

张亦然,刘廷玺,童新,等.基于多源遥感和机器学习方法的科尔沁沙地植被覆盖度反演[J].中国沙漠,2022,42(3):187-195.

基于多源遥感和机器学习方法的科尔沁沙地植被覆盖度反演

张亦然^a,刘廷玺^{ab},童新^{ab},段利民^{ab},贾天宇^a,季亚新^a

(内蒙古农业大学 a.水利与土木建筑工程学院, b.内蒙古自治区水资源保护与利用重点实验室, 内蒙古 呼和浩特 010018)

摘要: 植被覆盖度是监测生态系统及其功能的关键参数,如何提高大区域植被覆盖度的反演精度,对生态脆弱区环境可持续发展至关重要。本研究基于人工神经网络、支持向量回归和随机森林等机器学习方法,利用无人机、Worldview-2与Landsat 8 OLI遥感数据,对科尔沁沙地植被覆盖度进行多尺度反演。结果表明:随机森林模型比人工神经网络、支持向量回归模型表现佳,可在单元(试验区)、区域(研究区)尺度上较高精度地反演沙地的植被覆盖度,反演值与无人机实测值均在线性水平上呈显著相关($P<0.01$);在单元、区域尺度上,构建的植被覆盖度反演模型测试集 R^2 分别为0.84、0.80, MSE 分别为0.0145、0.0370,一致性指数 d 分别为0.9576、0.8991。利用多源遥感数据和机器学习方法,通过局部区域的高精度反演逐步实现低空间分辨率遥感影像的大区域植被覆盖度反演,不仅可有效提高沙地植被覆盖度的反演精度($R^2=0.78$,大于0.63),也为区域生态环境监测与生态系统健康评价提供支持。

关键词: 植被覆盖度;多源遥感;机器学习;科尔沁沙地

文章编号: 1000-694X(2022)03-187-09

DOI: 10.7522/j.issn.1000-694X.2021.00149

中图分类号: Q948.1

文献标志码: A

0 引言

植被覆盖度(fractional vegetation cover, FVC)是反映植被生长状况的直接定量指标,是估算和监测生态系统及其功能的关键参数^[1-2],可用于揭示荒漠化地区生态系统中存在的植被退化、风沙侵蚀、过度放牧等问题^[3-5]。快速、有效、精确地估算植被覆盖度对了解短期内区域生态环境状况、促进资源环境的可持续发展和管理都具有重要意义。

当前,估算植被覆盖度的方法主要为地面实测调查和应用遥感技术^[6]。前者存在空间范围小、费时费力、易受人为因素影响等缺点^[7],相比之下,遥感技术可以在短时间内、高效获取大范围的地表植被覆盖度,尤其在下垫面复杂的地区,遥感技术的优势更加突出^[8]。其中,无人机(unmanned aerial vehicle, UAV)遥感作为航天遥感与地面调查的桥梁,用以弥补采样点与航天遥感像元间空间匹配问题,

同时,也为构建植被覆盖度反演模型提供大量样本^[9-10]。

近年来,机器学习方法在植被覆盖度反演研究中得到广泛应用^[11-12]。陈黔等^[13]基于机器学习方法对毛乌素沙地的灌木覆盖度进行了估算;Higginbottom等^[14]基于随机森林面模型和Landsat卫星构建了半干旱草原的植被覆盖度反演模型;Ge等^[15]运用MODIS数据和支持向量回归方法对高寒地区植被覆盖度进行了反演。然而,学者们大多基于单一遥感数据源进行植被覆盖度的估算反演^[16-17],具有显著局限性,尤其对于植被稀疏、破碎度较大的荒漠化地区,无法可靠、有效地反演区域植被覆盖度。另外,由于野外数据和卫星遥感数据的覆盖范围相差较大,模型构建精度得不到充分保证。

因此,本研究基于航天、无人机遥感影像并辅以地面调查,通过局部区域的高精度反演,逐步实

收稿日期:2021-09-06; 改回日期:2021-12-13

资助项目:国家自然科学基金项目(51620105003,51809141);内蒙古自然科学基金项目(2018ZD05,2018BS05001);教育部创新团队发展计划项目(IRT_17R60);科技部重点领域科技创新团队(2015RA4013);内蒙古自治区草原英才产业创新创业人才团队项目;内蒙古农业大学寒旱区水资源利用创新团队项目(NDTD2010-6)

作者简介:张亦然(1996—),女,内蒙古巴彦淖尔人,硕士研究生,主要从事生态环境遥感研究。E-mail: 534805685@qq.com

通信作者:刘廷玺(E-mail: txliu1966@163.com)

现基于低空间分辨率遥感数据的大区域植被覆盖度的精确估算。以期通过多源遥感数据和机器学习方法,对科尔沁沙地进行单元、区域尺度的反演,提高植被覆盖度的准确性,以及探寻机器学习模型的适用性,为该地区生态系统健康评价提供科学参考。

1 研究区与研究方法

1.1 研究区概况

研究区位于科尔沁沙地东南部,是一个内陆闭合流域,行政隶属于内蒙古自治区通辽市科尔沁左

翼后旗,为典型荒漠化的农牧交错区($42^{\circ}40'—43^{\circ}42'N$ 、 $122^{\circ}00'—123^{\circ}20'E$,海拔150—650 m,图1),总面积约8 000 km²。属于半干旱温带大陆性季风气候,受蒙古冷高压和太平洋暖低压消长变化影响,冬季多刮偏北风或偏西风,夏季则以偏南和西南风为主,年降水量300—400 mm,70%—80%发生于7—9月。植被呈斑块化分布,空间异质性较强,沙丘以沙生灌木、半灌木为主,草地以菊科、禾本科多年生植被为优势种^[18]。整个研究区由多个沙丘-草甸-湖泊相间的地貌单元组成(图1),在研究区内选取一典型单元作为试验区($43^{\circ}18'48''—43^{\circ}21'24''N$ 、 $122^{\circ}33'00''—122^{\circ}41'00''E$,图1),面积约55 km²。

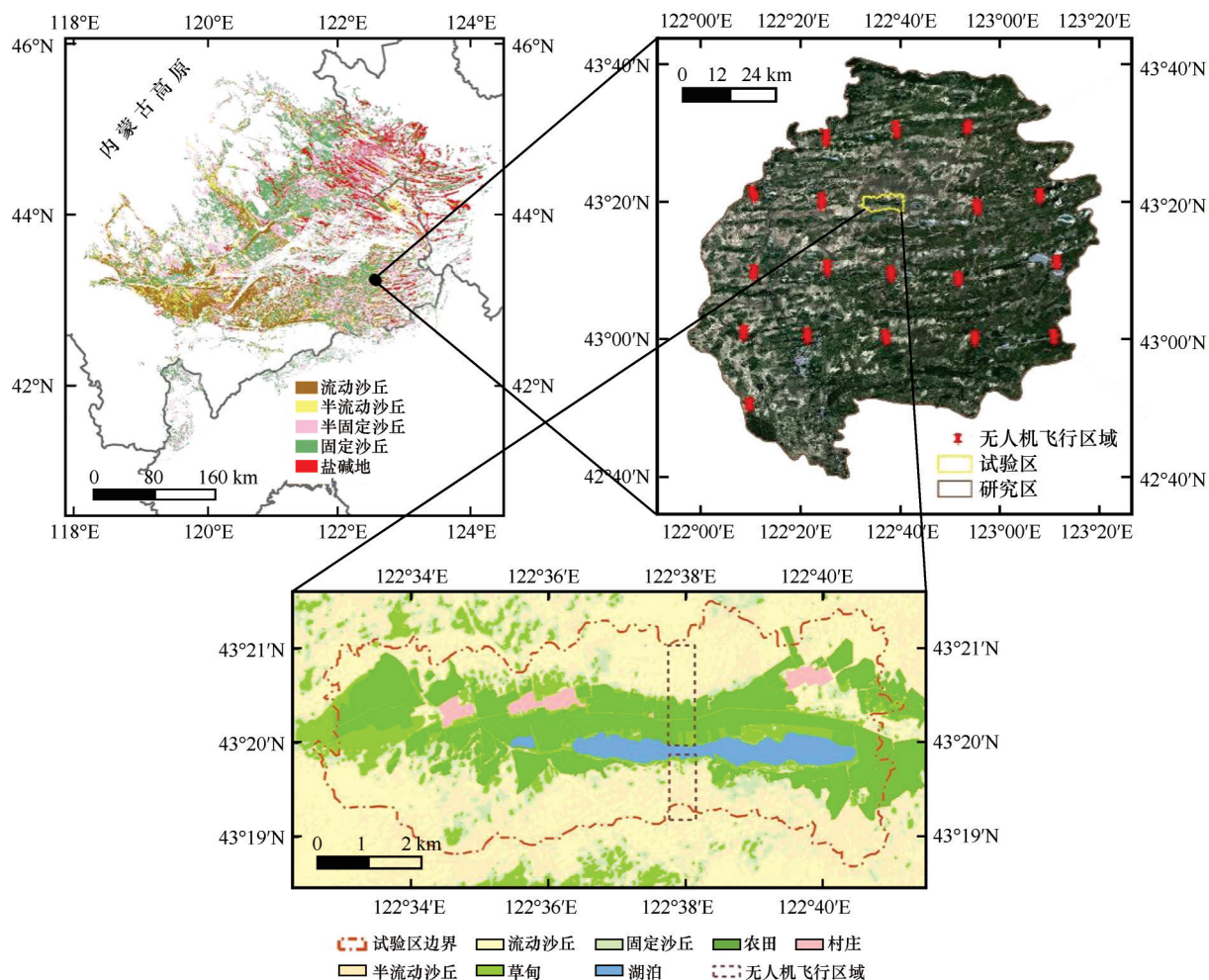


图1 研究区、试验区位置及无人机航拍区域

Fig.1 Location of research area, test area and UAV aerial photography area

1.2 数据获取与预处理

1.2.1 遥感数据

在试验区选取一条典型沙丘-草甸相间带(图1),于2019年7月28—31日使用无人机携带的可见

光(RGB)相机进行了影像拍摄;在研究区内以试验区为核心,经线每隔15'(约20.3 km),纬线每隔10'(约18.5 km),选取5条经线、5条纬线的交汇点(18个)进行无人机影像的拍摄。无人机型号为DJI

Phantom 4 Pro V2.0, 试验区设置飞行样地大小为北部(HB)500 m×2 000 m、南部(HN)500 m×1 500 m; 研究区设置飞行样地大小均为500 m×2 000 m。飞行高度80 m, 旁向重叠率65%, 航向重叠率80%。每个飞行样地采集图像800—850张。样地内均匀布设地面控制点(总误差<0.5 m)便于无人机正射影像地理配准矫正。然后, 利用Agisoft PhotoScan Professional软件进行照片匹配、建立密集点云、生成网格、纹理、数字高程、正射影像以及矫正地面控制点7项处理, 获得无人机飞行样带正射影像, 空间分辨率为0.03 m。

Worldview-2多光谱遥感影像于2019年7月30日获取, 包括0.5 m空间分辨率的全色波段和1.8 m分辨率的多光谱波段。其中, 新增的近红外2波段反射率受大气影响较少, 可增强植被特征。利用ENVI5.3软件对遥感影像进行辐射定标、大气校正、几何校正预处理。

Landsat 8 OLI多光谱遥感影像于2019年7月27日获取, 来源于中国科学院计算机网络信息中心(<http://www.Gscloud.cn>)。利用ENVI 5.3软件进行上述相同方法的预处理, 其空间分辨率为30 m。由于该景影像东南侧的部分被厚云遮挡, 运用ArcGIS 10.5对厚云及湖泊进行掩膜处理。

1.2.2 地面数据

地面调查工作于7月28日至8月1日进行。在无人机飞行的样带上共布置40个调查点(HB:25个, HN:15个), 为同卫星、无人机影像的像元大小相匹配, 每个调查点布设一个大-小嵌套式样方, 即先布设一个30 m×30 m的调查大样方(灌木), 然后在样方的四角和中心布设5个2 m×2 m(半灌木)或1 m×1 m(草本)的调查小样方。调查内容为植被的密度、冠幅(灌木)。对于分布规整的农田和乔木林做简单踩点调查。地面调查数据用于无人机影像植被覆盖度提取的阈值确定与验证。

1.3 研究方法

1.3.1 植被信息提取

鉴于无人机获取的正射影像与可见光波段差异植被指数(visible-band difference vegetation index, $VDVI$) 在植被与非植被间具有高分离度^[19], 本文选用 $VDVI$ 进行无人机影像植被覆盖度信息的提取。结合地面调查数据及目视解译, 经反复调整确定 $VDVI$ 的阈值为0.035, 即当 $VDVI>0.035$ 时, 代表

该区域为植被像元, 反之亦然。

$$VDVI = \frac{2 \times \rho_{\text{green}} - \rho_{\text{red}} - \rho_{\text{blue}}}{2 \times \rho_{\text{green}} + \rho_{\text{red}} + \rho_{\text{blue}}} \quad (1)$$

式中: ρ_{red} 、 ρ_{green} 、 ρ_{blue} 分别为红、绿、蓝3个波段对应的反射率。

1.3.2 遥感特征变量的选取

用于单元尺度(试验区)、区域尺度(研究区)模型构建的遥感变量包括Worldview-2、Landsat 8 OLI波段、主成分分析(Principal Component Analysis, PCA)的前3个分量($PCA1$ 、 $PCA2$ 、 $PCA3$)、窗口大小为3×3的纹理方差(Variance, Var)和归一化植被指数($NDVI$)。其中, 结合前人研究, 基于Worldview-2影像特有的近红外2(NIR_2)波段, 试验区模型构建的遥感变量增加了一个归一化植被指数2($NDVI_2$)^[20]。研究利用Python中sklearn库, 采用递归特征消除(recursive feature elimination, RFE)–交叉验证(cross validation, CV)的方法对遥感变量进行特征选择^[21]。具体步骤为: 在5折交叉验证的循环过程中, 计算每个特征量的重要性, 移除当前最不重要的特征, 不断重复递归, 直至找到最优的特征数量与特征变量, 在此过程中, 特征变量被消除的次序就是特征的排序, 特征的重要性值越大则输入特征的重要性程度越高^[22]。

纹理方差Var计算公式:

$$Var = \sum_{i,j=0}^{N-1} i \cdot P_{i,j} \left(i - \sum_{i,j=0}^{N-1} i \cdot P_{i,j} \right)^2 \quad (2)$$

式中: $P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$, $V_{i,j}$ 为第*i*行第*j*列位置上的像元

值; N 为计算纹理测量度时移动的串口大小, 本研究取 $N=3$ 。记 $Var(1 \cdots 7/8)$ 为Landsat 8 OLI/Worldview-2遥感影像各波段对应的Var值。

$NDVI$ 和 $NDVI_2$ 的计算公式如下:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3)$$

$$NDVI_2 = \frac{NIR_2 - R}{NIR_2 + R} \quad (4)$$

式中: NIR 、 R 分别为Worldview-2、Landsat 8 OLI影像中近红外波段、红色波段; NIR_2 为Worldview-2影像中近红外2波段。

1.3.3 模型构建

本研究选用人工神经网络(back propagation neural network, BP-ANN)、支持向量回归(support vector regression, SVR)与随机森林(random forest,

RF)等机器学习算法构建模型。其中,对于SVR和RF算法,80%的样本作为训练集,其余20%的样本作为测试集;对于BP-ANN算法,80%的样本作为训练集,10%的样本作为测试集,另外10%的样本作为验证集。

人工神经网络算法具有较强的柔韧性、可塑性和函数逼近能力,主要结构包括输入层、隐含层、输出层^[23]。隐含层和神经元的数量决定学习精度的高低和泛化能力的强弱,本研究根据试错过程确定其最优数量,并选用Levenberg-Marquardt函数对所建模型进行神经网络训练。利用MATLAB语言Neural Network工具箱实现BP-ANN模型的构建与验证。

支持向量回归算法在非线性及高维模式识别问题中具有优势,其高维空间中构建一组超平面,在超平面上,利用核函数对样本进行非线性到线性的转换,并选择最优线性分离器完成回归和其他任务^[24]。本研究以径向基函数(RBF)为核函数,利用MATLAB语言libsvm工具箱中的网格搜索法确定重要参数,完成SVR模型的构建与验证。

随机森林算法泛化能力强,不会过度拟合,性能稳定。其基于bootstrap抽样,每个训练样本用于构建决策树(classification and regression trees, CART),按照节点不纯度最小化原则不断被选择,直到形成一个完整的树。当形成多个决策树时,通过投票来获得最终的预测结果^[25]。本研究利用MATLAB语言RF-MexStandalone-v0.02工具箱中的RF函数进行模型的构建与验证。

1.4 模型评价指标

模型精度评价采用决定系数(coefficient of de-

termination, R^2)、均方误差(mean square error, MSE)和一致性指数(index of agreement, d)。当 R^2 、 d 越接近1, MSE 越小时,表示模型精度越高。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$d = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}_i| + |y_i - \bar{y}_i|)^2} \quad (7)$$

式中: y_i 为实测值; \hat{y}_i 为模型预测值; \bar{y}_i 为实测值的平均值; n 为样本数。

2 结果与分析

2.1 遥感特征向量的选择

运用递归特征消除-5折交叉验证法对Worldview-2、Landsat 8 OLI遥感影像提取的变量进行特征选择(图2),对于Worldview-2遥感影像,优选出6个最优特征变量,分别为 $NDVI$ 、 $NDVI_2$ 、蓝波段($B2$)、绿波段($B3$)、红波段($B5$)和绿波段的纹理方差($Var3$),其余变量的重要程度依次减弱;对于Landsat 8 OLI遥感影像, $NDVI$ 、 $PCA1$ 、绿波段($B3$)、近红外波段($B5$)和短波红外2波段的纹理方差($Var7$)5个特征变量的重要性最高,均优于其他特征。因此,本文选取Worldview-2遥感影像的 $NDVI$ 、 $NDVI_2$ 、 $B2$ 、 $B3$ 、 $B5$ 、 $Var3$ 特征变量进行单元尺度(试验区)的植被覆盖度反演模型的建立;选取Landsat 8 OLI遥感影像的 $NDVI$ 、 $PCA1$ 、 $B3$ 、 $B5$ 、 $Var7$

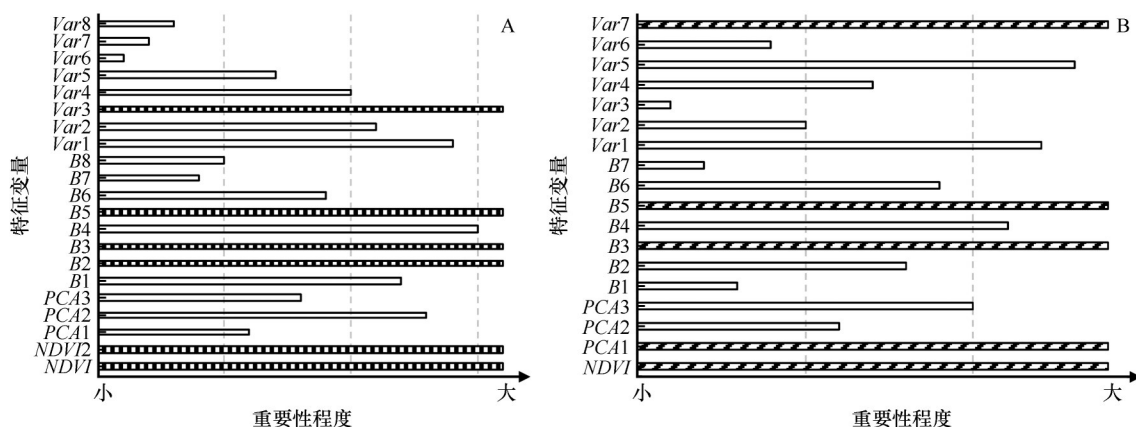


图2 Worldview-2(A)、Landsat8 OLI(B)遥感影像特征变量选择

Fig.2 Variables selection for Worldview-2 (A) and Landsat 8 OLI (B)

特征变量进行区域尺度(研究区)的植被覆盖度反演模型的构建。

2.2 单元尺度(试验区)植被覆盖度反演

将基于无人机正射影像提取的样带植被覆盖度信息与基于 Worldview-2 遥感影像进行叠加,运用 ArcGIS 软件的 Zonal Statistics 工具,统计无人机正射影像各对应格网单元内的植被覆盖度(图3)。共提取 270 个(HB:160 个;HN:110 个)样本点用于单元尺度(试验区)植被覆盖度反演模型的建立与验证。

模型的输入变量为样本点对应的无人机影像提取的植被覆盖度值与 Worldview-2 遥感影像的 6 个特征值($NDVI$ 、 $NDVI_2$ 、 B_2 、 B_3 、 B_5 、 $Var3$)。基于人工神经网络、支持向量回归和随机森林算法分别构建了植被覆盖度反演模型,其评价结果示于表 1。由表 1 可知,在整个数据集下各项精度评价指标,随机森林模型均表现最佳,训练集、测试集 R^2 分别为 0.94、0.84, MSE 分别为 0.0054、0.0145, d 均达到 0.95 以上(0.9838、0.9576),支持向量回归模型次之,人工神经网络模型最低。因此,本文选取随机森林模型对整个试验区的植被覆盖度进行反演估算(图4)。

进一步运用无人机遥感影像提取的 270 个样本

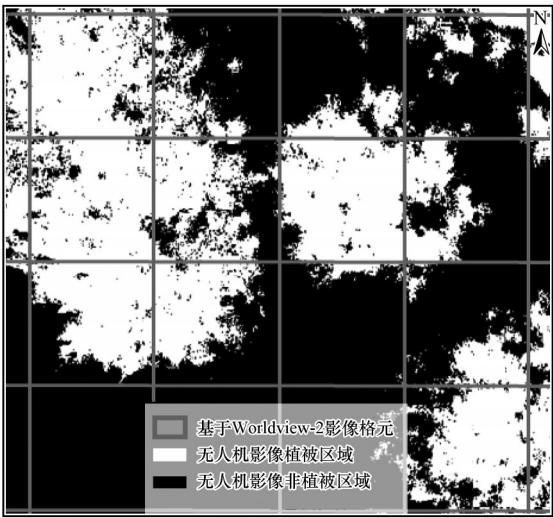


图3 以 Worldview-2 遥感影像格网为统计单元计算植被覆盖度

Fig.3 Using Worldview-2 remote sensing image grid as statistical unit to calculate vegetation coverage

点对反演的植被覆盖度做精度验证,基于 world-view-2 遥感影像构建的随机森林模型反演的植被覆盖度与基于无人机影像提取的植被覆盖度具有较好的相关性($P<0.01$)。空间上,试验区北部多为低覆盖度的流动沙丘、紧邻村庄和湖附近则为高覆盖度的农田和草甸;南部为中覆盖度的半固定、固定

表 1 单元尺度(试验区)不同模型精度评估

Table 1 Accuracy evaluation of different models in unit scale (the test area)

方法	训练集			测试集		
	决定系数 R^2	均方差 MSE	一致性指数 d	决定系数 R^2	均方差 MSE	一致性指数 d
BP-ANN	0.76	0.0816	0.9302	0.73	0.0886	0.9217
SVR	0.81	0.0706	0.945	0.80	0.0831	0.9401
RF	0.94	0.0054	0.9838	0.84	0.0145	0.9576

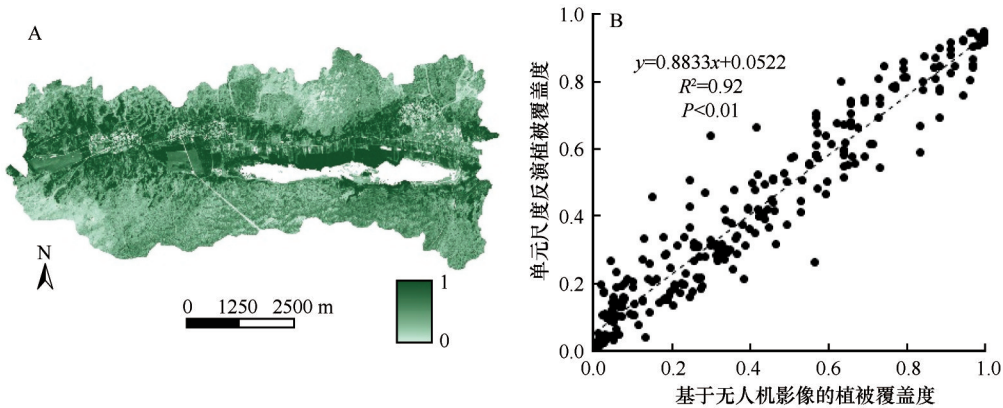


图4 Worldview-2 遥感影像反演的试验区植被覆盖度(A)和精度验证(B)

Fig.4 Vegetation coverage (A) and accuracy verification (B) of Worldview-2 remote sensing image inversion in the test area

沙丘,与野外实况相符。

综上所述,随机森林算法可适用于单元尺度(试验区)植被覆盖度的反演,构建的随机森林模型具有较好的精度,且从其反演的植被覆盖度中选取样本进行区域尺度(研究区)的植被覆盖度模型构建与反演是可靠的。

2.3 区域尺度(研究区)植被覆盖度反演

将单元尺度反演的植被覆盖度进行升尺度转换:重采样形成空间分辨率为 30 m 的影像信息,以便与 Landsat 8 OLI 遥感影像相匹配。鉴于 Worldview-2 遥感影像反演的植被覆盖度具有较高的精度,因此从单元尺度反演的植被覆盖度中随机选取 500 个样本点(剔除 4 个无效样本点)用于区域尺度(研究区)植被覆盖度反演模型的建立与验证。其中模型的输入变量为样本点对应的单元尺度反演植被覆盖度值与 5 个特征变量($NDVI$ 、 $PCA1$ 、 $B3$ 、 $B5$ 、 $Var7$)值。由于随机森林算法在单元尺度(试验区)植被覆盖度模型构建中表现最佳,因此,本研究

基于随机森林算法构建了区域尺度(研究区)植被覆盖度反演模型,同时与人工神经网络、支持向量回归算法进行比较。其评价结果示于表 2。在整个数据集下,各项精度评价指标中,随机森林模型较支持向量回归模型、人工神经网络模型也均表现最佳, R^2 分别为 0.87(训练集)、0.80(测试集), MSE 分别为 0.0082(训练集)、0.0370(测试集), d 分别为 0.9595(训练集)、0.8991(测试集)。因此选用随机森林模型对整个研究区的植被覆盖度进行反演估算,结果如图 5 所示。

表 2 区域尺度(研究区)不同模型精度评估
Table 2 Accuracy evaluation of different models in regional scale (the research area)

方法	训练集			测试集		
	R^2	MSE	d	R^2	MSE	d
BP-ANN	0.52	0.1599	0.8113	0.49	0.1959	0.7869
SVR	0.68	0.1447	0.8322	0.61	0.1503	0.8121
RF	0.87	0.0082	0.9595	0.80	0.0370	0.8991

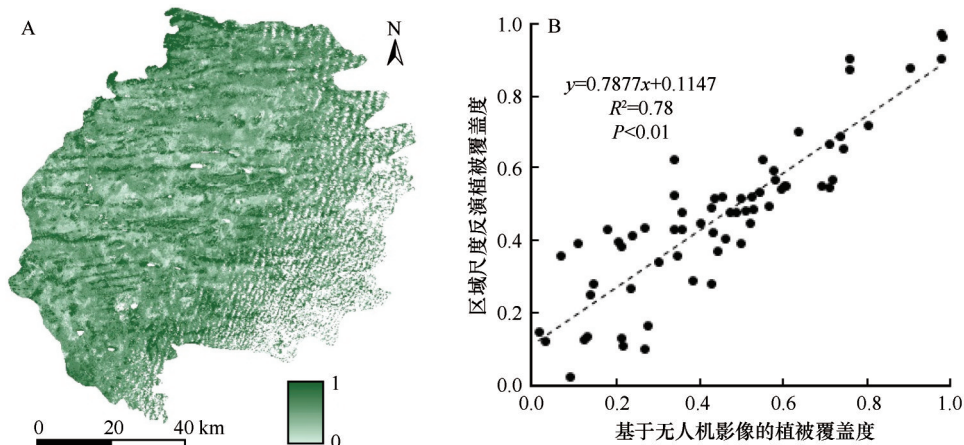


图 5 Landsat 8 OLI 遥感影像反演的研究区植被覆盖度(A)和精度验证(B)

Fig.5 Study area vegetation coverage (A) and accuracy verification (B) of Landsat 8 OLI remote sensing image inversion

将基于 18 个样带提取的无人机正射影像与基于 Landsat 8 OLI 遥感影像进行叠加,统计无人机正射影像各对应格元内的植被覆盖度。共提取 65 个(每样带 3—4 个)样本点用于区域尺度(研究区)植被覆盖度反演模型的验证。整体而言,基于随机森林模型反演的研究区植被覆盖度与无人机影像提取的植被覆盖度具有较高的相关性, R^2 为 0.78($P<0.01$),误差较低,反演结果相对稳定。空间上,研究区植被覆盖度呈高—中—低特征化单元相间分布,高植被覆盖区多为条带状,广泛分布于湖泊、村庄附近,中、低植被覆盖区占比较大,其中,研究区南部

植被覆盖度较北部高。

为进一步验证基于多源遥感数据和机器学习方法,逐步实现区域尺度植被覆盖度的反演精度,使用单一 Landsat 8 OLI 遥感影像进行植被覆盖度反演进行对比分析(图 6)。对比图 5、6 可知,本研究反演的结果精度较单一使用 Landsat 8 OLI 遥感影像有了显著提高($R^2=0.78$,大于 0.63)。

3 讨论

植被指数能够较好地反映绿色植被的相对丰

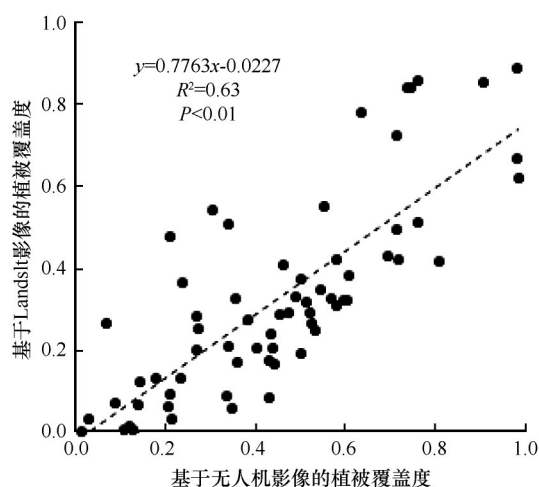


图6 使用单一遥感影像(Landsat 8 OLI)反演的植被覆盖度精度验证

Fig.6 Accuracy verification of vegetation coverage inversion using a single remote sensing image (Landsat 8 OLI)

富度^[26],在Worldview-2、Landsat 8 OLI遥感影像特征向量选择中,*NDVI*重要性均表现最高;由于植被冠层或叶片所含叶绿素对绿色的强反射作用,荒漠植被具有显著的“绿峰”特征^[27],故特征向量绿波段也均被优选出用于反演模型的构建;纹理方差是遥感影像灰度空间分布模式上提取的方差值,可进一步反映植被与非植被间的差异性^[28],作为优选的特征变量有助于植被覆盖度反演模型的建立。

在对植被覆盖度进行反演估算时,多源遥感数据的应用比单一遥感源数据更为合理。基于单一遥感源所构建的模型精度较低,稳定性较差,且具有较大的空间异质性^[29]。以往的研究表明,在下垫面复杂、植被斑块化显著的区域,直接应用低空间分辨率遥感影像进行植被覆盖度的估算反演,所得到的结果与实际情况往往存在较大偏差^[30]。此外,像元混合问题无法有效减弱^[31]。因此,本研究基于多源遥感数据(无人机、高、低空间分辨率遥感影像(Worldview-2、Landsat 8 OLI)),通过局部区域的高精度反演,弥补了空间尺度带来的偏差性,实现大区域的植被覆盖度高精度反演。如图5、6所示,反演结果的精度较单一使用Landsat 8 OLI遥感影像有了显著提高($R^2=0.78$,大于0.63)。

研究结果表明,基于随机森林算法较人工神经网络、支持向量回归算法构建的模型具有较好的反演精度,测试集精度(单元尺度 $R^2=0.84$,区域尺度 $R^2=0.80$)近似于已有的相关研究^[32-34],在一定程度上说明该模型具有较强的适用性。但在低植被覆

盖度(小于0.3)区间内^[35],应用随机森林算法反演的区域尺度(研究区)植被覆盖度值存在部分高估现象(图5),造成这种情况发生的原因可能有:一是本研究选取的样本点均在无人机飞行区域内,区域尺度(研究区)上植被覆盖度的极大、极小反演值超出了样本范围,进而产生一定误差,且样本点选取的数量多少对结果也具有一定的影响;二是在使用多源遥感数据进行植被要素反演时,本研究采用的重采样方法为局部平均法,该方法具有过度平滑的缺点,使大部分低植被覆盖度的区域在尺度转换后的相邻像元中都有一定比例的植被覆盖度,造成区域尺度反演的植被覆盖度值被高估。但整体而言,随机森林模型还是具有较好的鲁棒性和稳定性的,且可有效提高大区域植被覆盖度的反演精度。

4 结论

本文基于多源遥感数据,利用机器学习方法,实现了科尔沁沙地植被覆盖度的多尺度反演,为沙地植被恢复提供有效方法。

随机森林算法较人工神经网络、支持向量回归算法所构建的植被覆盖度反演模型更优,其中,单元尺度(试验区)模型 $R^2=0.84$ 、 $MSE=0.0145$ 、 $d=0.9576$;区域尺度(研究区)模型 $R^2=0.80$ 、 $MSE=0.0370$ 、 $d=0.8991$ 。随机森林模型更适用于荒漠化地区植被覆盖度的准确提取。

基于多源遥感数据和随机森林模型,通过局部高精度反演逐步实现区域低空间分辨率遥感影像植被覆盖度的反演方法,可以有效提高区域尺度植被覆盖度反演的精度($R^2=0.78$,大于0.63)。

本研究仅对一种植被要素(植被覆盖度)进行反演,然而,反映区域生态环境现状的参变量还有很多,如叶面积指数、地上生物量等,未来也应进一步从多参量协同解译、反演考虑,实现对区域生态系统全方面、多角度的科学评估。

参考文献:

- [1] Feng L L, Jia Z Q, Li Q X, et al. Spatiotemporal change of sparse vegetation coverage in Northern China[J]. Journal of the Indian Society of Remote Sensing, 2019, 47(2): 18-26.
- [2] Feng H, Zou B, Luo J. Coverage-dependent amplifiers of vegetation change on global water cycle dynamics[J]. Journal of Hydrology, 2017, 550: 220-229.
- [3] Gao J, Liu Y S. Determination of land degradation causes in

- Tongyu County, Northeast China via land cover change detection[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2009, 12(1): 9–16.
- [4] 宋超, 余琦殷, 王瑞霞, 等. 基于植被覆盖度的宁夏灵武白芨滩自然保护区防风固沙功能时空变化研究[J]. *生态学报*, 2021, 41(8): 1–13.
- [5] Zhou Q, Wei X, Zhou X, et al. Vegetation coverage change and its response to topography in a typical karst region: the Lianjiang River Basin in Southwest China[J]. *Environmental Geology*, 2019, 78(6): 191–201.
- [6] 万红梅, 李霞, 董道瑞. 基于多源遥感数据的荒漠植被覆盖度估测[J]. *应用生态学报*, 2012, 23(12): 3331–3337.
- [7] Curran P J, Williamson H D. Sample size for ground and remotely sensed data[J]. *Remote Sensing of Environment*, 1986, 20(1): 31–41.
- [8] Meng B P, Gao J L, Liang T G, et al. Modeling of alpine grassland cover based on unmanned aerial vehicle technology and multi-factor methods: a case study in the East of Tibetan Plateau, China[J]. *Remote Sensing*, 2018, 10(2): 320–339.
- [9] 刘婵, 赵文智, 刘冰, 等. 基于无人机和MODIS数据的巴丹吉林沙漠植被分布特征与动态变化研究[J]. *中国沙漠*, 2019, 39(4): 92–102.
- [10] 蔡宗磊, 苗正红, 常雪, 等. 基于无人机大样方数据及国产卫星反演草地植被覆盖度方法研究[J]. *草地学报*, 2019, 27(5): 1431–1440.
- [11] 贾坤, 姚云军, 魏香琴, 等. 植被覆盖度遥感估算研究进展[J]. *地球科学进展*, 2013, 28(7): 774–782.
- [12] Jia K, Liang S L, Gu X F, et al. Fractional vegetation cover estimation algorithm for Chinese GF-1 wide field view data[J]. *Remote Sensing of Environment*, 2016, 177: 184–191.
- [13] 陈黔, 李晓松, 修晓敏, 等. 基于Google Earth Engine与机器学习的大尺度30 m分辨率沙地灌木覆盖度估算[J]. *生态学报*, 2019, 39(11): 4056–4069.
- [14] Higginbottom T P, Symeonakis E, Meyer H, et al. Mapping fractional woody cover in semi-arid savannahs using multi-seasonal composites from Landsat data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 139: 88–102.
- [15] Ge J, Meng B P, Liang T G, et al. Modeling alpine grassland cover based on MODIS data and support vector machine regression in the headwater region of the Huanghe River, China[J]. *Remote Sensing of Environment*, 2018, 218(9): 162–173.
- [16] 刘垚霖, 曾鹏, 张然, 等. 基于GEE和BRT的1984–2019年长三角生态绿色一体化发展示范区植被覆盖度变化[J]. *应用生态学报*, 2021, 32(3): 1033–1044.
- [17] 张亦然, 刘廷玺, 董新, 等. 基于U型神经网络的沙丘–草甸相间地区无人机影像植被覆盖度提取及其影响因素[J]. *中国沙漠*, 2021, 41(3): 16–24.
- [18] Zuo X A, Zhao X Y, Zhao H L, et al. Scale dependent effects of environmental factors on vegetation pattern and composition in Horqin Sandy Land, Northern China[J]. *Geoderma*, 2012, 174(8): 1–9.
- [19] 汪小钦, 王苗苗, 王绍强, 等. 基于可见光波段无人机遥感的植被信息提取[J]. *农业工程学报*, 2015, 31(5): 152–158.
- [20] 张泽民, 吕昌河, 谢苗苗, 等. 基于WorldView-2影像的矿区植被重建效果评估[J]. *生态学报*, 2018, 38(4): 1301–1310.
- [21] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 28–35.
- [22] 朱婉雪, 孙志刚, 李彬彬, 等. 基于无人机遥感的滨海盐碱地土壤空间异质性分析与作物光谱指数响应胁迫诊断[J]. *地球信息科学学报*, 2021, 23(3): 536–549.
- [23] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533–536.
- [24] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273–297.
- [25] Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [26] Li M Y, Liu T X, Luo Y Y, et al. Fractional vegetation coverage downscaling inversion method based on land remote-sensing satellite (System, Landsat-8) and polarization decomposition of Radarsat-2[J]. *International Journal of Remote Sensing*, 2021, 42(9): 3255–3276.
- [27] 王惠宁, 靳瑰丽, 范燕敏, 等. 不同盖度下伊犁绢蒿荒漠草地光谱特征及盖度反演精度研究[J]. *中国草地学报*, 2019, 41(2): 51–57.
- [28] Wang C, Du H Q, Xu X J, et al. Multi-scale crown closure retrieval for moso bamboo forest using multi-source remotely sensed imagery based on geometric-optical and Erf-BP neural network models[J]. *International Journal of Remote Sensing*, 2015, 36(21): 5384–5402.
- [29] Wang H, Mu Y, Jiang L. Landscape-level vegetation classification and fractional woody and herbaceous vegetation cover estimation over the dryland ecosystems by unmanned aerial vehicle platform[J]. *Agricultural and Forest Meteorology*, 2019, 278: 107–128.
- [30] Leprieux C, Kerr Y H, Mastorchio S, et al. Monitoring vegetation cover across semi-arid regions: comparison of remote observations from various scales[J]. *International Journal of Remote Sensing*, 2000, 21(2): 281–300.
- [31] 古丽·加帕尔, 陈曦, 包安明. 干旱区荒漠稀疏植被覆盖度提取及尺度扩展效应[J]. *应用生态学报*, 2009, 20(12): 2925–2934.
- [32] Tang L, He M, Li X. Verification of fractional vegetation coverage and NDVI of desert vegetation via UAVRS technology[J]. *Remote Sensing*, 2020, 12(11): 1742–1754.
- [33] 王光镇, 王静璞, 邹学勇, 等. 基于像元三分模型的锡林郭勒草原光合植被和非光合植被覆盖度估算[J]. *生态学报*, 2017, 37(17): 5722–5731.
- [34] Zhang W, Yang X, Manlike A, et al. Comparative study of remote sensing estimation methods for grassland fractional vegetation coverage: a grassland case study performed in Hi prefecture, Xinjiang, China[J]. *International Journal of Remote Sensing*

ing, 2019, 40(6): 2243–2258.

及其生态效应评估[J]. 自然资源学报, 2019, 34(8): 1748–

[35] 张文强, 孙从建, 李新功. 晋西南黄土高原区植被覆盖度变化

1758.

Inversion of vegetation coverage based on multi-source remote sensing data and machine learning method in the Horqin Sandy Land, China

Zhang Yiran^a, Liu Tingxi^{ab}, Tong Xin^{ab}, Duan Limin^{ab}, Jia Tianyu^a, Ji Yaxin^a

(a. College of Water Conservancy and Civil Engineering / b. Inner Mongolia Key Laboratory of Protection and Utilization of Water Resource, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: Fractional vegetation coverage is a key parameter for monitoring ecosystems and its functions. How to improve the retrieval accuracy of fractional vegetation coverage in large areas is very important for the sustainable development of the environment in ecologically fragile areas. Based on machine learning methods such as back propagation neural network (BP-ANN), support vector regression (SVR) and random forest (RF), this study uses UAV, Worldview-2 and Landsat 8 OLI remote sensing data to carry out multi-scale inversion of fractional vegetation coverage in Horqin Sandy Land. The results show that: (1) The RF model performs better than the BP-ANN and SVR model. It can invert the vegetation coverage of sandy land with higher accuracy on the scale of unit (experimental area) and regional (research area). There was a significant correlation between the inversion value and the measured value of UAV at the linear level (P is less than 0.01). (2) On the unit and regional scales, the test set of the constructed vegetation coverage inversion model are R^2 of 0.84 and 0.80, MSE of 0.0145 and 0.0370, and the consistency index d of 0.9576 and 0.8991 respectively. The method of gradually realizing the large area fractional vegetation coverage inversion of low spatial resolution remote sensing images by using multi-source remote sensing data and machine learning method, which can not only effectively improve the inversion accuracy of fraction vegetation coverage in sandy land ($R^2=0.78$, less than 0.63), but also provide support for regional ecological environment monitoring and ecosystem health assessment.

Key words: fractional vegetation cover; multi-source remote sensing; machine learning method; Horqin Sandy Land