doi:10.3969/j.issn.1003-2029.2025.06.001

基于点云空间特征的随机森林海面光子提取方法

王晓明[1]，张豪帅[2]，胡文慧[1]，亓 超[2,3]

（1. 中交星宇科技有限公司，北京 102200；2. 山东科技大学测绘与空间信息学院，山东 青岛 266590；
3. 自然资源部海洋测绘重点实验室，山东 青岛 266590）

摘 要：冰、云和陆地高程卫星 2 号（Ice, Cloud, and land Elevation Satellite-2，ICESat-2）搭载的高灵敏度光子计数激光雷达为海面探测开辟了新途径。然而，受到传感器对光子信号的极高敏感性及海气跨介质反射特性的双重约束，原始数据中表现出海面有效光子与噪声光子混杂的特点，限制了数据的直接应用。为此，本文提出了一种基于多维度空间特征挖掘的随机森林分类框架，建立从粗提取到精确分类的两级处理策略。首先基于高程的光子数量直方图的信噪比分析，实现海面光子的粗提取，继而通过人工标注构建高置信度训练数据集；提取 7 维特征参量训练分类模型以进一步优化分类结果。为验证方法优势，利用上述方法提取青岛近海区域日夜间数据海面光子，并与具有噪声的基于密度的聚类算法（Density-Based Spatial Clustering of Applications with Noise，DBSCAN）进行对比分析。结果表明：DBSCAN 算法虽能实现海面光子提取（搜索半径 1 m 时准确率较高），但其性能受参数敏感性影响，鲁棒性较低；而本文提取算法通过构建光子空间特征，在日间和夜间数据中分别取得 96.92%与 99.75%的总体分类精度，效果优于传统方法。同时，引入高斯拟合算法提取一类水体（夏威夷拉奈岛）与二类水体（青岛近海）数据海面点并与本文方法对比，验证了本文方法在不同海域的鲁棒性。因此，本方法有效解决了强噪声背景下信号提取难题，证实机器学习策略在单光子遥感数据处理中的显著优势，为星载光子计数数据的海洋学应用提供了新的方法论支撑。

关键词：ICESat-2；海面探测；随机森林；DBSCAN
中图分类号：P229 **文献标识码：**A **文章编号：**1003-2029（2025）06-0001-10

海面探测在海洋科学研究、气候变化监测和海洋资源管理等领域具有重要意义。全球海平面变化是气候变化的重要指标之一[1]，高精度的海面高程测量对于研究冰盖消融、海洋热膨胀及极端气候事件具有关键作用[2]。此外，海面形态的动态变化直接影响海洋动力过程，如洋流、潮汐和风暴潮的演变，对沿海防灾减灾和海洋工程规划至关重要[3]。借助卫星测高技术，科学家能够突破传统测量手段的局限，实现对全球海洋的高分辨率、长时间序列监测[4-5]。然而，海洋环境的复杂性和遥感技术的局限性使得海面光子数据的精准提取成为核心难题。因此，研究高效、准确的海面信号光子提取方法，不仅有助于提升卫星遥感数据的应用价值，也将为全球海洋观测和科学研究提供重要技术支撑。

随着卫星遥感技术的发展，星载激光雷达成为获取全球高精度海面高程信息的关键工具[6]。尤其是 2018 年美国国家航空航天局（National Aeronautics and Space Administration，NASA）发射的ICE-

收稿日期：2025-04-25
基金项目：2024 年中交集团青年创新项目（PJ2024033354）；山东省自然科学基金资助项目（ZR2024QD062）
作者简介：王晓明（1994—），男，硕士，工程师，主要从事卫星导航定位及海洋测绘数据处理研究。E-mail：sdust_wxm@163.com
通信作者：亓超（1994—），男，博士，副教授，主要从事海洋测绘与遥感水陆地形地貌探测研究。
E-mail：qichoice007@sdust.edu.cn

Sat-2 卫星，其搭载的先进地形激光测高系统（Advanced Topographic Laser Altimeter System，ATLAS）首次引入了微脉冲光子计数技术，实现了海面高程测量从传统雷达到单光子激光雷达的跨越式发展[7-9]。与传统雷达相比，单光子激光雷达对光子信号更敏感，能够识别返回时间近似，能量微弱的光子，故而空间分辨率极高[10]。但其信号光子数据中常混杂大量噪声光子（如太阳背景噪声、大气散射噪声）[11-12]，为数据处理带来了显著挑战。

海面光子提取本质是解决动态波浪调制下信号–噪声在时空域的混叠问题。现有方法可分为三类：①统计滤波，依赖全局分布假设而忽略局部波浪结构；②物理模型，受限于海浪理论的理想化假设；③空间聚类，虽能适应波浪非均匀性但对参数敏感。基于第一类方法，许艺腾等[13]通过累加距离直方图的方式实现了噪声的剔除；陈博伟等[14]、夏少波等[15]通过局部距离统计对多高度计光束实验激光雷达（Multiple Altimeter Beam Experimental Lidar，MABEL）数据进行去噪。此类方法大多基于陆源数据进行实验，在海洋环境中的提取效果未得到有效证明。为此，基于第三类方法，利用信号光子空间分布密集而噪声光子分布相对稀疏的特点，通过密度聚类算法如点序搜索（Ordering points to identify the clustering structure，OPTICS）、DBSCAN 等[8]，直接将单光子数据分类为噪声光子点和信号光子点[16-17]，成为研究热点。张少华等[18]先对单光子数据进行粗去噪，随后基于改进的 DBSCAN 算法进行精去噪，去噪效果极好；ZHANG J 等[19]基于改进的椭圆搜索形状的 DBSCAN 密度聚类算法对 MABEL 数据进行分类；周智标等[20]通过椭圆邻域搜索及自适应密度阈值粗略提取海面点，随后利用高斯拟合高程直方图去除错分点的方法提取精确的海面信息；杨鑫科等[21]通过高斯拟合确定海面高程上下界的方法确定海面光子。目前基于物理理论的提取方法研究相对较少，其中 MA Y 等[22]通过联合北海波浪计划（Joint North Sea Wave Project，JONSWAP）海浪谱对海面轮廓的建模研究提取海面光子，取得了较好的结果。尽管已有研究取得了一定进展，但现有方法在海面光子提取方面仍存在诸多挑战。一方面，基于陆源优化的算法难以适应波浪调制下的动态变化的光子分布特征；另一方面，密度聚类算法在应对密度不均匀与连接性弱的海面光子点时[23]，缺乏足够的鲁棒性，限制了其在高动态海况中的适用性。因此，针对海面光子提取方法的进一步优化，仍是当前亟待解决的研究问题。

针对单光子数据处理中对海面单光子数据提取研究较少且精度不高的问题，本文提出了一种基于点云空间特征的随机森林海面光子提取方法，通过提取七类单光子数据的空间特征以判断光子点是否为海面信号。使用该方法对青岛近海区域 ICESat-2 的数据进行处理，并将分类结果与 DBSCAN 算法的提取结果做对比，使用人工标记的海面点作为真值验证本文算法识别精度。

1 研究区域及数据

1.1 研究区域

本文选取青岛市黄海沿岸海域作为研究区域，裁剪出区域内长沿轨距离日间轨道数据与夜间轨道数据。日间数据为 2019-09-18，RGT#1253，GT1R（图 1(a)红色实线为研究区日间数据），夜间数据为 2022-09-17，RGT#1337，GT1R（图 1(b)绿色实线为研究区夜间数据）。

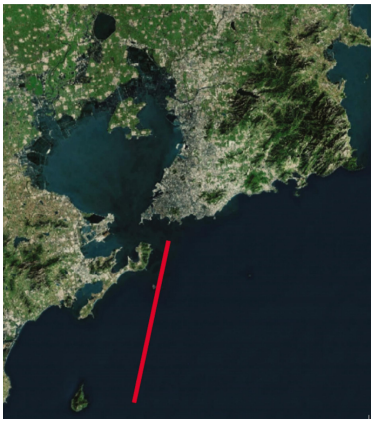
研究区域地形平坦，海洋环境稳定，为开展高精度单光子激光雷达技术研究提供了优越的实验条件。海域内海面波动幅度小，绝大多数海浪高度低于 0.5 m，有效避免了大范围波浪和强海流对激光测量的干扰，为验证单光子激光雷达高程数据处理方法及分类算法的精确性提供了理想环境。

1.2 ICESat-2 高程数据

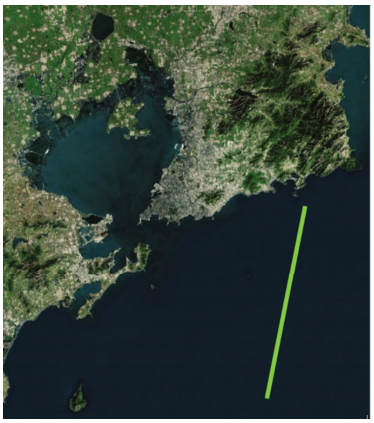
本文基于 ICESat-2/ATLAS 激光雷达系统提供的 ATL03 产品开展研究。该数据集记录了光子的时间标签、高程、经纬度及置信度等参数信息。其中，高程和经纬度信息基于世界大地测量系统（World Geodetic System，WGS84）参考椭球系统[24]，具有较高的精度。因此只需有效分离海面光子与噪声光子，即可获得高精度的海面高程信息。

本文选取了青岛市黄岛区东侧近海海域的日间和夜间两轨数据，以此评估太阳背景噪声对海面光子提取的影响。针对日间轨道数据中包含部分陆地光子信号的情况，为避免其对研究结果产生干扰，

对数据进行了裁剪，提取出纬度小于 36°N 的单光子数据作为研究数据。图 2(a)展示了裁剪后的日间轨道数据点云图像，覆盖了近两万米长的沿轨距离区域。图 2(b)显示了完整的夜间数据点云图像。从这两幅原始图像中可以清晰地观察到，海域单光子数据在垂直维度上呈现“双峰分布”，主峰对应平均海平面，次峰反映水下噪声与波浪飞沫干扰；在水平维度，光子分布呈准周期性聚集，局部密度变化反映波浪破碎过程，表现了单光子数据对海面的高敏感性。

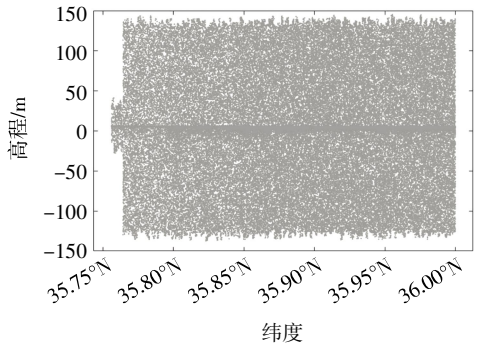


(a) 日间数据

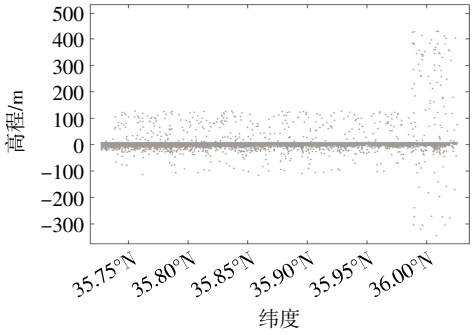


(b) 夜间数据

图 1　研究区域示意



(a) 日间数据



(b) 夜间数据

图 2　两轨数据的原始点云图像

2　海面光子提取算法

本文提出的海面光子信号提取方法流程如图 3 所示，主要包含单光子数据置信度评估、随机森林模型的构建与训练、两种海面光子信号提取算法及其结果的对比分析等。

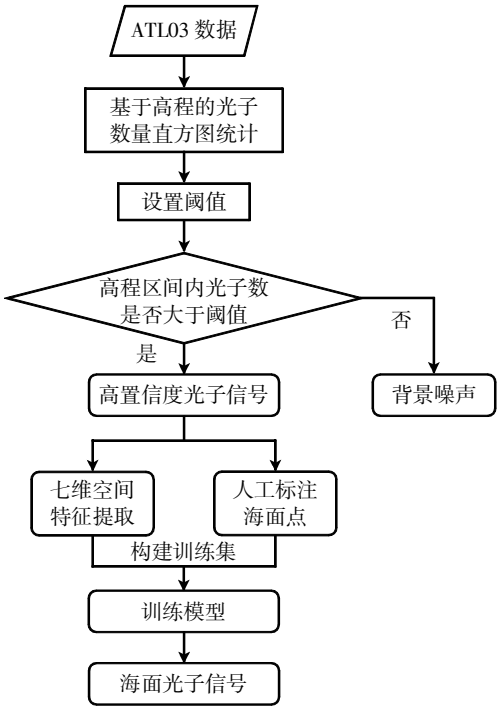


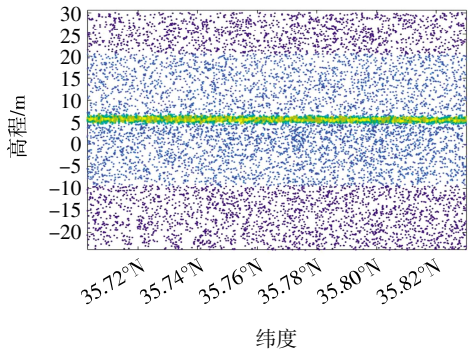
图 3　本文方法流程

2.1　海面光子粗提取

ATL03 将每个光子事件分类为可能的信号光子事件或背景光子事件，并提供对这些分类的置信度评估。这种分类是基于光子事件的空间分布统计特性，通过分析光子密度随高度的变化区分海面信号与背景噪声。具体方法中，光子事件首先按高度坐

标分层（如 15 m 或 30 m 的垂直区间），统计每个区间内的光子数量并生成直方图[22]，海面信号通常表现为局部光子密度的显著峰值，而背景噪声则呈低密度均匀分布。通过计算每个高度区间的信噪比（Signal-to-Noise Ratio，SNR）实现光子分类：设定动态阈值（通常高置信度阈值范围固定在 3~5 区间内）后，光子事件被划分为高、中、低三个置信度等级[25]，高置信度光子对应 SNR 显著超过阈值且空间聚集性强的地表反射信号，中置信度光子需结合时空连续性验证以排除孤立噪声，低置信度光子则直接归为背景噪声[6]（如大气散射或探测器噪声）。

如图 4 所示，以日间数据为例，低置信度光子数为 7 948，中置信度光子数为 8 798，高置信度光子数为 5 819，图 4 中可以看出利用置信度进行信号粗提取存在明显信号分类错误的情况。剔除其中的噪声点（紫色点）与低置信度光子点（蓝色点），可显著降低机器学习中人工标记真实海面光子点的工作量。



蓝色为低置信度（7 948 个光子）；绿色为中置信度（8 798 个光子）；黄色为高置信度（5 819 个光子）。

图 4 点云信号粗提取示意图

2.2 基于点云空间特征的随机森林算法提取海面光子

光子信号分类的研究早期主要基于非监督分类方法进行点云滤波。近年来，机器学习方法在激光雷达点云分类中的应用取得了较好效果。本文算法通过随机选择数据和特征构建多个决策树，并将其结果整合的方式提取海面信号光子。由于本文需要对单光子数据点云进行监督分类，因此需要获取单光子数据的真实分类结果作为训练和验证的基础。为此，本文采用人工判读的方法，对单光子数据中的信号属性进行逐一标记，包括海面光子与噪声光子类别。这些人工标记的真实结果为后续基于点云空间特征的监督分类研究提供了可靠的数据支持。

随机森林的随机性体现在两个方面。第一，训练决策树时，从总训练数据中随机地抽取数据。第二，决策树判断使用的特征是从总特征中随机抽取小于总特征个数的特征。其具体步骤如下。

（1）提取训练样本与测试样本：打乱下载的研究区域单光子数据的顺序，提取其中的前 80%数据中的光子高程、经纬度信息作为训练样本，其余的 20%作为测试样本。

（2）特征构建和提取：根据海面光子信号的空间特征及单光子数据自身的特点计算 7 种空间特征，主要包括：单光子数据高程，局部邻域内各点与该邻域高程均值、中值及不同的高度分位数（10、25、50、75）之差[14]。根据研究区域光子数据空间分布，波浪垂直波动范围在 5 m 内（图 4），10 m 窗口可覆盖 2 倍标准差范围，确保特征稳定性。因此，局部窗口范围设置为 10 m，各空间特征定义如表 1 所示。

（3）数据整理：将提取出的特征放入同一矩阵中为接下来的随机森林分类准备。将经过人工判读的海面信号点云和噪声点云也放入矩阵中作为真实结果。

（4）决策树训练：随机抽取一定个数的特征构建多个决策树，从训练样本中随机抽取一定数量数据输入到决策树中进行训练，得到初步结果。将初步分类结果与人工标定的真实结果进行对比，根据对比结果修正分类模型，设置预测精度阈值，输出预测精度最高的决策树模型。将所有决策树集合起来，得到随机森林模型。

表 1 单光子数据空间特征[14]

编号	特征名称	特征含义
1	Hight	单光子数据高程
2	Mean	局部邻域内各点与该邻域高程均值差值
3	Median	局部邻域内各点与该邻域高程中位数差值
4	Median 10	局部邻域内各点与该邻域 10%分位数高程差值
5	Median 25	局部邻域内各点与该邻域 25%分位数高程差值
6	Median 50	局部邻域内各点与该邻域 50%分位数高程差值
7	Median 75	局部邻域内各点与该邻域 75%分位数高程差值

（5）模型验证：将测试样本输入模型，输出模型预测的分类结果并与真实分类结果对比，以此对训练出的随机森林模型进行验证。

（6）性能评价：对分类结果进行精度评价，利用多种精度评价公式进行计算，验证模型分类性能，如 Kappa 系数等，计算公式如下。

$$P_0 = \frac{a_1 + a_2 + \cdots + a_n}{N} \tag{1}$$

$$P_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \cdots + a_n \times b_n}{N \times N} \tag{2}$$

$$k = \frac{P_0 - P_e}{1 - P_e} \tag{3}$$

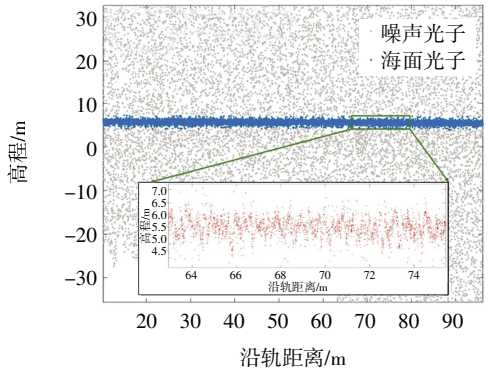
式中，a_1，a_2，…，a_n 为每一类别中的真实样本个数；b_1，b_2，…，b_n 为模型计算后预测的每一类的样本个数；N 为数据总个数；n 为类别个数；P_0 为总体分类精度，是每一类正确分类的样本数量之和除以总样本数。Kappa 系数的值域为[−1，1]，但通常结果为[0，1]。

3　结果与讨论

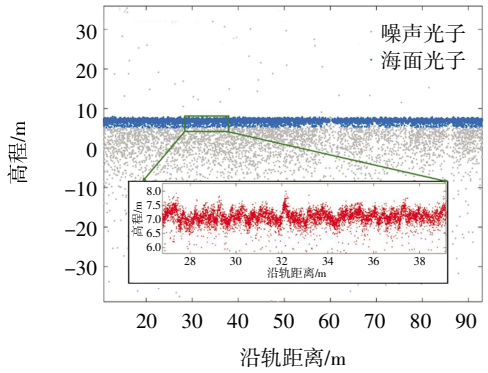
3.1　分类结果对比分析

本文分别利用 DBSCAN 算法和基于点云空间特征的随机森林算法对单光子数据的海面光子进行提取。DBSCAN 算法作为密度聚类的代表性算法，其被广泛用于 ICESat-2 数据处理；相较于基于统计的 OPTICS、基于波形的高斯拟合等方法，DBSCAN 更适配海面光子的空间连续性特征。

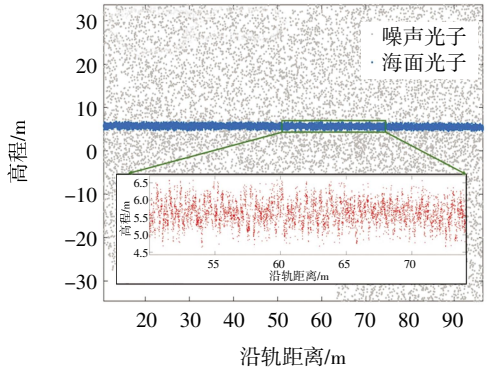
通过网格搜索测试半径参数在 0.1~2 m 半径范围内 DBSCAN 算法分类效果，发现半径 R 为 1 m时提取效果最接近人工标注的波浪聚集尺度（图 5(a)和图 5(b)）。从 DBSCAN 算法提取结果图（图 5）可知，该算法在一定程度上可有效提取海面光子，但其性能高度依赖于输入参数的设置。例如，当搜索半径为 1 m 时，算法在日夜间数据上表现较好（图 5(a)和图 5(b)），能够完整地识别海面光子，但也出现了较为明显的海面点误判现象；当搜索半径变至 0.5 m 时（图 5(c)和图 5(d)），受限于海面光子信号的非均匀分布，海面点提取后存在空白区域，夜间数据表现明显（图 5(d)）。相比之下，本文算法提取结果图（图 6）清晰展现出海面的瞬时波浪形态特征，能够有效区分噪声光子与海面光子。从夜间数据的海面光子提取结果（图 6(a)和图6(c)）与 DBSCAN 提取结果（图 5(b)至图 5(d)）对比中可以看出，本文算法针对水面以下贴近水面



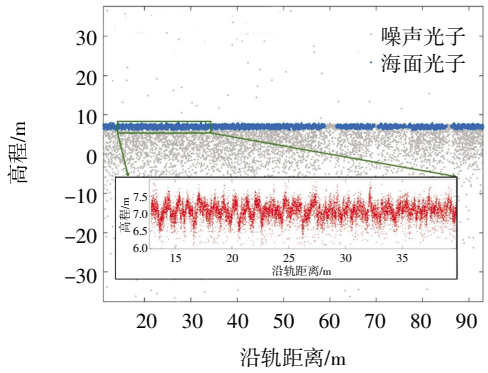
(a) 日间数据海面点提取结果（R=1）



(b) 夜间数据海面点提取结果（R=1）

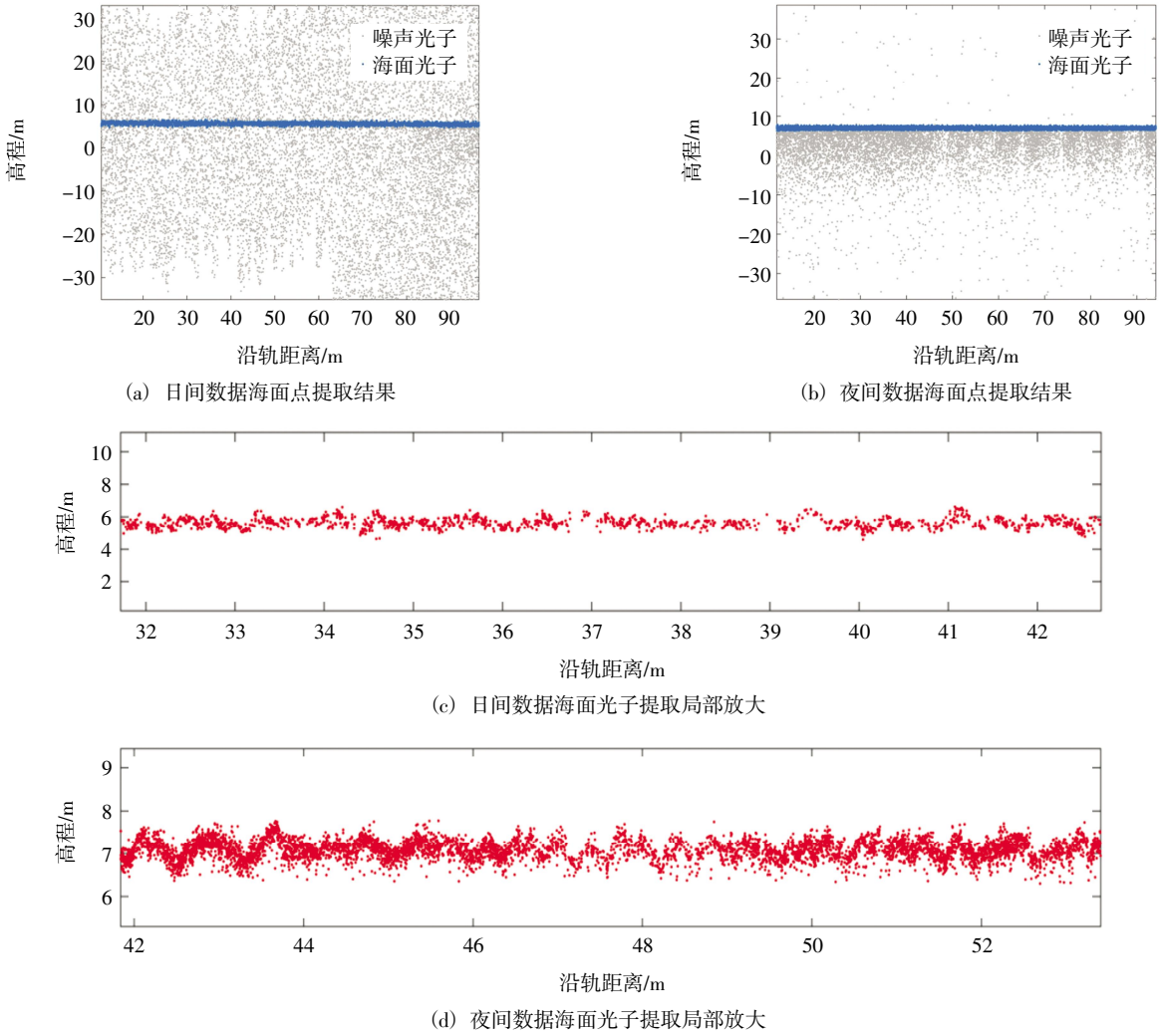


(c) 日间数据海面点提取结果（R=0.5）



(d) 夜间数据海面点提取结果（R=0.5）

图 5　DBSCAN 海面光子提取结果



(a) 日间数据海面点提取结果

(b) 夜间数据海面点提取结果

(c) 日间数据海面光子提取局部放大

(d) 夜间数据海面光子提取局部放大

图 6 日夜间数据基于点云空间特征的海面光子提取结果

处噪声光子的识别效果明显。与 DBSCAN 算法相比，本文所提算法在分类精度和稳定性方面表现出显著优势，尤其是在水下贴近水面处反射光子点的干扰下仍能保持较高的分类准确性。

3.2 分类结果定量分析

以本文算法对青岛近海海域两个研究区域数据的提取结果为例，利用多种精度评价指标对随机森林模型的性能进行定量衡量。通过混淆矩阵计算得出生产者精度（Producer's Accuracy，PA）、用户精度（User's Accuracy，UA）、总精度（Overall Accuracy，OA）及 Kappa 系数等指标（表 2）。日间数据海面点分类总体精度为 96.92%，Kappa 系数为 0.90；夜间数据海面点分类总体精度为 99.75%，Kappa 系数为 0.98。同时，本文结合去噪精确度（precision）、去噪召回率（recall）和 F1 指标（F1-score）对分类结果进行进一步评价。实验结果表明，日间数据模型有效性为 0.92，夜间数据为 0.99，这些结果共同表明太阳光光子信号等复杂背景噪声对分类精度影响显著。其他精度指标也呈现类似趋势，进一步验证了夜间环境下该算法的优越性。

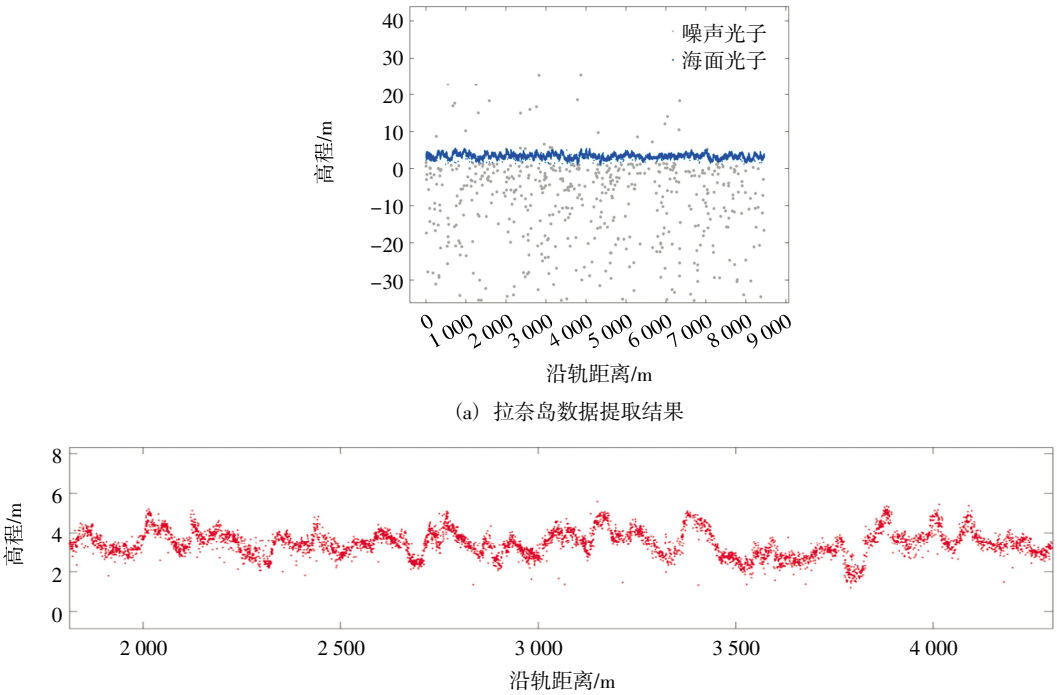
表 2 海面光子提取结果及各精度指标

研究区	光子总数/个	海面光子点/个	噪声光子点/个	预测海面光子数/个	预测噪声光子数/个	PA 值/%	UA 值/%	OA 值/%	Kappa 值	F1-score
日间数据研究区	27 000	5 553	21 447	5 386	20 782	96.99	89.01	96.92	0.90	0.92
夜间数据研究区	15 622	13 649	1 973	13 631	1 952	99.86	98.83	99.75	0.98	0.99

4　讨　论

为了验证所提方法在不同海域的适用性，本文选取二类水体（青岛近海，如 1.2 节所示）与一类水体（夏威夷海域）典型海况下单光子数据，系统分析了本文算法的环境适应能力。如图 6 所示，在青岛近海复杂光学环境下，算法成功实现了 96.92%的分类精度（Kappa 值为 0.90），特别是对波浪破碎区光子簇的完整保留与镜面反射噪声的有效剔除。转向夏威夷开阔大洋环境，如图 7 所示，算法通过调整邻域窗口阈值（10 m→15 m），在波浪动态范围变化剧烈的挑战下仍保持 94.17%的分类精度（Kappa 值为 0.88），验证了本文算法的鲁棒性。

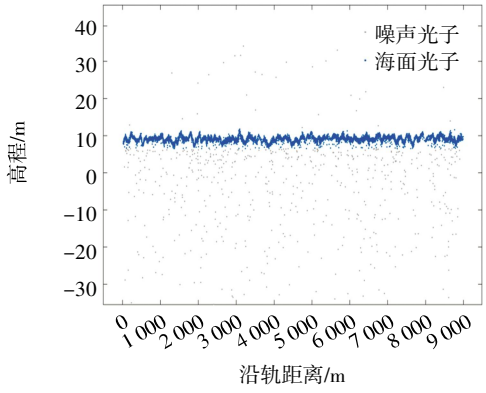
为全面评估随机森林算法的优势，本文引入高斯拟合方法作为传统方法的代表进行对比。如图 8 所示，高斯拟合方法在提取出海面点的同时对水下噪声不能有效去除（图 8(b)椭圆所示）。同时，使用高斯拟合提取青岛近海夜间数据的海面点，如图 9 所示，可以看出在二类水体区域受水质影响，噪声光子显著增强，对于由波浪飞沫与水体造成的噪声光子无法剔除，鲁棒性较差。通过分析本文算法对拉奈岛与青岛近海数据分类结果并与高斯拟合在两类区域的分类结果对比，可以有效论证本文算法在提取海面点时的稳定性。



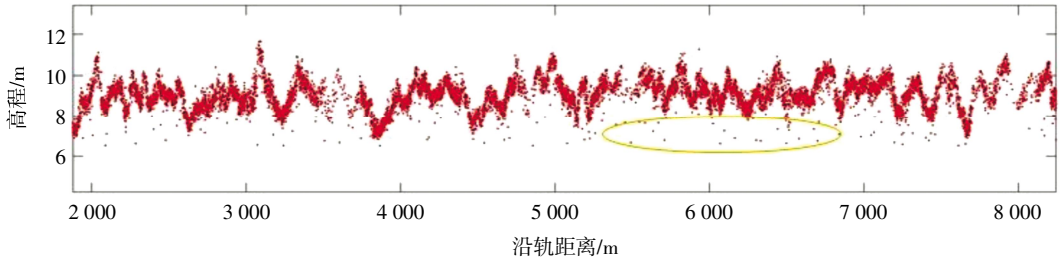
(a) 拉奈岛数据提取结果

(b) 拉奈岛数据提取结果局部放大

图 7　拉奈岛数据基于点云空间特征的海面光子提取结果

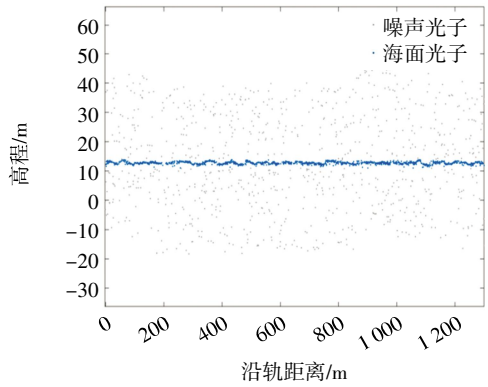


(a) 拉奈岛数据高斯拟合提取结果

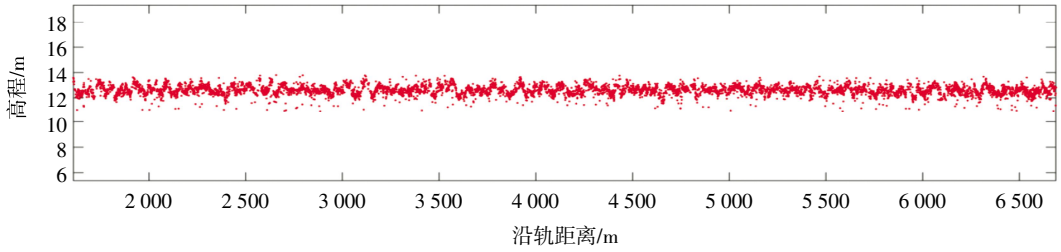


(b) 拉奈岛数据高斯拟合提取结果局部放大

图 8 拉奈岛数据基于高斯拟合算法的海面光子提取结果



(a) 青岛近海数据高斯拟合提取结果



(b) 青岛近海数据高斯拟合提取结果局部放大

图 9 青岛近海数据基于高斯拟合算法的海面光子提取结果

5 结 论

本文针对星载单光子数据中海面光子与噪声光子混杂的难题，提出了一种基于随机森林的多维度空间特征分类框架。通过计算数据中各光子点与其有限区域内点云形成的七类空间特征构建多种决策树，并整合为随机森林分类模型，以此提取高精度的海面光子信号。与多次优化后的最佳 DBSCAN 分类结果和高斯拟合结果进行对比，分析了不同算法间的优劣性及太阳噪声等复杂背景噪声在日夜间数据上表现等，取得了一系列成果，主要研究结论如下。

(1) 基于点云空间特征的海面光子提取算法通过构建光子高程及局部统计特征（如均值、分位数差等）的 7 维特征空间，实现了海面光子的高精度分离。日间与夜间数据的总体分类精度分别达到 96.92%和 99.75%，提取效果（夜间 F1-score 达0.99）显著优于传统 DBSCAN 算法，验证了机器学习在强噪声背景下信号提取的优越性。

(2) 相较于 DBSCAN 算法对参数设置的敏感性，本文算法通过特征建模有效适应了海面光子的动态分布，在复杂噪声（如日间太阳背景噪声）下仍保持高鲁棒性，解决了密度聚类算法在非均匀光子分布场景中的局限性。同时，该算法可有效剔除水面上下贴近水面处的噪声光子。

(3) 基于信噪比分析的粗提取有效降低了人工标注成本，而多维度特征驱动的随机森林模型则通过监督学习优化了分类边界，为海洋动力参数反演、海平面变化监测等应用奠定了高精度数据基础。

参考文献：

[1] CAZENAVE A，LLOVEL W. Contemporary sea level rise[J]. Annual Review of Marine Science，2010，2：145–173.

[2] STAMMER D，CAZENAVE A，PONTE R M，et al. Causes for contemporary regional sea level changes[J]. Annual Review of Marine Science，2013，5：21–46.

[3] NICHOLLS R J，TOL R S J. Impacts and responses to sea level rise: A global analysis of the sres scenarios over the twenty-first century[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences，2006，364（1841）：1073–1095.

[4] 马毅，张杰，张靖宇，等. 浅海水深光学遥感研究进展[J]. 海洋科学进展，2018，36(3)：331–351.

[5] 李建成，金涛勇. 卫星测高技术及应用若干进展[J]. 测绘地理信息，2013，38(4)：1–8.

[6] CAO B，WANG J，HU Y，et al. ICESAT-2 shallow bathymetric mapping based on a size and direction adaptive filtering algorithm[J]. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing，2023，16：6279–6295.

[7] MARKUS T，NEUMANN T，MARTINO A，et al. The ice, cloud, and land elevation satellite-2(ICESat-2): Science require-ments, concept, and implementation[J]. Remote Sensing of Environment，2017，190：260–273.

[8] 朱笑笑，王成，习晓环，等. ICESat-2 星载光子计数激光雷达数据处理与应用研究进展[J]. 红外与激光工程，2020，49（11）：76–85.

[9] ZHAO X，QI C，ZHU J，et al. A satellite-derived bathymetry method combining depth invariant index and adaptive logarithmic ratio: A case study in the Xisha Islands without in-situ measurements[J]. International Journal of Applied Earth Observation and Geoinformation，2024，134：104232.

[10] NEUMANN T A，MARTINO A J，MARKUS T，et al. The ice, cloud, and land elevation satellite-2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system[J]. Remote Sensing of Environment，2019，233：111325.

[11] MA Y，XU N，LIU Z，et al. Satellite-derived bathymetry using the ICESat-2 lidar and Sentinel-2 imagery datasets[J]. Remote Sensing of Environment，2020，250：112047.

[12] 庞勇，李增元，陈博伟，等. 星载激光雷达森林探测进展及趋势[J]. 上海航天，2019，36(3)：20–27.

[13] 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)：148–157.

[14] 陈博伟，庞勇，李增元，等. 基于随机森林的光子计数激光雷达点云滤波[J]. 地球信息科学学报，2019，6：898–906.

[15] 夏少波，王成，习晓环，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.

[16] WANG X，PAN Z，GLENNIE C. A novel noise filtering model for photon-counting laser altimeter data[J]. IEEE Geoscience and Remote Sensing Letters，2016，13(7)：947–951.

[17] 王振华，陈诗贤，孔伟，等. 光子计数激光雷达中光子点云的滤波方法比较与分析[J]. 激光与光电子学进展，2023，60(6)：1–16.

[18] 张少华，施银迪，谭莲红，等. 基于改进 DBSCAN 的单光子激光点云去噪算法[J]. 矿山测量，2022，50(1)：32–37.

[19] ZHANG J，KEREKES J. An adaptive density-based model for extracting surface returns from photon-counting laser altimeter data[J]. IEEE Geoscience and Remote Sensing Letters，2015，12(4)：726–730.

[20] 周智标，周辉，马跃，等. ICESat-2 激光雷达海面信号提取和海浪要素计算[J]. 红外与激光工程，2023，52(2)：212–220.

[21] 杨鑫科，赵辽英，张华国. ICESat-2 激光雷达数据海面信号光子提取和海浪要素反演方法[J]. 应用海洋学学报，2025，44(3)：483–492.

[22] MA Y，LIU R，LI S，et al. Detecting the ocean surface from the raw data of the MABEL photon-counting lidar[J]. Optics Express，2018，26(19)：24752–24762.

[23] HUANG G，DONG Z，LIU Y，et al. An optimized denoising method for ICESat-2 photon-counting data considering heterogeneous density and weak connectivity[J]. Optics Express，2023，31(25)：41496–41517.

[24] 谢丛霜，陈鹏，潘德炉. 主被动遥感融合辐射传输卷积神经网络水深反演方法[J]. 光子学报，2024，53(8)：46–58.

[25] NEUMANN T，BRENNER A，HANCOCK D，et al. Algorithm theoretical basis document（ATBD）for global geolocated photons[R]. Washington：National Aeronautics and Space Administration，2021.

A Random Forest-Based Method for Extracting Sea Surface Photons Based on Point Cloud Spatial Features

WANG Xiaoming[1], ZHANG Haoshuai[2], HU Wenhui[1], QI Chao[2,3]

(1. CCCG Xingyu Technology Co., Ltd., Beijing 102200, China; 2. College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China; 3. Key Laboratory of Oceanic Surveying and Mapping, Ministry of Natural Resources, Qingdao 266590, China)

Abstract: The highly sensitive photon-counting lidar carried by the ICESat-2 (Ice, Cloud, and land Elevation Satellite-2) has opened up new avenues for sea surface detection. However, due to the dual constraints of the sensor's extremely high sensitivity to photon signals and the cross-medium reflection characteristics between the sea and the atmosphere, the original data shows the feature of a mixture of effective photons from the sea surface and noise photons, which limits the direct application of the data. To this end, this paper proposes a random forest classification framework based on multi-dimensional spatial feature mining and establishes a two-level processing strategy from coarse extraction to precise classification. Firstly, based on the signal-to-noise ratio analysis of the photon quantity histogram of elevation, the coarse extraction of sea surface photons is achieved, and then a high-confidence training dataset is constructed through manual annotation. Extract 7-dimensional feature parameters to train the classification model to further optimize the classification results. To verify the advantages of the method, the above-mentioned method was used to extract the sea surface photons of the day-night data in the coastal area of Qingdao and conduct a comparative analysis with the DBSCAN algorithm. The results show that although the DBSCAN algorithm can achieve photon extraction from the sea surface (with a relatively high accuracy rate when the search radius is 1 m), its performance is affected by parameter sensitivity and its robustness is relatively low. The extraction algorithm in this paper achieved overall classification accuracies of 96.92% and 99.75% respectively in daytime and nighttime data by constructing photon spatial features, and its performance was superior to that of traditional methods. Meanwhile, the Gaussian fitting algorithm was introduced to extract the sea surface points of Class I water bodies (Lanai Island, Hawaii) and Class Ⅱ water bodies (near the coast of Qingdao), and compared with the method in this study, verifying the robustness of this method in different sea areas. Therefore, this method effectively resolves the challenge of signal extraction in a strong noise background, confirming the significant advantages of machine learning strategies in the processing of single-photon remote sensing data, and providing new methodological support for the oceanographic application of spaceborne photon counting data.

Keywords: ICESat-2; marine surface detection; random forest; DBSCAN