

冷杉叶片氮含量高光谱反演技术研究

宋雪莲¹, 王志伟¹, 张文^{1,2}, 张威³, 丁磊磊¹, 柳嘉佳¹, 阮玺睿¹, 王普昶⁴

(1. 贵州省农业科学院草业研究所, 贵州 贵阳 550006; 2. 贵州阳光草业科技有限责任公司, 贵州 贵阳 550006; 3. 贵州省水利水电勘测设计研究院有限公司, 贵州 贵阳 550002; 4. 贵州省农业科学院, 贵州 贵阳 550006)

摘要:氮是植物生长的基本营养元素,对氮含量的监测有助于及时了解植物的代谢过程和健康状况。高光谱技术能够对叶片氮含量进行无损快速的检测,已成为叶片氮含量检测的重要手段之一。采用高光谱技术对冷杉叶片氮含量进行估算,首先分别采用相关系数法、连续投影法、LASSO算法对4种形式的光谱变量 $R, R', \log(1/R), \log(1/R)'$ 进行敏感波段筛选,对筛选后的敏感波段分别采用偏最小二乘、随机森林、支持向量机进行建模反演。结果显示:两种导数形式变量的反演误差最小;且相关系数 R 与 $\log(1/R)'$, LASSO与 $\log(1/R)'$ 能够有效筛选出叶片氮含量的敏感波段组合,无论采取何种建模方法,其估算效果在几种筛选方法组合中最好, $R^2 > 0.84$, RMSE在0.19~0.24,估算效果明显优于全波段建模;随机森林算法对相关系数法和LASSO算法筛选出的 $\log(1/R)'$ 形式变量的反演误差最小,但对其他形式变量的反演误差变化范围较大,表现并不稳定。3种变量筛选方法筛选出的 R 形式的变量与前人研究相符,连续投影算法能筛选出更多与叶片其他化学含量相关的波段,证明了3种变量筛选方法的有效性。

关键词:高光谱;冷杉叶片;氮含量;筛选波段

中图分类号: S758 **文献标志码:** A **文章编号:** 1009-5500(2021)06-0139-09

DOI: 10.13817/j.cnki.cyycp.2021.06.020



收稿日期: 2020-12-11; **修回日期:** 2021-03-25

基金项目:贵州省农业科学院2017年度学术新苗培养及创新探索专项“基于人工智能的牧草营养高光谱反演技术研究”(黔农科院青年基金[2018]91号);贵州省科技计划项目“高光谱遥感技术评价牧草饲用价值的研究”(黔科合支撑[2018]2371);贵州省自然科学基金重点项目“喀斯特山区不同土地利用方式土壤养分转化的微生物驱动机制”(黔科合基础[2018]1419);贵州省农业科学院2021年院基本科研业务发展专项“贵州高原草地生态系统多功能性对不同管理措施的响应”(黔农科院青年科技基金[2021]31号)

作者简介:宋雪莲(1989-),女,湖北随州人,硕士,研究实习员,主要从事植被高光谱遥感方面研究。

E-mail: 1002848850@qq.com

王志伟为通信作者。

E-mail: wzw1206@163.com;

张文为通信作者。

E-mail: zhangwen0708@163.com

氮是植物生长的基本营养元素,是陆地生态系统植物生长的主要限制因子,与植物的光合作用和细胞生长分类等重要生理活动相关^[1]。缺氮会影响植物的光合作用能力,导致叶片生理及形态结构的相应变化,因而引起叶片光谱反射特性的变化^[2]。传统的化学方法能够精确地监测出植物氮含量,但具有高损耗、复杂、时滞性等缺点,不能进行全面、大范围、快速的营养诊断。高光谱光谱分辨率高,能够获得连续的光谱信息,已成为监测植物氮含量的有效手段。

目前有不少学者采用敏感波段筛选和植被指数进行高光谱植物氮含量检测的相关研究。朱艳等^[4]研究表明,单波段光谱在610 nm和680 nm处的水稻冠层反射率与叶片氮含量具有较好的相关性,提出采用回归系数来提高水稻叶片氮含量估测的准确性。刘冰峰等^[5]指出720 nm处的反射光谱以及550,720 nm处的一阶光谱导数是夏玉米氮含量的敏感参数。Tarp-ley等^[6]研究认为,利用红边位置和短波近红外波段光

谱反射率比值可预测棉花叶片氮浓度。Pacheco—Labrador 等^[7]采用归一化指数 NDIS 和三波段指数 TBIs 有效估算 Holm oak 叶片氮含量; Abdel—Rahman 等^[8]采用一阶导数的 SR 指数(波段 743, 1 316 nm)(波段 743, 1 317 nm)(波段 741, 1 323 nm)估算的叶片氮含量的决定系数在 0.75 左右; Yao 等^[9]采用归一化指数 NDSI 和比值指数 RSI 一阶导数对叶片氮的累积量进行了估算, 其决定系数高达 0.81; Ullah 等^[10]利用 MERIS 数据采用植被指数 (NDVI, NBDI, SAVI, TSAVI, REIP, MTCI) 以及波段深度参数对草地生物量及氮含量进行了估算和对比。Sanches 等^[11]采用反射率, 吸光度及其衍生形式和偏最小二乘方法对植物氮磷钾进行估算, 结果表明一阶导数形式的估算精度最高。另有学者着重从算法角度进行了植物氮含量高光谱反演技术研究, Zhang 等^[12]利用单变量线性回归、逐步多元线性回归、偏最小二乘对湿地植物芦苇氮含量进行估算, 并采用留一交叉验证比较 3 种模型的精度, 认为氮的敏感波段集中在红波段和绿波段, 单变量线性回归对 665 nm 和 680 nm 处的归一化植被指数效果最好, 3 种方法中逐步线性回归的精度最高。Yao 等^[13]采用原始反射率、一阶导数、植被指数等变量利用 SMLR、PLSR、ANNS、SVMs 等反演了冠层叶片氮含量, 结果表明一阶导数的支持向量机方法精度更高。Wang 等^[3]验证了多核支持向量机在估算小麦叶片氮含量的有效性, 并与多元线性回归, 偏最小二乘, 人工神经网络, 单核支持向量机进行了对比。总体来说, 现有的关于高光谱植物氮含量监测的大部分研究集中在植被指数反演氮含量以及敏感波段筛选方面, 关于建模算法的研究较少^[3]。且已有的研究或集中于建模之前敏感特征的筛选或集中于建模算法的比较, 本文尝试将不同特征提取方法与建模方法相结合, 通过不同方法间的组合, 筛选出一套完整的能有效估算叶片氮含量的流程。

1 材料和方法

1.1 光谱数据与氮含量数据获取

数据集采用 Accelerated canopy chemistry program (ACCP) (<https://doi.org/10.3334/ORN-LDAAC>)^[14], 该数据集旨在研究不同生态系统中植被冠层氮含量及木质素含量遥感反演的理论基础, 包含野外样品的实验室化学分析数据、光谱数据、小型冠层

试验的化学分析数据及冠层建模数据。本研究选择 ACCP 中 91 组道格拉斯冷杉幼苗期新鲜叶片光谱数据及相应的实验室化学分析数据, 采集于 1992 年 11 月。光谱数据为 400~2 498 nm 处的吸光度 [$\log(1/\text{反射率})$], 光谱间隔为 2 nm, 分辨率为 10 nm。在采集光谱数据后的 6 h, 采集叶片测量叶片氮及叶绿素含量。采用常规的实验室方法, 对每棵树幼苗期的叶片样本进行干燥、研磨和全氮、全叶绿素(a+b)分析。样品经硫酸—氧化汞催化剂(Perstorp 分析)在块状消化器中消化后, 用 Alpkem 连续流自动分析仪测定总氮, 氮含量的单位为单位干重百分比。

1.2 数据分析

利用 Matlab 对试验数据进行了异常数据剔除、计算了原始反射率、一阶导数 R' 、反射率倒数的对数 $\log(1/R)$ 及其一阶导数 $\log(1/R)'$, 以及这些参数与氮含量的相关系数, 采用连续投影算法、LASSO、相关系数筛选以上 4 种参数形式的特征波段, 并采用随机森林、偏最小二乘以及支持向量机建立氮含量反演模型。其中, 相关系数、连续投影算法筛选波段组合, 偏最小二乘建模在 Matlab 中完成, LASSO 算法筛选变量组合、随机森林和支持向量机建模在 R 中完成。

1.2.1 相关系数法 相关系数法是最简单直接能够帮助理解特征与相应变量之间关系的方法, 衡量的是变量之间的线性相关性, 其值绝对值越大表示相关性越强。很多研究通过选择最大相关性的变量作为特征变量进行建模研究。 $\lambda_1 - \lambda_2$, λ_1/λ_2 表示两变量间的差异性, 其与因变量间的相关系数表示两变量组合对因变量的解释量, 选取相关系数最大的两个波长即为特征波长。

1.2.2 连续投影算法 连续投影算法(SPA)是一种矢量空间共线性最小化的前向变量选择算法, 能够有效从全波段中提取特征波段, 消除原始光谱矩阵中的冗余信息, 降低模型的复杂度, 在波长选取中取得了较好的效果^[15]。连续投影算法是一种前向循环选择算法, 设样本集 M 和波段数 K 组成一个光谱矩阵 $X_{M \times K}$, 分别记 $X_{k(0)}$ 和 N 为初始的迭代向量和需要提取的波段个数。从一个波长开始, 每次循环计算它在未选入的波长上的投影, 将投影向量最大的波长引入波长组合, 直到循环 N 次, 每一个新入选的波段, 都与前一个线性关系最小。通过循环会得到 $N \times K$ 对波段组合, 对每一对 $X_{k(0)}$ 和 N 所决定的组合分别建立多元回归

模型,并用预测均方根误差 RMSE 来决定所建模型的优劣,最小的 RMSE 对应的 $X_{k(o)}$ 为最佳的波段组合。

1.2.3 LASSO 算法 Least Absolute Shrinkage and Selection Operator 是 Robert Tibshirani 提出的一种变量选择方法。Lasso 是在一般线性最小二乘的前提下添加惩罚约束,要求各系数的绝对值之和比某一常数要小,因此有可能使得一些变量的系数为零,最终将无效变量删除达到筛选的目的^[16]。

1.2.4 随机森林 随机森林是一种并行式集成学习方法,以决策树为基础,并在决策树训练过程中引入随机属性选择。它通过自助法重采样技术,从原始训练样本集 N 中有放回地重复随机抽取 K 个样本生成新的训练样本集合,然后根据自助样本集生成 K 个分类树组成随机森林^[17-18]。

1.2.5 支持向量机 支持向量机是一种基于统计学习理论的机器学习算法,它建立在 VC(Vapnik-Chervonenkis Dimension)维理论和结构风险最小原理基础上,能较好地解决小样本、非线性、高维数和局部极小点等实际问题,同时能获得较好的泛化能力^[19]。

1.2.6 偏最小二乘 偏最小二乘回归是一种新型的多变量回归分析方法,可以实现回归建模、简化数据结果和分析两组变量见得相关性,给多元数据统计分析带来极大便利,它能在回归建模过程中采用数据将为、信息综合和筛选技术,提取对系统最佳解释能力的新综合成分^[20]。

1.3 建模方法与精度评价

分别采用偏最小二乘,随机森林与支持向量机建立叶片氮含量反演模型,并采用决定系数 R^2 、均方根误差来评价各个模型的精度。 R^2 表示相关密切程度, RMSE 用来衡量估测值与真实值之间的偏差程度。 R^2 越大, RMSE 的绝对值越小,表示模拟结果精度更高。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - y_i)^2}{n}}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

式中: Y_i 和 y_i 分别为实测叶片氮含量和模拟叶片氮含量, \bar{Y} 为实测氮含量的平均值, n 为样本数。

2 结果与分析

2.1 不同光谱形式与氮含量的相关性分析

原始反射率 R 与氮含量呈负相关关系,在 500~730 nm 波段的相关系数超过 -0.6,在 550、702 nm 左右相关性达到最大,且差异显著 ($P < 0.05$),在 550 nm 左右相关性最大是由于叶绿素对绿光的反射作用,702 nm 位于红边区域,有相关研究发现红边波段与植物叶绿素含量、生物量等参数存在显著相关性;氮含量与原始光谱的相关性在近红波段较小(表 1)。

一阶导数 R' 能够有效消除叶面积变化的影响,反射率的一阶微分与氮含量的相关性变化较为复杂,在 490~544、556~710 和 736~800 nm 超过 0.6,其中 R' 在 530、574、626、694 nm 达到最大,近红波段的相关性较小,但在 1 200、1 672、2 164、2 310 nm 相关性分别达到峰值 0.6 以上(图 1)。

$\log(1/R)$ 相关性曲线与原始反射率的相关性曲线线型一致,但相关性相反。 $\log(1/R)'$ 与氮含量的相关性变化也较为复杂,在 490~546、620~670、672~694、702~804 nm 相关性在 0.6 以上,在红外波段 1 204、1 672、2 158 nm 的相关性形成峰值,达到 0.6 以上。

应用 $\lambda_1 - \lambda_2$, λ_1 / λ_2 与氮含量的相关系数筛选出的波段组合如表 1 所示, R 与 $\log(1/R)$ 筛选出的波段集中在可见光波段,两种导数形式筛选出的波段集中在近红波段;原始波段反射率及其变换形式筛选出的波段组合都包含 522、526 nm;导数形式筛选出的波段组合均包含 2 158、2 044 和 2 070 nm 3 个波段,虽然这几个单独的波段与氮含量的相关性并不强,但其波段组合与氮含量的相关性达到 0.86 以上,说明该波段组合包含了氮含量的大部分信息。

2.2 连续投影算法筛选敏感波段特征筛选

ACCP 叶片数据集共 91 组鲜叶片数据,随机选取其中 65 组数据进行连续投影算法。连续投影方法将 65 组数据分为训练集和校正集,以校正集的预测均方根误差来确定最佳的光谱变量总数,均方根误差和决定系数来评价最佳的波段变量组合。对以上 4 种形式的光谱参数分别采用连续投影算法,确定氮含量反演的最佳光谱变量(表 2)。

表 1 相关系数法筛选的波段组合

Table 1 Band combinations selected by correlation coefficient method

光谱参数	波段位置/nm	相关系数
R	522,526;528,530	-0.888 2,0.868 8
R'	1 876,2 158;2 044,2 070	0.945 6,0.922 2
log(1/R)	654,692;522,526	0.878 9,0.888 8
log(1/R)'	1 416,2 158;2 044,2 070	0.932 6,0.932 5

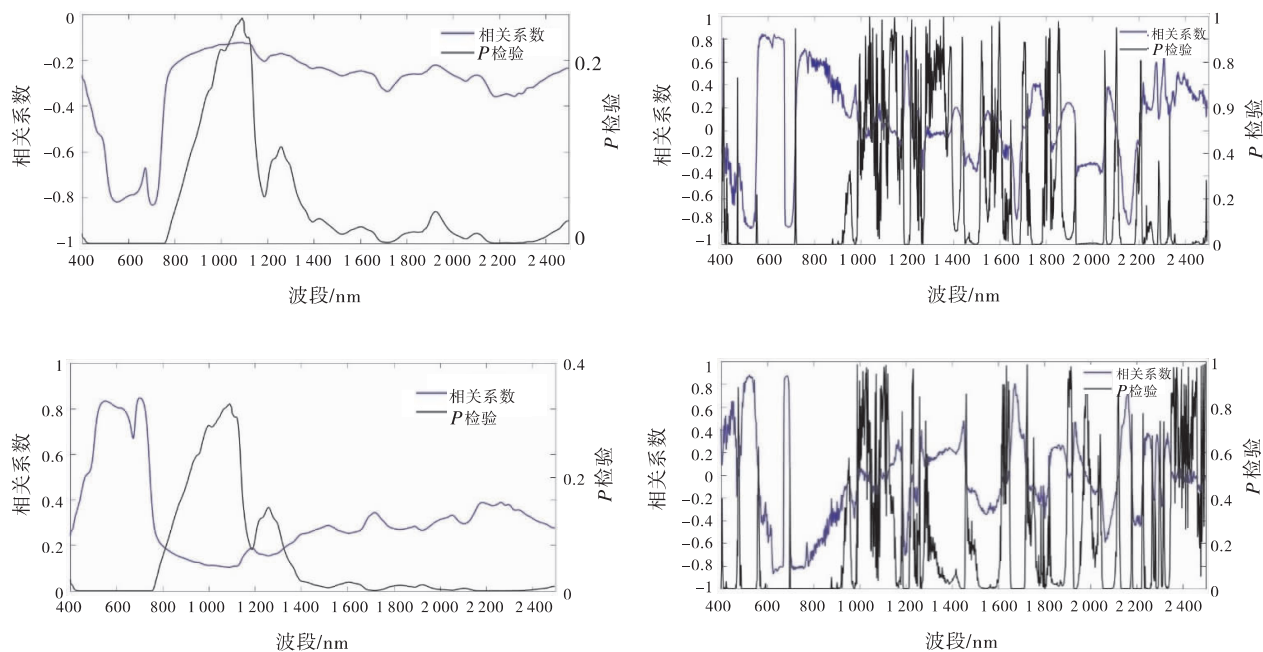


图 1 R,R',log(1/R),log(1/R)'与氮含量的相关性

Fig. 1 The correlation between R,R',log(1/R),log(1/R)' and leaf nitrogen content

表 2 连续投影法筛选的波段组合

Table 2 Band combinations selected by successive projection algorithm

光谱参数	波段数	波段位置/nm	RMSE
R	16	404,460,522,544,716,740,776,978,1 188,1 320,1 408,1 504,1 666,1 716,2 086,2 226	0.230 3
R'	3	694,1 886,2 162	0.234 3
log(1/R)	7	704,766,872,1 852,2 12 6,2 17 8,2 39 6	0.203 5
log(1/R)'	3	1 366,2 164,2 208	0.211 0

原始反射率所选的波段在波段范围内分布较为均匀(图 2),位于可见光的波段有 6 个, 2 个波段位于红边区域,位于波峰或波谷处的波段有 6 个。一阶导数只选择了 3 个变量,其中一个位于可见光波段。Log(1/R)投影算法选择了两个红光波段,其余均为近红外波段。log(1/R)'仅选择了 3 个红外波段。校正数据集 RMSE 最低的是 Log(1/R)形式, RMSE 最高的是 一阶导数形式。

2.3 LASSO 算法

采用 Lasso 算法对 4 种光谱形式进行变量筛选,各光谱形式筛选出的变量数均比连续投影筛选出的变量数多。原始光谱反射率筛选出 22 个波段,其中有 12 个波段位于可见光区。Log(1/R)筛选出的变量数为 16 个,其中 9 个可见光波段。两种导数形式均筛选出 23 个变量数,一阶导数筛选出的可见光波段 8 个, Log(1/R)'筛选出的可见光波段 6 个(表 3)。

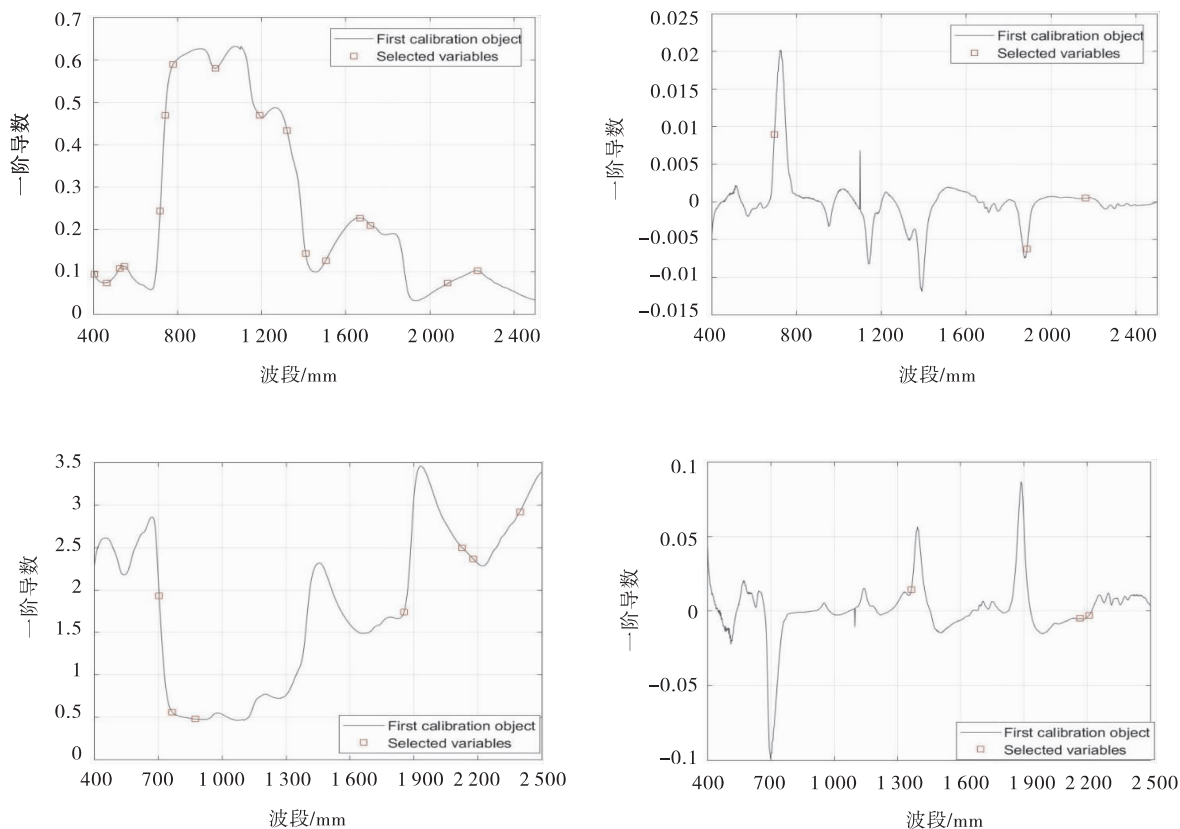


图 2 连续投影算法筛选出的波段位置

Fig. 2 The band wavelength selected by successive projection algorithm

表 3 LASSO 筛选的波段组合

Table 3 Band combinations selected by LASSO algorithm

光谱形式	波段数	筛选的光谱位置/nm
R	23	400,406,410,688,690,692,694,696,700,724,730,734,1 360,1 362,1 368,1 602,1 604,1 606,1 608,1 610,2 180,2 182,2 184
R'	23	400,420,436,442,444,536,540,544,802,838,856,866,1 196,1 198,1 200,1 666,1 668,1 672,2 070,2 158,2 164,2 216,2 218
Log(1/R)	16	410,696,698,700,702,706,730,734,740,1 362,1 368,1 604,1 608,2 180,2 182,2 184
Log(1/R)'	23	442,524,536,540,544,674,800,810,818,838,1 192,1 196,1 198,1 200,1 666,1 668,1 672,2 064,2 158,2 164,2 204,2 214,2 218

2.4 模型预测结果与对比

分别将相关系数筛选法、连续投影算法、LASSO 算法筛选出的波段作为偏最小二乘、随机森林、支持向量机的输入来估算叶片氮含量,以 65 组数据为建模数据,26 组数据为验证数据。不同方法组合的反演结果如表 4 及图 3 所示。

3 种变量筛选方法和 3 种建模方法的组合中,

RMSE 在 0.19~0.38, R^2 在 0.6~0.89, 其中,随机森林算法对相关系数法和 LASSO 算法筛选出的 $\text{Log}(1./R)'$ 形式变量的反演误差最小, RMSE 在 0.19~0.20, 能够解释氮含量的变化, 但该算法对筛选出的其他形式的变量反演误差较大, 表现并不稳定; 不论采用何种建模方法, 对于相关系数法, 其 $\text{log}(1./R)'$ 形式的变量反演效果较好; 对于连续投影法, 其筛选出的 R'

变量,反演误差较小;对于 LASSO 算法,其筛选出的 R' 和 $\log(1/R)'$ 形式的变量反演效果要明显优于其余两种变量形式。同时,分别采用偏最小二乘,随机森林、支持向量机进行全波段的建模,反演结果分别为 $R^2=0.605$ 、 $RMSE=0.4542$, $R^2=0.659$ 、 $RMSE=0.$

3352 , $R^2=0.7304$ 、 $RMSE=0.3484$ 。表明,相关系数 $\&Log(1./R)'$, LASSO $\&Log(1./R)'$, SPA $\&R'$ 能够有效筛选出冷杉叶片氮含量的敏感波段组合,且前两种方法组合效果更优。

表 4 不同方法组合的反演结果

Table 4 Inversion results from the combination of different methods

波段形式		偏最小二乘		随机森林		支持向量机	
		R^2	Rmse	R^2	Rmse	R^2	Rmse
相关系数法	R	0.7049	0.3280	0.6028	0.3804	0.6618	0.3510
	R'	0.8413	0.2405	0.6628	0.3505	0.8577	0.2276
	$\log(1/R)$	0.7702	0.2894	0.6104	0.3768	0.7430	0.3060
	$\log(1/R)'$	0.8722	0.2158	0.8904	0.1998	0.8790	0.2099
连续投影法	R	0.7826	0.2815	0.6780	0.3425	0.7578	0.2970
	R'	0.8533	0.2312	0.8414	0.2404	0.8612	0.2248
	$\log(1/R)$	0.7532	0.2999	0.7113	0.3243	0.7582	0.2968
	$\log(1/R)'$	0.7764	0.2855	0.7126	0.3236	0.8217	0.2549
LASSO 算法	R	0.6932	0.3344	0.6992	0.3311	0.6785	0.3423
	R'	0.8305	0.2486	0.8517	0.2324	0.8646	0.2221
	$\log(1/R)$	0.6176	0.3733	0.6715	0.3459	0.7275	0.3151

3 讨论

3.1 不同变量筛选方法和建模方法的组合效果比较

从图 3 可以看出两种导数形式估算出的氮含量与实测氮含量 1:1 的对应关系更好,散点基本上沿着 $y=x$ 线分布,具有较好的线性关系。对于 R 及 $\log(1/R)$ 形式的反演,三种建模方法的结果与实测值相比普遍偏高,大部分点分布在 $y=x$ 线以上。两种导数形式的反演结果较为均匀紧致地分布在 $y=x$ 线两侧,反演结果与实测数据更接近。从实测数据与反演结果的拟合图可以看出,相关系数法 $\&\log(1/R)'$ 以及 LASSO $\&\log(1/R)'$ 中实测数据与反演结果的拟合函数具有较大的斜率,较小的截距,且 R^2 较大, $RMSE$ 较小,具有较高的反演精度(图 3)。

3.2 3 种筛选方法的敏感波段分析

两种最优敏感参数筛选方法组合筛选出的最优敏感参数均为其导数形式。以往的研究中,位于红边区域的 715 nm 和 731 nm 经常被认为是氮的敏感波段,这一区间植物的反射率是由叶绿素在红波段的强吸收及叶肉组织在近红外波段的散射引起的,叶片中的氮含量与叶绿素含量密切相关^[7]。717, 730 nm, R'_{715}/R'_{705} , 732 nm 也被认为是氮含量的敏感波段^[21-24]。3 种变量筛选方法中,连续投影算法筛选出 716 nm 为氮

含量的敏感波段, Lasso 算法筛选出 730nm 为氮的敏感波段,这与前人的研究相一致,说明两种算法筛选氮敏感波段的有效性。相关系数法筛选出的 R 形式的敏感波段在 522~530 nm。在 450~720 nm, 氮含量与其反射率成负相关关系,氮含量越高,反射率越低,这一区间的反射率主要受叶片色素吸收的影响^[25],而氮素含量的变化会影响植物叶绿素含量,因而当植株缺氮时,反射率升高^[26-27]。连续投影算法筛选出的 460, 978, 1188, 1408, 1504, 2086 nm 均与叶片其他含量相关,其中 460 nm 处与叶绿素 b 相关, 978 nm 与叶片水分和淀粉相关, 1188 nm 与水分、纤维素、淀粉、木质素相关, 1408 nm 与叶片水分相关, 1504 nm 与蛋白质和氮含量相关, 2086 nm 与糖分和淀粉相关^[28]。这些波段大部分不是前人研究中氮的吸收波段或者是敏感波段,一个可能的原因是氮的吸收特征与叶片其他化学含量存在相关性,而与这些化学含量的吸收特征是重叠的;另一个原因可能是连续投影算法只是筛选出了与氮含量具有相关性的波段,但未考虑其特定的物理学意义。相比之下, LASSO 算法筛选出的 2180 nm 与蛋白质和氮含量相关,其余波段与叶片化学含量均不相关,可能的原因是该算法只是筛选出与氮含量数学上相关的特征组合,而未能考虑到其物理化学意义。

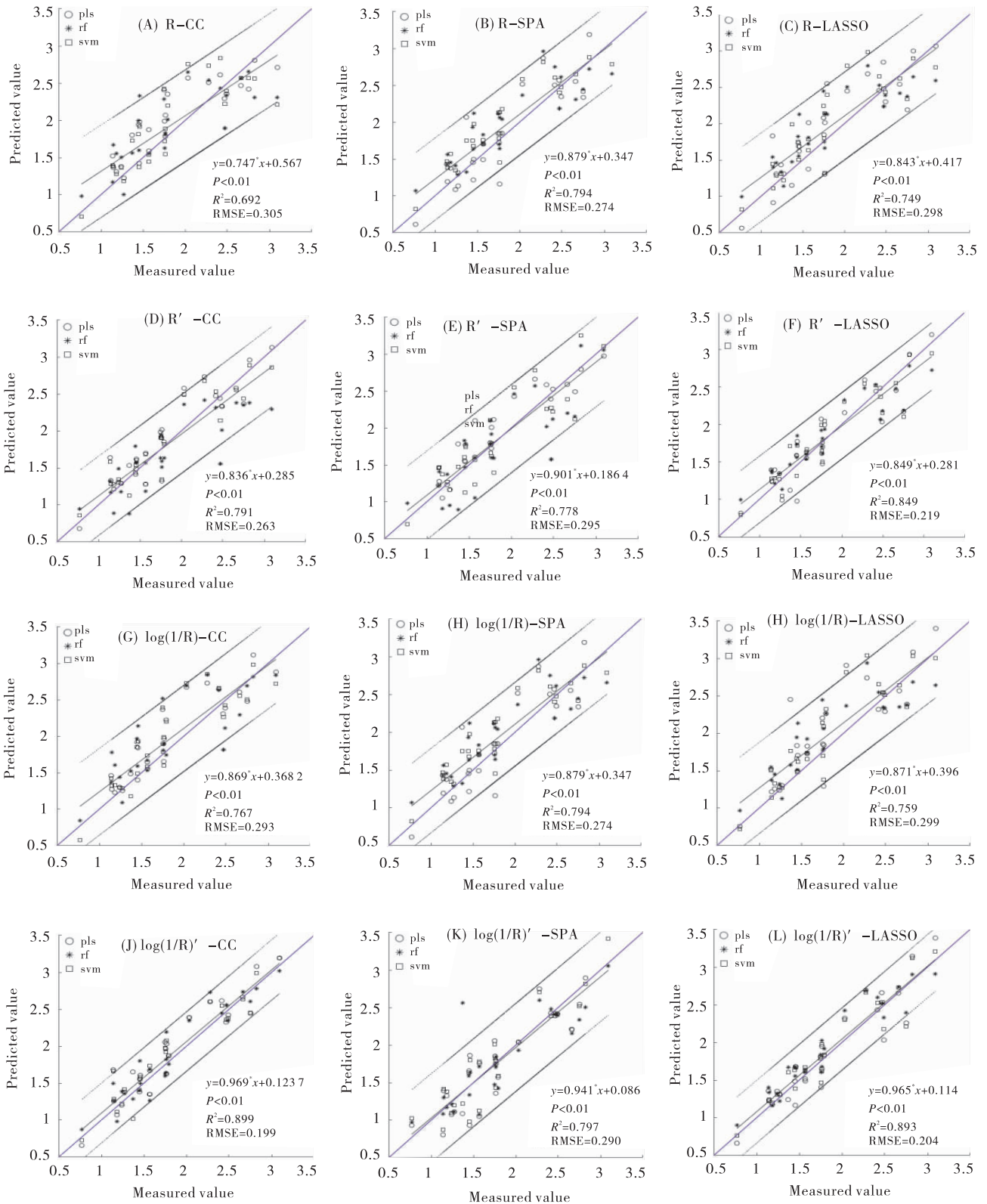


图 3 实测氮含量与不同方法反演氮含量散点图

Fig. 3 Scatter plots of measured and predicted values based on different methods

注:蓝色线为 $y=x$,黑色实线为拟合线,黑色虚线为 95%置信区间

4 结论

采用相关系数法、连续投影法、LASSO 算法对 4

种形式的光谱变量 $R, R', \log(1/R), \log(1/R)'$ 进行敏感波段筛选,对筛选后的敏感波段分别采用偏最小二乘、随机森林、支持向量机对氮含量进行建模反演,得

出以下结论:

1)两种导数形式变量的反演误差最小。

2)相关系数 $\& \cdot \text{Log}(1/R)'$, LASSO $\& \cdot \text{Log}(1/R)'$

能够有效筛选出冷杉叶片氮含量的敏感波段组合,无论采取何种建模方法,其估算效果在几种筛选方法组合中最好, $R^2 > 0.84$,RMSE 在 0.19~0.24,估算效果明显优于采用 3 种建模方法进行全波段建模。

3)随机森林算法对相关系数法和 LASSO 算法筛选出的 $\text{Log}(1/R)'$ 形式变量的反演误差最小,但对其他形式筛选出的变量反演结果的误差变化范围较大,表现并不稳定。

4)3 种变量筛选方法筛选出的 R 形式的变量与前人研究相符,连续投影算法能筛选出更多与叶片其他化学含量相关的波段,证明了 3 种变量筛选方法的有效性。

参考文献:

[1] Lambers H, Chapin F S, Pons T L. Plant Physiological Ecology[M]. Springer New York, 1998.

[2] 张霞, 刘良云, 赵春江, 等. 利用高光谱遥感图像估算小麦氮含量[J]. 遥感学报, 2003, 7(3): 176—181.

[3] Wang L, Zhou X, Zhu X, *et al.* Estimation of leaf nitrogen concentration in wheat using the MK—SVR algorithm and satellite remote sensing data[J]. Computers & Electronics in Agriculture, 2017, 140: 327—337.

[4] 朱艳, 李映雪, 周冬琴, 等. 稻麦叶片氮含量与冠层反射光谱的定量关系[J]. 生态学报, 2006, 26(10): 3463—3469.

[5] 刘冰峰, 李军, 赵刚峰, 等. 夏玉米叶片全氮含量高光谱遥感估算模型研究[J]. 植物营养与肥料学报, 2012, 18(4): 41—52.

[6] Tarpley L, Reddy K R, Sassenrath—Cole G F. Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration[J]. Crop Science, 2000, 40(6): 1814.

[7] Pacheco—Labrador J, Gonz lez—Casc n, Rosario, Mart n M Pilar, *et al.* Understanding the optical responses of leaf nitrogen in Mediterranean Holm oak (*Quercus ilex*) using field spectroscopy[J]. International Journal of Applied Earth Observation and Geoinformation, 2014, 26(2): 105—118.

[8] Abdel—Rahman E M, Ahmed F B, Berg M V D. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy[J]. international journal of applied earth observations & geoinformation, 2010, 12(S1): 52—57.

[9] Yao X, Zhu Y, Tian Y C, *et al.* Exploring hyperspectral bands and estimation indices for leaf nitrogen accumulation in wheat[J]. International Journal of Applied Earth Observations & Geoinformation, 2010, 12(2): 1—100.

[10] Ullah S, Si Y, Schlerf M, *et al.* Estimation of grassland biomass and nitrogen using MERIS data[J]. International Journal of Applied Earth Observation and Geoinformation, 2012, 19: 196—204.

[11] Sanches I, Tuohy M, Hedley M, *et al.* Seasonal prediction of in situ pasture macronutrients in New Zealand pastoral systems using hyperspectral data[J]. Int J Remote Sens, 2013, 34: 276—302.

[12] Zhang M, Li M, Liu W, *et al.* Analyzing the performance of statistical models for estimating leaf nitrogen concentration of *Phragmites australis* based on leaf spectral reflectance[J]. Spectroscopy Letters, 2019, 52(9): 1—9.

[13] Yao Xia, Huang Yu, Shang Guiyan, Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration.

[14] Yoder B J, Johnson L F. Seedling Canopy Reflectance Spectra, 1992—1993 (ACCP). ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC>

[15] 陈斌, 孟祥龙, 王豪. 连续投影算法在近红外光谱校正模型优化中的应用[J]. 分析测试学报, 2007(1): 66—69.

[16] 高少龙. 几种变量选择方法的模拟研究和实证分析[D]. 济南: 山东大学, 2014.

[17] Svetnik V. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling[J]. J chem inf comput sci, 2003, 43(6): 1947—1958.

[18] Rodriguez—Galiano V, Mendes M P, Garcia—Soldado M J, *et al.* Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain)[J]. Science of the total environment, 2014, 476—477(1): 189—206.

[19] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2—10.

[20] 付元元, 王纪华, 杨贵军, 等. 应用波段深度分析和偏最小二乘回归的冬小麦生物量高光谱估算[J]. 光谱学与光谱分析, 2013, 32(5): 1315—1319.

[21] Vogelmann J, Rock B, Moss D. Red edge spectral measurements from sugar maple leaves[J]. Int J Remote

- Sens, 1993, 14, 1563–1575
- [22] Corp L A, Middleton E M, Campbell P E, *et al.* Spectral indices to monitor nitrogen – driven carbon uptake in field corn[J]. *J Appl Remote Sens*, 2010, 4: 043555.
- [23] Tian Y, Yao X, Yang J, *et al.* Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground-and space-based hyperspectral reflectance[J]. *Field Crop Res*, 2011, 120: 299–310.
- [24] Ramoelo A, Skidmore A K, Schlerf M, *et al.* Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations[J]. *ISPRS J Photogramm, Remote Sens*, 2011, 66: 408–417
- [25] Gitelson A A, Merzlyak M N. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, 1996, 148(3/4): 483–492.
- [26] 张雪红, 刘绍民, 何蓓蓓. 不同氮素水平下油菜高光谱特征分析[J]. *北京师范大学学报(自然科学版)*, 2007, 43(3): 245–249.
- [27] 刘根华, 黄坚钦, 潘春霞, 等. 基于反射光谱的山核桃幼苗氮素营养状况分析[J]. *林业科学*, 2011, 47(1): 165–171.

Study on hyperspectral Inversion technology of Nitrogen content in *Abies fabri* leaves

SONG Xue-lian¹, WANG Zhi-wei¹, ZHANG Wen^{1,2}, DING Lei-lei¹, LIU Jia-jia¹,
RUAN Xi-rui¹, ZHANG Wei³

(1. *Institute of prataculture, Guizhou Academy of Agricultural Sciences, Guiyang 550006, CHina*; 2. *Guizhou Sunshine Grass Technology Co., Ltd., Guiyang 550006, China*; 3. *Guizhou Survey & Design Research Institute for Water Resources and Hydropower Guizhou, Guiyang 550002, China*)

Abstract: Nitrogen is an essential element for plant growth. Monitoring nitrogen content is helpful to understand the metabolic process and health status of plants. Hyperspectral technology, a nondestructive and rapid method to detect leaf nitrogen content, has become one of the important methods for nitrogen analysis. In this paper, hyperspectral technology was used to estimate leaf nitrogen content in *Abies fabri*. Correlation coefficient method, successive projections algorithm and LASSO algorithm were used to select sensitive bands of four spectral variables including R , R' , $\log(1/R)$ and $\log(1/R)'$. Partial least square, random forest and support vector machine were used to conduct modeling inversion for the selected sensitive bands. The results showed that the inversion error of the two derivatives was the smallest. The correlation coefficient including $\log(1/R)'$ and LASSO & $\log(1/R)'$ could be used effectively to select the combined sensitive band for leaf nitrogen content. The estimation effects using the combination of those two variables gave the best results among several selection methods with $R^2 > 0.84$ and RMSE ranged from 0.19 ~ 0.24, and the estimation effect was significantly better than that of all the three modeling methods using the whole bands. The inversion errors derived from random forest algorithm in combination with correlation coefficient and $\log(1/R)'$, and random forest algorithm in combination with LASSO & $\log(1/R)'$ were smallest. But the error range of the random forest method to other forms of variables was large and its performance was not stable. The variables in the form of R screened by the three variable screening methods were consistent with previous studies. The successive projection algorithm could screen more bands related to other chemical content of leaves, which proved the effectiveness of the three variable screening methods.

Key words: hyperspectrum; *Abies fabri* leaves; nitrogen content; band selection