A ROBUST SUPERVISED METHOD TO ESTIMATE CHLOROPHYLL AB CONTENT FROM SPECTRAL REFLECTANCE.

Bikram Koirala, Paul Scheunders

Imec-Visionlab, University of Antwerp, Belgium

ABSTRACT

Leaf chlorophyll ab content is an important indicator of vegetation physiological status and is generally obtained from spectral reflectance. For non-destructive estimation of chlorophyll ab content, physical leaf reflectance models, such as the PROSPECT model and supervised methods have been applied. While the former generally does not perform optimal, the latter only performs well when trained on similar data. In this work, we developed a robust supervised method that overcomes this problem. The method derives a proxy for chlorophyll ab content as the relative position of a leaf reflectance spectrum on the arc spanned by the two extremes, containing high and low chlorophyll ab content. This proxy is found to be unaffected by spectral variability, caused by environmental and acquisition conditions. The relation between this proxy and the actual chlorophyll ab content is obtained by a supervised regression model, that is trained on a single leaf reflectance dataset, and that is transferable to other datasets. The proposed method is validated on seven real hyperspectral datasets.

1. INTRODUCTION

Retrieval of leaf biochemical parameters such as chlorophyll ab content (C_{ab}) is of great interest due to its direct connection with plant health and growth [1, 2, 3]. Although destructive techniques can accurately estimate C_{ab} [4], they are labor-intensive and expensive [1]. Non-destructive methods employ leaf spectral reflectance. The Normalized Difference Vegetation Index (NDVI) is the most popular method to estimate C_{ab} in a non-destructive manner [5]. This method uses two wavebands, one correlated with C_{ab} (red) and the other uncorrelated (near-infrared). To utilize more than two wavebands, shape indices have been developed [6]. Even though these methods are computationally not expensive, they cannot physically interpret the estimated parameters.

To describe the optical properties of plant leaves, several physical models have been developed [7]. In remote sensing applications, the PROSPECT model describes the leaf spectral reflectance as a function of a number of parameters, i.e., the specific absorption spectra of the biochemical parameters (e.g., C_{ab}) and a wavelength-dependent refractive index of the plant leaf [2]. The leaf biochemical parameters are estimated from the spectral reflectance by inverting the PROSPECT model.

On the other hand, several advanced machine learning regression techniques (MLRTs) have been developed to directly retrieve C_{ab} from spectral reflectance [3, 8]. These methods are supervised methods and require a training set of reflectance spectra and ground-truth information of C_{ab}. As the relationship between reflectance spectra and C_{ab} is nonlinear, nonlinear regression algorithms, such as Gaussian Process Regression (GP) have been utilized [9]. MLRTs were found to outperform the PROSPECT model, when trained and validated on the same dataset [10]. One particular problem with the MLRTs is that estimated values of the leaf parameters do not necessarily fall within the physical range of C_{ab}, and even can become negative [10]. To tackle this challenge, in [10], we developed a method that combines the physical interpretability of the PROSPECT model with the flexibility of the regression methods.

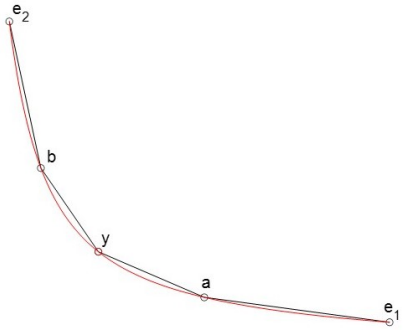
The performance of the supervised methodologies is found to be relatively poor when algorithms are trained and validated on independent datasets [10]. This is because they can not capture the intrinsic nonlinear relationship between the reflectance spectra and the target variable. In this work, we developed a robust supervised method to accurately estimate C_{ab} from spectral reflectance. We derive a proxy for C_{ab} as the relative position of a leaf reflectance spectrum on the arc spanned by the two extremes, containing high and low C_{ab}, and make it invariant to changes in environmental and acquisition conditions. The relation between this proxy and the actual C_{ab} can then be learned by a supervised regression method, that is trained on a single leaf reflectance dataset, and is transferable to other datasets. The proposed method will be validated experimentally on seven real hyperspectral datasets.

2. METHODOLOGY

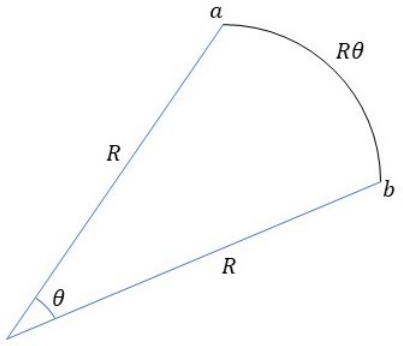
In this work, we develop a robust supervised method to estimate C_{ab} accurately from spectral reflectance, by deriving a representation (proxy) for C_{ab} that is invariant to environmental and acquisition conditions. We assume that the data manifold, sampled by a number of spectra with varying C_{ab} is a curve in spectral space between two extremes (endmem-

bers), i.e., spectra with low C_{ab} ($\mathbf{R}_L$) and high C_{ab} ($\mathbf{R}_H$). The proxy for C_{ab} can then be represented by the relative arc length of the plant leaf spectrum on the curve between $\mathbf{R}_H$ and $\mathbf{R}_L$.

The most obvious approach to approximate the arc length of a curve is by approximating the curve as a piece-wise linear curve and by summing up the Euclidean distances between neighboring samples (Fig. 1(a)). The more samples are available on the curve, the better the approximation [11].



(a)



(b)

Fig. 1: (a) Red: A curve connecting two endmembers $\mathbf{e}_1$ and $\mathbf{e}_2$; Black: the piece-wise linear approximation. Here $\mathbf{a}$, $\mathbf{y}$, and $\mathbf{b}$ denote data points on the curve. (b) The arc formed by connecting $\mathbf{a}$ and $\mathbf{b}$ on a hypersphere. Here R is the radius of the sphere, and $\theta = \arccos\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{||\mathbf{a}||||\mathbf{b}||}\right)$ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

In practical situations, only one spectrum of the plant leaf may be available, so that the piece-wise approximation for the arc length leads to large errors. On a hypersphere however, the arc length between two data points can be computed by multiplying the radius of the sphere (R) with the angle (θ) between them (Fig. 1(b)). To utilize this property of the hypersphere, we project all spectra onto the unit sphere ($R = 1$), by dividing each spectrum by its length ($\mathbf{b} \rightarrow \frac{\mathbf{b}}{||\mathbf{b}||}$). In this way, the relative arc length of any spectrum on the curve is obtained by just calculating the angles between the spectrum and $\mathbf{R}_H$ and $\mathbf{R}_L$.

However, after projection on the unit hypersphere, it is not guaranteed that all data points lie on the arc connecting the two endmembers. To correct for this, the following optimization is performed:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\|\mathbf{d}_E^{-1}\mathbf{d}_y - \mathbf{a}\right\|^2 s.t. \sum_{j=1}^{p} a_j = 1, \forall j : a_j \geq 0 \tag{1}$$

where $p = 2$ is the number of endmembers and $\mathbf{a}$ is the vector containing the 2 arc lengths, corrected to lie on the arc between the endmembers, and made relative (between 0 and 1) by dividing with the arc length between the endmembers. $\mathbf{d}_y = \begin{bmatrix}\arccos(\mathbf{R}_L^{\mathrm{T}}\mathbf{y}) \\ \arccos(\mathbf{R}_H^{\mathrm{T}}\mathbf{y})\end{bmatrix}$ is the vector containing the arc lengths between the spectrum of the plant leaf ($\mathbf{y}$) and endmembers, $\mathbf{d}_E = \begin{bmatrix}\arccos(\mathbf{R}_L^{\mathrm{T}}\mathbf{R}_L) & \arccos(\mathbf{R}_L^{\mathrm{T}}\mathbf{R}_H) \\ \arccos(\mathbf{R}_H^{\mathrm{T}}\mathbf{R}_L) & \arccos(\mathbf{R}_H^{\mathrm{T}}\mathbf{R}_H)\end{bmatrix}$ is the matrix containing the arc lengths between the endmembers, the operation T is the transpose operator. Besides the advantage that the arc lengths can be calculated for each spectrum of the plant leaf separately, any random scaling of the measured samples, due to external variability in e.g. illumination conditions is automatically resolved.

The estimated relative arc length is nonlinearly related to the true C_{ab}. To learn this relation, a supervised regression model is applied. In this work, we choose the Gaussian process as a regression model [9]. The estimated map of the relative arc lengths of N test samples ($\hat{\mathbf{A}}_t = \{\hat{\mathbf{a}}_i\}_{i=1}^N$) to their C_{ab} ($\mathbf{\Theta}_t = \{\hat{\theta}_i\}_{i=1}^N$) is given by:

$$\mathbf{\Theta}_t = \mathbf{\Theta}_{\mathcal{D}}(K(\hat{\mathbf{A}}_{\mathcal{D}}, \hat{\mathbf{A}}_{\mathcal{D}}) + \sigma_n^2\mathbf{I})^{-1}K(\hat{\mathbf{A}}_t, \hat{\mathbf{A}}_{\mathcal{D}})^T \tag{2}$$

Here, $\mathbf{\Theta}_{\mathcal{D}}$ is a row vector containing training C_{ab}, $K(\hat{\mathbf{A}}_{\mathcal{D}}, \hat{\mathbf{A}}_t)$ is the matrix of kernel functions between the training arc lengths ($\hat{\mathbf{A}}_{\mathcal{D}}$) and the test arc lengths, and $K(\hat{\mathbf{A}}_t, \hat{\mathbf{A}}_t)$ is the matrix of kernel functions between the test arc lengths. σ_n^2 is the noise variance of C_{ab} in the training set. The kernel function is computed by the following equation:

$$k(\hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j) = \sigma_f^2 \exp\left(-\sum_{b=1}^{p} \frac{\left(\hat{a}_i^b - \hat{a}_j^b\right)^2}{2l_b^2}\right) \tag{3}$$

where σ_f^2 is the variance of the input relative arc length, $p = 2$ is the number of endmembers and l_b is a characteristic length-scale for each endmember. The hyperparameters of this kernel function are optimized by minimizing the negative log marginal likelihood of the training dataset (-log($p(\mathbf{\Theta}_{\mathcal{D}}^T|\hat{\mathbf{A}}_{\mathcal{D}}^T)$)).

In the remaining of this work, we will refer to this method as the supervised method for estimating C_{ab} from spectral reflectance ($\mathrm{Cab_S}$).

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental set-up and evaluation statistics

The proposed method $\mathrm{Cab_S}$ is validated and compared to:

- the PROSPECT model,
- NDVI: a method that uses the spectral index NDVI as a proxy for C_{ab}, and learns a mapping between both,
- GP: a direct mapping between the reflectance spectrum and C_{ab} [3],
- GP_Linear: a methodology that combines PROSPECT with GP [10].

As both GP and GP_Linear utilize the Gaussian process as a regression algorithm, for a fair comparison, we selected the Gaussian process as a learning algorithm for both NDVI and Cab_S. To compute arc lengths, Cab_S requires endmembers. These two endmembers were manually selected from the ANGERS leaf optical properties database [12], as the spectra with the most extreme values of C_{ab}. The experiment was limited to the wavelength region 600-800 nm, because other biochemical parameters (e.g., carotenoid) do not have absorption features in that region, and thus will not influence the results. We did not observe a significant difference in the performance of the PROSPECT model when the reflectance values between 400-800 nm were utilized.

Quantitative comparisons are provided by the normalized root mean squared error (NRMSE) between the estimated C_{ab} ($\hat{\theta}$) and the ground truth C_{ab} (θ):

$$\text{NRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\theta}_i - \theta_i}{\max(\theta)}\right)^2} \times 100 \qquad (4)$$

All the methods are validated on the following seven hyperspectral datasets (see Table 1): The ANGERS dataset [12], and the Ecosis_cedarcreek, Ecosis_soybean_aphid and Yang_Pheno [13] datasets were acquired by utilizing a ASD field spectroradiometer while the Divittorio [14], LOPEX [15] and Ngee_arctic [16] datasets were acquired by using an Ocean Optics USB 2000 spectrometer, a Perkin Elmer Lambda 19 spectrophotometer, and a SVC HR-1024i field spectroradiometer respectively.

To obtain ground truth C_{ab}, chlorophyll pigment was extracted either using ethanol [12] or acetone [15],[14],[13],[16] by grinding fresh leaf disks in a chilled mortar. C_{ab} was then obtained by analyzing the absorption spectra of the solution (see [4] for more details) that were acquired by using a dual-beam scanning UV-Vis spectrophotometer. All datasets can be downloaded from the following link: `https://github.com/ashiklom/spectra_db`.

3.2. Results and discussion

We selected the ANGERS dataset to train NDVI, GP, GP_Linear, and Cab_S. The obtained models were then validated on the ANGERS, Divittorio, Ecosis_cedarcreek, Ecosis_soybean_aphid, LOPEX, Ngee_arctic and Yang_Pheno datasets. Results are shown in Table 2. As expected, all methods performed very well on the ANGERS dataset. The low error of GP on the ANGERS dataset suggests that it was able to fit the training dataset perfectly.

Table 1: Summary of the dataset. Here N and N_{GT} indicate the total number of reflectance spectra and the number of reflectance spectra with ground truth C_{ab} respectively.

Dataset	N	N_{GT}	Sensor
ANGERS [12]	276	276	asd-fs
Divittorio [14]	504	504	oo-2000
Ecosis_cedarcreek	831	831	asd-fs3
Ecosis_soybean_aphid	1131	1131	asd-fs4
LOPEX [15]	330	320	Pe-l19
Ngee_arctic [16]	615	104	SVC HR-1024i-fs
Yang_Pheno [13]	688	656	asd-fs3

Table 2: The results of different C_{ab} estimation techniques in terms of NRMSE (%). The best-performing technique is shown in bold.

Dataset	NDVI	PROSPECT	GP	GP_Linear	Cab_S
ANGERS	11.94	10.00	**3.50**	3.65	5.23
Divittorio	25.38	13.65	17.36	19.65	**12.20**
Ecosis_cedarcreek	32.04	48.41	38.62	40.66	**25.08**
Ecosis_soybean_aphid	57.16	55.89	42.61	41.85	**41.32**
LOPEX	25.07	30.96	18.53	18.70	**16.94**
Ngee_arctic	57.51	26.67	15.41	**12.76**	21.27
Yang_Pheno	35.46	88.36	43.41	34.32	**19.03**

GP_Linear was the best performer for estimating C_{ab} on the Ngee_arctic dataset. It is interesting to observe that the PROSPECT model outperformed NDVI for estimating C_{ab} in four among the seven datasets. Although the performance of NDVI is low in most of the cases, the advantage of this method is that it is invariant to the environmental conditions (e.g., varying illumination conditions). This is especially beneficial when spectra are acquired outdoors. The proposed method performed the most consistent and the best on five of the seven datasets. On the Ecosis_soybean_aphid dataset, none of the techniques performed well. This should be investigated in more detail.

4. CONCLUSIONS

In this paper, we proposed a data-driven approach to accurately estimate the chlorophyll ab content from spectral reflectance. The method derives the relative position of the reflectance spectrum acquired from the plant leaf on the arc connecting two endmembers as a proxy for the chlorophyll ab content. Similar to NDVI, this method is invariant to environmental conditions. The chlorophyll ab content can then be estimated by applying a supervised regression model. The approach was validated and compared to a number of methods from the literature on a large number of leaf reflectance datasets. The experimental results show that this method is

very promising.

Acknowledgement

The research presented in this paper is funded by the Research Foundation-Flanders - project G031921N.

5. REFERENCES

[1] Jin Xu, Timothy A. Volk, Lindi J. Quackenbush, and Stephen V. Stehman, "Estimation of shrub willow leaf chlorophyll concentration across different growth stages using a hand-held chlorophyll meter to monitor plant health and production," *Biomass and Bioenergy*, vol. 150, pp. 106132, 2021.

[2] J.-B. Féret, A.A. Gitelson, S.D. Noble, and S. Jacquemoud, "Prospect-d: Towards modeling leaf optical properties through a complete lifecycle," *Remote Sensing of Environment*, vol. 193, pp. 204 – 215, 2017.

[3] Shari Van Wittenberghe, Jochem Verrelst, Juan Pablo Rivera, Luis Alonso, José Moreno, and Roeland Samson, "Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset," *Journal of Photochemistry and Photobiology B: Biology*, vol. 134, pp. 37 – 48, 2014.

[4] Hartmut K. Lichtenthaler, "[34] chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," in *Plant Cell Membranes*, vol. 148 of *Methods in Enzymology*, pp. 350 – 382. Academic Press, 1987.

[5] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the Great Plains with ERTS.," in *Proceedings of the Third ERTS Symposium*. NASA, Dec. 1973, vol. 1, pp. 309–317.

[6] Jochem Verrelst, Gustau Camps-Valls, Jordi Muñoz-Marí, Juan Pablo Rivera, Frank Veroustraete, Jan G.P.W. Clevers, and José Moreno, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 108, pp. 273 – 290, 2015.

[7] Stéphane Jacquemoud and Susan L. Ustin, "Leaf optical properties: A state of the art," 2000.

[8] Bo Liu, Yue-Min Yue, Ru Li, Wen-Jing Shen, and Ke-Lin Wang, "Plant leaf chlorophyll content retrieval based on a field imaging spectroscopy system," *Sensors*, vol. 14, no. 10, pp. 19910–19925, 2014.

[9] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, New York, 2006.

[10] Bikram Koirala, Zohreh Zahiri, and Paul Scheunders, "A machine learning framework for estimating leaf biochemical parameters from its spectral reflectance and transmission measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7393–7405, 2020.

[11] Bikram Koirala, Zohreh Zahiri, Alfredo Lamberti, and Paul Scheunders, "Robust supervised method for nonlinear spectral unmixing accounting for endmember variability," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[12] Jean-Baptiste Feret, Christophe François, Gregory P. Asner, Anatoly A. Gitelson, Roberta E. Martin, Luc P.R. Bidel, Susan L. Ustin, Guerric le Maire, and Stéphane Jacquemoud, "Prospect-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 3030 – 3043, 2008.

[13] Xi Yang, Jianwu Tang, John F. Mustard, Jin Wu, Kaiguang Zhao, Shawn Serbin, and Jung-Eun Lee, "Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests," *Remote Sensing of Environment*, vol. 179, pp. 1–12, 2016.

[14] Alan V. Di Vittorio, "Enhancing a leaf radiative transfer model to estimate concentrations and in vivo specific absorption coefficients of total carotenoids and chlorophylls a and b from single-needle reflectance and transmittance," *Remote Sensing of Environment*, vol. 113, no. 9, pp. 1948–1966, 2009.

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

[16] Shawn Serbin, Wil Lieberman-Cribbin, Kim Ely, and Alistair Rogers, "Ngee arctic leaf spectral reflectance and transmittance, barrow, alaska, 2014-2016," 2 2019.