

# 基于波形形态特征的单频机载激光雷达测深全波形数据分类

李莹莹<sup>1</sup>, 刘子维<sup>1</sup>, 张静坤<sup>2</sup>, 吴琳琳<sup>1</sup>, 纪雪<sup>1,3,4\*</sup>, 王明常<sup>1\*</sup>

(1. 吉林大学地球探测科学与技术学院, 吉林 长春 130000;

2. 上海工程技术大学城市轨道交通学院, 上海 201620;

3. 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉 430079;

4. 自然资源部海洋遥测技术创新中心, 山东 青岛 266061)

**摘要:** 单频机载激光雷达测深系统凭借低成本、低负载、高采样率等优势成为大范围海岸带地形地貌探测的理想选择。然而如何解决单频局限,在不依赖辅助传感器情况下实现全波形数据的准确划分为精确点位坐标解算的关键环节。目前基于全波形形态特征进行波形分类研究缺乏系统性评估分析和普遍性结论。该研究尝试从全波形空间形态入手,细化了波形类别(异常波形、过拟合波形、陆地波形、海面波形和测深波形),在已有波形特征基础上,系统分析了不同类别波形的形态特征差异,有针对性地提取了 24 维波形特征并基于随机森林特征选择和分类模型完成了各特征分类性能及最佳特征组合评估与定量分析。研究证明,包括相邻两点间振幅偏差、震荡主频等在内的 6 维特征组合对 5 种波形的分类效果最好,总体分类精度可达 98.55%, $Kappa$  系数为 0.9820。为了验证特征的普适性,另外选取了一块实验区域进行验证,得到水陆分类的总体精度为 96.81%。

**关键词:** 机载激光雷达测深; 全波形; 波形特征; 波形分类

**中图分类号:** TN958.98      **文献标志码:** A      **DOI:** 10.3788/IRLA20230096

## 0 引言

机载激光雷达测深系统 (Airborne LiDAR Bathymetry, ALB) 是集激光系统、全球定位系统和惯性导航系统等于一体的主动遥感测深系统<sup>[1-3]</sup>,属于全波形系统 (Full Waveform System),记录了激光脉冲从发射到接收整个过程的强度(振幅)时间剖分<sup>[1,4]</sup>。根据完整的波形可以得到目标物与传感器之间的距离、目标物的几何形状和物理特征等信息。ALB 系统有单频系统和双频系统两种,其中双频系统的近红外波段 (1064 nm) 无法穿透水体,需与蓝绿波段 (532 nm) 相结合提高水深测量的精度。单频 ALB 系统仅采用波长为 532 nm 的蓝绿激光作为激光器发射光源<sup>[5-6]</sup>,简化了系统结构、降低了设备成本,在海洋浅水测绘中

得到广泛应用<sup>[5]</sup>。然而,由于单频机载激光雷达接收的信息有限,在缺乏辅助数据的情况下识别陆地和水域波形困难,是单频 ALB 数据处理中迫切需要解决的问题<sup>[7]</sup>。

目前,针对 ALB 波形数据处理的研究主要集中在双频 ALB 波形分解和波峰提取上<sup>[8]</sup>,对波形分类研究较少。有研究借助辅助数据进行海陆波形分类,如增加拉曼 (647 nm) 散射通道法<sup>[9-11]</sup>、影像或视频辅助<sup>[12-13]</sup>和增加近红外通道进行辅助<sup>[14]</sup>。王金虎等<sup>[15]</sup>利用广义高斯函数方法对后向散射波形进行分解,提取了组分距离、回波宽度、振幅和后向散射截面 4 个特征对植被、建筑物进行点云分类。范士俊等<sup>[16]</sup>对波形进行分解,提取了振幅、回波宽度和回波次数这 3 个特征,利用随机森林实现地物分类。纪雪等<sup>[10]</sup>利

收稿日期:2023-02-27; 修订日期:2023-04-06

基金项目:武汉大学测绘遥感信息工程国家重点实验室开放基金项目 (22S02);自然资源部海洋遥测技术创新中心开放基金项目 (2022002);国家自然科学基金项目 (22FAA01871, 42301505, 52305445);吉林省自然科学基金项目 (20210101098JC)

作者简介:李莹莹,女,硕士生,主要从事海洋测绘方面的研究。

导师(通讯作者)简介:王明常,男,教授,博士,主要从事地图学理论、遥感图像分析与变化检测、空天地一体化信息感知等方面的研究。

通讯作者:纪雪,女,讲师,博士,主要从事海洋测绘、声学遥感、光学遥感领域的数据分析、处理与应用等方面的研究。

用 Levenberg-Marquardt(LM-GGM) 算法的广义高斯优化模型对波形进行分解, 提取了 38 维波形参数, 结合随机森林特征选择模型得到的 10 维重要特征输入到随机森林分类器中实现过拟合波形、孤立波形、海洋波形和陆地波形这 4 种波形的分类。王丹茜等<sup>[17]</sup>对远离海陆和海陆交界处的回波进行了分类研究, 对于远离海陆交界处的回波采用了绿光通道的首末回波生成的点云进行初分类, 海陆交界处的回波则提取了绿光通道的波形特征结合点云高程特征筛选相似回波进行分类。

然而, 已有方法的波形特征提取缺乏对于不同波形形态的精细化分析和性能评估。针对这个问题, 文中开展了基于波形形态特征的单频机载激光雷达测深全波形数据分类研究。通过波形可视化分析细化了波形类别, 由传统的水陆两类波形细分为异常波形、过拟合波形、陆地波形、海面波形和测深波形 5 类; 基于不同类别波形形态特征, 有针对性地提取了 24 维波形特征, 利用随机森林特征选择模型(Random Forest Feature selection Model, RFFS) 对特征波形分类性能进行评估; 通过不同特征组合的波形分类精度反馈确定最优特征组合; 最后, 基于不同实验区采集的数据采用随机森林分类模型对所提方法的有效性和普适性进行评价。

## 1 研究方法

### 1.1 方法流程

文中方法具体工作流程如图 1 所示, 可划分为以下 6 个步骤:

- 1) 读取波形数据。对 ALB 获取的 LAS 1.3 格式文件进行数据解析, 得到波形数据和点云数据。
- 2) 划分波形类别。结合实际影像和波形的形态, 明确要分类的波形类型: 异常波形、过拟合波形、陆地波形、海面波形和测深波形 5 种, 波形的具体形态见图 2。

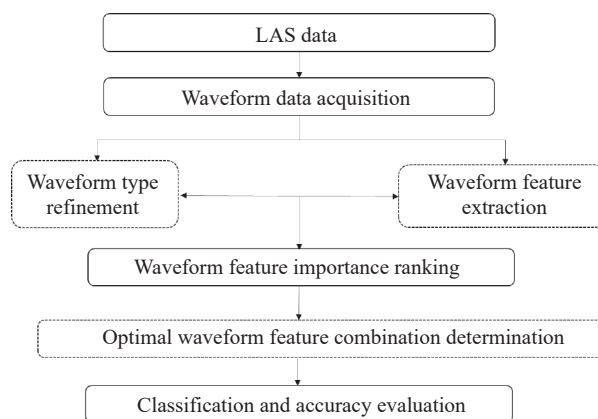


图 1 文中方法流程

Fig.1 Methodology flow of this paper

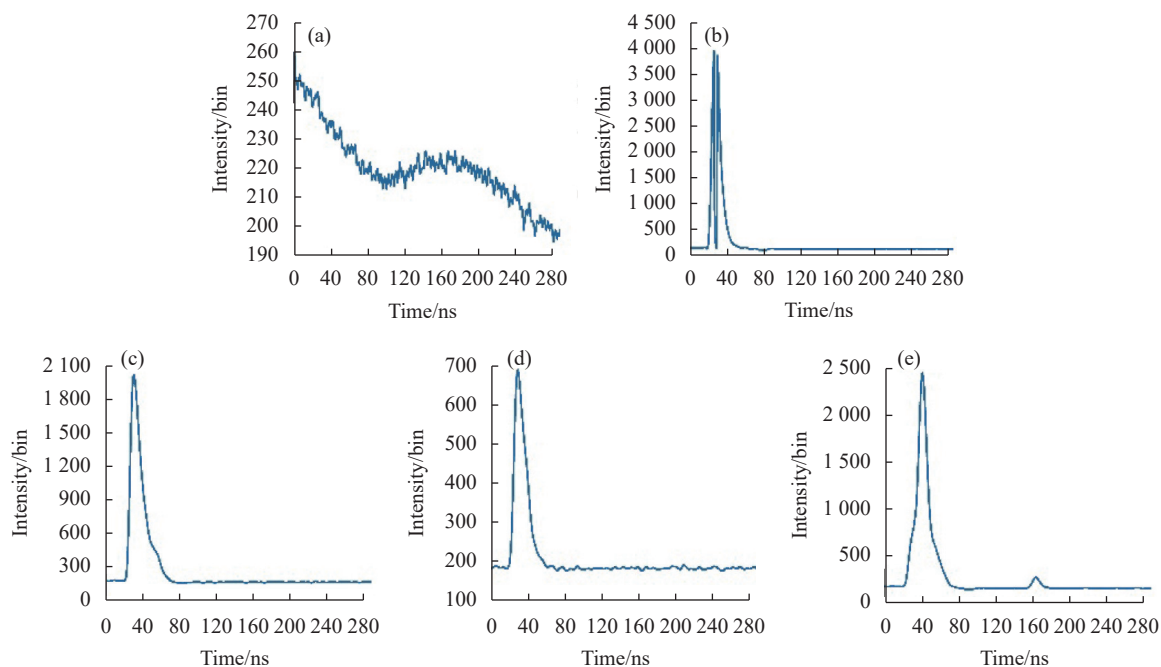


图 2 5 种波形分布图

Fig.2 Five types of waveform distribution

3) 提取波形形态特征。分析不同波形的形态特点,对波形进行特征提取。

4) 评估波形特征。利用 RFFS 模型对提取的特征进行特征重要性排序。

5) 自适应确定最佳波形特征组合。将得到的特征重要性排名以不同的组合输入分类器中,确定最佳的特征组合。

6) 波形分类实验及评估。将最佳特征组合输入分类器进行分类实验和精度评估检验。

总体方法流程如图 1 所示。

### 1.2 波形类别确定

目前全波形分类集中在海陆波形的划分<sup>[10,17]</sup>。纪雪等<sup>[10]</sup>尝试对除水、陆地波形以外的过拟合波形和孤立波形进行分类。但是,分类过程复杂,特征也较为常规,未针对不同类别的波形空间形态差异进行针对性提取。因此,文中在水陆分类过程中,尝试对干扰波形进行识别和剔除,以减少后续对信号的干扰和坐标解算的计算量。结合实际影像、波形形态,明确了要分类的 5 种波形类型:异常波形、过拟合波形、陆地波形、海面波形和测深波形,下面将对其分别进行介绍。

1) 异常波形,如图 2(a) 所示,其形态整体呈现了先下降后上升再下降的趋势;水域、陆地范围都有分布,较为随机,没有规律。

2) 过拟合波形,如图 2(b) 所示,分布在激光透射性较差或亮反射区域,如建筑物、柏油马路和光滑的海面(镜反射)等。从形态上可以直观地看出:相邻两点间振幅发生了剧烈的变化,这是由于目标物的回波信号超过了传感器的接收阈值。

3) 陆地波形,如图 2(c) 所示,指激光最后发射到陆地再将信息反射回接收器接收到的完整波形。对于植被等透射性较好的地物,陆地回波可能出现多个波峰。

4) 海面波形,如图 2(d) 所示,这里定义为未穿透水体,直接从海面返回接收器的波形信号,因此,海面波形只有一个波峰。

5) 测深波形,如图 2(e) 所示,是激光到达海底所接收的波形信息。测深波形一般会有两个波峰,第一个波峰是海面反射的回波信号,一般峰值较高,第二

个回波峰值是由海底反射的,一般较小。

### 1.3 波形特征提取

波形特征提取是海陆波形分类的重要步骤。文中在常规特征参数的基础上总结了不同波形的形态和时域特征,并依据这些特征进行了创新和改进,共选取了 24 维特征参数。

#### 1) 最大最小振幅时间差

其中,异常波形在形态上与其他波形具有显著的差异。强度的最大值和最小值一般分布在首、尾位置,因此,最大最小振幅时间差  $T_{\max-\min}$  可以较好地区分异常波形。

#### 2) 相邻两点间的振幅偏差

过拟合波形在形态上出现“截断”现象,有两个尖锐的波峰,如图 3 所示,基于此,提出了相邻两点间振幅偏差为  $\Delta A$ 。

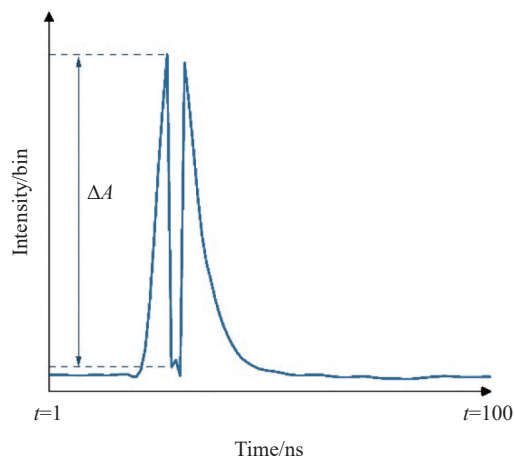


图 3 相邻两点间振幅偏差 ( $\Delta A$ )

Fig.3 Deviation of amplitude between two adjacent points ( $\Delta A$ )

#### 3) 最大强度

通常情况下,首个回波的信号强度最大,选取最大强度  $W/I$  特征,可以反映不同地物的属性。陆域回波信号通常具有更高的回波强度<sup>[17]</sup>,此特征有助于区分陆地和海洋波形。

#### 4) 最大强度位置

激光脉冲回波信号,除了信号幅值外,脉冲的时间也是反映回波信号的主要特征量<sup>[18]</sup>。最大强度位置  $It$  表示回波信号最大强度所出现的时间。

#### 5) 峰值数量

不同目标物具有的峰值数量<sup>[10]</sup>  $Num$  不尽相同。

例如,陆地波形多为一个,当遇到植被等由于光束透射会有多个回波;测深波形为两个峰值;海面回波波形有一个峰值。

6) 首个回波峰值

文献 [10] 中的 10 维最重要特征主要分布在首个回波之中。因此,文中提取首个回波峰值  $Fpk$  表示多个回波峰值中的第一个回波的峰值数值。

7) 首个回波峰值的位置

首个回波峰值的位置  $Ft$  主要描述了首个回波峰值所出现的时间,以进一步了解激光首次接触地物的时刻。

8) 首个回波偏度

首个回波偏度  $S_{first}$  用于描述波形的偏移程度。观察测深波形形态多呈右偏态<sup>[19]</sup>,因此提取此特征可用于区别其他波形。

9) 半峰宽

此特征的方法是先找到最大的峰值计算出峰值高度一半的数值,然后根据这个数值计算出其出现的第一个和最后一个时间,作差即可得到半峰宽  $FWHM$ <sup>[20]</sup>。

10) 回波波宽

回波波宽  $Wf_w$  是以数据形式表达不同地物波形的展宽程度。相比于陆地回波而言,海洋浅水和海洋深水的回波展宽程度较为明显<sup>[8]</sup>。其表达式为:

$$Wf_w = \frac{FWHM}{2\sqrt{2\ln 2}} \quad (1)$$

11) 衰减常数

衰减常数  $a$ <sup>[21]</sup>,即首个峰值  $A_1$  与第二个峰值  $A_2$  之差,再将差值与首个峰值  $A_1$  作商而得到的。如图 4 所示。此特征表征激光在水体中传输时的衰减<sup>[22]</sup>,考虑到首个振幅、第二个振幅对区分波形的影响,其公式为:

$$a = \frac{A_1 - A_2}{A_1} \quad (2)$$

12) 峰值之比

峰值之比  $R_p$ ,即将最小峰值与最大峰值进行作商,一般首个峰值即为最大波峰,随着信号的逐渐减弱,最小峰值一般会出现在最后一个回波,因此,为探究不同波形的吸收、衰减作用提出了此特征。

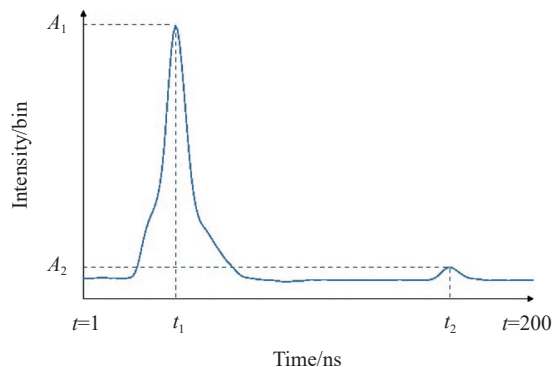


图 4 衰减常数 (a)

Fig.4 Decay constant (a)

13) 震荡主频

震荡主频  $f$ ,如图 4 所示,采用回波波形中的首个峰值和第二个峰值所出现的时间差取倒数而得到的<sup>[21]</sup>。其公式可表示为:

$$f = \frac{1}{t_2 - t_1} \quad (3)$$

式中:  $t_1$  为首个峰值出现的时间;  $t_2$  为第二个峰值出现的时间。

14) 波形总能量

一条波形是由激光接触到地物所产生的回波和激光传输过程中产生的背景噪声所组成的,不同地物产生的能量各不相同。因此,波形总能量  $S_{all}$  表征回波中所有能量的累积。 $S_{all}$  的计算公式为:

$$S_{all} = \frac{1}{2} \sum_{t=1}^{t=288} (Wf(t) + Wf(t+1)) \quad (4)$$

式中:  $Wf(t)$  为采用的原始波形;  $t=1$  表示波形包含噪声的开始时间为 1 ns;  $t=288$  表示波形包含背景噪声的结束时间是 288 ns。

15) 有效回波能量

有效回波能量  $S_w$ ,即除背景噪声外组成目标物的信号总能量<sup>[19]</sup>。此特征主要用来区分陆地、海面 and 测深回波所产生的有效能量。其公式可以表示为:

$$S_w = \frac{1}{2} \sum_{t=t_s}^{t=t_0} (Wf(t) + Wf(t+1)) \quad (5)$$

式中:  $t_s$  表示波形有效范围的开始时间;  $t_0$  表示波形有效范围的结束时间。

16) 波形有效范围

波形的有效范围  $L$  是描述回波信号的持续时间,在一定程度上反映信号的重叠情况<sup>[17]</sup>,如图 5 所示。

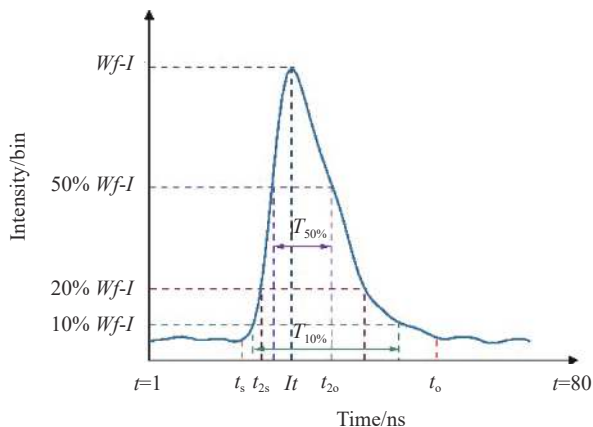


图 5 波形特征分布

Fig.5 Waveform feature distribution

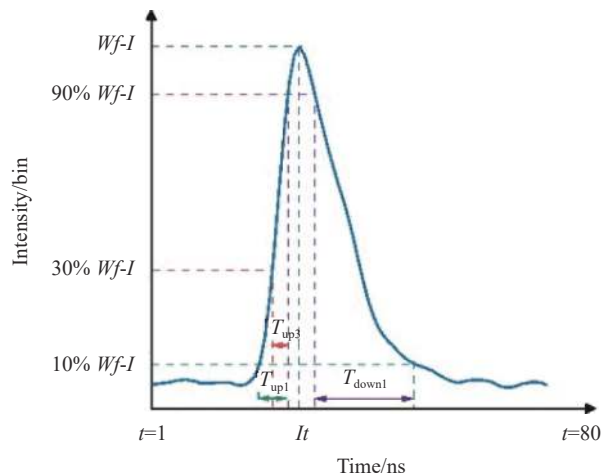


图 6 波形特征分布

Fig.6 Waveform feature distribution

$L$  的计算公式为:

$$L = t_o - t_s \quad (6)$$

17) 峰值 20% 以上的面积

选取最大峰值的 20% 这部分回波的能量<sup>[17]</sup>, 如图 5 所示, 其公式可以表示为:

$$S_{20\%} = \int_{t=t_{2s}}^{t=t_{2o}} Wf(t) dt \quad (7)$$

式中:  $t_{2s}$  表示波形峰值 20% 的开始时间;  $t_{2o}$  表示波形峰值 20% 的结束时间。

18) 最大振幅的 10% 所持续的时间

最大振幅的 10% 所持续的时间<sup>[23]</sup> $T_{10\%}$ , 即利用最大振幅的 10% 作为阈值, 确定阈值范围的时间差即为  $T_{10\%}$ , 如图 5 所示。

19) 最大振幅的 50% 所持续的时间

最大振幅的 50% 所持续的时间<sup>[23]</sup> $T_{50\%}$ , 与最大振幅的 10% 所持续时间  $T_{10\%}$  相似, 如图 5 所示。

20) 脉冲上升沿 10%~90% 的时间

选取波形数据中第一个大于最大峰值 10% 和第一个大于最大峰值 90% 的数据所对应的时间差值作为脉冲上升沿 10%~90% 的时间<sup>[24]</sup> $T_{up1}$ , 如图 6 所示。

21) 脉冲上升沿 10%~90% 的陡度

脉冲上升沿 10%~90% 的陡度<sup>[24]</sup> $ST_1$  选用了 80% 的回波信号最大强度与脉冲上升部分的 10%~90% 的时间差作商。

$$ST_1 = \frac{0.8 \times Wf-I}{T_{up1}} \quad (8)$$

式中:  $Wf-I$  表示回波信号最大强度;  $T_{up1}$  表示脉冲上升沿 10%~90% 的持续时间。

22) 脉冲下降沿 10%~90% 的时间

脉冲下降沿 10%~90% 的时间<sup>[24]</sup> $T_{down1}$  可理解为波形下降的部分选取最大幅值的 10%~90% 对应的时间列的时间差即为脉冲下降沿 10%~90% 的时间  $T_{down1}$ , 如图 6 所示。

23) 脉冲上升沿 30%~90% 的时间

此特征与  $T_{up1}$  特征相似, 即将 10%~90% 换为 30%~90% 即为脉冲上升沿 30%~90% 的时间<sup>[24]</sup> $T_{up3}$ , 如图 6 所示。

24) 脉冲上升沿 30%~90% 的陡度

选取脉冲上升波形部分的 30%~90% 的数据计算其数值变化情况<sup>[24]</sup>,  $ST_3$  的公式为:

$$ST_3 = \frac{0.6 \times Wf-I}{T_{up3}} \quad (9)$$

式中:  $T_{up3}$  表示脉冲上升沿 30%~90% 的持续时间。

#### 1.4 基于随机森林的特征重要性评估

随机森林 (Random Forest, RF) 是 Breiman 于 2001 年提出的一种统计学习理论方法<sup>[25-27]</sup>, 属于基于决策树的分类器<sup>[25]</sup>。目前, 随机森林在分类、回归和降维等问题方面得到了广泛应用。

RFFS 算法<sup>[26]</sup> 对于重要测度指标的计算主要有两种: 基于 Gini 指数的平均不纯度减少指标 (Mean Decrease Impurity, MDI) 和基于 OOB 数据置换的平均精度准确率减少指标 (Mean Decrease Accuracy, MDA)<sup>[27]</sup>。利用第二种方式计算模型中各特征的重要性, 首先在原始训练样本中通过 Bootstrap 的方式有放回地进行

样本的抽取。假设训练数据中共有  $N$  个对象的数据, 从这些数据样本中每次有放回地取出一些样本数据, 每次取出的并不完全相同, 这些样本将作为决策树的训练数据组, 这些训练数据组被分为  $n$  个样本集, 每个样本集生成一棵决策树; 另外 1/3 的原始样本并未被抽取, 这部分样本称为袋外 (Out of Bag, *OOB*) 数据, 可用于预测错误率  $E_1$ ; 然后随机置换变量  $i$  的观测值构建决策树, 计算袋外数据的预测错误率  $E_{2i}$ , 最后计算 *OOB* 误差 ( $OOB_{error}$ ), 其公式为:

$$OOB_{error} = E_{2i} - E_1 \quad (10)$$

使用 *OOB* 误差进行模型评估以确定模型最优特征数量。通过计算不同特征组合模型的 *OOB* 误差, 综合考虑模型精度和计算复杂度来确定最优特征个数<sup>[28]</sup>并得出变量重要性排序。特征  $i$  的置换重要性 *importance*( $i$ ) 公式为:

$$importance(i) = \sum OOB_{error} / T \quad (11)$$

式中:  $T$  表示分类树的数量。

### 1.5 基于随机森林的分类模型

随机森林是一种非参数分类与回归方法, 不需要先验知识, 易于使用, 能在保证良好精度的前提下并行化处理高维海量数据, 是一种高效的机器学习算法<sup>[28]</sup>。

随机森林决策树构建中的一个关键步骤——节点分裂时的特征选择: 假设每个数据点都有  $M$  个特征, 则可以在每个节点处随机选择  $m(m < M)$  个特征, 选用信息熵、信息增益或者基尼指数来选择最好的特征进行节点分裂。然后将  $m$  分成两部分, 用最佳分裂方案对节点进行分裂, 决策树生长期内  $m$  大小始终不变, 每棵树均进行完全的生长且不被修剪。最后, 将上述步骤进行重复, 通过  $k$  次样本抽取和样本训练得到  $k$  个决策树模型, 利用集成学习理论将这  $k$  个决策树进行线性组合, 其中每个决策树占相等的权重。组合多棵决策树, 从每个决策树中获取预测结果, 对每个预测结果进行投票, 通过投票得出最终的预测结果。

分类问题中 CART 使用 *Gini* 值作为评判标准。定义为:

$$Gini = 1 - \sum (P(i) \times P(i)) \quad (12)$$

式中:  $P(i)$  为当前节点上数据集中第  $i$  类样本的比例。

## 2 实验结果与分析

### 2.1 研究区域与数据

文中的研究区域位于中国海南省三亚市崖州区的东锣岛, 其中扫描条带覆盖东锣岛范围, 如图 7 所示。东锣岛陆域面积约 0.1180 km<sup>2</sup>, 地理坐标为北纬 18°19'33.20", 东经 108°59'24.13", 岸线长约 1.432 km, 其附近的水深约为 4.3~10.2 m。研究区数据由 Optech 公司设计的 Aquarius 型号机载激光雷达进行采集。此型号的机载激光雷达专门为沿海和内河航道测绘而设计。

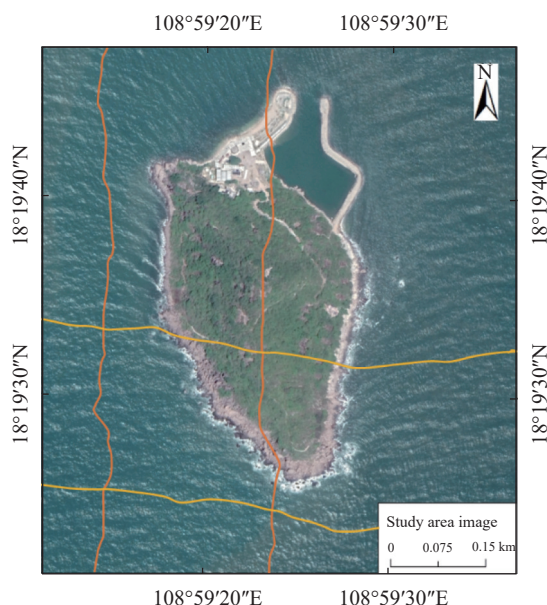


图 7 ALB 研究区域位置

Fig.7 ALB study area location

### 2.2 特征重要性评估

文中在总数据集中随机选取了 10 万个样本数据, 每个类别各随机选取了 1 万个训练集, 1 万个测试集。对每个波形都提取了 24 个特征参数, 生成了 1×24 的特征矩阵, 输入到随机森林特征重要性评估模型中进行特征重要性排序。为避免特征冗余, 增强所选特征的可靠性, 随机选取每类波形的样本数据各 1 万条进行了 10 次随机森林特征重要性实验, 根据 10 次实验的权重系数取平均值得到最终的特征重要性排名, 如图 8 所示。

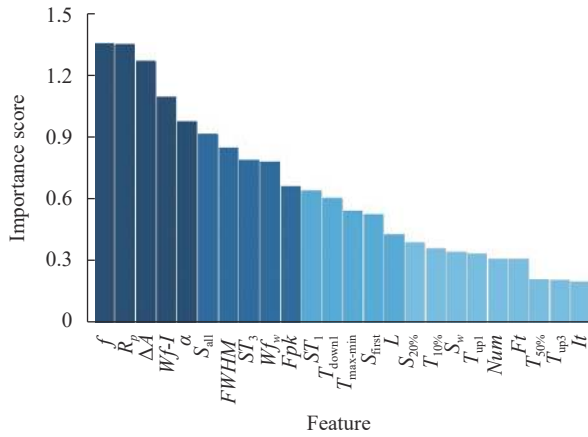


图 8 特征重要性排序结果

Fig.8 Feature importance ranking result

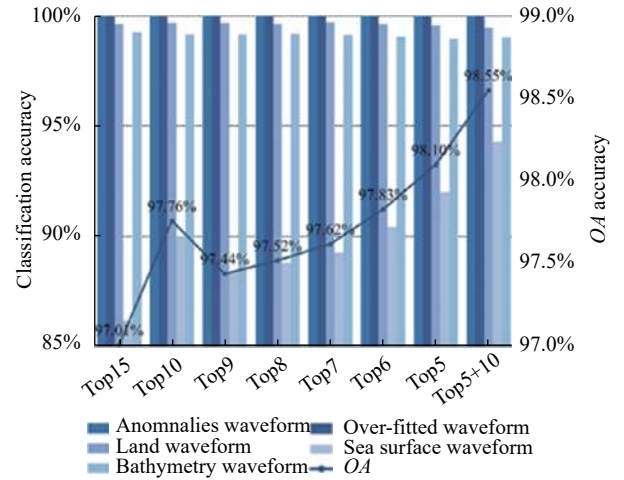


图 9 不同特征组合精度统计

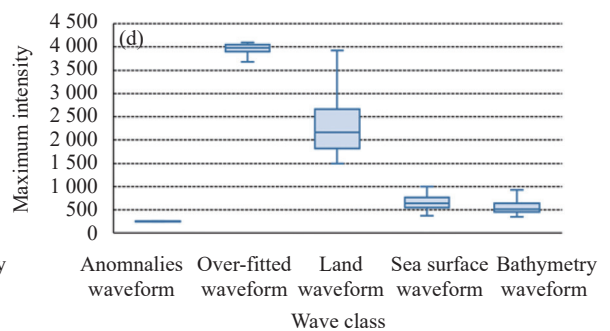
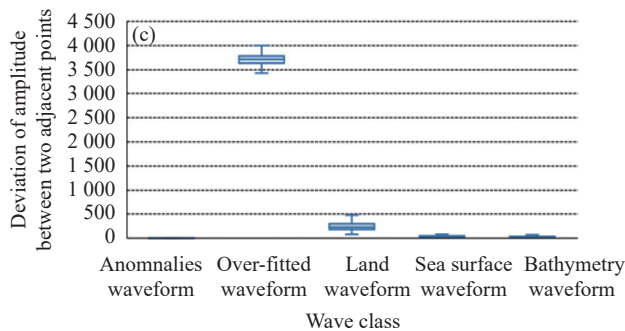
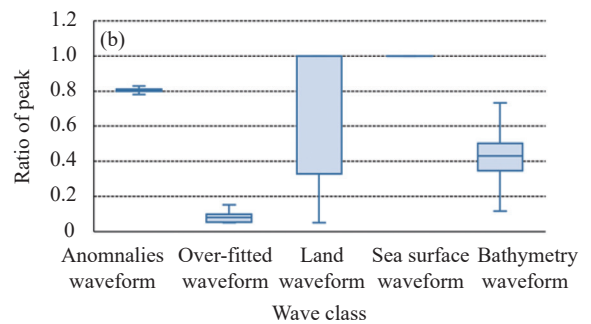
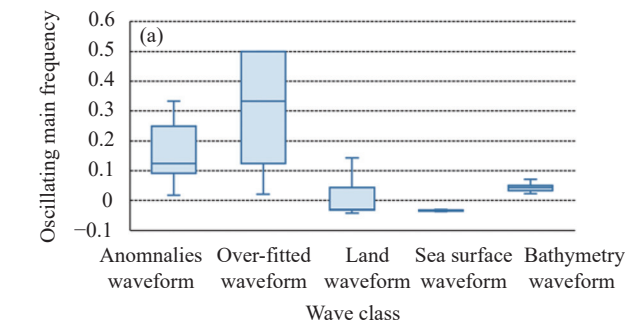
Fig.9 Accuracy statistics for different feature combinations

2.3 最佳特征组合确定

在大量特征中选取最可靠的特征,是提高分类精度和表达地物类型空间分布的关键。根据特征重要性排序结果选择不同特征进行组合、分类和精度评估以确定最优的特征组合,如图 9 所示。

从图 9 可以看出,Top15 由于特征的冗余,精度最低;Top10 较 Top15 精度有所上升,说明 11~15 维特征对于 5 种波形的区分效果不明显;Top9 与 Top10 相

较,精度有 0.32% 的下降,这说明第 10 维特征对此随机数据的精度有贡献;Top9~Top5 随着特征数量的降低,精度在逐渐提高,最大可达到 98.10%。特征重要性前 5 维特征和第 10 维特征对分类的贡献最高,重新进行随机实验,精度达到了 98.55%。震荡主频  $f$ 、峰值之比  $R_p$ 、相邻两点间振幅偏差  $\Delta A$ 、最大强度  $Wf-I$ 、衰减常数  $a$ 、首个回波峰值  $Fpk$  作为最终的特征优选结果。其在 5 种波形上的箱型分布如图 10 所示。



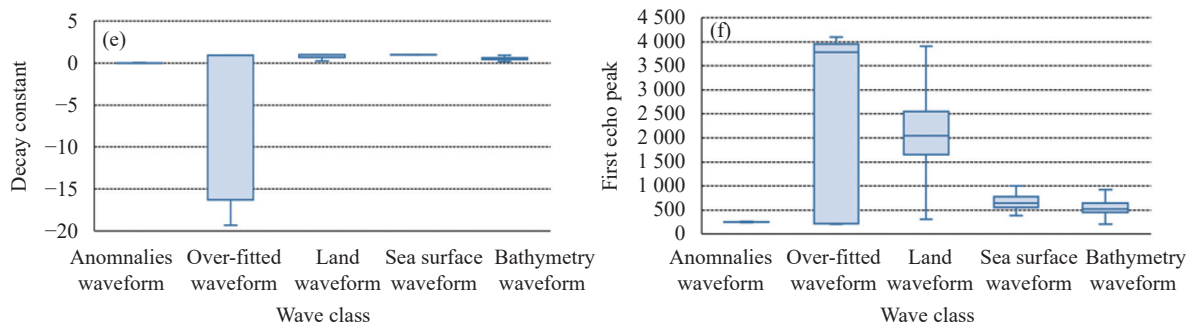


图 10 波形重要特征分布箱型图

Fig.10 Box figures of waveform important feature distributions

2.4 波形分类及精度评估

为了进一步检验随机森林分类方法的性能,选取了 K 最邻近值 (KNN) 和支持向量机 (SVM) 分类算法进行对比,分类结果如表 1 所示。

实验结果表明:RF 的分类效果最好,总体分类准确率 (OA) 为 98.55%。其中,异常波形和过拟合波形

的分类精度均为 100%,陆地波形的分类精度为 99.46%,海面波形的分类精度为 94.29%,测深波形的分类精度为 99.02%, Kappa 系数为 0.9820。OA 和 Kappa 系数的计算公式分别为:

$$OA = \frac{\sum_{i=1}^n n_{ii}}{N} \tag{13}$$

$$Kappa = \frac{N \sum_{i=1}^n n_{ii} - \sum_{i=1}^n n_i \cdot n_i}{N \times N - \sum_{i=1}^n n_i \cdot n_i} \tag{14}$$

表 1 不同分类方法结果对比

Tab.1 Comparison of the results of different classification methods

Waveform types	RF	SVM	KNN
Anomalies waveform	100%	100%	100%
Over-fitted waveform	100%	100%	100%
Land waveform	99.46%	95.9%	93.42%
Sea surface waveform	94.29%	84.5%	72.86%
Bathymetry waveform	99.02%	90%	88.77%
OA	98.55%	94%	91.01%

式中: N 是波形总数;  $n_{ii}$  代表对角线上的总波形数;  $n_i$  则是第 i 行的总波形数;  $n_i$  则代表第 i 列的总波形数。

测试数据的分类结果相对应的混淆矩阵如图 11 所示。研究区的波形分类结果如图 12 所示。其中,

True values \ Predicted values	Anomalies waveform	Over-fitted waveform	Land waveform	Sea surface waveform	Bathymetry waveform	Sample correct rates Sample error rates
	Anomalies waveform	10 000 20%	0 0%	0 0%	0 0%	
Over-fitted waveform	0 0%	10 000 20%	0 0%	0 0%	0 0%	100% 0%
Land waveform	0 0%	0 0%	9 946 19.89%	0 0%	54 0.11%	99.46% 0.54%
Sea surface waveform	0 0%	0 0%	0 0%	9 429 18.86%	571 1.14%	94.29% 5.71%
Bathymetry waveform	0 0%	0 0%	97 0.19%	1 0%	9 902 19.80%	99.02% 0.98%
Waveform correct rates Waveform error rates	100% 0%	100% 0%	99.03% 0.97%	99.99% 0.01%	94.06% 5.94%	98.55% 1.45%

图 11 5 种波形分类的混淆矩阵

Fig.11 Confusion matrix for the five waveform classifications



过拟合波形分布呈现出一定的航带中央集中性,这主要是由于水域范围内,过拟合波形的产生主要是由于发生海面镜面反射造成的高返回信号,这与激光发射角度有一定关系,随着发射角增大,镜面反射现象逐渐弱化,因此记录的过拟合波形减少;而在陆地附近的过拟合波形分布主要是由于地物类别引起的,对于沙滩、裸露礁石、建筑物、柏油马路等地物得到高返回信息,因此过拟合波形较多,这使得在陆地区域不

存在过拟合波形航带中央集中特性。

为了验证文中分类的准确性、模型的迁移性和普适性,另外选择了一个实验区域的条带数据进行波形分类,结果如图 13 所示。从图中可以看出:过拟合波形主要分布在裸露地面和建筑物上,异常波形的分布比较均匀。任意选取 10 块区域进行海陆分类精度验证,海洋的分类精度为 97.43%,陆地的分类精度为 96.2%,水陆总体分类精度为 96.81%,满足生产需求。

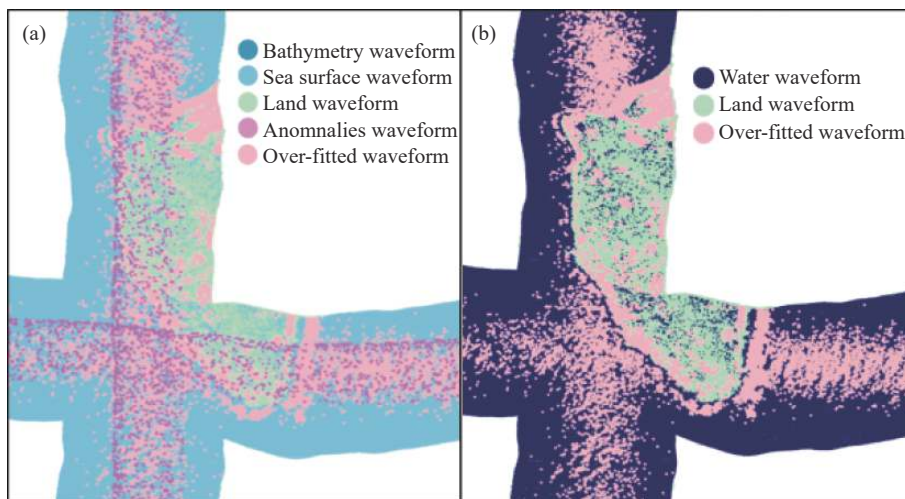


图 12 研究区波形分类结果

Fig.12 Results of waveform classification in the study area

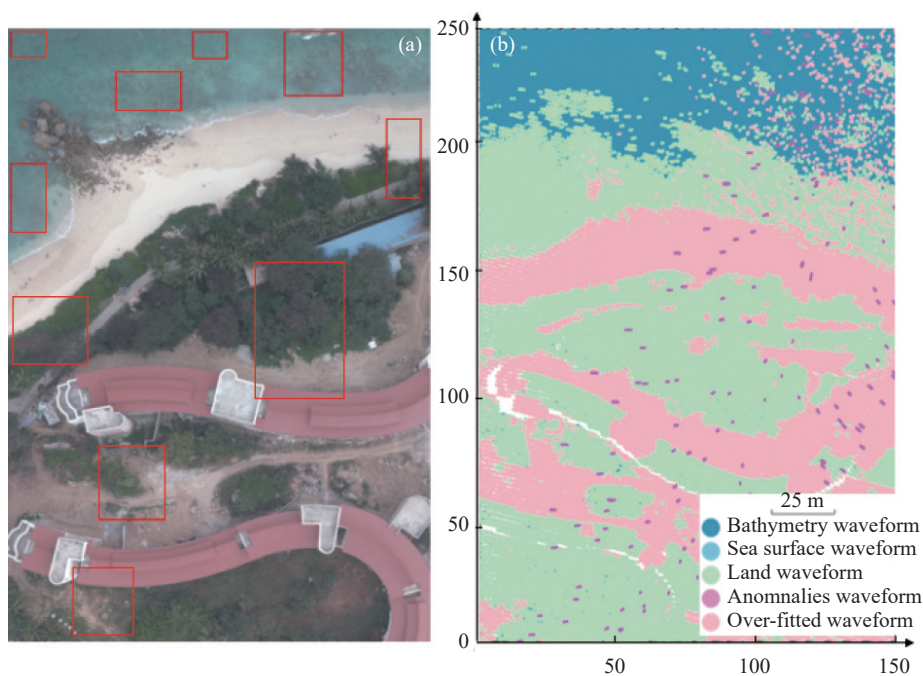


图 13 实验区波形分类结果图

Fig.13 Figure of waveform classification results in the experimental area

但是,精度发生了大约 2% 的降低,主要原因是研究区和实验区的数据并不是同一时间、同一区域采集的,而且飞行条件和飞行高度等参数设置不一致、区域的地物类别引起的地物振幅的变化等造成了一定的偏差。

### 3 结 论

如何在不依赖辅助传感器情况下,实现单频 ALB 对水陆波形进行分类,是一个亟需解决的问题。为精确识别波形,系统分析了不同类型波形的形态差异,有针对性地提取了 24 维特征参数,经最佳特征组合和分类性能评估等过程,发现震荡主频  $f$ 、峰值之比  $R_p$ 、相邻两点间振幅偏差  $\Delta A$ 、最大强度  $WfI$ 、衰减常数  $a$ 、首个回波峰值  $Fpk$  这 6 维特征,可以有效区分这 5 种波形。总体分类精度可达 98.55%,  $Kappa$  系数为 0.9820。其中,异常和过拟合波形的 100% 提取,强有力地证实了形态特征的针对性和有效性。更换实验区后,水陆分类精度达到了 96.81%,证明所采用的特征、方法具有迁移性和普适性,满足生产的需求。

更换实验区后尽管也得到了较为理想的波形分类结果,但精度下降了约 2%。这主要是由于不同实验区,设备飞行参数、作业参数设置不同,地物类别存在差异,使得样本存在一定局限。所以,要保持波形分类精度可以结合具体实验区适当补充样本数据。此外,深入研究了波形形态,而对于信号接收系统等内在因素对回波的影响,需要进一步实验验证。后续还将增加研究的数据量并对信号探测进行研究。

#### 参考文献:

- [1] Liu Yongming, Deng Ruru, Qin Yan, et al. Data processing methods and applications of airborne LiDAR bathymetry [J]. *Journal of Remote Sensing*, 2017, 21(6): 982-995. (in Chinese)
- [2] Chen Jian, Jin Xianglong. Advances in and application of airborne laser bathymetry technology [J]. *Marine Science Bulletin*, 2002(6): 75-82, 90. (in Chinese)
- [3] Zhai Guojun, Wang Keping, Liu Yuhong. Technology of airborne laser bathymetry [J]. *Hydrographic Surveying and Charting*, 2014, 34(2): 72-75. (in Chinese)
- [4] Li Hongpeng, Li Guoyuan, Cai Zhijian, et al. Full-waveform LiDAR echo decomposition method [J]. *Journal of Remote Sensing*, 2019, 23(1): 89-98. (in Chinese)
- [5] Liu Zhimin, Yang Anxiu, Yang Fanlin, et al. The feasibility analysis for the airborne LiDAR bathymetry application in marine surveying and charting [J]. *Hydrographic Surveying and Charting*, 2018, 38(4): 43-47. (in Chinese)
- [6] Zhao Jianhu, Ouyang Yongzhong, Wang Aixue. Status and development tendency for seafloor terrain measurement technology [J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(10): 1786-1794. (in Chinese)
- [7] Hu Haiying, Hui Zhenyang, Li Na. Airborne LiDAR point cloud classification based on Multiple-entity eigenvector fusion [J]. *Chinese Journal of Lasers*, 2020, 47(8): 0810002. (in Chinese)
- [8] Hu Tiancheng, Tao Bangyi, Mao Zhihua, et al. Classification of sea and land waveform based on multi-channel ocean lidar [J]. *Chinese Journal of Lasers*, 2017, 44(6): 0610002. (in Chinese)
- [9] Deng Qian, Wu Decheng, Kuang Zhiqiang, et al. 532 nm/660 nm dual wavelength lidar for self-calibration of water vapor mixing ratio [J]. *Infrared and Laser Engineering*, 2018, 47(12): 1230004. (in Chinese)
- [10] Ji X, Tang Q, Xu W, et al. Island feature classification for single-wavelength airborne lidar bathymetry based on full-waveform parameters [J]. *Applied Optics*, 2021, 60(11): 3055-3061.
- [11] Zhang Yang, Huang Weidong, Dong Changzhe, et al. Research on the development of the detection satellite technology in oceanographic lidar [J]. *Infrared and Laser Engineering*, 2020, 49(11): 20201045. (in Chinese)
- [12] Fernandez-Diaz J C, Glennie C L, Carter W E, et al. Early results of simultaneous terrain and shallow water bathymetry mapping using a single-wavelength airborne LiDAR sensor [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(2): 623-635.
- [13] Cao Bincai, Zhu Shulong, Qiu Zhengge, et al. Water-land classification for LiDAR bathymetric data based on echo waveform characteristics [J]. *Hydrographic Surveying and Charting*, 2018, 38(3): 12-16. (in Chinese)
- [14] Guenther G C, LaRocque P E, Lillycrop W J. Multiple surface channels in Scanning Hydrographic Operational Airborne Lidar Survey (SHOALS) airborne lidar [C]//Proceedings of SPIE, Ocean Optics XII, 1994, 2258: 1-9.
- [15] Wang Jinhui, Li Chuanrong, Zhou Mei. Analysis of airborne full-waveform LiDAR data for supervised point cloud classification [J]. *Remote Sensing Information*, 2013, 28(5): 21-27. (in Chinese)
- [16] Fan Shijun. Research on classification for airborne full waveform Lidar data based on random forest [J]. *Journal of*

- Capital Normal University (Natural Science Edition)*, 2013, 34(5): 71-73, 78. (in Chinese)
- [17] Wang Dandi, Xing Shuai, Xu Qing, et al. Automatic sea-land waveform classification method for single-wavelength airborne LiDAR bathymetry [J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51(5): 750-761. (in Chinese)
- [18] Yang Yuchuan, Tan Bitao, Long Chao, et al. Time-domain characteristics of active detecting echo laser pulse [J]. *Infrared and Laser Engineering*, 2013, 42(12): 3228-3233. (in Chinese)
- [19] Zuo Zhiqiang, Tang Xinming, Li Guoyuan, et al. Adaptive Gaussian filtering of the full waveform of GF-7 satellite laser altimeter [J]. *Infrared and Laser Engineering*, 2020, 49(11): 20200251. (in Chinese)
- [20] Xu Zhiniu, Hu Zhiwei, Zhao Lijuan, et al. Highly accurate key parameters extraction algorithm for Brillouin scattering spectrum using Voigt profile [J]. *Infrared and Laser Engineering*, 2018, 47(S1): S122004. (in Chinese)
- [21] Wu Tian, Fu Daorui, Peng Yong, et al. Characteristic parameters of pulse waveform of potential transfer current for live working on tension tower of 1000 kV transmission line [J]. *Power System Technology*, 2022, 46(5): 2009-2016. (in Chinese)
- [22] Xu Peituo, Tao Yuting, Liu Zhipeng, et al. Comparison of oceanic lidar experiments and simulation results [J]. *Infrared and Laser Engineering*, 2020, 49(2): 0203007. (in Chinese)
- [23] Wang Guoli, Hao Yanpeng, Li Yanming. Pulse current parameter evaluation of typical partial discharge models in oil [J]. *Power System Technology*, 2001(10): 34-37, 46. (in Chinese)
- [24] Zhao Yang, Qie Xiushu, Kong Xiangzhen, et al. Analysis on the parameters of the current waveforms of triggered lightning [J]. *Acta Physica Sinica*, 2009, 58(9): 6616-6626. (in Chinese)
- [25] Sun Jie, Lai Zulong. Airborne LiDAR feature selection for urban classification using random forests [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(11): 1310-1313. (in Chinese)
- [26] Xiong Yan, Gao Renqiang, Xu Zhanya. Random forest method for dimension reduction and point cloud classification based on airborne LiDAR [J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(4): 508-518. (in Chinese)
- [27] Song Shufang, He Ruyang. Importance measure index system based on random forest [J]. *Journal of National University of Defense Technology*, 2021, 43(2): 25-32. (in Chinese)
- [28] He Yun, Huang Chong, Li He, et al. Land-cover classification of random forest based on Sentinel-2A image feature optimization [J]. *Resources Science*, 2019, 41(5): 992-1001. (in Chinese)

## Classification of full waveform data for monochromatic airborne LiDAR bathymetry based on waveform morphological features

Li Yingying<sup>1</sup>, Liu Ziwei<sup>1</sup>, Zhang Jingkun<sup>2</sup>, Wu Linlin<sup>1</sup>, Ji Xue<sup>1,3,4\*</sup>, Wang Mingchang<sup>1\*</sup>

(1. Academy of Earth Exploration Science and Technology, Jilin University, Changchun 130000, China;

2. Academy of Urban Rail Transportation, Shanghai University of Engineering and Technology, Shanghai 201620, China;

3. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China;

4. Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources, Qingdao 266061, China)

### Abstract:

**Objective** Monochromatic airborne LiDAR bathymetry becomes considerably favorable for topography and geomorphology detection over coastal area by means of its low cost, low load and high sampling rate. However, addressing the limitation of single wavelength to realize the accurate division of full waveform data independently from auxiliary sensor becomes the critical part of coordinate calculation. Given the existing literatures, there is a lack of systematic evaluation analysis and general conclusions for waveform classification contraposing to full waveform morphological features.

**Methods** In view of the latest development of waveform features extraction, refined waveform categories (anomalies, over-fitted, land, sea surface and bathymetry waveforms), 24-dimensional waveform features are

designed and calculated upon systematic analysis on morphological characteristics of different waveforms, and then their classification performance and optimal feature combination are evaluated and quantitatively analyzed utilizing random forest feature selection and classification model.

**Results and Discussions** The results proved that the combination of 6-dimensional features (Fig.8-11), including deviation of amplitude between two adjacent points and oscillating main frequency, is the most effective in classifying five waveforms, with an overall classification accuracy of 98.55% and a Kappa coefficient of 0.982 0 (Fig.9-10, Fig.12, Tab.1). To verify the universality of the features, an additional experimental area was selected for validation and the overall accuracy of water and land classification was 96.81% (Fig.13).

**Conclusions** To accurately identify waveforms, a systematic analysis was conducted to determine the morphological differences between different types of waveforms, and 24-dimensional feature parameters were extracted. After the optimal feature combination and classification performance evaluation, it was found that the 6-dimensional features of oscillating main frequency  $f$ , ratio of peak  $R_p$ , deviation of amplitude between two adjacent points  $\Delta A$ , maximum intensity  $Wf-I$ , decay constant  $a$ , and first echo peak  $Fpk$  were highly effective in distinguishing these five types of waveforms, where 100% of the anomalies and over-fitted waveforms were extracted, strongly confirming the relevance and validity of the morphological features. After replacing the experimental area, the accuracy of the water and land classification reached 96.81%, proving that the features and methods used were adaptable and generalizable, and could meet the production requirements. The 2% decrease in waveform classification accuracy after changing the study area is mainly due to the varying equipment parameter settings and coverage feature categories in different experimental areas. Limited sample selection further compounds this issue. To maintain accuracy, the sample data can be appropriately supplemented according to the actual situation in the experimental area. Although the waveform morphology has been studied thoroughly, additional experimental evidence is necessary to ascertain the impact of intrinsic factors such as signal reception systems on the echoes. To this end, future research will focus on expanding the study's data and signal detection.

**Key words:** airborne LiDAR bathymetry; full waveform; waveform characteristics; waveform classification

**Funding projects:** Open Research Fund Program of LIESMARS (22S02); Open Fund of Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources (2022002); National Natural Science Foundation of China (22FAA01871, 42301505, 52305445); Natural Science Foundation of Jilin Province (20210101098JC)