Journal of Photochemistry and Photobiology B: Biology 134 (2014) 37–48

Contents lists available at ScienceDirect

Journal of Photochemistry and Photobiology B: Biology

journal homepage: www.elsevier.com/locate/jphotobiol

Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset

Shari Van Wittenberghe [a,*], Jochem Verrelst [b], Juan Pablo Rivera [b], Luis Alonso [b], José Moreno [b], Roeland Samson [a]

[a] Department of Bioscience Engineering, Faculty of Sciences, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium
[b] Image Processing Laboratory, University of Valencia, C/ Catedrático José Beltrán 2, E-46980 Paterna, Valencia, Spain

ARTICLE INFO

Article history:
Received 30 October 2013
Received in revised form 13 February 2014
Accepted 6 March 2014
Available online 25 March 2014

Keywords:
Machine learning algorithm
Spectral features
Hyperspectral
Parameter retrieval
Chlorophyll
Leaf structure
Specific leaf area
Leaf water content

ABSTRACT

Biochemical and structural leaf properties such as chlorophyll content (Chl), nitrogen content (N), leaf water content (LWC), and specific leaf area (SLA) have the benefit to be estimated through nondestructive spectral measurements. Current practices, however, mainly focus on a limited amount of wavelength bands while more information could be extracted from other wavelengths in the full range (400–2500 nm) spectrum. In this research, leaf characteristics were estimated from a field-based multi-species dataset, covering a wide range in leaf structures and Chl concentrations. The dataset contains leaves with extremely high Chl concentrations (>100 $\mu g\,cm^{-2}$), which are seldom estimated. Parameter retrieval was conducted with the machine learning regression algorithm Gaussian Processes (GP), which is able to perform adaptive, nonlinear data fitting for complex datasets. Moreover, insight in relevant bands is provided during the development of a regression model. Consequently, the physical meaning of the model can be explored. Best estimates of SLA, LWC and Chl yielded a best obtained normalized root mean square error of 6.0%, 7.7%, 9.1%, respectively. Several distinct wavebands were chosen across the whole spectrum. A band in the red edge (710 nm) appeared to be most important for the estimation of Chl. Interestingly, spectral features related to biochemicals with a structural or carbon storage function (e.g. 1090, 1550, 1670, 1730 nm) were found important not only for estimation of SLA, but also for LWC, Chl or N estimation. Similar, Chl estimation was also helped by some wavebands related to water content (950, 1430 nm) due to correlation between the parameters. It is shown that leaf parameter retrieval by GP regression is successful, and able to cope with large structural differences between leaves.

1. Introduction

Spectroscopic retrieval of vegetation biophysical parameters is of general interest due to their explicit connection with the vegetation's physiological function and the possibility to extrapolate the technique to remote sensing. Absorption and reflectance spectra of fresh leaves are thereby able to provide estimations of chlorophyll (Chl) content [1], water content [2,3], leaf structure [4] and leaf structural components [5,6]. This led to the development of empirical relationships between those leaf properties and the corresponding leaf spectra. To estimate leaf biochemical and structural leaf and canopy properties, hyperspectral narrowband data from the visible to the shortwave infrared (SWIR) (400–2500 nm), is nowadays often used. In comparison with broadband data, hyperspectral narrowband data makes use of a continuous spectral range, revealing specific absorption features of biochemicals [5]. Relationships between reflectance spectra and leaf characteristics have been often derived through multiple regression and other least-squares statistical methods [4,5,7]. More common, however, is the approach to estimate leaf or vegetation parameters by the use of spectral vegetation indices [2,8,9]. These are simple formulations that consist of a combination of two or more reflectance wavelengths and its outcome is then related to a parameter of interest. All kinds of indices (e.g. simple ratio's, normalized difference ratio) have been developed to calibrate with leaf parameters [9,10]. They typically involve only two wavelengths from the whole spectrum, one correlated and the other uncorrelated to the biochemical or structural parameter. For instance, the most widely used Normalized Difference Vegetation Index [11] uses a reflectance band at the Chl absorption maxima

* Corresponding author. Tel.: +32 3 265 3569; fax: +32 3 265 3225.
E-mail address: Shari.VanWittenberghe@ua.ac.be (S. Van Wittenberghe).

http://dx.doi.org/10.1016/j.jphotobiol.2014.03.010
1011-1344/

in the red (sensitive to Chl) and a reflectance band in the near-infrared (NIR) (insensitive to Chl) as a reference band. Even more sensitive to Chl are the indices based on the red edge spectral region (680–740 nm) since this region is less subjected to the absorption saturation in the red [1,12,13]. Red edge based indices have been proven to perform well for the estimation of low (0–40 $\mu g\,cm^{-2}$) to medium (40–80 $\mu g\,cm^{-2}$) Chl content [1], but concentrations exceeding 80 $\mu g\,cm^{-2}$ have been hardly estimated.

An alternative way to estimate Chl content is based on Chl fluorescence [14–16]. Chl fluorescence is the plant-specific red and far-red emission by Chl *a* upon absorption of photosynthetic active radiation. Light energy emitted as Chl fluorescence competes with photochemical and non-photochemical de-excitation of excited Chl. Similar to reflectance indices, a ratio of two fluorescence wavelengths is typically used, containing one wavelength (usually the red peak at 690 nm) influenced by the re-absorption of fluoresced light by Chl, and the other wavelength (usually the far-red peak at 740 nm) not influenced by re-absorption resulting in a red/far-red ratio. However, similar to reflectance in the red, strong re-absorption at the level of the red fluorescence emission peak makes it hard to estimate higher levels of Chl content [14,17].

Both reflectance and fluorescence indices based on a few wavelengths therefore pose limitations for species with a high Chl content. The same holds for estimation of other parameters related to leaf structure such as specific leaf area. In fact, using only a few wavelengths highly underutilizes the potential of full range spectral data to correlate with leaf parameters [18]. Moreover, spectral variation in the NIR (750–1400 nm) and the SWIR (1400–2500 nm) could provide more insight about the functioning of plants compared to the wavelengths which are absorbed for photosynthesis [19]. Light in these regions can penetrate deeper into the leaf (or canopy), being less subjected to leaf absorption saturation as in the visible. Therefore, the full spectrum might contribute to parameter retrieval in order to establish broadly applicable relationships.

Nonparametric models are adjusted to predict a parameter of interest by using a training dataset of input–output data pairs which come from concurrent measurements of the parameter (e.g., Chl content) and the corresponding spectral observation (e.g., reflectance or fluorescence). They thus hold the possibility of generating valuable information from a complete spectral profile. Several nonparametric regression algorithms are available in the statistics literature, and recently have been introduced for biophysical parameter retrieval [20,21]. Particularly, the family of Machine Learning Regression Algorithms (MLRAs) emerged as a powerful nonparametric approach for delivering biophysical parameters. MLRAs have the potential to generate adaptive, robust relationships and, once trained, they are fast to apply [22]. Typically, they are able to cope with the strong nonlinearity of the functional dependence between the biophysical parameter and the spectral observation. One of the promising emerging MLRAs is Gaussian Processes (GP) regression [23]. It belongs to the family of the kernel methods and has been demonstrated as simpler and more robust than other MLRAs, such as neural networks, given an in situ training dataset [21,21]. GP regression has been successfully used for retrieval of chlorophyll content, leaf area index and fraction of vegetation cover [24,25].

The aim of this research is to apply and validate the GP regression method for leaf parameter retrieval from a wide range spectral dataset of structurally different species. The objectives thereby are (i) to find powerful estimates of Chl content, leaf water content, nitrogen content and specific leaf area that do not suffer from saturation effects, and (ii) to explain the physical meaning of the wavelengths chosen in the GP models.

2. Materials and methods

2.1. Data collection and leaf parameters

Leaves from four tree species, which substantially differ in leaf structure and pigment content, were collected and measured during two field campaigns in the period August 10–31 in 2011 in the city of Valencia. In total 320 green leaves were equally taken from the urban trees European nettle tree (*Celtis australis* L.), White mulberry (*Morus alba* L.), London plane (*Platanus x acerifolia* (Aiton) Willd.), and Canary Island date palm (*Phoenix canariensis* Chabaud) for simultaneous measurements of solar-induced reflectance, absorbance and fluorescence spectra and leaf parameters. The latter species has a equifacial leaf type, while the others possess a bifacial leaf type. A first part of the dataset originated from a campaign focusing on trees across the city exposed to either a low or high traffic intensity whereby about seven trees per species were sampled from the bottom canopy layer [17]. A second part of the dataset contained leaves taken from twelve trees, whereby three different canopy heights were sampled from three trees per species [26]. All trees were sun-exposed and grown at different locations within the urban environment (e.g., street side, park). During both field campaigns, branches from each sampling site were transported to the laboratory with their stems submerged in water. At the laboratory, an additional collection of 45 leaves for each species was sampled for measurement leaf thickness on images of microscopic cross sections [27] to illustrate structural differences among species. Total chlorophyll content (Chl $a + b$) was calculated based on readings from a chlorophyll content meter (SPAD-502Plus, Konica Minolta Optics Inc., Osaka, Japan) calibrated for the four species by a power fitting (Table 1) [28].

Specific leaf area (SLA) was determined as the ratio of fresh leaf area over dry leaf weight. The latter was determined after drying of leaves at 50–60 °C for five days in a drying oven. Leaf area was determined on the scanned leaf surface. Leaf water content (LWC) was calculated as the difference between fresh and dry leaf weight, divided by fresh leaf area. Each leaf's total nitrogen content (N) was measured using Kjeldahl method (Büchi 430 Kjeldahl Digester and Büchi 321 Kjeldahl Distiller, Büchi Labortechnik AG, Flawil, Switzerland).

2.2. Leaf reflectance, absorbance and Chl fluorescence

From each leaf, apparent reflectance, transmittance and steady-state Chl fluorescence (F) was measured under natural light conditions outside the laboratory using a spectroradiometer (FieldSpec 3 Hi-Res, Analytical Spectral devices (ASD) Inc., Boulder, USA) coupled with the FluoWat leaf clip (for details see [17,29]). Reflectance and transmittance are measured by inserting the fiber optic of the spectroradiometer respectively in an upper or lower leaf clip opening, without repositioning the leaf. Subsequently the leaf clip is positioned towards the sun with a 45° inclination between the sun's elevation and the leaf. Further, the leaf clip contains a short-pass filter cutting off light >650 nm, which can slide over the light opening in order to measure solar-induced Chl F in the

Table 1
Power functions and coefficients of determination (R^2) for the fit between Chl $a + b$ laboratory measurements ($\mu g\,cm^{-2}$) and SPAD measurements based on 80 leaves for each tree species (Adapted from [28]).

Species	Formula	R^2
Celtis australis	Chl $a + b = 0.02618 * SPAD^{2.052}$	0.92
Morus alba	Chl $a + b = 0.0001616 * SPAD^{3.472}$	0.90
Phoenix canariensis	Chl $a + b = 0.01818 * SPAD^{2.157}$	0.87
Platanus x acerifolia	Chl $a + b = 0.05325 * SPAD^{1.870}$	0.91