Evaluation and exploitation of retrieval algorithms for estimating biophysical crop variables using Sentinel−2, Venμs and PRISMA satellite data

CASA Raffaele [1], UPRETI Deepak[1], PALOMBO Angelo[2], PASCUCCI Simone[2], YANG Hao[3], YANG Guijun[3], HUANG Wenjiang[4], PIGNATTI Stefano[2]

1. *Department of Agriculture Forestry Nature and Energy, Università degli Studi della Tuscia , Via San Camillo de Lellis,* 01100 *Viterbo, Italy*;
2. *Institute of Methodologies for Environmental Analysis, Italian National Research Council, C.da Santa Loja, Tito Scalo,* 85050 *Potenza, Italy*;
3. *Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing* 100097*, China*;
4. *Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing* 100094*, China*

Abstract: This paper is devoted to the development and testing of the optimal procedures for retrieving biophysical crop variables by exploiting the spectral information of current multispectral optical satellite Sentinel-2 and Venμs and in view of the advent of the new Sino-EU hyperspectral satellite (e.g., PRISMA, EnMAP, and GF-5). Two different methodologies devoted to the estimation of biophysical crop variables Leaf area index (LAI) and Leaf chlorophyll content (Cab) were evaluated: (a) non-kernel-based and kernel-based Machine Learning Regression Algorithms (MLRA); (b) Sentinel-2 and Venμs data comparison for the analysis of the durum wheat-growing season. Results show that for Sentinel-2 data, GPR (Gaussian Process Regression) was the best performing algorithm for both LAI (R^2 = 0.89 and RMSE = 0.59) and Cab (R^2 = 0.70 and RMSE = 8.31). Whereas, for PRISMA simulated data the Kernel Ridge Regression (KRR) was the best performing algorithm among all the other MLRA (R^2 = 0.91 and RMSE = 0.51) for LAI and (R^2 = 0.83 and RMSE = 6.09) for Cab, respectively. Results of Sentinel-2 and Venμs data for durum wheat-growing season were consistent with ground truth data and confirm also that SWIR bands, which are used as tie-points in the PROSAIL inversion, are extremely useful for an accurate retrieving of crop biophysical parameters.

Key words: biophysical crop parameters, PRISMA, GF-5, Sentinel 2, Venμs

Citation format: **Casa R,Upreti D,Palombo A,Pascucci S,Yang H,Yang G J,Huang W J and Pignatti S. 2020. Evaluation and exploitation of retrieval algorithms for estimating biophysical crop variables using Sentinel−2, Venμs and PRISMA satellite data. Journal of Remote Sensing(Chinese). 24(S1): 25−32**

1 INTRODUCTION

Within the ESA-MOST DRAGON4 initiative, the Sino - Italian project "Combined exploitation of Sino - EU earth observation data for supporting the monitoring and management of agricultural resources" aims at expanding the current capability of monitoring biophysical and physiological variables related to agronomical processes optimizing the contribution of EU and Sino EO (Earth Observation) optical data to the monitoring capability".

Within this project framework, a specific research activity (developed within subproject 1 - Topic1) has been conducted with the title "Algorithm Development Exploiting Multitemporal and multisensor Satellite data for improving crop classification, biophysical and agronomic variables retrieval and yield prediction" (ADEMS). The main project objective is to develop and test new algorithms, as well as the improvement of existing ones, exploiting the wealth of information that can be obtained when assimilating multi-temporal and multi-sensor (i. e. multi and hyperspectral optical data and high-resolution satellite data) into agricultural crop models. This goal can be achieved through an unprecedented frequency of medium and high spatial resolution data that can be obtained within a crop growing season, when European and Chinese existing satellite acquisitions are combined together (e. g., GF series, ZY-3, HJ, SENTINEL-2/3, Landsat-8 and Venμs). This is also fostering the next generation satellite hyperspectral data of PRISMA, EnMAP, and GF-5, useful to overcome spectral/spatial and temporal defi-

Received: 2019-10-21; **Accepted:** 2020-04-16

Foundation: The present manuscript describes the achievement of two research activities performed within the subproject that at the mid-term of the project have produced the most interesting results. They deal with two methodological aspects related to:

First author biography: Casa Raffaele (1964—), male, researcher, his research interest is applications of remote sensing to retrieval of agronomic variables. E-mail: rcasa@unitus.it

Corresponding author biography: biography: Pignatti Stefano (1965—), male, researcher, his research interest is applications of hyperspectral remote sensing to land. E-mail: stefano.pignatti@cnr.it

ciencies of the actual satellite data.

In particular, the research activity of the subproject ADEMS has been structured to (a) develop new models to exploit the potential of multi-temporal remote sensing data in retrieving vegetation canopy biophysical variables; (b) to investigate the optimal way to incorporate, into data assimilation algorithms, uncertainties and errors both in the remote sensing observations, in the retrieval of biophysical variables and in the crop models; (c) to estimate crop yield and quality (e.g., grain protein content in cereal crops).

(a) the analysis of the performance of non-kernel-based and kernel-based MLRA (Machine Learning Regression Algorithms) for the retrieval of biophysical variables, LAI (Leaf Area Index) and leaf chlorophyll content (Cab), in order to explore the possible operational algorithms for the recent and forthcoming hyperspectral missions in China (GF-5 already in orbit) and in Europe (PRISMA, launched on March 2019, that will operate as ESA TPM; EnMap);

(b) the comparison of two optical satellite sensors (Sentinel 2 and Venµs) for the analysis of the durum wheat growing season in order to evaluate both the SWIR and the different spatial resolution (respectively 10 and 5 m), which research covers the durum wheat growing season in specific fields of the Maccarese Farm located in central Italy close to Rome.

Concerning the first research activity, biophysical variables such as LAI and Cab are of crucial importance for different agricultural applications (Goffart et al., 2008; Zhang and Kovacs, 2012; Cilia et al., 2014; Jay et al., 2017; Casa et al., 2019). There are different methods for the quantification of these variables from remote sensing data, starting from simple statistical-based vegetation indices (Rouse et al., 1974; Rondeaux et al., 1996) to the more advanced physically-based approaches (Pasolli et al., 2012; Verrelst et al., 2015b). The current state of the art hybrid methods make use of all the available spectral information and rely on the generation of the spectra simulated from physically-based Radiative Transfer Model (RTM), followed with the training of intelligent non-parametric regression algorithms or MLRA (Durbha et al., 2007; Lázaro-Gredilla et al., 2014). Although hybrid approaches of biophysical variables retrieval offer several advantages and are widely accepted as the state of the art methods, there are some drawbacks. Noteworthy is that the size of RTM simulations depends on its parameters and the sensor spectral information. In time, there have been significant improvements in MLRA for such tasks. Most of the operational image processing packages implement Neural Networks (NN) for such tasks (Chai et al., 2008; Weiss and Baret, 2016). The problem of training RTM based simulated spectra using NN is the tuning of multiple hyperparameters and the black-box nature of NN (Verrelst et al., 2015a). Due to these disadvantages, in recent years, a lot of attention is put on other alternatives (as compared to NN), i.e., kernel-based methods.

In comparison to NN, Kernel-based methods are simple to train and requires tuning of very few parameters, these algorithms also have the flexibility to train the algorithm with different kernel functions from very simple RGB kernel to very sophisticated kernels (Williams and Rasmussen, 2006; Verrelst et al., 2016). Some kernel-based methods such as GPR (Gaussian Process Regression) also provide uncertainty estimates with the mean value of estimation (Verrelst et al., 2015a, 2016, 2018). The drawback of the kernel-based algorithms is that these are computationally very demanding and it is not feasible to train them with large datasets as simulations generated by the RTM require a reasonable amount of time. To make the training feasible with the kernel-based algorithms active learning techniques have been proposed that optimally sample the simulations generated by RTM in a manner that the training of kernel-based MLRA is performed on the most informative samples from the simulation set. This way the redundancy in the RTM simulated spectra can be avoided and kernel-based MLRA can be trained in a short time.

With respect to the second methodological research aspect, it relays on the importance for both scientific and practical purposes to have an accurate and timely spatial classification of the main crop biophysical variables based on Sentinel-2 and Venµs satellite remote sensing data. The potential of deriving an accurate crop phenology information at medium/high spatial resolution is still a challenge because of the actual revisiting time of the multispectral constellation and the inevitable cloud contamination. The actual research on croplands, in fact, has moved from the analysis of single static imagery to the assortment of time-series images and multispectral information.

In this framework, the use of Sentinel-2 and Venµs satellite multispectral data were evaluated and compared for durum wheat crop monitoring. To this aim, we applied the PROSAIL model to retrieve from Sentinel-2 and Venµs satellite remotely sensed data the ensemble of the model input variables (both biophysical and biochemical). The crop parameters retrieval was performed by using both SNAP, image processing application developed by ESA (European Space Agency) and in house developed IDL code based on the minimization of a cost function (Rivera et al., 2013) in order to allow the inversion of a generic remote sensing dataset (i. e. proximal or remote, multispectral or hyperspectral and vegetation library). The efficiency of the IDL inversion methods was assessed in terms of variables estimation and leaf spectra fitting by using RMSE (Root Mean Square Error) and Coefficient of determination (R^2) for the spectra fitting along the proper satellite bands. In more detail, the aim of this activity is to check the suitab ility of Venµs bands in retrieving biophysical variables in comparison with Sentinel-2 results (Frampton, W. J., et al. 2013). In this way, Venµs could provide an opportunity for estimating the Red REP (Edge Position) at higher spatial and temporal resolutions with respect to Sentinel-2 data, thus increasing the ability to estimate more accurate crop phenology.

2 DATA & METHODS

Evaluation of retrieval algorithms suitable for the next hyperspectral mission

MLRA and kernel and non-kernel methods for retrieval biophysical variables (LAI and Cab) have been considered for the upcoming hyperspectral missions, i.e. Italian PRISMA (Stefano Pignatti et. al., 2013), German EnMAP (Guanter et al., 2015) and the Chinese GaoFen-5 with very similar characteristics with respect to PRISMA (Table 1).

Taking into account the numerous similarities among the hyperspectral sensors presented in Table 1, we have tested the retrieval algorithms for the PRISMA sensor for which we have available the spectral curves and other ancillary information useful for the resampling in the PROSAIL simulations.

PROSAIL (Jacquemoud et al., 2009) model has been widely used for generating simulated spectra and biophysical variables. The inputs for the PROSAIL model used in this work are presented in Table 2 (Weiss and Baret, 2016). The PROSAIL parameters were sampled using a full orthogonal experimental plan (Bacour et al., 2002), by using a component of BV-NNET (Biophysical Variable Neural NETworks) tool developed by Baret et al. (2003). The combination of a different number of classes in Table 1 defines the total number of simulated spectra, 41472 in this case.

Table 1 Upcoming hyperspectral sensors characteristics

Imager	Spectral bands	Spectral range/nm	FWHM/nm	SNR
PRISMA	247	400/2500	7/11	600 @ 650 nm > 400 @ 1550 nm > 200 @ 2100 nm
EnMAP	242	420/2450	10	> 500 @495 nm > 150 @2200 nm
AHSI-GF5	—	388/2518	4.5/8.6	702 @600 nm 506 @1300 nm 251 @2200 nm

Table 2 Input Parameters of PROSAIL model used for generating Sentinel-2 and PRISMA simulated spectra.

	Variable	Minimum	Maximum	Mode	Std	Nb_Class	Law
Canopy	LAI	0.0	15.0	2.0	2.0	6	log_normal
	ALA/(°)	30	80	60	20	3	gauss
	Crown Cover	1.0	1.0	0.8	0.4	1	uni
	HsD	0.1	0.5	0.2	0.5	1	gauss
Leaf	N	1.20	1.80	1.50	0.30	3	gauss
	Cab/($\mu g.cm^{-2}$)	20	90	45	30	4	gauss
	Cdm/($g.m^{-2}$)	0.0030	0.0110	0.0050	0.0050	4	gauss
	Cw_Rel	0.60	0.85	0.75	0.08	4	uni
	Cbp	0.00	2.00	0.00	0.30	3	gauss
Soil	Bs	0.50	3.50	1.20	2.00	4	gauss

Different MLRA have been tested with simulated spectra. As the kernel-based algorithms are computationally very demanding when trained on the large simulated spectra (Pasolli et al., 2012; Verrelst et al., 2016), in order to perform the retrieval with kernel-based MLRA, different random subsets from the simulated spectra were extracted and active learning strategies were used to perform the retrieval efficiently. We have used to compare the performance of all the proposed algorithms identical sets of random subsets of 1000 simulations obtained from the 41472 simulations in PROSAIL. Non-kernel based algorithms require a large dataset for training in order to obtain good performances, but on the other side, it is not possible to train kernel-based algorithms with large datasets in a reasonable amount of time. Therefore, we have used ten different random subsets of 1000 spectra extracted from PROSAIL simulated spectra.

Non Kernel based MLRA, NN (Hagan and Menhaj, 1994), BagT (Bagging Trees) (Breiman, 1996), PCR (Principal Component Regression) (Wold et al., 1987), RFTB (Random Forest Tree Bagger) (Breiman, 2001), LSLR (Least Square Linear regression) (Draper and Smith, 2014), PLSR (Partial Least Square Regression) (Geladi and Kowalski, 1986), BooT (Boosting Trees) (Friedman et al., 2000), RegT (Regression Trees) (Breiman et al., 2009), RFFE (Random Forest Fit Ensemble) (Breiman, 2001) and Kernel-based MLRA, GPR (Williams and Rasmussen, 2006), VHGPR (Variational Heteroscedastic Gaussian Process Regression) (Lázaro-Gredilla et al., 2014), RVM (Relevance Vector Machines) (Saarela et al., 2010), SVR (Support Vector Regression) (Vapnik et al., 1997), KRR (Kernel Ridge Regression) (Suykens and Vandewalle, 1999) and ELM (Extreme Learning Machines) (Huang et al., 2015) have been trained with randomly extracted 3/4 of the simulated spectra and cross validated with the other 1/4 part of the PROSAIL simulated spectra. The training and cross validation of PROSAIL simulated spectra have been performed using ARTMO (Automated Radiative Transfer Models Operator) toolbox (v 3.24) (Verrelst et al., 2011).

The retrieval is validated against the 1/4 part of the simulated spectra for both multispectral and hyperspectral sensor settings, i.e., Sentinel-2 (8 bands) and PRISMA (213 bands). The goodness of fit measures used in this work is R^2 RMSE.

Multispectral sensors synergy for monitoring durum wheat growing season

The experimental Maccarese farm is located north to Rome (Italy) close to the Fiumicino Airport. Eight Sentinel-2 cloud-free images from January 2017 to May 2017 were acquired on this site, and also four Venµs images were acquired for the same period (Table 3). Contemporary to the Sentinel-2 acquisitions, field measurements were performed on the Maccarese experimental fields to assess LAI, biomass, and chlorophyll content of the crops canopies. Field campaigns were held from January 2018 to April 2018 for different growing conditions of the crop of interest. The variables were sampled according to the ESU (Elementary Sampling Unit) scheme, to capture the variability within and among fields. Each ESU consisted of a 20m by 20m pixel size to easily accommodate Sentinel-2 10m pixel resolution. A total of 15 ESU, placed aat different locations were employed at different sampling dates. Each of the ESU contains nine points where LAI and Cab were measured using LAI 2000 and Dualax leaf clip reader (Upreti et al., 2019).

The characteristics of Sentinel-2 and Venµs multispectral sensors applied are given in Table 4. The main differences between the two sensors relay in: (a) the revisit time capability (5 days for S2 and 2 days for Venµs), (b) the spatial resolution (10—60 m for S2 and 5.3 m for Venµs) and (c) the covered spectral region (VNIR for Venµs and VNIR plus SWIR for Sentinel 2).

Table 3 List of images and data available for the analysis

Sentinel-2		Venμs
2018-01-29	2018-04-21	2018-01-28
2018-02-10	2018-05-19	2018-04-20
2018-02-13	2018-05-26	2018-04-26
2018-04-06	2018-05-31	2018-05-20

In order to retrieve leaf biochemical parameters from Sentinel-2 and PRISMA satellite data, a mathematical inversion of the PROSAIL physically-based canopy RTM has been implement in IDL. Among the possible inversion methods, the CF (Cost Function) has been implemented as inversion approaches so that the optimal parameter (opt_par) can be formulated as a semi-parametric problem where D is the function difference between satellite spectra (S) and the PROSAIL reflectance (P) determined by varying the input parameters Eq.1:

$$opt_{\mathrm{par}} = \operatorname{argmin} D\left[S(\lambda_j, P_i(\lambda_j)\right]; j = 1,2,\cdots.,n \quad (1)$$

Table 4 Sentinel-2 and Venμs main sensor characteristics

	Sentinel-2	Venμs
Spatial resolution	MS：10，20，60 m	MS：5.3 m
Spectral resolution (Center W.- Bandwidth)	B1：0.44-0.20 μm B2：0.49-0.65 μm B3：0.56-0.35 μm B4：0.66-0.30 μm B5：0.71-0.15 μm B6：0.74-0.15 μm B7：0.78-0.20 μm B8：0.84-1.15 μm B8a：0.86-0.20 μm B9：0.95-0.20 μm B10：1.38-0.30 μm B11：1.61-0.90 μm B12：2.2-1.80 μm	B1：0.42-0.40 μm B2：0.44-0.40 μm B3：0.49-0.40 μm B4：0.55-0.40 μm B5：0.62-0.40 μm B6：0.62-0.40 μm B7：0.67-0.30 μm B8：0.70-0.24 μm B9：0.74-0.16 μm B10：0.78-0.16 μm B11：0.86-0.40 μm B12：0.91-0.20 μm
Swath width	69 km with 2 cameras	27.6 km
Revisit capability	5 days	2 days

This procedure yields the biophysical and leaf biochemical parameters of PROSAIL by minimizing the summed differences between simulated and measured reflectance for the entire spectrum. RTM model is invertible if the solution of the inverse problem exists and is unique, it follows that, several combinations of canopy biophysical and leaf biochemical parameters can lead to very similar solutions (i.e. equivalence of the different parameters solutions for a single modelled canopy reflectance). To minimize this problem and to enhance the robustness of the retrievals, the developed code applies for the multiple best solutions instead of the single one (i.e. the mean of three runs for each spectra to speed the procedure).

The cost function procedure has been applied to retrieve the eight PROSAIL input parameters (i.e. *Cab*, C_{ar}, C_{brown}, C_{w}, C_{m}, N_{s}, *ALA,* LAI and *psoil* when LAI < 1), while sun zenith, view zenith and sun-sensor azimuth angle were fixed considering the acquisition geometry. Statistical metrics like based on R^2, RMSE, NRMSE between the PROSAIL retrieved spectra and the true image spectra were used to evaluate the goodness of the inversion process.

3 RESULTS

Evaluation of retrieval algorithms suitable for the next generation hyperspectral mission

Ten different random subsets of 1000 simulations were extracted from the entire simulated spectra, i.e., 41472 simulations. The retrieval is performed 10 times with different subsets, to reach unbiased and statistical reliable results. Table 5 shows the mean value of estimation and standard deviation as the measure of uncertainty for both Sentinel-2 and PRISMA sensor settings. Non-kernel based MLRA were trained with 3/4 part of the simulated spectra and validated against the other 1/4 part of the spectra. In case of kernel based MLRA, we have implemented an Active Learning technique based on the diversity criteria in order to make the training feasible and possible in a reasonable amount of time.

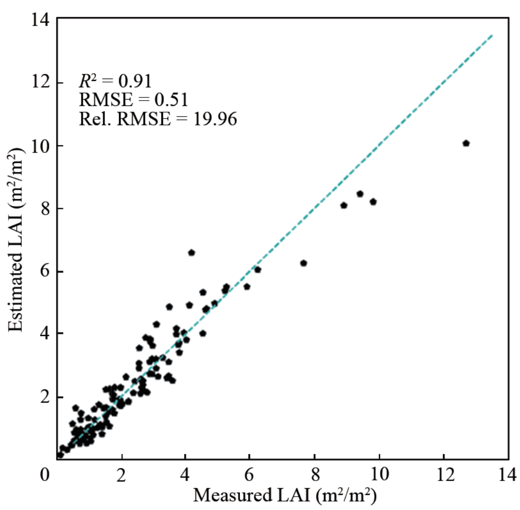
For this reason, we started our training with a random set of 50 simulations and used 1000 simulated spectra in a data pool. In an iterative loop of 100 iterations, we added 20 samples per iteration to the initial training data of 50 simulations. The samples are added based on the Euclidean distance-based diversity criteria (Douak et al., 2013, Verrelst et al., 2016), i.e., the samples in the data pool that are at the farthest distance from the initial training data and which improve the performance of kernel based MLRA were added to the initial training data. In this way, only the most relevant and informative and diverse samples are added to the initial training data and regression is performed on this updated dataset. The results of the retrieval are presented in Table 5 and Figure 1. As can be seen from Table 5, in case of Sentinel-2 the GPR is the best performing algorithm for both LAI (R^2 = 0.89 and RMSE = 0.59 m^2 m^{-2}) and Cab (R^2 = 0.70 and RMSE = 8.31 μg cm^{-2}). Whereas, for PRISMA simulated data the KRR is the best performing among all the other MLRA, LAI (R^2 = 0.91 and RMSE = 0.51 m^2 m^{-2}) and Cab (R^2 = 0.83 and RMSE = 6.09 μg cm^{-2}). It appears from these results that there is a moderate advantage of PRISMA over Sentinel-2 in terms of accuracy of the retrievals for both LAI and Cab (Fig. 1).

Table 5 Cross Validation results of LAI and Cab retrieval using different Machine Learning Regression Algorithms for Sentinel−2 and PRISMA sensors

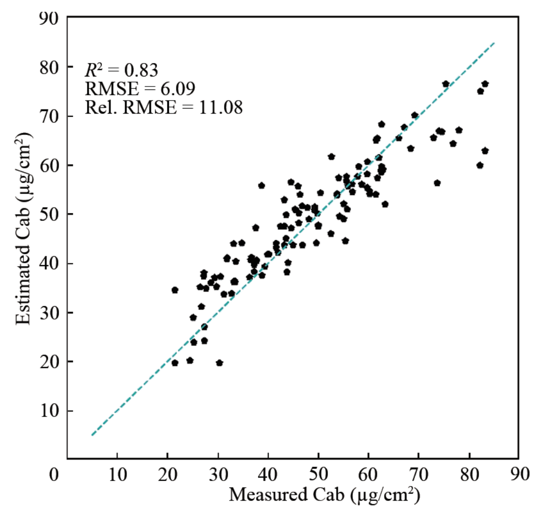
	SENTINEL−2 (8 Bands)				PRISMA (213 Bands)			
MLRA	LAI		Cab		LAI		Cab	
	R^2(SD)	RMSE(SD)	R^2(SD)	RMSE(SD)	R^2(SD)	RMSE(SD)	R^2(SD)	RMSE(SD)
GPR	**0.89 (0.06)**	**0.59 (0.36)**	**0.70 (0.13)**	**8.31 (2.05)**	0.89 (0.08)	0.53 (0.23)	0.79 (0.10)	6.82 (2.00)
VHGPR	0.86 (0.11)	0.70 (0.43)	0.71 (0.12)	8.31 (2.07)	0.89 (0.05)	0.55 (0.20)	0.75 (0.15)	7.08 (1.62)
RVM	0.87 (0.08)	0.63 (0.27)	0.66 (0.10)	9.21 (1.95)	0.88 (0.07)	0.60 (0.27)	0.75 (0.12)	7.57 (2.58)
SVR	0.88 (0.07)	0.65 (0.37)	0.65 (0.08)	9.02 (2.02)	0.86 (0.08)	0.70 (0.37)	0.78 (0.10)	8.60 (2.85)
KRR	0.89 (0.09)	0.58 (0.21)	0.68 (0.12)	8.51 (1.74)	**0.91 (0.05)**	**0.51 (0.29)**	**0.83 (0.11)**	**6.09 (2.25)**
ELM	0.86 (0.07)	0.69 (0.27)	0.64 (0.12)	8.85 (3.47)	0.81 (0.08)	0.81 (0.31)	0.61 (0.11)	9.16 (1.60)
NN	0.71 (0.04)	1.00 (0.15)	0.51 (0.17)	10.15 (3.58)	0.81 (0.03)	0.82 (0.12)	0.61 (0.04)	9.61 (0.54)
BagT	0.74 (0.05)	0.96 (0.13)	0.44 (0.18)	11.01 (3.90)	0.80 (0.03)	0.84 (0.11)	0.45 (0.05)	11.54 (0.62)
PCR	0.67 (0.07)	1.06 (0.22)	0.48 (0.17)	10.42 (3.70)	0.78 (0.03)	0.89 (0.10)	0.59 (0.02)	9.85 (0.46)
RFTB	0.72 (0.06)	0.98 (0.13)	0.45 (0.19)	10.79 (3.81)	0.79 (0.03)	0.86 (0.11)	0.45 (0.06)	11.46 (0.60)
LSLR	0.67 (0.07)	1.07 (0.20)	0.51 (0.17)	10.11 (3.60)	0.72 (0.04)	1.04 (0.11)	0.49 (0.02)	11.24 (0.64)
PLSR	0.67 (0.07)	1.07 (0.19)	0.51 (0.17)	10.11 (3.60)	0.72 (0.04)	1.04 (0.11)	0.48 (0.03)	11.39 (0.69)
BooT	0.63 (0.06)	1.13 (0.13)	0.42 (0.17)	11.75 (4.62)	0.62 (0.06)	1.24 (0.11)	2.87 (7.78)	12.28 (0.45)
RegT	0.57 (0.09)	1.28 (0.17)	0.27 (0.20)	13.92 (4.88)	0.60 (0.06)	1.26 (0.13)	0.16 (0.04)	16.34 (0.68)
RFFE	0.56 (0.09)	1.35 (0.21)	0.25 (0.20)	14.56 (5.47)	0.60 (0.09)	1.30 (0.14)	0.16 (0.05)	16.77 (0.84)

Multispectral sensors synergy for monitoring durum wheat growing season

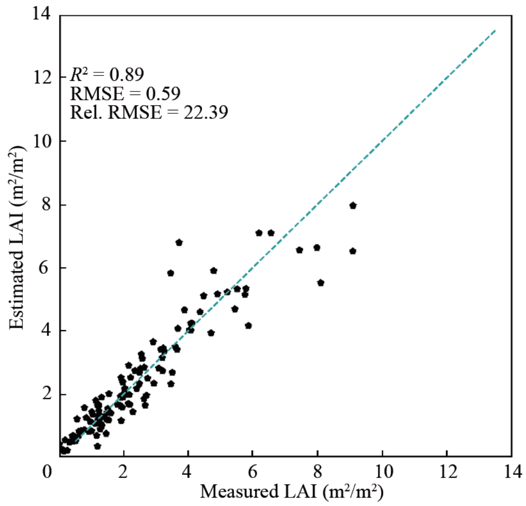
All the available data set have been inverted on the field quadrats on which LAI and chlorophyll have been measured for the evaluation of the results. This procedure has been adopted to reduce the processing time, as for the entire images the process it very time consuming, being pixel-by-pixel. The retrieved LAI and Cab parameters from both Sentinel 2 and Venμs data, were compared on a timely base with: (a) the two parameters measured in the field (Figure 2(a) and 2(b)); (b) other relevant biophysical crop relevant parameters, like C_m (dry matter content expressed in g/cm^2) and C_w (equivalent water thickness expressed in cm) (Fig. 2(c) and 2(d)). This is to check the interoperability of the two sensor (Sentinel-2 and Venμs) in monitoring the durum wheat biophysical parameters along the growing season. As examples, Fig. 2 shows the time plots of quadrat Q7 in field B041 (Maccarese farm, durum wheat cultivar) for LAI and Chlorophyll both retrieved by an own IDL (values identified by estimated − Est. in the legend) code for Sentinel 2 and Venμs using SNAP only for Sentinel-2.



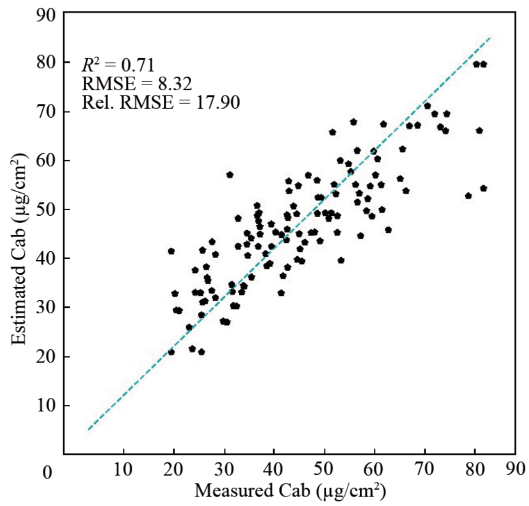
(a) LAI from PRISMA



(b) Cab from PRISMA

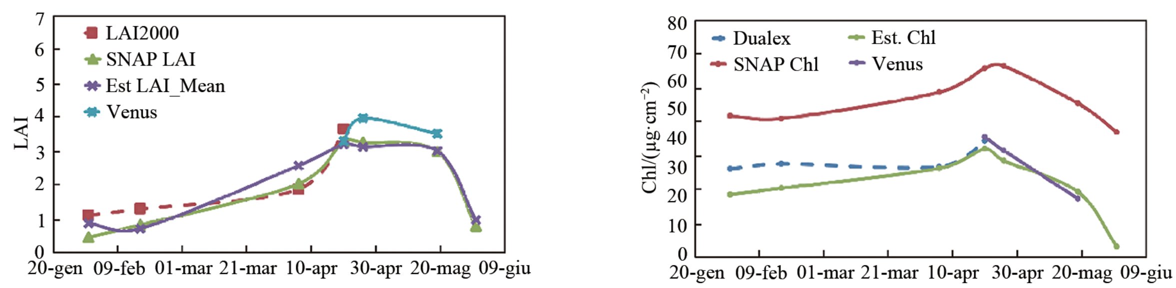


(c) LAI from Sentinel-2

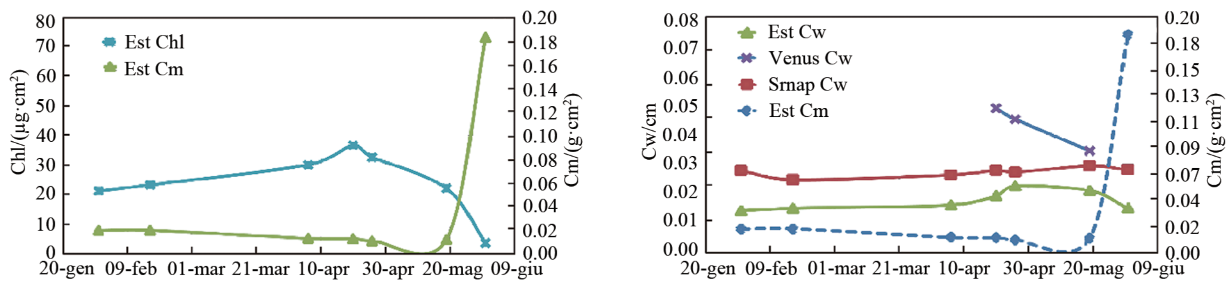


(d) Cab from Sentinel-2

Fig.1 LAI and Cab measured vs estimated using KRR for PRISMA and GPR for Sentinel-2



(a) LAI (b) Chl



(c) C_m (d) C_w

Fig.2 Examples for field B041-Q7 of time plots of (a) LAI and (b) Chl (Cab) retrieved using Sentinel-2 and Venμs data compared with field measured values. Time plots (c) and (d) show other biophysical parameters (C_m and C_w) estimated using PROSAIL

4 DISCUSSIONS

Evaluation of retrieval algorithms suitable for the next hyperspectral mission

The forthcoming hyperspectral mission PRISMA opens opportunities to implement retrieval algorithms in operational processing chains. The interest is on retrieval algorithms that are fast, accurate and robust, by exploiting the opportunities offered by the 247 available spectral channels of PRISMA. MLRA's and their continuous improvements and developments over time can cope with this strong non-linearity inherent data. Some previous works have demonstrated the use of kernel based GPR as a promising regressor in terms of processing speed and accuracy when using a local training dataset (Verrelst et al., 2012b). A review of different MLRA for biophysical variables retrieval is presented in Verrelst et al. (2012a, 2012b) where GPR has been proven competitive for the retrieval of LAI and Cab for Sentinel-2 data among other potential retrievals MLRA's.

In this work, we have simulated spectra and biophysical variables, LAI and Cab, using widely accepted leaf and canopy RTM PROSAIL, for currently operational multispectral sensor and for the forthcoming hyperspectral mission PRISMA. Different MLRA were tested on these simulated datasets.. Kernel based MLRA outperforms non kernel based MLRA in most cases. In case of PRISMA, however, bands had not been selected optimally and according to the vegetation properties. Band selection and dimensionality reduction techniques would reduce redundancy and spectral noise. This way, it can be expected that the retrieval and the results could be more accurate (Rivera-Caicedo et al., 2017).

Multispectral Sensors synergy for monitoring durum wheat growing season

In the Mediterranean area, in which optimal acquisition, in clear sky condition images, can be problematic, especially in the coastal areas like Maccarese, the possibility to have the synergy of two missions can be of real interest for monitoring the agronomical variables that vary rapidly. It is, therefore, of interest for the agronomical community to verify if Sentinel-2 and Venµs can be mutually applied for monitoring winter cultivar like durum wheat. For this purpose, we have inverted with PROSAIL on 18 measuring points (quadrats) more than 600 spectra to verify the retrieval outputs along the growing season.

By analyzing the LAI trend along the growing season, as presented in Fig. 2 we noticed that both inversion codes (SNAP and IDL) applied on Sentinel-2 images retrieved similar results in terms of RMSE values (see the 1.1 line, of 0.74 and 0.77 respectively for SNAP and the IDL code). In the date when we have at disposal contemporary Sentinel-2 and Venµs acquisitions, the retrieved LAI shows an R^2 of 0.79 between the two sensors. As regards Cab, instead, quite different values were retrieved by SNAP and IDL implemented code with respect to the ground measurements of SPAD. SNAP overestimates the Chlorophyll values with an RMSE of 17.02 µg. cm^{-2}, while the IDL code produces values with an RMSE of 10.94 µg. cm^{-2} with respect to the SPAD values. In the date of contemporary Sentinel-2 and Venµs acquisitions, the Cab retrieval has produced an R^2 of 0.92 between the two sensors, maintaining similar RMSE values with respect to ground true. Based on SNAP and IDL code accordance, we have also plotted the C_m and C_w versus time even though no samples have been collected for this purpose. Being not possible the quantitative assessment of these variables, only a qualitative evaluation of both trends were performed along the crop-growing season. It can be observed that C_m remains constant until chlorophyll remains at high values (e. g., higher than 20 µg. cm^{-2}), while C_m significantly increases when senescence starts (i. e. at the end of May). Similarly, C_w remains constant until the end of May, while chlorophyll decreases, dry matter increase so that water content starts to decrease. As regarding the retrieval of C_w using Sentinel-2 data, the values obtained using both SNAP and IDL are comparable while the ones obtained using Venµs data, are significantly high and with a trend, even though negative as it should be, too much steep with respect to a realistic one. This result confirms that SWIR bands are extremely useful for an accurate retrieving of crop biophysical parameters, as they act as tie-points in the PROSAIL inversion.

5 CONCLUSIONS

In this work, we have presented two activities performed within the ADEMS sub-project dealing with the identification of the optimal methods for the retrieval of crop biophysical parameters by using the current available multispectral imaging Sentinel-2 and Venµs satellite data or the next generation of hyperspectral satellite mission PRISMA.

We have explored different MLRA from the non-kernel and kernel based family for the retrieval of biophysical variables: LAI and Cab for Sentinel-2 and upcoming hyperspectral mission PRISMA. The results of retrieval of these variables are better in the case of kernel based MLRA as compared to non-kernel based MLRA. Kernel based MLRA's allows for a fast and accurate variable prediction, and they retain attractive properties to become next generation operational hybrid retrieval schemes. However, the problem of training with large simulated spectra generated by RTM persists and it can be triggered with active learning techniques. It can be suggested to implement these algorithms in operational chains in order to generate biophysical products from next generation satellite hyperspectral remote sensing data.

We have developed our own IDL code to invert PROSAIL applicable to any EO satellite data set using a squared error function. Tests have been performed using as input both on Sentinel-2 and Venµs data acquired on the experimental farm of Maccarese (Italy), having also at disposal LAI, Chlorophyll and Biomass crop measurements at the same date of satellite acquisitions. Results of IDL inversion code were compared with the SNAP code showing equivalent RMSE values for LAI, while for chlorophyll IDL code has produced a lower RMSE than SNAP results. Retrieval results were consistent with ground true data for both Sentinel-2 and Venµs data. Time plots have shown that Sentinel-2 and Venµs data are suitable to be combined together to follow the dynamic of the crop during the growing season allowing for higher points (crop biophysical variables estimations) density. The results show also that SWIR bands, even if spectrally broad (in Sentinel-2 data), are extremely useful to retrieve other biophysical variables like C_w, as they act as abovementioned as spectral tie points in the inversion process.

REFERENCES

Bacour C., Jacquemoud S., Tourbier Y., Dechambre M., Frangi J. -P., 2002. Design and analysis of numerical experiments to compare four canopy reflectance models. Remote Sens. Environ. 79, 72 – 83.

Baret F., Leroy M., Roujean J.L., Knorr W., Lambin E., Linderman M., 2003. CYCLOPES User Requirement Document. INRA-CSE Avignon.

Breiman L., 2001. Random forests. Mach. Learn. 45, 5 – 32.

Breiman L., 1996. Bagging predictors. Mach. Learn. 24, 123 – 140.

Breiman L., Friedman J.H., Olshen R.A., 2009. Stone, cj (1984) classification and regression trees. Wadsworth Belmont Calif.

Casa R., Upreti D., Pelosi F., 2019. Measurement and estimation of leaf area index (LAI) using commercial instruments and smartphone-based systems. Presented at the IOP Conference Series: Earth and Environmental Science, IOP Publishing, p. 012006.

Chai L., Qu Y., Zhang L., Wang J., 2008. Lai retrieval from cyclopes and modis products using artificial neural networks, in: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International. IEEE, pp. III-1064-III – 1067.

Cilia C., Panigada C., Rossini M., Meroni M., Busetto L., Amaducci S., Boschetti M., Picchi V., Colombo R., 2014. Nitrogen status assessment for variable rate fertilization in maize through hyperspectral imagery. Remote Sens. 6, 6549 – 6565.

Douak F., Melgani F., Benoudjit N., 2013. Kernel ridge regression with active learning for wind speed prediction. Appl. Energy 103, 328 – 340.

Draper N.R., Smith H., 2014. Applied regression analysis. John Wiley & Sons.

Durbha S.S., King R.L., Younan N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote Sens. Environ. 107, 348 – 361.

Friedman J., Hastie T., Tibshirani R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann. Stat. 28, 337 – 407.

Geladi P., Kowalski B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1 – 17.

Goffart J.P., Olivier M., Frankinet M., 2008. Potato crop nitrogen status assessment to improve N fertilization management and efficiency: past – present – future. Potato Res. 51, 355 – 383.

Guanter L., Kaufmann H., Segl K., Foerster S., Rogass C., Chabrillat S., ... & Straif, C. (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sensing, 7(7), 8830-8857.

Hagan M.T., Menhaj M.B., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. 5, 989–993.

Huang G., Huang, G.-B., Song S., You K., 2015. Trends in extreme learning machines: A review. Neural Netw. 61, 32–48.

Jacquemoud S., Verhoef W., Baret F., Bacour C., Zarco-Tejada P.J., Asner G.P., François C., Ustin S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66.

Jay S., Maupas F., Bendoula R., Gorretta N., 2017. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. Field Crops Res. 210, 33–46.

Lázaro-Gredilla M., Titsias M.K., Verrelst J., Camps-Valls G., 2014. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. IEEE Geosci. Remote Sens. Lett. 11, 838–842.

Pasolli E., Melgani F., Alajlan N., Bazi Y., 2012. Active learning methods for biophysical parameter estimation. IEEE Trans. Geosci. Remote Sens. 50, 4071–4084.

Rivera-Caicedo J.P., Verrelst J., Muñoz-Marí J., Camps-Valls G., Moreno J., 2017. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. ISPRS J. Photogramm. Remote Sens. 132, 88–101.

Rondeaux G., Steven M., Baret F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107.

Rouse J., Haas R., Schell J., Deering D., Harlan J., 1974. Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFC, Type III, ReportFinal. Greenbelt 371.

Saarela M., Elomaa T., Ruohonen K., 2010. An analysis of relevance vector machine regression, in: Advances in Machine Learning I. Springer, pp. 227–246.

Stefano P., Angelo P., Simone P., Filomena R., Federico S., Tiziana S., Umberto A., Vincenzo C., Acito N., Marco D., 2013. The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring, in: Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International. IEEE, pp. 4558–4561.

Suykens J.A., Vandewalle J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9, 293–300.

Upreti D., Huang W., Kong W., Pascucci S., Pignatti S., Zhou X., Ye H., Casa R., 2019. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. Remote Sens. 11, 481.

Vapnik V., Golowich S.E., Smola A., 1997. Support vector method for function approximation, regression estimation, and signal processing. Presented at the Advances in Neural Information Processing Systems, pp. 281–287.

Verrelst J., Alonso L., Camps-Valls G., Delegido J., Moreno J., 2012a. Retrieval of vegetation biophysical parameters using Gaussian process techniques. IEEE Trans. Geosci. Remote Sens. 50, 1832–1843.

Verrelst J., Camps-Valls G., Muñoz-Marí J., Rivera J.P., Veroustraete F., Clevers J.G., Moreno J., 2015a. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS J. Photogramm. Remote Sens. 108, 273–290.

Verrelst J., Dethier S., Rivera J.P., Muñoz-Marí J., Camps-Valls G., Moreno J., 2016. Active learning methods for efficient hybrid biophysical variable retrieval. IEEE Geosci. Remote Sens. Lett. 13, 1012–1016.

Verrelst J., Malenovský Z., Van der Tol C., Camps-Valls G., Gastellu-Etchegorry J.-P., Lewis P., North P., Moreno J., 2018. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surv. Geophys. 1–41.

Verrelst J., Muñoz J., Alonso L., Delegido J., Rivera J.P., Camps-Valls G., Moreno J., 2012b. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3. Remote Sens. Environ. 118, 127–139.

Verrelst J., Rivera J., Alonso L., Moreno J., 2011. ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. Presented at the Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK, pp. 11–13.

Verrelst J., Rivera J.P., Veroustraete F., Muñoz-Marí J., Clevers J.G., Camps-Valls G., Moreno J., 2015b. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. ISPRS J. Photogramm. Remote Sens. 108, 260–272.

Weiss M., Baret F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER.

Williams C.K., Rasmussen C.E., 2006. Gaussian processes for machine learning. MIT Press 2, 4.

Wold S., Esbensen K., Geladi P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.

Zhang C., Kovacs J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. Precis. Agric. 13, 693–712.