

基于可见/近红外反射光谱的大豆叶绿素含量估算方法比较

汤旭光^{1,2}, 宋开山¹, 刘殿伟^{1*}, 王宗明¹, 张 柏¹, 杜 嘉¹, 曾丽红¹, 姜广甲^{1,2}, 王远东^{1,2}

1. 中国科学院东北地理与农业生态研究所, 吉林 长春 130012
2. 中国科学院研究生院, 北京 100049

摘 要 作物叶绿素含量的估测可以为精准农业提供技术支持。该文利用 PROSAIL 模型模拟了不同叶绿素水平下的大豆冠层光谱反射率, 而后针对多期实测高光谱及相应的叶绿素数据, 在对响应波段进行小波能量系数提取的基础上, 分别采用多元线性回归、BP 神经网络和 RBF 神经网络、以及偏最小二乘法进行估算, 并进行了比较分析。研究结果表明, 基于小波分析的三种回归模型都取得了较好的估算效果, 验证模型的 R^2 分别为 0.634, 0.715, 0.873 和 0.776, 其中 RBF 神经网络方法和基于高斯核函数的 PLS 模型精度最好, 能够全面稳定地估算叶绿素含量。

关键词 可见/近红外光谱; 叶绿素含量; 小波分析; 神经网络; 偏最小二乘

中图分类号: TP731; TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2011)02-0371-04

引 言

叶绿素含量是植被重要的生理生化参数, 不仅是作物营养发育状况、光合作用能力的良好指示剂, 同时也是评价作物健康状态及受环境胁迫程度的关键指标^[1,2]。对于大范围农田系统而言, 作物叶绿素含量的估测还可以为合理灌溉、及时施肥、病虫害控制等田间操作提供技术支持。叶绿素的常规测定是使用分光光度计, 这种方法需要破坏植株, 并且在样本运输与制备过程中造成损失^[3]。由于植被反射光谱在可见光范围主要受植被色素的影响, 因此可以用植被冠层和叶片的反射光谱来估算其生化参数特别是色素含量^[4]。高光谱遥感以其连续的波段和高光谱分辨率(数 nm), 无损、快速、高效等方面优势拥有极大的潜力应用于植被的生化参数提取, 可以作为实验室方法的重要补充。

利用高光谱数据进行叶绿素估算通常有三种方法: 一是多元统计方法, 包括光谱反射率、导数光谱、植被指数等^[5,6]; 二是基于特征光谱位置的分析技术, 包括红边、绿峰位置等^[7,8]; 三是辐射传输模型方法^[9]。光谱数据和叶片的生化参量往往在单一试验环境中获得, 能否有效解决波段间和参量间的相关性影响并且充分利用有用光谱信息是反演成败的关键。为此, 一些新的数据处理方法如主成分分析^[10]、小波分析^[11]等方法开始被应用到高光谱信息处理中来, 同

时神经网络^[12]、偏最小二乘法^[13]等技术也逐步得到应用以提高反演精度。然而, 直到目前, 应用高光谱数据估算叶绿素等参数大多使用某一种研究方法, 方法间缺乏集成, 有必要检验现有方法的有效性以及探索发展新方法。本文将在光谱分析的基础上, 同时又利用 PROSAIL 光学辐射传输模型构建了四个层次叶绿素含量的冠层光谱数据, 选择相关波段进行小波分析, 而后再分别采用多元线性回归、神经网络以及偏最小二乘法探讨小波能量系数与相应的叶绿素含量之间的关系, 并进行比较分析。

1 数据来源与处理

本研究的数据于 2006, 2007 年在吉林农业大学试验地进行测量获取, 两个年度共获得 66 组有效数据, 采样时间遍布整个生长季的关键时期。在所有选中的小区里, 用标签确定光谱采集点位置, 用 ASD 野外光谱仪测定大豆冠层光谱反射率, 每个测试点采集 10 个光谱数据, 取其平均值作为一个样本的结果。所有的观测均选择在晴朗无风天气, 每次测定时间在北京时间 10:00~14:00 之间(太阳高度角大于 45°)。在使用 ASD 高光谱仪采集光谱数据时, 探头要保持垂直向下的状态, 探头与冠层顶相距 2.5 m, 每次采集位置相同。叶绿素含量由实验室内通过 UV2550 津岛分光光度计对野外采集的叶片样本进行分析计算获得。

收稿日期: 2010-05-10, 修订日期: 2010-09-25

基金项目: 中国科学院知识创新工程重点项目(KZCX2-YW-QN305), 国家重点基础研究发展计划专题(2009CB421103)资助

作者简介: 汤旭光, 1986 年生, 中国科学院东北地理与农业生态所在读博士研究生 e-mail: tang11100@163.com

* 通讯联系人 e-mail: liudianwei@neigae.ac.cn

ASD FieldSpec 光谱仪可以获取 350~2 500 nm 的光谱数据, 本实验采用了 400~1 350 nm 波长范围内的光谱反射率, 同时将实测的冠层反射率重采样至 10 nm 分辨率。小波分析、神经网络、PLS 模型均在 Matlab 7.0 中通过算法实现, 多元线性回归通过 SPSS11.5 实现, 作图通过 Origin7.5 软件实现。

2 结果与分析

2.1 叶绿素含量与冠层光谱反射率的关系

植被冠层反射光谱在不同的波段有明显的特征, 主要由叶子色素、细胞结构以及叶片含水量所造成, 另外还受植被冠层结构、下垫面情况等影响。为了探究单纯叶绿素对大豆冠层反射光谱的影响, 本文利用 PROSAIL 光学辐射传输模型构建了四个层次的与叶绿素含量相应的光谱曲线, 如图 1 可知, 叶绿素主要影响 400~700 nm 反射率, 特别是反射峰和吸收谷等重要位置变化尤为显著。研究表明光谱微分技术可以有效去除部分线性或接近线性的背景噪声并提高重叠光谱分辨率, 而一阶导数光谱是叶片叶绿素含量较好的评价器^[14]。将其中一条反射光谱进行一阶微分光谱分析, 如图 2 所示, 在可见光波段, 随着反射光谱变化, 导数光谱也会呈现相应的谷或峰值。这也为大豆叶绿素含量遥感监测及定量

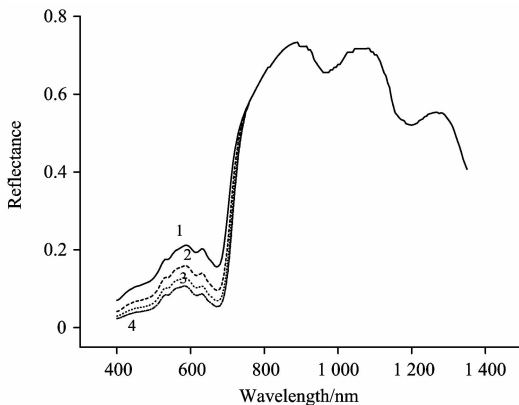


Fig. 1 Soybean canopy spectral reflectance of different chlorophyll concentration

1: $50 \text{ mg} \cdot \text{m}^{-2}$; 2: $100 \text{ mg} \cdot \text{m}^{-2}$;
3: $150 \text{ mg} \cdot \text{m}^{-2}$; 4: $200 \text{ mg} \cdot \text{m}^{-2}$

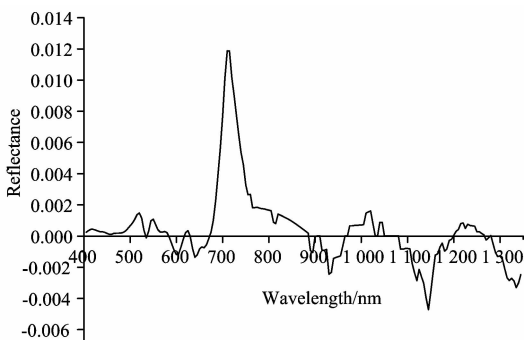


Fig. 2 The first derivative of soybean canopy spectral reflectance

信息提取在敏感波段选取上提供参考和依据。为充分利用光谱信息, 本次研究将受叶绿素影响的 400~700 nm 全部波段进行小波分析。

2.2 光谱数据小波能量系数提取

高光谱遥感虽然可以提供目标地物的精细光谱信息, 但同时也为数据分析与处理带来了挑战, 即实现降维同时又尽可能保持地物原有信息是关键, 小波分析则基本满足了上述要求, 成为高光谱数据处理的有力工具^[12]。然而作为新的光谱分析手段, 小波分析法在植被生理生化参数高光谱提取方面还处于初始阶段, 本文将采用 Daubechies 小波系的 db3 作为基小波而后在 Matlab 下对每个样本的反射率数据进行分解, 共获取 6 个小波能量系数(即数据维数降至 6 维), 而后分别采用多元线性回归、神经网络以及偏最小二乘回归法探讨小波能量系数与相应的叶绿素含量之间的关系。

2.3 大豆叶绿素含量的多元线性回归

小波能量系数提取有类似主成分分析的作用, 各能量系数之间能够实现优势互补, 因此在此项研究中对各小波能量系数做多元线性回归, 从采样数据中随机选取 33 组数据作为训练集建立模型, 这里把各个小波系数依次定义为 x_1, x_2, \dots, x_6 ; 把大豆叶绿素含量定义为 y , 则回归方程如下

$$y = 4.055 - 0.724x_1 - 49.879x_2 + 1.210x_3 + 7.180x_4 - 13.332x_5 - 18.784x_6$$

叶绿素估算模型的确定性系数 R^2 为 0.634, F 值为 7.395, Sig 值为 0.000, 表明模型处于极显著水平, RMSE 为 0.385。利用上式对剩余样本进行预测, 如图 3 所示。验证模型的 R^2 为 0.578, RMSE 为 0.812。由图可以看出, 当叶绿素含量较低时, 实测值与预测值拟合效果较好, 紧密分散在 1:1 对角线附近, 而当大豆叶绿素含量达到 2.5 附近时, 会出现饱和现象, 验证点较分散, 尤其是到了成熟期后, 实测值与预测值误差最大, 小波能量系数对叶绿素的敏感度降低, 影响了估算叶绿素的精度。

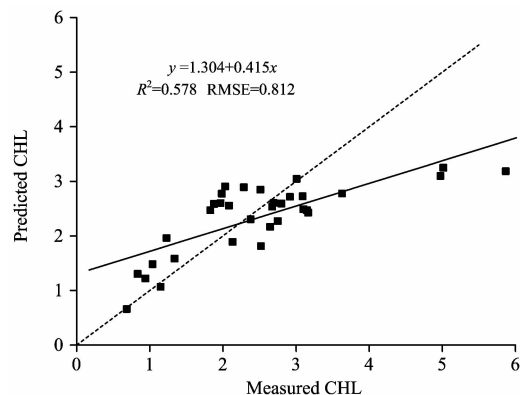


Fig. 3 Relationship between measured CHL and simulated CHL by multiple linear regression

2.4 大豆叶绿素含量的神经网络估算

神经网络算法因其在数据拟合与分类中良好的预测性及实用性广泛应用到各个领域。因此本文将尝试分别采用 BP 神经网络与 RBF 径向基神经网络, 将 6 个小波能量系数作为输入矢量来预测大豆叶绿素含量的变化。估算过程均在

Matlab 中通过算法实现, BP 神经网络训练函数采用 Trainlm, 本文通过多次试验证明网络目标误差为 0.01, 隐藏层神经元为 5 时可以得到较好的估算结果, 利用剩余样本进行预测, 结果如图 4, 实测叶绿素数据和预测值之间 R^2 达 0.715, RMSE 为 0.643。对于 RBF 神经网络, 经过多次试验表明, 当误差目标为 0.01, SPREAD 为 0.1 时可以取得较好的结果, 验证结果如图 5, 实测叶绿素数据和预测值之间 R^2 达 0.873, RMSE 为 0.468。两种神经网络估算模型都取得了较好的效果, 尤其是 RBF 神经网络, 整个生长期的验证点拟合效果都很好。

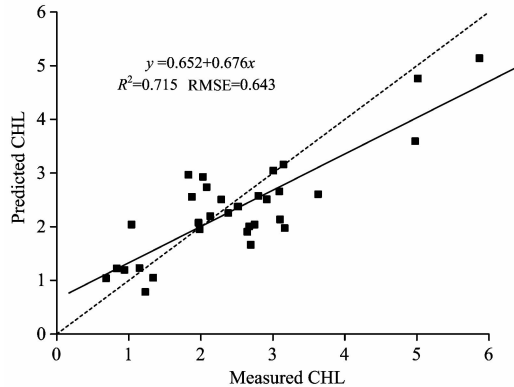


Fig. 4 Relationship between measured CHL and simulated CHL by BP-NN

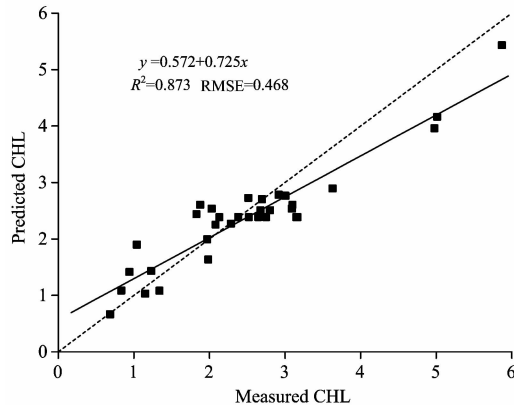


Fig. 5 Relationship between measured CHL and simulated CHL by RBF-NN

2.5 大豆叶绿素含量的偏最小二乘回归

偏最小二乘法 (PLS) 考虑自变量集和因变量集整体的关联关系, 基于寻找自变量集和因变量集的最大相关为目的抽取自变量和因变量各自的主要成分, 从而能够在自变量和因变量间都存在严重多重相关性的条件下进行回归建模。本研究采用的是基于高斯核函数的 PLS 模型, 当各分量分段数取 6、提取的主成分数为 8 时, 模型的预测结果如图 6 所示, 实测叶绿素数据与预测值之间拟合确定性系数 R^2 达 0.776, RMSE 为 0.557, 预测效果较好。

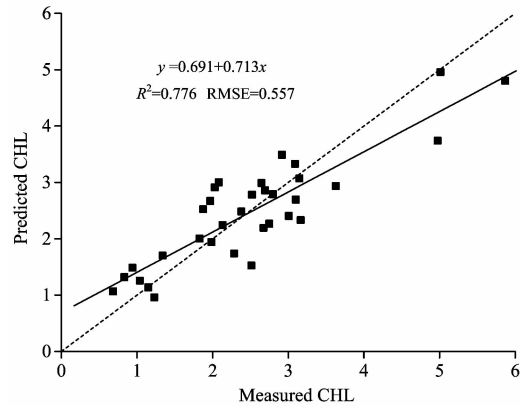


Fig. 6 Relationship between measured CHL and simulated CHL by PLS

2.6 三种模型估算结果比较

对比图 3~图 6 可看出, 多元线性回归、神经网络法、偏最小二乘回归法估算大豆叶绿素含量时都取得了较好的效果, 后两者尤佳, 验证点预测值与实测值密集分布于 1:1 对角线附近, 一致性较好。两种神经网络预测模型验证 R^2 分别为 0.715 和 0.873, RMSE 分别为 0.643 和 0.468, RBF 神经网络的局部优化特性优于 BP 神经网络, 但相对多元线性回归 (R^2 为 0.578, RMSE 为 0.812) 都有非常明显的提高。基于高斯核函数的 PLS 模型验证 R^2 为 0.776, RMSE 为 0.557, 也明显优于传统的多元线性回归模型, 仅次于 RBF 神经网络, 但它有自己的优势, 预测过程神经网络模型相对不稳定, 可能会出现局部异值。

3 结论与讨论

本文通过多期地面实测数据, 研究了基于小波分析的多元线性回归、神经网络、偏最小二乘回归方法估算大豆叶绿素含量的效果。本文中所有数据均是通过地面实测获取, 数据测试工作在多个年份内完成, 测量数据包括多种大豆品种类型, 水肥条件也不尽相同, 而且土壤类型多样化, 数据获取来源于不同的背景条件和生长时期, 可以有效地避免单一条件因素的影响, 因此建立的模型普适性较好, 能够应用于相对较广的环境条件下。

研究表明将小波分析引入高光谱数据特征量提取, 证明小波能量系数与大豆叶绿素含量关系密切, 用来建立反演模型是可行且成功的。三种方法都取得了较好的估算效果, 验证模型的 R^2 分别为 0.578, 0.715 和 0.873, 0.776, RMSE 分别为 0.812, 0.643 和 0.468, 0.557, 其中 RBF 神经网络方法和基于高斯核函数的 PLS 模型精度最好, 能够全面稳定地估算叶绿素含量。偏最小二乘法有其独有的优势, 还可以通过对 PLS 建模的结果进行分析, 来评价光谱指数构建的合理性, 在今后的研究中将进一步应用到其他生理生化参数反演。

References

- [1] Champagne C M, Staenz K, Bannari A, et al. *Remote Sensing of Environment*, 2003, 87: 148.
- [2] Batt B, Mcvicar T R, Vanniel T G, et al. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41: 1246.
- [3] Madaeira A C, Mendonca A, Ferreira M E, et al. *Communications in Soil Science and Plant Analysis*, 2000, 31: 634.
- [4] Thomas J R, Gausman H W. *Agronomy Journal*, 1977, 60(6): 799.
- [5] JIAO Quan-jun, ZHANG Xia, ZHANG Bing, et al(焦全军, 张霞, 张兵, 等). *Remote Sensing for Land & Resources(国土资源遥感)*, 2006, 68(2): 26.
- [6] GAO Yan-hua, CHEN Liang-fu, ZHOU Xu, et al(高彦华, 陈良富, 周旭, 等). *Journal of Remote Sensing(遥感学报)*, 2009, 13(4): 623.
- [7] YAO Fu-qi, ZHANG Zhen-hua, YANG Run-ya, et al(姚付启, 张振华, 杨润亚, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2009, 25(Supp. 2): 123.
- [8] ZHANG Jin-heng, WANG Juan, HAN Chao(张金恒, 王娟, 韩超). *Chinese Agricultural Science Bulletin(中国农学通报)*, 2010, 26(7): 78.
- [9] Dawson T P, Curran P J, Plummer S E. *Remote Sensing of Environment*, 1998, 65: 50.
- [10] Sasmita S, Dadhwal V K. *Journal of the Indian Society of Remote Sensing*, 2005, 33(2): 323.
- [11] SONG Kai-shan, ZHANG Bai, WANG Zong-ming, et al(宋开山, 张柏, 王宗明, 等). *Chinese Agricultural Science Bulletin(中国农学通报)*, 2006, 22(9): 101.
- [12] LIANG Liang, LIU Zhi-xiao, YANG Min-hua, et al(梁亮, 刘志霄, 杨敏华, 等). *J. Infrared Millim. Waves(红外与毫米波学报)*, 2009, 28(5): 353.
- [13] SHEN Zhang-quan, WANG Ke, Xuwen HUANG(沈掌泉, 王珂, Xuwen HUANG). *J. Infrared Millim. Waves(红外与毫米波学报)*, 2010, 29(1): 32.
- [14] ZHANG Jin-heng, WANG Ke, WANG Ren-chao(张金恒, 王珂, 王人潮). *Journal of Shanghai Jiaotong University(上海交通大学学报)*, 2003, 21(1): 74.

Comparison of Methods for Estimating Soybean Chlorophyll Content Based on Visual/Near Infrared Reflection Spectra

TANG Xu-guang^{1,2}, SONG Kai-shan¹, LIU Dian-wei^{1*}, WANG Zong-ming¹, ZHANG Bai¹, DU Jia¹, ZENG Li-hong¹, JIANG Guang-jia^{1,2}, WANG Yuan-dong^{1,2}

1. Northeast Institute of Geography and Agricultural Ecology, Chinese Academy of Sciences, Changchun 130012, China
2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The estimation of crop chlorophyll content could provide technical support for precision agriculture. Canopy spectral reflectance was simulated for different chlorophyll levels using radiative transfer models. Then with multiperiod measured hyperspectral data and corresponding chlorophyll content, after extracting six wavelet energy coefficients from the responded bands, an evaluation of soybean chlorophyll content retrieval methods was conducted using multiple linear regression, BP neural network, RBF neural network and PLS method. The estimate effects of the three methods were compared afterwards. The result showed that the three methods based on wavelet analysis have an ideal effect on the chlorophyll content estimation. R^2 of validated model of multiple linear regression, BP neural network, RBF neural network and PLS method were 0.634, 0.715, 0.873 and 0.776, respectively. PLS based on Gaussian kernel function and RBF NN methods were better with higher precision, which could estimate chlorophyll content stably.

Keywords Visual/near infrared spectra; Chlorophyll content; Wavelet analysis; Neural network; PLS

* Corresponding author

(Received May 10, 2010; accepted Sep. 25, 2010)