第28卷,第5期 光 谱 学 与 光 谱 分 析 Vol. 28,No. 5,pp1098-1101
2008年5月 Spectroscopy and Spectral Analysis May, 2008

基于光谱波段自相关的水稻信息提取波段选择

王福民,黄敬峰,徐俊锋*,王秀珍

浙江大学农业遥感与信息技术应用研究所,浙江大学农业信息科学与技术中心,
浙江省农业遥感与信息技术重点研究实验室,浙江 杭州 310029

摘 要 通过大田试验,使用ASD光谱仪测量水稻不同生育期的冠层光谱,将光谱以10 nm为步长进行合并,再将不同日期光谱的所有波段组合计算相关系数平方(R^2),生成R^2矩阵,并绘制R^2分布图。根据R^2越大,光谱波段之间冗余信息越多,R^2越小,水稻光谱波段信息含量越多的原则,在所有测量日期中选择出前100个R^2最小值对应的波段,将这些波段进行统计分析。结果表明,可见光区域各个波段之间和红外(近红外和短波红外)区域各个波段之间都含有大量冗余信息。水稻信息量丰富的波段主要集中在可见光的长波波段,红边波段,近红外第一和第二峰值波段,以及短波近红外第一峰前区(1 530 nm附近)和第二峰值区(2 215 nm附近)。比较水稻与其他植被对于最优波段的选择,400～410,630～650和1 520～1 540 nm三个波段区间表现为水稻信息提取较为独特的波段。

关键词 波段自相关;水稻;波段选择
中图分类号:S127 **文献标识码**:A **文章编号**:1000-0593(2008)05-1098-04

引 言

遥感技术被认为是估算植被生物物理和生物化学参数的一种可靠方法[1]。高光谱遥感具有大量、连续的窄波段通道,可以获取地物的连续光谱曲线。因此,它可以提供更精确描述地物反射特征的能力。大量的研究表明高光谱数据在估算覆盖度、生物量、叶面积指数等方面比多光谱数据具有一定的改进和提高[2-5],特别是在生物化学参数估算方面具有明显的优势[6-10]。可是,高光谱数据比多光谱数据复杂得多,数据量大并且存在大量冗余[11]。

为了解决高光谱波段冗余问题,有必要根据不同的遥感目标物,不同的应用目的,从大量高光谱波段中选择出一些有用的波段进行遥感信息提取。波段选择方法主要包括逐步回归选择法[12],简单波段自相关选择法[13]和主成分系数权重选择法[14]等。对于水稻而言,由于其生长在水田中,土壤表面的水层改变了土壤植被系统的光谱反射率[15],需要针对这一特点进行分析研究,确定水稻信息提取的合适波段。本文试图利用水稻冠层光谱数据,采用一种包含所有波段自相关的分析方法,来获取水稻信息提取的光谱波段。

1 材料和方法

1.1 试验设计

试验水稻包括秀水110、嘉育293、嘉早312、嘉早324、协优9308五个品种。设置了3个氮素水平,每公顷分别为0,120,240 kg。多品种、多氮素水平的目的是使波段选择不是只针对一种水稻而进行,而使选择出的波段能更好的代表水稻这个整体。

1.2 测量方法

观测使用的仪器为美国ASD(analytical spectral device)公司的ASD FieldSpe Pro FR™光谱仪。于2002年7月17日,7月23日,7月30日,8月5日,8月22日,8月31日,9月20日和10月03日,10:00～14:00测定了水稻冠层光谱反射率。测量时基本无云无风,传感器探头垂直向下,视场角为25°,距离冠层顶端约为0.7 m。

1.3 波段选择方法

这里选用基于波段自相关系数的波段选择方法。由于临近光谱波段高度相关,因此在350～2 500 nm光谱范围内从350 nm开始以10 nm为波段宽度逐步平均,构成215个10 nm宽度的波段值,从而与第一个星载高光谱传感器Hyperion的波段设置相一致。去除由于水汽影响的36个波段,剩

收稿日期:2007-01-16,**修订日期**:2007-04-19
基金项目:国家自然科学基金项目(40571115)和科技部科技基础性工作专项(2003DEA2C010-13)资助
作者简介:王福民,1978年生,浙江大学环资学院博士研究生 *通讯联系人 e-mail:xjf11@zju.edu.cn

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

下的 179 个波段用于水稻信息提取的波段选择。将这些波段，两两组合计算相关系数的平方(R^2)，所有这些波段组合构成一个 179 ×179 的 R^2 矩阵，共 32 041 个元素。一般来说，R^2 越小，表示两个波段之间冗余信息越少[14]。

2　结果与分析

2.1　波段自相关系数平方的图形表示

将所有计算的 R^2 矩阵用图形表示，由于矩阵是对称的，因此只用一个三角矩阵就可以表示全部数据。图 1 给出的是由 7 月 30 日和 8 月 05 日水稻冠层光谱计算生成的波段自相关 R^2 图。其他日期的 R^2 这里没有给出。图中每一个点分别对应的两个波段(横、纵坐标)的 R^2。不同深度的颜色表示 R^2 值处于不同的范围，最黑的点表示前 200 个最小的 R^2，稍微次黑的点表示第 201 个到第 500 个最小 R^2 值，依次类推，一直到第 5 000 个之后都用较为偏白的点表示。波段自相关图可以用来确定信息含量丰富的波段。R^2 越大，表示这两个波段相关性越高，则含有的冗余信息越多；而 R^2 越小，则意味着两个波段冗余信息越少，意味着这两个波段分别表征了水稻的不同信息，即它们是信息含量丰富的波段。由图 1 可见，可见光波段区域各个波段之间和红外波段(近红外和短波近红外)区域各个波段之间都具有很高的相关性，这意味着可见光波段之间和红外波段之间本身都含有大量的冗余信息。但可见光与红外波段之间的某些组合具有较小的 R^2，因此这些光谱波段包含着大量的水稻信息。

2.2　在波段自相关系数平方(R^2)矩阵中最佳波段的选择

由图 1 可以得到一个对信息含量丰富的波段所在位置的整体了解。然而要具体确定哪些波段含有了水稻大量的信息，还有必要做更深入细致的分析。首先，在相关系数平方矩阵中选出的前 100 个最小的 R^2 所对应的两两组合的波段值，然后再将这些波段值相同的进行合并，统计出不同波段

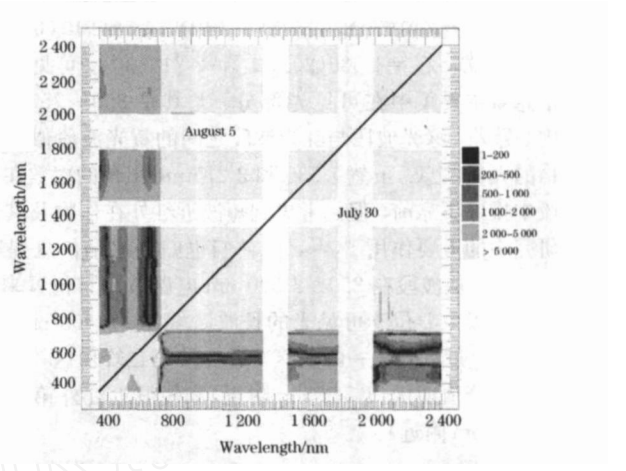

Fig 1　Distribution plot of the square of correlation coefficients among different reflectance wavebands of rice canopy (The black points (1-200) denote the first 200 minimum R^2 values)

值出现的次数(表 1)。由于数据表过大，只给出前 10 个出现次数最多的波段值。由表 1 可见，不同生育期水稻信息含量丰富波段不同，整体而言，主要集中 405 nm 附近、565～705，885～945，1 045～1 125 和 1 525 nm 附近和 2 215 nm 附近几个区间。其中，出现在 565～705 nm 光谱区域的波段占总波段数的 32.94 %，1 045～1125 和 885～945 nm 区域出现的波段分别占总波段数的 12.19 %，10.31 %，位于短波红外 2 个光谱区间的波段数量都相对较少。在这些选出的波段中，将相邻位置的波段合并，就构成了波段宽度，而这个宽度的平均值即为波段的中心位置。从中选择出 17 个波段作为含有水稻大量信息的光谱波段，它们的波段中心位置和波段宽度(nm)分别是：λ_1 = 405(10)，λ_2 = 565(10)，λ_3 = 585(30)，λ_4 = 605(10)，λ_5 = 620(20)，λ_6 = 640(20)，λ_7 = 660(20)，λ_8 = 680(20)，λ_9 = 695(10)，λ_{10} = 705(10)，λ_{11} = 720

Table 1　Wavebands selected and their occurrences for different growth stages

日期	排序	1	2	3	4	5	6	7	8	9	10
7 月 17 日	波段	565	645	655	705	1 525	1 535	1 545	2 215	2 225	2 235
	次数	11	5	5	29	8	7	5	6	5	5
7 月 23 日	波段	355	365	405	605	645	695	755	1 085	1 735	2 215
	次数	3	4	5	35	6	34	4	2	2	2
7 月 30 日	波段	515	565	575	585	595	605	615	635	695	2 015
	次数	6	13	19	9	11	6	9	3	6	7
8 月 5 日	波段	405	635	645	665	675	685	775	845	865	915
	次数	15	24	10	16	15	15	4	4	4	4
8 月 22 日	波段	705	715	725	895	905	915	925	935	945	975
	次数	5	5	4	11	16	15	5	5	6	5
8 月 31 日	波段	715	745	885	895	905	915	925	935	945	1 995
	次数	6	11	5	9	10	9	6	5	6	7
9 月 20 日	波段	625	635	645	1 065	1 075	1 085	1 095	1 105	1 115	1 125
	次数	11	11	11	10	11	12	11	10	12	12
10 月 3 日	波段	605	615	625	635	645	655	935	1 075	1 085	1 095
	次数	8	14	14	16	25	23	6	6	6	6

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

(20), λ_{42} = 740 (20), λ_{43} = 865 (40), λ_{44} = 910 (60), λ_{45} = 1 085(90), λ_{46} = 1 530(20), λ_{47} = 2 215(30)。可见水稻光谱信息量主要集中在可见光部分,尤其是 550～750 nm 范围内。另外,绿光波段与红光波段之间的黄光波段似乎含有水稻的大量信息。虽然 1 530 和 2 215 nm 出现的次数在所有波段中排序不靠前,但是考虑到短波近红外在作物及其他植被研究中的重要作用[16,17],从它们也归到水稻信息提取的有用波段。从波段在 350～2 500 nm 电磁波分布位置来看,选出的波段主要位于可见光的长波区域(除了 405 nm 外),红边波段,近红外第一峰值区,近红外第二峰值区,短波近红外第一峰值前区(1 530 nm 附近)和短波近红外第二峰值区(2 215 nm 附近)。

3 结论和讨论

本研究通过分析选择出 17 个不同波段宽度的波段区间,这些波段位置都与水稻的某些参数具有密切的关系。λ_1 是蓝光区域较短的波段,与色素吸收密切相关;λ_2 是靠近绿峰的波段,对叶绿素比较敏感;λ_3, λ_4, λ_5, λ_6, λ_7 是反射率由绿峰向红谷减小的波段,对于叶绿素,生物量,LAI 都比较敏感;λ_8, λ_9 分别是红谷波段和红谷之后反射率开始增加的波段,与生物量和 LAI 有密切关系;λ_{10}, λ_{11}, λ_{12} 是反射率迅速增加的红边波段,与作物胁迫密切相关;λ_{13}, λ_{14} 和 λ_{15} 是与细胞结构及冠层结构有关的近红外高反射平台波段,对叶绿素总量和蛋白质比较敏感;λ_{16} 是短波近红外第一峰反射率上升波段,与淀粉含量有关;λ_{17} 是短波近红外第二峰的峰值波段,对木质素、生物量和淀粉比较敏感。比较这些波段与其他研究给出的波段[14,16,17],可见既有相似的波段,也有不同的波段。其区别主要在 λ_1, λ_5, λ_{16} 这三个波段区间,或许这正是水稻与其他植被不同之处。另外,本文选出的 17 个波段,只是表示包含水稻信息量大的波段,并不是包含了所有水稻信息。

参 考 文 献

[1] Hung T N, Byun-Woo L. European Journal of Agronomy, 2006, 24: 349.
[2] Broge N H, Leblanc E. Remote Sensing of Environment, 2001, 76: 156.
[3] Mutanga O, Skidmore A K. Journal of Applied Earth Observation and Geoinformation, 2004, 5: 87.
[4] Tang Y L, Wang R C, Huang J F. Pedosphere, 2004, 14: 467.
[5] Liu Z Y, Huang J F, Wu X H, et al. Journal of Integrative Plant Biology, 2007, 49: 299.
[6] LI Jun-xia, MIN Shun-geng, ZHANG Hong-liang, et al(李君霞, 闵顺耕, 张洪亮, 等). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2006, 26(5): 833.
[7] O'Neill A L, Kupiec J A, Curran P J. Remote Sensing of Environment, 2002, 80: 134.
[8] Zhao D L, Reddy K R, Kakani V G, et al. European Journal of Agronomy, 2005, 22: 391.
[9] Zarco-Tejada P J, Miller J R, Morales A. Remote Sensing of Environment, 2004, 90: 463.
[10] WEI Liang-ming, JIANG Hai-ying, LI Jun-hui, et al(魏良明, 姜海鹰, 李军会, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(9): 1404.
[11] Becher B L, Lusch D P, Qi Jiaguo. Remote Sensing of Environment, 2005, 97: 238.
[12] Etchegorry J P, Bruniquel-Pinel V. Remote Sensing of Environment, 2001, 76: 1.
[13] Huang Z, Turner B J, Dury S J. Remote Sensing of Environment, 2004, 93: 18.
[14] Thenkabail P S, Enclona E A, Ashton M S Bauke V D. Remote Sensing of Environment, 2004, 91: 354.
[15] Casanova D, Epema G F, Goudriaan J. Field Crops Research, 1998, 55: 83.
[16] Lars E, Lars H, Andres K. Remote Sensing of Environment, 2001, 78: 239.
[17] Thenkabail, P S, Ward A D, & Lyon J G. International Journal of Remote Sensing, 1994, 15: 49.

Wavebands Selection for Rice Information Extraction Based on Spectral Bands Inter-Correlation

WANG Fu-min, HUANG Jing-feng, XU Jun-feng*, WANG Xiu-zhen
Institute of Agriculture Remote Sensing & Information System Application, Zhejiang University, Hangzhou 310029, China

Abstract The hyperspectral remote sensing data usually involve hundreds or even thousands of narrow bands, which may be crucial for providing additional information with significant improvements over broad bands in quantifying biophysical and biochemical variables of agricultural crop. However, the huge data generated by hyperspectral systems, and the problems this presents for storage and analysis, have far prevented the routine use of such data. The objective of the present research was to identify the spectral bands in the visible and near-infrared range that were suitable for the study of rice. The hyperspectral reflectance

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

of canopy in different development stages was measured in experimental field using a 1 nm-wide spectroradiometer but was aggregated to 10 nm-wide bandwidths to match the first spaceborne hyperspectral sensor, Hyperion. The correlation coefficients (r) between all the combinations of spectral bands were computed, and then they were converted to R^2, which constituted R^2 matrices. The matrices were plotted against wavebands. The criterion of band selection is that the lower the R^2 value, the less the redundancy between two wavebands while the higher R^2 indicates that there is redundant information between two wavebands. According to the criterion, the wavebands corresponding to the first 100 minimum R^2 values were selected from all canopy spectra collected on different dates. And then these bands were analyzed. The results indicate that the visible and infrared (NIR and SWIR) themselves contain redundant information. The wavebands containing abundant information of rice are located in specific bands in the longer wavelength portion of the visible region, with secondary clusters in red edge region, in strongly reflective near-infrared region with relatively higher reflectance, in one particular section of short wave near-infrared (SWIR) (1 530 nm) and in the second maximum reflectance region of SWIR (2 215 nm). Compared with the selected bands with other vegetation, rice seems to have three spectral regions of 400-410 nm, 630-650 nm and 1 520-1 540 nm, which exclusively depict the characteristics of rice. Moreover, this research identified 17 spectral bands in the visible and near-infrared region, which were 405, 565, 585, 605, 620, 640, 660, 680, 695, 705, 720, 740, 865, 910, 1 085, 1 530 and 2 215 nm. These bands contain the majority of the rice information content. A reduction in band number without significant information loss is important because it makes it possible to achieve fine spatial resolution without sacrificing the ability to characterize rice status.

Keywords Bands inter-correlation; Rice; Bands selection

(Received Jan. 16, 2007; accepted Apr. 19, 2007)

* Corresponding author

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net