Simple and robust methods for remote sensing of canopy chlorophyll content: a comparative analysis of hyperspectral data for different types of vegetation

Yoshio Inoue1, Martine Guérif2, Frédéric Baret2, Andrew Skidmore3, Anatoly Gitelson4, Martin Schlerf5, Roshanak Darvishzadeh3 & Albert Olioso2

1National Institute for Agro-Environmental Sciences, Tsukuba, Japan, 2INRA, UMR1114, EMMAH, F-84914, Avignon, France, 3University of Twente, Enschede, The Netherlands, 4University of Nebraska, Lincoln, NE, USA and 5Luxembourg Institute of Science and Technology, Belval, Luxembourg

ABSTRACT

Canopy chlorophyll content (CCC) is an essential ecophysiological variable for photosynthetic functioning. Remote sensing of CCC is vital for a wide range of ecological and agricultural applications. The objectives of this study were to explore simple and robust algorithms for spectral assessment of CCC. Hyperspectral datasets for six vegetation types (rice, wheat, corn, soybean, sugar beet and natural grass) acquired in four locations (Japan, France, Italy and USA) were analysed. To explore the best predictive model, spectral index approaches using the entire wavebands and multivariable regression approaches were employed. The comprehensive analysis elucidated the accuracy, linearity, sensitivity and applicability of various spectral models. Multivariable regression models using many wavebands proved inferior in applicability to different datasets. A simple model using the ratio spectral index (RSI; R815, R704) with the reflectance at 815 and 704 nm showed the highest accuracy and applicability. Simulation analysis using a physically based reflectance model suggested the biophysical soundness of the results. The model would work as a robust algorithm for canopy-chlorophyll-metre and/or remote sensing of CCC in ecosystem and regional scales. The predictive-ability maps using hyperspectral data allow not only evaluation of the relative significance of wavebands in various sensors but also selection of the optimal wavelengths and effective bandwidths.

Key-words: photosynthesis; reflectance; spectral index.

INTRODUCTION

Systematic monitoring, diagnostics and predictions of photosynthetic productivity are essential for plant and environmental sciences as well as agricultural applications (Roy et al. 2001; IPCC 2014; Way & Long 2015). Leaf chlorophyll concentration (LCC) and/or green leaf area index (LAI) have been used in various photosynthesis studies (Nobel 2005). The greenness of crop leaves has been used for fertilizer management owing to the proportional relationship of LCC with nitrogen content (Ferwerda et al. 2005; Houlès et al. 2007; Rorie et al. 2011; Inoue et al. 2012). The significant contribution of nitrogen to photosynthesis can be explained by the high nitrogen content in the photosynthetic apparatus (Sinclair & Sinclair & Horie 1989). In rice leaves, for example, 75–85% of the total nitrogen is included in chloroplast throughout the growing period (Morita 1978).

However, in situ measurement of both LCC and LAI values representative of a canopy is not only an easy task but also prone to uncertainty (Jonckheere et al. 2004; Parry et al. 2014). Quantification of LAI is affected by the threshold between green and non-green elements, but it is unclear (Jonckheere et al. 2004). More essentially, the canopy-scale productivity is driven by the total photosynthetically active radiation (APAR) absorbed by all chlorophyll pigments that are distributed in 3D within a canopy (De Pury & Farquhar 1997). These facts suggest that accurate spatial assessment of canopy chlorophyll content (CCC) by remote sensing is vital for a wide range of ecophysiological and agricultural applications.

Several spectral indices have been proposed specifically for the assessment of chlorophyll content on leaf or canopy scales (e.g. Broge & Leblanc 2000; Daughtry et al. 2000; Dash & Curran 2004; Gitelson et al. 2005; Delegido et al. 2008; le Maire et al. 2008). However, the optimal model specifications and their general applicability remain unclear because predictive performances of spectral models are affected by scales, species and various confounding factors. Accordingly, preceding studies have strongly suggested a comprehensive analysis based on diverse canopy-scale datasets to explore accurate and robust models (Richardson et al. 2002; Delegido et al. 2008; le Maire et al. 2008).

Here, our close international collaborations have enabled such a comprehensive analysis based on high-quality hyperspectral datasets for different types of vegetation acquired by various sensors at diverse locations. The objectives of this study were to explore accurate and robust algorithms...
for remote sensing of CCC and to elucidate the predictive ability and relative advantage/limitations of various spectral models from the aspects of predictive accuracy, robustness, applicability and simplicity.

**MATERIALS AND METHODS**

**Datasets**

Datasets for six types of vegetation (rice, wheat, corn, soybean, sugar beet and natural grass) were analysed in this study. The datasets were obtained independently in four locations, that is Japan, France, USA and Italy during a 1990–2012 period (Table 1). These vegetation canopies include high variability in size and shape of leaves as well as in canopy geometry. All spectral data were obtained by using calibrated hyperspectral sensors of comparable specifications based on normal measuring configurations for acquisition of representative canopy reflectance signatures (Thenkabail et al. 2011).

Among the six datasets, the rice canopy dataset was used for exploring the new spectral models because the advanced sensor specifications (e.g. spectral resolution, dynamic range and sensitivity), direct determination of CCC and high canopy homogeneity favoured this exploratory analysis. The other datasets were used for comprehensive validation studies. Accordingly, the rice dataset is explained in detail in the succeeding texts while the basic information on the other datasets is summarized in Table 1 with reference documents for details.

**Dataset for model exploration**

The rice dataset was obtained in the experimental fields of National Institute for Agro-Environmental Studies (NIAES; Tsukuba, Japan) in 2009. A rice variety (*Oryza sativa* L., japonica; variety: Koshihikari) was grown in 10 of 10 × 10 m experimental plots. A bundle of four seedlings (hill) of about 15 cm long were transplanted at a spacing of 30 × 15 cm. In addition to the standard level of nitrogen application (10 g m⁻²), four different N levels (2, 6, 14 and 16 g m⁻²) were applied to induce a wide range of LAI and CCC.

Four times during the vegetative and reproductive growth stages (26 June, 14 July, 21 July and 3 August), plant height, stem density, leaf area, dry biomass, water content and chlorophyll content were determined for each plot. Five hills per plot were sampled randomly for destructive measurements and chemical analysis in the laboratory. The plant growth within each nitrogen treatment was uniform (Coefficient of Variance (CV) for plant height was 2–5%) throughout the season. LAI was determined by using optical area metre (LI-3100C, Li-COR, Lincoln, NE, USA) after careful removal of senescent leaf parts. The water content of each part was determined after desiccation in an oven at 70°C for 48 h. The chlorophyll pigments were extracted with 90% acetone from all leaves detached from an intermediate hill in each plot, and the concentration per dry weight was determined by absorption spectroscopy using a spectrophotometer (UV-1600GLP, Shimadzu, Kyoto, Japan). CCC was then obtained by multiplying the

<table>
<thead>
<tr>
<th>Location</th>
<th>Reference</th>
<th>Year</th>
<th>Plots</th>
<th>Sensor and spectral range (nm)</th>
<th>LAI (m² m⁻²)</th>
<th>CCC (g m⁻² m⁻³)</th>
<th>Data range</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vegetation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rice</td>
<td>This paper</td>
<td>2009</td>
<td>64</td>
<td>FS-Pro (ASD)</td>
<td>0.03–2.13</td>
<td>0.08–6.73</td>
<td>350–2500</td>
<td>spectrophotometer</td>
</tr>
<tr>
<td>Wheat (Triticum aestivum)</td>
<td>Houlès et al. 2007</td>
<td>2001</td>
<td>144</td>
<td>CASI-3 (Hres Inc.)</td>
<td>0.05–7.66</td>
<td>0.40–8.65</td>
<td>400–1050</td>
<td>spectrophotometer</td>
</tr>
<tr>
<td>Corn (Zea mays)</td>
<td>Olib et al. 2005</td>
<td>2003</td>
<td>118</td>
<td>USB2000 (Ocean Optics)</td>
<td>0.03–3.61</td>
<td>0.17–5.52</td>
<td>400–1050</td>
<td>spectrophotometer</td>
</tr>
<tr>
<td>Soybean (Glicine max (L.) Merr)</td>
<td>Darvishzadeh et al. 2008</td>
<td>2005</td>
<td>100</td>
<td>GER3700 (GERC)</td>
<td>0.03–2.79</td>
<td>0.14–5.45</td>
<td>400–1050</td>
<td>spectrophotometer</td>
</tr>
<tr>
<td>Sugar beet (Beta vulgaris L.)</td>
<td>Darvishzadeh et al. 2008</td>
<td>2005</td>
<td>100</td>
<td>GER3700 (GERC)</td>
<td>0.03–2.79</td>
<td>0.14–5.45</td>
<td>400–1050</td>
<td>spectrophotometer</td>
</tr>
</tbody>
</table>

leaf chlorophyll \((Ca + Ch)\) concentration by the biomass of green leaves per \(m^2\) in the canopy (Morita 1978).

Canopy reflectance spectra were obtained under clear-sky conditions around midday (10:00–13:00 Local Standard Time (LST)) using a portable spectroradiometer (FieldSpec-Pro, ASD, Boulder, CO, USA). The spectral range of the sensor was 350–2500 nm. The spectral resolution was 3 nm for the 350–1000 nm wavelength region and 10 nm for the 1000–2500 nm wavelength region. The field of view of the sensor was 25°. Reflectance measurements were taken at a nadir-looking angle from 2 m above the canopy. More than 30 spectra were obtained using the same instrument under the controlled laboratory environment. The soils had a variety of colours and carbon contents ranging from 0.16 to 19.8%.

**Analytical methods**

**Spectral index approach**

We applied the normalized difference spectral index (NDSI)-map and ratio spectral index (RSI)-map approaches to explore the optimal indices for assessment of CCC using the entire hyperspectral data (Eqns 1 and 2; Inoue et al. 2008; Inoue et al. 2012). The definitions of the NDSI are given by the following equation:

\[
\text{NDSI} (x, y) = \frac{y-x}{x+y},
\]

where \(x\) and \(y\) are reflectance \((R_i\) and \(R_j)\) or first derivative \((D_i\) and \(D_j)\) values at \(i\) and \(j\) nm over the whole hyperspectral regions (Liu et al. 2003; Mutanga & Skidmore 2004; Schlerf et al. 2005; Inoue et al. 2008; Inoue et al. 2012). Similarly, the RSI is defined as follows:

\[
\text{RSI} (x, y) = \frac{x}{y}.
\]

Here, both \(R\) and \(D\) values were used for \(x\) and \(y\), respectively.

The NDSI and RSI maps are created as a contour map of statistical indicator such as coefficient of determination \(\left(r^2\right)\) between the target variable and spectral indices (NSDIs or RSIs). The reflectance \((R)\) spectra at 2 and 3 nm intervals were derived for all available wavelength range for six datasets, and first derivative \((D)\) spectra were generated from these reflectance spectra. For the comprehensive evaluation, 32 major SIs from the literature (see the note for Fig. 5) were included in a comparison of predictive ability. Note that not all of them were necessarily designed for assessment of CCC.

**Multivariable regression methods**

Partial least-squares regression (PLSR) and interval partial least-squares regression (iPLSR) were used. The PLSR is able to reduce the multi-collinearity problem for hyperspectral data without losing the information about the contribution of individual wavebands. The iPLSR is an improved version of PLSR with iterative waveband selection processes to minimize the residual error (Norgaard et al. 2000).

The theoretical details for the PLSR are given as follows:

\[
y_i = \beta_0 + \sum_{k=1}^{r} \beta_k T_{ik} + e_i (i = 1, \ldots, n)
\]

\[
T_{ik} = \sum_{j=1}^{m} c_{kj} x_{ij} (k = 1, \ldots, r).
\]

where \(y_i\) is target variable (dependent variable), \(x_{ij}\) is spectral reflectance (independent variables), \(n\) is number of samples, \(m\) is number of spectral bands, \(e_i\) is error, \(\beta_k\) is regression coefficients, \(T_{ik}\) is latent variable (LV), \(r\) is number of latent variables, \(c_{kj}\) is coefficient for LV.

The set of coefficients \(c_{kj}\) is determined so that the covariance between \(T_k\) and \(y\) is maximized. The number of latent variables is determined to minimize the prediction error through cross validation.

**Physically based canopy reflectance model**

The radiative transfer model PROSAIL was used to simulate canopy reflectance spectra under various plant and soil conditions. The PROSAIL model can calculate the canopy reflectance as a function of seven input parameters including LCC, leaf structural parameter, equivalent water thickness, sun zenith angle, background reflectance, LAI and leaf angle distribution (LAD) (Jacquemoud et al. 2009). Here, we simulated the reflectance spectra by changing LAI, LCC, LAD and soil spectra to compare the response of spectral reflectance with the differences in canopy size, leaf chlorophyll, plant type and soil background, respectively.

**Comparative methods for assessment of model performance**

To compare the predictive ability of spectral models using SIs, PLSR and iPLSR, we employed statistical indicators such as \(r^2\), root mean square error (RMSE) or normalized RMSE (NRMSE) and discrepancy of slope (DS) or normalized DS (NDS). DS indicates the discrepancy of the slope of regression line between measured and predicted values from 1:1 line (slope = 1). These indicators are defined as follows:

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}
\]

\[
\hat{y}_i \quad \text{predicted values}
\]

\[
y_i \quad \text{measured values}
\]
\[ NRMSE = \frac{(RMSE/m)}{range} \] (6)

\( m \) mean value for each dataset
\( range \) range of \( NRMSE \) values for all spectral models

In \( NRMSE \), \( RMSE \) was normalized by both mean and range to take account of their differences in each dataset. The \( DS \) is defined by

\[ DS = |s - 1| \] (7)

\( s \) slope of the regression line between measured and predicted data

\[ NDS = 1 - DS/range \] (8)

\( range \) range of \( DS \) for all spectral models

In \( NDS \), \( DS \) was normalized by the range only because the slope is independent from the mean in each dataset. Consequently, these three indicators \( NRMSE \), \( NDS \) and \( r^2 \) vary between 0 and 1 and represent the overall scattering including bias, sensitivity (slope) and linearity of the model, respectively (Gauch et al. 2003). In addition, another statistical indicator \( d_r \) (Willmott et al. 2012) was calculated for checking the robustness of the evaluation of model performances.

\[ d_r = 1 - \frac{\sum |\hat{y}_i - y_i|}{2\sum |y_i - \bar{y}|} \text{ when } \sum |\hat{y}_i - y_i| \leq 2\sum |y_i - \bar{y}| \]

\[ = \frac{2\sum |y_i - \bar{y}|}{\sum |y_i - y_i|} - 1, \text{ when } \sum |\hat{y}_i - y_i| > 2\sum |y_i - \bar{y}| \] (9)

\( \hat{y}_i \): predicted values \( y_i \): measured values \( \bar{y} \): mean of measured values.

RESULTS AND DISCUSSION

Predictive models derived from rice dataset

Exploring best spectral indices using NDSI-map and RSI-map approach

Overall, the reflectance spectra showed the typical response to green vegetation, that is a positive relationship with increasing CCC in the near-infrared wavelength region and a negative response in the red region. Figure 1 shows some examples of

![Figure 1](image-url)

**Figure 1.** Typical reflectance spectra (a) and derivative spectra (b) for rice canopies with different canopy chlorophyll contents (CCCs). Data in two wavelength regions around 1400 and 1900 nm are eliminated because of the low incoming solar energy due to atmospheric water vapour.

reflectance spectra (a) and derivative spectra (b) in rice canopies with different levels of CCC. The range of CCC (0.01–2.13 g m⁻²) and LAI (0.08–6.73 m² m⁻²) covered nearly maximum values in normal conditions, respectively. In the derivative spectra, D values show obvious peaks at around 710, 1130 and 1320 nm corresponding to the increasing CCC. The shift of peak-wavelength in the red-edge region (so-called blue-shift; e.g. Vogelmann et al. 1993; Filella & Peñuelas 1994) was noticeable.

Several spectral indices were found to be correlated very well with the changing CCC. Figure 2 shows the NDSI and RSI maps (contour maps of $r^2$ between CCC and SIs) using reflectance values for the 400–1100 nm wavelength region. In the NDSI map (Fig. 2a), the most significant spot ($r^2 = 0.940$) was found around the peak at NDSI (R740, R761). This significant area was narrow (approximately 10 nm) along 740 nm (Ri) but wide over 750–830 nm (Rj). In the RSI map (Fig. 2b), the most significant spot ($r^2 = 0.946$) was found around the peak at RSI (R815, R704), and the second significant spot ($r^2 = 0.940$) was at RSI (R815, R578). These results clearly indicate the critical role of R815 for the spectral assessment of CCC and the excellent predictive ability of its combination with a red-edge band (R704) or a green band (R578). Another important finding is that RSIs would have more robust predictive performance than NDSIs because the size of significant areas is much broader in RSIs. Generally, spectral models using SIs at broader spot are less affected by the uncertainties in wavelength calibration or other sensor specifications such as spectral resolution or bandwidth.

In case that derivative values (D) were applied to the NDSI and RSI maps (Fig. 3), the most significant peak ($r^2 = 0.918$) in the NDSI map was at NDSI (D689, D755). Two most signifi-
cant peaks in RSI maps were found at RSI (D734, D542) ($r^2 = 0.948$) and RSI (D734, D683) ($r^2 = 0.945$), respectively. These results confirm the important role of red-edge and green wavebands as reported by many papers (e.g. Vogelmann et al. 1993; Gitelson et al. 2005). However, the useful spots were much narrower in both NDSIs and RSIs than those using reflectance values (Fig. 2). Therefore, these indices strongly require high spectral resolution and high accuracy of wavelength position, which would be one of the critical constraints for wider applicability (Lee et al. 2008).

Multivariable regression models using hyperspectra

Figure 4 shows the comparison of measured CCC with predicted values by iPLSR using reflectance (R) and derivative (D) spectra. The iPLSR using D was better than that using R. In the iPLSR model using reflectance spectra, wavebands around the red-edge wavelength regions (720–820 nm) were selected to compose seven latent variables. For the iPLSR model using derivative spectra, 10 wavebands in blue, green and red-edge regions were selected for seven latent variables. These wavelength regions are closely related to the chlorophyll pigment and green biomass. The statistical indicators $r^2$ (0.947 and 0.954) and RMSE (0.115 and 0.106) for the cross-validation suggest the excellent performance of iPLSR models. It is interesting that the predictive ability of iPLSR model was superior to PLSR model even though PLSR uses much larger number of wavebands (233). These results confirmed the limited applicability of PLSR, which has been recognized in laboratory chemometry (Grossman et al. 1996) and in field applications (Inoue et al. 2012).

Comparison of spectral models

The predictive ability of selected spectral models is compared in Fig. 5. Both iPLSR models using reflectance and derivative spectra were ranked first and second. The index models using RSI (D734, D542) and RSI (R815, R704) were ranked third and fourth, respectively. It is remarkable that the RSI (R815, R704) using only two narrow wavebands has predictive ability comparable with the better-ranked models using larger number of wavebands. Overall, previous indices proposed for CCC using red-edge and green wavebands, such as VOG-3, GMI-2, CIred-edge, CIgreen and MTCI, proved to have moderate to excellent predictive ability. The $\lambda_{rep}$ model using the red-edge position had extremely high RMSE, but it was attributable to a few extreme points beyond the range of red-edge. This may suggest that $\lambda_{rep}$ would not be able to cover a sufficient range of CCC. Generally, spectral models proposed for leaf-scale variables (e.g. MCA, PRI and SIPI) did not show good predictive ability for assessment of CCC as suggested in preceding studies (le Maire et al. 2008).

Comprehensive validation of the predictive performance of spectral models using independent datasets

Comparison of selected spectral models

The applicability of the 40 spectral models to the different datasets is depicted by the position in the x–y space of linearity ($r^2$) versus overall error (RMSE) (Fig. A1). The positions of the four top-ranked models reveal the superior performance of the RSI (R815, R704). Clearly, iPLSR models are not always excellent when applied to different datasets. This may be attributable to the over-fitting of the calibration dataset. Their performance was poor, especially in different plant types (soybean and sugar beet). The applicability to the natural grass dataset was relatively low in all models (Fig. A1f).

The predictive performance of all spectral models was compared using the average values of linearity ($r^2$), normalized overall error (NRMSE) and normalized discrepancy of sensitivity (NDS) for the six datasets (Fig. A2). The linearity ($r^2$)
was highest in RSI (R815, R704) and second in CIred-edge, which utilizes the ratio of R840 ~ 870 nm and R720 ~ 730 nm. The difference of the linearity was relatively minor among the majority of models using the red-edge wavelengths (700 ~ 750 nm). The effect of using more than two wavebands was not clear. The overall scattering error (NRMSE) was smallest in VOG-3 using four wavebands (R715, R720, R734 and R747) within the red-edge region. The second best was the RSI (R815, R704), and the third was CIred-edge. The SIs were developed specifically to detect leaf-scale variables such as PRI, and MCARI had relatively large error. The NDS (from 1:1 line) was smallest in CIred-edge2, which utilizes the ratio of R750 ~ 800 nm and R695 ~ 740 nm. The RSI (R815, R704) was the second, and VOG-3 was the third. The large variability of NDS for the majority of models using only the red-edge wavelengths suggests their instability because of the high sensitivity to the position of selected wavebands within the narrow region. The applicability of iPLSR models proved poor in all statistical indicators (NRMSE, $r^2$ and NDS). This would be attributable not only to the multi-collinearity issue but also to higher necessity for accurate absolute reflectance. Another index $d$, for evaluation of model performance showed close negative relationships with mean-absolute error (MAE), RMSE and DS and a close positive relationship with $r^2$. Accordingly,
the ranking of models by $d$, was similar to those by other statistical indicators.

Finally, the overall model performance was compared by using the mean of the three statistical indicators ($r^2$, 1-NRMSE and 1-NDS) in Fig. 6. The model using RSI (R815, R704) was best followed by VOG-3 and ZM. Accordingly, the new spectral model using RSI (R815, R704) would be most promising in the aspects of linearity, robustness, sensitivity and applicability. The RSI (R815, R704) is also superior in simplicity to other more complex models that utilize larger number of wavebands. Figure 7 shows the scatter plots between predicted and measured CCC in all six datasets for the best model, that is RSI (R815, R704). A systematic difference of slope for the dataset of natural grass is observed commonly in all scatter plots, which might be explained in part by the differences in LAI measuring method and the complexity of ecosystems. Another systematic bias observed for the higher range of wheat data was obvious especially in VOG-3 (Fig. A3a) and ZM (Fig. A3b). In these models, the use of wavelengths only within the red-edge region might have lowered the applicability to wider conditions. Consequently, the spectral model $CCC_{sp}$ using the RSI (R815, R704) proved to be most suitable for assessment of CCC;

$$CCC_{sp} (\text{g m}^{-2}\text{land}) = 0.325 \text{RSI}(R815, R704) - 0.358. \quad (10)$$

This equation can be rearranged to be $0.325 [\text{RSI} (R815, R704) - 1] - 0.033$, which implies the proportionality of [RSI (R815, R704) - 1] to CCC. This proportionality supports the assumptions by Gitelson et al. (2005) proposed for development of spectral indices. These results suggest that the $CCC_{sp}$ model would be applicable to various types of vegetation without modification.

Although few studies were concerned about the bandwidths for SIs (e.g. Gitelson et al. 2005; Inoue et al. 2008), our results (Fig. 2) suggest that bandwidths would also be critical for high predictive ability and robustness. For the $CCC_{sp}$ model, 5 nm for R704 and 10 nm for R815, respectively would be optimal to achieve the highest predictive ability. However, the model would still have relatively high performance even with wider bandwidths such as 10 and 30 nm, respectively.

Multivariable regression methods (e.g. PLSR) and machine learning methods (e.g. support vector machine and

![Figure 7. Scatter plot of predicted and measured CCC values for the best model in Fig. 6, RSI (R815, R704). Model and model parameters determined for the rice datasets were applied to all six vegetation types. The $r^2$ and RMSE values are for the combined dataset.](image)
artificial neural network) can be applied to hyperspectral data (Hansen & Schjoerring 2003; Ali et al. 2015). However, applicability of multivariable regression models to different sensors and/or different types of vegetation proved unstable.

Accuracy and applicability of data-driven models by machine learning methods are highly dependent on the size and quality of the training datasets (Doktor et al. 2014; Ali et al. 2015). Accordingly, the SI method has some unique

Figure 8. Reflectance spectra simulated by a physically based canopy reflectance model PROSAIL under a wide range of plant and soil conditions, that is different leaf angle distribution (a), LCC (b), LAI (c) and soil colour and water content (d).
advantages in simplicity, interpretability, robustness and applicability compared with these methods. Additionally, the SI contour-map approach using hyperspectral data has another advantage that SI maps can provide clear overview for selecting optimal wavebands and bandwidths for various sensors.

Investigating the biophysical soundness of the experimental results using a canopy reflectance model

Figure 8a depicts the effects of difference in plant types (spherical, planophile, erectophile and plagio-plate) on reflectance under fixed LCC, LAI and soil background. Surprisingly, the effect of plant type (LAD) is minimal at around 704 nm whereas the other wavebands including R815 nm are highly affected by plant types. This implies that the applicability of the CCCsp model would be constrained by vegetation types. However, this simulation also suggests that the difference of canopy geometry can be adjusted by changing the single coefficient because of the insensitivity of R704. The predictive ability of the model in individual types can be improved by optimizing the parameters in Eqn 10.

Interestingly, the reflectance in near-infrared region is nearly insensitive to the change of LCC at constant LAI (3) whereas the green to red-edge regions are moderately sensitive to LCC (Fig. 8b). The effects of increasing LAI is negative in visible regions and positive in near-infrared regions while minimal at around 740 nm (Fig. 8c). Accordingly, spectral models using near-infrared bands with the band in the region from green to the left side of red-edge would be most significant when seasonal data for a single plant type are used. Because the response of R704 and R815 to LAI is consistent in inverse directions, these two bands would have an effective role for quantification of CCC under changing LAI. The effect of different soil backgrounds is minor at around 704 nm whereas it is relatively large in near-infrared region including 815 nm (Fig. 8d). Therefore, the diversity of soils in our datasets from different locations would be another reason why the model using R704 and R815 was superior in this comparative analysis.

Consequently, these simulation results suggest that the waveband selection for SI models in this study was reasonable. Interestingly, the specific wavelength of chlorophyll absorption around 650 nm was not necessarily selected for CCC assessment in canopy scales whereas it is prerequisite for leaf-scale chlorophyll meter such as SPAD-502 (Konica Minolta, Tokyo, Japan) (Inada 1985). Nevertheless, for more realistic canopy-scale simulations, the strong interactions between LAD, LCC and LAI would have to be incorporated into canopy reflectance model via the combination with plant growth model (Olioso et al. 2005; Baret et al. 2007).

Confounding factors and applications

The physical accuracy and consistency of spectral signatures are affected by sensor quality (sensitivity and dynamic range), measuring configurations (sun and viewing angles, distance) and atmospheric conditions (aerosol and water vapour) (Verhoef & Bach 2012). Accordingly, acquisition of accurate top-of-the-canopy reflectance is essential for precise assessment of CCC. Nevertheless, note that the semi-empirical models using simple SIs are affected by such factors, although the disturbance by atmospheric and/or measuring conditions can be reduced to some extent through normalization (Huete 1988; Myneni & Asrar 1994; Bachmann et al. 2015).

The CCC is not only the essential ecophysiological variable for photosynthetic productivity but also closely related to the other physiological and structural status. Therefore, the CCC model would be useful to infer directly or indirectly the biotic and abiotic stresses such as nitrogen (Inoue et al. 2012), light absorptivity and light use efficiency (Gamon et al. 1997; Inoue et al. 2008), biomass (Kawamura et al. 2010), diseases (Apan et al. 2004), and water deficit (Ceccato et al. 2002).

CONCLUSIONS

This comprehensive study revealed the relative advantages and disadvantages of the majority of spectral models for remote estimation of CCC in the aspects of accuracy, linearity, sensitivity, robustness, simplicity and versatility.

An SI-based model using RSI (R815, R704), that is the ratio of reflectance values at 815 and 704 nm, was found to be superior to all other models in overall predictive ability of CCC. The soundness of the model (CCCsp) was supported by simulation analyses using a physically based reflectance model under various canopy conditions including plant types (canopy geometry), LCC, LAI and soil background. The PLSR and iPLSR models using much larger number of wavebands proved to be inferior to the index-based models, especially in versatility. The CCCsp would be used as a simple and robust algorithm for the canopy scale chlorophyll-metre and/or remote sensing of CCC in ecosystem and regional scales.

Geospatial and timely assessment of CCC is vital for a wide range of agricultural and ecological applications such as diagnostics for precision farming, ecosystem health and carbon cycle sciences. Upcoming hyperspectral satellite sensors such as EnMAP and HyspIRI would provide great opportunities for spectral assessment of ecophysiological variables (Staenz & Held 2012). The NDSI-map and RSI-map methods proved useful for overall evaluation of the relative significance of wavelengths as well as for selecting the optimal wavebands and effective bandwidth. Our analytical approaches and results as well as the new models would provide useful information and insights for the assessment of ecophysiological functioning of terrestrial vegetation.

ACKNOWLEDGMENTS

This work was supported in part by MEXT, JSPS and CSTI-SIP, Japan, respectively. A.G. thanks Marie Curie Incoming International Fellowship supported his work on this paper.
REFERENCES


© 2016 John Wiley & Sons Ltd, Plant, Cell and Environment, 39, 2609–2623
Assimilation of remote sensing data into crop simulation models and SVAT models. *Irrigation and Drainage Systems* 19, 377–412.


Received 28 April 2016; accepted for publication 2 August 2016
Figure A1. Applicability of 40 spectral models determined for rice dataset to the other plant species as indicated by coefficient of determination ($r^2$) and root mean square error (RMSE). The best four spectral models in Fig. 5 obtained for the rice dataset are indicated by symbols with number.
Figure A2. Predictive ability of spectral indices based on the average values of coefficient of determination ($r^2$), NRMSE and NDS. This graph compares the mean values for six different datasets (rice, wheat, corn, soybean, sugar beet and natural grass). Spectral models with ● are explored in this study, and those with * are proposed in the past. Numbers in graphs indicate the relative ranking.
Figure A3. Scatter plot of predicted and measured CCC values for the second and third best models in Fig. 6, VOG-3 and ZM. Model and model parameters determined for the rice datasets are applied to all six vegetation types. The $r^2$ and RMSE values are for the combined dataset.