ISPRS Journal of Photogrammetry and Remote Sensing 91 (2014) 72–84

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Different units of measurement of carotenoids estimation in cotton using hyperspectral indices and partial least square regression

Qiuxiang Yi [a,*], Guli Jiapaer [a], Jingming Chen [b], Anming Bao [a], Fumin Wang [c]

[a] Xinjiang Institute of Ecology and Geography Chinese Academy of Sciences, 818 Beijing South Road, Urumqi, Xinjiang 830011, PR China
[b] Department of Geography and Program in Planning University of Toronto, 100 St. George St., Toronto, Ontario M5S 3G3, Canada
[c] Institute of Hydrology and Water Resources, Zhejiang University, Hangzhou, Zhejiang 310058, PR China

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form 9 January 2014
Accepted 14 January 2014
Available online 15 March 2014

Keywords:
Carotenoids
Different units
Hyperspectral indices
Wavelength band selection
Partial least square regression
Cotton

ABSTRACT

This paper examines the use of canopy reflectance for different units of measurements of carotenoids estimation. Field spectral measurements were collected over cotton in different intensive field campaigns organized during the growing seasons of 2010 and 2011. Three units of measurement were evaluated carotenoids expressed as a mass per unit soil surface area (g/m^2), a mass per unit leaf area ($\mu g/cm^2$), and a mass per unit fresh leaf weight (mg/g), respectively. Four methods were compared to retrieve amount of carotenoids: stepwise multiple linear regression (SMLR), published spectral indices, band combination indices, and partial least square regression (PLSR). Results show that maximum sensitivity of reflectance to variation in different units of measurement of carotenoids was found in the green region at 515–550 nm, and at 715 nm and 750 nm regions in the far-red wavelengths. The predictive accuracies of Car (g/m^2), Car ($\mu g/cm^2$) and Car (mg/g) were tested on a validation data set and the results show that the highest R^2 values between estimations and observations were 0.468 for Car (g/m^2), 0.563 for Car ($\mu g/cm^2$), and 0.456 for Car (mg/g), with relative root mean square error (RMSE%, RMSE/mean) of 48.72%, 22.07% and 21.07%, respectively. Compared to Car (g/m^2) and Car (mg/g), the model performance indices for Car ($\mu g/cm^2$) show a high degree of consistency among the R^2 values and RMSE% and MAE% values. Further comparison were performed among the estimation accuracies of different unit carotenoids and among the different approaches used in the study by a paired-*t*-test. The results indicate that although the best estimation results for Car ($\mu g/cm^2$) and Car (mg/g) were both obtained based on PLSR, they can be estimated by all four adopted methods without significant differences ($P > 0.1$). Whereas for Car (g/m^2), the best estimation results were obtained based on published vegetation indices CIred-edge, which were significantly better than the estimation results based on SMLR ($P < 0.000$). In summary, the results of this study show that even the carotenoids expressed on concentration (mg/g) or content ($\mu g/cm^2$) basis at leaf level can be estimated with the same prediction accuracies to the carotenoids expressed as a mass per unit surface area (g/m^2) at canopy level using reflectance measurement at canopy level.

1. Introduction

Remote sensing techniques are a prominent tool for determining the plant physiological state (Zur et al., 2000). Leaf chemical constituents are determining indicators of plant physiology and other functional processes up to the ecosystem level. Among them, plant pigments are the most studied traits (Blackburn, 2007; Ustin et al., 2009). Carotenoids (Car) and chlorophylls (Chl) are the main pigments of green leaves (Gitelson et al., 2002). Measurement of total chlorophyll content (Cab) and carotenoids content (Car) has many applications in agriculture, ecology, and Earth science. The methods for remote estimation of chlorophylls a and b (Cab) have been quite well established (Gitelson et al., 2009; Haboudane et al., 2002; Le Maire et al., 2004, 2008; Malenovský et al., 2013; Schlerf et al., 2010; Zarco-Tejada et al., 2004), However, for carotenoids (Gitelson et al., 2002, 2006; Hernández-Clemente et al., 2012), are still not well developed.

Carotenoids are also important photosynthetic pigments (Demmig-Adams and Adams, 1992). Carotenoids have several physiological functions associated with photosynthesis, including structural role in the organization of photosynthetic membranes,

* Corresponding author. Tel.: +86 9917823131.
E-mail address: yiqx@ms.xjb.ac.cn (Q. Yi).

http://dx.doi.org/10.1016/j.isprsjprs.2014.01.004
0924-2716/

participation in light harvesting, energy transfer, as well as quenching of Chl excited states and photoprotection (Demmig-Adams et al., 1996; Edge et al., 1997; Horton et al., 1996; Peterman et al., 1997; Young and Frank, 1996).

During the last decade, some attempts have been undertaken to develop nondestructive techniques for Car assessment (Asner and Martin, 2008; Blackburn, 1998, 1999; Chappelle et al., 1992; Datt, 1998; Féret et al., 2008; Gitelson et al., 2002, 2006; Sims and Gamon, 2002; Thomas and Gausman, 1977; Zur et al., 2000). In these studies, different units have been used to express amounts of carotenoids, e.g. nmol/cm^2 (Gitelson et al., 2002), mg/cm^2 (Datt, 1998), μg/cm^2 (Hernández-Clemente et al., 2012; Zarco-Tejada et al., 2013). Data has been expressed in units of area, or mass. The most common three units of measurement were evaluated cartenoids expressed as a mass per unit surface area (g/m^2), a mass per unit leaf area (μg/cm^2), and a mass per unit fresh leaf weight (mg/g). These three units of measurement are different but are often used in remote sensing.

Grossman et al. (1996) showed that band selection using stepwise multiple linear regression, depended on whether the chemical data were expressed on a concentration (g/g) or content (g/m^2) basis. For a given chemical, similar bands were selected on a concentration or a content basis less than 6%. Datt (1998) indicated that the use of content (a mass per unit leaf area) rather than concentration (a mass per unit leaf mass) has been found to be more suitable for remote sensing applications because it is a better representation of amount of matter interacting with light per unit surface area. Furthermore, literatures also demonstrated that in order to be compatible with remotely sensed canopy reflectance, the leaf level chemical measurements were normally multiplied by biomass or LAI to be upscaled to the canopy level, and the canopy properties were expressed as a mass per unit surface area (g/m^2) (Homolová et al., 2013). However, it still remains unclear how significant differences are when a chemical expressed on concentration (mg/g) or content (μg/cm^2) basis at leaf level was estimated using reflectance measurement at canopy level, and few efforts have been made for comparing the accuracy of a chemical estimation based on different unit expressions, concentration (mg/g), content (μg/cm^2) and density (g/m^2), using experimental canopy reflectance measurement.

Some previous studies have compared retrieval capability for a given unit measurement of chemical at different remote sensing scales. For instance, Bian et al. (2013) have established the relationships between the concentrations (mg/g) of some key biochemical compounds of tea and the reflectance at three different levels: powder, fresh leaf and canopy levels. But until now, the relationships between the different units of measurement of carotenoids or any other biochemical compounds and the reflectance measurement at a given level (canopy-level in this study) have not been well discussed.

Additionally, most of the algorithms or vegetation indices for carotenoids reported in these literatures have been developed using leaf reflectance measurements carried out on a few deciduous and coniferous species from the northern hemisphere (Hernández-Clemente et al., 2012; Zarco-Tejada et al., 2013). Such algorithms and vegetation indices need to be applied to other species from different geographical and climatic regions of the world to see if they are indeed general. The cotton in Xinjiang, China, is different from the vegetation types usually described in remote sensing literature for carotenoids, where carotenoids content and chlorophyll content are relatively higher than forest sites. Furthermore, the continental arid climate of Xinjiang is characterized by aridity, rich sunlight and rare rainfall. The previous studies have proved that the photoprotection system plays a critical role in plants adapted to high temperature, high irradiation levels and drought (Faria et al., 1996; Hernández-Clemente et al., 2011). Very few high spectral resolution reflectance studies been carried out on cotton in China, and none of these have investigated the relationships between canopy reflectance and amount of carotenoids. This study is the first to investigate, in detail, the relationship between different units of measurement of carotenoids and experimented canopy reflectance measurement for cotton.

The present study investigated the relationship between cotton canopy reflectance measurement and the amounts of carotenoids by comparing carotenoids expressed as a mass per unit area (Car (g/m^2) and Car (μg/cm^2)) and as a mass per unit mass (Car (mg/g)) using stepwise multiple linear regression (SMLR), published vegetation indices, band selection indices and partial least square regression (PLSR) approaches. The main aims of the study were (1) to analyze the relationships between canopy hyperspectral reflectance and different units of measurement of carotenoids; (2) to compare the estimation accuracy of the different units of measurement of carotenoids; (3) to assess the prediction capability of SMLR, published vegetation indices, band-selection indices and PLSR in different units of measurement of carotenoids estimation.

2. Material and methods

2.1. Field data collection

The field experiment was conducted in June–September 2010 and 2011 at agricultural belts in Shihezi, Xinjiang, Northwest of China (85°59′E, 44°19′N), where cotton is the dominate crop. The continental arid climate of Xinjiang is characterized by aridity, rich sunlight and rare rainfall, with sharply defined seasons, high annual and diurnal fluctuations in air temperature, and low precipitation. Cotton is generally planted in April–May, and harvested in September–October. The whole growth period is about 180 days. The medium loam soil at the experiment area had the following properties: the field moisture capacity at depth of 10 cm was 0.33 g/cm^3, the volumetric water content at depth of 10 cm was 1.59 g/cm^3, and the saturation moisture content was 0.44 g/cm^3.

Field data collection were conducted in June–September 2010–2011 for eight times from seedling stage until boll stage (the actual dates were 12 June, 14 July, 8 August, and 8 September, 2010; 24 June, 11 July, 28 July, and 17 August, 2011, respectively). This procedure ensured that the normally occurring variation due to growth stage and measurement factors was included in the models, giving a more realistic basis for model development.

Canopy reflectance was obtained using an Analytical Spectral Devices, FieldSpec Full Range (ASD FieldSpec FR, Analytical Spectral Devices, Inc., Boulder, CO, USA) that acquires continuous spectra from 350 to 2500 nm. All canopy spectral measurements were taken on clear days with no visible cloud cover between 10:00 am and 14:00 pm (Beijing local time) since during this period the weather conditions and sunlight conditions were generally at the most stable state. In each plot, representative plants were selected for canopy spectral measurement. Taking into account the impact of soil background, in the first field campaign, the sensor head was placed about 0.3 m vertically above the canopies. This resulted in a spot size of 13 cm in diameter in each measurement since the ASD sensor has a field of view of 25 degrees. In the other field campaigns, the sensor head was placed approximately 1 m vertically above the canopies, leading to a spot size of approximately 44 cm in diameter on the canopies.

The reflectance of a white Spectralon panel (BaSO4) was measured before every reflectance was taken, then the reflectance was calculated as the ratio between energy reflected by the canopy and energy incident on the canopy. Every reflectance was an average of ten repeated scans that were automatically acquired by the FieldSpec.

Leaves were sampled from the top of the canopy, and were placed in paper bags and then stored in a freezer at −4 °C prior to carotenoids determination. Leaves healthy and homogenous color without visible symptoms of damage were used in the experiments. The wet lab extraction technique was used to determine the carotenoids concentration. A 1.8 cm circle was cut from each leaf sample. After weighing the fresh leaf weight, the samples were grinded in 10 ml extract (80% acetone: waterless-ethanol: distilled water = 4.5:4.5:1, volume proportion), and then another 10 ml extract were added for a total of 20 ml in each tube. Tubes were stored in the dark at 25 °C for 48 h prior to spectrophotometer measurements. Each sample for pigment determination was filtered and placed in a cuvette, and absorbance measured at 470 nm, 645 nm and 663 nm using BECKMAN722S spectrophotometer. Chlorophyll a (Ca) chlorophyll b (Cb) and carotenoids (Car) concentration were calculated using the extinction coefficient derived by Jiang and Zhu (1999), and absorbance measured at 470 nm, 645 nm and 663 nm with Eqs. (1)–(3):

$$\mathrm{Ca}\ (\mathrm{mg/L}) = 9.784 * A_{663} - 0.990 * A_{645} \quad (1)$$

$$\mathrm{Cb}\ (\mathrm{mg/L}) = 21.426 * A_{645} - 4.650 * A_{663} \quad (2)$$

$$\mathrm{Car}\ (\mathrm{mg/L}) = 4.695 * A_{470} - 0.268 * (\mathrm{Ca} + \mathrm{Cb}) \quad (3)$$

where A_x is the absorbance of the extract solution at wavelength x. The unit of the carotenoids was subsequently converted to concentration (a mass per unit leaf area, μg/cm^2), content (a mass per unit leaf fresh weight, mg/g) and density (a mass per unit soil surface area, g/m^2), respectively, using data on the volume of leaf pigment extract, the leaf fresh weight and the leaf disc area, with Eqs. (4) and (5):

$$\mathrm{Car}\ (\mathrm{mg/g}) = \mathrm{Car}\ (\mathrm{mg/L}) * \mathrm{vol.\ solvent}\ (\mathrm{ml})/1000/\mathrm{leaf\ fresh\ weight}\ (\mathrm{g}) \quad (4)$$

$$\mathrm{Car}\ (\mu\mathrm{g/cm^2}) = \mathrm{Car}\ (\mathrm{mg/g}) * 1000 * \mathrm{leaf\ fresh\ weight}\ (\mathrm{g})/\mathrm{leaf\ area}\ (\mathrm{cm^2}) \quad (5)$$

where vol. solvent is the volume of leaf pigment extract solution. If the concentration or content at leaf level is multiplied by biomass or LAI, one obtains a canopy value (i.e. canopy property, density) expressed as a mass per unit surface area (Homolová et al., 2013). In our study, the carotenoids density (Car (g/m^2)) was calculated on the basis of carotenoids concentration (Car (mg/g)) and the amount of fresh leaves per unit surface area m^2:

$$\mathrm{Car}\ (\mathrm{g/m^2}) = \mathrm{Car}\ (\mathrm{mg/g})/1000 * \mathrm{leaf\ fresh\ weight\ per\ unit\ surface\ area}\ (\mathrm{g/m^2}) \quad (6)$$

2.2. Published spectral vegetation indices for carotenoids estimation

The literatures have proposed many spectral indices for Car estimation using diverse combinations of wavelengths. However, which kind of indices is an appropriate one for cotton carotenoids has not been discussed. This part compared the performance of different indices applied on cotton canopy dataset and different units of measurement of carotenoids to aid in the selection of an adapted one. A detailed summary of the narrow-band vegetation indices in this study is shown in Table 1. Candidate Car optical indices have been grouped into two main categories based on the spectral region used: visible ratios (Gamon et al., 1992; Garrity et al., 2011; Gitelson et al., 2003a; Gitelson et al., 2006; Hernández-Clemente et al., 2011) and visible/NIR ratios (Blackburn, 1998; Chappelle et al., 1992; Datt, 1998; Merzlyak et al., 1999; Peñuelas et al., 1995). In the visible region, Gitelson et al. (2002) showed that the sensitivity of reciprocal reflectance to carotenoids content was maximal in a spectral range around 510 nm, and the 550 and 700 nm bands were used to remove the effect of chlorophyll, proposing the CCI as (1/515)–(1/550) and (1/515)–(1/700). Gamon et al. (1992) proposed PRI and calculated PRI with the 570 nm band as a reference (PRI570), and later with 515 nm band as a reference (PRI515) and has been found to minimize structural effects (Hernández-Clemente et al., 2011). In Table 1, Nos. 1–8 vegetation indices belong to this type. The main spectral bands proposed for carotenoids estimation in the visible/NIR region are based on band ratios in bands around $R_{700–770}$ and the green region (500 and 550 nm) (Chappelle et al., 1992; Merzlyak et al., 1999). A few examples are CRI700, and ratio indices R_{760}/R_{550} (Nos. 9–13 in Table 1). Other indices have been formulated as combinations of R_{800} or R_{860} with visible bands (470, 680, 635 nm) (Blackburn, 1998; Peñuelas et al., 1995). Some examples are $(R_{800} - R_{470})/(R_{800} + R_{470})$ (Blackburn, 1998), $((R_{800} - R_{750})/(R_{750} - R_{670}))$ (Dash and Curran, 2004) and R_{800}/R_{510} (Datt, 1998).

2.3. Specific waveband selection

As shown by Le Maire et al. (2004,2008) testing all possible combinations of wavebands may help to build better models. In this section, two families of indices were tested: the normalized difference (ND) (Eq. (7)) and the difference of reciprocal reflectance index (1/R), (Eq. (8)). The reciprocal index (1/R) is selected because some previous studies found that reciprocal reflectance at wavelengths out of the pigment absorption maxima could be successfully used for assessment of chlorophyll content in leaves of a number plant species (Gitelson et al., 1999, 2002; Gitelson and Merzlyak, 1994, 1996, 1997). This feature of leaf reflectance has been used to estimate Chl content in leaves accurately (Gitelson and Merzlyak, 1996, 1997). However, it should be examined whether this approach can be used for Car retrieval from the canopy reflectance spectra. We test this technique for carotenoids estimation. Moreover, ND indices and reciprocal indices normally showed better results than other indices such as simple reflectance and difference of reflectance.

$$\mathrm{NDVI} = (R_{\lambda1} - R_{\lambda2})/(R_{\lambda1} + R_{\lambda2}) \quad (7)$$

$$1/R = (1/R_{\lambda1}) - (1/R_{\lambda2}) \quad (8)$$

All possible two-band combinations of 650 (350–1000 nm) wavelengths (422,500 combinations) were used in Eqs. (7) and (8). A linear regression was performed in order to determine the correlation coefficient (R^2). All the R^2 values were plotted in a matrix plot and the plot revealed a characteristic pattern with a number of "hot spots" with relatively high coefficients of determination. These "hot spots" were selected by choosing the wavelength combinations that showed an R^2 exceed 0.5. The bandwidths for each of the selected spots were determined by fitting a rectangle that could the spot interest inside its limits. The advantages of the matrix plots are that they give a quick overview of thousands of wavelength combinations and make it possible to detect wavelengths of interest for further analysis Le Maire et al. (2004, 2008).

2.4. Partial least squares regression (PLSR)

PLSR is a bilinear calibration method using data compression by reducing the large number of measured collinear spectral variables to a few non-correlated latent variables. It is an extension of multiple linear regressions modeling that statistically determines the relative contribution of each chemical constituent to reflectance (Asner et al., 2009, 2011). The PLSR approach could model several

Table 1
Hyperspectral vegetation indices proposed in other studies.

No.	Vegetation index	Index ID	Formulations	References
1	Carotenoid concentration index	CRI550	$(1/R_{515}-1/R_{550})$	Gitelson et al. (2003a, 2006)
2	Carotenoid concentration index	CRI700	$(1/R_{515}-1/R_{700})$	Gitelson et al. (2003a, 2006)
3	Carotenoid concentration index	RNIR * CRI550	$((1/R_{515}-1/R_{550}) * R_{770})$	Gitelson et al. (2003a, 2006)
4	Carotenoid concentration index	RNIR * CRI700	$((1/R_{515}-1/R_{700}) * R_{770})$	Gitelson et al. (2003a, 2006)
5	Photochemical reflectance index	PRI	$(R_{570}-R_{530})/(R_{570}+R_{530})$	Gamon et al. (1992)
6	Modified photochemical reflectance index	PRIm1	$(R_{515}-R_{530})/(R_{515}+R_{530})$	Hernández-Clemente et al. (2011)
7	Carotenoid/chlorophyll ratio index	PRI * CI	$((R_{570}-R_{530})/(R_{570}+R_{530}) * (R_{760}/R_{700}-1))$	Garrity et al. (2011)
8			R_{515}/R_{570}	Hernández-Clemente et al. (2012)
9	Ratio analysis of reflectance spectra	RARS	(R_{746}/R_{513})	Chappelle et al. (1992)
10	Chlorophyll index red edge	CIred-edge	(R_{750}/R_{710})	Haboudane et al. (2002); Meggio et al. (2010)
11			R_{760}/R_{500}	Chappelle et al. (1992)
12	Plant senescencing reflectance Index	PSRI	$(R_{680}-R_{500})/R_{750}$	Merzlyak et al. (1999)
13	Modified chlorophyll absorption in reflectance index	MCARI	$((R_{700}-R_{670})-0.2(R_{700}-R_{550})) * (R_{700}/R_{670})$	Daughtry et al. (2000)
14	Green chlorophyll index	CIgreen	$((R_{800}/R_{550})-1)$	Gitelson et al. (2003a, 2003b, 2005)
15	Red-edge chlorophyll index	CIRed_edge	$((R_{800}/R_{750})-1)$	Gitelson et al. (2003a, 2003b, 2005)
16	Structure insensitive pigment index	SIPI	$(R_{800}-R_{445})/(R_{800}+R_{680})$	Gitelson and Merzlyak (1996); Peñuelas et al. (1995); Zarco-Tejada et al. (2004)
17	MERIS terrestrial chlorophyll index	MTCI	$(R_{800}-R_{750})/(R_{750}-R_{670})$	Dash and Curran (2004)
18	Pigment-specific simple ratio	PSSRc	(R_{800}/R_{470})	Blackburn (1998)
19	Pigment-specific normalized difference	PSNDc	$(R_{800}-R_{470})/(R_{800}+R_{470})$	Blackburn (1998)
20			R_{800}/R_{510}	Datt (1998)

response variables simultaneously while effectively dealing with strongly collinear and noisy independent variables (Wold et al., 2001). A large number of studies reported successful results and potential applications of PLSR in various disciplines (Hansen and Schjoerring, 2003; Hatonen et al., 1999; Lindberg et al., 1983). The PLSR algorithm has inferential capability, which can be used to model a possible linear relationship between carotenoids and measured canopy reflectance spectra. PLSR implicitly incorporates all wavebands in a linear fashion and finds a single equation typically via leave-one-out cross-validation (LOOCV) (Atzberger et al., 2010; Wold et al., 2001). The number of factors entering into a final model needs to be carefully chosen to avoid excessive overfitting (Zhao et al., 2013). In this study, reflectance data in wavelength range between 350 and 1000 were used for the PLSR analysis. A leave-one-out cross-validation (LOOCV) scheme was employed to determine the number of chosen factors by minimizing the predicted residual sums of squares (PRESS) (Chen et al., 2004). The basic PLSR algorithm will not be described in this paper, but further information on the PLSR model can be obtained in (Lindberg et al., 1983). Detailed description of the PLSR technique can be found in Geladi and Kowalski (1986); Wold et al. (2001). In present paper, the PLSR modeling was performed using MATLAB (version 7.8).

2.5. Quantitative of model performance

To quantify performance of carotenoid-reflectance models based on stepwise multiple linear regression (SMLR), published vegetation indices, band selection indices, and PLSR, various parameters between estimated and measured carotenoids were calculated: root mean square error (RMSE), relative RMSE (RMSE/mean, RMSE%), mean and relative mean absolute errors (MAE and MAE%). Both RMSE and MAE indicate absolute estimation errors. RMSE provides a direct estimate of the modeling error expressed in original measurement units (Kvalheim, 1987), and MAE indicates degree of agreement between measured and estimated values (Nash and Sutcliffe, 1970). The equations for RMSE, RMSE%, MAE and MAE% are given in Eqs. (9)–(12):

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}/n \tag{9}$$

$$\mathrm{MAE} = \sum_{i=1}^{n}|\hat{y}_i - y_i|/n \tag{10}$$

$$\mathrm{RMSE}\% = 100 * \mathrm{RMSE}/\bar{y}_i \tag{11}$$

$$\mathrm{MAE}\% = 100 * \mathrm{MAE}/\bar{y}_i \tag{12}$$

where $\hat{y}_i$ and y_i are the estimated and measured crop variables, respectively, and n is the number of samples.

Besides, coefficient of determination (R^2) value was calculated from each data set (calibration and validation). R^2 indicates strength of statistical correlation between measured and estimated values for models. Detailed descriptions and definitions of these model performance parameters are given in Cacuci (2005); Taylor (1997).

3. Results and discussion

3.1. Carotenoids content and composition

The data set (n = 193) was divided into a calibration set (n = 130) and a validation set (n = 63). To describe the distribution of different units carotenoids, a number of statistics such as maximum, minimum, mean and standard deviation were calculated. Basic statistics for the three different units of measurement of carotenoids were summarized in Table 2.

The amount of carotenoids ranges from 0.036 to 1.097 for Car (g/m^2), from 6.90 to 44.69 for Car ($\mu g/cm^2$), and from 0.256 to 1.009 for Car (mg/g), respectively. The wide range in the investigated crop variables can make the relationship between plant performance and reflectance measurements as realistic and universal as possible.

The distribution of each unit measurement of carotenoids is shown in Fig. 1. As seen, the distribution of Car (mg/g) can be fitted with a Gaussian distribution, while the Car ($\mu g/cm^2$) and Car (g/m^2) represented by a lognormal distribution. The most values were centered at 0–0.2 for Car (g/m^2), at 10–20 for Car ($\mu g/cm^2$), and at 0.3–0.5 for Car (mg/g).

The correlations between different units of measurement of carotenoids are shown in Fig. 2. Car (mg/g) and Car ($\mu g/cm^2$) were

Table 2
Basic statistics computed on the three units of measurement of Car.

Group	*n*	g/m^2			μg/cm^2			mg/g		
		Mean	Range	SD	Mean	Range	SD	Mean	Range	SD
All	195	0.271	0.036–1.097	0.207	18.54	6.90–44.69	7.206	0.471	0.256–1.009	0.152
Calibration	130	0.281	0.044–1.097	0.222	18.73	6.90–44.69	7.560	0.469	0.269–1.009	0.149
Validation	63	0.239	0.036–0.681	0.156	18.16	9.35–38.04	6.493	0.458	0.256–0.844	0.131

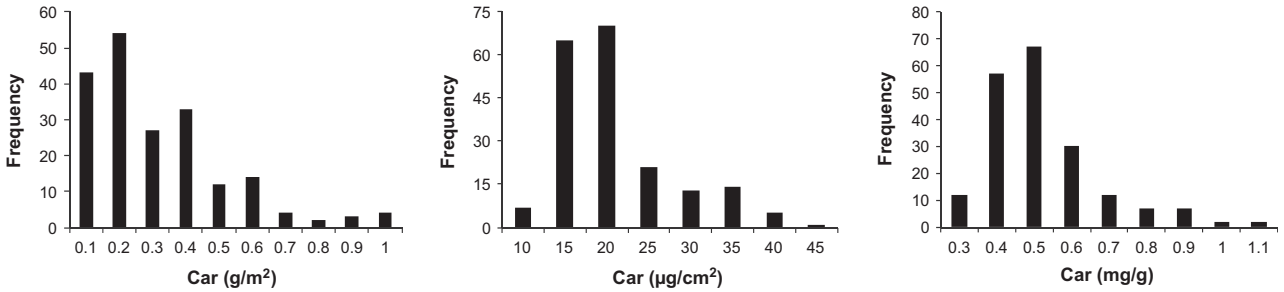

Fig. 1. Distributions of Car (g/m^2), Car (μg/cm^2) and Car (mg/g) in the entire data set.

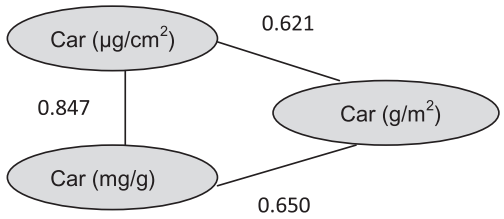

Fig. 2. Correlation (*r*) between different units of measurement of carotenoids.

highly correlated. A low correlation was observed between Car (μg/cm^2) and Car (g/m^2). The reason why Car (g/m^2) was more correlated to Car (mg/g) than to Car (μg/cm^2) probably because Car (g/m^2) was obtained using Car (mg/g) multiplied by leaf biomass per unit surface area.

3.2. Correlation between simple reflectance and carotenoids

In order to develop better algorithms for estimating carotenoids, the wavelength bands with maximum sensitivity to carotenoids were identified from correlogram plots showing the correlation coefficients between different units of measurement of carotenoids and canopy reflectance at all wavelengths. The correlogram is shown in Fig. 3.

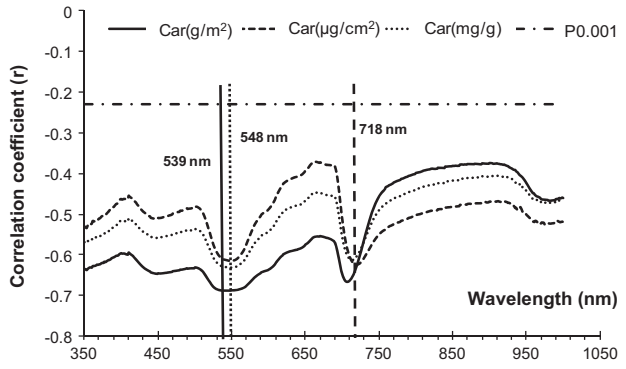

Fig. 3. Correlogram showing the correlation coefficient between amount of carotenoids and simple reflectance (350–1000 nm) at each wavelength band.

The correlation between individual wavelengths and amount of carotenoids were differed with carotenoids units expression, as weight-normalized values (i.e. mg carotenoids per g fresh leaf weight) or as area-normalized values (i.e. μg carotenoids per cm^2 fresh leaf area and g carotenoids per m^2 surface area). As can be seen, the correlation coefficients of three units carotenoids were visually similar in shape and absorption band positions, and all the coefficients were reached a significant at $P < 0.001$ (horizontal line represents the value significant at $P = 0.001$). The wavelengths of maximum sensitivity to carotenoids were indicated by big negative correlation coefficient, as reflectance decreases with increasing carotenoids at these wavelengths. The vertical lines show the wavelength locations that have the maximum correlation coefficients value: Car (g/m^2): 539 nm, Car (μg/cm^2): 718 nm, and Car (mg/g): 548 nm. In each case, there were two troughs in negative correlation, at about 550 and 710 nm, between reflectance and carotenoids. These two regions have been found to high sensitivity to pigment content in many plant species (Chappelle et al., 1992; Gitelson and Merzlyak, 1996, 2002). In the visible region, there was a stronger linear relationship between reflectance and Car (g/m^2). At wavelengths between about 350 nm and 720 nm, the relationships between reflectance and carotenoids were consistently greater for (g/m^2) (maximal correlation = −0.688 at 539 nm) and (mg/g) (maximal correlation = −0.625 at 548 nm) compared with (μg/cm^2) (maximal correlation = −0.630 at 718 nm). However, in the NIR region, at wavelengths between 720 nm and 1000 nm, the reflectance showed a higher correlation with Car (μg/cm^2).

In this section, stepwise multiple linear regression (SMLR) between 350 nm and 1000 nm with a maximum of two regressors was performed with canopy reflectance measurement and three units of measurement of carotenoids. The following algorithm equations (Eqs. (13)–(15)) were developed for the estimation of three units of measurement of carotenoids using SMLR:

$$\text{Car}\ (\text{g/m}^2) = 0.787 - 62.69 * R_{536} + 52.76 * R_{560} \quad (13)$$

$$\text{Car}\ (\mu\text{g/cm}^2) = 23.09 - 715.86 * R_{549} + 574.82 * R_{700}; \quad (14)$$

$$\text{Car}\ (\text{mg/g}) = 0.547 - 14.84 * R_{548} + 10.12 * R_{703} \quad (15)$$

Above equations shown that although the wavelengths selected by the SMLR were different for all three units carotenoids, the selected bands were all related to known absorption bands, around

Table 3
Correlation of vegetation indices with three units of measurement of carotenoids (n = 195).

No.	Vegetation index	Car (g/m^2)	Car (μg/cm^2)	Car (mg/g)
1	$(1/R_{515}-1/R_{550})$	0.567	0.289	0.405
2	$(1/R_{515}-1/R_{700})$	**0.709**	0.547	0.603
3	$((1/R_{515}-1/R_{550})*R_{770})$	0.224 (ns)	−0.058 (ns)	0.068 (ns)
4	$((1/R_{515}-1/R_{700})*R_{770})$	0.44	0.191 (ns)	0.295
5	$(R_{570}-R_{530})/(R_{570}+R_{530})$	0.428	0.559	0.473
6	$(R_{515}-R_{530})/(R_{515}+R_{530})$	0.144 (ns)	0.404	0.268
7	$((R_{570}-R_{530})/(R_{570}+R_{530})*(R_{760}/R_{700}-1))$	0.655	0.558	0.562
8	R_{515}/R_{570}	−0.026 (ns)	0.252	0.116 (ns)
9	(R_{746}/R_{513})	0.333	0.071 (ns)	0.183 (ns)
10	(R_{750}/R_{710})	0.316	0.076 (ns)	0.172 (ns)
11	R_{760}/R_{500}	0.33	0.068 (ns)	0.178 (ns)
12	$(R_{680}-R_{500})/R_{750}$	0.036	0.255 (ns)	0.137 (ns)
13	$((R_{700}-R_{670})-0.2(R_{700}-R_{550}))*(R_{700}/R_{670})$	−0.368	−0.467	−0.398
14	$((R_{800}/R_{550})-1)$	0.526	0.324	0.401
15	$((R_{800}/R_{750})-1)$	0.693	0.686	**0.672**
16	$(R_{800}-R_{445})/(R_{800}+R_{680})$	0.435	0.217 (ns)	0.312
17	$((R_{800}-R_{750})/(R_{750}-R_{670}))$	0.672	**0.706**	0.671
18	(R_{800}/R_{470})	0.391	0.14 (ns)	0.243
19	$(R_{800}-R_{470})/(R_{800}+R_{470})$	0.412	0.185	0.285
20	R_{800}/R_{510}	0.374	0.117 (ns)	0.225 (ns)

"ns" indicates not significant at P = 0.001. The values in boldface indicate the highest correlation coefficients for three units of carotenoids.

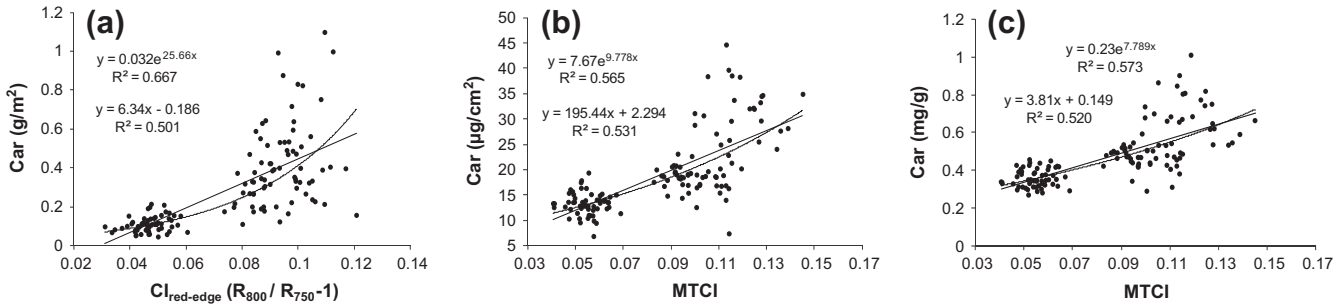

Fig. 4. The exponential and linear relationship between the best published indices and (a): Car (g/m^2); (b): Car (μg/cm^2); and (c): Car (mg/g).

535 nm, 550 nm and 700 nm. The first selected wavelengths for all three units carotenoids were all at green region, and the selected bands for Car (μg/cm^2) and Car (mg/g) were almost same. The calibration and validation statistics for Eqs. (13)–(15) were given in the Section 3.6.

3.3. Sensitivity of published vegetation indices to different unit carotenoids

The relationship between different units carotenoids and published vegetation indices was examined by linear regression. The correlation coefficients between these indices and carotenoids are given in Table 3. All correlation coefficient reported were significant at $P \leqslant 0.001$ and those not significant were indicated by "ns". The results show that the published vegetation indices based on band ratios in bands around $R_{700}-R_{770}$ and the green region (500 and 550 nm) (i.e. Nos. 9–13, Table 3) were not directly applicable to estimation of carotenoids in cotton. The indices such as $((1/R_{515}-1/R_{550})*R_{770})$, R_{515}/R_{570} and $(R_{800}-R_{445})/(R_{800}+R_{680})$ were also poorly correlated with carotenoids. The indices with the highest correlation coefficient (r = 0.709) for Car (g/m^2) was a reciprocal type index, i.e. $(1/R_{515}-1/R_{700})$, but it was not the optimal index for estimating of Car (μg/cm^2) and Car (mg/g). The normalized differences type index $(R_{800}-R_{750})/(R_{750}-R_{670})$ showed the best correlation with Car (μg/cm^2) and Car (mg/g), with r equal to 0.706 and 0.671, respectively. Among the three units of carotenoids, Car (g/m^2) and Car (μg/cm^2) showed a stronger and stable correlation with most of the published vegetation indices.

A preliminary comparison of these indices indicated that the CIred-edge index was the best index for Car (g/m^2), and MTCI was the best for both Car (μg/cm^2) and Car (mg/g) estimation. Scatterplots were constructed to examine the correlation between the selected vegetation indices and carotenoids, and to check for linear and nonlinear relations. The linear and exponential relationship between these indices and carotenoids were plotted in Fig. 4. As seen, more variation was explained if the selected indices were used in an exponential relationship to all three units carotenoids.

3.4. Band selection and development of indices

ND and $1/R$ narrow band vegetation indices were calculated from the measured canopy reflectance spectra using all possible two-band combinations. The R^2 between these narrow band vegetation indices and carotenoids were computed. The illustrations R^2 between the narrow band vegetation indices and carotenoids are shown in Fig. 5.

The "hot spots" were selected by choosing the wavelength combinations that showed an R^2 exceed 0.5 (Table 4). As can be seen, for Car (g/m^2), the difference of reciprocal reflectance type index performed better than NDVI type index, but for Car (μg/cm^2) and Car (mg/g), NDVI type index was more efficient. Band combinations provided by λ1 at red edge spectral region from 600 to 750 nm combined with λ2 in the green area 450–550 nm were rep-

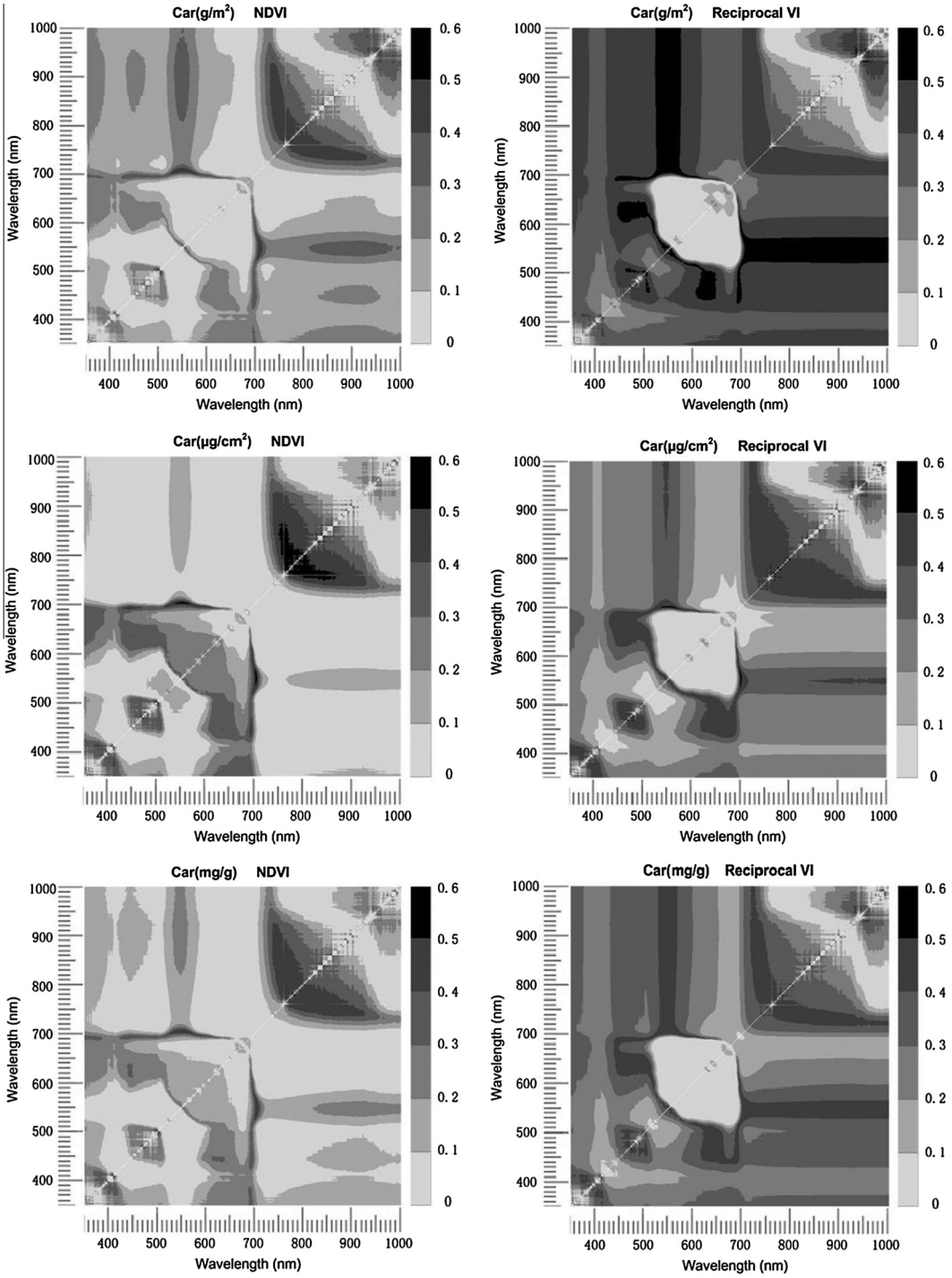

Fig. 5. Coefficient of determination (R^2) for the relation of all combinations of wavelengths used for a linear regression analysis of normalized vegetation index = $(R_{\lambda 1} - R_{\lambda 2})/(R_{\lambda 1} + R_{\lambda 2})$ and reciprocal index = $(1/R_{\lambda 1} - 1/R_{\lambda 2})$ against Car (g/m^2), Car (μg/cm^2), and Car (mg/g), respectively. A total number of $650 * 650 = 422500$ combinations were investigated.

resented in 50% of all selected hot spots for all three units of measurement of carotenoids. Another effective band combination with wavelengths was in the 750–800 nm and 780–860 nm. The red edge region was almost present for all three units measurement of cartenoids. So the determination coefficient (R^2) spectra of band combination indices at red-edge (750 nm) were detailed.

Spectra of the determination coefficient (R^2) of the relationship between the difference of reciprocal reflectance, $(1/R_{750})$–$(1/R_\lambda)$, normalized difference index, $(R_{750} - R_\lambda)/(R_{750} + R_\lambda)$, and different units of measurement of carotenoids are shown in Fig. 6. For Car (g/m^2), very high correlation was observed in green region (around 550 nm), reciprocal type index. However, for Car (μg/cm^2) and Car (mg/g), a broad flat maximum of the correlation (R^2 = 0.4–0.5) was observed in the NIR region of the spectrum between 800 and 900 nm, both in reciprocal type index and the normalized difference type index. For all three units measurement of caroteonids, correlations in the 600–700 nm range were lower than other regions.

Based on R^2 values in the 2-D correlation plots, band combinations that formed the best indices were determined for Car. The

Table 4
Band width of hot spots with the coefficient of determination bigger than 0.5 ($R^2 > 0.5$).

Units	Band width	1/R			NDVI	
		Hot spot1 (nm)	Hot spot2 (nm)	Hot spot3 (nm)	Hot spot1 (nm)	Hot spot2 (nm)
Car (g/m^2)	λ1	500–510	600–650	700–1000		
	λ2	460–500	450–510	520–570		
Car (μg/cm^2)	λ1	797–799			750–800	700–705
	λ2	800–801			780–860	535–565
Car (mg/g)	λ1				750–780	410–411
	λ2				800–870	407–409

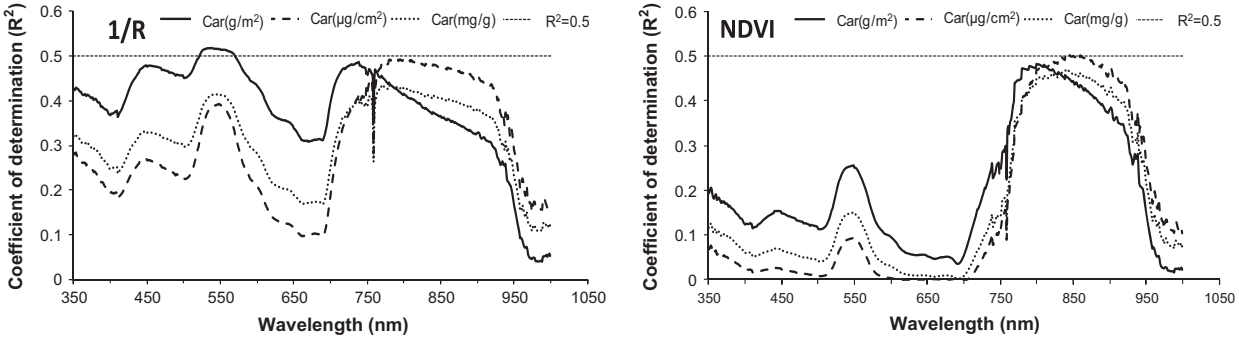

Fig. 6. The spectra of the determination coefficient (R^2) of the relationship (a) between $(1/R_{750})$–$(1/R_\lambda)$ and carotenoids; and (b) between $(R_{750} - R_\lambda)/(R_{750} + R_\lambda)$ and carotenoids.

Table 5
Band positions and R^2 values between the optimal indices and three different units of measurement of carotenoids.

Units	Optimal indices	R^2
Car (g/m^2)	$1/R_{532} - 1/R_{707}$	0.575
Car (μg/cm^2)	$(R_{800} - R_{762})/(R_{800} + R_{762})$	0.584
Car (mg/g)	$(R_{800} - R_{762})/(R_{800} + R_{762})$	0.545

best performing indices and the band positions are tabulated in Table 5.

Both linear ($y = ax + b$) and exponential ($y = a * \exp^{bx}$) fitting procedures were tested using these new indices and their equations and the values of R^2 are shown on the graphs (Fig. 7). Same to the results of published vegetation indices, more variation was explained if the selected indices were used in an exponential relationship to all three units carotenoids. Among the three units carotenoids, the coefficient of determination for the relationship between the optimal band selection index $(1/R_{532})$–$(1/R_{707})$ and Car (g/m^2) was greater than the coefficients of determination for Car (μg/cm^2) and Car (mg/g).

3.5. Performance of partial least square regression (PLSR)

The relationship between the numbers of latent variables and the prediction residual error sum of squares (PRESS) statistics were shown in Fig. 8. It can be learned that the optimal number of latent variables in the final models for Car (g/m^2), Car (μg/cm^2) and Car (mg/g) were 10, 8 and 9, respectively (see Fig. 8).

The contribution of each wavelength can be visualized by analysis of the computed loading weights (LW) (Fig. 9). Computed LWs stress further the results of 2-D correlograms analysis that spectral bands useful for carotenoids were those in 750, 715 and 550 nm regions. These zones showed a significant peak. The first three LWs showed some common trends, reflecting the relationship between the spectral data and the canopy. Two zones at approximately 550 and 700–740 nm of major importance for the PLSR models could be identified.

Calibration equation coefficients (*B*-coefficients) were also used to determine the importance of spectral bands (Haaland and Thomas, 1988; Hansen and Schjoerring, 2003; Wold et al., 2001; Zhao et al., 2013) (Fig. 10). In the PLSR model, *B*-coefficients represent the contribution of each predictor (waveband) to the model. The closer to zero a coefficient is, the less useful the associated band is. Strong co-variation between Car and canopy reflectance in the green (520–550 nm) and red-edge regions (680–750 nm) can also be visualized in the form of negative and positive peaks of the regression coefficient *B*.

The accuracy of the estimated carotenoids values and the performance of predictive models were deduced based on the performance statistics derived from measured and estimated carotenoids values and on the calculated model performance indices (Table 6). The results show that accurate predictions of cartenoids can be made based on PLSR models. As can be seen, the difference in R^2 values inside each of the three units measurements of carotenoids were very small (2–3%). The best coefficient of determination (R^2) was obtained for estimates of Car (μg/cm^2). The highest RMSE% and MAE% values were obtained for Car (g/m^2). The high R^2 value indicates the strong linear relationship between the measured Car (g/m^2) and canopy reflectance, but the high RMSE% and MAE% confirm the large difference between measured and estimated Car (g/m^2) values. For Car (mg/g), the smallest R^2 value, RMSE%, and MAE% were all obtained. This discrepancy (poor correlationship but low estimation errors) probably suggests non-linear relationship between Car (mg) and canopy reflectance, but the PLSR made a good linear fit. Among these three units measurement of carotenoids, the model performance indices for Car (μg/cm^2) has a high degree of agreement among R^2, RMSE% and MAE%.

3.6. Comparison of the estimation accuracies of different units carotenoids

The predictive accuracies of SMLR, published vegetation indices, optimal waveband selection indices and PLSR models were assessed against independent validation data sets ($n = 63$) not

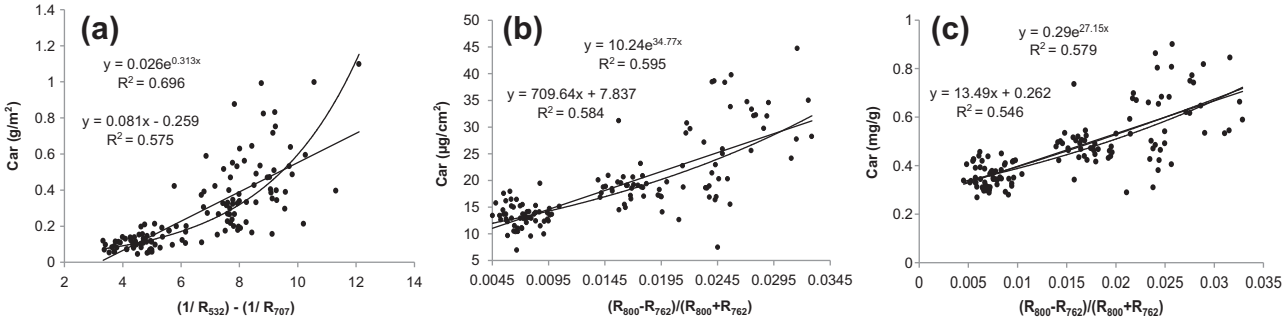

Fig. 7. The exponential and linear relationship between the best band selection indices and (a): Car (g/m^2); (b): Car (μg/cm^2); and (c): Car (mg/g).

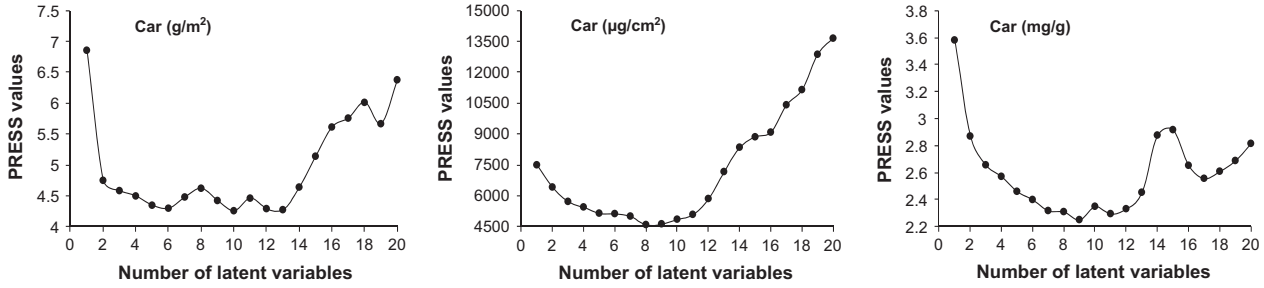

Fig. 8. The relationships between the numbers of latent variables and the predicted residual sums of squares (PRESS).

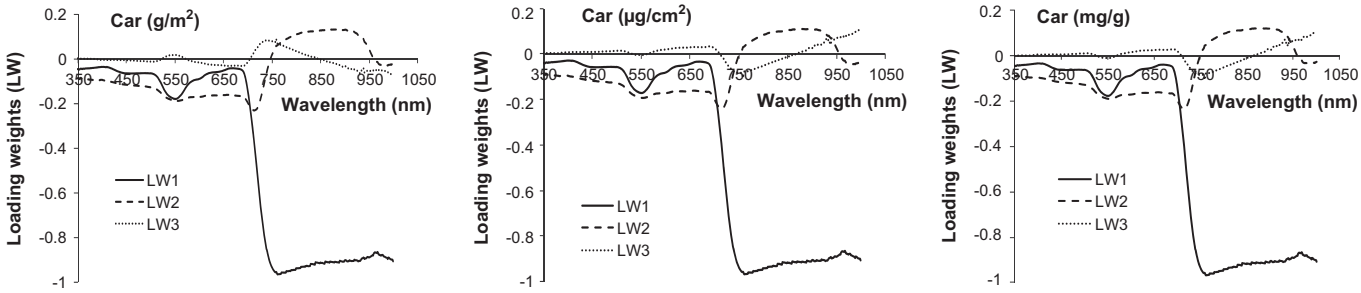

Fig. 9. Loading weights (LW) for the PLSR models calibrated to estimate to Car (g/m^2), Car (μg/cm^2), and Car (mg/g). High numerical values of regression coefficients indicate high importance of the reflected wavelength in the PLSR analysis.

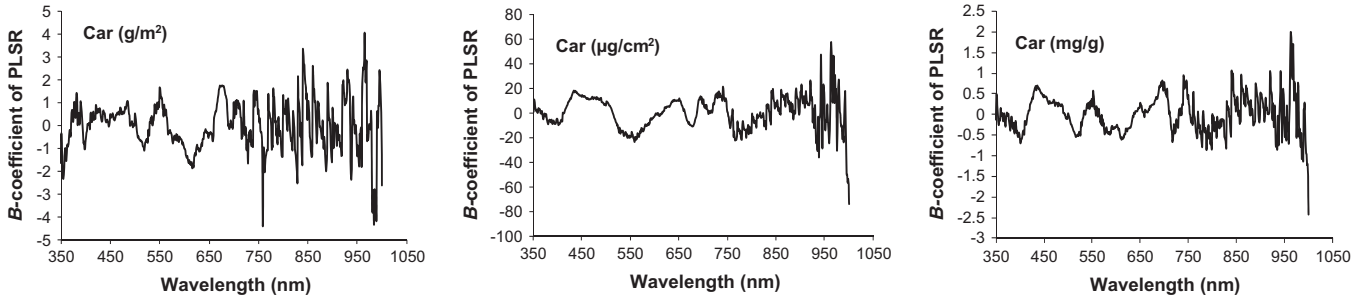

Fig. 10. Regression coefficients of *B* derived from PLSR calibration model.

concluded in the calibration phase, and the results were also compared with calibration data set ($n = 130$). The values of R^2, RMSE, RMSE%, MAE and MAE% are given in Table 7. Because the model performance indices for Car (μg/cm^2) has a high degree of consistency, the values of R^2, RMSE% and MAE% for Car (g/m^2) and Car (mg/g) were all compared with those for Car (μg/cm^2). The differences between R^2, RMSE% and MAE% for Car (g/m^2) and Car (mg/g) and those for Car (μg/cm^2) are shown in brackets (Table 7).

For calibration data set, the average R^2 between estimations and observations was 0.554 for Car (g/m^2), 0.592 for Car (μg/cm^2), and 0.547 for Car (mg/g). In generally, higher explanations of variation (R^2 values) were obtained by expressing the carotenoids on the basis of content (g/m^2) or (μg/cm^2) than by expressing the data on the basis of concentration (mg/g). The highest R^2 values were always obtained for Car (μg/cm^2) by all four methods. The differences of R^2 values between Car (g/cm^2) and Car (μg/cm^2) and

Table 6
Performance statistics of PLSR predictive models.

	R^2	RMSE	RMSE%	MAE	MAE%
Car (g/m^2)	0.645	0.132	46.95	0.094	33.52
Car (μg/cm^2)	0.665	4.359	23.27	3.24	17.29
Car (mg/g)	0.635	0.089	19.07	0.061	12.96

between Car (mg/g) and Car (μg/cm^2) decreased 3.8% and 4.5%, respectively. The highest R^2 values of three units carotenoids were all obtained by PLSR, with R^2 = 0.665 for (μg/cm^2), R^2 = 0.645 for (g/m^2), and R^2 = 0.635 for (mg/g), respectively (Table 7). However, the distributions of RMSE% and MAE% values for three units carotenoids were not coincident with R^2 values. As seen, the average relative root mean square errors (RMSE%) were 53.27% for Car (g/m^2), 25.79% for Car (μg/cm^2), and 21.11% for Car (mg/g) with average MAE% of 35.98%, 18.32% and 15.03%, respectively. Compared to the results yielded for Car (μg/cm^2), the average RMSE% and MAE% for Car (g/m^2) increased 27.48% and 17.66%, while for Car (mg/g), both RMSE% and MAE% decreased 4.68% and3.29%, respectively. The lower R^2, RMSE% and MAE% values were all observed for Car (mg/g). The relatively low R^2 value between measured and estimated Car (mg/g) values probably indicated poor linear relationship between the measured Car (mg/g) and canopy reflectance. Whereas, the low RMSE% and MAE% values suggested considerable similarity between measured and estimated Car (mg/g). Just as mentioned before, this discrepancy (poor correlationship but low estimation errors) probably suggests non-linear relationship (e.g. exponential relationship mentioned in Sections 3.4 and 3.5) between Car (mg/g) and canopy reflectance, but the algorithm made a good linear fit.

For validation data set, the variation tendency of all model performance indices was a little different from that for calibration data set. First, the highest R^2 values by different methods were not always obtained for Car (μg/cm^2), especially when published indices and SMLR were used (Table 7). Also worth noting, the highest RMSE% and MAE% values were still always obtained by Car (g/m^2). This result indicated the large difference between measured and estimated Car (g/m^2) values. Second, for the calibration data set, the highest R^2 and the lowest RMSE% and MAE% for Car (g/m^2), Car (μg/cm^2) and Car (mg/g) were all derived from PLSR. However, the comparison of performances obtained with validation data set shows that PLSR cannot lead to the highest R^2 value for Car (g/m^2) anymore. PLSR and published vegetation indices performed better than those optimal band selection indices, which was especially true for Car (g/m^2) and Car (mg/g). Those results showed that the main disadvantage of band selection indices probably was their empirical base, which can result in a lack of generality (Féret et al., 2011). Finally, the summary statistics (see Table 7) showed a high degree of agreement among the R^2 values and RMSE% and MAE% values for Car (μg/cm^2) for both calibration and validation data sets.

The algorithm equations were used to predict the carotenoids with reflectance measurement or vegetation indices calculated from the validation samples. The good agreement between the estimated and measured values of carotenoids was indicated by the scatterplots in Fig. 11. The values of slope and intercept were shown on the graphs. Ideally, when the slope = 1 and intercept = 0, it should be a perfect match.

The results in Fig. 11 and Table 7 both indicated that better results were obtained by expressing the carotenoids on the basis of content (g/m^2) or (μg/cm^2) than by expressing the data on the basis of concentration (mg/g). As can be seen, the highest slope value was obtained by Car (g/m^2) using band selection VI approach, with slope = 0.701. However, it was very apparent from Fig. 11, although the estimated and measured Car (g/m^2) showed a perfect linear relationship, the estimated Car (g/m^2) values were normally greater than measured Car (g/m^2), which could explain why the higher R^2 values and higher RMSE% and MAE% values were all observed for Car (g/m^2) as indicated in Table 7. Besides, it was worth noting that carotenoids were generally underestimated on higher values. These saturation-like patterns were found in the relationships between all three units carotenoids and vegetation indices or spectral reflectance. The same saturation effects were also found in the relationship between leaf carotneoids content and canopy R_{515}/R_{570} for LAI values above 3 by Zarco-Tejada et al. (2013). More investigations are needed to indentify the saturation relationships between carotneoids and vegetation indices.

In addition, a detailed comparison of how significant the differences are among the estimation accuracies of those three units carotenoids was performed by a paired-*t*-test (see Table 8). Paired-*t*-test (α = 0.05) was performed on the relative residuals val-

Table 7
Performance statistics of predictive models.

Methods	Data sets	Calibration (n = 130)					Validation (n = 63)				
	Units	R^2	RMSE	RMSE%	MAE	MAE%	R^2	RMSE	RMSE%	MAE	MAE%
SMLR	(g/m^2)	0.521 (−2.8%)	0.153	54.54 (27.53)	0.111	39.49 (19.32)	0.431 (2.2%)	0.139	53.10 (24.83)	0.107	40.92 (21.42)
	(μg/cm^2)	0.548	5.060	27.01	3.778	20.17	0.410	5.134	28.27	3.541	19.50
	(mg/g)	0.516 (−3.2%)	0.103	21.96 (−5.05)	0.077	16.50 (−3.67)	0.346 (−6.4%)	0.129	27.23 (−1.04)	0.086	18.18 (−1.32)
Published VI	(g/m^2)	0.465 (−8.8%)	0.164	58.50 (31.24)	0.105	37.41 (19.13)	0.468 (4.3%)	0.115	48.72 (21.83)	0.088	37.30 (19.05)
	(μg/cm^2)	0.553	5.105	27.25	3.425	18.28	0.426	4.800	26.88	3.260	18.25
	(mg/g)	0.523 (−3.0%)	0.103	21.96 (−5.30)	0.072	15.37 (−2.91)	0.451 (2.6%)	0.099	21.55 (−5.33)	0.069	15.12 (−3.14)
Band selection indices	(g/m^2)	0.559 (−4.4%)	0.149	53.08 (27.44)	0.094	33.51 (15.99)	0.390 (−7.0%)	0.145	60.92 (34.68)	0.098	40.99 (23.89)
	(μg/cm^2)	0.603	4.803	25.64	3.282	17.52	0.459	4.685	26.24	3.053	17.10
	(mg/g)	0.543 (−6.0%)	0.101	21.45 (−4.19)	0.072	15.29 (−2.23)	0.365 (−9.4%)	0.108	23.59 (−2.65)	0.073	16.04 (−1.06)
PLSR	(g/m^2)	0.645 (−2%)	0.132	46.95 (23.68)	0.094	33.52 (16.23)	0.466 (−9.7%)	0.129	53.63 (31.56)	0.010	41.37 (25.35)
	(μg/cm^2)	0.665	4.359	23.27	3.24	17.29	0.563	3.99	22.07	2.89	16.02
	(mg/g)	0.635 (−3%)	0.089	19.07 (−4.2)	0.061	12.96 (−4.33)	0.456 (−10.7%)	0.097	21.07 (−1.00)	0.068	14.69 (−1.33)

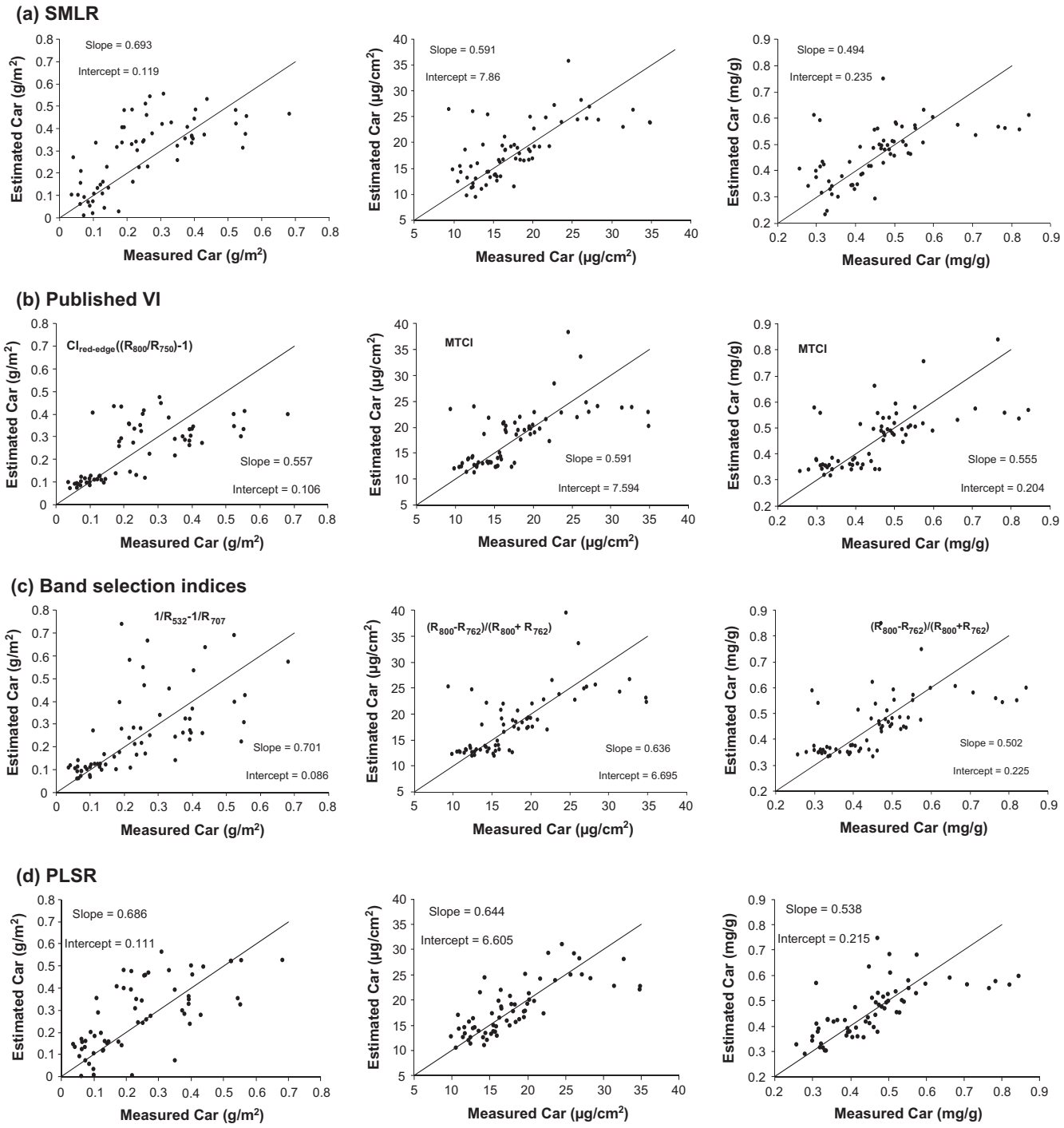

Fig. 11. Scatter plots of measured vs. estimated Car (g/m^2), Car (μg/cm^2) and Car (mg/g) derived from (a) SMLR; (b) the best published vegetation indices; (c) the best band selection indices; and (d) PLSR.

Table 8
P-values of paired-*t* test between the relative residuals of Car (g/m^2), Car (μg/cm^2) and Car (mg/g).

	SMLR		Published VI		Band selection VI		PLSR	
	Car (μg/cm^2)	Car (mg/g)	Car (μg/cm^2)	Car (mg/g)	Car (μg/cm^2)	Car (mg/g)	Car (μg/cm^2)	Car (mg/g)
Car (g/m^2)	0.020	0.010	0.100	0.035	0.056	0.039	0.002	0.002
Car (μg/cm^2)		0.493		0.574		0.67		0.97

ues from Car (g/m^2), Car (μg/cm^2) and Car (mg/g). It can be learned from Table 8, the differences of estimation accuracies between Car (μg/cm^2) and Car (mg/g) were not significant no matter what kind methods were used. Whereas, the differences of relative residuals values between Car (g/m^2) and Car (μg/cm^2) and between Car (g/m^2) and Car (μg/cm^2) were always significant ($P < 0.05$), which

Table 9
P-values of paired-*t* test between estimated carotenoids values by SMLR (M1), Published VI (M2), Band selection indices (M3) and PLSR (M4).

	Car (g/m^2)			Car (μg/cm^2)			Car (mg/g)		
	M2	M3	M4	M2	M3	M4	M2	M3	M4
M1*	0.000	0.006	0.184	0.830	0.245	0.912	0.985	0.356	0.614
M2		0.423	0.000		0.055	0.787		0.423	0.459
M3			0.042			0.350			0.192

* M1: SMLR; M2: published VI; M3: band selection VI; M4: PLSR.

was especially true when PLSR were used. These results indicated that Car (μg/cm^2) and Car (mg/g) can be equally estimated without significant differences, but the estimation accuracies of Car (g/m^2) were significantly influenced by the used methods.

Finally, the performances of individual models based on different methods have been assessed using statistical parameters described in Table 7, but the performance differences among those adopted methods have not been quantified. In order to statistically compare the differences of all four methods in the estimation of different units carotenoids, a paired-*t*-test ($\alpha = 0.05$) was also performed on estimated Car (g/m^2), Car (μg/cm^2) and Car (mg/g) values obtained by SMLR (M1), published vegetation indices (M2), band selection vegetation indices (M3), and PLSR (M4) (see Table 9).

The results in Table 9 showed that four methods can be equally used for Car (μg/cm^2) and Car (mg/g) without significant differences ($P > 0.05$). Whereas, the estimation results for Car (g/m^2) were significant different when different methods were used. Combined with the results in Table 7, it can be inferred that the estimation results of Car (g/m^2) based on published vegetation indices (M2) were significantly better than the estimation results based on SMLR (M1) and PLSR (M4) ($P < 0.000$). These results indicated that the estimation accuracies of Car (g/m^2) were strongly influenced by what kind of methods was used.

4. Conclusions

This paper examines SMLR, published vegetation indices, band selection indices and PLSR modeling for estimation of carotenoids on the basis of content (g/m^2) and (μg/cm^2), and on the basis of concentration (mg/g), from canopy spectral data obtained at experiment. The focus of the study was to systematically compare the accuracy of carotenoids estimation based on different unit expressions at leaf level and canopy level. The results allow drawing the following conclusions:

(1) The relationship between the remotely sensed observations and the amounts of carotenoids was investigated by comparing carotenoids expressed as a mass per unit area (Car (g/m^2) and Car (μg/cm^2)) and as a mass per unit mass (Car (mg/g)) with (i) reflectance between 350 and 1000 nm; (ii) all possible combinations of wavebands between 350 and 1000 nm; (iii) published vegetation indices and (iv) partial least squares regression. All the results indicated that spectral bands useful for carotenoids were those in 515–550 nm, 715 and 750 nm regions, which matched wavelengths described by previous studies. Furthermore, the differences of the correlationship between these three units carotenoids and canopy reflectance were small, normally the coefficients of correlation for Car (g/m^2) and Car (μg/cm^2) were slightly greater than those for Car (mg/g). Additionally, the band combination of the differences of reciprocal reflectance type indices were found much more efficient than the normalize indices for Car (g/m^2), while for Car (μg/cm^2) and Car (mg/g), the normalized indices performed better.

(2) The assessment of prediction models suggested that no matter carotenoids expressed as concentration (mg/g) or content (μg/cm^2) at leaf level or content at canopy level (g/m^2), all can be estimated by reflectance measurement at canopy level without significant differences. Compared to Car (g/m^2) and Car (mg/g), a high degree of consistency was observed among the model performance indices, i.e. R^2, RMSE% and MAE%, for Car (μg/cm^2).

(3) The best relationships between canopy reflectance measurements and Car (μg/cm^2) and Car (mg/g) were both obtained by PLSR, with the coefficient of determination $R^2 = 0.56$ and $R^2 = 0.46$, respectively. Car (g/m^2) was best estimated by published vegetation index CIred-edge, with coefficient of determination $R^2 = 0.47$. Further comparison by a paired-*t*-test indicated that Car (μg/cm^2) and Car (mg/g) can be equally estimated by all four adopted methods without significant differences. Whereas, the estimation results for Car (g/m^2) obtained by published vegetation indices were significantly better than those obtained by SMLR ($P < 0.000$).

In summary, the results of this study show that even the carotenoids expressed on concentration (mg/g) or content (μg/cm^2) basis at leaf level can be estimated with almost the same prediction accuracies to the carotenoids expressed as a mass per unit surface area (g/m^2) at canopy level using reflectance measurement at canopy level. The conclusion in this study gives rise to positive expectations for predicting a given chemical expressed on a concentration (mg/g) or content (μg/cm^2) basis from airborne and space-borne sensors at the landscape or regional scales.

Acknowledgements

This work was funded by the following remote sensing programs: National Natural Science Foundation of China (41101430, 41171295, 51109183, and 41371393), the Doctoral Fund of Ministry of Education of China (20110101120036). The work assistances given by colleagues in Xinjiang Institute of Ecology and Geography Chinese Academy of Sciences are highly appreciated.

References

Asner, G.P., Martin, R.E., 2008. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. Remote Sens. Environ. 112, 3958–3970.
Asner, G.P., Martin, R.E., Ford, A.J., Metcalfe, D.J., Liddell, M.J., 2009. Leaf chemical and spectral diversity in Australian tropical forests. Ecol. Appl. 19, 236–253.
Asner, G.P., Martin, R.E., Tupayachi, R., Emerson, R., Martinez, P., Sinca, F., 2011. Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests. Ecol. Appl. 21, 85–98.
Atzberger, C., Guerif, M., Baret, F., Werner, W., 2010. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Comput. Electron. Agric. 73, 165–173.
Bian, M., Skidmore, A.K., Schlerf, M., Wang, T.J., Liu, Y.F., Zeng, R., Fei, T., 2013. Predicting foliar biochemistry of tea (Camellia sinensis) using reflectance spectra measured at powder, leaf and canopy levels. ISPRS J. Photogrammetry Remote Sensing 78, 148–156.
Blackburn, G.A., 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. Remote Sens. Environ. 66, 273–285.
Blackburn, G.A., 1999. Relationships between spectral reflectance and pigment concentrations in stacks of deciduous broadleaves. Remote Sens. Environ. 70, 224–237.
Blackburn, G.A., 2007. Hyperspectral remote sensing of plant pigments. J. Exp. Bot. 58 (4), 855–867.
Cacuci, D.G., 2005. Sensitivity and uncertainty analysis, vol. 1. Theory. Taylor & Francis e – Library.
Chappelle, E.W., Kim, M.S., McMurtrey, J.E., 1992. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. Remote Sens. Environ. 39, 239–247.

Chen, S., Hong, X., Harris, C.J., Sharkey, P.M., 2004. Sparse modeling using orthogonal forward regression with PRESS statistic and regularization. IEEE Trans. Syst., Man, Cyber. – Part B: Cyber. 34 (2), 898–911.
Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413.
Datt, B., 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a + b, and total carotenoids content in Eucalyptus leaves. Remote Sens. Environ. 66, 111–121.
Daughtry, C.S.T., Walthall, C.L., Kim, M.S., Brown de Colstoun, E., McMurtrey III, J.E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sens. Environ. 74, 229–239.
Demmig-Adams, B., Adams III, W.W., 1992. Photoprotection and other responses of plants to high light stress. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43, 599–626.
Demmig-Adams, B., Gilmore, A.M., Adams III, W.W., 1996. *In vivo* functions of carotenoids in higher plants. FASEB J. 10, 403–412.
Edge, R.D., McGarvey, J., Truscott, T.G., 1997. The carotenoids as anti-oxidants – a review. J. Photochem. Photobiol. B: Biol. 41, 189–200.
Faria, T., García-Plazaola, J.I., Abadia, A., Cerasoli, S., Pereira, J.S., Chaves, M.M., 1996. Diurnal changes in photoprotective mechanisms in leaves of cork oak (Quercus suber) during summer. Tree Physiol. 16, 115–123.
Féret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R., Bidel, L.P.R., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. Remote Sensing Environ. 112, 3030–3043.
Féret, J.B., François, C., Gitelson, A., Gregory, P.A., Barry, K.M., Panigada, C., Richardson, A.D., Jacquemoud, S., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sensing Environ. 115, 2742–2750.
Gamon, J.A., Peñuelas, J., Field, C.B., 1992. A narrow-wave band spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing Environ. 41, 35–44.
Garrity, S.R., Eitel, J.U.H., Vierling, L.A., 2011. Disentangling the relationships between plant pigments and the photochemical reflectance index reveals a new approach for remote estimation of carotenoid content. Remote Sensing Environ. 115, 628–635.
Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.
Gitelson, A.A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of Asculus hippocastanum and Acer platanoides leaves. Spectral features and relation to chlorophyll estimation. J. Plant Physiol. 143, 286–292.
Gitelson, A.A., Merzlyak, M.N., 1996. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. J. Plant Physiol. 148, 494–500.
Gitelson, A.A., Merzlyak, M.N., 1997. Remote estimation of chlorophyll content in higher plant leaves. Int. J. Remote Sensing 18, 2691–2697.
Gitelson, A.A., Buschmann, C., Lichtenthaler, H.K., 1999. The chlorophyll fluorescence ratio F735/F700 as an accurate measure of the chlorophyll content in plants — experiments with autumn chestnut and maple leaves. Remote Sensing Environ. 69, 296–302.
Gitelson, A.A., Zur, Y., Chivkunova, O.B., Merzlyak, M.N., 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochem. Photobiol. 75, 272–281.
Gitelson, A.A., Gritz, U., Merzlyak, M.N., 2003a. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282.
Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003b. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5), 1248.
Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. Geophys. Res. Lett. 32, L08403.
Gitelson, A.A., Keydan, G.P., Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin content in higher plant leaves. Geophys. Res. Lett. 33, L11402.
Gitelson, A.A., Chivkunova, O.B., Merzlyak, M.N., 2009. Non-destructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. Am. J. Bot. 96, 1861–1868.
Grossman, Y.L., Ustin, S.L., Jacquemoud, S., Sanderson, E.W., Schmuck, G., Verdeboutt, J., 1996. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. Remote Sensing Environ. 56, 182–193.
Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem. 60 (11), 1193–1202.
Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sensing Environ. 81, 416–426.
Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sensing Environ. 86, 542–553.
Hatonen, J., Hyotyniemi, H., Miettunen, J., Carlsson, L.E., 1999. Using image information and partial least squares method to estimate mineral concentrations in mineral flotation. In: IPMM 99: Proceedings of the second international conference on intelligent processing and manufacturing of materials. Honolulu, HI, USA, 10–15 July, vol. 1, pp. 459–464.
Hernández-Clemente, R., Navarro-Cerrillo, R.M., Suárez, L., Morales, F., Zarco-Tejada, P.J., 2011. Assessing structural effects on PRI for stress detection in conifer forests. Remote Sensing Environ. 115, 2360–2375.
Hernández-Clemente, R., Navarro-Cerrillo, R.M., Zarco-Tejada, P.J., 2012. Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT + DART simulations. Remote Sensing Environ. 127, 298–315.
Homolová, L., Malenovsky, Z., Clevers, J.G.P.W., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. Ecol. Complexity 15, 1–16.
Horton, P., Ruban, A.V., Walters, R.G., 1996. Regulation of light harvesting in green plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 655–684.
Jiang, D.A., Zhu, C., 1999. Guide for Plant Physiology Experiment. Chengdu University of Science and Technology Press, pp. 20–23.
Kvalheim, O.M., 1987. Latent-structure decompositions (projections) of multivariate data. Chemom. Intell. Lab. Syst. 2, 283–290.
Le Maire, G., François, C., Dufrêne, E., 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote Sensing Environ. 89, 1–28.
Le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailler, J.Y., Bréda, N., et al., 2008. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. Remote Sensing Environ. 112, 3846–3864.
Lindberg, W., Persson, J.A., Wold, S., 1983. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. Anal. Chem. 55, 643–648.
Malenovský, Z., Homolová, L., Zurita-Milla, R., Lukeš, P., Kaplan, V., Hanuš, J., Jean-Philippe, G.E., Schaepman, M.E., 2013. Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. Remote Sensing Environ. 131, 85–102.
Meggio, F., Zarco-Tejada, P.J., Núñez, L.C., Sepulcre-Cantó, G., Gonzalez, M.R., Martin, P., 2010. Grape quality assessment in vineyards affected by iron deficiency chlorosis using narrow-band physiological remote sensing indices. Remote Sensing Environ. 114, 1968–1986.
Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of leaf senescence and fruit ripening. Physiol. Plant. 106, 135–141.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models, 1: a discussion of principles. J. Hydrol. 10, 282–290.
Peñuelas, J., Baret, F., Filella, I., 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31, 221–230.
Peterman, E.J.G., Gradinaru, C.C., Calkoen, F., Borst, J.C., Grondelle van, R., Amerongen van, H., 1997. Xanthophylls in light-harvesting complex II of higher plants: light harvesting and triplet quenching. Biochemistry 36, 12208–12215.
Schlerf, M., Atzberger, C., Hill, J., Buddenbaum, H., Werner, W., Schüler, G., 2010. Retrieval of chlorophyll and nitrogen in Norway spruce (Picea abies L. Karst.) using imaging spectroscopy. Int. J. Appl. Earth Observation Geoinformation 12, 17–26.
Sims, D.A., Gamon, J.A., 2002. Relationship between pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sensing Environ. 81, 337–354.
Taylor, J.R., 1997. An introduction to error analysis: the study of uncertainties in physical measurements, second ed. University Science Books.
Thomas, J.R., Gausman, H.W., 1977. Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops. Agronomy J. 69, 799–802.
Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., et al., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing Environ. 113, S67–S77.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58 (2), 109–130.
Young, J., Frank, H.A., 1996. Energy transfer reactions involving carotenoids: quenching of chlorophyll fluorescence. J. Photochem. Photobiol. B: Biol. 36, 3–15.
Zarco-Tejada, P.J., Miller, J.R., Harron, J., Hu, B., Noland, T.L., Goel, N., et al., 2004. Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies. Remote Sensing Environ. 89, 189–199.
Zarco-Tejada, P.J., Guillén-Climent, M.L., Hernández-Clemente, R., Catalina, A., González, M.R., Martin, P., 2013. Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). Agric. Forest Meteorology 171–172, 281–294.
Zhao, K.G., Valle, D., Popescu, S., Zhang, X.S., Mallick, B., 2013. Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. Remote Sensing Environ. 132, 102–119.
Zur, Y., Gitelson, A.A., Chivkunova, O.B., Merzlyak, M.N., 2000. The spectral contribution of carotenoids to light absorption and reflectance in green leaves. In: Proc. 2nd International Conference on Geospatial Information in Agriculture and Forestry. Lake Buena Vista, FL, 10–12 January, vol. 2, pp. II-17–II-23.