 4 CBH bias changes with cloud thickness

综合以上分析,云分类、云顶高度和云光学性质等输入数据的误差会影响算法对云底高度估计结果的精度,尤其体现在多层云时算法估计值相较 CloudSat/CALIPSO 探测的云底高度偏小. 此外,本文使用了 3 个月的匹配数据对随机森林模型进行训练,未来若能使用更多的数据将会一定程度上提高算法的精度.

3.2 个例分析

图 5 左边为 2017 年 11 月 29 日 FY-4A 的全圆盘云顶高度,右边是由算法估计的云底高度,黑色实线代表 CloudSat 卫星的星下点轨迹. 图中对 FY-4A 云顶高度产品存在有效值的区域,都能通过算法得到云底高度的估计值. 为检验云底高度的估计结果,

选取图中纬度为 3°S 至 4°N 范围内 CloudSat 卫星探测的云廓线与相应的 FY-4A 云顶高度、云底高度作为典型个例,如图 6 所示. 图 7 中黑色代表 CloudSat 卫星探测的云廓线,红色代表 FY-4A 的云顶高度产品和估计的最上层云云底高度. 图中所示结果与上述分析相吻合, A 区域附近对单层云的云底高度估计结果较好,而对 B 区域的多层云, FY-4A 云顶高度小于 CloudSat 探测的云顶高度,由算法估计的云底高度也小于最上层云云底高度. 另外,当云层厚度较大时(C 区域), FY-4A 的云底高度明显大于 CloudSat 卫星探测的云底高度.

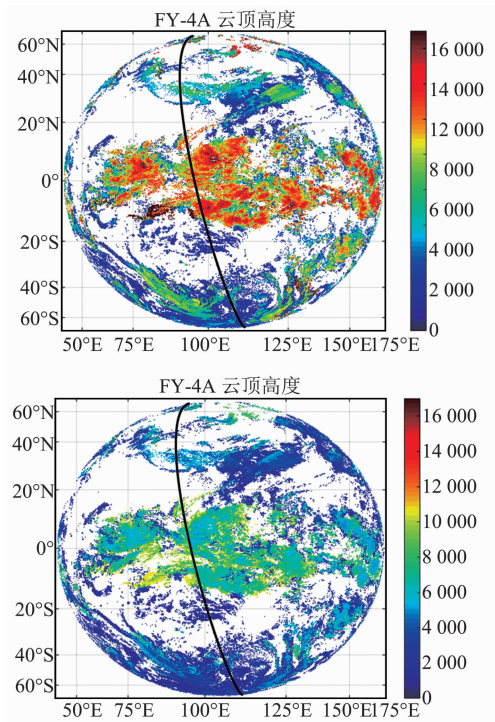


图5 FY-4A 云顶高度产品与算法估计的云底高度

Fig. 5 CTH product and CBH estimates of FY-4A

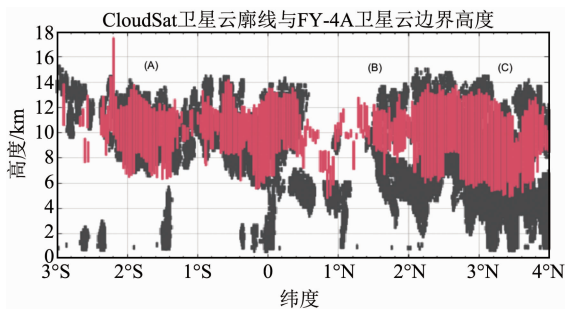


图6 CloudSat 卫星云廓线(灰色)与FY-4A 卫星云边界高度(红色)对比

Fig. 6 CloudSat cloud profile (grid) and FY-4A cloud boundary (red) at Nov. 29, 2017

4 结论

基于随机森林算法建立了 FY-4A 对最上层云云底高度的估计算法,利用 2017 年 8 月至 10 月的 FY-4A 与 A-Train 卫星资料的匹配数据对随机森林算法进行了训练,然后将算法应用至 11 月的 FY-4A 上游产品,得到对最上层云的云底高度,并用 CloudSat 和 CALIPSO 卫星探测的云底高度对结果进行了检验与误差分析,得到主要结论如下:

基于随机森林算法,利用 FY-4A 的云顶高度、

云光学厚度、云有效粒子半径、云液态水(冰)路径等上游产品,能有效估计云底高度. 结果与 CloudSat 或 CALIPSO 卫星探测的最上层云云底高度有较好的一致性,其中对水云、过冷云和混合云的云底高度估计结果相对较好,平均绝对偏差小于 1 km,相关系数均大于 0.8,对冰云的云底高度估计误差相对较大,平均绝对偏差为 1.72 km,相关系数为 0.77.

单层云的云底高度偏差分布主要集中在 0 附近,表明其云底高度误差相对较小. 当多层云存在时,云底高度偏差比较分散且主要集中在负数区,表明多层云时 FY-4A 的云底高度小于 CloudSat 或 CALIPSO 卫星探测的最上层云云底高度.

云顶高度的误差会显著影响云层厚度和云底高度的结果,对水云和冰云的统计分析表明:随着云顶高度误差增大,云层厚度和云底高度的误差也随之增大. 此外,冰云的云层厚度误差和云底高度误差相对水云较大,这也与前面的结论一致. 对云层厚度小于约 3 km 的云, FY-4A 的云底高度偏差相对较小. 但是当云层厚度大于 3 km 时,云底高度偏差随云层厚度的增加而逐渐增大. 对云层厚度较大的云,云底高度偏差一般为正值,即 FY-4A 的云底高度高于 CloudSat 或 CALIPSO 探测的云底高度.

需要指出的是,在本文的算法训练与结果评估中,都以 CloudSat 和 CALIPSO 卫星探测的云底高度为真值,而根据 Welliver 等人的研究结果,CloudSat 与 CALIPSO 对最底层云的云底高度探测结果与实际云底高度存在一定的误差^[22],而随机森林算法的训练效果十分依赖给定的目标输出. 因此,未来若能结合多站点的地基探测数据和 A-Train 系列卫星的星载主动探测结果构建更加准确的样本集,可以有效减小算法误差,借助 FY-4A 静止气象卫星的大范围,高时间分辨率的探测优势,可以得到更加精确的云底高度结果,对于气候变化研究、数值天气预报和航空安全等领域具有重要意义.

致谢:

衷心感谢国家卫星气象中心提供的 FY-4A 数据和技术文档,A-Train 卫星数据下载自美国国家航空航天局 NASA 地球数据网站.

References

- [1] Wang J H, Rossow. Effects of cloud vertical structure on atmospheric circulation in the GIS GCM [J]. *J. Climate*, 1998, 11(11): 3010-3029.
- [2] Weisz E, Li J, Menzel W P. Comparison of AIRS, MODIS, CloudSat and CALIPSO cloud top height retrievals

- [J]. *Geophys Res Lett.*, 2007, **34**(17):L17811.
- [3] YAN Wei, HAN Ding, LU Wen, *et al.* Research of cloud base height retrieval based on COSMIC occultation sounding data [J]. *Chinese J. Geophys* (严卫, 韩丁, 陆文, 等. 基于COSMIC 掩星探测资料的云底高反演研究. *地球物理学报*), 2012, **55**(1): 1-15.
- [4] YAN Wei, Han Ding, Zhao Xian-Bin, *et al.* Comparative research of cloud boundary heights based on millimeter-wave radar, radio occultation and radiosonde data [J]. *Chinese J. Geophys* (严卫, 韩丁, 赵现斌, 等. 基于毫米波雷达、无线电掩星和探空仪资料的云边界高度对比研究. *地球物理学报*), 2012, **55**(7): 2212-2226.
- [5] Menzel Coauthors. MODIS global cloud-top pressure and amount estimation: Algorithm description and results [J]. *J. Appl. Meteor. Climatol.*, 2008, **47**: 1175-1198.
- [6] Weisz E, Li Jun, Zhou D K, *et al.* Cloudy sounding and cloud-top height retrieval from AIRS alone single field-of-view radiance measurements [J]. *Geophys. Res. Lett.*, 2007, **34**: L12802.
- [7] Heidinger, Pavolonis M J. Gazing at cirrus clouds for 25 years through a split window. Part I: Methodology [J]. *J. Appl. Meteor. Climatol.*, 2009, **48**: 1100-1116.
- [8] Noh Y J, and Coauthors. Cloud-base height estimation from VIIRS. Part II: A statistical algorithm based on A-Train satellite data [J]. *J. Atmos. Oceanic.*, 2017, **34**: 585-598.
- [9] Sun, Xue-Jin, Li Hao-Ran, Barker H W, *et al.* Satellite-based estimation of cloud-base heights using constrained spectral radiance matching [J]. *Quart. J. Roy. Meteor. Soc.*, 2016, **142**: 224-232.
- [10] Hutchison, Wong E, Ou S C. Cloud base height retrieval during nighttime conditions with MODIS data [J]. *Int. J. Remote Sens.*, 2006, **27**: 2847-2862.
- [11] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [12] Vincenzi S, Zucchetto M, Franzoi P, *et al.* Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy [J]. *Ecological Modelling*, 2011, **222**: 1471-1478.
- [13] Grimm R, Behrens T, Marker M, *et al.* Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using Random Forests analysis [J]. *Geoderma*, 2008, **146**: 102-113.
- [14] Prasad A, Iverson L, Liaw A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction [J]. *Ecosystems*, 2006, **9**: 181-199.
- [15] Ismail R, Mutanga, O. A comparison of regression tree ensembles: predicting *Sirex noctilio* induced water stress in *Pinus patula* forests of KwaZulu-Natal, South Africa [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2010, **12**: S45-S51.
- [16] Abdel-Rahman, E M, Berg van den, Way M, *et al.* Hand-held spectrometry for estimating thrips (*Fulmekiola serrata*) incidence in sugarcane [J]. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, *IEEE, Cape Town, South Africa*, 2009, 268-271.
- [17] Onesimo Mutanga, Elhadi Adam, Moses Azong Cho. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2012, **18**: 399-406.
- [18] ZHANG Zhi-Qing, DONG Yao-Hai, DING Lei, *et al.* China's First Second-generation FY-4 Meteorological Satellite Launched. *Space International*, (张志清, 董瑶海, 丁雷, 等. 我国首颗第二代静止气象卫星风云-4 升空. *国际太空*) 2016, **12**: 6-12.
- [19] Miller S D, Coauthors. Estimating three-dimensional cloud structure via statistically blended satellite observations [J]. *J. Appl. Meteor. Climatol.*, 2014, **53**: 437-455.
- [20] Holz R E, Ackerman S A, Nagle F W, *et al.* Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIPSO [J]. *J. Geophys. Res.*, 2008, **113**: D00A19.
- [21] Platnick, *et al.* The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, **55**: 502-525.
- [22] Welliver E A. Remote detection of cloud base height using CloudSat and CALIPSO [master's thesis]. California: Naval postgraduate school, 2009: 25-40.