Hybrid approaches for the estimation of biophysical variables of agronomic interest from hyperspectral and multispectral data: applications for maize crops

[Marina Ranghetti]

Hunger and food insecurity have been a growing problem over the last years. In 2019 one in ten people in the world were exposed to severe levels of food insecurity, and an estimated two billion people in the world did not have regular access to safe, nutritious and sufficient food. Nowadays, the global megatrends (climate change, population growth, technological change) gradually caused the supply-demand balance to shift towards a not sufficient and unsustainable food production, with a potentially dramatic consequence for environmental and humanitarian aspects. This poses a challenge on the agriculture to increase the global production, while at the same time ensuring healthy products and sustainable practices. Nowadays, geoinformation products can be used to provide farmers with decision-supporting spatial information (crop traits maps), able to highlight within-field crop variability, as a fundamental tool to support site-specific management (i.e., precision farming). In this context, leaf and canopy biochemical and biophysical variables (BVs) can become essential parameters describing vegetation features and indicating agro-ecosystem conditions. Thus, monitoring BVs, such as chlorophyll and nitrogen content and the plant density is a good approach to provide spatio-temporal information needed for improving agricultural practices. In the last decades, remote sensing was employed in different ways for estimating these traits in a non-destructively and non-invasively way and for supporting crop monitoring and management across space (agricultural sites) and time (seasons). Recently, the availability of new generation hyperspectral sensors in space and the methodological advances in the retrieval schemes are forging ahead the possibility to obtain up-to-date information about the variability of vegetation traits across the globe. The imaging spectroscopy mission PRecursore IperSpettrale della Missione Applicativa (PRISMA), launched on March 22nd, 2019 by the Italian Space Agency, is a recent spaceborne initiative for EO devoted to test a new-generation hyperspectral sensor. With its high spectral resolution (~10 nm Full Width Half Maximum) in the solar spectrum range (400-2500 nm) and with a ground sampling distance (GSD) of 30 m, PRISMA opens new opportunities in multiple scientific domains and applications, including crop monitoring, sustainable agriculture and precision farming. The availability of spaceborne hyperspectral data allows the development of new approaches for the operational mapping of important vegetation BVs and potentially open new opportunity for added value information flow in management and monitoring agricultural system. The hybrid approach, recently introduced by the scientific community for BVs retrieval, represents a possible solution to this problem. This approach consists in the combination of radiative transfer models (RTM) and machine learning regression algorithms (MLRAs): the RTM generates a database of simulated vegetation spectra (input), related to the vegetation BVs (output), and the MLRA identifies a non-linear model between input-output pairs. Thus, hybrid methods inherit both the transferability guaranteed by the use of a physically-based method and the computational efficiency and flexibility provided by MLRAs.

The goal of this work is the evaluation of the hybrid approach and the comparison of multispectral (Sentinel-2) and hyperspectral (PRISMA) sensors for the estimation of different maize BVs, such as

Leaf Chlorophyll Content (LCC), Canopy Chlorophyll Content (CCC), Leaf Nitrogen Content (LNC), Canopy Nitrogen Content (CNC) and Leaf Area Index (LAI).

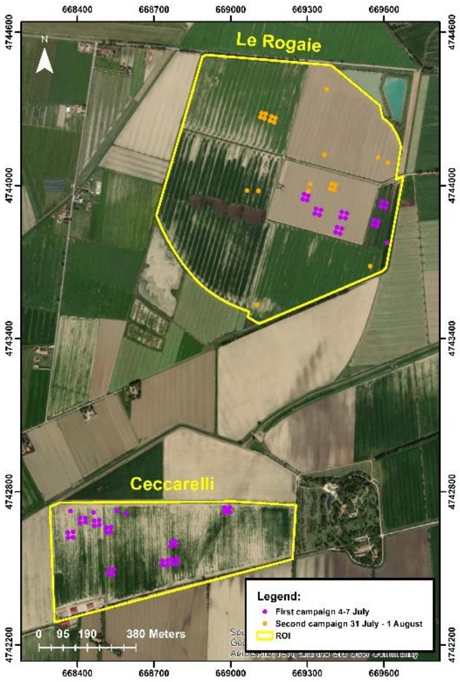

Figure 1. Study area and field measurements.

The study area (Figure 1) 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, LCC and LNC. CCC and CNC were calculated as LAI * LCC/LNC. The field activities included CAL/VAL radiometric measurements performed 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.

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 addition, real Sentinel-2 images were also available on the area of interest on 8th July and 2nd August. The spectral configuration for S2 dataset (simulated and real) includes 8 bands: B3, B4, B5, B6, B7, B8, B11, B12.

Figure 2 provides a synthetic representation of the general workflow of this study (top panel - A) and the method and data used for each step (bottom panel - B).

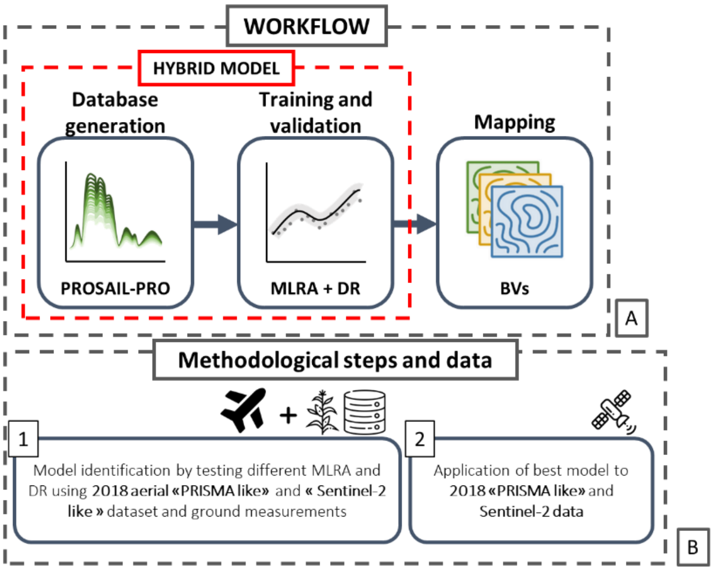
WORKFLOW
HYBRID MODEL
Database generation
Training and validation
Mapping
PROSAIL-PRO
MLRA + DR
BVs
A
Methodological steps and data
1
Model identification by testing different MLRA and DR using 2018 aerial «PRISMA like» and « Sentinel-2 like » dataset and ground measurements
2
Application of best model to 2018 «PRISMA like» and Sentinel-2 data
B

Figure 2. General workflow of the methodological phases followed in this study.

The first step consists in the configuration of PRISMA-like and S2-like hybrid models (database generation and training and validation of different MLRA with or without dimensionality reduction combinations) for the estimation of the selected BVs. PROSAIL-PRO was used to simulate canopy reflectances based on the combination of different input, characterising the crop, the soil and the sun-sensor geometry. It simulates canopy reflectances from 400 to 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. The training phase was performed using different MLRA for the retrieval of the BVs. 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). Dimensionality reduction, such as Principal Component Analysis (PCA) was tested for the PRISMA-like database. The algorithms were run by setting different PCA, such as 5,10,15,20. The trained models were then applied to PRISMA-like and S2-like datasets. 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 (step 1). Finally, the best performing algorithms were applied to the datasets (PRISMA-like and real S2) to generate maps of BVs over the investigated maize crops (step 2).

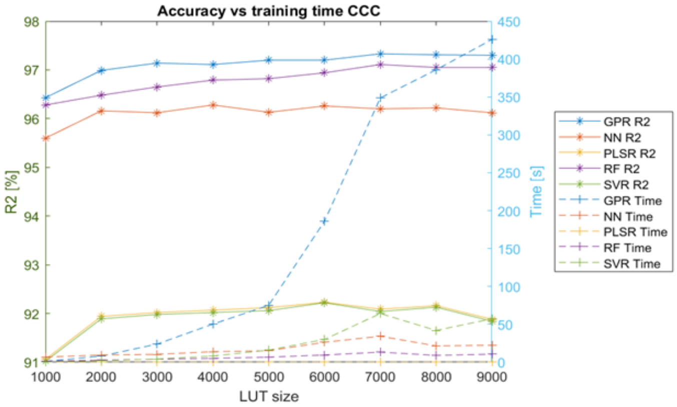

Figure 3. Impact of the training dataset size on the models' accuracy (R^2) and training time for CCC.

The impact of the dataset size on the retrieval performance was investigated for all the BVs. Several datasets ranging from 1000 to 9000 samples, with a 1000 samples step, were generated in the S2-like configuration. Figure 3 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.

The comparison of the best results for CCC, CNC and LAI estimated from PRISMA-like and real S2 dataset using the hybrid approach are resumed in Figure 4. At the leaf scale, not significant performances were obtained, for this reason, they are not reported. This behaviour confirms that the retrieval of variables at the leaf scale is still a challenging task that requires further studies and analyses. In general, at the canopy level, retrieval results for CCC, CNC and LAI show very good performances for both PRISMA-like and S2 dataset. For all BVs, S2 achieved slightly better performance than PRISMA-like, in terms of MAE (CCC: 0.2 for S2 and 0.32 for PRISMA; CNC: 0.73 for S2 and 0.91 for PRISMA; LAI: 0.39 for S2 and 0.52 for PRISMA). Even if PRISMA provided a better correlation coefficient (R^2 = 0.77) than S2 (R^2 = 0.73) for CCC, there is an overestimation of this BV of 20%. Moreover, it is worth noting that PRISMA-like gives better estimates than S2 at high values, highlighting a saturation and underestimation effect for S2. Among the tested ML algorithms GPR provided the best results for CCC and LAI retrieved from PRISMA-like and only for

CCC for S2. Whereas NN performed better for CNC estimated from PRISMA-like and for CNC and LAI estimated from S2.

The tests performed on different feature selection strategies within the hybrid approach did not show any conclusive results. Sometimes, as in the case of canopy level, the use of PCA gave slightly better results than using all bands. Anyway, the retrieval performances obtained using all available bands show results comparable to those obtained with PCA feature selection.

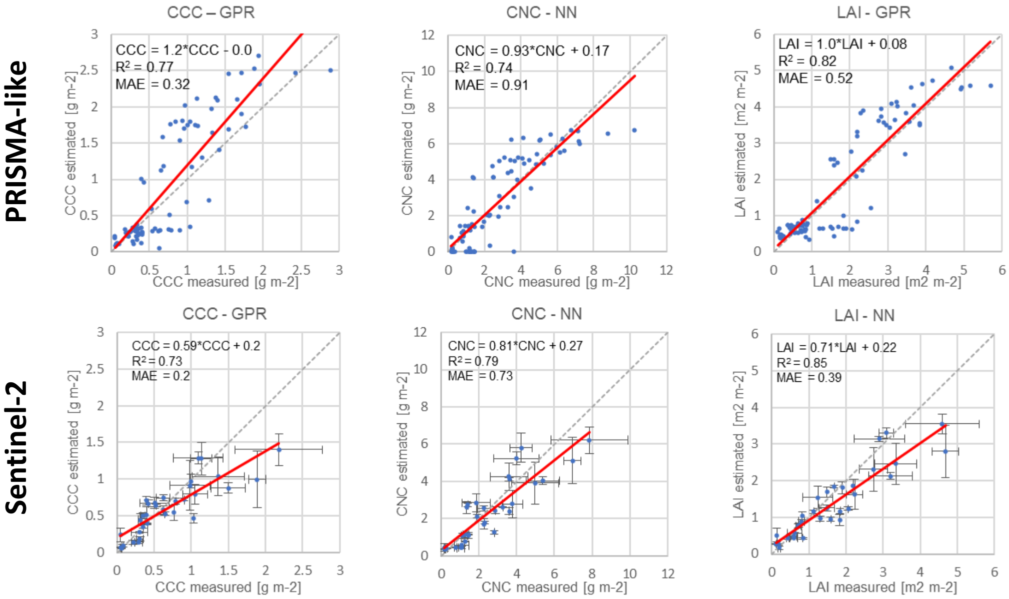

Figure 4. Comparison of CCC (left), CNC (middle) and LAI (right) estimations from PRISMA-like (top) and Sentinel-2 dataset using the hybrid approach.

Figure 5 shows, as example, the comparison of CNC maps estimated from PRISMA-like image (20 m spatial resolution) acquired on 30th July 2018 (left) and real S2 image (10 m spatial resolution) acquired on 2nd August 2018 (right) using the hybrid approach. For PRISMA-like image, it can be observed that the field patterns are well captured and the values are within consistent ranges. Inside Le Rogaie field, higher values of CNC are visible in S2 maps. This behaviour can be due to the time difference but also to the different geometric resolution of the images.

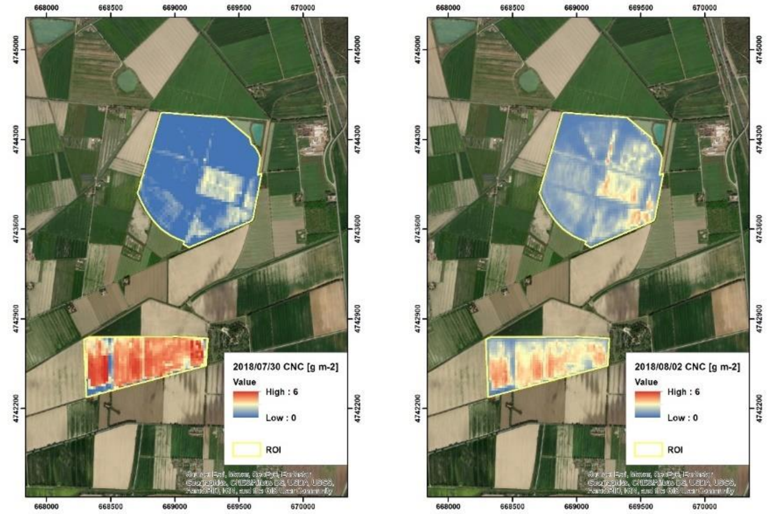

Figure 5. Comparison of CNC maps estimated from PRISMA-like image acquired on 30th July 2018 (left) and Sentinel-2 acquired on 2nd August 2018 (right) using the hybrid approach. ESRI image on the background.

This work proposed a hybrid method, which combines the radiative transfer model PROSAIL-PRO and several machine learning regression algorithms (PLSR, GPR, SVR, NN and RF), for the estimation of LCC, LNC, CCC, CNC 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 for the retrieval of CCC, CNC and LAI showed very good performances for PRISMA-like and real S2 dataset. For all BVs, S2 achieved slightly better performance than PRISMA-like (in terms of MAE), even though S2 estimates showed a saturation and underestimation effect visible at high CCC, CNC and LAI values. The capability of hyperspectral and multispectral data to perform retrieval of more problematic parameters, such as chlorophylls and nitrogen at the leaf scale, needs to be further investigated by, for example, optimizing the spectral sampling for each specific BV.

Result obtained from this study show how crop BVs, in particular maize traits, can be estimated from space using data from new-generation hyperspectral sensors. The hybrid approach, which is fully independent from field measurements and datasets, led to good performance and accuracy during the validation phase against the test site data.