

DOI: 10.5846/stxb201311042664

郭超凡, 段福洲, 郭道宇, 赵文吉, 刘克. 基于最佳波段判别的湿地植物叶片全氮反演研究. 生态学报, 2014, 34(17): 4839-4849.

Guo C F, Duan F Z, Guo X Y, Zhao W J, Liu K. Estimating Wetland plant leaf total nitrogen content based on optimal bands of reflectance from wetland vegetation. Acta Ecologica Sinica, 2014, 34(17): 4839-4849.

基于最佳波段判别的湿地植物叶片全氮反演研究

郭超凡^{1,2,3,4}, 段福洲^{1,2,3,4}, 郭道宇^{1,2,3,4,*}, 赵文吉^{1,2,3,4}, 刘克⁵

(1. 首都师范大学资源环境与旅游学院, 北京 100048;

2. 北京市城市环境过程与数字模拟重点实验室-省部共建国家重点实验室培育基地, 北京 100048;

3. 三维信息获取与应用教育部重点实验室, 北京 100048;

4. 资源环境与地理信息系统北京市重点实验室, 北京 100048; 5. 北京市地质研究所, 北京 100120)

摘要: 利用高光谱遥感技术定量估测湿地植被叶片全氮含量, 对于监测和诊断湿地植被的生理状况及生长趋势具有重要意义。但叶片氮素遥感诊断研究多存在反演模型过拟合、入选波段与生化参量间因果关系不明确和入选变量间“多重共线性”等局限。以芦苇(*Phragmites australis*)和香蒲(*Typha angustifolia*)叶片全氮含量作为研究对象, 通过谱带分区, 分区最佳波段选取和偏最小二乘回归相结合的方法构建芦苇和香蒲叶片全氮含量反演模型, 并利用交叉验证决定系数(R_{cv}^2)和均方根误差(RMSE_{cv})对模型精度进行检验, 尝试克服传统反演方法中的不足。结果表明, 不同湿地植物类型相比, 利用芦苇反射光谱建立的预测模型精度都高于香蒲。不同回归模型相比, 一阶导数光谱-偏最小二乘回归模型(FDS-PLSR)精度远高于原始光谱-偏最小二乘回归模型(OS-PLSR)。芦苇最佳模型交叉验证决定系数(R_{cv}^2)达到了0.84, 方根误差(RMSE_{cv})为0.10, 香蒲最佳模型交叉验证的决定系数(R_{cv}^2)达到了0.66, 方根误差(RMSE_{cv})为0.13, 是构建湿地植物芦苇和香蒲光谱与叶片全氮含量关系的最佳模型。在不降低湿地植物叶片氮含量反演精度的基础上, 有效地避免了传统地物高光谱模型反演中的局限性, 是无损害遥感探测方面的有益尝试。

关键词: 湿地植物; 高光谱; 全氮; 波段识别; 偏最小二乘法

Estimating Wetland plant leaf total nitrogen content based on optimal bands of reflectance from wetland vegetation

GUO Chaofan^{1,2,3,4}, DUAN Fuzhou^{1,2,3,4}, GUO Xiaoyu^{1,2,3,4,*}, ZHAO Wenji^{1,2,3,4}, LIU Ke⁵

1 College of Resources Environment and Tourism, Capital Normal University, Beijing 100048, China

2 Urban Environmental Processes and Digital Modeling Laboratory, Beijing 100048, China

3 Laboratory of 3D Information Acquisition and Application, MOST, Beijing 100048, China

4 Beijing Municipal Key Laboratory of Resources Environment and GIS, Beijing 100048, China

5 Beijing Institute for Geological Engineering, Beijing 100120, China

Abstract: Utilization of hyperspectral remote sensing technology to estimate wetland plant leaf nitrogen content quantitatively in large area is important to monitor and diagnosis of physiological condition and growth trend of wetland vegetation. However, there are many limitations such as over-fitting of inversion model, indeterminate causal relationship between the selected bands and biochemical parameters, “multicollinearity” of selected bands in leaf nitrogen diagnosis remote sensing research. The total nitrogen content of leaves of typical wetland plants, *Phragmites australis* and *Typha angustifolia*, was selected as our study objects. These plants grow in South Wetland purification system in the Olympic Park

基金项目: 国家自然科学基金(40901281); 北京市教育委员会科技计划面上项目(KM201310028012)

收稿日期: 2013-11-04; **修订日期:** 2014-07-01

* 通讯作者 Corresponding author. E-mail: xiaoyucnu@126.com

in Beijing, a typical wetland using reclaimed water. The leaf reflectance spectra of main wetland plants were acquired by means of an ASD FieldSpec 3 spectrometer (350—2500nm). Leaf total nitrogen content was determined by Kjeldahl nitrogen measurement method after acquiring the leaf reflectance spectra. The method, “Give a cross correlation analysis”, was used to build the correlations between leaf nitrogen content and the original spectrum, first derivative spectrum. Then, the selected bands were divided into some areas according to QiSeGuang spectral range and the interval of selected bands. Those bands which are high frequency, frequency more, ranked high frequency are chosen as representative bands of different areas and considered to be the optimal bands used to build the regression model. Finally, partial least squares was used to build inversion model. The accuracy of this model was tested with cross-validated coefficient of determination (R_{cv}^2) and cross-validated root mean square error ($RMSE_{cv}$). The results show that the first derivative transformation can effectively improve the sensitivity of the original spectrum leaf nitrogen content inversion, and fully reflect the sensitivity of near infrared wave band representing leaf total nitrogen content. The accuracy of regression model based on the first derivative and the partial least-squares was much higher than that of the original spectra. In the regression model of the reed, verification accuracy (R_{cv}^2) reached 0.84, square root error ($RMSE_{cv}$) was 0.11, in the regression model of the cattail, verification accuracy (R_{cv}^2) reached 0.66, square root error ($RMSE_{cv}$) was 0.13, which were the optimal models to estimate leaf total nitrogen content. The determination of parameters in “Give a cross correlation analysis” provided a scientific basis for building the model of eliminating “singularband” and reducing “multicollinearity” problem. Spectrum zone division provides a scientific basis for revealing the causal relationships between optimal band and biochemical parameters. And partial least squares regression method was used to avoid “multicollinearity” of selected bands. The result from this study can not only fill the gaps in the detection of leaf nitrogen using remote sensing, but also provide a strong scientific basis for the nitrogen content monitoring and management of urban wetlands using reclaimed water. At last, Partial least squares regression method was used to avoid “multicollinearity” of selected bands.

Key Words: wetland plant; hyperspectrum; total nitrogen; band identification; partial least squares

植物叶片的光合作用、呼吸、蒸散发和分解等生物过程与叶片的生化参量如叶片叶绿素、水分、氮和纤维素等含量具有密切的关系^[1],而植物受叶片内叶肉细胞、叶绿素、水分含量、氮素含量以及其他生理生化参量的影响表现出不同的反射光谱曲线^[2-3]。大量文献成功借助于植物冠层或叶片光谱特征对植物生化参量进行实时监测和快速诊断,其结果对于理解大范围的生态系统功能特性具有重要的意义^[4]。由于叶绿素含量的变化会引起光谱趋势改变的观点已得到普遍认同^[5-6],同时有文献证明叶片叶绿素含量与叶片氮含量具有正相关关系^[7],为基于叶片光谱特征的叶氮含量反演提供了科学依据。早期的叶片氮素遥感诊断研究多集中于借助指数模型判别氮素的敏感波段^[8]及基于光谱数据的变换提高光谱吸收特征可辨性^[9]。但指数模型在实际应用中容易受植物品种、生育期、生长环境的影响^[8],其反演精度存在不稳定性和普适性较差的问题。其次指数模型对于高光谱数据丰富的光谱信息而言极其有

限,容易造成重要信息的丢失^[10]。为克服指数模型的局限性,许多学者通过原始光谱变换和寻求各种统计方法构建基于光谱特征的植物氮素含量估算模型。其中采用逐步回归的方法来判别生物生化参量的敏感波段得到了广泛的应用^[11-12],该方法具有一定的有效性和稳定性,但叶片氮素遥感诊断研究多存在反演模型过拟合、入选波段的“多重共线性”和入选波段与生化参量间因果关系不明确^[9,13]。随后有学者提出通过光谱吸收特性的连续移除法、带深标准化分析和多元逐步线性回归相结合的模型构造方法并获得了比较一致的、适用于多种物种且精度较高的氮素及其他化学组分含量的预测方法^[13],但该方法容易受到冠层结构及叶片水分等因素的影响,并不适宜用于湿生叶片生化参量的估测^[11]。偏最小二乘法已被证明是从植被光谱数据提取叶面属性的最有效的经验方法,开始较多地应用于对植被和农作物的遥感监测中^[14]。

本课题组前期借助偏最小二乘法进行了植物叶

片全氮含量的高光谱反演研究,因该方法具有较多的光谱参数,考虑了全谱区各波长点的光谱参数,并解决了多元线性回归中变量多重共线性等问题,因而具有较高反演精度^[14]。但由于偏最小二乘法全谱区所有波段参与模型反演,同样无法避免受土壤背景、传感器信噪比、大气吸收、冠层结构及叶片水分等因素对反演精度的影响。同时基于该方法所构建的模型同样无法明确入选波段与生化参量间的因果关系。基于此,本研究以奥林匹克公园南园湿地优势植物芦苇和香蒲为研究对象,选取 400—1000 nm 波段范围的原始光谱及其派生的一阶导数光谱为光谱数据源,通过舍一交叉法选取与叶氮含量相关性较高且相关频率较大的波段作为叶片全氮含量反演模型的最佳波段集,并利用偏最小二乘法构建这些波段集与叶片全氮含量的反演模型,研究结果预期在反演模型构建中避免模型入选波段多重共线性、明确入选波段与生化参量间因果关系、剔除构建模型中的奇异波段和消除模型的过拟合等方面进行有益的尝试,为湿地植物生长的遥感监测提供科学依据。

1 研究区概况

奥林匹克公园南园湿地地处北京市区北部,位于奥林匹克公园内。该区属暖温带半干旱半湿润大陆性季风气候。多年平均气温 11—12 ℃,极端最高气温 41.6 ℃,极端最低气温 -21.2 ℃。多年平均水面蒸发量在 1200 mm 左右,多年平均降雨量约 600 mm。降雨年内分配不均,多集中在汛期 6—9 月,约占全年的 85%。奥林匹克公园以清河、北小河再生水处理厂的出水为补水水源,并循环利用,且补水入湖前流经南园湿地以改善水质。南园湿地以再生水处理湿地和循环水处理湿地为核心,植物氧化塘、生态氧化塘为辅助,混合生态功能区为补充,形成多层次的湿地净化系统(图 1)^[15]。该净化系统的各个部分均布有大量水生植物,主要包括芦苇、香蒲、水葱、千屈菜等。通过植物的净化作用出水水质可以达到地表Ⅲ类水水质标准(GB3838—2002)。

2 研究方法

2.1 湿地植物光谱采集及 TN 含量的测定

选择研究区主要的湿地植物芦苇和香蒲为光谱

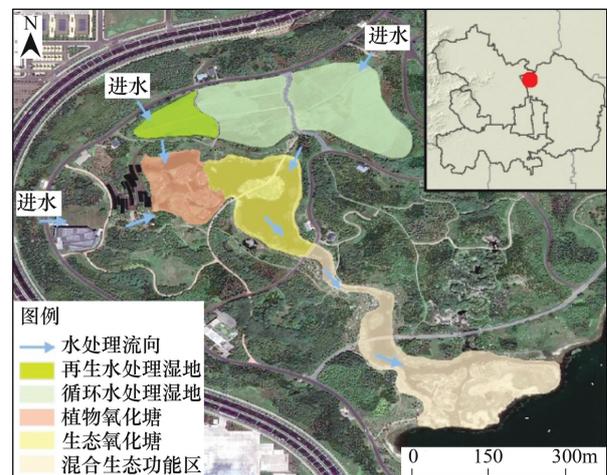


图 1 奥林匹克公园南园湿地地理位置及净化系统布局图^[15]

Fig.1 Location and distribution of the purification system in the South Wetland^[15]

采集对象。根据两种湿地植物在每个水质净化子系统(再生水处理湿地,循环水处理湿地,植物氧化塘,生态氧化塘,混合生态功能区)的分布情况,设置芦苇(*Phragmites australis*)光谱采样点 28 个,香蒲(*Typha angustifolia*)光谱采样点 20 个。于 2010 年 8 月进行湿地植物叶片光谱测量,测量仪器采用美国 ASD (Analytical Spectral Device) 公司 FieldSpec 3 便携式地物波谱仪。该仪器探头视场角度为 10°,光谱范围为 350—2500 nm,光谱分辨率在 700 nm 时为 3 nm,在 1400 nm 时为 8.5 nm,在 2100 nm 时为 6.5 nm,光谱采样间隔在 350—1000 nm 为 1.4 nm,在 1000—2500 nm 为 2 nm。野外测量选择晴朗无风的天气,测量时间为 10:00—14:00,根据天气条件进行优化,至多每 20 min 利用白板进行 1 次。测量时选择植株从顶部下数第 1 片完全展开的新叶,探头距离叶片中间部位约 5 cm,保证探头的视野范围落在叶片上,同时注意避开叶脉位置,每个样点一次保存 10 条光谱。与光谱测量同步,取相应叶片,在 105 ℃ 下杀青 30 min 后,再于 80 ℃ 下烘干至恒量,用 H₂SO₄-H₂O₂ 消煮后,利用凯氏法测定叶片全氮含量。

2.2 数据处理

2.2.1 光谱数据预处理

本文选取的波段范围为:400—1000 nm,对每个样点采集的光谱进行平均。为减少数据冗余度,对光谱进行重采样(仪器输出光谱自动重采样为 1 nm 的分辨率),采样间隔为 5 nm,并采用 Savitzky-Golay 方法进行平滑,该方法为窗口移动多项式最小二乘

平滑,能够保留光谱细微特征并减少随机噪声,提高光谱信噪比^[16]。

2.2.2 光谱数据的一阶导数变换

光谱微分技术是高光谱遥感数据最主要的分析技术之一,利用光谱微分技术可以消除大气和背景噪声的影响,突出光谱特征和规律。一阶导数的计算方法如公式(1):

$$R'(\lambda_i) = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{\lambda_{i+1} - \lambda_{i-1}} \quad (1)$$

式中, λ_i 为每个波段的波长, $R'(\lambda_i)$ 为波长 λ_i 的一阶导数。

2.2.3 模型建立及精度验证

基于 SPSS, SIMCA-P 11.5 等软件,利用相关性分析和统计分析的方法选取与叶片全氮含量关系密切的敏感波段,并运用偏最小二乘回归法分别构建原始最佳光谱集与叶片全氮含量的回归模型(Original Spectrum-Partial Least Squares Regression, OS-PLSR)和一阶导数光谱最佳波段集与叶片全氮含量的回归模型(First Derivative Spectrum-Partial Least Squares Regression, FDS-PLSR)。模型精度检验采用舍一交叉验证方法,评价指标为交叉验证决定系数(R_{cv}^2)和交叉验证均方根误差(RMSE_{cv}),并绘制实测值与估计值之间的 1:1 关系图。

3 结果与讨论

3.1 相关性分析

分别对芦苇和香蒲的所有样点的原始光谱及其一阶导数光谱和叶片全氮含量进行相关性分析(图 2)。从原始光谱相关性分析可以看出芦苇和香蒲的相关性系数曲线相似,叶片氮含量与光谱(400—1000 nm)反射率呈负相关,并在 550 nm 和 715 nm 附近出现波谷。第一个波谷的形成是由于叶绿素对于绿光的反射作用,并且在 550 nm 附近达到最大值,第二个波谷位于“红边”区域,有相关研究发现“红边”波段与植物叶绿素含量、生物量等参数间存在显著相关性^[17]。以上结论与 barbara.J.Yoder 等人的研究成果一致。说明在原始光谱中,反映芦苇和香蒲叶片氮含量的敏感光谱主要集中在可见光波段。其中芦苇叶片全氮含量与光谱在 400—980 nm 范围内相关性达到显著水平($P < 0.05$),在 440—840 nm 范围内到极显著水平($P < 0.01$);而香蒲叶片全氮

含量与光谱的相关性在 455—665、685—770 nm 和 910—1000 nm 达到显著水平($P < 0.05$),在 515—605 nm 和 695—730 nm 达到极显著水平($P < 0.01$)。从一阶导数相关性分析,芦苇叶片全氮含量与可见光波段反射率的相关性水平较高,在 405—455、485—545 nm 和 555—725 nm 波段范围内达到显著水平($P < 0.05$),在 405—455、485—535 nm 和 555—675 nm 波段范围内达到极显著水平($P < 0.01$)。香蒲在可见光波段与芦苇的相关性曲线表现出一致的变化趋势,在 425—455、485—545、555—645 nm 和 680—710 nm 波段范围内达到显著($P < 0.05$),在 485—530、555—600 nm 和 680—710 nm 波段范围内达到极显著水平($P < 0.01$),但对比发现,与芦苇叶片全氮含量达到显著和极显著水平的波段范围更广。在近红外波段,芦苇叶片全氮含量与光谱在 820、875、910 nm 和 940 nm 等附近的也出现了相关性较高的峰值,而香蒲在近红外波段范围内不存在与叶片氮含量相关性较好的波段。这些高相关波段为湿地植物芦苇和香蒲全氮含量的估测提供了保障。

3.2 OS-PLSR 模型构建及精度评价

3.2.1 原始叶片氮反演最佳波段选择

为选择基于地物高光谱特征的湿地植物叶片全氮反演的最佳波段,本文尝试采用“舍一交叉法”。具体流程如下:将原始光谱随机划分为 10 份,每次取其中的九份进行与叶氮含量的相关性分析,选出每组中相关系数绝对值最大的 20 个极值点(共 200 个)。之后对入选波段按其相关性绝对值的大小进行排序(1—20 名),并统计每个波段出现的频率、频率最高排名和频率最高排名的次数(表 1)。实验中认为当波段出现的频率小于 5 时(出现频率小于 50%),表示该波段为不稳定波段,应该删除。最终确定芦苇和香蒲各有 21 个,这些入选波段为潜在的湿地植物全氮含量预测的最佳波段。按照表中波段出现的间断,并结合光谱分区特征,对入选波段进行谱带划分。芦苇的入选波段主要分布在红波波段(I)、黄光波段(II)和绿光波段(III),香蒲的入选波段主要分布在红光波段(I)、黄光波段(II)和绿光波段(III)。其中红光波段反映了叶绿素对红光的吸收特征;黄光谱带反映了叶绿素和类胡萝卜素对黄光的反射特征;绿光波段反映了植物叶绿素和类胡萝卜素对绿光的反射特征。为进一步确定叶片氮含量

反演的最佳波段,在各个谱带内选取频率最高、出现频率次数最大且频率排名最高的波段作为该谱带的代表波段。结果如下:芦苇的最佳波段集为 550、580、715 nm,芦苇的最佳波段集为 555、580、715 nm。芦苇和香蒲叶片全氮含量反演的最佳波段集几乎一致,说明 550(555 nm)、580、715 nm 波段是叶氮反演的关键波段,在芦苇和香蒲叶氮反演中具有普适性。此外这些入选波段不仅与叶片全氮含量具有高度相关性,而且可从生化参数的角度解释其与叶片氮含量高度相关的意义,同时在谱区范围划分及谱区范围分区中各参数选择则可以有效地避免同一谱区范

围内“多重共线性”问题对反演模型的影响。从研究结果还可以看出在原始光谱叶片氮含量反演最佳波段判别过程中,不同植物类型对于叶片全氮含量的敏感波段差异不明显,进一步说明入选的波段是基于原始光谱特征的反演植物叶片氮含量的特征波段。其中 550(555 nm)和 715 nm 是能有效反映植物叶片全氮含量的敏感波段已被许多学者证实^[17-18],580 nm 波段由于受到 550 nm 波段的影响,在传统模型中该波段对于叶片全氮含量的估测能力往往不能充分体现。

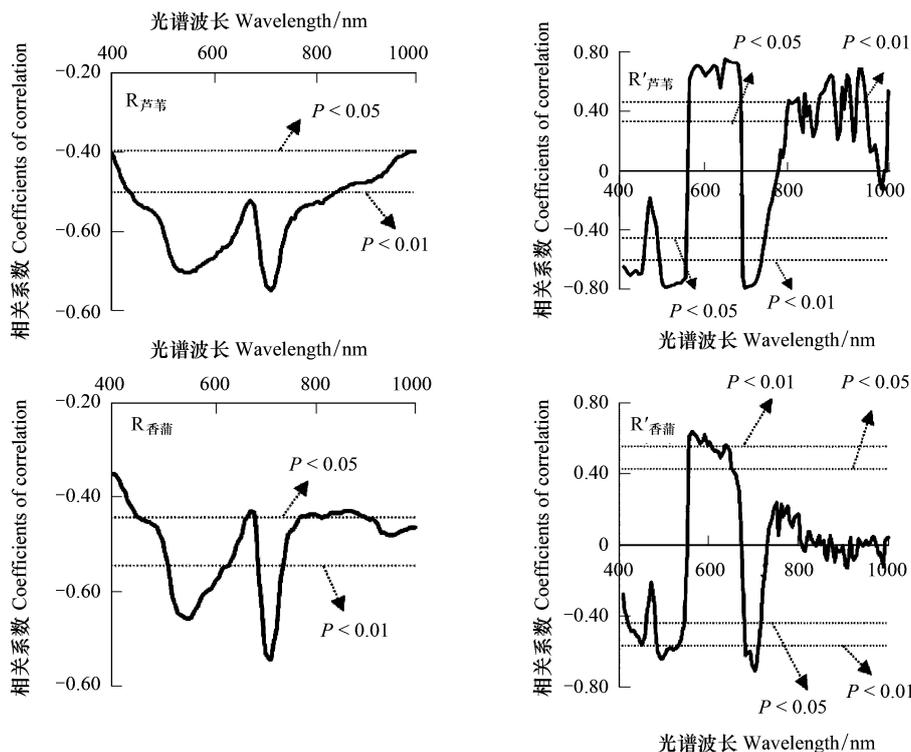


图2 芦苇、香蒲原始光谱和一阶导数光谱与 TN 含量的相关系数图

Fig.2 Coefficients correlation relating the spectrum of original and first-derivation to concentrations of nitrogen

3.2.2 模型及精度验证

经过统计分析选出的最佳波段集代表了各个谱带内与叶片氮含量相关性最好的波段,且相关性绝对值均大于 0.6,均达到极显著水平。说明这些光谱特征变量与叶片尺度的全氮含量具有较高的关联性。因此利用偏最小二乘法构建叶片全氮含量和最佳波段集的回归模型,构建的回归模型如表 3。根据交叉有效性原理,得到芦苇和香蒲所提取的成分维数分别为 1 维和 2 维。建立最佳估算模型后,为了检验估算模型的可靠性和实用性,采用了交叉验证

决定系数 (R_{cv}^2) 和均方根误差 ($RMSE_{cv}$) 2 个指标对估算模型进行评定,并绘制了实测值与预测值之间的 1:1 关系图,以直观的展示估算模型的拟合度和可靠性。验证的结果如图 3 所示,在构建的回归模型中芦苇模型的 R_{cv}^2 为 0.58, $RMSE_{cv}$ 为 0.16,香蒲模型的 R_{cv}^2 为 0.56, $RMSE_{cv}$ 为 0.16。在原始光谱构建的回归模型中,芦苇模型精度略高于香蒲,但差异不大。虽然该模型在一定程度上能够反映芦苇和香蒲叶片的全氮含量,但是精度较低,稳定性有待提高。

表 1 湿地植物原始光谱潜在最佳波段统计表

Table 1 Potential optimal bands of original spectrum from wetland vegetation

芦苇 <i>Phragmites australis</i>							
谱带划分 Spectrum division	入选波段 Selected band	频率 Frequency	频率最高排名 Magnitude tier	频率最高 排名次数 Highest frequency number	谱带分区 Spectral zone	最佳波段 Optimum band	
红光波段 Red spectrum	730	5	6	1	I	715	
	725	9	4	1	I		
	720	10	2	1	I		
	715	10	1	7	I		
	710	10	1	3	I		
	705	10	3	4	I		
	700	10	4	1	I		
黄光波段 Yellow spectrum	695	8	6	1	I	580	
	585	5	15	1	II		
	580	9	14	4	II		
	绿光波段 Green spectrum	575	10	13	2		III
		570	10	12	2		III
		565	10	10	2		III
		560	10	8	2		III
		555	10	6	2		III
		550	10	6	3		III
		545	10	8	1		III
		540	10	9	1		III
		535	10	11	2		III
		530	10	10	1		III
525	10	6	1	III			
香蒲 <i>Typha angustifolia</i>							
谱带划分 Spectrum division	入选波段 Selected band	频率 Frequency	频率最高排名 Magnitude tier	频率最高 排名次数 Highest frequency number	谱带分区 Spectral zone	最佳波段 Optimum band	
红光波段 Red spectrum	730	7	7	1	I	715	
	725	8	4	1	I		
	720	10	2	1	I		
	715	10	1	6	I		
	710	10	1	3	I		
	705	10	3	4	I		
	700	10	1	1	I		
黄光波段 Yellow spectrum	695	9	7	1	I	580	
	585	5	16	1	II		
	580	7	15	2	II		
	绿光波段 Green spectrum	575	7	14	4		III
		570	9	12	1		III
		565	10	10	2		III
		560	10	6	1		III
		555	10	6	3		III
		550	10	6	1		III
		545	10	9	7		III
		540	10	11	7		III
		535	10	11	1		III
		530	9	10	1		III
525	9	16	3	III			

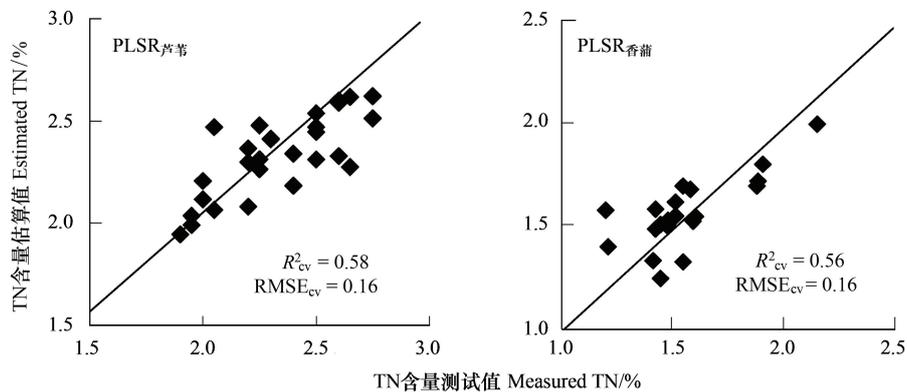


图3 原始光谱的偏最小二乘回归模型的交叉验证实测值和估计值的1:1关系图

Fig.3 The 1:1 relationship between the estimated value and measured value in cross validation of spectral OS-PLSR model

3.3 FDS-PLSR 模型构建及精度评价

3.3.1 一阶导数叶片全氮反演最佳波段选择

相同的方法应用于一阶导数光谱的最佳波段选取。产生的极值点(共200个)所对应的波段即为叶片全氮含量反演的潜在敏感波段。对入选波段按其相关性绝对值大小进行排序(1—20名),并统计每个波段出现的频率、频率最高排名和频率最高排名的次数(表2)。实验中认为当波段出现的频率小于5(频率<50%)时,表示该波段为不稳定波段,应该删除。统计发现芦苇有21个波段入选,香蒲有23个波段入选。按照表中波段出现的间断,并结合光谱分区特征,对入选波段进行谱带划分。芦苇的入选波段主要分布在近红外波段(I)、红光波段(II、III)、绿光波段(IV)和蓝光波段(V),香蒲的入选波段主要分布在红光波段(I)、黄光波段(II)、绿光波段(III、IV)和蓝光波段(V)。其中蓝光波段,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,黄光光谱反映了叶绿素和类胡萝卜素对黄光的反射特征,近红外波段(760—1000 nm)反映了植物叶子的多孔薄壁组织(海绵组织)对近红外波段的强烈反射。虽然尚未发现该波段与叶片全氮含量的直接相关的生化意义,但氮素作为植物生长、发育所必须的元素,氮含量的多少影响植物叶子的多孔薄壁组织(海绵组织)结构^[14],进而影响光谱反射率。为进一步确定湿地植物叶片氮含量反演的最佳波段,在每个谱带分区中各选取一个出现频率最高,出现频率次数最大和频率最高排名次数最多的波段作为该分区中的代表波段即最佳波段。其中芦苇的最佳波段集为490、500、635、680 nm和940 nm,香蒲的最佳波段集为490、495、565、590 nm和700 nm。

与原始光谱最佳波段判别结果相比较,一阶导数光谱判别中获得的谱区范围更广。芦苇与香蒲的最佳谱带中出现了红外光和蓝光等原始光谱中不存在的谱带,这些谱带作为原始光谱的斜率对叶片全氮含量的敏感性被充分挖掘。不同植物类型相比较,芦苇和香蒲二者不仅在谱区范围的划分上存在较大的差异,而且在入选波段上有所变异。在可见光范围内,相同入选波段(490、495 nm/500 nm)是经过一阶导数变换后植物叶片全氮含量反演的稳定特征波段,这些关键波段对于芦苇和香蒲叶片全氮含量反演研究具有普遍适用性。不同的入选波段则是通过一阶导数转换有效的从光谱角度反映出不同植物类型的生理生化参量的差异与组织结构的关系的反应。

3.3.2 模型及精度验证

经过统计分析选出的最佳波段集代表了各个谱带内与叶片氮含量相关性最好的波段,且相关性均大于0.65,说明这些光谱特征变量与叶片尺度的全氮含量具有较高的关联性。因此利用偏最小二乘法构建叶片全氮含量和最佳波段集的回归模型,构建的回归模型如表3。根据交叉有效性原理,得到芦苇和香蒲所提取的成分维数分别为2维和3维。建立最佳估算模型后,为了检验估算模型的可靠性和实用性,采用了交叉验证决定系数(R_{cv}^2)和均方根误差($RMSE_{cv}$)2个指标对估算模型进行评定,并绘制了实测值与预测值之间的1:1关系图,以直观的展示估算模型的拟合度和可靠性。验证的结果如图4所示,芦苇的回归模型的 R_{cv}^2 为0.82, $RMSE_{cv}$ 为0.11。香蒲的回归模型的 R_{cv}^2 为0.66, $RMSE_{cv}$ 为0.13。相对

于 OS-PLSR 模型, R_{cv}^2 分别增加了 0.11 和 0.05, RMSE_{cv} 分别减少了 0.03 和 0.02, 精度有了很大的提高。此外, 研究结果表明一阶导数模型中, 仍然是芦苇所建立的模型精度较高。

表 2 湿地植物一阶导数光谱潜在最佳波段统计表
Table 2 Potential optimal bands of first derivative spectrum from wetland vegetation

芦苇 <i>Phragmites australis</i>						
谱带划分 Spectrum division	入选波段 Selected band	频率 Frequency	频率最高排名 Magnitude tier	频率最高 排名次数 Highest frequency number	谱带分区 Spectral zone	最佳波段 Optimum band
近红外光谱 Near infrared	940	7	1	1	I	940
红光光谱 Red spectrum	705	5	12	1	II	680
	700	10	9	3	II	
	695	10	4	1	II	
	690	10	5	3	II	
	685	10	3	4	II	
	680	10	1	9	II	
	645	6	18	3	III	635
	640	10	13	1	III	
	635	10	12	1	III	
绿光光谱 Green spectrum	535	10	11	1	IV	500
	530	10	13	3	IV	
	525	10	9	1	IV	
	520	10	8	1	IV	
	515	10	6	2	IV	
	510	10	4	2	IV	
	505	10	3	5	IV	
	500	10	2	9	IV	
	495	10	3	1	IV	
蓝光光谱 Blue spectrum	490	8	10	1	V	490
	485	2	16	1	V	
香蒲 <i>Typha angustifolia</i>						
谱带划分 Spectrum division	入选波段 Selected band	频率 Frequency	频率最高排名 Magnitude tier	频率最高 排名次数 Highest frequency number	谱带分区 Spectral zone	最佳波段 Optimum band
红光光谱 Red spectrum	705	10	2	2	I	700
	700	10	1	8	I	
	695	10	2	4	I	
	690	7	6	1	I	
	685	10	3	1	I	
	680	7	2	1	I	
黄光光谱 Yellow spectrum	595	10	5	1	II	590
	590	8	12	1	II	
绿光光谱 Green spectrum	575	7	13	2	III	565
	570	10	6	1	III	
	565	10	4	1	III	
	560	9	2	1	III	
	555	6	5	1	III	
	530	6	11	1	IV	495
	525	4	12	1	IV	
	520	5	14	1	IV	
	515	4	16	1	IV	
	510	3	18	3	IV	
	505	8	5	1	IV	
	500	9	7	1	IV	
	495	10	4	3	IV	
蓝光波段 Blue spectrum	490	10	3	2	V	490
	485	7	2	1	V	

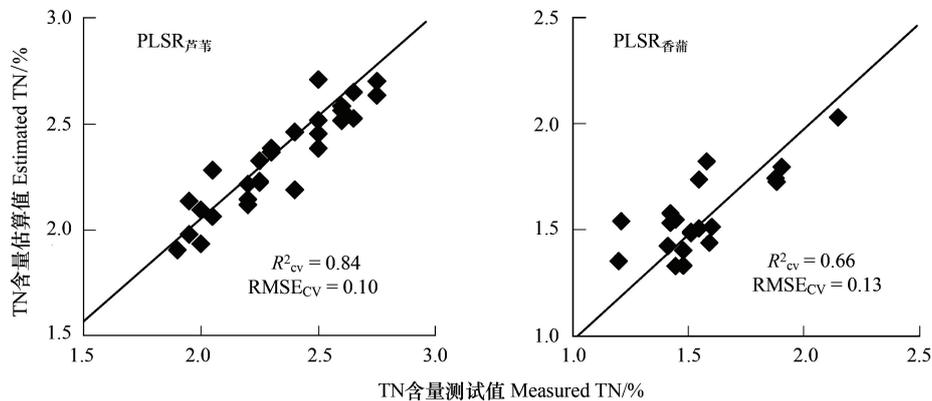


图4 一阶导数的偏最小二乘回归模型的交叉验证实测值和估计值的1:1关系图

Fig.4 The 1:1 relationship between the estimated value and measured value in cross validation of spectral FDS-PLSR model

表3 叶片全氮含量与最佳波段集构建的回归模型

Table 3 Regression model between nitrogen content and the best band set

湿地植物类型 Species	模型类型 Model types	回归方程 Regression equation
芦苇	OS-PLSR 模型	$Y = 0.11R_{550} + 1.85R_{580} - 1.71R_{715} + 0.60$
<i>Phragmites australis</i>	FDS-PLSR 模型	$Y = -76.06R_{490} - 52.21R_{500} + 86.28R_{635} - 50.72R_{680} + 126.81R_{940} + 0.59$
香蒲	OS-PLSR 模型	$Y = -0.10R_{555} + 0.16R_{580} + 0.57R_{715} + 0.51$
<i>Typha angustifolia</i>	FDS-PLSR 模型	$Y = -354.74R_{490} - 164.50R_{495} - 230.31R_{565} + 489.95R_{590} + 35.15R_{700} + 0.49$

4 结论与讨论

为了探索植物生化参量遥感诊断的可能性,20世纪70年代以来有关科学家就进行了大量的基础研究,探索植物生化参量反演的最佳反演模型。然而,研究发现这些回归模型存在入选波段的稳定性差、波段生化参量含义不明确、共线性和回归模型的过拟合等问题。虽然这些问题被许多学者反复指出并做了大量解决方法的尝试^[11,13,19],但并未取得较为理想和统一的效果。本文结合了“舍一交叉相关分析”、谱带划分、统计分析和偏最小二乘回归等多种方法,构建了湿地植物芦苇和香蒲叶片全氮含量的回归模型。其中,“舍一交叉相关分析”处理能够有效剔除模型中的“奇异波段”,降低了参量间的“多重共线性”,谱段范围的划分解释了入选波段与生化参量间的因果关系,偏最小二乘法进一步消除了入选波段间“多重共线性”。此外,由于构建的模型是多参量方程,在一定程度上避免了指数模型易受环境等其他干扰因素影响的缺点。研究结果表明上述处理方法对于构建的叶片全氮含量的回归模型具有较好精度。理论上,该方法也可应用于其他生物参量如叶绿素、水分含量的估测。本研究的方法

和模型针对其他生理生化参量反演的适用性有待今后进一步研究。

本文分别用原始光谱和一阶导数光谱进行了叶片全氮含量回归模型的构建。结果表明,FDS-PLSR模型比OS-PLSR模型具有更好的稳定性和精度。有研究表明经过一阶导数转换能有效提高光谱与生物参量的相关性^[20],即原始光谱的斜率更能高效的反映出生物参量与光谱间的关系。此外,本课题组前期已有学者进行了芦苇和香蒲叶片全氮含量估测最优模型的构建^[21],与其结果相比较,芦苇和香蒲的FDS-PLSR模型精度都高于其所构建的最优模型PLSR。虽然,FDS-PLSR模型是湿地植物叶片全氮含量估测的最优模型,尤其对于芦苇叶片全氮的反演具有较高的精度,大量文献研究中发现不同的信息提取方法对于构建的模型精度有很大的影响。目前,已有许多学者尝试从光谱转换,提取隐藏信息这一角度进行回归模型的构建^[22-23]。不同的方法在不同类型的生理生化参量反演中发挥了不同的作用。应用其他的转化方法取代一阶导数转化进行回归模型的构建也是一个值得深入探讨的问题。

从湿地植物类型看,模型精度在不同湿地植物类型之间存在差异,利用芦苇光谱的模型精度均高

于香蒲。这与该组早期的研究结果一致^[20]。有文献表明芦苇对氮的吸收能力高于香蒲^[24],说明芦苇更能反映所处的环境特征,这可能是其回归模型精度较高的原因。本文最终确定了芦苇叶片全氮含量反演的最佳波段组合集为 490、500、635、680 nm 和 940 nm,香蒲叶片全氮含量反演的最佳波段组合集为 490、495、565、590 nm 和 700 nm。在反演模型中,芦苇和香蒲相同的人选波段为 490 nm 和 495/500 nm,这些关键波段对于芦苇和香蒲叶片全氮含量反演研究具有普遍适用性。而不同的人选波段则反映出不同植物类型的生理生化参数的差异与组织结构对于光谱反射值得影响。此外,研究发现芦苇在近红外范围 940 nm 出现了最佳波段而香蒲在该区域不存在最佳波段。研究表明在 922—990 nm 植物光谱会受到植物水分差异的影响^[14],芦苇和香蒲水分的差异和组织结构的不同可能是造成这种现象的原因,其深层次的原因还有待进一步研究。本文主要对湿地植物芦苇和香蒲全氮含量估测模型构建的可能性进行了探讨,在不影响回归精度的前提下,有效避免了传统生物参量光谱反演中存在的一些问题。研究成果可以为湿地植物全氮含量的大面积监测提供有力的科学依据。

致谢: 林川、张翼然、段光耀、阿多、杨典华、袁德阳、尹川、朱先芳、熊薇参与了大量野外采样和室内化验分析工作,特此致谢。

References:

- [1] Goetz S J, Prince S D. Remote sensing of net primary production in boreal forest stands. *Agricultural and forest Meteorology*, 1996, 78(3/4): 149-179.
- [2] Wang J H, Zhao C J, Huang W J. *Fundamentals and Applications of Quantitative Remote Sensing in Agriculture*. Beijing: Science Press, 2008: 4-5.
- [3] Lin C, Gong Z N, Zhao W J. Hyperspectral estimation models for plant community water content at both leaf and canopy levels in Wild Duck Lake wetland. *Acta Ecologica Sinica*, 2011, 31(22): 6645-6658.
- [4] Dawson T P, Curran P J, North P R J, Plummer S E. The propagation of foliar biochemical absorption features in forest canopy reflectance: a theoretical analysis. *Remote Sensing of Environment*, 1999, 67(2): 147-159.
- [5] Yoder B J, Pettigrew-Crosby R E. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400—2500nm) at leaf and canopy scales. *Remote Sensing of Environment*, 1995, 53(22): 199-211.
- [6] Gitelson A A, Merzlyak M N. Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, 1997, 18(12): 2691-2697.
- [7] Zhu Z Y, Bao Y D, Huang M, Feng L. Study of the relationship between the chlorophyll and the nitrogen content of oilseed rapeseeds. *Journal of Zhejiang University: Agriculture and Life Sciences*, 2006, 32(2): 152-154.
- [8] Xue L H, Luo W H, Cao W X, Tian Y C. Research progress on the water and nitrogen detection using spectral reflectance. *Journal of Remote Sensing*, 2003, 7(1): 73-80.
- [9] Curran P J, Dungan J L, Peterson D L. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the kokaly and clark methodologies. *Remote Sensing of Environment*, 2001, 76(3): 349-359.
- [10] Wang Y Y, Li G C, Zhang L J, Fan J L. Retrieval of leaf water content of winter wheat from canopy hyperspectral data using partial least square regression. *Spectroscopy and Spectral Analysis*, 2010, 30(4): 1070-1074.
- [11] Bortolot Z J, Wynne R H. A method for predicting fresh green leaf nitrogen concentrations from shortwave infrared reflectance spectra acquired at the canopy level that requires no in situ nitrogen data. *International Journal of Remote Sensing*, 2003, 24(3): 619-624.
- [12] Yan C Y, Niu Z, Wang J H, Liu L Y, Huang W J. The assessment of spectral indices applied in chlorophyll content retrieval and a modified crop canopy chlorophyll content retrieval model. *Journal of Remote Sensing*, 2005, 9(6): 742-750.
- [13] Kokaly R F, Clark R N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, 1999, 67(3): 267-287.
- [14] Liu K, Zhao W J, Guo X Y, Hu D Y, Gong Z N, Long J. Spectral bands of typical wetland vegetation in the Wild Duck Lake. *Acta Ecologica Sinica*, 2010, 30(21): 5853-5861.
- [15] Deng Z Z. *On Water System and Rainwater Harvesting System in Beijing Olympic Green*. Beijing: China Water Power Press, 2009: 54-88.
- [16] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [17] Li B Z, Li M X, Zhou X, Zhang L S, Zhang H Y. Hyperspectral estimation models for nitrogen contents of apple leaves. *Journal of Remote Sensing*, 2010, 14(4): 767-780.
- [18] Song S L, Li P X, Gong W, Wang J P. PLS wavelength selection by hyperspectral remote sensing in rice. *Geomatics and Information Science of Wuhan University*, 2010, 35(2): 219-223.
- [19] Pasternak H, Edan Y, Schmilovitch Z. Overcoming multicollinearity by deducting errors from the dependent variable.

- Journal of Quantitative Spectroscopy & Radiative Transfer, 2001, 69(6): 761-768.
- [20] Han L H. Estimating chlorophyll-a concentration using first-derivative spectra in coastal water. *International Journal of Remote Sensing*, 2005, 26(23): 5235-5244.
- [21] Liu K, Zhao W J, Guo X Y, Wang Y H, Miao Q. Estimating total nitrogen content in wetland vegetation based on measured reflectance spectra. *Spectroscopy and Spectral Analysis*, 2012, 32(2): 465-471.
- [22] Blackburn G A, Ferwerda J G. Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment*, 2008, 112(4): 1614-1632.
- [23] Malenovsky Z, Homolová L, Zurita-Millad R, Lukeš P, Kaplane V, Hanuš J, Gastellu-Etchegorry J P, Schaepman M E. Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. *Remote Sensing of Environment*, 2013, 131: 85-102.
- [24] Chen Y H, Wu X F, Jiang L J, Chen M L, Zeng M, Lei D, Zhang Z N, Yang S Y. Screening and evaluation of plant purification potential for phytoremediation of sanitary sewage. *Acta Scientiae Circumstantiae*, 2008, 28(8): 1549-1554.
- 的高光谱估算模型. *生态学报*, 2011, 31(22): 6645-6658.
- [7] 朱哲燕, 鲍一丹, 黄敏, 冯雷. 油菜叶绿素与氮含量关系的试验研究. *浙江大学学报: 农业与生命科学版*, 2006, 32(2): 152-154.
- [8] 薛利红, 罗卫红, 曹卫星, 田永超. 作物水分和氮素光谱诊断研究进展. *遥感学报*, 2003, 7(1): 73-80.
- [10] 王圆圆, 李贵才, 张立军, 范锦龙. 利用偏最小二乘回归从冬小麦冠层光谱提取叶片含水量. *光谱学与光谱分析*, 2010, 30(4): 1070-1074.
- [12] 颜春燕, 牛铮, 王纪华, 刘良云, 黄文江. 光谱指数用于叶绿素含量提取的评价及一种改进的农作物冠层叶绿素含量提取模型. *遥感学报*, 2005, 9(6): 742-750.
- [14] 刘克, 赵文吉, 郭道宇, 胡德勇, 宫兆宁, 龙娟. 野鸭湖典型湿地植物光谱特征. *生态学报*, 2010, 30(21): 5853-5861.
- [17] 李丙智, 李敏夏, 周璇, 张林森, 张海燕. 苹果树叶片全氮含量高光谱估算模型研究. *遥感学报*, 2010, 14(4): 767-780.
- [18] 宋沙磊, 李平湘, 龚威, 汪金平. 基于水稻高光谱遥感数据的 PLS 波长选择研究. *武汉大学学报: 信息科学版*, 2010, 35(2): 219-223.
- [21] 刘克, 赵文吉, 郭道宇, 王翊虹, 孙永华, 苗茜. 基于地面实测光谱的湿地植物全氮含量估算研究. *光谱学与光谱分析*, 2012, 32(2): 465-471.
- [24] 陈永华, 吴晓英, 蒋丽鹃, 陈明利, 曾敏, 雷电, 张珍妮, 阳石英. 处理生活污水湿地植物的筛选与净化潜力评价. *环境科学学报*, 2008, 28(8): 1549-1554.

参考文献:

- [2] 王纪华, 赵春江, 黄文江. 农业定量遥感基础与应用. 北京: 科学出版社, 2008: 4-5.
- [3] 林川, 宫兆宁, 赵文吉. 叶冠尺度野鸭湖湿地植物群落含水量