

## 基于随机森林的 ICESat-2 卫星数据地表覆盖分类

李彬彬<sup>1</sup>, 谢欢<sup>1\*</sup>, 童小华<sup>1</sup>, 叶丹<sup>1</sup>, 孙凯鹏<sup>2</sup>, 李铭<sup>3</sup>

1. 同济大学测绘与地理信息学院 上海市航天测绘遥感与空间探测重点实验室, 上海 200092;
2. 上海卫星工程研究所, 上海 201109;
3. 中国科学院上海技术物理研究所, 上海 200083)

**摘要:** 该研究将 ICESat-2 卫星激光测高数据作为地表覆盖分类的新数据源, 提出了一种基于随机森林的 ICESat-2 卫星地表覆盖分类方法, 探索了光子计数卫星激光测高在地表覆盖分类中的应用潜力。该方法采用光子数目、不同类型光子水平和垂直分布比例、信噪比、太阳条件、大气条件作为分类的输入, 并在中国长三角地区开展了多类地表覆盖类型分类实验进行了验证。实验结果表明, ICESat-2 卫星的强波束和弱波束的激光数据在水体、森林、低植被以及城市/裸地四类地表的总体分类精度均能达到优于 85%; 在水体、森林以及低植被/城市/裸地三类地表的总体分类精度能达到优于 90% 的水平。

**关键词:** 地表覆盖分类; ICESat-2; 光子计数; 随机森林

中图分类号: P237 文献标志码: A DOI: 10.3788/IRLA20200292

## Land cover classification using ICESat-2 data with random forest

Li Binbin<sup>1</sup>, Xie Huan<sup>1\*</sup>, Tong Xiaohua<sup>1</sup>, Ye Dan<sup>1</sup>, Sun Kaipeng<sup>2</sup>, Li Ming<sup>3</sup>

1. Shanghai Key Laboratory for Planetary Mapping and Remote Sensing for Deep Space Exploration, College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China;
2. Shanghai Institute of Satellite Engineering, Shanghai 201109, China;
3. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China)

**Abstract:** ICESat-2 data was considered as a new land cover classification data source, and a method was proposed to classify land cover using ICESat-2 data with random forest, to explore the application potential of the space-borne photon counting lidar in the land cover classification. The method used the photon number, the proportion of horizontal and vertical distribution of different types of photons, signal-to-noise ratio, solar conditions and atmospheric conditions as the input of classification, and was verified by the experiment of multi-category land cover in China's Yangtze River Delta. For four categories of water, forest, low vegetation and urban/barren, the classification results show that the overall accuracy of strong beam and weak beam is better than 85%. For three categories of water, forest, and low vegetation/urban/barren, the classification results show that the overall accuracy of strong beam and weak beam is better than 90%.

**Key words:** land cover classification; ICESat-2; photon counting; random forest

收稿日期: 2020-07-01; 修订日期: 2020-09-22

基金项目: 国家自然科学基金 (41822106); 上海市科技创新行动计划 (18511102100); 上海市教育发展基金会曙光计划 (18SG22); 湖南省自然资源调查与监测工程技术研究中心开放课题 (2020-8)

作者简介: 李彬彬 (1991-), 男, 博士生, 主要从事卫星激光测高等方面的研究。Email: libinbin@tongji.edu.cn

通讯作者: 谢欢 (1983-), 女, 教授, 博士, 主要从事卫星激光测高与高光谱遥感等方面的研究。Email: huanxie@tongji.edu.cn

## 0 引言

地表覆盖分类对于地球系统科学、全球环境变化和可持续发展研究具有重要意义。传统研究大多采用光学等遥感图像作为土地覆盖分类的数据源。而探索地表覆盖分类新的数据源,如星载激光测高、SAR 影像等,则可作为光学遥感图像的有益补充。

ICESat/GLAS 是首台对地球进行连续全球观测的激光测高仪,通过以 40 Hz 的激光重复频率发射一束波长为 1 064 nm 激光束对地球表面进行探测,最终在沿卫星轨道地面上形成直径约 65 m 的激光足印和 172 m 的激光足印采样间隔<sup>[1]</sup>。以 ICESat 卫星数据作为土地覆盖分类数据源的研究主要有四种:(1) 基于回波波形参数分析,如 Doung 等人依据回波高斯分解后的波形参数,对荷兰实验区域内四类地表(高树/森林、城市、水和裸地/低植被)进行分类,分类结果总精度为 73%<sup>[2]</sup>。(2) 基于机器学习的方法<sup>[3-4]</sup>,如 Ghosh 在印度 Uttarakhand 实验地区,通过使用 K 均值,中心点聚类算法和模糊 C 均值法进行土地覆盖类型分类,分类准确率最高可达 89.41%<sup>[3]</sup>。(3) 波形理论分类器<sup>[5-6]</sup>,或结合机器学习的方法与波形理论分类器<sup>[6]</sup>,如 Li 等人通过结合波形理论分类器和支持向量机分类器,对格陵兰区域的水体和海冰进行分类,分类结果的总体准确率达到 98.21%<sup>[6]</sup>。(4) 联合遥感影像处理提高分类精度,如 Liu 等人通过联合使用 ICESat 卫星数据与 Landsat 多光谱图像,提高中国河南省实验区域内四类地表(耕地、林地、灌丛和水体/裸地)的分类精度,从单一数据源约 85% 分类总体准确率提升到 91%<sup>[7]</sup>。

ICESat 于 2010 年退役,其第二代卫星 ICESat-2 已于 2018 年 9 月由 NASA 发射升空。ICESat-2/ATLAS 能以 10 kHz 的激光重复率发射 3 对激光束(532 nm 波长)对地球表面进行探测。每对波束都包含一个强波束和弱波束(能量比为 4:1),并在沿卫星轨道地面上形成直径约 17 m 的激光足印和 0.7 m 的激光足印采样间隔<sup>[8-10]</sup>。ICESat-2 采用了光子计数的新技术<sup>[11-12]</sup>,在对地观测时具备更高的空间分辨率和更多的地表轮廓描述信息。因此,ICESat-2/ATLAS 有望进一步开发星载激光测高仪在土地覆盖分类中的应用潜力。但由于 ICESat-2 卫星发射时间较晚,基于该卫星数据进行土地覆盖分类的应用研究较少。如 Zhang 等人

根据 ICESat-2 卫星数据在冰雪覆盖区与裸地背景噪声的差异,提出了一种在高原山区中对雪域与裸地进行分类的方法,并在阿克赛钦和帕米尔高原附近的研究区域对该方法进行验证,验证结果表明该方法的总体准确率在 93% 以上<sup>[13]</sup>;Liu 等人基于随机森林和逻辑回归,在北加利福尼亚州和新墨西哥州西部森林区域,进行燃烧区和未燃烧区分类,随机森林和逻辑回归的分类精度分别达到 83% 和 76%<sup>[14]</sup>。

为此,文中提出了一种基于随机森林的 ICESat-2 卫星地表覆盖分类方法,创新性地采用 ICESat-2 卫星产品中光子数目、不同类型光子水平和垂直分布比例、信噪比、太阳条件、大气条件作为分类的输入,并在中国长三角地区开展了多类地表覆盖类型分类实验以进一步探索该类卫星激光测高在多类型地表覆盖分类领域的应用潜力。

## 1 ICESat-2 卫星数据地表覆盖随机森林分类方法

### 1.1 ICESat-2 卫星数据特点

从 ATLAS 发射出的小部分光子在完成首次穿越大气层后,经地球表面反射,再次穿越大气层返回进入 ATLAS 望远镜中,最终被仪器电子设备记录下来。在此过程中,来自 532 nm 波长的太阳光的背景光子可能会到达探测器,且其中的一些背景光子也会被 ATLAS 记录下来。ATLAS 记录到达时间的任何光子称为光子事件,该记录过程与光子的来源无关。因此,ATLAS 记录的光子事件的数量取决于地球表面的几何结构和反射率、太阳条件以及大气中的散射和衰减<sup>[15]</sup>。

ICESat-2 卫星官方通过划分一定间隔的沿轨距离,对所探测到的光子进行分析和处理。不同等级的产品具有不同的沿轨距离间隔,如 ATL03 为 20 m<sup>[16]</sup>,ATL08 为 100 m<sup>[15]</sup>。文中所开展的地表覆盖分类的空间最小分辨率正是各级产品对应的单个沿轨距离间隔。通过分析分布于每个沿轨距离间隔内的探测光子在空间上的分布情况,可以识别出不同的地表覆盖。而每个沿轨距离间隔内有效探测光子(反射于地表的探测光子)的比例则将决定地表分类正确性。

### 1.2 数据预处理

ICESat-2 卫星官方在产品发布前所进行的滤波

工作,可以很大程度地抑制背景噪声光子的影响,但并非完全消除,特别当沿轨距离间隔内的有效探测光子数量过低时。为了更好地保证地表分类的准确性,文中依据了 ICESat-2 卫星产品中的信噪比标签,通过设定门限值对低信噪比 (<3) 的数据进行剔除。

同时,对每个沿轨距离间隔内的有效探测光子进行分析时,希望有效探测光子是均匀分布在整段沿轨距离间隔内。因此,文中对沿轨距离间隔内出现水平方向上超过 20% 有效探测光子缺失的数据进行了剔除。

### 1.3 分类器属性选择

如 1.1 节所介绍的,ICESat-2 卫星数据对地表采用了光子级灵敏度的探测。同时,对不同类别的地表分类时是通过分析分布于每个沿轨距离间隔内的探测光子在空间上的分布情况而进行的。因此,在对不同地表进行分类时,关于分类器属性的选择,文中主要考虑以下几个方面。

首先,在每个沿轨距离间隔内,具有不同反射率的地表所返回的探测光子数是不同的。例如,雪和冰表面在 532 nm 波长的激光束处具有高反射率(典型的朗伯反射系数介于 0.8 和 0.98 之间),而树冠和地形表面的反射率要低得多(土壤通常约为 0.3,植被约为 0.1),预计从地球表面返回的光子数量是冰雪表面的 1/3 到 1/9<sup>[9, 15]</sup>。基于此,文中采用产品在每个沿轨距离间隔内所记录的光子数量作为分类器的属性之一。该属性可以直接通过 ICESat-2 卫星产品中的标签获得。

其次,对于每个沿轨距离间隔,不同地表的探测光子在垂直方向中的不同高度层内,具有不同的比例。例如,当 ICESat-2 卫星对植被区域进行探测时,由于植被冠层的遮挡,分布于地面(或水面)的探测光子比例将低于裸地区域(或水域);以及 ICESat-2 卫星对于森林区域与灌木等低植被区域的冠层内部和顶部进行探测时,由于植被遮挡情况的不同(树冠的茂密程度不同),两者的冠层内部和顶部具有不同的探测光子比例。基于此,文中在每个沿轨距离间隔内,计算被分类为地面光子、被分类为冠层光子、以及被分类为冠层顶层光子各自相对于全部光子的比例作为分类器的属性。上述前三类光子的数量可以直接通过 ICESat-2 产品中的标签获得。该类属性是关于

地表垂直方向上的描述信息,为传统遥感影像所不具有的信息,将有望辅助传统光学遥感图像分类以进一步提高地表分类精度。

另外,对于每个沿轨距离间隔,不同地表的探测光子在水平方向中具有不同的分布特性(或比例)。例如,由于植被冠层的遮挡,裸地(或水体)区域内被分类为地面光子在水平方向上的分布相比于植被区域的地面光子更为均匀。基于此,文中主要在每个沿轨距离间隔内,计算被分类为地面的光子,以及被分类为冠层或冠层顶部的光子在水平方向上的分布比例作为分类器的属性。以 ICESat-2 卫星的 ATL08 产品为例,产品对于每个沿轨距离间隔,在水平方向上均划分为五段等长子间隔,并对每段子间隔内的光子类别标识从而构建指示标签<sup>[9, 15]</sup>。通过此类标签,可以计算地面的光子,和冠层或冠层顶部的光子在水平方向上的分布比例。

最后,考虑到信噪比、太阳条件以及大气条件对有效探测光子的影响<sup>[15]</sup>,文中将这些影响因素纳入分类器属性集内。这些属性可以直接通过 ICESat-2 产品中的标签所获得。

综上所述,文中构建分类器的属性如表 1 所示。

在完成数据预处理和分类器的属性选择后,通过选择训练样本进行随机森林训练可得到分类器模型,而测试样本可以通过该模型实现样本类别的预测,整个过程如图 1 所示。

表 1 分类器属性

Tab.1 Classifier attribute

Attribute	Description
Photon number	Product label
Proportion of terrain photons in vertical direction	Computed
Proportion of canopy photons in vertical direction	Computed
Proportion of top canopy photons in vertical direction	Computed
Proportion of terrain photons horizontal direction	Computed
Proportion of canopy/top canopy photons horizontal direction	Computed
SNR	Product label
Solar elevation	Product label
Solar azimuth	Product label
Cloud confidence flag	Product label

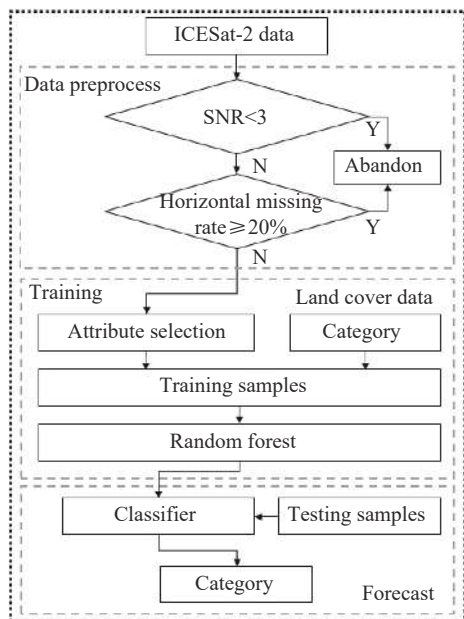


图 1 基于随机森林的 ICESat-2 卫星数据地表覆盖分类流程图

Fig.1 Flowchart of classifying land cover using ICESat-2 data with random forest

## 2 实验分析

### 2.1 实验区域与数据

文中的实验区域位于中国的江苏省、上海市和浙江省,实验数据为 ICESat-2 卫星的 ATL08 产品<sup>[9, 15]</sup>。产品中包含的 IGBP 土地覆盖分类数据可作为参考的地表覆盖数据<sup>[15]</sup>。该数据是通过 2012 年的 MODIS 数据生成的,其空间分辨率为 500 m,涵盖了从森林、

灌木、草场、水体以及城市等 17 个不同类别。考虑到卫星激光测高在地表纹理、反射率等信息上表征的不足从而导致相似地表覆盖分类性能中存在局限性的情况,以及结合相关学者关于高树/森林、城市、水和裸地/低植被四类地表覆盖分类的研究<sup>[2]</sup>,文中首先将水体、森林地表覆盖作为两类地表类别,然后从植被与非植被地表区分角度,初步对低植被,城市和裸地三类地表覆盖划分为低植被和城市/裸地地表覆盖作为后两类地表类别,最后依据上述四类地表覆盖分类结果的混淆矩阵,进一步合并混淆度较大的地表类别以进一步分析分类器性能。依据 IGBP 17 类地表的混淆矩阵<sup>[17]</sup>,可以估算文中采用的水体、森林、低植被以及城市/裸地四类地表的总体分类精度约为 93%。ICESat-2 官方已对 ATL08 产品数据与该数据作了最佳的地理位置匹配<sup>[15]</sup>,可以直接作为该实验的验证数据。

依据 1.2 节中所描述的数据预处理,首先通过 ATL08 产品中每个沿轨距离间隔内(100 m)的探测光子的信噪比标签(SNR),对信噪比低于 3 的数据进行剔除。然后,通过 ATL08 产品中的每个沿轨距离间隔内(100 m)的探测光子在水平方向上 5 段等长子间隔(20 m)类别分布标签(subset\_te\_flag),对超过 20% 有效探测光子缺失的数据进行剔除。剔除后的数据分布如图 2 所示。最后,将土地覆盖分类数据(IGBP 土地覆盖分类数据)重新分为水体、森林、低植被以

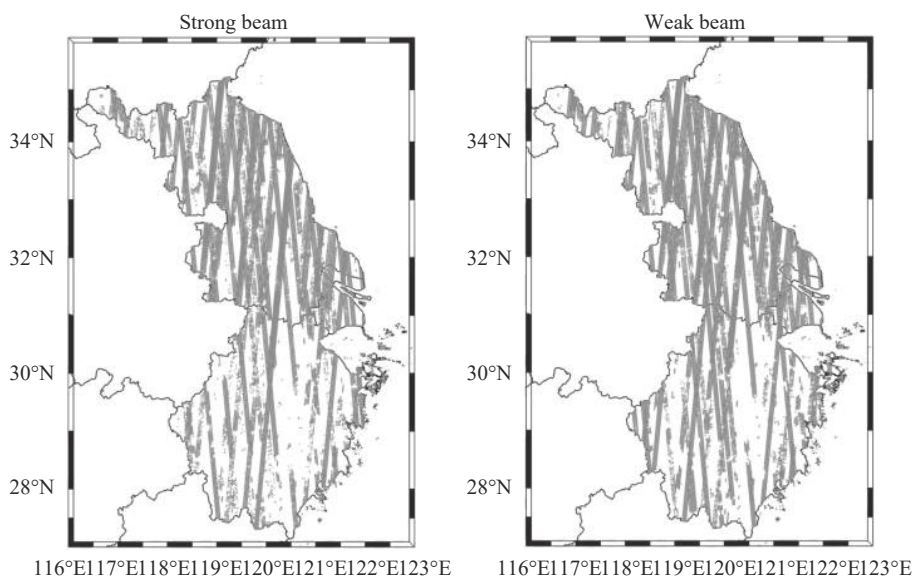


图 2 实验数据的分布

Fig.2 Distribution of experimental data

及城市/裸地四类。四种类别地表的样本数量如表 2 所示。

表 2 四类地表样本数量

Tab.2 Number of four kinds of land cover samples

Class	Strong beam	Weak beam
Water	30 794	31 205
Forest	51 385	80 315
Low vegetation	241 955	279 090
Urban/barren	18 522	19 305

2.2 数据训练与预测处理流程

依据 1.3 节, 完成了属性集的构建, 然后, 对表 2 中的强波束和弱波束数据分别按以下步骤进行处理:

- (1) 分别从各个类别的样本中随机抽取 25% 作为训练样本。
- (2) 对训练样本进行训练, 获得随机森林分类器。
- (3) 以剩余的样本作为测试样本, 通过步骤 (2) 获得的随机森林分类器进行对样本类别进行预测。
- (4) 重复步骤 (1)~(3) 五次, 得到五组随机分类结果。

训练过程中, 随机森林的参数中的 ntrees 设置为

500, mtry 设置为 4。

2.3 地表分类结果

按照 2.2 节的数据处理步骤进行处理后, 可以获得五组强波束和弱波束数据关于四类地表的分类结果, 如图 3 所示。其中, 第一组强波束和弱波束各自的混淆矩阵如图 4 所示。

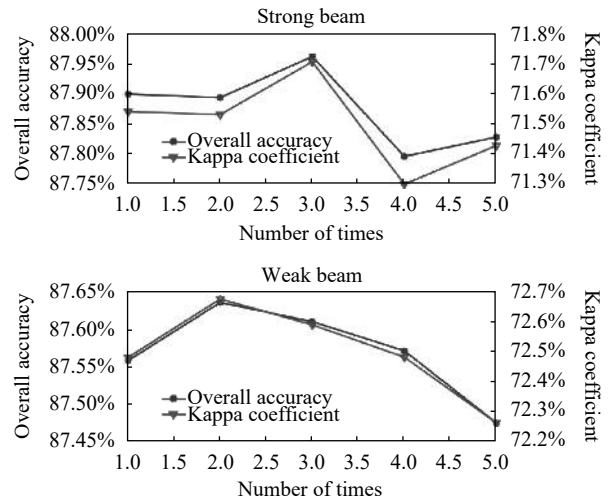


图 3 强波束和弱波束数据关于四类地表的分类结果

Fig.3 Classification results for four kinds of land cover using strong beam and weak beam altimetry data

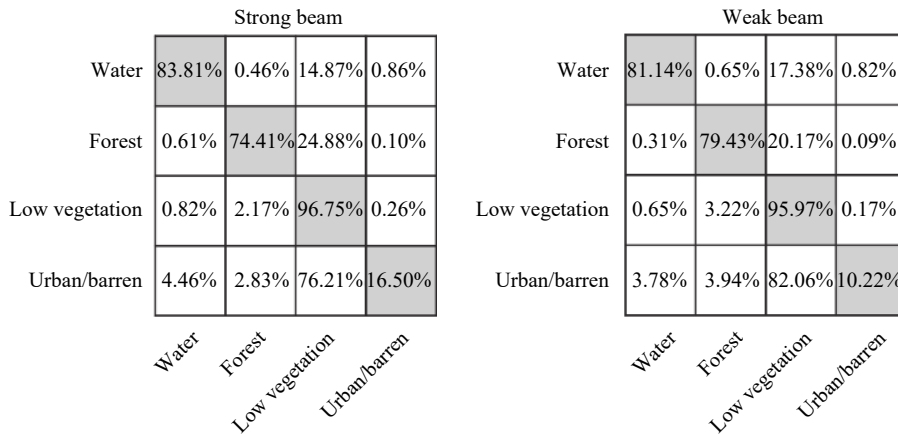


图 4 强波束和弱波束测高仪数据关于四类地表的分类混淆矩阵

Fig.4 Classification confusion matrix of four kinds of land cover using strong beam and weak beam altimetry data

从图 3 中可以看出, 强波束和弱波束的数据在水体、森林、低植被以及城市/裸地四类地表的分类结果较好, 总体分类精度均优于 85%, Kappa 系数均优于 70%。从图 4 中可以发现, 低植被的分类精度最佳, 往后依次为水体、森林和城市/裸地。其中, 城市/裸地类别的地表与低植被类被的地表混淆最大(强波

束: 76.21%, 弱波束: 82.06%), 这可能是部分低植被区域的冠层覆盖率较低, 信号(在这种情况下是树冠)光子可能被错误地识别为数据产品内的太阳背景噪声而被过滤掉<sup>[15]</sup>, 从而与裸地(或城市中空旷区域)混淆。此外, 城市中稀疏的建筑与植被交错区域也将是潜在的混淆源。为了进一步提高分类器的性

能,笔者等尝试将低植被类别的地表与城市/裸地类别的地表作为相同类别的地表,从而开展水体、森林、以及低植被/城市/裸地三类地表分类实验。该地表分类实验结果如图 5 所示。从图 5 中可以看出,分类器的性能得到进一步的提高,强波束和弱波束的总体精度均提高至 90% 以上, Kappa 系数均提高至 79% 以上。

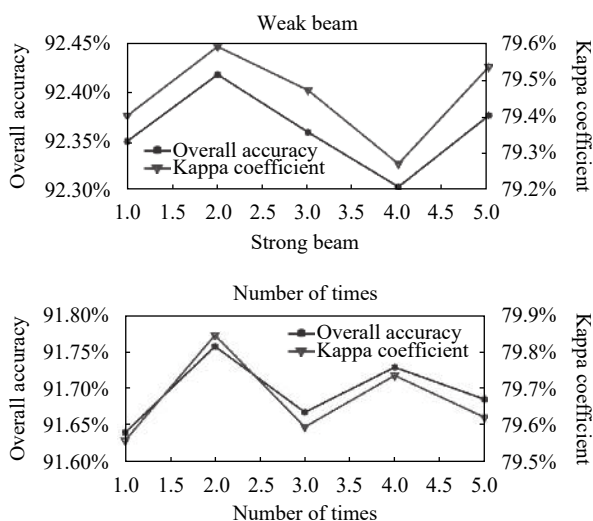


图 5 强波束和弱波束测高仪数据关于三类地表的分类结果

Fig.5 Classification results for three kinds of land cover using strong beam and weak beam altimetry data

### 3 结论

文中提出了一种基于随机森林的 ICESat-2 卫星地表覆盖分类方法,并通过在中国长三角地区开展的多类地表覆盖类型分类实验对该方法进行了验证。验证结果表明,在水体、森林、低植被以及城市/裸地四类地表中,强波束和弱波束的总体精度均优于 85%, Kappa 系数均优于 70%;水体、森林以及低植被/城市/裸地三类地表中,强波束和弱波束的总体精度均提高至 90% 以上, Kappa 系数均提高至 79% 以上。该方法总体上所具有的良好分类性能,特别是分类过程中所兼顾的地表垂直方向上的描述信息(探测光子在垂直方向中的不同高度层内的分布比例),将有望辅助传统光学遥感图像分类以进一步提高地表分类精度。更多地,该方法在进行四类地表分类中,城市/裸地该类别的地表分类精度表现性能较差,主要是因为与低植被类别的地表混淆较大。这是在未来的工作中需要进一步研究改进的方向。

### 参考文献:

- [1] Schutz B E, Zwally H J, Shuman, C A, et al. Overview of the ICESat mission [J]. *Geophysical Research Letters*, 2005, 32(21): L21S01.
- [2] Duong H, Pfeifer N, Lindenbergh R. Full waveform analysis: ICESat laser data for land cover classification[C]//In proceedings: ISPRS Mid-term Symposium, Remote Sensing, 2012, 36: 31-35.
- [3] Ghosh S, Nandy S, Patra S, et al. Land cover classification using ICESat/GLAS full waveform data [J]. *Journal of the Indian Society of Remote Sensing*, 2017, 45(2): 327-335.
- [4] Quan Xuefeng, Tang Xinming, Li Guoyuan, et al. Land cover classification application of satellite laser altimetry data: a case study in Beijing, China [J]. *Remote Sensing Information*, 2019, 34(6): 6-11. (in Chinese)  
权学烽, 唐新明, 李国元, 等. 基于激光测高卫星全波形数据的地表覆盖分类应用—以北京市城区为例[J]. *遥感信息*, 2019, 34(6): 6-11.
- [5] Ma Yue, Zhang Wenhao, Zhang Zhiyu, et al. Sea and sea-ice waveform classification for the laser altimeter based on semi-analytic model [J]. *Infrared and Laser Engineering*, 2018, 47(5): 0506005. (in Chinese)  
马跃, 张文豪, 张智宇, 等. 基于半解析模型的激光测高回波海水海冰波形分类方法[J]. *红外与激光工程*, 2018, 47(5): 0506005.
- [6] Li S, Zhang W H, Ma Y, et al. Theoretical surface type classifier based on a waveform model of a satellite laser altimeter and its performance in the north of Greenland [J]. *Applied Optics*, 2018, 57(10): 2482.
- [7] Liu C X, Huang H B, Gong P, et al. Joint use of ICESat/GLAS and landsat data in land cover classification: a case study in Henan Province, China [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(2): 511-522.
- [8] Markus T, Neumann R, Martino A, et al. The ice, cloud, and land elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation [J]. *Remote Sens Environ*, 2017, 190, 260: 273.
- [9] Neuenschwander A, Pitts K. The ATL08 land and vegetation product for the ICESat-2 Mission [J]. *Remote Sensing of Environment*, 2019, 221: 247-259.
- [10] Xia Shaobo, Wang Cheng, Xi Xiaohuan, et al. Point cloud filtering and tree height estimation using airborne experiment data of ICESat-2 [J]. *Journal of Remote Sensing*, 2014, 18(6):

- 1199-1207. (in Chinese)
- 夏少波, 王成, 刁晓环, 等. ICESat-2机载试验点云滤波及植被高度反演[J]. 遥感学报, 2014, 18(6): 1199-1207.
- [11] Xie Feng, Yang Gui, Shu Rong, et al. An adaptive directional filter for photon counting lidar point cloud data [J]. *Journal of Infrared and Millimeter Waves*, 2017, 36(1): 107-113. (in Chinese)
- 谢锋, 杨贵, 舒嵘, 等. 方向自适应的光子计数激光雷达滤波方法[J]. 红外与毫米波学报, 2017, 36(1): 107-113.
- [12] Xu Yiteng, Li Guoyuan, Qiu Chunxia, et al. Single photon laser data processing technology based on terrain correlation and least square curve fitting [J]. *Infrared and Laser Engineering*, 2019, 48(12): 1205004. (in Chinese)
- 许艺腾, 李国元, 邱春霞, 等. 基于地形相关和最小二乘曲线拟合的单光子激光数据处理技术[J]. 红外与激光工程, 2019, 48(12): 1205004.
- [13] Zhang Z, Xu N, Ma Y, et al. Land and snow-covered area classification method based on the background noise for satellite photon-counting laser altimeters[J]. *Optics Express*, 2020, 28: 16030-16044.
- [14] Liu M, Popescu S, Malambo L. Feasibility of burned area mapping based on ICESAT-2 photon counting data [J]. *Remote Sensing*, 2019, 12(1): 24.
- [15] Neuenschwander A, Pitts K, Jelley B, et al. Ice, Cloud, and Land Elevation Satellite 2 (ICESat-2) algorithm theoretical basis document (ATBD) for land - vegetation along-track products (ATL08)[R]. ICESat-2 Algorithm Theoretical Basis Document, 2020.
- [16] Neumann T, Brenner A, Hancock D, et al. Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2) project algorithm theoretical basis document (ATBD) for global geolocated photons ATL03[R]. ICESat-2 Algorithm Theoretical Basis Document, 2019.
- [17] Friedl M A, Sulla-Menashe D, Tan B, et al. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets[J]. *Remote Sensing of Environment*, 2010, 114: 168-182.