

大豆叶面积指数的高光谱估算方法比较

杨飞^{1,2}, 张柏^{1*}, 宋开山¹, 王宗明¹, 刘殿伟¹, 刘焕军^{1,2},
李方¹, 李凤秀^{1,2}, 国志兴^{1,2}, 靳华安^{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 100039

摘要 叶面积指数(leaf area index, LAI)是重要的生物物理参数,亦是各种生态模型、生产力模型以及碳循环研究等的重要生物物理参量,因此具有重要的研究意义。通过分析大量实测数据,选用归一化植被指数(normalized difference vegetation index, NDVI)和比值植被指数(ratio vegetation index, RVI)、主成分分析(principal component analysis, PCA)、神经网络(neural network NN)三种方法对大豆LAI进行了估算,比较分析了三种方法的估算结果。研究表明,植被指数法(NDVI, RVI),主成分分析,神经网络方法LAI都取得了较为理想的结果,验证模型的不确定性系数分别达0.758和0.753,0.954,0.899,其中主成分分析方法和神经网络方法精度较高,主成分分析方法验证模型的稳定性更好,其验证模型的RMSE为0.267,明显低于两个植被指数(NDVI和RVI的RMSE分别为0.594和0.616)和神经网络(RMSE=0.413)。当叶面积指数较小时,植被指数能够较好地去除土壤、大气等条件影响,并精确估算LAI;当叶面积指数较大时,主成分分析能够弥补植被指数饱和的缺陷,得到很好的LAI估算效果。神经网络受LAI大小的影响效果居中,其对叶面积指数较小和较大时具有一致的估算效果,具有较好的应用潜力。

关键词 大豆; LAI; NDVI; RVI; 主成分分析; 神经网络

中图分类号: TP731, TP79 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2008)12-2951-05

引言

叶面积指数可为植物冠层表面最初能量交换描述提供结构化定量信息,是生态系统研究中最重要结构参数之一^[1]。叶面积指数也是重要的植物生理参数,是生态系统功能、净初级生产力等模型重要的陆地特征参量^[2-4]。

采用各种植被指数是早期LAI遥感估算研究的常用方法^[5-7],植被指数法的准确性和普适性难以保证。物理模型被证明能得到较好的LAI反演结果和具有广阔的应用前景,如以三维辐射传输模型为基础的MODIS LAI产品已成为研究、应用等重要且方便的数据源^[8,9]。但物理模型要输入较多的参数,算法复杂,运算较慢,限制了其广泛应用。

随着高光谱技术的发展,较多研究者开始利用高光谱遥感方法来反演叶面积指数、叶绿素等生理参数^[10-15]。高光谱数据可以提供丰富的精细光谱信息,同时也对数据处理提出

了挑战。为充分利用光谱信息,一些新的技术手段,如用神经网络技术^[12,16-18]等逐步得到应用。主成分分析方法具有较好的压缩数据量和降低数据维数功能,在地学中的应用多集中在粗略地识别光谱差异显著的矿物和岩性识别等方面^[19],目前国内外还鲜见其在高光谱数据处理来估算LAI等生理参数方面研究。本文引入主成分分析方法,利用高光谱数据对LAI进行了估算,并与植被指数法和神经网络法估算LAI的效果进行了比较。

1 数据来源与处理

1.1 数据来源

本研究的数据于2003、2004、2006年3个年度分别在长春市郊区、中国科学院海伦农业生态试验地、吉林农业大学试验地进行测量获得,分别获取32,112,19组有效数据,3个年度共获得163组数据。

收稿日期: 2007-05-28, 修订日期: 2007-09-08

基金项目: 中国科学院东北振兴科技行动计划重点项目, 国家自然科学基金项目(40401003), 中国科学院知识创新工程重要方向项目(KZCX3-SW-356)和中国科学院资源环境领域野外台站基金项目资助

作者简介: 杨飞, 1981年生, 中国科学院东北地理与农业生态研究所读博士研究生 e-mail: yangf00_1@163.com

*通讯联系人 e-mail: zhangbai@neigae.ac.cn

在所有选中的小区里，每次测量前用标签进行定点，以便光谱测量和叶面积测量位置一致。2003 和 2004 年使用 ASD 公司生产的 FieldSpec Pro VNIR 高光谱仪测定大豆冠层光谱反射率，光谱范围为 350 ~ 1 050 nm，视场角为 25°，重采样间隔为 1.4 nm；2006 年使用 FieldSpec Pro FR 高光谱仪，光谱范围为 350 ~ 2 500，视场角为 25°，重采样间隔为 1 nm。在每个样点采集光谱之前都进行白板标定，每个测试点采集 10 个光谱数据。所有的观测均选择在晴朗无风天气，每次测定时间为北京时间 10:00-14:00（太阳高度角大于 45°）。测量时保持探头垂直向下，与冠层顶相距约 2.5 m 左右。

大豆叶面积的测定采用 LFCOR 公司生产的 LAF2000 植被冠层分析仪进行。所有叶面积测量时间都选择在当日 6:30 ~ 9:00 或 16:30 ~ 19:00，LAF2000 冠层分析仪设置 1 个天空光，6 个测量目标值，探头佩带 45° 张角的镜头盖 (cap)，避免直射光的影响。

1.2 数据处理

植被指数法是估算 LAI 等生物物理及其他生物化学参数常用的方法，本文选择 NDVI 和 RVI 两个最常用的植被指数估算 LAI^[20, 21]。

$$NDVI = (r_{NIR} - r_R) / (r_{NIR} + r_R)$$

$$RVI = r_{NIR} / r_R$$

式中， r_{NIR} 为近红外波段反射率， r_R 为红波段反射率。

植被指数、主成分分析及其模型计算均通过 SPSS 11.5 实现，神经网络分析在 MATLAB 6.5 中通过 Neural Network Toolbox 实现，作图通过 Origin Pro7.5 软件实现。

2 结果与分析

2.1 LAI 的高光谱植被指数估算研究

通过对数据分析发现，大豆 NDVI 和 RVI 两个植被指数与 LAI 分别呈较好的指数关系和线性关系，估算模型的确定

性系数分别为 0.845 和 0.710。NDVI 和 RVI 实测数据和模型预测数据同样有较好的拟合效果， r^2 分别达到 0.758 和 0.753。由图 1(a) 可看出，当 LAI 小于 3 时，植被指数具有非常理想的去除土壤、大气等噪声影响的效果，模拟和实测值紧密分布在 1:1 附近。但当 LAI 大于 3 时，实测值和模拟值之间的误差较大，验证点分布较分散，NDVI 和 RVI 变化对 LAI 的敏感度降低，因此影响了植被指数估算 LAI 的精度。

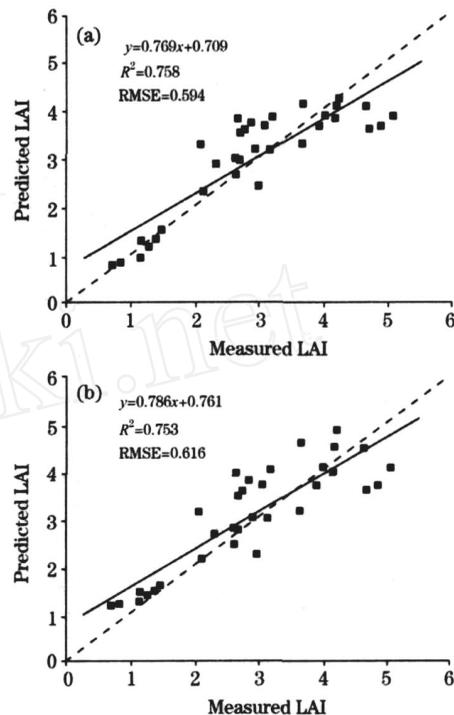


Fig 1 Validated model for LAI estimation of vegetation indices

(a) : Validated model of NDVI;

(b) : Validated model of RVI

Table 1 Estimating LAI Models of NDVI and RVI

植被指数	估算模型 (n=131)			模型验证 (n=32)		
	拟合函数	R ²	RMSE	拟合函数	R ²	RMSE
NDVI	y = 0.029x ^{5.402}	0.845	0.621	y = 0.769x + 0.709	0.758	0.594
RVI	y = 0.182x + 0.421	0.710	0.648	y = 0.786x + 0.763	0.753	0.616

2.2 高光谱数据主成分分析估算 LAI 研究

本文选择 MODIS 传感器的前 15 个可见光和近红外波段中心波长的高光谱反射率数据，对 15 个变量偏相关系数进行检验，得知 KMO (Kaiser-Meyer-Olkin) 检验值为 0.869，根据 Kaiser 度量标准可知，原有 15 个变量适合进行因子分析。

数据处理结果表明，在 412 ~ 678 nm 范围内的 9 个波长在第 1 个因子上有较高的载荷，第 1 个因子主要解释了这几个波长的信息，这几个波长均位于可见光波段范围内，因此将第 1 个因子解释为可见光因子；剩余 6 个波长在第 2 个因子上有较高的载荷，同理第 2 个因子可解释为近红外因子。

Table 2 Principal components of hyperspectral reflectance from PCA

主成分	特征值	比重/ %	累积百分比/ %
1	7.861	52.407	52.407
2	6.432	42.878	95.285

由表 2 可看出，这两个主成分保留了原来 15 个波长的 95.285% 的信息量，信息量丢失很少，基本可以代替原来的 15 个变量。利用上述两个主成分对 LAI 进行估算，结果可用下式表示

$$Y = -14.139f_1 + 15.116f_2 - 1.654, R^2 = 0.904,$$

$$\text{Sig} = 0.000, F = 597.275, \text{RMSE} = 0.373 \quad (1)$$

式中 Y 代表叶面积指数, f_1 和 f_2 分别为主成分分析得到的可见光因子和近红外因子, Sig 表示模型显著水平, F 为 F 检验值。

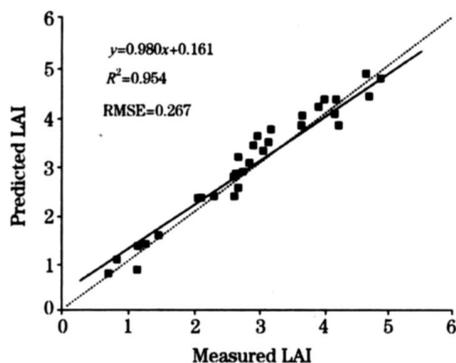


Fig. 2 Relationship between measured LAI and simulated LAI by PCA model

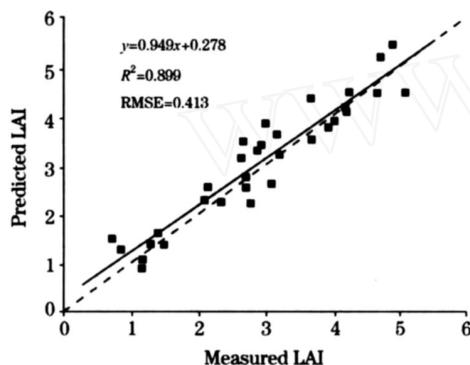


Fig. 3 Relationship between measured LAI and simulated LAI by BP-NN model

通过对高光谱反射率数据进行主成分变换得到可见光和近红外两个主成分因子, 随机选择 131 个样品点建立 LAI 估算模型, 剩余 32 个样品用来验证模型精度, 取得了理想的结果如 (1) 式和图 2。LAI 估算模型的确定性系数 R^2 达到 0.904 ($F = 597.275$, $\text{Sig} = 0.000$), RMSE 为 0.373。验证模型的 R^2 也达到 0.954, RMSE 为 0.267。由此可看出, 主成分分析得到的可见光和近红外两个因子, 可以较好地全面反映 LAI 的变化, 并可对 LAI 进行准确的估算。

2.3 LAI 的神经网络估算

本文采用 BP (Pack-Propagation) 神经网络反向传播算法, 网络共有 3 层, 分别为输入层, 隐藏层, 输出层。输入层中有 15 个输入变量, 分别为 MODIS 传感器的前 15 个可见光和近红外波段中心波长的光谱反射率; 隐藏层和输出层的传输函数分别为“tansig”和“purelin”, 输出层输出 LAI, 训练函数采用 Trainlm, 网络目标误差为 0.001, 训练迭代次数为 1 000 次, 本文通过多次试验证明, 隐藏层神经元个数分别为 5 时可以得到较好的 LAI 估算结果。在本文 163 组数据中, 随机选取 131 组数据用于训练 BP 网络, 剩余的 32 组数据用

来验证网络训练结果。利用 32 组数据对网络训练结果进行验证, 结果如图 3。实测 LAI 数据和神经网络模拟值之间拟合确定性系数 R^2 达 0.899, RMSE 为 0.413, 直线斜率为 0.949, 说明实测和模拟 LAI 值之间具有非常好的变化一致性。

2.4 三种方法估算 LAI 结果比较

对比图 1~图 3 和表 3 可看出, 两种植被指数法、主成分分析法、神经网络法估算大豆的叶面积指数都取得了较好的效果。主成分分析方法估算 LAI 的模型确定性系数达到 0.904, 明显高于植被指数法的估算效果。主成分分析和神经网络方法的验证模型的确定性系数分别为 0.954 和 0.899, 明显优于植被指数法, NDVI 和 RVI 两指数验证模型的 R^2 分别为 0.758 和 0.753。主成分分析方法和神经网络方法验证模型的精度较植被指数法也有非常明显的提高, 前两者的 RMSE 分别为 0.267 和 0.413, 而两个植被指数验证模型的 RMSE 却高达 0.594 和 0.616。比较图 1~图 3 也可看出, 图 2 和图 3 中的点密集分布在 1:1 的对角线附近, 相比较而言, 图 1 中点分布较为分散。主成分分析和神经网络方法的直线斜率分别为 0.980 和 0.949, 植被指数法则分别为 0.769 和 0.786, 由此可看出, 前两种方法验证模型的直线斜率较接近于 1, 即模型模拟 LAI 值和实测数据较一致。

当 LAI 较小时, 植被指数法和主成分分析法的模拟值和实测值之间一致性较好; 但当 LAI 较大 (大于 3 或 4) 时, 植被指数法的模拟值和实测值相差较大, 距离 1:1 斜线较远, 而主成分分析方法仍然表现出较好的稳健性; 神经网络法则一直较为稳定, 但相对较分散。总体来看, 主成分分析方法估算 LAI 可以取得较为稳定的估算结果, 且模型精度最高。

Table 3 Validation models comparison of 4 methods

名称	验证模型	R^2	RMSE	斜率
植被指数 NDVI	$y = 0.769x + 0.709$	0.758	0.594	0.769
植被指数 RVI	$y = 0.786x + 0.763$	0.753	0.616	0.786
PCA	$y = 0.980x + 0.161$	0.954	0.267	0.980
NN	$y = 0.949x + 0.278$	0.899	0.413	0.949

3 讨论

本文通过多期地面实测数据, 研究了主成分分析方法和神经网络方法估算植被 LAI 的效果。植被指数等方法也可以较好地估算 LAI, 但由于其一般仅利用几个波长的信息, 模型稳定性难以保证。主成分分析等方法可以充分挖掘高光谱各波段信息, 实现各波段之间信息互补, 较好地降低了波段数少而带来的随机干扰, 因此主成分分析方法的可靠性和普适性也较好。神经网络方法也能取得与主成分分析方法相近的估算 LAI 的效果, 但对训练样本数量要求较高, 一般适用于大样本数据, 普适性相对稍差。主成分分析方法较神经网络方法简单易用, 并为各领域研究者或工作人员熟悉, 具有更好的推广性。通过主成分分析得到可见光和近红外两个因子, 本文仅利用 1 050 nm 波长之前的高光谱数据, 而其他波

段数据尚未分析。因此利用主成分分析和神经网络方法, 综合 1 050 nm 波长之后的各波段(如到 2 500 nm 左右) 高光谱数据估算 LAI 的效果仍需进一步研究。

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, 有厚、薄黑土和草甸土等, 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。

4 结 论

本文通过对东北地区三个不同试验地的多个品种大豆的实测高光谱数据和叶面积数据进行分析, 对比了植被指数、主成分分析、神经网络三种方法估算大豆 LAI 的效果。研究表明, 应用主成分分析方法处理高光谱数据, 来估算大豆 LAI 是成功的。植被指数(NDVI, RVI)、主成分分析、神经网络方法估算大豆 LAI 都取得了较好的效果, 验证模型的

R^2 分别为 0.758 和 0.753; 0.954 和 0.899, 其中主成分分析方法和神经网络方法精度较好, 这是因为这两种方法用到较多波段的高光谱数据信息, 能够全面稳定地估算 LAI。

经主成分分析转换得两个主成分, 分别解释为可见光因子和近红外因子, 这两个因子可以包括 95.285% 的高光谱信息量。PCA 方法具有理想的 LAI 估算效果(估算模型 r^2 达 0.904), 验证模型的精度明显高于植被指数法和神经网络法。

当叶面积指数很小, 植被指数法可以有效的去除土壤噪声的干扰, 实测数值和模型预测值聚集于 1:1 斜线附近; 但当叶面积指数较大时, 主成分分析方法仍然是较为理想的估算方法, 可以有效弥补植被指数易饱和的缺陷。

神经网络在高光谱数据处理方面的应用, 属于一个相对较新的领域, 尚处于探索阶段, 输入层、隐藏层等数目的选择及其与学习速率等的最佳组合, 都对高光谱数据处理结果产生较大影响, 有待于进一步研究。从本文数据分析结果可看出, 神经网络用于 LAI 生理参数估算的预测值与实测值较一致(斜率为 0.949), 具有较好的应用潜力。

参 考 文 献

- [1] WANG Xi-qun, MA Lu-yi, JIA Zhong-kui, et al(王希群, 马履一, 贾忠奎, 等). Chinese Journal of Ecology(生态学杂志), 2005, 24(5): 537.
- [2] Sellers P J, Dickinson R E, Randall D A, et al. Science, 1997, 275: 502.
- [3] Cramer W, Kicklighter D W, Bondeau A. Global Change Biology, 1999, 5(suppl. 1): 1.
- [4] Running S W, Nemani R R, Heinsch F A, et al. Bioscience, 2004, 54(6): 547.
- [5] Chen J M, Cihlar J. Remote Sensing of Environment, 1996, 55: 153.
- [6] Turner P D, Warren B C, Robert E K. Remote Sensing of Environment, 1999, 70: 52.
- [7] Chen J M, Pavlica G, Brown L, et al. Remote Sensing of Environment, 2002, 85: 165.
- [8] Myneni R B, Nemani R R, Running S W. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(66): 1380.
- [9] Wythers K R, Reich P B, Turner D P. Tree Physiology, 2003, 23: 1171.
- [10] WANG Xiu-zhen, HUANG Jing-feng, LI Yun-mei, et al(王秀珍, 黄敬峰, 李云梅, 等). Journal of Remote Sensing(遥感学报), 2004, 8(1): 81.
- [11] JI Hai-yan, WANG Peng-xin, YAN Tai-lai, et al(吉海彦, 王鹏新, 严泰来, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2007, 27(3): 514.
- [12] SONG Kai-shan, ZHANG Bai, WANG Zong-ming, et al(宋开山, 张 柏, 王宗明, 等). Scientia Agricultura Sinica(中国农业科学), 2006, 39(6): 1138.
- [13] Gong P, Pu R L, Biging G S, et al. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(6): 1355.
- [14] Haboudane D, Miller J R, Pattey E, et al. Remote Sensing of Environment, 2004, 90(3): 337.
- [15] Gupta R K, Vijayan D, Prasad T S. Advances in Space Research, 2006, 38(10): 2212.
- [16] Gong P. International Journal of Remote Sensing, 1999, 20(1): 111.
- [17] Bacour C, Baret F, Béal D, et al. Remote Sensing of Environment, 2006, 105: 313.
- [18] Fang H, Liang S. Remote Sensing of Environment, 2005, 94: 405.
- [19] PU Rui-liang, GONG Peng(浦瑞良, 宫 鹏). Hyperspectral Remote Sensing and Its Application(高光谱遥感及其应用), 2000.
- [20] Pearson R L, Miller L D. In Proc 8th International Symposium on Remote Sensing of Environment, ERIM, Ann Arbor, MI, 1972. 1357.
- [21] Rouse J W, Jr Hasas R H, Schell J A, et al. Progress Report 1978-1, NTIS No. E73-10693, Remote Sensing Center, Texas A & M University, College Station, 1973. 93.

Comparison of Methods for Estimating Soybean Leaf Area Index

YANG fei^{1,2}, ZHANG Bai^{1*}, SONG Kai-shan¹, WANG Zong-ming¹, LIU Dian-wei¹, LIU Huan-jun^{1,2}, LI Fang¹, LI Feng-xiu^{1,2}, GUO Zhi-xing^{1,2}, JIN Hua-an^{1,2}

1. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130012, China

2. Graduate School of Chinese Academy of Sciences, Beijing 100039, China

Abstract Leaf area index (LAI) is an important biophysical parameter, and is the critical variable in many ecology models, productivity models and carbon circulation study. Based on the field experiment data, an evaluation of soybean LAI retrieval methods was conducted using NDVI (normalized difference vegetation index) and RVI (ratio vegetation index), principle component analysis (PCA) and neural network (NN) methods, and the estimate effects of three methods were compared. The results showed that the three methods have an ideal effect on the LAI estimation. R^2 of validated model of vegetation indices, PCA, NN were 0.753 (NDVI), 0.758 (RVI), 0.883, 0.899. PCA and NN methods were better with higher precision, and PCA method was the best, as its RMSE (0.202) was slower than the two vegetation indices (RMSEs of NDVI and RVI were 0.594 and 0.616) and NN (RMSE was 0.413) method. While the LAI was small, vegetation indices were obvious for removing the noise from soil and atmospheric effect and obtained the good evaluation result. PCA showed better effect for all LAI. LAI affected the estimating result of NN method moderately. As for the NN method, modeled LAI value and measured LAI regression formula slope was the nearest to 1 with R^2 of 0.949, which showed a great potential for LAI estimating. As a whole, PCA and NN methods were the prior selection for LAI estimation, which should be attributed to the application of hyperspectral information of many bands.

Keywords Soybean; LAI; NDVI; RVI; PCA; NN

(Received May 28, 2007; accepted Sep. 8, 2007)

* Corresponding author