

基于高光谱数据季相特征的山地草甸植被分类识别

郑奕^{1,2,3}, 王瑶^{1,2}, 刘艳^{1,2*}

1. 中国气象局乌鲁木齐沙漠气象研究所, 新疆 乌鲁木齐 830002
2. 中亚大气科学研究中心, 新疆 乌鲁木齐 830002
3. 新疆树木年轮生态实验室, 新疆 乌鲁木齐 830002

摘要 新疆天山北坡山地草甸是天山山区草地生产力最高的草地类型, 草地退化情况较为严重。对草地植被进行分类与识别, 监测草地生态系统本底状况, 可以快速、准确、有效的评价草地退化动态与程度, 是进行生态重建的关键。为了探索适合草地植被的分类方法, 选择天山北坡中段山地草甸植被作为研究对象, 利用高光谱成像光谱仪(SOC710VP)获取了典型植被多季相(4个关键生育期)的原始反射光谱数据, 通过多项式卷积平滑(S-G)及最小噪声分离(MNF)变换对光谱数据进行平滑去噪及降维处理, 分别采用支持向量机(SVM)、BP神经网络(BP-ANN)及波谱角填图(SAM)三种方法建立分类模型, 并对分类结果进行了对比分析。结果表明: 使用S-G滤波及MNF变换预处理方法可以有效的对草地植被高光谱数据进行降维除噪, 获得较平滑的光谱曲线, 减少了数据的冗余程度并缩短了分类时间。不同季相山地草甸植被的“绿峰”、“红谷”及“红边”等参数差异较大, 在植被生长旺盛期(4月—5月)的光谱曲线特征比黄枯期的光谱曲线特征更容易区分, 这个时期分类精度较高。SVM分类模型在返青期(4月)和分蘖(枝)期(5月)总体分类精度均超过了90%, Kappa系数也超过了0.9; 利用SVM方法进行分类时, 在植物生长旺盛期(4月—5月)Polynomial核函数分类精度较高, 植物成熟期(6月—9月)径向基核(RBF)函数分类精度较高。BP-ANN在分蘖(枝)期分类精度较高, 总体分类精度为91.07%, Kappa系数为0.89, 其他时期分类效果一般, 虽然在MNF变换降维后能极大的缩短数据处理时间, 但分类时间还是较SVM时间要长。SAM分类速度最快, 但在各生育期的分类精度都较低, 最高值为分蘖(枝)期的总体分类精度77.80%, Kappa系数为0.73。因此, 利用Polynomial核函数的SVM分类模型适合对山地草甸植被进行分类识别, 分类结果类别完整, 准确度高, 误分、错分现象相对较少, 相比BP-ANN及SAM等高光谱数据分类方法具有较大的优势。

关键词 高光谱数据; 多季相; 山地草甸; 分类识别

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)06-1939-09

引言

近年来, 由于气候变化及人为干扰强度增加, 天山山区草地生态严重失衡, 天然草地面积逐渐减少, 部分草地类型发生退化。天山北坡山地草甸是全区草地生产力最高的草地类型, 也是退化较为严重^[1]、草畜矛盾突出的区域。草地退化是一个由量变到质变的发展过程, 表现为草地生态系统的物种个体, 种群特征、群落组成等发生着不同程度的演变。因此, 对草地植被进行分类识别, 监测草地生态系统本底状

况, 可以快速、准确、有效的评价草地退化动态与程度, 是进行生态重建的关键。传统的草地资源实地考察的监测方法难以实现对偏远地区监测及快速资源调查评估的需要; 多光谱遥感监测精度较低, 无法对植被种类进行识别, 难以获取草地持续退化过程中优势种和群落组成变化的精细信息, 对于部分毒杂草丛生或“逆向”演替的退化草地类型具有明显的局限性。因此, 寻求一种适用的草地植被分类识别方法具有十分重要的意义。

高光谱遥感技术的发展, 为草地监测提供了前所未有的数据源。高光谱遥感数据具有光谱分辨率高、信息量大的特

收稿日期: 2021-05-12, 修订日期: 2021-08-16

基金项目: 中亚大气科学研究基金项目(CAAS202011), 中央级公益性科研院所基本科研业务费项目(IDM2021004)和国家自然科学基金项目(42071075)资助

作者简介: 郑奕, 1975年生, 中国气象局乌鲁木齐沙漠气象研究所副研究员 e-mail: zhengyi@idm.cn

* 通讯作者 e-mail: liuyan@idm.cn

点,能够更为全面、细致的获取地物光谱特征及其差异性,从而大幅度提高地物分类的类别精细度和准确度^[2]。目前国内的相关分类识别研究主要集中在利用地面实测高光谱数据,通过波段选择和特征提取,结合支持向量机(support vector machine, SVM)^[3]、决策树^[4]、逻辑回归^[5]及极限学习机(extreme learning machine, ELM)^[6]等分类器对植被进行分类识别,取得了较好的分类效果。如林川等分析和对比了不同植物生态类型的光谱特征,选定用于识别植物生态类型的光谱特征变量,最后应用非线性的反向传播人工神经网络(BP-ANN)与线性判别分析(FLDA)的类型识别方法,对湿地植物生态类型进行了识别^[7]。肖波等采集了6个草地早熟禾品种新鲜叶片的高光谱图像,提取了叶片的光谱信息,运用逐步判别分析法选择特征波段,采用 Fisher 线性判别法,构建草地早熟禾品种的判别分析模型,为快速识别草地早熟禾品种提供了一种新的方法^[8]。但是,高光谱遥感精细分类中仍存在“维数灾难”(Hughes 现象)及椒盐噪声严重影响分类结果的应用^[2]。有效降低高光谱数据的维数、选择特征波段以及提高光谱图像的信噪比和质量仍是高光谱数据处理和应用中亟待解决的关键问题^[9]。另外,由于草地植被的光谱曲线特征差异性不大,存在“同物异谱”和“异物同谱”现象,加上植被在生长发育的不同阶段,植被的外部形态特征到内部的叶绿素含量、含水量都会发生变化,其光谱特征也会随之变化,都会影响分类识别的精度。现阶段的研究大多采取植被成熟时期的光谱数据进行植被识别,而缺少时间过程的多时相地物光谱数据综合分析、目标识别研究。如何使植被类型间的差异加大,增强植被的可分性,提高分类精度

还有待更多的研究实例进行验证。

基于此,选择新疆天山北坡中段山地草甸典型植被作为研究对象,采集了多季相(4个关键生育期)植被原始反射光谱数据,然后通过多项式卷积平滑(Savitzky-Golay, S-G)对实测高光谱数据进行平滑除噪,再利用最小噪声分离(minimum noise fraction, MNF)变换进行降维处理,构建多种分类模型进行植物识别;一方面研究了不同季相草地植被的辨识精度,另一方面通过选用不同的分类识别算法,探索适用于新疆山地草甸植被的高光谱分类识别方法。

1 研究方法

1.1 研究区概况

研究区位于天山北麓中段白杨沟的中天山草地生态气象监测野外基地,海拔 2 050.0 m,年平均气温 2.4 °C,年平均总日照时数 2 429 h,年降水量 468 mm,年平均蒸发量 1 194 mm,土壤为栗钙土。该基地草场类型属山地草甸,建群种主要有禾本科牧草(老芒麦 *Elymus sibiricus* L.、草地早熟禾 *Poa pratensis* L. 等)、豆科牧草(黄芪 *Astragalus penduliflorus* Lam. 等)及典型的退化指示种(草原糙苏 *Phlomis pratensis* kar. et Kir.、草原老鹳草 *Geranium pratense* L.、绿草莓 *Fragaria viridis* Duch.、委陵菜 *Potentilla chinensis* Ser.、火绒草 *Leontopodium leontopodioides* (Willd.) Beauv.)等,局部地段生长有中生灌木,其气候特点和自然植被类型具有较好的代表性。

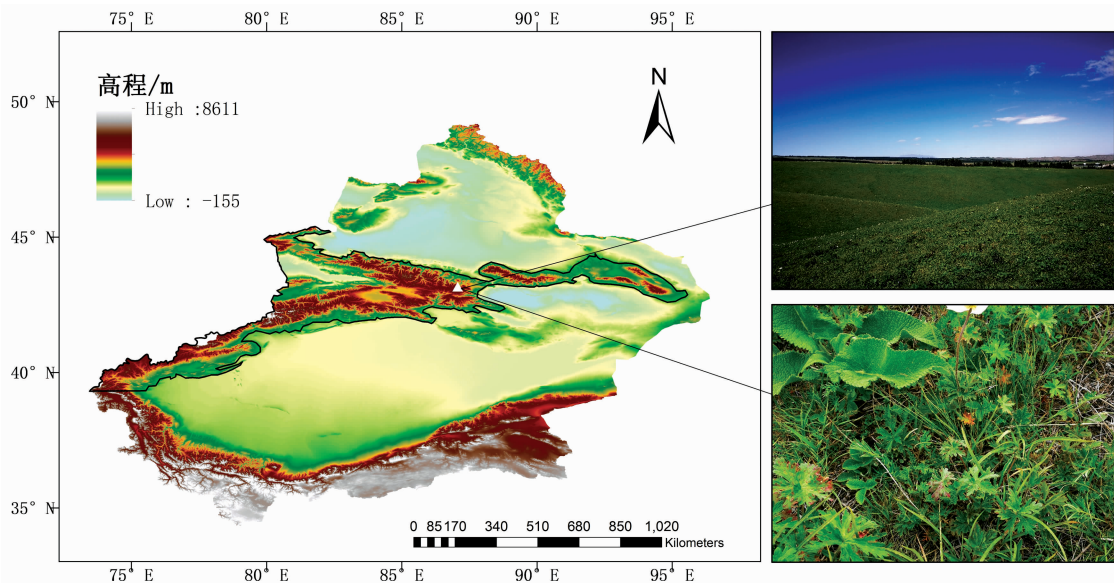


图 1 研究区地理位置及典型植被组成图(图中黑框区域为天山山区)

Fig. 1 Location and typical vegetation composition in study area (Tianshan Mountains marked by black border in the map)

1.2 高光谱数据的采集

植物光谱观测采用美国的高光谱成像光谱仪(SOC710VP),波长范围是 400~1 000 nm,光谱分辨率为 4.7 nm,波段数为 128 个,所获取的高光谱图像具有图谱合

一的特点,图像上每一个像元点都包含着丰富的光谱信息,构成光谱数据立方体。选择晴朗、无云、光照条件较好的天气进行测量,时间为 12:00—14:00。测定植被冠层光谱时,应保证测定区域内植被类型具有代表性。观测时传感器方向

垂直向下，高度 0.6 m，同时记录观测时间、天气状况、样地经纬度、植被类型情况。测量的时点包含山地草甸典型植被的返青期(4 月)、分蘖(枝)期(5 月)、花期(6 月—7 月)及黄枯期(9 月)。采用定点观测，每个测点采集 5 景高光谱图像，

一个测点整个生长季共采集了 20 景高光谱图像，选取其中质量最好的 4 景作为实验样本，优势植物主要为老芒麦、草原糙苏、草原老鹳草和绿草莓。

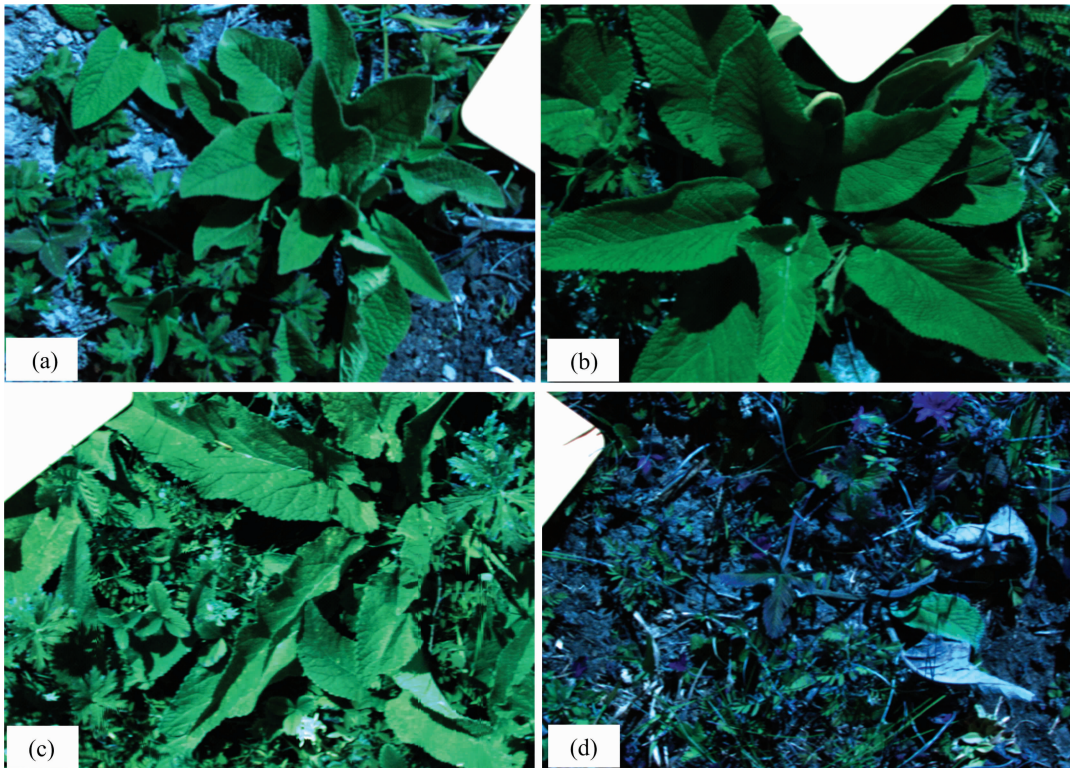


图 2 植被不同时期高光谱图像

(a): 返青期; (b): 分蘖(枝)期; (c): 花期; (d): 黄枯期

Fig. 2 Hyperspectral images of vegetation in different periods

(a): Green-up date; (b): Tillering stage; (c): Flowering phase; (d): Withering date

1.3 高光谱数据处理方法

高光谱数据收集后，剔除数据有明显异常的波段，截取的波段范围为 406~997 nm，共计 115 个波段，先利用 SRA-nal710 软件进行高光谱数据标定及反射率转换处理，再用 ENVI5.3 软件进行数据分析。

1.3.1 平滑去噪处理

采用多项式卷积平滑(S-G)对高光谱数据进行平滑去噪，它是一种在时域内基于局部多项式最小二乘拟合的滤波方法，该方法的最大优点在于滤波噪声的同时可以确保信号的形状和宽度不变，较好地保留了图像细节。由于该算法不受图像数据本身限制，因而提高了平滑去噪的适用性。其计算公式如式(1)

$$Y_j^* = \frac{\sum_{i=-m}^m C_i Y_{j+i}}{N} \tag{1}$$

式(1)中， Y_j^* 是函数拟合值， Y_{j+i} 为原始数据， N 为卷积数，数值上等于平滑窗口的大小($2m+1$)， C_i 表示第 i 个点的权重。

1.3.2 降维处理

由于高光谱数据波段数较多，有些波段质量不佳，将其

直接作为模型输入将会导致数据过大，影响速度和精度。同时为了克服 Hughes 现象，选用最小噪声分离(MNF)变换对高光谱数据进行降维和波段优选以提高建模速度、减少计算量。MNF 是针对一组多元随机变量构造线性变换，得到一组相互正交的结果变量，变换的目标是使结果变量的信噪比最大化^[9]。其本质上是两次叠置的主成分变换，第一次变换(基于估计的噪声协方差矩阵)用于分离和重新调节数据中的噪声，这步操作使变换后的噪声数据只有最小的方差且没有波段间的相关；第二步是对噪声白化数据(noise-whitened)进行主成分变换，具体相关推导公式见文献^[9]。

1.4 分类识别方法

分别利用支持向量机分类(support vector machine, SVM)、人工神经网络分类(artificial neural network, ANN)及波谱角填图(spectral angle mapper, SAM)这三种常用的高光谱数据分类方法对研究区植被进行分类识别，并对分类精度进行综合评价与比较分析。

1.4.1 支持向量机

近期的研究表明 SVM 的方法在高光谱遥感数据监督分类方面具有相当大的潜力，能获得比神经网络、决策树和最大似然等概率分类方法更高的精度，其分类原理及具体公式

见文献。在 SVM 中需要指定一个重要参数为惩罚系数 C (表示对误差的宽容度), C 越大, 说明越不能容忍出现误差, 容易过拟合, 即训练集准确率可能很高而测试集准确率不高; C 越小, 容易欠拟合^[10]。另外, SVM 中的核函数及其参数的选择也很重要, 直接影响到 SVM 的分类效果, 常用的核函数有线性核(linear)、多项式核(polynomial)、Sigmoid 核与径向基核(radial basis function)4 种。其公式如下:

(1) 线性核(linear)

$$k(x_i, x_j) = x_i x_j \quad (2)$$

(2) 多项式核(polynomial)

$$k(x_i, x_j) = (x_i x_j + c)^d \quad (3)$$

(3) Sigmoid 核

$$k(x_i, x_j) = \tanh[\mu(x_i x_j) + \nu] \quad (4)$$

(4) 径向基核(radial basis function)

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (5)$$

其中, c, d, μ, ν 和 σ^2 均为参数。

1.4.2 人工神经网络

ANN 指用计算机模拟人脑的结构, 用许多小的处理单元模拟生物的神经元, 用算法实现人脑的识别、记忆、思考过程。其中误差反向传播算法(back propagation, BP)应用最为广泛, 它通常指多层前向神经网络, 含有输入层、输出层以及处于输入输出层之间的隐含层, 如果进行非线性分类, 输入的区域并非线性分类或需要两个超平面才能区分类别的时候, 隐含层数设置为大于或等于一。隐含层的状态影响输入与输出之间的关系, 改变隐含层的权系数, 可以改变整个多层神经网络的性能, 本文使用的是算法较成熟的前馈三层 BP 网络模型(BP-ANN), 有关模型的算法详见文献[7, 10]。

1.4.3 波谱角填图

SAM 是高光谱图像分类中较常用的方法, 该方法从考虑光谱维的信息出发, 强调了光谱的形状特征, 通过计算待测光谱与参考光谱之间的“角度”来确定两者之间的相似性。每个光谱都看成 n 维空间的一个向量, 第 i 个向量和第 j 个向量之间存在一个夹角 θ_{ij} , 夹角的余弦值称为相似系数, 如式(6)

$$\cos(\theta_{ij}) = \frac{\sum_{k=1}^n X_{ik} X_{jk}}{\sqrt{\sum_{k=1}^n X_{ik}^2} \sqrt{\sum_{k=1}^n X_{jk}^2}} \quad (6)$$

式(6)中, i 和 j 均为 n 维空间向量, X 为向量对应的值。光谱角值越小, 其余弦值越接近 1, 待测光谱与参考光谱就越相似, 地物特征信息也就越相似, 归为一类的可能性就越高^[11]。

1.5 精度评价

衡量分类算法性能的 2 个指标分别为总体分类精度、Kappa 系数, 值域范围均为 $[0, 1]$, 2 个指标的值越接近 1, 说明算法的性能越好。总体分类精度为所有正确识别的样本数与总验证样本数的比值, 其公式如式(7)^[12]

$$\text{总体分类精度} = \frac{1}{N} \sum_{i=1}^m P_{ii} \quad (7)$$

式(7)中, N 为样本总数, m 为地物类型数, P_{ii} 为混淆矩阵第 i 行第 i 列的像元数(即地物 i 被正确识别的样本量)。

Kappa 系数是另一种衡量图像整体分类精度的指标, 既考虑了对角线上被正确分类的像元, 又考虑了不在对角线上的各种漏分和错分误差, 其值越大, 分类结果与真实地物类型的一致性越好, 其公式如式(8)

$$\text{Kappa 系数} = \frac{N \sum_{i=1}^m P_{ii} - \sum_{i=1}^m (P_{i+} P_{+i})}{N^2 - \sum_{i=1}^m (P_{i+} P_{+i})} \quad (8)$$

式(8)中, P_{i+} 和 P_{+i} 分别为混淆矩阵第 i 行和第 i 列的像元总数。

2 结果与讨论

2.1 原始光谱曲线经 Savitzky-Golay 滤波后效果

图 3 为几种优势植物不同季相光谱曲线平滑去噪处理前后对比, 可以看出, 草地植被未经平滑去噪处理的原始光谱, 其锯齿状的波动不利于光谱特征的判断和提取, 经过 Savitzky-Golay 滤波算法处理后, 对比原始曲线, 大部分噪音得以去除, 且滤波后的曲线平滑度较高。平滑处理在剔除干扰值的同时, 保持了原始曲线的基本特征, 并突出刻画了波峰与波谷相交替的状态, 较好的体现了植物生长的变化特征。

2.2 不同季相优势植物光谱特征变化

图 4 反映了草原糙苏、绿草莓、草原老鹳草、老芒麦等 4 种山地草甸优势植物在不同季相的光谱曲线响应变化, 在可见光波段 400~420 nm 处的吸收谷, 称为“蓝谷”, 反映了植物光合作用色素(叶绿素、类胡萝卜素)对蓝光的吸收特征。

在可见光波段 550 nm 附近是叶绿素的绿色强反射区, 形成一个反射峰, 即“绿峰”, 返青期的“绿峰”幅值最高, 说明这个时期植被叶绿素含量较高。在植物分蘖(枝)期及花期, “绿峰”均不同程度向蓝光方向偏移, 这是因为随着植被的生长, 叶绿素浓度增加, 植物光合作用增强。黄枯期这几种植物的叶片开始黄化, 生物量减少, 叶绿素缺乏使植物在叶绿素吸收带上的吸收减少, “绿峰”又向红光方向偏移。如绿草莓、草原老鹳草黄枯期的“绿峰”位置偏移到 640 nm 左右。表现最为明显的是草原糙苏, 黄枯期地上部分大多凋落或枯黄, 绿色叶片大量减少, 光谱曲线无明显“绿峰”和植物光谱特征, 已呈现出土壤的光谱反射特征, 只是反射率要高于土壤, 这是由于草原糙苏枯叶表面比较光滑, 在测量光谱过程中, 入射光照射到枯叶表面会有一部分产生反射, 使其反射率比土壤的反射率要大, 这和已有的一些研究结论相同^[13]。

在可见光波段 670~700 nm 处的吸收谷, 称为“红谷”, 是植物光合作用吸收最强的波段范围。随着植被的生长, 有“红谷”蓝移现象。由图 4 还可以看出, 分蘖(枝)期(5 月)为植被生长旺期, 叶绿素含量较高, 几种植物对红光的吸收较强, “红谷”幅值基本都是全生长季最小。

在近红外波段 700~750nm 区域, 植被叶面反射率急剧

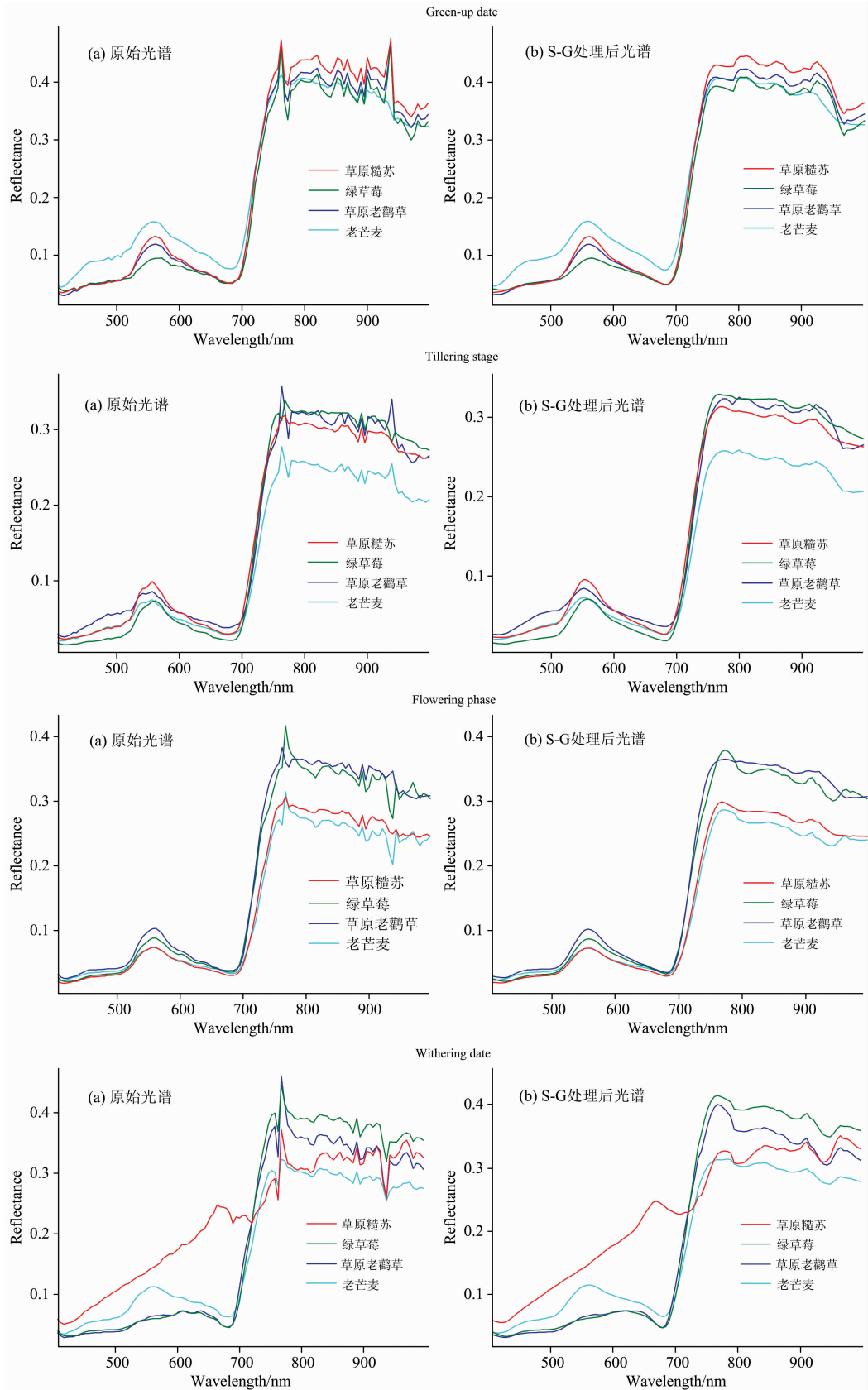


图 3 不同季相优势植物光谱曲线平滑去噪处理前后对比

Fig. 3 Comparison of spectral curves of dominant vegetation in different seasons before and after smooth denoising

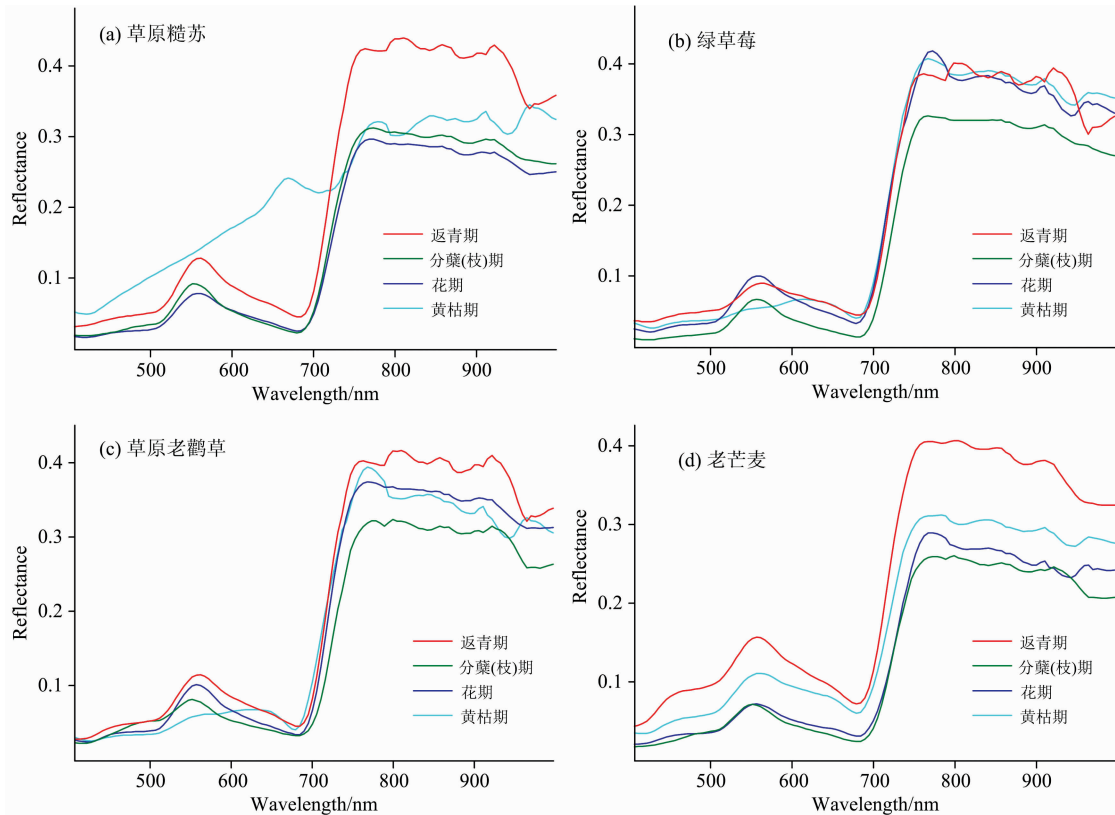


图 4 优势植物在 4 个关键生育期的光谱曲线对比图

Fig. 4 Comparison of spectral curves of dominant vegetation in four key growth periods

增大, 出现反射“陡坡”, 称为“红边”。“红边”斜率主要与植被覆盖度或叶面积指数有关, 覆盖度越高, 叶绿素含量越高, 红边的斜率越大。另外, 植被生长旺盛时期, 植被生物量、色素含量高, 红边位置会向红光方向偏移, 到黄枯期, 又向蓝光方向偏移。

近红外波段 750~950 nm 是一个相对平坦、反射率较高的区域, 它主要由植被叶片内部结构决定。在这个范围内, 不同植被在不同季节, 其光谱反射率是不同的。几种植物返青期草地光谱反射率值均不同程度高于其他时期的草地反射率值, 主要是因为返青期草地生长状况良好, 叶面积指数较大, 叶绿素含量较高。黄枯期则因植被地上部分大多凋落或枯黄, 绿色叶片大量减少, 反射率受到土壤光谱干扰而呈现数值偏高的现象。

通过以上分析, 可以看出几种植物虽然具有相同的反射率变化特征, 即波峰、波谷等特征出现的位置基本一致。但在生长的不同阶段, 由于叶绿素含量差异影响到植物光合作用效率, 叶绿素含量越低, 蓝、红波段吸收减弱, 可见光波段反射率升高, 近红外反射率减弱, 反之叶绿素含量越多, 蓝、红波段吸收增强, 可见光波段反射率降低, 近红外反射率增强, 导致光谱特征存在细微差别, 主要体现在“绿峰”、“红谷”及“红边”等参数差异较大, 其位置会根据植被健康状况、成熟状态等发生偏移, 因此可以通过不同季相光谱曲线的细小差别实现对植被的精细分类。

2.3 多种方法分类结果与精度评价

基于同一样地采样数据, 在对原始数据进行降维除噪后, 分别采用 SVM, BP-ANN 和 SAM 三种方法进行分类。采用 ENVI5.3 软件提取高光谱图像的感兴趣区域(region of interest, ROI), 为了保证精度验证的准确性, 建模样本与验证样本必须彼此独立, 重新选取 ROI 作为验证样本来计算混淆矩阵。不同季相不同植物建模样本和验证样本的比例大致为 2:1。

2.3.1 SVM 分类器的构建与评价

筛选出 Polynomial 核函数和 RBF 核函数构建 SVM 分类器, 径向基核函数除了受惩罚系数 C 影响, 还受核函数宽度参数 σ^2 影响, σ^2 间接地决定了数据映射到新特征空间后的分布状况, 值越大, 支持向量越少, 可能会造成过拟合; 值越小, 支持向量越多, 可能出现大的平滑效应, 无法在训练集上得到特别高的准确率, 影响最终测试集的分类准确率。因此, C 与 σ^2 的选择需要维持一种动态平衡, 不仅要有较高的测试集分类准确率, 而且还要保证分类器的通用性、泛化性等性能。SVM 模型的分类及检验结果如表 1 所示。

结果表明, 不管是采用多项式核函数还是采用高斯径向基核函数, 利用 SVM 对山地草甸植被高光谱数据进行分类均取得了较理想的效果。尤其在返青期和分蘖(枝)期分类精度较高, 总体分类精度均超过了 95%, Kappa 系数也超过了 0.94, 其中分蘖(枝)期 Polynomial-SVM 分类精度达到

97.86%，Kappa 系数为 0.97。利用 SVM 方法进行分类时，在植物生长旺盛期(4 月—5 月)Polynomial 核函数分类精度较高，植物成熟期(6 月—9 月)径向基核(RBF)函数分类精度较高。

表 1 高光谱数据支持向量机建模结果

Table 1 Modeling results of support vector machine by using hyperspectral data

生育期	核函数	总体分类精度/%	Kappa 系数
返青期	RBF	95.88	0.95
	Polynomial	96.91	0.96
分蘖(枝)期	RBF	95.19	0.94
	Polynomial	97.86	0.97
花期	RBF	91.01	0.89
	Polynomial	86.62	0.84
黄枯期	RBF	88.27	0.86
	Polynomial	87.80	0.86

2.3.2 其他分类方法的比较

为了验证 SVM 对高光谱数据分类的优势，又利用 BP-ANN 及 SAM 分类方法进行了比较分析，具体分类结果见表 2。

表 2 多种方法分类结果与评价

Table 2 Classification results and evaluations of multiple methods

生育期	分类方法	总体分类精度/%	Kappa 系数
返青期	BP-ANN	79.50	0.77
	SAM	65.42	0.61
分蘖(枝)期	BP-ANN	91.07	0.89
	SAM	77.80	0.73
花期	BP-ANN	88.87	0.87
	SAM	75.39	0.72
黄枯期	BP-ANN	89.90	0.88
	SAM	46.00	0.38

在常用的高光谱数据分类方法中，BP-ANN 为学习型机制，通过误差逆传播算法不断训练模型实现分类，但是针对高光谱数据的不确定性、高维数等分类效果不够理想。由表 2 可以看出，除了分蘖(枝)期分类精度较高，总体分类精度为 91.07%，Kappa 系数为 0.89，其他时期分类效果一般。另外，在样本数量较大时，BP-ANN 模型训练较为耗时，分类效率较低，虽然在 MNF 变换降维后能极大的缩短数据处理时间，但分类时间还是较其他两种分类方法都长。该方法在黄枯期的分类精度略高于 SVM 和 SAM 分类方法，总体分类精度为 89.90%，Kappa 系数为 0.88。

SAM 分类也是高光谱分类的常用方法，虽然有训练样本数量少，分类速度快的优点，但是在植被各季相的分类结

果精度均不高，最高值为分蘖(枝)期的总体分类精度 77.80%，Kappa 系数为 0.73。这是因为 SAM 算法把光谱角测度作为一种全局性的训练样本对整个光谱向量集进行相似性测量，当地物类别内部存在一定的差异时，这种方法对局部特征和辐射强度并不敏感，且易受噪声影响，导致分类精度下降^[14]。另外，由于所有高光谱的波段都参与影像分类，也容易产生 Hughes 现象。

比较以上三种分类方法，可以看出，在植被生长旺盛期(4 月—5 月)的光谱曲线特征比黄枯期的光谱曲线特征更容易区分，因此这个时期比较适合进行草地植被分类，精度较高。三种方法对花期的分类精度都有所降低，可能是因为这个时期部分植物的叶片受叶黄素、花青素等色素影响有变色现象；另外，花期草甸植被植株较其他时期明显增高，受采集时风速的影响，高光谱影像上部分区域出现些许条纹偏移，也是致使花期植被分类精度降低的原因。这几种常用的高光谱分类方法对于植被黄枯期的分类精度均不高，原因大致是因为黄枯期植被大多呈现枯黄状态，地上部分凋落或者萎蔫，叶片叶绿素含量减少，植被稀疏及土壤裸露，植被光谱反射率受到下垫面反射光谱影响造成分类精度降低，因此，植被黄枯期不适合对植被进行分类与识别。

综上所述，SVM 算法以数学统计方法和优化技术为基础，分类器结构简单、容易训练、收敛速度快、具有很高的分类精度，与神经网络方法和波谱角填图方法相比具有一定的优势，是一种有效的草地植被分类识别方法，但是在核函数和最优参数组合的选择方面仍然是一个难点。植物生长旺盛期基于 Polynomial 核函数的 SVM 分类器精度较高，植物成熟期基于 RBF 核函数的 SVM 分类器精度较高。

3 结 论

针对草地植被光谱相似度高、难以分类识别等特点，利用 S-G 滤波及 MNF 变换预处理结合不同核函数的 SVM 分类方法，对比 BP-ANN 及 SAM 等几种常用的高光谱数据分类方法，统计分析得出以下结论：

(1)使用 S-G 滤波及 MNF 变换预处理方法可以有效的对高光谱数据进行降维除噪，获得较平滑的光谱曲线，减少了数据的冗余程度并缩短了分类时间。

(2)不同季相的植被因结构特征、色素含量和含水量等不同造成光谱曲线差异，“绿峰”、“红谷”及“红边”参数等均存在差异。在植被生长旺盛期(4 月—5 月)的光谱曲线特征比黄枯期的光谱曲线特征更容易区分，因此这个时期分类精度较高，这也反映出植物的叶绿素含量是进行山地草甸植被分类的主要决定因素，植被黄枯期不适合对植被进行精细化分类与识别。

(3)SVM 算法可以实现对山地草甸植被的精细分类，分类结果类别完整，准确度高，误分、错分现象相对较少，相比 BP-ANN 及 SAM 等常用的高光谱数据分类方法具有较大的优势。植物生长旺盛期(4 月—5 月)Polynomial-SVM 分类精度较高，植物成熟期(6 月—9 月)RBF-SVM 分类精度较高。

References

- [1] LIU Yan, NIE Lei, YANG Yun(刘 艳, 聂 磊, 杨 耘). Ecology and Environment Science(生态环境学报), 2018, 27(5): 802.
- [2] TONG Qing-xi, ZHANG Bing, ZHANG Li-fu(童庆禧, 张 兵, 张立福). Journal of Remote Sensing(遥感学报), 2016, 20(5): 689.
- [3] YANG Kai-ge, FENG Xue-zhi, XIAO Peng-feng(杨凯歌, 冯学智, 肖鹏峰). National Remote Sensing Bulletin(遥感学报), 2016, 20(3): 409.
- [4] CHAI Ying, RUAN Ren-zong, CHAI Guo-wu, et al(柴 颖, 阮仁宗, 柴国武, 等). Remote Sensing for Land & Resource(国土资源遥感), 2016, 28(3): 86.
- [5] WU Fu-yu, WANG Xue, DING Jian-wei, et al(武复宇, 王 雪, 丁建伟, 等). National Remote Sensing Bulletin(遥感学报), 2020, 24(4): 439.
- [6] MOU Duo-duo, LIU Lei(牟多铎, 刘 磊). Remote Sensing Technology and Application(遥感技术与应用), 2019, 34(1): 115.
- [7] LIN Chuan, GONG Zhao-ning, ZHAO Wen-ji, et al(林 川, 宫兆宁, 赵文吉, 等). Acta Ecologica Sinica(生态学报), 2013, 33(4): 1172.
- [8] XIAO Bo, MAO Wen-hua, LIANG Xiao-hong, et al(肖 波, 毛文华, 梁小红, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2012, 32(6): 1620.
- [9] LIU Dai-zhi, HUANG Shi-qi, WANG Yi-ting, et al(刘代志, 黄世奇, 王艺婷, 等). Hyperspectral Remote Sensing Image Processing and Application(高光谱遥感图像处理与应用). Beijing: Science Press(北京: 科学出版社), 2016. 2.
- [10] YU Xu-chu, FENG Wu-fa, YANG Guo-peng, et al(余旭初, 冯伍法, 杨国鹏, 等). Analysis and Application of Hyperspectral Image(高光谱影像分析与应用). Beijing: Science Press(北京: 科学出版社), 2013. 171.
- [11] DAI Xiao-ai, JIA Hu-jun, ZHANG Xiao-xue, et al(戴晓爱, 贾虎军, 张晓雪, 等). Remote Sensing for Land & Resources(国土资源遥感), 2016, 28(3): 174.
- [12] XU Yong-ming(徐永明). ENVI Software Comprehensive Practice Course(ENVI 遥感软件综合实习教程). Beijing: Science Press(北京: 科学出版社), 2019. 130.
- [13] SHI Gui-hua, WANG Ying-shun, HOU Qiong, et al(师桂花, 王英舜, 侯 琼, 等). Chinese Journal of Agrometeorology(中国农业气象), 2013, 34(1): 114.
- [14] YANG Ke-ming, LIU Fei, SUN Yang-yang, et al(杨可明, 刘 飞, 孙阳阳, 等). Journal of Image and Graphics(中国图象图形学报), 2015, 20(6): 0836.

Study on Classification and Recognition of Mountain Meadow Vegetation Based on Seasonal Characteristics of Hyperspectral Data

ZHENG Yi^{1, 2, 3}, WANG Yao^{1, 2}, LIU Yan^{1, 2*}

1. Institute of Desert Meteorology, China Meteorological Administration, Urumqi 830002, China

2. Center for Central Asia Atmosphere Science Research, Urumqi 830002, China

3. Xinjiang Key Laboratory of Tree-Ring Ecology, Urumqi 830002, China

Abstract The mountain meadow on the north slope of the Tianshan Mountains has the highest grassland productivity, and the grassland degradation is serious. The classification and recognition of grassland vegetation play an important role in monitoring the background status of the grassland ecosystem. It is also the key to carrying out ecological reconstruction, which can quickly, accurately, and effectively evaluate grassland the dynamics and degree of grassland degradation. In this paper, we explored the classification method in grassland vegetation of the typical mountain meadow vegetation in the middle section of the north slope of Tianshan Mountain in Xinjiang. Firstly, a hyperspectral imaging spectrometer obtained original reflectance spectra of typical vegetation in four key growth periods (SOC710VP). Then, Savitzky-Golay filtering and the minimum noise fraction transformation (MNF) were used to smooth and reduce the dimensions of the spectrum data. Thirdly, classification models were established by the support vector machine (SVM), the backpropagation artificial neural network (BP-ANN) and the spectral angle mapper (SAM). Finally, a comparative analysis of the classification results from three models was made. The results showed that the dimension reduction and noise removal of grassland vegetation hyperspectral data could be effectively carried out by using the S-G filter and MNF transform preprocessing method. This processing reduced the redundancy of data and shortened the classification time while obtaining a smoother classification image. The parameters such as “green peak”, “red valley”, and

“red edge” of mountain meadow vegetation varied greatly in different seasons. The spectral curve characteristics in the vigorous vegetation growth period (from April to May) were easier to distinguish than those in the withering date. Thus, the classification accuracy was higher in this period. The overall classification accuracy of the SVM model exceeded 90%, and the Kappa coefficient exceeded 0.9 in the green-up date (April) and tillering stage (May). Based on the SVM model, the classification accuracy of the polynomial kernel function was higher in the vigorous growth period (from April to May), and the radial basis function (RBF) showed better performance in the mature period (from June to September). BP-ANN had higher classification accuracy in the tillering stage, the overall classification accuracy was 91.07%, and the kappa coefficient was 0.89. However, the classification effect was general in other periods. Moreover, the classification time was still longer than that of SVM although after the reduction of MNF transformation dimensionality. SAM had the fastest classification speed, but the classification accuracy was low in each growth stage. The highest value was 77.80% of the overall classification accuracy in tillering stage, and the kappa coefficient was 0.73. Therefore, the SVM classification model using the polynomial kernel function was suitable for classifying and recognising mountain meadow vegetation, which had complete classification category results, higher accuracy and relatively few misclassification. It was a better classification method than BP-ANN and SAM.

Keywords Hyperspectral data; Multi-seasonal; Mountain meadow; Classification and recognition

(Received May 12, 2021; accepted Aug. 16, 2021)

* Corresponding author