

10-2011

Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling

Jean-Baptiste Féret

Carnegie Institution for Sciences, jbferet@stanford.edu

Christophe François

Université Paris-Sud, Orsay, France, christophe.francois@u-psud.fr

Anatoly A. Gitelson

University of Nebraska-Lincoln, agitelson2@unl.edu

Gregory P. Asner

Carnegie Institution for Sciences, gpa@stanford.edu

Karen M. Barry

University of Tasmania, Karen.Barry@utas.edu.au

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/natrespapers>

 Part of the [Natural Resources and Conservation Commons](#)

Féret, Jean-Baptiste; François, Christophe; Gitelson, Anatoly A.; Asner, Gregory P.; Barry, Karen M.; Panigada, Cinzia; and Jacquemoud, Stéphane, "Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling" (2011). *Papers in Natural Resources*. 311.
<http://digitalcommons.unl.edu/natrespapers/311>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Jean-Baptiste F  ret, Christophe Fran  ois, Anatoly A. Gittelso, Gregory P. Asner, Karen M. Barry, Cinzia Panigada, and St  phane Jacquemoud

Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling

Jean-Baptiste Féret,^{1,2,3} Christophe François,² Anatoly Gitelson,⁴ Gregory P. Asner³
Karen M. Barry,^{5,6} Cinzia Panigada,⁷ Andrew D. Richardson,⁸ and Stéphane Jacquemoud¹

1. Institut de physique du globe de Paris – Sorbonne Paris Cité, Université Paris Diderot,
UMR CNRS 7154, Case 7011, 35 rue Hélène Brion, 75013 Paris, France

2. Laboratoire Ecologie, Systématique et Evolution (UMR 8079), Université Paris-Sud, Orsay, France

3. Department of Global Ecology, Carnegie Institution for Sciences, 260 Panama Street, Stanford, CA 94305, USA

4. Center for Advanced Land Management Information Technologies, School of Natural Resources,
University of Nebraska–Lincoln, Lincoln, NE, USA

5. School of Agricultural Science, University of Tasmania, Private Bag 12, Hobart, Tasmania 7001, Australia

6. CRC for Forestry, Private Bag 12, Hobart, Tasmania 7001, Australia

7. Remote Sensing of Environmental Dynamics Lab., Dip. Scienze dell’Ambiente e del Territorio,
Università Milano-Bicocca, 20126 Milano, Italy

8. Harvard University, Department of Organismic and Evolutionary Biology,
HUH, 22 Divinity Ave, Cambridge MA 02138, USA

Corresponding author – J.-B. Féret, Department of Global Ecology, Carnegie Institution for Sciences, 260 Panama Street, Stanford, CA 94305, USA; tel 650 462-1047, fax 650 462-5968, email jbferet@stanford.edu

Other email addresses: C. François: christophe.francois@u-psud.fr; A. Gitelson: agitelson2@unl.edu; G. P. Asner: gpa@stanford.edu; K. M. Barry: Karen.Barry@utas.edu.au; C. Panigada: cinzia.panigada@unimib.it; A. D. Richardson: arichardson@oeb.harvard.edu; S. Jacquemoud: jacquemoud@ipgp.fr

Abstract

We used synthetic reflectance spectra generated by a radiative transfer model, PROSPECT-5, to develop statistical relationships between leaf optical and chemical properties, which were applied to experimental data without any readjustment. Four distinct synthetic datasets were tested: two unrealistic, uniform distributions and two normal distributions based on statistical properties drawn from a comprehensive experimental database. Two methods used in remote sensing to retrieve vegetation chemical composition, spectral indices and Partial Least Squares (PLS) regression, were trained both on the synthetic and experimental datasets, and validated against observations. Results are compared to a cross-validation process and model inversion applied to the same observations. They show that synthetic datasets based on normal distributions of actual leaf chemical and structural properties can be used to optimize remotely sensed spectral indices or other retrieval methods for analysis of leaf chemical constituents. This study concludes with the definition of several polynomial relationships to retrieve leaf chlorophyll content, carotenoid content, equivalent water thickness and leaf mass per area using spectral indices, derived from synthetic data and validated on a large variety of leaf types. The straightforward method described here brings the possibility to apply or adapt statistical relationships to any type of leaf.

Keywords: leaf optical properties, PROSPECT, hyperspectral data, pigment content, water content, leaf mass per area, spectral indices, partial least squares regression

1. Introduction

Leaf chemical constituents are determining indicators of plant physiology and other functional processes up to the ecosystem level. Chlorophylls *a* and *b* enable light harvesting for photosynthesis (Anderson, 1986; Lichtenthaler et al., 1981), while carotenoids (carotenes and xanthophylls) and anthocyanins afford protection from excess light, for example during leaf development or abiotic stress (Close and Beadle, 2003; Gould et al., 2008). Measurement of total chlorophyll content (C_{ab}) and carotenoid content (C_{xc}) has many applications in agriculture, ecology, and Earth science. Beside photosynthetic pigments, leaf water content, expressed as Equivalent Water Thickness (EWT), and dry matter content, expressed as Leaf Mass per Area (LMA), are also critical variables in plant ecology, especially in forest fire risk assessment, water stress analysis, net ecosystem exchange and carbon storage computation.

Remote sensing is particularly adapted to study these leaf constituents because of their strong influence both on leaf and canopy reflectance. This study focuses on the assessment of vegetation chemical properties at the leaf level. Until now, C_{ab} and EWT have been the most extensively investigated chemical constituents at the leaf scale, because of their strong absorption features in the visible and shortwave infrared domains, respectively (Seelig et al., 2008; Ustin et al., 2009). Quantifications of C_{xc} and LMA are much more challenging, because they are masked by stronger absorbers like C_{ab} in the visible and EWT in the infrared (Gitelson et al., 2002; Kokaly et al., 2009; Zur et al., 2000). The number of methods to retrieve carotenoid content is limited (Asner and Martin, 2008; Chapelle et al., 1992; Féret et al., 2008; Gitelson et al., 2002, 2006; Sims and Gamon, 2002; Zur et al., 2000) and there are even less focusing on anthocyanins (Asner and Martin, 2008; Gitelson et al., 2001, 2006, 2009; Sims and Gamon, 2002). LMA gathers several types of constituents, which show strongly

overlapping optical properties. These constituents, mainly cellulose, lignin, nitrogen and proteins, have been investigated by numerous authors (reviewed by Kokaly et al., 2009), and a few studies considered direct assessment of LMA using remote sensing (Asner et al., 2011; Baret and Fourty, 1997; Jacquemoud et al., 1996; le Maire et al., 2008).

Several methods have been applied to assess the chemical content of plant leaves from their optical properties. There have been three main approaches: *i*) semi-empirical methods based on spectral indices adjusted to experimental datasets (e.g. Danson and Bowyer, 2004; Gitelson et al., 2006; Sims and Gamon, 2002); these indices may be calibrated and validated on synthetic datasets simulated using leaf optical models (le Maire et al., 2004, 2008), *ii*) statistical multivariate methods based on linear (e.g. Asner and Martin, 2008; Asner et al., 2009; Blackburn and Ferwerda, 2008; Jacquemoud et al., 1995; Jørgensen et al., 2007; Li et al., 2007) and nonlinear regression models (e.g. Shi & Sun, 2007), *iii*) radiative transfer model (RTM) inversion (e.g. Barry et al., 2009; Di Vittorio, 2009; Féret et al., 2008; Jacquemoud et al., 1996).

Whatever the method, its success depends upon the quality of the training dataset, the selection of the wavelengths and the availability of an independent dataset for the method validation. There is a large variety of published spectral indices for leaf C_{ab} retrieval, but inter-comparison studies intended to find the "best" index in terms of reliability have underscored several problems (Gitelson et al., 2003; Richardson et al., 2002; Sims and Gamon, 2002). These authors noted that reaching consensus is difficult because the calibration and validation datasets are generally limited to one or a few closely related species. As a consequence, methods based on small-sized datasets often lead to specialized indices that perform poorly when applied to a wide variety of plant leaves and conditions. Moreover the datasets were often not generic enough in terms of pigment composition and distribution, phenological stage and leaf structure. Finally they noted that an independent validation dataset is an exception to the rule.

To address this problem, le Maire et al. (2004, 2008) proposed a method which purports to be generic at two levels: the design of a comprehensive calibration database and the choice of a set of wavelengths. The calibration database, simulated with the leaf optical properties RTM PROSPECT (Jacquemoud et al., 1996), was used to select optimal combinations of narrow spectral bands through systematic exploration of the wavelength space to build new indices dedicated to C_{ab} and LMA retrieval. With the same approach, Danson and Bowyer (2004) compared the correlation between EWT and different spectral indices, adjusted both on experimental and simulated data. Shi and Sun (2007) similarly trained an artificial neural network for C_{ab} and EWT retrieval, but the contribution of simulations has not been clearly demonstrated and only one species was studied. Blackburn and Ferwerda (2008) proposed a method to retrieve leaf chlorophyll content from reflectance using wavelet analysis; this approach based on PROSPECT simulations gave promising results but was not validated with experimental data.

The application of radiative transfer models is also generic by its nature since it does not depend on a specific calibration database: models are calibrated once, thus their inversion does not require any prior adjustment for a training dataset. The retrieval of leaf chemistry by PROSPECT inversion gave good results when applied to various species and phenological stages (Barry et al., 2009; Féret et al., 2008). Although the mathematical libraries for optimization routines and the increasing computer performance are making model inversion more tractable, it still requires much spectral information, and its implementation is not as common or straightforward as are statistical methods and spectral indices.

Using simulations in place of experimental data may be worthwhile because it increases flexibility and control when building the training dataset. For example it has the advan-

tage of avoiding laboratory analytical errors in deriving chemical constituents. Moreover model simulations are more easily reproducible and can facilitate comparison studies based on a standardized approach.

Our goal is to compare several methods of sampling and to provide a reproducible framework for the design of an optimal synthetic dataset using the model PROSPECT-5. Four sampling strategies differing by their intrinsic variance and covariance matrices are tested to build two types of statistical models, spectral indices and PLS, with the intention of assessing C_{ab} , C_{xc} , EWT, and LMA retrieval at the leaf scale. The models adjusted to synthetic data are then compared to those adjusted to experimental data, in terms of chemical retrieval performance. They are also examined against PROSPECT-5 inversions. All methods face the same problem when it comes to validation, i.e., they all require a large and representative database. Here the main attribute of our validation dataset is that it maximizes spectral and biochemical variability. To achieve this, plant species from different ecosystems, grown under various conditions and at different phenological stages have been selected.

2. Material and methods

We first describe the experimental dataset used in this study and the PROSPECT-5 model that is run to simulate leaf reflectance. The different strategies adopted to build a synthetic dataset are then compared. Finally, the methods used to assess leaf chemical constituents are introduced: spectral indices, PLS, and radiative transfer model inversion.

2.1. Description of the experimental datasets

Seventeen independent datasets including a wide range of leaf spectral, chemical, and structural properties were incorporated into this study. They encompass 1417 leaves corresponding to about 120 different species from various growing conditions and developmental stages (Table 1). Exhaustive information about the methods used to measure optical and chemical properties of leaves can be found in the references given in Table 1. These datasets contain leaf directional-hemispherical reflectance spectra and, when available, transmittance spectra measured in the solar domain (1 nm increment) with laboratory spectrophotometers or field spectroradiometers equipped with integrating spheres, except BIRCH 2 for which leaf optical properties are measured using a leaf clip. The spectral range varies according to the datasets but they all include measurements in the visible (VIS, 400–700 nm) and in part or total of the near-infrared (NIR, 700–1000 nm) domains. Some also include the shortwave-infrared (SWIR, 1000–2500 nm). The datasets generally share a pool of chemical constituents expressed as a mass per unit leaf area, thus comparable: C_{ab} and C_{xc} ($\mu\text{g}\cdot\text{cm}^{-2}$), EWT ($\text{g}\cdot\text{cm}^{-2}$ or cm), and LMA ($\text{g}\cdot\text{cm}^{-2}$). Basic statistics for the chemical constituents are displayed in Table 2.

The first step in compiling leaf datasets was to develop synthesis information about the distribution of each leaf property and correlations between leaf constituents. As shown in Figure 1, the distribution of C_{ab} , C_{xc} and LMA can be fitted with a Gaussian distribution, while EWT is better represented by a lognormal distribution. Despite our efforts to include as many vegetation types as possible, this database does not represent all types of vegetation since it mainly contains deciduous trees, cereal crops and fruit trees, and a very few humid tropical species. Nevertheless, it is one of the largest that has been compiled to date, and it provides the possibility to validate our modeling results on more than a hundred plant species. Although the distributions may not be suitable to all types of ecosystems, the various growing conditions—from natural growth to fertilized growth under green houses—and physiological stages—from young to senescent leaves and healthy to stressed leaves—create a dataset with wide-ranging leaf traits.

Table 1. All available datasets.

Name	Number of samples	Number of species	Vegetation type	Spectral domain (nm)	R	T	Leaf biochemistry			
							C_{ab}	C_{xc}	EWT	LMA
ANGERS (1)	276	49	Temperate species and crops	400-2450	×	×	×	×	×	×
BIRCH 1 (2)	140	1	Temperate species	400-835	×	×	×	×	×	×
BIRCH 2 (3)	98	1	Temperate species	400-1000	×		×			×
Maple and chestnut 1 (4)	49	2	Temperate species	400-750	×	×	×	×		
Beech, elm, vine (5)	66	3	Temperate species	400-780	×		×	×		
Maple and chestnut 2 (6)	45	2	Temperate species	400-750	×		×	×		
Maize, soybean (7)	40	2	Crops	400-800	×		×	×		
Oak 1 (8)	142	1	Temperate species	400-1600	×	×	×	×	×	×
Oak 2 (9)	112	1	Temperate species	400-2200	×	×	×			×
Ese (10)	99	11	Temperate species	410-785	×		×			×
Eucalyptus (11)	64	1	Temperate species	400-1650	×	×	×	×	×	×
Fig (12)	60	1	Temperate species	400-800	×	×	×	×		
Hawaii (13)	41	41	Tropical species	400-2500	×	×	×	×	×	×
Beech (14)	46	1	Temperate species	400-800	×	×	×	×		
LOPEX (15)	64	58	Temperate species and crops	400-2400	×	×			×	×
Chestnut 3 (16)	22	1	Temperate species	400-780	×	×	×	×		
Vine (17)	53	1	Temperate species	400-950	×		×	×		

(1) Féret et al., 2008; (2) C. Panigada (personal communication); (3) Richardson et al., 2002; (4) Gitelson et al., 2002; (5) Gitelson et al., 1999; (6) Gitelson et al., 2003; (7) Gitelson et al., 2006; (8) Rossini et al., 2006; (9) J. Louis (personal communication); (10) le Maire et al., 2004; (11) Barry et al., 2009; (12) Gitelson & Merzlyak, 1997; (13) Féret et al., 2008; (14) Gitelson et al., 2002; (15) Hosgood et al., 1994; (16) Gitelson & Merzlyak, 1994; (17) Steele et al., 2008.

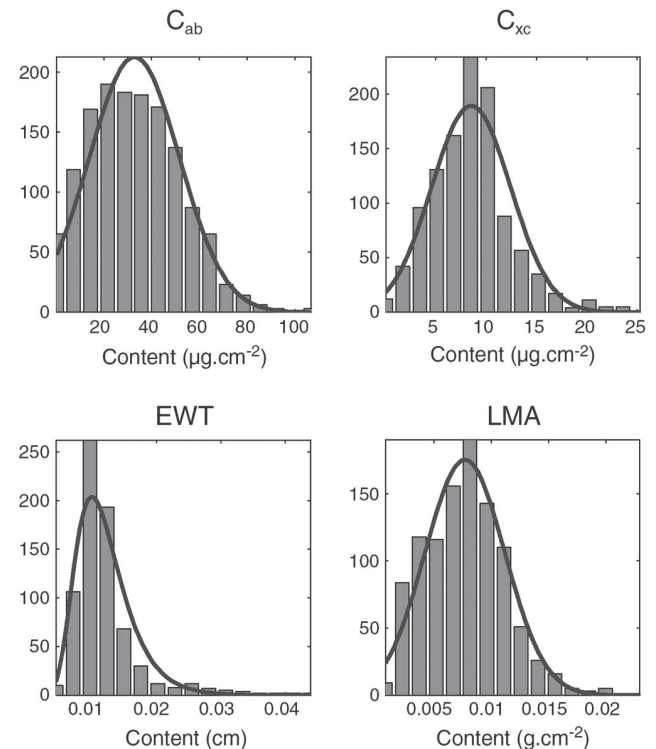
Table 2. Basic statistics computed on the whole dataset (1417 leaves).

	C_{ab} ($\mu\text{g}\cdot\text{cm}^{-2}$)	C_{xc} ($\mu\text{g}\cdot\text{cm}^{-2}$)	EWT (cm)	LMA ($\text{g}\cdot\text{cm}^{-2}$)
Number of samples	1417	1106	725	1034
Minimum value	0.30	0.04	0.0043	0.0008
Maximum value	106.72	25.30	0.0713	0.0331
Mean value	32.81	8.51	0.0129	0.0077
Standard deviation	18.87	3.92	0.0073	0.0035

This database is also used at the validation stage to evaluate the performance of the methods detailed hereafter for C_{ab} , C_{xc} , EWT, and LMA retrievals from leaf optical measurements. The 17 experimental datasets vary both in terms of the spectral domain covered and the foliar attributes quantified, as shown in Table 1. Therefore, the number of leaf samples available for the validation stage varies as well, depending upon the spectral domain required and on the chemical constituents related to leaf optical properties. We underline that it is important that the validation dataset is common to all the methods, and thus its size is constrained by the spectral domain required to apply the method which is spectrally the most demanding. For example the dataset BIRCH 1, which comprises EWT and LMA measurements but spectral measurements from 400 to 835 nm only, is not included in the validation dataset used with these two constituents because their assessment by model inversion requires SWIR data. As a result, the validation is performed using 821 reflectance spectra measured from 400 to 800 nm for C_{ab} , 707 reflectance spectra measured from 400 to 800 nm for C_{xc} , 587 reflectance spectra measured from 800 to 1600 nm for EWT, and 381 reflectance spectra measured from 800 to 1800 nm for LMA. The SWIR domain is narrowed to 1600 nm for EWT and to 1800 nm for LMA in order to keep as much experimental data as possible.

2.2. Radiative transfer modeling of leaf reflectance

PROSPECT-5, the latest version of the PROSPECT model (Jacquemoud & Baret, 1990), simulates leaf directional-hemispherical reflectance and transmittance, referred to as reflectance and transmittance hereafter, from 400 to 2500 nm with five input variables: C_{ab} , C_{xc} , LMA, EWT, and N, the leaf structure parameter (Féret et al., 2008). The adding of carotenoids to the photosynthetic pigments taken into account by the

**Figure 1.** Distributions of C_{ab} , C_{xc} , EWT, and LMA in the entire dataset compiled from Table 1.

model is the main advance of the new version. PROSPECT is a generalized plate model, i.e. it describes the leaf as a stack of N homogeneous, absorbing elementary layers, separated by N-1 air spaces. The leaf structure parameter corresponds to the number of layers. Although N is most conceivable as a whole number, accounting for subtle variations in leaf structure requires to feed the model with continuous values. Therefore, N is a real number ranging between 1 and 3. Jacquemoud and Baret (1990) proposed a relationship between N and the LMA, but our attempts to obtain similar relationship showed that it was unsuitable for our data. Thus the characteristic value of N for each leaf that is used in this study results from the inversion of PROSPECT-5.

2.3. Design of a synthetic dataset

Four different sampling strategies intended to design synthetic leaf optical properties datasets are tested. The probability distributions of the input variables of the model and the number of simulations required to create the synthetic dataset distinguish each approach.

2.3.1. Sampling #1

This approach, proposed by le Maire et al. (2004, 2008) involves defining a range of variation for each input variable of PROSPECT and, within the given range, a fixed number of levels following a uniform distribution. All possible combinations are then simulated. The range for each input variable is derived from experimental observations. Le Maire et al. (2004) simulated 11,583 leaf reflectances with PROSPECT, in order to adjust optimal indices aiming at assessing C_{ab} . Following a similar principle, le Maire et al. (2008) built indices applicable to C_{ab} and LMA, and they extended their investigation to the canopy level. Two datasets were generated: one contained 6006 leaf reflectance spectra obtained by crossing four variables (C_{ab} , EWT, LMA, and N) and the other was made of 149,688 canopy reflectance spectra obtained by crossing seven variables. They showed that C_{ab} is well retrieved by the spectral indices adjusted at the leaf scale, while LMA is poorly determined. The advantage of this sampling strategy is its basic implementation and the wide range of values taken by the chemical and structural properties, to encompass as many leaf types as possible, thus making the index applicable on a large range of species and growing conditions. The main restrictions lie in the size of the synthetic dataset which grows exponentially with the number of variables and the generation of unrealistic data. In this study, the distribution of le Maire et al. (2008) is reproduced for N, C_{ab} , EWT and LMA, with the addition of C_{xc} for which eleven levels are uniformly selected between 2.5 and 25 $\mu\text{g}\cdot\text{cm}^{-2}$ (Table 3). A total of 66,066 leaf reflectance spectra make up the first dataset.

2.3.2. Sampling #2

Design of experiment (DOE) is a statistical method that allows defining a structured and restricted number of simulations where all input variables vary at the same time. It maintains the statistical properties of a uniformly distributed sampling (sampling #1), but greatly decreases the number of simulations. This method is applied at the leaf scale by Pavan et al. (2004) in order to perform sensitivity studies with the PROSPECT model. Here we use a Hyper Graeco Latin Geometric sampling scheme (Benoist et al., 1994) which allows seven values equi-distributed within a definition range for each of the five input variables of PROSPECT-5, leading to 2401 simulations. A distribution range similar to sampling #1 has been chosen here (Table 3). The very reasonable size of sampling #2 allows us to apply statistical methods using information from the whole spectral domain such as PLS, whereas the application of these methods to the dataset obtained from sampling #1 result in a computationally intensive task.

2.3.3. Sampling #3

To study the relationship between spectral indices and leaf chemistry, some authors create synthetic datasets with ac-

Table 3. Leaf variables used to build a 66,066 spectra synthetic dataset based on sampling #1.

	Minimum	Maximum	Step
C_{ab} ($\mu\text{g}\cdot\text{cm}^{-2}$)	10	110	10
C_{xc} ($\mu\text{g}\cdot\text{cm}^{-2}$)	2.5	25	2.5
EWT (cm)	0.004	0.024	0.004
LMA ($\text{g}\cdot\text{cm}^{-2}$)	0.002	0.014	0.001
N	1.1	2.3	0.2

tual distributions. Ceccato et al. (2001) performed a sensitivity analysis of the Moisture Stress Index (MSI) (Hunt & Rock, 1989) using PROSPECT. The variation for all the input variables was taken from the LOPEX dataset of actual leaf measurements (Hosgood et al., 1994), and a log-normal distribution is allocated to EWT. Ceccato et al. (2001) then applied the EFAST (Extended Fourier Amplitude Sensitivity Test) method that requires 9987 simulations for only three input variables. Danson and Bowyer (2004) underwent a similar effort to adjust spectral indices sensitive to EWT. The main advantage of this method is that the user can define the number of simulations needed to create the dataset. For instance, these authors simulated 335 leaf reflectance spectra, the same number as in the experimental dataset. The results confirmed the high correlation between MSI and EWT regardless of whether the dataset is experimental or synthetic. However, they did not provide similar results for all the indices, highlighting possible inaccuracies in the model. Note that no noise is applied to the data, which may explain the overestimation of the correlation between the spectral indices adjusted on simulated data and EWT. Based on the means and standard deviations of Table 2, we used 2500 samples randomly selected, following a normal distribution for C_{ab} , C_{xc} , LMA and N, and a log-normal distribution for EWT. Note that another trial using a smaller dataset containing 500 samples did lead to similar results, but we decided to set a size comparable to that used in sampling #2.

2.3.4. Sampling #4

The fourth sampling strategy is similar to sampling #3, except that it includes actual correlations between leaf constituents. Danson and Bowyer (2004) did not include covariations between leaf constituents and concluded on the necessity to investigate their influence on the accuracy of the relationship linking chemical and optical properties based on simulations. Indeed, this is a critical point as ignoring covariance among leaf traits may become an issue as it is likely that unrealistic combinations of chemical constituents are simulated with samplings #1 to #3. Table 4 shows the different degree of correlation of leaf constituents, with highest correlation found between C_{ab} and C_{xc} . The probability to find leaves which deviate from this straight line exists, for example in senescent leaves that show a very low $C_{ab}:C_{xc}$ ratio (Gitelson et al., 2003, 2006). However, leaves with a very high ratio are physiologically unlikely in nature. Avoiding such unrealistic combinations, which would result in the simulation of meaningless leaf optical properties and induce some bias in the relationship, is then justified. This is achieved by the generation of 2500 samples using a multivariate normal distribution based on the mean and standard deviation from Table 2 and correlations from Table 4.

In summary, the synthetic datasets produced by the four sampling strategies vary in size (66,066 samples for sampling #1, 2401 samples for sampling #2, 2500 samples for samplings #3 and #4), as well as in structure. For each dataset, we calculate the associated reflectance spectra from 400 to 2500 nm (1 nm increments). Gaussian noise corresponding to 2% of the reflectance is then applied to the spectra, in order to decrease the effect of auto-correlation and possible minor model inaccuracies. This value used for Gaussian noise corresponds to the RMSE between the measured and modeled reflectances calculated by Féret et al. (2008).

Table 4. Correlation matrix for the five leaf variables obtained from the compilation of leaf datasets (C_{ab} , C_{xc} , EWT and LMA), or by inversion of PROSPECT (N).

	C_{ab}	C_{xc}	EWT	LMA	N
C_{ab}	1				
C_{xc}	0.86	1			
EWT	0.19	0.27	1		
LMA	0.19	0.43	0.63	1	
N	-0.15	-0.02	0.28	0.13	1

2.4. Methods to estimate leaf chemicals from optical properties

Three methods for leaf chemical estimation differing by the amount of spectral information needed are compared. They are presented in increasing order of spectral information required. Their performance in retrieving leaf properties was assessed by means of the RMSE between measured and retrieved values.

2.4.1. Spectral indices

Spectral indices present the easiest and most popular method to estimate leaf chemical constituents. Based on a limited number of wavebands, usually between two and four, they are fairly accurate when estimating C_{ab} , C_{xc} , EWT, and LMA. Simple reflectance, reflectance difference and ratio, and normalized difference are the most common forms. There is a body of literature regarding C_{ab} retrieval using spectral indices, as recently reviewed by Ustin et al. (2009), highlighting the recognized importance of chlorophyll to physiological processes (Khamis et al., 1990) and its major influence on leaf optical properties in the VIS. Thus C_{ab} assessment has various applications in agriculture, forestry, ecology, and environmental science. The literature offers many spectral indices using diverse combinations of wavelengths so that the choice of an appropriate one is difficult. Moreover, most indices are tested on only one species or one type of vegetation, thus they are not widely applicable due to the dependence of the relationship upon an experimental dataset. Some publications compare the performance of different indices applied to the same dataset to aid in the selection of an adapted one (e.g. Gitelson and Merzlyak, 2004; Gitelson et al., 2003; le Maire et al., 2004; Richardson et al., 2002; Sims and Gamon, 2002). The main disadvantage of spectral indices is their empirical base, which can result in a lack of generality. Figure 2 illustrates this issue: The relationship found by Richardson et al. (2002) for C_{ab} retrieval based on the $NDVI_{red\ edge}$ with red edge wavelength at 705 nm (Gitelson and Merzlyak, 1994, 1997) is first applied to BIRCH 2, the dataset used in the original publication, and then to the entire database. The quasi-linear relationship found by the authors is well adapted to BIRCH 2 but cannot be applied to the whole database with much higher C_{ab} contents. It means that either index (i.e. $NDVI_{red\ edge}$) or spectral bands used were not optimal for whole database.

Two interesting approaches may improve these spectral indices. The first consists of searching for the optimal set of spectral bands that apply to a given type of index (le Maire et al., 2004, 2008). The second, a semi-analytical method, is based on

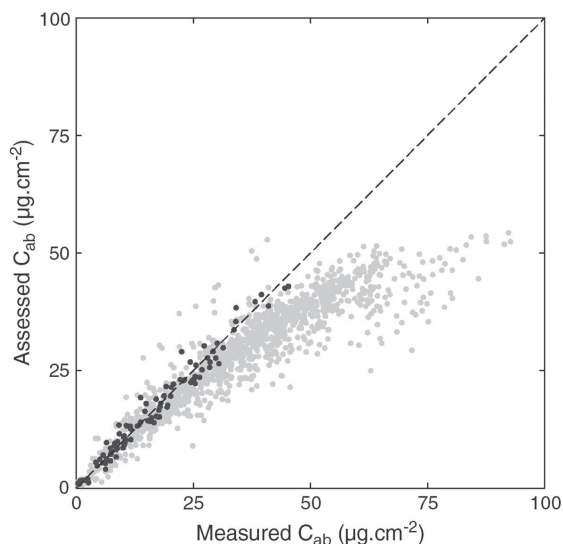


Figure 2. Assessment of C_{ab} using $NDVI_{red\ edge}$ adjusted on BIRCH 2 (black dots, 98 samples) and applied to the compiled leaf database (grey dots, 1417 samples). Dashed line = 1:1.

the prior study of correlations between chemical and optical properties of leaves (Gitelson et al., 2003, 2006; Gitelson and Merzlyak, 2004). Unlike most of the spectral indices that are based on narrow-band reflectances, those resulting from this method use wide spectral bands that integrate the spectral variability between species. This approach produces indices that have proven particularly well suited for C_{ab} and C_{xc} estimation. A preliminary comparison of several indices from the literature also shows good results in terms of pigment retrieval when applied to our experimental dataset. In this paper, we selected $Chl_{red\ edge}$ and $Car_{red\ edge}$ two indices published by Gitelson et al. (2006). It seems that there are no broadband indices for EWT and LMA. Two studies (Ceccato et al., 2001; Danson and Bowyer, 2004) referred to MSI as a very suitable index to estimate EWT, so we selected it. Finally, to estimate LMA from leaf reflectance, le Maire et al. (2008) created a normalized difference index called ND_{LMA} and applied it to the ANGERS dataset with moderate success. They stressed the challenge of such a task, particularly when using spectral indices adjusted with a synthetic dataset. The four spectral indices selected to estimate leaf chemical constituents are shown in Table 5. Gitelson et al. (2006) noted that $Chl_{red\ edge}$ and $Car_{red\ edge}$ are quasi-linearly correlated to C_{ab} and C_{xc} , respectively, le Maire et al. (2008) fitted the relationship ND_{LMA} versus LMA by a second-order polynomial, and Danson and Bowyer (2004) preferred to use a power function between MSI and EWT.

For each sampling strategy presented in Section 2.3, a statistical relationship (polynomial or power function according to the type of index) is established between the leaf chemical content and the spectral index. This relationship is then applied to the experimental database for validation. In addition, a leave-one-out cross-validation (LOOCV) is implemented to predict the performance of the spectral indices when designed using observations. Assume that a single sample is temporarily removed from the n data points of the experimental dataset: the relationship is trained on the remaining $n - 1$ data points and the error on the validation sample is reported. This is repeated n times such that each observation is used once as a validation sample. The average RMSE is then calculated as a criterion to judge the performance of the relationship. The LOOCV is expected to give the best results since the relationships are directly fitted on the experimental dataset. It is presented here only as a reference, keeping in mind the fact that the generic nature of the indices fitted this way is not guaranteed.

2.4.2. Partial least square (PLS) regression

PLS is an extension of multiple linear regression modeling that statistically determines the relative contribution of each chemical constituent to leaf reflectance (or transmittance) (Asner et al., 2009; Asner and Martin, 2008). PLS utilizes the continuous, full-range spectrum rather than a band-by-band analysis. The number of measured spectra required is much higher than for spectral indices, and the resulting model associates one coefficient to each wavelength. When applied to experimental data, the general approach of PLS theoretically allows a better assess-

Table 5. Spectral indices selected for assessment of leaf chemical constituents.

Biochemical constituent	Spectral index	Formula	Relationship	Reference
C_{ab}	$Chl_{red\ edge}$		Polynomial	Gitelson et al. (2003, 2006)
C_{xc}	$Car_{red\ edge}$		Polynomial	Gitelson et al. (2003, 2006)
EWT Rock	MSI		Power function	Hunt and (1989)
LMA	ND_{LMA}		Polynomial	le Maire et al. (2008)

ment of chemistry than do spectral indices, but there is a risk of overfitting that would lead to inaccurate relationships when checking the model on a very different dataset to the training one. To decrease this risk, the prediction residual error sum of squares (PRESS) statistic is minimized to determine the number of factors in the PLS analysis (Chen et al., 2004). For each chemical constituent and the spectral domains defined in Section 2.1, PLS-PRESS is performed with three synthetic datasets (samplings #2 to #4), as well as the experimental data. No result is available with sampling #1 because of the large size of this dataset that prevents the application of such a method.

2.4.3. Model inversion

This approach already proved to be accurate in estimating leaf chemical constituents from leaf spectra. Model inversion aims at finding the optimal set of input variables (θ), namely C_{ab} , C_{xc} , EWT, LMA and N, by comparing modeled to measured leaf optical properties. This problem is addressed using a minimization algorithm based on the merit function $J(\theta)$ defined by:

$$J(\theta) = \sum_{\lambda_{\min}}^{\lambda_{\max}} [(R^*(\lambda) - R_{\text{mod}}(\lambda, \theta))^2 + (T^*(\lambda) - T_{\text{mod}}(\lambda, \theta))^2] \quad (1)$$

Where R^* and T^* are the measured and R_{mod} and T_{mod} are the modeled reflectance and transmittance. A quasi-Newton optimization routine (E04JAF, NAG Fortran library) is applied to minimize $J(\theta)$ and the accuracy of the inversion and the fit can be assessed by the RMSE. When transmittance is not available, the second term in Equation (1) disappears but the problem becomes ill-posed and N cannot be assessed properly, which impacts on the accuracy of the other retrieved variables. For that reason, inversion is only performed on reflectance and transmittance, when available simultaneously. In our study, various combinations of initial guesses have been tested during model inversion. Although the convergence toward the global minimum is not guaranteed by the quasi-Newton algorithm, the minimum search performed when inverting reflectance and transmittance is not influenced by the initial guess for leaf chemical and structural properties, and in practice, global optimization does not improve the retrieval of leaf chemistry. Optical measurements in the VIS and the NIR are necessary to assess leaf pigment, while EWT and LMA require measurements in the NIR and the SWIR.

3. Results and discussion

3.1. Comparison of published spectral indices, PLS PRESS, and model inversion

In this section we calculate the four vegetation indices of Table 5 using the reflectance spectra generated according to the four strategies presented in Section 2.3, and the experimental reflectance spectra. We fit the relationships between the indices and the leaf chemical constituents. Their validation is then assessed using the experimental database. The performance of spectral indices is also compared to that of PLS-PRESS, PROSPECT-5 inversion, and LOOCV. This assists in defining and selecting the best overall strategy. Table 6 summarizes the results obtained with these different methods for leaf chemical constituents' retrieval.

Table 6 shows that the normal distribution and the normal multivariate distribution (samplings #3 and #4) allow better assessment of leaf chemical constituents than the uniform distribution (samplings #1 and #2). This observation is valid both when applying spectral indices and PLS-PRESS. A pairwise t -test ($\alpha = 0.05$) is applied to compare the residuals from sampling #3 and sampling #4. The difference between these two samplings is not significant, except for the assessment of C_{xc} based on PLS-PRESS: using correlation between leaf constituents for the dataset simulation allows improvement in C_{xc} retrieval.

Table 6. RMSE obtained for the retrieval of four leaf constituents (C_{ab} , C_{xc} , EWT and LMA) using different methods (spectral index, PLS-PRESS, model inversion) and sampling strategies. The spectral indices, presented in Table 5, and PLS-PRESS are calibrated on the synthetic datasets, and the LOOCV experimental dataset for reference. The validation dataset for all methods is the whole experimental dataset presented in Table 2.

	Synthetic data				Experimental data
	Sampling #1	Sampling #2	Sampling #3	Sampling #4	LOOCV
C_{ab} ($\mu\text{g}\cdot\text{cm}^{-2}$)					
Spectral index	7.85	7.91	6.31	6.33	6.32
PLS-PRESS	n/a	14.12	8.62	8.01	5.38
Model inversion			5.84		
C_{xc} ($\mu\text{g}\cdot\text{cm}^{-2}$)					
Spectral index	4.11	4.20	3.50	3.47	3.27
PLS-PRESS	n/a	4.37	3.20	2.46	1.90
Model inversion			2.69		
EWT (cm)					
Spectral index	0.0053	0.0051	0.0043	0.0044	0.0037
PLS-PRESS	n/a	0.0036	0.0034	0.0032	0.0025
Model inversion			0.0027		
LMA ($\text{g}\cdot\text{cm}^{-2}$)					
Spectral index	0.0027	0.0025	0.0022	0.0022	0.0021
PLS-PRESS	n/a	0.0025	0.0021	0.0020	0.0007
Model inversion			0.0031		

Therefore, the optical dataset simulated by sampling #4 is chosen to assess optimal indices in the next section. The pairwise t -test ($\alpha = 0.05$) is also used to compare results using the synthetic datasets obtained by the different sampling strategies with the result of the LOOCV applied to the experimental dataset. It shows that there is a significant difference in the retrieval of chemical traits when the relationship used has been adjusted on experimental data or synthetic data, except for C_{ab} . Each chemical constituent result reported in Table 6 is discussed hereafter.

3.1.1. Chlorophyll content

For this constituent only, the PLS-PRESS is less accurate than the spectral indices when trained on synthetic data. This suggests that PROSPECT-5 lacks accuracy in some spectral domains: increasing the level of noise decreases the RMSE. This higher noise may lessen the significance of some spectral domains which make the PLS-PRESS model too sensitive to variations in C_{ab} . Best performances are shown when PLS-PRESS is trained on experimental data. This is not observed when the index $\text{Chl}_{\text{red,edge}}$ is used, suggesting the reliability of PROSPECT-5 to simulate accurate reflectances in the spectral domains used by this index (690–720 nm and 760–800 nm). It is also notable that a unique relationship derived from a spectral index provides a fairly good assessment of C_{ab} for various species and physiological status represented in the experimental dataset. This result confirms that statistical relationships obtained by spectral indices and ground truth should be calibrated on a wide set of data in order to be robust and widely applicable as already underlined by Figure 2.

3.1.2. Carotenoid content

The difference between the spectral index and the PLS-PRESS previously observed with C_{ab} tends to diminish (sampling #2) and is even opposite (samplings #3 and #4) in favor of PLS-PRESS. Moreover, increasing the level of noise does not change the results obtained with PLS-PRESS, contrary to the observations made when assessing C_{ab} . This suggests that the level of noise applied to the simulations decreases the significance of wavelengths which show excessively high correlation with C_{xc} on synthetic data compared to experimental data. $\text{Car}_{\text{red,edge}}$ may not use an optimal combination of wavelengths, as PLS outperforms it. However, finding out such a combination is quite a complicated process and a significant improve-

ment of the results may require more than two or three spectral bands. The main assumption of the three-band model for $Car_{red\ edge}$ is equal specific C_{ab} absorption coefficients in the red edge range and in the green around 520 nm. It also assumes that C_{ab} and C_{xc} are independent variables. Very strong correlation between C_{xc} and C_{ab} (Table 4) probably makes impossible independent retrieval of C_{xc} and C_{ab} using $Car_{red\ edge}$. The difference between the PLS-PRESS trained on experimental data and PROSPECT-5 inversion is also significant. The strong correlation between C_{ab} and C_{xc} in the experimental data may induce an overestimation of PLS performances in C_{xc} retrieval. The doubt is that PLS is evidencing mainly the bands correlated with C_{ab} . This leads to a good estimation of both the leaf constituent in case of correlation between C_{ab} and C_{xc} , but leads to a decrease of C_{xc} estimation accuracy in case of senescent leaves characterized by a low $C_{ab}:C_{xc}$ ratio.

3.1.3. Equivalent water thickness

The first attempt to find a good MSI vs. EWT relationship using synthetic data was unsuccessful. A reanalysis of the initial synthetic datasets showed that the noise applied to the reflectance spectra was too strong, leading to overestimated scattering compared to experimental data. By reducing it to 0.5% reflectance, a better relationship could be obtained (Table 6). The performance of PROSPECT inversion is now similar to that obtained with the PLS-PRESS trained using experimental data. The spectral domain is restricted to 800–1600 nm, so that some water absorption features are not taken into account. Improved results are to be expected if the spectral domain used to perform PLS-PRESS is extended to 800–2500 nm. The better performance of PLS-PRESS compared to MSI when adjusted on synthetic data also indirectly suggests that other optimal wavelengths could be used to build a better spectral index.

3.1.4. Leaf mass per area

As for EWT, the noise is decreased from 2% to 0.5% of reflectance to adjust a relationship derived from spectral indices. It is notable that model inversion performs poorly compared to synthetic models based on PROSPECT simulation. Contrary to le Maire et al. (2008) who showed that ND_{LMA} is badly adjusted on synthetic data, the RMSE obtained with ND_{LMA} trained on synthetic (sampling #3 and #4) and experimental data are not significantly different. The readjustment applied to the noise level may be the main explanation for this better relationship, as the assessment of LMA based on sampling #2 is fair. Therefore we can expect improved results if a new attempt to select the optimal wavelengths is done, based on an updated sampling method and an appropriate noise level. The PLS-PRESS directly applied to experimental data shows far better results than any other method, and notably PROSPECT inversion, which is usually not very different from PLS applied to experimental data. This certainly indicates inaccuracies inherent to PROSPECT when modeling the influence of LMA on the leaf optical properties.

3.2. Design of optimal spectral indices

As shown by le Maire et al. (2004, 2008), testing all possible combinations of wavebands may help to build better models. In this section, we use the leaf optical data simulated by sampling #4, to design optimized spectral indices, i.e. the set of wavelengths that best explain the variation of leaf biochemical constituents, using the criterion of minimum RMSE between assessed and actual content. Two families of indices are tested: the ratio index (RI) and the normalized difference (ND). These indices are selected because only two wavelengths are required to compute them, which make systematic computing of all possible combinations of wavelengths achievable in a reasonable time. Moreover these indices showed better results than other indices such as simple reflectance and difference of reflectance. Besides these

classical indices, a semi-analytical spectral index is tested for C_{ab} and C_{xc} retrieval, based on the results obtained by Gitelson et al. (2003, 2006). It is built using three large spectral bands, then six optimal wavelengths are required for the lower and upper bounds of each band. The three spectral domains explored to obtain a semi-analytical spectral index are 680–730 nm, 750–800 nm, and 750–800 nm for C_{ab} ; 480–530 nm, 650–800 nm, and 750–800 nm for C_{xc} . These semi-analytical spectral indices are not tested on EWT and LMA because of the large spectral domains to explore: the selection of six optimal wavelengths, defining the lower and upper bound for three spectral domains would be too time-consuming.

Table 7 summarizes the RMSE resulting from the application to the experimental data of all the relationships derived from optimal spectral indices adjusted on synthetic data. The level of noise is the same as in Section 3.1. These results show that the optimal spectral indices perform as well (LMA) or better (C_{ab} , C_{xc} and EWT) than those taken from the literature. The optimal combination of wavelengths is based on the minimum RMSE obtained with the calibration dataset, however the comparison of performances obtained with $Chl_{red\ edge}$ (Table 6) and $Chl_{RE\ opt}$ (Table 7) shows that this combination does not necessarily match with the combination leading to the minimum RMSE obtained when applying the relationships derived from simulated dataset to the experimental data. This may be due to some left over inaccuracies inherent to PROSPECT and could be corrected, for example, by adapting a specific level of noise to each wavelength, based on the uncertainties of the model. Attempts to adjust the global level of noise were not conclusive. However, this problem does not occur with the optimal ND index, which is in reality $NDVI_{red\ edge}$ with red edge band at 712 nm. The optimal combination of wavelengths allows adjusting the relationship showing the best performance when applied to experimental data (Table 7), contrary to the relationships obtained with narrow band RI and wide band $Chl_{red\ edge}$ spectral indices. It shows that optimizing spectral bands allows significant improvement performances of vegetation indices. The use of spectral band at 712 nm instead of 705 nm in $NDVI_{red\ edge}$ (presented in Figure 2) allowed to increase the accuracy of C_{ab} estimation. Also worth noting, that all three optimal models for C_{ab} estimation use red edge range of the spectrum as term sensitive to C_{ab} . Remarkably, reflectance at wavelengths up to 730 nm remains sensitive to C_{ab} ; this spectral region is 60 nm far from range of maximal C_{ab} absorption *in situ* around 670 nm. In two of three optimal models obtained for the assessment of C_{xc} (RI and ND) only one band sensitive to C_{xc} (around 530 nm) is employed. However, reflectance in this spectral region is affected by both C_{ab} and C_{xc} . It means that very strong correlation between C_{xc} and C_{ab} was a main factor affecting choice of optimal spectral band for C_{xc} retrieval. The results obtained for the two remaining chemical constituents are satisfying, and the wavelengths selected for spectral indices applied to LMA are very close to those obtained by le Maire et al. (2008). This confirms that a correct level of noise, combined with an appropriate sampling strategy, can improve the fit between modeled and experimental data, extending the generic nature of the results. Results obtained for carotenoid retrieval with optimal indices are also significantly better than those obtained using indices from the literature. The coefficients given in Table 7 correspond to the polynomial relationship linking the chemical constituent content to the value of the index.

4. Conclusion

This study aimed to assess the ability of synthetic datasets generated by PROSPECT-5 to provide statistical relationship between leaf reflectance and the main leaf chemical constituents: C_{ab} , C_{xc} , EWT, and LMA. Four synthetic datasets distinct from one another by the distributions of leaf chemical and structural properties were tested: two uniform distributions totally dis-

Table 7. Performance of different optimal indices derived from sampling #4, for each of the 4 studied biochemical constituents. The polynomial coefficients correspond to the relationship $[bioch] = a \times index^2 + b \times index + c$.

Name	Formula	Polynomial coefficients			RMSE ($\mu\text{g}\cdot\text{cm}^{-2}$)
		<i>a</i>	<i>b</i>	<i>c</i>	
C_{ab}					
$Chl_{RE\ opt}$	$\left(\frac{1}{R_{680-730}} - \frac{1}{R_{780-800}}\right) \times R_{755-780}$	3.96	23.86	- 3.31	6.53
$RI_{708,775}$	R_{708}/R_{775}	96.08	- 209.76	115.08	6.60
$ND_{780,712}$	$(R_{780} - R_{712}) / (R_{780} + R_{712})$	40.65	121.88	- 0.77	6.24
C_{xc}					
$Car_{RE\ opt}$	$\left(\frac{1}{R_{510-530}} - \frac{1}{R_{680-730}}\right) \times R_{760-780}$	- 0.13	3.35	0.94	2.65
$RI_{530,800}$	R_{530}/R_{800}	24.96	- 39.89	17.74	2.59
$ND_{800,530}$	$(R_{800} - R_{530}) / (R_{800} + R_{530})$	8.09	11.18	- 0.38	2.57
EWT					
$RI_{1062,1393}$	R_{1062}/R_{1393}	- 0.0006	0.019	- 0.0187	0.0031
$ND_{1062,1393}$	$(R_{1062} - R_{1393}) / (R_{1062} + R_{1393})$	0.0910	0.0164	0.0016	0.0032
LMA					
$RI_{1368,1722}$	R_{1368}/R_{1722}	- 0.0353	0.1305	- 0.0994	0.0021
$ND_{1368,1722}$	$(R_{1368} - R_{1722}) / (R_{1368} + R_{1722})$	- 0.1004	0.1286	- 0.0044	0.0021

connected from the experimental database (only used to define the leaf constituent range of variation), and two normal distributions based on the statistical properties of the experimental database. The experimental database was used for the validation and comparison of the different methodology tested.

Our main findings are that a synthetic dataset based on a normal distribution from measured experimental data, sampling #4, was the best to adjust the relationship between any leaf constituent and a spectral index. Including correlations between leaf constituents also improved the assessment of C_{xc} by removing unrealistic combinations of constituents. These relationships and the optimal relationships derived from experimental data performed very similarly. The selection of optimal wavelengths used to build spectral indices improved the assessment of C_{ab} , C_{xc} , EWT, and LMA.

It is remarkable that the performance of the optimal indices fitted on the synthetic dataset reached and even exceeded that obtained with published indices directly fitted on the LOOCV experimental dataset for any leaf constituent analyzed. Beside this good result, PLS-PRESS applied to the experimental dataset gave the best results. This study also showed that the level of noise applied to the modeled reflectance is a critical issue, whatever the model used. A value of 2% seemed to be correct to adjust a relationship to retrieve carotenoid content, whereas model inaccuracies could lead to the selection of non-optimal wavelengths for chlorophyll retrieval, and more than 0.5% noise prevented from a correct adjustment when trying to retrieve EWT and LMA with MSI and ND_{LMA} . Further studies to adapt the level of noise, for instance as a function of PROSPECT inaccuracies, are required.

As for the spectral indices, PLS-PRESS worked better when applied to normally distributed synthetic datasets. Meanwhile, the results obtained with the PLS-PRESS applied to experimental data indicated slight but significant discrepancies in PROSPECT, which were particularly noticeable when studying LMA retrieval. The results obtained with LMA showed that a relationship based on leaf reflectance and properly adjusted gave very accurate LMA assessment ($0.007\text{ g}\cdot\text{cm}^{-2}$), 60% lower than any other method based on PROSPECT modeling, suggesting

possible difficulties to build an optimal spectral index based on model simulations.

The main criticism of the method developed in this paper is that the synthetic dataset is based on experimental distributions and correlations, which may not correspond to distribution and correlations existing among certain types of vegetation. Despite our efforts to gather as many leaf species as available, collected in various ecosystems to maximize the natural spectral and biochemical variability, our experimental database is too small to represent universal distributions and correlations for leaf chemical properties. Therefore it is possible that the relationships derived here do not apply to all leaves. However, we expect a good agreement for the types of vegetation included in our dataset that already represent a wide range of terrestrial ecosystems. Our method is simple enough to update the distributions with observed or expected values and create a new synthetic dataset prior to readjustment of the relationship if it appears that our models cannot be applied to specific data.

Acknowledgments – This work was funded by several remote sensing programs: Programme National de Télédétection Spatiale (INSU), Terre Océan Surfaces Continentales Atmosphère (CNES), Earth Observing System, Carnegie Spectranomics Project, and Terrestrial Ecology (NASA). We also thank the Carnegie Institution for Science, the California Space Institute (CalSpace), the USDA Hatch program, Yves Tourbier (RENAULT technocentre) for his valuable help on designs of experiment, and Brigitte Berthelemot for support. We are also particularly grateful to researchers who contributed to this work by sharing their datasets: B. Hosgood (Joint Research Center) and J. Louis (Université Paris-Sud). IPGP contribution no. 3190.

References

- Anderson, 1986 • J. M. Anderson, Photoregulation of the composition, function, and structure of thylakoid membranes. *Annual Review of Plant Physiology*, **37** (1986), pp. 93–136.
- Asner and Martin, 2008 • G. P. Asner and R. E. Martin, Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment*, **112** (2008), pp. 3958–3970.
- Asner et al., 2009 • G. P. Asner, R. E. Martin, A. J. Ford, D. J. Metcalfe, and M. J. Liddell, Leaf chemical and spectral diversity in Australian tropical

- forests. *Ecological Applications*, **19** (2009), pp. 236–253.
- Asner et al., 2011** • G. P. Asner, R. E. Martin, R. Tupayachi, R. Emerson, P. Martinez, F. Sinca, *et al.* Taxonomy and remote sensing of leaf mass per area (LMA) in humid tropical forests. *Ecological Applications*, **21** (2011), pp. 85–98.
- Baret and Fourty, 1997** • F. Baret and T. Fourty, Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie*, **17** (1997), pp. 455–464.
- Barry et al., 2009** • K. M. Barry, G. J. Newnham, and C. Stone, Estimation of chlorophyll content in *Eucalyptus globulus* foliage with the leaf reflectance model PROSPECT. *Agricultural and Forest Meteorology*, **149** (2009), pp. 1209–1213.
- Benoist et al., 1994** • D. Benoist, Y. Tourbier, and S. Germain-Tourbier, *Plans d'expériences: construction et analyse*, Coll. Tec & Doc, Lavoisier, Paris (1994), 693 pp.
- Blackburn and Ferwerda, 2008** • G. A. Blackburn and J. G. Ferwerda, Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment*, **112** (2008), pp. 1614–1632.
- Ceccato et al., 2001** • P. Ceccato, S. Flasse, S. Tarantola, S. Jacquemoud, and J. M. Grégoire, Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, **77** (2001), pp. 22–33.
- Chappelle et al., 1992** • E. W. Chappelle, M. S. Kim, and J. E. McMurtrey, Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. *Remote Sensing of Environment*, **39** (1992), pp. 239–247.
- Chen et al., 2004** • S. Chen, X. Hong, C. J. Harris, and P. M. Sharkey, Sparse modeling using orthogonal forest regression with PRESS statistic and regularization. *IEEE Transactions on Systems, Man, and Cybernetics*, **34** (2004), pp. 898–911.
- Close and Beadle, 2003** • D. C. Close and C. L. Beadle, The ecophysiology of foliar anthocyanin. *The Botanical Review*, **69** (2003), pp. 149–161.
- Danson and Bowyer, 2004** • F. M. Danson and P. Bowyer, Estimating live fuel moisture content from remotely sensed reflectance. *Remote Sensing of Environment*, **92** (2004), pp. 309–321.
- Di Vittorio, 2009** • A. V. Di Vittorio, Enhancing a leaf radiative transfer model to estimate concentrations and in vivo specific absorption coefficients of total carotenoids and chlorophylls *a* and *b* from single-needle reflectance and transmittance. *Remote Sensing of Environment*, **113** (2009), pp. 1948–1966.
- Féret et al., 2008** • J. B. Féret, C. François, G. P. Asner, A. A. Gitelson, R. Martin, and L. P. R. Bidet, *et al.* PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, **112** (2008), pp. 3030–3043.
- Gitelson et al., 1999** • A. A. Gitelson, C. Buschmann, and H. K. Lichtenthaler, The chlorophyll fluorescence ratio F735/F700 as an accurate measure of the chlorophyll content in plants — Experiments with autumn chestnut and maple leaves. *Remote Sensing of Environment*, **69** (1999), pp. 296–302.
- Gitelson et al., 2009** • A. A. Gitelson, O. B. Chivkunova, and M. N. Merzlyak, Non-destructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany*, **96** (2009), pp. 1861–1868.
- Gitelson et al., 2003** • A. A. Gitelson, Y. Gritz, and M. N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, **160** (2003), pp. 271–282.
- Gitelson et al., 2006** • A. A. Gitelson, G. P. Keydan, and M. N. Merzlyak, Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, **33** (2006), p. L11402.
- Gitelson and Merzlyak, 1994** • A. A. Gitelson and M. N. Merzlyak, Spectral reflectance changes associated with autumn senescence of *Asculus hippocastanum* and *Acer platanoides* leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, **143** (1994), pp. 286–292.
- Gitelson and Merzlyak, 1997** • A. A. Gitelson and M. N. Merzlyak, Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, **18** (1997), pp. 2691–2697.
- Gitelson and Merzlyak, 2004** • A. A. Gitelson and M. N. Merzlyak, Non-destructive assessment of chlorophyll, carotenoid and anthocyanin content in higher plant leaves: Principles and algorithms; in S. Stamatiadis, J. M. Lynch, and J. S. Schepers, editors, *Remote Sensing for Agriculture and the Environment*, Ella, Greece (2004), pp. 78–94.
- Gitelson et al., 2001** • A. A. Gitelson, M. N. Merzlyak, and O. B. Chivkunova, Optical properties and non-destructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, **74** (2001), pp. 38–45.
- Gitelson et al., 2002** • A. A. Gitelson, Y. Zur, O. B. Chivkunova, and M. N. Merzlyak, Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology*, **75** (2002), pp. 272–281.
- Gould et al., 2008** • K. Gould, K. Davies, and C. Winefield, editors, *Anthocyanins. Biosynthesis, Functions, and Applications*, Springer (2008) 330 pp.
- Hosgood et al., 1994** • B. Hosgood, S. Jacquemoud, G. Andreoli, J. Verdebout, G. Pedrini, and G. Schmuck, Leaf Optical Properties EXperiment 93 (LOPEX93), European Commission — Joint Research Centre EUR 16095 EN, Ispra (Italy) (1994), <http://ftp-gem.jrc.it/verstmj/loplex/>; 20 pp.
- Hunt and Rock, 1989** • E. R. Hunt and B. N. Rock, Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*, **30** (1989), pp. 43–54.
- Jacquemoud and Baret, 1990** • S. Jacquemoud and F. Baret, PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, **34** (1990), pp. 75–91.
- Jacquemoud et al., 1996** • S. Jacquemoud, S. L. Ustin, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood, Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment*, **56** (1996), pp. 194–202.
- Jacquemoud et al., 1995** • S. Jacquemoud, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood, Investigation of leaf biochemistry by statistics. *Remote Sensing of Environment*, **54** (1995), pp. 180–188.
- Jørgensen et al., 2007** • R. N. Jørgensen, L. K. Christensen, and R. Bro, Spectral reflectance at sub-leaf scale including the spatial distribution discriminating NPK stress characteristics in barley using multiway partial least squares regression. *International Journal of Remote Sensing*, **28** (2007), pp. 943–962.
- Khamis et al., 1990** • S. Khamis, T. Lamaze, Y. Lemoine, and C. Foyer, Adaptation of the photosynthetic apparatus in maize leaves as a result of nitrogen limitation: Relationships between electron transport and carbon assimilation. *Plant Physiology*, **94** (1990), pp. 1436–1443.
- Kokaly et al., 2009** • R. F. Kokaly, G. P. Asner, S. V. Ollinger, M. E. Martin, and C. A. Wessman, Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, **113** (2009), pp. S78–S91.
- le Maire et al., 2004** • G. le Maire, C. François, and E. Dufrêne, Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, **89** (2004), pp. 1–28.
- le Maire et al., 2008** • G. le Maire, C. François, K. Soudani, D. Berveiller, J. Y. Pontailler, and N. Bréda, *et al.*, Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, **112** (2008), pp. 3846–3864.
- Li et al., 2007** • L. Li, S. L. Ustin, and D. Riano, Retrieval of fresh leaf fuel moisture content using genetic algorithm partial least squares (GA-PLS) modeling. *IEEE Geoscience and Remote Sensing Letters*, **4** (2007), pp. 216–220.
- Lichtenthaler et al., 1981** • H. K. Lichtenthaler, C. Buschmann, M. Doll, H. J. Fietz, T. Bach, and U. Kozel, *et al.* Photosynthetic activity, chloroplast ultrastructure, and leaf characteristics of high-light and low-light plants and of sun and shade leaves. *Photosynthesis Research*, **2** (1981), pp. 115–141.
- Pavan et al., 2004** • G. Pavan, S. Jacquemoud, L. Bidet, C. François, G. de Rosny, and J. P. Rambaut, *et al.* RAMIS: a new portable field radiometer to estimate leaf biochemical content, *Proc. 7th International Conference on Precision Agriculture and Other Precision Resources Management, Minneapolis (MN)* (2004), p. 15 (25–28 July 2004).
- Richardson et al., 2002** • A. D. Richardson, S. P. Duigan, and G. P. Berlyn, An evaluation of noninvasive methods to estimate foliar chlorophyll content. *The New Phytologist*, **153** (2002), pp. 185–194.
- Rossini et al., 2006** • M. Rossini, C. Panigada, M. Meroni, and R. Colombo, Assessment of oak forest condition based on leaf biochemical variables and chlorophyll fluorescence. *Tree Physiology*, **26** (2006), pp. 1487–1496.
- Seelig et al., 2008** • H. D. Seelig, A. Hoehn, L. S. Stodieck, D. M. Klaus, W. W. Adams, and W. J. Emery, Relations of remote sensing leaf water indices to leaf water thickness in cowpea, bean, and sugarbeet plants. *Remote Sensing of Environment*, **112** (2008), pp. 445–455.
- Shi and Sun, 2007** • R. H. Shi and J. Sun, Estimating leaf biochemical information from leaf reflectance spectrum using artificial neural network. *Proc. 6th International Conference on Machine Learning and Cybernetics, Hong Kong* (2007), pp. 2224–2228.
- Sims and Gamon, 2002** • D. A. Sims and J. A. Gamon, Relationship between pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, **81** (2002), pp. 337–354.
- Steele et al., 2008** • M. R. Steele, A. A. Gitelson, and D. C. Rundquist, A comparison of two techniques for nondestructive measurement of chlorophyll content in grapevine leaves. *Agronomy Journal*, **100** (2008), pp. 779–782.
- Ustin et al., 2009** • S. L. Ustin, A. A. Gitelson, S. Jacquemoud, M. Schaepman, G. P. Asner, J. A. Gamon, *et al.*, Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, **113** (2009), pp. S67–S77.
- Zur et al., 2000** • Y. Zur, A. A. Gitelson, O. B. Chivkunova, and M. N. Merzlyak, The spectral contribution of carotenoids to light absorption and reflectance in green leaves. *Proc. 2nd International Conference on Geospatial Information in Agriculture and Forestry, Lake Buena Vista (Florida)* (2000), pp. II.17–II.23.