Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Optimal waveband identification for estimation of leaf area index of paddy rice*

Fu-min WANG[1], Jing-feng HUANG†‡[1], Qi-fa ZHOU[2], Xiu-zhen WANG[3]

([1]*Institute of Agricultural Remote Sensing and Information Application, Zhejiang University, Hangzhou 310029, China*)

([2]*College of Life Sciences, Zhejiang University, Hangzhou 310058, China*)

([3]*Zhejiang Meteorological Institute, Hangzhou 310004, China*)

†E-mail: hjf@zju.edu.cn

Received July 6, 2008; revision accepted Nov. 3, 2008; CrossCheck deposited Nov. 12, 2008

Abstract: The objectives of the study were to select suitable wavebands for rice leaf area index (LAI) estimation using the data acquired over a whole growing season, and to test the efficiency of the selected wavebands by comparing them with feature positions of rice canopy spectra. In this study, the field experiment in 2002 growing season was conducted at the experimental farm of Zhejiang University, Hangzhou, China. Measurements of hyperspectral reflectance (350~2500 nm) and corresponding LAI were made for a paddy rice canopy throughout the growing season. And three methods were employed to identify the optimal wavebands for paddy rice LAI estimation: correlation coefficient-based method, vegetation index-based method, and stepwise regression method. This research selected 15 wavebands in the region of 350~2500 nm, which appeared to be the optimal wavebands for the paddy rice LAI estimation. Of the selected wavebands, the most frequently occurring wavebands were centered around 554, 675, 723, and 1633 nm. They were followed by 444, 524, 576, 594, 804, 849, 974, 1074, 1219, 1510, and 2194 nm. Most of them made physical sense and had their counterparts in spectral known feature positions, which indicates the promising potential of the 15 selected wavebands for the retrieval of paddy rice LAI.

Key words: Rice, Hyperspectral reflectance, Leaf area index (LAI), Wavebands identification
doi:10.1631/jzus.B0820211 **Document code:** A **CLC number:** S127

INTRODUCTION

Rice is the staple food of mankind and provides 35%~60% of the dietary calories consumed by 3 billion people, making it inarguably the most important crop worldwide (Confalonieri and Bocchi, 2005). Measuring the leaf area index (LAI, leaf area per unit ground area) of paddy rice fields provides information on crop growth dynamics, and has the potential to be a good indicator of the status of paddy rice throughout the growing season (Xiao *et al*., 2002). Besides, it is highly correlated with rice biomass and productivity (Dobermann and Pampolino, 1995). Moreover, the LAI monitoring of paddy rice is crucial in outlining an efficient water management policy in dry areas because paddy rice is grown on flooded soils.

Direct estimation of LAI using destructive field measurements are extremely labor-intensive and are, for practical reasons, limited to a few samples (Eklundh *et al*., 2001). Remote sensing technique allows fast, non-destructive and relatively cheap characterization of crop status (Bouman, 1995), and it has been recognized as a reliable method for retrieval of canopy LAI, especially using hyperspectral remote sensing that generally measures reflectance in many narrow, continuous spectral bands (Strachan *et al*., 2002). The hyperspectral remote sensing data usually involve

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 40571115), the Hi-Tech Research and Development Program (863) of China (No. 2006AA120101), and the National Basic Research Program (973) of China (No. 2006BAD10A09)

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 (Mutanga and Skidmore, 2004; Haboudane *et al*., 2004). Some previous studies also have demonstrated the usefulness and potential of hyperspectral data for LAI estimation of paddy rice (Shibayama and Akiyama, 1989; Casanova *et al*., 1998; Vaesen *et al*., 2001; Yang *et al*., 2007).

While hyperspectral data provide the opportunity for a more detailed analysis of on-ground materials, the hyperdimensional data generated by hyperspectral sensors create a new challenge not only for conventional spectral data analysis techniques (Jiminez and Landgrebe, 1999), but also for storage and transport of such data. Another issue associated with hyperspectral data is band redundancy, which means that adjacent wavebands often contain redundant information (Thenkabail *et al*., 2004a; Becker *et al*., 2005). In addition, the greater number of wavebands requires larger sets of training data, which is not practical since one must take into account that as the number of dimensions increases, the sample size of the training data needs to increase exponentially in order to have reliable multivariate statistics for estimating specific parameters (Koger *et al*., 2003). One of the solutions to these problems is to select a relatively few spectral wavebands that are optimized for a specific target, for instance paddy rice LAI estimation, thus reducing the dimensionality of the hyperspectral data.

The selection of wavebands has been performed in numerous studies. Thenkabail *et al*.(2000) recommended 12 specific narrow bands that provide optimal crop information in the 350 to 1050 nm spectral range. Then in the broader range from 400 to 2500 nm, 22 optimal bands were selected to best characterize vegetation and agricultural crops (Thenkabail *et al*., 2004b). Based on the second-derivative analysis, Becker *et al*.(2005) identified 8 optimal bands in the visible and near-infrared (NIR) wavelength regions, which appear to contain the majority of the coastal wetland information content of the full spectral region. Through a variety of waveband selection techniques, the original hyperspectral dataset can be distilled by removing the redundant bands and keeping only those that contain the valuable information. These selection techniques include correlation coefficient-based method (Curran *et al*., 2001; Huang *et al*., 2004), vegetation index-based method (Thenkabail *et al*., 2000) and stepwise regression method (Grossman *et al*., 1996; Lee *et al*., 2004), and so on.

Although numerous studies on wavebands selection have been carried out as mentioned above, few attempts have been made so far to select optimal wavebands suitable for rice LAI estimation using data acquired over a whole growing season. With this in mind, the objective of this research was to select suitable wavebands for rice LAI estimation using the data acquired over a whole growing season and to test the efficiency of the selected wavebands by comparing them with feature positions of rice canopy spectra.

MATERIALS AND METHODS

Experimental site

The field experiment was conducted at the experiment farm of Zhejiang University, Hangzhou, Zhejiang Province, China, located at 30°14′ N, 120°10′ E. The experiment area is characterized by a monsoon climate with a hot summer, a cool winter, and marked seasonal variations in precipitation. The soil type is sandy loam paddy soil with pH of 5.7, organic matter of 16.5 mg/g and total nitrogen (N) of 1.02 mg/g.

Experimental design

The experiment was carried out during the 2002 growing season (July to October, 2002). Five rice varieties (Xiushui 110, Xieyou 9308, Jiayu 293, Jiazao 312 and Jiazao 324) were selected for the investigation. Among these varieties, Xieyou 9308, Jiayu 293, Jiazao 312 and Jiazao 324 are indicia rice, and Xiushui 110 is japonica rice. Xieyou 9308 is hybrid rice; the others are common rice. Three nitrogen treatments were set, with the concentrations of 0, 140, and 240 kg/ha, respectively, using urea as the nitrogen fertilizer. Each treatment had four replications. Therefore, we used a completely randomized design consisting of 60 plots of 4.81 m×4.68 m. Phosphorus (P) and potassium (K) were supplied in adequate amounts according to the general nutrient status of the field.

Data collection

The rice canopy reflectance data and the corresponding LAIs were collected from early stem elongation until ripe, at one- to three-week intervals.

1. Reflectance measurements

The spectroradiometer used in this study is an ASD Fieldspec FR spectroradiometer (ASD, Boulder, USA), which covers the range from 350 to 2500 nm. The sampling interval is 1.4 nm with a resolution of 3 nm over the 350~1050 nm range, and about 2 nm with a spectral resolution between 10 and 12 nm over the 1050~2500 nm range. Hyperspectral reflectance was measured within each plot at randomly selected locations with a nadir view of 25° from a height of 1.0 m above canopy, which produced a circular instantaneous-field-of-view of 0.44 m diameter and approximately 0.154 m^2 in canopy surface area. Prior to each plant reflectance measurement, reflectance of a white standard panel coated with $BaSO_4$ was taken. The spectral files were recorded as the average of 10 readings to minimize the impact of target variability.

In all measurement campaigns, measurements were made as close to solar noon as possible on days, with clear sky conditions to minimize the effects of solar angle and contributions from diffuse light. The measurements were performed on July 17, July 23, July 30, Aug. 5, Aug. 22, Aug. 31, Sept. 20, and Oct. 4, and the corresponding growth stages of rice are presented in Table 1 (Wang *et al*., 2008). It can be seen that Xiushui 110 and Xieyou 9308 need a longer growing time than Jiazao 312, Jiayu 293, and Jiazao 324.

Table 1 Rice growth stages of different varieties in 2002 growing season (Wang *et al*., 2008)*

Development stages	Var1	Var2	Var3	Var4	Var5
Seeding	02/06	02/06	02/06	02/06	08/06
Emergence	05/06	05/06	05/06	05/06	11/06
Three leaves	09/06	09/06	09/06	09/06	15/06
Transplanting	25/06	25/06	25/06	25/06	25/06
Tillering	05/07	04/07	04/07	04/07	05/07
Elongation	05/08	17/07	17/07	17/07	05/08
Booting	26/08	21/07	21/07	21/07	23/08
Heading	02/09	28/07	28/07	04/08	31/08
Flowering	06/09	31/07	31/07	07/08	02/09
Milk ripe	19/09	15/08	15/08	24/08	15/09
Ripe	15/10	05/09	05/09	10/09	15/10
Harvest	25/10	12/09	12/09	12/09	25/10

Var1: Xiushui 110; Var2: Jiazao 312; Var3: Jiayu 293; Var4: Jiazao 324; Var5: Xieyou 9308. *Data expressed as day/month

2. Leaf area index (LAI)

For each plot, a representative area was selected for reflectance measurements, and after that the same samples were cut and taken for laboratory analysis. In the lab, plants were separated into leaves, stems, and roots. The leaf areas of samples were measured, and then the LAI was determined.

Methods of selection of wavebands

We implemented a series of methods and techniques in selecting suitable wavebands for paddy rice LAI estimation. These methods are presented below.

1. Correlation coefficient-based method

Spectra were smoothed using a mean filter smoothing algorithm prior to the derivative computation procedure, and the derivative approximation method presented by Tsai and Philpot (1998) was used to transform the original spectra into the first- and second-spectra. Then, the correlation coefficients (r) between the spectral derivatives and in situ paddy rice LAI were calculated. The wavebands corresponding to the extreme values of r were selected for paddy rice LAI estimation.

2. Vegetation index-based method

All possible narrow-band normalized difference vegetation indices (NDVIs) and ratio vegetation indices (RVIs) were computed from the combinations of wavebands in the region of 350 to 2500 nm. Due to severe noises from the water absorption, the spectral regions of 1330 to 1480 nm, 1780 to 1990 nm, and 2400 to 2500 nm were excluded from the analysis. So, there were 1690 discrete narrow bands available for the computation of $N\times N$=2856100 NDVIs and RVIs. Coefficients of determination (R^2) between all possible two-narrow-band vegetation indices and measured LAI were acquired to generate a R^2 matrix ($N\times N$), which could be illustrated in a contour plot with several grades of R^2 values. The plots revealed a characteristic pattern of a number of 'hot spots' with relatively high R^2 values. According to the spectral distribution of 'hot spots', we determined several vegetation indices that produced the relatively high R^2 values than other vegetation indices in the immediate spectral vicinity. The wavebands, which constructed the better-performing vegetation indices, were selected as suitable wavebands for paddy LAI estimation.

3. Stepwise regression method

In order to select the optimal spectral wavebands

for paddy rice LAI estimation, a stepwise multiple regression analysis which related the measured paddy rice LAI to the hyperspectral reflectance data was also performed:

$$LAI = a_0 + \sum_{i=1}^{N} a_i R_i, \quad (1)$$

where LAI is measured rice LAI; R is reflectance in bands i (i=1 to N with N=1690); a_i is the coefficient for reflectance in band i; a_0 is the constant term. The regression model was constructed by adding one independent variable at a time. The first step was to choose the single waveband which was the best statistical predictor; and then the second independent waveband was added to the regression equation, which provided the best fit in conjunction with the first variable. Further variables were then added in the recursive way (Jacquemoud *et al*., 1995). According to the regression equation, the optimum wavebands were determined for paddy rice LAI estimation.

RESULTS

Optimal wavebands identification

At first, the suitable wavebands for rice LAI estimation were selected using the dataset from 2002 growing season with the following methods.

1. Optimal wavebands identified by correlation coefficient analysis

The first- and second-derivative spectra were correlated with measured rice LAI obtained on the eight sampling dates. The correlogram provided a clear picture of which wavebands produced the relatively high r values. Fig.1 illustrates the variations of correlation coefficients with wavelengths for the first- (Fig.1a) and second-derivative spectra (Fig.1b), respectively, using dataset acquired on July 23, 2002. The remaining seven correlograms are not presented here. From the figure, it is obvious that higher r values occur at some wavebands but not at other wavebands in the immediate spectral vicinity. So the wavebands with the relatively high r values were selected for paddy rice LAI estimation.

The wavebands selected through extreme values of correlation coefficients are shown in Table 2. The number of wavebands selected through the second-derivative spectra is greater than that selected using the first-derivative spectra. For the first-derivative spectra, some selected wavebands are the same or very similar for different sampling dates. For example, the red minimum is 675 nm for July 17, 677 nm for July 23, 676 nm for July 30, 680 nm for Aug. 22, 681 nm for Aug. 31, 676 nm for Sept. 20, and 674 nm for Oct. 3. For the second-derivative spectra, the selected wavebands also have a similar pattern as the case of the first-derivative spectra.

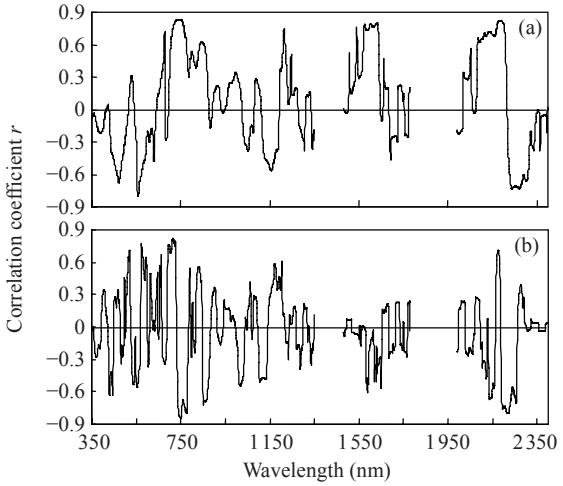

Fig.1 Correlograms between (a) the first- and (b) second-derivative spectra and measured LAI obtained on July 23, 2002

2. Optimal wavebands identified by vegetation indices

The R^2 matrix was generated from the linear regression models between measured paddy rice LAI and NBNDVI and NBRVI (NB=narrow band). Due to the symmetry of the R^2 matrix for NBNDVI, it is feasible to display the results only below the diagonal of the matrix (Fig.2a), while the R^2 matrix of NBRVI requires the whole matrix to be displayed (Fig.2b). Fig.2 is presented as an example, so more plots of R^2 matrices will not be given here. In this figure, several 'hot spots' can be observed. The wavelengths, which was used to construct the narrow-band vegetation indices in the central 'hot spots', were selected for estimation of paddy rice LAI. For example, the vegetation index $[(R_{864}-R_{715})/(R_{864}+R_{715})]$, which had the maximum R^2 value for all the nearby wavebands, was located near the central position in the 'hot spot' where wavelengths of the red edge (700~735 nm) coincided with narrow bands between 700 and 1150 nm (Fig.2a). The wavelengths 715 and 864 nm

Table 2 Wavebands selected using correlograms between the first- and second-derivative spectra and LAIs at different sampling dates

Sampling date	Derivative spectra	Wavebands (nm)
July 17 (n=60)	First	423, 462, 527, 553, 631, 675, 685, 736, 1153, 1214, 1509, 1541, 1600, 1594, 1634, 2052, 2084, 2116, 2239, 2269, 2302
	Second	397, 430, 445, 465, 514, 534, 557, 576, 596, 678, 711, 748, 1171, 2226
July 23 (n=60)	First	470, 557, 677, 735, 800, 844, 1159, 1215, 1508, 1542, 1576, 1603, 1633, 2054, 2181, 2236
	Second	430, 445, 497, 520, 538, 555, 573, 593, 664, 712, 725, 750, 777, 831, 852, 1018, 1172, 1203, 2145, 2175, 2195, 2219, 2256
July 30 (n=60)	First	475, 558, 676, 733, 845, 1161, 1215, 1245, 1509, 1540, 1579, 2122, 2190, 2308, 2333
	Second	430, 466, 520, 540, 556, 573, 593, 663, 677, 700, 748, 779, 832, 853, 1135, 1175, 1204, 1588, 2219, 2255
Aug. 5 (n=38)	First	472, 559, 880, 1542, 1572, 1637, 2118, 2202, 2275, 2301
	Second	428, 520, 544, 561, 580, 622, 663, 696, 715, 726, 747, 775, 1023, 1203, 1589, 1619, 2159, 2174, 2229, 2257
Aug. 22 (n=60)	First	453, 549, 615, 661, 680, 753, 823, 884, 940, 1215, 1308, 1575, 1640, 1660
	Second	413, 448, 459, 501, 525, 572, 602, 627, 642, 656, 679, 710, 725, 765, 776, 807, 833, 863, 897, 929, 978, 1031, 1071, 1103, 1170, 1203, 1591, 1632
Aug. 31 (n=60)	First	448, 493, 553, 647, 665, 681, 756, 825, 956, 1019, 1077, 1137, 1212, 1659, 2209
	Second	412, 448, 502, 520, 571, 602, 628, 640, 652, 656, 711, 725, 750, 760, 776, 806, 845, 895, 913, 979, 1007, 1032, 1044, 1069, 1104, 1160, 1203, 1527, 1592, 1612, 1663, 2160, 2190, 2258, 2304
Sept. 20 (n=24)	First	495, 558, 676, 687, 747, 858, 955, 1029, 1077, 1132, 1215, 1572, 1608, 1631, 2206
	Second	429, 461, 502, 520, 572, 638, 662, 701, 711, 720, 751, 799, 807, 845, 895, 910, 978, 1008, 1069, 1087, 1111, 1154
Oct. 3 (n=24)	First	386, 564, 674, 730, 861, 1086, 1161, 1215, 1556, 1573, 1626, 2267, 2325, 2370
	Second	431, 576, 654, 699, 716, 725, 744, 777, 804, 853, 900, 991, 1022, 1081, 1136, 1171, 1650, 2229, 2254, 2288, 2348

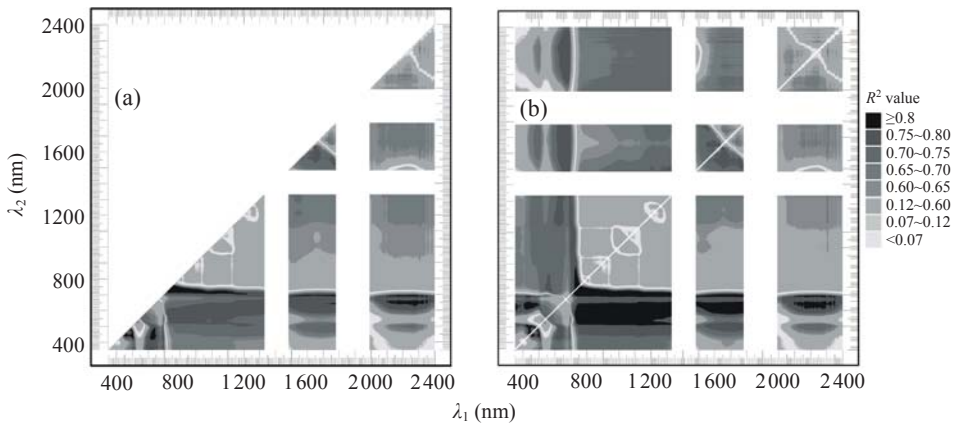

Fig.2 Distribution plot of coefficients of determination (R^2) for wavebands selection. The R^2 values were obtained from the linear fitted regressions between measured paddy rice LAI and all two-band combinations in (a) narrow-band NDVI vegetation index $[(R_{\lambda_2} - R_{\lambda_1})/(R_{\lambda_1} + R_{\lambda_2})]$ and (b) narrow-band RVI vegetation index $(R_{\lambda_2}/R_{\lambda_1})$ (λ is wavelength). The data used were acquired on Aug. 31, 2002

were selected as the suitable wavebands for LAI estimation. The procedure was repeated for each 'hot spot' in the R^2 distribution plots to select the various optimal wavebands.

In the study, seven narrow-band NDVIs were determined for different sampling dates (Table 3). Because of the symmetry of NDVI, the number of NDVIs was less than that of the RVIs which was 9 (Table 4). As can be seen from Tables 3 and 4, the most frequently occurring wavebands are located in the red-edge region, followed by the longer wavelength visible portion (550~700 nm) and a particular section of the short-wave infrared (1600~1650 nm). When comparing the wavebands selected using

NBNDVI with those selected using NBRVI, we found that there were several the same or nearly the same wavebands for both NBNDVI and NBRVI.

3. Optimal wavebands identified by stepwise regression

The stepwise regression models were applied to establish the relationship between measured paddy rice LAIs and the spectral wavebands. The stepwise criteria were: $P \leq 0.05$ for entry and $P > 0.10$ for removal. In general, as more wavebands are entered into stepwise regression equation, the coefficients of determination will increase, but the risk of overfiting the training data may also increase. So, to improve the generalization capability while minimizing the risk of model overfitting, the ratio of the number of independent variables or the number of bands to the total number of field samples should be limited within 0.15~0.20 (Thenkabail *et al*., 2000).

Table 3 Wavebands corresponding to seven best NBNDVIs for the linear NBNDVI models at different sampling dates

Sampling date	Wavelength	Waveband (nm)						
		Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7
July 17	λ_1	724	712	704	603	704	702	607
	λ_2	787	832	952	809	1318	1648	1592
July 23	λ_1	718	525	710	700	564	592	460
	λ_2	1020	607	1630	2194	1630	2194	471
July 30	λ_1	567	701	696	559	583	592	697
	λ_2	718	1630	2108	1630	2292	2017	788
Aug. 5	λ_1	458	602	590	453	731	614	611
	λ_2	477	2314	2002	480	1660	1530	2194
Aug. 22	λ_1	715	710	550	571	670	503	670
	λ_2	951	1218	817	1023	1629	1678	2314
Aug. 31	λ_1	722	714	439	441	694	690	650
	λ_2	853	995	629	479	1671	2233	2224
Sept. 20	λ_1	667	736	726	548	476	731	725
	λ_2	678	965	1123	963	507	893	1144
Oct. 3	λ_1	838	742	727	549	440	714	550
	λ_2	976	1246	1629	2194	514	2318	2320

NDVI vegetation index: $(R_{\lambda_2} - R_{\lambda_1})/(R_{\lambda_1} + R_{\lambda_2})$

Table 4 Wavebands corresponding to nine best NBRVIs for the linear NBRVI models at different sampling dates

Sampling date	Wavelength	Waveband (nm)								
		Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8	Index 9
July 17	λ_1	738	730	791	850	724	812	839	938	775
	λ_2	753	816	721	720	802	559	718	708	605
July 23	λ_1	726	722	716	708	523	574	901	1074	598
	λ_2	990	1233	1630	2189	613	2003	726	723	525
July 30	λ_1	689	617	519	514	593	517	691	578	721
	λ_2	1515	1515	1529	2366	1629	1320	979	721	565
Aug. 5	λ_1	674	669	508	667	666	668	508	617	1658
	λ_2	2086	1554	1515	2195	1201	1317	985	541	734
Aug. 22	λ_1	742	802	712	575	589	672	1043	922	1214
	λ_2	775	738	1062	923	1223	1628	714	556	569
Aug. 31	λ_1	715	704	581	699	632	693	639	634	803
	λ_2	864	1157	1031	1630	1630	2223	2208	449	733
Sept. 20	λ_1	666	736	845	971	739	727	983	893	1127
	λ_2	680	963	737	732	891	1152	546	731	719
Oct. 3	λ_1	727	825	743	441	550	860	977	1243	1629
	λ_2	1630	972	1247	514	1506	922	838	742	727

RVI vegetation index: $R_{\lambda_2} / R_{\lambda_1}$

For the 8 datasets from 2002 growing season, the obtained regression models explained 76%~88% of the LAI variability. From the regression equations, the important wavebands were also determined. These wavebands, listed in Table 5, are regarded as the most suitable wavebands for estimating paddy rice LAI. From Table 5, a shift in importance among wavebands is noted for different sampling dates, which may be attributed to the differences in growth stages.

Table 5 Wavebands selected using stepwise regression method at different sampling dates

Sampling date	Wavebands (nm)
July 17	384, 386, 437, 609, 612, 640, 692, 2049, 2337
July 23	711, 758, 969, 1127, 2001, 2304, 2307, 2326, 2334
July 30	529, 1505, 1508, 1650, 1658, 1662, 2284, 2333, 2347
Aug. 5	1128, 1648, 1677, 1690, 2346, 2348
Aug. 22	703, 710, 719, 723, 999, 1003, 1007, 1117, 1131
Aug. 31	509, 589, 591, 703, 713, 1124, 1137, 1140, 1147
Sept. 20	536, 1069, 1074, 2017
Oct. 3	813, 1230, 1235, 1298

4. Comprehensive analysis of the selected wavebands

Based on the same datasets collected during 2002 growing season, the wavebands, which appeared to be suitable for estimating rice LAI, were selected using three methods: correlation coefficient-based method, narrow-band vegetation index-based method, and stepwise regression method. Higher occurrence of a given waveband indicates the importance of the waveband in estimating paddy rice LAI. However, the occurrence positions of a given waveband might vary a little in a narrow range due to the effects of external factors such as growth stages, nitrogen treatment, canopy structure, etc. In addition, the higher correlation between bands in the immediate spectral vicinity also may result in a little change in the positions of the selected wavebands.

In the study, a comprehensive analysis of the selected wavebands was performed to determine the final optimum wavebands. At first, all the wavebands from the three methods were put together. Then, the frequencies of waveband occurrences were calculated in a narrow range (10 nm) centered at each waveband in the spectral region of 350~2500 nm. From Fig.3, it is clear that the most frequently occurring wavebands are centered at 554, 675, 723, and 1633 nm. They are followed by 444, 524, 576, 594, 804, 849, 974, 1074, 1219, 1510, and 2194 nm. The wavebands at 524, 576, 594, and 1510 nm correspond to notable slopes within the reflectance spectra (zones of significant change in reflectance). The wavebands at 804, 849, and 1074 nm correspond to local maxima within the high reflectance of vegetation in near-infrared. The wavebands at 444, 554, 675, 723, 974, 1219, 1633, and 2194 nm are closely associated with some useful spectral features as discussed blow.

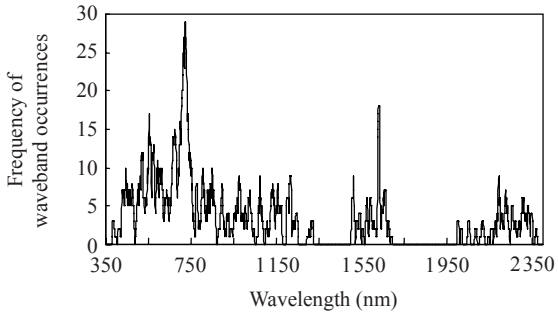

Fig.3 Frequency of the selected hyperspectral waveband occurrences in a 10 nm-range centered at each waveband in the spectral region of 350~2500 nm

It can be said that these wavebands contain an overwhelming fraction of the rice LAI information in the full spectrum, and they will be compared with the wavelengths of spectral features in the following subsection.

Feature extraction of spectra and comparison between feature positions and selected wavebands

1. Feature extraction of spectra

Feature extraction was performed on all the spectra from the eight campaigns, and statistical analysis of these features was carried out. There are 12 spectral features involved in the study as shown in Table 6.

The 'blue edge', 'yellow edge' and 'red edge' positions are determined by the maximum value of the first-derivative of the reflectance spectra in the ranges of 490~530, 540~580, and 680~760 nm, respectively. The 'green peak' and 'red trough' denote the green maxima and red minima of the reflectance spectra, respectively. The 'NIR peak 1', 'NIR trough 1', 'NIR peak 2', 'NIR trough 2', 'NIR peak 3', 'SWIR peak 1', 'SWIR peak 2' positions are located

Table 6 Variation and standard deviation of mean positions at 12 main spectral features for different sampling dates

Spectral feature	Variation of mean positions (nm)									Standard deviation of positions (nm)								
	July 17	July 23	July 30	Aug. 5	Aug. 22	Aug. 31	Sept. 20	Oct. 3	Ave*	July 17	July 23	July 30	Aug. 5	Aug. 22	Aug. 31	Sept. 20	Oct. 3	Ave
Blue edge	521.5	520.8	521.0	520.7	520.8	520.7	521.8	521.0	521.0	2.2	1.4	1.2	1.0	1.5	1.5	1.3	1.6	1.5
Green peak	555.4	554.6	554.6	555.0	555.4	556.2	554.9	556.8	555.4	1.9	1.4	1.1	0.9	1.7	2.4	0.9	1.1	1.4
Yellow edge	574.0	574.0	570.9	571.1	572.0	570.0	571.0	569.8	571.6	13.1	15.5	0.4	0.3	9.0	4.9	0.4	6.1	6.2
Red minimum	673.9	673.8	673.8	673.7	673.5	673.2	674.8	674.2	673.9	1.1	1.1	0.8	0.8	0.9	1.2	1.0	0.8	1.0
Red edge	711.6	721.3	727.0	726.1	729.2	724.3	719.9	721.9	722.7	10.0	9.5	4.0	4.2	8.2	13.8	1.9	10.2	7.7
NIR peak 1	807.0	821.8	870.7	837.8	902.1	905.7	888.9	884.8	864.9	1.6	24.1	28.0	31.1	15.2	16.6	8.6	22.5	18.4
NIR trough 1	969.1	970.5	972.3	970.7	967.6	971.4	968.9	981.4	971.5	11.2	7.9	6.8	5.6	7.0	8.5	6.5	4.9	7.3
NIR peak 2	1073.6	1071.0	1072.8	1074.2	1080.7	1083.7	1074.3	1080.4	1076.3	15.6	7.7	3.8	3.9	9.2	13.8	3.2	6.0	7.9
NIR trough 2	1173.4	1185.3	1194.4	1194.0	1195.1	1192.3	1191.0	1198.5	1190.5	21.7	16.1	6.5	7.3	3.0	11.1	10.4	1.7	9.7
NIR peak 3	1255.4	1261.2	1266.4	1266.6	1267.6	1272.0	1271.1	1274.4	1266.8	12.3	5.7	4.0	2.8	3.9	5.5	3.5	3.0	5.1
SWIR peak 1	1666.9	1667.4	1668.7	1669.1	1666.4	1666.0	1666.0	1665.0	1666.9	11.5	4.7	3.8	3.5	4.3	5.2	4.3	4.8	5.3
SWIR peak 2	2202.5	2213.5	2199.7	2201.0	2193.1	2210.9	2214.0	2213.2	2206.0	29.0	9.7	32.2	28.7	30.8	8.4	7.9	7.6	19.3

*Ave: Average

in the 700~950, 950~1050, 1050~1130, 1130~1225, 1225~1330, 1600~1750, and 2100~2300 nm, respectively.

All the mean feature positions and the standard deviations of them are shown in Table 6. It can be seen that 11 of the 12 spectral feature positions vary within 10 nm, especially for 'blue edge', 'green peak' and 'red minimum', which had a variation range less than 2 nm. 'NIR peak 1' has the most widely varying range of all the spectral features because it is sensitive to a number of external factors such as rice canopy structure, sun/viewing geometry, background effect, and so on. The changes in LAI cause less variations in feature positions than in magnitudes of reflectance at corresponding feature positions. So, waveband selection, which determines the most sensitive wavelengths, is very important for paddy rice LAI estimation.

2. Comparison between feature positions and selected wavebands

The mean feature positions were calculated from all the spectra. When comparing the 15 selected wavebands with the spectral feature positions, we found that most of the 15 selected wavebands had their counterparts in the positions of 12 main spectral features (Table 7), and there was a high degree of coincidence between the selected waveband and the spectral position of corresponding spectral feature, especially for wavebands at 524, 554, 576, 675, 723, 974, and 1074 nm, where the differences between feature positions and selected wavebands were less than 5 nm (Table 7). This suggests that these spectral features might contain more information than other parts of the spectrum, and shows the importance of spectral features in estimating paddy rice LAI. Since most of spectral positions make biophysical sense, the selected wavebands may also make physical sense, as discussed in the following section.

Table 7 Comparison between the mean spectral feature positions and selected wavebands

Spectral feature	Mean position (nm)	Selected waveband (nm)	Difference (nm)
Blue edge	521.0	524	3.0
Green peak	555.4	554	1.4
Yellow edge	571.6	576	4.4
Red minimum	673.9	675	1.1
Red edge	722.7	723	0.3
NIR peak 1	864.9	849	15.9
NIR though 1	971.5	974	2.5
NIR peak 2	1076.3	1074	2.3
NIR though 2	1190.5	1219	28.5
NIR peak 3	1266.8		
SWIR peak 1	1666.9	1633	33.9
SWIR peak 2	2206.0	2194	12.0
Others		444	
		594	
		804	
		1510	

DISCUSSION

Effect of methods on the selection of wavebands

The correlation coefficient-based method, vegetation index-based method, and stepwise regression

method were employed to select the suitable wavebands for rice LAI estimation. The different methods resulted in some differences in both the number and wavelength positions of the selected wavebands. The number of wavebands from the correlation coefficient-based method was larger than those from the other two methods. Notably, for a specific dataset, the wavebands selected by the different methods were not the same, especially for the stepwise regression method. That might be because only the stepwise regression method takes into account the relationships among spectral bands, while the other two methods just select the qualified wavebands regardless of the correlation between spectral wavebands.

Although the selected wavebands appear to be method-dependent, the combination of wavebands obtained from all the three methods may give a relatively reasonable result compared with that from only one method. So, in the study, the wavebands selected using all the three methods were combined to make the final determination of optimum wavebands.

Effect of growth stages on selection of wavebands

In the combined set, 15 wavebands were selected for estimating paddy rice LAI. The large number of the selected wavebands might be explained by the fact that the waveband positions sensitive to changes in LAI varied when factors such as background effect, nitrogen treatment, irrigation management, biophysical and biochemical parameters changed with growth stages and sampling dates. For example, in the early stage, reflectance at 675 nm was very sensitive to change in LAI when LAI was relatively low. But with the development of rice, reflectance at that waveband might approach a saturation level asymptotically for a certain range of LAI (Zhao *et al*., 2007), and was not as useful for prediction of LAI as the reflectance at slightly longer or shorter wavelengths, which indicated that a shift of sensitive waveband occurred. So, the difference in growth stages would cause a significant change in the selection of wavebands. To make a comprehensive selection of wavebands, the data from early to late growth stages were collected to select the suitable wavebands for paddy rice LAI estimation. This procedure ensured that the normally occurring variations due to growth stages would be captured by the selected wavebands set.

Physical meanings of the selected wavebands

Although the selection of wavebands for rice LAI estimation was purely based on the statistical methods, many of the optimum wavebands also make biophysical sense given what we know about plant chemistry, canopy structure, and plant reflectance. The most frequently occurring waveband is located in the region centered at 723 nm, which is in the red edge region. The red edge region is considered to contain more information on biomass quantity and LAI as compared to other parts of the electromagnetic spectrum (Mutanga and Skidmore, 2004). So that spectral region is important in LAI estimation, which is consistent with a few studies results (Thenkabail *et al*., 2000; 2004a). The 554 nm band is located at the green reflectance peak in the visible region, where chlorophyll reflection is maximal. The reflectance at 554 nm is more sensitive to LAI in a broad range than that in blue and red spectral regions. So, Gitelson *et al*. (1996) constructed a new vegetation index, green NDVI, by replacing the red band with the green band, and the index was proved to be an efficient measurement of vegetation greenness. The waveband at 675 nm corresponds to the absorption maximum of chlorophyll *a*, and the reflectance at 675 nm is very sensitive to LAI when LAI value is less than 2 (before canopy closure), but beyond that, it tends to reach a saturation level (Gitelson, 2004). The band at 1633 nm corresponds to the 1st overtone of N-H absorption and the 3rd overtone of NH_3+NH deformation (Huang *et al*., 2004), in the short wave infrared region, which is sensitive to lignin, and has been shown to be important for vegetation parameter retrievals in some studies (Brown *et al*., 2000).

Despite the absence of NIR wavebands in the first four most frequently occurring bands, the importance of wavebands in NIR region should not be neglected. Their absence was because that the selection of wavebands concentrated on several wavebands such as 804, 849, 974, 1074, and 1219 nm rather than on one or two wavebands. The reflectance at 804, 849, and 1074 nm are mainly determined by the arrangement of cells within the mesophyll layer of leaves and by canopy structure, especially the number of leaf layers along the vertical. The 974 and 1219 nm bands are associated with narrow water absorption regions, which are sensitive to moisture content in leaves (Thenkabail *et al*., 2004b).

The other selected wavebands also contain biophysical information and have been used in some studies. The waveband at 444 nm corresponds to the convoluted absorption bands of chlorophyll *b* and carotenoids (Sims and Gamon, 2002). The 524 and 576 nm bands are located beside the green peak. The band at 576 nm coincides with one of the absorption bands of chlorophyll *a* (Chappelle *et al*., 1993). The 1510 nm band is dominated by absorption features due to N-H bond stretching (Curran, 1989). The band at 2194 nm is located at the second peak of SWIR region, and related to protein absorption.

Comparison of the selected wavebands with those from other studies

A few studies have performed the selection of wavebands for vegetation monitoring (Thenkabail *et al*., 2000; 2004a; 2004b; de Jong *et al*., 2003; Becker *et al*., 2005). When comparing the 15 selected wavebands in this study with those selected in the previous studies, we found that most of them have the exact same or very similar wavelengths with those reported previously, especially for those bands at 554, 675, 723, and 1633 nm, as well as 524, 576, 804, 849, 974, 1074, 1219, 1510, and 2194 nm. The wavebands 444, 594, and 1510 nm selected in the study were seldom selected as useful wavebands for monitoring green vegetation in previous studies. This might suggest that those wavebands are uniquely applicable for rice LAI estimation.

CONCLUSION

In this paper, 15 wavebands were identified as optimal wavebands for estimating paddy rice LAI by using 3 different methods: correlation coefficient-based method, vegetation index-based method, and stepwise regression method. Of the selected wavebands, the most frequently occurring wavebands were centered at 554, 675, 723, and 1633 nm. They were followed by 444, 524, 576, 594, 804, 849, 974, 1074, 1219, 1510, and 2194 nm. Compared with the positions of known vegetation reflectance spectral features, most of the selected wavebands had their counterparts in these spectral features. These wavebands will be very useful for paddy rice LAI estimation using remotely sensed data.

References

Becker, B.L., Lusch, D.P., Qi, J., 2005. Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sens. Environ.*, **97**(2):238-248. [doi:10.1016/j.rse.2005.04.020]

Bouman, B.A.M., 1995. Crop modeling and remote sensing for yield prediction. *Neth. J. Agric. Sci.*, **43**:143-161.

Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. *Remote Sens. Environ.*, **71**(1):16-25. [doi:10.1016/S0034-4257(99)00035-8]

Casanova, D., Epema, G.F., Goudriaan, J., 1998. Monitoring rice reflectance at field level for estimating biomass and LAI. *Field Crop. Res.*, **55**(1-2):83-92. [doi:10.1016/S0378-4290(97)00064-6]

Chappelle, E.W., Kim, M.S., Mcmurtrey, J.E., 1993. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of concentration of chlorophyll *a*, chlorophyll *b*, and carotenoids in soybean leaves. *Remote Sens. Environ.*, **39**(3):239-247. [doi:10.1016/0034-4257(92)90089-3]

Confalonieri, R., Bocchi, S., 2005. Evaluation of CropSyst for simulating the yield of flooded rice in northern Italy. *Eur. J. Agron.*, **23**(4):315-332. [doi:10.1016/j.eja.2004.12.002]

Curran, P.J., 1989. Remote sensing of foliar chemistry. *Remote Sens. Environ.*, **30**(3):271-278. [doi:10.1016/0034-4257(89)90069-2]

Curran, P.J., Dungan, J.L., Peterson, D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the Kokaly and Clark methodologies. *Remote Sens. Environ.*, **76**(3):349-359. [doi:10.1016/S0034-4257(01)00182-1]

de Jong, S.M., Pebesma, E.J., Lacaze, B., 2003. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. *Int. J. Remote Sens.*, **24**(7):1505-1520. [doi:10.1080/01431160210145560]

Dobermann, A., Pampolino, M.F., 1995. Indirect leaf area index measurement as a tool for characterizing rice growth at the field scale. *Commun. Soil Sci. Plant Anal.*, **26**(9):1507-1523. [doi:10.1080/00103629509369387]

Eklundh, L., Harrie, L., Kuusk, A., 2001. Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sens. Environ.*, **78**(3):239-251. [doi:10.1016/S0034-4257(01)00222-X]

Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.*, **161**(2):165-173. [doi:10.1078/0176-1617-01176]

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.*, **58**(3):289-298. [doi:10.1016/S0034-4257(96)00072-7]

Grossman, Y.L, Ustin, S.L., Jacquemoud, J., 1996. Critique of

stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sens. Environ.*, **56**(3):182-193. [doi:10.1016/0034-4257(95)00235-9]

Haboudane, D., Mille, J.R., Pattey, E., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sens. Environ.*, **90**(3):337-352. [doi:10.1016/j.rse.2003.12.013]

Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.*, **93**(1-2):18-29. [doi:10.1016/j.rse.2004.06.008]

Jacquemoud, S., Verdebout, J., Schmuck, G., Andreoli, G., Hosgood, B., 1995. Investigation of leaf biochemistry by statistics. *Remote Sens. Environ.*, **54**(3):180-188. [doi:10.1016/0034-4257(95)00170-0]

Jiminez, L., Landgrebe, D., 1999. Hyperspectral data analysis and supervised feature reduction via project pursuit. *IEEE Trans. Geosci. Remote Sensing*, **37**(6):2653-2667. [doi:10.1109/36.803413]

Koger, C.H., Bruce, L.M., Shaw, D.R., Reddy, K.N., 2003. Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (*Ipomoea lacunosa*) in soybean (*Glycine max*). *Remote Sens. Environ.*, **86**(1):108-119. [doi:10.1016/S0034-4257(03)00071-3]

Lee, K., Warren, B.C., Robert, E.K., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sens. Environ.*, **91**(3-4):508-520. [doi:10.1016/j.rse.2004.04.010]

Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.*, **25**(19):3999-4014. [doi:10.1080/01431160310001654923]

Shibayama, M., Akiyama, T., 1989. Seasonal visible, near-infrared and mid-infrared spectra of rice canopies in relation to LAI and above-ground dry phytomass. *Remote Sens. Environ.*, **27**(2):119-127. [doi:10.1016/0034-4257(89)90011-4]

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.*, **81**(2-3):337-354. [doi:10.1016/S0034-4257(02)00010-X]

Strachan, I.B., Pattey, E., Boisvert, J.B., 2002. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sens. Environ.*, **80**(2):213-224. [doi:10.1016/S0034-4257(01)00299-1]

Thenkabail, P.S., Smith, R.B., Pauw, E.D., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.*, **71**(2):158-182. [doi:10.1016/S0034-4257(99)00067-X]

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Legg, C., Dieu, M.J.D., 2004a. Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sens. Environ.*, **90**(1):23-43. [doi:10.1016/j.rse.2003.11.018]

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Meer, B.V.D., 2004b. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.*, **91**(3-4):354-376. [doi:10.1016/j.rse.2004.03.013]

Tsai, F., Philpot, W., 1998. Derivative analysis of hyperspectral data. *Remote Sens. Environ.*, **66**(1):41-51. [doi:10.1016/S0034-4257(98)00032-7]

Vaesen, K., Gilliams, S., Nackaerts, K., Coppin, P., 2001. Ground-measured spectral signatures as indicators of ground cover and leaf area index: the case of paddy rice. *Field Crop. Res.*, **69**(1):13-25. [doi:10.1016/S0378-4290(00)00129-5]

Wang, F.M., Huang, J.F., Wang, X.Z., 2008. Identification of optimal hyperspectral bands for estimation of rice biophysical parameters. *J. Integr. Plant Biol.*, **50**(3):291-299. [doi:10.1111/j.1744-7909.2007.00619.x]

Xiao, X., He, L., Salas, W., Li, C., Moore, B., Zhao, R., Frolking, S., Boles, S., 2002. Quantitative relationships between field-measured leaf area index and vegetation index derived from VEGETATION images for paddy rice fields. *Int. J. Remote Sens.*, **23**(18):3595-3604. [doi:10.1080/01431160110115799]

Yang, X., Huang, J., Wang, J., Wang, X., Liu, Z., 2007. Estimation of vegetation biophysical parameters by remote sensing using radial basis function neural network. *J. Zhejiang Univ. Sci. A*, **8**(6):883-895. [doi:10.1631/jzus.2007.A0883]

Zhao, D., Huang, L., Li, J., Qi, J., 2007. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS-J. Photogramm. Remote Sens.*, **62**(1):25-33. [doi:10.1016/j.isprsjprs.2007.01.003]