

星载激光雷达云和气溶胶分类反演算法研究

李明阳^{1,2}, 范萌^{1*}, 陶金花¹, 苏林¹, 吴桐^{1,3}, 陈良富¹, 张自力⁴

1. 中国科学院遥感与数字地球研究所遥感科学国家重点实验室, 北京 100101
2. 中国科学院大学遥感与数字地球研究所, 北京 100049
3. 山东科技大学测绘科学与工程学院, 山东 青岛 266510
4. 浙江省环境监测中心, 浙江 杭州 310007

摘要 激光探测对于获取云和气溶胶的垂直廓线, 研究大气中云和气溶胶的垂直分布特征以及对全球气候变化的影响意义重大。而星载大气激光雷达云气溶胶分类算法的研究, 对于激光雷达数据的参数反演及应用极为重要。针对激光条件下探测的云和气溶胶特有的光学信息和空间分布, 结合概率统计与机器学习算法, 提出了一种对于云/气溶胶、云相态及气溶胶子类型识别的分类算法, 实现了星载激光雷达的大气特征层快速、有效分类。算法采用中国地区 2016 年 CALIOP 的观测数据作为样本数据, 主要由三部分组成: (1) 基于激光探测的云和气溶胶层不同的光学特性以及地理空间分布特征, 分别构建了云和气溶胶的 γ_{532} , χ , δ , Z 和 lat 的五维概率密度函数, 以此为基础构建云气溶胶的分类置信函数, 并基于此实现了云和气溶胶类型的反演; (2) 选取支持向量机(SVM)作为随机朝向冰晶粒子(ROI)和水云分类的算法模型基础, 结合云层的 γ_{532} , χ , δ Z 和云顶温度 T 的概率密度函数构建 ROI, 水平朝向冰晶粒子(HOI)和水云的分类置信函数以修正 SVM 误分的特征层以及筛选出水云中少部分的 HOI 冰云, 获得云相态的分类结果; (3) 以各气溶胶子类型的光学以及空间分布特性为基础, 采用决策树策略的气溶胶子类型识别算法实现了对气溶胶子类型的区分, 完成气溶胶子类型的识别。利用现有 CALIOP 观测结果作为样本数据构建分类数据库, 避免了对于地面以及航测数据的依赖, 而机器学习则大大简化了算法的实现过程, 使得云气溶胶分类更加高效。算法结果与正交极化云气溶胶激光雷达垂直特征层分布数据(CALIPSO VFM)产品对比分析: 云层有 98.51% 一致性, 气溶胶有 88.43% 的一致性, 且白天比夜间一致性高。对于云相态分类, 可以有效区分出水云和冰云, 其中二者水云一致性高达 93.44%。在气溶胶子类型反演结果中, 可以准确识别出大多数气溶胶特征层子类型。霾、沙尘以及晴空三种典型情况下的反演结果均与 CALIOP VFM 产品数据具有较好的一致性。其中, 霾天的大部分煤烟型以及污染型(污染沙尘以及污染大陆)气溶胶反演结果与 VFM 具有较好的一致性。沙尘天也能够获得较好的沙尘以及污染沙尘的结果。晴空为数不多的气溶胶层也取得了较为一致的结果。对于实现的星载大气激光雷达特征层分类算法, 针对 CALIOP 激光测量的云气溶胶层的分类进行了重要的改进, 在保证一定精度的基础上, 简化了算法, 提高了数据处理的效率, 在下一步工作中, 将分别构建不同时段和季节的分类模型以及提高两种不同偏振特性的冰云和气溶胶子类型的分类精度。

关键词 星载激光雷达; 云和气溶胶分类; 概率密度函数; 支持向量机; 决策树

中图分类号: P407.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)02-0383-09

引言

激光雷达是探测大气中云和气溶胶类型及时空分布的重

要技术手段之一, 对于研究气候变化、环境污染和人类健康至关重要。相比于地基激光雷达观测, 星载激光雷达可以获取大区域范围内不同高度和类型的云、气溶胶分布, 对于研究云和气溶胶的微物理特性、辐射强迫效应、污染物传输以

收稿日期: 2017-12-12, 修订日期: 2018-04-23

基金项目: 国家重点研发计划(2016YFC0200404), 国家自然科学基金项目(41501373, 41571347), 浙江省科技厅重点研发计划项目(2017C03037)资助

作者简介: 李明阳, 1990 年生, 中国科学院遥感与数字地球研究所遥感科学国家重点实验室硕士研究生 e-mail: limy2015@radi.ac.cn

* 通讯联系人 e-mail: fanmeng@radi.ac.cn

及污染源的解析等具有不可替代的作用。在激光雷达云和气溶胶光学参数反演中,基于云和气溶胶类型的识别结果进而设定不同大气情况下的雷达比也是其关键步骤之一^[1]。因此,星载激光雷达云和气溶胶类型反演算法的研究对于全球气候变化过程的认识和人类生存环境的改善有着重要的现实意义。

搭载有星载激光雷达 CALIOP 的 CALIPSO 卫星由美国国家航天局(NASA)与法国国家航天中心合作于 2006 年发射,已获取了十年的全球云和气溶胶类型的垂直分布数据,为全球气候和环境变化研究提供了重要资料^[2-3]。目前 CALIOP 已超期服役,国内外均已开始了新一代星载激光雷达的研制,旨在持续获取全球云、气溶胶类型和光学特性的

垂直分布信息,支持全球气候变化、云与气溶胶相互作用等研究。其中,欧洲与日本计划于 2018 年合作发射的 EARTH CARE 卫星上搭载有 ALID(Atmosphere Lidar)载荷^[4-5],美国 NASA 计划在 2020 年实施的 ACE(Advanced Composition Explorer)卫星发射计划也将搭载新一代激光雷达系统。国内,航天科技集团九院 704 所研制中的星载多波束激光雷达,将作为我国首个星载激光雷达,预计于 2020 年发射。由上海光机所负责研发的具有中国自主知识产权的星载激光雷达 ACDL 也预计在 2020 年发射。

国内外基于星载激光雷达的云和气溶胶类型反演算法的研究进展如表 1 所示,主要内容包括三个部分:云/气溶胶分类、云相态反演和气溶胶子类型反演。

表 1 云/气溶胶分类算法研究进展

Table 1 The introduction of the Cloud and Aerosol Classification Algorithms

研究进展介绍	
云/气溶胶区分	(1) CALIOPV1 算法以三参数(532nm 后向散射系数、色度比和特征层中间高度)的概率密度函数构建置信函数进行云/气溶胶分类; (2) Chen 等提出主被动遥感相结合的方法(CLIM),即引入 CALIPSO 卫星红外成像辐射计 IIR 红外波段数据以解决沙尘气溶胶的误判问题 ^[6] ; (3) Liu 等加入退偏比和纬度参数优化构建的置信函数提高分类精度 ^[7] ; (4) Ma 等利用支持向量机来解决在有限样本数据条件下,实现云/气溶胶分类 ^[8] ; (5) Naeger 等利用 CALIOP 结合 MODIS 的红外波段数据解决云和沙尘气溶胶的误判 ^[9]
云相态分类	(1) CALIOP 以云层顶温度,层底温度以及层积分退偏比构建云相态分类置信函数,将云粗略分为冰云、水云以及混合相态的云; (2) Hu 等基于多年 CALIOP 观测数据的云相态特征综合分析结果,利用层积分散射系数和层积分退偏比,并参考云顶温度将云分为水云以及水平朝向冰晶粒子(HOI)、随机朝向冰晶粒子(ROI)的冰云,提高了云相态反演精度 ^[10]
气溶胶子类型分类	结合 AERONET 地面站点数据,利用聚类分析的方法,将气溶胶子类型分为沙尘、煤烟型、洁净大陆、污染大陆、海洋、污染沙尘和其他

目前,在基于 CALIOP 激光测量的云和气溶胶类型反演算法中,所使用的样本都是 NASA 前期气溶胶观测实验得到的结果,样本数量有限,且对于初始信号大气特征层的判断多依赖于人为判断。在 NASA 公开的 CALIOP 反演算法文档中,许多参数的设定并未提及,算法实现过程比较复杂,而国内星载大气激光雷达的研究还处于起步阶段,相关研究非常有限。因此,本工作针对中国地区,通过直接应用现有 CALIOP 观测结果来选择类别明确的大气特征层作为样本数据以提高分类精度,并将支持向量机和决策树方法引入云和气溶胶类型反演算法,该算法可以在保证分类精度的同时降低验证样本的需求数量,简化算法。

1 方法与数据

NASA 对于 CALIOP 大气特征层分类以监督分类方法为基础,根据训练样本学习分类规则,对未知类别的样本进行分类,即构建一个分类器尽可能准确地预测特征属性空间中每个点的类别。用 n 个概率来描述包含 k 个特征属性的向量 $X = (x_1, \dots, x_k)$ 归属于 n 个类别的可能性,即

$$f_i(X) = \Pr(y = i | X) \quad (1)$$

式(1)中, $\Pr()$ 为概率函数, $f_i(X)$ 反映了向量 X 属于第 i 类别的概率大小,这是 NASA CALIOP 大气特征层分类算法的理论基础。但在实际分类过程中, $f_i(X)$ 是未知的,只能通过训练和学习已知样本得到 $f_i(X)$ 的近似估计函数 $\bar{f}_i(X)$ 。因此,在考虑到 NASA 已有算法本身的局限性和计算复杂性,提出了一种基于现有 CALIOP 观测结果,结合支持向量机与决策树方法来实现云/气溶胶分类、云相态反演、气溶胶子类型反演的新思路。

为研究星载激光雷达在中国地区的云/气溶胶类型反演问题,选取了中国地区 2016 年共 3376 轨 CALIOP L2 级 5 km 云/气溶胶产品。对其中置信度绝对值为 100 的大气特征层作为有效特征层进行提取,共提取到 1 610 111 个云特征层和 270 389 个气溶胶特征层。

2 云和气溶胶类型反演算法

2.1 云/气溶胶分类算法

基于云和气溶胶的 γ_{532} , χ , δ , Z 和 lat 五个参数的不同特点,构建云/气溶胶的置信函数 f_{5D} 来进行云和气溶胶分类,即

$$f_{5D}(\gamma_{532}, \chi, \delta, Z, \text{lat}) = \frac{P_c(\gamma_{532}, \chi, \delta, Z, \text{lat}) - P_a(\gamma_{532}, \chi, \delta, Z, \text{lat})}{P_c(\gamma_{532}, \chi, \delta, Z, \text{lat}) + P_a(\gamma_{532}, \chi, \delta, Z, \text{lat})} \quad (2)$$

式(2)中, P_c 和 P_a 分别为云和气溶胶的概率密度函数。 f_{5D} 为置信函数, 取值范围为 $-1 \sim 1$, 表征云或气溶胶。云/气溶胶分类算法流程如图 1 所示, 具体步骤为:

(1) 采用 2016 年 CALIOP 的云/气溶胶数据, 构建云和气溶胶分类置信函数[式(2)];

(2) 对于 CALIOP 激光雷达探测到的大气特征层, 获取并计算其光学参数($\gamma_{532}, \chi, \delta$)和地理参数(Z, lat);

(3) 采用对流层高度辅助数据, 提取对流层内大气特征层进行分类计算;

(4) 利用提取的特征参数计算 f_{5D} , 若函数值小于 0, 则分类为气溶胶, 函数值大于 0, 则分类为云, 函数值为 0, 则类别不确定, 并取函数值($\times 100\%$)作为特征层类别的置信度 CAD_Score。

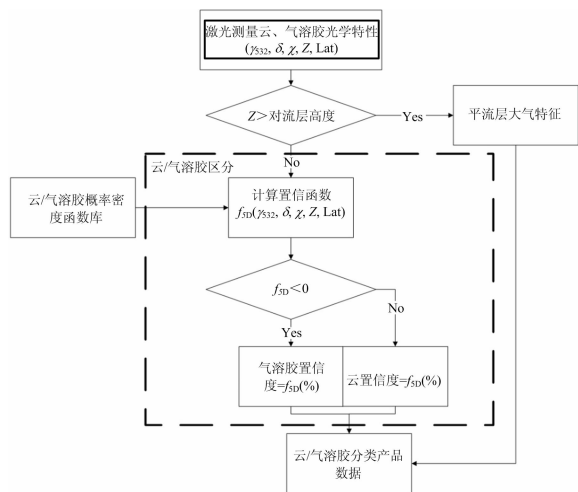


图 1 云/气溶胶分类算法流程图

Fig. 1 Flowchart of cloud/aerosol discrimination algorithm

2.2 云相态反演算法

基于 2016 年 CALIPSO 分类产品, 引入支持向量机方法(SVM), 其优点在于以直接观测数据为样本, 实现云相态的高效分类。

采用 1 339 201 个云特征层数据(1 080 496 个 ROI 特征层, 8 777 个 HOI 特征层, 249 928 个水云特征层)[图 2(a)], 研究不同相态云的 δ 和 γ_{532} 之间的关系[图 2(b)], ROI 和水云呈现出明显的线性可分性, HOI 则相对分散地夹杂二者之间。基于此, 选用线性核函数的 SVM 对云相态进行反演。

对于 SVM 的分类结果[图 2(b)中黑色实线为分类线], 其误判频数统计结果如图 3(a), 主要集中在分类线周围。为此, 结合分类置信函数对 SVM 进行改进, 即在 SVM 的基础上, 结合水云和 HOI 的 $\gamma_{532}, \chi, \delta, Z$ 和云顶温度 T 构建置信函数[见式(3)]

$$f_{5D} = \frac{P_{\text{水云}}(\gamma_{532}, \delta, \chi, Z, T) - P_{\text{ROI}}(\gamma_{532}, \delta, \chi, Z, T)}{P_{\text{水云}}(\gamma_{532}, \delta, \chi, Z, T) + P_{\text{ROI}}(\gamma_{532}, \delta, \chi, Z, T)} \quad (3)$$

改进后的结果中误判特征层数量显著减少[图 3(b)]。误判总数量减少了 3 948, 其中, ROI 误判减少了 889, 水云误判减少了 3 059(表 2)。

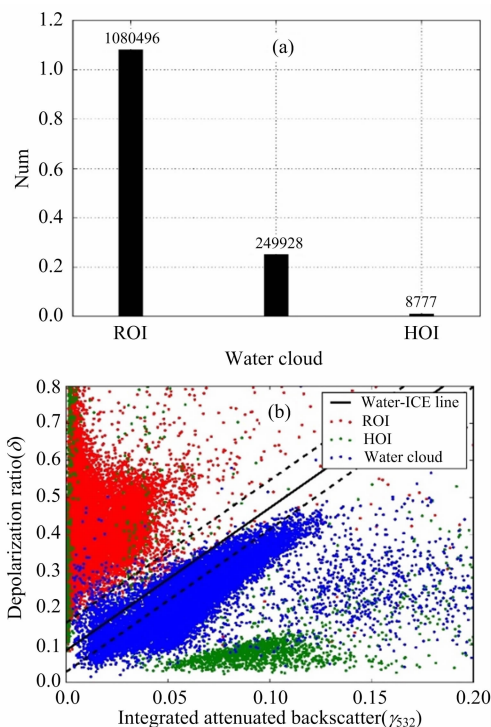


图 2 (a) 云特征层的相态数量统计, (b) 不同相态云的层积分退偏比与层积分后向散射系数关系图

Fig. 2 (a) Numbers of three cloud phases and (b) relationships between γ_{532} and δ for clouds with different phases

表 2 ROI/水云分类结果

Table 2 The result of ROIs and water clouds classification

	ROI 识别为水云的数量	水云识别为 ROI 的数量	总的误识别数量
SVM(改进前)	5 110	4 297	9 407
SVM + f_{5D} (改进后)	4 221	1 238	5 459

为了提高分类效率且考虑到 ROI 和 HOI 的雷达比设定相同, 仅需区分出水云中的 HOI。与 ROI 和水云分类计算类似, 通过构建水云和 HOI 的置信函数[公式(4)]筛选水云中的 HOI。

$$f_{5D} = \frac{P_{\text{水云}}(\gamma_{532}, \delta, \chi, Z, T) - P_{\text{HOI}}(\gamma_{532}, \delta, \chi, Z, T)}{P_{\text{水云}}(\gamma_{532}, \delta, \chi, Z, T) + P_{\text{HOI}}(\gamma_{532}, \delta, \chi, Z, T)} \quad (4)$$

云相态反演算法流程如图 4 所示, 具体步骤包括:

(1) SVM 核函数及其参数的设定: 线性核函数简单实用并且在选择处罚和损失函数时具有较大的灵活性, 因此本文选用线性核函数作为 SVM 分类时的核函数。此外考虑到样本中 ROI 和水云之间数量不对等, 需要根据样本数据量的大小自动设置类别权重; 利用 SVM 分类器初步区分水云和 ROI;

(2) 针对 SVM 误判的结果, 依据误判特征的直方图分布

设定阈值范围建立不确定区[范围即图 2(b)黑色虚线之间], 并通过构建 ROI 和水云的置信函数[式(3)]对区间内进行重新分类; 在不确定区外, 对置信度大于 80% 的误判特征层进行修正;

(3) 构建 HOI 和水云的置信函数[式(4)]将 HOI 从水云中分离出来;

(4) 依据水云和冰云的温度特征, 将云相态反演结果中云顶温度高于 0 °C 的冰云修正为水云, 将云顶温度低于 -40 °C 的水云修正为冰云(ROI)。

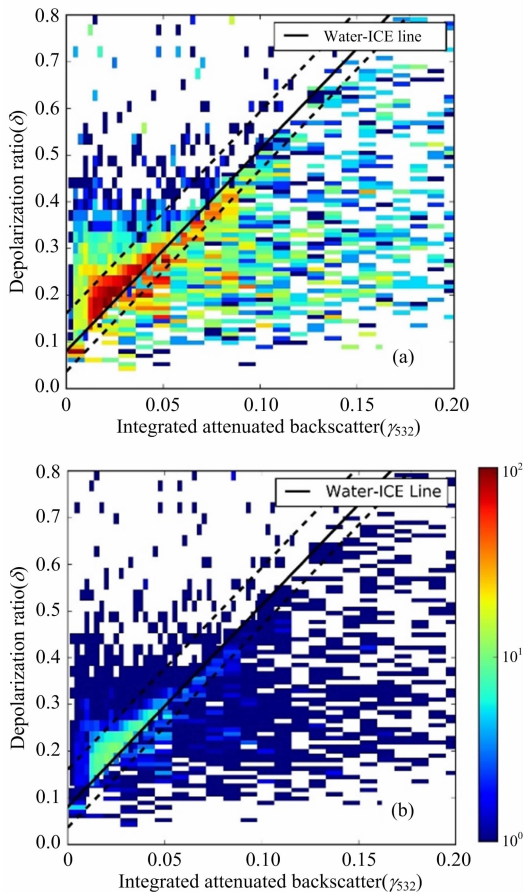


图 3 对 SVM 结果 (a)修正前和 (b)修正后频数分布
Fig. 3 Frequency distributions of (a) SVM and (b) SVM + f_{5D} classification results

2.3 气溶胶子类型反演算法

基于观测实验以及 AERONET 的长期观测结果, CALIPSO 产品将气溶胶分为沙尘、煤烟、洁净大陆、污染大陆、海洋、污染沙尘和其他七个类型。

在此基础上, 利用不同气溶胶子类型的光学和空间分布特征, 增加决策树分类策略对气溶胶子类型进行反演。沙尘型气溶胶一般具有较大的 δ 且主要分布在沙漠地区, 以 0.3 为阈值, 首先识别出主要的沙尘气溶胶, 之后参考 IGBP 土地类型, 筛选出洁净大陆型; 然后依据 IGBP, 识别出沙漠地区的沙尘型和污染沙尘型气溶胶并参考 δ 将二者区分, 以及主要分布在海洋上的洁净海洋型气溶胶; 此外, 针对不同类型气溶胶的垂直分布, 利用层中间高度信息, 将气溶胶分为

主要分布在 ~1.5 km 以下的污染大陆型和污染沙尘型气溶胶和 ~2 km 以上的高空的煤烟型和洁净大陆型气溶胶; 最后参考气溶胶层顶温度区分出污染大陆性和污染沙尘型气溶胶, 并根据 γ_{532} 和 χ 的主要分布范围, 识别出煤烟型和洁净

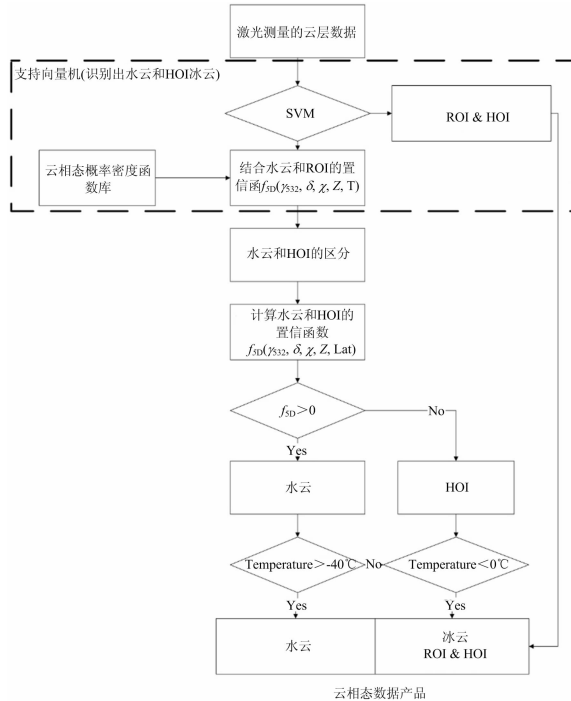


图 4 云相态分类流程图
Fig. 4 Flowchart of ice/water cloud algorithm

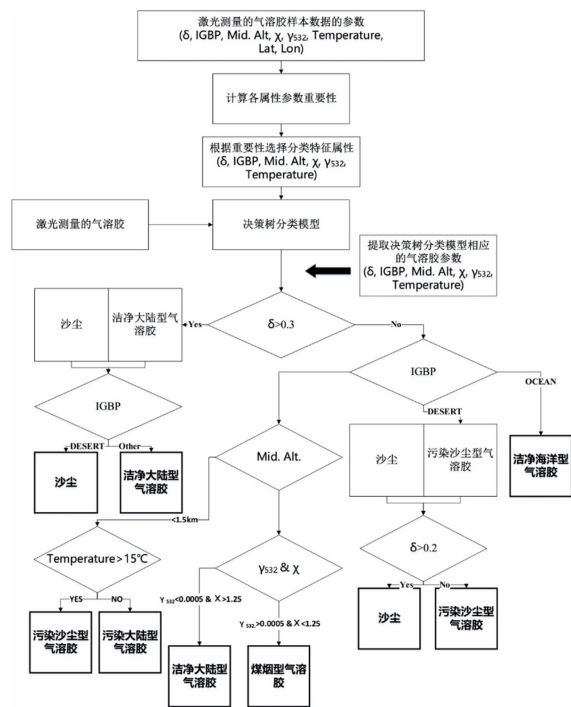


图 5 决策树气溶胶子类型分类流程图
Fig. 5 Flowchart of decision tree algorithm for aerosol subtype classification

大型气溶胶,完成气溶胶子类型分类。具体流程图如图 5 所示。

3 反演结果与分析

3.1 基于 CALIOPVFM 产品的比对分析

考虑到 CALIPSO 官方数据自 2006 年起已通过多次实验验证^[11],选取 2016 年 12 月 CALIOP 垂直特征层数据与本文结果进行一致性比较,以说明本文算法的可靠性和适用性。

表 3 为云/气溶胶分类结果统计,本算法实现了大多数云和气溶胶的识别,无法识别的比例均小于 1%,具有较高的适用性。此外,与 CALIOP VFM 产品具有较高的一致性,其中差异性主要为,本文 1.49%云在 VFM 产品中被判定为气溶胶,11.57%的气溶胶被判定为云。且白天云和气溶胶的反演一致性普遍高于夜晚,主要因为白天的训练样本数据量多,使得训练得到的分类模型对白天特征层的表征能力高于夜晚。图 6 为本文算法(红色)与相对应的 CALIOP(蓝色)置信度 CAD_Score 的比较结果。其置信度频率分布与 VFM 产品基本一致,尤其正值区域的云分类结果。在 CAD_Score = -100 处,本文云/气溶胶分类算法对于气溶胶识别率高于 VFM 产品,但在其他地方,尤其是当 CAD_Score = 0 ± 5 时,云和气溶胶的混合特征层较多,识别率远低于 VFM 产品,其中,当 CAD_Score = 0 时,本文和 CALIOP VFM 的频率密度差值达到 0.024。

表 3 云/气溶胶分类结果

Table 3 The results of cloud and aerosol discrimination

总的数量	气溶胶识别为云的百分比/%	云识别为气溶胶的百分比/%	无法识别的百分比/%
374 150 (day and night)	11.57	1.49	0.58
210 047(day)	5.49	1.1	0.39
164 103(night)	15.40	1.8	0.73

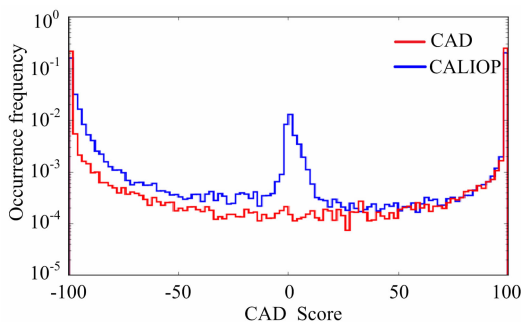


图 6 本算法与 CALIOP VFM 产品的云/气溶胶分类置信度分布直方图

Fig. 6 Statistical histograms of CAD_Score for our algorithm retrievals and CALIOP VFM products

通过将云相态反演结果与 CALIOP VFM 产品比对(图 7

和表 4),二者对水云识别一致性达到 93.44%。二者的差异主要体现在对 ROI 和 HOI 两种冰云的判识上,其中 ROI 的一致性为 88.04%,HOI 仅为 51.11%。但由于两者均为冰云,雷达比近似相同(CALIOP 产品中将二者的雷达比均设为 25),因此,两种冰云之间的误判对于云消光系数的反演结果影响较小。

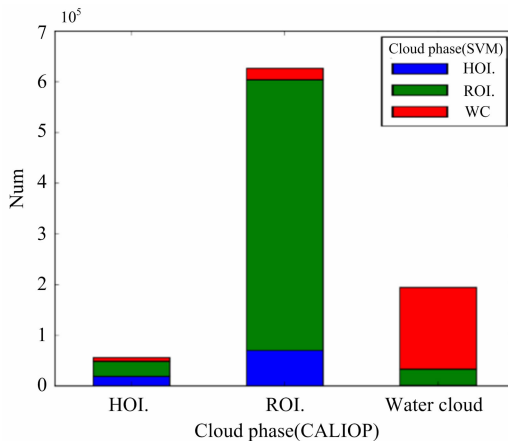


图 7 不同相态云分类结果柱状图

Fig. 7 The result of cloud phase classification

表 4 云相态分类结果

Table 4 The results of cloud phase classification

总的数量	识别一致/%	识别不一致/%	无法识别/%
水云	93.44	6.54	0.02
ROI	88.04	11.64	0.31
HOI	51.11	48.44	0.45

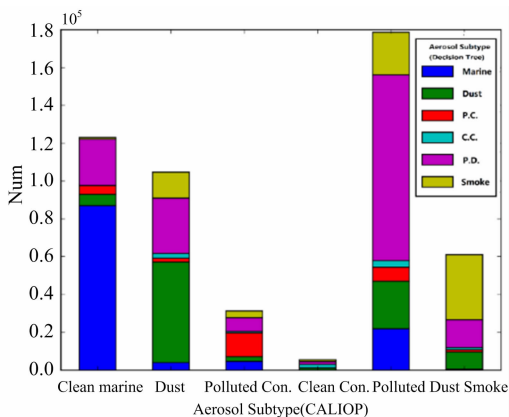


图 8 气溶胶子类型分类结果柱状图

Fig. 8 The result of aerosol subtype classification

图 8 为本文气溶胶子类型反演结果与 CALIOP VFM 产品的比对结果。结果表明,海洋型气溶胶一致性可以达到 84%,煤烟型则可以达到 72%。沙尘型则有 46%的一致性,但有 37%被标记为污染沙尘型。污染大陆型有 43%的一致性,有 29%分为海洋型。对于洁净大陆型,总体数量较少,一致性可以达到 42%,有 38%的分为煤烟型。污染沙尘型的一致性有 54%,有 17%分为煤烟型。总体来说,海洋型、煤

烟型以及污染沙尘型具有较高的一致性；部分沙尘型被标记为污染沙尘型以及污染大陆型被标记为海洋型，原因在于污染沙尘型和污染大陆型气溶胶都被定义为沙尘或大陆型与其他气溶胶混合的类型，因此决策树算法会出现识别混淆。

3.2 不同大气条件下反演结果分析

以华北地区为例，分别选择霾(2017年1月25日)、沙尘(2017年5月4日)和晴天(2017年10月29日)三种大气条件下的反演结果与 CALIOP VFM 产品进行对比, CALIOP 过境轨迹与对应的 MODIS 真彩影像如图 9 所示。

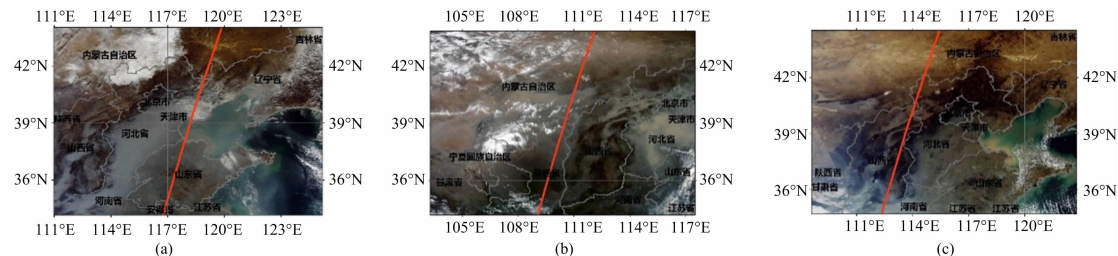


图 9 不同大气条件下 CALIOP 轨迹及 MODIS 真彩图

(a): 2017 年 1 月 25 日, 霾; (b): 2017 年 5 月 4 日, 沙尘; (c): 2017 年 10 月 29 日, 晴空

Fig. 9 CALIOP trajectory and the corresponding MODIS true color image under different atmosphere conditions

(a): Jan 25, 2017, haze; (b): May 4, 2017, dust; (c): October 29, 2017, clear

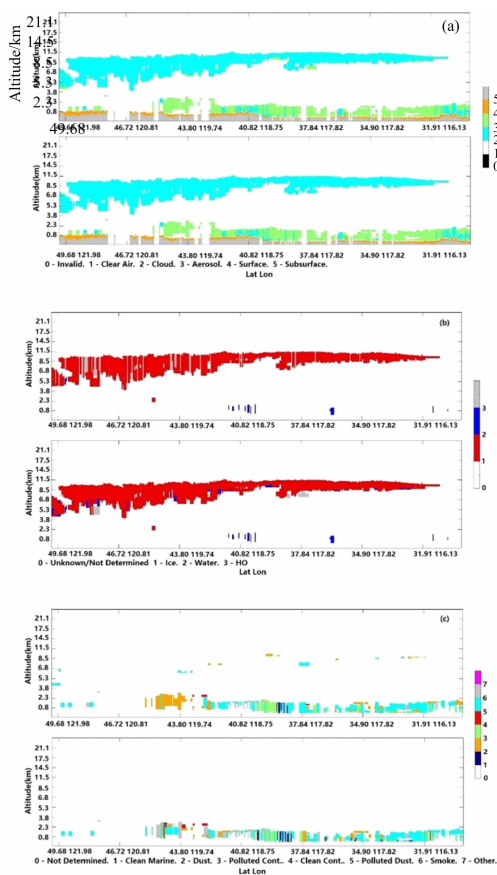


图 10 2017 年 1 月 25 日霾天气 CALIPSO 数据的 (a) 云/气溶胶分类、(b) 云相态分类、(c) 气溶胶子 (VFM 产品(上), 本文算法结果(下))

Fig. 10 The CALIPSO VFM (upper) and the result of this paper (lower) of cloud and aerosol discrimination (a), cloud phase (b) and aerosol subtype classification (c) on Jan 25, 2017

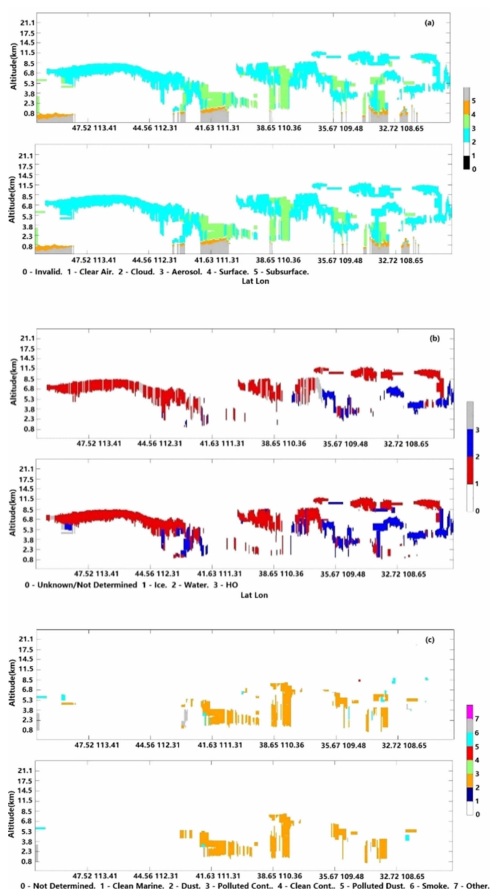


图 11 2017 年 5 月 4 日沙尘天气 CALIPSO 数据的 (a) 云/气溶胶分类、(b) 云相态分类、(c) 气溶胶子类型分类 (VFM 产品(上), 本文算法结果(下))

Fig. 11 The CALIPSO VFM (upper) and the result of this paper (lower) of cloud and aerosol discrimination (a), cloud phase (b) and aerosol subtype classification (c) on May 4 2017

图 10 为 2017 年 1 月 25 日华北地区重污染天气条件下的分类结果对比图,可以看出,云和气溶胶特征层的反演结果与 VFM 产品基本一致。云和气溶胶获得了较好的区分[图 10(a)],但在 5~8 km 处的云下或云中沙尘和污染沙尘气溶胶被识别为云;在云相态分类[图 10(b)]中,水云具有较高的一致性,对于 ROI 和 HOI 两种冰云区分效果不理想;霾天条件的气溶胶子类型识别[图 10(c)]极为重要,从图中可以看出,5 km 以下的气溶胶层的污染沙尘型以及污染大陆性识别较一致,但在 2.5 km 左右(43°N, 119°E)处,本文反演结果为煤烟型气溶胶,而在 VFM 产品中则被标记为沙尘,但华北地区冬季少沙尘而多因供暖烧煤产生的霾,因此需参考地面数据进一步分析。

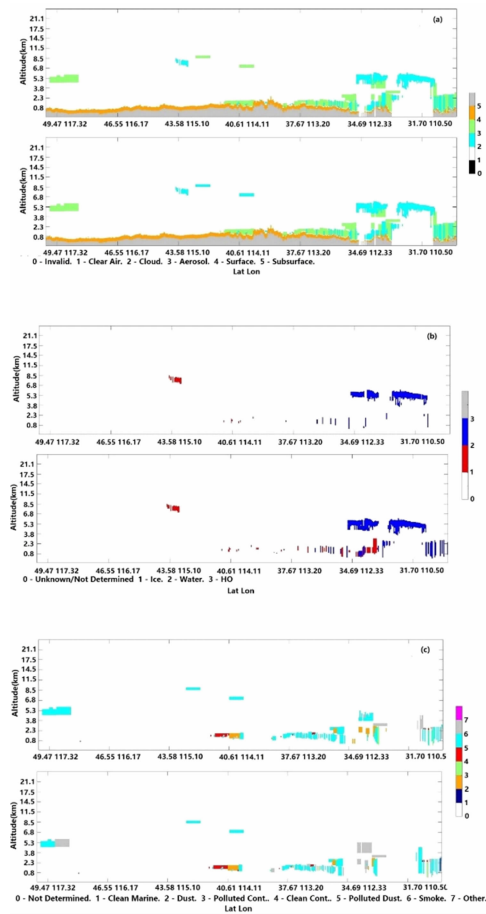


图 12 2017 年 10 月 29 日晴空天气 CALIPSO 数据的 (a) 云/气溶胶分类、(b) 云相态分类、(c) 气溶胶子 (VFM 产品(上), 本文算法结果(下))

Fig. 12 The CALIPSO VFM (upper) and the result of this paper (lower) of cloud and aerosol discrimination (a), cloud phase (b) and aerosol subtype classification (c) on October 29, 2017

2017 年 5 月 3 日—5 日在中国北部地区出现一次强沙尘过程,选择 2017 年 5 月 4 日陕西境内的一组数据作为沙尘天气条件下的典型例子与 VFM 产品对比分析,结果如图 11 所示:云/气溶胶分类结果[图 11(a)]与 VFM 产品基本一

致,且有效反演出沙尘的垂直空间分布[图 11(c)];云相态分类中[图 11(b)],冰云和水云区分效果较好;在气溶胶分类中,大多数云下或云中沙尘气溶胶被误识别为云,下一步工作将着重解决云气溶胶混合层的识别和分解。

选择 2017 年 10 月 29 日北京地区分类结果图进行晴空条件下的对比分析,从图中可以看出,晴空条件下云层以及气溶胶层数量都比较少,云/气溶胶分类结果[图 12(a)]、云相态分类结果[图 12(b)]以及气溶胶分类结果[图 12(c)]与 CALIPSO VFM 产品数据的一致性明显高于重污染天气。但在研究区域的北部内蒙古境内出现的污染沙尘型气溶胶,被判别为煤烟型,以及南部山西境内 2.5 km 处,本文算法结果中识别的 ROI 冰云, VFM 产品中却未能进行成功分类。

4 结 论

发展了一种基于 CALIPSO 2016 年中国地区数据所构建的激光探测的云和气溶胶的光学及空间特征数据库,实现了星载激光雷达的大气特征层快速、有效分类,并将本文反演结果与 CALIPSO 垂直特征层数据进行了对比分析,主要得出以下结论:

(1) 基于激光探测的云和气溶胶层不同的光学以及地理空间分布特征,分别构建云和气溶胶的五维概率密度函数和分类置信函数,基于此实现了云和气溶胶类型的反演。通过将反演结果与 CALIPSO VFM 产品对比,二者的云层反演一致性达到 97.93%,气溶胶层一致性达到 87.85%,得到了较好的效果。选取 SVM 作为 ROI 和水云分类的算法模型基础,结合云层的 γ_{532} , χ , δ , Z 和云顶温度 T 的概率密度函数构建云相态的分类置信函数以修正 SVM 误分的特征层以及筛选出水云中少部分的 HOI 冰云。通过对比分析,水云一致性为 93.44%, ROI 为 88.04%, HOI 为 51.11%。其中 ROI 和 HOI 的雷达比设定相同,二者的混淆对于后续消光系数反演影响极小。将基于决策树算法的气溶胶子类型反演结果与 CALIPSO 产品对比分析,二者海洋型气溶胶一致性为 84%,煤烟型为 72%,沙尘型有 46% 的一致性,污染沙尘型一致性则有 54%,污染大陆型有 43% 的一致性,对于洁净大陆型,总体数量较少,一致性可以达到 42%。二差异性主要是混合型气溶胶的误判,混合型气溶胶定义为多种气溶胶类型的混合类,容易导致决策树分类算法误分。利用现有 CALIPSO 观测结果作为样本数据构建分类数据库,避免了对于地面以及航测数据的依赖,利用机器学习可以自动获取云、气溶胶分类策略,不需要计算过多的属性参数,大大简化了算法的实现过程,使得云气溶胶分类更加高效。

(2) 霾、沙尘以及晴空三种典型情况下的反演结果均与 CALIPSO VFM 产品数据具有较好的一致性。其中,霾天的大部分煤烟型气溶胶反演结果与 VFM 具有较好的一致性。沙尘天能够识别出大多数的沙尘以及污染沙尘。晴空为数不多的气溶胶层结果也较为一致。但是,对于云下以及云中的云/气溶胶混合层的识别还需要加强。

(3) 对于本文实现的星载大气激光雷达特征层分类算法,在下一步工作中,将针对不同时段(日间/夜间)和季节云和

气溶胶表现出不同的特性, 分别构建分类模型以提高分类精度。本文算法分类结果基于雷达比的设定与 CALIOP 产品获得了较好的一致性, 基本满足了对于大气特征层垂直消光系数廓线反演的要求, 但仍需要进一步分析不同相态的云以及对于不同气溶胶子类型光学和空间分布特性来提高分类精

度, 以保证云和气溶胶不同类型的垂直分布特性研究以及应用的精度要求。

致谢: 感谢位于 NASA Langley Research Centre 的 CALIPSO 卫星组提供的 CALIOP 激光雷达探测数据。

References

- [1] Cuesta J, Flamant J P H. *Applied Optics*, 2010, 49(12): 2232.
- [2] Pandit A K, Gadhavi H S, Ratnam M V, et al. *Atmospheric Chemistry and Physics*, 2015, 15(24): 13833.
- [3] Khaykin S M, Godin-Beekmann S, Keckhut P, et al. *Atmospheric Chemistry and Physics*, 2017, 17(3): 1829.
- [4] Sugimoto N, Nishizawa T, Shimizu A, et al. *AEROSOL Classification Retrieval Algorithms for Earth CARE/ATLID, CALIPSO/CALIOP, and Ground-Based Lidars*. in *Geoscience and Remote Sensing Symposium*. 2011.
- [5] Illingworth A J, Barker H W, Beljaars A, et al. , *Bulletin of the American Meteorological Society*, 2015, 31(1): 197.
- [6] Chen B, Minnis P, Hu Y, et al. *Atmos. Chem. Phys.* , 2010, 10: 4241.
- [7] Liu Z Y, Vaughan M, Winker D, et al. *Journal of Atmospheric and Oceanic Technology*, 2009, 26(7): 1198.
- [8] Ma Y Y, Gong Y W. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2012, 5(6): 1849.
- [9] Naeger A R, Christopher S A, Ferrare R, et al. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(1): 642.
- [10] Hu Y X, Winker D, Vaughan M, et al. *Journal of Atmospheric and Oceanic Technology*, 2009, 26(11): 2293.
- [11] Burton S P, Ferrare R A, Vaughan M A, et al. *Atmospheric Measurement Techniques*, 2013, 6(5): 1397.

The Space-Borne Lidar Cloud and Aerosol Classification Algorithms

LI Ming-yang^{1,2}, FAN Meng^{1*}, TAO Jin-hua¹, SU Lin¹, WU Tong^{1,3}, CHEN Liang-fu¹, ZHANG Zi-li⁴

1. State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, Beijing 100101, China
2. Institute of Remote Sensing and Digital Earth, University of Chinese Academy of Sciences, Beijing 100049, China
3. College of Geomatics, Shandong University of Science and Technology, Qingdao 266510, China
4. Zhejiang Environment Monitoring Centre, Hangzhou 310007, China

Abstract LIDAR plays significant roles in monitoring the vertical distribution characteristics of clouds and aerosols and studying their impacts on the global climate change. For the space-born LIDAR, discrimination between clouds and aerosol is the first step of cloud/aerosol vertically optical property retrieve, and to a great extent, the retrieval precision depends on the accuracy of cloud and aerosol classification algorithm. Based on the optical and geographic characteristics of aerosols and clouds observed by LIDAR, in this study, the CALIOP aerosol and cloud products over China in the year of 2016 were trained as the sample sets. An effective cloud/aerosol classification algorithm was developed by combining the support vector machines (SVM) and decision tree methods. Our algorithm includes 3 parts: cloud and aerosol discrimination, ice-water cloud classification and aerosol subtype classification. (1) The cloud and aerosol were discriminated by the classification confidence functions of 5-D probability density function (PDF) with parameters of γ_{532} , χ , δ , Z and lat. (2) Randomly oriented ice (ROI) and water cloud were classified based on the SVM. And by constructing the PDFs with γ_{532} , χ , δ , Z and T , feature layers misclassified by SVM were corrected, and a small portion of the horizontally oriented ice (HOI) clouds were removed from the water clouds. (3) Based on the optical and geographic characteristics of aerosol subtypes, decision tree classification was used for the determination of aerosol subtypes. Our retrieval results showed a good agreement with the CALIOP VFM products. For the cloud and aerosol discrimination results, the consistency ratios between our retrieves and VFM products for aerosol and cloud are up to 98.51% and 88.43%, respectively. And the consistency ratios in the day are higher than those at night. For the cloud phase retrieval results, water clouds can be well separated, and the consistency ratio of water cloud between our retrieves and VFM products is as high as 93.44%. The consistency ratio of HOI is low due largely to the confusion between HOI and ROI. For the aerosol subtype classification, most aerosol subtypes could be well recognized by our algorithm. However, the consistency ratios of the mixed subtypes (e. g. polluted continental and polluted dust) between retrieval results and VFM products are relatively lower. Moreover, the cloud/aerosol, cloud phase and aerosol subtype classifications were also compared with the VFM products under three typical

air conditions, i. e. haze, dust and clean. Under the haze condition, our results for most of the smoke aerosols agree quite well with the corresponding results from VFM. Under the duststorm condition, our algorithm can effectively discriminate the most of dust and polluted dust aerosols. For the clear day, our results for the few existing cloud and aerosol layers are quite consistent with the VFM results. This paper is an important improvement of the cloud and aerosol classification algorithms, which can simplify the processing and improve efficiency with satisfactory accuracy. In the future work, we will build day/night and seasonal training sample sets, and consider more ice cloud phases and aerosol properties in the cloud/aerosol classification retrieval algorithm.

Keywords Space-borne lidar; Cloud and aerosol classification; Probability density function; SVM; Decision tree classification

(Received Dec. 12, 2017; accepted Apr. 23, 2018)

* Corresponding author