Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: potential and limitations of physical modeling and machine learning.


To cite this version:


HAL Id: hal-02939160
https://hal.inrae.fr/hal-02939160
Submitted on 15 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: potential and limitations of physical modeling and machine learning.


1 TETIS, Irstea, AgroParisTech, CIRAD, CNRS, Université Montpellier, Montpellier, France
2 CIRAD, UMR ECO&SOLS, Montpellier, France
3 Eco&Sols, Univ Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, Montpellier, France.
4 Interdisciplinary Center of Energy Planning (NIPE), UNICAMP, 13083-896 Campinas, Brazil
5 Aix Marseille Univ, CNRS, Centrale Marseille, Institut Fresnel, F-13013 Marseille, France
6 Ecologie Systematique Evolution, University of Paris-Sud, CNRS, AgroParisTech, Université Paris
7 Saclay, F-91400 Orsay, France
8 ITAP, Irstea, Montpellier SupAgro, Université Montpellier, Montpellier, France
9 School of Agricultural Engineering - FEAGRI, University of Campinas, São Paulo, Brazil
10 Instituto Nacional de Pesquisas Espaciais, Sao Jose dos Campos 12227-010, Brazil
11 Optics of Photosynthesis Laboratory, INAR/Forests, Faculty of Agriculture and Forestry, 00014 University of Helsinki, Finland
12 Laboratoire Evolution et Diversité Biologique UMR 5174, CNRS, Université Paul Sabatier, Toulouse, France
13 University of Sao Paulo, ESALQ/USP, Piracicaba 13418-900, Brazil
14 AMAP, IRD, CIRAD, CNRS, INRA, Univ. Montpellier, Montpellier, France
15 GEOSMIT, French Institute of Pondicherry, Pondicherry, India
16 Centre d’Etudes Spatiales de la Biosphère, Toulouse 31400, France

CNES
Abstract

Leaf mass per area (LMA) and leaf equivalent water thickness (EWT) are key leaf functional traits providing information for many applications including ecosystem functioning modeling and fire risk management. In this paper, we investigate two common conclusions generally made for LMA and EWT estimation based on leaf optical properties in the near-infrared (NIR) and shortwave infrared (SWIR) domains: (1) physically-based approaches estimate EWT accurately and LMA poorly, while (2) statistically-based and machine learning (ML) methods provide accurate estimates of both LMA and EWT.

Using six experimental datasets including broadleaf species samples of more than 150 species collected over tropical, temperate and boreal ecosystems, we compared the performances of a physically-based method (PROSPECT model inversion) and a ML algorithm (support vector machine regressions, SVM) to infer EWT and LMA based on leaf reflectance and transmittance. We assessed several merit functions to invert PROSPECT based on iterative optimization and investigated the spectral domain to be used for optimal estimation of LMA and EWT. We also tested several strategies to select the training samples used by the SVM, in order to investigate the generalization ability of the derived regression models.

We evidenced that using spectral information from 1700 to 2400 nm leads to strong improvement in the estimation of EWT and LMA when performing a PROSPECT inversion, decreasing the LMA and EWT estimation errors by 55% and 33%, respectively. The comparison of various sampling strategies for the training set used with SVM suggests that regression models show limited generalization ability, particularly when the regression model is applied on data fully independent from the training set. Finally, our results demonstrate that, when using an appropriate spectral domain, the PROSPECT inversion outperforms SVM trained with experimental data for the estimation of EWT and LMA. Thus we recommend that estimation of
*LMA* and *EWT* based on leaf optical properties should be physically-based using inversion of reflectance and transmittance measurements on the 1700 to 2400 nm spectral range.

1. **INTRODUCTION**

Global climate change and biodiversity loss strongly impact species and ecosystem functions, which directly influence processes at landscape and regional scales, and disrupts global biogeochemical cycles (Chapin, 2003). These ecosystem functions are tightly connected with species composition and can be partly described and explained using plant traits (Diaz and Cabido, 2001; Eviner and Chapin, 2003). By definition, plant traits correspond to morphological, physiological or phenological features measurable at the individual level, and functional traits are defined as these features impacting individual fitness via their effects on growth, reproduction and/or survival, the three components of individual performance (Violle et al., 2007). Therefore, our understanding of the interactions between climate, human activity and ecosystem functioning strongly depends on our capacity to monitor critical functional traits across space and time (Asner and Martin, 2016).

Leaf mass per area (*LMA*) is defined as the ratio of leaf dry mass (*DW*) to leaf area (*A*):

\[
LMA = \frac{DW}{A} \text{ (mg cm}^{-2}\text{)} 
\]

It is a plant functional trait widely used as an indicator of plant functioning and ecosystem processes. In the leaf economic spectrum theory, the biophysical constraints explain the high coordination between organs properties and available resources: for instance, plants that have high trunk water conductivity generally have high stomatal conductance, low *LMA* and high photosynthetic capacities, developed root system and nutrient uptake, high turnover rate of resource acquisition organs, high growth rates. *LMA* is therefore a very significant trait because it correlates with key
plant functional properties (de la Riva et al., 2016; Oren et al., 1986; Reich et al., 1997), therefore capturing a great proportion of the functional variation in the ecosystem. 

*LMA* is important for the description of plant strategies and photosynthetic capacity over various vegetation types and climates (Asner et al., 2011; Gratani and Varone, 2006; Osnas et al., 2013; Puglielli et al., 2015; Reich et al., 1997, 1998; Weng et al., 2017). It is also a predictor of relative growth rate (Antúnez et al., 2001; Rees et al., 2010) and is usually correlated with mass-based maximum photosynthetic rate (Wright et al. 2004). At broader scales, it is also identified as a critical plant trait for the global monitoring of functional diversity, and for the determination of species fitness in their environment, affecting various ecosystem processes (Poorter et al., 2009; Schimel et al., 2015). Measurement of *LMA* is also relevant for many other applications, such as fire risk assessment (Cornelissen et al., 2017). Finally, *LMA* allows the conversion of traits expressed on an area basis into mass basis and vice versa. This is important since physical models usually express leaf constituent content per surface unit, whereas ecologists and plant physiologists may use constituent content per surface unit or per mass unit (Osnas et al., 2013; Wright et al., 2004).

The second important functional trait discussed in this study is the equivalent water thickness (*EWT*), defined as:

\[
EWT = \frac{FW - DW}{A} \quad (mg.cm^{-2})
\]

with *FW* the leaf fresh mass. *EWT* is the area-weighted moisture content. It is related to a range of physiological and ecosystem processes, including leaf-level tolerance to dehydration, and ecological strategy. Indeed, species with large *EWT* tend to have lower construction costs, and are predominantly fast-growing and pioneer species (Wright et al., 2004).
The ability to accurately estimate both EWT and LMA is also critical for applications such as fire danger assessment: fuel moisture content (FMC, Chuvieco et al., 2002), also referred to as gravimetric water content (GWC, Datt, 1999), is a critical variable affecting fire interactions with fuel (Yebra et al., 2013). The accurate estimation of FMC is usually limited by the uncertainty associated to the estimation of LMA (Riano et al., 2005). Destructive measurements of LMA and EWT are time-consuming and logistically complex in remote environments. Alternative methods based on leaf spectroscopy have showed good performances for the estimation of various constituents (Asner et al., 2011, 2009; Ceccato et al., 2001; Colombo et al., 2008; Feilhauer et al., 2015; Féret et al., 2017; Fourty and Baret, 1998). Two main types of methods have been developed for the estimation of vegetation properties from their optical properties (including leaf chemistry but also canopy biophysical properties): physically-based methods and data-driven methods, also referred to as “radiometric data-driven approaches” and “biophysical variable driven approaches” respectively, by Baret and Buis (2008). In this study, we will only use the terms physically-based methods and data-driven methods in order to avoid confusion.

Physically-based methods are based on radiative transfer models (RTM) providing a mechanistic link between leaf traits and their optical properties. They aim at minimizing the residuals between measured and modeled radiometric data (hence the term “radiometric data-driven approach” by Baret and Buis, 2008). The PROSPECT model (Jacquemoud and Baret, 1990; Féret et al., 2017) is the most widespread model, due to its relative simplicity and computational efficiency combined with excellent modeling performances for a broad range of leaf types. Several retrieval algorithms have been developed to estimate leaf chemistry from their optical properties, taking advantage of physical modeling. These include look-up-table (LUT) methods (Ali et al., 2016) and iterative optimization based on minimization algorithms (Jacquemoud et al., 1996). Physically-based methods do not require calibration data, but they are computationally demanding.
Data-driven methods use a calibration dataset of measured leaf optical properties and traits in order to adjust regression models for the estimation of leaf chemistry (Verrelst et al., 2016). These include regression models derived from spectral indices, one of the most classic approaches (Gitelson et al., 2006; Main et al., 2011). More complex multivariate methods such as partial least square regression (Asner et al., 2011), and machine learning algorithms (ML) are also extensively used in the domain of remote sensing. These include support vector machine (SVM, Cortes and Vapnik, 1995; Drucker et al., 1996), random forest (Breiman, 2001), and artificial neural networks (Hornik et al., 1989). ML algorithms have been extensively used for remote sensing applications during the past decades, most of them at the canopy level when it comes to the estimation of biochemical constituents (Brown et al., 2000; Gualtieri, 2009; Lardeux et al., 2009; le Maire et al., 2011; Schmitter et al., 2017; Stumpf and Kerle, 2011; Zhang et al., 2017), and a limited number of studies focusing on the leaf/needle scale (Conejo et al., 2015; Dawson et al., 1998; le Maire et al., 2004). ML algorithms usually show good performances in terms of prediction ability and high computational efficiency.

The capacity of data-driven approaches to accurately predict leaf chemistry from their optical properties is inherently dependent on the dataset used to train the algorithm and regression model. The experiments performed in this study aim at quantifying this assertion over an extensive experimental dataset. This implies that correct implementation of data-driven methods using experimental data for training requires substantial efforts for the measurement of leaf optical properties and chemical constituents with destructive methods, whereas physical modeling only requires leaf optical properties.

Note that a third type of approach, namely, hybrid methods, could also be mentioned here (Verrelst et al., 2015). Such methods use data-driven algorithms trained with spectral properties simulated with physical models. These methods are particularly developed at the canopy scale, and combine
the advantages of physically-based and data-driven methods: they do not require destructive measurements to build an experimental training dataset, and they are computationally efficient. 

*LMA* and *EWT* both influence leaf optical properties in the near-infrared (NIR) and shortwave infrared (SWIR) domains (Bowyer and Danson, 2004). However, physically-based methods have often been reported to perform poorly for the estimation of *LMA* (Colombo et al., 2008; le Maire et al., 2008; Riano et al., 2005; Wang et al., 2011). Several reasons have been mentioned in the literature, including suboptimal modeling (Qiu et al., 2018), optical data collection (Merzlyak et al., 2004) or inversion (Colombo et al., 2008; Qiu et al., 2018; Riano et al., 2005; Sun et al., 2018; Wang et al., 2011, 2015).

A first reason related to modeling is that the influence of *LMA* on the optical properties modeled by PROSPECT is defined by a single specific absorption coefficient (SAC), although various non-pigment organic materials (cellulose, hemicellulose, lignin, proteins, starch) influence leaf optics individually (Jacquemoud et al., 1996). Therefore, this single SAC assumes that the relative proportion of each of these single constituents is constant among leaves, which may not be the case. Another reason may be due to an imperfect modeling of light propagation within the leaf. From that perspective, Qiu et al. (2018) proposed a refined version of PROSPECT (named PROSPECT-g) including an anisotropic-scattering factor in order to improve the estimation of *LMA*, and developed an iterative inversion procedure specifically dedicated to this model.

Experimental uncertainty should also be considered when discrepancies between measurements and simulations are observed. Indeed, accurately measuring leaf optical properties remains challenging despite the high performances of field and lab spectroradiometers, leading to possible experimental bias which is usually unaccounted for. As an example, Merzlyak et al. (2004) reported the difficulty to accurately measure leaf optical properties in the NIR domain due to incomplete collection of the light leaving the highly scattering tissue. They proposed a correcting factor for
transmittance based on the hypothesis that leaf absorption in the NIR domain is negligible. For these reasons, the relevance of systematically using the full spectral domain (especially the NIR domain) can be questioned. Finally, several authors suggested that classical least-squares inversion based on the use of leaf reflectance and transmittance over the full spectral domain was suboptimal for physically-based estimation of LMA, especially due to the lower influence of LMA on leaf optical properties in the SWIR domain as compared to EWT (Colombo et al., 2008; Riano et al., 2005). More elaborated inversion procedures have thus been proposed to improve LMA estimation. Some of them are based on complex iterative procedures consisting in successively estimating different PROSPECT parameters using unweighted merit functions computed over specific spectral domains (Qiu et al., 2018; Li and Wang, 2011; Wang et al., 2015). When using the full spectral domain from 400 to 2500 nm, Sun et al. (2018) showed that LMA estimation based on PROSPECT inversion and an unweighted merit function was more accurate when using only reflectance or only transmittance instead of reflectance plus transmittance. When using bidirectional reflectance measurements, Li et al. (2018) developed an approach (PROCWT) coupling PROSPECT with continuous wavelet transform in order to suppress surface reflectance effects. PROCWT was shown to perform better than PROSPECT and a simplified version of PROCOSINE (Jay et al., 2016) for the estimation of LMA. All of these studies demonstrate the complexity of a direct estimation of LMA from leaf optical properties using physically-based methods, and the difficulty to clearly identify the origin of current limitations. In the case of data-driven methods, the estimation of LMA has seldom been investigated comprehensively: training and test data are usually collected following a unique protocol specific to a unique set of equipment and by the same team of operators. This means that possible experimental biases due to protocol, equipment and/or operators may be embedded into the
resulting regression model, leading to poor generalization ability when applied to independent datasets collected under different conditions or with different equipment.

The objective of this study is to assess the relative performances of physically-based and data-driven approaches for the estimation of LMA and EWT based on leaf optical properties. Our working questions are (1) what are the limitations of PROSPECT for LMA and EWT estimation, and is there any solution to overcome these limitations, and (2) what is the generalization ability of data-driven approaches when independent datasets are used for training and validation? We gathered six datasets in temperate, tropical and boreal ecosystems, with joint measurements of broadleaf optical properties, LMA and EWT (Section 2). Then, we designed specific protocols to address questions (1) and (2), and to perform an objective comparison of their performances (Section 3). This includes the selection of specific spectral information for PROSPECT inversion, and different strategies for the sampling of the training dataset for ML algorithms. Section 4 presents the results obtained with the different approaches, including a comparison of the validation with the six experimental datasets. Finally, section 5 discusses the potential and current limitations of the approaches and section 6 provides a conclusion.

2. MATERIALS

a. Global description of the datasets

For this study, six datasets were collected over various ecoregions, ranging from tropical forests, to temperate and boreal ecosystems (Table 1). LOPEX and ANGERS are publicly available and used in many publications. HYYTIALA, ITATINGA, NOURAGUES and PARACOU are unpublished datasets.

- The ANGERS\textsuperscript{1} dataset was collected in 2003 at INRA (Institut national de la recherche agronomique) in Angers (France). It encompasses physical measurements and biochemical

\textsuperscript{1}http://opticleaf.ipgp.fr/index.php?page=database
analyses collected over 43 species and varieties of woody and herbaceous plants. ANGERS was used for the calibration of the SAC for chlorophylls, carotenoids and anthocyanins in the latest versions of PROSPECT (Féret et al., 2017, 2008).

- The Leaf Optical Properties Experiment (LOPEX) dataset was collected in 1993 in Italy during a campaign conducted at the Joint Research Centre (Ispra, Italy) (Hosgood et al., 1994). It encompasses physical measurements and biochemical analyses collected over more than 50 species of woody and herbaceous plants, and has been widely used by the remote sensing community (Bowyer and Danson, 2004; Féret et al., 2008; Mobasher and Fatemi, 2013; Romero et al., 2012). The full LOPEX dataset includes dry and fresh samples and was used for the calibration of the SAC of $LMA$ (Féret et al., 2008), as well as broadleaf and needleleaf samples. However, only broadleaf samples were used in the current study, all fresh leaves except for one set of five dry maize leaf samples.

- The HYYTIALA dataset was collected in July 2017 at the Hyytiälä Forestry Field Station in Southern Finland in the frame of the Fluorescence Across Space and Time (FAST) campaign. This station is located in the boreal belt and is dominated by mixed forest of Scots pine, Norway spruce and silver birch. This dataset encompasses physical measurements and biochemical analyses collected over various native and non-native broadleaf species located in the field station.

- The ITATINGA dataset was collected in October 2015 as part of the IPEF-Eucflux project and HYPERTROPIK project (TOSCA, CNES, France), from experimental $Eucalyptus$ stands planted in November 2009 near the University of São Paulo forestry research station at Itatinga Municipality (São Paulo State, southeastern Brazil). ITATINGA includes sixteen genotypes and four species of $Eucalyptus$, eventually with hybrids, provided by different forestry companies in

[^2]: http://teledetection.ipgp.jussieu.fr/opticleaf/lopx.htm
different regions of Brazil. For each genotype, leaves corresponding to various developmental stages were collected, from juvenile to mature to senescent, and various locations within the crown (shaded leaves from the lower part of the crown, leaves from mid crown and sunlit leaves from the upper part of the crown). This dataset is the only genus-specific dataset. Hence, in spite of the large variability in terms of developmental stages, the ranges of $LMA$ and $EWT$ show significantly lower variability than those observed for the other datasets (Table 1). See Oliveira et al. (2017) for more details.

- The NOURAGUES dataset was collected at the CNRS Nouragues experimental research station, French Guiana, in September 2015, in the frame of the HYPERTROPIK project. This site is a lowland Amazonian forest, protected since 1996 by a Natural Reserve status. This dataset includes four to ten leaf samples from 38 emerging tropical tree species, collected from both shaded and sunlit parts of the crown. The Nouragues station is also a pilot site for remote sensing studies of tropical ecosystems (Réjou-Méchain et al., 2015).

- The PARACOU dataset was collected at the CIRAD-INRA Paracou experimental research station, French Guiana, in September 2015 (HYPERTROPIK project). This dataset includes four to ten leaf samples from 28 emerging tropical tree species, collected from both shaded and sunlit parts of the crown. Paracou is located in coastal lowland Amazonian forest. Various experiments are ongoing, including disturbance experiments, CO2 flux experiments, fertilization and long-term studies in forest dynamics and biodiversity.

b. Measurements of leaf optical properties

For all the samples, directional-hemispherical reflectance and transmittance (Schaepman-Strub et al., 2006) of the upper surface of the leaves were measured with a spectroradiometer and an integrating sphere in the visible (VIS), NIR and SWIR domains between 400 and 2500 nm. Here, we
used the infrared domain ranging from 900 to 2400 nm, due to the low influence of $LMA$ and $EWT$

on leaf optical properties below 900 nm, and to the low signal-to-noise ratio (SNR) beyond 2400 nm.

All datasets shared the same protocol for the measurement of leaf optical properties, and included spectral calibration for stray light in order to correct the imperfect collimation of the lamp beam as well as compensation for the optical properties of the coating of the integrating sphere when measuring leaf reflectance and transmittance (Asner et al., 2009; Carter and Knapp, 2001). The datasets were collected by different operators, and using different devices. Despite efforts to share a unique protocol for the acquisition of leaf optical properties, this diversity of operators, equipment and conditions of acquisition, is a possible source of bias that we discuss here.

c. Measurements of $LMA$ and $EWT$

The measurement of $EWT$ and $LMA$ shared the same protocol among experimental datasets. Leaf samples were collected in the field, stored in a cooler and measured in an experimental facility equipped with a precision scale and a drying oven. Minutes after measuring the leaf optical properties, disks of fresh leaf material were sampled using a cork borer, and immediately weighted using the precision scale to obtain $FW$ (Eq. 2). The disks were then placed in a drying oven at 85°C for at least 48 hours until constant mass was attained, and immediately weighted when out of the oven in order to determine $DW$ (Eq. 1 and Eq. 2) (Cornelissen et al., 2003; Pérez-Harguindeguy et al., 2013). $EWT$ and $LMA$ were then computed based on Eq. 1 and Eq. 2.

Table 1 summarizes basic statistics and information for each dataset. $LMA$ and $EWT$ were systematically measured for each sample in each dataset, except for the PARACOU dataset which only includes $LMA$ measurements. Similarly to optical properties, various sources of uncertainty may have affected $EWT$ and $LMA$ measurements, including errors in the area sampled on leaf material due to imperfect circular sampling disks, loss in water content between leaf optics measurements.
and weighting of fresh mass, or rehydration between drying and weighting of dry mass. However, care was paid to standardize data collection, so as to minimize the influence of these possible biases. 

\( EWT \) and \( LMA \) show no correlation for ITATINGA, weak correlation for LOPEX, moderate correlation for HYYTIALA and NOURAGUES, and strong correlation for ANGERS. A moderate correlation of 0.44 is measured when pooling all samples together.

Table 1. Summary of the main properties of the experimental datasets. Basic statistics for each dataset (minimum and maximum value, mean and standard deviation) are given for \( EWT \) and \( LMA \), as well as their correlation \( r(EWT, LMA) \).

<table>
<thead>
<tr>
<th></th>
<th>ANGERS</th>
<th>LOPEX</th>
<th>HYYTIALA</th>
<th>ITATINGA</th>
<th>NOURAGUES</th>
<th>PARACOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Samples</td>
<td>308</td>
<td>330</td>
<td>96</td>
<td>415</td>
<td>262</td>
<td>272</td>
</tr>
<tr>
<td>#Species/genotypes</td>
<td>43 sp.</td>
<td>46 sp.</td>
<td>10 sp.</td>
<td>4 sp. /16 gt.*</td>
<td>38 sp.</td>
<td>28 sp.</td>
</tr>
<tr>
<td>( EWT ) (mg.cm(^{-2}))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min – Max</td>
<td>4.40 – 34.00</td>
<td>0.29 – 52.48</td>
<td>3.68 – 23.73</td>
<td>2.20 – 20.20</td>
<td>3.20 – 38.10</td>
<td>N/A</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>11.47 ± 4.70</td>
<td>11.13 ± 6.97</td>
<td>9.16 ± 2.98</td>
<td>14.44 ± 2.09</td>
<td>11.73 ± 4.86</td>
<td>N/A</td>
</tr>
<tr>
<td>( LMA ) (mg.cm(^{-2}))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min – Max</td>
<td>1.66 – 33.10</td>
<td>1.71 – 15.73</td>
<td>2.76 – 15.77</td>
<td>6.90 – 14.70</td>
<td>3.10 – 21.10</td>
<td>5.28 – 25.56</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>5.12 ± 3.53</td>
<td>5.29 ± 2.47</td>
<td>6.27 ± 3.04</td>
<td>10.24 ± 1.62</td>
<td>10.81 ± 3.89</td>
<td>12.32 ± 4.06</td>
</tr>
<tr>
<td>( r(EWT, LMA) )</td>
<td>0.72</td>
<td>0.28</td>
<td>0.40</td>
<td>0.03</td>
<td>0.51</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Four species from Eucalyptus genus, corresponding to sixteen genotypes

3. METHODS

a. PROSPECT model: general presentation

PROSPECT is based on the generalized plate model (Allen et al., 1969, 1970) and was initially developed by Jacquemoud and Baret (1990). This model simulates the leaf directional-hemispherical...
reflectance and transmittance (Schaepman-Strub et al., 2006) with a limited number of input biophysical and biochemical variables, including various absorbing compounds and a unique leaf structure parameter, named $N$. Many versions have been developed since the first version, in order to include more absorbing compounds (Féret et al., 2017, 2008; Jacquemoud et al., 1996) or to adapt to specific conditions and leaf types, such as needle-shaped leaves (Malenovský et al., 2006). In this study, we used the latest version of PROSPECT, named PROSPECT-D (Féret et al., 2017). As we focused on leaf optical properties in the 900 – 2400 nm range, the capability of PROSPECT in terms of separation of pigments was not critical as no pigment absorbs in this spectral domain, but the refractive index differs from the one used on PROSPECT-5 (Féret et al., 2008). Brown pigments were not retrieved during the inversion, as including them showed no significant difference in the results obtained for any of the strategies tested here.

The $N$ parameter corresponds to the number of uniform compact plates separated by $N - 1$ air spaces. The value of $N$ represents the complexity of the leaf internal structure, with low $N$ values corresponding to moderate complexity such as in monocots, and higher $N$ values corresponding to higher complexity, a characteristic of dicots. To date, no protocol exists to experimentally estimate $N$ from leaf samples, other than using leaf optical properties. $N$ influences leaf scattering and shows negligible impact on leaf absorption: increasing $N$ values increase reflectance and decrease transmittance, and $N$ shows particularly strong effects in domains with low absorption, such as the NIR domain. Recently, Qiu et al. (2018) found an extremely strong correlation between $N$ and the ratio between reflectance and transmittance on simulated data.

PROSPECT can be run in forward or inverse mode. The forward mode aims at simulating leaf optical properties based on a full set of biophysical and biochemical properties (leaf chemistry and $N$). The inverse mode aims at identifying the optimal set of biophysical and biochemical properties that minimize a merit function (or goodness-of-fit criterion) based on a comparison between measured
and simulated leaf optics. A common inversion procedure is based on the numerical minimization of the sum of weighted square errors over all spectral bands available. The corresponding merit function $M$ is expressed as follows when using both reflectance and transmittance:

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

Eq. 3  

with $N$ the leaf structure parameter, $p$ the number of chemical constituents accounted for by PROSPECT and retrieved during the inversion, $C_i$ the biochemical content per leaf surface unit for constituent $i$, $\lambda_1$ and $\lambda_n$ the first and last wavebands investigated for inversion, $R_{\lambda}$ and $T_{\lambda}$ the experimental reflectance and transmittance measured at waveband $\lambda$, $R_{\lambda}$ and $T_{\lambda}$ the reflectance and transmittance simulated by PROSPECT with $\{N, \{ C_i \}_{i=1:p}\}$ as input variables, $W_{R,\lambda}$ the weight applied to the squared difference between experimental and simulated reflectances, and $W_{T,\lambda}$ its equivalent for transmittance. Eq. 3 can be used to estimate the full set of input variables, or a limited subset if prior information or arbitrary value is set for some variables.

b. Estimation of $EW_T$ and $LMA$ through iterative optimization

The large majority of the studies focusing on leaf scale model inversions through iterative optimization used Eq. 3 with unweighted merit function over the full spectral domain available ($W_{R,\lambda} = W_{T,\lambda} = 1$). This merit function provides accurate estimates of leaf pigments and $EW_T$ (Féret et al., 2017; Jacquemoud et al., 1996; Newnham and Burt, 2001), but several studies reported poor results for $LMA$ estimation (Féret et al., 2008; Riano et al., 2005). Colombo et al. (2008) used an alternative weighting, with $W_{R,\lambda} = (R_{\lambda})^{-2}$ and $W_{T,\lambda} = (T_{\lambda})^{-2}$, which is otherwise unused in the literature when inverting leaf models, and not so common when inverting canopy models (Baret and Buis, 2008). In practice, implementing such a merit function requires precaution as high sensor noise
(in particular in the SWIR domain) may result in close-to-zero reflectance and transmittance, leading to exaggerated importance of the corresponding spectral bands. This merit function then needs to be adapted to exclude these spectral bands. Colombo et al. (2008) reported fair performances of this merit function for the estimation of EWT, but poor performances for LMA. However, the SWIR domain beyond 1600 nm was not measured for their study, in spite of its importance for the estimation of LMA (Asner et al., 2011, 2009; le Maire et al., 2008). Therefore a fair comparison between this merit function and the unweighted merit function including the full spectral range is required.

As mentioned in the introduction, LMA estimation could also be improved by focusing on optimal spectral ranges (Li and Wang, 2011; Qiu et al., 2018; Wang et al., 2015). This amounts to choosing the weights such that \( W_{R,\lambda} = W_{T,\lambda} = 1 \) in the considered range, and \( W_{R,\lambda} = W_{T,\lambda} = 0 \) elsewhere. Note that such a procedure is relatively straightforward and could potentially be applied to the canopy scale in a similar way.

In this study, three inversion procedures were applied to the six independent experimental datasets, and their relative performances were compared. These inversion procedures correspond to “one-step” procedures, aiming at estimating EWT, LMA and \( N \) simultaneously from both reflectance and transmittance:

- **Iterative optimization 1 (IO1)** uses an unweighted merit function \( (W_{R,\lambda} = W_{T,\lambda} = 1) \) with reflectance and transmittance defined from 900 nm to 2400 nm.
- **Iterative optimization 2 (IO2)** uses a weighted merit function as defined by Colombo et al. (2008) \( (W_{R,\lambda} = (R_{\lambda})^{-2} \text{ and } W_{T,\lambda} = (T_{\lambda})^{-2}) \) with reflectance and transmittance defined from 900 nm to 2400 nm.
- **Iterative optimization 3 (IO3)** uses a weighted merit function defined by \( W_{R,\lambda} = W_{T,\lambda} = 1 \) over an optimal contiguous spectral domain \([\lambda_1, \lambda_n]\) defined between 900 and 2400 nm, and
\[ W_{R,\lambda} = W_{T,\lambda} = 0 \] elsewhere. This optimal spectral domain is adjusted in the present study and is the same for both reflectance and transmittance, and for all experimental datasets.

In the case of IO3, the exhaustive comparison of all combinations of spectral domains or spectral bands is computationally too demanding and extremely inefficient given the strong correlations between neighboring spectral domains. In order to reduce the computational cost, we focused on contiguous spectral domains defined by partitioning the initial spectral domain into 15 evenly-sized segments of 100 nm from 900 to 2399 nm. The choice of 100 nm segments is driven by constraints in terms of computation and by the ability to identify the main absorption features of EWT and LMA individually. The performances of PROSPECT inversion for the estimation of LMA and EWT were tested with all continuous spectral domains that can be generated from these 15 spectral segments, leading to 120 continuous segments. Finally, the spectral domain leading to the minimum RMSE averaged for all experimental datasets and for the estimation of both LMA and EWT from PROSPECT inversion was selected and defined as the optimal spectral range used in IO3.

For IO1, IO2 and IO3, \( N, EWT \) and LMA were simultaneously estimated using a constrained nonlinear optimization algorithm, i.e., the Sequential Quadratic Programming algorithm implemented within the Matlab function fmincon. The lower bounds selected for the three parameters to be optimized were defined to respect the condition of strict positivity and include minimum values observed for experimental data, whereas the upper bounds were set in order to include the maximum values observed for experimental data, with significant margins: EWT values were investigated between 0.01 and 80 mg.cm\(^{-2}\); LMA values were investigated between 0.01 and 40 mg.cm\(^{-2}\); \( N \) values were investigated between 0.5 and 4. No correlation constraints between EWT and LMA were included in the inversion procedure, since such correlation was not systematic between datasets.
The performances of data-driven methods inherently depend on the training data. In most cases, these performances are reported after splitting an experimental dataset into training and validation subsets, and the resulting regression models are not validated on fully independent datasets. In the perspective of operational applications, this raises the question of the possibility to share regression models adjusted with ML algorithms on public experimental datasets, and to use leaf spectroscopy operationally with no destructive measurements required to adjust dataset-specific regression models. With increasing use of machine learning, software packages including already trained regression models may be shared the same way statistical models derived from spectral indices have been proposed in the scientific literature (Féret et al., 2011). We want to answer the following questions related to data-driven methods: do regression models trained with one or several experimental datasets perform well when applied on independent datasets, or should training data systematically include samples from the validation dataset? To answer these questions, three strategies for the composition of a training dataset were tested, and the performances of data-driven methods were compared with PROSPECT inversions:

- **Training sampling 1 (TS1):** A single dataset was used as training data and the regression model was then applied on each of the remaining datasets.

- **Training sampling 2 (TS2):** All but one experimental datasets were used as training data, and the regression model was then applied on the remaining dataset.

- **Training sampling 3 (TS3):** All experimental datasets were pooled into a single one, and 300 samples (comparable in size to individual datasets) were randomly selected for training. Validation was then performed on the remaining samples (1668 samples for LMA, and 1396 samples for EWT), and performances (in terms of RMSE) were evaluated per individual dataset and globally. In each case, to account for possible sampling bias, random sampling of training
dataset was repeated 20 times and the distribution of RMSE values across all samplings was calculated.

Here, these three strategies used to define the training dataset were used with support vector machine (SVM) regression algorithm corresponding to the *Matlab* implementation of the *LibSVM* library (Chang and Lin, 2011). Reflectance and transmittance measurements from 900 to 2400 nm were stacked in a unique vector, resulting in $n_\lambda = 3002$ predictor spectral variables for each sample. Reflectance and transmittance were scaled between 0 and 1 for each spectral band, as well as leaf chemical constituent of interest (*LMA* and *EWT*). The radial basis function (RBF) kernel was selected, which implies optimizing two free parameters, $C$ and $\gamma$. $C$ is a cost parameter used to trade error penalty for stability and common to any SVM model. $\gamma$ is specific to RBF kernels and it corresponds to the inverse of the radius of influence of samples selected by the model as support vectors. The $C$ and $\gamma$ parameters were optimized using an exhaustive grid search ($C \in [10^{-2}; 10^{-1}; \ldots; 10^{+2}], \gamma \in [10^{-5}; 10^{-4}; \ldots; 10^{+1}]$ in order to include the default values recommended by Chang and Lin (2011) and a five-fold cross validation over the training data for each combination of $C$ and $\gamma$. The optimal $C$ and $\gamma$ values were then used with the full training data to adjust a regression model.

4. RESULTS

This section is divided into three subsections. The first subsection aims at identifying the optimal spectral domain to be used with *IO3*. This first section is a prerequisite to the second section, which then focuses on the comparison between the three types of iterative optimization, and the two types of training samplings based on the integrality of experimental datasets, *TS1* and *TS2*. Finally, the third section compares the performances of *TS3*, which is based on a random sampling among all
experimental datasets, with the performances of IO3 and TS2, when the validation samples are identical to those used in TS3.

a. Influence of spectral domain used for the estimation of EWT and LMA with PROSPECT inversion (optimization of IO3 method)

Figure 1 and Figure 2 show the results obtained for the estimation of EWT and LMA, respectively, when inverting PROSPECT over each dataset and each of the 120 spectral domains defined in Section 3.b with the IO3 method. For the sake of comparison, for each dataset, the RMSE was normalized by the RMSE obtained when using the spectral information from 900 to 2400 nm, and this normalized RMSE (NRMSE) was expressed as a percentage. In the case of EWT, the optimal spectral domain excluded the NIR domain under 1300 nm for all datasets, but no unique optimal spectral domain common to each dataset could be identified. The relative improvement induced by the reduction of the spectral domain was also strongly dataset-dependent: NRMSE was reduced by 23 % (LOPEX) to 56 % (NOURAGUES).
Figure 1. Normalized RMSE (NRMSE, in %) obtained for EWT with PROSPECT inversion method IO3 over each dataset and each reduced spectral domains bounded by a starting wavelength $\lambda_1$ (y-axis) and an ending wavelength $\lambda_2$ (x-axis). The normalization is specific to each dataset based on the performances of IO1 (NRMSE=100%, lower right corner). The green star indicates the spectral segment producing the best results.

In the case of LMA, both optimal spectral domain and relative improvement or degradation showed stronger consistency among datasets than for EWT (Figure 2). For all datasets, excluding information from 1500 nm and beyond led to strong degradations of the performances. In the case
of LOPEX and HYYTIALA, estimation of $LMA$ could be improved only when using spectral domains with ending wavelength between 2100 and 2400 nm, except when using a narrow spectral domain from 1600 to 1800 nm. For the four other datasets, extended spectral combinations led to improved estimation of $LMA$, as most of the combinations excluding the domain from 900 to 1200 nm led to improved estimation of $LMA$, except when using a reduced spectral domain ranging from 1800 to 2100 nm only, which corresponds to one of the main absorption features of water. Overall, the optimal spectral range excluded the NIR domain and included spectral information until 2400 nm for all datasets. The relative improvement induced by the selection of an optimal specific for each dataset ranged from 60 (ITATINGA) to 67 % (NOURAGUES).
Figure 2. Normalized RMSE (NRMSE, in %) obtained for LMA estimation with PROSPECT inversion method IO3, over each dataset and each reduced spectral domains bounded by a starting wavelength $\lambda_1$ (y-axis) and an ending wavelength $\lambda_2$ (x-axis). The normalization is specific to each dataset based on the performances of IO1 (NRMSE=100%, lower right corner). The green star indicates the spectral segment producing the best results.

These figures provide a visual representation of the spectral domains leading to improved or decreased performances compared to full spectral information. They confirm that selecting the appropriate spectral information during inversion strongly influences for the estimation of leaf constituents.
Figure 3 provides NRMSE for the estimation of $EWT$ and $LMA$ averaged over all datasets, and confirms suboptimal performances obtained when using NIR information only. Overall, the spectral domain ranging from 1700 to 2400 nm was found to be optimal when estimating $EWT$ and $LMA$ simultaneously (mean NRMSE was reduced by 33% for $EWT$ and by 55% for $LMA$), and was used hereafter within the IO3 method.

**Figure 3.** Mean normalized RMSE values (NRMSE, in %) obtained for the estimation of $EWT$ (left), $LMA$ (center), and both constituents (right), after PROSPECT inversion over all experimental datasets pooled and each of the 120 spectral domains defined in Section 3.b. The green star indicates the spectral segment producing the best results.

b. Comparison of PROSPECT inversion methods and ML algorithms for the estimation of $LMA$ and $EWT$: training ML with independent datasets

The performances obtained for the estimation of $EWT$ when using TS1 and TS2 for ML regression, and IO1, IO2 or IO3 (with the 1700 – 2400 nm spectral range) for PROSPECT inversion are reported in Table 2. Overall, IO2 and IO3 produced the most consistent results, systematically outperforming the
other methods. ML regressions performed particularly poorly compared to \textit{IO2} and \textit{IO3}, and \textit{TS2} led to the better results than \textit{TS1} (except form HYYTIALA). \textit{TS1} led to very inconsistent results, with 175\% increase compared to \textit{IO2} and \textit{IO3} on average, and up to 500\% increase in RMSE compared to \textit{PROSPECT} inversion \textit{IO2} when estimating \textit{EWT} from ITATINGA after training with LOPEX.

\textbf{Table 2.} RMSE values (in mg.cm\textsuperscript{-2}) obtained for the estimation of \textit{EWT} with SVM and training strategies \textit{TS1} and \textit{TS2}, and with \textit{IO1}, \textit{IO2} and \textit{IO3}. For each column (validation dataset), the minimum RMSE is indicated in bold, and colors correspond to the level of performances, from green color for minimum RMSE to red color for maximum RMSE.

<table>
<thead>
<tr>
<th>Method</th>
<th>Valid</th>
<th>ANGERS</th>
<th>LOPEX</th>
<th>HYYTIALA</th>
<th>ITATINGA</th>
<th>NOURAGUES</th>
<th>PARACOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{TS1} ANGERS</td>
<td>-</td>
<td>4.82</td>
<td>1.90</td>
<td>3.31</td>
<td>2.49</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>LOPEX</td>
<td>3.14</td>
<td>-</td>
<td>3.23</td>
<td>6.73</td>
<td>2.32</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HYYTIALA</td>
<td>3.79</td>
<td>5.40</td>
<td>-</td>
<td>2.84</td>
<td>3.92</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ITATINGA</td>
<td>3.38</td>
<td>5.82</td>
<td>3.03</td>
<td>-</td>
<td>3.43</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>NOURAGUES</td>
<td>2.54</td>
<td>5.04</td>
<td>3.15</td>
<td>2.47</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>PARACOU</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>\textit{TS2} All but 1</td>
<td>2.47</td>
<td>4.54</td>
<td>2.68</td>
<td>2.08</td>
<td>2.10</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>\textit{IO1} \textit{PROSPECT}</td>
<td>2.07</td>
<td>2.03</td>
<td>1.72</td>
<td>1.93</td>
<td>3.44</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>\textit{IO2} \textit{PROSPECT}</td>
<td>1.48</td>
<td>1.68</td>
<td>1.44</td>
<td>1.13</td>
<td>1.21</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>\textit{IO3} \textit{PROSPECT}</td>
<td>1.41</td>
<td>1.70</td>
<td>1.21</td>
<td>1.20</td>
<td>1.66</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

25
Figure 4 provides scatterplots for the results showed in Table 2 and corresponding to IO1, IO2, IO3 and SVM regression with sampling strategy TS2. Overall, IO2 showed the best performances for the estimation of $EWT$, and SVM regression produced the lowest performances, mainly because of the strong error obtained for extreme values on LOPEX.

The performances obtained for the estimation of $LMA$ when using training samplings TS1 and TS2 for ML regression, and IO1, IO2 or IO3 for PROSPECT inversion are reported in Table 3. IO3 outperformed the other methods for all datasets except HYYTIALA and ITATINGA: IO2 slightly outperformed IO3 for ITATINGA only and TS2 outperformed IO3 for HYYTIALA and ITATINGA. However, the difference in RMSE between IO3 and the optimal method remained less than 20% for these two datasets. The relative performances obtained with IO1 and IO2 differed among datasets:
while using IO2 led to significantly improved estimation of LMA compared to IO1 for five datasets (from a 26% decrease in RMSE for LOPEX to more than 50% for ITATINGA, NOURAGUES and PARACOU), and slightly degraded estimation compared to IO3 for four datasets, the performances obtained for HYYTIALA were degraded by more than 75% compared to IO1, with systematic strong overestimation (Figure 5). On the other hand, the RMSE corresponding to estimation of LMA using IO3 decreased by 60% compared to IO1. ML regression trained with TS2 performed better than IO1 overall but was outperformed by IO2 and IO3. As for EWT, ML trained with TS1 led to very inconsistent results, and was strongly outperformed by IO2, IO3 and ML regressions trained with strategy TS2 in most cases.

**Table 3.** RMSE values (in mg.cm$^{-2}$) obtained for the estimation of LMA with SVM and training samplings TS1 and TS2, and with IO1, IO2 and IO3. For each column (validation dataset), the minimum RMSE is indicated in bold, and colors correspond to the level of performances, from green color for minimum RMSE to red color for maximum RMSE.
Figure 5 provides scatterplots for the results showed in Table 3 and corresponding to IO1, IO2, IO3 and SVM regression with training sampling TS2. Overall, IO3 produced the most accurate estimation of LMA. IO2, IO3 and TS2 respectively resulted in 33%, 55% and 27% decreases in RMSE for the estimation of LMA when compared to IO1.

Figure 5. LMA estimation results obtained using PROSPECT inversion (IO1, IO2, IO3) and SVM regression (training sampling TS2).
c. Comparison of PROSPECT inversion methods and ML algorithms for the estimation of \( LMA \) and \( EWT \): training ML with pooled datasets

Table 4 and Table 5 summarize the performances of SVM regression for the estimation of \( EWT \) and \( LMA \) when TS3 is selected as training strategy (i.e. all datasets are pooled together and 300 calibration samples are randomly selected). The performances corresponding to IO3 and TS2 were computed for the same validation samples as with TS3 for each of the 20 repetitions in order to ensure fair comparison.

The mean performances reported in Table 4 and Table 5 were very similar to those reported in Table 2 and Table 3 for both IO3 and TS2, which means that IO3 systematically outperformed TS2 on individual datasets, except for the estimation of \( LMA \) for HYTTIALA and ITATINGA. TS3 outperformed TS2 in most cases for the estimation of both \( EWT \) and \( LMA \). Still, TS3 was outperformed by IO3 when estimating \( EWT \), the overall RMSE increasing by 44% (and by 99% when using TS2). When estimating \( LMA \), TS3 and IO3 showed very similar overall performances, with less than 6% increase of RMSE for TS3 when compared to IO3. IO3 and TS3 showed very similar average RMSE for LOPEX, HYTTIALA and NOURAGUES, TS3 showed higher RMSE for ANGERS and PARACOU, and lower RMSE for ITATINGA. However, the standard deviations associated with these performances highlight the strong effect of training and validation sampling on the performances of the ML algorithm: the standard deviation computed over 20 repetitions was 5 to 20 times higher for TS3 than IO3 when estimating \( EWT \), while it was 2.5 to 10 times higher when estimating \( LMA \). The standard deviations related to the performances of TS2 were generally similar to those obtained for IO3, suggesting that the strong differences in performance between regression models were induced by the selection of the training samples.
Table 4. Mean RMSE and standard deviation of RMSE (both in mg.cm\(^{-2}\)) of the estimation of \(EWT\) using SVM regression (TS2 and TS3) and PROSPECT inversion (IO3) on the validation samples used for TS3. Best mean performances are indicated in bold.

<table>
<thead>
<tr>
<th></th>
<th>ANGERS</th>
<th>LOPEX</th>
<th>HYYTIALA</th>
<th>ITATINGA</th>
<th>NOURAGUES</th>
<th>PARACOU</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS3</td>
<td>1.76±0.24</td>
<td>2.77±0.47</td>
<td>2.08±0.37</td>
<td>1.64±0.33</td>
<td>2.18±0.37</td>
<td>-</td>
<td>2.12±0.26</td>
</tr>
<tr>
<td>TS2</td>
<td>2.47±0.04</td>
<td>4.55±0.28</td>
<td>2.66±0.07</td>
<td>2.08±0.04</td>
<td>2.1±0.05</td>
<td>-</td>
<td>2.97±0.10</td>
</tr>
<tr>
<td>IO3</td>
<td>1.43±0.04</td>
<td>1.70±0.07</td>
<td>1.21±0.04</td>
<td>1.21±0.02</td>
<td>1.65±0.05</td>
<td>-</td>
<td>1.47±0.02</td>
</tr>
</tbody>
</table>

Table 5. Mean RMSE and standard deviation of RMSE (both in mg.cm\(^{-2}\)) of the estimation of \(LMA\) using SVM regression (TS2 and TS3) and PROSPECT inversion (IO3) on the validation samples used for TS3. Best mean performances are indicated in bold (differences in mean RMSE < 1% are considered equivalent).

<table>
<thead>
<tr>
<th></th>
<th>ANGERS</th>
<th>LOPEX</th>
<th>HYYTIALA</th>
<th>ITATINGA</th>
<th>NOURAGUES</th>
<th>PARACOU</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS3</td>
<td>1.70±0.28</td>
<td>1.98±0.56</td>
<td>1.56±0.22</td>
<td>1.12±0.29</td>
<td>1.59±0.19</td>
<td>2.01±0.18</td>
<td>1.64±0.18</td>
</tr>
<tr>
<td>TS2</td>
<td>2.24±0.21</td>
<td>4.05±0.06</td>
<td>1.33±0.04</td>
<td>1.23±0.02</td>
<td>2.13±0.03</td>
<td>2.43±0.05</td>
<td>2.31±0.05</td>
</tr>
<tr>
<td>IO3</td>
<td>0.92±0.03</td>
<td>2.00±0.07</td>
<td>1.54±0.08</td>
<td>1.45±0.04</td>
<td>1.58±0.06</td>
<td>1.77±0.07</td>
<td>1.54±0.03</td>
</tr>
</tbody>
</table>

5. DISCUSSION

a. Differences in performances among merit functions

Our study shows that \(IO1\), the most commonly used merit function, is actually outperformed by a less common merit function (\(IO2\)) when estimating \(EWT\) and \(LMA\) from PROSPECT inversion using reflectance and transmittance in the NIR/SWIR domain (900-2400 nm). These results are in agreement with the results obtained when investigating the optimal spectral domain to be used with
Figure 3 shows that, in most cases, selecting a spectral domain including NIR information leads to suboptimal estimation of both $EWT$ and $LMA$. Therefore, the application of a weight inversely proportional to the square of the reflectance and transmittance ($IO2$) reduce the importance of spectral domains showing higher reflectance and transmittance values such as the NIR domain. The improvement is particularly strong for the estimation of $LMA$, as reported in Figure 3. The particularly low performances obtained for the estimation of $LMA$ on HYYTIAA were also investigated. The leaf optical properties measured for this dataset showed low SNR, particularly in the SWIR domain for wavelengths of 2300 nm and beyond. The estimation of $LMA$ with $IO2$ was strongly improved on this dataset when applying a Savitzky-Golay smoothing filter and restricting the spectral domain from 1700 to 2300 nm. The exclusion of the spectral domain beyond 2300 nm was responsible for the strongest improvement. Finally, the RMSE obtained for HYYTIAA when using the merit function used in $IO2$ and these preprocessing reached 1.97 mg.cm$^{-2}$, which is still 30% higher than the RMSE obtained with $IO3$. Therefore using $IO2$ is strongly discouraged when the signal to noise ratio of leaf optical properties is not sufficient, while $IO3$ based on the 1700-2400 nm spectral range appears to be reliable even with low SNR.

b. Physical interpretation of the performances obtained with PROSPECT inversion

As highlighted in the previous section, the SNR of leaf optical properties can become a strong limitation when estimating leaf constituents using PROSPECT inversion if the spectral domain and merit functions are not carefully chosen. However, this SNR is not the main limiting factor explaining the poor performances of $IO1$ for the estimation of $LMA$ and its suboptimal performances for the estimation of $EWT$. Indeed, the NIR domain is theoretically characterized by a higher signal to noise ratio for leaf material but still appears to be the main limitation for an accurate estimation of these
leaf constituents. Therefore, we attempt here to list possible explanations for such poor performances.

i. Predominant water absorption

The main reason cited to explain the poor retrieval of $LMA$ is the predominant water absorption in the SWIR domain. Indeed, Figure 3 shows that $LMA$ is poorly estimated when the spectral domains used for inversion mainly include domains with strong water absorption, such as the domain from 1800 to 2100 nm. However Figure 3 also shows that $LMA$ can still be estimated accurately even if most of the spectral information corresponds to domains with predominant water absorption. Our results show that the main limitation with $IO1$ is actually caused by the NIR domain between 900 and 1300 nm: most of the spectral domains excluding such wavebands resulted in improved estimation of $LMA$. The 900-1300 nm range does not show predominant water or dry matter absorption, so the poor retrieval of $LMA$ cannot be explained by absorption features hidden by water absorption or any other constituent.

ii. Approximations of PROSPECT

As any model, PROSPECT is based on a number of approximations. Although some of these approximations are possible sources of inaccuracy in specific situations, they guarantee good overall performances given a minimum number of descriptors of leaf biophysical properties. Model discrepancies in the simulation of leaf optical properties may be explained by inaccurate physical description at three levels: surface effects, volume scattering and volume absorption. Surface effects strongly depend on the presence of waxes or trichomes, and Barry and Newnham (2012) reported how epicuticular waxes affect PROSPECT inversion. Surface effects mostly influence leaf reflectance in the domains characterized by strong absorption where the leaf reflectance is minimum (Bousquet et al., 2005; Jay et al., 2016). In the NIR/SWIR spectral range, these domains mainly depend on water absorption. The sensitivity analysis performed by Jay et al. (2016) with
similar EWT values showed that surface effects have the largest influence beyond 1800 nm, this
domain being close to the one leading to optimal PROSPECT inversion results with IO3 (1700-2400
nm). Such a result thus tends to indicate that surface effects had a limited detrimental influence on
estimation performance.

Volume scattering is modeled by multiple factors in PROSPECT, including leaf structure with the $N$
parameter, and the refractive index. The unique value of the refractive index is a well-identified
simplification of PROSPECT, as it does not agree with the Kramers-Kronig relations stating that the
real (refractive index) and imaginary (absorption coefficient) parts of the complex refractive index of
a medium are physically linked (Lucarini et al., 2005). Qiu et al. (2018) developed PROSPECT-g, a
modified version of PROSPECT including an additional wavelength-independent factor specific to
each leaf and aiming at representing first-order effects of anisotropic scattering, which are not
included through the N structural parameter of the original PROSPECT model. They also proposed a
multistage inversion to be used with PROSPECT-g. This inversion procedure may strongly increase
computing time, and the applicability of PROSPECT-g inversion at the canopy scale does not seem
straightforward as additional parameters may increase the ill-posedness of canopy models such as
PROSAIL (Jacquemoud et al., 2009). However, they reported promising results, including improved
estimation of $LMA$ and improved simulation of both reflectance and transmittance in the NIR
domain when compared to PROSPECT-5.

Volume absorption is defined by the SACs which are adjusted based on experimental data during the
calibration of PROSPECT (Féret et al., 2008, 2017). We attempted a recalibration of the SAC for $LMA$
in order to reduce the inaccuracies observed between experimental and simulated data, and
improve the estimation of $LMA$. This did not lead to any improvement when including the NIR
domain. Moreover, the incorrect definition of the SAC corresponding to $LMA$ would lead to
systematic underestimation or overestimation of absorption when running PROSPECT in direct
mode. However, the analysis of the residuals between measured leaf optical properties and their simulated counterparts obtained with PROSPECT in direct mode did not result in systematic errors (results not shown). The SAC corresponding to LMA in PROSPECT integrates the optical influences of various organic constituents, which may also lead to inaccuracies if leaf samples include strong variations in stoichiometry. However, the data required to test this possible source of inaccuracy was not available.

iii. Bias in the leaf optical measurements

As highlighted in the introduction, the uncertainty associated to leaf optical measurements in the NIR domain may be increased because of the incomplete collection of the light leaving the highly scattering tissue (Merzlyak et al., 2002). Merzlyak et al. (2004) proposed a correcting factor for transmittance based on the hypothesis that leaf absorption in the NIR domain from 780 to 900 nm is negligible for healthy leaves. However this correcting factor is not adopted as a standard correction by the community. In order to detect possible uncertainty in the optical measurements in the NIR domain with our data, we tested our ML approach with TS1 (training with a unique dataset) and spectral information either from 1700 to 2400 nm or from 1400 to 2400 nm (results not showed). For both LMA and EWT, the regression models applied on independent datasets performed similarly for the two spectral domains considered, but systematically performed better than the regression models trained with the spectral information from 900 to 2400 nm. However, they were still outperformed by PROSPECT inversion. Such a result thus tends to confirm that leaf optical measurements in the NIR domain might be affected by some experimental uncertainty. The poor performances reported for the estimation of LMA with PROSPECT inversion using IO1 are therefore mainly explained by the use of the NIR domain, which is subject to inaccuracies, from a modeling and/or from an experimental point of view. Based on our study, we cannot conclude on the relative importance of one or the other factor. These two possibilities should then be considered
and tested using the methods proposed in the literature (Merzlyak et al., 2004; Qiu et al., 2018).

Finally, the difference between directional hemispherical measurements and bidirectional measurements should be systematically accounted for and appropriate physical models should be used with the type of data they are expected to simulate.

c. Influence of the sampling of the training dataset on machine learning algorithms

Our results highlight the strong influence of the training dataset on the performances of ML methods, which is not an original result \textit{per se}. However, the different training strategies tested here show that regression models should be used with extreme care when they are applied on data which were not collected in the exact same conditions as training data. Finally, the optimal training strategy in our case, TS3, requires that each campaign aiming at collecting leaf optical properties in order to estimate constituent content based on statistical/ML methods should include destructive measurements to be used during the training step. This means that publicly available datasets such as ANGERS and LOPEX should not be used as the only training datasets for the estimation of leaf chemistry based on spectroscopy from independent datasets. The origin of the suboptimal performances obtained in particular with TS2 and TS3 should also be investigated. ML algorithms are currently mainly used for their predictive capacity. However, they can also as part of a descriptive framework. Feilhauer et al. (2015) proposed an interesting illustration as they suggested combining multiple methods in order to identify the most relevant spectral bands related to leaf chemistry, based on both experimental and simulated data. Following the same method, the identification of the spectral bands maximizing the generalization ability of ML algorithms by discarding spectral domains prompt to experimental uncertainty or model approximations could be considered. Finally, hybrid methods using simulated data during the training stage of a ML algorithm appear as an interesting alternative to data-driven methods purely based on experimental data, and further
investigation is needed in order to define the proper strategy to generate such training dataset and combine the generalization ability of physically-based approaches with the computational efficiency of data-driven approaches.

d. Relevance of these results for leaf trait monitoring

The results obtained in this study contribute to a better understanding of the optimal remotely-sensed monitoring of $LMA$ and $EWT$, two key vegetation traits that convey multiple information about the spatial and temporal variation in ecological and functional diversity of terrestrial ecosystems. This can possibly contribute to facilitating the study of plant functions and their interactions with and responses to the environment. As an example, Feihauer et al. (2018) provide a good illustration of the interest of remotely-sensed $LMA$ for ecological analysis of wetland vegetation, in particular for the better understanding of the effect of long-term drought on ecosystem functions. They focused on $LMA$ because of its plasticity in response to variable environmental conditions, and its relationship with potential growth rate.

The estimation of these traits at the leaf scale now needs to be further investigated at the canopy scale. In order to test the applicability of our approach at the canopy scale, the first step will consist in working with a simulated dataset obtained with canopy reflectance models such as SAIL (Jacquemoud et al., 2009; Verhoef, 1984) and DART (Gastellu-Etchegorry et al., 1996, 2015). The direct application of model inversion based on iterative optimization restricts the complexity of the canopy model, hence the type of vegetation to be investigated: the adaptation of our method should be relatively straightforward when using PROSAIL on homogeneous canopy covers, but hybrid methods should be considered when using DART simulations and working on heterogeneous canopy covers.
An important challenge for the applicability of our results at the canopy scale is the low intensity of the solar radiation in the optimal SWIR domain identified in this study, which usually leads to low signal to noise ratio. Currently, hyperspectral information is mainly available from airborne imaging spectroscopy (Asner et al., 2012; Schaepman et al., 2015). Asner et al. (2015) obtained accurate estimation of LMA based on multivariate statistical methods applied on imaging spectroscopy for heterogeneous canopies in tropical ecosystems, and they also concluded on the importance of the spectral domain from 2000 nm to 2500 nm for a proper calibration of the regression models. Recently, Feilhauer et al. (2018) reported good suitability of airborne imaging spectroscopy analyzed with a hybrid method (Random forest trained with PROSAIL simulations) for LMA mapping in natural ecosystems. Hyperion is the only spaceborne sensor, but the signal to noise ratio is known to be relatively low (le Maire et al., 2008). The contribution of modeling through sensitivity studies performed at canopy scale may therefore provide insightful information for the instrumental specifications of future satellites dedicated to the monitoring of vegetation and environment such as EnMAP, and for the development of algorithms (Jetz et al., 2016; Lee et al., 2015; Leitão et al., 2015).

6. CONCLUSIONS

In this paper, we compared the performances of various methods for the estimation of EWT and LMA based on leaf reflectance and transmittance in the spectral domain ranging from 900 to 2400 nm. These methods included PROSPECT inversion based on iterative optimization with various merit functions and machine learning (ML) algorithms with different training strategies. Six independent datasets acquired from various vegetation types, including temperate, boreal and tropical ecosystems were used in order to validate our results.

Our results showed that the poor performances of PROSPECT inversion reported in many studies for the estimation of LMA could be dramatically improved when excluding spectral information in the...
NIR domain from 900 to 1300 nm. We investigated the performances of PROSPECT inversion for the estimation of $EWT$ and $LMA$ using multiple spectral subdomains, and identified an optimal spectral domain ranging from 1700 to 2400 nm. Overall, PROSPECT inversion performed on this spectral domain provided more accurate $LMA$ and $EWT$ estimates than ML algorithms trained on experimental datasets. Unlike ML algorithms, PROSPECT inversion showed strong generalization ability. Despite numerous studies showing the poor performances of PROSPECT for the estimation of $LMA$, our study shows that model inversion using iterative optimization can outperform other methods with an appropriate merit function, with no need for recalibration or training stage. By this study, we therefore confirm the strong potential and accuracy of PROSPECT on critical spectral domains. We also identified weaknesses which can be attributed either to physical modeling and experimental acquisition of leaf optical properties in the NIR domain.

These results motivate further investigation involving hybrid methods for the estimation of $LMA$ and $EWT$, in order to take advantage of the computational efficiency of data-driven algorithms and overcome limitations inherent to suboptimal experimental sampling of training data. Implications of these results for the optimal estimation of $LMA$ and $EWT$ at the canopy scale will also be investigated, as $LMA$ and $EWT$ are both key traits when monitoring ecosystem functions.

7. ACKNOWLEDGMENTS

The authors warmly thank Luc Bidel, Christophe François and Gabriel Pavan who collected the ANGERS dataset. This work was funded by the TOSCA program grant of the French Space Agency (CNES) (HyperTropik project); the International Network for Terrestrial Research and Monitoring in the Arctic (INTERACT) (FLUO-SYNTHESIS project); the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES). ITATINGA genotype test is funded in part by the EUCFLUX project of Forestry Science and Research Institute (IPEF) and in part by SOERE F-ORE-T, which is
supported annually by Ecofor, Allenvi, and the ANAEE-F. The authors also thank Valentine Alt, Samuel Counil, and Philippe Gaucher for tree climbing and data collection, as well as Anna Grandchamp for her help during data collection in Nouragues field station.

8. REFERENCES


de la Riva, E.G., Olmo, M., Poorter, H., Ubera, J.L., Villar, R., 2016. Leaf Mass per Area (LMA) and its Relationship with Leaf Structure and Anatomy in 34 Mediterranean Woody Species along a


9. LIST OF FIGURE CAPTIONS

Figure 1. Normalized $RMSE$ ($NRMSE$, in %) obtained for $EWT$ with PROSPECT inversion method $IO3$ over each dataset and each reduced spectral domains bounded by a starting wavelength $\lambda_1$ (y-axis) and an ending wavelength $\lambda_2$ (x-axis). The normalization is specific to each dataset based on the performances of $IO1$ ($NRMSE=100\%$, lower right corner). The green star indicates the spectral segment producing the best results.

Figure 2. Normalized $RMSE$ ($NRMSE$, in %) obtained for $LMA$ estimation with PROSPECT inversion method $IO3$, over each dataset and each reduced spectral domains bounded by a starting wavelength $\lambda_1$ (y-axis) and an ending wavelength $\lambda_2$ (x-axis). The normalization is specific to each dataset based on the performances of $IO1$ ($NRMSE=100\%$, lower right corner). The green star indicates the spectral segment producing the best results.

Figure 3. Mean normalized RMSE values ($NRMSE$, in %) obtained for the estimation of $EWT$ (left), $LMA$ (center), and both constituents (right), after PROSPECT inversion over all experimental datasets pooled and each of the 120 spectral domains defined in Section 3.b. The green star indicates the spectral segment producing the best results.

Figure 4. $EWT$ estimation results obtained using PROSPECT inversion ($IO1$, $IO2$, $IO3$) and ML regression (training sampling $TS2$).

Figure 5. $LMA$ estimation results obtained using PROSPECT inversion ($IO1$, $IO2$, $IO3$) and SVM regression (training sampling $TS2$).