doi:10.6041/j.issn.1000-1298.2025.01.007

基于多源数据与丰度信息融合的森林生物量估算研究

林洁雯[1,2]　陈　建[1]

(1. 中国农业大学工学院, 北京 100083; 2. 中国科学院沈阳自动化研究所机器人学国家重点实验室, 沈阳 110016)

摘要: 森林是维持碳平衡的重要组成部分,精确的森林生物量探测对环境改善和相关政策制定均有重要的推动作用。本文探索了将多源数据及丰度信息融合分析实现森林生物量反演。首先,采用 MOPSOSCD 获取研究区域的端元束,并获得每组树木端元的丰度信息,然后在 Landsat 8 OLI 及 ASTGTM DEM 中提取单波段因子、植被指数、地形因子、纹理特征等 46 个指标,测试融合丰度前后模型拟合效果。通过多元线性回归和 BP 神经网络模型进行生物量反演试验发现,采用多元线性回归模型时,优化前生物量均方根误差(RMSE)和决定系数(R^2)分别为 41.09 mg/hm^2、0.40,优化后最优 RMSE 和 R^2 分别为 38.66 mg/hm^2、0.44。采用 BP 神经网络模型时,优化前生物量 RMSE 和 R^2 分别为 32.73 mg/hm^2、0.56,最优 RMSE 和 R^2 分别为 32.07 mg/hm^2、0.57。添加丰度后 BP 神经网络模型具有最优反演效果。通过试验验证了 MOPSOSCD 算法提取端元束对应的丰度在提升模型生物量反演精度的有效性。同时,试验证明端元的提取精度越高,对应模型生物量反演效果越好。

关键词: 多源数据融合; 森林生物量; 端元束; 混合像元分解; 反演

中图分类号: TP751.1; S771.8　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)01-0065-09　　**OSID**:

Forest Biomass Estimation Based on Multi-source Data and Abundance Information Fusion

LIN Jiewen[1,2]　CHEN Jian[1]

(1. *College of Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *State Key Laboratory of Robotics*, *Shenyang Institute of Automation*, *Chinese Academy of Sciences*, *Shenyang* 110016, *China*)

Abstract: Forests are essential in maintaining carbon balance, and accurate detection of forest biomass plays a crucial role in promoting environmental improvement and related policy formulation. The comprehensive analysis of multi-source data and abundance information was explored to achieve forest biomass inversion. Firstly, MOPSOSCD was used to obtain the endmember bundle of the study area, and the abundance information of each group of tree endmembers was obtained. Totally 46 indicators, including single band factor, vegetation index, terrain factor, texture feature, etc., were extracted from Landsat 8 OLI and ASTGTM DEM to test the model fitting effect before and after fusion abundance through biomass inversion experiments using multiple linear regression and BP neural network models. It was found that when using the multiple linear regression model, among the seven optimized models extracted for abundance, two groups had better RMSE than the multiple linear regression model without added abundance, and seven groups had better R^2 than the multiple linear regression model without added abundance. The RMSE and R^2 before optimization were 41.09 mg/hm^2 and 0.40, and the optimal RMSE and R^2 were 38.66 mg/hm^2 and 0.44, respectively. When using the BP neural network model, all BP models with added abundance showed an improvement in RMSE, with six groups having better R^2 than those without added abundance. The RMSE and R^2 before optimization were 32.73 mg/hm^2 and 0.56, and the optimal RMSE and R^2 were 32.07 mg/hm^2 and 0.57, respectively. The BP neural network model with added abundance had the best inversion effect. The MOPSOCCD algorithm was used to extract the

收稿日期: 2024-11-01　修回日期: 2024-11-29

基金项目: 国家自然科学基金项目(52472463)、国家重点研发计划项目(2022YFD2001405)和机器人学国家重点实验室开放基金项目(2024-001)

作者简介: 林洁雯(1995—),女,博士生,主要从事遥感信息处理研究,E-mail: b20203070536@cau.edu.cn

通信作者: 陈建(1983—),男,教授,博士生导师,主要从事无人系统具身智能研究,E-mail: jchen@cau.edu.cn

abundance corresponding to the endmember bundle, and abundance was used as the inversion factor for biomass inversion. The experiment demonstrated the effectiveness of abundance in improving biomass inversion accuracy. Meanwhile, the higher the accuracy of extracting endmembers was, the better the corresponding model biomass inversion effect was.

Key words: multi-source data fusion; forest biomass; endmember bundle; mixed pixel unmixing; inversion

0 引言

森林是重要的碳池,碳平衡森林地上生物量(AGB)可用于测量森林碳储量,准确量化森林 AGB 的空间格局和规模,对于理解从区域到全球尺度的碳循环过程[1]、管理生物多样性以及制定碳抵消和缓解战略至关重要[2]。有效进行生物量估算可以推断森林退化的排放量,还可以为国家制定相关政策提供支持[3]。

传统生物量通过砍伐或标注木法等直接测定。但是耗时费力,且只能获得有限的地面点信息[4]。遥感技术能够准确、快速地获取各个尺度的“面”上连续分布的森林信息,避免消耗大量的经济成本[5-6]。通常光学遥感主要通过提取纹理信息和植被指数,通过回归分析[7]、随机森林、回归向量机[8]、模糊聚类[9]等方法实现模型拟合。多源数据融合具有包容性,其优势在于可以充分利用数据的互补性,提高地物分类[10]和地貌信息提取[11-12]的准确性和可靠性,如:光谱信息和高程信息融合。融合后的数据可以更充分地描述地表特征,可以得到更全面、更准确的地物特征,为生物量反演提供更加精准的信息。LI 等[13]采用 PALSAR-2(ALOS-2)和 Sentinel-2 图像中提取的 48 个变量,输入随机森林回归模型实现森林生物量估算,证明了多传感器信息可以提高森林生物量模型稳健性。与此同时,光学传感器与 SAR[14]、SAR 与 LiDAR 的融合[15]也被广泛应用于森林生物量反演研究中。在多源数据融合方面,YANG 等[16]融合冠层高度、光谱指数、地形数据、气候数据等信息估计了分辨率 30 m 下的中国森林 AGB 模型,该高分辨率森林 ABG 地图可作为重要的替代数据来源。

混合像元分解[17]技术常被应用于地表覆盖类型[18-19]和信息提取[20]、作物种类识别及矿产勘探等领域[21]。端元及丰度信息对于一些指标的反演具有一定的潜力,通常作为反演指标或者构建新的参数被用于反演中[22]。于此同时,丰度信息通常被作为加权用在产量预测、反演及关键参数提取中[23-24]。丰度指标作为加权或构建新的指标具有较大的潜力,少部分研究将混合像元融入森林生物量反演中[25-26]。将丰度信息融合至森林生物量反演的必要性有:①无论是机载传感器还是星载传感器,其成像受到分辨率的限制,不可避免地产生混合像元现象。该种情况下,从像元中提取的信息代表了混合像元的所有信息,每个反映出的信息不准确,造成生物量反演结果不准确。②森林生长环境、形态各异,通常与多种地物类型共处以维持生态的稳定性,若需要以森林单一地物为研究对象提取其生物量信息,需要更精细的丰度信息。在现有研究中,一般均采用丰度信息与 VI 指数加权或融合,很少作为独立的指标探索与反演信息及模型的关系。

光谱影像的光谱变异性普遍存在,现有研究中缺乏探讨同物异谱情况下丰度对反演模型的作用。本文首次探索将星载传感器数据和端元束所对应的丰度信息进行融合实现森林生物量提取。首先,从 Landsat 8 OLI 数据和 ASTGTM DEM 数据中提取单波段因子、植被指数、地形因子及纹理特征等 46 个指标,并获取主要的相关性指标。其次,考虑到丰度信息的可用性及光谱变异性,采用 MOPSOSCD 提取的端元束构建的新丰度信息作为新反演因子融合至生物量反演模型中,并在多元线性回归模型和 BP 神经网络模型中验证丰度信息对于森林生物量反演的有效性。

1 实验区与数据预处理

选择北京市房山区四海镇为研究区域,位于 116°22′~116°25′E、40°30′~40°33′N,属于热带季风气候,夏季湿润多雨,冬季降雨较少,温度较低,年平均气温为 -2.8℃,年平均降水量为 93.0 mm。研究区域包含大量的植被覆盖,同时包含土地、屋顶等典型地物类型。采用 Landsat 8 OLI 陆地资源卫星数据获取上述研究区域多光谱数据(图 1),Landsat 8 OLI 陆地成像仪包括 9 个波段,空间分辨率为 30 m。将图 1 数据作为采样点及模型初步训练及预测数据。同时,为了验证方法有效性,在北京区域(LC81230322019245LGN00、LC81230332019245LGN00、LC81240322019236LGN00)随机选择与该地貌和森林分布相似的数据,根据研究区域行政区划、时间和含云量进行遥感影像的下载,尽量保证少云分布。同时,下载该研究区域的 ASTGTM DEM 数据,ASTGTM DEM 数据采用 UTM/

WGS94 投影，数据值域范围为 -152 ~8 806 m，空间分辨率为 30 m。

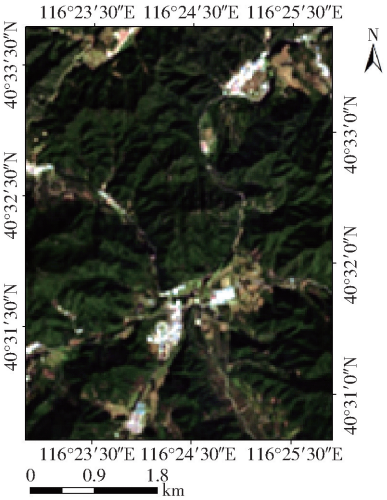

图 1　研究区域

Fig. 1　Research regional

为了验证方法的有效性，以两个数据作为辅助数据：①2020 年全国植被覆盖类型 GLC_FCS30 数据[27]，将该数据用于查询研究区域内森林分布信息，排除农田、草本、不透水面、水体等类别外，北京地区包括封闭落叶阔叶林、封闭常绿针叶林、封闭落叶针叶林 3 种类型林，因此，进行森林生物量反演模型的训练主要考虑针叶林和阔叶林两种类型的数据。② 2019 年中国生物量制图数据[16]，2019 年中国生物量数据制图如图 2 所示，该数据收集 8 000 多份地面数据，该数据 AGB 测绘结果与独立的地面清查评估数据集均方根误差为 0.75 mg/hm^2，作为本文验证方法采样数据和反演对比支撑。

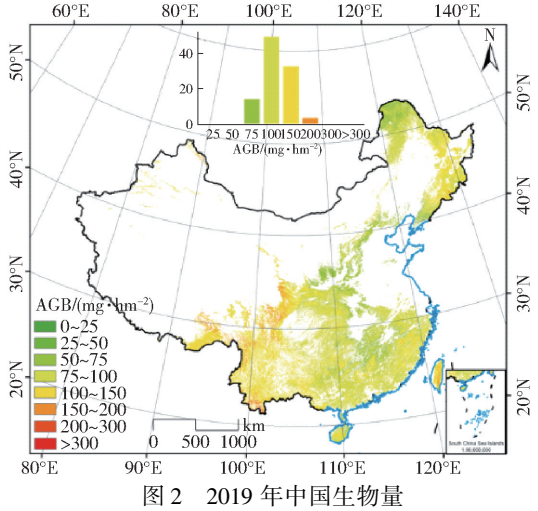

图 2　2019 年中国生物量

Fig. 2　2019 China biomass

上述数据均为空间分辨率 30 m 的数据。由于端元提取及生物量反演均依赖于数据的稳定性，为避免误差产生，对多源数据进行预处理。采用 ENVI 5.6 软件对 Landsat 8 OLI 进行辐射定标消除传感器本身误差，然后进行 FLAASH 大气校正消除大气造成的辐射误差，从而反演地物真实的表面反射率。为克服不同数据源的空间偏差，文中模型训练和测试的区域及样点均基于几何校正和空间配准后的数据进行提取。

2　多源数据和丰度信息融合的生物量反演模型

本文通过融合多光谱、DEM 和丰度信息，提出基于多源数据和丰度信息融合的森林生物量估算研究，具体路线如图 3 所示。首先，对 ROI 区域数据进行预处理。然后，针对预处理后的数据提取端元束，并得到每组树木端元的丰度信息；根据多源数据提取波段因子、植被指数、地形因子及纹理特征 4 类综合参数，并进行相关性分析。最后结合强相关性因子和丰度信息，在多元线性回归模型及 BP 神经网络模型中进行生物量反演测试，并对比丰度在生物量反演模型中的作用。

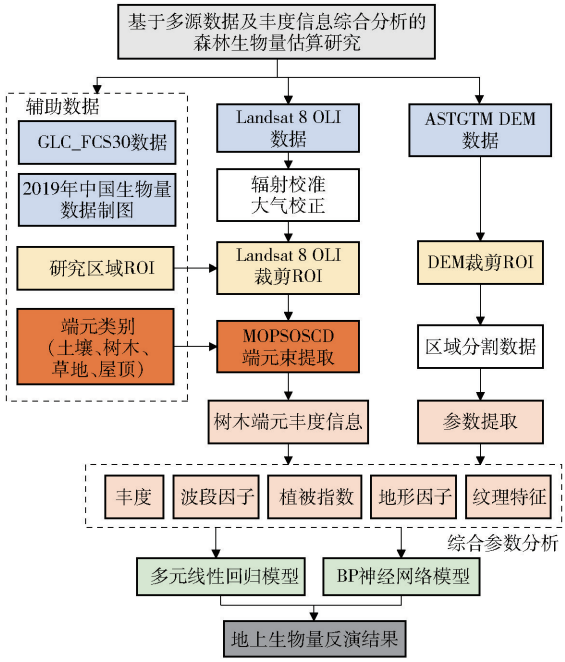

图 3　技术路线图

Fig. 3　Technical roadmap

2.1　混合像元分解模型

受到传感器分辨率的影响，影像中大部分单一像元内包含不同的地物覆盖，形成混合像元现象。通常反演因子通过像元进行提取，而像元内不仅包含了一种地物（混合像元），每一个像元所提取的反演因子是多个混合后的像元结果。因此，在进行森林生物量反演过程中需要提取出地物的光谱特征，并获取每种成分的比例。

混合像元分解分为端元提取和丰度估计。线性光谱混合模型[28]（Linear spectral mixing model，

LSMM)是太阳入射辐射只与一种地物表面发生作用,且地物与地物之间没有相互作用的模型假设。对于包含 N 个像元的 l 个波段的高光谱影像 $R=\{r_i\}(i=1,2,\cdots,N)$,LSMM 问题通常描述为

$$r_i=\sum_{j=1}^{q}e_j a_{ij}+\varepsilon_j \tag{1}$$

式中 e_j——端元

a_{ij}——端元在像元中所占比例,即丰度

ε_j——误差项

q——端元数量

由于 a_{ij} 的物理意义是端元在像元中的面积比,需满足“非负约束”以及“和为 1 约束”2 个条件,即

$$a_{ij}\geqslant 0 \tag{2}$$

$$\sum_{j=1}^{q}a_{ij}=1 \tag{3}$$

根据 LSMM 假设,方程数量 L 大于未知数的数量 m,属于超定方程组,可以找到一个 $\boldsymbol{a}_i$ 使得 $\boldsymbol{E}\boldsymbol{a}_i^{\mathrm{T}}$ 与 r_i 的差距尽可能小,这个差距通常用向量 2 范数的平方表示,因此,求解该方程组可以转换为优化问题

$$\begin{cases}\min\limits_{a_i}(\boldsymbol{E}\boldsymbol{a}_i^{\mathrm{T}}-r_{2i}^2)\\ \text{s.t. }\ \boldsymbol{a}_i\in\mathbf{R}^m\end{cases} \tag{4}$$

采用 MOPSOSCD[30] 算法进行端元束提取,并采用全约束最小二乘法进行丰度提取,解混流程如图 4 所示。

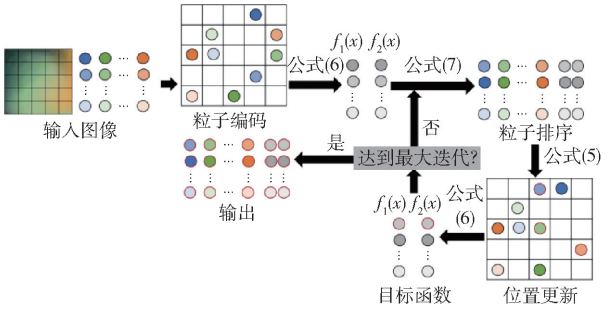

图 4 MOPSOSCD 算法流程图

Fig. 4 MOPSOSCD algorithm flowchart

MOPSOSCD 算法考虑到光谱变异性的影响,通过智能优化算法迭代获取最能够表达不同地物的光谱信息。对于任一像素点编号集合 $\boldsymbol{Z}$ 进行种群进化时,存在某组可行解 $X=\{x_1,x_2,\cdots,x_q\}(x_i\in\boldsymbol{Z})$,则对应一组端元为 $\boldsymbol{E}=(e_1,e_2,\cdots,e_q)^{\mathrm{T}}$,通过方程实现可行解的位置更新。

$$x_i(t+1)=\begin{cases}\mathrm{ceil}(x_i(t)+v_i(t+1)) & (i\notin\mathrm{Range}(\boldsymbol{\beta}))\\ \mathrm{ceil}(x_r(t+1)) & (i\in\mathrm{Range}(\boldsymbol{\beta}))\end{cases} \tag{5}$$

式中 ceil( )——取整函数

$x_i(t)$——粒子 i 在第 t 代的位置更新

$\boldsymbol{\beta}$——粒子随机更新矩阵

Range(·)——矩阵元素

$x_r(t+1)$——区间 Z 内的随机整数

$v_i(t+1)$——粒子的更新速度

更新后的位置通过无约束最小二乘法和全约束最小二乘法进行解丰度求解,再将端元 e_j 与丰度估计 $\hat{a}_{ij}$ 重新合成返混影像 $\hat{R}$,求解均方根误差 E_{RMSE} 作为目标函数。

$$E_{\mathrm{RMSE}}=\sqrt{\frac{\sum_{i=1}^{N}(\hat{R}-R)^2}{N}} \tag{6}$$

对于迭代过程中产生的粒子空间,通过拥挤距离 S_i 进行个体优劣排序。

$$S_i=\begin{cases}\max(C_{i,x},C_{i,f(x)}) & (C_{i,x}>C_{avg,x}\ \text{or}\ C_{i,f(x)}>C_{avg,f(x)})\\ \min(C_{i,x},C_{i,f(x)}) & (C_{i,x}\leqslant C_{avg,x}\ \text{or}\ C_{i,f(x)}\leqslant C_{avg,f(x)})\end{cases} \tag{7}$$

式中 $C_{i,x}$——决策空间距离

$C_{avg,x}$——决策距离平均值

$C_{i,f(x)}$——目标空间距离

$C_{avg,f(x)}$——目标距离平均值

通过目标空间与决策空间的综合排序可在迭代完成后获得有效的帕累托最优解集。因此,采用端元束提取结果,并通过全约束最小二乘法获得每组树木端元丰度信息。

2.2 模型建立与评价指标

为了验证端元束所提取的丰度信息在生物量反演的作用,分别选择一种典型的线性模型和非线性模型进行生物量反演:多元线性回归(Multivariate linear regression, MLR)模型和 BP 神经网络[30](Back propagation, BP)模型。

2.2.1 多元线性回归模型

线性关系指主要影响因素作为自变量解释因变量的变化过程,在地上生物量估算中,线性模型最具代表性的为多元线性回归方法。

2.2.2 BP 神经网络模型

通过多个指标综合评价生物量是一个复杂的问题,可以将该问题转换为一个多输入单输出的问题。BP 神经网络根据观测值与预测值进行误差分析,进而修改权值和阈值。本文构建隐藏层为单层的 BP 神经网络,输入为丰度信息和具有强相关性的反演因子,输出为植被地上生物量预测值。设有 d 个输入神经元、l 个输出神经元、q_1 个隐藏层神经元,给定训练集 $D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_{m1},y_{m1})\}$,则第 h 个隐藏层神经元输入为

$$\partial_h=\sum_{i=1}^{d}v_{ih}x_i \tag{8}$$

式中　v_{ih}——输入层第 i 个神经元与隐藏层第 h 个神经元之间的连接权重

第 j 个输出神经元的输入表示为

$$\beta_j = \sum_{h=1}^{q_1} w_{hj} b_h \tag{9}$$

式中　w_{hj}——隐藏层连接权重

b_h——隐藏层神经元输出

构建的 BP 神经网络采用 Sigmoid 函数。神经网络的输出为

$$\hat{y}_j^k = f(\beta_j - \theta_j) \tag{10}$$

式中　θ_j——输出层第 j 个神经元的阈值

2.2.3　评价指标

采用均方根误差(RMSE)及决定系数 R^2 作为线性模型和非线性模型评价指标。

3　试验结果与分析

以北京市房山区四海镇 ROI 区域的森林作为研究对象,探究该区域的生物量估测结果。为了拟合模型并测试模型的有效性,首先对 ROI 进行丰度提取,并获取 ROI 区域内的 4 类反演因子,然后在 ROI 内随机选择 100 个抽样点以比例 8∶2 划分为训练样本和测试样本,用于选择最优的丰度信息。最后采用最优丰度进行反演模型对比,获取最优反演模型。

3.1　端元束及丰度提取结果

本文使用 Matlab 2020b 平台部署 MOPSOSCD 算法进行端元束提取,并从每组端元组合中提取丰度信息,代码位于:https://github.com/Rosalindjiewen/MOPSOSCD/tree/master。研究区域主要包含森林、草地、土壤及屋顶 4 种主要地物,设置 MOPSOSCD 算法的端元种类为 4。同时,定向选择概率 $p_m = 0.2$、粒子数量 $P = 20$、迭代次数 $M = 100$。通过算法迭代共获取 7 组端元组合构成端元束。本文将所提取的端元束中的端元组合命名为 EM_1 ~ EM_7。采用全约束最小二乘法解丰度并反混后与原图的 E_{RMSE} 分别为:0.09、0.06、0.07、0.06、0.06、0.05、0.06。对应树木丰度如图 5 所示,将树木端元通过全约束最小二乘法解丰度后的丰度信息命名为 EM_abu1 ~ EM_abu7。

通过 E_{RMSE}结果可发现,所提取的 7 组端元组合的 E_{RMSE} 差异不大,均为有效值。其中 EM_6 的 E_{RMSE}处于最优的状态,其次是 EM_2 和 EM_4。结合图 5 树木丰度的可视化结果可发现 EM_abu6 和 EM_abu2 效果更好,能够将 ROI 区域内东部及中部的非植被部分和植被部分有效区分,且整体的植被丰度值较高。将 7 组树木丰度作为反演因子添加至反演模型中,进一步探讨添加丰度信息对生物量反演模型的影响。

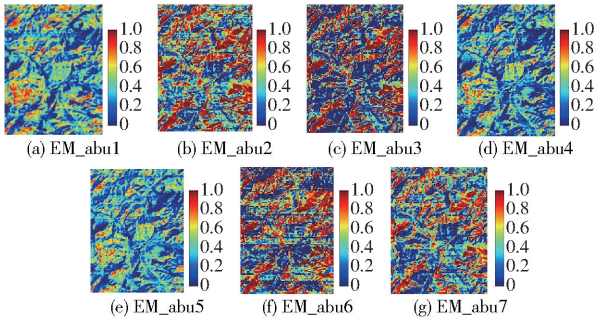

图 5　各组端元的丰度可视化结果

Fig. 5　Abundance results of each end member group

3.2　反演因子提取及 Pearson 相关性分析结果

通过 Landsat－8 OLI 影像及 ASTGTM DEM 模型提取 4 类反演因子[31－32],如表 1 所示。主要包括 B2 ~ B7 内的 6 个波段因子、13 个植被指数、高程及相关坡度和坡向地形因子、3 种窗口构建的纹理特征(3×3/5×5/7×7),共 46 个。

表 1　反演因子

Tab. 1　Inversion factors

因子类别	反演因子	数量
波段因子	B2 ~ B7(B_2、B_3、B_4、B_5、B_6、B_7)	6
植被指数	NDVI(V_{NDVI})、DVI(V_{DVI})、RVI(V_{RVI})、ARVI(V_{ARVI})、EEVI2(V_{EEVI2})、ENDVI(V_{ENDVI})、EVI(V_{EVI})、GDVI(V_{GDVI})、GNDVI(V_{GNDVI})、GRVI(V_{GRVI})、NDVI43(V_{NDVI43})、OSAVI(V_{OSAVI})、SAVI(V_{SAVI})	13
地形因子	高程(DEM)、坡度(D_{slope})、坡向(D_{aspect})	3
纹理特征	均值(Me)、方差(Var)、对比度(Con)、信息熵(En)、协同性(Homo)、相异性(Dis)、二阶矩(Sec)、相关性(Cor)	24

对划分为训练集的 80 个抽样点的建模因子进行 Pearson 相关性分析,获得 10 个强相关性反演因子的相关性系数,如表 2 所示。采用 2019 年生物量数据制图[26]提取生物量样点进行试验,因此在高程和坡度方面具有较强的相关性,其次是 NDVI、ENDVI、EEVI2、RVI、ARVI、GNDVI 等相关的生理生化指标,在纹理特征方面,5×5 和 7×7 的相关性(Cor)指标相关性较高。

表 2　Pearson 分析结果

Tab. 2　Pearson analysis results

序号	1	2	3	4	5
反演因子	DEM	D_{slope}	V_{NDVI}	V_{ENDVI}	V_{EEVI2}
相关性系数	0.53	0.52	0.47	0.47	0.47
序号	6	7	8	9	10
反演因子	V_{RVI}	V_{ARVI}	$Cor_{7\times7}$	V_{GNDVI}	$Cor_{5\times5}$
相关性系数	0.47	0.44	0.35	0.35	0.31

3.3 多元线性回归模型测试结果

为了验证所提取的 7 组树木端元对于多元线性回归模型的影响,本文首先对在多元线性回归模型中设置未添加丰度及添加丰度的反演模型试验。未添加丰度的多元线性回归模型采用表 2 中 10 项高相关性的因子进行生物量反演。添加丰度的多元线性回归模型在采用表 2 中 10 项高相关性因子的基础上,将树木端元采用全约束最小二乘法解丰度的值单独作为一项因子添加到模型中,即 11 项因子。采用划分为训练集的 80 个样点构建多元线性回归模型,未添加丰度的多元线性回归模型以及 7 组添加丰度的多元线性回归模型为:

MLR

$$y = 168.70 + 0.01DEM + 0.19D_{\text{slope}} - 384\,251.05V_{NDVI} + 383\,641.56V_{ENDVI} + 265.52V_{EEVI2} + 8.50V_{RVI} + 116.30V_{ARVI} + 4.22Cor_{7\times7} + 6.35V_{GNDVI} + 0.28Cor_{5\times5}$$

MLR + EM_abu1

$$y = 164.17 + 0.01DEM + 0.18D_{\text{slope}} - 441\,624.36V_{NDVI} + 441\,021.87V_{ENDVI} + 257.83V_{EEVI2} + 8.54V_{RVI} + 111.62V_{ARVI} + 4.38Cor_{7\times7} + 20.96V_{GNDVI} + 0.29Cor_{5\times5} - 0.98abu$$

MLR + EM_abu2

$$y = 116.81 + 0.01DEM + 0.17D_{\text{slope}} - 471\,161.83V_{NDVI} + 470\,923.93V_{ENDVI} - 16.81V_{EEVI2} + 7.17V_{RVI} + 61.40V_{ARVI} + 4.27Cor_{7\times7} + 56.17V_{GNDVI} + 0.60Cor_{5\times5} + 3.86abu$$

MLR + EM_abu3

$$y = 115.23 + 0.01DEM + 0.16D_{\text{slope}} - 525\,910.99V_{NDVI} + 525\,609.41V_{ENDVI} + 45.81V_{EEVI2} + 7.27V_{RVI} + 54.73V_{ARVI} + 4.73Cor_{7\times7} + 70.34V_{GNDVI} + 0.42Cor_{5\times5} + 4.18abu$$

MLR + EM_abu4

$$y = 146.5 + 0.01DEM + 0.17D_{\text{slope}} - 444\,966.72V_{NDVI} + 444\,985.91V_{ENDVI} - 385.52V_{EEVI2} + 10.28V_{RVI} + 67.48V_{ARVI} + 4.11Cor_{7\times7} + 89.99V_{GNDVI} + 0.68Cor_{5\times5} - 6.03abu$$

MLR + EM_abu5

$$y = 157.21 + 0.01DEM + 0.17D_{\text{slope}} - 457\,606.53V_{NDVI} + 457\,223.08V_{ENDVI} + 48.44V_{EEVI2} + 8.88V_{RVI} + 90.78V_{ARVI} + 4.35Cor_{7\times7} + 38.08V_{GNDVI} + 0.43Cor_{5\times5} - 2.98abu$$

MLR + EM_abu6

$$y = 157.21 + 0.01DEM + 0.17D_{\text{slope}} - 457\,606.53V_{NDVI} + 457\,223.08V_{ENDVI} + 48.44V_{EEVI2} + 8.88V_{RVI} + 90.78V_{ARVI} + 4.35Cor_{7\times7} + 38.08V_{GNDVI} + 0.43Cor_{5\times5} - 2.98abu$$

MLR + EM_abu7

$$y = 133.44 + 0.01DEM + 0.16D_{\text{slope}} - 646\,197.68V_{NDVI} + 645\,926.49V_{ENDVI} - 31.71V_{EEVI2} + 8.33V_{RVI} + 65.18V_{ARVI} + 4.85Cor_{7\times7} + 72.15V_{GNDVI} + 0.34Cor_{5\times5} + 5.21abu$$

对添加丰度信息前后多元线性回归模型进行测试,结果如表 3 所示。添加丰度信息作为反演因子后,7 组添加丰度后模型 R^2 均优于 MLR,说明添加丰度信息可以有效提高多元线性回归模型反演精度,其中 MLR + EM_abu4 效果最优,R^2 为 0.56。MLR + EM_abu6 在 RMSE 指标上获得最优效果,RMSE 为 2.70 mg/hm^2。但是,7 组添加丰度模型的整体 RMSE 结果与 MLR 的 RMSE 接近,优化效果不明显。

表 3 多元线性回归模型测试结果

Tab. 3 Multiple linear regression model test results

模型	R^2	RMSE/(mg·hm^{-2})
MLR	0.51	2.75
MLR + EM_abu1	0.51	2.77
MLR + EM_abu2	0.54	2.75
MLR + EM_abu3	0.52	2.90
MLR + EM_abu4	0.56	2.75
MLR + EM_abu5	0.53	2.72
MLR + EM_abu6	0.53	2.70
MLR + EM_abu7	0.52	2.90

3.4 BP 神经网络模型测试结果

构建研究区域的 BP 神经网络反演模型。同样采用 Pearson 强相关性 10 个因子进行 BP 神经网络模型训练,为了验证丰度信息在地上生物量反演中的有效性,将 7 组丰度信息分别添加到 BP 神经网络模型中训练。试验将 80 个样点作为训练集,20 个样点作为测试集,统一设定迭代次数为 50 000,隐藏层节点为 15,初始学习率为 0.001。BP 神经网络模型以及添加 7 组树木丰度后进行生物量反演 BP 神经网络模型测试结果如表 4 所示。

表 4 BP 神经网络模型测试结果

Tab. 4 BP neural network model test results

模型	R^2	RMSE/(mg·hm^{-2})
BP	0.49	3.12
BP + EM_abu1	0.50	2.67
BP + EM_abu2	0.55	2.59
BP + EM_abu3	0.50	2.73
BP + EM_abu4	0.53	2.80
BP + EM_abu5	0.43	3.06
BP + EM_abu6	0.55	2.50
BP + EM_abu7	0.50	2.86

通过表 4 可发现,采用 MOPSOSCD 算法所提取

的 7 组树木端元的丰度值可以有效提高 BP 神经网络模型的生物量反演精度。模型训练过程中观测值和预测值关系如图 6 所示。BP + EM_abu1 ~ BP + EM_abu7 模型，在 RMSE 指标上均优于采用 BP 神经网络模型（图 6a）的测试结果，当采用 EM_abu6（图 6b）作为反演因子添加到 BP 模型时测试的 R^2 和 RMSE 达到最优，分别为 0.55、2.50 mg/hm^2，其次为 EM_abu2（图 6c）和 EM_abu4（图 6d）。

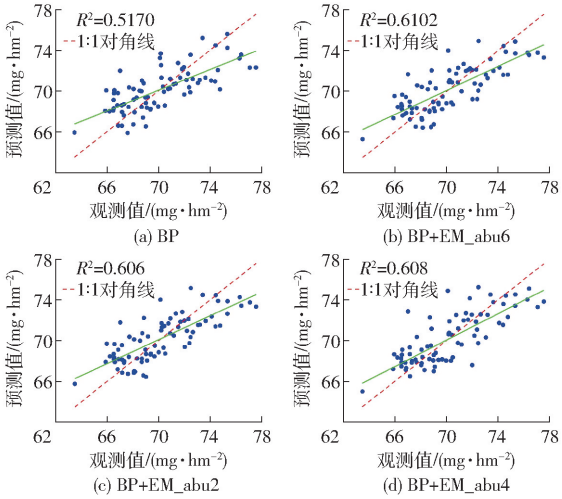

图 6　BP 神经网络模型观测值和预测值关系

Fig. 6　Relationship between observed and predicted values of BP model

3.5　丰度与反演结果关系分析

综合分析 E_{RMSE} 和各组端元的反演结果可发现，在多元线性回归中，R^2 最优的 3 个模型从高到低依次为：MLR + EM_abu4、MLR + EM_abu2、MLR + EM_abu5，RMSE 效果最优的 3 个模型从高到低依次为：MLR + EM_abu6、MLR + EM_abu5、MLR + EM_abu2。BP 神经网络中，R^2 和 RMSE 的结果排序基本保持一致，EM_6 作为最优的 E_{RMSE} 值，在 BP 神经网络模型中对应的 BP + EM_abu6 也获得了最优的效果。因此，所提取的端元组合中，较优的 E_{RMSE} 更有利于提高多元线性回归模型和 BP 神经网络模型的生物量反演精度。

进一步，对 ROI 区域内的 ROI 反演效果进行可视化。多元线性回归模型 MLR、MLR + EM_abu6 结果如图 7a、7b 所示。MLR 与 MLR + EM_abu6 反演的 RMSE 分别为 41.09、38.66 mg/hm^2，R^2 分别为 0.40、0.44。BP 神经网络模型 BP、BP + EM_abu6 结果如图 7c、7d 所示。BP 与 BP + EM_abu6 反演的 RMSE 分别为 32.73、32.07 mg/hm^2，R^2 分别为 0.56、0.57。

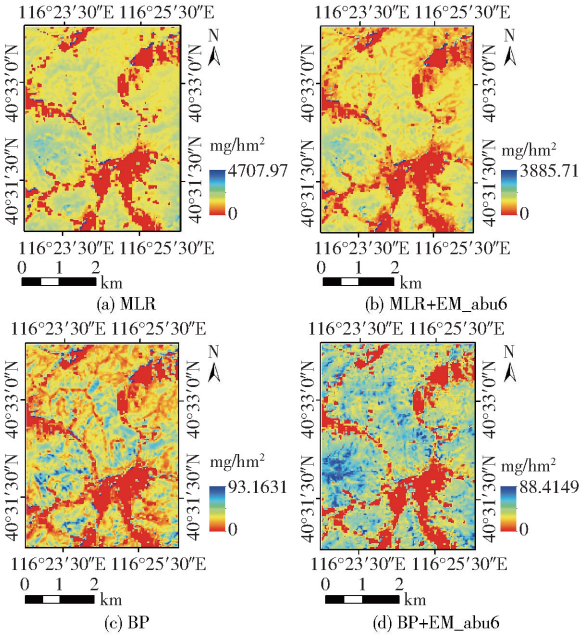

图 7　ROI 生物量反演结果对比

Fig. 7　Comparison of ROI biomass inversion results

综合反演结果及反演效果，采用丰度作为反演因子可以提高 MLR 的精度，但采用 MLR 方法进行生物量反演的精度较低，反演效果存在部分的奇异值。BP 神经网络模型及丰度优化后的模型获得最优反演效果。因此，可优先考虑从端元束中选择效果最优的端元提取丰度信息，并将其作为反演指标进行 BP 神经网络模型的训练及生物量反演。

为了验证最优丰度信息在 BP 神经网络模型的有效性，本文进一步在所研究的 ROI 之外从北京地区随机选取 4 个 ROI 验证。随机选择的区域同时包含针叶林和阔叶林 2 种森林类型。在最优端元提取时采用 3.1 节描述的方法，设置各区域端元数量为 4，同样选择最优丰度结果进行模型测试，4 个 ROI 测试结果及反演效果如表 5 所示。

表 5　北京地区不同 ROI 的测试结果

Tab. 5　Test results of different ROIs in Beijing area

模型	ROI1		ROI2		ROI3		ROI4	
	RMSE/(mg·hm^{-2})	R^2	RMSE/(mg·hm^{-2})	R^2	RMSE/(mg·hm^{-2})	R^2	RMSE/(mg·hm^{-2})	R^2
BP	32.97	0.59	39.25	0.43	35.06	0.49	27.63	0.73
BP + EM_abu	32.00	0.61	40.41	0.43	34.95	0.50	26.15	0.74

通过表 5 可发现，添加了丰度信息的 BP 神经网络模型在 4 个 ROI 反演的整体效果较好，丰度信息添加在 BP 神经网络模型前后 R^2 均值为 0.56 和 0.57，RMSE 均值为 33.73、33.37 mg/hm^2。尽管

4 个 ROI 区域没有采样参与模型训练,但是丰度信息依旧在模型中起到正向的作用。4 个 ROI 的生物量反演效果如图 8 所示。在随机选取的 4 个 ROI 的反演结果中不存在奇异值,其结果与模型在训练 ROI 的结果一致,证明了模型在随机区域的有效性。

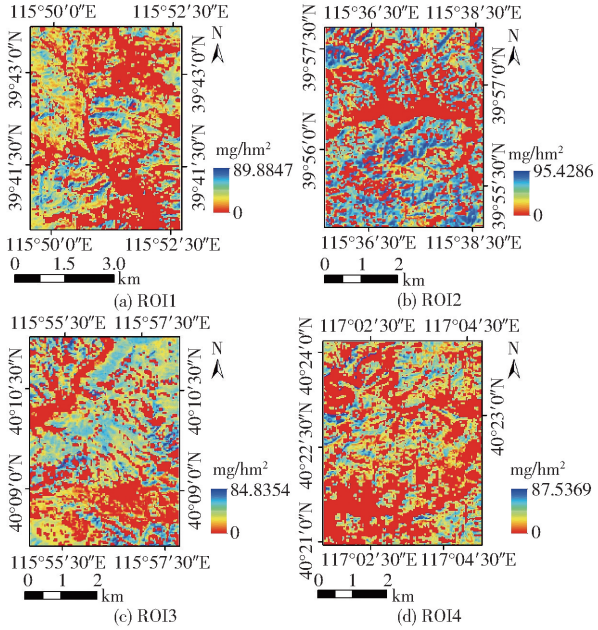

图 8 随机选取的 4 个 ROI 生物量反演结果

Fig. 8 Biomass inversion results of four randomly selected ROIs

4 结束语

提出基于多源数据及丰度信息融合的森林地上生物量估算方法,通过试验验证了丰度信息具有提升生物量反演精度的作用。尽管丰度信息并不直接表征植被的光合信息及结构信息,但作为亚像素维度的参数对于生物量反演起到正向作用。通过综合分析研究结果可发现,端元及丰度提取的精度和地上生物量的反演精度具有较高的相关性。采用 MLR 模型并将丰度信息作为指标进行网络训练时,最优 RMSE 为 38.66 mg/hm^2,R^2 为 0.44。采用 BP 神经网络模型并将丰度信息作为指标进行网络训练时,最优 RMSE 为 32.07 mg/hm^2,R^2 为 0.57。两种模型均比不采用丰度进行生物量反演时效果更好。同时,通过试验证明,采用丰度作为反演指标的前提下,BP 神经网络模型效果优于多元线性回归模型。

参 考 文 献

[1] LE T T, QUEGAN S, DAVIDSON M W J, et al. The BIOMASS mission: mapping global forest biomass to better understand the terrestrial carbon cycle[J]. Remote Sensing of Environment, 2011, 115(11): 2850 - 2860.

[2] PHILLIPS O L, MALHI Y, HIGUCHI N, et al. Changes in the carbon balance of tropical forests: evidence from long-term plots[J]. Science, 1998, 282(5388): 439 - 442.

[3] SANTORO M, CARTUS O, CARVALGAIS N, et al. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations[J]. Earth System Science Data, 2021, 13(8): 3927 - 3950.

[4] BROWN S. Measuring carbon in forests: current status and future challenges[J]. Environmental Pollution, 2002, 116(3): 363 - 372.

[5] 孙钊,谢运鸿,王宝莹,等. 基于无人机多维数据集的森林地上生物量估测模型研究[J]. 农业机械学报,2024,55(6):186 - 195,236.
SUN Zhao, XIE Yunhong, WANG Baoying, et al. Development of forest aboveground biomass estimation model based on multidimensional dataset of UAV[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(6): 186 - 195, 236. (in Chinese)

[6] 周小成,王佩,谭芳林,等. 基于多时相无人机影像的高郁闭度森林采伐生物量估算[J]. 农业机械学报,2023,54(6):168 - 177.
ZHOU Xiaocheng, WANG Pei, TAN Fanglin, et al. Biomass estimation of high-density forest harvesting based on multi-temporal UAV images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 168 - 177. (in Chinese)

[7] NAIK P, DALPONTE M, BRUZZONE L. Prediction of forest aboveground biomass using multitemporal multispectral remote sensing data[J]. Remote Sensing, 2021, 13(7): 1282.

[8] ZHANG B, ZHANG L, XIE D, et al. Application of synthetic NDVI time series blended from Landsat and MODIS data for grassland biomass estimation[J]. Remote Sensing, 2015, 8(1): 10.

[9] 孙懿慧. 基于 GIS 多源数据融合的湖北省中稻增产潜力及影响因子的研究[D]. 武汉:华中农业大学,2015.
SUN Yihui. Study of production potential and the affect factors of middle-season rice in Hubei Province using GIS and multi-source data[D]. Wuhan: Huazhong Agricultural University, 2015. (in Chinese)

[10] 冯权泷,牛博文,朱德海,等. 基于多核主动学习和多源数据融合的农田塑料覆被分类[J]. 农业机械学报,2022,53(2):177 - 185.
FENG Quanlong, NIU Bowen, ZHU Dehai, et al. Classification of agricultural plastic cover based on multi-kernel active learning and multi-source data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 177 - 185. (in Chinese)

[11] 魏永霞,杨军明,吴昱,等. 基于多源数据融合模型的水稻面积提取[J]. 农业机械学报,2018,49(10):300 - 306.
WEI Yongxia, YANG Junming, WU Yu, et al. Rice planting area extraction based on multi-source data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(10):300 - 306. (in Chinese)
[12] 孙亚楠,李仙岳,史海滨,等. 基于多源数据融合的盐分遥感反演与季节差异性研究[J]. 农业机械学报,2020,51(6):169 - 180.
SUN Ya'nan, LI Xianyue, SHI Haibin, et al. Remote sensing inversion of soil salinity and seasonal difference analysis based on multi-source data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(6):169 - 180. (in Chinese)
[13] LI H, KATO T, HAYASHI M, et al. Estimation of forest aboveground biomass of two major conifers in Ibaraki Prefecture, Japan, from PALSAR - 2 and Sentinel - 2 data[J]. Remote Sensing, 2022, 14(3): 468.
[14] ZHU Y, FENG Z, LU J, et al. Estimation of forest biomass in Beijing (China) using multisource remote sensing and forest inventory data[J]. Forests, 2020, 11(2): 163.
[15] MUSTHAFA M, SINGH G. Improving forest above-ground biomass retrieval using multi-sensor L-and C-band SAR data and multi-temporal spaceborne LiDAR data[J]. Frontiers in Forests and Global Change, 2022, 5: 822704.
[16] YANG Q, NIU C, LIU X, et al. Mapping high-resolution forest aboveground biomass of China using multisource remote sensing data[J]. GIScience & Remote Sensing, 2023, 60(1): 2203303.
[17] KESHAVA N, MUSTARD J F. Spectral unmixing[J]. IEEE Signal Processing Magazine, 2002, 19(1): 44 - 57.
[18] DU M, LI M, NOGUCHI N, et al. Retrieval of fractional vegetation cover from remote sensing image of unmanned aerial vehicle based on mixed pixel decomposition method[J]. Drones, 2023, 7(1): 43.
[19] WU J, GAO Z, LIU Q, et al. Methods for sandy land detection based on multispectral remote sensing data[J]. Geoderma, 2018, 316: 89 - 99.
[20] 刘雅婷,龚龑,段博,等. 多时相 NDVI 与丰度综合分析的油菜无人机遥感长势监测[J]. 武汉大学学报(信息科学版),2020,45(2):265 - 272.
LIU Yating, GONG Yan, DUAN Bo, et al. Combining multi-temporal NDVI and abundance from UAV remote rensing data for oilseed rape growth monitoring[J]. Geomatics and Information Science of Wuhan University, 2020,45(2):265 - 272. (in Chinese)
[21] ZHANG T, LIU D. Estimating fractional vegetation cover from multispectral unmixing modeled with local endmember variability and spatial contextual information[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 209: 481 - 499.
[22] DU H, FAN W, ZHOU G, et al. Retrieval of canopy closure and LAI of Moso bamboo forest using spectral mixture analysis based on real scenario simulation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(11): 4328 - 4340.
[23] 龚龑,肖洁,侯金雨,等. 基于无人机遥感混合光谱分析的油菜估产模型[J]. 测绘地理信息,2017,42(6):40 - 45.
GONG Yan, XIAO Jie, HOU Jinyu, et al. Rape yields estimation research based on spectral analysis for UAV image[J]. Journal of Geomatics, 2017,42(6):40 - 45. (in Chinese)
[24] WANG W, ZHENG H, WU Y, et al. An assessment of background removal approaches for improved estimation of rice leaf nitrogen concentration with unmanned aerial vehicle multispectral imagery at various observation times[J]. Field Crops Research, 2022, 283: 108543.
[25] SWATANRTAN A, DUBAYAH R, ROBERTS D, et al. Mapping biomass and stress in the Sierra Nevada using LiDAR and hyperspectral data fusion[J]. Remote Sensing of Environment, 2011, 115(11): 2917 - 2930.
[26] KIM T W, WE G J, SUH Y C. Correlation analysis with vegetation indices and vegetation-endmembers from airborne hyperspectral data in forest area[J]. Journal of the Korean Association of Geographic Information Studies, 2012, 15(3): 52 - 65.
[27] ZHANG X, LIU L, CHEN X, et al. GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery[J]. Earth System Science Data, 2021, 13(6): 2753 - 2776.
[28] 董乐. 高光谱遥感影像解混算法研究[D]. 西安:中国科学院西安光学精密机械研究所,2021.
DONG Le. Research of unmixing algorithm for hyperspectral remote sensing image[D]. Xi'an: Institute of Optics & Precision Mechanics, Chinese Academy of Sciences, 2021. (in Chinese)
[29] 林洁雯,陈建,罗婷文,等. 基于多模态多目标优化的端元束提取方法研究[J]. 农业机械学报,2023,54(7):234 - 242.
LIN Jiewen, CHEN Jian, LUO Tingwen, et al. Endmember bundle extraction method based on multi-modal and multi-objective optimization[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023,54(7):234 - 242. (in Chinese)
[30] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533 - 536.
[31] 魏雪梅. 多源数据支持下的森林地上生物量估算方法[J]. 武汉大学学报(信息科学版),2019,44(9):1385 - 1390.
WEI Xuemei. Estimation of forest aboveground biomass based on multi-source data[J]. Geomatics and Information Science of Wuhan University, 2019,44(9):1385 - 1390. (in Chinese)
[32] 李康杰,胡中岳,刘萍,等. 基于 Landsat - 8 Oli 影像的珠三角森林生物量估测[J]. 中南林业科技大学学报,2023,43(3):73 - 81.
LI Kangjie, HU Zhongyue, LIU Ping, et al. Estimation of forest aboveground biomass in the Pearl River Delta based on Landsat - 8 Oli remote sensing image data[J]. Journal of Central South University of Forestry & Technology, 2023,43(3): 73 - 81. (in Chinese)