

机器学习视角下的多光谱卫星浅海水深反演研究

刘瑾璐¹, 孙德勇^{1,2*}, 孔得雨³, 潘锡山³, 焦红波⁴, 李正浩¹, 王胜强^{1,2}, 何宜军^{1,2}¹南京信息工程大学海洋科学学院, 江苏 南京 210044;²自然资源部空间海洋遥感与应用重点实验室, 江苏 南京 210044;³江苏省海涂研究中心, 江苏 南京 210044;⁴国家海洋信息中心, 天津 300171

摘要 利用覆盖典型海岛的 Landsat-8 OLI 多光谱卫星遥感影像和收集到的水深数据, 分别采用传统多元线性回归模型、机器学习中的 back propagation (BP) 神经网络模型和随机森林模型对目标海域水深进行整体反演, 并对三种方法的反演精度进行评价。结果表明: 相比于多元线性回归模型, 机器学习方法的水深反演精度更高; 随机森林模型的水深反演精度最高, 平均绝对误差为 1.94 m, 平均绝对百分比误差为 18.29%, 模型的鲁棒性更加出色, 整体精度较多元线性回归模型有明显提高。本研究比较三种方法构建的浅海水深模型的性能, 为后续更加高效地获取高精度浅海水深信息提供参考价值。

关键词 Landsat-8 OLI; 机器学习; 水深反演; 多光谱影像

中图分类号 TP79

文献标志码 A

DOI: 10.3788/LOP220584

Shallow Water Depth Inversed Using Multispectral Satellite Based on Machine Learning

Liu Jinlu¹, Sun Deyong^{1,2*}, Kong Deyu³, Pan Xishan³, Jiao Hongbo⁴, Li Zhenghao¹, Wang Shengqiang^{1,2}, He Yijun^{1,2}¹School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;²Key Laboratory of Space Ocean Remote Sensing and Application, Ministry of Natural Resources, Nanjing 210044, Jiangsu, China;³Jiangsu Provincial Marine Environment Monitoring Engineering Technology Research Center, Nanjing 210044, Jiangsu, China;⁴National Marine Data and Information Service, Tianjin 300171, China

Abstract Using the Landsat-8 OLI multispectral satellite remote sensing images covering typical islands and collected water depth data, this study comprehensively invert the water depth of the target sea area using the traditional multiple linear regression model, back propagation neural network model and random forest model in machine learning. The inversion accuracy of the three methods is evaluated. The results show that compared with the multiple linear regression model, machine learning methods have higher water depth inversion accuracy. The water depth inversion accuracy of the random forest model is the highest with a mean absolute error of 1.94 m and a mean absolute percentage error of 18.29%, and the robustness of the model is better, and the overall accuracy is significantly improved compared with that of the multiple linear regression model. This study compares the performance of shallow water bathymetric models built using the three methods, providing reference value for subsequent research on obtaining high-precision shallow water bathymetric information more efficiently.

Key words Landsat-8 OLI; machine learning; water depth inversion; multispectral image

收稿日期: 2022-01-20; 修回日期: 2022-02-15; 录用日期: 2022-02-25; 网络首发日期: 2022-03-09

基金项目: 国家自然科学基金(42176179, 41876203, 42176181)、江苏省基础研究计划(自然科学基金)项目(BK20211289, BK20210667)

通信作者: *sundeyong@nuist.edu.cn

1 引言

大多数海洋活动和海洋空间规划都需要深度数据,水深对海岸建设、航行安全、资源勘探、近岸和潮间带生态系统研究至关重要^[1-4]。传统的水深测量主要采用船载声呐测量方式与主动非成像机载激光雷达测量方式,精度较高,但费时费力,经济成本高昂,且易受天气影响,无法开展大范围连续测量^[5-7]。遥感技术的发展为水深测量带来了新思路,利用卫星影像可以进行大范围、低成本的重叠观测,并且可以对难以到达的海域进行水下地形检测^[8-9],一定程度上弥补了现场测量的不足。

随着海洋卫星遥感技术的不断发展,利用卫星多光谱数据反演浅海水深已逐渐在水体测深遥感方面发展,形成了理论解析模型、经验模型和半理论半经验模型。理论解析模型基于辐射传输过程建立水深反演方程,需要大量的水体参数参与运算,但实际应用中难以获取,因此极少应用^[10-14]。在理论解析模型基础上,利用辐射能量在水体中传输的衰减特性,并结合一定数量的水深实测数据作为先验值,形成了半理论半经验模型^[15-16],该方法减少了中间参数的量化过程。经验模型则将卫星影像的遥感反射率及各类波段组合作为自变量,水深值作为因变量,借助统计回归等方法来实现水深反演^[17-18]。随着计算机技术的发展,除了简单的线性回归外,国内外学者也逐渐将机器学习算法应用于水深反演上,机器学习算法在解决多变量、非线性复杂问题等方面具有更好的拟合能力,能够克服人为

因素的局限性。

1998年,Sandidge等^[19]在高光谱影像上利用back propagation(BP)神经网络模型反演水深。王艳姣等^[20]利用动量BP神经网络建立水深反演模型,并对长江口南港河段水深进行了反演。Manessa等^[21]利用随机森林和线性回归分别对吉里曼特拉岛和庞岗岛地区进行水深反演,发现随机森林的性能明显优于线性回归。Zhang等^[22]利用黄岩岛地区的高分辨率遥感影像及相应实测水深,采用随机森林算法构建水深反演模型,发现训练集中的光谱特征越多,遥感反演的效果越好。温开祥等^[23]也对比了三种机器学习方法,在反演梅州水库时发现随机森林模型的反演效果最佳。

综上所述,目前大多数水深遥感反演研究主要是针对单一区域建立独立模型(分区域模型)的,而将多个区域视为整体进行数据整合的整体建模(联合建模)鲜有研究。本研究将以多个典型海岛为例,进行联合建模,分别通过多元线性回归模型、BP神经网络模型、随机森林模型来对反演结果的精度进行验证和比较,得出最佳反演模型,同时通过对比各海岛独立建模的结果来研究大水深数据量对模型精度的增益,以期对相关研究提供一定的参考依据。

2 材料与方法

2.1 研究区域

研究的区域如图1所示,分别为如图1(a)所示的位于孟加拉湾与缅甸海之间、十度海峡之北的南安达

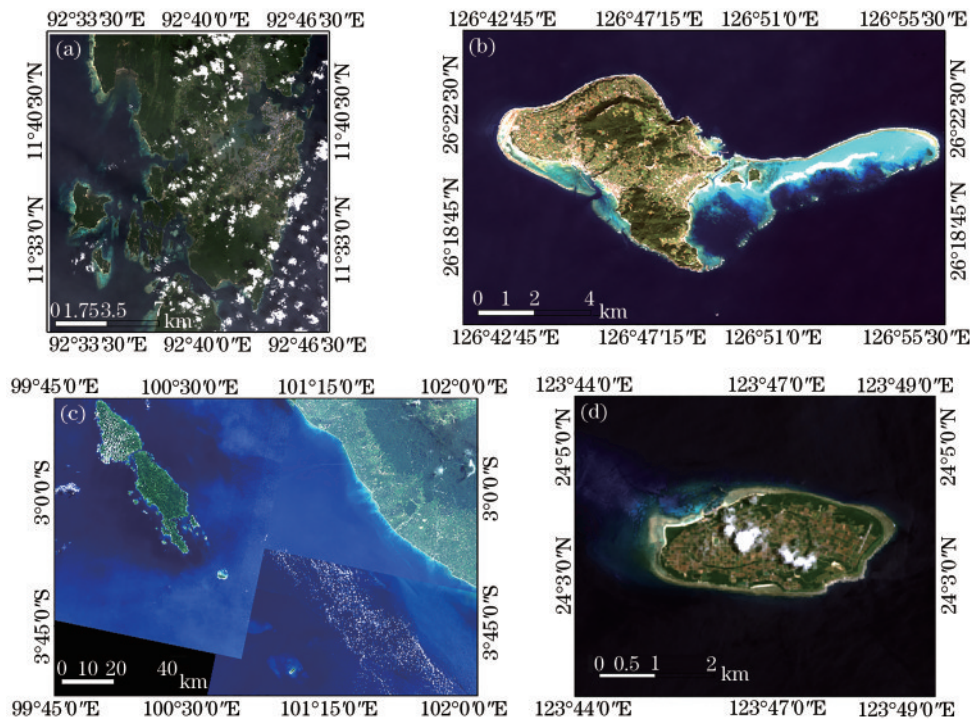


图1 研究区域。(a)南安达曼群岛;(b)久米岛;(c)明打威群岛;(d)波照间岛

Fig. 1 Study area. (a) South Andaman islands; (b) Kume-jima; (c) Mentawai islands; (d) Hateruma-jima

曼群岛 (South Andaman islands), 如图 1(b) 所示的冲绳群岛最西端的久米岛 (Kume-jima), 如图 1(c) 所示的明古鲁西部附属岛链的明打威群岛 (Mentawai islands) 及如图 1(d) 所示的琉球列岛八重山群岛最南端的波照间岛 (Hateruma-jima)。4 个岛屿近岸海域水体清澈, 均为一类水体, 受水体质量影响小, 具有一定典型海岛的代表性, 适宜于作为水深反演的典型区域, 可以更好地展现水深反演的效果; 同时, 这些海岛的定量水深反演工作开展较少, 可以一定程度上弥补研究区域水深数据的不足, 为海岛冲浪等海上活动增加一份安全保障。其中, 南安达曼群岛位于 11.46° N~11.75° N, 92.51° E~92.78° E; 久米岛位于 26.26° N~26.42° N, 126.67° E~126.09° E; 明打威群岛位于

2.49° S~4.16° S, 99.11° E~102.36° E; 波照间岛位于 24.02° N~24.10° N, 123.72° E~123.83° E, 呈椭圆形。研究区域的具体位置分布和真彩色合成影像如图 1 所示。本文所研究的浅海主要指岛屿周围海水深度小于等于 20 m 的区域。

2.2 数据及预处理

共收集覆盖 4 个研究区域的清晰无云 Landsat-8 OLI 影像数据 (<https://glovis.usgs.gov/>), 影像时间和传感器参数如表 1 所示。通过商业途径获取研究区水深栅格数据及电子海图数据 (ENC) 作为实测水深数据集, 其中水深栅格数据分辨率为 90 m。通过卫星数据与水深栅格或海图数据的空间匹配, 提供后续的建模与验证样本数据集, 整体技术流程如图 2 所示。

表 1 数据源介绍
Table 1 Data source introduction

Sensor	Wavelength / μm	Central wavelength / μm	Resolution / m	Imaging time
Landsat-8 OLI	Coastal: 0.430-0.450 (B ₁)	Coastal: 0.440	30	Hateruma-jima 2013-06-05
	B: 0.450-0.510 (B ₂)	B: 0.480		Kume-jima 2015-12-13
	G: 0.550-0.590 (B ₃)	G: 0.570		South Andaman 2021-03-10
	R: 0.640-0.670 (B ₄)	R: 0.655		Mentawai 2019-05-02,
	NIR: 0.850-0.880 (B ₅)	NIR: 0.865		2019-05-27, 2019-01-26
	SWIR1: 1.570-1.650 (B ₆)	SWIR1: 1.610		
	SWIR2: 2.110-2.290 (B ₇)	SWIR2: 2.200		

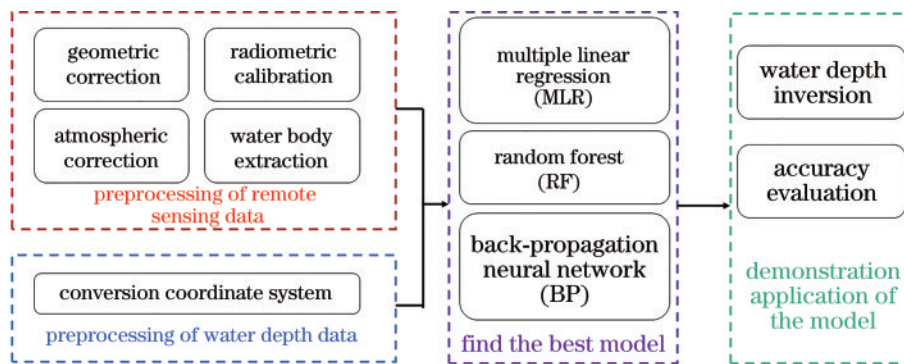


图 2 技术路线图

Fig. 2 Technology roadmap

对收集的 Landsat-8 OLI 数据进行辐射校正和几何校正处理, 并运用 simplified and robust surface reflectance estimation method (SREM)^[24] 实现大气校正, 无需气溶胶参数即可得到接近真实的地表反射率。再借助归一化水体指数 (NDWI) 对影像进行水陆分离, 提取水体区域。最后, 对遥感影像与水深数据进行配准, 对水深数据进行投影转换, 使其与遥感影像的投影信息相一致, 再通过 ArcGIS 空间分析功能提取样本点处的遥感反射率, 得到水深样本。4 个研究区域累积共获得 7105 个样本点, 基于此数据集开展联合建模。

2.3 浅海水深模型构建

2.3.1 多元线性回归模型

多元线性回归 (MLR) 模型中经典的统计学方法旨在建立遥感数据与收集的水深值之间的线性关系模型, 通过分析遥感反射率单波段或波段组合 (X) 与水深数据的相关性, 选取最优反演因子或反演因子组合, 其中 X 主要包括 B_i、Ln(B_i)、B_i/B_j、Ln(B_i)/Ln(B_j) 四种形式, B_i 与 B_j 分别代表遥感反射率数据的第 i 与第 j 个波段, 构建反演因子与水深数据回归模型。

多元线性回归模型形式为

$$Z = A_0 + A_1 X_1 + A_2 X_2 + \dots + A_n X_n, \quad (1)$$

式中: X_i 为反演因子; A_i 为 X_i 的系数; A_0 为常数项。

2.3.2 BP神经网络模型

BP神经网络是人工神经网络的一个常见分支,是利用预测值和真实值之间误差的反向传播来进行训练的一种多层前馈神经网络,工作原理是模拟脑神经组织,将结构分为输入层、输出层和隐藏层,主要特点是误差反向传播和信号前向反馈^[25-26]。在前向传播过程中,输入数据集自输入层经过隐藏层进行逐层处理,最后至输出层,每层神经元的状态仅影响下一层,如果在输出层未获得期望输出值时,则误差信号将反向传播,沿着原始连接路径返回,并修改每个神经元的权重以最小化误差信号。在水深遥感反演中,BP神经网络的输入层负责接收反演因子,隐藏层对输入层传来的数据进行处理后送到输出层,获得水深值,结构如图3所示。这个过程主要通过调用Matlab2016中的net函数实现,在进行网络训练前需要设置参数,包括激活函数、隐含层个数、训练目标误差、学习率及训练次数等。

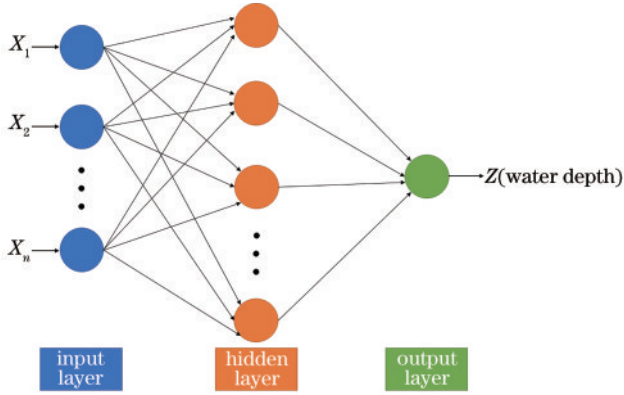


图3 BP神经网络示意图

Fig. 3 Schematic of BP neural network

2.3.3 随机森林算法

随机森林(RF)是一种监督学习算法,为了克服使用单个决策树时过拟合和不稳定的缺点,随机森林在原始训练数据集的随机子集上独立构建多个决策树。这些单棵树的预测平均值用于提高模型的鲁棒性^[27-28]。对每个决策树执行自抽样,并使用袋外的样本数据分析错误。利用Matlab2016的TreeBagger函数,设置树个数和叶子节点数等参数。

设训练数据中有 M 个样本单元和 N 个变量。随机森林模型结构如图4所示,构建过程如下。

1) 从原始训练数据中随机抽取 M 个样本进行采样,生成训练集。未收集到的样本数据构成袋外数据集并生成决策树。

2) 在每棵树的每个节点上,随机选取小于 N 个变量作为节点分段的候选变量。每个节点上的变量树是一致的。

3) 递归分割每棵树,至该节点全部的训练样本均属同一类,不再进行分割。最后平均所有树的预测结

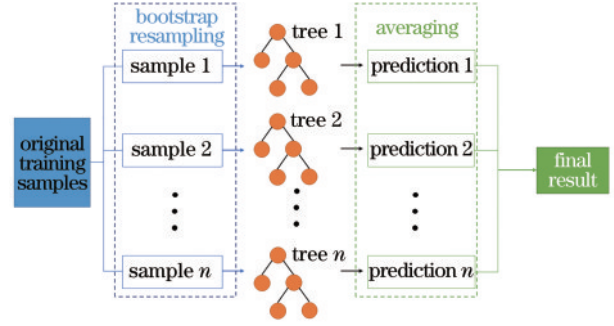


图4 随机森林示意图

Fig. 4 Schematic of random forest

果,得到预测值。

2.4 精度评价参数

本研究所使用的精度评价指标为平均绝对百分比误差(MAPE)、平均绝对误差(MAE)、决定系数(R^2),表达式为

$$E_{\text{MAP}} = \frac{\sum_{i=1}^n (\Delta Z_i / Z_i)}{n}, \quad (2)$$

$$E_{\text{MA}} = \frac{\sum_{i=1}^n (\Delta Z_i)}{n}, \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n Z_i Y_i - \frac{\sum_{i=1}^n Z_i \sum_{i=1}^n Y_i}{n}}{\sqrt{\left[\sum_{i=1}^n Z_i^2 - \frac{(\sum_{i=1}^n Z_i)^2}{n} \right] \times \left[\sum_{i=1}^n Y_i^2 - \frac{(\sum_{i=1}^n Y_i)^2}{n} \right]}}, \quad (4)$$

式中: $\Delta Z_i = |Z_i - Y_i|$, Z_i 和 Y_i 分别为第 i 个检查点的实测水深值和反演水深值; n 为水深检查点数量。此外,本研究中使用深度学习方法构建浅海水模型时,根据水深样本数值分布,随机均匀选择 5000 条水深数据作为训练样本,剩下的 2105 条数据作为验证样本,训练样本和验证样本的比例约为 7:3。

3 模型建立与验证

3.1 模型建立及评价

所建立的三种浅海水深反演模型的主要信息如下所示。

1) 多元线性回归模型。首先将联合建模训练集中 5000 个样本的遥感反射率及所有波段组合输入 SPSS 软件中,利用逐步回归方法确定反演因子,再利用最小二乘法,基于建模样本,进行反演因子参数 (a_i) 优化,确定模型形式。最终确定的多元线性回归模型形式为

$$Z = 136.16 \times \frac{\ln b_3}{\ln b_4} - 60.26 \times \frac{b_4}{b_2} - 3.07 \times \frac{b_1}{b_4} - 71.98, \quad (5)$$

式中: b_i 为第 i 波段的遥感反射率。

2)BP神经网络模型。将联合建模训练集的5000个样本的遥感反射率及所有波段组合作为输入,对比不同的训练参数,选取最优参数。经多次训练后发现当隐含层设置为20、输入层到隐含层的传递函数为tansig、隐含层到输出层的传递函数为purelin、最大训练次数为1000、训练要求精度为0.001、学习速率为0.05时,模型效果最佳。

3)随机森林模型。与BP神经网络相似,将联合建模训练集中5000个样本的遥感反射率及所有波段组合作为输入,选取不同参数进行训练,确定最优参数,发现当最优叶子节点个数为5、最优树个数为800时,模型效果最佳。

各模型的建模精度结果如表2所示。多元线性回归模型的MAE为2.37 m,MAPE为35.48%, R^2 仅有0.68,拟合精度较低,建模效果最差;BP神经网络模型的MAE为1.93 m,MAPE为17.12%, R^2 为0.76,各精

表2 各模型建模精度比较

Table 2 Comparison of modeling accuracy of all models

Model	MAE /m	MAPE /%	R^2
MLR	2.37	35.48	0.68
BP	1.93	17.12	0.76
RF	1.01	9.02	0.93

度评价参数较多元线性回归模型有明显的提升,整体反演精度提高;随机森林模型的MAE为1.01 m,MAPE为9.02%, R^2 为0.93,精度较BP神经网络有明显提升,各参数为三种模型中最高,具有最好的建模效果。

同时为了分析不同模型的建模精度差异,采用散点图进行对比分析,如图5所示。图中实线为1:1线,当样本点位于实线上方时表示预测值大于实测值,反之则表示预测值小于实测值,越接近1:1线则代表预测值与水深值的差距越小。图中虚线为2 m误差线,样本点在虚线中代表误差小于2 m,反之则表示大于2 m。三种模型的水深反演散点图均分布在1:1线附近。如图5(a)所示,多元线性回归模型的密度中心位于5~15 m的连续区间上,但反演水深值被高估,且有多数散点位于误差线外,表明整体建模效果的预测与实测相比偏高,在15 m以深区域预测值偏低,有大多数样本点出现在下误差线外。如图5(b)所示,BP神经网络模型的预测值分布更加紧凑,两密度中心均处在误差线范围内,整体预测值分布较为平均。从图5(c)可以看到,随机森林模型建模分布则与前两种模型存在明显差异,散点密度中心连续且均处于1:1线上,散点均紧密地围绕在1:1线周围,绝大多数样本点处于2 m误差线内,明显优于前两种模型,建模精度最高。

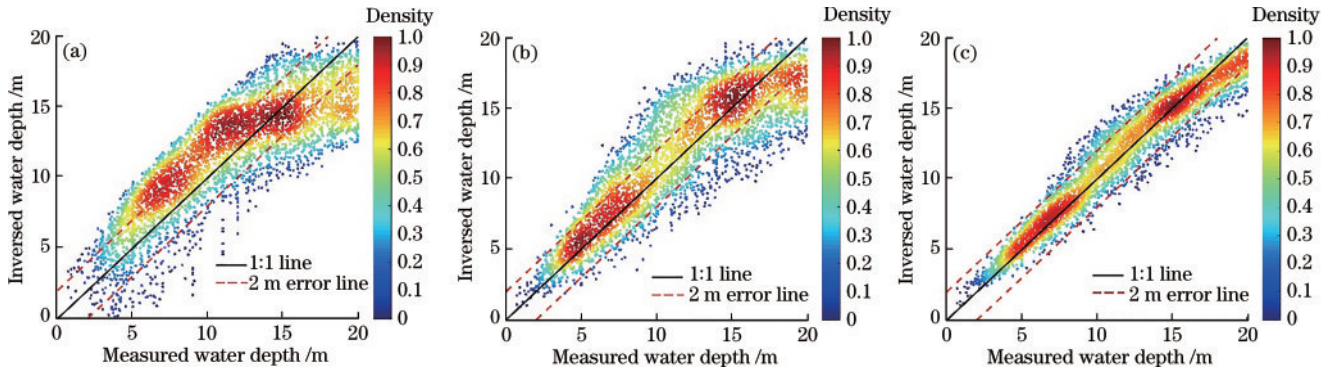


图5 建模精度对比。(a)多元线性回归;(b)BP神经网络;(c)随机森林

Fig. 5 Comparison of modeling accuracy. (a) MLR; (b) BP neural network; (c) RF

3.2 模型验证及评价

为了对比各模型的反演精度,分别将联合建模的验证集的2105样本遥感反射率及所有波段组合输入训练好的3种模型中,结果如表3所示。与建模精度分布的情况相似,多元线性回归模型的MAE为2.45 m,MAPE为25.38%, R^2 仅有0.66,表明该模型拟合精度较低,反演效果最差;BP神经网络模型的MAE为2.05 m,MAPE为19.43%, R^2 为0.74,其中MAE较多

表3 各模型验证精度比较

Table 3 Comparison of validation accuracy of all models

Model	MAE /m	MAPE /%	R^2
MLR	2.45	25.38	0.66
BP	2.05	19.43	0.74
RF	1.94	18.29	0.75

元线性回归模型降低了0.4 m,MAPE降低了5.95个百分点, R^2 提高了0.08,整体反演精度较高;随机森林模型的MAE为1.94 m,MAPE为18.29%, R^2 为0.75,为三种模型里精度最高的一种,具有最好的反演效果。

首先,进行整体的模型结果观察,3种模型的水深反演散点都均匀地分布在1:1线附近,水深在0~20 m区间较为均匀,三种模型均未见极大异常点;多元线性回归模型分布较两种机器学习模型更为分散,其密度中心位于5~15 m的连续区间上,但偏离1:1线,更靠近于误差线,以误差线为中心分布,多点超出误差线偏上,导致整体预测值偏高;BP神经网络模型的预测值分布较多元线性回归更加紧凑,有两个密度中心位于5~10 m及15~20 m区间上,且更接近于1:1线,第一个密度中心处于1:1线与误差线之间,第二个密度中

心则更加偏向于误差线,样本点在误差线内密度最大,可见反演结果较好;随机森林模型预测值分布与BP神经网络相似,也呈现两密度中心,但均以1:1线为中心分布,且大多数点落在误差线内。其次,对反演模型的具体细节进行分析,图6(a)所示的多元线性回归模型在0~10 m深度范围内反演效果较差,有点离1:1线距离较大,在10 m后则整体紧凑,无过多散落点,15 m

以深预测值偏低;图6(b)显示BP神经网络模型在0~10 m的浅水区域预测值较接近真实值,但10~15 m水深区域预测值呈发散趋势,与1:1线距离偏大,15 m以深预测值偏低;图6(c)显示随机森林模型的分布与BP神经网络的分布相近,但在0~10 m处的分布较BP神经网络的分布更加紧凑。

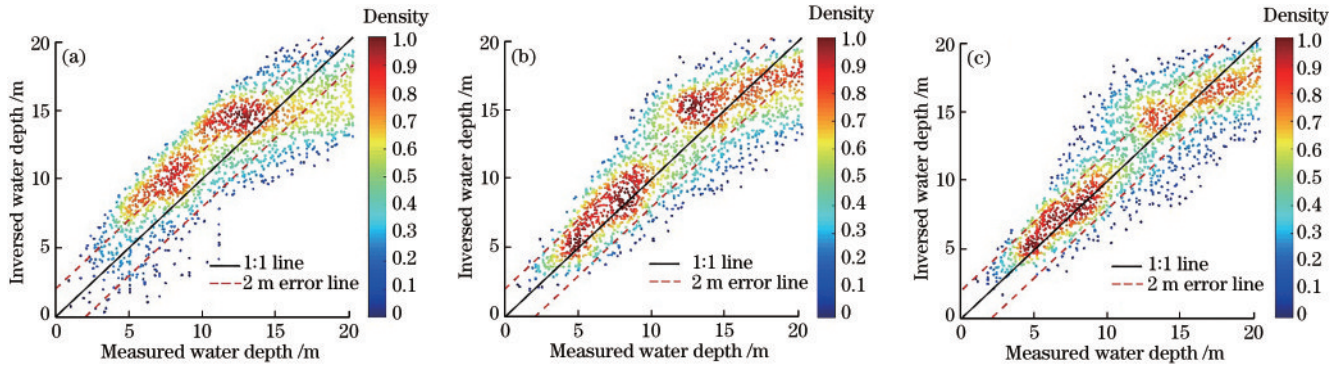


图6 验证精度对比。(a)多元线性回归;(b)BP神经网络;(c)随机森林

Fig. 6 Comparison of validation accuracy. (a) MLR; (b) BP neural network; (c) RF

进一步对比模型的残差分布情况,比较不同模型的优劣差异。图7展示了不同模型的残差分布情况,其中多元线性回归模型的残差分布如图7(a)所示,与散点图的密度分布相似,残差密度中心连续且分布在水深5~15 m,值处在-2~4 m,水深10 m以深的样本的残差值随深度增加逐渐上升,整体趋势由负值转向正值,这导致深水区域的反演结果偏低。BP神经网络

模型的残差分布则更为集中,如图7(b)所示,两不连续的密度中心处于-2~0 m,同样在10 m后残差有整体向上的趋势,在15 m后多为正值。相比以上两模型,随机森林算法的两个残差密度中心出现在5~10 m及15~20 m,且残差值更集中于0 m附近,如图7(c)所示,表明模型拟合效果更好。

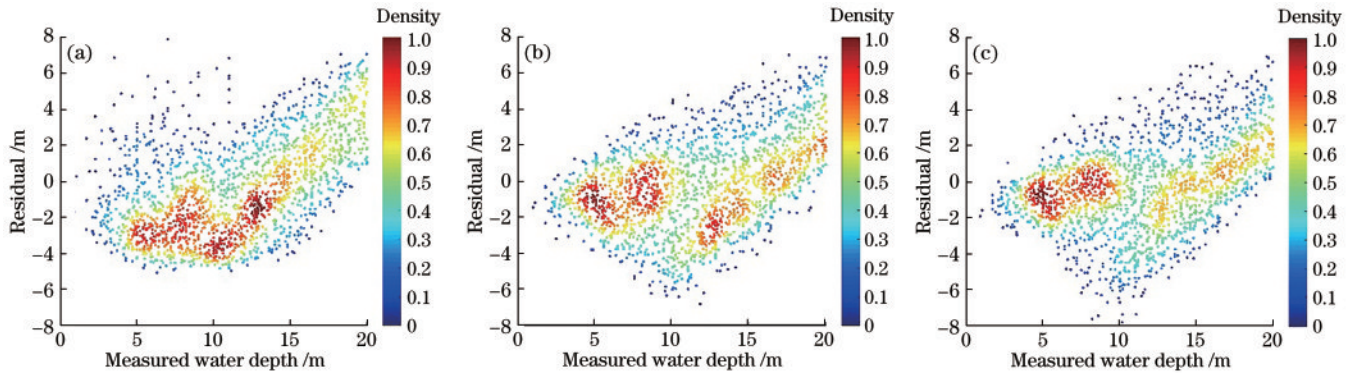


图7 残差分布。(a)多元线性回归;(b)BP神经网络;(c)随机森林

Fig. 7 Residual distribution. (a) MLR; (b) BP neural network; (c) RF

综上所述,多元线性回归模型对水深的预测值偏高,而两种机器学习算法则对水深预测更准确。水深反演结果表明:多元线性回归受限于线性表达,精度最低;BP神经网络与随机森林可以处理复杂的非线性关系。对比两种方法进一步可以看出,建模过程中随机森林算法的性能明显优于BP神经网络,而在验证数据集中两者之间的精度差异较小,协同考虑建模与验证数据集,可以看出随机森林模型的反演精度更佳。

4 讨 论

通过对比发现,在视4个研究区域为1个整体进行整体建模验证时,随机森林模型的精度最高。但上文只进行了整体建模的精度比较,并未建立每个研究区域的分区模型,为了进一步对比分区建模与整体建模的精度,将利用随机森林模型,结合相同的建模验证样本,建立分区模型并分析模型在各区域的反演精度。各区域模型验证精度结果如表4中标注的分区所

表 4 分区模型及整体模型精度对比

Table 4 Precision comparison of sub-region model and whole-region model

Index	Region	South Andaman	Hateruma-jima	Kume-jima	Mentawai
R^2	Sub-region	0.74	0.87	0.71	0.64
	Whole-region	0.74	0.87	0.72	0.66
MAE /m	Sub-region	1.95	1.34	2.30	1.78
	Whole-region	1.95	1.36	2.11	1.80
MAPE /%	Sub-region	18.32	11.60	23.03	18.12
	Whole-region	18.37	11.40	21.19	16.79

示;进一步计算整体模型在各区域的具体精度,如表 4 中标注的整体所示。可以看出,针对不同岛屿,分区模型与整体模型的精度相近,两种方式建立的波照间岛区域模型精度最高,其次是南安达曼群岛,久米岛与明打威群岛略低。如图 8 所示:整体模型与分区模型的决定系数在四个区域相差细微,针对明打威和久米岛区域的整体模型略高于分区模型;MAE 与 MAPE 也是久米岛区域差异较大,数值上分区模型明显高于整体模型,可以反映出整体模型的精度更高,不排除与水深样本质量和相关区域的底质类型复杂程度存在一定的联系。由此可见,将目标区域作为整体来进行水深反演可以一定程度上提高反演精度。

基于上述分析,将建立的随机森林模型应用于目标区域卫星影像上,得到研究区域的水深分布图,如图 9 所示。发现各区域水深整体分布情况类似,都呈现出由近岸向远海逐步上升的趋势。波照间岛在中西部存在浅水区域,与杨晓彤等^[29]对波照间区域得出的水深空间分布一致,如图 9(a)所示。而久米岛浅水主

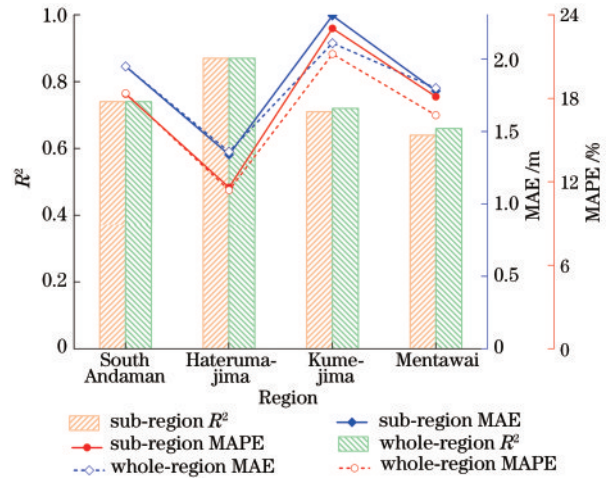


图 8 分区模型与整体模型的精度对比

Fig. 8 Precision comparison between sub-region model and whole-region model

要分布在中东部海域,在东北部整体水深较浅,此处对应久米岛著名观光景点恨之浜,表现为被四周露出的

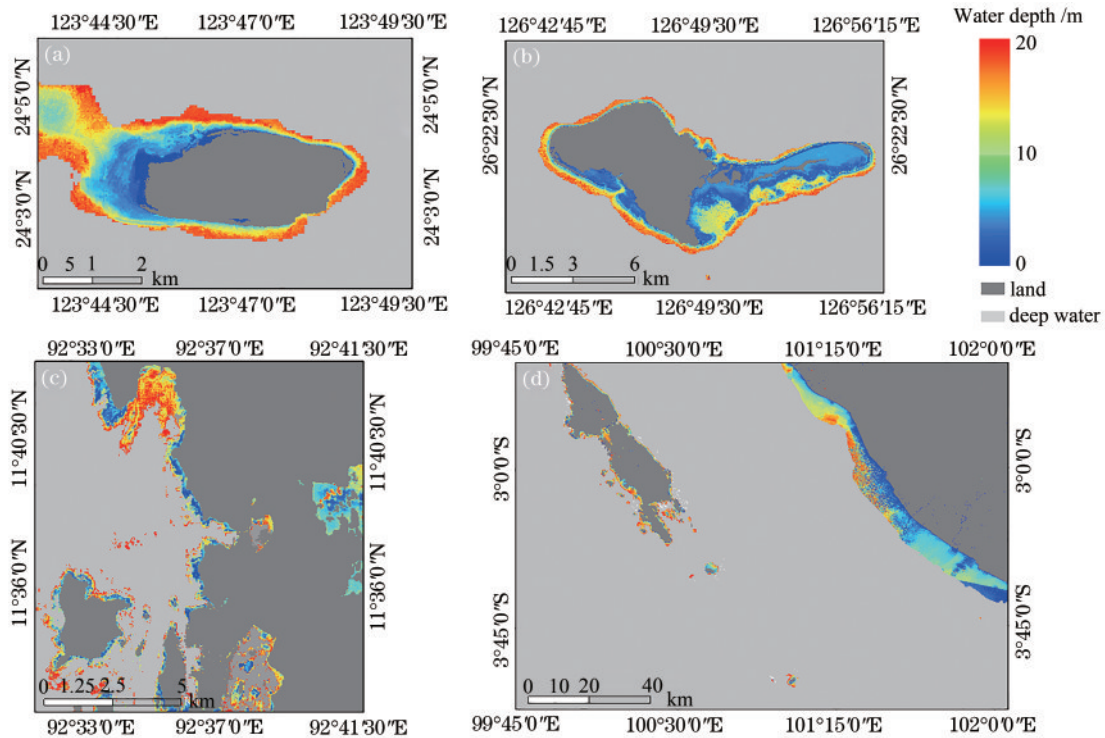


图 9 反演结果图。(a)波照间;(b)久米岛;(c)南安达曼群岛;(d)明打威群岛

Fig. 9 Inversion result. (a) Hateruma-jima; (b) Kume-jima; (c) South Andaman islands; (d) Mentawai islands

沙滩包裹住的一片浅海区域,反演结果符合实际水深空间分布,如图 9(b)所示。南安达曼群岛的水深沿主岛海岸线由浅及深均匀变化,小岛周围也呈现相同趋势,外部深水带呈包围趋势,如图 9(c)所示。明打威地区水深从近岸向远海逐渐变深,符合客观事实,且水深梯度变化均匀,西部岛屿外部被深水带包围,东部区域水深沿岸线向外呈加深状态,整体分布均匀,如图 9(d)所示。综合来看,各区域的浅水水深空间分布都较为相近,结果可信。

本研究针对 4 个不同研究区域(南安达曼群岛、波照间岛、久米岛、明打威群岛),完成了随机森林模型的构建。在实际应用中,发现随机森林算法可以很好地适用于全部区域,具有较高精度。同时与分区域的模型的反演结果相比,发现分区域模型与全区模型精度差异不大,全区模型精度略高。关于机器学习反演水深已进行了很多相关的研究,总体来看,机器学习算法反演水深的精度均高于线性回归,更加适宜于大数据情况。所用的 Landsat-8 OLI 影像数据分辨率仅为 30 m,若借助更高分辨率的卫星影像,如 Superview 系列(精度为 2 m)、Worldview 系列(2 m)等,可以实现更加精准的数据采集,可以更细致地描述研究区域的海底地形特征,更好地服务于多区域浅水水深反演模型建立过程。

5 结 论

本研究结合典型海岛的多光谱影像数据与收集到的水深数据,基于机器学习算法分别构建 BP 神经网络和随机森林模型,并与传统的多元线性回归模型的预测结果进行比较,得到以下结论。基于机器学习的水深反演算法在精度上明显优于多元线性回归,可以更好地拟合非线性关系,描述真实的水深分布情况。通过水深深度划分,可以看出 10 m 以浅的反演精度较 15 m 以深更高,受到天气、水质、悬浮颗粒物等因素的影响,辐射信号对水体的穿透程度会发生改变,后续可进行水深分段反演来提高反演精度。对比分区域水深建模结果发现联合建模的效果与分区域模型的效果相近,优势高于分区域建模效果,可见联合建模可以在一定程度上提高建模效率。同时,对于机器学习而言,更多的训练样本在一定程度上能提升模型的精度。

参 考 文 献

- [1] Dahlgren T G, Schläppy M L, Šaškov A, et al. Assessing the impact of windfarms in subtidal, exposed marine areas[M]//Shields M A, Payne A I L. Marine renewable energy technology and environmental interactions. Humanity and the sea. Dordrecht: Springer Netherlands, 2014: 39-48.
- [2] McCombs M P, Mulligan R P, Boegman L. Offshore wind farm impacts on surface waves and circulation in Eastern Lake Ontario[J]. Coastal Engineering, 2014, 93: 32-39.
- [3] 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.
Ma Y, Zhang J, Zhang J Y, et al. Progress in shallow water depth mapping from optical remote sensing[J]. Advances in Marine Science, 2018, 36(3): 331-351.
- [4] 潘信亮, 杨仁辉, 江涛, 等. 基于 Bi-LSTM 的近岸水体深度反演[J]. 光学学报, 2021, 41(10): 1012003.
Pan X L, Yang R H, Jiang T, et al. Depth inversion of coastal waters based on Bi-LSTM[J]. Acta Optica Sinica, 2021, 41(10): 1012003.
- [5] Coleman J B, Yao X B, Jordan T R, et al. Holes in the ocean: filling voids in bathymetric lidar data[J]. Computers & Geosciences, 2011, 37(4): 474-484.
- [6] Collin A, Archambault P, Long B. Mapping the shallow water seabed habitat with the SHOALS[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(10): 2947-2955.
- [7] 赵建虎, 欧阳永忠, 王爱学. 海底地形测量技术现状及发展趋势[J]. 测绘学报, 2017, 46(10): 1786-1794.
Zhao J H, Ouyang Y Z, Wang A X. Status and development tendency for seafloor terrain measurement technology[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1786-1794.
- [8] Robinson J A, Feldman G C, Kuring N. Data fusion in coral reef mapping: working at multiple scales with SeaWiFS and astronaut photography[C]//Proceedings of the 6th International Conference on Remote Sensing for Marine and Coastal Environments, May 1-3, 2000, Charleston, South Carolina, USA. [S. l.: s. n.], 2000: 473-483.
- [9] Gao J. Bathymetric mapping by means of remote sensing: methods, accuracy and limitations[J]. Progress in Physical Geography, 2009, 33(1): 103-116.
- [10] Lyzenga D R. Passive remote sensing techniques for mapping water depth and bottom features[J]. Applied Optics, 1978, 17(3): 379-383.
- [11] Lee Z, Carder K L, Mobley C D, et al. Hyperspectral remote sensing for shallow waters. 2. deriving bottom depths and water properties by optimization[J]. Applied Optics, 1999, 38(18): 3831-3843.
- [12] Eugenio F, Marcello J, Martin J. High-resolution maps of bathymetry and benthic habitats in shallow-water environments using multispectral remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(7): 3539-3549.
- [13] Kanno A, Koibuchi Y, Isobe M. Statistical combination of spatial interpolation and multispectral remote sensing for shallow water bathymetry[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(1): 64-67.
- [14] 田庆久, 王晶晶, 杜心栋. 江苏近海岸水深遥感研究[J]. 遥感学报, 2007, 11(3): 373-379.
Tian Q J, Wang J J, Du X D. Study on water depth extraction from remote sensing imagery in Jiangsu coastal zone[J]. Journal of Remote Sensing, 2007, 11(3): 373-379.
- [15] Philpot W D. Bathymetric mapping with passive multispectral imagery[J]. Applied Optics, 1989, 28(8): 1028007-8

- 1569-1578.
- [16] Stumpf R P, Holderied K, Sinclair M. Determination of water depth with high-resolution satellite imagery over variable bottom types[J]. *Limnology and Oceanography*, 2003, 48(1): 547-556.
- [17] 党福星, 丁谦. 利用多波段卫星数据进行浅海水深反演方法研究[J]. *海洋通报*, 2003, 22(3): 55-60.
Dang F X, Ding Q. A technique for extracting water depth information from multispectral scanner data in the South China Sea[J]. *Marine Science Bulletin*, 2003, 22(3): 55-60.
- [18] 王纪坤, 陈正华, 余克服, 等. 珊瑚礁区多光谱遥感水深反演研究[J]. *遥感技术与应用*, 2018, 33(1): 61-67.
Wang J K, Chen Z H, Yu K F, et al. Water depth information extraction with multi-spectral remote sensing in coral reefs region[J]. *Remote Sensing Technology and Application*, 2018, 33(1): 61-67.
- [19] Sandidge J C, Holyer R J. Coastal bathymetry from hyperspectral observations of water radiance[J]. *Remote Sensing of Environment*, 1998, 65(3): 341-352.
- [20] 王艳姣, 张鹰. 基于 BP 人工神经网络的水体遥感测深方法研究[J]. *海洋工程*, 2005, 23(4): 26-35.
Wang Y J, Zhang Y. Study on remote sensing of water depth based on BP artificial neural networks[J]. *The Ocean Engineering*, 2005, 23(4): 26-35.
- [21] Manessa M D M, Haidar M, Hartuti M, et al. Determination of the best methodology for bathymetry mapping using spot 6 imagery: a study of 12 empirical algorithms[J]. *International Journal of Remote Sensing and Earth Sciences*, 2018, 14(2): 127-136.
- [22] Zhang J, Li S J, Wang M. Water depth inversion based on Landsat-8 data and random forest algorithm[J]. *Journal of Physics: Conference Series*, 2020, 1437(1): 012073.
- [23] 温开祥, 李勇, 王华, 等. 基于遥感和机器学习的内陆水体水深反演技术[J]. *热带地理*, 2020, 40(2): 314-322.
Wen K X, Li Y, Wang H, et al. Estimating inland water depth based on remote sensing and machine learning technique[J]. *Tropical Geography*, 2020, 40(2): 314-322.
- [24] Bilal M, Nazeer M, Nichol J E, et al. A simplified and robust surface reflectance estimation method (SREM) for use over diverse land surfaces using multi-sensor data[J]. *Remote Sensing*, 2019, 11(11): 1344.
- [25] 李朝静, 唐幼纯, 黄霞. BP 神经网络的应用综述[J]. *劳动保障世界(理论版)*, 2012(8): 71-74.
Li Z J, Tang Y C, Huang X. Review of application of BP neural network[J]. *Labor Safeguard World*, 2012(8): 71-74.
- [26] Chen L M, Jagota V, Kumar A. Research on optimization of scientific research performance management based on BP neural network[J]. *International Journal of System Assurance Engineering and Management*, 2021: 1-10.
- [27] 王鑫, 贝祎轩, 李连阳, 等. 基于机器学习的水深反演方法研究[J]. *国土资源信息化*, 2021(5): 22-28.
Wang X, Bei Y X, Li L Y, et al. Research on bathymetry retrieval method based on machine learning[J]. *Land and Resources Informatization*, 2021(5): 22-28.
- [28] Belgiu M, Drăguț L. Random forest in remote sensing: a review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31.
- [29] 杨晓彤, 焦红波, 李艳雯, 等. 两种浅海水深快速反演方法对比研究[J]. *测绘科学*, 2017, 42(11): 177-183.
Yang X T, Jiao H B, Li Y W, et al. Comparative research of two methods for fast water depth retrieval for shallow water[J]. *Science of Surveying and Mapping*, 2017, 42(11): 177-183.