

玉米叶绿素含量高光谱反演的线性模型研究

孙阳阳 汪国平 杨可明* 卓伟 张婉婉

(中国矿业大学(北京)地球科学与测绘工程学院 北京 100083)

摘要: 叶绿素含量是衡量植被生长状况的一个重要指标,高光谱数据具有较高的光谱分辨率,利用其光谱信息建立叶绿素含量的关系模型,已成为监测植被长势的一种有效手段。传统叶绿素含量线性回归模型的输入因子是植被特征提取参数,由于高光谱数据波段间的冗余度较高,导致一般的线性模型的反演精度较低。主成分分析可以减少数据的维数,简化网络结构,得出能反映原始信息的综合变量。本文以盆栽玉米为研究对象,利用植被特征和主成分分析方法提取光谱反演参数,根据所提取的参数建立玉米叶片叶绿素含量的一元线性和多元线性回归模型。结果表明,利用绿峰峰值和近红外反射率均值两参数可在一元线性模型中较好地反演玉米叶片叶绿素含量;而利用分波段提取的主成分能够在多元线性回归模型中更好地反演叶绿素含量,反演精度较高。

关键词: 玉米; 叶绿素含量; 光谱参数; 主成分分析; 线性回归模型

中图分类号: S513.01 **文献标识码:** A **文章编号:** 1001-4942(2015)07-0117-05

Research on Linear Model of High Spectral Retrieval of Chlorophyll Content in Corn

Sun Yangyang, Wang Guoping, Yang Keming*, Zhuo Wei, Zhang Wanwan

(College of Geoscience and Surveying Engineering, China University of Mining & Technology(Beijing), Beijing 100083, China)

Abstract The chlorophyll content is an important indicator to measure the vegetation growth status, and the hyperspectral data have fine spectral resolution, so to establish the chlorophyll content model by using the spectral information has become an effective means for monitoring vegetation growth. The input factors of the conventional linear regression model for chlorophyll content were the feature extraction parameters of vegetation, so high redundancy between the bands of hyperspectral data resulted in lower accuracy of the general linear model. The principal component analysis (PCA) can effectively reduce the data dimensions and simplify the network structure to obtain comprehensive variables reflecting the original information. Using the potted corn as research object, the linear and multiple linear regression models of corn leaf chlorophyll content were established with the spectral parameters extracted by vegetation characteristics and principal component method. The results showed that the green peak value and the mean of near-infrared reflectivity were better for the inversion of chlorophyll content in corn leaves with linear mode, while the principal components extracted from different brands were better for multiple linear regression model.

Key words Corn; Chlorophyll content; Spectral parameters; Principal component analysis; Linear regression model

叶绿素通过光合作用从光中吸收能量,合成一些有机化合物,从而将光能转变为化学能并为

植物提供养料。植被叶绿素含量与其光合能力、生长环境、发育阶段以及营养状况有较大的相关

收稿日期: 2015-05-08; 修回日期: 2015-06-15

基金项目: 国家自然科学基金项目(41271436); 中央高校基本科研业务费专项资金(2009QD02)

作者简介: 孙阳阳(1991-), 山东单县人, 硕士, 主要从事高光谱遥感、矿山形变控制等研究。E-mail: yaogansunyangyang@163.com

* 通讯作者: 杨可明(1969-), 男, 博士, 教授, 主要从事高光谱与多光谱遥感、矿山形变控制与矿山地理信息等研究。E-mail: ykm69@163.com

性,是衡量植被发育情况的显著性标志。由于植被在可见光范围内反射光谱主要受植被色素的影响,因此可以利用植被的反射光谱来估算植被的色素含量,进而预测植被的生长状态。高光谱遥感技术具有光谱分辨率高、连续性强、蕴含丰富的光谱信息等优势,为监测植被长势及生理状态提供了强有力的技术支撑。

目前众多学者通过高光谱遥感技术,对植被叶绿素含量与光谱信息的相关性进行实验和分析,提取光谱特征参数进行叶绿素含量的估测和反演,取得了一定的成果。姚付启等^[2]利用植被的红边参数进行叶绿素含量反演,得出利用神经网络反演的结果精度较高,是一种良好的高光谱反演模式;陈婉婧等^[3]提取与叶绿素含量相关性较高的敏感波段和特征光谱参数,得到一个基于神经网络算法的叶绿素含量最优模型;陈君颖等^[4]通过研究不同水稻叶片光谱特性与叶绿素含量之间的关系,利用多元逐步回归分析方法建立了水稻叶片叶绿素含量的反演模型;宋开山等^[5]采用叶绿素敏感波段和相关性较高的波段对植被叶绿素含量进行估算,得出基于神经网络的反演模型能提高预测模型的精度。本文以实验室盆栽玉米为研究对象,利用光谱仪测量玉米叶片光谱数据,提取与叶绿素含量相关性较高的光谱特征参数:蓝谷和红谷深度、绿峰峰值、红边参数、植被指数和近红外平均反射率,利用主成分分析提取叶片光谱的综合信息,并用回归模型对提取的光谱参数进行反演,得到预测叶绿素含量的最佳多元线性回归模型。

1 数据处理与方法

1.1 实验数据的获得

实验对象为“中糯 1 号”玉米,培养期为 2014 年 5 月 6 日~7 月 17 日,培养结束对玉米植株的不同部位叶片分别取样,测量样品的叶片光谱反射率和叶绿素含量。

采样日期为 2014 年 7 月 17 日,用 SVC HR-1024I 高性能地物光谱仪(光谱范围 350~2 500 nm)、50 W 的卤素灯光源和 4°视场角的探头进行叶片光谱数据采集,探头垂直于叶片表面 40 cm,每个叶片样本测 3 次数据,并拍照记录,最后输出的光谱线为 3 条原始扫描光谱自动平均所得。光

谱反射系数经专用平面白板标准化。在光谱数据采集后,用 SPAD-502 叶绿素仪测定玉米叶片不同位置的叶绿素含量值,每个样本重复测 5 次,求平均得到叶片的叶绿素含量值。

1.2 基于植被特征提取光谱参数

在可见光波段(380~760 nm)内,植被叶绿素对光谱具有较强的吸收和反射作用,在蓝光波段(430~500 nm)和红光波段(620~760 nm)附近,叶绿素对光谱的吸收较强,分别形成蓝吸收谷和红吸收谷;而在绿光波段(500~560 nm)内,叶绿素对光谱反射较强,从而在蓝谷和红谷之间形成绿反射峰;在红光波段(620~760 nm),因叶绿素对光谱强烈的吸收作用和近红外波段光谱的多次散射致使反射率急剧上升形成爬升脊的红边^[6],因此红边参数与植被的叶绿素有着较强的相关性,可以利用红边参数估算叶绿素含量。在近红外波段(760~2 500 nm),因叶片内细胞壁间的多重反射形成近红外反射坪,从而使近红外波段对叶片光谱反射率的变化相对敏感,可通过计算近红外反射率的均值来评定植被的生长状况^[7]。归一化植被指数 NDVI 是通过近红外波段和红光波段计算得到并能够反映植被生长状况的特征指数^[8-9],它对植被叶片特征的变化非常灵敏,可用于监测植被的长势状况,计算公式为:

$$NDVI = (R_{760} - R_{700}) / (R_{760} + R_{700}) \quad (1)$$

式中 R_{760} 和 R_{700} 为实验室测量的玉米叶片在 760 nm 和 700 nm 的光谱反射率值。

使用光谱微分方法提取红边位置,即通过对红光波段反射率计算一阶导数,寻找其最大值来确定红边位置^[10-12]。导数光谱计算表达式为:

$$A_i = \frac{\ln(R_{i+1}) - \ln(R_i)}{\lambda_{i+1} - \lambda_i} \quad (2)$$

式中 A_i 、 R_i 和 R_{i+1} 、 λ_i 和 λ_{i+1} 分别为测点 i 的光谱导数值、光谱反射率值和中心波长。

连续统去除法是一种常用的光谱分析方法,该方法能有效去除背景吸收的影响,突出分析叶片光谱的吸收和反射特征^[13,14]。在 ENVI 中使用连续统去除法对光谱数据进行处理,提取蓝谷和红谷位置,经连续统去除后,峰值点上的值为 1,非峰值点上的值均小于 1^[15,16],使光谱特征归一化到一致的背景上,有利于与其他光谱曲线特征

进行数值比较,从而提取光谱特征位置及其参数供分析使用。波段深度计算公式:

$$BD = 1 - R'(\lambda_i) \quad (3)$$

式中, λ_i 为波长位置, $R'(\lambda_i)$ 为对应的连续统去除反射率。

提取的各光谱特征参数值见表 1。

表 1 光谱特征参数

样本	蓝谷	绿峰	红谷	红边	NDVI	近红外均值
1	0.5730	0.1745	0.7412	0.2911	0.8577	0.5017
2	0.4766	0.2322	0.6951	0.2899	0.7808	0.3827
3	0.4745	0.2118	0.7179	0.2841	0.7423	0.4183
...
43	0.5658	0.1695	0.7776	0.2658	0.7731	0.4847
44	0.4691	0.2608	0.7010	0.3037	0.9346	0.3579
45	0.5642	0.1692	0.7428	0.2607	0.7538	0.4773

1.3 主成分分析提取光谱反演参数

主成分分析是一种常见的多元统计分析方法,在数据特征提取、压缩、降维等方面有广泛的应用。其基本思想是通过分析将原来众多的具有相关性的指标,重新组合成一组较少的互不相关的综合指标,既可以反映出原始信息量又减少信息的重叠,简化了运算的复杂度。高光谱数据的波段数据丰富,但波段间的冗余度较高,因此可以通过主成分变换,减少数据的维数,得出能反映原始信息的综合变量。

利用主成分分析法对获得的光谱数据进行两种不同方式的处理。第一种是在光谱的蓝光波段、绿光波段、红光波段和近红外波段提取主成分,每个波段的主成分都需要**含原始光谱 95% 以上的信息**,提取的主成分见表 2,可以看出红光和近红外需要分别提取两个主成分才能符合要求;第二种是对原始光谱进行分析,提取包含原始信息 95% 以上的主成分,提取的主成分见表 3。

表 2 主成分变换 1 特征值及方差贡献率

主成分变量	特征值	方差贡献率 (%)	累计方差贡献率 (%)
蓝光(PCB1)	47.56	99.08	99.08
绿光(PCG1)	39.42	98.55	98.55
红光 1(PCR1)	92.39	89.70	89.70
红光 2(PCR2)	5.76	5.59	95.29
近红外 1(PCRI1)	571.32	83.53	83.53
近红外 2(PCRI2)	90.70	13.26	96.79

表 3 主成分变换 2 特征值及方差贡献率

主成分变量	特征值	方差贡献率 (%)	累计方差贡献率 (%)
PC1	764.10	78.53	78.53
PC2	100.88	10.37	88.90
PC3	60.28	6.20	95.09
PC4	24.10	2.48	97.57
PC5	7.67	0.79	98.36
PC6	4.42	0.45	98.81

2 特征参数与叶绿素值的相关性分析与一元线性模型

2.1 相关性分析

把提取的光谱参数与叶绿素含量进行相关性分析(见表 4),得出基于植被特征提取的参数与叶绿素值具有较强的相关性,除红边参数外,其他参数的相关性绝对值均达到 0.4 以上,其中,蓝谷、绿峰、NDVI 和近红外反射率均值均达到 0.5 以上,说明这些特征参数对叶绿素含量的敏感度较好。利用主成分分析提取的特征参数,分波段提取的主成分相关性整体较高,说明每个波段提取的主成分信息与植被的叶绿素含量有一定的相关性;而 350 ~ 2 500 nm 整体提取的主成分与叶绿素含量的相关性整体偏低,不能较好地反演叶片的叶绿素含量。可见,前两组特征参数都能较好地反映出与叶绿素含量的相关性,具有一定的代表性,对建立叶绿素含量的线性回归模型具有一定的帮助。

表 4 特征参数与 SPAD 值的相关性

参数类型	1	2	3	4	5	6
I	0.52	-0.74	0.44	-0.21	-0.52	0.82
II	0.39	-0.71	-0.55	-0.60	-0.25	0.18
III	0.34	-0.06	-0.72	-0.33	-0.13	-0.06

注:参数类型 I、II、III 分别代表植被特征提取参数、主成分分析 I 提取参数和主成分分析 2 提取参数。参数 I 中的 1~6 分别指蓝谷、绿峰、红谷、红边参数、归一化植被指数和近红外反射率均值;参数 II、III 中 1~6 分别指主成分分析提取的 6 个主成分。

2.2 一元线性反演模型

运用回归分析法建立光谱特征参数与叶绿素含量之间的数学关系模型,依据表 4 中的相关性,选取光谱特征参数中与叶绿素含量相关性较大的绿峰峰值(R_{gp})和近红外平均反射率($R_{760-2500}$),

主成分分析1中相关性较大的PCG1、PCR2主成分,主成分分析2中的PC1和PC3主成分,分别进行一元线性回归分析,并以模型的的决定性系数(R^2)和均方根误差(RMSE)作为评定模型精度的指标。从表5结果中看出,基于植被特征参数的一元线性回归模型精度较好,而基于主成分的模型整体表现较差,不适宜用一元线性模型反演。所以在一元线性回归模型中应优先考虑用近红外反射率均值($R_{760-2500}$)和绿峰峰值(R_{gp})反演叶绿素含量。

表5 不同光谱参数的一元线性模型及精度

参数类型	拟合曲线方程	R^2	RMSE
R_{gp}	$y = -99.998x_{R_{gp}} + 53.819$	0.53	3.55
$R_{760-2500}$	$y = 68.66x_{R_{760-2500}} + 3.956$	0.66	2.97
PCG1	$y = -0.606x_{PCG1} + 34.249$	0.50	3.69
PCR2	$y = -1.302x_{PCR2} + 34.249$	0.33	4.21
PC1	$y = 0.067x_{PC1} + 34.249$	0.12	4.93
PC3	$y = 0.490x_{PC3} + 34.249$	0.50	3.62

注:表中的一元线性回归方程式均在置信水平 $\alpha = 0.01$ 下显著。

3 多元线性回归分析

3.1 多元线性回归模型的原理

玉米叶片叶绿素含量的变化往往受几个重要参数的影响,此时就需要用两个或两个以上的影响因素作为自变量来解释因变量的变化。多元线性回归模型能够将多个变量综合起来解释因变量,所以本文使用多元线性回归分析的方法建立反演模型。

建立模型的步骤:

(1) 多元线性模型的基本结构形式为:

$$y = a_1 x_{1i} + a_2 x_{2i} + \cdots + a_k x_{ki} + a_0 + \varepsilon_i \quad (4)$$

式中 $a_0, a_1, a_2, \cdots, a_k$ 为待定参数; ε_i 为随机变量;若用 $b_0, b_1, b_2, \cdots, b_k$ 表示 $a_0, a_1, a_2, \cdots, a_k$ 的拟合值,则有回归方程为

$$\hat{y} = b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + b_0 \quad (5)$$

用最小二乘法原理可求得式中参数,其中 b_0 为常数, b_1, b_2, \cdots, b_k 为偏回归系数。

(2) 模型的显著性检验:因变量 y 的观测值 y_1, y_2, \cdots, y_n 之间的波动差异,是由两个因素引起的,一是自变量 x_1, x_2, \cdots, x_n 的取值不同,二是其他随机因素的影响。为了从 y 的总变差中分离出

两个变量,就需要对回归模型进行方差分析,将 y 的总变差 $S_{总}$ 分解成两部分,即回归平方和 U 与剩余平方和 Q ,有下列计算公式:

$$S_{总} = U + Q \quad (6)$$

$$L_{iy} = \sum_{j=1}^n (x_{ji} - \bar{x}_i) (y_j - \bar{y}_i) \quad (7)$$

$$U = \sum_{j=1}^n (\hat{y}_j - \bar{y})^2 = \sum_{i=1}^k b_i L_{iy} \quad (8)$$

$$Q = \sum_{j=1}^n (y_j - \hat{y}_j)^2 = S_{总} - U \quad (9)$$

按照F统计量,则

$$F = \frac{U/k}{Q/(n-k-1)} \quad (10)$$

在显著性水平 α 下,若 $F > F_{\alpha}(k, n-k-1)$,则认为回归方程在 α 水平下是显著的。

3.2 特征参数的反演模型

首先将建模数据归一化,选取45个样本数据的35个作为建模数据建立反演模型,剩余10个作为检验样本检验模型精度,模型精度的检验选取决定性系数(R^2)、均方根误差(RMSE)和平均相对误差(e)。将三组建模样本分别输入SPSS中进行回归分析,得到表6拟合方程。

表6 光谱参数的多元线性拟合方程

参数类型	拟合曲线方程
I	$y_1 = -10.065x_1 - 264.458x_2 - 77.127x_3 + 151.629x_4 + 19.213x_5 - 11.699x_6 + 95.043$
II	$y_2 = -0.814x_1 - 8.177x_2 + 9.558x_3 + 0.094x_4 - 1.564x_5 + 0.075x_6 + 34.174$
III	$y_3 = 0.076x_1 - 0.420x_2 - 0.576x_3 - 0.448x_4 - 0.258x_5 - 0.235x_6 + 34.256$

注:表中多元回归方程式均在置信水平 $\alpha = 0.01$ 下显著。

从图1可以看出三种模型的叶绿素预测值与实测值走向大体一致,具有一定的反演能力,能较好地预测出叶绿素含量。

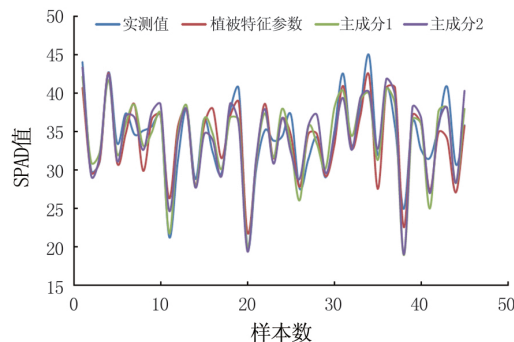


图1 多元线性回归模型实测值与预测值

根据三种模型的建模精度和预测精度(见表7)可以看出,模型Y2的总体精度最高,其次为模

型 Y3 ,两者的建模和检验精度均比较接近 ,说明分波段进行主成分分析虽能提高分析精度 ,但提升幅度较小; 模型 Y1 的建模精度最低 ,表明主成分在综合光谱信息方面具有较突出的优势 ,能使模型更加优化 ,从而得到更为理想的反演结果。

表7 三种参数模型的建模及检验精度

精度类型	反演模型	R ²	RMSE	e
建模精度	Y1	0.78	2.57	6.10
	Y2	0.83	2.20	5.00
	Y3	0.82	2.32	5.5
检验精度	Y1	0.65	3.47	8.04
	Y2	0.80	3.20	8.00
	Y3	0.79	3.58	9.80

4 结论

本文以盆栽玉米为试验对象 ,研究了基于主成分分析构建线性回归模型反演叶片叶绿素含量的可行性。结论如下:

(1) 玉米叶片叶绿素含量与绿峰峰值和近红外反射率均值相关性较高 ,利用这两个参数能够较好地在一元线性反演模型中预测玉米叶片叶绿素含量; 而以主成分进行一元线性模型反演 ,精度较差。

(2) 以主成分作自变量构建的多元线性回归模型反演精度优于以植被特征提取的光谱参数构建的模型精度; 而主成分中 ,利用在蓝光、绿光、红光和近红外各波段提取的主成分建立的多元线性回归模型 ,反演精度优于利用在 350 ~ 2 500 nm 光谱反射波段内提取的主成分建立的反演模型 ,但两者差异较小。通过实测值与预测值的对比以及建模精度与检验精度的比较得出 ,基于主成分分析的多元线性回归模型具有较强的玉米叶片叶绿素含量反演能力 ,反演精度较高 ,能够定量分析植被的叶绿素含量 ,从而为利用光谱数据分析植被长势和污染状况等提供了简易的反演模型。

参 考 文 献:

- [1] Thomas J R , Gausman H W. Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops [J]. *Agronomy Journal* , 1977 , 69(5) : 799 - 802.
- [2] 姚付启 , 张振华 , 杨润亚 , 等. 基于红边参数的植被叶绿素含量高光谱估算模型 [J]. *农业工程学报* , 2009 , 25(S2) : 123 - 129.
- [3] 陈婉婧 , 阎广建 , 吕琳 , 等. 大豆叶片水平叶绿素含量的高光谱反射率反演模型研究 [J]. *北京师范大学学报: 自然科学版* , 2012 , 48(1) : 60 - 65.
- [4] 陈君颖 , 田庆久 , 施润和. 水稻叶片叶绿素含量的光谱反演研究 [J]. *遥感信息* , 2005(6) : 12 - 16.
- [5] 宋开山 , 张柏 , 王宗明 , 等. 大豆叶绿素含量高光谱反演模型研究 [J]. *农业工程学报* , 2006 , 22(8) : 16 - 21.
- [6] 谭昌伟 , 郭文善 , 王纪华 , 等. 浅析遥感光谱特征参量的原理及基本方法 [J]. *遥感技术与应用* , 2010 , 25(1) : 155 - 159.
- [7] 江南. 农作物重金属污染胁迫信息遥感提取方法研究 [D]. 北京: 中国地质大学 , 2009.
- [8] 唐延林 , 王秀珍 , 黄敬峰 , 等. 水稻微分光谱和植被指数的作用探讨 [J]. *农业工程学报* , 2003 , 19(1) : 145 - 150.
- [9] 朱高龙 , 柳艺博 , 居为民 , 等. 4 种常用植被指数的地形效应评估 [J]. *遥感学报* , 2013 , 17(1) : 210 - 234.
- [10] 徐瑞松 , 马跃良 , 陈彧. 资源环境遥感探测 [M]. 合肥: 中国科学技术大学出版社 , 2012.
- [11] 胡红. 矿区植物胁迫作用与遥感信息提取 [D]. 青岛: 山东科技大学 , 2010.
- [12] 浦瑞良 , 宫鹏. 高光谱遥感及其应用 [M]. 北京: 高等教育出版社 , 2000.
- [13] 张金恒. 基于连续统去除法的水稻氮营养光谱诊断 [J]. *植物生态学报* , 2006 , 30(1) : 78 - 82.
- [14] 彭杰 , 迟春明 , 向红英 , 等. 基于连续统去除法的土壤盐分含量反演研究 [J]. *土壤学报* , 2014 , 51(3) : 459 - 469.
- [15] Kokaly R F , Clark R N. Spectroscopic determination of leaf biochemistry using band depth analysis of absorption features and stepwise multiple linear regression [J]. *Remote Sensing of Environment* , 1999 , 67(3) : 267 - 287.
- [16] Zhi H , Brian J T , Stephen J D , et al. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis [J]. *Remote Sensing of Environment* , 2004 , 93(10) : 18 - 29.