

DOI: 10.5846/stxb202106251691

陈航,王颖,张昕,曹利.秦岭南坡子午河中下游流域土地利用/土地覆被信息提取及其应用.生态学报,2022,42(22):9239-9249.

Chen H, Wang Y, Zhang X, Cao L. Extraction and application of land use/land cover information in the middle and lower reaches of Ziwu River on the south slope of Qinling Mountains. Acta Ecologica Sinica, 2022, 42(22): 9239-9249.

秦岭南坡子午河中下游流域土地利用/土地覆被信息提取及其应用

陈 航¹, 王 颖^{1,*}, 张 昕², 曹 利¹

¹ 西安理工大学, 西安 710048

² 陕西省引汉济渭工程建设有限公司, 西安 710086

摘要:精确的土地利用/土地覆被数据不仅可以反映区域的生态环境状况,为环境部门提供决策支持,也为实现区域生态环境高质量发展发挥重要作用。以子午河中下游流域为研究区,利用多源多时相的 landsat8 卫星遥感数据,结合地面调查数据、文献调研等,探讨并研究支持向量机分类法(SVM)和随机森林模型(RFM)对该区的植被类型和土地利用现状类型进行识别,对两种方法的分类精度进行对比,并分析和评价光谱特征变量对模型的重要性和适用性。利用满足要求的土地利用现状数据,再结合修正的通用土壤流失方程 RUSLE 模型进一步计算出研究区的土壤侵蚀模数,绘制研究区土壤侵蚀分布图,结合土地利用/植被覆盖信息计算研究区的生态环境状况指数,从宏观上对子午河中下游流域进行生态环境评价。结果表明:①随机森林模型可以有效利用样本的特征因子,并与地形约束因子结合,从而对植被和土地利用类型进行分类,分类总体精度均达到 80%以上, kappa 系数分别为 0.73 和 0.86,与传统的 SVM 方法相比,RFM 方法均提高了森林类型和土地利用类型的分类精度。②研究区总体生态环境状况指数为 87.12,生态环境状况为优,其中水源区附近由于土壤侵蚀流失量相对较大,所以生态环境状况为良,占研究区总面积的 15.69%。

关键词:遥感解译;随机森林模型;多源多时相;植被类型图;生态环境

Extraction and application of land use/land cover information in the middle and lower reaches of Ziwu River on the south slope of Qinling Mountains

CHEN Hang¹, WANG Ying^{1,*}, ZHANG Xin², CAO Li¹

¹ Xi'an University of Technology, Xi'an 710048, China

² Shaanxi Han Weihe Water Diversion Engineering Construction Company Limited, Xi'an 710086, China

Abstract: Accurate land use/land cover data can not only reflect the regional ecological environment and provide decision support for environmental departments, but also play an important role in achieving higher quality development of regional ecological environment. Taking the middle and lower reaches of Ziwu River Basin as the study area, using multi-source and multi temporal Landsat 8 satellite remote sensing data, combined with ground survey data and literature survey, this study discusses and studies the support vector machine classification (SVM) and random forest model (RFM) to identify the vegetation types and land use status types in this area, and compares the classification accuracy of the two methods. The importance and applicability of spectral characteristic variables to the model are analyzed and evaluated. Based on the data of land use status, combined with the modified general soil loss equation RUSLE model, the soil erosion modulus of the study area was further calculated, the distribution map of soil erosion in the study area was drawn, and the ecological

基金项目:陕西省自然科学基金基础研究计划-引汉济渭工程水源区湿地生态修复技术研究(2019JLM-63)

收稿日期:2021-06-25; **网络出版日期:**2022-07-12

* 通讯作者 Corresponding author. E-mail: wangying@xaut.edu.cn

environment index of the study area was calculated according to the land use vegetation cover information, and the ecological environment evaluation of the tourism basin under meridian river insects was carried out from a macro point of view. To the extent that it is not possible to make a difference between the two groups. The results show that: ① the random forest model can effectively use the characteristic factors of samples and combine them with topographic constraints to classify vegetation and land use types. The overall classification accuracy is more than 80%, and the kappa coefficients are 0.73 and 0.86 respectively. Compared with the traditional SVM method, RFM method improves the classification accuracy of forest types and land use types. ② The overall Eco-environmental Status Index of the study area is 87.12, and the eco-environmental status is excellent. Due to the relatively large amount of soil erosion near the water source area, the eco-environmental status is good, accounting for 15.69% of the total area of the study area.

Key Words: remote sensing interpretation; random forest; multi-source and multi-temporal phase; vegetation type map; ecological environment

土地利用/土地覆被信息提取在研究区域整体生态环境状况方面起着重要作用^[1]。遥感技术以其快速、同步监测等特点,成为提取土地利用/植被覆盖信息高效的技术手段。近年来,随着遥感技术的快速发展,将多源遥感数据应用于植被覆盖类型和土地利用现状变化监测研究,以及植被物候监测已经非常普遍。基于决策树法,结合光谱特征和物候信息,对植被覆盖类型提取和土地资源监测的技术已经非常成熟^[2-9]。如韩东等基于无人机和决策树算法对浑善达克沙地的榆树疏林草原草本和草本植被类型实现了自动划分和植被覆盖度的估算^[10]。姚永慧等为提取太白山各级植被类型,通过资源三号、高分二号的多源多时相高分辨率遥感数据与地形约束因子、地面调查数据和以往研究资料相结合,实现了大中比例尺植被类型图的编制^[11-13]。WANG 等采用支持向量机分类方法,利用北京 1 号卫星数据提取了鄱阳湖的 NDVI 指数和 NDWI 指数,对鄱阳湖湿地植被功能型进行了分类^[14]。贾明明等基于面向对象分类方法,通过获取 MODIS 时序数据的物候特征并提取了 HJ-1 CDD 的光谱和纹理特征,将三者结合起来,实现了对森林植被类型的提取,结果表明加入物候特征可大幅提高模型的分类精度^[15]。随机森林分类方法(Random forest classification)是由美国科学家 Breiman 提出的一种基于决策树组合的机器学习新型高效的分类器,其优越的性能广泛应用于国外诸多领域^[16-26]。然而,国内学者对于 RF 模型的研究与应用相对较少。

本研究通过详尽的实地调查,获取研究区植被覆盖类型多样性方面的基础数据,建立地形约束因子,揭示了子午河中下游流域的植被类型、分布格局、群落结构特征,并基于样方数据和遥感地面调查点数据,结合影像的光谱特征、地形因子以及纹理特征建立随机森林模型,对子午河中下游流域的植被覆盖类型和土地利用类型进行分类,同时,与传统的 SVM 模型进行精度对比分析,评价并分析模型特征变量的重要性,研究基于多源信息的 RF 模型在植被覆盖类型和土地利用现状分类上应用的可行性,基于修正的通用土壤流失方程 RUSLE 模型,利用满足要求的分类数据进一步计算研究区土壤侵蚀模数,并利用 ArcGIS 软件进行可视化处理,得到研究区的土壤侵蚀模数分布图,结合上述土地利用/植被覆盖等信息,从宏观角度出发对子午河中下游流域进行生态环境质量评价,量化该地区的生态环境状况。

1 研究区概况

子午河中下游流域位于陕西省汉中市佛坪县、洋县与安康市宁陕县、石泉县四县交界地带,地理位置为 107°46'8"E—108°18'59"E, 33°10'38"N—33°41'2"N,属秦岭南麓,面积约为 2476.9km²,地势上北高南低,构造形迹以褶皱挤压变形为主,属相对稳定地区。区内主要河流有子午河、蒲河和椒溪河等;由于冬季受到北方冷气团控制,夏季受到南方副高压带的影响,导致气候、地形均呈现差异性变化,研究区南部地区多年平均降雨量为 806mm,多年平均气温为 14.5℃,多年平均蒸发量为 1065.6mm,而北部地区多年平均降雨量为 901mm,多年平均蒸发量为 1213mm,属于亚热带北缘山地暖温带湿润季风气候。研究区山多地少,农业基本条件差,

主要农作物为稻谷、小麦、玉米、油料和大豆等^[27];土地利用类型以草地、农耕地和林地为主;研究区林地以针叶林和混交林为主,有较少的含常绿阔叶树的落叶阔叶林,其中针叶林主要为油松林、马尾松林,混交林主要是油松-栎类混交林,阔叶林以青冈栎林、栓皮栎林、板栗林、核桃林、槲栎林、化香林等为主^[28—29]。区内海拔一般 500—2500m,植被垂直地带分布较复杂,自下而上分布有灌草丛带、落叶阔叶林带、针阔混交林带和针叶林带(图 1)。

2 研究方法

本研究基于 Landsat 8 OLI_TIRS 影像的光谱特征、纹理特征以及地形因子建立随机森林模型,对研究区土地利用/土地覆被进行分类,与传统的 SVM 分类方法进行精度对比分析,评价并分析模型特征变量的重要性。

同时,基于修正的通用土壤流失方程 RUSLE 模型,利用满足要求的分类数据进一步计算研究区土壤侵蚀模数,绘制该区域的土壤侵蚀分布图,结合上述土地利用/土地覆被信息,从宏观上量化子午河中下游流域的生态环境状况,具体流程如图 2 所示。

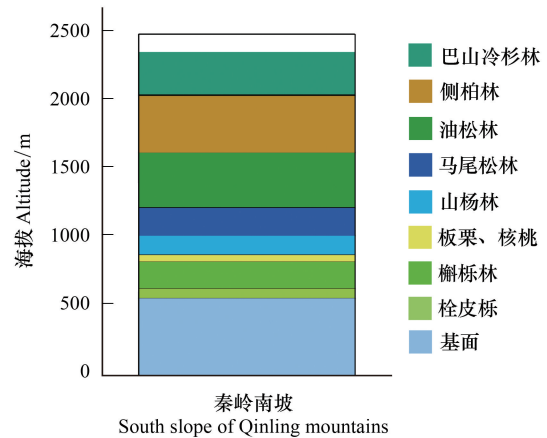


图 1 研究区植被垂直带谱

Fig.1 Vertical zone spectrum of vegetation in the study area^[30—35]

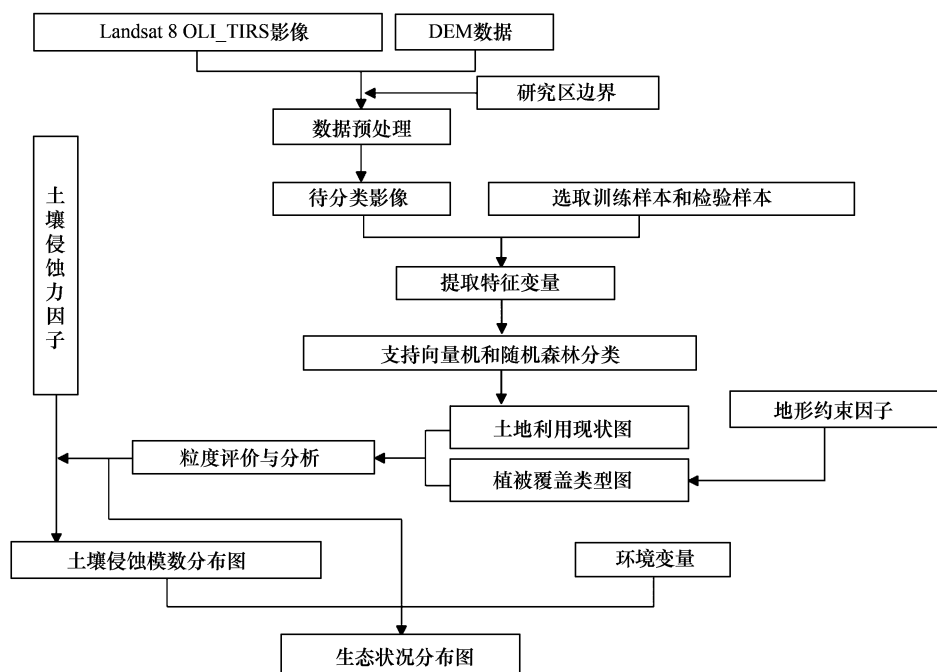


图 2 技术路线图

Fig.2 Technical roadmap

2.1 遥感数据来源及预处理

本文以 Landsat 8 OLI_TIRS 影像(15m 分辨率)为基础数据源,下载自地理空间数据云数据共享网站(<http://www.gscloud.cn/search>)。选取 2015—2019 年覆盖整个生长季(5—9 月)共 14 景影像,所选影像云量均少于 5%。首先,利用遥感图像处理平台对遥感图像进行辐射定标处理,消除传感器本身产生的误差;然后

利用 ENVI 5.3 软件的 MODTRAN 模型对遥感影像进行 Flassh 大气校正,将其保存为 TIFF 格式,再利用 ArcGIS 的数据处理模块对遥感影像进行镶嵌拼接与裁剪。为充分利用 landsat8 OLI_TIRS 影像丰富的波段光谱信息和高分辨率,还需利用 ENVI 5.3 软件将 30m 分辨率的多光谱影像与 15m 分辨率的全色波段影像进行图像融合处理,最终得到覆盖整个研究区既具有多光谱信息且分辨率为 15m 的完整遥感影像图^[34]。其他辅助数据包括 DEM(digital elevation model)高程数据、降雨侵蚀力、土壤可蚀性、水资源量、氨氮、化学需氧量、氮氧化物、二氧化硫、烟尘等数据,其中降雨侵蚀力和土壤可蚀性来源于《生产建设项目土壤流失量测算导则》(SL 773-2018),水资源量来源于陕西省水文水资源勘测中心出版的《2019 年陕西省水资源公告》,氨氮、化学需氧量等污染负荷数据来源于长江水资源保护科学研究所提供的《陕西省引汉济渭工程环境影响报告书》。

此外,本文还运用到了遥感地面样方调查数据。该数据是研究团队于 2020 年 9 月 7—11 日和 2021 年 4 月 26—30 日在调查区展开实地调查,通过 OMAP 结合遥感影像与 GPS 采样,依据典型性原则,样方在能够代表整个样地植被、地形及土壤等特征的区域随机布设,获取了研究区各海拔梯度上共 350 个植被类型调查点和 626 个土地类型调查点(图 3)。每个调查点设置 1 个样方,样方大小为 10m×10m。植被类型调查点涵盖研究区各海拔梯度上主要的 4 个植被型组,即灌草丛、落叶阔叶林、针阔混交林和针叶林;土地利用类型调查点覆盖了研究区 6 种主要的土地利用类型,即植被、居民地、裸地、建筑物、水域和农田。

2.2 制图方法

首先根据 30m 空间分辨率的 DEM 数据利用 ArcGIS 软件的 3D Analyst 模块,计算研究区的坡向并提取相应垂直带的海拔高度数据,结合研究区植被的垂直带谱信息生成研究区植被类型图的地形约束因子(图 4),分析并提取区内针叶林、阔叶林、针阔混交林、

灌丛草甸等植被型组。然后基于实地样点调查数据和遥感地面调查数据构建 RF 模型和 SVM 模型,对比两种分类结果的总体精度(Overall Accuracy)和 Kappa 系数,筛选得到符合精度要求的研究区域的植被和土地利用现状初步分类图。在地形约束因子的辅助下,通过目视解译对分类结果进行初步校正,确定植被型组。然后结合专家校正以提高分类精度,对最终分类结果进行精度评价,并分析和评价光谱特征变量对模型的重要性和适用性。随机森林模型是一种基于决策树组合的机器学习新型高效的分类器,通过 Weka 软件中的随机森林模块和 RStudio 软件进行预测和计算。利用 RF 模型进行分类通常只需要设定 2 个参数,分别是生长树的数量(ntree)和特征变量的数量(mtry);本文通过实验发现,进行植被覆盖和土地利用类型分类,当 ntree ≥ 500, mtry 分别为 5 和 4 时,分类结果的袋外误差值(OOB)最小,其值分别为 19.69%和 11.4%。SVM 模型是基于广义肖像算法的一种新型机器学习算法,利用有限的样本数量,求出学习样本的最优超平面,将不同类型的数据分开。

考虑到 Landsat 8 OLI_TIRS 遥感影像的特点,充分利用 1—7 波段的光谱信息;并且考虑到研究区位于秦巴山区,地形复杂多样,本研究选用了研究中常用的相关地形因子,如坡度(slope)、坡向(aspect)以及高程(elevation)。归一化植被指数(NDVI)、比值植被指数(RVI)、土壤调节植被指数(SAVI)等可以反映不同的植被类型,且利用多波段信息监测植被比单波段更为稳定。此外,纹理特征也是区分不同地物的重要变量。最终,共提取出 17 项分类特征变量。

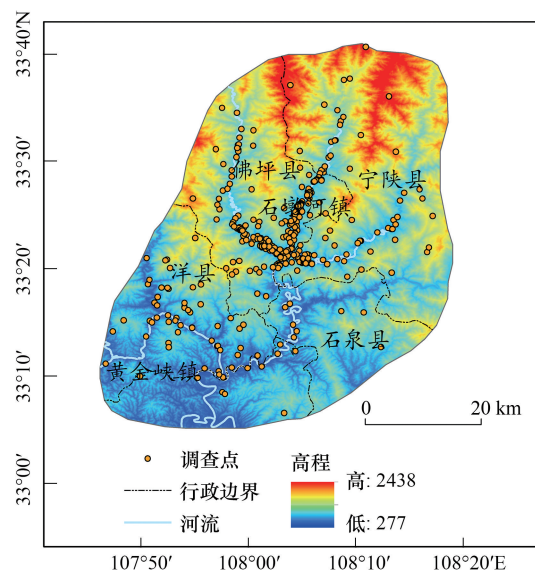


图 3 研究区概况及遥感地面调查样点分布

Fig. 3 Outline of the study area and the distribution of ground survey samples

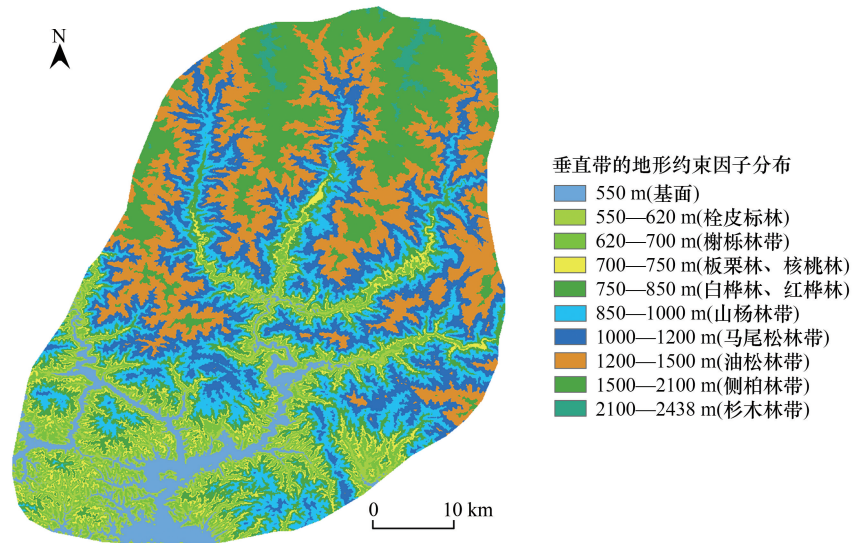


图 4 基于垂直带的地形约束因子分布图

Fig.4 Terrain constraint factor map with mountain altitudinal belt information

根据实地样点调查和遥感地面调查,共得到 350 个植被样点数据和 626 个土地样点调查数据,对两类调查数据分别随机选取其中 2/3 作为训练样本,在 Wake 软件中设置模型参数,建立 RF 模型,另外 1/3 未被选中的样点用于分类结果的精度评价。本文选用混淆矩阵(Confusion matrix)的方法对分类结果进行精度评价,以 Kappa 系数和总体精度(Overall Accuracy)作为评价指标。初步分类结果经专家校正后,在 Arcgis10.4 软件平台并进一步修正以提高分类精度。

结合研究区植被和土地利用现状分类结果,利用美国农业部提出的修正的通用土壤流失方程 RUSLE 模型^[36],基于 Arcgis10.4 软件平台 Spatial Analyst 工具模块中的栅格计算器计算出研究区的土壤侵蚀模数。其表达式为:

$$A = fR \times K \times L \times S \times C \times P \quad (1)$$

式中, A 为土壤侵蚀模数($t \text{ hm}^{-2} \text{ a}^{-1}$); R 为降雨和径流因子($\text{MJmm} \text{ hm}^{-2} \text{ h}^{-1} \text{ a}^{-1}$); K 为土壤可蚀性因子($t \text{ hm}^2 \text{ MJ}^{-1} \text{ h} \text{ mm}^{-1} \text{ hm}^{-2}$); L 为坡长因子; S 为坡度因子; C 为植被覆盖因子; P 为水土保持措施因子, C 与 P 为无量纲单位。 f 为使 A 的单位(因 R 、 K 使用美用单位)转换为国际通用单位的常数,其值为 $224.2^{[36]}$ 。

2.3 生态环境状况评价

根据《生态环境评价技术规范》(HJ192—2015),生态环境状况评价利用一个综合指数(生态环境状况指数,EI)反映区域生态环境整体的状况,指标体系包括生物丰度指数、植被覆盖指数、水网密度指数、土地胁迫指数、污染负荷指数和一个环境限制指数,各项评价指标的权重如表 1 所示。

表 1 各项指标权重

Table 1 Weight of each index

指标 Index	生物丰度指数 Biological abundance index	植被覆盖指数 Vegetation coverage index	水网密度指数 Water network density index	土地胁迫指数 Land stress index	污染负荷指数 Pollution load index	环境限制指数 Environmental limit index
权重 Weight	0.35	0.25	0.15	0.15	0.1	约束性指标

生态环境状况计算方法如下:生态环境状况指数(EI)= $0.35 \times$ 生物丰度指数+ $0.25 \times$ 植被覆盖指数+ $0.15 \times$ 水网密度指数+ $0.15 \times (100 - \text{土地胁迫指数}) + 0.1 \times (100 - \text{污染负荷指数}) + \text{环境限制指数}$,其中各指标计算方法如表 2 所示:

表 2 各项指标计算方法
Table 2 Weight of each index

指标 Index	计算方法 Method of calculation	备注 Remarks
生物丰度指数 Biological abundance index	$A_{\text{bio}} \times (0.35 \times \text{林地} + 0.21 \times \text{草地} + 0.28 \times \text{水域湿地} + 0.11 \times \text{耕地} + 0.04 \times \text{建设用地} + 0.01 \times \text{未利用地}) / \text{区域面积}$	A_{bio} 为生物丰度指数的归一化系数, 取值为 520.35
植被覆盖指数 Vegetation coverage index	$A_{\text{veg}} \times \left(\sum_{i=1}^n P_i / n \right)$	P_i 为 5—9 月像元 NDVI 月最大值的均值, n 为区域像元数, A_{veg} 为植被覆盖指数的归一化系数, 取值为 0.012
水网密度指数 Water network density index	$A_{\text{riv}} \times \text{河流长度} / \text{区域面积} + A_{\text{lak}} \times \text{水域面积} (\text{湖泊、水库、河渠}) / \text{区域面积} + A_{\text{res}} \times \text{水资源量} / \text{区域面积} / 3$	A_{riv} 为河流长度的归一化系数, 取值为 85.89, A_{lak} 为水域面积的归一化系数, 取值为 602.44, A_{res} 为水资源量的归一化系数, 取值为 87.95
土地胁迫指数 Land stress index	$A_{\text{ero}} \times (0.4 \times \text{重度侵蚀面积} + 0.2 \times \text{中度侵蚀面积} + 0.2 \times \text{建设用地面积} + 0.2 \times \text{其它土地胁迫}) / \text{区域面积}$	A_{ero} 为土地胁迫指数的归一化系数, 取值为 240.29
污染负荷指数 Pollution load index	$0.2 \times A_{\text{COD}} \times \text{COD 排放量} / \text{区域年降水总量} + 0.2 \times A_{\text{NH}_3} \times \text{氨氮排放量} / \text{区域年降水总量} + 0.2 \times A_{\text{SO}_2} \times \text{SO}_2 \text{ 排放量} / \text{区域面积} + 0.1 \times A_{\text{YFC}} \times \text{烟(粉)尘排放量} / \text{区域面积} + 0.2 \times A_{\text{NO}_x} \times \text{氮氧化物排放量} / \text{区域面积} + 0.1 \times A_{\text{SOL}} \times \text{固体废物丢弃量} / \text{区域面积}$	A_{COD} 为 COD 的归一化指数, 取值为 4.47, A_{NH_3} 为氨氮的归一化系数, 取值为 40.9, A_{SO_2} 为 SO_2 的归一化系数, 取值为 0.065, A_{YFC} 为烟尘的归一化系数, 取值为 4.16, A_{NO_x} 为氮氧化物的归一化系数, 取值为 0.52, A_{SOL} 为固体废物的归一化系数, 取值为 0.076

归一化系数 A 为一系列数据中最大值的倒数的 100 倍^[37]

根据生态环境状况指数, 将生态环境分为 5 级, 即优、良、一般、较差、和差, 如表 3 所示。

表 3 生态环境状况分级
Table 3 Classification of ecological environment

级别 Level	优 Excellent	良 Good	一般 General	较差 Poor	差 Bad
指数 Index	$EI \geq 75$	$55 \leq EI < 75$	$35 \leq EI < 55$	$20 \leq EI < 35$	$EI < 20$
描述 Describe	植被覆盖度高, 生物多样性丰富, 生态系统稳定。	植被覆盖度较高, 生物多样性较丰富, 适合人类居住。	植被覆盖度中等, 生物多样性水平一般, 较适合人类生活, 但有不适合人类生活的制约性因子出现。	植被覆盖较差, 严重干旱少雨, 物种较少, 存在着明显限制人类生活的因素。	条件较恶劣, 人类生活受到限制。

EI 为生态环境状况指数 Eco-environmental Status Index

3 实验结果与分析

3.1 分类结果与精度检验

本文分别采用 RF 模型和 SVM 模型对秦岭南坡子午河中下游流域的各种地物类型进行识别, 初步得到研究区土地利用/土地覆被类型图; 然后在地形约束因子的辅助下, 通过目视解译对分类结果进行初步校正, 确定植被型组; 最后结合专家校正得到研究区的土地利用现状和植被类型结果图(图 5, 6)。随机森林的植被分类结果显示, 研究区 59.21% 的植被类型为针叶林, 25.82% 的植被为针阔混交林, 落叶阔叶林和灌草丛仅占 14.97%; 从土地利用现状分类结果来看, 研究区大部分区域是林地, 占 70.05%, 建筑所占的百分比为 6.26%, 水域占 10.37%。支持向量机模型的植被分类结果表明, 针叶林占林地总面积的 61.33%, 针阔混交林占 29.22%, 灌草丛和落叶阔叶林分别占 5.06 和 4.39%; 土地利用的分类结果表明, 林地面积占研究区总面积的 87.59%, 建筑占 1.16%, 水域面积占 8.99%。

为了定量评价两种分类方法的分类精度, 本文分别计算模型的总体精度 (Overall accuracy)、Kappa 系数以及林地的使用者精度 (User accuracy) 和生产者精度 (Producer's accuracy) (表 4, 5)。与传统的 SVM 方法相

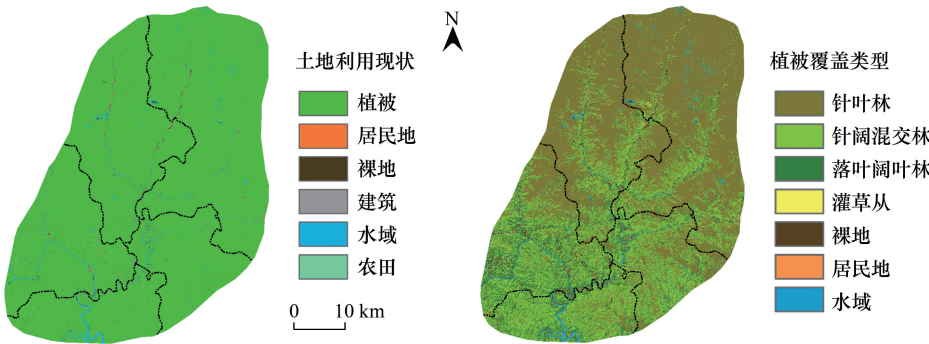


图 5 随机森林分类法

Fig.5 Random forest classification

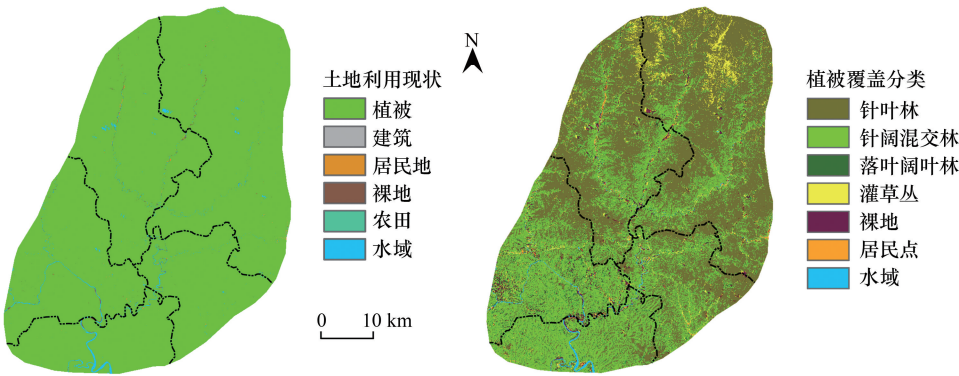


图 6 支持向量机分类法

Fig.6 Support vector machine classification

比,利用 RF 模型对植被覆盖类型进行分类的总体精度为 80.31%,提高了 4.37%;Kappa 系数为 0.734,提高了 4.2%;对于土地利用分类的总体精度为 90%,提高了 5.4%,Kappa 系数为 0.864,提高了 5.9%。通过利用样点对分类结果图进行精度验证,两种分类方法对于针叶林和灌草丛的识别度较高,对针阔混交林和落

叶阔叶林的分类精度相对较低,原因在于针阔混交林的构成较为复杂,与落叶阔叶林的光谱特征易被混淆。土地利用现状分类结果的精度较高,但两种分类方法对植被和农田的识别度较低,原因是研究区的主要作物水稻、蔬菜、玉米,其光谱特征与植被很难区分。对于模型的生产者精度和使用精度,随机森林分类法都比支持向量机分类法有了明显提升,由此可见,基于光谱特征变量和地形因子以及纹理特征的随机森林分类法,比支持向量机分类方法

在各方面精度都有了明显的提高,随机森林相对于支持向量机分类法有更好的分类效果。

表 4 两种分类方法对植被分类精度的对比

Table 4 Comparison of accuracy of two classification methods for vegetation

分类方法 Classification	总体精度% Overall accuracy	Kappa 系数 Kappa coefficient	生产者精度% Producer's accuracy	使用者精度% User accuracy
随机森林模型 Random forest classification	80.31	0.734	75.36	77.62
支持向量机分类 Support vector machine classification	75.94	0.692	70.42	71.51

表 5 两种分类方法对土地利用现状分类精度的对比

Table 5 Comparison of accuracy of two classification methods for land use status

分类方法 Classification	总体精度% Overall accuracy	Kappa 系数 Kappa coefficient	生产者精度% Producer's accuracy	使用者精度% User accuracy
随机森林模型 Random forest classification	90	0.864	89.4	90.8
支持向量机分类 Support vector machine classification	84.7	0.805	83.9	84.9

3.2 特征变量的重要性

对于多源数据的分类,在各种不同类型特征变量数量较多时,需要了解这些特征变量对分类结果的贡献,以便挑选出较重要的特征变量。本文选取的特征变量及其重要性(图 7)。总体上看,在进行植被分类时,地形因子中高程因子的贡献率占据主导地位,说明高程对于研究区的植被类型的影响极高,与实地调查和文献资料中的结果完全一致,光谱特征变量中的土壤调节植被指数(SAVI)贡献率较大;进行土地利用现状分类时,对分类结果起决定性作用的因子是归一化植被指数(NDVI),原因是 NDVI 是指示植被生长状态和植被覆盖程度的指示因子,其次就是纹理特征中的均值、水体归一化植被指数(NDWI)以及高程因子对土地利用现状分类的贡献率较大。综上分析,在利用随机森林模型分类时应尽量选择使样本差异性较大的特征变量,要结合研究区的地形特点进行选取。

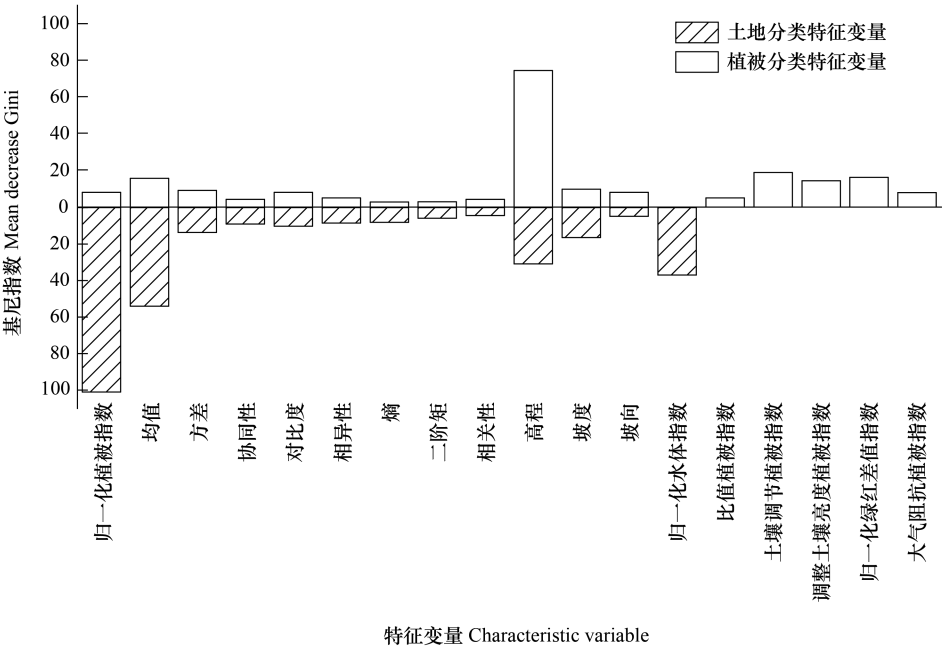


图 7 随机森林模型的特征变量及其重要性

Fig.7 Characteristic variables and their importance in Random Forest Classification

3.3 模型结果应用

3.3.1 土壤侵蚀模数计算

基于修正的通用土壤流失方程 RUSLE 模型,结合上述符合要求的土地利用/土地覆被分类结果,利用 ArcGIS 10.4 软件计算研究区的土壤侵蚀模数(图 8)。计算结果显示,研究区 91.56%的区域为微度侵蚀,土壤侵蚀模数均小于 200t km⁻² a⁻¹,仅有水源区附近侵蚀相对较严重,其中输水沿线土壤侵蚀模数为 1000—1500t km⁻² a⁻¹,属于轻度侵蚀,水源区土壤侵蚀模数较大,大部分区域土壤侵蚀模数大于 2000t km⁻² a⁻¹,主要侵蚀方式为水流侵蚀,风力侵蚀的可能性较小。只有局部区域由于河道变窄,水流流速增大导致土壤侵蚀模数为

2500—5000t km⁻² a⁻¹,属于中度侵蚀。

3.3.2 生态环境状况评价

根据《生态环境评价技术规范》(HJ192—2015),结合土地利用现状、植被覆盖类型和土壤侵蚀模数以及地面生态环境等相关参数,计算出研究区的生态环境状况指数,从而对研究区的生态环境状况进行定量化评价,得到研究区的生态环境状况分布(图9)。可以看出,研究区平均生态环境状况指数(EI)为87.12,其中74.31%的区域生态环境状况指数都大于75,生态环境状况为优;仅有15.69%的区域生态环境状况指数为55—75,生态环境状况为良,这些区域多分布在水源区附近以及洋县黄金峡水利枢纽附近。经查阅文献和实地调研,总结出水源区附近和洋县下游与周边的生态环境状况产生差异性的原因主要有以下两点:第一,水源区附近土壤受到水流侵蚀,水土流失相对较严重;第二,由于水库蓄水导致水源区附近植被遭到破坏,生物丰度下降,另外由于黄金峡水利枢纽正处于施工期,有大量污染物排放,会使下游水质变差,且洋县上游存在生活污水排放,致使片区生态环境状况略微受到影响。

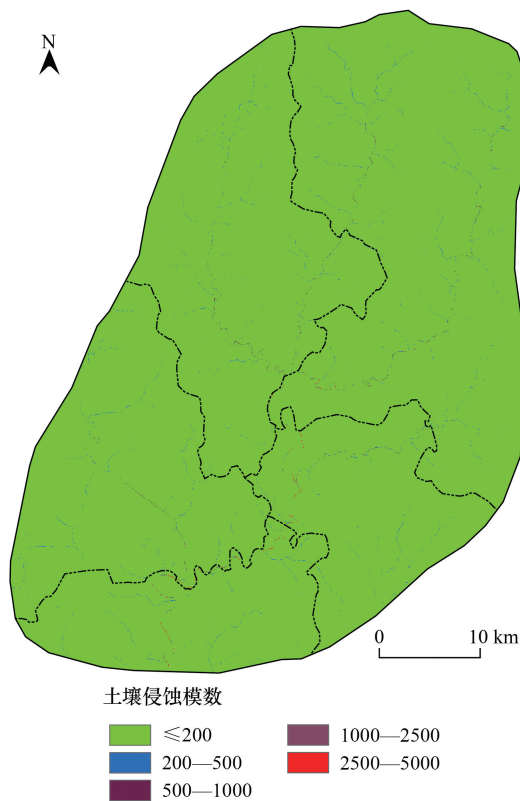


图8 研究区土壤侵蚀模数分布图

Fig.8 Distribution map of soil erosion modulus in the study area

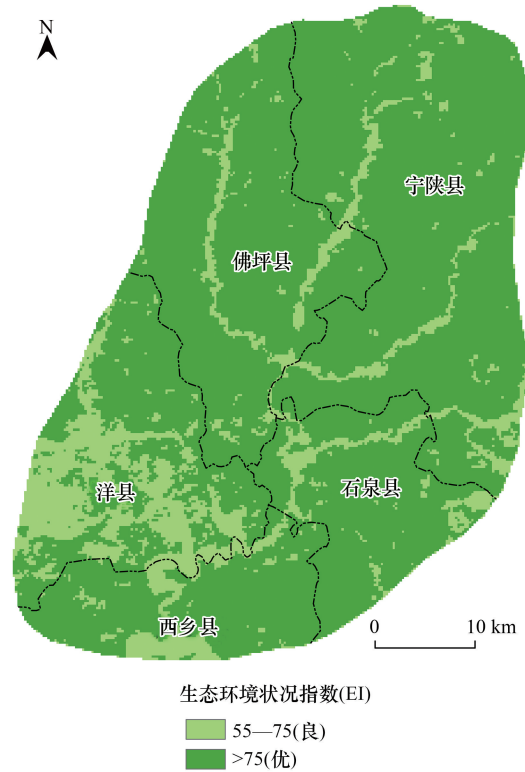


图9 研究区生态环境状况分布

Fig.9 Distribution of ecological environment in the study area

EI 为生态环境状况指数 Eco-environmental Status Index

4 结论与讨论

4.1 结论

(1)利用 Landsat 8 OLI_TIRS 遥感影像对土地利用/土地覆被信息提取的实验中,随机森林(random forest)分类法能够很好地适用于秦岭南坡子午河中下游流域土地利用/土地覆被信息提取,分类结果表明研究区59.21%的植被类型为针叶林,25.82%的植被为针阔混交林,落叶阔叶林和灌草丛仅占14.97%;从土地利用现状分类结果来看,研究区大部分区域是林地,占70.05%,建筑所占的百分比为6.26%,水域占10.37%。

(2)随机森林模型比支持向量机法对土地利用/土地覆被的信息提取更具优势,分类速度更快、精度更

高,是一种有效的分类方法。模型精度验证结果显示,随机森林模型对植被覆盖分类的总体精度比支持向量机模型提高了 4.37%;Kappa 系数提高了 4.2%;对于土地利用分类的总体精度前者比后者提高了 5.4%,Kappa 系数提高了 5.9%。

(3) 研究区 91.56% 的区域土壤侵蚀强度为微度侵蚀,输水沿线土壤侵蚀强度较高,属于轻度侵蚀,主要侵蚀方式为水流侵蚀,风力侵蚀的可能性较小;局部河道由于断面变窄,水流流速增大,导致土壤侵蚀较为严重,属于中度侵蚀。

(4) 研究区内整体生态环境状况为优,平均生态环境状况指数为 87.12,仅有 15.69% 区域生态环境状况为良,各个区域生态环境状况评价结果与实际调研结果大致相符,说明随机森林模型的分类结果能够应用于子午河中下游流域的生态环境状况评价。

4.2 讨论

(1) 就土地利用/土地覆被信息提取方法而言,以往类似的研究多基于朴素贝叶斯、CART 决策树、支持向量机和面向对象模型等,如孙亚楠、刘睿和张露洋等选取不同的特征变量基于决策树、支持向量机和面向对象模型分别对河套灌区永济灌域、新疆石河子市和辽河平原地区的土地利用类型进行了分类,分类精度相对于传统的像元相似度分析方法有了较大的提升^[38-40]。然而,这些方法存在耗时长且容易发生过拟合现象。本文以光谱特征、地形特征和纹理特征作为特征变量,构建的随机森林模型和支持向量机模型对秦岭南坡子午河中下游流域土地利用/土地覆被信息进行了提取,将两者的分类精度进行对比分析,结果表明 RF 模型相对于 SVM 模型分类速度更快,且模型的总体精度、生产者精度和使用者精度均优于 SVM 模型。

(2) 本文的局限性主要体现在以下几个方面:首先,本文的实验研究仍采用监督分类结合目视解译的方法进行的,在利用随机森林模型进行监督分类时采用的是 Landsat 8 OLI_TIRS 遥感影像,受到其分辨率和波段数量的限制,并且子午河中下游流域地形复杂,只是简单地结合光谱特征、纹理特征和地形因子等特征变量构建随机森林模型也具有一定的局限性^[41]。其次,本文虽然分析了特征变量对随机森林模型的重要性,但没有具体分析和评价各特征变量对随机森林模型分类精度会产生多大的影响^[42]。因此,后续的研究中,需要使用高分辨率的高光谱影像,并结合研究区的水文气象因子以及影像的几何特征等特征变量构建更加精确的随机森林模型,并对模型的特征参数进行更加深入的研究,实现更高精度的遥感信息提取,从而使生态环境评价更加切合实际,为有关部门提供更加准确的数据参考。

参考文献 (References):

- [1] 郭亚红, 阿布都热合曼·哈力克, 魏天宝, 木卡达斯·阿不都热合曼. 基于土地利用变化的和田地区生态系统服务价值分析. 生态学报, 2021, 41(16): 6363-6372.
- [2] 宫兆宁, 陆丽, 金点点, 邱华昌, 张强, 关晖. 土地利用/覆被变化扎龙湿地蒸散发量及生态需水量的遥感估算. 生态学报, 2021, 41(9): 3572-3587.
- [3] 菅永峰, 韩泽民, 黄光体, 王熊, 李源, 周靖靖, 佃袁勇. 基于高分辨率遥感影像的北亚热带森林生物量反演. 生态学报, 2021, 41(6): 2161-2169.
- [4] 邓晨晖, 白红英, 马新萍, 黄晓月, 赵婷. 2000—2017 年秦岭山地植被物候变化特征及其南北差异. 生态学报, 2021, 41(3): 1068-1080.
- [5] 彭羽, 王越, 马江文, 范敏, 白岚, 周涛. 基于实地调查和高光谱数据的浑善达克沙地中部植物 alpha 多样性遥感估测. 生态学报, 2019, 39(13): 4883-4891.
- [6] 帅爽, 张志, 吕新彪, 陈思, 马梓程, 谢翠容. 矿山恢复治理区植被物候与健康状况遥感监测. 农业工程学报, 2021, 37(4): 224-234.
- [7] 包旭, 李杰, 刘锋, 胡金明. 1995—2015 年缅甸土地利用/覆被变化对生态系统服务价值的影响. 生态学报, 2021, 41(17): 6960-6969.
- [8] 税燕萍, 卢慧婷, 王慧芳, 严岩, 吴钢. 基于土地覆盖和 NDVI 变化的拉萨河流域生境质量评估. 生态学报, 2018, 38(24): 8946-8954.
- [9] 李全红, 李雷, 李纯斌, 吴静, 常秀红. 基于残差 U-Net 的遥感影像土地覆盖类型自动分割. 中国土地科学, 2021, 35(1): 98-106.
- [10] 韩东, 王浩舟, 郑邦友, 王锋. 基于无人机和决策树算法的榆树疏林草原植被类型划分和覆盖度生长季动态估计. 生态学报, 2018, 38(18): 6655-6663.
- [11] 姚永慧, 张俊瑶, 索南东主. 南北过渡带 1:5 万植被类型图遥感制图案例研究. 地理学报, 2020, 75(3): 620-630.
- [12] 姚永慧, 寇志翔, 胡宇凡, 张百平. 秦巴山区马尾松林和油松林的空间分布及亚热带与暖温带界线划分. 地理学报, 2020, 75(11):

- 2298-2306.
- [13] 张俊瑶, 姚永慧, 索南东主, 郜丽静, 王晶, 张兴航. 基于垂直带谱的太白山区山地植被遥感信息提取. 地球信息科学学报, 2019, 21(8): 1284-1294.
- [14] Wang L, Dronova I, Gong P, Yang W B, Li Y R, Liu Q. A new time series vegetation-water index of phenological-hydrological trait across species and functional types for Poyang Lake wetland ecosystem. Remote Sensing of Environment, 2012, 125: 49-63.
- [15] 贾明明, 任春颖, 刘殿伟, 王宗明, 汤旭光, 董张玉. 基于环境星与 MODIS 时序数据的面向对象森林植被分类. 生态学报, 2014, 34(24): 7167-7174.
- [16] Leo B. Random Forests. Machine. Learning, 2001, 45(1): 5-32.
- [17] Pal M. Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 2005, 26(1): 217-222.
- [18] Gislason P O, Benediktsson J A, Sveinsson J R. Random Forests for land cover classification. Pattern Recognition Letters, 2006, 27(4): 294-300.
- [19] 王雪娜. 基于 SPOT6 卫星影像和随机森林模型的土地利用精细分类研究. 科技创新与应用, 2021, 11(17): 19-21.
- [20] 张雁. 基于机器学习的遥感图像分类研究[D]. 北京: 北京林业大学, 2014.
- [21] 张晓羽, 李凤日, 甄贞, 赵颖慧. 基于随机森林模型的陆地卫星-8 遥感影像森林植被分类. 东北林业大学学报, 2016, 44(6): 53-57, 74.
- [22] 邢晓语, 杨秀春, 徐斌, 金云翔, 郭剑, 陈昂, 杨东, 王平, 朱立博. 基于随机森林算法的草原地上生物量遥感估算方法研究. 地球信息科学学报, 2021, 23(7): 1312-1324.
- [23] 马玥, 姜琦刚, 孟治国, 李远华, 王栋, 刘骅欣. 基于随机森林算法的农耕地土地利用分类研究. 农业机械学报, 2016, 47(1): 297-303.
- [24] 王猛, 张新长, 王家耀, 孙颖, 箭鸽, 潘翠红. 结合随机森林面向对象的森林资源分类. 测绘学报, 2020, 49(2): 235-244.
- [25] 金宇, 周可新, 高吉喜, 穆少杰, 张小华. 基于随机森林模型的国家重点保护陆生脊椎动物物种优先保护区的识别. 生态学报, 2016, 36(23): 7702-7712.
- [26] 李治, 杨晓梅, 孟樊, 范文义. 物候特征辅助下的随机森林宏观尺度土地覆盖分类方法研究. 遥感信息, 2013, 28(6): 48-55.
- [27] 国家统计局农村社会经济调查司. 中国农业统计资料-1949-2019. 北京: 中国统计出版社, 2020.
- [28] 吴征镒, 孙航, 周浙昆, 李德铎, 彭华. 中国种子植物区系地理. 生物多样性, 2011, 19(1): 148.
- [29] 陈倩. 秦岭山地天然次生林群落物种多样性与功能多样性研究[D]. 杨凌: 西北农林科技大学, 2019.
- [30] 张巧明. 秦岭南坡中段主要植物群落及物种多样性研究[D]. 杨凌: 西北农林科技大学, 2012.
- [31] 夏列钢. 耦合“图—谱”特征的遥感影像自动分类方法研究[D]. 杭州: 浙江工业大学, 2011.
- [32] 岳明. 秦岭植物垂直带谱完整复杂. 森林与人类, 2015(2): 76-81.
- [33] 方正, 高淑贞. 秦岭太白山南北坡的植被垂直带谱. 植物生态学与地植物学丛刊, 1963, 1(S1): 162-163.
- [34] 姚永慧, 寇志翔, 胡宇凡, 张百平. 秦巴山区马尾松林和油松林的空间分布及亚热带与暖温带界线划分. 地理学报, 2020, 75(11): 2298-2306.
- [35] 李海宁. 太白山北坡植物物种多样性及其垂直分布格局研究[D]. 西安: 陕西师范大学, 2007.
- [36] 郭思琪, 韩磊, 赵永华, 员学锋, 王达, 李蕴琪. 秦岭地区土壤侵蚀时空变化及景观格局. 生态学杂志, 2019, 38(7): 2167-2176.
- [37] 中华人民共和国环境保护部. 生态环境状况评价技术规范: HJ 192—2015[S]. 北京: 中国环境科学出版社, 2015.
- [38] 孙亚楠, 李仙岳, 史海滨, 崔佳琪, 马红雨, 王维刚. 基于特征优选决策树模型的河套灌区土地利用分类. 农业工程学报, 2021, 37(13): 242-251.
- [39] 刘睿, 王志勇, 高瑞. 时序 SAR 影像的干旱地区土地利用分类应用. 测绘科学, 2021, 46(10): 90-97.
- [40] 张露洋, 雷国平, 郭一洋, 路中. 基于 Landsat 影像的面向对象土地利用分类——以下辽河平原为例. 应用基础与工程科学学报, 2021, 29(2): 261-271.
- [41] 张晓东, 朱文博, 张静静, 朱连奇, 赵芳, 崔耀平. 伏牛山地森林植被物候及其对气候变化的响应. 地理学报, 2018, 73(1): 41-53.
- [42] 汤敏. 基于面向对象和随机森林方法的 SPOT6 湟水流域土地利用/土地覆被信息提取[D]. 西宁: 青海师范大学, 2020.