



Applicability of the PROSPECT model for estimating protein and cellulose + lignin in fresh leaves

Zhihui Wang^{a,b,*}, Andrew K. Skidmore^a, Tiejun Wang^a, Roshanak Darvishzadeh^a, John Hearne^b

^a Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

^b School of Mathematical and Geospatial Sciences, RMIT University, GPO Box 2476, Melbourne, Victoria 3001, Australia

ARTICLE INFO

Article history:

Received 28 January 2015

Received in revised form 23 June 2015

Accepted 3 July 2015

Available online 24 July 2015

Keywords:

Leaf biochemicals

PROSPECT

Radiative transfer model

Cellulose

Lignin

Protein

Hyperspectral data

ABSTRACT

Hyperspectral remote sensing of leaf biochemicals is critical for understanding many biochemical processes. Leaf biochemical contents (e.g., protein, cellulose and lignin) in fresh and dry leaves have been quantified from hyperspectral data using empirical models. However, they cannot be retrieved for fresh leaves by inverting radiative transfer models. We demonstrated the applicability of PROSPECT leaf optical properties model in the separation of specific absorption coefficients for protein and cellulose + lignin following a newly proposed algorithm, and evaluated the feasibility in estimating leaf protein and cellulose + lignin content through model inversion. Assessment was performed across a large variety of plant species benefiting from the Leaf Optical Properties Experiment (LOPEX) dataset. To alleviate ill-posed problems, inversion was performed over different spectral subsets. The PROSPECT model with newly calibrated specific absorption coefficients was able to accurately reconstruct leaf reflectance and transmittance. Leaf protein and cellulose + lignin were estimated at moderate to good accuracies for both fresh and dry leaves. The spectral subset of 2100–2300 nm yielded the most accurate estimation of leaf cellulose + lignin ($R^2 = 0.70$, $RMSE = 5.21E-04$ g/cm²) and protein ($R^2 = 0.47$, $RMSE = 2.75E-04$ g/cm²) in fresh leaves, which were comparable with those obtained from stepwise multiple linear regressions (protein: $R^2 = 0.83$, $RMSE = 3.91E-04$ g/cm²; cellulose + lignin: $R^2 = 0.66$, $RMSE = 2.02E-04$ g/cm²). Our results confirm the importance of selecting a proper spectral subset that contains sufficient information for a successful inversion. For the first time, we provide promising estimations of leaf protein in fresh leaves through inversion of a radiative transfer model, which can be applied at canopy level for regional mapping if coupled with a canopy reflectance model and air- or space-borne hyperspectral imaging.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Leaf protein, cellulose and lignin provide critical information about many biochemical processes, such as photosynthesis, respiration, and litter decomposition (Peterson & Hubbard, 1992). Proteins are the main nitrogen-containing biochemical constituent in plants (Kokaly, Asner, Ollinger, Martin & Wessman, 2009), and nitrogen is often a limiting factor for plant growth and important in terrestrial ecosystem carbon dynamics (Ollinger & Smith, 2005; Smith, Ollinger, Martin, Aber, Hallett & Goodale, 2002). Cellulose and lignin (hereafter referred to as “cellulose + lignin”) is highly correlated with the total carbon in leaf (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996), and play a significant role in forest litter decomposition and nutrient cycling (Aber & Federer, 1992; Steudler, Bowden, Melillo & Aber, 1989). Therefore, remote estimation of leaf protein and cellulose + lignin is essential in producing an

equivalent ratio of carbon/nitrogen (C/N), which will further improve the understanding of biogeochemical processes.

Two different techniques have been investigated for retrieving leaf protein and cellulose + lignin using hyperspectral data. Empirical methods have been the dominant techniques used to estimate the leaf parameters, focusing on building statistical regression models from the high correlations between their content and reflectance or its derivatives (Curran, 1989; Kokaly & Clark, 1999; Skidmore, Ferwerda, Mutanga, Van Wieren, Peel, Grant, et al., 2010; Zhao, Valle, Popescu, Zhang & Mallick, 2013). However, it proved difficult to transfer the empirical relationships across species or sites (LaCapra, Melack, Gastil & Valeriano, 1996; Martin & Aber, 1997). Moreover, the wavebands selected by statistical analysis from different studies have, on occasion, been inconsistent depending on using reflectance or transmittance, dry or fresh leaves, and have deviated from known absorption bands of leaf compounds (Curran, Dungan & Peterson, 2001; Huang, Turner, Dury, Wallis & Foley, 2004; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996).

Radiative transfer models (RTMs) describe the interaction of solar radiation in leaves based on laws of optics, thus offer advantages in

* Corresponding author at: Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
E-mail address: z.wang-1@utwente.nl (Z. Wang).

Table 1

Statistics of the LOPEX dataset used in this study.

Parameter	Unit	Min	Max	Mean	Std.
Equivalent water thickness (EWT) ^a	cm	0.0046	0.0405	0.0114	0.0067
Equivalent water thickness (EWT) ^b	cm	1.85E−05	0.0009	0.0002	0.0001
Protein	g/cm ²	2.92E−04	0.0020	0.0010	0.0003
Cellulose + lignin	g/cm ²	1.47E−04	0.0046	0.0013	0.0010
Leaf mass per area (LMA) ^a	g/cm ²	0.0019	0.0137	0.0053	0.0024
Leaf mass per area (LMA) ^b	g/cm ²	0.0024	0.0165	0.0059	0.0026

^a and ^b were for parameters in fresh and dry leaves, respectively.

robustness and transferability compared to empirical models (Darvishzadeh, Atzberger, Skidmore & Schlerf, 2011; Jacquemoud & Baret, 1990; Schlerf & Atzberger, 2006). RTMs have been widely applied to retrieve several leaf biochemical parameters (e.g., leaf chlorophyll, dry matter and water content) from remotely sensed data (Baret & Fourty, 1997a; Darvishzadeh, Skidmore, Schlerf & Atzberger, 2008; Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008; Omari, White, Staenz & King, 2013), but they have not been well developed for leaf protein and cellulose + lignin. Several studies (Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996) attempted to incorporate leaf protein, cellulose + lignin into the absorption and scattering processes in radiative transfer models, such as the PROSPECT leaf optical properties model (Jacquemoud & Baret, 1990). Leaf protein and cellulose + lignin for dry leaves could be moderately well estimated ($R^2 = 0.49$ – 0.84) through PROSPECT model inversion (Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996), but so far, these parameters have not been successfully estimated for fresh leaves using the PROSPECT model (Botha, Zebarth & Leblon, 2006; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996; Kokaly, Asner, Ollinger, Martin & Wessman, 2009). On fresh leaves, leaf reflectance and transmittance are insensitive to protein because of its small percentage in the leaf mass (Baret & Fourty, 1997b; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). Moreover, the high covariance with water and other nitrogen-containing compounds, such as chlorophyll, has led to inconsistencies in retrieving leaf protein via PROSPECT inversion (Jacquemoud & Baret, 1990; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996; Kokaly, Asner, Ollinger, Martin & Wessman, 2009). The idea of incorporating leaf protein, and cellulose + lignin into the PROSPECT model was therefore abandoned in the 1990s (Jacquemoud & Baret, 1990; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996), and they were represented by a more general parameter of “dry matter content” (Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996).

The retrieval of leaf parameters (e.g. protein and cellulose + lignin) in fresh leaves by inversion of PROSPECT remains challenging. The first issue needed to be resolved is the determination of specific absorption coefficients of leaf biochemicals, which describes the absorption property of a leaf, together with the content of these biochemicals (Jacquemoud & Baret, 1990). The specific absorption coefficients are considered to be inherent properties, which are constant across samples and species (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008). It is difficult to determine these coefficients for the model; the coefficient for pure liquid water and chlorophyll has been well investigated (Buiteveld, Hakvoort & Donze, 1994; Kou, Labrie & Chylek, 1993; Wieliczka, Weng & Querry, 1989), but gaps still exists for leaf pigments (i.e. chlorophyll a and b) and other cell wall constituents (i.e. protein, cellulose + lignin) (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008). The obstacles lie in measuring the coefficients in vivo, as well as extracting each biochemical constituent for its measurement in vitro (Porra, 2002). Therefore, the in vivo specific absorption

coefficients are normally calibrated from intact leaves via model inversion using the measured spectra and biochemical contents (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008; Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996; Jacquemoud & Baret, 1990; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996).

The “ill-posed” inverse problem of radiative transfer model is the second challenge in the retrieval of leaf protein and cellulose + lignin. Different combinations of leaf parameters can generate similar leaf spectra. And the leaf parameters which make greater contribution to leaf reflectance, such as leaf structure parameter and leaf water, bring more uncertainties in the retrieval of protein and cellulose + lignin. Several studies have demonstrated that selecting spectral subsets will return a higher accuracy than using full wavelengths in inverting radiative transfer models (Darvishzadeh, Skidmore, Schlerf & Atzberger, 2008; Meroni, Colombo & Panigada, 2004; Schlerf & Atzberger, 2006). The reasons for this can be poorly measured spectra, some wavelengths may not be well described in the model (Schlerf & Atzberger, 2006), or extra bands may add noise without adding significant information (Weiss, Baret, Myneni, Pragnère & Knyazikhin, 2000). In addition, higher accuracies could be obtained for estimating leaf parameters if the inversion approach would consider a specific merit function assigned for each parameter and incorporate sensitive spectral bands determined by sensitivity analysis (Li & Wang, 2011; Zhao, Guo, Huang, Reddy, Lee, Fletcher, et al., 2014).

The radiative transfer model PROSPECT has been recalibrated with a newly proposed algorithm which can simulate leaf optical properties with a higher accuracy than previous versions (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008). In their study, the incidence angle of incoming radiation was reassessed, and the specific absorption coefficients of leaf constituents and the refractive index were recalibrated for two new version models (PROSPECT-4 and PROSPECT-5). PROSPECT-5 is the same as PROSPECT-4 except the separation of total chlorophylls and total carotenoids in the visible range of 400–750 nm. The algorithm has been tested for accurate simulation of leaf reflectance (Li & Wang, 2011; Ma, Chen, Wang, Li & Jiapaerl, 2012), and has been further adapted to incorporate tannin as an input parameter for assessing the quality of tea in Tea-PROSPECT (Bian, 2013). However, to our knowledge, the newly proposed algorithm has not been evaluated for determining the specific absorption coefficients for leaf constituents, such as protein, cellulose, hemicellulose, and lignin. And the potential of PROSPECT-5 for retrieving leaf constituents through model inversion needs to be further explored.

The estimation of leaf constituents (e.g. protein and cellulose + lignin) using fresh leaf spectra through the radiative transfer model PROSPECT has not been studied extensively (Botha, Zebarth & Leblon, 2006; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996), probably because of the notion that there may be little sensitivity of leaf optical properties to the leaf constituents. The aim of our study was therefore to evaluate the most recent PROSPECT model (version 5) for estimating leaf protein, cellulose + lignin by recalibrating their specific absorption coefficients, using the calibration algorithm proposed by Feret, Francois, Asner, Gitelson, Martin, Bidet, et al. (2008). Our specific objectives were: (1) to test if the specific absorption coefficients could be robustly estimated across species from a model calibration procedure; (2) to test if the PROSPECT model with the recalibrated coefficients could successfully simulate leaf reflectance

Table 2

Ranges of the input parameters used in the recalibrated PROSPECT-5 model.

Parameter	Unit	Min	Max
Leaf structure parameter (<i>N</i>)	–	1	3
Equivalent water thickness (EWT) ^a	cm	0.001	0.05
Equivalent water thickness (EWT) ^b	cm	0.00001	0.001
Protein	g/cm ²	0.0001	0.002
Cellulose + lignin	g/cm ²	0.0001	0.005

^a and ^b were for parameters in fresh and dry leaves, respectively.

Table 3

Selected spectral subsets for model inversion to estimate leaf protein and cellulose + lignin content based on absorption features from the literature, global sensitivity analysis and stepwise multiple linear regressions.

Spectral subset	Wavelengths (nm)
1	1100–2500
2	1600–1800, 2100–2300
3	2100–2300
4	1020, 1510, 1730, 1980, 2060, 2130, 2180, 2240, 2300
5	1120, 1200, 1420, 1450, 1490, 1540, 1780, 1820, 2100, 2270, 2280, 2340
6 ^a	1345, 2120, 2295, 2300
6 ^b	1320, 1450, 2015, 2290
7 ^a	1150, 1185, 1220, 1295
7 ^b	2000, 2150, 2170, 2255

Spectral subset 2 and 3 are selected based on global sensitivity analysis. 4 and 5 are associated with protein absorption features and cellulose + lignin absorption features, respectively. See (Curran, 1989; Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996). 6^a and 7^a are selected spectral subsets for protein and cellulose + lignin, respectively, based on stepwise multiple linear regressions on fresh leaves. 6^b and 7^b are selected spectral subsets for protein and cellulose + lignin, respectively, based on stepwise multiple linear regressions on dry leaves.

and transmittance in forward mode; (3) to evaluate if selected spectral subsets could improve the estimation accuracy of leaf protein and cellulose + lignin, when compared to using the full wavelengths of 1100–2500 nm; and (4) to compare the accuracy of estimated leaf protein and cellulose + lignin from PROSPECT model inversion and stepwise multiple linear regression.

2. Materials and methods

2.1. Description of the dataset

The Leaf Optical Properties Experiment (LOPEX) dataset was established by the Joint Research Center (JRC) of the European Commission (Ispra, Italy) in 1993. The dataset includes 50 woody and herbaceous species, representing a large variety of leaf internal structure, biochemical composition and spectra. Hemispherical reflectance and transmittance over 400–2500 nm were measured with a Perkin Elmer Lambda 19 Spectrophotometer equipped with an integrating sphere. Spectralon references were used for calibration. For each sample, reflectance and transmittance from five leaves were obtained, and averaged after calibration. The fresh weight (FW, g) of a 4.10 cm² (leaf area, LA) disk taken on each leaf using a cork borer was measured first, and reweighed for dry weight (DW, g) after the disk was placed in a drying oven at 85 °C for 48 h. Leaf mass per area (LMA, g/cm²) and equivalent

water thickness (EWT, cm) were then calculated as: $LMA = DW / LA$, and $EWT = (FW - DW) / LA$. The biochemical constituents, including protein, cellulose, lignin and starch, were determined by two independent laboratories in France and Belgium. Standard wet chemistry analysis were performed for measuring protein using Kjeldahl method (AOAC, 1970), cellulose (Van Soest & Wine, 1991; Weende, 1985), lignin (Van Soest & Masson, 1967; Van Soest & Wine, 1991) and starch (Ewerts, 1985). Measurements from the two laboratories were quite consistent for protein and the sum of cellulose and lignin, and the data from the French laboratory were used.

In this study we only considered protein, cellulose and lignin, since starch is a minor constituent. Protein, cellulose and lignin were expressed as grams per unit fresh leaf area (g/cm²). In total, 63 fresh leaf samples and 49 dry leaf samples were studied. The statistical characteristics of these samples are presented in Table 1, and details of the dataset used and their measurements can be found in Hosgood, Jacquemoud, Andreoli, Verdebout and Schmuck (1995) and Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996).

2.2. Calibration of the PROSPECT model

The PROSPECT leaf optical properties model was developed to simulate leaf directional–hemispherical reflectance and transmittance over the optical domain from 400 nm to 2500 nm (Jacquemoud & Baret, 1990). The PROSPECT model considers the leaf as a stack of elementary homogeneous layers, with each layer as an element with absorbing and scattering properties (Jacquemoud & Baret, 1990). The scattering process is described by the number of stacked elementary homogeneous layers, leaf structure parameter N and the refractive index $n(\lambda)$. The absorption process is described by the absorption coefficient of one elementary layer $k(\lambda)$. The directional–hemispherical reflectance $R_{mod}(\lambda)$ and transmittance $T_{mod}(\lambda)$ at each wavelength λ can be simulated as:

$$[R_{mod}(\lambda), T_{mod}(\lambda)] = \text{PROSPECT}(N, k(\lambda), n(\lambda)) \quad (1)$$

where $k(\lambda)$ is linear combination of the content of leaf constituents C_i (e.g., chlorophyll, water, and dry matter), and their corresponding specific absorption coefficients $k_i(\lambda)$:

$$k(\lambda) = \sum_i k_i(\lambda) \times \frac{C_i}{N}. \quad (2)$$

Then Eq. (1) can be simplified as follows by using only four input parameters:

$$[R_{mod}(\lambda), T_{mod}(\lambda)] = \text{PROSPECT}(N, \text{Chl}, \text{LMA}, \text{EWT}) \quad (3)$$

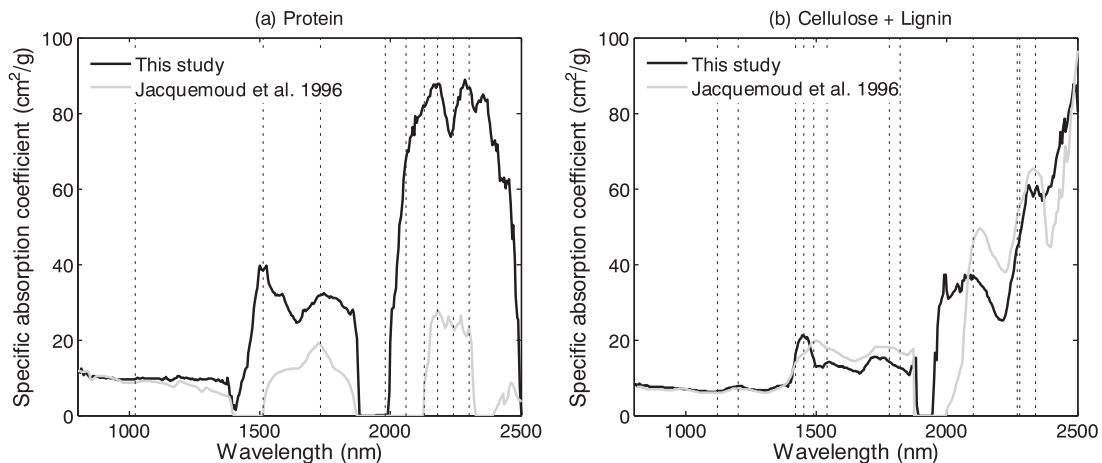


Fig. 1. Recalibrated specific absorption coefficients of protein (a) and cellulose + lignin (b) determined on fresh leaves. The vertical dashed lines correspond to published absorption features of protein (a) and cellulose + lignin (b) (see details in Tables 4 and 5).

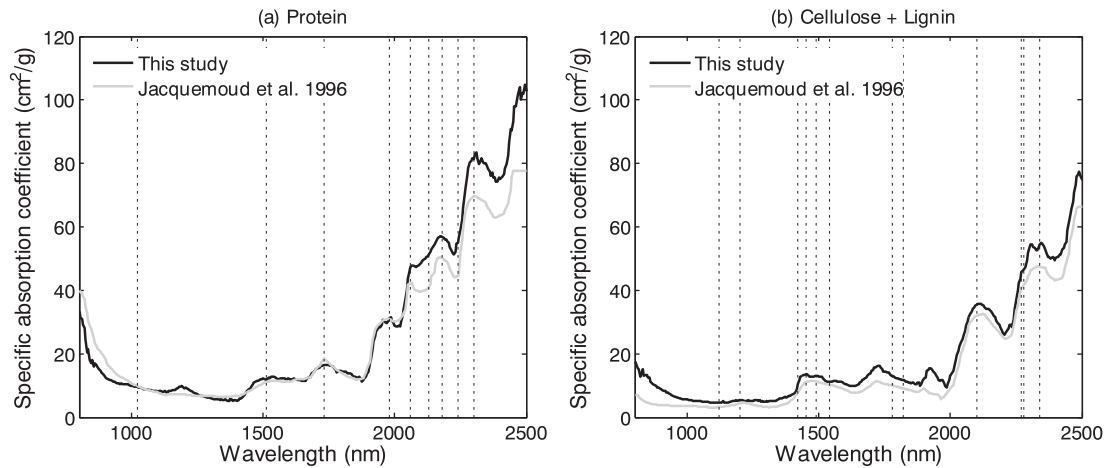


Fig. 2. Recalibrated specific absorption coefficients of protein (a) and cellulose + lignin (b) determined on dry leaves. The vertical dashed lines correspond to published absorption features of protein (a) and cellulose + lignin (b) (see details in Tables 4 and 5).

where the refractive index and specific absorption coefficients are fixed as constant values from the model calibration. The parameters are leaf chlorophyll content (Chl, $\mu\text{g}/\text{cm}^2$), LMA (g/cm^2), and EWT (cm), respectively. Note that Eq. (1) differs from Eq. (3), because the influences of Chl, LMA and EWT on leaf optical properties were represented by the parameter $k(\lambda)$, using Eq. (2).

The improved (1-nm resolution) and recalibrated version, PROSPECT-5, which was published in 2008 (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008), is used in this study. We performed the recalibration in three steps following the algorithm presented in Feret, Francois, Asner, Gitelson, Martin, Bidet, et al. (2008). The first step was to determine the leaf structure parameter through inversion of the PROSPECT model using the “three wavelengths method” (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). The maximum incident angle α was set to a more realistic value of 40° in this study, as suggested by PROSPECT-5 (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008). In the second step, we calculated the absorption coefficient of one elementary layer $k(\lambda)$ and the refractive index $n(\lambda)$ by inverting the model for a second time and using the leaf structure parameter determined in the first step. The derived refractive index in this study was close to that in PROSPECT-5, and therefore the refractive index from PROSPECT-5 was adopted for later analysis, as

that calibration involved more comprehensive datasets (Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008).

In the third step, we deduced the specific absorption coefficients of leaf protein and cellulose + lignin from Eq. (2), using the absorption coefficient of one elementary layer $k(\lambda)$ computed in the second step, together with the measured leaf water, protein, cellulose + lignin content from the LOPEX dataset. As suggested by Feret, Francois, Asner, Gitelson, Martin, Bidet, et al. (2008), the specific absorption coefficient of water did not need recalibration, so we used the values in PROSPECT-5. In this step, we divided the dry matter into protein, cellulose and lignin, and chose to study the 800–2500 nm region, because water, protein, cellulose and lignin absorb solar radiation mostly in this range (Baret & Fourty, 1997a; Curran, 1989; Feret, Francois, Asner, Gitelson, Martin, Bidet, et al., 2008). We combined the cellulose and lignin as one constituent (“cellulose + lignin”). Different combinations of leaf constituents were tested as input parameters to PROSPECT by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996), who found that the combination of water, protein, and cellulose + lignin best explained the variation in leaf optical properties in the 800–2500 nm range (Jacquemoud, personal communication).

The inversion was performed using a bounded optimization package, FMINSEARCHBND.M in MATLAB (The MathWorks, Inc.). The MATLAB code is available online provided by (D’Errico, 2012). The

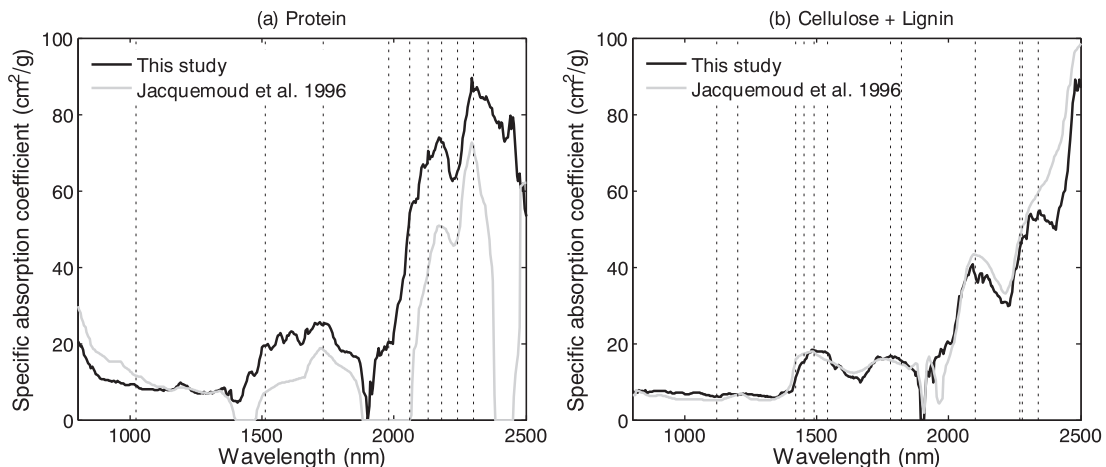


Fig. 3. Recalibrated specific absorption coefficients of protein (a) and cellulose + lignin (b) determined on fresh + dry leaves. The vertical dashed lines correspond to published absorption features of protein (a) and cellulose + lignin (b) (see details in Tables 4 and 5).

Table 4

Absorption peaks of protein reported in Curran (1989) and Fourty, Baret, Jacquemoud, Schmuck and Verdebout (1996), and in recalibrated specific absorption coefficients (SAC) from fresh, dry and fresh + dry leaves.

Wavelength (nm) Literature	Wavelength (nm) SAC from fresh leaves	Wavelength (nm) SAC from dry leaves	Wavelength (nm) SAC from fresh + dry leaves	Absorption mechanisms
1020	–	–	–	N–H stretch
1510	1520	1510	1520	N–H stretch, 1st overtone
1730	1750	1730	1730	C–H stretch
1980	–	1980	–	N=H asymmetry
2060	–	2060	–	N=H bend, 2nd overtone, N=H bend, N–H stretch
2130	–	–	–	N–H stretch
2180	2165	2180	2170	N–H bend, 2nd overtone, C–H stretch, C–O stretch, C=O stretch C–N stretch
2240	–	–	–	C–H stretch
2300	2295	2300	2300	N–H bend, C=O stretch, C–H bend, 2nd overtone

solution method only guarantees a local minimum, but after extensive tests with sets of initial values it was concluded that a global minimum had been found. The calibration process was performed separately for fresh, dry and fresh + dry leaves. The recalibrated specific absorption coefficients of protein, and cellulose + lignin from each group were compared with previous results reported by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996).

2.3. Sensitivity analysis

Sensitivity analysis (SA) helps to identify the contribution of the variability in input parameters to the variability in the output reflectance or transmittance. Two types of SA have been listed in literature, local SA and global SA. Local SA provides information on how the variation of each input parameter individually explains the variation in the model output and ignores the interactions between model parameters (Asner, 1998; Saltelli, 1999), while global SA provides information of how the variation of each input parameter individually and their interactions with each other accounts for the variation of model output. Global SA is required when searching the full range of model parameter space, including the simultaneous variations of model parameters (Bowyer & Danson, 2004), therefore, a global SA was adopted in this study.

SA calculates the fractional contribution of a given input variable X to the variance of Y . Supposing a model with three input parameters (Ceccato, Flasse & Grégoire, 2002), then the total variance $V(Y)$ of the model output is decomposed as follows

$$V = V_1 + V_2 + V_3 + V_{12} + V_{13} + V_{23} + V_{123} \quad (4)$$

where V_1 is the variance of input parameter 1, V_{12} is the variance of interaction between parameters 1 and 2, and V_{123} is the variance of interaction between parameters 1, 2 and 3, and so on.

The first-order sensitivity index S_1 for parameter 1 is defined as

$$S_1 = V_1/V. \quad (5)$$

Similarly, the second-order sensitivity index S_{12} can be calculated from

$$S_{12} = V_{12}/V. \quad (6)$$

And the total-order indices ST for parameter 1 is measured as follows

$$ST = (V_1 + V_{12} + V_{13} + V_{123})/V. \quad (7)$$

The first-order sensitivity indices represent the independent effect of the corresponding parameters while the second and higher order terms give the interaction effects of each parameter with others. The Fourier amplitude sensitivity test (FAST) method calculates the contribution of each input parameter to the variation in the output efficiently (Saltelli & Bolado, 1998; Xiao, Zhao, Zhou & Gong, 2014), which refers to first-order sensitivity indices. The method of Sobol (Sobol, 1993) was superior to FAST in computing the higher interaction terms (Dejonge, Ascough, Ahmadi, Andales & Arabi, 2012; Xiao, Zhao, Zhou & Gong, 2014). A Matlab software tool (GSAT) (Cannavó, 2012) was applied in this study to perform global sensitivity analysis, in which both FAST analysis and Sobol's analysis were implemented to calculate first order sensitivity coefficients and joint global sensitivity coefficients for model parameter combinations. The global sensitivity analysis was separately performed for fresh and dry leaves using the ranges of input parameters presented in Table 1.

Table 5

Absorption peaks of cellulose + lignin reported in Curran (1989) and Fourty, Baret, Jacquemoud, Schmuck and Verdebout (1996), and in recalibrated specific absorption coefficients (SAC) from fresh, dry and fresh + dry leaves.

Wavelength (nm) Literature	Wavelength (nm) SAC from fresh leaves	Wavelength (nm) SAC from dry leaves	Wavelength (nm) SAC from fresh + dry leaves	Absorption mechanisms
1120	–	–	–	C–H stretch, 2nd overtone
1200	1200	1200	1200	O–H bend, 1st overtone
1420	–	–	–	C–H stretch, C–H deformation
1450	1450	1450	1460	O–H stretch, 1st overtone, C–H stretch, C–H deformation
1490	–	1490	1470	O–H stretch, 1st overtone
1540	–	1540	–	O–H stretch, 1st overtone
1780	1760	–	–	C–H stretch, 1st overtone, O–H stretch, H–O–H deformation
1820	–	–	–	O–H stretch, C–O stretch, 2nd overtone
2100	2100	2100	2100	O=H bend, C–O stretch, C–O–C stretch, 3rd overtone
2270	–	–	–	C–H stretch, O–H stretch, CH ₂ bend, CH ₂ stretch
2280	–	–	–	C–H stretch, CH ₂ deformation
2340	2340	2340	2340	C–H stretch, O–H deformation, C–H deformation, O–H stretch

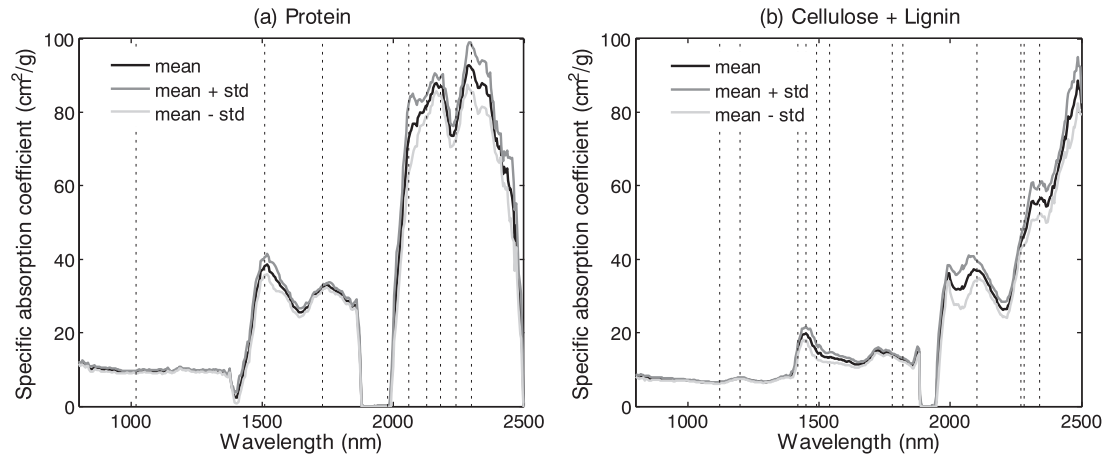


Fig. 4. Mean and variability of the recalibrated specific absorption coefficients of protein and cellulose + lignin from 100 repetitions of k-fold cross-validation (fresh leaves).

2.4. Validation of the recalibrated PROSPECT-5 model in forward mode

The k-fold cross validation was performed due to the limited number of samples, which enables independent data set used in the calibration and validation. The commonly used 10-fold cross-validation was adopted in this study. Detailed descriptions are presented as follows.

- (1) The original samples were randomly partitioned into k equal size subsamples ($k = 10$).
- (2) The $k - 1$ subsamples were used in the calibration to determine the specific absorption coefficients of protein, and cellulose + lignin (see Section 2.2).
- (3) The remaining one subsample was retained as the validation data for testing the performance of the recalibrated model in forward modeling leaf reflectance and transmittance, by using the calibrated specific absorption coefficients from the previous step, the measured content of constituents (e.g., water, protein, cellulose + lignin) and estimated leaf structure parameter in the calibration.
- (4) The cross-validation procedure was then repeated k times with each one of the k subsamples used once as the validation data, after which the modeled leaf reflectance and transmittance of all samples can be obtained.

The k-fold cross validation (steps 1–4) was repeated 100 times to avoid biased results (Schlerf, Atzberger & Hill, 2005). We used the

measured content of constituents (e.g., water, protein, cellulose + lignin), their corresponding recalibrated specific absorption coefficients, and the estimated leaf structure parameter to simulate the reflectance and transmittance in the forward mode. Simulated leaf reflectance and transmittance were also obtained by using the specific absorption coefficients presented by Jacquemoud et al. (1996). For comparison, this simulation was also implemented using the measured content of constituents (e.g., water, dry matter), their specific absorption coefficients in PROSPECT-5, and the estimated leaf structure parameter.

The root mean square error of prediction (RMSEP), the bias (BIAS), and the standard error of prediction corrected from the bias (SEPC) between the modeled and measured reflectance and transmittance were calculated for each wavelength. The averaged RMSEP, BIAS, SEPC of the 100 repetitions were used to evaluate the model performance using calibrated specific absorption coefficients from this study. Let y_j and y_j' be the measured and estimated values, respectively, $\bar{y}$ be the mean of the measured values, and n be the number of measurements:

$$\text{RMSEP} = \sqrt{\frac{\sum_{j=1}^n (y_j' - y_j)^2}{n}} \quad (8)$$

$$\text{BIAS} = \frac{\sum_{j=1}^n (y_j' - y_j)}{n} \quad (9)$$

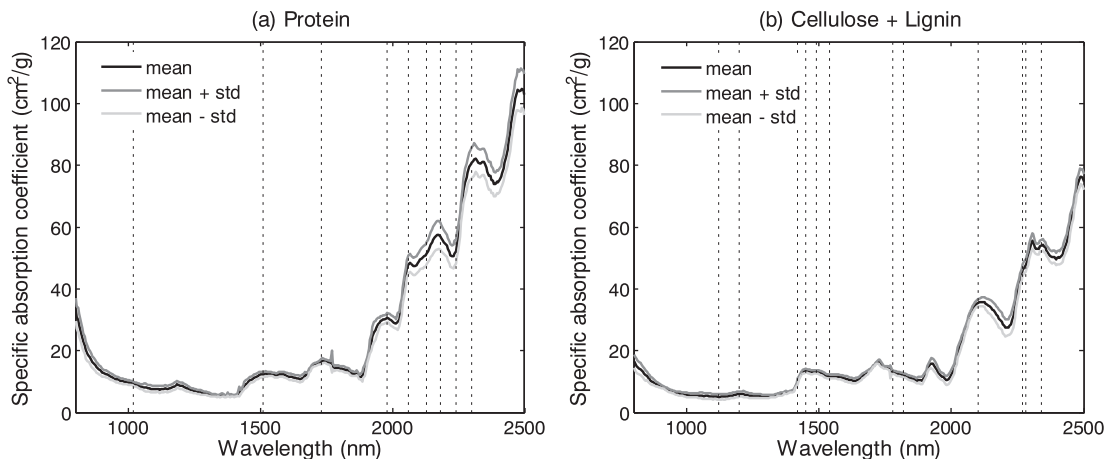


Fig. 5. Mean and variability of the recalibrated specific absorption coefficients of protein and cellulose + lignin from 100 repetitions of k-fold cross-validation (dry leaves).

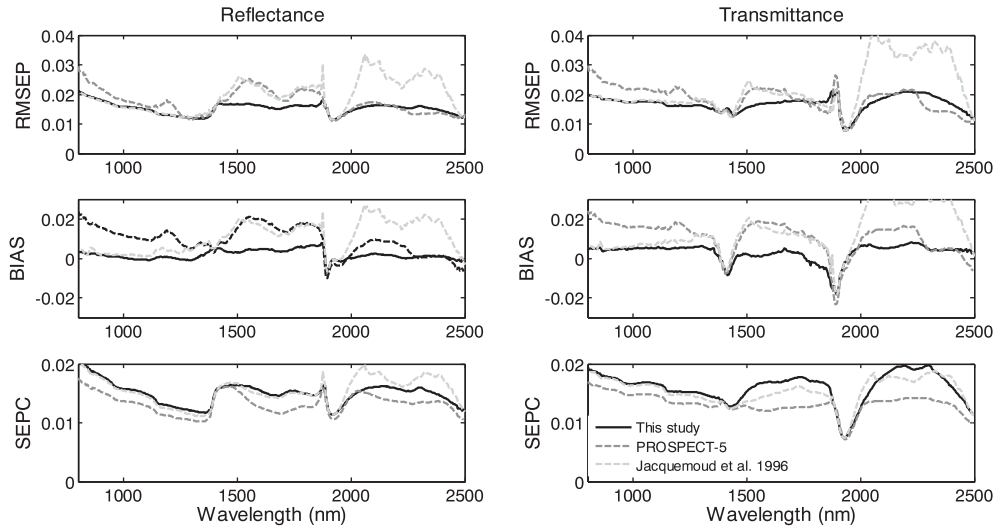


Fig. 6. RMSEP, BIAS and SEPC of the modeling of the reflectance and transmittance of fresh leaves by this study, Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996) and PROSPECT-5. RMSEP, BIAS and SEPC of this study were averaged values from 100 repetitions of k-fold cross validation.

$$SEPC = \sqrt{\frac{\sum_{j=1}^n (y_j' - y_j - \text{BIAS})^2}{n}} \quad (10)$$

2.5. Estimation of leaf parameters

2.5.1. PROSPECT model inversion

The input model parameters included leaf structure parameter (N), EWT, protein content (C_p), and cellulose + lignin content (C_{cl}). The inversion process was to find the parameter vector $\theta = [N, \text{EWT}, C_p, C_{cl}]^T$, which minimizes the merit function:

$$J(\theta) = \sum_{\lambda_1}^{\lambda_2} (R_{mes}(\lambda) - R_{mod}(\lambda, \theta))^2 + (T_{mes}(\lambda) - T_{mod}(\lambda, \theta))^2 \quad (11)$$

where λ is the wavelength, R_{mes} and T_{mes} are respectively the measured reflectance and transmittance, and R_{mod} and T_{mod} are the modeled

values. The ranges of the input parameters were determined based on the dataset used in this study (Table 1) and are given in Table 2. The parameters were estimated through model inversion using an iterative optimization algorithm. Specific absorption coefficients of protein, and cellulose + lignin determined by the pooled samples were utilized to involve the variations of leaf optical and biochemical properties across species.

To alleviate the “ill-posed” inverse problems, a specific merit function was assigned for each parameter defined on its own sensitive spectral bands, which has been proved to improve the performance of inversion results (Li & Wang, 2011; Wang et al., 2015; Zhao, Guo, Huang, Reddy, Lee, Fletcher, et al., 2014). Thus, we performed the model inversion for retrieving leaf parameters not only over the full wavelength range (1100–2500 nm, subset 1), but also over another six spectral subsets (Table 3). Subset 2 was the sensitive wavelength region for protein and cellulose + lignin, according to the global sensitivity analysis for fresh leaves (see Section 2.3). Subset 3 was also defined for protein and cellulose + lignin, but based on the global SA for dry leaves. Subset 4 was a collection of protein absorption bands from

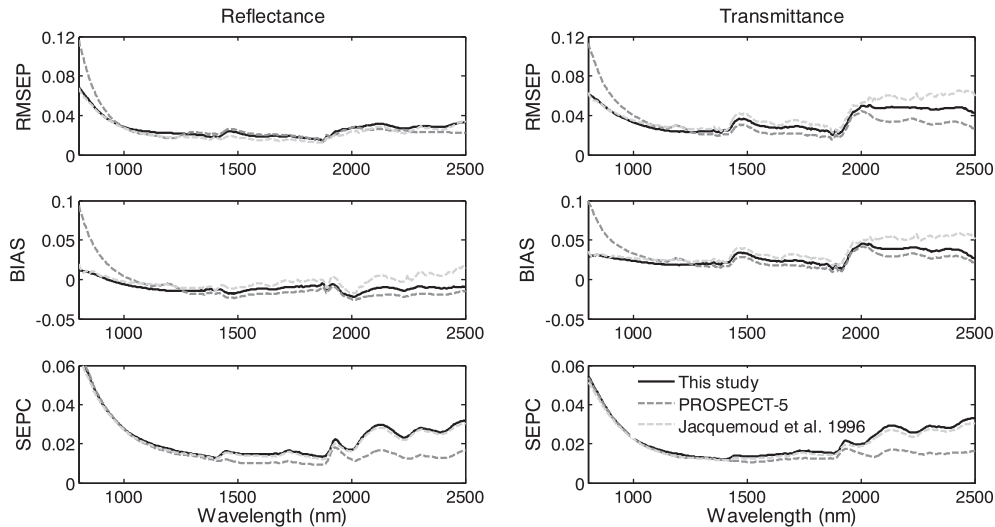


Fig. 7. RMSEP, BIAS and SEPC of the modeling of the reflectance and transmittance of dry leaves by this study, Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996) and PROSPECT-5. RMSEP, BIAS and SEPC of this study were averaged values from 100 repetitions of k-fold cross validation.

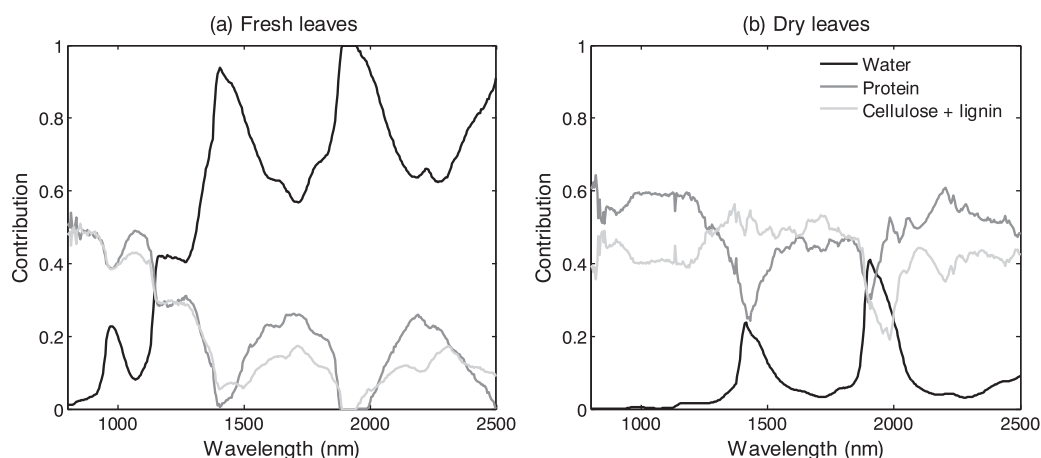


Fig. 8. Contribution of water, protein and cellulose + lignin to the total leaf absorption (800–2500 nm) in fresh (a) and dry (b) leaves.

previous findings (Curran, 1989; Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996). Subset 5 was a collection of cellulose + lignin absorption bands (Curran, 1989; Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996). Subset 4 and 5 were involved to assess the plausibility of the calibrated specific absorption coefficients of protein and cellulose + lignin, and to test whether the subsets could achieve an equal or even improved estimation accuracy of leaf protein and cellulose + lignin compared to using the full wavelengths.

2.5.2. Stepwise multiple linear regression

We also applied a classical statistical technique, stepwise multiple linear regression (SMLR) to select sensitive spectral subsets for protein and cellulose + lignin, and to compare its performance of estimation with that of PROSPECT model. Though a number of limitations have been pointed out for SMLR (Majeke, van Aardt & Cho, 2008), it serves as a straightforward way for comparison with inversion models. The stepwise multiple linear regression relates the protein or cellulose + lignin content with the reflectance or transmittance. The model starts with no independent variables (e.g. reflectance), adds the variable which was the best statistical predictor, and adds the second variable which gives the best fit together with the first variable. Further variable was added to provide an optimal fit in a recursive way (Jacquemoud, Verdebout, Schmuck, Andreoli & Hosgood, 1995). The SMLR was performed using MATLAB (The MathWorks, Inc.) over the 1100–2500 nm

in order to compare results with those from PROSPECT inversion. We restricted the number of selected wavelengths to four to have a simple model and avoid over-fitting. Subsets 6 and 7 in Table 3 were determined on the selected wavelengths by SMLR for predicting protein and cellulose + lignin, respectively.

The accuracies of the retrieved parameters (EWT , C_p , C_{cl}) via PROSPECT model inversion and SMLR were evaluated by the coefficient of determination (R^2), the root mean square error (RMSE), and the normalized RMSE ($NRMSE = RMSE / range$) between the estimated and measured values.

3. Results

3.1. Recalibrated specific absorption coefficients for protein and cellulose + lignin

The recalibrated specific absorption coefficients for protein and cellulose + lignin over 800–2500 nm determined on fresh leaves, dry leaves, and fresh + dry leaves are shown in Figs. 1, 2, and 3, respectively. For dry leaves, most of the absorption peaks of protein and cellulose + lignin were well represented (Fig. 2), and are consistent with findings by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996). For fresh leaves, fewer absorption peaks were displayed (see dashed vertical lines in Fig. 1), and discrepancies were found between

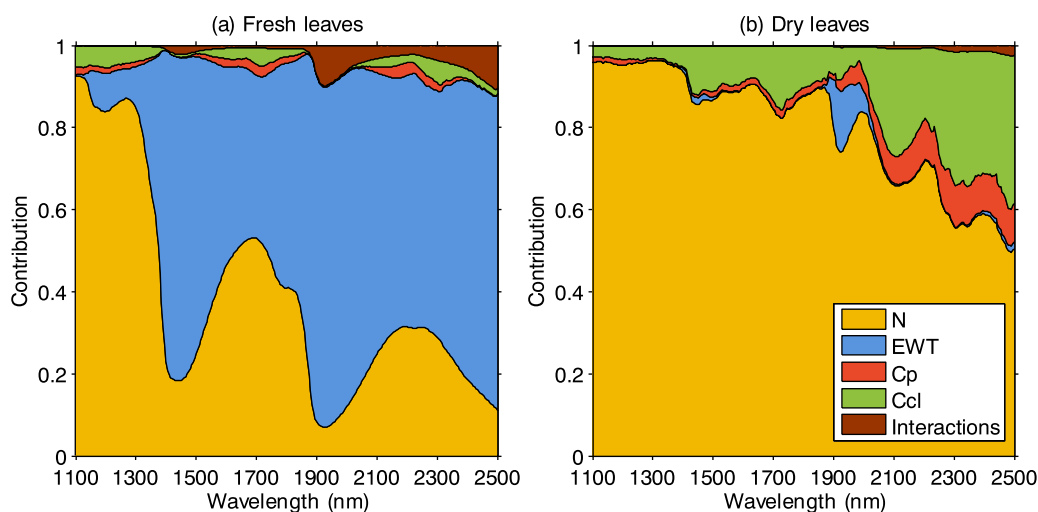


Fig. 9. Results of FAST first order sensitivity coefficients and interactions to leaf reflectance for global sensitivity analysis with PROSPECT model in (a) fresh leaves, (b) dry leaves. The legend is given in (b), where N is the leaf structure parameter, EWT is the equivalent water thickness, C_p is the protein content, and C_{cl} is the cellulose + lignin content.

Table 6

Validation of equivalent water thickness (EWT), protein and cellulose + lignin content retrieval for fresh leaves using recalibrated PROSPECT-5.

Spectral subset	EWT (cm)			Protein (g/cm ²)			Cellulose + lignin (g/cm ²)		
	R ²	RMSE	NRMSE	R ²	RMSE	NRMSE	R ²	RMSE	NRMSE
Subset 1	0.95	0.0015	0.04	0.45	3.16E−04	0.19	0.40	8.66E−04	0.20
Subset 2	0.89	0.0022	0.06	0.26	3.81E−04	0.23	0.44	9.19E−04	0.21
Subset 3	0.91	0.0020	0.06	0.47	2.75E−04	0.17	0.70	5.21E−04	0.12
Subset 4	0.94	0.0017	0.05	0.44	3.11E−04	0.19	0.46	7.77E−04	0.18
Subset 5	0.96	0.0014	0.04	0.48	3.12E−04	0.19	0.42	8.24E−04	0.19
Subset 6 ^a	0.83	0.0030	0.08	–	8.13E−04	0.49	0.33	9.51E−04	0.22
Subset 7 ^a	0.91	0.0029	0.08	–	8.68E−04	0.52	0.61	8.48E−04	0.19

the absorption peaks in specific absorption coefficients and absorption features in literature (Tables 4 and 5). Generally, absorption positions were consistent for coefficients determined on fresh leaves, dry leaves and fresh + dry leaves, except for those masked by the strong water absorption at 1800–2000 nm. 100 repetitions of k-fold cross-validation generated quite consistent specific absorption coefficients for protein and cellulose + lignin (Figs. 4 and 5). The coefficients of protein and cellulose + lignin for dry leaves display minor differences with results in a previous study (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). However, the recalibrated specific absorption coefficient of protein for fresh leaves yielded significantly different results. The coefficients of protein for fresh leaves in this study are much higher (with a peak value of over 80 cm²/g) than results reported by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996).

3.2. Validation of the performance of recalibrated PROSPECT model in forward mode

The modeling of leaf reflectance and transmittance as performed in this study is clearly more accurate than that with the original PROSPECT-5 (Figs. 6 and 7), despite the fact that the sum of protein and cellulose + lignin occupied two-thirds to three-quarters of the dry leaf mass, which was also confirmed by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996). The spectral RMSEP's were generally less than 0.03, which confirms the capability of the PROSPECT model when recalibrated for simulating leaf optical properties using only a few parameters (Jacquemoud & Baret, 1990; Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). For fresh leaves (Fig. 6), the recalibrated specific absorption coefficients greatly improved the performance of modeling leaf optical properties over the range of 1400–1800 nm and 2100–2500 nm. The modeled fresh leaf spectra using coefficients from Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996) gave the least accurate estimate in the range of 2100–2500 nm.

3.3. Contribution of different constituents to the total leaf absorption

The contribution of each constituent to the total leaf absorption coefficient was computed as (Baret & Fourty, 1997a; Fourty, Baret,

Jacquemoud, Schmuck & Verdebout, 1996):

$$k_i(\lambda) \cdot \bar{C}_i / \left(\sum_i k_i(\lambda) \times \bar{C}_i \right) \quad (12)$$

where $\bar{C}_i$ is the average content of the constituent i , and $k_i(\lambda)$ the corresponding specific absorption coefficient. This can provide information on identifying the sensitive bands for each constituent used in model inversion. Fig. 8 demonstrates the contribution of water, protein and cellulose + lignin to the total leaf absorption coefficient. It was not surprising to see the largest contribution of water content in fresh leaves over the whole range, and two peaks were located around 1400–1600 nm and 1800–2000 nm. Protein and cellulose + lignin have a relatively larger contribution over 1600–1800 nm and 2100–2300 nm. The contribution of protein was generally larger than that of cellulose + lignin. The protein and cellulose + lignin contributed approximately 40% to the total leaf absorption in these spectral ranges. For dry leaves, the contribution of water showed similar trends (in terms of spectral regions), but was generally lower (less than 50% contribution to total leaf absorption) than that of the other two constituents. Cellulose + lignin contributed most to leaf absorption over 1300–1800 nm, while protein contributed more after 1900 nm, which is consistent with reports by Fourty, Baret, Jacquemoud, Schmuck and Verdebout (1996).

3.4. Global sensitivity analysis

Fig. 9 demonstrates the results of FAST first order sensitivity coefficients and interactions to leaf reflectance for global sensitivity analysis with PROSPECT model. Global sensitivity analysis showed that the sum of FAST first order sensitivity coefficients (N , EWT, C_p , C_d) was close to 100% for both fresh and dry leaves, which indicated that a minor contribution was from interactions of these parameters to the leaf reflectance variability. For fresh leaves, leaf structure parameter N had the greatest contribution, followed by EWT. In contrast, the influence of cellulose + lignin and protein was limited, and was relatively greater in the spectral ranges of 800–1300 nm, 1600–1800 nm and 2100–2300 nm. Compared to fresh leaves, the influence of both protein and cellulose + lignin on the leaf spectra for dry leaves was much greater in wavelengths longer than 2100 nm, and the sum of their contributions amounted to around 40% (Fig. 9(b)). Leaf structure parameter still

Table 7

Validation of equivalent water thickness (EWT), protein and cellulose + lignin content retrieval for dry leaves using recalibrated PROSPECT-5.

Spectral subset	EWT (cm)			Protein (g/cm ²)			Cellulose + lignin (g/cm ²)		
	R ²	RMSE	NRMSE	R ²	RMSE	NRMSE	R ²	RMSE	NRMSE
Subset 1	0.51	9.67E−05	0.11	0.50	3.96E−04	0.20	0.59	6.81E−04	0.13
Subset 2	–	4.96E−04	0.56	0.37	4.83E−04	0.25	0.41	8.24E−04	0.15
Subset 3	–	5.16E−04	0.58	0.57	3.55E−04	0.18	0.79	5.05E−04	0.09
Subset 4	0.32	1.06E−04	0.14	0.52	3.90E−04	0.20	0.68	5.72E−04	0.11
Subset 5	0.27	1.55E−04	0.18	0.43	4.84E−04	0.25	0.21	1.11E−03	0.21
Subset 6 ^b	0.32	1.43E−04	0.16	0.57	3.71E−04	0.19	0.62	6.88E−04	0.13
Subset 7 ^b	0.49	1.18E−04	0.13	0.11	7.61E−04	0.39	0.74	7.23E−04	0.14

Table 8

Validation of protein and cellulose + lignin content retrieval for fresh and dry leaves using stepwise multiple linear regression.

	Protein (g/cm ²)			Cellulose + lignin (g/cm ²)		
	R ²	RMSE	NRMSE	R ²	RMSE	NRMSE
Fresh leaves	0.66	2.02E-04	0.12	0.83	3.91E-04	0.09
Dry leaves	0.74	2.32E-04	0.12	0.79	4.50E-04	0.08

acted as the dominant role in driving the leaf reflectance variability, but the influence of water was only noticeable in 1880–2000 nm. Similar features were observed for transmittance in global sensitivity analysis (results not shown).

3.5. Accuracy of estimated leaf parameters

The accuracy of the estimation for EWT, protein and cellulose + lignin content using the recalibrated PROSPECT-5 inversion is shown in Table 6 (fresh leaves) and Table 7 (dry leaves). For fresh leaves, EWT was estimated well by all seven spectral subsets, with an R^2 around 0.90, and RMSE of 0.0020 cm. Spectral subset 3 gave the most accurate estimation of cellulose + lignin ($R^2 = 0.70$; RMSE = 5.21E-04 g/cm²), followed by subset 7^a ($R^2 = 0.61$; RMSE = 8.48E-04 g/cm²). The estimation of protein based on spectral subset 5 yielded the highest R^2 value of 0.48 and a RMSE of 3.12E-04 g/cm², followed by that based on the subset 3 ($R^2 = 0.47$; RMSE = 2.75E-04 g/cm²). For dry leaves, EWT was generally not well assessed using all the spectral subsets, and cellulose + lignin values were estimated with a higher accuracy than protein values. Subset 3 with the region of 2100–2300 nm

provided the most accurate retrieval for both cellulose + lignin ($R^2 = 0.79$; RMSE = 5.05E-04 g/cm²) and protein ($R^2 = 0.58$; RMSE = 3.55E-04 g/cm²). The subset 7 based on selected wavelengths for predicting cellulose + lignin using SMLR, gave a robust estimation for both fresh and dry leaves when applying the subset in model inversion. The selected subset 6^b for protein based on SMLR on dry leaves also provide robust estimation in model inversion, but it was not valid for the subset 6^b on fresh leaves. Table 8 presents the accuracy of predicted protein and cellulose + lignin using stepwise multiple linear regression. Generally, the stepwise multiple linear regression outperformed in estimations of protein and cellulose + lignin over PROSPECT inversion, especially for protein in fresh leaves (Figs. 10 and 11).

4. Discussion

The leaf protein and cellulose + lignin content were moderately well estimated for both fresh and dry leaves with the recalibrated specific absorption coefficients when PROSPECT-5 was inverted. Comparable accuracies were obtained for estimating cellulose + lignin on fresh and dry leaves through PROSPECT model inversion and stepwise multiple linear regression, irrespective of the water content. Despite of the low amount of leaf protein in fresh leaves, it can still be retrieved via model inversion at promising accuracies. Estimating leaf protein in fresh leaves, using inversion of a radiative transfer model, has been considered impossible since the 1990s (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996), although empirical methods have achieved relatively good accuracies (Kokaly, 2001; Kokaly, Asner, Ollinger, Martin & Wessman, 2009; Kokaly & Clark, 1999; Martin, Plourde, Ollinger, Smith & McNeil, 2008). To our knowledge, this is the

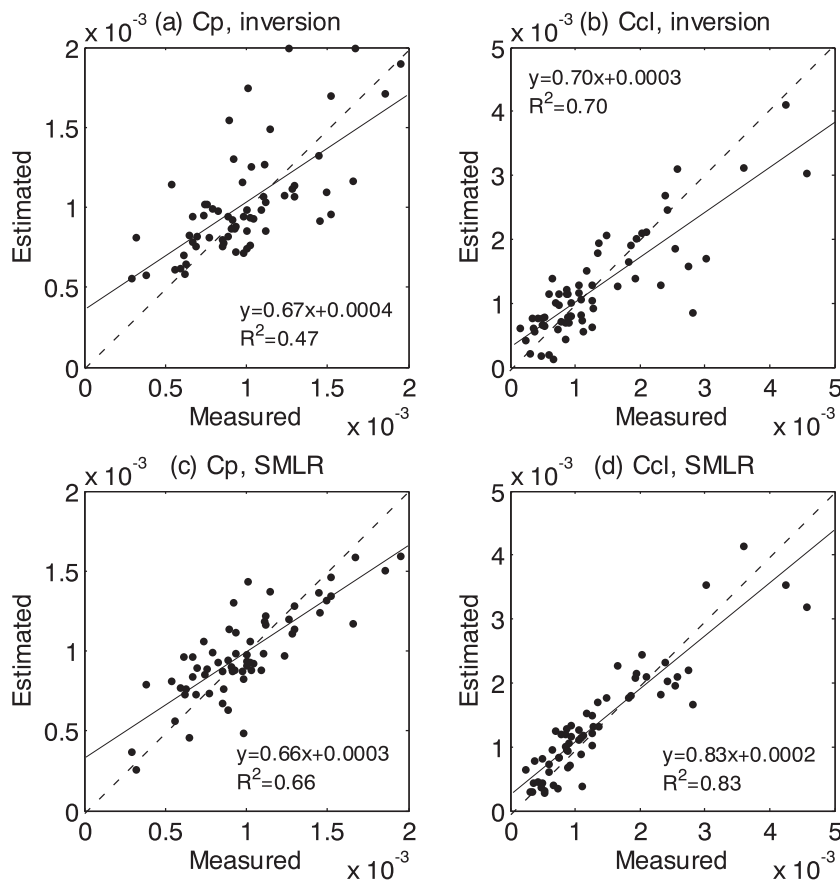


Fig. 10. Measured versus estimated protein content (Cp, g/cm²) and cellulose + lignin content (Ccl, g/cm²) for fresh leaves using PROSPECT-5 inversion with recalibrated specific absorption coefficients and stepwise multiple linear regression (SMLR).

first study to prove the feasibility of estimating leaf protein and cellulose + lignin in fresh leaves at a moderate to good accuracy using a radiative transfer model.

The specific absorption coefficients of leaf protein and cellulose + lignin determined on fresh, dry and fresh + dry leaves displayed similar magnitudes (Figs. 1, 2 and 3). The coefficients were also quite consistent across species (Figs. 4 and 5) which can be noted from a repetitive 10-fold cross validation, in each repetition a different data set of species was randomly selected. The findings demonstrated the robustness of the recalibrated specific absorption coefficients, which holds great potential in the development for PROSPECT model given the difficulties in measuring in vivo coefficients. Absorption peaks found in the coefficients were in good agreement with known absorption features from literature regardless of the leaf water status (Figs. 1, 2 and 3, Tables 4 and 5). This holds promise for building a general model across a large variety of species and leaf water. Most of known absorption peaks of protein and cellulose + lignin were displayed well in coefficients determined on dry leaves, while fewer absorption features were found in those determined on fresh leaves. The stronger absorption of leaf water on fresh leaves masks some absorption features of protein and cellulose + lignin, such as 1940, 1980 and 2060 nm. And it also led to the off-center wavelength (5–20 nm) of absorption peaks in coefficients from known absorption features in literature, such as 1510, 1730, 2180 nm for protein, and 1780 nm for cellulose + lignin (Tables 4 and 5).

The recalibrated specific absorption coefficients showed similar features for protein (in dry leaves) and cellulose + lignin (in both fresh, dry and fresh + dry leaves) to a previous study (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). However, our

results showed great differences in the behavior of the coefficient of protein for fresh leaves to those reported by Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996). The difference lays not just in the shape of the coefficient curve but also in its magnitude. Our absorption peaks of protein reached approximately 80 cm²/g rather than their value of 20 cm²/g. This new insight for protein in fresh leaves provides a solid foundation for estimating leaf protein via model inversion. Such improvements can be largely attributed to the new calibration algorithm proposed by Feret, Francois, Asner, Gitelson, Martin, Bidet, et al. (2008). The improvements of the coefficients were further confirmed when applying them in PROSPECT model for forward simulating leaf reflectance and transmittance. The leaf optical properties for fresh leaves were more accurately modeled, especially over the ranges 1400–1800 nm and 2100–2500 nm. This means the constituents of protein and cellulose + lignin could explain the variations observed in leaf optical properties, although they only contribute a relatively small proportion to the dry matter. For dry leaves, a small improvement was noted when modeling leaf reflectance, however, the performance deteriorated when transmittance was modeled. This might be due to the limitation of the PROSPECT model in terms of dry leaves, which does not consider inter-correlations between model parameters (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). For example, as a leaf dries, the leaf structure changes with the lower water content.

The selected spectral subsets generally provided equally or more accurate estimation of cellulose + lignin and protein, compared with using full wavelengths of 1100–2500 nm. This confirmed that selected spectral subsets often present better results than using full wavelengths (Darvishzadeh, Skidmore, Schlerf & Atzberger, 2008; Meroni, Colombo

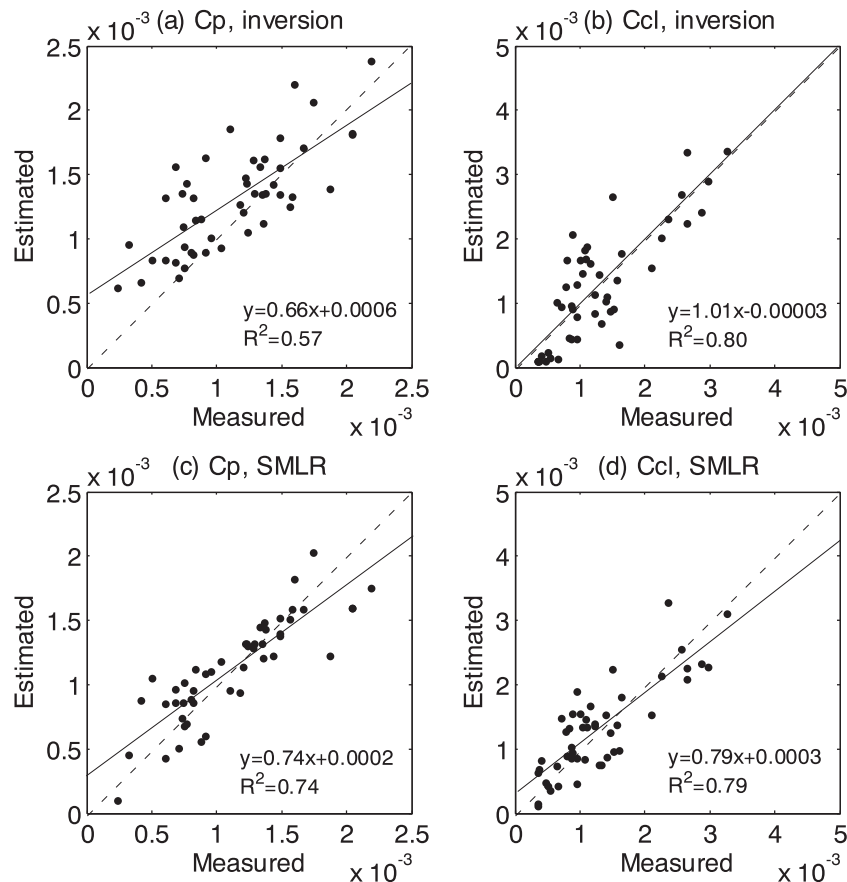


Fig. 11. Measured versus estimated protein content (Cp, g/cm²) and cellulose + lignin content (Ccl, g/cm²) for dry leaves using PROSPECT-5 inversion with recalibrated specific absorption coefficients and stepwise multiple linear regression (SMLR).

& Panigada, 2004; Schlerf & Atzberger, 2006). However, the spectral subset (Subset 4) defined based on known absorption features of protein did not guarantee more accurate estimation for protein than cellulose + lignin, and the same phenomenon was observed for a specific spectral subset (Subset 5) for cellulose + lignin. This can be explained by the overlapping absorption features due to their common C—H and O—H bonds (Curran, 1989; Fourty, Baret, Jacquemoud, Schmuck & Verdebout, 1996), and their overlapping absorption features with other secondary compounds (e.g. hemicellulose, starch) which may be in high content for some plant species. The subset 3 of 2100–2300 nm proved to be the most effective spectral range for estimating cellulose + lignin and protein on both fresh and dry leaves. This can be attributed to the less influence caused by water absorption, as well as the high values of specific absorption coefficients in this range.

The performances of PROSPECT model inversion and stepwise multiple linear regression were comparable in retrieving cellulose + lignin for both fresh leaves and dry leaves. However, less accurate estimations were obtained for protein when using model inversion other than SMLR. A reason could be the greater abundance of cellulose + lignin than protein in leaves. The relationship between estimated and measured values from model inversion appeared to be less linear than that from SMLR, which indicated larger uncertainties when retrieving parameters using PROSPECT. The uncertainties can be partly explained by the inherent differences lying in radiative transfer models and statistical models. All parameters were simultaneously retrieved through model inversion, thus the parameters making greater contribution to leaf optical properties, such as leaf structure parameter and EWT, inevitably affects the parameters having less influence, namely cellulose + lignin and protein. In contrast, the statistical model was developed exclusively for one parameter, which is less sensitive to other parameters.

We made efforts to alleviate the “ill-posed” inverse problem and uncertainties in estimation by selecting specific spectral subsets for cellulose + lignin and protein, but the improvement for retrieving protein was limited. Li and Wang (2011) proposed an approach of retrieving parameters in the order of their contributions to leaf reflectance, which is leaf structure parameter, leaf chlorophyll content, equivalent water thickness and leaf mass per area. In that study, one parameter was retrieved at each step by assigning a specific merit function using its own sensitive wavelengths. They first retrieved leaf structure parameter, then estimated the next parameter using the computed values from previous steps as fixed, and continued until all parameters were obtained. The “step-by-step” inversion approach improved the accuracies of estimated leaf chlorophyll content, equivalent water thickness and leaf mass per area, compared to simultaneously determining all parameters using a single merit function (Li & Wang, 2011; Zhao, Guo, Huang, Reddy, Lee, Fletcher, et al., 2014). However, the approach did not work well for estimating cellulose + lignin and protein in this study. Different spectral transformation methods have been used to minimize the influence of water on estimating leaf biochemical constituents in statistical studies, such as first or second derivatives, continuum removal (Curran, Dungan & Peterson, 2001; Huang, Turner, Dury, Wallis & Foley, 2004; Kokaly & Clark, 1999; Mutanga & Skidmore, 2004; Mutanga, Skidmore, Kumar & Ferwerda, 2005), and water removal through spectral matching (Gao & Goetz, 1994; Ramoelo, Skidmore, Schlerf, Mathieu & Heitkönig, 2011; Schlerf, Atzberger, Hill, Buddenbaum, Werner & Schueler, 2010). It is worthwhile to investigate such techniques specifically for inversion of a radiative transfer model.

From a practical perspective, the estimation of leaf biochemical content for fresh leaves could have wide application when using airborne or satellite platforms for sensing green canopy foliage (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996). So far, there has been a long-held perception, based on the results of Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli and Hosgood (1996), that leaf protein cannot be retrieved from fresh green leaf spectra when using radiative transfer models. Our results validate one radiative transfer

model, i.e. PROSPECT, for estimating leaf biochemical content for fresh leaves, which confirms the statistical models (Asner, Martin, Knapp, Tupayachi, Anderson, Carranza, et al., 2011; Martin, Plourde, Ollinger, Smith & McNeil, 2008; Mutanga & Skidmore, 2004; Skidmore, Ferwerda, Mutanga, Van Wieren, Peel, Grant, et al., 2010) that have consistently demonstrated leaf biochemical content being successfully retrieved from hyperspectral imagery. The ability to model leaf protein, and *ipso facto* leaf nitrogen content, holds the promise that eventually global maps of foliar nitrogen content could be generated when hyperspectral satellites such as Sentinel (Malenovsky, Rott, Cihlar, Schaepman, Garcia-Santos, Fernandes, et al., 2012), EnMAP (Kaufmann, Segl, Guanter, Hofer, Foerster, Stuffer, et al., 2008) and HypsIRI (<http://hypsiri.jpl.nasa.gov/>) become operational. In turn, such monitoring would support the development and demonstration of Essential Biodiversity Variables (EBVS), which are used for assessing biodiversity (Pereira, Ferrier, Walters, Geller, Jongman, Scholes, et al., 2013).

5. Conclusions

This study demonstrates, for the first time, that leaf protein and cellulose + lignin can be successfully estimated in fresh leaves at a moderate to good accuracy through inversion of a radiative transfer model. The specific absorption coefficients developed for PROSPECT-5 were recalibrated specifically for protein and cellulose + lignin using the new calibration algorithm proposed by Feret, Francois, Asner, Gitelson, Martin, Bidel, et al. (2008). The recalibrated coefficients were robustly estimated across species and showed a greater accuracy when reconstructing leaf reflectance and transmittance. Absorption peaks in coefficients correspond well with known protein and cellulose + lignin absorption features regardless of the leaf water status. The accuracy of retrieved protein and cellulose + lignin from PROSPECT model inversion was comparable with that from stepwise multiple linear regression. The selected spectral subsets improved the accuracy of estimating leaf protein and cellulose + lignin compared with using full wavelengths.

The use of a radiative transfer model offers a robust approach that should be transferable across species and test sites when compared with empirical methods (Darvishzadeh, Skidmore, Schlerf & Atzberger, 2008). The LOPEX dataset tested in this study contained plant species with a large variety of leaf structure, leaf biochemistry and leaf optical properties, further confirming the robustness and transferability of this approach. The statistical models can be used to support the development of radiative transfer models in selecting spectral subsets, which provided robust estimations in model inversion. This study was conducted at leaf level, which could be applied at the canopy level for regional or global mapping if coupled with a canopy reflectance model and suitable hyperspectral satellite imagery. However, the confounding factors, such as canopy structure, solar and viewing geometry, and soil background need to be fully considered before upscaling. Canopy level protein (equivalent to nitrogen) (Jacquemoud, Ustin, Verdebout, Schmuck, Andreoli & Hosgood, 1996) and cellulose + lignin (approximation of carbon) would not only yield valuable information about photosynthesis and net primary production, but would also allow the C:N ratio to be estimated, which can be used for evaluating the nutrient cycle.

Acknowledgments

This work of the first author was supported by the China Scholarship Council under Grant 201204910232 and co-funded by ITC Research Fund under Grant 93003032 from the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands. We gratefully acknowledge the open access to the PROSPECT model as well as to the LOPEX dataset. We thank Jackie Senior for editing the manuscript and the three anonymous reviewers for their valuable comments.

References

- Aber, J., & Federer, C.A. (1992). A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia*, 92, 463–474.
- AOAC (1970). *Official methods of analysis* (11th ed.). Washington, DC: Association of Official Analytical Chemists.
- Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64, 234–253.
- Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C., Carranza, L., et al. (2011). Spectroscopy of canopy chemicals in humid tropical forests. *Remote Sensing of Environment*, 115, 3587–3598.
- Baret, F., & Fourty, T. (1997a). Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie*, 17, 455–464.
- Baret, F., & Fourty, T.H. (1997b). Radiometric estimates of nitrogen status of leaves and canopies. In G. Lemaire (Ed.), *Diagnosis of the Nitrogen Status in Crops* (pp. 201–227). Springer Berlin Heidelberg.
- Bian, M. (2013). *Assessing the quality of tea by hyperspectral techniques* (pp. 143). Enschede, the Netherlands: University of Twente Faculty of Geo-Information and Earth Observation (ITC).
- Botha, E.J., Zebbarth, B.J., & Leblon, B. (2006). Non-destructive estimation of potato leaf chlorophyll and protein contents from hyperspectral measurements using the PROSPECT radiative transfer model. *Canadian Journal of Plant Science*, 86, 279–291.
- Bowyer, P., & Danson, F.M. (2004). Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. *Remote Sensing of Environment*, 92, 297–308.
- Buiteveld, H., Hakvoort, J.H.M., & Donze, M. (1994). The optical properties of pure water. *Ocean Opt.*, 13, 2258, 174–183.
- Cannavó, F. (2012). Sensitivity analysis for volcanic source modeling quality assessment and model selection. *Computers & Geosciences*, 44, 52–59.
- Ceccato, P., Flasse, S., & Grégoire, J.M. (2002). Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. *Remote Sensing of Environment*, 82, 188–197.
- Curran, P.J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30, 271–278.
- Curran, P.J., Dungan, J.L., & Peterson, D.L. (2001). Estimating the foliar biochemical concentration of leaves with reflectance spectrometry testing the Kokaly and Clark methodologies. *Remote Sensing of Environment*, 76, 349–359.
- Darvishzadeh, R., Atzberger, C., Skidmore, A., & Schlerf, M. (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: A comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 894–906.
- Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, 112, 2592–2604.
- Dejonge, K.C., Ascough, J.C., Ahmadi, M., Andales, A.A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, 231, 113–125.
- D'Errico, J. (2012). *Fminsearchbnd*, *fminsearchcon* (<http://www.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon/content/FMINSEARCHBND/fminsearchbnd.m>).
- Ewerts (1985). 3ème Directive avec rectification (J.O.C.E. du 27 novembre 1980). *Recueil des Normes Françaises-AFNOR*.
- Feret, J.-B., Francois, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidet, L.P.R., et al. (2008). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, 112, 3030–3043.
- Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., & Verdebout, J. (1996). Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sensing of Environment*, 56, 104–117.
- Gao, B.-C., & Goetz, A.F.H. (1994). Extraction of dry leaf spectral features from reflectance spectra of green vegetation. *Remote Sensing of Environment*, 47, 369–374.
- Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J.P.G., & Schmuck, G. (1995). *Leaf optical properties experiment 93 (LOPEX93) report EUR-16095-EN*. Ispra (Italy) European Commission – Joint Research Centre.
- Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., & Foley, W.J. (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment*, 93, 18–29.
- Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34, 75–91.
- Jacquemoud, S., Ustin, S.L., Verdebout, J., Schmuck, G., Andreoli, G., & Hosgood, B. (1996). Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, 56, 194–202.
- Jacquemoud, S., Verdebout, J., Schmuck, G., Andreoli, G., & Hosgood, B. (1995). Investigation of leaf biochemistry by statistics. *Remote Sensing of Environment*, 54, 180–188.
- Kaufmann, H., Segl, K., Guanter, L., Hofer, S., Foerster, K.P., Stuffer, T., et al. (2008). Environmental mapping and analysis program (EnMAP) – Recent advances and status. *IEEE International Geoscience and Remote Sensing Symposium*, 2008 (pp. IV – 109–IV – 112). IGARSS.
- Kokaly, R.F. (2001). Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. *Remote Sensing of Environment*, 75, 153–161.
- Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., & Wessman, C.A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, 113, S78–S91.
- Kokaly, R.F., & Clark, R.N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, 67, 267–287.
- Kou, L., Labrie, D., & Chylek, P. (1993). Refractive indices of water and ice in the 0.65– to 2.5 micrometer spectral range. *Applied Optics*, 32, 3531–3540.
- LaCapra, V.C., Melack, J.M., Gastil, M., & Valeriano, D. (1996). Remote sensing of foliar chemistry of inundated rice with imaging spectrometry. *Remote Sensing of Environment*, 55, 50–58.
- Li, P.H., & Wang, Q. (2011). Retrieval of leaf biochemical parameters using PROSPECT inversion: A new approach for alleviating ill-posed problems. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 2499–2506.
- Ma, Z.G., Chen, X., Wang, Q., Li, P.H., & Jiapaer, G. (2012). Retrieval of leaf biochemical properties by inverted PROSPECT model and hyperspectral indices: an application to *Populus euphratica* polymorphic leaves. *Journal of Arid Land*, 4, 52–62.
- Majeke, B., van Aardt, J.A.N., & Cho, M.A. (2008). Imaging spectroscopy of foliar biochemistry in forestry environments. *Southern Forests*, 70, 275–285.
- Malenovsky, Z., Rott, H., Cihlar, J., Schaepman, M.E., Garcia-Santos, G., Fernandes, R., et al. (2012). Sentinels for science: Potential of sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, 120, 91–101.
- Martin, M.E., & Aber, J.D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications*, 7, 431–443.
- Martin, M.E., Plourde, L.C., Ollinger, S.V., Smith, M.L., & McNeil, B.E. (2008). A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sensing of Environment*, 112, 3511–3519.
- Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, 92, 195–206.
- Mutanga, O., & Skidmore, A.K. (2004). Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment*, 90, 104–115.
- Mutanga, O., Skidmore, A.K., Kumar, L., & Ferwerda, J. (2005). Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, 26, 1093–1108.
- Ollinger, S.V., & Smith, M.L. (2005). Net primary production and canopy nitrogen in a temperate forest landscape: An analysis using imaging spectroscopy, modeling and field data. *Ecosystems*, 8, 760–778.
- Omari, K., White, H.P., Staenz, K., & King, D.J. (2013). Retrieval of forest canopy parameters by inversion of the PROFLAIR leaf-canopy reflectance model using the LUT approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6, 715–723.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., et al. (2013). Essential biodiversity variables. *Science*, 339, 277–278.
- Peterson, D.L., & Hubbard, G.S. (1992). Scientific issues and potential remote-sensing requirements for plant biochemical content. *Journal of Imaging Science and Technology*, 36, 446–456.
- Porra, R. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. *Photosynthesis Research*, 73, 149–156.
- Ramoelo, A., Skidmore, A.K., Schlerf, M., Mathieu, R., & Heitkönig, I.M.A. (2011). Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 408–417.
- Saltelli, A. (1999). Sensitivity analysis: Could better methods be used? *Journal of Geophysical Research, [Atmospheres]*, 104, 3789–3793.
- Saltelli, A., & Bolado, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis*, 26, 445–460.
- Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, 100, 281–294.
- Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 95, 177–194.
- Schlerf, M., Atzberger, C., Hill, J., Buddenbaum, H., Werner, W., & Schueler, G. (2010). Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. Karst.) using imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 12, 17–26.
- Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Van Wieren, S.E., Peel, M., Grant, R.C., et al. (2010). Forage quality of savannas – Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment*, 114, 64–72.
- Smith, M.L., Ollinger, S.V., Martin, M.E., Aber, J.D., Hallett, R.A., & Goodale, C.L. (2002). Direct estimation of aboveground forest productivity through hyperspectral remote sensing of canopy nitrogen. *Ecological Applications*, 12, 1286–1302.
- Sobol, I.M. (1993). Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1, 407–414.
- Steudler, P.A., Bowden, R.D., Melillo, J.M., & Aber, J.D. (1989). Influence of nitrogen fertilization on methane uptake in temperate forest soils. *Nature*, 341, 314–316.
- Van Soest, P.J., & Masson, V.C. (1967). The influence of the Maillard reaction upon the nutritive values of fibrous feed. *Feed Science and Technology*, 32, 45–53.
- Van Soest, P.J., & Wine, R.H. (1991). Use of detergents in the analysis of fibrous feeds. IV. Determination of plant cell-wall constituents. *Journal of the Association of Official Analytical Chemists*, 50, 50–55.
- Wang, Z., Skidmore, A. K., Darvishzadeh, R., Heiden, U., Heurich, M., & Wang, T. (2015). Leaf nitrogen content indirectly estimated by leaf traits derived from the PROSPECT model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP(99), 1–11 <http://dx.doi.org/10.1109/JSTARS.2015.2422734>.
- Weende, P. (1985). *NF V03-400*. Recueil des Normes Françaises-AFNOR.

- Weiss, M., Baret, F., Myneni, R.B., Pragnère, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20, 3–22.
- Wieliczka, D.M., Weng, S., & Querry, M.R. (1989). Wedge shaped cell for highly absorbent liquids: Infrared optical constants of water. *Applied Optics*, 28, 1714–1719.
- Xiao, Y., Zhao, W., Zhou, D., & Gong, H. (2014). Sensitivity analysis of vegetation reflectance to biochemical and biophysical variables at leaf, canopy, and regional scales. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 4014–4024.
- Zhao, F., Guo, Y., Huang, Y., Reddy, K.N., Lee, M.A., Fletcher, R.S., et al. (2014). Early detection of crop injury from herbicide glyphosate by leaf biochemical parameter inversion. *International Journal of Applied Earth Observation and Geoinformation*, 31, 78–85.
- Zhao, K., Valle, D., Popescu, S., Zhang, X., & Mallick, B. (2013). Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. *Remote Sensing of Environment*, 132, 102–119.