

结合植被覆盖度指数的土地覆盖分类方法研究 ——以巴基斯坦信德地区为例^{*}

孙娜 赵祥[†] 穆宝慧 赵嘉诚 刘乃精

(北京师范大学地理科学学部, 遥感科学国家重点实验室, 北京市陆表遥感数据产品工程技术研究中心, 100875, 北京)

摘要 基于 Landsat 8 OLI 反射率数据, 结合定量遥感反演植被覆盖度(fractional vegetation cover, FVC)提取的植被物候特征数据, 对比了神经网络、支持向量机和随机森林 3 种土地覆盖分类方法。结果表明: 随机森林分类方法具有较好的结果, 反射率结合植被特征数据的分类方法的总体精度为 85.52%, Kappa 系数为 0.821 2, 比仅用反射率的土地覆盖分类总体精度提高了 3.45 百分点, Kappa 系数提高 0.042 9; 植被覆盖度提取的植被特征数据能有效改善耕地、草地和裸地的制图精度和用户精度, 对林地与水体的用户精度分别提高了 7.79 百分点与 1.81 百分点, 灌木与人造地表的制图精度分别提升了 7.69 百分点与 0.59 百分点。整体来看, 结合植被覆盖度及其派生植被特征进行土地覆盖信息的提取, 在简单易行的同时, 为提高分类精度提供了有效支持。

关键词 土地覆盖; 植被覆盖度; 随机森林; 精度分析
中图分类号 F301.2; P237

DOI: 10.12202/j.0476-0301.2021224

0 引言

土地覆盖是地球表面各种物质类型与自然属性特征的综合体, 例如土地表层覆盖的植被、水体、沙漠和人造地表等^[1]。土地覆盖分类是区分土地覆盖组成单元的过程^[2]。由于人类活动的加强, 人与各种自然因子共同影响土地覆盖^[3], 土地覆盖变化同时也是影响全球和区域气候和陆表生态系统等变化的关键因子。因此, 研究与准确掌握土地覆盖情况对社会经济发展、监测生态环境和预测自然灾害等具有重要意义^[4]。

20 世纪 70 年代起, 反射率数据被广泛运用在土地覆盖分类^[5-6]。目前, 由反射率数据为主的土地覆盖类型信息提取技术主要是在像元尺度上进行分类^[7-8]。但存在像元的空间分辨率较粗和地物的复杂性问题, 因此像元具有混合光谱的特点^[9]。同时, 由于地物的光谱曲线存在“同物异谱”和“异物同谱”, 利用反射率数据进行土地覆盖分类会存在错分和漏分等现象, 包括耕地和人造地表的混淆、林地信息的提取误差等^[10], 这些问题仍然是基于遥感土地覆盖分类中需要解决的难题。

近些年, 遥感技术发展迅速, 定量遥感产品成为

全球环境监测和进行土地覆盖分类分析的重要数据集^[11-12]。定量遥感产品, 如植被指数、植被覆盖度等数据所具有的植被特性与土地覆盖类型具有紧密联系, 宫攀等^[13]、张栋等^[14]使用植被指数产品结合机器学习进行了土地覆盖分类和分析。在关中地区, EVI (enhanced vegetation index) 对植被的识别能力较 NDVI 高, 更有利于土地覆盖分类^[15]。在延河流域, 研究者参考当地年植被变化特征, 根据土地覆盖类型的 NDVI 时序曲线特征进行土地覆盖分类^[16]。定量遥感产品可以突出地物的物理和生物地球化学属性, 从而可以在一定程度上辅助区分土地覆盖类型^[11-12, 17-19]。

本文研究目的在于探索植被覆盖度在土地覆盖分类中的适用性。本研究在 Landsat 8 OLI 反射率基础上, 根据植被覆盖度(fractional vegetation cover, FVC)变化提取植被物候特征, 并将此作为分类依据。根据已有的土地覆盖数据, 如土地覆盖数据(finest resolution observation and monitoring of global land cover, FROM-GLC)^[20]、全球 30 m 地表覆盖遥感数据产品(Globe-Land30)^[21]和谷歌地球影像数据选取样本, 比较神经网络、支持向量机和随机森林共 3 种土地覆盖分类方法, 并探讨 FVC 用于土地覆盖分类的优势。

^{*} 国家重点研发计划资助项目(2016YFB0501404)

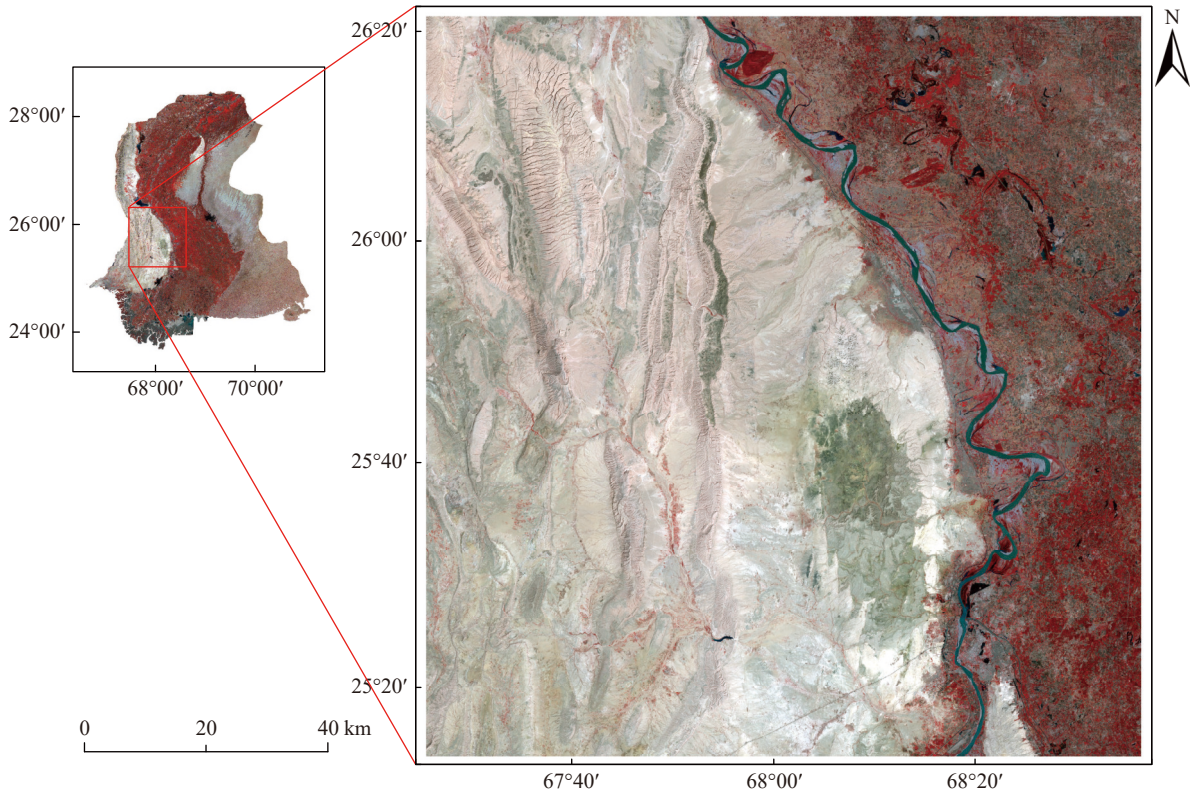
[†] 通信作者: 赵祥(1972—), 男, 博士, 教授。研究方向: 资源环境遥感应用。E-mail: zhaoxiang@bnu.edu.cn

收稿日期: 2021-09-10

1 材料与方 法

1.1 研究区 研究区位于南亚次大陆西北部的巴基斯坦信德省中部,如图 1 所示.信德省面积为 14.09 万 km²,地处印度河下游平原,属于亚热带气候,气候干旱;经济以农业为主,主要种植小麦、棉花和甘蔗;西部是山地为主,主要土地覆盖类型为裸地,植被较

少;东部是平原,以耕地为主.总体上,研究区位于印度河流域,地形上由西向东为山地—河谷—平原,经济上包含居住聚集区、乡村种植区和远郊地区,呈现出人类活动由强到弱的变化,非常有利于探究植被覆盖度用于区分林灌草耕地等土地覆盖类型、优化分类算法,从而评估植被覆盖度产品能否应用并提高地表覆盖分类精度.



Landsat 8 OLI 标准假彩色图, R、G、B 通道分别为近红外波段 5、红色波段 4 和绿色波段 3.

图 1 研究区范围

1.2 数据

1.2.1 土地覆盖分类数据及样本点 土地覆盖分类数据包括土地覆盖数据(FROM-GLC)和全球 30 m 地表覆盖遥感数据产品(GlobeLand30).FROM-GLC 数据分辨率为 10 m,数据时间为 2017 年.GlobeLand30 由国家基础地理信息中心牵头研制,数据时间为 2000、2010 和 2020 年.2 个数据集都包含 10 个主要的土地覆盖类型,分别是耕地、森林、草地、灌木地、湿地、水体、苔原、人造地表、裸地、冰川和永久积雪.

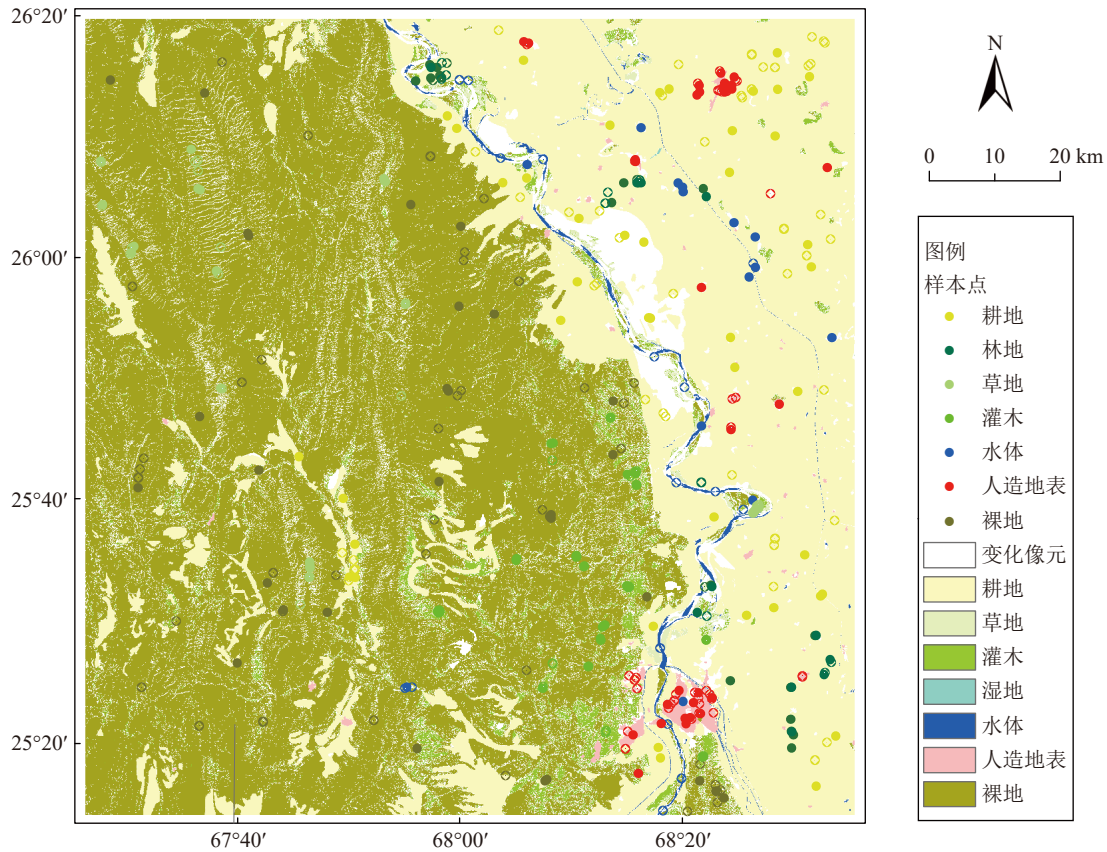
分类样本点参考 2017 年的 FROM-GLC 数据和 2000、2010 与 2020 年的 GlobeLand30 数据进行选择.对于 3 期 GlobeLand30 影像中,土地覆盖类型一致的像元,认为该像元是该类型的固定像元,具体空间分布参见图 2.对于林地等没有一致像元的土地覆盖类型,则参考 FROM-GLC 数据和 2016 年的谷歌地球影像进行选择.最后,样本点参考谷歌影像数据进行筛

选和补充,保证样本点的正确性.如表 1 所示,各土地覆盖分类类型的训练样本约占总样本的 70%,验证样本点约占总样本的 30%.

1.2.2 Landsat 8 OLI 数据与 NDVI 数据 Landsat 8 OLI 影像源于美国地质调查局网站(<http://earthexplore.usgs.gov/>),空间分辨率为 30 m,表 2 显示了 OLI 数据波段范围与主要用途.选取的影像条带号为 h152 和 v42,时间为 2016 年 4 月 13 日,在研究区范围内无云覆盖,品质较好.Landsat 8 OLI 遥感影像采用最近邻重采样方法将影像分辨率采样至 16 m,和植被覆盖度数据分辨率一致,并选用 2~7 波段参与分类.

NDVI 数据由 Landsat 8 OLI 第 4 红波段和第 5 红外波段计算得到.

1.2.3 植被覆盖度数据 植被覆盖度数据根据高分一号的宽覆盖多光谱相机(GaoFen-1 Wide Field View, GF-1 WFV)获取的地表反射率数据估算得到,并采用



底图为 2000、2010、2020 年的 GlobeLand30 数据的土地覆盖类型不变像元。

图 2 样本点空间分布

表 1 样本点占比

土地覆盖类型	训练样本像元	验证样本像元	总样本像元	训练样本占比/%	验证样本占比/%
耕地	748	324	1072	69.78	30.22
林地	175	72	247	70.85	29.15
草地	180	74	254	70.87	29.13
灌木	158	65	223	70.85	29.15
水体	244	106	350	69.71	30.29
人造地表	402	172	574	70.03	29.97
裸地	489	202	644	70.77	29.23
总计	2396	968	3364	70.24	29.76

表 2 Landsat 8 OLI 2-7 波段信息

波段	波长范围/ μm	主要用途
2-蓝波段	0.45~0.52	水体穿透, 分辨土壤和植被
3-绿波段	0.53~0.60	分辨植被
4-红波段	0.63~0.68	观测道路、裸土和植被种类等
5-近红外波段	0.85~0.89	植被信息提取
6-短波红外1	1.57~1.65	分辨道路, 裸土和水
7-短波红外2	2.11~2.29	岩石、矿物和植被覆盖的分辨能力很好

PROSAIL 植被冠层辐射传输模型和随机森林算法相结合的方法生产数据^[22]. PROSAIL 模型根据植被理化参数, 生成 FVC 和 GF-1 WFV 在绿光、红光和近红外波段的反射率的模拟数据集. 以模拟数据集的反射率数据作为输入, 以 FVC 作为输出训练随机森林回归模型. 植被覆盖度数据的分辨率为 16 m, 时间为 2016 年的逐月数据.

1.3 方法

1.3.1 FVC 指数 耕地、林地、草地和灌木的 FVC 变化曲线不同, 因此构建 FVC 指数来提取 4 种植被类型的 FVC 特征, 使得 4 种植被更易区分. 其中, 耕地

和林地的 FVC 变化曲线较为相似, 在 4—8 月份的 FVC 升高, 1—4 月份与 9—10 月份的 FVC 降低, 但林地的 FVC 整体要比耕地较高. 草地和灌木的 FVC 全年变化很小, 但灌木要比草地 FVC 数值整体较高. 研究选用非生长季的 4 月份与生长季的 8 月份的 FVC 数据构建 2 种 FVC 指数, 构建的 F_a 指数在 4 种植被类型上具有数据区分度, 有助于区分植被与非植被区域, 同时提高耕地、林地、草地和灌木 4 种植被类型的区分度. 构建的 F_b 指数也辅助区分耕地、林地、草地和灌木 4 种植被类型.

$$F_a = (F_{\text{AUG}} + F_{\text{APR}}) / 2, \quad (1)$$

$$F_b = (F_{\text{AUG}} - F_{\text{APR}}) / 2, \quad (2)$$

式中: F_a 和 F_b 代表 2 种 FVC 指数; F_{AUG} 和 F_{APR} 分别代表 8 月份和 4 月份的 FVC.

1.3.2 神经网络 神经网络分类是计算机采用一系列处理单元模拟神经元并实现神经元之间的信号传递,通过算法模拟人脑识别图像分类的过程.神经网络分类方法含有输入层、输出层以及连接输入层和输出层的隐含层,采用反向传播神经网络通过最小化误差函数来完成训练^[23].

1.3.3 支持向量机 支持向量机是根据统计学习理论基础建立的机器学习方法.支持向量机能最大化区分不同类别样本,根据自动寻找最优间隔超平面的支持向量构建分类器^[24, 25].支持向量机分类可以实现不同类别之间的最大化间隔,从而实现样本的正确分类.

1.3.4 随机森林 随机森林算法是由 Breiman 等^[26]于 2001 年提出的.它通过集成学习的思想将若干决策树集成在一起,利用随机抽样训练多棵决策树,并采用投票方式决定样本的分类结果.随机森林算法可以输入高维度和多特征的数据,不需要进行数据降维和特征选择,可以判断各个特征在分类上的重要性和不同特征间的相互影响.

随机森林算法通过不同样本集来增加各决策树模型间的差异,从而提高整个分类模型的外推预测能力.在 N 轮训练得到 1 个分类模型序列 $(h_1(x), h_2(x), \dots, h_N(x))$, 并通过投票方式决定最后的分类结果.

$$H(x) = \operatorname{argmax}_Y \sum_{i=1}^N I(h_i(x) = Y), \quad (3)$$

式中: $H(x)$ 表示分类模型; h_i 表示单个决策树分类模型; Y 表示输出变量; $I()$ 为示性函数.

采用随机森林方法进行土地覆盖分类.与其他算法相比,随机森林具有极好的准确率,并且参数调试少,训练速度较快,易于操作与实现^[27-28].

1.3.5 分类精度评价 共选取总体分类精度、Kappa 系数、制图精度和用户精度共 4 个指标来评价土地覆盖分类效果.总体分类精度指被正确分类的像元占像元总数的比例;制图精度指假定地表真实为 A 类,分类器能将 1 幅图像的像元归为 A 的概率;用户精度指假定分类器将像元归到 A 类时,相应的地表真实类别是 A 的概率;Kappa 系数(k)是一种比例系数,代表着分类与完全随机的分类产生错误减少的比例. k 为 0~1, k 越大表示分类精度越高.4 个指标具体计算方法参见表 3.

表 3 4 种分类精度评价指标

分类精度指标	公式	作用和意义
总体分类精度	$O = \sum_{i=1}^k N_{ii} / N \times 100\%$	直接反映总体正确的比例,计算简单
Kappa 系数	$k = \left(N \sum_{i=1}^k N_{ii} - \sum_{i=1}^k N_{i+} N_{+i} \right) / \left(N^2 - \sum_{i=1}^k N_{i+} N_{+i} \right)$	用于一致性检验和衡量土地覆盖分类精度
制图精度	$P_{Ai} = \frac{N_{ii}}{N_{+i}} \times 100\%$	是制图者对分类精度的一个度量,即实际分类与参考类一致性的程度
用户精度	$U_{Ai} = \frac{N_{ii}}{N_{ii}} \times 100\%$	用户最感兴趣的一种分类精度的度量

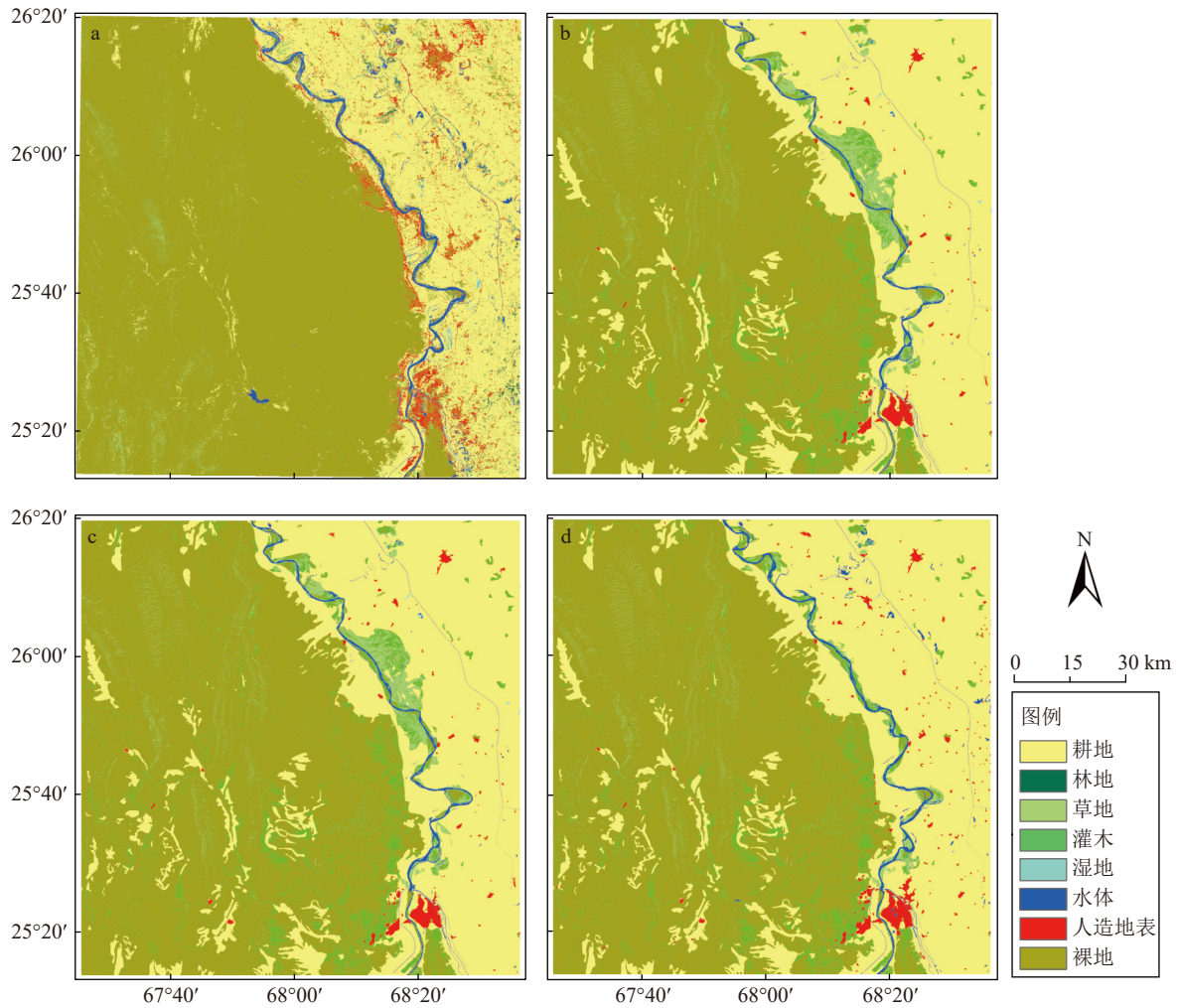
式中: N 代表总样本数; k 代表总类别数; N_{ii} 代表被分到正确类别的样本数; N_{i+} 和 N_{+i} 是第 i 类的真实样本数和预测为第 i 类的样本数.

2 研究结果

2.1 已有土地覆盖分类数据对比 研究区以耕地和裸地为主,印度河流经其中.如图 3 所示,研究区的土地类型共有耕地、林地、草地、灌木、湿地、水体、人造地表和裸地 8 种. FROM-GLC 数据和 GlobeLand30 数据具有一致性.在 2 种土地覆盖类型数据中,研究区西部为山地,其中大部分为裸地,分布着零星的草地、灌木和耕地.研究区东部为平原,是人类活动的主要区域,大部分区域为耕地,小部分为林地和人造地表.通过谷歌地球影像目视解译, FROM-GLC 数据和 GlobeLand30 数据在该区域的分类都有不足之处. FROM-GLC 数据在研究区西边存在草地和灌木的漏

分,而 GlobeLand30 在河岸边存在人工地表的漏分,并且二者都存在林地的漏分.

如表 4 所示, 2 种土地覆盖分类数据在林地、草地、湿地、水体和人造地表的覆盖类型占比相对一致,在耕地、灌木和裸地的覆盖类型占比差异较大.裸地占比最高, 2 种土地覆盖分类数据都 >50%, 而 FROM-GLC 数据比 GlobeLand30 数据裸地的占比多约 11 个百分点,因为在 GlobeLand30 数据的裸地中分布着更多的耕地、草地和灌木.耕地占比次之,均 >25%, 而 FROM-GLC 数据比 GlobeLand30 数据的耕地占比要少约 10 个百分点.因为在 FROM-GLC 数据的耕地中伴随着零散的人工地表和林地,而 GlobeLand30 数据分类结果中耕地更加整体.林地、草地、水体、湿地



a. 2017 年 10 m FROM-GLC; b. 2000 年 30 m GlobeLand; c. 2010 年 30 m GlobeLand; d. 2020 年 30 m GlobeLand.

图 3 不同土地覆盖数据

表 4 不同土地覆盖分类数据中各土地类型在研究区占比

类型	FROM-GLC	GlobeLand30		
		2000年	2010年	2020年
耕地	26.3765	35.0501	35.0448	36.3027
林地	0.3433	0	0	0.0005
草地	3.4592	2.8436	2.8439	1.7290
灌木	0.5456	6.8172	6.8208	6.2853
湿地	0.2210	0.0167	0.0164	0.0148
水体	1.3729	0.8527	0.8522	1.1261
人造地表	3.0736	0.6979	0.6971	1.0729
裸地	64.6080	53.7218	53.7248	53.4687

和人造地表在 FROM-GLC 占比更多, 而灌木在 GlobeLand30 数据中比 FROM-GLC 数据中占比要多, 这可能是二者分类的样本选择、模型选取与设置等因素不同造成的差异。

2.2 不同植被类型的植被覆盖度特征提取 由于现

有土地覆盖分类数据存在漏分和错分现象, 本文将融合 FVC 变化特征和 Landsat 8 OLI 数据对研究区进行分类, 提高分类精度. 研究区主要存在的植被类型为耕地、林地、草地和灌木. 4 种植被类型的 FVC 变化如图 4-a 所示, 耕地和林地的变化曲线较为相似, 在 3—6、11 月份呈下降趋势, 数值较低, 但林地的 FVC 数值整体要比耕地较高. 草地和灌木的 FVC 变化曲线一致, 都存在较小的波动, 但灌木要比草地 FVC 数值整体较高。

基于 4 种植被类型的 FVC 变化曲线, 构建 FVC 指数来进一步区分 4 种植被. 选用 4、8 月份的 FVC 数据构建了 2 种 FVC 指数. 据图 4-b、c 所示, 2 种 FVC 指数能在一定程度上区分耕地、林地、草地和灌木 4 种植被类型. 根据构建的 FVC 指数, 有助于区分植被与非植被区域, 并且可以提高不同植被类型的区分度。

2.3 分类方法比较 利用神经网络、支持向量机和随机森林分类方法对研究区进行土地覆盖分类, 并比较这 3 种分类方法在 Landsat 8 OLI 和结合 NDVI 数据

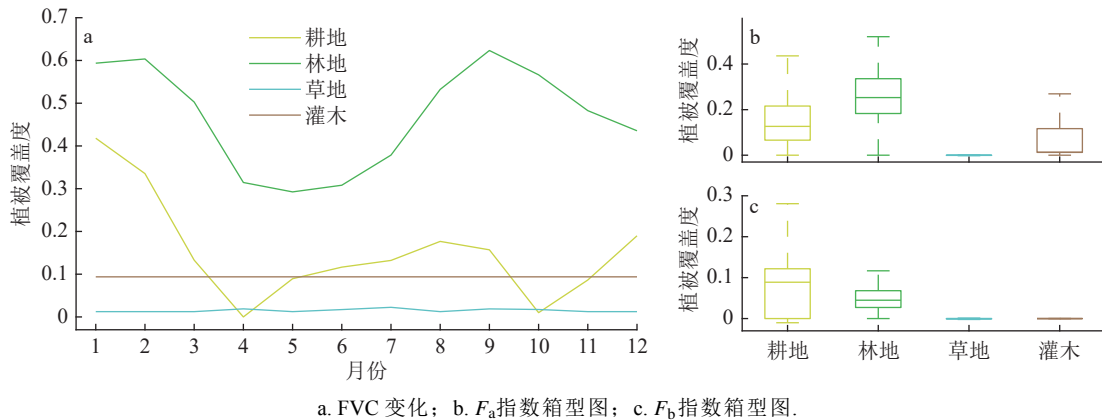


图 4 4种植被覆盖度变化曲线和箱型图

与 FVC 指数数据后的精度. 如表 5 所示, 3 种分类方法中, 随机森林方法最优, 支持向量机次之, 神经网络分类结果最差. 增加 FVC 指数数据, 3 种分类方法能在 Landsat 8 OLI 数据基础上精度有所提升, 总体精度最大提升 3.45 百分点, k 最大提高 0.0429. 另外增加 NDVI 数据, 分类精度增幅较低. 在神经网络分类中, 增加

NDVI 数据参与分类, 虽然总体精度提升了 0.1 百分点, 但 k 降低了 0.0016, 这是因为土地类别的样本数量不一样, 总体精度很容易偏向占比较大的类别, 而 k 能纠正这种偏向性. 因为随机森林分类方法中精度最高, FVC 指数比 NDVI 在模型中对于分类精度提升较多, 因此用随机森林分类方法做进一步分析.

表 5 不同组合分类数据在 3 种分类方法中的精度评价

分类数据	总体精度/%			k		
	神经网络	支持向量机	随机森林	神经网络	支持向量机	随机森林
Landsat 8 OLI	77.73	80.49	82.07	0.7178	0.7577	0.7783
Landsat 8 OLI+FVC 指数	79.11	80.99	85.52	0.7397	0.7638	0.8212
Landsat 8 OLI+NDVI+FVC 指数	79.21	81.67	86.21	0.7381	0.7720	0.8293

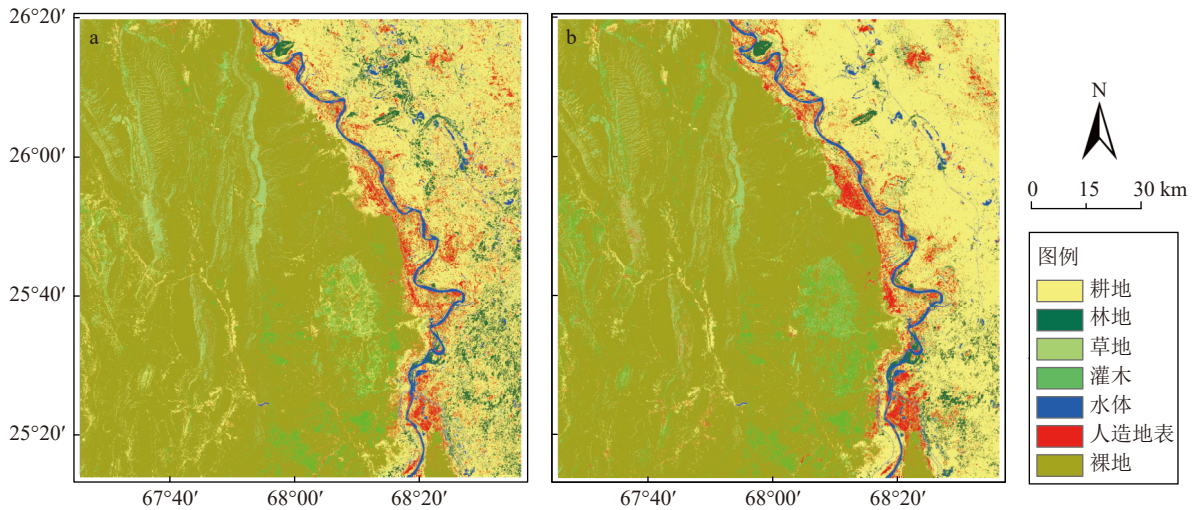
2.4 分类结果 采用随机森林分类方法, 利用 Landsat 8 OLI 数据进行研究区土地覆盖分类, 再用 Landsat 8 OLI 数据结合构建的 FVC 指数数据进行土地覆盖分类. 土地覆盖分类类型为耕地、林地、草地、灌木、水体、人造地表与裸地. 因为湿地在研究区分布较少, 因此不选择湿地作为主要土地覆盖分类类型. Landsat 8 OLI 结合构建的 FVC 指数后, 研究区的土地覆盖分类总体精度从 82.07% 提升到 85.52%, k 从 0.7783 提高到 0.8212. 结果表明融合了植被 FVC 变化特征的 FVC 指数数据有助于提升总体分类精度. 如图 5 和表 6 所示, 增加 FVC 指数作为分类特征数据, 具体提高了耕地、草地和裸地的制图精度和用户精度, 对林地与水体的用户精度分别提高了 7.89 与 1.81 百分点, 灌木与人造地表的制图精度分别提升了 7.69 百分点与 0.59 百分点.

3 讨论

土地覆盖数据集是土地信息研究的基础, 然而不同的土地覆盖数据来自不同的科研团队, 采用不同的

数据源及分类方法, 导致分类结果在不同地物类别上具有差异性^[29]. GlobeLand30 数据景观多样性较低, 地表景观类别均质性较强, 而 FROM-GLC 数据的地表景观异质性较强, 破碎化程度较高^[30]. 在中国区域, 对于耕地类型的数量和空间准确度, GlobeLand30 产品与 FROM-GLC 产品存在差别^[31]. FROM-GLC 的 2017 数据集在中国裸地、林地和耕地的精度相对较高, 灌丛的精度较低^[32]. 定量研究土地覆盖数据的一致性和差异性, 对于之后土地覆盖分类具有重要意义.

在土地覆盖分类中, 增加 FVC 指数对于每个土地覆盖类型的精度有不同的影响. 增加 FVC 指数参与分类与仅用反射率数据相比, 提高了耕地、草地和裸地的制图精度和用户精度, 提升了林地与水体的用户精度和灌木与人造地表的制图精度. 根据表 6、图 5 和 6 所示, FVC 指数数据使得图像右下角耕地中的林地像元更明显, 使得原先错分成林地的像元重新分为耕地, 因此提升了林地 7.79 百分点的用户精度. 在河岸中段, FVC 数据能纠正错分成耕地的人造地表, 使得人造地表更加聚集, 因此提升了人造地表 0.59 百分



a. Landsat 8 OLI Land Cover; b. Landsat 8 OLI 和 FVC 指数.

图 5 土地覆盖分类

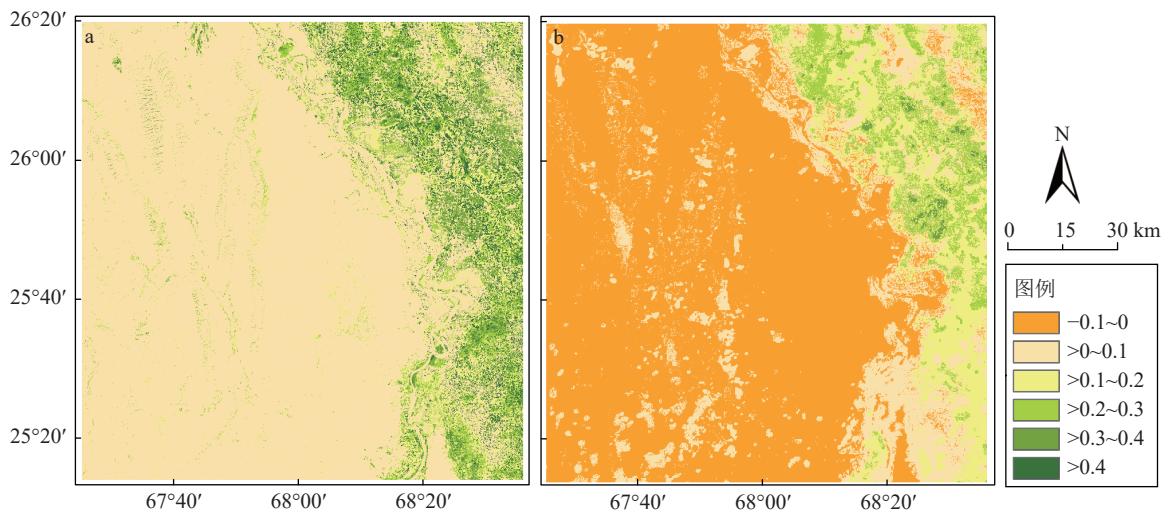
表 6 分类精度 %

地物类型	Landsat		Landsat + FVC 指数	
	制图精度	用户精度	制图精度	用户精度
耕地	76.23	81.25	80.25	88.14
林地	91.67	59.46	91.67	67.35
草地	74.32	79.71	91.89	90.67
灌木	63.08	87.23	70.77	85.19
水体	98.11	96.30	98.11	98.11
人造地表	76.74	79.04	77.33	75.14
裸地	93.07	89.95	94.55	90.95
总计	81.89	81.85	86.37	85.08

点的制图精度, 林地和人造地表的纠正也提升了耕地 4.02 个百分点和 6.89 个百分点的制图精度和用户精度. 在研究区的西边, FVC 指数数据使得错分成草地的灌

木得到纠正, 提高了灌木 7.69 个百分点的制图精度, 但也提升了其他类别错分成灌木, 因此灌木的用户精度降低了 2.04 百分点. FVC 指数表明在研究区左下角耕地占比较少, 加入 FVC 指数纠正了耕地的分类, 使得草地和裸地的分类更加精准, 因此草地的制图精度和用户精度分别提升了 17.57 个百分点和 10.96 百分点, 裸地的制图精度和用户精度分别提高了 1.48 百分点和 1 百分点.

利用植被指数数据参与分类, 在整体分类精度上提升了分类精度, 改善了分类器的不足, 但是对于不同的土地覆盖类型, 分类精度有提升, 但是仍然存在误判现象^[33]. 用 MODIS NDVI 数据参与分类, 对于零星的植被区域和城镇也存在误判^[34]. 研究者增加 MODIS EVI 数据进行植被物候参数提取并用于改善 MODIS 土地覆盖产品, 但在林地的制图精度和用户



a. F_a 指数数据, $F_a = (F_{AUG} + F_{APR}) / 2$; b. F_b 指数数据, $F_b = (F_{AUG} - F_{APR}) / 2$.

图 6 研究区植被覆盖度指数空间分布

精度有所降低,草地分类的用户精度也降低 59 百分点^[5]。由于混合像元的存在,FVC 提取的部分像元的植被物候特征不明显或者部分非植被像元提取到植被物候特征而被错分,对各种土地覆盖类型的分类精度产生不同影响,研究有待进一步改进,之后可以引入其他的定量遥感产品,进一步提高各个土地覆盖类型的分类精度。

4 结论

本文分析了高分辨率的土地覆盖分类数据,并利用 Landsat 8 OLI 与植被覆盖度提取的植被特征数据,运用神经网络、支持向量机和随机森林进行土地覆盖分类。通过以上研究发现:

1)土地覆盖分类数据中,FROM-GLC 数据和 GlobeLand30 数据具有一致性。参考 2016 年谷歌影像数据,FROM-GLC 数据在研究区西边存在草地和灌木的漏分,而 GlobeLand30 在河岸存在人工地表的漏分,并且二者都存在林地的漏分。

2)耕地、林地、草地和灌木的植被覆盖度变化曲线不同,使用不同季节的植被覆盖度能有效区分不同植被类型的植被特征,能够辅助分类。

3)3 种土地覆盖分类方法中,随机森林明显优于其他 2 种,支持向量机次之,神经网络分类结果最差。另外,在神经网络和随机森林分类模型中,增加 FVC 指数数据比 NDVI 数据,分类精度增幅更大。

4)植被覆盖度辅助下的土地覆盖分类方法的总体精度为 85.52%,Kappa 系数为 0.8212,比仅用反射率的土地覆盖分类总体精度提高了 3.45 百分点,Kappa 系数提高 0.0429。

5)对于具体的土地类型,提高了耕地、草地和裸地的制图精度和用户精度,对林地与水体的用户精度和灌木与人工地表的制图精度有提升作用,改善了图像的错分与漏分现象,表明植被覆盖度提取的 FVC 指数能有效辅助分类,可以提升分类精度。

5 参考文献

- [1] 陈军,陈晋,宫鹏,等.全球地表覆盖高分辨率遥感制图[J].*地理信息世界*,2011,9(2):12
- [2] 史培军.土地利用/覆盖变化研究的方法与实践[M].北京:科学出版社,2000
- [3] 张新荣,刘林萍,方石,等.土地利用、覆被变化(LUCC)与环境变化关系研究进展[J].*生态环境学报*,2014,23(12):2013
- [4] SHRESTHA U S. Land use and land cover classification from ETM sensor data: a case study from Tamakoshi River Basin of Nepal[J]. *ISPRS-International Archives of the Photogrammetry & Remote Sensing and Spatial Information Sciences*, 2014, XL-8(1): 943
- [5] YUAN F, SAWAYA K E, LOEFFELHOLZ B C, et al. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing[J]. *Remote Sensing of Environment*, 2005, 98(2/3): 317
- [6] FUKUE K, SHIMODA H. Global land cover classification using modis surface reflectance products[J]. *ISPRS-International Archives of the Photogrammetry & Remote Sensing and Spatial Information Sciences*, 2016, XLI-B7(1): 219
- [7] 李恒凯,王利娟,肖松松.基于多源数据的南方丘陵山地土地利用随机森林分类[J].*农业工程学报*,2021,37(7):244
- [8] 杜明义,武文波,郭达志.多源地学信息在土地荒漠化遥感分类中的应用研究[J].*中国图象图形学报*,2002,7(7):117
- [9] 汤国安.遥感数字图像处理[M].北京:科学出版社,2004
- [10] 陈姝,居为民.遥感图像分类方法及研究进展[J].*河北农业科学*,2009,13(1):143
- [11] 詹福雷.基于面向对象的高分辨率遥感影像信息提取[D].长春:吉林大学,2014
- [12] 陈嘉,赵祥,唐荣云,等.基于定量遥感产品的土地覆盖分类方法研究[J].*北京师范大学学报(自然科学版)*,2018,54(3):405
- [13] 宫攀,陈仲新,唐华俊,等.基于MODIS温度/植被指数的东北地区土地覆盖分类[J].*农业工程学报*,2006,22(9):94
- [14] 张栋,徐丹丹,吴永波,等.融合物候特征对城市土地覆被进行分类及覆被变化分析:以金坛区为例[J].*长江流域资源与环境*,2021,30(7):1614
- [15] 张盼,马孝义,赵龙,等.基于不同植被指数和分类回归树的关中地区土地利用分类[J].*水土保持研究*,2018,25(3):310
- [16] 陆广勇,杨勤科,王海江.基于MODIS-NDVI时序数据的黄土丘陵区土地覆盖分类研究[J].*水土保持研究*,2011,18(2):112
- [17] LU L L, KUENZER C, GUO H D, et al. A novel land cover classification map based on a MODIS time-series in Xinjiang, China[J]. *Remote Sensing*, 2014, 6(4): 3387
- [18] FRIEDL M A, MCIVER D K, HODGES J C F, et al. Global land cover mapping from MODIS: algorithms and early results[J]. *Remote Sensing of Environment*, 2002, 83(1/2): 287
- [19] 廖顺宝,岳艳琳.基于时序NDVI图谱库提高土地覆盖分类精度的方法[J].*农业工程学报*,2018,34(7):241
- [20] GONG P, LIU H, ZHANG M N, et al. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017[J]. *Science Bulletin*, 2019, 64(6): 370
- [21] 陈军,廖安平,陈晋,等.全球30 m地表覆盖遥感数据产品-Globe Land30[J].*地理信息世界*,2017,24(1):1

- [22] TAO G F, JIA K, ZHAO X, et al. Generating high spatio-temporal resolution fractional vegetation cover by fusing GF-1 WFV and MODIS data[J]. *Remote Sensing*, 2019, 11(19): 2324
- [23] 黄孝斌, 沈军, 李国新, 等. 遗传算法优化确定BP神经网络的遥感图像分类[J]. *现代电子技术*, 2020, 43(12): 47
- [24] 员永生. 基于支持向量机分类的面向对象土地覆被图像分类方法研究[D]. 杨凌: 西北农林科技大学, 2010
- [25] 李玲, 王红, 刘庆生, 等. 基于纹理特征和支持向量机的ALOS图像土地覆被分类[J]. *国土资源遥感*, 2011, 23(4): 58
- [26] BREIMAN L. Random forest[J]. *Machine Learning*, 2001, 45(1): 5
- [27] 王李娟, 孔钰如, 杨小冬, 等. 基于特征优选随机森林算法的农耕地土地利用分类[J]. *农业工程学报*, 2020, 36(4): 244
- [28] 马慧娟, 高小红, 谷晓天. 随机森林方法支持的复杂地形区土地利用/土地覆被分类研究[J]. *地球信息科学学报*, 2019, 21(3): 359
- [29] 周晖. 融合多源遥感影像的森林资源变化监测研究: 以黄土高原为例[D]. 北京: 北京林业大学, 2020
- [30] 邵明超, 宋宏利, 尚明, 等. 多源土地覆被数据类别一致性及其景观格局差异性: 以京津冀区域为例[J]. *生态学杂志*, 2020, 39(11): 3881
- [31] 陆苗, 吴文斌, 张莉, 等. 不同耕地数据集在中国区域的对比研究[J]. *中国科学: 地球科学*, 2016, 46(11): 1459
- [32] 郭紫甜, 王春梅, 刘欣, 等. 基于小流域抽样单元的中国FROM-GLC30数据精度评价[J]. *地球信息科学学报*, 2021, 23(3): 524
- [33] 张霞, 孙睿, 张兵, 等. 基于MODIS植被指数时间谱的华北平原土地覆盖分类[J]. *农业工程学报*, 2006, 22(12): 128
- [34] 何彬方. 基于MODIS数据土地覆盖分类研究[D]. 合肥: 中国科学技术大学, 2008
- [35] 康峻, 侯学会, 牛铮, 等. 基于拟合物候参数的植被遥感决策树分类[J]. *农业工程学报*, 2014, 30(9): 148

Land cover classification combined with fractional vegetation cover

SUN Na ZHAO Xiang MU Baohui ZHAO Jiacheng LIU Naijing

(State Key Laboratory of Remote Sensing Science, Beijing Engineering Center for Global Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, 100875, Beijing, China)

Abstract Land cover is closely related to local ecological environment. Remote sensing technology can quickly and accurately extract ground feature information, and plays an important role in land cover classification. Singular classification data source, mixed pixels, few quantitative remote sensing products, all leave plenty room for further improvements in existing land cover classification methods, and in the accuracy of present classifications. Landsat 8 OLI reflectance data were combined with vegetation phenological feature data (extracted by quantitative remote sensing inversion of Fractional Vegetation Cover - FVC), and the existing three land cover classification methods of neural network, support vector machine and random forest were compared. The random forest classification method showed good results. The overall accuracy of the classification method combining reflectance with vegetation feature data is 85.52%, and the Kappa coefficient is 0.8212, 3.45pct higher than the overall accuracy of land cover classification using reflectance alone, and the Kappa coefficient is increased by 0.0429. The vegetation feature data extracted by vegetation coverage can effectively improve mapping accuracy and user accuracy of cultivated land, grassland and bare land. User accuracy of woodland and water bodies was found to have increased by 7.79pct and 1.81pct respectively. Mapping accuracy of shrubs and artificial ground was found to have increased by 7.69pct and 0.59pct, respectively. Overall, extraction of land cover information combined with vegetation coverage and derived vegetation characteristics provides effective support for improving classification accuracy while being simple and easy to implement.

Keywords land cover; fractional vegetation cover; random forest; accuracy analysis

【责任编辑: 刘先勤】