PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents

Jean-Baptiste Féret, Katja Berger, Florian de Boissieu, Zbyněk Malenovský

PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents

Jean-Baptiste Féret[1], Katja Berger[2], Florian de Boissieu[1], Zbyněk Malenovský[3]

[1] TETIS, INRAE, AgroParisTech, CIRAD, CNRS, Université Montpellier, Montpellier, France

[2] Department of Geography, Ludwig-Maximilians Universität München, Luisenstr. 37, 80333 Munich, Germany

[3] School of Technology, Environments and Design, College of Sciences and Engineering, University of Tasmania, Private Bag 76, Hobart 7001, Australia

Keywords

Leaf protein content, leaf optical properties, spectroscopy, nitrogen assessment, radiative transfer modeling, PROSPECT

Abstract

Models of radiative transfer (RT) are important tools for remote sensing of vegetation, allowing for forward simulations of remotely sensed data as well as inverse estimation of biophysical and biochemical traits from vegetation optical properties. Estimation of foliar protein content is the key to monitor the nitrogen cycle in terrestrial ecosystems, in particular to assess the photosynthetic capacity of plants and to improve nitrogen management in agriculture. However, until now physically based leaf RT models have not allowed for proper spectral decomposition and estimation of leaf dry matter as nitrogen-based proteins and other carbon-based constituents (*CBC*) from optical properties of fresh and dry foliage. Such an achievement is the key for subsequent upscaling to canopy level and for development of new Earth observation applications.

Therefore, we developed a new version of the PROSPECT model, named PROSPECT-PRO, which separates the nitrogen-based constituents (proteins) from *CBC* (including cellulose, lignin, hemicellulose, starch and sugars). PROSPECT-PRO was calibrated and validated on subsets of the LOPEX dataset, accounting for both fresh and dry broadleaf and grass samples. We applied an iterative model inversion optimization algorithm and identified the optimal spectral ranges for retrieval of proteins and *CBC*. When combining leaf reflectance and transmittance within the selected optimal spectral domains, PROSPECT-PRO inversions revealed similarly accurate *CBC* estimates of fresh and dry leaf samples (respective validation R^2 = 0.96 and 0.95, *NRMSE* = 9.6% and 13.4%), whereas a better performance was obtained for fresh than for dry leaves when estimating proteins (respective validation R^2 = 0.79 and 0.57, *NRMSE* = 15.1% and 26.1%). The accurate estimation of leaf constituents for fresh samples is attributed to the optimal spectral feature selection procedure.

We further tested the ability of PROSPECT-PRO to estimate leaf mass per area (*LMA*) as the sum of proteins and *CBC* using independent datasets acquired for numerous plant species. Results showed that both PROSPECT-PRO and PROSPECT-D inversions were able to produce comparable *LMA* estimates across

an independent dataset gathering 1685 leaf samples (validation R^2 = 0.90 and *NRMSE* = 16.5% for PROSPECT-PRO, and R^2 = 0.90 and *NRMSE* = 18.3 % for PROSPECT-D). Findings also revealed that PROSPECT-PRO is capable of assessing the carbon-to-nitrogen ratio based on the retrieved *CBC*-to-proteins ratio (R^2 = 0.87 and *NRMSE* = 15.7% for fresh leaves, and R^2 = 0.65 and *NRMSE* = 28.1% for dry leaves). The performance assessment of newly designed PROSPECT-PRO demonstrates a promising potential for its involvement in precision agriculture and ecological applications aiming at estimation of leaf carbon and nitrogen contents from observations of current and forthcoming airborne and satellite imaging spectroscopy sensors.

1. INTRODUCTION

Nitrogen (N) is a major nutrient for all living plant organisms, cultivated as well as wild forms. In agriculture, crop yield quality is primarily dependent on protein content, with the N availability being the most critical factor of actual grain protein content (Brown et al., 2005). N limitation in soil and plants generally restricts the development and growth of roots, suppresses lateral root initiation, increases the carbon-to-nitrogen (C:N) ratio within the plant, reduces photosynthesis, and results in early leaf senescence (Kant et al., 2011; Paul and Driscoll, 1997; Wingler et al., 2006). On the other hand, N over-fertilization is undesirable for quality of both crops and environment. Excess of N reduces yield and decreases its quality (e.g., organoleptic quality), reduces the content of mineral nutrients and secondary metabolites, and increases nitrate content in leaves (Albornoz, 2016). From the environmental perspective, the human activity that altered the global N cycle by through N fertilization of farming systems has negative impacts on terrestrial and aquatic ecosystems (Davidson et al., 2011; Gruber and Galloway, 2008). The consequences include habitat eutrophication, acidification, and contribution to the accelerated loss of biodiversity caused by decreased competitive advantage of plants adapted to efficient use of nitrogen (Vitousek et al., 1997). Optimization of N management has, therefore, an important role

62 in mitigating such effects, while securing sufficient and sustainable food production. N concentration in
63 plants is, in general, considered as an important surrogate measure for plant photosynthetic capacity
64 (Evans, 1989), and its remote estimation is, therefore, of a great interest for plant biology and ecology.
65 Remotely sensed (RS) monitoring of N in vegetation is a prospective tool for N management improvement
66 and for reduction of negative impacts imposed by conventional farming. Decision support systems that
67 use RS information are mostly based on the relationship between leaf N and chlorophyll content. Such
68 monitoring has certain operational advantages, originating from strong chlorophyll *a*+*b* spectral
69 absorption features in the visible domain, but also from a great diversity of physically based, data driven
70 and hybrid methods designed to estimate chlorophylls from multi- and hyperspectral data (Baret et al.,
71 2007; Clevers and Gitelson, 2013; Malenovský et al., 2013; Verrelst et al., 2015). Although a significant
72 amount of literature reported a strong correlation between leaf N and chlorophyll content in crops (Baret
73 et al., 2007; Clevers and Kooistra, 2012; Vos and Bom, 1993; Yoder and Pettigrew-Crosby, 1995), this
74 relationship does not hold during their senescence and does not appear to be universal, as it is relatively
75 weak across species and ecosystems (Asner and Martin, 2009; Homolová et al., 2013). N is involved in
76 many leaf physiological processes, including photosynthesis, respiration, structural growth and storage
77 capacity building (Liu et al., 2018). This results in multiple N-based leaf biochemical constituents with
78 different physiological roles that are created throughout the plant life cycle in response to changing
79 environmental factors. Chlorophyll pigments contain only a small fraction of N, representing less than 2%
80 of the total leaf N (Kokaly et al., 2009). In comparison, proteins are the major nitrogen-containing
81 biochemical constituents, with Ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco enzyme)
82 holding 30–50% of N that is present in green leaves (Elvidge, 1990; Kokaly et al., 2009). Rubisco, the most
83 abundant protein on Earth, catalyzes the photosynthetic fixation of carbon dioxide (Sharwood, 2017).
84 Together with other photosynthesis-related proteins, it is the major source of N available for
85 remobilization among plant parts (Masclaux-Daubresse et al., 2010). For instance, N in oilseed rape

(*Brassica napus*) remobilizes from senescing to expanding leaves during the vegetative growth stage and from senescing leaves to seeds during the reproductive stage (Malagoli, 2005). This indicates that, unlike chlorophyll, plant nitrogen content does not decrease upon reaching mature growth stages, but is rather translocated to other organs, which makes the relationship between plant nitrogen and leaf chlorophyll content through the vegetation growth cycle nonlinear. Consequently, a quantitative non-destructive retrieval of leaf protein content is expected to be a more reliable proxy of nitrogen content (Berger et al., 2020b).

As reported in the pioneering studies from Curran (1989), Elvidge (1990) and Himmelsbach et al. (1988), the absorption features corresponding to proteins are caused mainly by N-H bond stretches. They are located in the shortwave infrared (SWIR) domain between 1500 and 2400 nm, with two additional features reported in the near infrared (NIR) domain at 910 and 1020 nm. The quantification of proteins from leaf optical properties is, however, challenging, because of their relatively low concentrations, and some of their specific absorption features being overlapped by absorption features from water or other dry matter constituents (Fourty et al., 1996; Jacquemoud et al., 1996). At the canopy scale, additional confounding factors (e.g., vegetation structure, geometry of acquisition or soil and atmosphere properties, etc.) also contribute to the reflectance signal measured by optical sensors. Multispectral systems with broad spectral bands and moderate spectral sampling are insufficient to correctly differentiate biochemical constituents with narrow and overlapping absorption features. The contiguous narrow spectral bands measured with imaging spectroscopy are more suitable to differentiate spectral features corresponding to the combination of multiple optically active constituents (Hank et al., 2019). Even subtle contributions of proteins to the hyperspectral signal may allow their accurate estimation, if using appropriate methods. Such methods include multivariate statistical and machine learning algorithms, physically based approaches or hybrid combinations of both (Verrelst et al., 2019a).

Physical models offer a certain number of advantages over empirical and machine learning approaches. The physically explicit representation of the interactions between electromagnetic radiation and vegetation structures enables forward simulation and inversion of reflectance signals acquired by any laboratory/field, close-range, airborne or space-borne spectroradiometer. Their main advantages, when compared to empirical methods, are robustness and transferability, although recent publications suggest that these advantages may not be as large as expected (Serbin et al., 2019). The definition of a physical model that includes proteins as an input requires the calibration of specific absorption coefficients for proteins and other constituents of dry matter in leaves, which has proven to be challenging (Botha et al., 2006; Kokaly et al., 2009). Jacquemoud et al. (1996) developed a version of the PROSPECT model including specific absorption coefficients for proteins and for different combinations of carbon-based constituents (*CBC*), but the model inversion resulted in moderate to good estimates of proteins (R^2 between 0.49 and 0.67) and different combinations of *CBC* (R^2 between 0.39 and 0.88) in dry leaves and poor to moderate accuracy for proteins (R^2<0.05) and *CBC* (R^2<0.50) in fresh leaves. Wang et al. (2015) updated a later version of PROSPECT to include proteins and lignin plus cellulose. They concluded on the importance of selecting specific spectral domains to obtain optimal results, which was later confirmed by Féret et al. (2019) when inverting PROSPECT for estimation of leaf dry mass per area (*LMA*) and equivalent water thickness (*EWT*). However, both Jacquemoud et al. (1996) and Wang et al. (2015) assumed that only proteins, lignin and cellulose, representing about 75% of *LMA*, contribute to leaf absorption. They excluded spectral contribution of non-structural carbohydrates (e.g., sugars and starch), which is a significant source of forward and inverse modelling uncertainties.

Our overall objective is to develop a new version of the PROSPECT model capable of differentiating and accurately estimating protein and *CBC* contents from leaf spectroscopic measurements. The new PROSPECT version, named PROSPECT-PRO, should be applicable to all types of bifacial leaves, including fresh green as well as senescent and dry leaves. As a secondary objective, we intend to identify optimal

spectral domains for quantitative estimation of leaf proteins through a PROSPECT-PRO inversion and validate its performance on independent datasets of leaf optical and biochemical measurements. The introduced improvements in PROSPECT-PRO leverage only shortwave infrared (SWIR, 1000-2500 nm) wavelengths, where protein and *CBC* absorption features are prominent, and therefore does not affect the existing functionality of PROSPECT with respect to foliar pigments.

We provide a general introduction to the PROSPECT model physical principles in Section 2. The data used for the calibration and validation of PROSPECT-PRO are described in Section 3. Explanation of the calibration procedure, including analytical tools for global sensitivity analysis, validation and identification of optimal retrieval spectral domains, is given in Section 4. Section 5 presents the results of the study. Finally, Section 0 discusses potential applications of and limitations to PROSPECT-PRO, are concluding findings are presented in Section 7.

2. General introduction of PROSPECT

PROSPECT is a physical model simulating leaf directional-hemispherical reflectance and transmittance (Schaepman-Strub et al., 2006) using a relatively low number of biophysical and biochemical input parameters. Its first version was developed by Jacquemoud and Baret (1990) as an extension of the generalized plate model of Allen et al. (1970, 1969), with later versions developed to include more absorbing constituents (Jacquemoud et al., 1996; Féret et al., 2008, 2017) or to adapt to specific conditions and leaf types, for example needle-shaped leaves (Malenovský et al., 2006). The PROSPECT model was also the starting point for development of independent extensions modelling RT of leaf chlorophyll fluorescence, such as FluorMODleaf (Pedrós et al., 2010) and Fluspect (van der Tol et al., 2019; Vilfan et al., 2018). PROSPECT can be used in forward mode to simulate leaf optical properties from the description of its biochemical and structural properties, or in inverse mode to estimate part or all of these

biochemical and structural properties based on measured leaf optical properties. Detailed description of these modes is provided in Section 4.

In addition to the leaf biochemical variables such as foliar pigments, *EWT* and *LMA*, PROSPECT requires a unique leaf mesophyll structure parameter N_{struct}. In a simplified leaf representation, described by the generalized plate model, it corresponds to the number of uniform compact plates separated by $N_{struct} - 1$ air spaces. N_{struct} describes the complexity of a leaf internal structure, where a low value (1-1.5) indicates a simpler compact mesophyll tissue (e.g., monocots) while a high value (1.5-2) indicates mesophyll of a greater complexity containing more intercellular air spaces (e.g., eudicots) (Boren et al., 2019). N_{struct} governs leaf internal light scattering, but it has a negligible impact on leaf absorption. Higher values of N_{struct} result in a greater reflectance and a decreased transmittance, which is obvious primarily in spectral domains of low absorption (e.g., NIR wavelengths). To date, N_{struct} is estimated indirectly from NIR leaf reflectance and transmittance measurements (Féret et al., 2019). Since we used the most recent model version PROSPECT-D as the basis for establishing a new PROSPECT-PRO, the wavelength dependent refractive index of leaf interior and the specific absorption coefficients for water remained identical to PROSPECT-D.

3. MATERIAL

a. Calibration and validation data to establish PROSPECT-PRO

The calibration and validation datasets must include directional-hemispherical leaf reflectance and transmittance and corresponding biochemical destructive measurements of the constituents used as model inputs, but only constituents with optical activity within the spectral range in which calibration is performed are needed. Since the new additions in PROSPECT-PRO utilize only the SWIR domain covering protein absorption features, contents of foliar pigments were not required for its calibration. The Leaf Optical Properties Experiment (LOPEX) dataset, established by the Joint Research Center (JRC) of the

European Commission (Ispra, Italy) in 1993 (Hosgood et al., 1994), contains optical, physical and biochemical measurements of more than 50 plant species collected around Ispra, Italy. Although this species diversity guarantees a certain variability in leaf optical and biochemical properties, the data used for the calibration of PROSPECT-PRO certainly does not cover the full range of variability across existing biomes. On the other hand, this dataset has, based on our search, the only publicly available data suitable for this calibration. We acknowledge that additional datasets obtained from various ecosystem types and growing conditions are required for a future PROSPECT-PRO verification. The optical properties of leaf directional-hemispherical reflectance and transmittance were measured with an integrating sphere from the visible (VIS) to shortwave infrared region (VSWIR, 400-2500 nm). The biochemical measurements of photosynthetic pigments, water (*EWT*) and generic dry matter (*LMA*) content, as well as carbon (C), hydrogen, oxygen, nitrogen, lignin, proteins, cellulose and starch content are expressed as a percentage of dry mass. The protein content in the original LOPEX dataset was estimated from the nitrogen content measured by the Kjeldahl method (Bradstreet, 1954; Sáez-Plaza et al., 2013) using the nitrogen-to-crude protein conversion factor of 6.25, which is widely used for food materials. We used the revised factor of 4.43, as suggested by Yeoh and Wee (1994) to be more representative of a broader range of vegetation types. This transformation of the protein content is one of the functional differences between our model and models calibrated in previous studies (Fourty et al., 1996; Jacquemoud et al., 1996; Wang et al., 2015). As the original version of LOPEX includes 120 samples, encompassing broad leaves, needles, stalks, and powders, we used only data corresponding to bifacial monocotyledon and eudicotyledon leaves. The five reflectance and transmittance measurements taken for each sample were averaged. For some samples, the measurements of optical properties were taken from both fresh and dry leaves. Therefore, we separated these measurements and produced two distinct datasets of dry and fresh samples. Chemical measurements were performed by two independent laboratories in Belgium and France (Verdebout et al., 1995). Although measurements of both laboratories were relatively consistent, we decided to use the

chemical analyses from the Belgian laboratory, leading to slightly improved overall results during calibration and validation stages. The chemical compositions measured several times over the same samples to test repeatability of lab measurement protocols were averaged. One sample of alder (*Alnus glutinosa*) with a particularly low SWIR transmittance (less than 1% on average between 1900 and 2500 nm and across spectral domains with less than 0.1% of transmittance) was discarded from the fresh samples, because its presence systematically prevented a proper calibration and validation across all tests of the data. Additionally, two fresh samples of beech (*Fagus sylvatica L.*) and poplar (*Populus canadensis*) leaves were placed in the validation data, because their presence in the calibration data resulted in systematically poor results. These three samples were all characterized by a high *EWT* > 0.030 cm (i.e. 30 $mg.cm^{-2}$). The final selection of the LOPEX dataset resulted in 66 fresh and 49 dry eligible samples.

To our best knowledge, LOPEX is the only open dataset that includes required information on leaf protein content for the calibration and validation of PROSPECT-PRO. Therefore, we split LOPEX into independent calibration and validation subsets. To minimize risks of an imbalanced distribution of protein content between calibration and validation sets, all dry and fresh samples were pooled together, and subsequently rank ordered based on increasing protein content. Every second sample among this pooled data was selected for calibration and the remaining samples were used for validation. The calibration datasets will be referred to as CALIBRATION while the validation datasets will be identified as VALIDATION, and the combined dataset will be referred to as LOPEX-CALVAL. Dataset mean and range values are provided in Table 1. The Pearson correlation coefficients for log-transformed biochemical contents of fresh leaf samples are presented in Figure 1 (Section 5.a).

225 Table 1. Statistical summary, mean values and ranges, of dry matter and protein contents and

226 concentration for fresh and dry samples included in the CALIBRATION and VALIDATION datasets.

Name	No. of samples	*LMA* (mg.cm^{-2})	Proteins (mg.cm^{-2})	Protein concentration (%DW)
		CALIBRATION		
Dry	23	5.84 (2.35-9.07)	0.78 (0.38-1.35)	14.32 (7.31-25.22)
Fresh	33	5.29 (2.58-13.69)	0.66 (0.17-1.23)	13.66 (5.02-26.06)
		VALIDATION		
Dry	26	5.89 (2.55-16.58)	0.77 (0.15-1.37)	13.98 (5.02-26.06)
Fresh	33	5.18 (1.88-10.88)	0.69 (0.29-1.22)	14.64 (6.97-28.94)

227

228 b. Data for estimation of *LMA* from PROSPECT-PRO inversion

229 Second dataset was assembled to test the capability of PROSPECT-PRO to estimate *LMA* as the
230 combination of leaf proteins plus *CBC* contents in comparison to the previous PROSPECT-D version (Féret
231 et al., 2017). For this, we combined six additional datasets that do not include destructive measurements
232 of leaf proteins: ANGERS, HYYTIALA, ITATINGA, NOURAGUES, PARACOU and LOPEX-Full (Féret et al., 2019)
233 (see Table 2 for *EWT* and *LMA* statistics). Note that LOPEX-Full includes all individual measurements of
234 leaf optical properties, i.e., 330 measurements of *LMA* and *EWT* for fresh leaves (66 fresh samples with
235 five repetitions), whereas the LOPEX CALIBRATION and VALIDATION datasets contain averages of these
236 five repetition and their corresponding mean protein contents (for more details see Hosgood et al., 1994).

237

238

Table 2. Statistical summary, mean values and ranges, of water and dry matter contents for experimental datasets used to validate *LMA* estimations from the PROSPECT-PRO inversion.

Name	No. of samples	*EWT* (mg.cm^{-2})	*LMA* (mg.cm^{-2})
ANGERS	308	11.47 (4.40 – 34.00)	5.12 (1.66 – 33.1)
HYYTIALA	96	9.16 (3.68 – 23.73)	6.27 (2.76 – 15.77)
ITATINGA	415	14.44 (2.20 – 20.20)	10.24 (6.90 – 14.70)
LOPEX-Full	330	11.13 (0.29 – 52.49)	5.30 (1.71 – 15.73)
NOURAGUES	262	11.73 (3.20 – 38.10)	10.81 (3.10 – 21.10)
PARACOU	272	*N/A* (*N/A* – *N/A*)	12.32 (5.28 – 25.56)

4. METHODS

a. PROSPECT forward modelling and inversion

In forward mode, PROSPECT simulates leaf optical properties based on a set of biophysical and biochemical properties (N_{struct} and leaf biochemistry). In inverse mode, the optimal set of biophysical and biochemical properties can be identified via a variety of methods, for example, using a merit function that minimizes the difference between measured and simulated LOP. A common inversion procedure is based on the numerical minimization of the sum of weighted square errors over all available spectral bands (Baret and Buis, 2008; Féret et al., 2019). The minimized merit function M, using both reflectance and transmittance properties, is expressed as follows:

$$M\left(N_{struct}, \{C_i\}_{i=1:p}\right) = \sum_{\lambda=\lambda_1}^{\lambda_n} \left[W_{R,\lambda} \times \left(R_\lambda - \hat{R}_\lambda\right)^2 + W_{T,\lambda} \times \left(T_\lambda - \hat{T}_\lambda\right)^2\right], \quad (1)$$

where p is the number of chemical constituents accounted for by PROSPECT and retrieved during the inversion, C_i the biochemical content per unit of leaf surface for a constituent i, λ_1 and λ_n are the first and last wavebands entering the inversion, R_λ and T_λ are the experimental reflectance and transmittance measured at waveband λ, $\hat{R}_\lambda$ and $\hat{T}_\lambda$ are the reflectance and transmittance simulated by PROSPECT with $\{N_{struct}, \{C_i\}_{i=1:p}\}$ as input variables, and $W_{R,\lambda}$ and $W_{T,\lambda}$ are the weights applied to the squared difference between experimental and simulated reflectance and transmittance, respectively. Eq. (1) can be used to estimate all input variables, or just their limited subset, if a prior information or arbitrary values of some variables are known. In this study, the values of $W_{R,\lambda}$ and $W_{T,\lambda}$ were set to 0 for non-selected and 1 for selected spectral bands, giving all the selected wavelengths the same importance.

b. Calibration of PROSPECT-PRO

The previous PROSPECT versions (Féret et al., 2017, 2008) had the specific absorption coefficients of *LMA* defined by implicitly accounting for various dry matter constituents. Since the distinction of all individual *LMA* constituents is beyond the scope of this study, our primary objective is to replace *LMA* by nitrogen-containing proteins and *CBC* as new leaf input constituents in PROSPECT-PRO. Lignin, cellulose, hemicellulose and non-structural carbohydrates (sugars and starch), were grouped in a single unique input called *CBC*, while the remaining nitrogen-based proteins represent the second standalone input. Please note that from hereafter we refer to the nitrogen-based proteins simply as proteins. Each constituent of *CBC* has a specific carbon content (Ma et al., 2018) but does not contain N. We used both dry and fresh leaf samples in the CALIBRATION dataset to calibrate the specific absorption coefficients corresponding to these two groups of leaf constituents, assuming that *LMA* can be split into protein and other *CBC* contents as follows:

$$LMA = Protein\ content + CBC\ content. \quad (2)$$

where *LMA*, protein content and *CBC* content are expressed in mass per leaf surface unit (mg.cm^{-2}). This ensures conservation of the mass of absorbing materials and allows us to invert PROSPECT-PRO for an estimation of *LMA* as the sum of leaf protein and *CBC* contents.

Absorption $k(\lambda)$ of a compact leaf layer at wavelength λ, for a given mesophyll structural parameter N_{struct}, is in every PROSPECT model defined as:

$$k(\lambda) = \frac{\sum_i K_{spe,i}(\lambda) \times C_i}{N_{struct}}, \tag{3}$$

where $K_{spe,i}(\lambda)$ is the specific absorption coefficient of a constituent i, and C_i is its corresponding content. In PROSPECT-D, only two input constituents contribute to absorption in the spectral region from 1000 to 2500 nm (focus of this study): water (*EWT*), with a negligible absorption before 1100 nm, and dry matter (*LMA*), with a constant absorption between 1000 and 1200 nm. Additional leaf constituents accounted for in PROSPECT-D and absorbing in the VIS-NIR spectral domain up to 1100 nm are brown pigments (Ustin and Jacquemoud, 2020). The brown pigments, observed in senescent leaves as result from oxidation and polymerization of cell constituents, are excluded from our analysis because they exhibit only a minor absorption between 1000 and 1100 nm. Therefore, Eq. (3) can be for the purpose of PROSPECT calibration between 1000 and 2500 nm written as follows:

$$k(\lambda) = \frac{K_{spe,EWT}(\lambda) \times C_{EWT} + K_{spe,LMA}(\lambda) \times C_{LMA}}{N_{struct}}. \tag{4}$$

Following the equivalence in Eq. (2), the contribution of *LMA* to the total absorption can then be decomposed into the proteins and *CBC* as:

$$K_{spe,LMA}(\lambda) \times C_{LMA} = K_{spe,PROT}(\lambda) \times C_{PROT} + K_{spe,CBC}(\lambda) \times C_{CBC}, \quad (5)$$

where $K_{spe,PROT}(\lambda)$ is specific absorption coefficient for proteins, $K_{spe,CBC}(\lambda)$ is specific absorption coefficient for the *CBC* (both in $cm^2.mg^{-1}$), and C_{PROT} and C_{CBC} are the corresponding contents (in $mg.cm^{-2}$), respectively. We assume that $K_{spe,LMA}$ in PROSPECT-D is accurately calibrated, and we use it as a constraint for the calibration of $K_{spe,PROT}$ and $K_{spe,CBC}$, based on Eq. (5).

The calibration of PROSPECT-PRO followed the commonly-used two-step process (Féret et al., 2017, 2008; Jacquemoud and Baret, 1990) that included the additional constraint, i.e. the decomposition of absorption for *LMA* into proteins and *CBC*. First, we determined the leaf structure parameter $N_{struct,j}$ of each leaf j in the calibration datasets. $N_{struct,j}$ was estimated based on a multivariate iterative optimization, simultaneously with three absorption coefficients, using reflectance and transmittance values measured at three wavelengths corresponding to the minimum absorptance (λ_1), maximum reflectance (λ_2), and maximum transmittance (λ_3) of the leaf (Jacquemoud et al., 1996). These values are generally located on the NIR reflectance and transmittance plateau. The iterative optimization was performed using the following merit function:

$$M_{leafN}\left(N_{struct,j}, k(\lambda_1), k(\lambda_2), k(\lambda_3)\right) = \sum_{l=1}^{3}\left[\left(R_{meas,j}(\lambda_l) - R_{mod}\left(N_{struct,j}, k(\lambda_l)\right)\right)^2 + \left(T_{meas,j}(\lambda_l) - T_{mod}\left(N_{struct,j}, k(\lambda_l)\right)\right)^2\right], \quad (6)$$

where $R_{meas,j}(\lambda_l)$ and $T_{meas,j}(\lambda_l)$ are measured directional-hemispherical reflectance and transmittance of leaf j at wavelength λ_l, R_{mod} and T_{mod} are the respective modeled values, and $k(\lambda)$ is the specific

absorption coefficient of a compact layer at the wavelength λ, which is being adjusted simultaneously with $N_{struct,j}$.

In the second step, $K_{spe,PROT}$ and $K_{spe,CBC}$ were computed by inverting PROSPECT-PRO and using the CALIBRATION dataset for each spectral band of interest independently. In order to include the constraint defined in Eq. (5), the minimization algorithm was executed in two consecutive phases, which were embedded (nested) in a unique iterative procedure for optimization of $K_{spe,PROT}(\lambda)$ and $K_{spe,CBC}(\lambda)$. During the first phase, the estimated value of $K_{spe,CBC}(\lambda)$ was computed by resolving a system of the following linear equations:

$$K_{spe,CBC}(\lambda)\begin{bmatrix} C_{CBC,1} \\ C_{CBC,2} \\ \cdot \\ \cdot \\ \cdot \\ C_{CBC,n} \end{bmatrix} = \begin{bmatrix} K_{spe,LMA}(\lambda) \times C_{LMA,1} - K_{spe,PROT}(\lambda) \times C_{PROT,1} \\ K_{spe,LMA}(\lambda) \times C_{LMA,2} - K_{spe,PROT}(\lambda) \times C_{PROT,2} \\ \cdot \\ \cdot \\ \cdot \\ K_{spe,LMA}(\lambda) \times C_{LMA,n} - K_{spe,PROT}(\lambda) \times C_{PROT,n} \end{bmatrix}, \tag{7}$$

where $K_{spe,PROT}(\lambda)$ is initially set to a user-defined value, then updated at each iteration. In the second phase, the optimal value of $K_{spe,PROT}(\lambda)$ was adjusted by following the strategy defined by the sequential quadratic programming algorithm (Fletcher, 2000). We minimized the following merit function J per wavelength (λ):

$$\begin{aligned} J\left(\{K_{spe,i}(\lambda)\}_{i=1:n}\right) \\ = \sum_{j=1}^{n}\Bigg[\left(R_{meas,j}(\lambda) - R_{mod,j}\left(N_{struct,j}, k(\lambda)\right)\right)^2 \\ + \left(T_{meas,j}(\lambda) - T_{mod,j}\left(N_{struct,j}, k(\lambda)\right)\right)^2\Bigg], \end{aligned} \tag{8}$$

with $k(\lambda)$ defined as follows:

$$k(\lambda) = \frac{K_{spe,EWT}(\lambda) \times C_{EWT} + K_{spe,PROT}(\lambda) \times C_{PROT} + K_{spe,CBC}(\lambda) \times C_{CBC}}{N_{struct}}, \quad (9)$$

where $K_{spe,PROT}(\lambda)$ is the only unknown term and $K_{spe,CBC}(\lambda)$ is taken from the previous phase. In other words, the $K_{spe,CBC}(\lambda)$ and $K_{spe,PROT}(\lambda)$ values were updated during each iteration until the procedure found the optimum for $K_{spe,PROT}(\lambda)$, and then the final value of $K_{spe,CBC}(\lambda)$ was obtained from Eq. (7), using the optimized $K_{spe,PROT}(\lambda)$.

This calibration procedure, expecting correctly defined $K_{spe,LMA}$, was performed within the spectral domain of 1000 to 2500 nm, using the same leaf refractive index and the specific absorption coefficients for *EWT* as defined in PROSPECT-D. The calibration of specific absorption coefficients in the NIR domain is challenging due to a generally low absorption and a possibly significant uncertainty in measured leaf optical properties within this spectral domain. When calibrating the specific absorption for *LMA*, Féret et al. (2008) set the specific absorption coefficients of *LMA* to a constant value for wavelengths < 1200 nm. Here, we fitted an exponential function to the start of absorption of proteins, ensuring smooth transition between the two non-absorptive and absorptive spectral domains, corresponding to the spectral domain between 1440 and 1490 nm. Subsequently, the specific absorption coefficients for *CBC* between 1000 and 1200 nm were adjusted according to the specific absorption coefficients of *LMA* from PROSPECT-D, by applying a multiplicative factor corresponding to the average ratio between *LMA* and *CBC* in the CALIBRATION dataset.

c. Global sensitivity analysis of PROSPECT-PRO

A global sensitivity analysis (GSA) was carried out for PROSPECT-PRO to quantify the contribution of proteins and *CBC* constituents to the overall spectral signal. Using a GSA the driving variables of a radiative transfer model can be identified by fully exploring the input parameter space (Verrelst et al., 2019b; Wang et al., 2015). The Matlab software tool GSAT (Cannavó, 2012), which includes Fourier amplitude sensitivity testing (FAST) analysis and Sobol's method for calculation of the first-order sensitivity coefficients was applied on PROSPECT-PRO simulations from 1000 nm to 2500 nm carried out with the following realistic input parameter ranges for fresh leaves: N_{struct}~ 1-2 (unitless), *EWT* ~ 0.001-0.015 cm, protein content (C_p) ~ 0-0.003 g/cm² and *CBC* content ~ 0-0.01 g/cm². The remaining PROSPECT-PRO input parameters, i.e. chlorophyll content, total carotenoid content, anthocyanin content and brown pigment content, were fixed to arbitrary values since they manifest no absorption between 1000-2500 nm.

d. Optimal spectral domains for estimation of protein and *CBC* content

In previous studies (Colombo et al., 2008; Féret et al., 2008; Jacquemoud et al., 1996), a PROSPECT model inversion was performed with an iterative optimization of the merit function defined in Eq. (8) over the entire optical domain or broad spectral intervals (e.g., VIS-NIR when retrieving leaf pigments and NIR-SWIR when retrieving *EWT* and *LMA*), using a uniform weight of 1 across all spectral bands. Féret et al. (2019) showed the importance of identifying optimal spectral domain for the accurate estimation of *LMA* and, to a lesser extent, *EWT*. Investigating a number of spectral ranges between 1000 and 2400 nm, they recommended determination of *EWT* and *LMA* with an iterative optimization of leaf optical properties between 1700 and 2400 nm. Although proteins are part of *LMA*, the hypothesis that proteins and *LMA* share the same optimal retrieval spectral domain needs appropriate testing. Moreover, the fact that protein absorption is expressed in a number of narrow SWIR spectral features (Curran, 1989; Fourty et al., 1996) suggests that proteins and *CBC* may have slightly different optimal retrieval spectral domains compared to *LMA*.

The procedure suggested by Féret et al. (2019) is limited to the identification of an optimal contiguous spectral domain. As such, it is unable to identify spectral features located in narrow non-contiguous domains separated by suboptimal spectral intervals of varying lengths. To be able to identify also the non-contiguous optimal spectral domains, we adapted a sequential forward feature selection (SFS) technique (Kudo and Sklansky, 2000; Marcano-Cedeno et al., 2010). SFS is a bottom-up search procedure that starts from an empty feature set and gradually adds features selected based on a minimization criterion. In our study, each spectral feature was defined as a set of 20 spectral bands of the original leaf optical properties with the 1 nm spectral sampling, which allowed for a large number of explored spectral bands and made the computation of the iterative optimization feasible. This way, we created 50 spectral features between 1400 and 2399 nm and applied SFS on these spectral features. We first identified the spectral feature leading to minimum root mean square error (*RMSE*) when estimating either proteins or *CBC* from the experimental data. Then we searched for the spectral feature leading to minimum *RMSE,* if combined with the previously identified spectral feature, until all spectral features were tested. At the final step, all features were sequentially added and ranked based on the search for minimum *RMSE*, and the full domain 1400 and 2399 nm was used for inversion with the original spectral data.

As the CALIBRATION dataset was dedicated to the calibration of the specific absorption coefficients of proteins and *CBC*, the accuracy and robustness of PROSPECT inversion for retrieval of these leaf constituents had to be assessed with the independent VALIDATION dataset. Use of the CALIBRATION dataset would be logically defrauded and scientifically incorrect. The selection of optimal spectral domains is not considered as a part of the calibration procedure but expected to be performed before or during a PROSPECT inversion. As such, its robustness would be best ensured when performed on an independent dataset. Unfortunately, due to a limited size of available experimental data, the sample pool could not be reasonably split into three independent CALIBRATION, FEATURE SELECTION and VALIDATION parts. Therefore, the optimal spectral domains were identified by applying SFS on the VALIDATION dataset.

Once the specific optimal spectral domains for estimation of proteins and *CBC* were identified, we compared their estimates with results obtained for the spectral domain between 1700 and 2400 nm, identified as optimal for the estimation of *LMA* and *EWT* by Féret et al. (2019). Finally, we compared the performances of PROSPECT-PRO and PROSPECT-D for the estimation of *LMA* using the 1700-2400 nm spectral domain, and by inverting PROSPECT-PRO over the SFS identified optimal spectral domains for proteins and *CBC* estimations and calculating *LMA* from Eq. (2). This model comparison was carried out for the six additional datasets listed in Table 2.

The normalized *RMSE* (*NRMSE* expressed in %) was computed to appraise the difference between the measured and estimated leaf constituents retrieved from the different datasets:

$$NRMSE = \frac{1}{\overline{X_{meas}}}\sqrt{\frac{\sum_{j=1}^{n}\left(X_{meas,j} - X_{mod,j}\right)^2}{n}}, \tag{10}$$

where $X_{meas,j}$ is the measured value and $X_{mod,j}$ is the values estimated by model inversion for a leaf j, $\overline{X_{meas}}$ is the mean value of the constituent, and n is the number of samples.

All inversions and optimal feature selections were performed with the `prospect` R package (Féret and Boissieu, 2020), which uses the nonlinear constrained multivariable minimization function of the `pracma` R package (Borchers, 2019).

e. Performances of PROSPECT-D and PROSPECT-PRO in forward modelling

We compared the performances of PROSPECT-D and PROSPECT-PRO for the forward simulation of leaf optical properties, using the structure parameter N_{struct} obtained from inversion and the corresponding biochemical constituents measured in laboratory. This statistical analysis was undertaken to reveal spectral domains impacted by high levels of uncertainty, which is relevant for possible future hybrid inversion applications involving machine learning algorithms trained with PROSPECT-PRO simulated data

(e.g., Verrelst et al., 2015 and 2019a). The comparison was performed by computing the per-wavelength spectral *RMSE* between measured and simulated reflectance and transmittance of fresh and dry samples from the VALIDATION dataset, and also systematic ($RMSE_S$) and unsystematic ($RMSE_U$) parts of *RMSE* (Willmott et al., 1985) defined as:

$$RMSE_S = \sqrt{\frac{\sum_{j=1}^{n}\left(\hat{X}_{meas,j} - X_{mod,j}\right)^2}{n}}, \tag{11}$$

and

$$RMSE_U = \sqrt{\frac{\sum_{j=1}^{n}\left(\hat{X}_{meas,j} - X_{meas,j}\right)^2}{n}}, \tag{12}$$

where $\hat{X}_{meas,j}$ is an ordinary least square estimate of $X_{meas,j}$, and n is the number of samples. Eqs. (11) and (12) are a complete partitioning of *RMSE* as follows:

$$RMSE^2 = {RMSE_U}^2 + {RMSE_s}^2 \tag{13}$$

The $RMSE_S$ corresponds to the linear bias of the estimate produced by the model itself, while the $RMSE_U$ corresponds to a measure of precision of the model and it is driven by uncertainties in input data.

f. Estimation of the carbon-nitrogen ratio

The carbon:nitrogen (C:N) ratio of plant canopies, crops and crop residues is of great importance for modelling C and N dynamics in natural ecosystem and agricultural systems, as it contains an indicative

information about plant growth rate and affects ecosystem response to CO_2 (Reich et al., 2006; Zheng, 2009). This C:N ratio is also an indicator of the relative allocation of resources in vegetation, an indicator of potential decomposition rate of litter and an important factor promoting soil organic carbon accumulation (Zhou et al., 2019). Thus, we tested the possibility of using the *CBC*:Proteins ratio, estimated from PROSPECT-PRO inversion, as a proxy for the C:N ratio of leaf samples in the LOPEX dataset. We established a linear model to estimate the C:N ratio based on the *CBC*:Proteins ratio as measured in the fresh samples of the CALIBRATION dataset. We then applied this linear relationship on *CBC*:Proteins ratio retrieved from leaf optical properties through PROSPECT-PRO inversion to estimate the C:N ratio for all samples in both CALIBRATION and VALIDATION datasets.

5. RESULTS

a. Correlations among biochemical constituents of fresh leaves in LOPEX-CALVAL data

Since descriptive statistics computed for leaf constituents of fresh samples in the LOPEX-CALVAL dataset revealed that a majority of them does not follow the Gaussian distribution, they were log-transformed by applying the natural logarithm. The subsequent correlation analysis performed on log-transformed data indicated potential relationships between individual biochemical compounds. Figure 1 shows the Pearson correlation coefficients (r) for tested constituents, including the C:N ratio. Proteins are not included, as they were derived directly from N measurements. The coefficients highlight strong and statistically significant relationships between carbon (C), hydrogen (H), oxygen (O), lignin, cellulose and *LMA*. Nitrogen (N) content is moderately correlated to chlorophyll *a*+*b* content (CHL), C, H, O, *LMA* and *EWT*. The moderate correlation between CHL and N (r = 0.51) indicates a modest capacity of CHL to estimate N across species in the LOPEX dataset. Finally, the C:N ratio was found to be moderately positively correlated with *LMA*, C, H, O and individual *CBC* except starch, and poorly negatively correlated with the N content (r = -0.36).

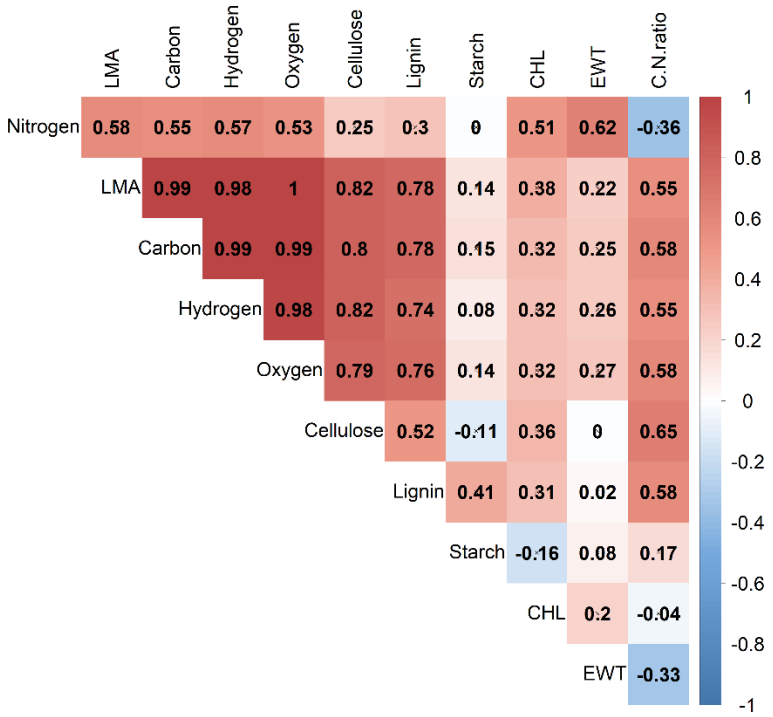

Figure 1. Pearson correlation coefficients computed among log-transformed contents of biochemical constituents of fresh leaf samples in the LOPEX-CALVAL dataset.

b. Calibration of PROSPECT-PRO

A Matlab version of the new PROSPECT-PRO model is downloadable from the following GitLab repository: https://gitlab.com/jbferet/prospect_pro_matlab. The R package is including PROSPECT-D is installable from the GitLab repository: https://jbferet.gitlab.io/prospect/. The specific absorption coefficients derived for leaf proteins and *CBC* are displayed in Figure 2. Most of the absorption features reported by Curran (1989) and Fourty et al. (1996) correspond with the local maxima of the obtained specific absorption of proteins, although some of them are spectrally shifted towards shorter or longer wavelengths.

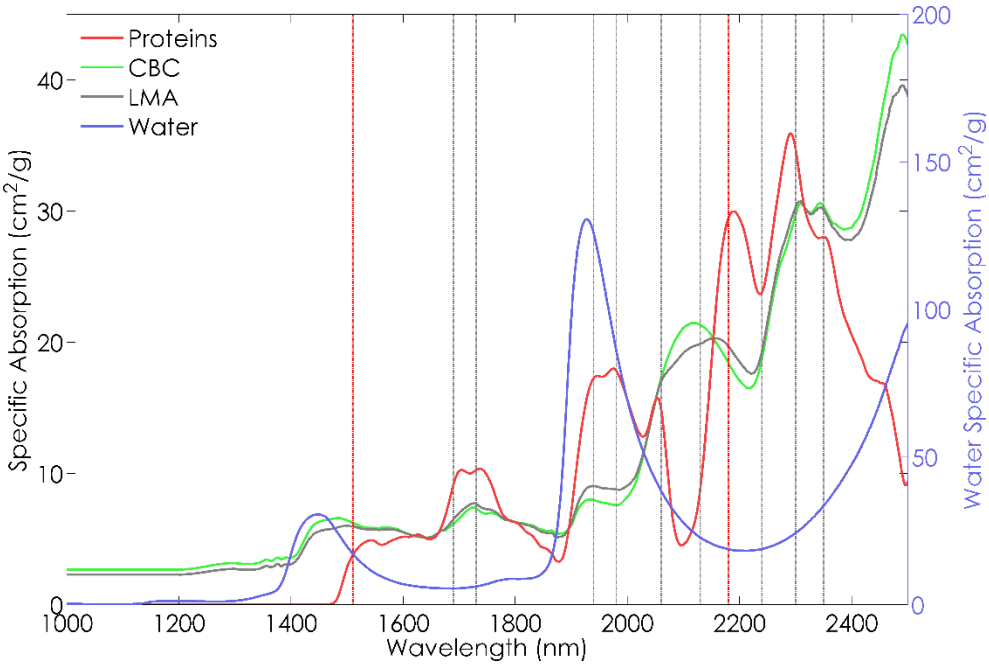

Figure 2. Specific absorption coefficients for proteins and *CBC* obtained from calibration of PROSPECT-PRO using the CALIBRATION dataset. The coefficient corresponding to *LMA* and water, used in PROSPECT-D, are displayed for comparison. Vertical dashed lines indicate wavelengths of absorption features linked to proteins by Curran (1989) and Fourty et al. (1996) (red = major and grey = minor absorption features).

c. Sensitivity of leaf optical properties to proteins and *CBC*

Results of GSA for PROSPECT-PRO simulated leaf reflectance and transmittance were nearly identical. Therefore, we present in Figure 3 only the outcomes for reflectance and absorptance of fresh and dry leaves. GSA identified the spectral regions that contain absorption peaks of proteins (> 1400 nm), but it also shows their relatively low contribution to the spectral signal in these wavelengths, in particular for fresh leaves. *CBC* play a larger role in driving leaf reflectance and absorptance, with their highest relevance in SWIR, especially above 2000 nm. Yet, the key driving input parameters of PROSPECT-PRO forward simulations of fresh leaf reflectance are the N_{struct} parameter and *EWT*. As expected, N_{struct} has no impact on leaf absorptance. Water and *CBC* are the dominant absorbents of fresh and dry leaves,

487 respectively, in the SWIR domain. Although absorption features of proteins between 1600 and 1800 nm
488 and between 2100 and 2300 nm are subtle, GSA confirms that these spectral domains have the greatest
489 potential for retrieval activities for both dry and fresh leaves.
490 Nevertheless, data of high spectral sampling and resolution with a sufficiently high signal-to-noise (SNR)
491 and an efficient identification of the most optimal retrieval wavelengths within these spectral domains
492 are required to enable the separation of all influencing constituents, especially in future efforts when
493 upscaling the retrieval methods from the leaf to the top-of-canopy level.
494

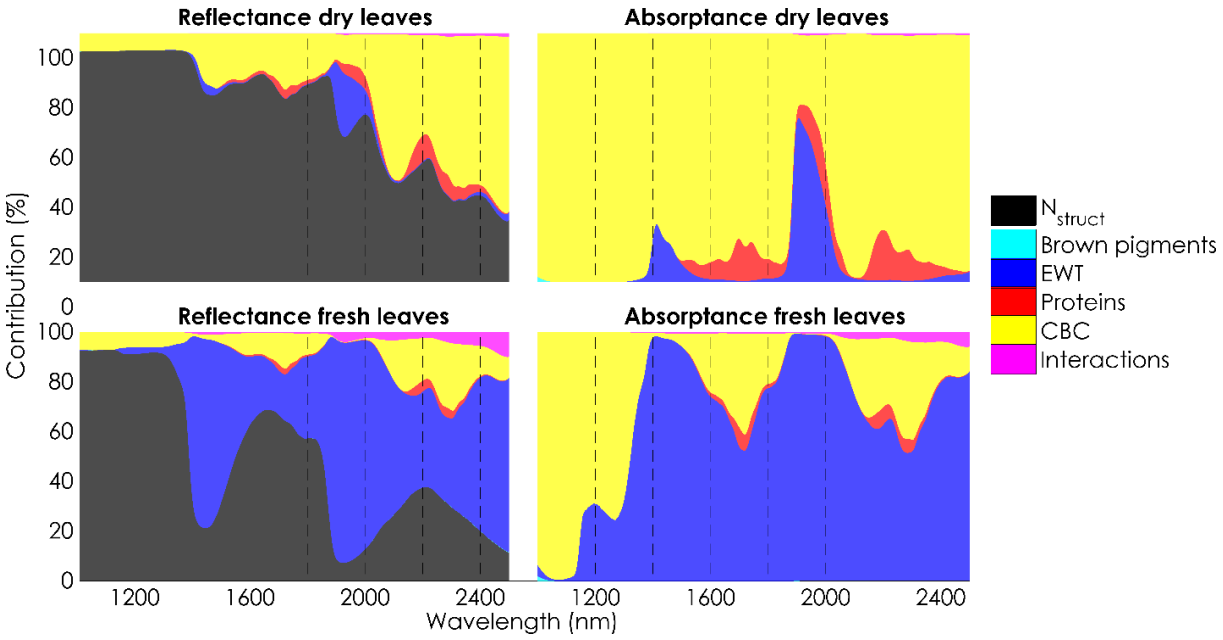

Figure 3. Global sensitivity analysis of PROSPECT-PRO input parameters, i.e. leaf structure (N_{struct}), leaf water content (*EWT*), protein content, carbon-based constituents (*CBC*) and brown pigments, simulating reflectance (left) and absorptance (right) of dry (top) and fresh leaves (bottom), including reciprocal interactions in a typical bifacial leaf. The y-axis ('contribution') quantifies the first-order effects, implying the contribution of each tested input to the modelled output variance.

495

496 d. Optimal wavelengths for PROSPECT-PRO retrieval of proteins and *CBC* content

Figure 4 shows the *NRMSE* for the estimation of protein and *CBC* contents by inverting PROSPECT-PRO over the VALIDATION dataset and applying the SFS optimization procedure on spectral features with 20 nm width within the spectral domain between 1400 and 2400 nm. The optimal estimation of proteins was obtained with three spectral features encompassing the spectral domains between 2100 and 2139 nm, and between 2160 and 2179 nm. The later one is located next to the strong protein absorption feature centered at 2180 nm, as noted by Curran (1989), Fourty et al. (1996) and Wang et al. (2015). The optimal estimation of *CBC* was obtained when selecting thirteen 20 nm wide spectral features, four of them located between 1480 and 1800 nm, and nine of them between 2040 and 2399 nm. Inclusion of additional spectral features did not lower accuracy for the *CBC* estimations, except for spectral domains between 1400 and 1439 nm, and between 1860 and 2000 nm that correspond to the two main water absorption features. In the case of proteins, inclusion of additional spectral information besides the identified optimal spectral features led to an increased *NRMSE.* The maximum *NRMSE* increase was obtained when including spectral information corresponding to the main absorption peak of water between 1880 and 2000 nm.

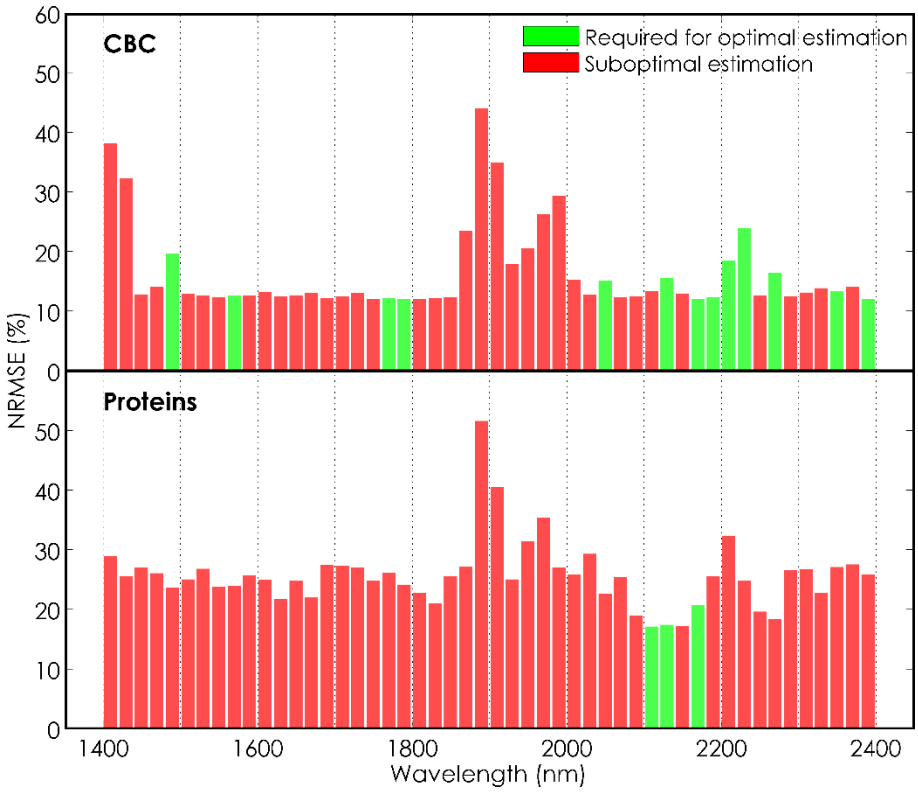

Figure 4. *NRMSE* (%) obtained for the estimation of protein and *CBC* contents with a PROSPECT-PRO inversion applied on the VALIDATION dataset using the SFS method (green = spectral features required to reach the minimal *NRMSE* and red = suboptimal spectral domains increasing *NRMSE*).

Figure 5 illustrates the evolution of *NRMSE* as the number of spectral features selected with SFS for the full VALIDATION dataset, as well as for the dry samples and fresh samples separately, increases. The *NRMSE* development fluctuates but stays relatively similar until 40 to 45 spectral features when estimating protein contents for both types of leaves. It dramatically increases for fresh leaves once spectral features located on the main absorption peak of water are included. Similar results were obtained for *CBC*. The errors remain relatively similar around the optimal performance until 40 spectral features, and also strongly increase when including spectral domains of water absorption.

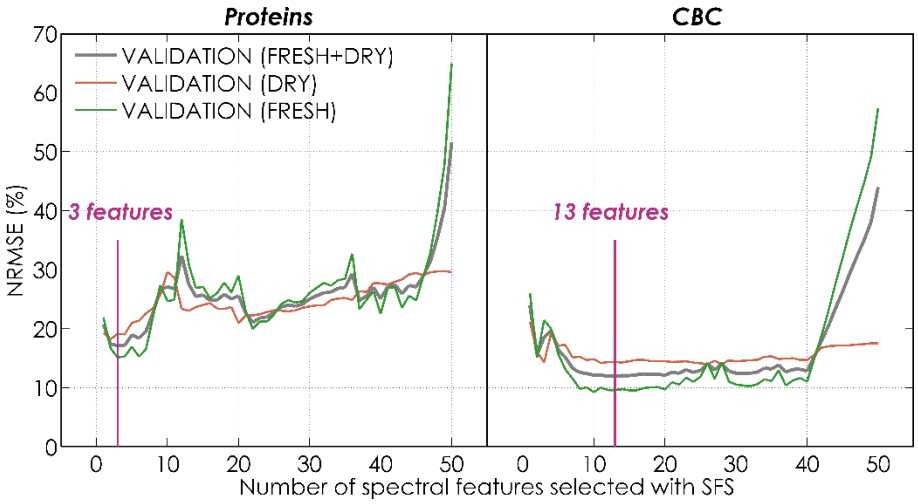

Figure 5. Evolution of *NRMSE* with an increasing number of spectral features selected with SFS when estimating proteins (left) and *CBC* (right) (grey line = *NRMSE* computed for all samples in the VALIDATION dataset, red line = *NRMSE* computed for dry samples and green line = *NRMSE* computed for fresh samples in the VALIDATION dataset. The number of features leading to minimum *NRMSE* for all VALIDATION samples is indicated with the violet vertical line.

e. PROSPECT-PRO validation by retrieval of leaf protein and *CBC* contents

Comparison between laboratory measured and PROSPECT-PRO estimated protein, *CBC* and *LMA* contents for the VALIDATION dataset is shown in Figure 6. The simultaneous retrievals of leaf protein and *CBC* contents from inversions were performed either over the spectral region from 1700 to 2400 nm or over the optimal spectral features identified with the SFS method. The R codes and data used to produce these results are available in the prospect R package (Féret and Boissieu, 2020).

The results illustrate the importance of selected optimal spectral features for accuracy of leaf constituent estimates, especially when analyzing fresh leaf samples. The estimation of proteins has a slightly higher uncertainty than the estimation of *CBC* and *LMA*, which can be explained by a lower contribution of proteins to the spectral signal (Figure 3). Additionally, a lower accuracy obtained for the estimation of protein content from dry leaves is caused mainly by a single sample of dry maple (*Acer pseudoplatanus L*)

leaf. The estimation of *LMA* based on the inversion of PROSPECT-D using the spectral domain between 1700 and 2400 nm produced results similar to those obtained with PROSPECT-PRO (*NRMSE* = 14.2% for dry leaves and *NRMSE* = 34.3% for fresh leaves).

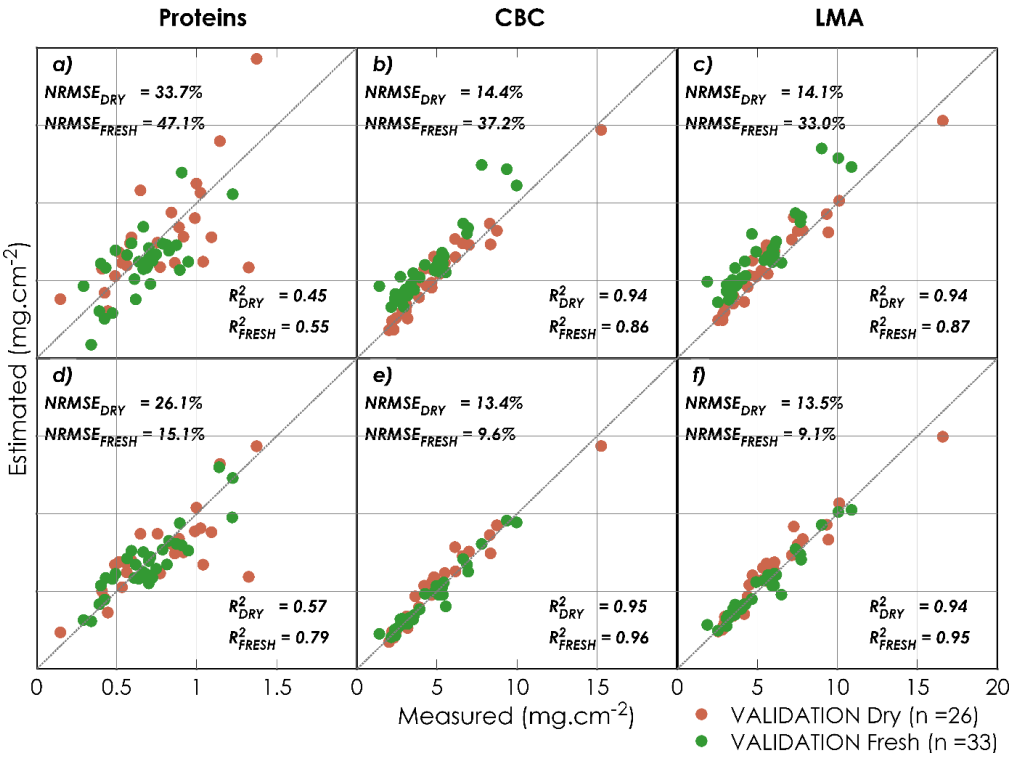

Figure 6. Comparison between laboratory measured and PROSPECT-PRO estimated leaf protein, *CBC* and *LMA* (proteins + *CBC*) contents obtained for the VALIDATION dataset using either the spectral range from 1700 to 2400 nm (a to c) or the optimal spectral features identified with the SFS method (d to f) (see Figure 4 in Section 5.d).

f. PROSPECT-PRO and PROSPECT-D compatibility assessed via estimation of *LMA* and *EWT*

Overall, the decomposition of *LMA* into protein and *CBC* contents estimated by the PROSPECT-PRO over the SFS optimized spectral regions slightly outperformed the *LMA* estimations obtained with the

PROSPECT-D inversion over the optimal spectral domain 1700-2400 nm identified by Féret et al. (2019) (Figure 7). When analyzing the results per dataset, inversion of PROSPECT-PRO using the optimal spectral domain for protein and *CBC* content retrievals resulted in a decreased *NRMSE* for *LMA* estimations for five out of eight datasets (including the VALIDATION Dry and VALIDATION Fresh datasets). Very similar performances were found for ITATINGA. The increase in *NRMSE* observed for ANGERS was caused by two samples of Holly osmanthus (*Osmanthus heterophyllus*), characterized by high *LMA* and *EWT* values, while the remaining samples showed comparable estimates. However, this influence of high *EWT* or *LMA* contents was not a general feature, as this effect was not observed in case of other datasets that include samples with high *LMA*. The less accurate performances observed for HYYTIALA corresponds to an increased uncertainty distributed among all samples. Finally, when combining all datasets described in Table 2, the indirect estimation of *LMA* from the inversion of PROSPECT-PRO using the optimal spectral features was slightly improved, with a 1.8% decrease in *NRMSE* and comparable R^2 across all data sets. Still, the differences observed among independent datasets suggest that the optimal spectral features computed for the VALIDATION dataset do not correspond exactly with the optimal spectral features for other datasets. Finally, the version of the model did not influence the estimation of *EWT* significantly. Inversions of PROSPECT-D and PROSPECT-PRO conducted over the same spectral domain (1700-2400 nm) resulted in similar outcomes (*NRMSE* = 11.9 and R^2 = 0.91 for both model inversions when combining all datasets). The results confirm the compatibility between PROSPECT-D and PROSPECT-PRO.

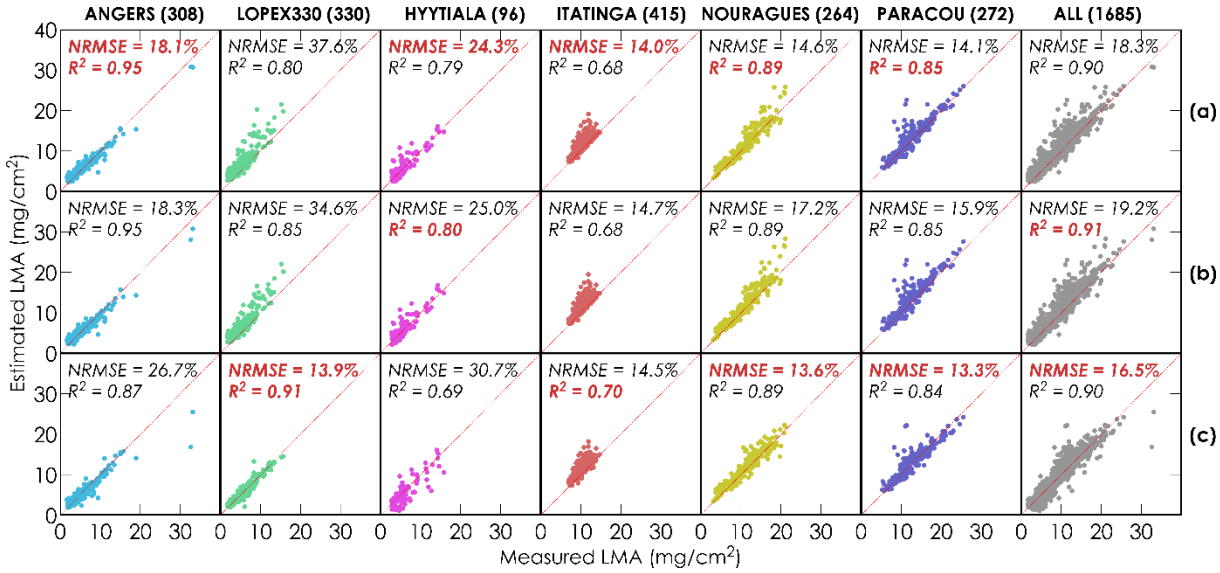

Figure 7. Comparison between measured *LMA* and its corresponding estimations by inversion of a) PROSPECT-D (1700-2400 nm), b) PROSPECT-PRO (1700-2400 nm), and c) PROSPECT-PRO (SFS optimized spectral features for proteins and *CBC*). Values in brackets show the number of samples and red fonts indicate the best achieved results per dataset.

g. Forward simulations of leaf optical properties

Figure 8 displays per-wavelength $RMSE$, $RMSE_S$ and $RMSE_U$ calculated between measured leaf reflectance and transmittance and their counterparts produced by PROSPECT-PRO for VALIDATION dry and fresh samples and also samples of the six independent datasets in Table 2 grouped together. The input biochemical constituents correspond to the values obtained from laboratory measurements, while the N_{struct} parameter was obtained from the inversion of PROSPECT-PRO using the spectral information between 1700 and 2400 nm. Since the independent datasets (Figure 8c and f) do not contain protein and *CBC* content measurements, the resulting statistical indicators are based on values of protein and *CBC* contents obtained from the model inversion using the optimal spectral features identified with SFS. Additionally, the PARACOU dataset was excluded from this analysis, as no measurements of *EWT* were available.

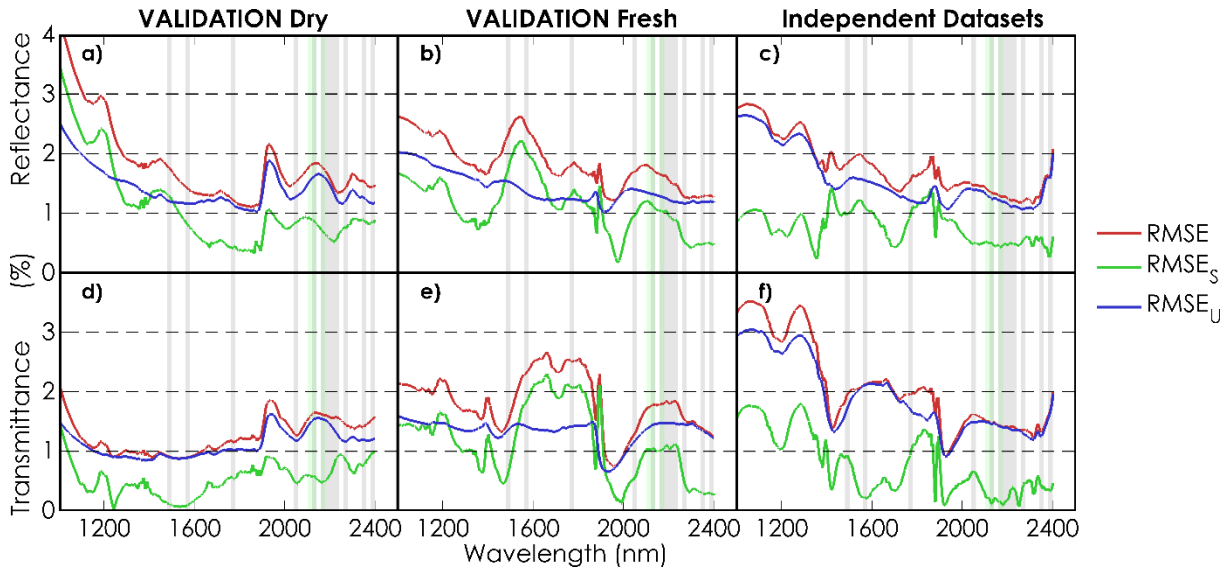

Figure 8. $RMSE$, $RMSE_S$ and $RMSE_U$ between measured and PROSPECT-PRO forward simulated leaf optical properties. The biochemical constituents were measured in laboratory and the N_{struct} parameter was derived by PROSPECT-PRO inversion using the spectral domain from 1700 to 2400 nm. The green and grey areas highlight the respective optimal spectral domains identified by SFS method for estimation of protein and *CBC* contents.

VALIDATION dry samples exhibited an $RMSE$ between 1 and 2% in the SWIR, increasing in the NIR to as high as 4% for reflectance and 2% for transmittance at 1000 nm. The increasing $RMSE$ at shorter wavelengths and also reflectance $RMSE_S$ higher than $RMSE_U$ may be due to the presence of constituents similar to brown pigments appearing after the drying process, which were not accounted for during the simulation, or by residual model inaccuracies at these wavelengths. $RMSE_U$ close to $RMSE$ in the SWIR region confirms acceptable model accuracy and moderate bias, and shows that the specific absorption coefficient of *LMA* as well as proteins and *CBC* are, in general, able to reassemble dry LOPEX data measurements very well.

In case of fresh VALIDATION samples, the $RMSE$ is generally slightly higher and fluctuates between 1 and 3% of reflectance and transmittance intensities. Nevertheless, $RMSE$ between 2000 and 2400 nm, where

583 most of the optimal spectral features for the estimation of protein and *CBC* contents are located, is
584 smaller, between 1 and 2%. Compared to dry samples, the reflectance $RMSE$ between 1000 and 1200 nm
585 is lower, which supports our interpretation of increased NIR $RMSE$ in dry samples due to the unaccounted
586 presence of absorbing constituents similar to brown pigment such as products from decay pigments
587 (Proctor et al., 2017). The higher $RMSE$ combined with corresponding high $RMSE_S$ between c. 1500 and
588 1800 nm suggest that the contribution of water absorption introduces a certain bias in the simulated fresh
589 leaf optical properties.
590 The $RMSE$ lower than 2.1% was found for both reflectance and transmittance of the independent
591 datasets for wavelengths > 1500 nm, which is slightly lower than the results obtained for fresh
592 VALIDATION samples. Relatively different and decoupled dynamics between $RMSE$ and $RMSE_S$ suggest
593 that the contribution of water absorption does not introduce the same bias as observed for fresh
594 VALIDATION samples. The high $RMSE$ and $RMSE_U$ for both reflectance and transmittance in NIR were
595 not observed for fresh VALIDATION samples, and they do not reassemble by shape the increases observed
596 for dry VALIDATION samples. This may be caused by discrepancies in the protocol for the measurement
597 of the leaf optical properties among datasets. The same analysis performed on the complementary
598 simulations from PROSPECT-D showed very similar results (results not shown).
599
600 h. Estimation of C:N from *CBC*:Proteins ratio retrieved from PROSPECT-PRO inversion
601 The correlation analysis displayed in Figure 1 shows that constituents of *CBC*, such as cellulose and lignin,
602 are strongly correlated with leaf C content. We applied the linear model fitted between *CBC*:Proteins and
603 C:N ratio of the fresh samples in the CALIBRATION dataset (Eq. (14)) to the *CBC*:Proteins ratio PROSPECT-
604 PRO estimates for the VALIDATION datasets:
605

$$C{:}N \ = 2.167 \times CBC{:}Proteins + 1.565. \quad (14)$$

The results show that the C:N ratio was derived from the *CBC* and proteins PROSPECT-PRO estimates with a *NRMSE* of 28.1% for dry samples, and an exceptionally low *NRMSE* of 15.7% (R^2 of 0.87) for fresh samples (Figure 9). The poorer performances for dry samples was strongly driven by a single sample of dry chestnut (*Castanea sativa*) leaf, and highlights the necessity of further independent verification of the C:N predictive regression models.

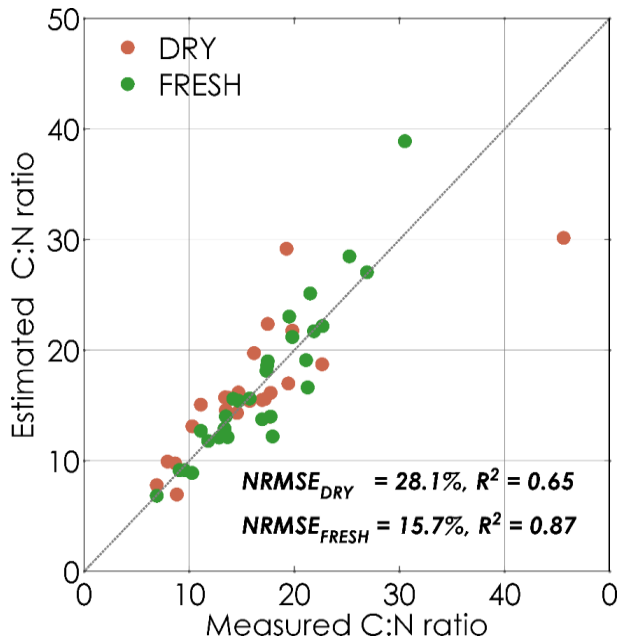

Figure 9. Comparison between the C:N ratio measured in laboratory and the same ratio derived from regression (Eq. (14)) established with PROSPECT-PRO estimated protein and *CBC* contents of dry and fresh samples in the VALIDATION dataset.

6. DISCUSSION

a. Limitations of experimental data available for PROSPECT-PRO calibration and validation

Calibration and validation uncertainties are related not only to the physical-empirical model design and mathematical inversion but also to model inputs, i.e. leaf biochemical and optical measurements (Malenovský et al., 2019). Although we used in this study only a single nitrogen-to-protein content conversion multiplicative factor of 4.43, this factor is not constant across all plant species. As reported by Yeoh and Wee (1994), it can range from 3.28 to 5.16, with an average and standard deviation of 4.43±0.40. This means that the protein content used for calibration and validation of PROSPECT-PRO contains an associated uncertainty that is proportional to the unaccounted variability of this conversion factor. This may also explain the moderately higher uncertainty observed in protein estimates when compared to *LMA* and *CBC* retrievals. Despite this, our results show that the specific absorption coefficients for in vivo proteins are consistent with absorption features derived from dried and ground leaves reported in literature (Curran, 1989). In addition, the protein content estimated through model inversion remained consistent and accurate.

Our study only includes one dataset with measured protein and *CBC* content. Therefore, the validation is performed on a limited number of samples (n=26 dry and 33 fresh). As such, the errors and uncertainties reported might be strongly affected by just few discrepancies in this low number of samples. As reported in Section 3.a, the presence or absence of a single sample in the calibration dataset significantly impacted the calibration process and the subsequent capability of the model to properly simulate leaf optical properties and to estimate leaf constituents. In the same way, the presence of just few validation samples showing a strong error may lead to difficulties for the statistical interpretation of results obtained from an inversion. In our case, we encountered a lower accuracy for proteins estimation on dry samples. It was caused by a single sample, for which either spectral or biochemical measurement error may have occurred. Additional datasets from various ecosystem types and growing conditions are, therefore, required to test further limitations of PROSPECT inversions, especially for high contents of *EWT* and *LMA*, as identified for a limited set of samples in this study. Finally, more public datasets containing reliable

VNIR and SWIR leaf optical properties and corresponding comprehensive and robust laboratory measurements of leaf biochemical constituents are strongly needed. They would also allow us to explore a potential differentiation and inclusion of new ecophysiologically important constituents.

b. Interpretation of GSA

Since N_{struct} has no effect on the leaf absorptance (Figure 3), the GSA of the model biochemical input parameters to leaf absorbance can be used to identify the most dynamic absorption regions of leaf constituents. However, it is important to mention that a strong influence of a given constituent does not mean removal of contributions from other constituents with a lower impact. Although *EWT* is the main driver of SWIR absorptance by fresh leaves, the impact of *CBC* is also significant and in case of dry leaves even dominant. Therefore, with the prerequisite of a clean spectral acquisition with a high spectral sampling and resolution, an appropriately parameterized model inversion procedure using selected optimal spectral features can be successful. Hereby, for the model inversion we took advantage of the spectral dynamics of absorbing constituents depicted in Figure 2 and Figure 3. There is a very pronounced increase of protein absorption within the 2100-2200 nm domain, while water absorption is decreasing, and *CBC* absorption reaches first a maximum peak and then starts to decrease. This unique contrasting spectral dynamic, i.e. a change from the local minimum to the local maximum of proteins in contrast to the other two main absorbers showing moderate changes in absorption, explains the high accuracy we achieved when estimating protein content from the identified optimal spectral domain. On the other hand, absorption of *EWT* and proteins both decrease in the spectral region around 2000 nm, and *CBC* and proteins both decrease beyond 2200 nm. These correlated behaviors negatively impact their retrieval accuracy through a model inversion.

c. Identification of the optimal spectral features

664 The optimal spectral features for estimation of *CBC* and proteins were defined based on the VALIDATION
665 dataset only. The analysis showed that the optimal spectral features selected across all tested datasets
666 were not the same, resulting in an increased *NRMSE* for *LMA* estimates of some experimental datasets
667 (Figure 7). Féret et al. (2019) reached a similar conclusion when identifying specific optimal subdomains
668 for estimation of *EWT* and *LMA*. While Féret et al. (2019) could eventually identify the optimal spectral
669 domain by combining all available datasets, the lack of protein and *CBC* content measurements did not
670 allow us to find the most optimal spectral features across the independent datasets in Table 2.
671 Finally, better performances obtained for the estimation of both proteins and *CBC* from fresh leaf
672 measurements are somewhat in a disagreement with the existing literature on this topic (Fourty et al.,
673 1996; Jacquemoud et al., 1996; Wang et al., 2015). This outcome can be explained by the new spectral
674 feature selection procedure applied in our study. The results in Figure 5 illustrate why systematically lower
675 performances were reported when attempting to estimate *LMA* and proteins by PROSPECT inversion
676 using a full contiguous spectral information containing spectral regions strongly confounded by water
677 absorption. They also provide the evidence that water absorption does not significantly interfere with the
678 spectral information selected as optimal for retrieval of protein and *CBC* contents in dry but also fresh leaf
679 samples in this study. Therefore, the PROSPECT-PRO protein and *CBC* estimation errors do not originate
680 from the model physical and spectral limitations but from the design of the inversion procedure and from
681 associated criterions of the minimization functions.
682
683 d. Complementarity of chlorophyll and protein estimates as proxies for nitrogen
684 content
685 Despite the known limitations of using chlorophyll *a*+*b* content as a proxy of N in remote sensing
686 monitoring applications, it has proved to be relatively successful in a certain number of cases (Baret et al.,
687 2007). The main advantage of estimating chlorophyll over protein content is its strong optical signal in the

VNIR (especially red-edge) domain, allowing for its accurate RS estimates even at the canopy level (e.g., Malenovský et al., 2013). In contrast, the SWIR domain, which is required to estimate protein content, but is characterized by lower solar energy flux and lower SNR (Guanter et al., 2015). Therefore, even if being physiologically more robust over a broader range of conditions and vegetation types, the estimation of N from protein content may be associated with a significantly higher uncertainty originating from a weaker SNR of the spectroscopic measurements. The enhanced capacity of PROSPECT-PRO to monitor vegetation C:N ratio and its seasonal changes through the separation of protein and *CBC* contents may prove useful, if systematically and rationally complemented by a RS chlorophyll monitoring.

e. Potential application for a canopy scale ecosystem nitrogen mapping

Physical RT modeling is a key component in revealing the underlying relations between quantitative vegetation properties and information encoded in RS optical data. In this study, we successfully calibrated and validated a new PROSPECT-PRO model that separates nitrogen-based protein constituents from other carbon-based constituents, i.e. cellulose, lignin, hemicellulose, sugars and starch. Unlike previous attempts, which either resulted in a poor protein content estimation performance (Jacquemoud et al., 1996) or suggested a limited accuracy of *LMA* predictions (Wang et al., 2015), the indirect estimation of *LMA* with PROSPECT-PRO (i.e., the sum of protein and *CBC* contents) was found to be fully comparable with the direct estimation of *LMA* using PROSPECT-D. The performances in forward simulations of leaf optical properties were also very similar for both model versions. Similar to Féret et al. (2019), an accurate estimation of *LMA* and its two components required selection of the appropriate spectral domains. The relatively narrow spectral domain identified as optimal for the retrieval of proteins (2100-2139 and 2160-2179 nm) must be considered in the future operational applications for nitrogen or *LMA* monitoring using field and air-/space-borne imaging spectroscopy. This may be incorporated by applying appropriate weights for the different spectral domains that optimize the sensitivity of retrieval algorithms to the

constituents of interest. The unsuitable spectral wavelengths can be identified and removed by hybrid band selection methods, feature extraction or band weighting procedures (Fassnacht et al., 2014; Feilhauer et al., 2015; Verrelst et al., 2015).

Results of this study offer a new opportunity for operational RS monitoring and consequent management of nitrogen in agricultural and natural ecosystems. However, applicability of PROSPECT-PRO for such a monitoring system is strongly dependent on scalability of the simulated leaf SWIR spectral signatures up to spatially and spectrally heterogeneous canopies. The potential of transferring a PROSPECT-PRO-based N estimating method into an operational application in the field is yet to be investigated. Although proximal remote sensing of small homogeneous canopies, based for instance on the PROCOSINE model (Jay et al., 2016; Morel et al., 2018), could be considered as an intermediate step, a certain number of challenges must be addressed first. These include the capacity to perform sufficiently accurate outdoor canopy measurements, a suppression of the vegetation canopy reflectance angular anisotropy (including spectral effects of background surfaces and leaf orientation), and an ability to achieve a sufficiently high signal-to-noise ratio in the SWIR domain. More generally, analyzing canopy reflectance requires accounting for multiple confounding factors, such as the structural properties of the canopy (i.e., leaf area index, leaf angle distribution and foliage clumping), spectral properties of soil, understory and atmosphere, and the sun-object-observer geometry at the time of data acquisition.

Our results indicate the importance of narrow SWIR domains, which will remain to be important also at the canopy level. Current multispectral spaceborne data (e.g., Landsat 8/9 and Sentinel-2 images) do not comply with the narrowband SWIR spectral requirements that we identified, and further investigations are necessary to conclude on feasibility and limitations of its potential use for N mapping using PROSPECT-PRO. An increasing number of available space-borne imaging spectroscopy data (Berger et al., 2020b) is bringing new opportunities in this field. Several satellite platforms are already operational or close to launch, e.g. PRISMA (Loizzo et al., 2019), Gaofen-5 (Liu et al., 2019), or EnMap (Guanter et al., 2015), and

few more candidate missions are in preparation, such as the Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) (Nieke and Rast, 2018) or NASA's EMIT (Green et al., 2019) and Surface Biology and Geology (SBG) missions (Committee on the Decadal Survey for Earth Science and Applications from Space et al., 2018; Hochberg et al., 2015). Data provided by these instruments holds a strong prospect for N monitoring. Yet, preparatory studies will be necessary to analyze the potential of PROSPECT-PRO in simulating sufficiently accurate imaging spectroscopy data of canopies when being coupled with canopy RT models, for instance with SAIL (Berger et al., 2018; Jacquemoud et al., 2009; Verhoef et al., 2007), SCOPE (van der Tol et al., 2009) , INFORM (Schlerf and Atzberger, 2006) or DART (Gastellu-Etchegorry et al., 2017, 2015). Berger et al. (2020a) studied the potential of PROSPECT-PRO coupled with SAIL for the estimation of crop nitrogen based on airborne imaging spectroscopy. They performed a sensitivity analysis identifying the most relevant spectral bands for this task and concluded on the importance of SWIR bands at 2124 and 2234 nm. Their most optimal spectral bands selected in the NIR and the first part of the SWIR spectra suggest that conclusions of this leaf scale study may differ from conclusions at the canopy scale. Accounting for the canopy reflectance confounding factors may need an additional spectral information, coming from different spectral domains than those required at the leaf scale.

7. CONCLUSIONS

This study introduces PROSPECT-PRO, a new version of the PROSPECT leaf RT model, capable of differentiating proteins from other carbon-based constituents as two independent components of *LMA*. The calibration of PROSPECT-PRO was based on the assumption that proteins and *CBC* are the two main spectrally active constituents of the leaf dry matter. We demonstrated that PROSPECT-PRO performs similarly in estimating protein and *CBC* content of both fresh and dry leaves, a marked improvement over

previous attempts. Errors computed between measured and simulated leaf optical properties were relatively low for both types of leaves.

Our results, based on leaf optical properties with the 1 nm spectral sampling, revealed that the optimal estimation of leaf protein content at the leaf scale is obtained when using two narrow spectral domains between 2100 and 2139 nm, and between 2160 and 2179 nm. The estimation of protein content, assessed by *NRMSE*, was found to be slightly less accurate than the estimation of *CBC* content or total *LMA*. Additionally, the C:N ratio was successfully estimated from the *CBC*:Proteins ratio retrieved by PROSPECT-PRO inversion. Despite these achievements, further investigations, that would be conducted on independent leaf-scale measurements of leaf optical properties, proteins, nitrogen and *LMA*, are still needed. Canopy-scale studies are also required to test the potential of this new model for operational airborne and space-borne applications. The capability of current satellite multispectral instruments (e.g., Sentinel-2 and Landsat-8/9) to estimate vegetation protein and *CBC* contents needs to be investigated in light of our findings. However, such estimations may remain extremely challenging, considering the coarse resolution and limited number of spectral bands of these instruments in the SWIR region. Spaceborne imaging spectroscopy missions with a higher SWIR spectral sampling may be of the critical importance for a future operational nitrogen-containing protein monitoring of agricultural and natural environments.

8. ACKNOWLEDGMENTS

Authors would like to thank Philip Townsend (University of Wisconsin – Madison), Jochem Verrelst (University of Valencia) and also four anonymous reviewers for their comments and advices improving scientific quality and readability of this manuscript. J.-B. Féret and F. de Boissieu acknowledge financial support from Agence Nationale de la Recherche (BioCop project—ANR-17-CE32-0001) and TOSCA program grant of the French Space Agency (CNES) (HyperTropik/HyperBIO project). Contribution of Z. Malenovský was supported by the Australian Research Council Future Fellowship 'Bridging Scales in

Remote Sensing of Vegetation Stress' (FT160100477). Katja Berger is funded within the EnMAP scientific preparation program under the DLR Space Administration with resources from the German Federal Ministry of Economic Affairs and Energy, grant number 50EE1923. Further, the research was supported by the Action CA17134 SENSECO (Optical synergies for spatiotemporal sensing of scalable ecophysiological traits) funded by COST (European Cooperation in Science and Technology, www.cost.eu).

9. REFERENCES

Albornoz, F., 2016. Crop responses to nitrogen overfertilization: A review. Sci. Hortic. 205, 79–83. https://doi.org/10.1016/j.scienta.2016.04.026

Allen, W.A., Gausman, H.W., Richardson, A.J., 1970. Mean effective optical constants of cotton leaves. J. Opt. Soc. Am. 60, 542–547. https://doi.org/10.1364/JOSA.60.000542

Allen, W.A., Gausman, H.W., Richardson, A.J., Thomas, J.R., 1969. Interaction of isotropic light with a compact plant leaf. J. Opt. Soc. Am. 59, 1376–1379. https://doi.org/10.1364/JOSA.59.001376

Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7, 269–276. https://doi.org/10.1890/070152

Baret, F., Buis, S., 2008. Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems, in: Liang, S. (Ed.), Advances in Land Remote Sensing. Springer Netherlands, Dordrecht, pp. 173–201.

Baret, F., Houles, V., Guerif, M., 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. J. Exp. Bot. 58, 869–880. https://doi.org/10.1093/jxb/erl231

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., 2018. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. Remote Sens. 10, 85. https://doi.org/10.3390/rs10010085

Berger, K., Verrelst, J., Féret, J.-B., Hank, T., Wocher, M., Mauser, W., Camps-Valls, G., 2020a. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. Int. J. Appl. Earth Obs. Geoinformation 92, 102174. https://doi.org/10.1016/j.jag.2020.102174

Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., Hank, T., 2020b. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens. Environ. 242, 111758. https://doi.org/10.1016/j.rse.2020.111758

Borchers, H.W., 2019. pracma: Practical Numerical Math Functions.

Boren, E.J., Boschetti, L., Johnson, D.M., 2019. Characterizing the Variability of the Structure Parameter in the PROSPECT Leaf Optical Properties Model. Remote Sens. 11, 1236. https://doi.org/10.3390/rs11101236

Botha, E.J., Zebarth, B.J., Leblon, B., 2006. Non-destructive estimation of potato leaf chlorophyll and protein contents from hyperspectral measurements using the PROSPECT radiative transfer model. Can. J. Plant Sci. 86, 279–291. https://doi.org/10.4141/P05-017

Bradstreet, R.B., 1954. Kjeldahl Method for Organic Nitrogen. Anal. Chem. 26, 185–187. https://doi.org/10.1021/ac60085a028

Brown, B., Westcott, M., Christensen, N., Pan, B., Stark, J., 2005. Nitrogen management for hard wheat protein enhancement (No. PNW 578). Washington State Univ. Coop. Ext.

Cannavó, F., 2012. Sensitivity analysis for volcanic source modeling quality assessment and model selection. Comput. Geosci. 44, 52–59. https://doi.org/10.1016/j.cageo.2012.03.008

Clevers, J.G., Kooistra, L., 2012. Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. Sel. Top. Appl. Earth Obs. Remote Sens. IEEE J. Of 5, 574–583.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. Int. J. Appl. Earth Obs. Geoinformation 23, 344–351. https://doi.org/10.1016/j.jag.2012.10.008

Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., Panigada, C., 2008. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. Remote Sens. Environ. 112, 1820–1834. https://doi.org/10.1016/j.rse.2007.09.005

Committee on the Decadal Survey for Earth Science and Applications from Space, Space Studies Board, Division on Engineering and Physical Sciences, National Academies of Sciences, Engineering, and Medicine, 2018. Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space. National Academies Press, Washington, D.C. https://doi.org/10.17226/24938

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

Davidson, E.A., David, M.B., Galloway, J.N., Goodale, C.L., Haeuber, R.A., Harrison, J.A., Howarth, R.W., Jaynes, D.B., Lowrance, R.R., Thomas, N.B., Peel, J.L., Pinder, R.W., Porter, E., Snyder, C.S., Townsend, A.R., Ward, M.H., 2011. Excess Nitrogen in the U.S. Environment: Trends, Risks, and Solutions.

Elvidge, C.D., 1990. Visible and near infrared reflectance characteristics of dry plant materials. Int. J. Remote Sens. 11, 1775–1795. https://doi.org/10.1080/01431169008955129

Evans, J.R., 1989. Photosynthesis and nitrogen relationships in leaves of C3 plants. Oecologia 78, 9–19. https://doi.org/10.1007/BF00377192

Fassnacht, F.E., Neumann, C., Forster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., Koch, B., 2014. Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 2547–2561. https://doi.org/10.1109/JSTARS.2014.2329390

Feilhauer, H., Asner, G.P., Martin, R.E., 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. Remote Sens. Environ. 164, 57–65. https://doi.org/10.1016/j.rse.2015.03.033

Féret, J.-B., Boissieu, F. de, 2020. prospect: an R package to run the leaf model PROSPECT in forward and inverse mode.

Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004

Féret, J.-B., le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J.C., Ponzoni, F.J., Solanki, T., de Boissieu, F., Chave, J., Nouvellon, Y., Porcar-Castell, A., Proisy, C., Soudani, K., Gastellu-Etchegorry, J.-P., Lefèvre-Fonollosa, M.-J., 2019. Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations

of physical modeling and machine learning. Remote Sens. Environ. 231, 110959. https://doi.org/10.1016/j.rse.2018.11.002

Fletcher, R., 2000. Practical Methods of Optimization: Fletcher/Practical Methods of Optimization. John Wiley & Sons, Ltd, Chichester, West Sussex England. https://doi.org/10.1002/9781118723203

Fourty, Th., Baret, F., Jacquemoud, S., Schmuck, G., Verdebout, J., 1996. Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. Remote Sens. Environ. 56, 104–117. https://doi.org/10.1016/0034-4257(95)00234-0

Gastellu-Etchegorry, J.-P., Lauret, N., Yin, T., Landier, L., Kallel, A., Malenovský, Z., Bitar, A.A., Aval, J., Benhmida, S., Qi, J., Medjdoub, G., Guilleux, J., Chavanon, E., Cook, B., Morton, D., Chrysoulakis, N., Mitraka, Z., 2017. DART: Recent Advances in Remote Sensing Data Modeling With Atmosphere, Polarization, and Chlorophyll Fluorescence. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 2640–2649. https://doi.org/10.1109/JSTARS.2017.2685528

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Cajgfinger, T., Gregoire, T., Grau, E., Feret, J.-B., Lopes, M., Guilleux, J., Dedieu, G., Malenovský, Z., Cook, B.D., Morton, D., Rubio, J., Durrieu, S., Cazanave, G., Martin, E., Ristorcelli, T., 2015. Discrete Anisotropic Radiative Transfer (DART 5) for Modeling Airborne and Satellite Spectroradiometer and LIDAR Acquisitions of Natural and Urban Landscapes. Remote Sens. 7, 1667–1701. https://doi.org/10.3390/rs70201667

Green, R., Mahowald, N., Thompson, D., Clark, R., Ehlmann, B., Ginoux, P., Kalashnikova, O., Miller, R., Okin, G., Painter, T., Perez, C., Realmuto, V., Swayze, G., Middleton, E., Guanter, L., Ben Dor, E., 2019. The Earth Surface Mineral Dust Source Investigation Planned for the International Space Station, in: Geophysical Research Abstracts. Presented at the European Geosciences Union, Vienna (Austria).

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451, 293–296. https://doi.org/10.1038/nature06592

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sens. 7, 8830–8857. https://doi.org/10.3390/rs70708830

Hank, T.B., Berger, K., Bach, H., Clevers, J.G.P.W., Gitelson, A., Zarco-Tejada, P., Mauser, W., 2019. Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. Surv. Geophys. 40, 515–551. https://doi.org/10.1007/s10712-018-9492-0

Himmelsbach, D.S., Boer, H., Akin, D.E., Barton, F.E.I., 1988. Solid-state 13C NMR, FTIR, and NIRS spectroscopic studies of ruminant silage digestion, in: Analytical Applications of Spectroscopy. Royal Society of Chemistry, London.

Hochberg, E.J., Roberts, D.A., Dennison, P.E., Hulley, G.C., 2015. Special issue on the Hyperspectral Infrared Imager (HyspIRI): Emerging science in terrestrial and aquatic ecology, radiation balance and hazards. Remote Sens. Environ. 167, 1–5. https://doi.org/10.1016/j.rse.2015.06.011

Homolova, L., Malenovskỳ, Z., Clevers, J.G., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. Ecol. Complex. 15, 1–16.

Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, A., Schmuck, G., 1994. Leaf Optical Properties Experiment 93 (LOPEX93) (European Commission No. EUR 16095 EN). Joint Research Centre, Institute for Remote Sensing Applications.

Jacquemoud, S., Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. Remote Sens. Environ. 34, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z

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 Sens. Environ. 56, 194–202. https://doi.org/10.1016/0034-4257(95)00238-3

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66. https://doi.org/doi:10.1016/j.rse.2008.01.026

Jay, S., Bendoula, R., Hadoux, X., Féret, J.-B., Gorretta, N., 2016. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. Remote Sens. Environ. 177, 220–236. https://doi.org/10.1016/j.rse.2016.02.029

Kant, S., Bi, Y.-M., Rothstein, S.J., 2011. Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J. Exp. Bot. 62, 1499–1509. https://doi.org/10.1093/jxb/erq297

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 Sens. Environ. 113, S78–S91.

Kudo, M., Sklansky, J., 2000. Comparison of algorithms that select features for pattern classifiers. Pattern Recognit. 33, 25–41. https://doi.org/10.1016/S0031-3203(99)00041-2

Liu, T., Ren, T., White, P.J., Cong, R., Lu, J., 2018. Storage nitrogen co-ordinates leaf expansion and photosynthetic capacity in winter oilseed rape. J. Exp. Bot. 69, 2995–3007. https://doi.org/10.1093/jxb/ery134

Liu, Y.-N., Zhang, J., Zhang, Y., Sun, W.-W., Jiao, L.-L., Sun, D.-X., Hu, X.-N., Ye, X., Li, Y.-D., Liu, S.-F., Cao, K.-Q., Chai, M.-Y., Zhou, W.-Y.-N., 2019. The Advanced Hyperspectral Imager: Aboard China's GaoFen-5 Satellite. IEEE Geosci. Remote Sens. Mag. 7, 23–32. https://doi.org/10.1109/MGRS.2019.2927687

Loizzo, R., Daraio, M., Guarini, R., Longo, F., Lorusso, R., Dini, L., Lopinto, E., 2019. Prisma Mission Status and Perspective, in: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Yokohama, Japan, pp. 4503–4506. https://doi.org/10.1109/IGARSS.2019.8899272

Ma, S., He, F., Tian, D., Zou, D., Yan, Z., Yang, Y., Zhou, T., Huang, K., Shen, H., Fang, J., 2018. Variations and determinants of carbon content in plants: a global synthesis. Biogeosciences 15, 693–702. https://doi.org/10.5194/bg-15-693-2018

Malagoli, P., 2005. Dynamics of Nitrogen Uptake and Mobilization in Field-grown Winter Oilseed Rape (Brassica napus) from Stem Extension to Harvest: I. Global N Flows between Vegetative and Reproductive Tissues in Relation to Leaf Fall and their Residual N. Ann. Bot. 95, 853–861. https://doi.org/10.1093/aob/mci091

Malenovský, Z., Albrechtová, J., Lhotáková, Z., Zurita-Milla, R., Clevers, J.G.P.W., Schaepman, M.E., Cudlín, P., 2006. Applicability of the PROSPECT model for Norway spruce needles. Int. J. Remote Sens. 27, 5315–5340. https://doi.org/10.1080/01431160600762990

Malenovský, Z., Homolová, L., Lukeš, P., Buddenbaum, H., Verrelst, J., Alonso, L., Schaepman, M.E., Lauret, N., Gastellu-Etchegorry, J.-P., 2019. Variability and Uncertainty Challenges in Scaling Imaging Spectroscopy Retrievals and Validations from Leaves Up to Vegetation Canopies. Surv. Geophys. 40, 631–656. https://doi.org/10.1007/s10712-019-09534-y

Malenovskỳ, Z., Homolová, L., Zurita-Milla, R., Lukeš, P., Kaplan, V., Hanuš, J., Gastellu-Etchegorry, J.-P., Schaepman, M.E., 2013. Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. Remote Sens. Environ. 131, 85–102.

Marcano-Cedeno, A., Quintanilla-Dominguez, J., Cortina-Januchs, M.G., Andina, D., 2010. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity

Neural Network, in: IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society. Presented at the IECON 2010 - 36th Annual Conference of IEEE Industrial Electronics, IEEE, Glendale, AZ, USA, pp. 2845–2850. https://doi.org/10.1109/IECON.2010.5675075

Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., Suzuki, A., 2010. Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. Ann. Bot. 105, 1141–1157. https://doi.org/10.1093/aob/mcq028

Morel, J., Jay, S., Féret, J.-B., Bakache, A., Bendoula, R., Carreel, F., Gorretta, N., 2018. Exploring the potential of PROCOSINE and close-range hyperspectral imaging to study the effects of fungal diseases on leaf physiology. Sci. Rep. 8. https://doi.org/10.1038/s41598-018-34429-0

Nieke, J., Rast, M., 2018. Towards the Copernicus Hyperspectral Imaging Mission For The Environment (CHIME), in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Valencia, pp. 157–159. https://doi.org/10.1109/IGARSS.2018.8518384

Paul, M.J., Driscoll, S.P., 1997. Sugar repression of photosynthesis: the role of carbohydrates in signalling nitrogen deficiency through source:sink imbalance. Plant Cell Environ. 20, 110–116. https://doi.org/10.1046/j.1365-3040.1997.d01-17.x

Pedrós, R., Goulas, Y., Jacquemoud, S., Louis, J., Moya, I., 2010. FluorMODleaf: A new leaf fluorescence emission model based on the PROSPECT model. Remote Sens. Environ. 114, 155–167.

Reich, P.B., Hobbie, S.E., Lee, T., Ellsworth, D.S., West, J.B., Tilman, D., Knops, J.M.H., Naeem, S., Trost, J., 2006. Nitrogen limitation constrains sustainability of ecosystem response to CO2. Nature 440, 922–925. https://doi.org/10.1038/nature04486

Sáez-Plaza, P., Michałowski, T., Navas, M.J., Asuero, A.G., Wybraniec, S., 2013. An Overview of the Kjeldahl Method of Nitrogen Determination. Part I. Early History, Chemistry of the Procedure, and Titrimetric Finish. Crit. Rev. Anal. Chem. 43, 178–223. https://doi.org/10.1080/10408347.2012.751786

Schaepman-Strub, G., Schaepman, M.E., Painter, T.H., Dangel, S., Martonchik, J.V., 2006. Reflectance quantities in optical remote sensing—definitions and case studies. Remote Sens. Environ. 103, 27–42. https://doi.org/10.1016/j.rse.2006.03.002

Schlerf, M., Atzberger, C., 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. Remote Sens. Environ. 100, 281–294. https://doi.org/10.1016/j.rse.2005.10.006

Serbin, S.P., Wu, J., Ely, K.S., Kruger, E.L., Townsend, P.A., Meng, R., Wolfe, B.T., Chlus, A., Wang, Z., Rogers, A., 2019. From the Arctic to the tropics: multibiome prediction of leaf mass per area using leaf reflectance. New Phytol. 224, 1557–1568. https://doi.org/10.1111/nph.16123

Sharwood, R.E., 2017. Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. New Phytol. 213, 494–510. https://doi.org/10.1111/nph.14351

Ustin, S.L., Jacquemoud, S., 2020. How the Optical Properties of Leaves Modify the Absorption and Scattering of Energy and Enhance Leaf Functionality, in: Cavender-Bares, J., Gamon, J.A., Townsend, P.A. (Eds.), Remote Sensing of Plant Biodiversity. Springer International Publishing, Cham, pp. 349–384. https://doi.org/10.1007/978-3-030-33157-3_14

van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 3109–3129. https://doi.org/10.5194/bg-6-3109-2009

van der Tol, C., Vilfan, N., Dauwe, D., Cendrero-Mateo, M.P., Yang, P., 2019. The scattering and re-absorption of red and near-infrared chlorophyll fluorescence in the models Fluspect and SCOPE. Remote Sens. Environ. 232, 111292. https://doi.org/10.1016/j.rse.2019.111292

Verdebout, J., Jacquemoud, S., Andreoli, G., Hosgood, B., Pedrini, A., Schmuck, G., 1995. Analysis of imaging spectrometer data to evaluate the biochemical content of vegetation, based on the results of a laboratory experiment, in: Mougin, E., Ranson, K.J., Smith, J.A. (Eds.), . Presented at the Satellite Remote Sensing, Rome, Italy, pp. 63–76. https://doi.org/10.1117/12.200746

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. IEEE Trans. Geosci. Remote Sens. 45, 1808–1822. https://doi.org/10.1109/TGRS.2007.895844

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS J. Photogramm. Remote Sens. 108, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., Moreno, J., 2019a. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surv. Geophys. 40, 589–629. https://doi.org/10.1007/s10712-018-9478-y

Verrelst, J., Vicent, J., Rivera-Caicedo, J.P., Lumbierres, M., Morcillo-Pallarés, P., Moreno, J., 2019b. Global Sensitivity Analysis of Leaf-Canopy-Atmosphere RTMs: Implications for Biophysical Variables Retrieval from Top-of-Atmosphere Radiance Data. Remote Sens. 11, 1923. https://doi.org/10.3390/rs11161923

Vilfan, N., Van der Tol, C., Yang, P., Wyber, R., Malenovský, Z., Robinson, S.A., Verhoef, W., 2018. Extending Fluspect to simulate xanthophyll driven leaf reflectance dynamics. Remote Sens. Environ. 211, 345–356. https://doi.org/10.1016/j.rse.2018.04.012

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H., Tilman, D.G., 1997. HUMAN ALTERATION OF THE GLOBAL NITROGEN CYCLE: SOURCES AND CONSEQUENCES. Ecol. Appl. 7, 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Vos, J., Bom, M., 1993. Hand-held chlorophyll meter: a promising tool to assess the nitrogen status of potato foliage. Potato Res. 36, 301–308. https://doi.org/10.1007/BF02361796

Wang, Z., Skidmore, A.K., Wang, T., Darvishzadeh, R., Hearne, J., 2015. Applicability of the PROSPECT model for estimating protein and cellulose+lignin in fresh leaves. Remote Sens. Environ. 168, 205–218. https://doi.org/10.1016/j.rse.2015.07.007

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. 90, 8995. https://doi.org/10.1029/JC090iC05p08995

Wingler, A., Purdy, S., MacLean, J.A., Pourtau, N., 2006. The role of sugars in integrating environmental signals during the regulation of leaf senescence. J. Exp. Bot. 57, 391–399. https://doi.org/10.1093/jxb/eri279

Yeoh, H.-H., Wee, Y.-C., 1994. Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species. Food Chem. 49, 245–250. https://doi.org/10.1016/0308-8146(94)90167-8

Yoder, B.J., Pettigrew-Crosby, R.E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. Remote Sens. Environ. 53, 199–211. https://doi.org/10.1016/0034-4257(95)00135-N

Zheng, Z.-L., 2009. Carbon and nitrogen nutrient balance signaling in plants. Plant Signal. Behav. 4, 584–591. https://doi.org/10.4161/psb.4.7.8540

Zhou, G., Xu, S., Ciais, P., Manzoni, S., Fang, J., Yu, G., Tang, X., Zhou, P., Wang, W., Yan, J., Wang, G., Ma, K., Li, S., Du, S., Han, S., Ma, Y., Zhang, D., Liu, J., Liu, S., Chu, G., Zhang, Q., Li, Y., Huang, W., Ren, H., Lu, X., Chen, X., 2019. Climate and litter C/N ratio constrain soil organic carbon accumulation. Natl. Sci. Rev. 6, 746–757. https://doi.org/10.1093/nsr/nwz045

1060