

i - ϕ -MaLe: A novel hybrid machine learning phasor-based approach to retrieve a full-set of solar-induced fluorescence metrics and biophysical parameters

R. SCODELLARO⁽¹⁾, I. CESANA⁽²⁾, L. D'ALFONSO⁽¹⁾, M. BOUZIN⁽¹⁾, M. COLLINI⁽¹⁾,
F. MIGLIETTA⁽³⁾, M. CELESTI⁽⁴⁾, D. SCHUETTEMAYER⁽⁵⁾, R. COLOMBO⁽²⁾,
G. CHIRICO⁽¹⁾, S. COGLIATI⁽²⁾ and L. SIRONI⁽¹⁾(*)

⁽¹⁾ *Laboratory of Advanced Bio-spectroscopy, Physics Department "G. Occhialini",
University of Milano-Bicocca - Piazza della Scienza 3, 20126 Milano, Italy*

⁽²⁾ *Remote Sensing of Environmental Dynamics Lab., DISAT, University of Milano-Bicocca
Piazza della Scienza 1, 20126 Milano, Italy*

⁽³⁾ *Institute of Bioeconomy (IBE), National Research Council (CNR) - Via Caproni 8, 50145
Florence, Italy*

⁽⁴⁾ *HE Space for ESA, ESTEC - Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands*

⁽⁵⁾ *ESA-ESTEC - Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands*

received 31 January 2023

Summary. — Solar-induced fluorescence (F) is crucial to monitor vegetation health, as it provides information about photosynthetic processes. Our new method, i - ϕ -MaLe, simultaneously estimates F spectra, Leaf Area Index (LAI), Chlorophyll Content (Cab), Absorbed Photosynthetic Active Radiation (APAR) and F Quantum Yield (F_{qe}) from canopy reflectance spectra by coupling the phasor approach with Machine Learning (ML) techniques. We validated i - ϕ -MaLe on simulations and spectra acquired for increasing spectrometer-canopy distances, up to 100 m (where O_2 bands are affected by atmospheric oxygen absorption). The reliability of i - ϕ -MaLe in such complex experimental scenarios paves the way to new perspectives concerning the real time monitoring of vegetation stress status on high scales.

1. – Introduction

Fluorescence (F), Photochemical (photosynthesis) and Non-Photochemical Quenching (heat) are the main paths exploited by vegetation to dissipate the absorbed energy. Consequently, the detection of F plays a pivotal role in Remote Sensing to investigate the terrestrial vegetation physiology. Actually, F is detected through spectrometers at

(*) Corresponding author. E-mail: laura.sironi@unimib.it

different spatial scales, from top of canopy (TOC) to satellite distances [1]: then, specialized algorithms [2] disentangle the acquired reflectance spectra R_{app} and the weak F signal. Unfortunately, these approaches analyze only high resolution spectra and, for high canopy-spectrometer distances (>100 m), reabsorption effects due to the telluric atmospheric oxygen strongly affects the R_{app} shape in the O_2 bands (O_2A , 760 nm and O_2B , 687 nm), requiring complex atmospheric correction models. These drawbacks are addressed by $i\text{-}\phi\text{-MaLe}$, hybrid Phasor-ML based approach, aimed to retrieve the F spectrum, the F_{RC} spectrum, obtained by applying the leaf/canopy Reabsorption Correction (RC) to F , and canopy biophysical parameters (LAI, Cab, APAR, Fqe) [3]. $i\text{-}\phi\text{-MaLe}$ compares the acquired R_{app} shape with a wide set of Radiative Transfer (RT) simulations obtained by coupling SCOPE and MODTRAN5 models [2] to detect the function which best matches with the experimental data. Then, the method associates the parameters characterizing the simulated function to the analyzed spectrum.

2. – Materials and methods

i-phi-MaLe algorithm. – $i\text{-}\phi\text{-MaLe}$ extends the standard phasor approach, widely used in biophysics [3-5], by analyzing consecutive fixed spectral windows. The method is composed of a training and a retrieving procedure, extensively discussed in [6]. Briefly, during the training phase, a dataset of simulated R_{app} spectra is generated by the SCOPE+MODTRAN5 RT model [2] by tuning the parameters related to different illumination conditions (*i.e.*, Solar Zenith Angle —SZA) or canopy growth stages (*i.e.*, LAI, Cab, Fqe) to cover a wide range of possible experimental scenarios. Each R_{app} is separated in N equally spaced spectral windows, where the Discrete Fourier Transform (DFT) is applied to obtain N points lying on N different phasor planes [6]. Each point is associated to the F , F_{RC} , Cab, LAI, APAR and Fqe characterizing R_{app} . The retrieval phase aims to estimate F , F_{RC} and the biophysical parameters from an experimental R_{app} spectrum, separated in the same windows exploited for the training set spectra. Since the signal to noise ratio (SNR) affects the points distributions by scattering their coordinates from their ideal position, $i\text{-}\phi\text{-MaLe}$ considers a circle of radius r , centered around the test point, scaling inversely with the SNR for each phasor plot. The Occurrence Occ and the Distance score D are associated to each training set R_{app} :

$$(1) \quad Occ = \sum_{i=1}^N \alpha_i, \quad D = \sum_{i=1}^N \frac{r}{d_i} \alpha_i, \quad \text{where } \alpha_i = 1 \text{ if } d_i \leq r, \quad 0 \text{ if } d_i > r,$$

where d_i is the training-test points Euclidean distance and N is the number of considered spectral windows. Finally, 3 configurations are set to retrieve the parameters: ALG1 considers the entire set of spectral windows, associating to the test spectrum tuple $P_T = [LAI, Cab, APAR, Fqe, F, F_{RC}]$ corresponding to the training spectrum with the highest Occ score. In the case of a tie, the spectrum with the highest D score is selected among those with the same Occ . If a tie is reached also for D , $i\text{-}\phi\text{-MaLe}$ computes their mean value; ALG2 retrieves each parameter independently by applying ALG1 only in the spectral windows characterized by the highest information content, as determined in [6]; ALG3 retrieves the parameter tuple P_T from spectra acquired at tower level by excluding the O_2 bands, affected by reabsorption phenomena.

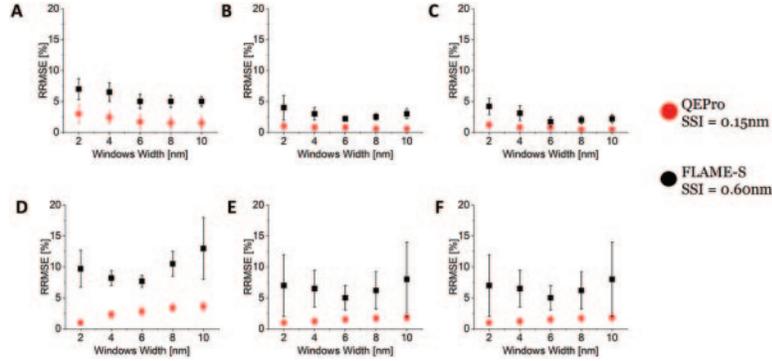


Fig. 1. – $i\text{-}\phi\text{-MaLe}$ retrieving accuracy in terms of Relative Root Mean Square Error (RRMSE) of LAI (A), Cab (B), APAR (C), F_{qe} (D), F (E) and F_{RC} (F) in dependence of the windows width for QEPro (red) and FLAME-S (black) SSI values.

3. – Results

$i\text{-}\phi\text{-MaLe}$ performances have been extensively quantified in [6] by varying 1) the spectral windows width, 2) the superposition between consecutive windows 3) the amount of noise affecting the analyzed spectra and 4) the training dataset dimensions. Here, we investigated two experimental scenarios, characterized by different acquisition spatial scales: TOC (forage and alfalfa crops in Italy) and tower-level (Downy oak forest in France). In both cases, we exploited the FLoX setup to acquire the R_{app} spectra during clear sky days. FLoX is equipped with the QEPro (Ocean Insight, US - Full Width Half Maximum (FWHM) of 0.30 nm, SNR of 1000, spectral sampling interval (SSI) of 0.15 nm) and the FLAME-S (FWHM of 1.7 nm, SNR of 250 SSI of 0.6 nm) spectrometers. In fig. 1, we investigated the effects of both spectrometers on the $i\text{-}\phi\text{-MaLe}$ retrieval by comparing the accuracy related to the SSIs characterizing QEPro and FLAME-S in dependence of the spectral width used for DFT computation. For low SSIs, accuracy slightly decreases, in particular for F_{qe} , F and F_{RC} (figs. 1(D)–(F)). Interestingly, the ratio between the

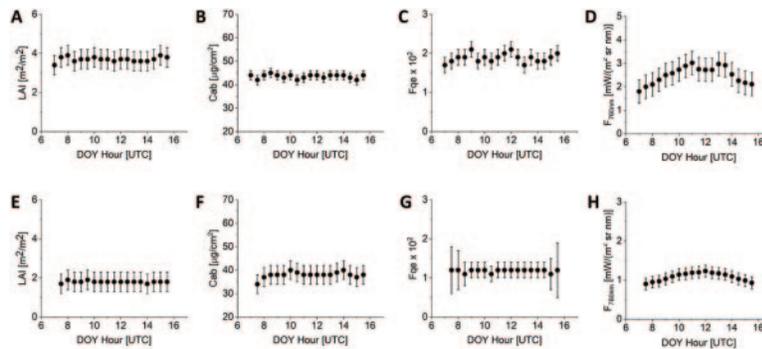


Fig. 2. – Diurnal trends of LAI ((A), (E)), Cab ((B), (F)) F_{qe} ((C), (G)) and F at 760 nm ((D), (H)) retrieved by $i\text{-}\phi\text{-MaLe}$ in top of canopy ((A)–(D), DOY 181) and tower-level ((E)–(H), DOY 231) conditions.

best window width and the SSI is similar between QEPro (13) and FLAME-S (10), suggesting that each spectral window is supposed to contain ~ 10 experimental points to achieve the best results. Figure 2 provides the daily (7:00–15:30 UTC) trends of biophysical and F parameters characterizing the spectra acquired in both TOC and tower level scenarios, analyzed by applying ALG2 and ALG3, respectively. As expected, LAI and Cab are almost constant and the stability of Fqe (figs. 2(C), (G)) in the absence of stress highlights the $i\text{-}\phi\text{-MaLe}$ reliability [6]. The parabolic trend of F (figs. 2(D), (H)) is related to the photosynthetic activity. A midday depression is visible at 11.00 UTC (solar noon) for F retrieved at TOC level (fig. 2(D)). The results in fig. 2(H) demonstrate that $i\text{-}\phi\text{-MaLe}$ provides meaningful (*i.e.*, positive) values without atmospheric correction models, in contrast to already developed algorithms. Even if ALG3 excludes the O_2 bands, which contain valuable information concerning fluorescence [1, 6], F and Fqe are retrieved with a reasonable accuracy (fig. 2(G), (H)). Biophysical parameters are extracted with almost the same error for both acquisition scales, since ALG2 and ALG3 exploit the same spectral windows. Higher uncertainties affect Fqe around 7:30 and 15:00 UTC (fig. 2(G)), since the SCOPE model is not accurate for high SZA values.

4. – Discussion and conclusion

$i\text{-}\phi\text{-MaLe}$ is the first algorithm able to simultaneously retrieve different biophysical parameters and full F and F_{RC} spectra. Indeed, other methods exploiting ML algorithms to analyze spectral data from vegetation retrieve few parameters [7, 8]. Moreover, $i\text{-}\phi\text{-MaLe}$ performs reliable estimates for low SSI values (fig. 1) and assures the best accuracy, since each considered parameter is retrieved on specific spectral windows endowed with the highest information content. This property allows avoiding the effect of the oxygen bands when relevant for data analysis. Indeed, $i\text{-}\phi\text{-MaLe}$ retrieves F also in scenarios where the O_2 absorption affects R_{app} measurements (fig. 2), differently from state of the art methods, totally dependent on the O_2 bands information content. Even if here we limited our analysis to few parameters tunable in the SCOPE model, our work can be extended by exploiting other canopy configurations and more sophisticated 3D RT models [9]. Further analysis will evaluate the compatibility between experimental data and different RT models at multiple spatial scales.

REFERENCES

- [1] MOHAMMED G. H. *et al.*, *Remote Sens. Environ.*, **231** (2019) 111177.
- [2] COGLIATI S. *et al.*, *Remote Sens.*, **11** (2019) 1840.
- [3] RADAELLI F. *et al.*, *Sci. Rep.*, **8** (2018) 6314.
- [4] SCODELLARO R. *et al.*, *Front. Oncol.*, **9** (2019) 527.
- [5] SCODELLARO R. *et al.*, *Nuovo Cimento C*, **44** (2021) 139.
- [6] SCODELLARO R. *et al.*, *Remote Sens. Environ.*, **280** (2022) 113196.
- [7] VERRELST Z. *et al.*, *Surv. Geophys.*, **40** (2019) 629.
- [8] BHADRA S. *et al.*, *Remote Sens.*, **12** (2020) 2082.
- [9] MALENOVSKY Z. *et al.*, *Remote Sens. Environ.*, **263** (2021) 112564.