RETRIEVAL OF MAIZE BIOPHYSICAL VARIABLES FROM MULTISPECTRAL AND HYPERSPECTRAL EO DATA USING A HYBRID APPROACH

M. Ranghetti [1]*, M. Boschetti [1], M. Gianinetto [2,1], G. Tagliabue [3], C. Panigada[3], G. Candiani[1]

[1] IREA, National Research Council, Via Bassini 15, 20133 Milano, Italy – (ranghetti.m, boschetti.m, candiani.g)@irea.cnr.it
[2] Department of Architecture, Built Environment and Construction Engineering (DABC), Politecnico di Milano, Via Ponzio 31, 20133 Milano, Italy – marco.gianinetto@polimi.it
[3] Remote Sensing of Environmental Dynamics Lab., Dip. Scienze dell'Ambiente e del Territorio, Università Milano-Bicocca, 20126 Milano, Italy – (giulia.tagliabue, cinzia.panigada)@unimib.it

KEY WORDS: Precision Farming, Radiative Transfer Modelling, Machine learning, BV Estimation, Hybrid Approach

ABSTRACT:

Assessing crops health and status is becoming relevant to support farmers' decisions and actions for a sustainable agriculture. The use of remote sensing techniques in agriculture has become widely popular during the past years. Earth Observing (EO) data can greatly contribute to constantly monitor crops phenology and to estimate important vegetation biophysical parameters.
This work presents a hybrid approach, which exploits the PROSAIL-PRO model and Machine Learning (ML) algorithms, to estimate maize biophysical variables, such as Canopy Chlorophyll Content (CCC) and Leaf Area Index (LAI). The test site is represented by two maize fields located near Grosseto (Tuscany, IT), where two field campaigns were carried out in July 2018. During the same period, the airborne sensor Hyplant-DUAL acquired two images of the test site. These images were used to simulate PRISMA and Sentinel-2 data in order to investigate the difference of the retrieval performance between hyperspectral and multispectral EO data. Results show similar performance between Sentinel-2 and PRISMA. The ML algorithms, providing the best performance (GPR and NN) within the hybrid framework, were then applied to actual Sentinel-2 images. The retrieval results for LAI and CCC were compared to estimations assessed through the ESA S2Toolbox. The comparison showed that the proposed method provides better results than those achieved through S2Toolbox, for both LAI (R^2 = 0.85 and MAE = 0.39; S2Toolbox: R^2 = 0.35 and MAE = 0.87) and CCC (R^2 = 0.73 and MAE = 0.20; S2Toolbox: R^2 = 0.29 and MAE = 0.68).

1. INTRODUCTION

Agricultural practices can lead to different environmental threats, such as water consumption, biodiversity loss, pollutants' leaching and emissions. The assessment of crops health and status can support farmers' decisions and actions for a sustainable agriculture. This is the context where precision farming becomes relevant.
Precision farming is a management strategy used to "apply the right treatment in the right place at the right time" (Gebbers & Adamchuk, 2010). It is an important component of smart farming, which aims at an information-driven optimization of all aspects of a farming system (Bach et al., 2016).
Precision farming thus holds the potential for increasing yields on limited land while saving resources and preventing environmental pollution (Plant et al., 2000).
In this context, assessing vegetation status and health through Earth Observation (EO) data become relevant for the determination of some important biophysical vegetation variables (BVs).
Different methods can be found in literature for BVs retrieval such as parametric and nonparametric, linear and nonlinear regression methods, as well as physically-based methods using radiative transfer models (RTMs) and, more recently, hybrid approaches.
During the last decades, simple empirical parametric models were employed for the retrieval of biophysical parameters by using narrowband vegetation indices. Since few regions of the electromagnetic spectrum are exploited by using vegetation indices, a limitation in the retrieval of specific vegetation parameters may arise.
To overcome this situation, nonparametric methods could be employed. In general, nonparametric methods seek to best fit the training data, whilst maintaining some ability to generalize the unseen data. Nonlinear nonparametric methods are also known as machine learning regression algorithms (MLRAs).
Physically-based methods represent an alternative to the regression methods. They use RTMs for simulating the reflectance of vegetation measured by an EO sensor.
Hybrid methods are increasingly used in literature (Berger, Verrelst, Féret, Hank, et al., 2020; Berger, Verrelst, Féret, Wang, et al., 2020; Verrelst et al., 2020). In this approach, RTM are used in combination with MLRAs. Thus, hybrid methods have the transferability guaranteed by the use of a physically-based method and the computationally efficiency and flexibility provided by the regression method.
The goal of this work is the evaluation of a hybrid approach for the estimation of BVs, such as Canopy Chlorophyll Content (CCC) and Leaf Area Index (LAI) of maize crops, from hyperspectral and multispectral sensors.
In particular, this study tested the most recent version of PROSAIL (PROSAIL-PRO, Féret et al., 2020) and several MLRAs. The PROSAIL-PRO model was used to generate a database with hundreds of simulated vegetation reflectance spectra (Look Up Table - LUT). This database was then used to train different combinations of ML algorithms and feature selection configurations to estimate crop BVs. The best performing combinations were then applied to actual Sentinel-2

* Corresponding author

data. The best results obtained from the hybrid approach were then compared to LAI and CCC estimated from S2Toolbox.

2. MATERIAL AND METHODS

The following sections describe the study area, field measurements, EO dataset and the steps involved in the hybrid approach for the BV retrieval.

2.1 Study area and fields measurements

The study area is located in Tuscany (42°49′47.02″ N 11°04′10.27″ E; elev. 2 m a.m.s.l.), central Italy, North of Grosseto and 20 km away from the coastline.
Within the study area, two maize crops, from two different farms, Le Rogaie (around 76 ha) and Ceccarelli (around 33 ha), were selected as test sites. These two fields feature different irrigation systems and different sowing dates.
During June and July 2018, two field campaigns were carried out on the two fields, in order to collect a comprehensive dataset of biochemical and biophysical parameters. In particular LAI was measured with a LAI-2200 (LI-COR Biosciences, USA) and Leaf Chlorophyll Content (LCC) was measured with SPAD (Konica Minolta, Japan) and DUALEX (Force-A, France). CCC was calculated as LAI * LCC. The field activities included CAL/VAL radiometric measurements performed with spectroradiometer SPECTRAL EVOLUTION SR-3500 (https://spectralevolution.com/products/hardware/compact-lab-spectroradiometers/sr-3500/) and vegetation measurements and sampling. 33 Elemental Sampling Units (ESU) of almost 20x20 m were identified for the field campaign. Each ESU includes up to 4 plots of 10x10 m, for a total of 87 plots.

2.2 Airborne and spaceborne EO dataset

The study area was acquired by the HyPlant-DUAL hyperspectral airborne sensor on 7th and 30th July 2018. HyPlant-DUAL dataset was spectrally resampled at PRISMA (PRISMA-like) and Sentinel-2 (S2-like) wavelengths. PRISMA-like spectra were compared to radiometric field measurements in order to remove noisy bands presenting a mean absolute error greater than 5%: the final spectral configuration includes 155 bands. In the case of both S2 datasets (actual and simulated form HyPlant-DUAL), B1, B2, B9, B10 were removed to be consistent with ESA S2Toolbox (Weiss & Baret, 2016), while B8a was removed by Sen2R (Ranghetti et al., 2020). Therefore, the final spectral configuration for S2-like dataset includes 8 bands: B3, B4, B5, B6, B7, B8, B11, B12.
In addition, actual Sentinel-2 images were also available on the area of interest. The Sentinel-2 images acquired on 8th July and 2nd August 2018 (the S2 images closest to the field campaign) were then downloaded with Sen2R (Ranghetti et al., 2020).
Table 1 resumes the EO dataset used in this study.

Dataset	Acquisition date
PRISMA-like	07/07/2018
S2-like	07/07/2018
Sentinel-2	08/07/2018
PRISMA-like	30/07/2018
S2-like	30/07/2018
Sentinel-2	02/08/2018

Table 1. Airborne and spaceborne EO datasets.

2.3 Hybrid approach

2.3.1 PROSAIL-PRO: The RTM tested in the hybrid approach proposed in this work is PROSAIL-PRO. It combines two RTMs: the leaf PROSPECT-PRO model and the canopy 4SAIL model. Starting from structural and biophysical parameters as inputs, PROSPECT-PRO simulates reflectance at the leaf level from 400 nm to 2500 nm. The outputs of the PROSPECT-PRO are used by the 4SAIL, together with other variables, such as plant structural parameters, viewing angles and soil background to simulate the vegetation reflectance at canopy level.

2.3.2 Development of the training LUT: The PROSAIL-PRO was used to simulate canopy reflectances based on the combination of different input, characterizing the crop, the soil and the sun-sensor geometry.
Assumptions on the distribution of the above input variables needs to be done. Each PROSAIL-PRO input was modelled according to Normal or Uniform distributions. For each variable, the ranges of these distributions (mean and standard deviation for Normal distribution; min and max values for Uniform distribution) were set according to both field measurements and literature (Weiss & Baret, 2016), in the case measurements were not available.
Those inputs were then randomly sampled according to their distribution and used to simulate, through the PROSAIL-PRO, canopy reflectances of maize crops in the range 400–2500 nm with a spectral resolution of 1 nm. These reflectances were then resampled at the selected PRISMA (155 bands) and Sentinel-2 (8 bands) wavelengths. The final training database includes both input variables and PRISMA-like or S2-like reflectance spectra.
A Gaussian white noise of 5% was also added to inputs and canopy reflectances, in order to get more realistic data.

2.3.3 MLRA training phase: The training phase was performed using different ML regression algorithms for the retrieval of LAI and CCC. The algorithms used in this study include Partial Least Square Regression (PLSR), Gaussian Process Regression (GPR), Support Vector Regression (SVR), Artificial Neural Networks (ANN) and Random Forests (RF).

2.3.4 Validation phase and maps generation: The trained models were then applied to the datasets reported in Table 1.
The 87 field measurements of BVs carried out in the two maize fields in Grosseto were used to validate the hybrid models, comparing measured and estimated BVs values.
Moreover, for Sentinel-2 images, the best retrieval results for LAI and CCC, from the above-mentioned algorithms, were compared to the estimates from ESA S2Toolbox (Weiss & Baret, 2016). Due to the different spatial resolution between HyPlant-DUAL and actual Sentinel-2 images, the validation statistics were computed by averaging the BV values belonging to the same ESU.
Finally, the best performing algorithms were applied to the datasets (PRISMA-like, S2-like and actual S2) to generate maps of LAI and CCC over the investigated maize crops.
The training, validation and generation of maps were performed using ARTMO (Verrelst et al., 2011).

3. RESULTS AND DISCUSSION

This section shows the results of the influence of LUT size, as well as the results of LAI and CCC estimations from the proposed hybrid approach and the S2Toolbox.

3.1 Impact of LUT size on retrieval performance

The impact of the database size on the retrieval performance was investigated for all the BVs. Several LUTs ranging from 1000 to 9000 samples, with a 1000 samples step, were generated in the S2-like configuration. These LUTs were then used to train retrieval models for LAI and CCC and were validated against field measurements.

Figure 1 shows, as example, the impact of the training database size on the accuracy and training time of the selected models for CCC. The increase in the size of the training dataset leads to a minor improvement in the model statistics. On the other hand, the training time rises significantly, in particular for GPR. A similar pattern was verified also for LAI. Therefore, a LUT of 2000 samples was considered a good trade-off between accuracy and time.

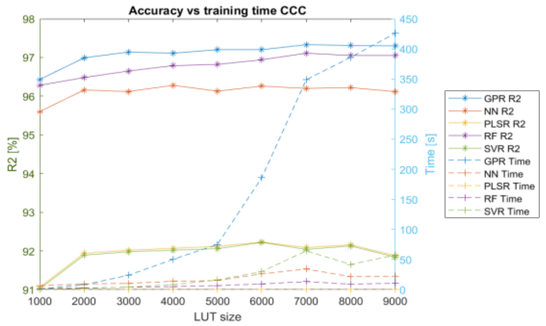

Figure 1. Impact of the training database size on the models' accuracy (R^2) and training time for CCC.

3.2 Hyperspectral vs multispectral BVs retrieval

The comparison of the best results for CCC and LAI estimated from PRISMA-like and S2-like dataset are resumed in Figure 2.

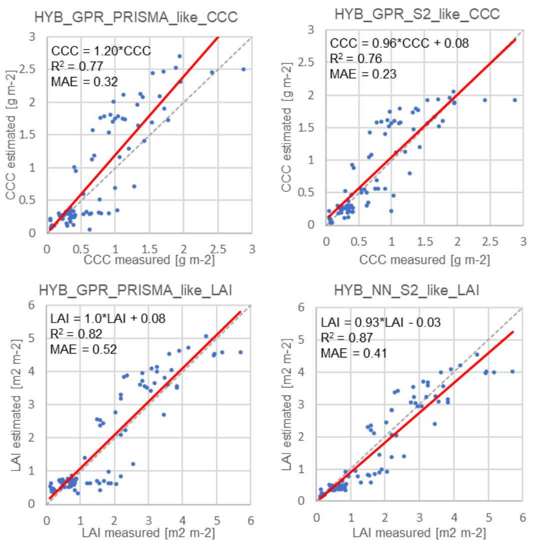

Figure 2. Comparison of CCC (top) and LAI (bottom) estimation from PRISMA-like (left) and S2-like (right) dataset using the hybrid approach (GPR and NN).

In general, retrieval results for CCC and LAI show very good performances for both PRISMA-like and S2-like dataset. For both BVs, S2-like achieved slightly better performance than PRISMA-like, in terms of MAE (CCC: 0.23 for S2 and 0.32 for PRISMA; LAI: 0.41 for S2 and 0.52 for PRISMA). Even if PRISMA provided a better correlation coefficient (R^2 = 0.77) than S2 (R^2 = 0.76) for CCC, there is an overestimation of this BV of 20%. Moreover, it is worth noting that PRISMA-like gives better estimates than S2-like at high values, highlighting a saturation effect for S2-like.

Among the tested ML algorithms GPR provided the best results for CCC and LAI retrieved from PRISMA-like. Whereas NN performed better for LAI estimated from S2-like.

3.3 Hybrid vs S2Toolbox BVs retrieval

The best performing algorithms, for the retrieval of CCC and LAI from S2-like, were applied to the actual Sentinel-2 images, acquired on 8th July and 2nd August 2018. S2Toolbox was applied to the same dataset to assess both BVs.

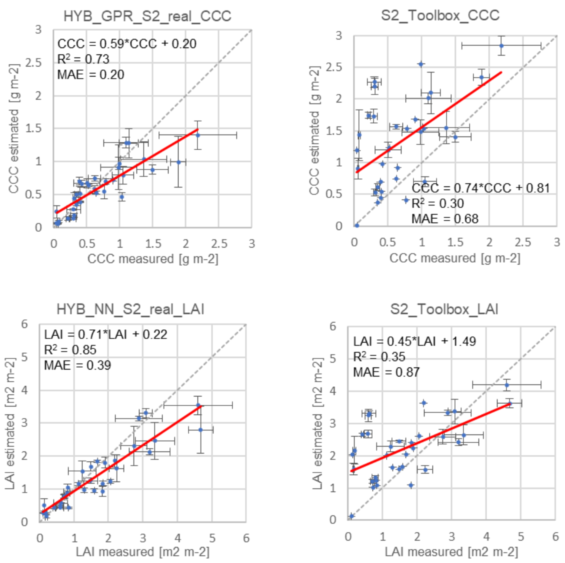

Figure 3. Comparison of CCC (top) and LAI (bottom) estimations from Sentinel-2 data using hybrid approach (left) and S2Toolbox (right).

Retrieval results, using the hybrid approach and S2Toolbox, are resumed in Figure 3. Considering the hybrid approach, from the scatterplots in Figure 2 and Figure 3, it is clear how the retrieval performed on real data leads to a general underestimation of the BVs: specifically, the angular coefficient of the regression line is generally lower than the slope for S2-like retrieval. This worsens the saturation effect observed in S2-like data. However, despite this issue, the hybrid approach provides good results for both CCC (R^2 = 0.73, MAE = 0.20) and LAI (R^2 = 0.85, MAE = 0.39).

Regarding ESA S2Toolbox, the estimations show much lower accuracy for both CCC (R^2 = 0.30, MAE = 0.68) and LAI (R^2 = 0.35, MAE = 0.87). These poor results could be explained considering that the NN algorithm in the S2Toolbox was trained with a comprehensive LUT which should be representative of the main vegetation types around the globe. While this global training data may be suitable for modelling the average vegetation status, they might not represent the status of specific areas. Thus, fine-tuned ad-hoc models, such as those proposed in this work, can lead to significantly better estimates.

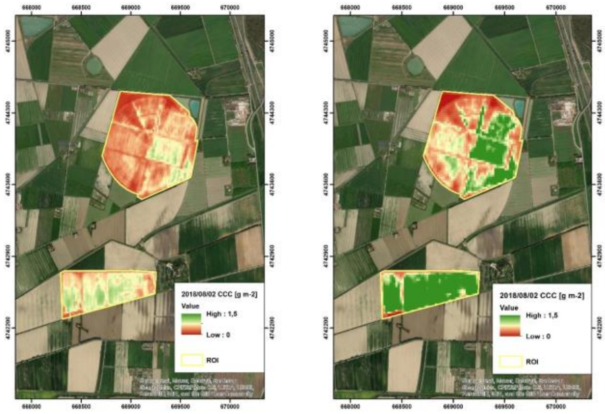

Figure 4. Comparison of CCC maps estimated from Sentinel-2 image acquired on 2nd August 2018 using the hybrid approach (left) and S2Toolbox (right).

Figure 4 shows the comparison of CCC maps estimated from Sentinel-2 image acquired on 2nd August 2018 using the hybrid approach (left) and S2Toolbox (right). As it was expected from the scatterplot, the map estimated from S2Toolbox presents many more saturated pixels especially for high CCC values.

4. CONCLUSIONS

This work proposed a hybrid method, which combines the radiative transfer model PROSAIL-PRO and several ML regression algorithms (PLSR, GPR, SVR, ANN and RF), for the estimation of CCC and LAI. The exploited EO dataset, acquired from both airborne and spaceborne sensors, includes both hyperspectral (PRISMA-like) and multispectral data (simulated and actual Sentinel-2 data).
The analysis on the impact of LUT size on retrieval performance showed that increments in LUT size have a minor impact on retrieval accuracy. On the other hand, an increase in the training time was observed, especially for GPR. For this reason, a LUT of 2000 samples was considered a good trade-off between accuracy and time.
The comparison between hyperspectral and multispectral data (simulated from the airborne imager HyPlant-DUAL) for the retrieval of CCC and LAI showed very good performances for PRISMA-like and S2-like dataset. For both BVs, S2-like achieved slightly better performance than PRISMA-like, even though S2-like estimates showed a saturation effect visible at high CCC and LAI values.
The best performing algorithms for S2-like were applied to actual Sentinel-2 data and compared to the results obtained using ESA S2Toolbox. The validation of these two approaches showed that the proposed hybrid method provides better estimates than S2Toolbox for both CCC and LAI, highlighting the need for specific algorithms tuned for specific areas.

ACKNOWLEDGEMENTS

The project activities have been framed in the CNR-DIPARTIMENTO DI INGEGNERIA, ICT E TECNOLOGIE PER L'ENERGIA E I TRASPORTI project “DIT.AD022.180 Transizione industriale e resilienza delle Società post-Covid19 (FOE 2020)”, sub project activity “Agro-Sensing”.
The research leading to these results was conducted within “ESA RFP/3-15502/18/NL/IA: CHIME RCS, Mission Requirement Consolidation Study” and its related CCN (Ref. HYS-CN-ISP-PM-0001), funded by the European Space Agency.

REFERENCES

Bach, H., Mauser, W., & Klepper, G. (2016). Earth observation for food security and sustainable agriculture. *European Space Agency, (Special Publication) ESA SP*.

Berger, K., Verrelst, J., Féret, J. B., Hank, T., Wocher, M., Mauser, W., & Camps-Valls, G. (2020). Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. In *arXiv*. https://doi.org/10.1016/j.jag.2020.102174

Berger, K., Verrelst, J., Féret, J. B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. In *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2020.111758

Féret, J. B., Berger, K., de Boissieu, F., & Malenovský, Z. (2020). Prospect-pro for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. In *arXiv*. https://doi.org/10.1016/j.rse.2020.112173

Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. In *Science* (Vol. 327, Issue 5967). https://doi.org/10.1126/science.1183899

Plant, R. E., Pettygrove, G. S., & Reinert, W. R. (2000). Precision agriculture can increase profits and limit environmental impacts. *California Agriculture*. https://doi.org/10.3733/ca.v054n04p66

Ranghetti, L., Boschetti, M., Nutini, F., & Busetto, L. (2020). “sen2r”: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers and Geosciences*, *139*. https://doi.org/10.1016/j.cageo.2020.104473

Verrelst, J., Berger, K., & Rivera-Caicedo, J. P. (2020). Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms. *IEEE Geoscience and Remote Sensing Letters*. https://doi.org/10.1109/lgrs.2020.3014676

Verrelst, J., Rivera, J. P., Alonso, L., & Moreno, J. (2011). ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. *Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK.*

Weiss, M., & Baret, F. (2016). S2ToolBox Level 2 products: LAI, FAPAR, FCOVER - Version 1.1. *Sentinel2 ToolBox Level2 Products*.