

基于 Sentinel-2A MSI 特征的毛竹林刚竹毒蛾危害检测

许章华^{1,2} 周鑫^{1,3} 姚雄⁴ 李巧斯⁵ 李增禄^{2,6} 郭孝玉²

(1. 福州大学环境与安全工程学院, 福州 350108; 2. 福建省资源环境监测与可持续经营利用重点实验室, 三明 365004;
3. 福州大学地理与生态环境研究院, 福州 350108; 4. 福建工程学院建筑与城乡规划学院, 福州 350118;
5. 香港大学地球科学系, 香港 999077; 6. 世纪大学教育学院, 哥打白沙罗 47810)

摘要: 为快速、准确地检测毛竹林刚竹毒蛾 (*Pantana phyllostachysae* Chao) 危害, 基于 Sentinel-2A MSI 数据分析不同刚竹毒蛾危害等级下毛竹林像元光谱的变化, 从叶损量、绿度、含水率等多个维度选择对刚竹毒蛾危害具有响应能力的 22 个 Sentinel-2A MSI 光谱衍生指标; 经单因素方差分析 (ANOVA) 以及递归特征消除法 (Recursive feature elimination, RFE) 优选后, 得到可用于刚竹毒蛾危害识别的 10 个遥感特征, 包括 LAI、RVI、NDMVI、EVI、NDVI₇₀₅、NDVI₇₈₃、RegVI₁、RegVI₂、GVMI 和 NDWI; 将上述指标作为自变量, 虫害等级作为因变量, 建立基于 XGBoost 模型的刚竹毒蛾危害检测模型。研究发现, Sentinel-2A MSI 数据波段 6、7、8、8a 对刚竹毒蛾危害具有较强的响应能力; 红边与近红外波段参与构建的指数有效反映了竹林的受害情况; XGBoost 模型对刚竹毒蛾危害识别的总精度为 83.70%, 对不同刚竹毒蛾危害等级的识别精度依次为 94.72%、72.06%、79.77%、92.41%。因此, 利用 ANOVA-RFE 筛选 Sentinel-2A MSI 光谱特征建立的 XGBoost 虫害检测模型, 具有较高的识别精度, 可为毛竹林刚竹毒蛾危害遥感监测提供技术支持。

关键词: 毛竹; 危害检测; 刚竹毒蛾; Sentinel-2A MSI 影像; 特征优选; XGBoost

中图分类号: TP79; S763.42 **文献标识码:** A **文章编号:** 1000-1298(2022)05-0191-10

OSID:



Severity Detecting of *Pantana phyllostachysae* Chao Infestation of Moso Bamboo by Selecting Optimal Sentinel-2A MSI Features

XU Zhanghua^{1,2} ZHOU Xin^{1,3} YAO Xiong⁴ LI Qiaosi⁵ LI Zenglu^{2,6} GUO Xiaoyu²

(1. College of Environment and Safety Engineering, Fuzhou University, Fuzhou 350108, China
2. Fujian Provincial Key Laboratory of Resources and Environment Monitoring & Sustainable Management and Utilization, Sanming 365004, China
3. Academy of Geography and Ecological Environment, Fuzhou University, Fuzhou 350108, China
4. College of Architecture and Planning, Fujian University of Technology, Fuzhou 350118, China
5. Department of Earth Sciences, The University of Hong Kong, Hong Kong 999077, China
6. Faculty of Education, SEGi University, Kota Damansara 47810, Malaysia)

Abstract: *Pantana phyllostachysae* Chao (PPC) is one of the most important leaf-eating pests of bamboo forests in China. It has become a major factor threatening the health of Moso bamboo forest and restricting the high quality and sustainable development of bamboo industry. It also has the characteristics of group-occurring, periodicity, and extremely serious harm, etc. How to quickly and accurately detect the damage of the Moso bamboo forest is a problem that needs to be solved at this stage. Whereas remote-sensing products can support the quickly, accurate, and comprehensive monitoring of forest health. Therefore, Sentinel-2A MultiSpectral Instrument (MSI) data, with three bands at the red-edge position, was of great significance for pest and disease detection in forests. By screening 22 spectrally derived indicators (e. g. leaf abscission, greenness and water content) using ANOVA combined with recursive RFE, totally 10 features were finally obtained to identify PPC damage. Based on the above

收稿日期: 2022-02-24 修回日期: 2022-03-14

基金项目: 国家自然科学基金项目 (42071300)、福建省自然科学基金项目 (2020J01504)、福建省资源环境监测与可持续经营利用重点实验室开放项目 (ZD202102)、福建省高校创新团队发展计划项目 (KC190002)、中国博士后科学基金项目 (2018M630728)、晋江市福大科教园区发展中心科研项目 (2019-JJFDKY-17) 和 3S 技术与资源优化利用福建省高校重点实验室开放项目 (fafugeo201901)

作者简介: 许章华 (1985—), 男, 校聘教授, 博士, 主要从事资源环境与林情遥感研究, E-mail: fafuxzh@163.com

results, the XGBoost detection model was established to detect PCC damage with high recognition accuracy. The results showed that Sentinel-2A MSI bands 6, 7, 8, and 8a exhibited strong responses to PPC damage; the index constructed by the red-edge and near-infrared bands effectively reflected the damage to bamboo forests; the overall detection accuracy of model was 83.70% compared with 94.72%, 72.06%, 79.77%, and 92.41% for ‘healthy’, ‘mildly damaged’, ‘moderately damaged’, and ‘severely damaged’ categories, respectively. These results indicated that the XGBoost detection model provided valuable support for the large-scale monitoring of pest damage to Moso bamboo forests.

Key words: Moso bamboo; damage detection; *Pantana phyllostachysae* Chao; Sentinel-2A MSI image; optimal features selection; XGBoost

0 引言

虫害一直威胁着竹林健康,导致竹林林分质量下降,价值受损。刚竹毒蛾(*Pantana phyllostachysae* Chao)是竹的主要食叶性害虫,虫害大爆发时可致竹林成片枯死,严重影响竹林生产和竹林生态系统的健康,造成重大经济损失,威胁生态安全,成为制约竹产业健康发展的主要因素。因此,开展大面积虫害等级检测对实现精准防控具有重大意义^[1]。

近年来,光学卫星遥感时空谱分辨率显著提升,在病虫害监测中得到广泛应用^[2]。一些学者基于 Landsat 系列卫星实现了云杉大小蠹(*Dendroctonus micans* Kug)、华山松大小蠹(*Dendroctonus armandi* Tsai et Li.)等森林虫害的遥感监测^[3-4],还有一些学者基于 MODIS 系列产品实现了春尺蠖(*Apocheima cinerarius* Erschoff)等虫害监测与时序变化分析^[5]。但是,在植被健康状态机理研究中发现,仅靠可见光(400~680 nm)以及红外(>750 nm)的光谱信息建立植被健康状态检测模型存在一定的局限性,而红边(680~780 nm)是植物病虫害识别中最明显的标志之一,红边左侧的反射率主要反映了植被叶绿素含量,右侧的反射率波动反映了叶内组织结构和植物体内含水量的变化,这就说明当植被叶绿素或者叶内组织结构和植物体内含水量发生变化时,“红边”会产生位移,从而揭示植被健康状态^[6]。现有研究表明基于红边位置构建的森林健康监测模型优于其他模型,并且与其他波段相比,红边波段和近红外波段更早监测到植物的应激反应^[7]。而常见的星载遥感,如 Landsat 系列、HJ 系列、SPOT 系列等卫星,只能获取到地物可见光以及红外的部分信息,难以获取到植被红边位置的光谱信息,Sentinel-2A 卫星解决了此问题,在林业相关领域得到了广泛应用,尤其是其红边位置信息,在森林火灾、森林冠层、森林健康诊断等方面均得到了良好应用^[8-10]。但是,Sentinel-2A MSI 在竹林虫害检测中的应用还鲜见报道。

近些年来,刚竹毒蛾危害下的寄主特征研究取

得了阶段性进展,学界对虫害胁迫下寄主的外部形态、内部生理现象以及响应机理已经有了一定的认识。相关研究发现^[11],703.43~898.56 nm 光谱对刚竹毒蛾危害具有较强的判别能力,并且受害叶片的光谱反射率明显低于健康叶片,虫害等级越高,其反射率越低。研究者基于刚竹毒蛾危害下寄主表征的变化,构建各种检测模型以反映毛竹叶片的受害程度,对比分析不同受害程度竹叶中相对叶绿素含量、含水量、含氮量以及光谱特征的差异,构建了叶片尺度的叶绿素反演模型以及虫害检测模型^[11-12]。但是,基于星载遥感实现刚竹毒蛾危害监测的研究还较为少见。

机器学习的快速发展也为虫害检测模型的构建提供了更多选择,如人工神经网络、支持向量机、随机森林模型等^[13-14]。一些研究者利用随机森林、Fisher 判别分析、XGBoost 等方法开展了不同尺度的虫害提取研究并获得了较好的成果^[15-17]。在对刚竹毒蛾危害下叶片尺度寄主特征研究的过程中,先后比较了 RF、XGBoost 和 Fisher 判别分析等模型的虫害判别能力,其中,XGBoost 模型的精度较高、稳定性较强^[12,17],并且 XGBoost 模型在目标函数中加入的正则项,可以有效地防止过拟合。因此,XGBoost 凭借其易用性以及高性能在刚竹毒蛾危害下叶片尺度寄主特征研究中取得优异表现^[18-19]。星载遥感为了实现分类,往往需要设置大量的特征,过多的数据特征可能会造成“Hughes”现象,即“维数灾难”^[20],因此,从大量原始特征中找到最优特征参与分类是首要问题。相关方法较多,像单因素方差分析(ANOVA)、递归特征消除法(Recursive feature elimination,RFE)等特征优选方法已经在诸多领域得到了应用,已证明其有效性^[21-22]。因此,在模型建立之前利用 ANOVA-RFE 对所选指标进行筛选可以有效提高分类效率和精度。

本文将叶片尺度下不同等级刚竹毒蛾危害的差异上升到像元尺度,分析像元尺度下不同虫害等级的光谱差异,优选出最优光谱衍生指标构建 XGBoost 虫害检测模型,以期竹林刚竹毒蛾危害

大尺度监测以及竹林防护提供理论依据以及技术支持。

1 材料与方法

1.1 试验区与数据

本研究以福建省南平市顺昌县为试验区。该县位于福建省西北部,坐标为 117°29′~118°38′E, 26°38′~27°12′N,县域面积 1 985 km²,以山地、丘陵地貌为主,林地面积约 17 万 hm²,其中竹林面积逾 4 万 hm²;受亚热带海洋性季风气候影响,干湿分明,气候温和。顺昌县是我国南方重点林区,是全国首批“竹子之乡”,国家木材战略储备基地县、森林质量精准提升示范县,并且完成了全国首单竹林碳汇交易。但是各类森林病虫害,诸如刚竹毒蛾、马尾松毛虫等害虫的发生面积居高不下,严重制约了当地林业经济的可持续发展。

于 2020 年 10 月在试验区设置 267 块 15 m × 15 m 的样地开展毛竹林调查,收集了不同刚竹毒蛾危害样本的林分结构参数、虫害状况等信息,并参照相关标准将虫害等级划分为无危害(Healthy, H)、轻度危害(Mildly damaged, Mi)、中度危害(Moderately damaged, Mo)与重度危害(Severely damaged, S)4 个等级^[13-14]。在欧空局哥白尼数据中心(<https://scihub.copernicus.eu/>)上获取覆盖试验区的 Sentinel-2A MSI 遥感影像 3 景,卫星过境时间为 2020 年 10 月 23 日,与野外调研时间接近。利用 SNAP 软件对其进行大气校正、超分辨率等处理,借助矢量边界数据裁剪试验区影像,并利用“基于片层-面向类”(Film-based & class-oriented, FB-CO)竹林提取算法^[23]提取试验区竹林信息(图 1)。

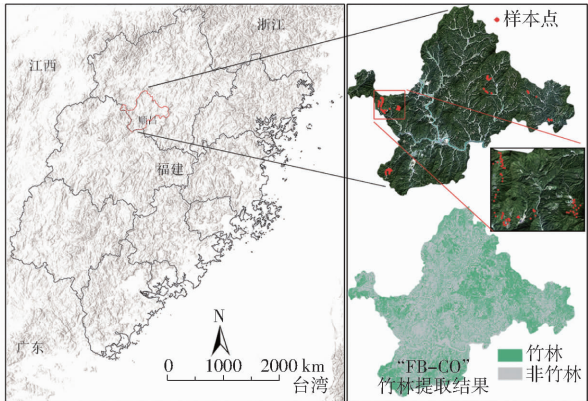


图 1 试验区及数据处理

Fig. 1 Study area and data-processing

1.2 特征优选方法

1.2.1 寄主特征及遥感指标分析

被刚竹毒蛾侵染后,毛竹林的生理特征发生改变。叶片尺度研究发现^[12,17-18],受害状态下的毛竹

叶片叶面积、叶片结构、叶绿素、含水率及光谱特征等均表现出与健康叶片的显著差异性;在利用无人机高光谱数据提取不同刚竹毒蛾危害等级时发现,与其他植被指数相比,SAVI、NDI、GNDVI 等指数对刚竹毒蛾危害具有更高的相关性^[24];而到了星载遥感上,LAI、NDVI 以及特征光谱构建的 RF 检测模型可以很好地反映遥感尺度下刚竹毒蛾危害程度。综合前期研究成果及 Sentinel-2A MSI 波段设计的特殊性,本研究从叶损量、绿度、含水率和红边 4 个维度选择了 22 个指标,分别为归一化差分红边植被指数(NDVI₇₀₅、NDVI₇₄₀、NDVI₇₈₃)^[8]、叶面积指数(LAI)^[25]、比值植被指数(RVI)^[26]、增强植被指数(EVI)^[26]、归一化植被指数(NDVI)^[26]、归一化差值山地植被指数(NDMVI)^[27]、转换型归一化差值植被指数(TNDVI)^[28]、土壤调整植被指数(SAVI_{0.1}, SAVI_{0.2}, ..., SAVI_{0.5})^[28]、修正的土壤调整植被指数(MSAVI)^[28]、缨帽变换湿度分量(WI)^[29]、全球植被湿度指数(GVMI)^[30]、归一化差分水体指数(NDWI)^[30]、改进归一化差分水体指数(MNDWI)^[30]、红边植被指数(Red edge vegetation index_{1,2,3}, RegVI₁, RegVI₂, RegVI₃)。挖掘像元尺度下的刚竹毒蛾危害遥感响应特征,同时为了统一所有指数的数量级,对其做了归一化处理,使得 22 个指标值统一为 0~1。其中,红边植被指数 RegVI₁、RegVI₂、RegVI₃ 计算公式分别为

$$RegVI_1 = \frac{R_{eg783} - R_{eg740}}{R_{eg783} + R_{eg740}} \quad (1)$$

$$RegVI_2 = \frac{R_{eg783} - R_{eg705}}{R_{eg783} + R_{eg705}} \quad (2)$$

$$RegVI_3 = \frac{R_{eg740} - R_{eg705}}{R_{eg740} + R_{eg705}} \quad (3)$$

式中 R_{eg705} 、 R_{eg740} 、 R_{eg783} ——Sentinel-2A 的 5、6、7 波段反射率

1.2.2 遥感响应特征优选

上述提取的众多特征并非都对不同虫害等级分类有意义,为此,通过特征选择剔除众多特征中对虫害响应相对不敏感的特征,找到最适合分类的特征子集,可以有效提高分类效率和精度。本研究利用单因素方差分析与递归特征消除法对所选指标进行优化。单因素方差分析(ANOVA)是一种假设检验,用来分析、分解试验数据的波动性,然后将某影响因素下的试验数据间可能存在的系统性波动与随机波动加以比较,推断各总体均值之间是否存在显著性差异,在均值不等的情况下,单因素方差分析可以很好地讨论因素对试验结果的影响,从而直观地展示各指标对虫害的响应能力^[31]。递归特征消除法

(Recursive feature elimination, RFE)是依据各指标在机器学习模型中的重要性排序,删去末位指标并反复迭代得出最优指标的方法^[32],算法中有 3 个关键参数,分别是“estimator”,“step”和“n_features_to_select”。“estimator”是 RFE 的基础评估器,以评估选定特征的性能(本研究为 XGBoost);“step”设置为 1,这意味着在每次迭代中添加或删除的特征数为 1;“n_features_to_select”表示最终选择的特征数量。

ANOVA 的结果只能够解释特征对虫害的响应能力,难以客观得出最佳特征数量,需要浪费大量的时间做排列组合实验;而 RFE 可以依据各指标在机器学习模型中的重要性排序,客观得出最佳特征数量。本文尝试将 ANOVA 与 RFE 结合,二者的结合既可以从机理上解释特征被选择的原因,又可以客观选择最佳特征数量,从而获取更高的精度。ANOVA-RFE 特征优选的主要步骤是:①利用 ANOVA 分析 22 个光谱衍生指标对不同刚竹毒蛾危害等级的响应能力。②将选出的光谱衍生指标放入 RFE 中,不断迭代确定参与分类的最优指标数量。③结合 ANOVA 与 RFE 的结果确定虫害响应的最优特征。

1.3 刚竹毒蛾虫害检测模型

极致梯度提升(Extreme gradient boosting, XGBoost)算法是一种集成学习模型,由梯度提升决策树算法(Gradient boosting decision tree, GBDT)改进而来,在机器学习和数据挖掘领域中有着深远的影响。传统的 GBDT 在优化时只用到了一阶导数,而 XGBoost 将目标函数泰勒展开至二阶,尽可能地保留原函数的有效信息,并添加正则化项以避免过拟合;此外, XGBoost 还设计了针对稀疏数据的算法,无需对确实特征进行填充预处理^[20]。该模型的理论框架及具体说明参见文献[33]。研究表明, XGBoost 经过适当调参后,可以得到比一般机器模型更好的分类结果。XGBoost 考虑正则化项,目标函数定义式为

$$F_{obj} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f(k)) \tag{4}$$

其中 $\Omega(f(k)) = \gamma T + \frac{1}{2} \lambda \| \omega \|^2 \tag{5}$

式中 $\sum_i l(y_i, \hat{y}_i)$ ——损失函数
 $\sum_k \Omega(f(k))$ ——正则化项
 $\hat{y}_i, y_i, T, \omega$ ——预测输出、label 值、叶子节点数、叶子权重
 $f(k)$ ——第 k 个树模型
 γ ——叶子树惩罚正则项,具有剪枝作用

λ ——叶子权重惩罚正则项,防止过拟合

1.4 试验设计与评价方法

按照 7:3 的比例将 1 066 个像元随机划分为训练集和测试集,并进行 6 次交叉试验。统计 6 次交叉验证结果,采用总体分类精度(Overall accuracy, OA)、分级生产者精度(Producer's accuracy, PA)、Kappa 系数(Kappa coefficient)进行精度评价^[12],OA 表示所有类别被正确分类的测试样本占样本总体的比例,反映了分类结果的系统准确性;PA 表示当前类被正确分类的样本数目与预测结果中该类总数的比值,反映了分类结果中当前类漏检情况;Kappa 系数表示预测结果内部的一致性。

2 结果与分析

2.1 Sentinel-2A MSI 遥感光谱特征分析

将实地调研涉及到的 1 066 个像元按照不同虫害等级分别计算反射率均值发现,在 Sentinel-2A MSI 中,毛竹林的光谱曲线与大部分绿色植物的光谱曲线特性一致,存在明显的峰谷特征,但随着刚竹毒蛾危害加重,其光谱特征会发生变化,呈现出“虫害等级越高,红边位置、近红外波段的反射率越低”的特征(图 2)。每年 2 月起,越冬代幼虫开始化蛹为虫,不断蚕食毛竹叶片,食出的缺口使得竹叶中的水分迅速流失,导致叶绿素的合成速率下降,从而影响光合作用速率。随着失叶程度的不断加剧,寄主光合作用的能力不断被削弱,当光合作用减弱到一定程度,光能吸收-释放的平衡会被打破,导致竹体内的水分无法被有效耗解,由此引发恶性循环,竹节内则不断干枯。同时,叶片的多孔薄壁细胞结构也会遭到破坏,多种因素共同导致光谱“红边”斜率和近红外波长反射率降低。刚竹毒蛾虫害的爆发往往是从源到面、从小到大的过程。因此,随着时间不断推移,越冬代、第一代、第二代刚竹毒蛾危害先后爆发,毛竹林受灾面积在 10 月下旬累积到极大值,且此时不同虫害等级之间的差异性最大。

总体而言,不同虫害等级之间距离越近可分性越差,距离越远可分性越好,其中 M_i 和 M_o 的光谱曲线最相似且区分难度最高。若只考虑样本均值,

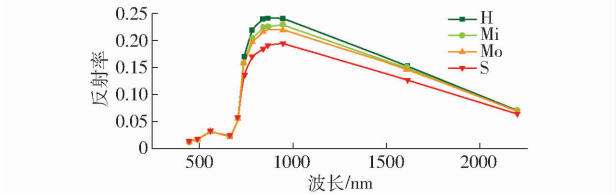


图 2 不同虫害等级像元的光谱反射率
Fig. 2 Reflectance spectra of bamboo forest with differing degrees of pest damage

不同虫害等级的波段 2 (Blue)、4 (Red)、6 (Red-edge₇₄₀, Reg₇₄₀)、7 (Red-edge₇₈₃, Reg₇₈₃)、8 (Near-infrared, NIR)、8a (Narrow near-infrared, narrow NIR) 和 11 (Short-wave infrared₁₆₁₀, SWIR₁₆₁₀) 均表现出明显的变化趋势(图 3)。随着虫害等级的提高,波段 2 以及波段 4 呈现明显的上升趋势,波段 6、7、8、8a 和 11 呈现出明显的下降趋势,这与不同波段表征的

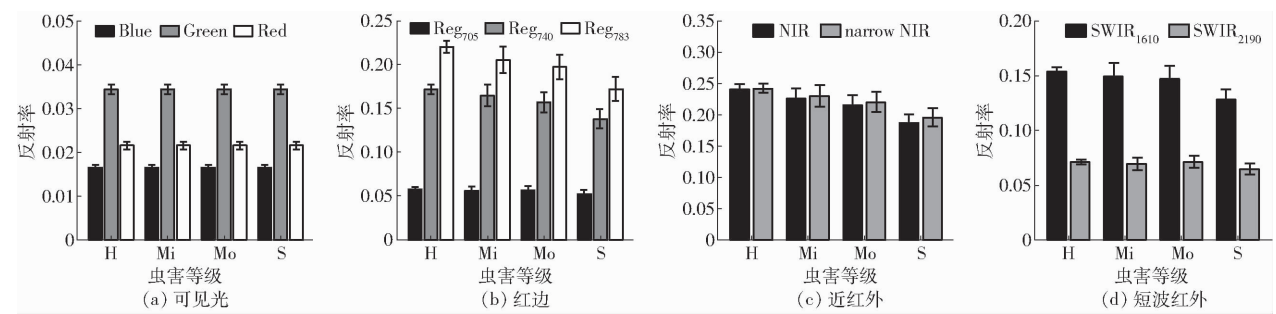


图3 原始光谱分组均值图
Fig.3 Raw spectral grouping means

2.2 基于 ANOVA-RFE 的遥感响应特征优选

2.2.1 遥感指标获取

大多数毛竹林刚竹毒蛾遥感响应特征都可以通过遥感影像直接获取,但是也有一些特征,如 LAI,需要对地面指标进行模型构建,再通过遥感进行反

演,且反演方法众多,不同方法具有不同的优势,相对而言,多元线性回归模型对于计算连续性变量具有简便、直观的优势,且更容易解释其机理;黑盒模型可能具有更高的估测精度^[27]。对所有指数进行归一化处理。具体指标如图 4 所示。

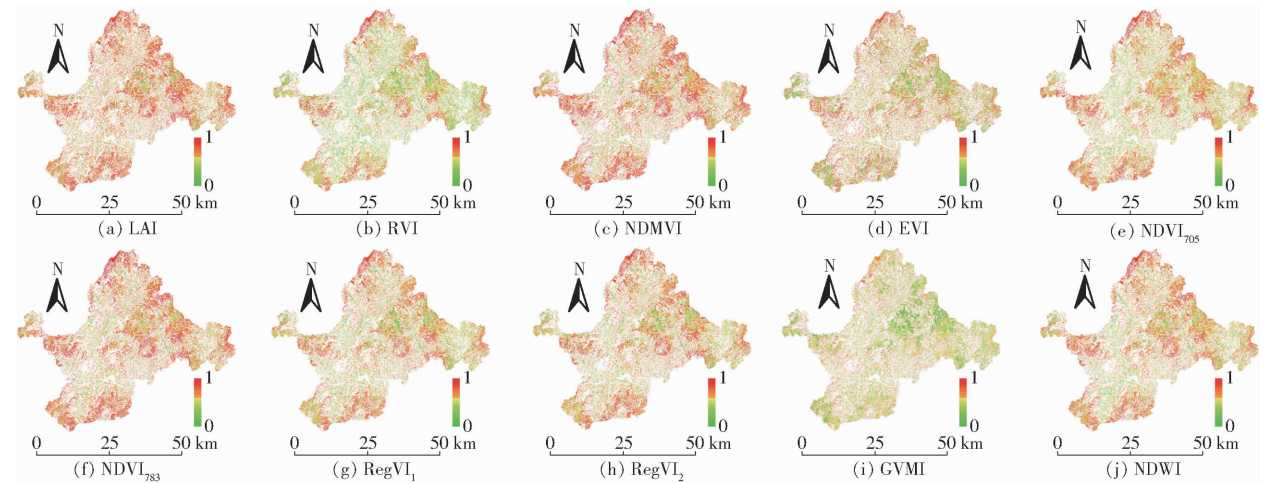


图4 遥感指标专题图
Fig.4 Thematic maps of remote sensing indices

2.2.2 基于 ANOVA-RFE 的特征优选与分析

利用 ANOVA 分析像元光谱对不同虫害等级的响应能力,结果显示(表 1),波段 6、7、8、8a 具有较强的虫害检测能力,对无危害-中度危害、无危害-重度危害、轻度危害-重度危害的响应能力达到了显著水平,其中无危害-重度危害达到了极显著水平。对指数而言(表 2),除了 MNDWI 外所选的 21 个指数均对虫害具有较好响应能力,均表现出虫害等级距离越近可分性越差、距离越远可分性越好的特点。其中以波段 4、5 (Red-edge₇₀₅, Reg₇₀₅)、6、7、8 构建的指数对虫害等级更敏感,如 NDVI、NDMVI、NDVI₇₀₅、

表1 原始光谱单因素方差分析结果

Tab.1 One-way ANOVA results of raw spectra						
波段	H-Mi	H-Mo	H-S	Mi-Mo	Mi-S	Mo-S
2	0.213 8	0.104 6	0.021 5 *	0.840 6	0.482 8	0.564 4
3	0.900 4	0.778 0	0.199 3	0.921 3	0.449 9	0.481 0
4	0.452 1	0.074 1	0.070 1	0.383 8	0.573 8	0.550 5
5	0.736 7	0.761 7	0.060 3	0.990 0	0.251 5	0.275 6
6	0.353 2	0.045 5 *	<0.000 1 **	0.481 9	0.020 6 *	0.086 1
7	0.120 6	0.014 0 *	<0.000 1 **	0.565 6	0.023 2 *	0.064 2
8	0.147 9	0.016 6 *	<0.000 1 **	0.529 7	0.014 3 *	0.056 4
8a	0.256 2	0.032 1 *	<0.000 1 **	0.526 2	0.029 3 *	0.094 5
11	0.521 0	0.343 5	0.000 1 **	0.867 0	0.054 3	0.066 0
12	0.654 7	0.917 6	0.049 5	0.734 2	0.349 8	0.207 0

注: * 表示 0.05 水平下显著, ** 表示 0.01 水平下极显著,下同。

NDVI₇₈₃等,各等级虫害均达到了极显著水平。所选的不同维度指标中,叶损量在所有虫害之间都有较好的响应能力,绿度指数对虫害均有较好的响应能力,但是对无危害-轻度危害、轻度危害-中度危害的

响应能力较差,而红边指数可以对此提供很好的补充,但是含水率指标下各指数对虫害的响应能力结果并不理想,较难单独作为虫害等级判别因子存在。

表 2 相关指数单因素方差分析结果
Tab.2 One-way ANOVA results of correlation indices

指数	H - Mi	H - Mo	H - S	Mi - Mo	Mi - S	Mo - S
LAI	0.021 2 *	<0.000 1 **	<0.000 1 **	0.016 0 *	<0.000 1 **	0.003 8 **
NDVI	0.006 1 **	<0.000 1 **	<0.000 1 **	0.000 6 **	<0.000 1 **	0.001 0 **
NDMVI	0.005 9 **	<0.000 1 **	<0.000 1 **	0.000 7 **	<0.000 1 **	0.000 5 **
TNDVI	0.006 5 **	<0.000 1 **	<0.000 1 **	0.000 7 **	<0.000 1 **	0.001 0 **
SAVI _{0.1}	0.007 7 **	<0.000 1 **	<0.000 1 **	0.209 3	0.000 7 **	0.018 9 *
SAVI _{0.2}	0.018 1 *	0.000 2 **	<0.000 1 **	0.305 8	0.002 0 **	0.028 9 *
SAVI _{0.3}	0.027 7 *	0.000 4 **	<0.000 1 **	0.347 6	0.003 0 **	0.033 0 *
SAVI _{0.4}	0.035 8 *	0.000 8 **	<0.000 1 **	0.370 4	0.003 7 **	0.035 2 *
SAVI _{0.5}	0.042 6 *	0.001 1 **	<0.000 1 **	0.384 5	0.004 2 **	0.036 4 *
RVI	0.001 4 **	<0.000 1 **	<0.000 1 **	0.000 5 **	<0.000 1 **	0.001 1 **
EVI	0.098 8	0.002 9 **	<0.000 1 **	0.3832	0.006 8 **	0.046 0 *
MSAVI	0.061 0	0.001 9 **	<0.000 1 **	0.391 0	0.004 9 **	0.036 1 *
WI	0.606 9	0.660 5	0.001 0 **	0.959 3	0.100 2	0.080 7
NDWI	0.251 6	0.158 5	<0.000 1 **	0.852 4	0.008 7 **	0.006 1 **
MNDWI	0.368 4	0.883 4	0.143 1	0.440 7	0.706 5	0.208 1
GVMi	0.364 2	0.004 0 **	0.000 2 **	0.080 9	0.013 3 *	0.332 8
NDVI ₇₀₅	0.004 2 **	<0.000 1 **	<0.000 1 **	0.020 3 **	<0.000 1 **	0.015 4 *
NDVI ₇₄₀	0.001 9 **	<0.000 1 **	0.260 3	0.000 8 **	<0.000 1 **	0.002 1 **
NDVI ₇₈₃	0.004 6 **	<0.000 1 **	<0.000 1 **	0.007 3 **	<0.000 1 **	0.002 2 **
RegVI ₁	0.259 0	0.006 2 **	<0.000 1 **	0.090 0	0.001 7 **	0.046 3 *
RegVI ₂	0.058 6	0.004 6 **	<0.000 1 **	0.335 5	0.010 0 *	0.061 7
RegVI ₃	0.003 7 **	0.088 2	0.020 0 *	0.325 0	0.782 3	0.530 0

将所有指标放入 RFE 中,结果如图 5 所示,分类精度随着参与分类的特征数目的增加呈现“先上升后降低逐渐趋于稳定”的特征,当特征数等于 10 时,分类精度最高,因此确定 10 个特征构建最优特征子集。

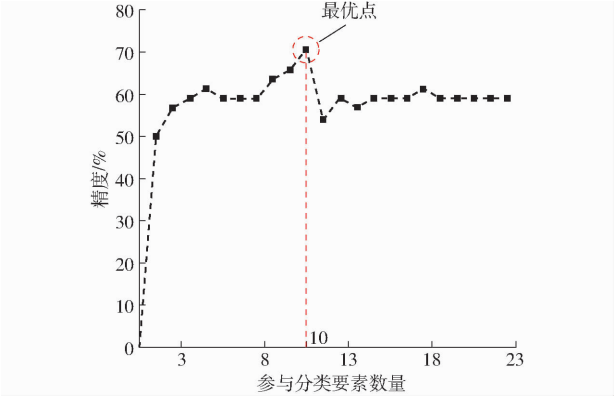


图 5 RFE 结果
Fig.5 Recursive feature elimination (RFE) results

根据实验 1 ~ 6 的重要性排序可知(图 6),10 个特征变量的贡献度由大到小依次为 RegVI₁、

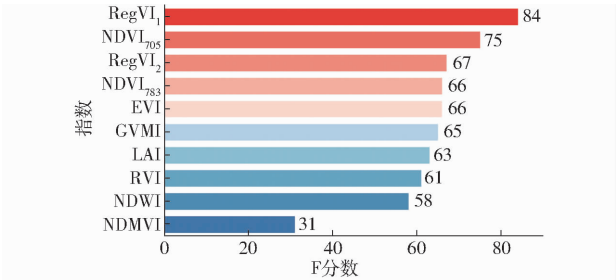


图 6 重要性排序
Fig.6 Importance ranking

NDVI₇₀₅、RegVI₂、NDVI₇₈₃、EVI、GVMi、LAI、RVI、NDWI 和 NDMVI。总的来说,选取的特征变量对刚竹毒蛾危害具有很好的响应能力,但是不同虫害等级的样本之间仍存在较大的差异性(图 7)。例如,随着虫害等级的加重,EVI 呈现出明显下降趋势而 NDWI 呈现出明显上升趋势;但从箱线图(图 7)看,即便相同等级的样本,其值也存在较大区别。不同维度的数据也表现出不同的趋势,绿度指标随着刚竹毒蛾危害等级的加重呈现下降的趋势;叶损量指标和红边指标与绿度一样,也表现出随着虫害的加

重而下降的趋势,且趋势更为明显;但含水率指标与其他指标明显不同,随着虫害的加重,该维度指标呈现总体上升局部下降的趋势。含水率指标表现出的

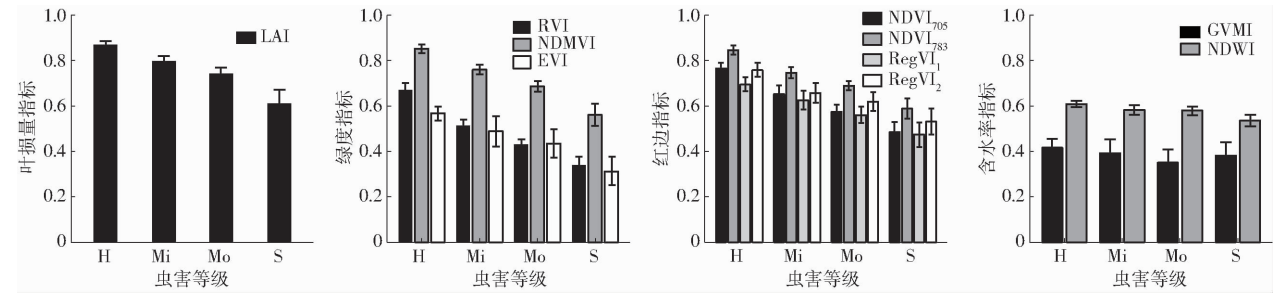


图 7 优选指标分组均值图
Fig. 7 Grouped mean values of selected optimal indices

2.3 基于 XGBoost 的刚竹毒蛾危害检测模型建立与评价

基于试验组样本数据,将上文特征优选后的 10 个指标放入 XGBoost 模型,以 10 个指标为自变量,无危害、轻度危害、中度危害及重度危害为因变量,并利用网格搜索交叉验证 (Grid search and cross validation, GridSearchCV) 调整模型重要参数 (n_estimator 为 40, max_depth 为 2, gamma 为 0.1, learning_rate 为 0.07)。

将验证集样本数据代入训练好的 XGBoost 虫害检测模型,并计算不同刚竹毒蛾危害等级的 OA、PA 和 Kappa 系数,结果如表 3 所示。结果显示,模型对刚竹毒蛾危害具有较好的检测效果,检测精度达到 81% 以上,Kappa 系数也超过了 0.81;试验 1~6 的检测精度分别为 82.93%、85.10%、82.75%、83.76%、84.48% 和 83.17%,总平均精度为 83.70%,最高和最低精度之间相差不到 3 个百分点,表明所建立模型具有较好的稳定性。细分各健康等级的分级结果(表 3),无危害的识别精度最高达到了 94.72%;其次为重度危害,识别精度为 92.41%;再次为中度危害,识别精度为 79.77%;轻度危害的识别精度最低,为 72.06%。在 6 个试验中,试验 2 总体精度最高,但是该次试验对轻度危害的识别精度较低,比 6 次试验轻度危害平均精度低了 3.56 个百分点;试验 3 的总体精度最低,但是轻度危害与中度危害的识别精度比该类危害下平均精度分别高了 0.66 个百分点和 0.56 个百分点。不同刚竹毒蛾危害等级的占比也不同,无危害占比最高,为 45.22%;其次为轻度危害,占比为 21.79%;最后是中度危害以及重度危害,占比分别为 17.89%、16.21%。刚竹毒蛾重度危害及无危害有着较为明显的集聚现象(图 8),重度危害竹林主要分布在西部以及中部偏东,轻度危害与中度危害围绕着重度危害离散分布,与刚竹毒蛾危害蔓延“由点到面、从

趋势与理论不一致,这可能是因为虫害等级的加深导致毛竹林冠层叶量减少,暴露出更多的林下植被,进而反映到该样本光谱上,使得光谱值异常抬高。

表 3 验证数据集的 OA、PA 和 Kappa 系数
Tab.3 Overall accuracy (OA), producer accuracy (PA), and Kappa coefficient of validation dataset

试验序号	PA/%				OA/%	Kappa 系数
	H	Mi	Mo	S		
1	94.74	73.33	84.01	87.84	82.93	0.8061
2	97.37	68.50	75.00	95.58	85.10	0.8149
3	92.31	72.73	80.33	91.14	82.75	0.8082
4	94.44	70.83	80.37	93.90	83.76	0.8026
5	95.35	73.33	81.08	94.74	84.48	0.8355
6	94.12	73.64	80.82	91.25	83.17	0.8089
平均值	94.72	72.06	79.77	92.41	83.70	0.8127

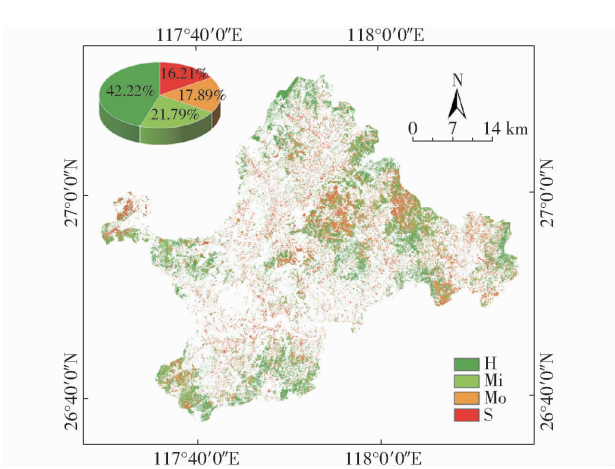


图 8 虫害提取结果

Fig. 8 Damage extraction results

小到大”特征相符,进一步佐证了提取结果的准确性。

总体而言,XGBoost 虫害检测模型对不同虫害等级的识别精度达到了预期,但对轻度危害的识别效果仍有待提高。

2.4 讨论

在虫害监测领域,早期研究者利用地面实测高光谱对农业、林业微观尺度的光谱变化做了大量的研究,发现虫害胁迫下的植被叶片和冠层反射辐射的总量及方向受到了影响,从而使得受胁迫植被的

光谱特性发生变化^[34]。现有研究表明,虫害胁迫下植被光谱曲线可见光反射率整体呈上升趋势,而在近红外区域呈现下降的特征;在可见光区的“蓝边”“绿边”“黄边”“红谷”及红光过渡到近红外的“红边”是描述植物健康状况的重要指示波段,其中“红边”是植物曲线最明显的特征^[35]。早在 20 世纪下半叶,就有研究者对植被红边波段开展研究,得出红边波段是叶绿素含量的最佳遥感描述特征之一,是植被光谱最明显的标志之一的结论,红边信息主要包括红边位置、红边峰值以及红边斜率,主要反映植被覆盖度或叶面积指数、植被叶绿素含量等,从而反映植物生长状况^[36]。之后,研究者不断拓宽红边信息的应用领域,红边信息逐渐在森林健康诊断、水体藻类等领域中大放异彩。

所有在轨卫星中,Sentinel-2 号卫星是唯一具有 3 个红边波段的卫星,而红边波段的组合可以很好地揭示森林的健康状态。在对云杉色卷蛾(*Choristoneura fumiferana*)胁迫下云杉受灾程度研究时发现,红边波段衍生的光谱指标相较于传统指标效果更好^[10],这在本研究中有体现,红边波段参与构建的指标在虫害检测中的整体表现优于传统指标。本文对比了不同刚竹毒蛾危害等级下 Sentinel-2A MSI 与地面高光谱数据(Ground hyper spectral data,GS)的反射率(图 9),发现 GS 的反射率均高于 Sentinel-2A MSI;当比较二者不同虫害等级对应波长反射率时发现(图 10),GS 数据的“绿峰红谷”现象比 Sentinel-2A MSI 更明显,并且不同虫害等级在 GS 数据 665 nm 比 Sentinel-2A MSI 红波段具有更明显的差异性(图 10);总体趋势上,二者在 1 300 nm 前趋势上均具有很明显的趋同性,但到了 1 500 nm 以后,二者在无危害、轻度危害以及中度危害的反射率变化基本保持一致,重度危害反射率却出现明显的差异,此时重度危害的 GS 反射率会高于其他虫害等级的反射率,这个现象与 Sentinel-2A MSI 表现结果正好相反。在虫害敏感波段上,GS 与 Sentinel-2A MSI 也具有极高的相似性,在早期地面研究中,发现 703 ~ 898 nm 是刚竹毒蛾危害下的毛竹叶片的 光谱特征波段^[1],而在本文的研究中,Sentinel-2A MSI 的 5、6、7、8、8a 这 5 个波段是刚竹毒

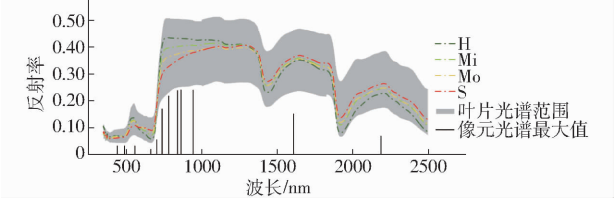


图 9 不同虫害等级的叶片反射率

Fig. 9 Leaf reflectance with different degrees of pest damage

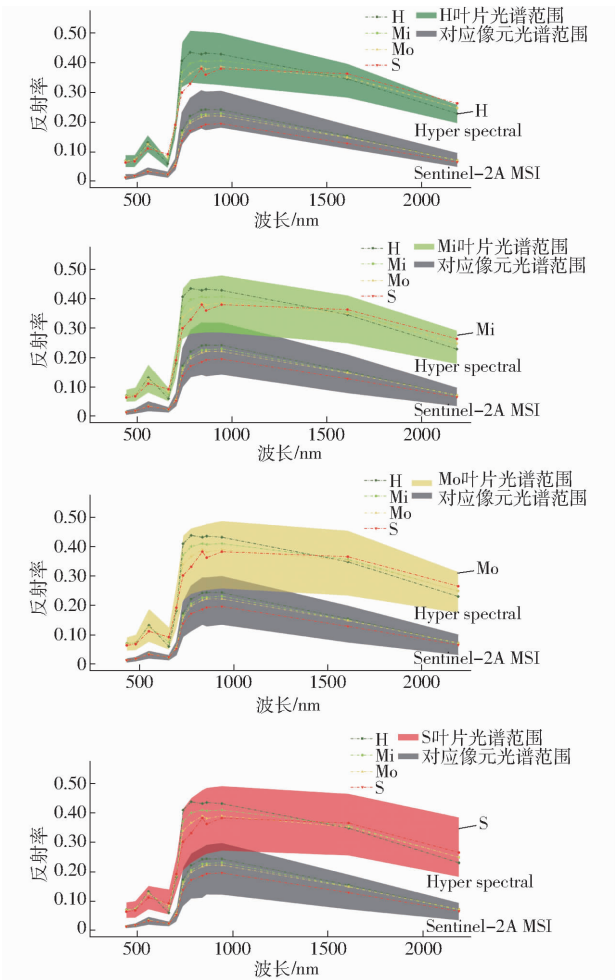


图 10 不同虫害等级的叶片反射率与像元反射率对比

Fig. 10 Comparison of leaf reflectance with image-scale reflectance for different infestation categories

蛾危害下像元尺度的敏感波段,也位于 703 ~ 898 nm;同时,研究发现部分基于 Sentinel-2A MSI 构建的植被指数在刚竹毒蛾危害的检测中会优于 GS 数据中对应指数,如 NDVI,基于 GS 构建的 NDVI 表现出的作用会劣于 Sentinel-2A MSI,这一结果与前人的研究结果^[37]相符。此外,在对比 GS 与 Sentinel-2A MSI 相同虫害等级的光谱反射率范围(图 10)时,发现在无危害等级下二者没有交集;但在受害情况下,二者出现明显的相交区域,轻度危害二者有 4 个波段相交,分别为波段 7、8、8a、9;中度危害和重度危害有 5 个波段相交,相较于轻度危害增加了波段 6;这 5 个波段均位于红边以及近红外波段范围内,从另一个角度也证明了红边以及近红外波段是刚竹毒蛾危害监测的重要波段。刚竹毒蛾是试验区毛竹林最主要的一种食叶性害虫,当然,竹林还可能受到竹蝗、竹蛭、枯梢病等病虫害的侵染,这些不同病虫害胁迫下的寄主外在表征可能具有一定的相似性,因此,在病虫害识别过程中,这种混淆情况的甄别将是未来的重要工作方向。

3 结论

(1)对于受刚竹毒蛾危害的样本而言,随着虫害等级的上升,其可见光波段呈现出“绿降红升”的趋势,红边波段反射率以及斜率不断降低。

(2)Sentinel-2A MSI 的原始波段具有一定的虫害响应能力,各虫害等级寄主竹林反射率在波段 6、7、8、8a 上表现的差异最为明显,对无危害-中度危害、无危害-重度危害、轻度危害-重度危害检测能力都达到了显著性水平。红边及近红外波段是虫害

检测的重要窗口,其原始光谱信息及相关指数对寄主竹林的健康状态极其敏感,在虫害检测中发挥极其重要的作用。

(3)根据模型给出的重要性排序,10 个特征变量的贡献度从高到低依次为 $RegVI_1$ 、 $NDVI_{705}$ 、 $RegVI_2$ 、 $NDVI_{783}$ 、 EVI 、 $GVMI$ 、 LAI 、 RVI 、 $NDWI$ 以及 $NDMVI$ 。从模型的检测效果来看,其对无危害、中度危害、重度危害的检测精度都达到了 75% 以上,甚至无危害、重度危害的检测精度达到了 85% 以上,但对轻度危害叶片的识别能力有所欠缺。

参 考 文 献

[1] 黄旭影,许章华,林璐,等. 刚竹毒蛾危害下的毛竹叶片光谱特征波长研究[J]. 光谱学与光谱分析,2018,38(12):3829-3838.
HUANG Xuying, XU Zhanghua, LIN Lu, et al. Spectral characteristic wavelengths of Moso bamboo leaves damaged by *Pantana phyllostachysae* Chao[J]. Spectroscopy and Spectral Analysis, 2018, 38(12):3829-3838. (in Chinese)

[2] 朱程浩,瞿帅,张晓丽. 油松毛虫灾害遥感监测及其影响因子分析[J]. 遥感学报,2016,20(4):653-664.
ZHU Chenghao, QU Shuai, ZHANG Xiaoli. *Dendrolimus tabulaeformis* disaster monitoring and analysis of its influencing factors through remote sensing technology[J]. Journal of Remote Sensing, 2016, 20(4): 653-664. (in Chinese)

[3] 马望,房磊,方国飞,等. 基于最大熵模型的神农架林区华山松大小蠹灾害遥感监测[J]. 生态学杂志,2016,35(8):2122-2131.
MA Wang, FANG Lei, FANG Guofei, et al. Mapping the infestation of *Dendroctonus armandi* in Shennongjia forested region using Landsat and MaxEnt model[J]. Chinese Journal of Ecology, 2016, 35(8): 2122-2131. (in Chinese)

[4] YE S, ROGAN J, ZHU Z, et al. Detecting subtle change from dense Landsat time series: case studies of mountain pine beetle and spruce beetle disturbance[J]. Remote Sensing of Environment, 2021, 263: 112560.

[5] 汪航,师苗,王岩,等. 基于 MODIS 时间序列数据的春尺蠖虫害遥感监测方法研究——以新疆巴楚胡杨为例[J]. 遥感技术与应用,2018,33(4):686-695.
WANG Hang, SHI Zhuo, WANG Yan, et al. A method for detecting the damage of *Apocheima cinerarius* Erschoff based on MODIS time series: case studies in *Bachu populus euphratica* forest of Xinjiang Province[J]. Remote Sensing Technology and Application, 2018, 33(4): 686-695. (in Chinese)

[6] 邹红玉,郑红平. 浅述植被“红边”效应及其定量分析方法[J]. 遥感信息,2010(4):112-116.
ZOU Hongyu, ZHENG Hongping. The effect and method of quantitative analysis of “red edge” of vegetation[J]. Remote Sensing Information, 2010(4): 112-116. (in Chinese)

[7] YU R, REN L, LUO Y. Early detection of pine wilt disease in *Pinus tabuliformis* in north China using a field portable spectrometer and UAV-based hyperspectral imagery[J]. Forest Ecosystems, 2021, 8: 44.

[8] FERNANDEZ-MANSO A, FERNANDEZ-MANSO O, QUINTANO C. Sentinel-2A red-edge spectral indices suitability for discriminating burn severity[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 50: 170-175.

[9] 李长春,陈伟男,王宇,等. 基于多源 Sentinel 数据的县域冬小麦种植面积提取[J]. 农业机械学报,2021,52(12):207-215.
LI Changchun, CHEN Weinan, WANG Yu, et al. Extraction of winter wheat planting area in county based on multi-sensor Sentinel data[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(12): 207-215. (in Chinese)

[10] BHATTARAI R, RAHIMZADEH B P, WEISKITTEL A, et al. Sentinel-2 based prediction of spruce budworm defoliation using red-edge spectral vegetation indices[J]. Remote Sensing Letters, 2020, 11: 777-786.

[11] 李凯,陈芸芝,许章华,等. 虫害胁迫下毛竹叶绿素含量高光谱估算方法[J]. 光谱学与光谱分析,2020,40(8):2578-2583.
LI Kai, CHEN Yunzhi, XU Zhanghua, et al. Hyperspectral estimation method of chlorophyll content in *Moso bamboo* under pests stress[J]. Spectroscopy and Spectral Analysis, 2020, 40(8): 2578-2583. (in Chinese)

[12] HUANG X Y, XU Z H, YANG X, et al. Monitoring the severity of *Pantana phyllostachysae* Chao on bamboo using leaf hyperspectral data[J]. Remote Sensing, 2021, 13: 4146.

[13] HAN N, DU H, ZHOU G, et al. Object-based classification using SPOT-5 imagery for Moso bamboo forest mapping[J]. International Journal of Remote Sensing, 2014, 35: 1126-1142.

[14] 向昌盛,周子英. ARIMA 与 SVM 组合模型在害虫预测中的应用[J]. 昆虫学报,2010,53(9):1055-1060.
XIANG Changsheng, ZHOU Ziyang. Application of ARIMA and SVM hybrid model in pest forecast[J]. Acta Entomologica Sinica, 2010, 53(9): 1055-1060. (in Chinese)

[15] YUAN Y, HU X. Random forest and objected-based classification for forest pest extraction from UAV aerial imagery[J/OL]. XXIII Isprs Congress, Commission I, 2016(XLI-BI):1093-1098. DOI:10.5194/isprarchives-XLI-BI-1093-2016.

[16] 许章华,刘健,陈崇成,等. 松毛虫危害下的马尾松林冠层光谱特征可辨性分析[J]. 国土资源遥感,2016,28(2):41-47.
XU Zhanghua, LIU Jian, CHEN Chongcheng, et al. Canopy spectral characteristics distinguish ability analysis of *Pinus*

- massoniana forests with *Pinus tabuliformis* Walker damage[J]. Remote Sensing for Land & Resources, 2016, 28(2): 41 – 47. (in Chinese)
- [17] 黄旭影, 许章华, 王小平, 等. 刚竹毒蛾危害下的毛竹叶片光谱特征及虫害等级检测研究[J]. 光谱学与光谱分析, 2021, 41(4): 1253 – 1259.
HUANG Xuying, XU Zhanghua, WANG Xiaoping, et al. Spectral characteristics of Moso bamboo leaves damaged by *Pantana phyllostachysae* Chao and monitoring of pest rating[J]. Spectroscopy and Spectral Analysis, 2021, 41(4): 1253 – 1259. (in Chinese)
- [18] JIA Y, JIN S, SAVI P, et al. GNSS – R soil moisture retrieval based on a XGBoost machine learning aided method: performance and validation[J]. Remote Sensing, 2019, 11: 1655.
- [19] JIN Q, FAN X, LIU J, et al. Estimating tropical cyclone intensity in the South China Sea using the XGBoost model and FengYun Satellite images[J]. Atmosphere, 2020, 11: 423.
- [20] LI G, LI W, TIAN X, et al. Short-term electricity load forecasting based on the XGBoost algorithm[J]. Smart Grid, 2017, 7: 274 – 285.
- [21] MU B Y, ZHANG S J, LI Z Z, et al. Early detection of cauliflower gray mold based on near-infrared spectrum feature extraction[J]. Spectroscopy and Spectral Analysis, 2021, 41: 2543 – 2548.
- [22] 周小成, 郑磊, 黄洪宇. 基于多特征优选的无人机可见光遥感林分类型分类[J]. 林业科学, 2021, 57(6): 24 – 36.
ZHOU Xiaocheng, ZHENG Lei, HUANG Hongyu. Classification of forest stand based on multi-feature optimization of UAV visible light remote sensing[J]. Scientia Silvae Sinicae, 2021, 57(6): 24 – 36. (in Chinese)
- [23] 余坤勇, 许章华, 刘健, 等. “基于片层-面向类”的竹林信息提取算法与应用分析[J]. 中山大学学报(自然科学版), 2012, 51(1): 89 – 95.
YU Kunyong, XU Zhanghua, LIU Jian, et al. Algorithm and application analysis of film-based & class-oriented for bamboo forest information extraction[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2012, 51(1): 89 – 95. (in Chinese)
- [24] 郑蓓君, 陈芸芝, 李凯, 等. 高光谱数据的刚竹毒蛾虫害程度检测[J]. 光谱学与光谱分析, 2021, 41(10): 3200 – 3207.
ZHENG Beijun, CHEN Yunzhi, LI Kai, et al. Detection of pest degree of *Phyllostachys* Chinese with hyperspectral data[J]. Spectroscopy and Spectral Analysis, 2021, 41(10): 3200 – 3207. (in Chinese)
- [25] 许章华, 刘健, 龚从宏, 等. 马尾松毛虫寄主有效叶面积指数遥感反演模型研究[J]. 中南林业科技大学学报, 2012, 32(10): 72 – 78.
XU Zhanghua, LIU Jian, GONG Conghong, et al. Effective leaf area index retrieving models for host of *Pinus tabuliformis* walker[J]. Journal of Central South University of Forestry & Technology, 2012, 32(10): 72 – 78. (in Chinese)
- [26] NESTOLA E, SCARTAZZA A, DI B D, et al. Are optical indices good proxies of seasonal changes in carbon fluxes and stress-related physiological status in a beech forest? [J]. Science of the Total Environment, 2018, 612: 1030 – 1041.
- [27] 吴志杰, 徐涵秋. 卫星影像数据构建山地植被指数与应用分析[J]. 地球信息科学学报, 2011, 13(5): 656 – 664.
WU Zhijie, XU Hanqiu. A new index for vegetation enhancements of mountainous regions based on satellite image data[J]. Journal of Geo-Information Science, 2011, 13(5): 656 – 664. (in Chinese)
- [28] 田庆久, 闵祥军. 植被指数研究进展[J]. 地球科学进展, 1998(4): 10 – 16.
TIAN Qingjiu, MIN Xiangjun. Advances in study on vegetation indices[J]. Advance in Earth Science, 1998(4): 10 – 16. (in Chinese)
- [29] SHI T T, XU H Q. Derivation of tasseled cap transformation coefficients for Sentinel – 2 MSI at-sensor reflectance data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12: 4038 – 4048.
- [30] 张海威, 张飞, 张贤龙, 等. 光谱指数的植被叶片含水量反演[J]. 光谱学与光谱分析, 2018, 38(5): 1540 – 1546.
ZHANG Haiwei, ZHANG Fei, ZHANG Xianlong, et al. Inversion of vegetation leaf water content based on spectral index[J]. Spectroscopy and Spectral Analysis, 2018, 38(5): 1540 – 1546. (in Chinese)
- [31] ANDERSON M J. A new method for non-parametric multivariate analysis of variance[J]. Austral Ecology, 2001, 26: 32 – 46.
- [32] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002, 46: 389 – 422.
- [33] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//KDD16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785 – 794.
- [34] CARTER G A, MILLER R L. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands[J]. Remote Sensing of Environment, 1994, 50: 295 – 302.
- [35] CURRAN P J, DUNGAN J L, GHOLZ H L. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine[J]. Tree Physiol, 1990, 7: 33 – 48.
- [36] 李少平, 吴正方, 赵云升. 长白山牡丹岭典型阔叶树叶变色期高光谱及红边特性研究[J]. 红外与毫米波学报, 2016, 35(5): 584 – 591.
LI Shaoping, WU Zhengfang, ZHAO Yunsheng. Hyperspectral and red-edge characteristics of typical hardwoods leaf coloring date in Mudan Valley, Changbai Mountain[J]. Journal of Infrared and Millimeter Waves, 2016, 35(5): 584 – 591. (in Chinese)
- [37] LANGE M, DECHANT B, REBMANN C, et al. Validating MODIS and Sentinel – 2 NDVI products at a temperate deciduous forest site using two independent ground-based sensors[J]. Sensors, 2017, 17: 1855.