

DOI: 10.20103/j.stxb.202208292458

赫晓慧, 张乐涵, 乔梦佳, 田智慧, 周广胜. 基于 PROSAIL 混合反演模型的 MODIS LAI 产品改进及评估. 生态学报, 2023, 43(22): 9328-9341.

He X H, Zhang L H, Qiao M J, Tian Z H, Zhou G S. Improvement and analysis of MODIS LAI product based on PROSAIL hybrid inversion model. Acta Ecologica Sinica, 2023, 43(22): 9328-9341.

基于 PROSAIL 混合反演模型的 MODIS LAI 产品改进及评估

赫晓慧^{1,2,*}, 张乐涵^{1,2}, 乔梦佳³, 田智慧^{1,2}, 周广胜²

1 郑州大学地球科学与技术学院, 郑州 450001

2 郑州大学中国气象科学研究院郑州大学生态气象联合实验室, 郑州 450001

3 郑州大学计算机与人工智能学院, 郑州 450001

摘要: 叶面积指数 (Leaf Area Index, LAI) 是定量陆地生态系统中光合作用、呼吸作用、蒸腾、碳和养分循环等过程中物质与能量交换的重要结构参数。目前大、中尺度的气候和生态水文建模使用的 LAI 产品主要来源于中分辨率成像光谱仪 (MODIS), 但由于其反演过程中的不确定性因素导致 MODIS LAI 产品在部分地区存在质量问题。以青海省复杂植被地区为研究区域, 基于实地考察与采样验证了区域内 MODIS LAI 所存在的质量问题分布, 并揭示了不确定因素的影响。与此同时, 提出了一种基于 PROSAIL 模型与深度神经网络 (DNN) 的混合建模技术, 针对 MODIS LAI 生成机制中地表分类数据、地表反射率数据和反演算法的不确定性进行改进, 并基于青海省大范围实测 LAI 数据评估了改进前后产品的准确度, 实测数据的验证结果发现: 改进模型的 LAI 准确度 (RMSE = 0.48, $R^2 = 0.64$) 显著高于 MODIS LAI (RMSE = 0.71, $R^2 = 0.56$), 预测结果与实测结果之间的偏差显著减少; 区域尺度上, 柴达木荒漠植被低覆盖典型区域、三江源高寒草甸中覆盖典型区域与青海湖牧场草地高覆盖典型区域的 RMSE 分别提高了 0.19、0.10、0.54, 改进方法有效解决了 MODIS LAI 产品中高覆盖植被饱和和效应导致的高估以及低覆盖植被未检索导致低估的质量问题, 改进结果分布连续, 更符合真实植被状况。基于以上研究, 充分证明了研究方法对 MODIS LAI 产品的改进具有可靠性, 能够在缺少实测样本数据的情况下有效提高 MODIS LAI 的质量, 为全球植被环境监测与生态建模提供重要的数据支持。

关键词: MODIS LAI; PROSAIL 模型; 叶面积指数; 深度神经网络

Improvement and analysis of MODIS LAI product based on PROSAIL hybrid inversion model

HE Xiaohui^{1,2,*}, ZHANG Lehan^{1,2}, QIAO Mengjia³, TIAN Zhihui^{1,2}, ZHOU Guangsheng²

1 School of Earth Science and Technology, Zhengzhou University, Zhengzhou 450001, China

2 Joint Laboratory of Eco-Meteorology, Zhengzhou University, Chinese Academy of Meteorological Science, Zhengzhou 450001, China

3 School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China

Abstract: Leaf Area Index (LAI) is an important structural parameter to quantify the exchange of matter and energy in the processes of photosynthesis, respiration, transpiration, carbon and nutrient cycling in terrestrial ecosystems. At present, the LAI products used for large-scale and medium-scale climate and eco-hydrological modeling are mainly derived from Moderate Resolution Imaging Spectroradiometer (MODIS), but due to the uncertain factors in the inversion process, the MODIS LAI products have quality problems in some areas. This study took the complex vegetation area of Qinghai Province as the research area, based on field investigation and sampling to verify the distribution of quality problems of MODIS LAI

基金项目: 第二次青藏高原科学考察计划 (2019QZKK0106); 河南省重大科技专项 (201400210900)

收稿日期: 2022-08-29; **采用日期:** 2023-04-07

* 通讯作者 Corresponding author. E-mail: hexh@zzu.edu.cn

products in the research area, and revealed the influence of uncertain factors of MODIS LAI products on their quality problems. At the same time, this paper proposed a hybrid modeling technique based on the PROSAIL radiative transfer model and deep neural network (DNN) model to improve the uncertainties of surface classification data, surface reflectance data, and inversion algorithm in the generation mechanism of MODIS LAI products. It is determined to make improvements and solve the quality problems of MODIS LAI products. In the end, based on the large-scale measured LAI data in Qinghai Province, the accuracy of the product before and after improvement was evaluated, and the validation results of the measured data showed that the LAI accuracy of the improved model ($RMSE=0.48$, $R^2=0.64$) was significantly higher than that of MODIS LAI products ($RMSE=0.71$, $R^2=0.56$), and the deviation between the predicted results and the measured results was significantly reduced. On a regional scale, the RMSE accuracy of typical areas of low cover type desert vegetation in the Qaidam region, typical areas of medium cover type alpine meadows in the Sanjiangyuan region, and typical areas of high cover type pasture grassland in the Qinghai Lake region have been increased by 0.19, 0.10, and 0.54, respectively. The improved method effectively solved the quality problems of overestimation caused by the saturation effect of high vegetation coverage and underestimation caused by the non-retrieval of low vegetation coverage in MODIS LAI products. The distribution of the improved results is continuous, which is more in line with the actual vegetation conditions. Based on the above research and analysis, it is fully proved that the research method in this paper are reliable for the improvement of MODIS LAI products, and can effectively improve the quality of MODIS LAI products in the absence of measured sample data, and provide important data support for global vegetation environmental monitoring and ecological modeling.

Key Words: MODIS LAI; PROSAIL model; leaf area index; deep neural network

叶面积指数定义为单位地表面上单面绿叶面积的总和,它与植物光合作用、蒸腾作用和生态系统生产力密切相关,是模拟陆地生态过程、生物地球化学循环以及水热循环的重要参数,在农业、林业、生态、气候变化等领域已经得到了广泛的应用^[1-2]。目前,卫星遥感技术已被证明是监测区域和半球/全球尺度 LAI 时空变化的有效方法^[3]。其中 MODIS LAI 产品作为成熟的大尺度 LAI 数据,能够获取全球大多数地区大范围、多时相的 LAI 分布^[4-5]。然而由于植被生长、地表异质性、影像数据等方面的差异,其应用到具体区域时存在明显的低估或者高估现象^[6-7]。国内外学者对 MODIS LAI 产品的验证进行了一系列研究,青海湖流域 LAI 验证中,MODIS LAI 产品对高寒草甸低估了 10%,对草原高估了 9%,对亚高山灌丛高估了 51%^[8]。通过验证锡林郭勒地区 MODIS LAI 发现,研究区的 MODIS LAI 产品(即 MOD15A2)被高估约为 25%^[9]。多项研究证明 MODIS LAI 产品的偏差问题是普遍存在的,对于植被环境差异较大的地区,这些偏差问题表现更为显著。

如今,较多学者对 MODIS LAI 产品进行了真实性检验与偏差分析,但针对 MODIS LAI 产品质量问题改进的研究很少。有学者利用经验反演模型建立实测数据与 Landsat 反射率数据之间的关系进行高空间分辨率的 LAI 反演,利用 Landsat 与 MODIS 之间的尺度转换提高 MODIS LAI 质量^[9-11]。这种方法更多侧重于 MODIS LAI 产品的验证,对于大尺度的 MODIS LAI 改进并不适用。数据同化技术能够将多源遥感数据与长时间序列数据结合用于实时获取叶面积指数,基于几何卡尔曼滤波与 LAI 多年变化数据利用数据同化方法能够实时地估计叶面积指数,解决在生长季节 MODIS LAI 缺失的问题^[12]。同样考虑到植被冠层的二向反射分布特征,利用多年际 MODIS 数据获取影像多角度观测信息修正观测几何条件的差异所导致的辐射信号不确定性,也能够提高 MODIS LAI 反演效果^[13-14]。

但以往研究是从数据机理方面对产品进行的补充,利用多源、多时相、多角度数据对 MODIS LAI 数据进行填补能够提高准确度。但复杂地区 MODIS LAI 的不确定性是长期存在的^[9, 15],仅外部数据补充无法完全解决 MODIS LAI 本身所存在的质量问题,需要针对产品生成机制内的问题进行改进。Ji 指出影响 MODIS LAI 产品质量的因素主要有 3 个方面,一是地表分类数据的不确定性,二是地表反射率数据的不确定性,三是用于构建查找表的反演算法的不确定性^[10]。近年来遥感反演技术的发展为解决 MODIS LAI 产品的不确定问题提供了思路。PROSAIL 模型通过耦合 PROSPECT 叶片模型和 SAIL 冠层结构模型来模拟植物冠层的反射与折

射^[16],能够准确生成大量的模拟反射率数据,广泛地应用于植被生物物理参数反演中^[17]。同样,深度学习方法也逐渐成为植被参数反演的热点,其中深度神经网络(deep neural network, DNN)已成功用于估算叶面积指数和其他全球植被产品,相比传统的统计方法和机器学习方法效率更高,适应性更强^[18-19]。

本研究以植被环境复杂的青海省为研究区域,基于实测数据、遥感数据以及考察经验,深入分析 MODIS LAI 产品在青海省存在的低估和高估问题的分布及产生原因,构建 PROSAIL 模型与深度神经网络(DNN)的混合模型对 MODIS LAI 产品生成机制中的数据、模型、算法的不确定性进行改进,最后对比改进前后 LAI 产品的质量。本研究着力于提高 MODIS LAI 产品精度,以期为全球尺度的生态建模与评估提供重要的数据支持。

1 材料与方法

1.1 研究区域概况

研究区域为中国青藏高原地区的青海省全域,位居中国西部,地处东部季风区、西北干旱区和青藏高原三大地理区域的交汇处,地理位置介于 $89^{\circ}35'—103^{\circ}04'E$, $31^{\circ}36'—39^{\circ}19'N$ 之间。青海省土地总面积 72.23 万 km^2 ,约占青藏高原总面积的 30%。如图 1 中青海省地势西高东低,南北高中间低,平均海拔 3000m 以上,地形复杂,地貌多样。属大陆性高原气候,年降水量为 15—750mm,降水量时空分布不均,从西北到东南渐增。年均温为 $-5.1—9.0^{\circ}C$,气温的垂直变化明显。

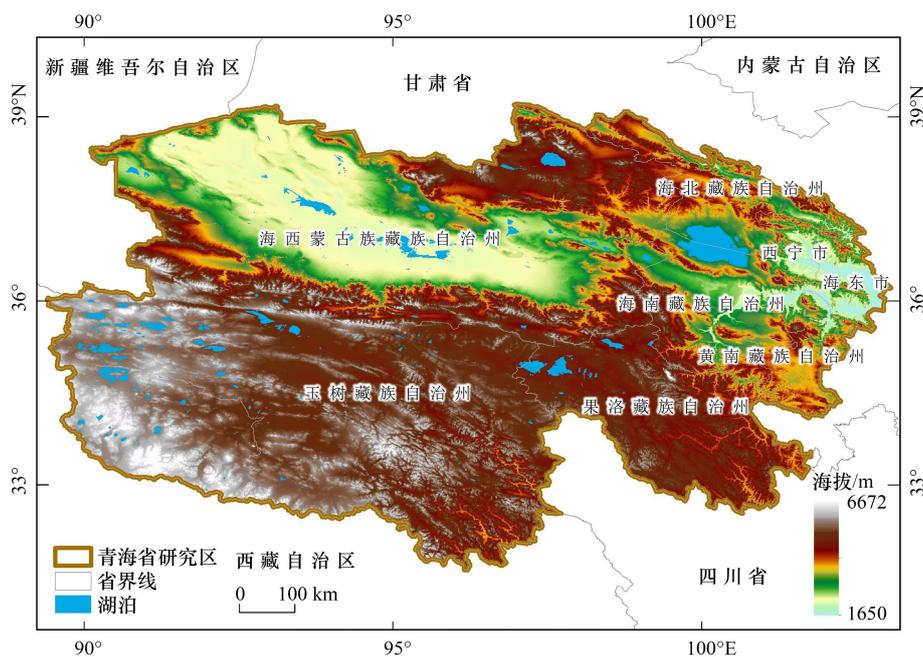


图 1 研究区域概况

Fig.1 Overview of the study area

青海省内的柴达木盆地、青海湖、三江源等生态保护区,在气候、地形地貌、生态环境、植被覆盖等多个方面差异较大。其中位于青海省西北部的柴达木盆地为典型的干旱半干旱地区,主要分布大范围的荒漠植被;东部的青海湖地区水土条件良好,分布有草甸草地、沼泽湿地以及祁连山的森林资源;南部的三江源地区属于高海拔高寒地带,分布有高寒草甸、高寒草原、高寒荒漠草原等多种典型草地植被^[20]。复杂多样的植被环境能够进行多类型植被的综合性研究。

1.2 数据源

1.2.1 实地数据调查与测量

在青海省开展植被叶面积指数野外调查,采集时间为 2021 年 7 月末—8 月初,该时段植被生长状况最

优。根据典型植被类型以及生长区域覆盖情况选择具有代表性的试验样点进行地面数据采集,每个样点区域内选择间隔均匀的三个不同位置设置样方点,测量 5 次取平均值代表其 LAI,整个样点区域由三个样方对比取值。使用的测量仪器为植物冠层分析仪,通过测定冠层上、下辐射以及消光系数等相关参数来计算叶面积指数,测量过程中记录样点坐标^[21]。

1.2.2 空间数据预处理

本研究获取的遥感影像数据为覆盖研究区且在野外调查时间内的 MODIS 卫星遥感数据,包括 MOD09A1 地表反射数据产品与 MOD15A2 叶面积指数产品。MOD09A1 反射率产品每 8 天生产一次,像素大小为 500m,其提供了 MODIS 数据 1—7 波段的地表光谱反射率,包含可见光和近红外波段范围,可以更好的反映植被信息。MOD15A2 是 MODIS 传感器基于三维传输模型,利用反射率数据与植被分类数据反演得到的 LAI 产品数据,像素大小为 500m。其主算法包括一个基于查找表(LUT)的程序,当主算法失效时使用归一化植被指数(NDVI)与冠层 LAI 之间的经验关系的备份算法进行计算^[22]。随产品附带较为详细的质量控制(QC)信息,描述 MODIS LAI 相关质量问题。本文对比选取了实测时间范围内质量最好的 MODIS 影像及对应的 LAI 产品,对其进行大气校正与去云处理,并根据质量控制信息对 MODIS LAI 质量进行了提取分类。

1.3 基于 PROSAIL 的混合反演模型

基于 PROSAIL 模型与深度神经网络(DNN)的混合建模技术对 MODIS LAI 生成过程进行改进。利用 PROSAIL 模型改进 MODIS LAI 算法中的三维辐射传输模型,建立准确、多样、具有代表性的模拟数据集^[23]。对于模拟数据与 MODIS 数据需要进一步预处理,通过数据匹配建立互补数据集以减少数据异质性。LUT 算法结合物理模型反演本质上是一个不适定问题^[24],大量的模拟数据导致 LUT 拟合效率与精度降低,利用 DNN 算法改进 MODIS LAI 算法中 LUT 算法,能够高效、快速的对数据进行非线性拟合。最后对预测结果进行精度验证以及改进效果分析(图 2)。

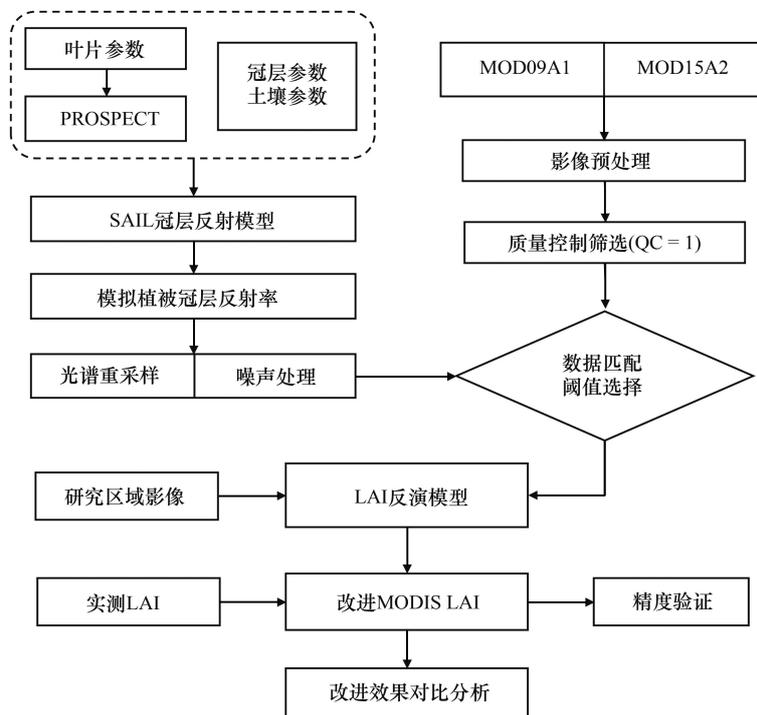


图 2 LAI 反演流程

Fig.2 LAI inversion process

QC:质量控制等级;MOD09A1:MODIS 地表反射数据;MOD15A2:MODIS 叶面积指数数据;PROSPECT:叶片反射率模型;SAIL:冠层反射率模型;LAI:叶面积指数

1.3.1 PROSAIL 模型构建

PROSAIL 辐射传输模型是通过耦合 PROSPECT 叶片模型和 SAIL 冠层结构模型得到的。在叶片层面, PROSPECT 通过叶片结构和生化参数模拟各种绿色单子叶和双子叶植物以及衰老叶片在 400—2500nm 太阳光谱范围内的反射和透射^[25]。之后 PROSPECT 模拟得到的叶片信息会作为 SAIL 冠层辐射模型的输入, SAIL 模型会把植被当作混浊介质,假设叶片方位角分布均匀,同时考虑任意的叶片倾角,最后精确模拟一定观测条件下的植被冠层反射率^[16]。本文采用的 PROSAIL 为后续改进版本 PROSPECT-5B 和 SAILH 耦合,其模型主要参数见表 1。

表 1 PROSAIL 模型参数内容
Table 1 PROSAIL model parameters

模型 Model	参数 Parameter	表述 Description	基础值 Base value	变化范围 Variable range
PROSPECT-5B	N	叶片结构参数	1.5	1—4
	Cab	叶片叶绿素含量/($\mu\text{g}/\text{cm}^2$)	50	0—80
	Car	类胡萝卜素含量/($\mu\text{g}/\text{cm}^2$)	10	0—15
	Cm	叶片干物质含量/(g/cm^2)	0.005	0.002—0.1
	Cw	叶片含水量/(g/cm^2)	0.015	0.005—0.04
SAILH	LAI	叶面积指数	3.5	0—10
	ALA	平均叶倾角/($^\circ$)	50	20—50
	hspot	热点参数	0.2	0.01—1
	Psoil	土壤亮度参数	0.1	0.01—3
	tts	太阳天顶角/($^\circ$)	30	-50—50
	tto	观察天顶角/($^\circ$)	0	-50—50
	Psi	观测相对方位角/($^\circ$)	0	0—180

模拟数据集的优劣直接会影响后续 LAI 的反演质量,由于 PROSAIL 模型输入参数较多,为了更好地确定参数类型和个数,在构建模型时需要利用敏感性分析找到驱动模型的关键变量^[26]。通过对各参数的敏感性分析可以寻找更符合研究区域的参数范围,提高模型反演精度。

1.3.2 数据集建立

遥感影像与模拟数据之间存在差异会导致反演过程病态^[27],模型模拟是通过微观尺度结构参数对植被进行观测,而遥感观测是将地表变量映射到遥感原始观测^[28]。通过建立模型与观测互补数据可以解决病态反演问题,首先利用 PROSAIL 模型建立大量的模拟 MODIS 数据,并对模拟数据进行噪声处理。将与波长无关的 2% 相对高斯白噪声添加到模拟反射率中,以解释来自大气校正算法的卫星表面反射率数据中的不确定性、残余云层影响和辐射传输模型的固有局限性,使模拟数据更接近真实影像状况^[29—30]。之后根据 MODIS LAI 质量分类选择 QC=1 的优等质量像元作为 MODIS 观测数据,通过欧式距离从 PROSAIL 模拟数据中匹配光谱和 LAI 相似的数据,设置合适的距离阈值使匹配结果既保留了相似性又保留部分不确定性。另外需要加入雪山、湖泊、沙漠等非植被区域数据作为补充,建立有效的数据集。

1.3.3 深度神经网络(DNN)

植被对光谱吸收与反射的有效辐射程度能够反应植被叶片面积的大小,从遥感影像中准确地挖掘光谱信息是叶面积指数预测的关键^[31]。深度学习技术在遥感特征提取中表现出了可靠的性能,由于模拟光谱为一维向量特征,通用的 CNN 卷积结构在特征提取时受到约束,而 DNN 模型中的全连接方式能够更加适用于深层挖掘纵向光谱特征^[32]。因此设计了一种 DNN 网络对模拟反射率与 LAI 之间的关系进行回归建模。

DNN 模型结构主要包括四个隐藏层和一个输出层(图 3),层与层之间的全连接结构建立了不同波段信息之间的权重连接^[19]。根据光谱数据和网络结构的特点,隐藏层中采用 relu 与 softsign 激活函数对光谱特征进行非线性拟合,输出层采用 linear 激活函数对 LAI 预测结果与标签进行线性修正。各网络层之间需经过标准化层对特征数据进行标准化处理,避免样本噪声导致 DNN 向错误方向更新^[18],能够提高模型的稳健性。

数据集中的光谱波段会被转化为特征向量输入到网络模型中进行训练与 LAI 预测。

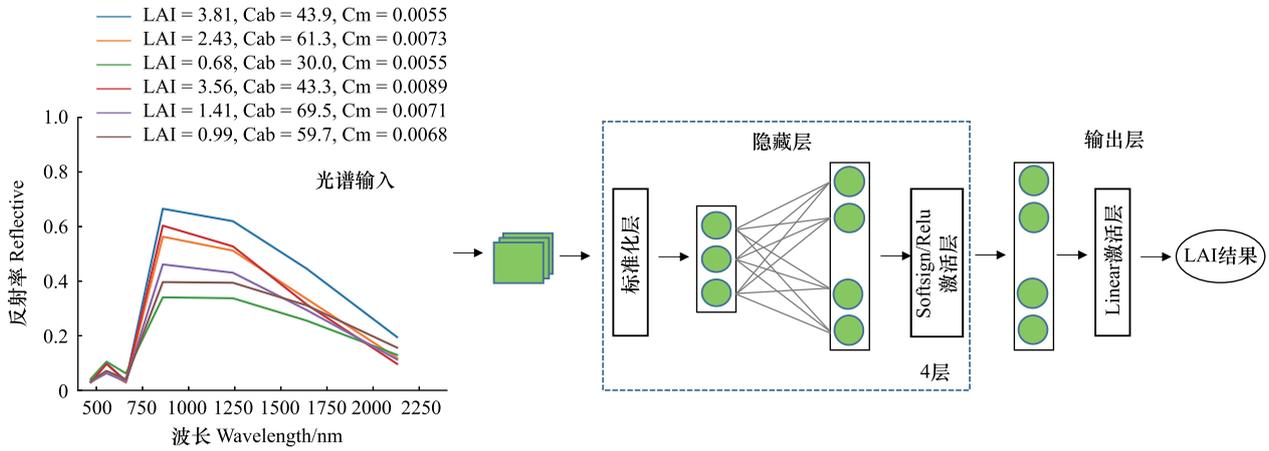


图3 DNN 网络结构

Fig.3 DNN network structure

Cab; 叶片叶绿素含量; Cm; 叶片干物质含量; DNN; 深度神经网络; Softsign、Relu、Linear; 深度学习模型中三种不同激活函数的名称;

1.3.4 精度评价

数据集包含 20000 组数据,随机分为训练集(80%)和验证集(20%),训练过程中使用 k-折叠交叉验证方法来获得模型的训练精度。对于精度验证的标准,主要采用决定系数(R^2)和均方根误差(RMSE)两种常用衡量回归方法的指标来评价模型精度。总体而言, R^2 越高, RMSE 越低,模型拟合越好。式中 y_i 为 LAI 预测值, \bar{y} 为预测 LAI 均值, \hat{y}_i 为 LAI 验证值, m 为预测点的数量。

$$R^2: \frac{\sum_{i=1}^m (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \tag{1}$$

$$RMSE: \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{2}$$

2 结果与分析

2.1 PROSAIL 模型参数率定

变量选择是减少模型过拟合,提高模型性能的关键。利用 EFAST 方法对 PROSAIL 模型中输入与输出进行全局敏感性计算^[26],其结果考虑到各个输入参数以及参数之间的相互作用与共同影响。通过 Monte Carlo 方法对 Cab、Car、Cw、Cm、N、ALA、LAI、Psoil 八个描述植被冠层的主要参数进行随机采样,使用 EFAST 方法对随机采样数据进行全局敏感性分析。

参数敏感性分析结果中(图 4), LAI、Cm、Cab、ALA 四个参数对反射率敏感程度较高,设置为主要的模型参数输入。模拟的高光谱数据利用 MODIS 光谱响应函数重采样到 MODIS 数据对应的 7 个波段。通过 Pearson's 相关系数计算模拟数据中波段与 LAI 值之间的相关性^[14],筛选出 B1(459—479nm)、B3(620—670nm)、B4(841—876nm)、B5(1230—1250nm)、B7(2105—2155nm)5 个对 LAI 贡献较大的波段,波段选择后减少了不相关数据的混淆,反演精度未受到显著影响,模型效率与稳定性得到提高。

2.2 MODIS LAI 产品质量

根据产品质量控制文件将研究区域内 MODIS LAI 划分为 4 种不同质量水平(QC),其中图 5 质量控制分布中灰色区域(QC=1)为未受到影响区域,产品质量较好,其余区域主要分布以下三种质量问题。

饱和效应问题(质量控制分布, QC=2, 红色区域)主要位于青海省的东部地区,典型区域选择为青海湖与

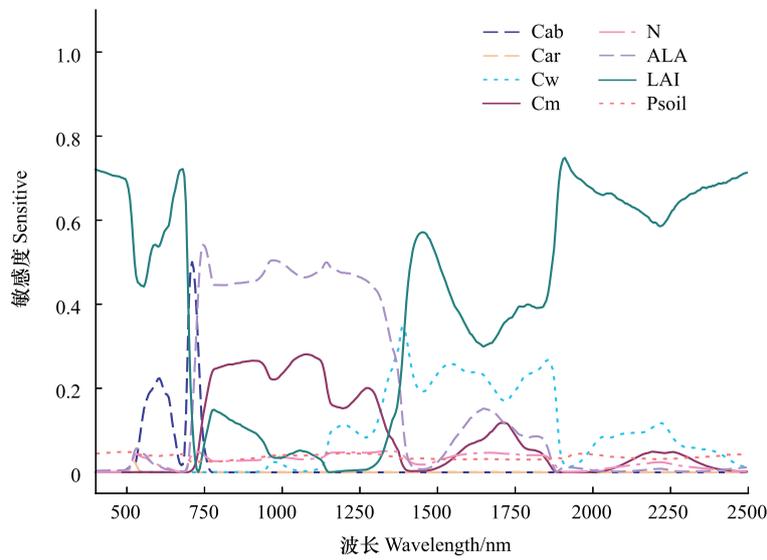


图 4 PROSAIL 主要参数的敏感性结果

Fig.4 Sensitivity of PROSAIL parameters

Cab:叶绿素含量;N:叶片结构参数;Car:类胡萝卜素含量;ALA:平均叶倾角;Cw:叶片含水量;LAI:叶面积指数;Cm:叶片干物质含量;Psoil:土壤亮度参数

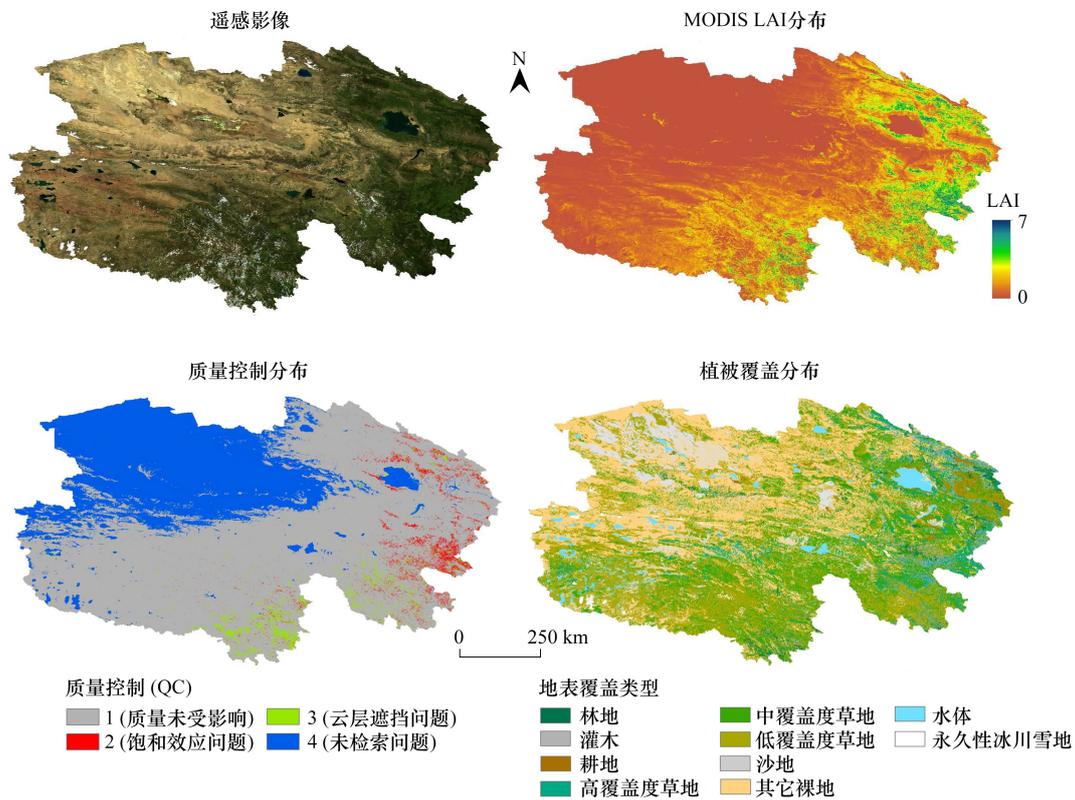


图 5 研究区域的相关专题地图

Fig.5 Relevant thematic maps of the study area

青海东部边界。该问题对应位置的植被类型主要分布为高覆盖度的牧场草地以及灌丛,而 MODIS LAI 所在像元区域整体大于 5,已超出草地、灌丛的 LAI 参数范围,存在明显的高估。对比分析得出,由于当植被 LAI

达到一定程度后,光谱反射率会趋于饱和^[33]。这导致在模型拟合的过程中,处于饱和范围的光谱会出现 LAI 因过饱和而被高估的问题。

无法检索问题(质量控制分布, QC=4, 蓝色区域)主要位于柴达木地区,分布于地表覆盖为沙漠、水体、常年积雪等非植被区域,MODIS 算法根据地表分类数据未进行 LAI 检索^[33-34]。该问题区域包含大范围的沙地与荒漠,20%的区域存在覆盖度低的荒漠植被。由于地表土壤裸露,MODIS 算法将该区域整体划分为非植被区域,这导致 MODIS LAI 出现无法检索问题,部分低覆盖的荒漠植被 LAI 被计算为 0。

云层遮挡问题(质量控制分布, QC=3, 绿色区域)主要分布于青海南部,如玉树附近,在反射率影像中该区域是存在大量云层导致算法失效,云层像元反射率与真实地表反射率差异过大^[35], LAI 无法进行有效预测,这造成所在区域 LAI 存在明显的不连续分布状态。但云层遮挡问题的出现不存在规律,主要与影像质量有关,因此不再进行过多讨论。

2.3 基于实测数据的验证

在青海省区域进行了实地考察与采样,由于实测样点获取难度较大,且植被叶面积指数对时间变化敏感,时效性较强^[4, 29, 36]。选取了对应时间范围内受云层影响最少的 MODIS 影像进行 LAI 预测。从实地采样中筛选出质量好、有代表性的 60 个 LAI 样点记录,这 60 个样点覆盖多种典型植被类型的区域,能够对改进结果与 MODIS LAI 产品进行有效的验证对比。

2.3.1 整体精度验证

利用实测样点对青海省整体区域的改进结果与 MODIS LAI 产品进行评估。在实测数据的验证中(图 6),MODIS LAI 的 RMSE 为 $0.71(\text{m}^2/\text{m}^2)$, R^2 为 0.56,整体上样点结果分布相较于 1:1 线分布离散,部分样点在高值范围的 LAI 偏高,低值范围的 LAI 偏低。改进结果准确度明显高于 MODIS LAI 产品, RMSE 提升到

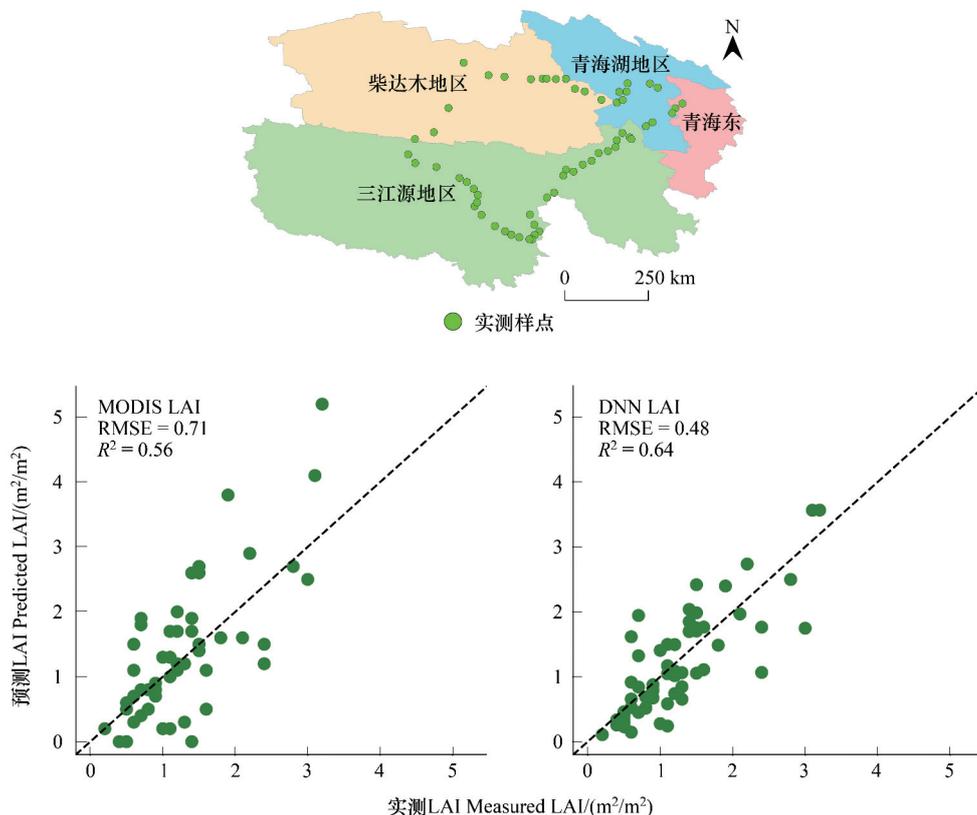


图 6 青海省实测样点分布与 LAI 结果验证

Fig.6 The distribution of measured sample points in Qinghai Province and the verification of LAI results
MODIS LAI; MODIS LAI 产品中的 LAI 结果; DNN LAI; 改进方法得到的 LAI 结果

0.48, R^2 提升到 0.64, 整体散点分布更集中于 1:1 线, 未出现严重的高估或低估问题。

计算各个样点实测结果与产品改进前后结果的偏差值, 以散点图形式分别展示 60 个样点位置的 MODIS LAI 产品偏差和 DNN 改进结果的偏差(图 7)。DNN 改进结果相对于 MODIS LAI 的低估与高估问题有显著的改善, 改进后 LAI 结果与真实值之间平均误差在 0.5 左右, 整体上限制在 0.5 以内, 偏差较大的样点较少, 相比 MODIS LAI 结果, 多于 60% 的样点精度都有明显提升。

2.3.2 区域精度分析

由于青海省区域广阔, 不同地区的地形气候、植被类型与覆盖存在较大差异。利用实测数据与影像结果对三个不同地区的改进结果进行区域精度验证(图 8)与典型区域对比分析(图 9), 图 9 中选择的 5 个不同区域为存在不同质量问题的典型区域。图 8 中 Error 为预测值与真实值之间偏差占真实值的比例, 代表各地区

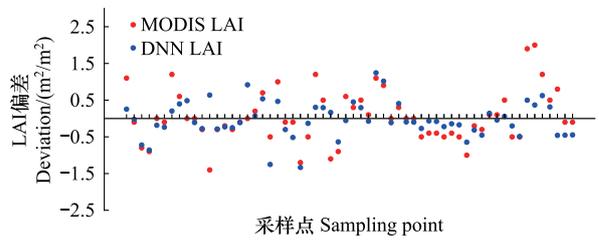


图 7 改进前后 LAI 结果与实测样点 LAI 的偏差
Fig.7 Deviation between LAI results and measured sample points before and after improvement

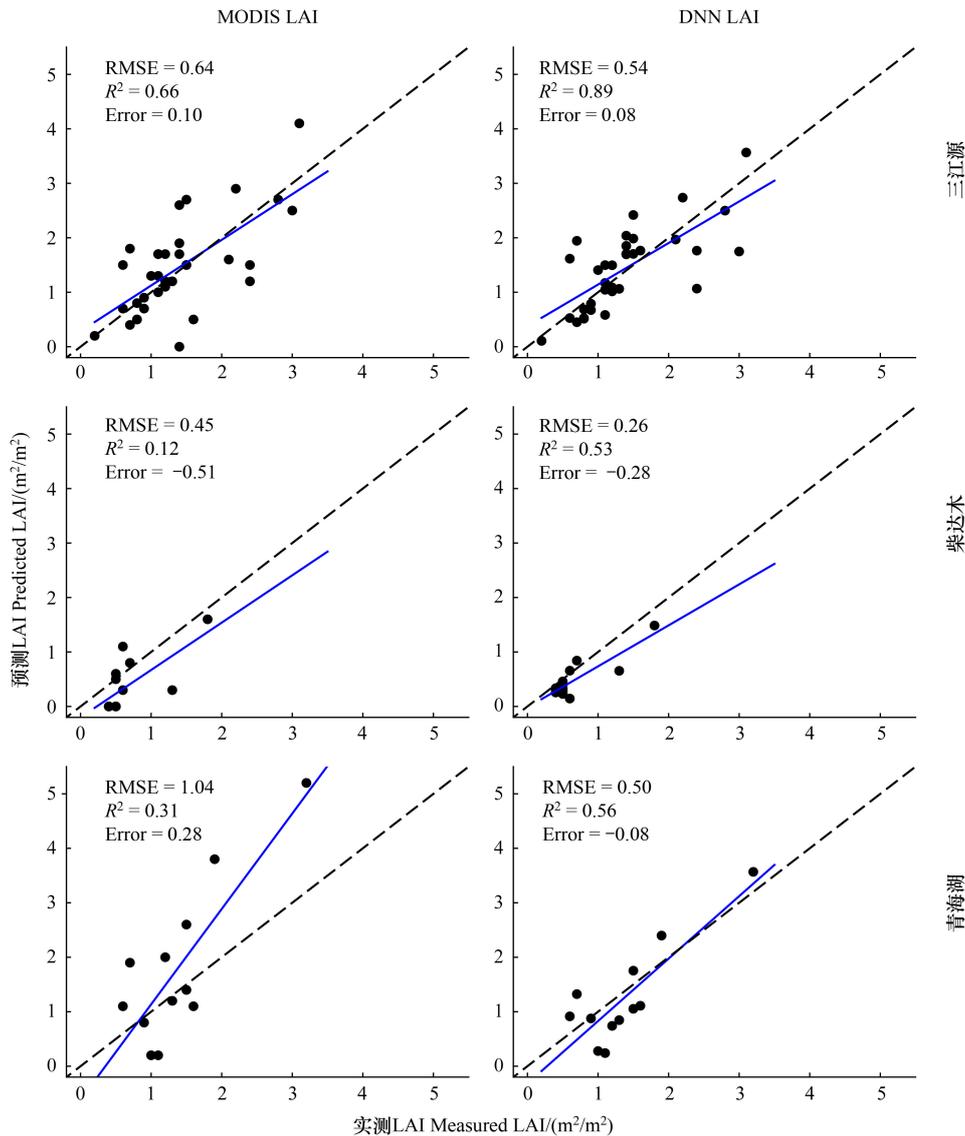


图 8 MODIS LAI 与 DNN LAI 的区域精度验证

Fig.8 Regional accuracy verification of MODIS LAI and DNN LAI

LAI 的相对误差水平,计算过程见下式。

$$\text{Error: } \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{\hat{y}_i} \tag{3}$$

式中, y_i 为 LAI 预测值, \hat{y} 为 LAI 真实值, n 为样点的数量。

三江源地区主要分布有中覆盖度的高寒草甸,该地区 DNN 改进模型预测得到的 LAI ($RMSE = 0.54, R^2 = 0.89$) 比 MODIS LAI 产品 ($RMSE = 0.64, R^2 = 0.66$) 更加准确,整体高估 Error 由 10% 降低到 8%。该地区整体存在的质量问题较少,典型区域三江源地区与玉树的影像结果中(图 9),由于三江源地区的整体质量未受影响 ($QC = 1$),改进前后差异不大,部分低估问题得到约束。玉树部分区域由于云层遮挡与饱和效应导致的 LAI 分布不连续问题得到改善, DNN LAI 结果整体分布更加聚集。

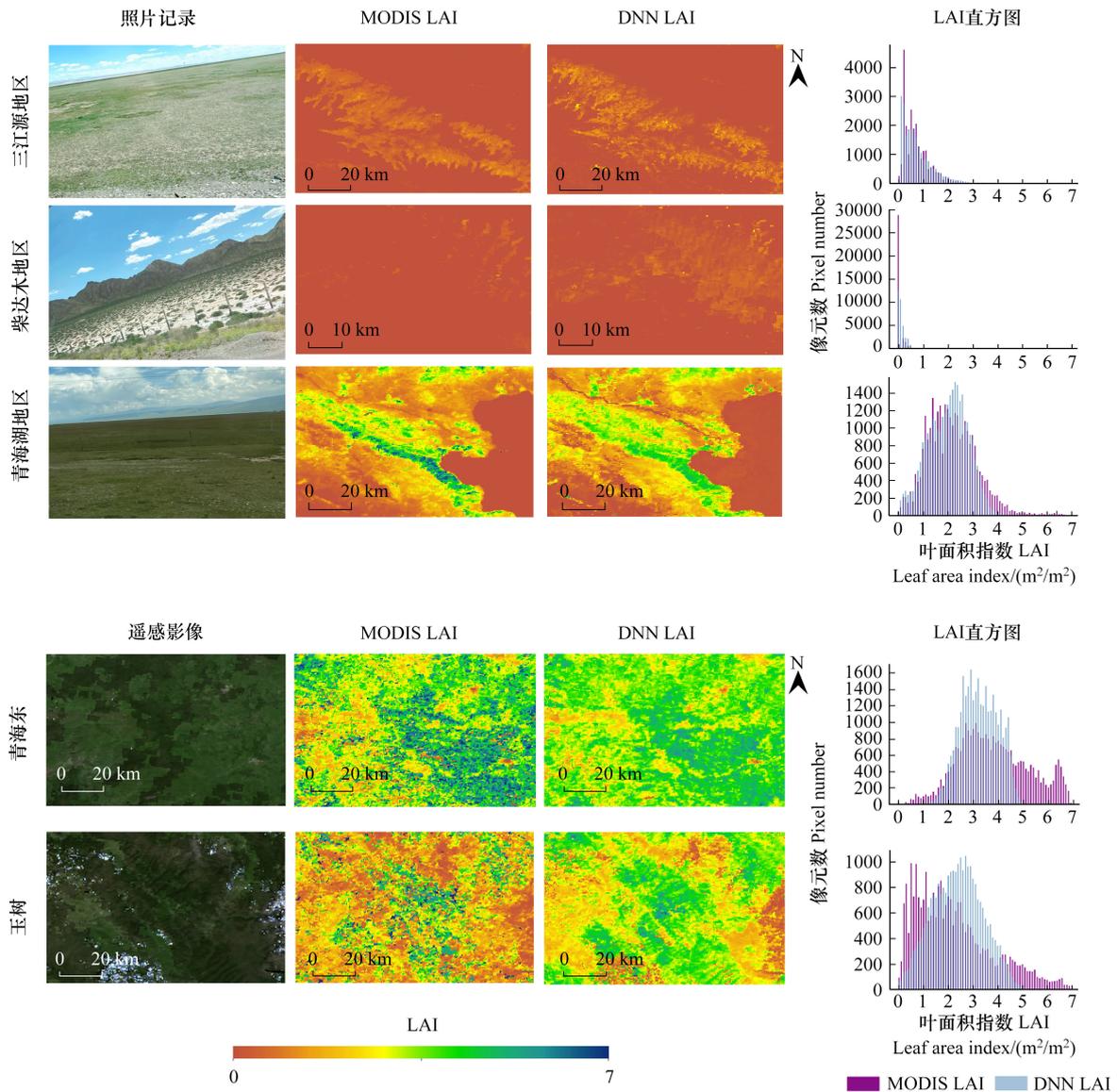


图 9 对比不同典型植被区域的改进结果

Fig.9 Comparison of improvement results in different typical vegetation regions

柴达木地区主要分布有低覆盖度的荒漠植被,该地区 DNN 改进模型预测得到的 LAI 准确度 ($RMSE = 0.26, R^2 = 0.53$) 显著高于 MODIS LAI 产品 ($RMSE = 0.45, R^2 = 0.12$), 低估问题由 51% 提高到 28%。该地区主

要分布的地表分类无法检索问题得到解决,典型区域的影像结果中 LAI=0 的像元显著减少,直方图总体分布于 0.3—0.7 范围的低覆盖植被程度。

青海湖地区主要分布有高覆盖度的牧场草地,该地区实测数据验证中 DNN LAI(RMSE=0.5, $R^2=0.56$) 预测性能高于 MODIS LAI(RMSE=1.04, $R^2=0.31$),整体高估 28% 的问题明显减少。该地区范围内的青海湖附近以及青海东两区域为受到饱和问题的典型区域,从影像结果中(图 9),可以看出饱和效应导致明显高估的杂乱像元得到解决。直方图中,饱和问题区域 LAI 大于 6 的高估像元显著减少,总体分布于 0—5,符合真实植被状况。

3 讨论

保证 MODIS LAI 产品质量是进行有效的大范围地表环境分析与评估的重要基础^[37-38]。在研究中发现 MODIS LAI 产品存在无法检索与饱和效应两大问题。无法检索问题导致地表裸露的低覆盖荒漠植被 LAI 未能预测,饱和问题导致高覆盖典型植被 LAI 因光谱过饱和出现明显的高估问题。为解决饱和问题与无法检索问题,减少 MODIS LAI 的高估与低估,本研究针对 MODIS LAI 产品中的三类不确定性因素进行改进。下文主要讨论改进方法的针对性与合理性:

地表分类数据的不确定性表现在低覆盖的荒漠植被区域被划分为非植被区域,使得 MODIS LAI 计算时未能进行检索。这是由于 MODIS LAI 中三维辐射传输模型需要考虑植被类别与光谱参数来确定 LAI 数值,而改进方法 PROSAIL 模型是从光学物理角度模拟植被生长特征,建立植被反射光谱与参数之间的物理机制^[34, 39]。该模型能够精确模拟冠层内光学辐射和热辐射的多重散射^[30],用于更好描述植被稀疏地区非均匀冠层的特征^[23, 40]。改进后的模型能够在不考虑植被类别的基础上,避免了低覆盖植被因地表分类不确定导致的 LAI 无法检索问题。有相关研究测试了 PROSAIL 模型在德国 EnMAP 数据中的适用性,认为该模型能够降低预测的不确定性,为未来参数反演提供巨大的潜力^[19, 41]。

反演算法的不确定性是由于 MODIS LAI 的 LUT 算法在复杂多样的模拟数据拟合过程中存在一定的局限性。由于植被饱和效应,使得光谱反射率与 LAI 之间存在复杂的非线性关系,LUT 线性拟合的精度与效率降低,导致出现高估问题。而学习能力更强的 DNN 算法可以更好的对复杂数据进行非线性拟合^[24, 41-42],能够在光谱特征中准确发现植被叶面积指数与光谱之间的饱和过程^[19]。通过植被指数与 LAI 的散点分布来表现光谱与 LAI 之间的关系(图 10),其中植被指数采用归一化植被指数(NDVI)。散点图中 MODIS LAI 像元分布散乱,在光谱特征相同的条件下,饱和像元的 LAI 出现明显高估。而对应的 DNN LAI 结果中的饱和像元分布

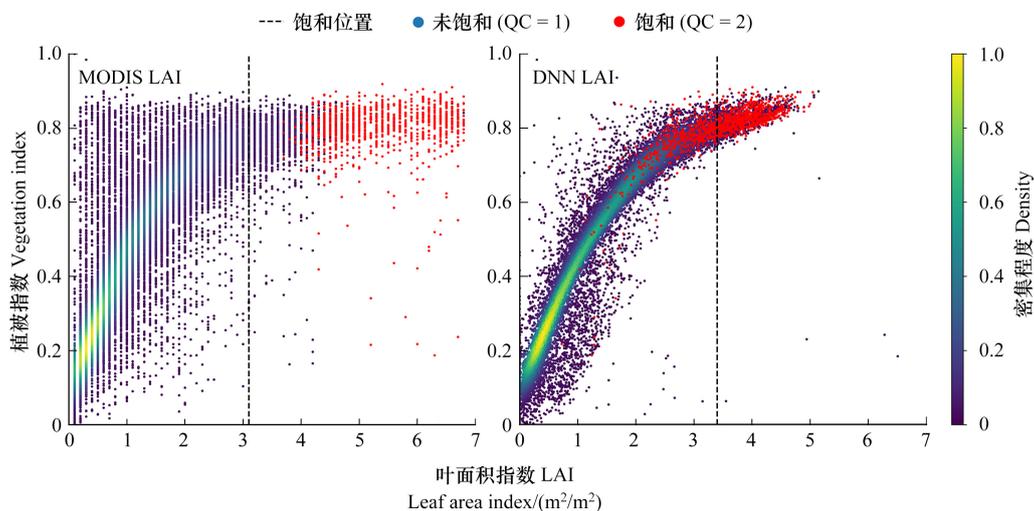


图 10 植被指数 NDVI 与叶面积指数 LAI 的关系散点图

Fig.10 Scatter plot of the relationship between vegetation index NDVI and leaf area index LAI

在饱和位置附近,整体拟合于未受影响的像元。这表明 DNN 算法对光谱与 LAI 之间的拟合程度更高,有效解决了 MODIS LAI 反演算法不确定导致饱和和效应高估问题。研究过程中对比了 SVR、RFR、GPR 机器学习算法的预测精度,DNN 模型具有最好的效果(表 2)。

对于地表反射率数据的不确定性,模型模拟与地表反射数据之间的信息差异也会影响 MODIS LAI 产品整体质量^[28]。为弥补这种差异,通过质量筛选与匹配生成互补数据集,利用遥感与植物生长模型的融合,实现优势互补、时空扩展,有助于改善因地表反射率数据不确定性导致的病态反演^[27-28]。在数据集训练对比中,互补数据集结果(表 2, RMSE=0.48, $R^2=0.64$)显著高于仅模拟数据集的结果(RMSE=0.83, $R^2=0.43$),互补信息的约束提高了反演过程的稳定性和合理性。

表 2 不同数据集下回归模型的性能对比

Table 2 Performance comparison of regression models under different datasets

模型 Model	互补数据集 Complementary dataset		模拟数据集 Simulated dataset	
	RMSE	R^2	RMSE	R^2
深度神经网络(DNN) Deep neural network	0.48	0.64	0.83	0.43
支持向量回归(SVR) Support vector regression	0.57	0.48	0.93	0.32
随机森林回归(RFR) Random forest regression	0.53	0.57	0.81	0.38
高斯过程回归(GPR) Gaussian process regression	0.52	0.60	0.88	0.41

综上所述,本文发现 MODIS LAI 产品在青海省内的低估问题集中于覆盖有荒漠稀疏植被的区域,高估问题主要分布在高覆盖度的草地和灌丛区域,高估值超出草地与灌丛 LAI 最大值。高估与低估情况与以往研究一致,存在高值区域高估、低值区域低估的现象。同时低估和高估问题分别与 MODIS LAI 质量等级中的未检索问题和饱和问题在空间分布上具有一致性(图 5),因此结合植被覆盖、地表反射等遥感数据分析发现由于地表分类数据的不确定性和反演算法的不确定性导致了未检索问题和饱和问题的出现。利用 PROSAIL 模型与 DNN 模型改进这两类不确定性问题,实验结果表明,DNN 模型更好的发现了植被的饱和现象,结果中饱和像元 LAI 与未受影响像元的达到一致(图 10)。PROSAIL 模型减弱了地表分类未检索的问题,低估与高估像元减少, RMSE 显著提高。整体实验过程中,地表反射率与模拟反射率之间的空间尺度异质性同样需要考虑,匹配建立的互补数据对此进行了很好的补充。针对 MODIS LAI 产品机制中的数据、模型、算法三方面进行改进,从改进结果验证分析与改进方法的对比讨论中可以证明本文研究方法是合理有效的,能够解决 MODIS LAI 中的具体质量问题,提高产品精度。

遥感应用产品的质量代表其是否能够准确反映地表真实状况,MODIS LAI 产品质量的保证对于大范围、多时相生态环境的建模与评估是必要的,高精度的 LAI 可以考虑代替 NDVI、FVC 作为监测与分析区域植被生长状况、植被物候的重要数据。同时本研究能够为大范围、缺少实测数据的多类型植被、农作物的 LAI 高精度遥感反演提供参考。基于本文研究内容,未来研究可以尝试将迁移学习方法应用在反演模型中,充分利用实测数据提高反演模型精度。

4 结论

MODIS LAI 产品作为大尺度、多时相生态环境建模的重要数据,其不同地区的质量问题导致存在明显的低估或高估现象^[8]。因此针对 MODIS LAI 存在的质量问题,从其生产过程的不确定性因素入手进行改进,同时考虑到了植被对时间的敏感变化,能够保证实时获取高精度的 MODIS LAI 产品。本文提出的一种基于 PROSAIL 模型与深度神经网络(DNN)的混合建模技术,不仅在训练数据上表现良好,而且可以在缺少实测样本的情况下进行有效的大范围反演。对于青海省地区,通过区域问题分析与实测数据验证,证明了改进方法有效地解决了 MODIS LAI 产品中因饱和效应导致的高估问题与因地表分类未检索导致的低估问题。与改进前 MODIS LAI 相比,改进后的 LAI 在整体精度和区域分布效果方面提供了更好的结果,能够作为一种通用的

改进方法获取更高质量的 MODIS LAI 产品为大范围生态过程建模与地区环境评估提供支持。

参考文献 (References):

- [1] Chen J M, Pavlic G, Brown L, Cihlar J, Leblanc S G, White H P, Hall R J, Peddle D R, King D J, Trofymow J A, Swift E, Van der Sanden J, Pellikka P K E. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sensing of Environment*, 2002, 80(1): 165-184.
- [2] Darvishzadeh R, Atzberger C, Skidmore A, Schlerf M. Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011, 66(6): 894-906.
- [3] 林杰, 潘颖, 杨敏, 佟光臣, 唐鹏, 张金池. 1988—2013 年基于 BP 神经网络的植被叶面积指数遥感定量反演. *生态学报*, 2018, 38(10): 3534-3542.
- [4] Hu Y G, Li H, Wu D, Chen W, Zhao X, Hou M L, Li A J, Zhu Y J. LAI-indicated vegetation dynamic in ecologically fragile region: a case study in the Three-North Shelter Forest program region of China. *Ecological Indicators*, 2021, 120: 106932.
- [5] Serbin S P, Ahl D E, Gower S T. Spatial and temporal validation of the MODIS LAI and FPAR products across a boreal forest wildfire chronosequence. *Remote Sensing of Environment*, 2013, 133: 71-84.
- [6] 孙晨曦, 刘良云, 关琳琳. 内蒙古锡林浩特草原 GLASS LAI 产品的真实性检验. *遥感技术与应用*, 2013, 28(6): 949-954.
- [7] 柳艺博, 居为民, 朱高龙, 陈镜明, 邢白灵, 朱敬芳, 周艳莲. 内蒙古不同类型草地叶面积指数遥感估算. *生态学报*, 2011, 31(18): 5159-5170.
- [8] Yang J L, Chen H Y, Borjigin N, Zhao M M, Zhou Y F, Huang Y M. Validation of the MODIS LAI product in Qinghai Lake Basin combined with field measurements using Landsat 8 OLI data. *Acta Ecologica Sinica*, 2017, 37(5): 322-331.
- [9] 孙晨曦, 刘良云, 关琳琳, 焦全军, 彭代亮. 锡林浩特草原区域 MODIS LAI 产品真实性检验与误差分析. *遥感学报*, 2014, 18(3): 518-536.
- [10] Yang W Z, Tan B, Huang D, Rautiainen M, Shabanov N V, Wang Y, Privette J L, Huemmrich K F, Fensholt R, Sandholt I, Weiss M, Ahl D E, Gower S T, Nemani R R, Knyazikhin Y, Myneni R B. MODIS leaf area index products: from validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(7): 1885-1898.
- [11] 方红亮. 我国叶面积指数卫星遥感产品生产及验证. *遥感技术与应用*, 2020, 35(5): 990-1003.
- [12] Ji J Y, Li X J, Du H Q, Mao F J, Fan W L, Xu Y X, Huang Z H, Wang J Y, Kang F F. Multiscale leaf area index assimilation for Moso bamboo forest based on Sentinel-2 and MODIS data. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 104: 102519.
- [13] 郭利彪, 刘桂香, 运向军, 张勇, 孙世贤. 基于数据机理的植被叶面积指数遥感反演研究. *遥感技术与应用*, 2020, 35(5): 1047-1056.
- [14] 潘颖, 丁鸣鸣, 林杰, 代侨, 郭赓, 崔琳琳. 基于 PROSAIL 模型和多角度遥感数据的森林叶面积指数反演. *林业科学*, 2021, 57(4): 90-106.
- [15] Wang Y J, Woodcock C E, Buermann W, Stenberg P, Voipio P, Smolander H, Häme T, Tian Y H, Hu J N, Knyazikhin Y, Myneni R B. Evaluation of the MODIS LAI algorithm at a coniferous forest site in Finland. *Remote Sensing of Environment*, 2004, 91(1): 114-127.
- [16] Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada P J, Asner G P, François C, Ustin S L. PROSPECT+SAIL models: a review of use for vegetation characterization. *Remote Sensing of Environment*, 2009, 113: S56-S66.
- [17] Danner M, Berger K, Wocher M, Mauser W, Hank T. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 173: 278-296.
- [18] Ou D P, Tan K, Lai J, Jia X P, Wang X, Chen Y, Li J. Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction. *Geoderma*, 2021, 385: 114875.
- [19] Li Y, Ren Y Z, Gao W L, Jia J D, Tao S, Liu X L. An enhanced spatiotemporal fusion method-Implications for DNN based time-series LAI estimation by using Sentinel-2 and MODIS. *Field Crops Research*, 2022, 279: 108452.
- [20] 刘凤, 曾永年. 近 16 年青海高原植被 NPP 时空格局变化及气候与人为因素的影响. *生态学报*, 2019, 39(5): 1528-1540.
- [21] 谢贤胜, 卢峰, 杨元征, 罗蔚生, 徐占勇, 苏宏新. 广西桉树人工林叶面积指数的估测及其校正. *生态学报*, 2022, 42(11): 4451-4462.
- [22] Myneni R, Park T, Yan K, Chen C, Xu B D, Knyazikhin Y. VIIRS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product algorithm theoretical basis document (ATBD). (2018-04-09) [2023-3-28]. https://viirsland.gsfc.nasa.gov/Products/NASA/LAI_FparESDR.html.
- [23] Liang L, Qin Z H, Zhao S H, Di L P, Zhang C, Deng M X, Lin H, Zhang L P, Wang L J, Liu Z X. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. *International Journal of Remote Sensing*, 2016, 37(13): 2923-2949.
- [24] Combal B, Baret F, Weiss M, Trubuil A, Macé D, Pragnère A, Myneni R, Knyazikhin Y, Wang L. Retrieval of canopy biophysical variables from

- bidirectional reflectance. *Remote Sensing of Environment*, 2003, 84(1): 1-15.
- [25] Jacquemoud S, Baret F. PROSPECT: a model of leaf optical properties spectra. *Remote Sensing of Environment*, 1990, 34(2): 75-91.
- [26] 王李娟, 牛铮. PROSAIL 模型的参数敏感性研究. *遥感技术与应用*, 2014, 29(2): 219-223.
- [27] 徐卫星, 薛华柱, 靳华安, 李爱农. 融合遥感先验信息的叶面积指数反演. *遥感技术与应用*, 2019, 34(6): 1235-1244.
- [28] Li X. Characterization, controlling, and reduction of uncertainties in the modeling and observation of land-surface systems. *Science China Earth Sciences*, 2014, 57(1): 80-87.
- [29] García-Haro F J, Campos-Taberner M, Muñoz-Marí J, Laparra V, Camacho F, Sánchez-Zapero J, Camps-Valls G. Derivation of global vegetation biophysical parameters from EUMETSAT Polar System. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 139: 57-74.
- [30] Fernández-Guisuraga J M, Verrelst J, Calvo L, Suárez-Seoane S. Hybrid inversion of radiative transfer models based on high spatial resolution satellite reflectance data improves fractional vegetation cover retrieval in heterogeneous ecological systems after fire. *Remote Sensing of Environment*, 2021, 255: 112304.
- [31] Schiefer F, Schmidlein S, Kattenborn T. The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. *Ecological Indicators*, 2021, 121: 107062.
- [32] Ai B, Wen Z, Jiang Y C, Gao S, Lv G N. Sea surface temperature inversion model for infrared remote sensing images based on deep neural network. *Infrared Physics & Technology*, 2019, 99: 231-239.
- [33] Knyazikhin Y, Myneni R B, Tian Y, Wang Y, Zhang Y. Estimation of vegetation canopy leaf area index and fraction of photosynthetically active radiation absorbed by vegetation from remotely sensed multi-angle and multi-spectral data. *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*. June 28-July 2, 1999, Hamburg, Germany. IEEE, 2002: 1872-1874.
- [34] Shabanov N V, Wang Y, Buermann W, Dong J, Hoffman S, Smith G R, Tian Y, Knyazikhin Y, Myneni R B. Effect of foliage spatial heterogeneity in the MODIS LAI and FPAR algorithm over broadleaf forests. *Remote Sensing of Environment*, 2003, 85(4): 410-423.
- [35] Huang Y, Liu H X, Yu B L, Wu J P, Kang E L, Xu M, Wang S J, Klein A, Chen Y N. Improving MODIS snow products with a HMRF-based spatio-temporal modeling technique in the Upper Rio Grande Basin. *Remote Sensing of Environment*, 2018, 204: 568-582.
- [36] Parker G G. Tamm review: leaf Area Index (LAI) is both a determinant and a consequence of important processes in vegetation canopies. *Forest Ecology and Management*, 2020, 477(1): 118496.
- [37] Propastin P, Erasmi S. A physically based approach to model LAI from MODIS 250m data in a tropical region. *International Journal of Applied Earth Observation and Geoinformation*, 2010, 12(1): 47-59.
- [38] Fang H L, Wei S S, Liang S L. Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sensing of Environment*, 2012, 119: 43-54.
- [39] Jacquemoud S, Verhoef W, Baret F, Zarco-Tejada P, Asner G, Francois C, Ustin S. PROSPECT SAIL: 15 years of use for land surface characterization. *2006 IEEE International Symposium on Geoscience and Remote Sensing. July 31-August 4, 2006, Denver, CO, USA. IEEE, 2007: 1992-1995.*
- [40] Verhoef W, Bach H. Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sensing of Environment*, 2007, 109(2): 166-182.
- [41] Berger K, Atzberger C, Danner M, D'Urso G, Mauser W, Vuolo F, Hank T. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: a review study. *Remote Sensing*, 2018, 10(2): 85.
- [42] Sinha S K, Padalia H, Dasgupta A, Verrelst J, Rivera J P. Estimation of leaf area index using PROSAIL based LUT inversion, MLRA-GPR and empirical models: case study of tropical deciduous forest plantation, North India. *International Journal of Applied Earth Observation and Geoinformation*, 2020, 86: 102027.