

基于海洋激光雷达和 BP 神经网络的叶绿素剖面反演算法

铁凝, 刘秉义*

中国海洋大学信息科学与工程学部海洋技术学院, 山东 青岛 266100

摘要 激光在海水传输过程中,接收视场逐渐增大、多次散射逐渐增强,导致叶绿素剖面反演误差较大。针对该问题,以反向传播(BP)神经网络为基础,建立了一个激光雷达回波信号反演叶绿素 a 浓度剖面的模型。数据集中包含标签-叶绿素 a 浓度与特征-激光雷达回波信号剖面。获得激光雷达回波剖面后,为了增强输入特征,使用随机抽样一致(RANSAC)算法剔除噪声,完成数据集搭建,进而以不同层网络的均方误差确定模型结构。反演结果显示:相比于传统的利用后向散射系数反演叶绿素浓度的方法,基于激光雷达回波的 BP 神经网络叶绿素剖面反演算法在验证集上的相对误差、均方根误差、平均误差分别降低了 34%、0.363 mg/m³ 和 0.213 mg/m³,相关系数提高了 0.18。传统方法在 50 m 深度水体的叶绿素浓度相对误差为 39%~93%,基于神经网络的剖面反演算法对应的相对误差为 17%~36%,反演精度具有较大提升。对于实测数据,LIMC-BPNN 反演结果的相对误差为 13%。结果表明,相比于传统叶绿素反演方法,基于深度学习的叶绿素剖面反演算法能够有效提取激光雷达回波特征,得到更好的反演结果。

关键词 海洋光学; 反向传播神经网络; 垂直剖面; 水体光学参数; 叶绿素 a

中图分类号 P714+.1

文献标志码 A

DOI: 10.3788/AOS230800

1 引言

海洋在地球气候和生态循环中扮演着重要角色,对于海洋的遥感探测,水色遥感能够大面积获取海洋次表层积分数据,而主动光学遥感是使用人工发射辐射源的探测技术,在工作中不依赖太阳光,不受昼夜、光照等外部环境因素限制,能够获得水体剖面信息。激光雷达作为一种主动遥感的的手段,支持多种搭载平台,能够实现对海洋生态系统的水平和垂直探测,是潜力较高的主动探测设备之一。海洋激光雷达能够执行赤潮监测^[1-3]、海洋生物检测^[4]、海洋数据探测^[5-6]等任务。

海水的组成成分复杂,多种元素以一定的物理化学形态存在,例如海水分子、颗粒物和可溶有机物等。浮游植物作为颗粒物的一种,是海洋生态系统中最重要初级生产者^[7-8],与海洋物理化学等过程存在联系,维持着海洋生态系统的能量流动、物质循环及信息传递,为海洋过程研究提供支撑。叶绿素能够表征水体的浮游植物含量,因此叶绿素研究工作的顺利开展是海洋生态研究领域的基础。

海洋激光雷达能够对叶绿素浓度进行探测。激光雷达回波反演叶绿素浓度通常利用激光雷达光学参数

与叶绿素浓度的经验关系进行反演。因此反演激光雷达光学参数是反演叶绿素浓度的基础。激光雷达光学参数包括消光系数与后向散射系数,消光系数的常用反演方法包括斜率法^[9]、Klett 法^[10]和 Fernald 法^[11],后向散射系数的常用反演方法包括扰动法(PR)^[12]。周雨迪^[13]对典型反演方法进行概述,总结其优缺点,表明上述算法都存在一定的限制。考虑到海洋激光雷达信号和叶绿素浓度之间是一个复杂的非线性关系,因此本文通过深度学习的方法,从回波信号中提取后向散射信号特征,建立一个高精度的叶绿素反演模型。

1943 年 McCulloch 等^[14]建立了人工神经元的数学模型,若干人工神经元以层进行排列连接,构建了人工神经网络。2012 年后深度学习领域快速发展,获得了 Dropout 方法^[15]、Adam 优化器^[16]、Transformer 模型^[17]等成果,提升任务结果质量的同时,也扩展了深度学习应用领域。遥感领域的研究者也逐渐使用深度学习开展反演工作,例如 2022 年吴文涵等^[18]建立了基于深度学习的云参量反演方法,2020 年刘林钰等^[19]建立的深度学习 PM_{2.5} 浓度反演算法,但基于深度学习和激光雷达回波反演叶绿素剖面的研究工作至今仍是一片空白,其阻碍在于数据集的构建。而光子传输模拟能够获得大量激光雷达回波数据,解决上述难题,进而可建

收稿日期: 2023-04-10; 修回日期: 2023-05-19; 录用日期: 2023-05-31; 网络首发日期: 2023-06-28

基金项目: 国家重点研发计划(2022YFB3901702,2016YFC1400905)

通信作者: *liubingyi@ouc.edu.cn

立基于深度学习和激光回波反演叶绿素剖面的模型。

本文以海洋激光雷达仿真回波信号作为研究对象,提出了一种基于海洋激光雷达和BP神经网络的叶绿素剖面反演算法。具体研究内容为:首先,基于反向传播(BP)神经网络,搭建了海洋激光雷达回波反演叶绿素浓度的网络模型。随后建立数据集,利用BGC-Argo (Biogeochemical-Argo)浮标中的叶绿素剖面数据与半解析蒙特卡罗仿真算法建立数据集的特征与标签。之后,将数据集运用在模型中进行训练,得到结果。最后定义模型评价标准,根据标准对模型结果进行对比分析。

2 基本原理

2.1 基于后向散射系数的叶绿素反演方法

考虑到数据集中数据多为大洋水体叶绿素剖面,并存在以下假设:水体可基于光学性质划分为 n 层。因此本文选用扰动法^[12]与颗粒物散射系数-叶绿素浓度的经验关系^[20],反演数据集特征,并与BP神经网络搭建的反演模型进行对比。本文将PR法反向后向散射系数求得叶绿素浓度的方法称为PR-Chla法。PR法不考虑水面等的影响,激光雷达在水深 z 处接收到的信号光子数为

$$N_s(z) = A[\beta_0 + \beta(z)] \exp\left[-2\alpha_0 z - 2 \int_0^z \alpha(y) dy\right], \quad (1)$$

其中光学参数分为两部分,变化量与非变化量。当忽略衰减系数变量的部分时,经过对数变换及距离校正后的变化回波信号 $S(z)$ 与不变的回波信号 $S_0(z)$ 可表示为

$$\begin{cases} S_0(z) = \ln(K\beta_0) - 2\alpha_0 z \\ S(z) = \ln[N(z) \times (nH + z)^2] \end{cases}, \quad (2)$$

式中: α_0 为最小二乘线性拟合后的斜率; K 为激光雷达

系统常数; β_0 表示体散射函数不变的部分,可根据探测面积计算获得; n 表示水体折射率; H 表示激光雷达高度。变化的体散射函数 $\beta(z)$ ^[21] 可表示为

$$\beta(z) = \frac{S(z)}{S_0(z)} \beta_0. \quad (3)$$

而 β_0 ^[22] 可表示为

$$\beta_0 = \frac{\exp[S(0)]}{A}, \quad (4)$$

式中: $S(0)$ 表示水面的回波信号; A 为激光雷达系统常数。1972年Petzold^[23]实测了水体的体积散射函数,结合海水后向散射系数可计算获得颗粒物散射相函数,并与颗粒物体散射函数求出颗粒物的后向散射系数 b_{bp} 。

2001年Morel等^[20]建立的颗粒物散射系数 b_p 与叶绿素浓度关系为

$$b_p(\lambda) = a \langle \text{chl} \rangle^b \left(\frac{550}{\lambda} \right), \quad (5)$$

式中: λ 表示波长; a 、 b 是与波长相关的常数。1983年Gordon等^[24]建立了适合一类水体反演的 a 、 b 参数,2001年Morel^[20]基于更大的数据集对其进行更新,使参数更适合。本文运用更新后的参数 $a=0.416$, $b=0.766$ 。

2.2 BP神经网络

1986年Rumelhart等^[25]提出了BP算法。此方法利用链式求导法则优化梯度,随后根据优化器对每一层节点的权重进行更新。BP神经网络在前馈神经网络的基础上添加了反向传播算法,由前向传播算法与反向传播算法共同构成模型结构。图1为BP神经网络拓扑结构,包括输入层、隐藏层和输出层,各层神经元仅与相邻神经元之间相互连接,同层内神经元之间无连接,构成具有层级结构的前馈神经网络。网络的隐藏层可根据问题需求进行扩展。图中, x_i 作为输入,经过隐藏层成为 y_i ,最后通过输出层得到结果 z_i 。

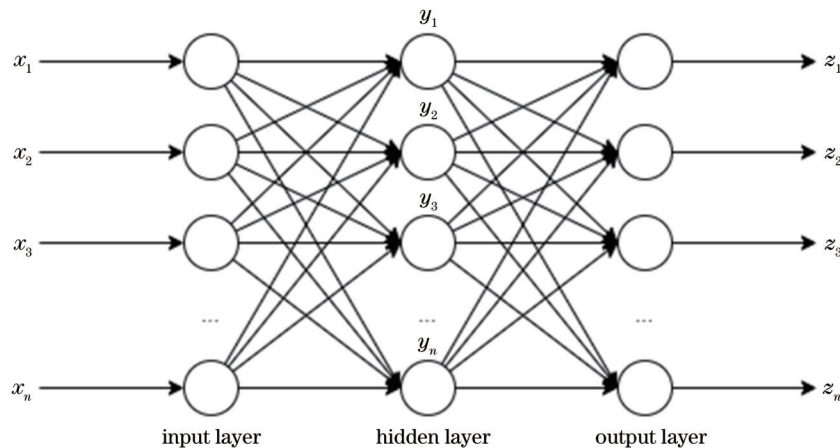


图 1 BP神经网络结构图

Fig. 1 Network structure of BP neural network

图中隐藏层节点的计算公式可描述为

$$y_i = \varphi \left(\sum_{i=1}^n w_i x_i + b_i \right), \quad (6)$$

式中: n 为节点的数量; φ 为激活函数; w_i 为第 i 层的参数权重; b_i 为第 i 层的偏置。

3 模型建立

3.1 LIMC-BPNN 结构

BP神经网络是前馈神经网络,激光雷达回波信号与叶绿素浓度是时间上独立的数据,因此前馈神经网络比反馈神经网络更加适合描述本文研究内容。本文将基于BP神经网络的海洋激光雷达叶绿素反演模型称为LIMC-BPNN(Lidar Inversion Model for Chlorophyll based on Back Propagation Neural Network)。LIMC-BPNN的网络结构如图2所示。

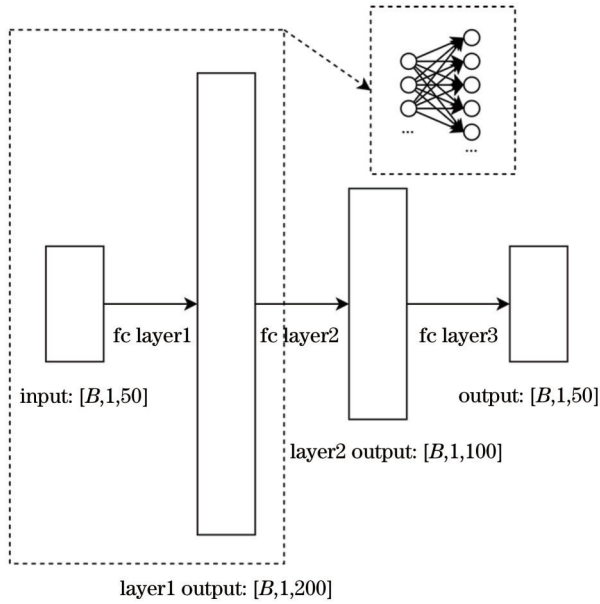


图 2 LIMC-BPNN 的网络结构图

Fig. 2 Network structure of LIMC-BPNN

LIMC-BPNN 的结构利用实验对比确定。表1为四组不同层数网络结构的均方误差结果,对比验证集的损失,找到合适的模型层数。如图2所示,LIMC-BPNN 的网络结构设计为一层输入层、两层隐藏层、一层输出层。输入数据维度为 $[1, 50]$, 代表深度为 50 m 的海洋激光雷达回波信号,训练以批处理形式进行, B 表示批处理大小,因此输入层的数据维度为 $[B, 1, 50]$, 隐藏层节点个数分别为 200、100, 经过隐藏层后输出维度为 $[B, 1, 100]$ 。最终经过输出层,数据维度变为 $[B, 1, 50]$, 代表输出深度为 0~50 m 的叶绿素 a 浓度剖面。LIMC-BPNN 反演叶绿素浓度的计算公式为

$$y = w^{(3)} \times \varphi \left[w^{(2)} \times \varphi \left(w^{(1)} \times x + b^{(1)} \right) + b^{(2)} \right] + b^{(3)}, \quad (7)$$

式中: φ 表示激活函数,用来增加模型的非线性。本文

选用线性整流(ReLU)函数,函数定义为

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases}, \quad (8)$$

式中: x 代表模型的输入; $w^{(i)}$ 表示第 i 层中神经元的权重; $b^{(i)}$ 表示第 i 层中神经元的偏执。均方差(MSE)函数作为损失函数,定义为

$$\text{MSE}(y, y') = \frac{1}{N} \sum_{i=1}^N (y_i - y'_i)^2, \quad (9)$$

式中: N 表示样本数量; y_i 表示数据集的标签值; y'_i 表示模型计算的结果。根据损失函数对模型参数进行优化。优化器选用自适应矩估计算法(Adam),定义为

$$w_{t+1} = w_t - \alpha \cdot \frac{\beta_1 \cdot m_{t-1} + (1 - \beta_1) g_t}{\sqrt{\beta_2 \cdot V_{t-1} + (1 - \beta_2) g_t^2}}, \quad (10)$$

式中: t 为 epoch 的序号; α 表示学习率; β_1, β_2 为一阶矩估计和二阶矩估计的衰减率; m_{t-1} 为 epoch= $t-1$ 时的一阶动量; V_{t-1} 为 epoch= $t-1$ 时的二阶动量; g_t 代表当前参数的梯度。

表 1 不同层数网络结构的 MSE

Table 1 MSE with different layer numbers of network structure

Layer number	2	3	4	5
Best MSE / (mg/m ³)	0.305	0.096	0.181	0.198

3.2 模型训练流程

模型训练首先以 7:2:1 划分训练集、验证集和测试集。模型批处理大小通过对比实验进行确定,表2为各批处理大小的结果。训练过程中利用验证集的误差保存最优模型,通过设置较大的迭代次数,不考虑训练时间,以找到最优的模型参数。学习率为可变学习率,初始学习率为 0.01,网络每迭代 100 次学习率下降 0.5。然后对网络权重参数赋初始值,将数据按照批处理大小进行划分,每次输入一组样本,根据 LIMC-BPNN 结构得到输出。完成每个训练周期后,计算训练集与验证集的损失函数,使用优化器对模型进行优化。完成训练集所有数据的训练后,在验证数据集上计算当前模型的损失大小,根据损失判断是否对模型进行保存。若当前验证集模型的均方误差小于在此之前训练中保存的模型,则对模型进行保存。训练完成后,根据保存的模型对测试集的回波信号进行反演,通过结果误差对模型进行评价。实验环境及编程语言的说明如表3所示。

表 2 不同批处理大小的验证集 MSE

Table 2 MSE with different batch sizes

Batch size	16	32	64	128
Best MSE / (mg/m ³)	0.202	0.147	0.177	0.272

为了加快训练速度,模型训练过程中采用图形处理器(GPU)进行加速。

表 3 模型搭建环境说明

Table 3 Model building environment description

Operating system	Programming language	Environment	Framework	GPU
Windows10	Python 3.8	Spyder	Pytorch	RTX A5000

4 数据集搭建过程

作为监督学习的一种,深度学习需要特征与标签共同构成数据集。基于任务特点,将叶绿素浓度数据作为标签,激光雷达回波数据作为特征。叶绿素浓度可通过 BGC-Argo^[26] 浮标获得。BGC-Argo 能够实现对多时空叶绿素浓度数据的全天候观测。蒙特卡罗算法可仿真激光在水中的传输过程,生成激光雷达回波数据。1968 年 Plass 等^[27] 设计了针对水色遥感的蒙特卡罗辐射传输仿真模型,1981 年 Poole 等^[28] 提出半解析方法对传统蒙特卡罗方法进行改进,通过计算位于接收视场角内的光子期望,此部分权重直接被探测器接收以提高算法效率,剩余部分继续在水中继续散射。近年来,中国海洋大学^[29-30],浙江大学^[31] 也相继开展关于蒙特卡罗仿真的多项工作。因此,蒙特卡罗仿真算法可被用于数据集搭建。使用蒙特卡罗算法时需要根据水体散射相函数计算散射量。针对此问题,本文将浮游植物含量与水分子作为影响激光散射的主要因素,而叶绿素浓度能够表征水体的浮游植物含量,因此利用叶绿素浓度数据可计算出叶绿素光学信息,进而获得水体散射量,用于激光雷达回波数据的仿真。

4.1 叶绿素获取

本文利用 BGC-Argo 官方网站 <http://www.Argodatamgt.org> 获得叶绿素浓度数据,样本大小为 19850。对叶绿素的时空分布进行统计,网格大小为经纬度 1°×1°,时间范围 2006 年至 2021 年全年。

同时,对叶绿素浓度的数据范围进行统计,结果如表 4 所示。可以看出,叶绿素峰值浓度多为低浓度范围的叶绿素剖面,虽然也存在高浓度的叶绿素数据,但

是样本数量远远小于低浓度样本数据。

表 4 叶绿素峰值浓度范围统计

Table 4 Chlorophyll peak range statistics

Chlorophyll concentration range	Value
(0.0, 5.0]	19183
(5.0, 10.0]	420
(10.0, 15.0]	100
(15.0, 20.0]	43
(20.0, 25.0]	20
(25.0, 30.0]	27
(30.0, 35.0]	6
>35.0	50

4.2 激光雷达回波获取

为了简化研究,本文忽略了天空背景光。2020 年,朱培志等^[32] 利用叶绿素浓度剖面计算得到叶绿素光学参数。本文采用相同的方法计算叶绿素剖面光学参数,用于激光雷达回波的仿真。具体流程如图 3 所示。1994 年 Lee 等^[33] 给出浮游植物吸收系数与 440 nm 处浮游植物吸收系数 $a_{ph}(440)$ 的关系 $a_{ph}(\lambda) = \{a_0(\lambda) + a_1(\lambda) \ln[a_{ph}(440)]\} a_{ph}(440)$, 其中, $a_0(\lambda)$ 、 $a_1(\lambda)$ 是与波长相关的经验系数,1998 年 Bricaud 等^[34] 提出了 $a_{ph}(440)$ 与叶绿素 a 浓度的经验关系 $a_{ph}(440) = 0.0378 \langle c_{chl} \rangle^{0.627}$, 即可计算叶绿素剖面的吸收系数。1997 年 Pope 等^[35] 利用积分球测量纯水的吸收系数,由此可计算水体吸收系数。1977 年,Morel 等^[36] 给出了纯水的散射相函数,1983 年 Gordon 和 Morel^[24] 提出了颗粒物散射系数与叶绿素浓度的关系,如式(5)所示,可求出颗粒物散射系数。根据颗粒物散射系数与纯水散射系数,可以求出水体剖面散射系数。水体剖面的光束衰减系数 c 可根据水体散射系数 b 与吸收系数 a 求和进行计算。Fournier-Fornad (FF) 相函数考虑了粒子折射率等参数,被广泛应用于数值模拟中^[37], 本文选用 FF 相函数作为颗粒物散射相函数。FF 散射相函数的表达式为

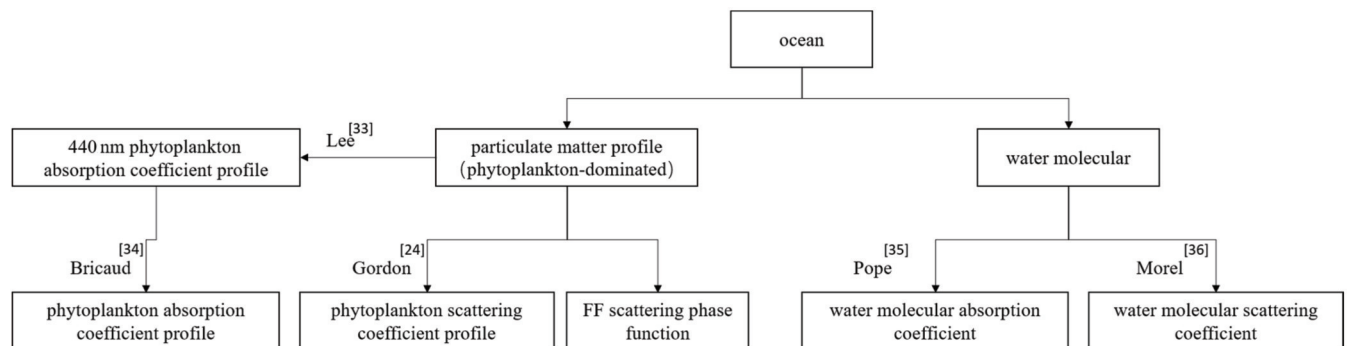


图 3 水体光学参数生成路线

Fig. 3 Flow chart of water optical parameters calculation

$$\tilde{\beta}_{FF}(\phi) = \frac{1}{4\pi(1-\delta)^2\delta^v} \cdot \left[v(1-\delta) - (1-\delta^v) + [\delta(1-\delta^v) - v(1-\delta)] \sin^{-2}\left(\frac{\phi}{2}\right) \right] + \frac{1-\delta_{180}^v}{16\pi(\delta_{180}-1)\delta_{180}^v} (3\cos^2\phi - 1), \quad (11)$$

式中： $v = \frac{3-\mu}{2}$ ； $\delta = \frac{4}{3(n-1)^2} \sin^2\left(\frac{\phi}{2}\right)$ ； n 为粒子的真实折射率，取值 1.138； μ 为双曲分布的斜率参数； δ_{180} 是在散射角为 180° 时的 δ 值。对方程进行积分，得到粒子后向散射分数，

$$B_{FF} = \frac{b_b}{b} = 1 - \frac{1 - \delta_{90}^{v+1} - 0.5(1 - \delta_{90}^v)}{(1 - \delta_{90})\delta_{90}^v}, \quad (12)$$

式中： δ_{90} 为在散射角为 90° 时的 δ 值；因 BGC-Argo 采样的水体多为大洋水体，所以 $\mu = 3.837$ 。

2020 年孔晓娟等^[30]建立了利用水体散射相函数的半解析蒙特卡罗算法。本文利用此方法仿真激光雷达回波。本文的实验假设水体是由 n 层光学性质不同的水体组成，光子在多层介质传输中的原理为 $-\ln(\xi) = \sum_i s_i c_i$ ，其中， c_i 与 s_i 分别是第 i 层的水体光束衰减系数和几何路径长度。随后以随机数产生的光程和水体光束衰减系数确定光子在多层介质中的几何路径长度。获得水体散射相函数后，假设发射光束为垂直向下的无限细准直光束，光子初始位置为 $p = (0, 0, 0)$ ，初始方向余弦为 $(u_x, u_y, u_z) = (0, 0, 1)$ ，光子的初始权重为 1，规定向下方向为正。每次仿真散射时光子权重会随着单次散射反照率进行更新。本文通过设

置散射次数判断光子是否存活。光子运动的几何路径 s 根据随机产生的光程进行计算。确定几何路径后，更新光子坐标，计算散射角与方位角，从而确定新的散射方向^[38-40]。当光子的纵坐标小于 0 时，需根据临界角判断光子是否发生全反射，发生全反射的光子重新回到水中继续运动，而出水的光子需计算出水坐标。同时确定传输方向及新的坐标位置，根据位置判断能否被探测器接收。表 5 为仿真需设置的参数。

表 5 半解析蒙特卡罗仿真参数

Table 5 Parameters for simulation of semi-analytic Monte Carlo

Parameter	Value
Laser wavelength /nm	486
Telescope diameter /m	0.1
Platform height /m	2000
Field of view /mrad	25
Phase function	Fourier-Fornad
Transmission photon counts	10^6
Maximum scattering times	10
Profile resolution /m	0.1

上述算法仿真中使用的叶绿素浓度与仿真得到的激光回波信号如图 4 所示。

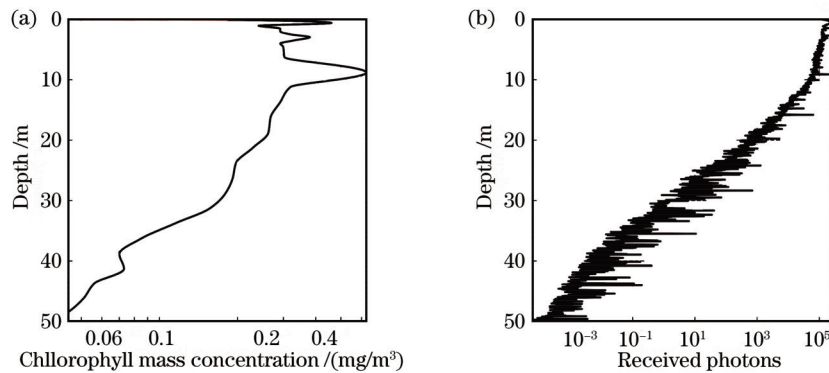


图 4 叶绿素浓度剖面与仿真激光雷达回波结果。(a)Argo-BGC 实测叶绿素浓度剖面图；(b)根据图 3(a)叶绿素剖面通过蒙特卡罗仿真模拟获得的激光雷达回波剖面

Fig. 4 Chlorophyll concentration profile and simulated lidar echo results. (a) Chlorophyll concentration profile from Argo-BGC; (b) lidar profile from Monte Carlo simulation using Fig. 3(a)

从图 4(a)中能看出，在 9 m 处叶绿素浓度激增，而图 4(b)中对应深度的激光雷达回波斜率发生变化。在 10 m 处叶绿素浓度骤减，对应的回波的斜率再次改变。由此可见，激光雷达回波中包含叶绿素浓度的特征，因此可通过激光雷达回波对叶绿素浓度进行反演。

4.3 数据预处理

蒙特卡罗仿真精度受到样本光子数的制约，样本光子数越多，仿真精度越高，随之仿真速度越慢。虽然前期已经通过优化算法、并行计算等技术来提高算法效率，但还需考虑样本数量和仿真精度。若为了缩短仿真时间而减少光子数量，则会因为样本点缺少而导

致噪声数据的产生。建立基于深度学习的叶绿素反演模型需要大量雷达回波信号数据作为数据集,因此需要权衡样本光子数、数据生成时间与生成数据误差。本文利用随机抽样一致(RANSAC)算法^[41]从生成回波信号中剔除缺少样本数导致的噪声,随后利用插值完成回波信号的预处理过程,此方式既缩短了回波信

号的仿真时间,又保证了结果精度。

如图 5(a)所示,通过 RANSAC 算法在 10^6 光子数下分离出内点和外点,对分离出的外点根据前一深度位置的值进行插值,得到最终处理的回波信号数据。从图 4(b)和图 5(b)的对比中能明显看出,激光雷达回波噪声减少。

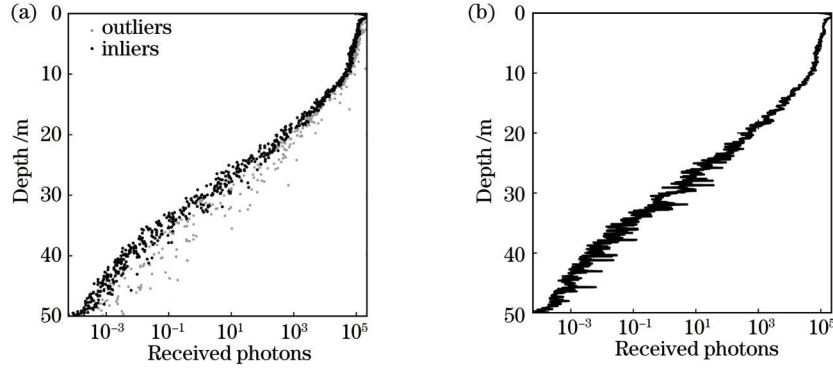


图 5 滤波算法处理前后对比。(a)RANSAC 的计算结果;(b)噪声处理后的激光雷达回波

Fig. 5 Comparison of filtering algorithms before and after processing. (a) Results of RANSAC; (b) lidar echoes after noise processing

5 实验结果与分析

数据集平均反演误差能够表征模型在验证集上的整体反演能力。验证集整体的误差如表 6 所示,包括相对误差(RE)、均方根误差(RMSE)和平均差(ME)。PR-Chla 的反演误差较大,LIMC-BPNN 的误差较小。随机选取两个个例,反演结果如图 6 所示。从图 6 中可以看出,LIMC-BPNN 的反演结果更接近 BGC-Argo 的叶绿素浓度。2017 年 Vaswani 等^[17]将反演得到的后向散射系数与 HSRL 系统进行对比,结果显示随着深度增加后向散射系数的反演误差也会增加,所以 PR-

Chla 方法反演叶绿素浓度误差也随之增加。因此,为了进一步衡量两个模型在不同深度的反演能力,以 10 m 为单位对反演结果进行划分,计算 PR-Chla 与 LIMC-BPNN 在不同深度的反演误差。

表 6 不同算法的平均误差

Table 6 Average errors of different algorithms

Algorithm	RE / %	RMSE / (mg/m ³)	ME / (mg/m ³)	R
PR-Chla	56.73	0.616	0.331	0.724
LIMC-BPNN	22.51	0.253	0.118	0.904

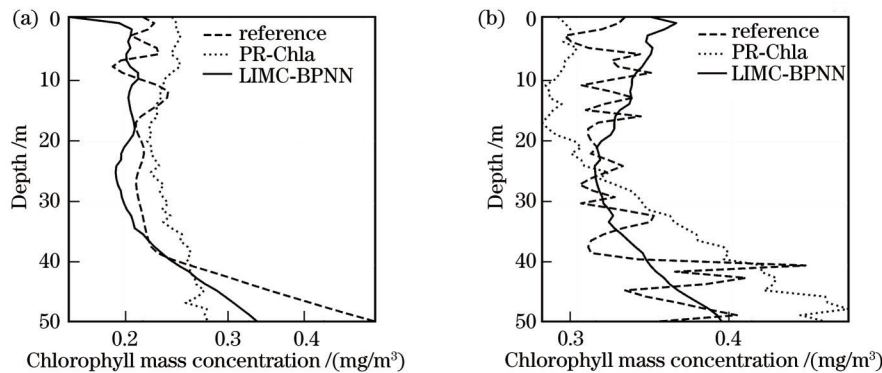


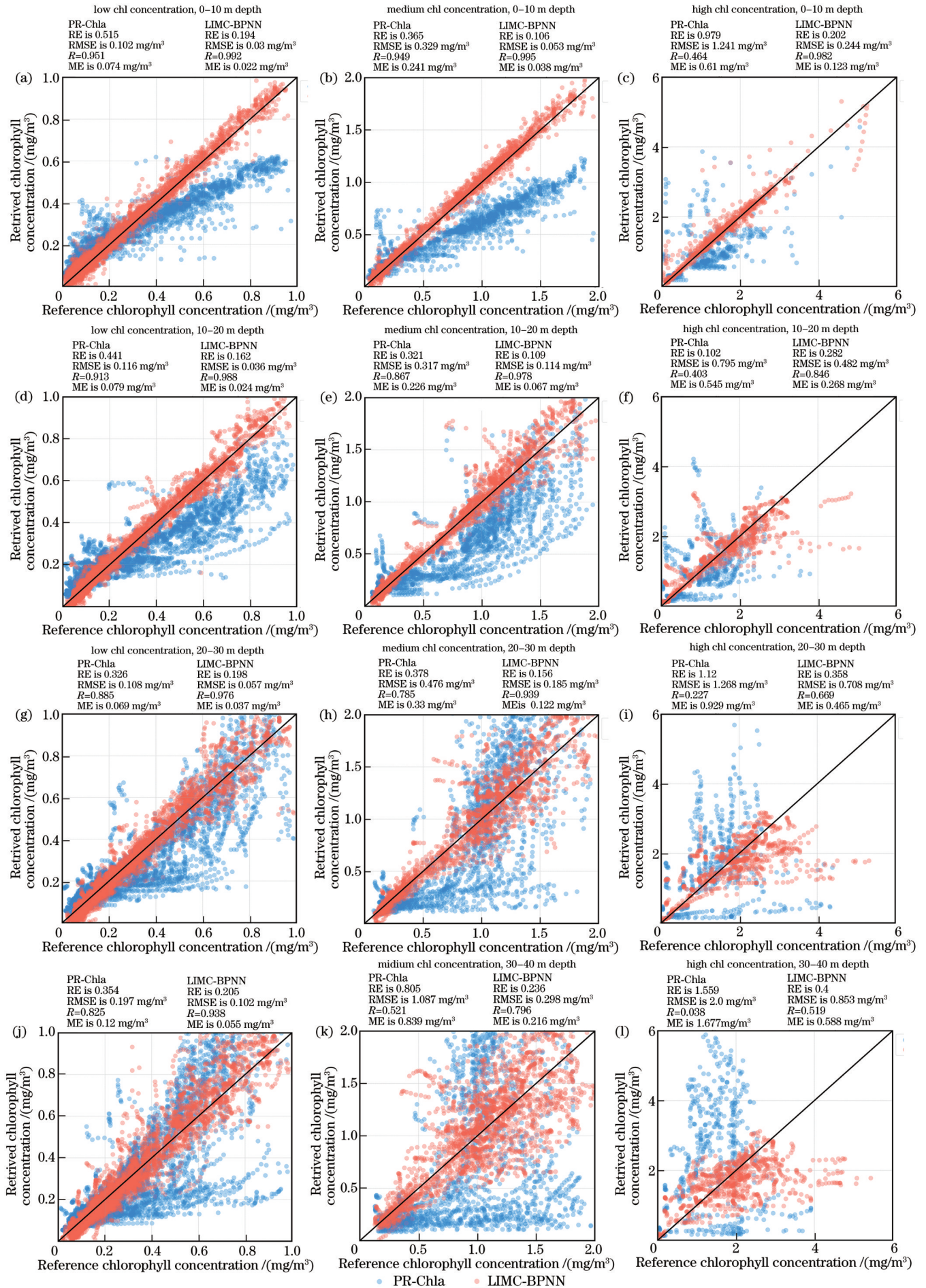
图 6 PR-Chla 法与 LIMC-BPNN 反演结果个例展示。(a)个例一;(b)个例二

Fig. 6 Cases displayed from PR-Chla and LIMC-BPNN. (a) Case1; (b) case2

数据集由激光雷达回波与叶绿素浓度组成,叶绿素浓度会造成激光散射,叶绿素浓度越高,散射的光子越多,因此不同范围的叶绿素浓度体现了样本多样性。为了研究各模型在不同样本的反演结果,本文将验证集按照叶绿素峰值浓度进行划分,设置低、中、高

三组,将浓度范围划分为 $0 \sim 1 \text{ mg/m}^3$ 、 $1 \sim 2 \text{ mg/m}^3$ 与大于 2 mg/m^3 ,在不同深度不同浓度上,将两个模型的反演结果与真值进行对比,并计算统计误差。

从图 7 中可以明显看出,LIMC-BPNN 在各个浓度范围上的相关系数均高于 PR-Chla,RE、RMSE 和



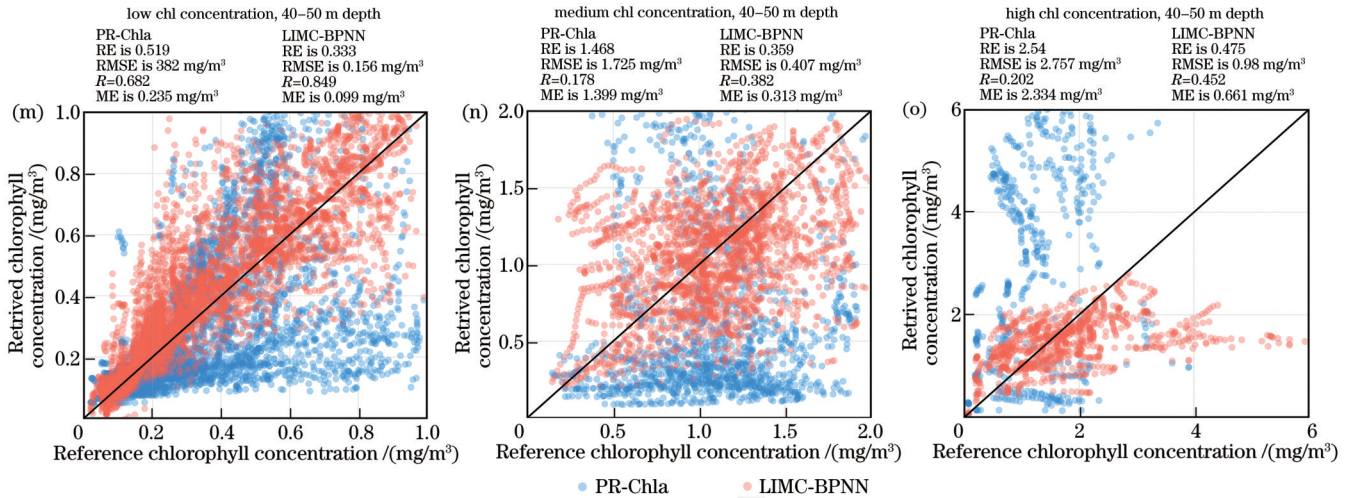


图 7 不同浓度范围、不同深度 LIMC-BPNN 与 PR-Chla 模型反演结果与数据集标签(叶绿素剖面)对比。(a)(d)(g)(j)(m)低浓度水体的反演结果;(b)(e)(h)(k)(n)中浓度水体的反演结果;(c)(f)(i)(l)(o)高浓度水体的反演结果

Fig. 7 Comparison of inversion results and dataset labels (chlorophyll profile) of LIMC-BPNN and PR Chla models with different concentration ranges and depths. (a) (d) (g) (j) (m) Inversion results of low mass concentration water; (b) (e) (h) (k) (n) inversion results of medium mass concentration water; (c) (f) (i) (l) (o) inversion results of high mass concentration water

ME 均低于 PR-Chla。低浓度水体的反演结果中, LIMC-BPNN 的三个反演误差较小, 因此在该浓度范围内, LIMC-BPNN 拥有较高的反演结果可信度与反演稳定性, 并且能够提供较小的反演差异。而 PR-Chla 的反演结果, 0~20 m 深度范围内, PR-Chla 的相关系数能保持在 0.9 以上, 在 20~40 m 范围内相关系数维持在 0.8 以上, 而 40~50 m 范围内相关系数约为 0.5, 因此 PR-Chla 在浅层水体的反演精度较高, 随着深度加深反演结果逐渐变差。中等浓度水体中: LIMC-BPNN 在 0~30 m 范围内的相关系数维持在 0.9 以上, 30~50 m 范围内的相关系数逐渐变大; RE、RMSE 和 ME 相比于低中浓度水体有一定程度的增大, 但是也基本维持稳定; 对于 PR-Chla, 反演结果与真值的相关系数呈逐渐下降趋势, 各误差也逐渐增大。高浓度水体反演结果中, LIMC-BPNN 在 0~10 m 的相关系数高于 0.9, 而下层反演精度随深度增加明显下降。PR-Chla 的相对系数在不同深度中均较低。综上所述, 对于不同叶绿素浓度范围的水体, LIMC-BPNN 的反演效果均好于 PR-Chla。但是在中高浓度水体中, LIMC-BPNN 的反演精度也出现了不同程度的下降, 其原因在于: 当叶绿素浓度较高时, 激光在浅层水体即发生大量散射, 导致深层水体的光子数量较少, 回波信号携带的叶绿素特征不足, 进而影响反演精度。

为了进一步观察不同模型反演误差随深度的变化趋势, 将上述分析中的不同浓度范围及验证集的平均 RE、RMSE、R 及 ME 在各深度上的结果进行展示, 如图 8 所示。平均结果如表 7 所示。对比不同浓度范围的反演结果, 两个模型在低浓度范围的反演结果均为最佳; 对比两个模型, LIMC-BPNN 的反演结果优于 PR-Chla; 对比两个模型反演误差随深度的变化趋势, LIMC-BPNN 的 RE、RMSE 和 ME 随深度增长较小, R

随深度下降程度也较少, PR-Chla 的 RE、RMSE 及 ME 随深度增长较大, R 随深度下降较为严重。综上所述, 相比于 PR-Chla, LIMC-BPNN 的反演结果可信度更高, 反演稳定性更强, 反演更贴近真实值, 并且数据差异小。然而, LIMC-BPNN 自身的反演精度也会随着深度增加而下降, 一方面激光在水下传输过程中的衰减导致信噪比降低, 信号包含的叶绿素特征较少, 影响反演结果。另一方面, 激光在水中的传输时, 叶绿素仅影响当前深度以下的激光传输, 所以网络当前节点应仅与深度下方节点相连。而 BP 神经网络层与层之间实现的是节点的全连接, LIMC-BPNN 的下层节点会连接上层节点, 并且由于激光在水中的衰减特性, 上层节点衰减, 值更大, 能够对下层节点产生更多的影响, 然而上层节点不应参与计算, 因此反演误差会随着深度增大而增大。

由于同站点的激光雷达信号与叶绿素浓度实测数据非常少, 本文基于三组实测数据, 对 LIMC-BPNN 的模型反演结果进行验证。图 9 为反演结果, 从图中能够看出, LIMC-BPNN 的反演结果与实测数据非常接近。表 8 为反演误差结果, RE、RMSE、R 与 ME 均较低, 表明 LIMC-BPNN 能够提供反演较高的精度。因此, 对于实测数据, LIMC-BPNN 也可以提供较高的反演精度。

6 结 论

本文通过 BGC-Argo 浮标获得 2015 年全月份、全球海域的叶绿素 a 浓度, 通过激光消光系数与叶绿素浓度的光学经验关系和蒙特卡罗仿真算法快速得到大量激光雷达回波信号数据, 并利用 RANSAC 对噪声进行剔除, 建立高质量的激光雷达回波数据集。

建立了基于 BP 神经网络的深度学习叶绿素反演

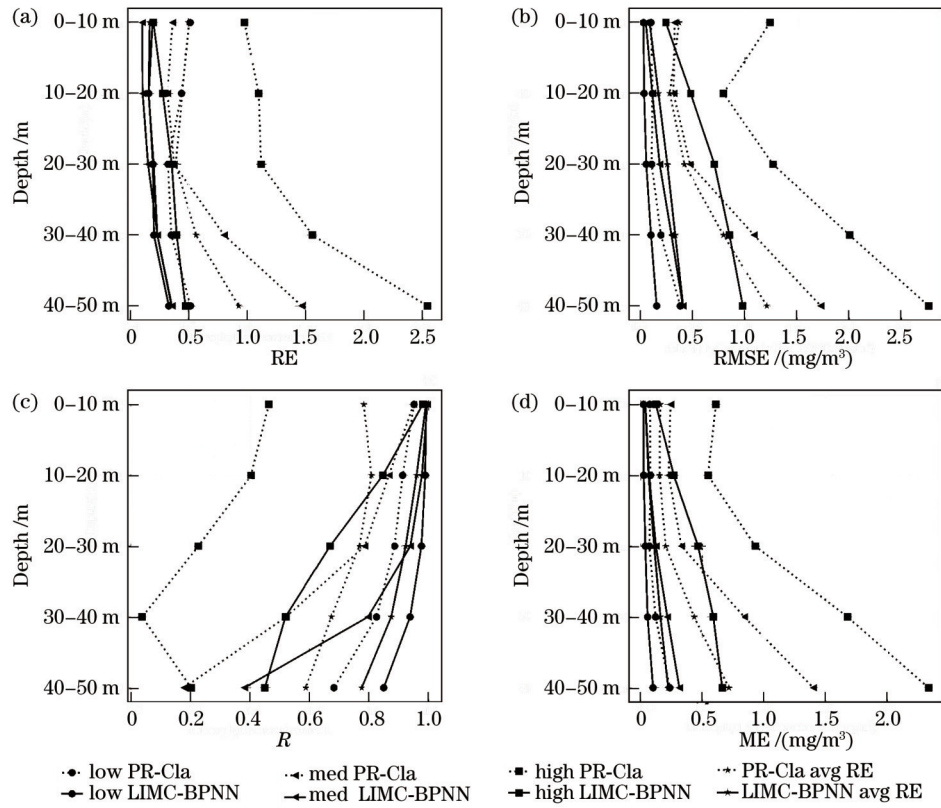


图 8 不同浓度范围及验证集整体的四个误差随深度变化结果。(a)RE 深度变化趋势;(b)RMSE 深度变化趋势;(c)R 深度变化趋势;(d)ME 深度变化趋势

Fig. 8 Errors of different conditions (depths, concentration ranges, algorithms). (a) RE; (b) RMSE; (c) R; (d) ME

表 7 两种模型不同深度平均误差对比

Table 7 Comparison of average errors between two models at different depths

Depth /m	RE /%		RMSE /(mg/m ³)		ME /(mg/m ³)		R	
	PR-Chla	LIMC-BPNN	PR-Chla	LIMC-BPNN	PR-Chla	LIMC-BPNN	PR-Chla	LIMC-BPNN
0-10	50.04	17.00	0.371	0.089	0.157	0.037	0.782	0.991
10-20	44.82	16.03	0.281	0.174	0.152	0.063	0.808	0.959
20-30	39.26	20.35	0.421	0.259	0.201	0.108	0.767	0.923
30-40	56.53	23.56	0.795	0.334	0.432	0.160	0.673	0.874
40-50	92.98	35.60	1.209	0.410	0.715	0.222	0.587	0.775

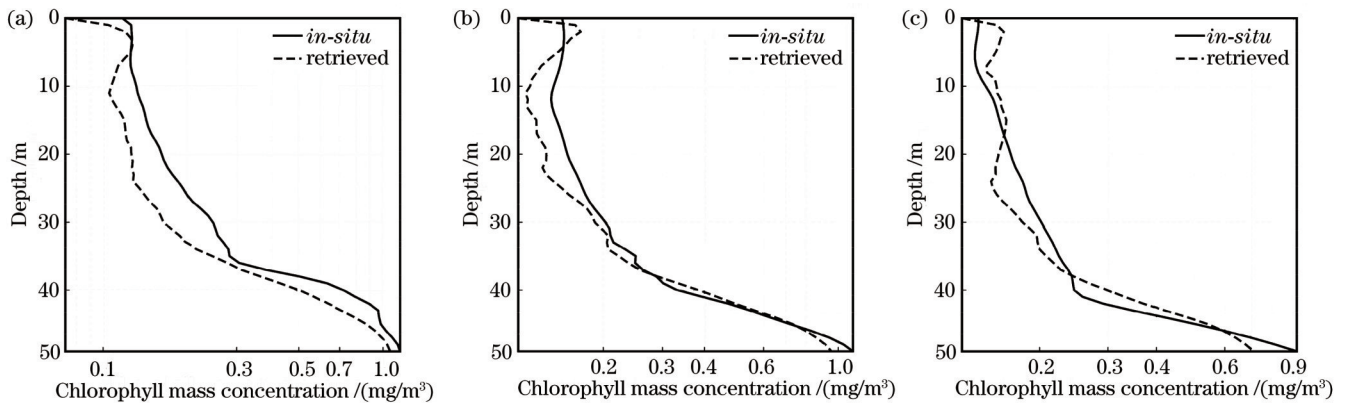


图 9 三个站点实测数据与反演结果对比。(a)B3 站点;(b)G1 站点;(c)G2 站点

Fig. 9 Comparisons between retrieved and *in-situ* in different sites. (a) B3 site; (b) G1 site; (c) G2 site

表 8 LIMC-BPNN 对于实测数据的反演结果
Table 8 Average errors of LIMC-BPNN retrieved to the measurement value in situ

Site	RE / %	RMSE / (mg/m ³)	ME / (mg/m ³)	R
B3	19.88	0.094	0.068	0.958
G1	10.24	0.037	0.026	0.988
G2	10.36	0.045	0.028	0.965

模型(LIMC-BPNN),并对模型进行训练,通过误差对反演结果进行量化,从而完成传统叶绿素反演算法与LIMC-BPNN反演能力的对比。实验结果表明,从验证集整体、不同深度范围及浓度范围的比较中,LIMC-BPNN的反演误差均小于PR-Chla,因此,LIMC-BPNN能够提供更高的反演精度。另外,如第2节计算方法所示,PR-Chla法需要先计算获得后向散射系数,在此过程需要使用拟合的计算方式,同时后向散射系数与叶绿素剖面浓度包含常数,此常数在全球范围的水体中运用。另外,PR-Chla法易受到多次散射的影响导致反演精度降低。LIMC-BPNN能在激光雷达回波与叶绿素浓度之间建立端到端的模型,进而解决了PR-Chla存在的问题。此外,LIMC-BPNN可以运用于实测数据,也可提供较好的反演结果。利用深度学习网络与激光雷达回波能够获得较为精确的叶绿素剖面浓度,为初级生产力、碳循环提供剖面研究数据,加深对全球生物化学循环和地球与气候的认识。

参 考 文 献

- [1] 马泳,林宏,冀航,等.基于机载激光雷达监测海洋赤潮模型研究[J].光子学报,2007,36(2):344-349.
Ma Y, Lin H, Ji H, et al. A scattering coefficient model for airborne lidar detection of red tide[J]. Acta Photonica Sinica, 2007, 36(2): 344-349.
- [2] 牟冰,崔廷伟,曹文熙,等.基于光学浮标的赤潮生消过程半分析监测方法[J].光学学报,2012,32(2):0201001.
Mu B, Cui T W, Cao W X, et al. A semi-analytical monitoring method during the process of red tide based on optical buoy[J]. Acta Optica Sinica, 2012, 32(2): 0201001.
- [3] 马泳,林宏,艾青,等.基于生物光学算法的海洋赤潮监测[J].光学学报,2008,28(1):7-11.
Ma Y, Lin H, Ai Q, et al. Red tide monitoring based on bio-optical algorithm[J]. Acta Optica Sinica, 2008, 28(1): 7-11.
- [4] 王晓彬,赵泓楷,周雨迪,等.偏振海洋激光雷达探测黄海水母特性[J].红外与激光工程,2021,50(6):20211038.
Wang X B, Zhao H K, Zhou Y D, et al. Characteristics of jellyfish in the Yellow Sea detected by polarized oceanic lidar[J]. Infrared and Laser Engineering, 2021, 50(6): 20211038.
- [5] 唐军武,朱培志,刘秉义,等.海洋剖面激光雷达探测中颗粒物偏振散射问题[J].光学学报,2022,42(12):1200001.
Tang J W, Zhu P Z, Liu B Y, et al. Polarization scattering of particles in lidar detection of ocean profile[J]. Acta Optica Sinica, 2022, 42(12): 1200001.
- [6] 吴航.双体制激光雷达探测与回波处理研究[D].北京:中国舰船研究院,2021.
Wu H. Research on detection and echo processing of dual-system lidar[D]. Beijing: China Ship Research and Development Academy, 2021.
- [7] 陈双,殷高方,赵南京,等.基于光合电子传递速率的浮游植物初级生产力测量[J].光学学报,2018,38(11):1126001.
Chen S, Yin G F, Zhao N J, et al. Measurement of primary productivity of phytoplankton based on photosynthetic electron transport rate[J]. Acta Optica Sinica, 2018, 38(11): 1126001.
- [8] 陈双,殷高方,赵南京,等.基于荧光动力学参数的浮游植物有效光合反应中心浓度分析[J].光学学报,2018,38(6):0630004.
Chen S, Yin G F, Zhao N J, et al. Analysis of phytoplankton concentration of functional reaction centers based on fluorescence dynamics parameters[J]. Acta Optica Sinica, 2018, 38(6): 0630004.
- [9] Collis R T H, Russell P B. Lidar measurement of particles and gases by elastic backscattering and differential absorption[M]// Hinkley E D. Laser monitoring of the atmosphere. Topics in applied physics. Berlin: Springer Berlin Heidelberg, 1976, 14: 71-151.
- [10] Klett J D. Stable analytical inversion solution for processing lidar returns[J]. Applied Optics, 1981, 20(2): 211-220.
- [11] Fernald F G. Analysis of atmospheric lidar observations: some comments[J]. Applied Optics, 1984, 23(5): 652-653.
- [12] Churnside J H, Marchbanks R D. Inversion of oceanographic profiling lidars by a perturbation to a linear regression[J]. Applied Optics, 2017, 56(18): 5228-5233.
- [13] 周雨迪.用于水体光学特性探测的海洋激光雷达研究[D].杭州:浙江大学,2020.
Zhou Y D. Research on marine lidar for detecting optical characteristics of water body[D]. Hangzhou: Zhejiang University, 2020.
- [14] McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity[J]. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.
- [15] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [16] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2014-12-22) [2023-03-02]. <https://arxiv.org/abs/1412.6980>.
- [17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12) [2023-03-02]. <https://arxiv.org/abs/1706.03762>.
- [18] 吴文涵,麻金继,孙二昌,等.基于深度学习的云参量反演方法研究[J].大气与环境光学学报,2022,17(4):453-464.
Wu W H, Ma J J, Sun E C, et al. Research on cloud parameter inversion method based on deep learning[J]. Journal of Atmospheric and Environmental Optics, 2022, 17(4): 453-464.
- [19] 刘林钰,张永军,李彦胜,等.基于深度学习的华东地区PM_{2.5}浓度遥感反演[J].环境科学,2020,41(4):1513-1519.
Liu L Y, Zhang Y J, Li Y S, et al. PM_{2.5} inversion using remote sensing data in Eastern China based on deep learning[J]. Environmental Science, 2020, 41(4): 1513-1519.
- [20] Morel A, Maritorena S. Bio-optical properties of oceanic waters: a reappraisal[J]. Journal of Geophysical Research: Oceans, 2001, 106(C4): 7163-7180.
- [21] James C, Johnathan H, Chris H, et al. Ocean backscatter profiling using high-spectral-resolution lidar and a perturbation retrieval[J]. Remote Sensing, 2018, 10(12): 2003.
- [22] Chen P, Jamet C, Zhang Z H, et al. Vertical distribution of subsurface phytoplankton layer in South China Sea using airborne lidar[J]. Remote Sensing of Environment, 2021, 263: 112567.
- [23] Petzold T J. Volume scattering functions for selected ocean waters[J]. Scripps Institution of Oceanography, 1972(3): 72-78.
- [24] Gordon H R, Morel A Y. Remote assessment of ocean color for interpretation of satellite visible imagery: a review[M]. New York: Springer US, 1983.
- [25] Rumelhart D E, Hinton G E, Williams R J. Learning

- representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [26] 邱国强, 王海黎, 邢小罡. BGC-Argo 浮标观测在海洋生物地球化学中的应用[J]. *厦门大学学报(自然科学版)*, 2018, 57(6): 827-840.
- Qiu G Q, Wang H L, Xing X G. Application of BGC-Argo floats observation to ocean biogeochemistry[J]. *Journal of Xiamen University (Natural Science)*, 2018, 57(6): 827-840.
- [27] Plass G N, Kattawar G W. Monte Carlo calculations of light scattering from clouds[J]. *Applied Optics*, 1968, 7(3): 415-419.
- [28] Poole L R, Venable D D, Campbell J W. Semianalytic Monte Carlo radiative transfer model for oceanographic lidar systems[J]. *Applied Optics*, 1981, 20(20): 3653-3656.
- [29] 李珂, 刘秉义, 杨倩, 等. 海洋激光雷达测量水体剖面偏振信号的仿真模拟[J]. *红外与激光工程*, 2021, 50(6): 20211035.
- Li K, Liu B Y, Yang Q, et al. Simulation of polarization profiles of water measured by oceanographic lidar[J]. *Infrared and Laser Engineering*, 2021, 50(6): 20211035.
- [30] 孔晓娟, 刘秉义, 杨倩, 等. 船载激光雷达测量水体光学参数的仿真模拟研究[J]. *红外与激光工程*, 2020, 49(2): 0205010.
- Kong X J, Liu B Y, Yang Q, et al. Simulation of water optical property measurement with shipborne lidar[J]. *Infrared and Laser Engineering*, 2020, 49(2): 0205010.
- [31] 刘航. 基于激光雷达回波的海洋光学参数反演研究[D]. 杭州: 浙江大学, 2020.
- Liu H. Research on ocean optical parameter inversion based on lidar echo[D]. Hangzhou: Zhejiang University, 2020.
- [32] 朱培志, 刘秉义, 孔晓娟, 等. 星载海洋激光雷达叶绿素剖面探测能力估算[J]. *红外与激光工程*, 2021, 50(2): 20200164.
- Zhu P Z, Liu B Y, Kong X J, et al. Study on the design and preparation technology of ultra-low profile wideband high reflection thin films[J]. *Infrared and Laser Engineering*, 2021, 50(2): 20200164.
- [33] Lee Z P. Visible-infrared remote sensing model and applications for ocean waters[D]. Tampa: University of South Florida, 1994.
- [34] Bricaud A, Morel A, Babin M, et al. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: analysis and implications for bio-optical models[J]. *Journal of Geophysical Research: Oceans*, 1998, 103(C13): 31033-31044.
- [35] Pope R M, Fry E S. Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements[J]. *Applied Optics*, 1997, 36(33): 8710-8723.
- [36] Morel A, Prieur L. Analysis of variations in ocean color[J]. *Limnology and Oceanography*, 1977, 22(4): 709-722.
- [37] Haltrin V I. Two-term Henyey-Greenstein light scattering phase function for seawater[C]//IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293), June 28-July 2, 1999, Hamburg, Germany. New York: IEEE Press, 2002: 1423-1425.
- [38] 崔晓宇, 陶雨婷, 刘群, 等. 采用半解析蒙特卡罗技术模拟星载海洋激光雷达回波信号的软件[J]. *红外与激光工程*, 2020, 49(2): 0203009.
- Cui X Y, Tao Y T, Liu Q, et al. Software to simulate spaceborne oceanic lidar returns using semianalytic Monte Carlo technique[J]. *Infrared and Laser Engineering*, 2020, 49(2): 85-91.
- [39] 程华. 激光雷达回波信号处理技术研究[D]. 成都: 中国科学院光电技术研究所, 2015.
- Cheng H. Study on the signal processing of lidar[D]. Chengdu: Institute of Optics and Electronics, Chinese Academy of Sciences, 2015.
- [40] 郭凯. 蒙特卡罗方法模拟水体光学特性技术研究[D]. 天津: 国家海洋技术中心, 2015.
- Guo K. Study on Monte Carlo method for simulating optical characteristics of water body[D]. Tianjin: National Ocean Technology Center, 2015.
- [41] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395.

Chlorophyll Profile Retrieval Algorithm Based on Oceanographic Lidar and BP Neural Network

Tie Ning, Liu Bingyi*

College of Marine Technology, Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, Shandong, China

Abstract

Objective Currents, sea waves, and climate changes are generated by air-sea interaction. Oceans cover more than 70 percent of the earth and play a significant role in the ecological environment. Therefore, various countries are researching oceans. A wide range of substances are present in oceans, of which phytoplankton are important primary producers in the marine ecosystem and are linked to a variety of oceanic processes. Chlorophyll a is an indicator to characterize the phytoplankton amount and plays an indispensable part in ocean research. Meanwhile, bio-optical parameters can be employed in various fields of oceanographic research and contribute to the rapid development of marine research. Chlorophyll concentration is an important prerequisite for the inversion of bio-optical parameters, and it directly affects the results of the bio-optical parameters. Active remote sensors with outstanding advantages have become one of the most rapidly growing and most effective remote sensors in recent years. Since active remote sensing technology does not depend on solar rays, it can obtain profile information with few detection limitations. As active remote sensing, oceanographic lidar can be mounted on a variety of platforms and obtain the profile concentration of chlorophyll. However, the traditional

methods of inverting chlorophyll from lidar signals have poor accuracy, because they are susceptible to multiple scattering. Therefore, high-precision chlorophyll inversion algorithms are essential for marine research. Since the echo signal and chlorophyll concentration have a complex nonlinear relationship, deep learning can be adopted to filter multiple scattering noises, extract the backward scattering signal features, and build a high-precision chlorophyll inversion model.

Methods Four steps are conducted as follows. Firstly, a dataset is built with two parts of label and feature. The label consisting of chlorophyll concentration profiles comes from BGC-Argo and the chlorophyll optical parameters are calculated by empirical relations. The lidar echo signals are simulated by a semi-analytic Monte Carlo algorithm and random sample consensus (RANSAC) algorithm is utilized to distinguish noises. Secondly, a network structure is constructed by Python. We build a lidar inversion model for chlorophyll based on backward propagation neural network (LIMC-BPNN) to solve the problem of multiple scattering effects degrading the accuracy. During the training, ReLU (linear rectification function) is adopted as the activation function, Adam (adaptive moment estimation) as the optimizer, and the epoch is 32. Python is an implementation language. Thirdly, chlorophyll concentration by PR-Chla is calculated to conduct a comparison between the two models. The perturbation retrieval (PR) proposed by Churnside can compute the lidar backscatter coefficient. Finally, relative error (RE), root mean square error (RMSE), correlation coefficient (R), and mean error (ME) are leveraged to quantify the results. The models are evaluated separately through three perspectives, including the average validation set, the validation set divided by water depth, and the set divided by chlorophyll concentration.

Results and Discussions First, a network structure of LIMC-BPNN is built to extract lidar echo features (Fig. 2), and its parameters are determined by experiments. Next, a feature of the dataset covers the five oceans, which is around twenty thousand. The label is created from the dataset by empirical formulations of chlorophyll optics and a semi-analytic Monte Carlo (Table 3). The data in Table 3 exhibit lidar echoes containing chlorophyll information (Fig. 4), and then a comparison before and after noise rejection is shown using RANSAC (Fig. 5). After training, the average errors of the validation set are shown (Table 5). Additionally, two cases are presented (Fig. 6), the results in various chlorophyll concentrations and depths are demonstrated (Fig. 7), and the error variation at different depths (Fig. 8) is discussed.

Conclusions The results of semi-analytic Monte Carlo can bring chlorophyll features, and RANSAC can filter outliers to enhance the dataset quality. In the ME of the validation dataset, LIMC-BPNN declines 34.22%, 0.363, and 0.213 in relative error, root mean square error, and mean error. The correlation coefficient is increased by 0.18, which indicates better credibility and stability of LIMC-BPNN to provide smaller data variances. Meanwhile, the error of LIMC-BPNN is lower than that of PR-Chla at different depths, which verifies the above findings. Additionally, in low concentration ranges, three errors of LIMC-BPNN are small. In 0–20 m, the traditional PR method performs well, but in 20–50 m RMSE and ME gradually grow larger. In medium and high concentration ranges, RE, RMSE, and ME are greater than those in low concentration ranges, with unchanged stability. Nevertheless, the PR-Chla is stable in 0–10 m and its error increased rapidly below 10 m. In conclusion, LIMC-BPNN is better than the PR-Chla for chlorophyll concentration. However, as the depth increases, errors of the two models are accumulated to demonstrate that the attenuation characteristics of the laser in water affect the accuracy.

Key words ocean optics; back propagation neural network; vertical profile; water optical parameters; chlorophyll a