LEAF WATER THICKNESS ESTIMATION FROM SPECTRAL REFLECTANCE

Ketaki Vinay Jambhali, Bikram Koirala, Paul Scheunders

Imec-Visionlab,Department of Physics, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp

ABSTRACT

Leaf Water Thickness, often described as equivalent water thickness, is an important trait in plant physiology and can be estimated non-destructively using the leaf's spectral reflectance. Leaf water thickness is defined as the depth of the water layer if all the water contained in a plant leaf were uniformly distributed over a given area. In this study, we propose a supervised machine learning approach based on polynomial regression, combined with our method, NRAL, to estimate leaf water thickness. The results are compared against those obtained using the PROSPECT model, a well-established physical model for simulating leaf optical properties. This methodology is evaluated on the LOPEX dataset, along with three additional hyperspectral datasets. In all cases, the proposed approach achieves a lower root mean squared error (RMSE) compared to the PROSPECT model. These findings highlight the potential of hybrid machine learning methods to outperform traditional models in accurately estimating leaf biophysical parameters.

Index Terms— Leaf water thickness, machine learning, polynomial regression, PROSPECT

1. INTRODUCTION

The biochemical parameters of the leaves are important indicators of the physiological status of a plant and play a key role in the determination of overall plant health and crop yield. These parameters include leaf mass per area, chlorophyll content, leaf water thickness (LWT), etc. These traits provide valuable insights into plant growth, photosynthetic activity, nutrient status, and stress responses.

Because the interaction of light with plant leaves depends on their biochemical content, spectral reflectance across various wavelengths, particularly in the visible, near-infrared, and shortwave infrared (SWIR) regions, provides a non-destructive and scalable means of estimating these parameters. Methods for retrieving leaf biochemical parameters from spectral data include empirical models based on vegetation indices (e.g., NDVI [1]), radiative transfer models (e.g., PROSPECT [2]), and, increasingly, machine learning techniques trained on field or laboratory reference data ([3]). These approaches enable high-quality monitoring of plant traits, supporting applications in precision agriculture, ecosystem modeling, and environmental monitoring. The PROSPECT model is a simple yet effective radiative transfer model, representing an improved version of Allen's generalized plate model. It views a leaf as a stack of 'N layers' of plant material, separated by 'N-1' air spaces. It predicts how leaves reflect and transmit light based on their biochemical and structural properties across the visible to SWIR wavelength regions (400 - 2500 nm). It uses key biochemical parameters such as chlorophyll content, carotenoid content, anthocyanin content, equivalent water thickness, and leaf mass per area to describe the optical properties of plant leaves [4]. To estimate leaf biochemical parameters from measured spectral reflectance, the PROSPECT model can be inverted by minimizing the reconstruction error between the measured spectra and those simulated by the model.

The PROSPECT is widely used in remote sensing of vegetation and the interpretation of hyperspectral data.

Conversely, several advanced machine learning regression techniques have been developed to directly estimate leaf biochemical parameters from spectral reflectance data (e.g., [3]). These are supervised methods that require a training dataset consisting of reflectance spectra paired with ground-truth measurements of leaf biochemical parameters. These techniques often map the measured reflectance spectra directly to leaf parameters, bypassing physics. The values of leaf parameters estimated by these techniques do not always lie within the physically plausible range and can even be negative. To address this challenge, in [5], we proposed a method that integrates the physical interpretability of the PROSPECT model with the flexibility of machine learning regression techniques. In contrast, in [6], we introduced a fully data-driven approach for accurately estimating chlorophyll content from reflected light. This method ensures that the estimated parameters remain within physically plausible bounds by representing chlorophyll content as the relative position of a leaf reflectance spectrum along an arc defined by two spectral extremes, corresponding to high and low chlorophyll concentrations.

Inspired by the methodology proposed in [6], in this work, we propose a method to accurately estimate LWT by analyzing the reflected light. In contrast to chlorophyll content whose absorption features are dominant only between 600-700 nm, water is highly dominant throughout the SWIR wavelength regions (1000-2500 nm). In that case, the re-

flectance of leaves with different LWT forms a curve between the two spectra with the most extreme LWT values. The relative position of a spectrum on this curve is then a proxy for the actual LWT.

The relationship between this proxy and the actual LWT can then be learned using a polynomial regression model. Such model can e.g., be trained on a single leaf reflectance dataset, with the ability to generalize across other datasets. The proposed method will be experimentally validated using four leaf optical property datasets that contain spectral reflectance and ground truth LWT measurements for different plant species.

2. DATASETS

In this study, four databases are used.

2.1. LOPEX

The Lopex93 data set (Leaf Optical Properties Experiment 1993) is a collection of optical and biochemical measurements of leaves, collected in the study area of Ispra, Italy, compiled by the Joint Research Center (JRC) of the European Commission in 1993. It has approximately 330 leaf samples (66 plants) of 45 different species. It contains spectral reflectance (400 to 2500 nm) and LWT. This dataset is widely used for validating radiative transfer models like PROSPECT [7].

2.2. ANGERS

The ANGERS Leaf Optical Properties Database was collected in 2003 in Angers, France. It contains optical and biochemical parameters of 276 leaf samples from 43 plant species. It contains reflectance from 400 to 2450 nm along with LWT [8].

2.3. ACCP

The ACCP data set was collected between 1992 and 1993 in San Jose, California. This data set is smaller, with a total of 34 samples from 18 species. It contains reflectances from 400 to 2500 nm, along with LWT.

2.4. NASA

The NASA dataset was collected in 2009 in the north-central and northeast of the United States [9]. It contains a total of 987 samples from 64 different species. The dataset contains spectral reflectance in the range 350-2500 nm, and the corresponding LWT for each sample.

Table 1 summarizes the datasets used.

3. METHODOLOGY

The proposed approach integrates the strengths of Normalized Relative Arc Length (NRAL) and the power of polynomial regression to accurately estimate LWT from spectral reflectance data.

3.1. NRAL

NRAL is a method that was previously designed for estimating fractional abundances of binary mixtures [10], and adapted for the estimation of moisture content [6] and water content in porous materials [11].

The core assumption of this method is that the data manifold, formed by numerous samples with a single varying parameter, traces a continuous curve connecting the spectral reflectances of samples with the minimal and maximal parameter values, referred to as endmembers. The relative position of a sample along this curve serves as a proxy for its corresponding parameter value. A supervised regression model is then used to learn the relationship between this proxy and the actual parameter.

NRAL addresses two key challenges in accurately estimating the relative position of a sample:
(a) Spectral variability: Variations in illumination and acquisition conditions often introduce scaling effects in the measured spectra. NRAL mitigates this issue by projecting each measured spectrum onto the unit hypersphere:

$$\mathbf{y} \rightarrow \frac{\mathbf{y}}{\|\mathbf{y}\|} \tag{1}$$

b) Inaccurate arc length estimation: Arc length is commonly approximated by dividing the curve into piecewise linear segments and summing the Euclidean distances between intermediate points. However, when only endmembers and a single spectrum from the sample are available, this approximation becomes unreliable. NRAL overcomes this limitation by employing the spherical law of cosines (see [6] for details), allowing for accurate arc length estimation and, consequently, precise computation of the sample's relative position.

3.2. Estimating LWT

Unlike fractional abundances, which are constrained between 0 and 1, LWT can vary across a broader physical range, from a minimum to a maximum thickness. To address this, the relative position estimated by NRAL must be calibrated. This calibration is performed using the following equation:

$$\text{CNRAL}(\mathbf{y}) = \frac{b_2}{b_1 + b_2}\text{LWT}(\mathbf{R}_{\text{min}}) + \frac{b_1}{b_1 + b_2}\text{LWT}(\mathbf{R}_{\text{max}}) \tag{2}$$

where b_1 and b_2 denote the true arc lengths between the sample and the endmembers. $\mathbf{R}_{\text{min}}$ and $\mathbf{R}_{\text{max}}$ denote the spec-

	LOPEX	ANGERS	ACCP	NASA
Date	1993	2003	1992-1993	2009
Number of Samples	330	276	34	987
Number of Species	45	43	18	64
Spectrophotometer/Spectroradiometer	Perkin Elmer Lambda 19	ASD FieldSpec	NIRS 6500	ASD FieldSpec3
Spectral Range	400–2500 nm	400–2450 nm	400 –2500 nm	350–2500 nm

Table 1. Specification of the different datasets.

tral reflectances corresponding to the minimum and maximum LWT endmembers, respectively. The function CNRAL $(\cdot)$ denotes the calibrated NRAL of the sample.

Due to the highly nonlinear nature of the curve connecting the two endmembers, the parameters obtained from Eq. (2) must be further mapped to the actual LWT values. Since this is a fairly simple one-dimensional mapping, a simple polynomial regression ($\hat{\text{LWT}}(\mathbf{y}) = a_0 + a_1\text{CNRAL}(\mathbf{y}) + a_2\text{CNRAL}^2(\mathbf{y})$) with a limited number of training samples will do the job.

To learn the parameters of the polynomial regression model (a_0, a_1, and a_2), the LOPEX dataset was used in this study. From this dataset, the two samples with the lowest and highest ground-truth LWT values were identified and designated as endmembers. The core assumption of the proposed method is that all other samples, representing varying LWT values, lie within the range defined by these two extremes. If this assumption is violated ($\text{LWT}(\mathbf{y}) < \text{LWT}(\mathbf{R}_{\text{min}})$ or $\text{LWT}(\mathbf{y}) > \text{LWT}(\mathbf{R}_{\text{max}})$), the proposed method will project the data onto one of the extremes.

4. EXPERIMENTS AND RESULTS

As described in the methodology section, the proposed approach uses the LOPEX dataset to learn the parameters of the polynomial regression model. The endmembers, corresponding to the lowest and highest LWT values in the LOPEX dataset, were identified, with LWT values of 20.98 μm and 524.88 μm, respectively. Using these two endmembers, the relative positions of all samples in all the datasets were computed. These relative positions, estimated by NRAL, were then calibrated using Eq. (2).

In the final step, the nonlinear relationship between the calibrated relative positions and the actual LWT was learned using polynomial regression. The regression model was trained using the LOPEX dataset as ground truth. To assess the complexity of the problem, we conducted two types of experiments. In the first experiment, all 320 samples from the LOPEX dataset were used to estimate the parameters of the polynomial regression model. In the second experiment, only 15 samples were used for training. This comparison aimed to evaluate the stability and performance of the polynomial fit under limited data conditions and to determine whether a small number of representative samples could still produce an accurate estimation model. Results showed that the model trained with only 15 samples exhibited less than a 1% deviation in RMSE compared to the model trained on the full dataset.

Using the polynomial regression model, the LWT is then estimated for all the datasets. Validation of the model is obtained by the RMSE of the estimation:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2} \tag{3}$$

where θ_i denote the ground truth LWT, $\hat{\theta}_i$ the estimated LWT, and n the number of samples.

The results using both the fully trained model and the reduced model are shown in Table 2. For comparison, the results of the PROSPECT model are shown as well. It can be observed that the proposed approach outperforms the PROSPECT model on all 4 datasets. Fig. 1 shows the scatterplot of the estimated versus ground truth LWT values on all four datasets when using the reduced regression model.

5. CONCLUSION

In this work, we proposed a data-driven supervised approach to estimate LWT from spectral reflectance. The method was validated and compared to the PROSPECT model on four datasets. It was observed that overall the proposed approach leads to better RMSE scores than the PROSPECT model, even with a very low training data size.

Acknowledgement

The research presented in this paper is funded by the Research Foundation-Flanders (project G031921N). Ketaki Vinay Jambhali is a grant holder supported by the Research Fund of the University of Antwerp. Bikram Koirala is a postdoctoral fellow of the Research Foundation Flanders, Belgium (FWO: 1250824N-7028)

	PROSPECT	NRAL - Training Sample Size 320	NRAL - Training Sample Size 15
Lopex	4.47	3.77	3.80
Anger	7.15	6.24	6.35
ACCP	13.11	12.99	12.87
Nasa	5.09	4.71	4.78

Table 2. Performance of LWT estimation in terms of RMSE for the different datasets.

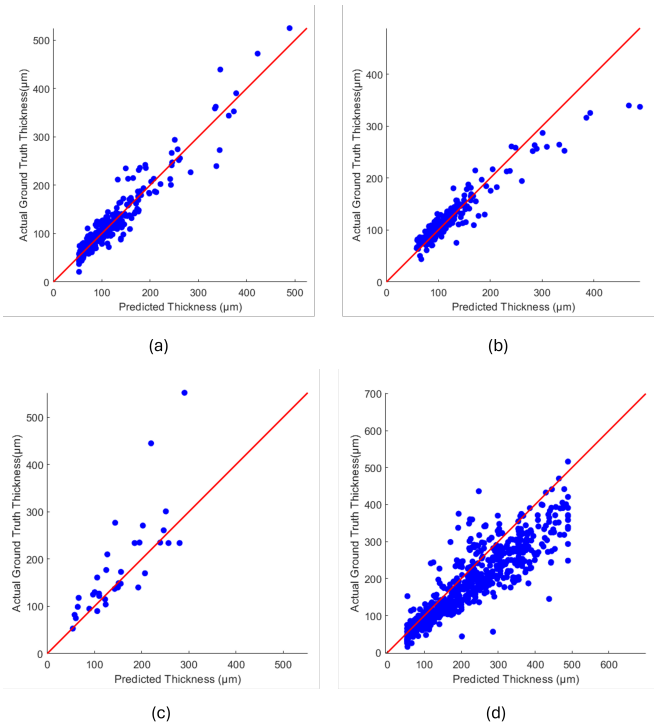

Fig. 1. Ground truth vs estimated Leaf Water Thickness for datasets (a) Lopex (b) Anger (c) ACCP and (d) Nasa.

6. REFERENCES

[1] 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.

[2] S. Jacquemoud and F. Baret, "Prospect: A model of leaf optical properties spectra," *Remote Sensing of Environment*, vol. 34, no. 2, pp. 75 – 91, 1990.

[3] S. Van Wittenberghe, J. Verrelst, J.P. Rivera, L. Alonso, J. Moreno, and R. 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] S. Jacquemoud and F. Baret, "Prospect: A model of leaf optical properties spectra," *Remote Sensing of Environment*, vol. 34, no. 2, pp. 75–91, 1990.

[5] B. Koirala, Z. Zahiri, and P. 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.

[6] B. Koirala and P. Scheunders, "A robust supervised method to estimate chlorophyll ab content from spectral reflectance," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 326–329.

[7] S. Jacquemoud, S.L. Ustin, J. Verdebout, G. Schmuck, G. Andreoli, and B. Hosgood, "Estimating leaf biochemistry using the prospect leaf optical properties model," *Remote Sensing of Environment*, vol. 56, no. 3, pp. 194–202, 1996.

[8] J.B. Feret, C. François, G.P. Asner, A. Gitelson, R.E. Martin, L.P.R. Bidel, S.L. Ustin, G. le Maire, and S. 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, June 2008.

[9] S.P. Serbin, A. Singh, B.E. McNeil, C.C. Kingdon, and P.A. Townsend, "Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species," *Ecological Applications*, vol. 24, no. 7, pp. 1651–1669, 2014.

[10] B. Koirala, B. Rasti, and P. Scheunders, "A supervised approach for estimating fractional abundances of binary intimate mixtures," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 8956–8966, 2024.

[11] B. Koirala and P. Scheunders, "An efficient method for water content estimation of building materials from spectral reflectance," *NDT & E International*, vol. 147, OCT 2024.