Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery

Giulia Tagliabue[a],*, **Mirco Boschetti[b]**, **Gabriele Bramati[a]**, **Gabriele Candiani[b]**, **Roberto Colombo[a]**, **Francesco Nutini[b]**, **Loredana Pompilio[b]**, **Juan Pablo Rivera-Caicedo[c]**, **Marta Rossi[a]**, **Micol Rossini[a]**, **Jochem Verrelst[d]**, **Cinzia Panigada[a]**

[a]Remote Sensing of Environmental Dynamics Laboratory, University of Milano - Bicocca, Milan, Italy

[b]Institute for Electromagnetic Sensing of the Environment, National Research Council, Milan, Italy

[c]Secretary of Research and Graduate Studies, CONACYT-UAN, Tepic, Nayarit, Mexico

[d]Image Processing Laboratory, University of Valencia, Valencia, Spain

Abstract

The recently launched and upcoming hyperspectral satellite missions, featuring contiguous visible-to-shortwave infrared spectral information, are opening unprecedented opportunities for the retrieval of a broad set of vegetation traits with enhanced accuracy through novel retrieval schemes. In this framework, we exploited hyperspectral data cubes collected by the new-generation PRecursore IperSpettrale della Missione Applicativa (PRISMA) satellite of the Italian Space Agency to develop and test a hybrid retrieval workflow for crop trait mapping. Crop traits were mapped over an agricultural area in north-east Italy (Jolanda di Savoia, FE) using PRISMA images collected during the 2020 and 2021 vegetative seasons. Leaf chlorophyll content, leaf nitrogen content, leaf water content and the corresponding canopy level traits scaled through leaf area index were estimated using a hybrid retrieval scheme based on PROSAIL-PRO radiative transfer simulations coupled with a Gaussian processes regression algorithm. Active learning algorithms were used to optimise the initial set of simulated data by extracting only the most informative samples. The accuracy of the proposed retrieval scheme was evaluated against a broad ground dataset collected in 2020 in correspondence of three PRISMA overpasses. The results obtained were positive for all the investigated variables. At the leaf level, the highest accuracy was obtained for leaf nitrogen content (LNC: r^2=0.87, nRMSE=7.5%), while slightly worse results were achieved for leaf chlorophyll content (LCC: r^2=0.67, nRMSE=11.7%) and leaf water content (LWC: r^2=0.63, nRMSE=17.1%). At the canopy level, a significantly higher accuracy was observed for nitrogen content (CNC: r^2=0.92, nRMSE=5.5%) and chlorophyll content (CCC: r^2=0.82, nRMSE=10.2%), whereas comparable results were obtained for water content (CWC: r^2=0.61, nRMSE=16%). The developed models were additionally tested against an independent dataset collected in 2021 to evaluate their robustness and exportability. The results obtained (i. *e.,*

LCC: r^2=0.62, nRMSE=27.9%; LNC: r^2=0.35, nRMSE=28.4%; LWC: r^2=0.74, nRMSE=20.4%; LAI: r^2=0.84, nRMSE=14.5%; CCC: r^2=0.79, nRMSE=18.5%; CNC: r^2=0.62, nRMSE=23.7%; CWC: r^2=0.92, nRMSE=16.6%) evidence the transferability of the hybrid approach optimised through active learning for most of the investigated traits. The developed models were then used to map the spatial and temporal variability of the crop traits from the PRISMA images. The high accuracy and consistency of the results demonstrates the potential of spaceborne imaging spectroscopy for crop monitoring, paving the path towards routine retrievals of multiple crop traits over large areas that could drive more effective and sustainable agricultural practices worldwide.

Keywords

Remote sensing; Earth Observation; Machine learning regression; Nitrogen content; Chlorophyll content; water content

1 Introduction

Reliable and accurate retrievals of vegetation traits are essential in numerous ecological applications such as precision agriculture (Lu et al.,2020), natural resource management (Oliver, 2016), flux modelling (van Bodegom et al., 2012; Scheiter et al., 2013) and functional diversity estimation (Tagliabue et al., 2020; Schweiger et al., 2017). In the context of agriculture, the retrieval of vegetation traits is key for crop status and growth monitoring, which is paramount within the current global change scenario to ensure sufficient and sustainable food production worldwide (Hank et al., 2019; Migdall et al., 2018).

An effective monitoring of the crop status requires characterizing leaf biochemical traits as well as canopy structural traits across space and over time. At the leaf level, the amount of photosynthetic pigments strongly determines the use of the incoming solar radiation for photosynthesis. Therefore, quantifying the leaf chlorophyll a and b content (LCC) is critical for monitoring the photosynthetic efficiency and early detecting crop stress (*e.g.*, chlorosis) (Estevez et al., 2021). Besides chlorophyll, the amount of leaf nitrogen content (LNC) is a critical trait because it is a proxy of the photosynthetic capacity. An optimal crop yield can only be obtained with a proper management of N, which in agricultural systems is applied through fertilisation. An insufficient N fertilisation inevitably leads to a decline of the crop yield in terms of quantity as well as grain quality (Wang et al., 2021). Conversely, overfertilisation causes plant stress (Albornoz, 2016) and environmental pollution, since the leftover N that remains in the soil can leach into groundwater bodies and accumulate in surface water bodies through infiltration and run-off, with well-known ecological consequences (Martmez-Dalmau et al., 2021). For these reasons, improved N fertilisation rates are instrumental in modern agriculture to maximise the crop yield while reducing the environmental impact. In this respect, LNC provides invaluable information about the actual crop nutritional status, which can be used to drive the N fertilisation and ensure an optimal crop production. Analogously, the quantification of leaf water content (LWC) provides useful information about the plant water status, which is needed to enhance the water use efficiency (Torres et al., 2019) through the adoption of optimal irrigation schemes. A more efficient and rational use of water is pushed by the future climatic

scenarios, which forecast more frequent and persistent drought events and heat-waves in the forthcoming decades (IPCC, 2021). Beyond these key leaf level traits, the leaf area index (LAI), defined as the one-sided green leaf area per ground surface unit, is one of the most important canopy traits to quantify since it represents the fundamental descriptor of vegetation density (Houborg et al., 2015; Haboudane et al., 2004). Since leaves are the main surface for mass and energy exchange between land and atmosphere, LAI largely controls the carbon, energy and water fluxes in the terrestrial ecosystems (Bonan, 1993; Chen and Black, 1991).

Remote sensing offers incomparable possibilities for quantitatively estimating these traits and supporting crop monitoring and management (Weiss et al., 2020; Gamon et al., 2019). However, mapping such traits from Earth Observation (EO) imagery is not straightforward for two main reasons. Firstly, the retrieval of vegetation properties from optical data is notoriously ill-posed, meaning that several combinations of traits might have the same effect on the optical properties of the canopy (Baret and Buis, 2008; Meroni et al., 2004; Combal et al., 2002). Secondly, the mapping of vegetation traits such as LNC, LCC and LWC inevitably requires a novel data stream of spaceborne data capable of better characterising the canopy reflectance (Berger et al., 2020; Liu et al., 2021; Hill et al., 2019).

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. While a number of studies already drew attention on the potential of spaceborne hyper-spectral imaging spectroscopy for the operational retrieval of vegetation traits (e.g., Rast and Painter, 2019; Nieke and Rast, 2018; Guanter et al., 2015), only a few investigated the actual performance of real satellite observations (Verger et al., 2011; Castaldi et al., 2016; Abdel-Rahman et al., 2013; Coops et al., 2003; Miphokasap and Wannasiri, 2018; Townsend et al., 2003). This is largely due to the rather limited availability of spaceborne sensors capable of providing hyperspectral observations so far. For the last twenty years, the only hyperspectral systems in space have been ESA's Compact High Resolution Imaging Spectrometer (CHRIS) on-board PROBA-1 and NASA's Hyperion onboard Earth Observing-1. Both satellite missions largely outreached their expected lifetime and provided invaluable indication of the advantage of spectrally contiguous spaceborne observations for the retrieval of vegetation traits. More recently, three more hyperspectral imaging spectrometers were launched into space. China's Advanced Hyper-spectral Imager (AHSI) onboard the Gao Fen-5 satellite and DLR' s Earth Sensing Imaging Spectrometer (DESIS) onboard the International Space Station (ISS) started operating since the mid of 2018 and ASI's PRe-cursore IperSpettrale della Missione Applicativa (PRISMA) was operational since the mid of 2019. However, barely no studies taking advantage of these new data streams for vegetation retrieval purposes are reported so far. To the best of our knowledge, only a recent study by Verrelst et al. (2021) exploited a PRISMA scene collected on an agricultural area in Germany to test the mapping capabilities of a canopy nitrogen content model developed using field spectroscopy data. In the near future, more spaceborne imaging spectroscopy missions such as ESA's Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) and FLuorescence EXplorer (FLEX), DLR's Environmental Mapping and Analysis Program (EnMAP), the Italian-Israeli Spaceborne Hyperspectral

Applicative Land and Ocean Mission (SHALOM) and NASA's Surface Biology and Geology (SBG) are expected to be launched in space, generating an unprecedented stream of new generation EO data.

The possibility to exploit contiguous spectral information from satellite imagery is particularly promising for the retrieval of leaf level traits which produce spectral changes in narrow regions of the electromagnetic spectrum. LNC is one of those, due to the presence of absorption features in the shortwave infrared (SWIR) spectral region. So far, a large number of studies exploiting optical data for the retrieval of the N content have estimated it indirectly through the relationship between N and chlorophyll content, given that these two traits are often correlated (Baret et al., 2007; Homolova et al., 2013). However, this assumption is flawed for two reasons: firstly, light harvesting complexes contain only a small amount of the total leaf N (~19%), while the largest amount (~70%) is contained in proteins (Chapin et al., 1987); secondly, the correlation between N and chlorophyll content does not hold across biomes and phenological stages (Berger et al., 2020; Hallik et al., 2009). For these reasons, LNC should be estimated through its relation with proteins rather than with chlorophyll molecules, as evidenced by previous studies (Berger et al., 2020). The presence of proteins produces multiple absorption features in the SWIR part of the spectrum (Curran, 1989). This broadly motivates the interest in hyperspectral data for leaf and canopy N quantification. Beyond N, other vegetation traits may benefit the forthcoming data stream of spaceborne hyperspectral imagery. Leaf and canopy water content might be one of those, since hyperspectral observations are able to capture subtle reflectance changes due to the presence of water (Pasqualotto et al., 2018; Das et al., 2017; Clevers et al., 2010). O–H bonds in water molecules in fact absorb light across the entire spectrum, but produce maximum absorption features at 970, 1200, 1400, 1940 and 2500 nm (Curran, 1989; Carter, 1991). Hence, contiguous spectral bands are essential to resolve these features and obtain accurate water content estimates. The usefulness of hyperspectral data for water content quantification is proven by the number of studies exploiting both proximal (Zhang et al., 2018; Zhang et al., 2021; Das et al., 2017; Mirzaie et al., 2014; Clevers et al., 2010) and remote sensing observations (Pasqualotto et al., 2018; Colombo et al., 2008). Also leaf chlorophyll content might be retrieved with enhanced accuracy exploiting contiguous spectral information from spaceborne sensors, as a few previous studies suggested. De Grave et al. (2020) estimated leaf chlorophyll content with a good accuracy from synthetic multispectral Ocean and Land Colour Instrument (OLCI) data, but the model performance improved considerably when using synthetic hyperspectral Fluorescence Imaging Spectrometer (FLORIS) data. Navarro-Cerrillo et al. (2014) compared hyperspectral and multispectral satellite sensors for estimating chlorophyll content in a Mediterranean pine plantation and found that the hyperspectral sensors yielded better results than the multispectral ones. Lu et al. (2019) obtained improved estimates of leaf chlorophyll content of grass species using hyperspectral data compared to multispectral data. Conversely, Lu et al. (2019) reported only a slight improvement in the estimation of canopy chlorophyll content using the hyperspectral data with respect to the multispectral ones.

Along with the feasibility to capture small variations in the spectral response of plants which is provided by hyperspectral observations, the quantitative estimation of vegetation traits considerably relies on the retrieval approach. So far, the majority of the studies

exploiting hyperspectral observations for the quantification of vegetation traits used statistical approaches, which rely on non-explicit relationships between the spectral data and the traits of interest (Verrelst et al., 2015). Non-parametric regression methods are particularly powerful and suited for hyperspectral data (jointly with a proper dimensionality reduction to mitigate multi-collinearity) since they leverage all the spectral information to build flexible models in a non-linear way. While these kinds of models are extremely effective, they have two major drawbacks: they require a large field dataset to properly train the model and they lack generalisation capacity. This second aspect is particularly relevant in view of future operational hyperspectral missions, since routine retrieval schemes have to be applicable across a wide range of vegetation types and conditions. While the previous studies exploiting hyper-spectral spaceborne observations provided indication that the retrieval of the traits of interest over large scales is possible, they all question the robustness and applicability of the developed statistical models in different contexts (Townsend et al., 2003; Coops et al., 2003; Abdel-Rahman et al., 2013). The hybrid approaches combine the efficiency and adaptivity of the non-parametric approaches and the generic properties of the physically-based methods, making them extremely promising candidates for the development of the processing chains for the retrieval of vegetation traits from satellite data (Verrelst et al., 2018).

In perspective of the upcoming operational hyperspectral spaceborne missions, the objective of this study was to demonstrate the feasibility to routinely retrieve a set of crop traits from space using a hybrid retrieval scheme. In this framework, we exploited for the first time multi-temporal hyperspectral imagery captured by PRISMA, which currently provides unparalleled opportunities to test new retrieval algorithms on spaceborne imagery. A set of PRISMA images acquired on an agricultural area in north-east Italy across the 2020 and 2021 vegetative seasons and corresponding ground validation data were exploited to: (i) optimise a hybrid-active learning retrieval scheme for the spaceborne retrieval of a complete set of leaf and canopy level crop traits; (ii) quantitatively assess the goodness of the developed models exploiting a broad ground dataset covering seven major crops; (iii) quantitatively assess the exportability of the developed models against an independent ground dataset collected in a different year; (iv) qualitatively evaluate the temporal consistency of the retrievals across the 2020 growing season and v) analyse the load of the spectral wavebands in the developed models and the contributing mechanisms in the retrieval of the investigated traits.

2 Data and methods

The steps of this work are described in the following sections. Firstly, we introduce the study site and field data collection occurred in 2020 and 2021. Secondly, we describe the PRISMA imagery acquisition focusing on the pre-processing approach used to remove artifacts and obtain smooth spectra. Finally, we describe the hybrid retrieval scheme used for crop trait mapping. The main steps of the retrieval workflow are summarized in Fig. 1.

2.1 Study site and field data collection

The site selected for this study is a large agricultural area located in north-east Italy, in the town of Jolanda di Savoia, FE (44.855061 N, 11.952233 E). The area belongs to Bonifiche Ferraresi S.p.A., a large farm of about 3850 ha which produces durum and soft wheat, barley, corn, sugar beet, alfalfa, soybean, rice and medicinal plants. The farm is managed by IBF-Servizi company, which provided geo-spatial information on the crops' distribution and on the agricultural practices. The site is characterised by a significant variability of the soil properties and by the massive presence of buried paleo-channels which are a testimony to the past of the area. The Po delta was in fact a former swamp area in which intense land reclamation activities were carried out in the late 1800s to expand the arable land. Fig. 2 provides an overview of the study site as seen by the PRISMA satellite. The 2020 land cover map is also shown. Fig. 3 shows the crop calendar of the investigated crops across the 2020 season to briefly provide an outlook of the development stage of each crop in correspondence of the PRISMA overpasses.

Three field campaigns were accomplished in the spring and summer of 2020 to collect spectral measurements and plant traits for the assessment of the satellite products and optimisation of the hybrid models. Two additional campaigns were performed in the late spring of 2021 to collect plant trait measurements to test the exportability of the developed models across different years. Field spectral measurements were collected on 14 July 2020 and 16 September 2020 on six homogeneous targets (*i.e.*, soybean, alfalfa, soil and crop residues) using a SR-3500 spectroradiometer (Spectral Evolution, Haverhill, MA, USA) covering visible, near-infrared and shortwave infrared wavelengths (350–2500 nm) for comparison against the PRISMA reflectance. The measurements were collected under clear sky conditions from approximately 1 m height above the canopy. For each target, 13 spectra collected within an area of 90×90 m^2 were averaged. The incident solar radiation was measured on a levelled, calibrated white reference panel (Spectralon; Labsphere, North Sutton, NH, USA) before each target measurement.

In 2020, plant traits were sampled in the field within three time windows scheduled close to three PRISMA acquisitions: 13–15 May (1st campaign), 1–2 July (2nd campaign) and 14–16 July 2020 (3rd campaign). The field measurements were collected within spatially homogeneous sampling areas of ~90×90 m^2 to ensure the comparability against PRISMA data. The sampling sites, hereafter referred to as elementary sampling units (ESUs), were geo-located with a commercial GPS. In total, 47 ESUs were sampled across the three campaigns. Six different crops (*i.e.*, durum and soft wheat, corn, alfalfa, sugar beet, soybean, and pea) were sampled to ensure a wide variability of the investigated traits due to morphological and phenological differences. The sampling protocol for each investigated trait (*i.e.*, leaf area index, leaf chlorophyll content, leaf water content, leaf nitrogen content) is described in the following paragraphs.

The leaf area index (LAI) was measured in the field using a LAI-2200 plant analyser (LI-COR Biogeosciences, Lincoln, NE, USA). In each ESU, four LAI-2200 readings were collected across four 10 m-transects (*i.e.*, starting from the center towards each corner of the ESU) with a ABBBBA sequence of above canopy (A) and below canopy (B) measurements.

The average LAI values were calculated using the dedicated FV2200 software (LI-COR Biogeosciences, Lincoln, NE, USA).

The leaf level traits were obtained from destructive measurements performed on leaf disk samples collected in the field using a 0.635 cm diameter leaf punch. In each ESU, we collected two sets of five Eppendorf tubes containing 12 leaf disks each. The first set was used for leaf chlorophyll content (LCC) extraction, while the second was used for leaf water content (LWC), leaf mass per area (LMA) and leaf nitrogen content (LNC) quantification. All the Eppendorf tubes were frozen and stored at –80 °C until the laboratory analysis.

For LCC quantification, the leaf disks contained in each Eppendorf tube ($Area_{tot}$=3.8 cm^2) were homogeneised in 5 ml of ice-cold methanol for 1 min (Ultra-Turrax; IKA-Werk, Staufen im Breisgau, Germany). The homogenate was kept at – 20 °C for 20 min. After centrifugation (4500 g, 10 min, 4 °C), the supernatant was recovered and kept at – 20 °C. The pellet was re-extracted in 5 ml of ice-cold methanol, kept at – 20 °C for 10 min and centrifuged. This operation was repeated twice. Finally, the three supernatants were merged, and 2 ml were filtered (0.45 μm PTFE syringe filter) and immediately analyzed for LCC quantification. The chlorophyll a (Chl_a) and b (Chl_b) concentrations (μg) were determined through spectrophotometry (V-630 UV–vis, Jasco, Pfungstadt, Germany) in a 100% methanol extract at 665.2 and 652.4 nm, respectively, while the turbidity was checked by measuring the absorbance at 750 nm. The Chl_a and Chl_b concentrations were calculated using the extinction coefficients proposed by Lichtenthaler and Buschmann (2001). The LCC content (μg cm^{-2}) was then calculated as LCC= (Chl_a+Chl_b)/$Area_{tot}$. All preparations and analyses were performed at low temperature and dim light.

For LWC and LMA quantification, the leaf disks were fresh and dry (after oven-drying at 80 °C for 48 h) weighted using an analytical balance with 0.0001 g sensitivity. The LWC (g cm^{-2}) and LMA (g cm^{-2}) were calculated according to the equations: LWC=($Weight_{fresh}$-$Weight_{dry}$)/$Area_{tot}$; LMA=$Weight_{dry}$/$Area_{tot}$.

The total N content (%) was determined by dry combustion with a N elemental analyzer (Flash EA 1112 NC-Soil; Thermo Fisher Scientific, Waltham, MA, USA). The LNC (g cm^{-2}) was then calculated as LNC=(N %/100) × LMA.

A summary of the descriptive statistics of the 2020 field data is provided in Table 1.

In 2021, we measured the same plant traits during two field campaigns performed on 25–27 May (4th campaign) and 22–24 June 2021 (5th campaign). During these two campaigns we sampled a total of 41 ESUs of four different crops, including rice, soybean, sugar beet and wheat. The plant traits were collected using the same protocol adopted in 2020 except for LCC, which was only measured indirectly in 2021. We used a MC-100 chlorophyll meter (Apogee Instruments Inc., Logan, UT, USA) which is factory-calibrated to measure chlorophyll concentration in leaves with units of μmol m^{-2}. For each ESU, a total of 60 MC-100 readings were performed on 15 fully expanded leaves drawn randomly. A conversion factor of 0.09 was applied to convert from μmol m^{-2} to μg cm^{-2} as proposed by Parry et al. (2014). Sugar beet was excluded from the analysis for leaf and canopy

chlorophyll content estimation since no calibration coefficients were available for this crop (Parry et al., 2014).

2.2 PRISMA imagery acquisition and pre-processing

In this study we exploited the newly available spaceborne hyper-spectral imagery captured by the PRISMA satellite of the Italian Space Agency (ASI), launched on 22 March 2019. PRISMA is a push broom imaging spectrometer featuring 240 spectral bands in the visible (VIS), near-infrared (NIR) and shortwave infrared (SWIR) spectral regions (400–2500 nm), with a nominal spectral sampling interval <11 nm and a bandwidth <15 nm. The 240 bands are resolved on 1000 across-track pixels with a 12-bit radiometric resolution. PRISMA has a ground spatial resolution of 30 m and a swath width of 30 km. The system is capable of off-nadir observations which are obtained through across-track or along-track roll operations. A recent study by Cogliati et al. (2021) performed a first analysis of the performance of the instrument, concluding that the results match the mission requirements and that the images are suitable for scientific applications.

Eight 30×30 km^2 PRISMA spot-images centered on Jolanda di Savoia fields were acquired between April and September 2020 (*i.e.*, 7 April, 11–17-23 May, 26 June, 14–31 July, 16 September 2020) to cover the entire phenological cycle of the investigated crops. Two additional PRISMA images were collected in the summer of 2021 (*i.e.*, 23 May and 21 June 2021) to be used as independent validation. The L2D PRISMA reflectance cubes (*i.e.*, geocoded at-surface reflectance) were downloaded from the PRISMA mission portal in HDF5 format and read using the `prismaread` tool (Busetto and Ranghetti, 2020). Due to atypical atmospheric conditions on 31 July 2020, the PRISMA image acquired on this date was re-processed using the ATCOR-3 software (ReSe Applications GmbH, Zurich, Switzerland). The atmospheric correction was performed on the L1 cube (*i.e.*, top-of-atmosphere spectral radiance in physical units) setting the atmosphere type and aerosol model to midlatitude summer and rural aerosol, respectively.

All at-surface reflectance cubes were then pre-processed in order to remove artifacts and obtain smooth spectra. The pre-processing was performed pixel-wise in R (R Core Team, 2020). Firstly, the spikes occurring along-track at specific wavelengths were removed using the `findpeaks` function implemented in the `pracma` package (Borchers, 2015) with a threshold of 0.018. Secondly, the spectral regions around the atmospheric gaseous absorptions presenting anomalous spikes and dips likely related to the spectral shift of the instrument were systematically excluded based on the comparison against top-of-canopy reflectance spectra collected in the field near simultaneously to PRISMA acquisitions (*i.e.*, 535–550, 755–780, 755–775, 810–855, 885–970, 1015–1050, 1080–1165, 1225–1285, 1330–1490,1685–1700, 1725–1750, 1780–1960, 1990–2030 nm). The remaining spectral bands were spline-smooth interpolated using the `SplineSmoothGapfilling` function included in the `FieldSpec-troscopyCC` package (Wutzler et al., 2016) with 60 degrees of freedom. Finally, the atmospheric water absorption regions (*i.e.*, 1350–1510 and 1795–2000 nm) and the last portion of the SWIR (i.e., 2320–2500 nm) were excluded. The resulting PRISMA spectra feature 170 smooth spectral bands from 400 to 2320 nm. An example of

the PRISMA spectra before and after the spline interpolation is displayed in Fig. 4, together with the corresponding field-based spectra acquired nearly simultaneously.

2.3 Hybrid retrieval of crop traits

The retrieval of crop traits from PRISMA imagery was based on hybrid modelling. The hybrid retrieval scheme relies on the use of physically based radiative transfer models for the forward simulation of a set of spectral responses as a function of the model input variables, and on the use of machine learning regression algorithms to learn the relationships between the simulated spectra and the model input variables (Berger et al., 2021; Berger et al., 2020; Verrelst et al., 2020). The model trained on the simulated dataset is finally applied to the real spectra for estimating the traits of interest.

For the generation of the simulated dataset we used the leaf level PROSPECT-PRO radiative transfer model (Feret et al., 2021) coupled with the canopy level 4SAIL model (Verhoef and Bach, 2007). PROSPECT-PRO is the latest development of the PROSPECT model (Jacquemoud and Baret, 1990), that includes the separation between the nitrogen-based (C_p) (*i.e.*, proteins) and the carbon-based components (CBC) (*i.e.*, cellulose, hemicellulose, lignin, starch and sugars) which constitute the leaf dry matter content (or leaf mass per area, LMA). This refinement allows the retrieval of a complete set of leaf level traits, including the CBC and the LNC, which can be estimated from the C_p using a protein-to-nitrogen factor of 4.43 proposed by Yeoh and Wee (1994) as previous studies suggested (Wang et al., 2018; Berger et al., 2020). The coupling with the 4SAIL model allows simulating the radiative transfer within the turbid medium canopy layer, thus providing a set of simulated reflectance spectra as a function of leaf (*e.g.*, leaf chlorophyll content, leaf water content, leaf protein content) as well as canopy-level (*e.g.*, leaf area index, average leaf angle) model input variables. The parameterisation of the coupled PROSPECT-PRO-4SAIL model is synthesized in Table 2. The model was run in forward mode to simulate a Look-Up Table (LUT) of 1000 spectra as a function of 1000 model input combinations drawn randomly with a uniform distribution within the specified ranges. The 1000 simulated spectra generated using the PROSPECT-PRO-4SAIL model and the corresponding model input variables were used to train a Gaussian processes regression (GPR) model. GPR is a powerful non-parametric probabilistic model based on the Bayesian theory (Rasmussen and Williams, 2006). GPR relies on Gaussian processes, which are described by a mean and covariance and essentially represent probability distributions over all the possible functions that are consistent with the observed data. As all the Bayesian approaches, the GPR starts with a prior distribution which is updated as more data points are observed, producing a posterior distribution. This characteristic allows the GPR to provide a standard deviation of the estimate together with the predictive mean. These two metrics allow the calculation of the coefficient of variation (CV) (*i.e.*, standard deviation/ mean×100), which provides a measure of the uncertainty of the retrieval and can be used as a criterion to exclude uncertain pixels. Another valuable feature of GPR is its capacity to provide information about the band relevance within the non-linear regression model which is built. The hyperparameter sigma of the GPR covariance function can in fact be interpreted as the inverse of the information content of each input band (Rasmussen and Williams, 2006). Hence, low sigma values indicate a high relevance of the considered band and vice versa.

The model was trained using as input the first 15 synthetic bands obtained through a Principal Component Analysis (PCA) in order to remove band collinearity and spectral information redundancy, as previous studies suggested (Rivera-Caicedo et al., 2017; De Grave et al., 2020). Besides the dimensionality reduction in the spectral domain, the reduction in the training sample size through active learning (AL) algorithms was explored as a tool for maximising the retrieval performance and speed. The AL helps reduce the size of the initial set of simulated data by extracting only the most explanatory spectra while discarding the unrealistic ones with respect to the real world conditions. Still, the model is built exclusively on the simulated data. The in situ data are only needed to select the best performing model out of the thousand models which are tested while reducing the dimensionality of the initial LUT. As a next step, the developed models were further tested on a completely independent field dataset in order to investigate their transferable capacity. A recent review from Berger et al. (2021) examined the studies exploiting AL techniques for optimising the training dataset in hybrid retrieval frameworks. Since different studies showed different convergence patterns depending on the investigated variable, both uncertainty *(i.e.,* variance-based pool of regressors (PAL) and residual regression active learning (RSAL)) and diversity (*i.e.*, Euclidean distance-based diversity (EBD), angle-based diversity (ABD) and clustering-based diversity (CBD)) based AL criteria were tested. For the optimisation of the training dataset through AL, an initial pool of 20 spectra (2% of the LUT) was drawn randomly from the LUT. Then, a new training sample was added to the initial pool according to the selected AL criterion, and the retrieval performance was evaluated against the 2020 ground dataset in terms of root mean square error (RMSE). If the added training sample reduced the RMSE, the training sample was kept, otherwise it was discarded. This process was repeated iteratively for each AL criterion 1000 times in order to ensure the evaluation of all the training samples of the LUT. The goodness-of-fit metrics for each added sample were recorded for each AL algorithm and trait to evaluate the sensitivity of the different AL algorithms to the investigated traits. The machine learning step of the hybrid workflow, including the evaluation of the AL criteria on the hybrid model performance, was performed within the Automated Radiative Transfer Models Operator (ARTMO) machine learning regression algorithms toolbox (https://artmo-toolbox.com) (Caicedo-Rivera et al., 2014).

The performance of the hybrid models in the retrieval of the traits of interest exploiting both the 2020 and 2021 datasets was evaluated in terms of standard goodness-of-fit statistics: coefficient of determination (r^2), RMSE, normalised RMSE (nRMSE) (*i.e.*, RMSE/range of measured values), bias (*i.e.*, mean of estimated values - mean of measured values) and relative bias (rbias) (*i.e.*, bias/mean of measured values) between measured and estimated values.

2.4   PRISMA multi-temporal mapping

The best performing models in terms of r^2 and RMSE were applied to the PRISMA dataset to map the spatio-temporal variability of the crop traits across the growing season. The output consisted of eight PRISMA maps for each investigated trait with the associated pixel-by-pixel uncertainty expressed in percentage as CV.

To check the model transferability and robustness when applying it to different conditions, the temporal evolution of the traits of the main crops in the farm was examined. At this purpose, the shapefile of the 2020 land cover provided by IBF-Servizi was used to extract from all fields the pixels belonging to the same crop type (*i.e.*, alfalfa, wheat, sugar beet, corn and soybean) on each date. An inner buffer of 60 m (*i.e.*, 2 PRISMA pixels) was applied to each field to avoid mixed pixels at the edges. The fields presenting bare soil, crop residues or low vegetation fractional cover were masked based on a threshold of 0.4 on the Normalised Difference Vegetation Index (NDVI) (Rouse et al., 1974). In addition, the fields presenting clouds or cloud shadows were excluded from the analysis. The multi-temporal distribution of the crop traits was then analysed in relation to the expected behaviour of each crop based on its management practice and seasonal development.

2.5 Analysis of the contribution of different spectral regions in the retrieval of the crop traits

To understand the importance of the different spectral regions in the retrieval of the examined traits, we delved into the contribution of each waveband in the developed models. This analysis was carried out exploiting the information encompassed in the sigma values provided by the GPR. For each model, the eigenvectors of the 15 PCA used as input of the machine learning regression were normalised by the inverse of the sigma values rescaled between 0 and 1. In other words, each PCA was multiplied by a coefficient proportional to its importance in the model (*i. e.*, the PCA with the highest informative contribution was multiplied by 1, the one with the lowest informative contribution was multiplied by 0). Then, the normalised eigenvectors were summed for each wavelength and averaged across three spectral domains: VIS (*i.e.,* 400–750 nm), NIR (*i.e.*, 750–1300 nm) and SWIR (*i.e.*, 1300–2500 nm).

3 Results

3.1 Crop trait retrievals from PRISMA imagery

The scatter plots showing the measured data against the estimates obtained from the 2020 PRISMA hyperspectral dataset with the hybrid GPR-AL model are displayed in Fig. 5. The summary of the goodness-of-fit statistics between the measured and estimated 2020 data are reported in Table 3. For each investigated trait, the best performing AL algorithm is reported together with the corresponding goodness-of-fit metrics.

A positive and statistically significant correlation was found for all the investigated traits. Among the leaf level traits, the best results were obtained for LNC (r^2=0.87, nRMSE=7.5%), while slightly worse results were obtained for LCC and LWC (r^2=0.67, nRMSE=11.7% and r^2=0.63, nRMSE=17.1%, respectively). LAI was estimated accurately (r^2=0.82, nRMSE=10.3%). Among the canopy level traits, CNC and CCC were retrieved with higher accuracy than the corresponding leaf level traits (r^2=0.92, nRMSE=5.5% and r^2=0.82, nRMSE=10.2%, respectively). Conversely, comparable results between LWC and CWC were observed (r^2=0.61, nRMSE=16%). The rbias was below 3% for all the investigated variables, evidencing the accuracy of the retrievals also in terms of absolute values.

Respect to the AL algorithm used for optimising the training dataset, the best results were obtained using different AL algorithms for the different traits, however in most cases other AL criteria yielded only slightly worse results. In general, the diversity based AL algorithms performed better than the uncertainty based ones. The only exceptions were CCC and CWC, but in both cases the results obtained using EBD and CBD were alike.

The results obtained applying the hybrid GPR-AL models optimised on the 2020 dataset on the independent data collected in 2021 are shown in Fig. 6. Also in this case the correlations between the measured and estimated plant traits were positive and statistically significant for all the considered traits. The retrieval accuracies were mostly comparable to the ones obtained in 2020, despite in 2021 the in situ data were not used to optimize the LUT used in the models. In particular, solid results were obtained for LAI (r^2=0.84) and for all the canopy level traits (r^2=0.79, r^2=0.62 and r^2=0.92 for CCC, CNC and CWC, respectively). At the leaf level, the results were accurate for LWC (r^2=0.74), while modest for LCC (r^2=0.62) and poor for LNC (r^2=0.35). For LCC, the coefficient of determination was in line with the one obtained using the 2020 data (r^2=0.67), however a general overestimation of the estimated values was observed, especially for low LCC values. The retrieval accuracy of LNC is inferior to the one obtained in 2020, however a lower range of variation of the measured trait is also observed. The summary of the goodness-of-fit statistics calculated between the measured and estimated 2021 data is reported in Table 4.

3.2 Multi-temporal crop trait maps from PRISMA imagery

The multi-temporal crop trait maps obtained applying the hybrid GPR-AL models to PRISMA imagery are displayed in Fig. 7 and Fig. 8. On each date, the fields with bare soil or crop residues (*i.e.*, winter crops after harvesting, summer crops before germination or alfalfa fields after mowing), clouds or cloud shadows were masked in order to extract only the vegetated fields. The maps show reliable patterns with a range of variation of the retrieved traits similar to the expected one. The interfield variability related to the presence of different crops and management practices is evident. A strong intra-field heterogeneity is also well observable on the maps. This variability is strongly related to the presence of buried paleo-channels and to the influence of the soil properties (Crema et al., 2020).

For each crop, the trait value distributions were extracted and plotted to analyse their temporal evolution across the 2020 phenological season. Fig. 9 shows the temporal course of LCC, LNC, LWC and LAI in the eight PRISMA dates for the main crops in the farm.

LAI (Fig. 9d) increases in wheat fields reaching its relative maximum on DOY 132, then it starts decreasing due to the incoming senescence stage (Fig. 3). Sugar beet fields show increasing LAI values from DOY 132 up to DOY 178, while they slightly decrease on DOY 196. The bimodal distribution observable in sugar beet is related to the presence of a small sugar beet field where the plants grew poorly. Corn fields are not mapped on DOY 132, 138 and 144 due to the very low signal related to the low fractional cover (*i.e.*, NDVI<0.4), while they are already at the maturity stage when captured on DOY 178 and 196. After that, they show decreasing LAI values on DOY 213 due to the plant drying before harvesting. In soybean fields, the LAI increases along with the plant growth until the plateau on DOY 196, then it declines due to senescence. Alfalfa fields are characterised by swinging LAI values

due to the mowing practices which happened four times across the season, on DOY 122, 165, 209 and 254. In general, the highest LAI values are observed around the peak of the development of the winter (*i.e.*, mid of May) and summer crops (*i.e.*, mid of July), while the lowest LAI values are observed at the transition between winter and summer crops and at the end of the vegetative season.

LCC (Fig. 9a) shows more stable values respect to LAI. The LCC values are relatively stable across the season within the same crop, while the main differences are observed among crops. Sugar beet, which presents bright green leaves, shows the lowest LCC values (*i.e.*, 25–40 –g cm^{-2}). In particular, very low LCC values (*i.e.*, ~25 µg cm^{-2}) were observed in the poorly grown field, where the leaves appeared yellowish. Wheat and corn, which present dark green leaves, are characterised by the highest LCC values (*i.e.*, 50–60 µg cm^{-2}). Sugar beet and soybean show a clear temporal evolution, with increasing LCC values from DOY 178 and 196, respectively. Conversely, wheat presents a slight decrease in LCC values over time due to the progressive degradation of chlorophyll pigments.

LWC (Fig. 9c) is quite similar among the different crops, with gaussian distributions centered around 0.006–0.01 g cm^{-2}. The only exception is represented by sugar beet, which shows a significantly higher water content compared to the other crops (*i.e.*, ~0.02 g cm^{-2}). Within the single crops, some temporal trends can be observed, for example wheat shows decreasing values from DOY 132 to DOY 144, which are due to the gradual wheat drying during the reproductive stage (Fig. 3).

LNC (Fig. 9b) is also fairly similar among the crops, with some exceptions. The highest LNC (*i.e.,* ~0.00025 g cm^{-2}) is observed in wheat fields. The other crops show similar LNC distributions, with lower values (*i.e.,* ~0.00018 g cm^{-2}) for corn and intermediate values for alfalfa and soybean (*i.e.,* ~0.00020 g cm^{-2}). Wheat shows decreasing LNC values over time, which are related to the nitrogen translocation happening in cereals during the maturity stage.

At the canopy level, the value distributions mirror the behaviour observed in the corresponding leaf level traits scaled by LAI.

Table 5 reports the average (±standard deviation) coefficients of variation of the estimates on the different PRISMA dates for the different traits. The CVs appear relatively stable over time, indicating that the model is consistent and robust across the phenological season. LCC presents CV values which are the lowest among the leaf level traits and similar to LAI values. Overall, the highest CV values are observed on DOY 260 for all traits. This is likely due to the influence of senescent vegetation and bare soil, for which the model was not properly trained.

3.3 Contribution of the spectral domains in the hybrid models developed

Fig. 10 shows the contribution (%) of the spectral domains in the retrieval of LAI, LCC, LNC and LWC through hybrid models. The VIS spectral region is critical for the retrieval of LCC, while the NIR and SWIR spectral regions are more relevant in the estimation of LAI and LNC. In particular, the SWIR region appears determining in the quantification of

LNC. NIR and SWIR are almost equally important for the estimation of LWC. The opposed contribution of the VIS and SWIR spectral regions to the quantitative estimation of LCC and LNC is shown more in detail in Fig. 11. The plot shows the load of each single spectral waveband in the LCC and LNC models, respectively. The contribution of the SWIR bands is consistently higher in the LNC model across the entire SWIR spectral domain. Conversely, the LCC model is characterised by a higher contribution of the VIS bands, though some influence of the VIS is also observed in the LNC model.

4   Discussion

4.1   Performance of crop trait retrievals from PRISMA imagery

Our study showed that spaceborne hyperspectral imaging spectroscopy can be used to track the temporal evolution of vegetation traits in agricultural ecosystems across a range of crop types and phenological stages. An unprecedented multi-temporal hyperspectral spaceborne dataset composed of eight images captured in 2020 by the newly available PRISMA satellite and of corresponding field measurements was exploited to retrieve a set of leaf and canopy level traits which are key in the context of crop monitoring and management.

LNC and CNC were estimated with the highest accuracy (r^2 of 0.87 and 0.92 and nRMSE of 7.5% and 5.5%, respectively) among all the investigated traits. The explained variance obtained in N retrieval is remarkably increased compared to the one reported in previous studies exploiting satellite data (*i.e.*, r^2=0.62-0.78) (Coops et al., 2003; Abdel-Rahman et al., 2013; Miphokasap and Wannasiri, 2018) and analogous or increased respect to the one obtained exploiting systematic ground-based measurements (Berger et al., 2018; Vigneau et al., 2011; Zhou et al., 2018), in which the influence of potentially disturbing factors such as the atmosphere and the observation geometry is minimised. In addition, it must be noted that most studies targeted single crops (mainly winter wheat, rice and corn) in which the N variability was either induced with ad hoc fertilisation treatments or related to the growing phase (*e.g.*, Berger et al., 2018; Zhou et al., 2018; Liang et al., 2018). This biases the model performance towards generally higher accuracy and strongly hinders the applicability of the developed models over other land covers regardless of the retrieval approach adopted (Lee et al., 2004; Townsend et al., 2003). Furthermore, it is well known that the variability of the biochemical traits at the canopy scale is mostly driven by the LAI variation related to the phenological development (Schiefer et al., 2021; Vohland et al., 2010; Sehgal et al., 2016). In our study instead, the model performance was assessed over seven different crop types grown in real conditions and representing some of the most widespread crops worldwide in order to ensure a better evaluation and applicability of the model.

Leaf and canopy water content were estimated with r^2=0.63, nRMSE=17.1% and r^2=0.61, nRMSE=16%, respectively. These were the most challenging variables to retrieve, leading to the least accurate results among the investigated traits. However, it must be noted that the distribution of the ground data is skewed, with a low range of variability of LWC. All the sampled crops in fact present LWC values between 0.0058 and 0.0123 g cm^{-2}, while only three ESUs sampled on pea and nine ESUs sampled on sugar beet show higher LWC values, ranging between 0.0131 and 0.0254 g cm^{-2}. This is partly related to the fact that the study site is an irrigated farming system that aims at maximising the crop production by

minimising potential water stresses. For this reason, no extreme LWC values are observed. Previous studies exploiting both proximal (Zhang et al., 2018; Zhang et al., 2021; Das et al., 2017; Mirzaie et al., 2014; Clevers et al., 2010) and remote sensing observations (Pasqualotto et al., 2018; Colombo et al., 2008) obtained different degrees of accuracy. Although a proper comparison among those studies is difficult due to different experimental conditions and retrieval approaches, the r^2 between measured and estimated water content values range between 0.59 and 0.84 for leaf water content and between 0.51 and 0.94 for canopy water content. While our results are in line with most of the studies conducted at satellite scale, the moderate r^2 and the higher coefficients of variation obtained indicate that the retrievals of LWC and CWC still need to be improved to become operational.

Conversely, leaf and canopy chlorophyll content were estimated accurately (r^2=0.67, nRMSE=11.7% and r^2=0.82, nRMSE=10.2%, respectively). The accuracy is higher for canopy chlorophyll content likely due to the compensating effect of LAI (Estevez et al., 2021; Xie et al., 2019), but the retrievals are also accurate at the leaf scale. This is further proved by the spatial patterns within the maps, by the temporal evolution of the investigated crops and by the low coefficients of variation of the estimates obtained. Previous studies exploiting hyper-spectral data acquired at different scales yielded diverse results. As an example, Doktor et al. (2014) retrieved leaf chlorophyll content of summer barley from field spectral measurements with r^2=0.89, Lu et al. (2019) estimated the leaf and canopy chlorophyll content of grass species from airborne data with r^2=0.45, RMSE=3.32 µg cm^{-2} and r^2=0.85, RMSE=12.4 µg cm^{-2}, respectively, Tagliabue et al. (2019) estimated leaf chlorophyll content of mixed broadleaf species from airborne data with r^2=0.65, RMSE=5.66 µg cm^{-2} and Navarro-Cerrillo et al. (2014) estimated leaf chlorophyll content of mediterranean pine from satellite images with r^2=0.56–0.65 depending on the sensor considered. Consistently with our findings, the leaf level retrievals were generally less accurate than the canopy level ones.

4.2 Hybrid retrieval scheme

The large majority of the studies exploiting real satellite observations for the retrieval of vegetation traits used non-parametric regression methods. Although these models are notably powerful, they severely lack transferable capacity therefore they can hardly be exploited in an operational framework. Conversely, hybrid models endow the exportability of the physically based models as well as the strength of statistical models, thus being ideal candidates for the implementation within operational retrieval schemes. This is of paramount importance in preparation of the forthcoming availability of hyperspectral data streams from new satellite missions, which will demand generic, robust and computationally fast models for producing high quality vegetation trait maps at the global scale.

Overall, the results obtained in this study using a hybrid approach are in line or improved compared to previous studies, depending on the considered trait. Most literature testing hybrid approaches focused on the retrieval of LAI, with r^2 ranging between 0.63 and 0.89 and RMSE ranging between 0.37 and 0.73 m^2 m^{-2} (Brown et al., 2019; Schlerf and Atzberger, 2006; Upreti et al., 2019; Verger et al., 2011; Doktor et al., 2014; Pipia et al., 2021; Estevez et al., 2021). Two studies also examined the use of AL algorithms

for the hybrid estimation of LAI: Upreti et al. (2019) tested different AL criteria for the retrieval of winter wheat LAI from Sentinel-2 data (r^2=0.78, RMSE=0.68 m^2 m^{-2}) and Pipia et al. (2021) used Sentinel-2 data to estimate the LAI green of various crops (r^2=0.63, RMSE=0.73 m^2 m^{-2}). Some studies also exploited hybrid approaches for the retrieval of LCC and CCC from satellite data. Zhou et al. (2020) used Landsat-8 data to estimate LCC of winter wheat through Gaussian processes regression trained on PROSAIL simulations (RMSE=12.43–16.44 µg cm^{-2}), Upreti et al. (2019) used Sentinel-2 data to estimate LCC of winter wheat using a hybrid AL approach (r^2=0.26, RMSE=8.88 µg cm^{-2}) and Estevez et al. (2021) exploited top-of-atmosphere Sentinel-2 data to retrieve LCC of corn and winter wheat using PROSAIL simulations and Gaussian processes regression (r^2=0.47, RMSE=6.48 µg cm^{-2}). In case of LCC, our retrieval scheme based on hybrid AL modelling applied on PRISMA imagery yielded largely improved results compared to previous literature (r^2=0.67, RMSE=5.88 µg cm^{-2}). Instead, our CCC retrievals (r^2=0.82, RMSE=0.36 g m^{-2}) are in line with previous studies exploiting hybrid approaches (r^2=0.69–0.85, RMSE=0.39–0.52 g m^{-2}) (Brown et al., 2019; Upreti et al., 2019; Estevez et al., 2021). Only a few studies exploited hybrid approaches for the retrieval of water content. Estévez et al. (2021) estimated LWC and CWC of corn and winter wheat from S2 images with r^2=0.23, RMSE=0.0074 g cm^{-2} and r^2=0.82, RMSE=139 g m^{-2}, respectively, and Trombetti et al. (2008) used MODIS data to retrieve the CWC of various vegetation canopies with r^2=0.7 and RMSE=0.0063 g cm^{-2}. In this case, the retrievals we obtained from PRISMA data were superior at the leaf scale and poorer at the canopy scale.

4.3   Model exportability

The results obtained in the studies exploiting hybrid approaches for the retrieval of vegetation traits evidence that they are in some cases less performing than the pure machine learning approaches (Zhou et al., 2020). The use of AL techniques enhances the effectiveness of the hybrid models due to the fact that only the most informative spectra are pooled out of larger LUTs which may contain redundant information and unrealistic combinations (Berger et al., 2021). Though, the increase in the retrieval accuracy has the natural drawback of a decrease in the model exportability, since ground data are required for an optimal sample selection. Within this study, a broad ground dataset including several crop types at various development stages was gathered. Future operational retrieval schemes can build upon this and other similar datasets to become more and more flexible and reliable on a broad variety of vegetation types and conditions. In our study, the model exportability was investigated in two ways: i) a quantitative assessment was performed by applying the developed models on an independent dataset collected in 2021 and ii) a qualitative assessment was performed by applying the model on multi-temporal images collected in 2020 at different phenological stages of the target crops. In both cases the models appeared to be fairly robust and exportable. Positive results were obtained applying the models to the 2021 dataset for all the investigated traits (r^2=0.35–0.92). Especially promising results were obtained for the canopy level traits, while the leaf level ones (*e.g.*, LNC) still present room for improvement. It is worth noting that the 2021 dataset also includes rice, which was not present in the 2020 dataset used for the model optimisation. This further suggests that the developed models are not much dataset dependent. Additionally, the analysis of the temporal evolution of the investigated crops across the 2020 season showed consistency

with the expected behaviour based on the crop management and consequent phenological development. This happens although the models were optimised based on the ground data collected in correspondence of just three out of the eight PRISMA images captured over the site in 2020. The consistency of the retrievals across all the images is also proved by the fact that the uncertainties (*i.e.*, the coefficients of variation of the estimates) for the different traits are relatively stable over time for the different crops (Table 5). This indicates that the retrieval is neither sensitive to the different scene conditions (*i.e.*, sunsensor geometries, crop phenological stage, atmospheric conditions) nor related to the presence of ground data for the AL tuning on specific dates. Yet, from an operational perspective, the ground dataset should be further broadened by including more vegetation types and conditions in order to build more generic and solid models which may be applied at larger spatio-temporal scales.

4.4 Spectral information in the developed hybrid models

The analysis of the contribution of the different spectral domains in the developed hybrid models enlightened some of the mechanisms underlying the remote assessment of the investigated crop traits. The contribution of the SWIR spectral region is key for the quantitative estimation of the LNC, which is in line with recent literature about the retrieval of nitrogen from remotely sensed data (Perich et al., 2021; Féret and de Boissieu, 2020; Berger et al., 2020). Instead, the VIS part of the electromagnetic spectrum was the most relevant in the LCC model, as well established in literature (Gitelson et al., 2003; Van Wittenberghe et al., 2014). The contrasting behaviour we found in the developed LCC and LNC models (*i.e.,* highest information content in the VIS for LCC, lowest information content in the VIS for LNC and the other way round in the SWIR, see Fig. 10 and Fig. 11) suggests that the retrieval of LNC is effectively driven by the protein rather than by the pigment content. The former in fact mostly absorbs in the SWIR due to the N-H bonds in the amino acids (Curran, 1989), while the latter influences the VIS due to the electron transition in the chlorophyll a and b molecules (Curran, 1989). This strongly evidences the need to exploit the SWIR region for the quantitative estimation of LNC and elucidates the advantage of hyperspectral imaging spectroscopy.

The contiguous spectral information provided by the hyperspectral sensors also facilitates the retrieval of the water content (Van Witten-berghe et al., 2014), which produces absorptions from 970 up to 1940 nm (Curran, 1989). Indeed, the analysis of the spectral information content in the LWC model evidenced that the wavebands located in the NIR and SWIR spectral regions are the ones the retrieval is most sensitive to (Fig. 10).

5 Conclusions

In this study we exploited for the first time multi-temporal images collected by the PRISMA hyperspectral satellite to develop and test novel approaches for the large-scale mapping and monitoring of key traits in agriculture. The comparison between the estimates and the ground data collected near simultaneously three PRISMA acquisitions confirmed the powerfulness of the hybrid approaches in the retrieval framework. All the investigated traits were in fact estimated accurately (r^2=0.61–0.92, nRMSE=5.5–17.1%). Moreover, the application of the developed models on a completely independent dataset collected in the

following year allowed to verify their transferability in different conditions (r^2=0.35–0.92, nRMSE=14.5–28.4%). The consistency of the temporal courses extracted with respect to the expected phenology of the investigated crops suggested a high reliability of the models across a wide range of vegetation and environmental conditions. These promising results demonstrate that the retrieval of a broad set of leaf and canopy traits from space using hybrid retrieval schemes is feasible. This paves the way for future algorithms for the routine mapping of vegetation traits from operational spaceborne sensors such as CHIME, which is expected to provide global products in support of agriculture and natural ecosystem management.

Acknowledgements

References

Abdel-Rahman EM, Ahmed FB, Ismail R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int J Remote Sens. 2013; 34: 712–728. DOI: 10.1080/01431161.2012.713142

Albornoz F. Crop responses to nitrogen overfertilization: A review. Sci Hortic. 2016; 205: 79–83. DOI: 10.1016/j.scienta.2016.04.026

Baret, F, Buis, S. Advances in Land Remote Sensing. New York, USA: 2008. Estimating canopy characteristics from remote sensing observations: review of methods and associated problems; 173–201.

Baret F, Houles V, Guerif M. Quantification of plant stress using remote sensing observations and crop models: The case of nitrogen management. J Exp Bot. 2007; 58: 869–880. DOI: 10.1093/jxb/erl231 [PubMed: 17220515]

Berger K, Caicedo Rivera JP, Martino L, Wocher M, Hank T. A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data. Remote Sens. 2021; 13: 1–23.

Berger K, Verrelst J, Feret JB, Hank T, Wocher M, Mauser W, Camps-Valls G. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. Int J Appl Earth Obs Geoinf. 2020; 92 102174 doi: 10.1016/j.jag.2020.102174

Berger, K; Wang, Z; Danner, M; Wocher, M; Mauser, W; Hank, T. Simulation ofspaceborn hyperspectral remote sensing to assist crop nitrogen content monitoringin agricultural crops; IEEE International Geosciences and Remote Sensing Symposium; 2018. 3809–3812.

van Bodegom PM, Douma JC, Witte J, Ordonez J, Bartholomeus R, Aerts R. Going beyond limitations of plant functional types when predicting global ecosystem-atmosphere fluxes: Exploring the merits of traits-based approaches. Glob Ecol Biogeogr. 2012; 21: 625–636. DOI: 10.1111/j.1466-8238.2011.00717.x

Bonan GB. Importance of Leaf Area Index and Forest Type When EstimatingPhotosynthesis in Boreal Forests. Remote Sens Environ. 1993; 314: 303–314.

Borchers H. Pracma: practical numerical math functions. 2015.

Brown LA, Ogutu BO, Dash J. Estimating forest leaf area index and canopy chlorophyll content with Sentinel-2: An evaluation of two hybrid retrieval algorithms. Remote Sensing. 2019; 11 doi: 10.3390/rs11151752

Busetto L, Ranghetti L. prismaread: A tool for facilitating access and analysis of prisma 11/12 hyperspectral imagery. 2020. https://lbusett.github.io/prismaread/

Caicedo-Rivera JP, Verrelst J, Munoz-man J, Moreno J, Camps-valls G, Member S. Toward a Semiautomatic Machine Learning Retrieval ofBiophysical Parameters. IEEE Journal of Selected Topics in Applied EarthObservations and Remote Sensing. 2014; 7: 1249–1259.

Carter GA. Primary and Secondary Effects of Water Content on the Spectral Reflectance of Leaves. Am J Bot. 1991; 78: 916–924. DOI: 10.1002/j.1537-2197.1991.tb14495.x

Castaldi F, Castrignano A, Casa R. A data fusion and spatial data analysis approach for the estimation of wheat grain nitrogen uptake from satellite data. Int J Remote Sens. 2016; 37: 4317–4336. DOI: 10.1080/01431161.2016.1212423

Chapin FS, Bloom AJ, Field CB, Waring RH. Plant Responses to MultipleEnvironmental Factors. Bioscience. 1987; 37: 49–57.

Chen JM. Black TA. Measuring leaf area index on plant canopies with brancharchitecture. Agric For Meteorol. 1991; 57: 1–12.

Clevers JG, Kooistra L, Schaepman ME. Estimating canopy water content using hyperspectral remote sensing data. Int J Appl Earth Obs Geoinf. 2010; 12: 119–125. DOI: 10.1016/jjag.2010.01.007

Cogliati S, Sarti F, Chiarantini L, Cosi M, Lorusso R, Lopinto E, Miglietta F, Genesio L, Guanter L, Damm A, Perez-Lopez S, Scheffler D, Tagliabue G, Panigada C, Rascher U, Dowling T, Giardino C, Colombo R. The PRISMA imaging spectroscopy mission: overview and first performance analysis. Remote Sens Environ. 2021; 262 doi: 10.1016/j.rse.2021.112499

Colombo R, Meroni M, Marchesi A, Busetto L, Rossini M, Giardino C, Panigada C. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. Remote Sens Environ. 2008; 112: 1820–1834. DOI: 10.1016/j.rse.2007.09.005

Combal B, Baret F, Weiss M, Trubuil A, Mace D, Pragnere A, Myneni RB, Knyazikhin Y, Wang L. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. Remote Sens Environ. 2002; 84: 1–15. DOI: 10.1016/S0034-4257C02000354

Coops NC, Smith ML, Martin ME, Ollinger SV. Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. IEEE Trans Geosci Remote Sens. 2003; 41: 1338–1346. DOI: 10.1109/TGRS.2003.813135

Crema A, Boschetti M, Nutini F, Cillis D, Casa R. Influence of soil properties on maize and wheat nitrogen status assessment from Sentinel-2 data. Remote Sensing. 2020; 12 doi: 10.3390/rs12142175

Curran PJ. Remote Sensing of Foliar Chemistry. Remote Sens. Environ. 1989; 30: 271–278.

Das B, Sahoo RN, Pargal S, Krishna G, Verma R, Chinnusamy V, Sehgal VK, Gupta VK. Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. Biosyst Eng. 2017; 160: 69–83. DOI: 10.1016/jbiosystemseng.05.007

De Grave C, Verrelst J, Morcillo-Pallares P, Pipia L, Rivera-Caicedo JP, Amin E, Belda S, Moreno J. Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. Remote Sens Environ. 2020; 251 112101 doi: 10.1016/rse.112101

Doktor D, Lausch A, Spengler D, Thurner M. Extraction of plant physiological status from hyperspectral signatures using machine learning methods. Remote Sensing. 2014; 6: 12247–12274. DOI: 10.3390/rs61212247

Estévez J, Berger K, Vicent J, Rivera-caicedo JP. Top-of-AtmosphereRetrieval of Multiple Crop Traits Using Variational Heteroscedastic GaussianProcesses within a Hybrid Workflow. Remote Sensing. 2021. 1–26.

Feret JB, Berger K, De Boissieu F, Malenovsky Z. Prospect-pro for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens Environ. 2021; 252. doi: 10.1016/j.rse.2020.112173

Feret JB, De Boissieu F. biodivMapR: An r package for a-and ^-diversity mapping using remotely sensed images. Methods Ecol Evol. 2020; 11: 64–70. DOI: 10.1111/2041-210X.13310

Gamon JA, Somers B, Malenovsky Z, Middleton EM, Rascher U, Schaepman ME. Assessing Vegetation Function with Imaging Spectroscopy. Surveys in Geophysics. 2019; 40: 489–513. DOI: 10.1007/s10712-019-09511-5

Gitelson AA, Gritz Y, Merzlyak MN. Relationships between leaf chlorophyllcontent and spectral reflectance and algorithms for non-destructive chlorophyllassessment in higher plant leaves. J Plant Physiol. 2003; 282: 271–282.

Guanter L, Kaufmann H, Segl K, Foerster S, Rogass C, Chabrillat S, Kuester T, Hollstein A, Rossner G, Chlebek C, Straif C, Fischer S. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sensing. 2015; 7: 8830–8857. DOI: 10.3390/rs70708830

Haboudane D, Miller JR, Pattey E, Zarco-Tejada PJ, Strachan IB. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens Environ. 2004; 90: 337–352. DOI: 10.1016/j.rse.2003.12.013

Hallik L, Kull O, Niinemets U, Aan A. Contrasting correlation networks between leaf structure, nitrogen and chlorophyll in herbaceous and woody canopies. Basic Appl Ecol. 2009; 10: 309–318. DOI: 10.1016/j.baae.2008.08.001

Hank, TB, Berger, K, Bach, H, Clevers, JG, Gitelson, A, Zarco-Tejada, P, Mauser, W. Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. Vol. 40. Springer; Netherlands: 2019.

Hill, J, Buddenbaum, H, Townsend, PA. Imaging Spectroscopy of Forest Ecosystems: Perspectives for the Use of Space-borne Hyperspectral Earth Observation Systems. Vol. 40. Springer; Netherlands: 2019.

Homolová L, Malenovsky Z, Clevers JG, Garaa-Santos G, Schaepman ME. Review of optical-based remote sensing for plant trait mapping. Ecological Complexity. 2013; 15: 1–16. DOI: 10.1016/j.ecocom.2013.06.003

Houborg R, McCabe M, Cescatti A, Gao F, Schull MA, Gitelson A. Joint leaf chlorophyll content and leaf area index retrieval from Landsat data using a regularized model inversion system (REGFLEC. Remote Sens Environ. 2015; 159: 203–221. DOI: 10.1016/j.rse.2014.12.008

IPCC. In: Masson-Delmotte, V; Zhai, P; Pirani, A; Connors, SL; Pean, C; Berger, S; Caud, N; Chen, Y; Goldfarb, L; Gomis, MI; Huang, M; Leitzell, K; , et al., editors. Climate change 2021: The physical science basis; Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; 2021. 1–3949.

Jacquemoud S, Baret F. PROSPECT: A model of leaf optical properties spectra. Remote Sens. Environ. 1990; 34: 75–91. DOI: 10.1016/0034-4257(90)90100-Z

Lee KS, Cohen WB, Kennedy RE, Maiersperger TK, Gower ST. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. Remote Sens Environ. 2004; 91: 508–520. DOI: 10.1016/j.rse..04.010

Liang L, Di L, Huang T, Wang J, Lin L, Wang L, Yang M. Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm. Remote Sensing. 2018; 10 doi: 10.3390/rs10121940

Lichtenthaler HK, Buschmann C. Chlorophylls and Carotenoids: Measurementand Characterization by UV-VIS Spectroscopy. Current Protocols in Food AnalyticalChemistry. 2001; 4: 1–8.

Liu N, Townsend PA, Naber MR, Bethke PC, Hills WB, Wang Y. Hyperspectral imagery to monitor crop nutrient status within and across growing seasons. Remote Sens Environ. 2021; 255 112303 doi: 10.1016/j.rse..112303

Lu B, Dao PD, Liu J, He Y, Shang J. Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sensing. 2020; 12: 1–44. DOI: 10.3390/RS12162659

Lu B, He Y, Dao PD. Comparing the performance of multispectral and hyperspectral images for estimating vegetation properties. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2019; 12: 1784–1797. DOI: 10.1109/JSTARS.2019.2910558

Martmez-Dalmau J, Berbel J, Ordonez-Fernandez R. Nitrogen fertilization. A review of the risks associated with the inefficiency of its use and policy responses. Sustainability (Switzerland). 2021; 13: 1–15. DOI: 10.3390/su13105625

Meroni M, Colombo R, Panigada C. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. Remote Sens Environ. 2004; 92: 195–206. DOI: 10.1016/j.rse.2004.06.005

Migdall, S, Bruggemann, L, Bach, H, Brunner, C, editors. Satellite-Based Earth Observation: Trends and Challenges for Economy and Society. Springer; 2018. 1–288.

Miphokasap P, Wannasiri W. Estimations of Nitrogen Concentration in sugarcane using hyperspectral imagery. Sustainability (Switzerland). 2018; 10: 1–16. DOI: 10.3390/su10041266

Mirzaie M, Darvishzadeh R, Shakiba A, Matkan AA, Atzberger C, Skidmore A. Comparative analysis of different uni- and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. Int J Appl Earth Obs Geoinf. 2014; 26: 1–11. DOI: 10.1016/jjag.2013.04.004

Navarro-Cerrillo RM, Trujillo J, de la Orden MS, Herneandez-Clemente R. Hyperspectral and multispectral satellite sensors for mappingchlorophyll content in a Mediterranean Pinus sylvestris L. plantation. Int J Appl Earth Obs Geoinf. 2014; 26: 88–96. DOI: 10.1016/jjag.2013.06.001

Nieke, J. Towards the Copernicus Hyperspectral Imaging Mission ForThe Environment (CHIM); International Geoscience and Remote Sensing Symposium (IGARSS); 2018. 157–159.

Oliver TH. How much biodiversity loss is too much? Science. 2016; 353: 220–221. [PubMed: 27418489]

Parry C, Blonquist JM, Bugbee B. In situ measurement of leaf chlorophyll concentration: Analysis of the optical/absolute relationship. Plant Cell and Environment. 2014; 37: 2508–2520. DOI: 10.1111/pce.12324

Pasqualotto N, Delegido J, Van Wittenberghe S, Verrelst J, Rivera JP, Moreno J. Retrieval of canopy water content of different crop types with two new hyperspectral indices: Water Absorption Area Index and Depth Water Index. Int J Appl Earth Obs Geoinf. 2018; 67: 69–78. DOI: 10.1016/jjag.2018.01.002

Perich G, Aasen H, Verrelst J, Argento F, Walter A, Liebisch F. Cropnitrogen retrieval methods for simulated sentinel-2 data using in-field spectrometer data. Remote Sensing. 2021; 13 doi: 10.3390/rs13122404

Pipia L, Amin E, Belda S, Salinero-Delgado M, Verrelst J. Green lai mapping and cloud gap-filling using gaussian process regression in google earth engine. Remote Sensing. 2021; 13: 1–25. DOI: 10.3390/rs13030403

R Core Team. R Foundation for Statistical Computing. Vienna, Austria: 2020. https://www.R-project.org/

Rasmussen, CE. Gaussian Processes for Machine Learning. TheMIT Press; 2006.

Rast M, Painter TH. Earth Observation Imaging Spectroscopy for Terrestrial Systems: An Overview of Its History, Techniques, and Applications of Its Missions. Surveys in Geophysics. 2019; 40: 303–331. DOI: 10.1007/s10712-019-09517-z

Rivera-Caicedo JP, Verrelst J, Munoz-Man J, Camps-Valls G, Moreno J. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. ISPRS Journal of Photogrammetry and Remote Sensing. 2017; 132: 88–101. DOI: 10.1016/j.isprsjprs.2017.08.012

Rouse, J; Haas, J; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS; Third Earth Resources Technology Satellite-1 Symposium - Volume I: Technical Presentations; Washington D.C. 1974. 309–317.

Scheiter S, Langan L, Higgins SI. Next-generation dynamic global vegetationmodels: learning from community ecology. New Phytol. 2013; 198: 957–969. [PubMed: 23496172]

Schiefer F, Schmidtlein S, Kattenborn T. The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. Ecol Ind. 2021; 121 107062 doi: 10.1016/j.ecolind.2020.107062

Schlerf M, Atzberger C. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. Remote Sens Environ. 2006; 100: 281–294. DOI: 10.1016/j.rse.2005.10.006

Schweiger AK, Schutz M, Risch AC, Kneubuhler M, Haller R, Schaepman ME. How to predict plant functional types using imaging spectroscopy: linking vegetation community traits, plant functional types and spectral response. Methods Ecol Evol. 2017; 8: 86–95. DOI: 10.1111/2041-210X.12642

Sehgal VK, Chakraborty D, Sahoo RN. Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. Information Processing in Agriculture. 2016; 3: 107–118. DOI: 10.1016/j.inpa.2016.04.001

Tagliabue G, Panigada C, Celesti M, Cogliati S, Colombo R, Migliavacca M, Rascher U, Rocchini D, Schuttemeyer D, Rossini M. Sun-induced fluorescence heterogeneity as a measure of functional diversity. Remote Sens Environ. 2020; 247 111934 doi: 10.1016/j.rse.2020.111934

Tagliabue G, Panigada C, Dechant B, Baret F, Cogliati S, Colombo R, Migliavacca M, Rademske P, Schickling A, Schuttemeyer D, Verrelst J, Rascher U, Ryu Y, et al. Exploring the spatial relationship between airborne-derived red and far-red sun-induced fluorescence and process-based GPP estimates in a forest ecosystem. Remote Sens Environ. 2019; 231 doi: 10.1016/j.rse.2019.111272

Torres I, Sanchez MT, Benlloch-Gonzalez M, Perez-Marm D. Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. Biosyst Eng. 2019; 180: 50–58. DOI: 10.1016/j.biosystemseng.2019.01.016

Townsend PA, Foster JR, Chastain RA, Currie WS. Application of imaging spectroscopy to mapping canopy nitrogen in the forest of the central Appalachian mountains using hyperion and AVIRIS. IEEE Trans Geosci Remote Sens. 2003; 41: 1347–1354. DOI: 10.1109/TGRS.2003.813205

Trombetti M, Riano D, Rubio MA, Cheng YB, Ustin SL. Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA. Remote Sens Environ. 2008; 112: 203–215. DOI: 10.1016/j.rse.2007.04.013

Upreti D, Huang W, Kong W, Pascucci S, Pignatti S, Zhou X, Ye H, Casa R. A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2. Remote Sensing. 2019; 11 doi: 10.3390/rs11050481

Van Wittenberghe S, Verrelst J, Rivera JP, Alonso L, Moreno J, Samson R. Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. J Photochem Photobiol, B. 2014; 134: 37–48. DOI: 10.1016/jjphotobioL2014.03.010 [PubMed: 24792473]

Verger A, Baret F, Camacho F. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens Environ. 2011; 115: 415–426. DOI: 10.1016/j.rse.2010.09.012

Verhoef W, Bach H. Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. Remote Sens Environ. 2007; 109: 166–182. DOI: 10.1016/.rse.12.013

Verrelst J, Berger K, Rivera-Caicedo JP. Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms. IEEE Geosci Remote Sens Lett. 2020; 1–5. DOI: 10.1109/lgrs.2020.3014676

Verrelst J, Camps-Valls G, Munoz-Man J, Rivera JP, Veroustraete F, Clevers JG, Moreno J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. ISPRS Journal of Photogrammetry and Remote Sensing. 2015; 108: 273–290. DOI: 10.1016/j.isprsjprs.2015.05.005

Verrelst J, Malenovsky Z, van der Tol C, Camps-Valls G, Gastellu-Etchegorry JP, Lewis P, North P, Moreno J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surveys in Geophysics. 2018; 1–41. DOI: 10.1007/s10712-018-9478-y

Verrelst J, Rivera-Caicedo JP, Reyes-Munoz P, Morata M, Amin E, Tagliabue G, Panigada C, Hank T, Berger K. Mapping landscape canopy nitrogen content from space using PRISMA data. ISPRS Journal of Photogrammetry and Remote Sensing. 2021; 178: 382–395. DOI: 10.1016/j.isprsjprs.2021.06.017

Vigneau N, Ecarnot M, Rabatel G, Roumet P. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. Field Crops Research. 2011; 122: 25–31. DOI: 10.1016/j.fcr.2011.02.003

Vohland M, Mader S, Dorigo W. Applying different inversion techniques to retrieve stand variables of summer barley with PROSPECT + SAIL. Int J Appl Earth Obs Geoinf. 2010; 12: 71–80. DOI: 10.1016/jjag.2009.10.005

Wang Z, Chen J, Zhang J, Fan Y, Cheng Y, Wang B, Wu X, Tan X, Tan T, Li S, Raza MA, Wang X, Yong T, et al. Predicting grain yield and protein content using canopy reflectance in maize grown under different water and nitrogen levels. Field Crops Research. 2021; 260 107988 doi: 10.1016/j.fcr.2020.107988

Wang Z, Skidmore AK, Darvishzadeh R, Wang T. Mapping forest canopy nitrogen content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. Agric For Meteorol. 2018; 253-254: 247–260. DOI: 10.1016/j.agrformet.2018.02.010

Weiss M, Jacob F, Duveiller G. Remote sensing for agricultural applications: A meta-review. Remote Sens Environ. 2020; 236 111402 doi: 10.1016/j.rse.111402

Wutzler T, Migliavacca M, Julitta T. FieldSpectroscopyCC: R package forCharacterization and Calibration of spectrometers. R package. 2016; 5: 227.

Xie Q, Dash J, Huete A, Jiang A, Yin G, Ding Y, Peng D, Hall CC, Brown L, Shi Y, Ye H, Dong Y, Huang W. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. Int J Appl Earth Obs Geoinf. 2019; 80: 187–195. DOI: 10.1016/jjag.2019.04.019

Yeoh HH, Wee YC. Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species. Food Chem. 1994; 49: 245–250. DOI: 10.1016/0308-8146(94)90167-8

Zhang C, Liu J, Shang J, Cai H. Capability of crop water content for revealing variability of winter wheat grain yield and soil moisture under limited irrigation. Sci Total Environ. 2018; 631-632: 677–687. DOI: 10.1016/j.scitotenv..03.004 [PubMed: 29539596]

Zhang J, Zhang W, Xiong S, Song Z, Tian W, Shi L, Ma X. Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content. Plant Methods. 2021; 17: 1–14. DOI: 10.1186/s13007-021-00737-2 [PubMed: 33407638]

Zhou K, Cheng T, Zhu Y, Cao W, Ustin SL, Zheng H, Yao X, Tian Y. Assessing the impact of spatial resolution on the estimation of leaf nitrogen concentration over the full season of paddy rice using near-surface imaging spectroscopy data. Front Plant Sci. 2018; 9: 1–18. DOI: 10.3389/fpls.2018.00964 [PubMed: 29410674]

Zhou X, Zhang J, Chen D, Huang Y, Kong W, Yuan L, Ye H, Huang W. Assessment of leaf chlorophyll content models for winter wheat using Landsat-8 multispectral remote sensing data. Remote Sens. 2020; 12 doi: 10.3390/RS12162574
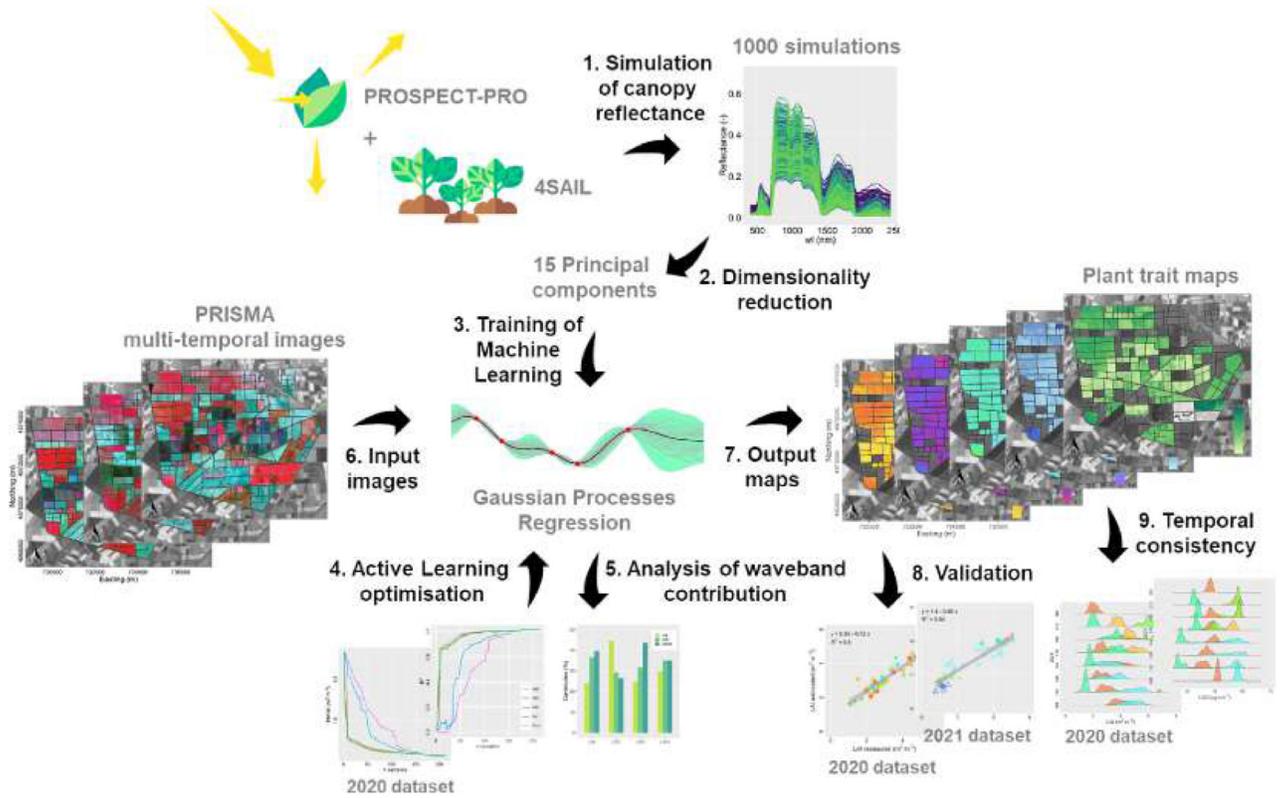
Fig. 1. Flowchart summarizing the main steps of the plant trait retrieval workflow.
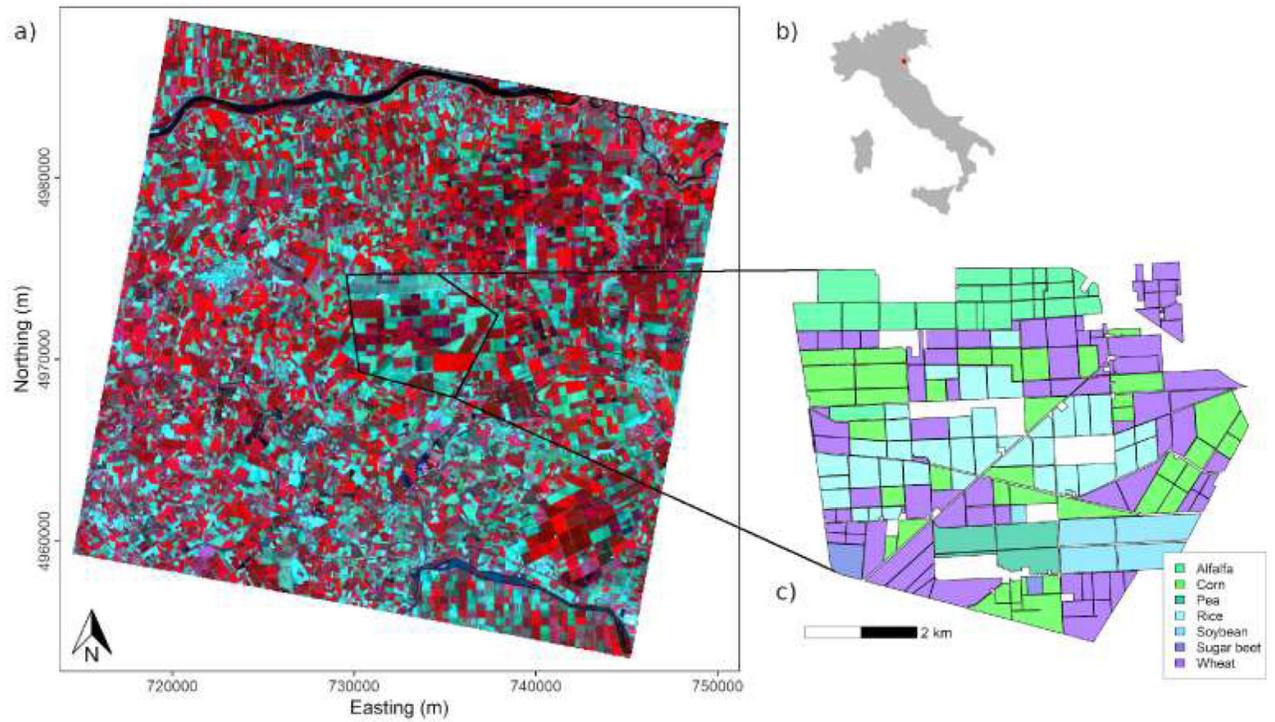
Fig. 2.

a) False colour composite (R=854.9 nm, G=650.5 nm, B=546.3 nm) of the PRISMA image captured over Jolanda di Savoia site on 31 July 2020. The map projection is UTM zone 32N with datum WGS84; b) Location of the study site in north-east Italy (red dot); c) Land cover of the Bonifiche Ferraresi farm in 2020.
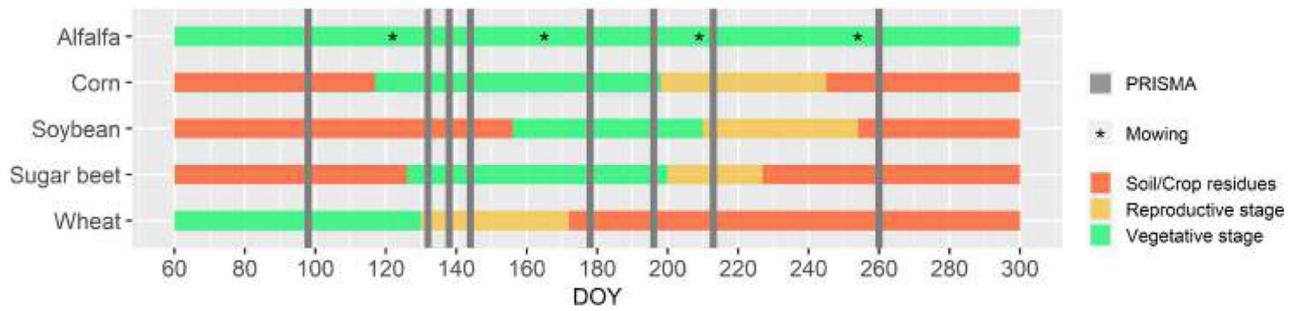
Fig. 3.
Crop calendar of the investigated crops across the 2020 season (DOY=Day Of the Year). The vegetative stage refers to the period from the crop emergence to the maximum development stage. The reproductive stage refers to the period from the maturity to the harvesting. The vertical gray bars indicate the PRISMA overpasses over the study area.
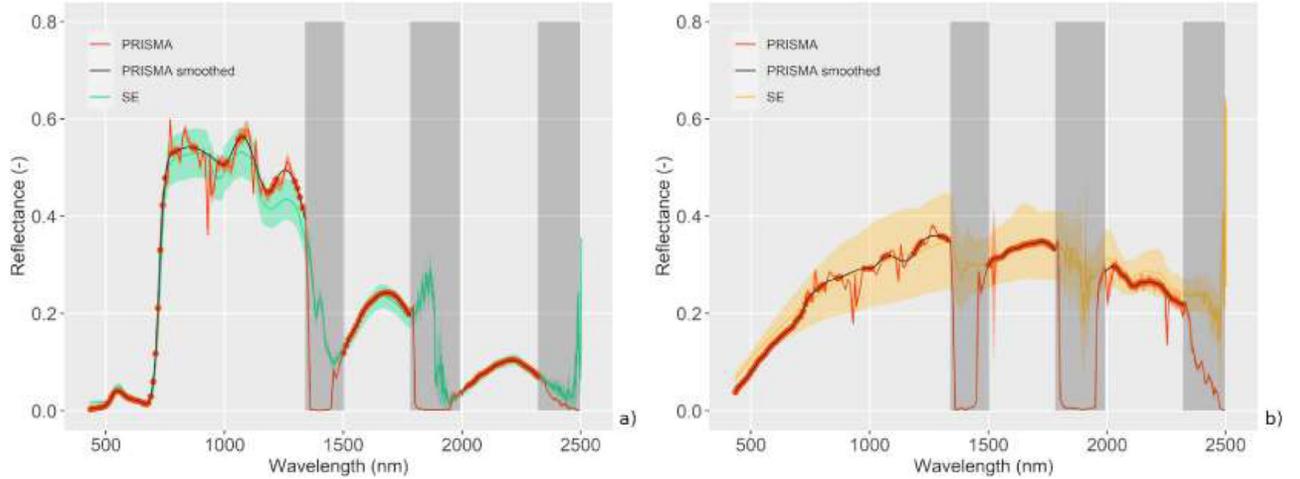
Fig. 4.

Examples of PRISMA spectra acquired on a) soybean and b) crop residues before and after the smoothing. The PRISMA spectra were extracted from regions of interest of 3×3 pixels centered on the sites where the Spectral Evolution (SE) measurements were performed. The dots mark the wavelengths which were used for the spline interpolation. The shaded coloured areas indicate the standard deviation of the PRISMA (number of samples=9) and SE (number of samples=13) reflectance spectra for each wavelength. The shaded grey areas indicate the spectral regions that were excluded after the smoothing to obtain the final PRISMA spectra.
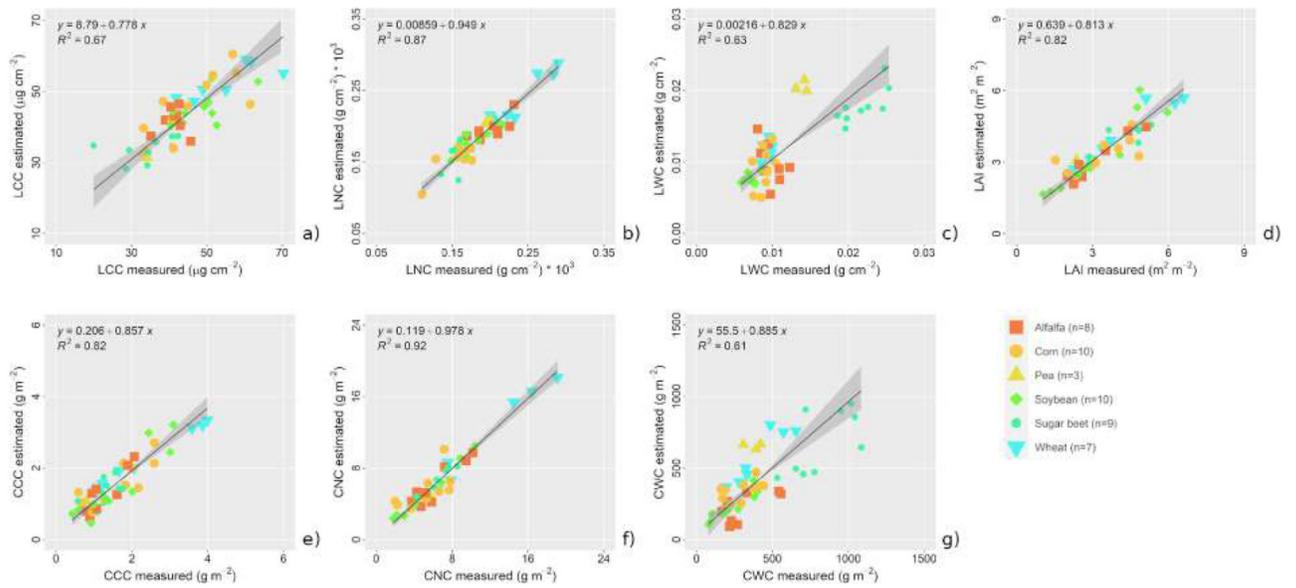
Fig. 5.
Scatter plots displaying the ground measured traits against the estimates obtained from PRISMA hyperspectral data using a hybrid GPR-AL model: a) leaf chlorophyll content (LCC), b) leaf nitrogen content (LNC), c) leaf water content (LWC), d) leaf area index (LAI), e) canopy chlorophyll content (CCC), f) canopy nitrogen content (CNC), g) canopy water content (CWC). The points are coloured according to the crop type. The number of samples (n) for each crop is also reported in the legend.
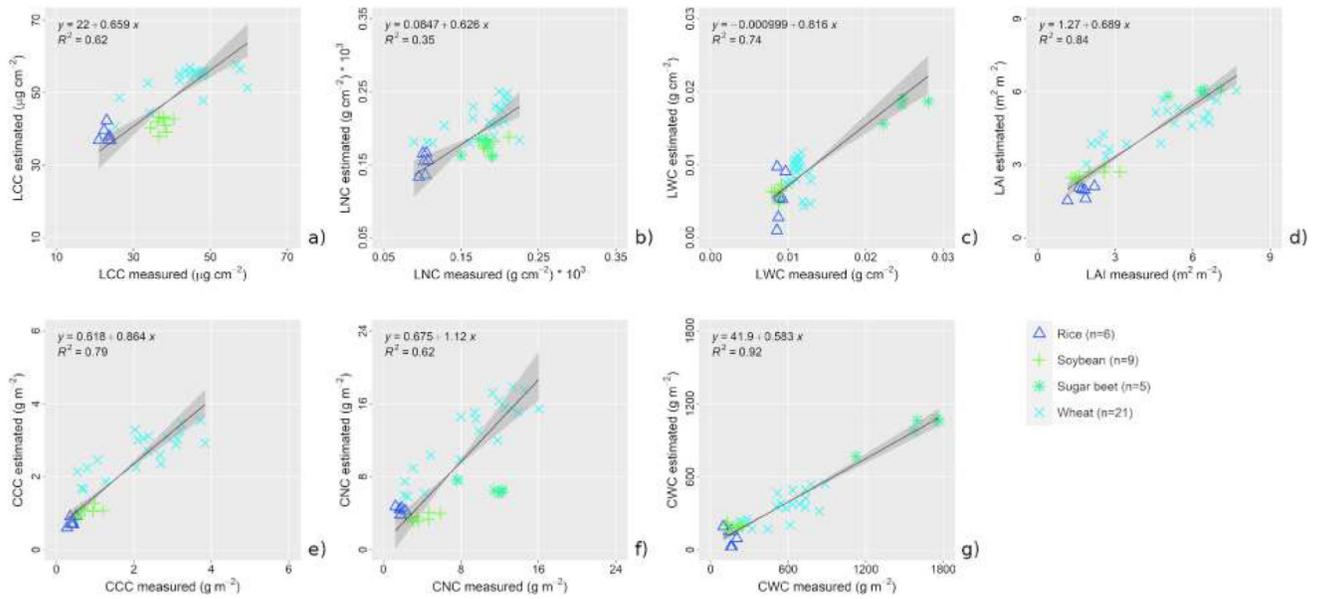
Fig. 6.

Scatter plots displaying the 2021 ground measured traits against the estimates obtained from PRISMA hyperspectral data by applying the hybrid GPR-AL model optimised on the 2020 dataset: a) leaf chlorophyll content (LCC), b) leaf nitrogen content (LNC), c) leaf water content (LWC), d) leaf area index (LAI), e) canopy chlorophyll content (CCC), f) canopy nitrogen content (CNC), g) canopy water content (CWC). The points are coloured according to the crop type. The number of samples (n) for each crop is also reported in the legend.
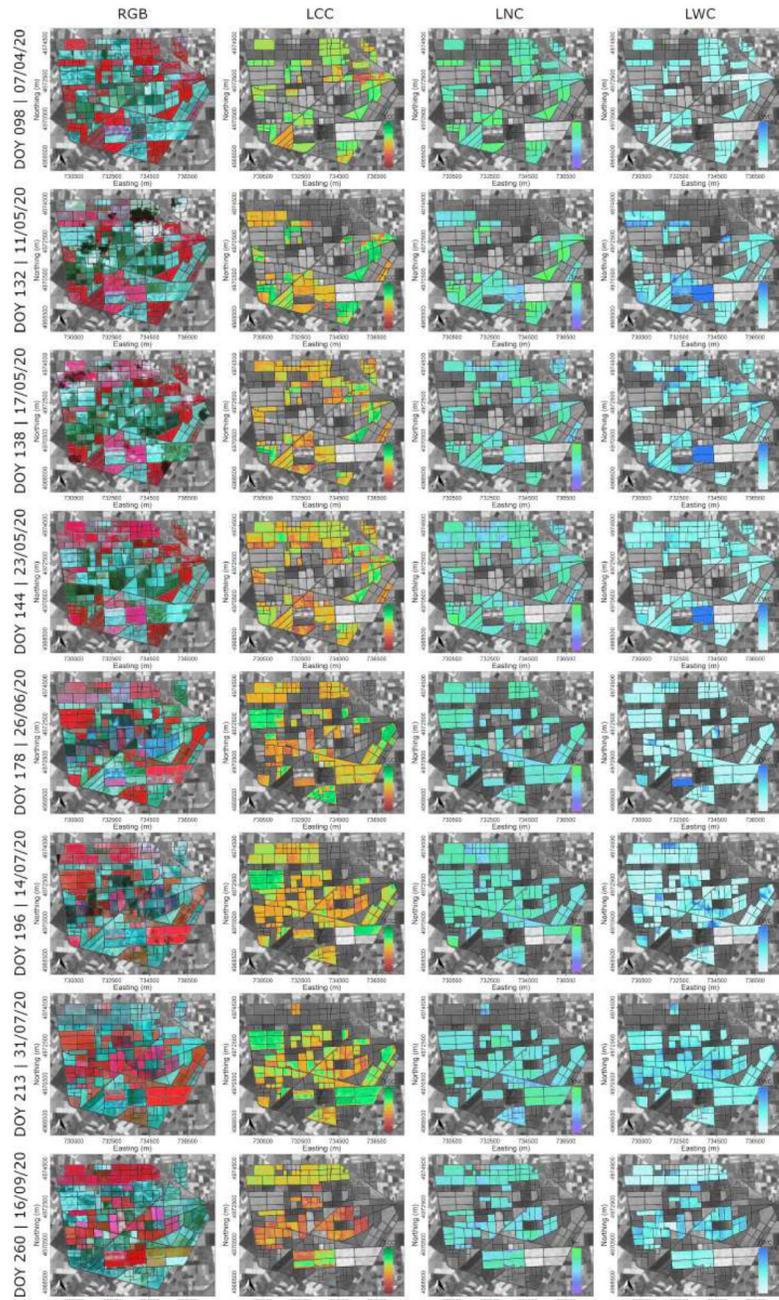
Fig. 7.

Multi-temporal crop trait maps obtained from PRISMA hyperspectral imagery using a hybrid Gaussian processes regression model. RGB=PRISMA false colour composition (R=854.9 nm, G=650.5 nm, B=546.3 nm), LCC=Leaf chlorophyll content ($\mu g\ cm^{-2}$), LNC=Leaf nitrogen content ($g\ cm^{-2}$)$\times 10^3$, LWC=Leaf water content ($g\ cm^{-2}$). In the crop trait maps, only the cloud-free vegetated fields are mapped. The map projection is UTM zone 32N with datum WGS84
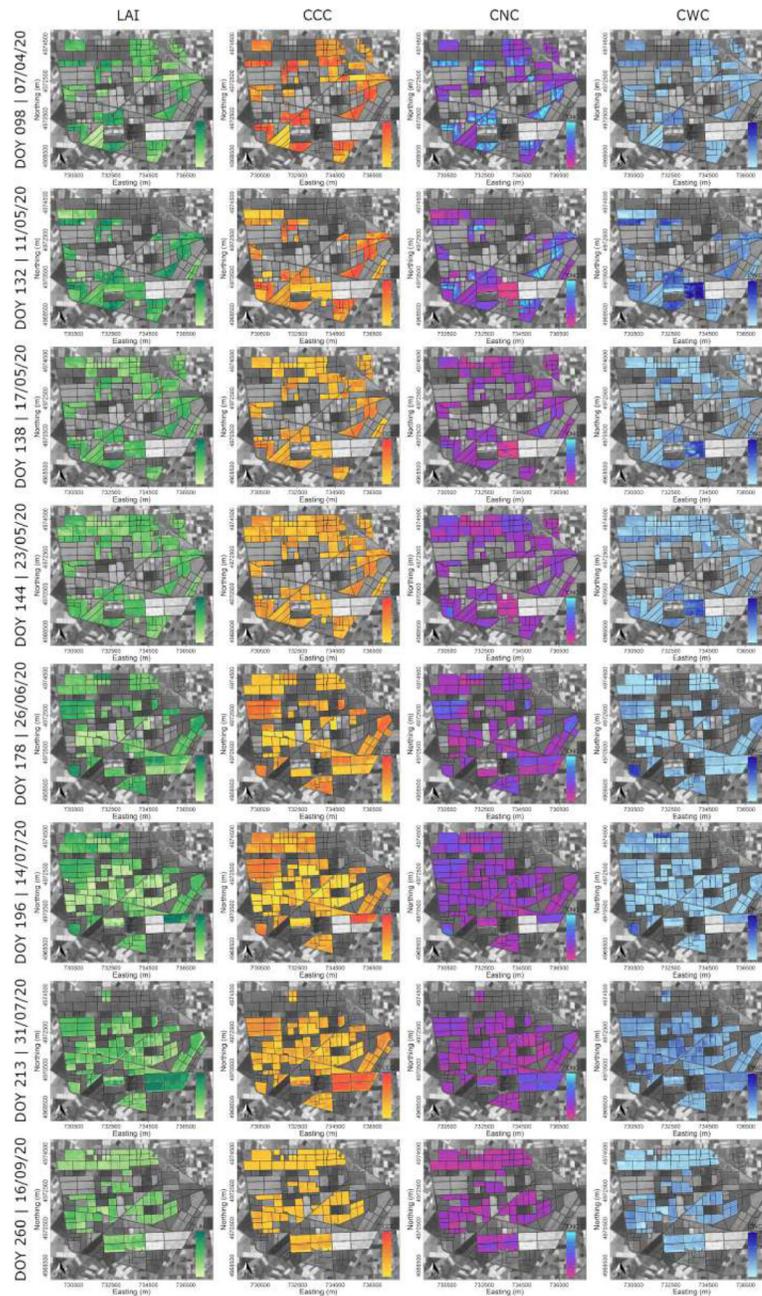
Fig. 8.
Multi-temporal crop trait maps obtained from PRISMA hyperspectral imagery using a hybrid Gaussian processes regression model. LAI=Leaf area index (m^2 m^{-2}), CCC=Canopy chlorophyll content (g m^{-2}), CNC=Canopy nitrogen content (g m^{-2}), CWC=Canopy water content (g m^{-2}). In the crop trait maps, only the cloud-free vegetated fields are mapped. The map projection is UTM zone 32N with datum WGS84.
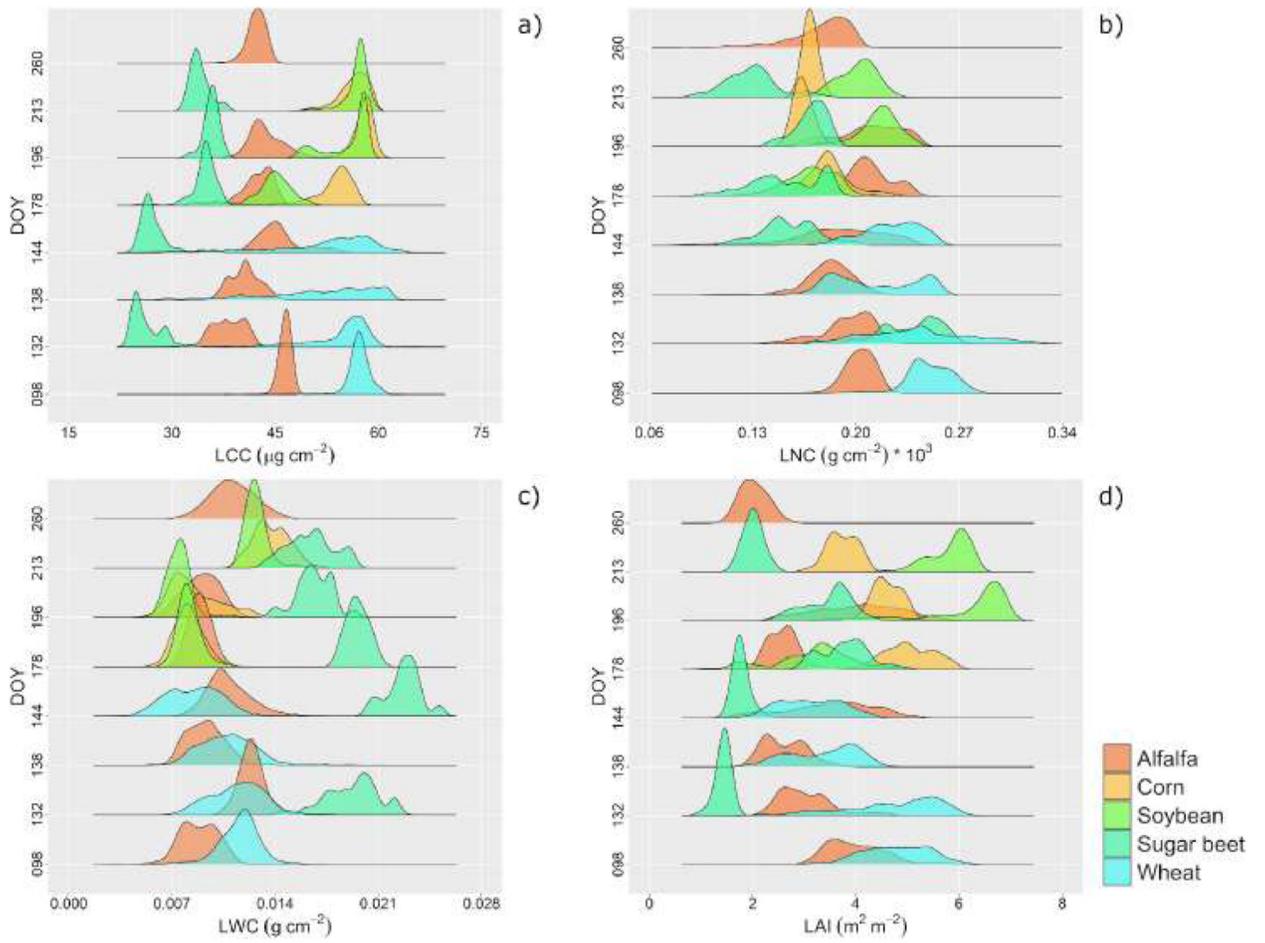
Fig. 9.
Density plots showing the temporal evolution of the traits of the five main crops in the study site across the growing season (DOY=Day Of the Year): a) leaf chlorophyll content (LCC), b) leaf nitrogen content (LNC), c) leaf water content (LWC), d) leaf area index (LAI). Each density curve represents the values extracted for each crop from all pure PRISMA pixels within the farm.
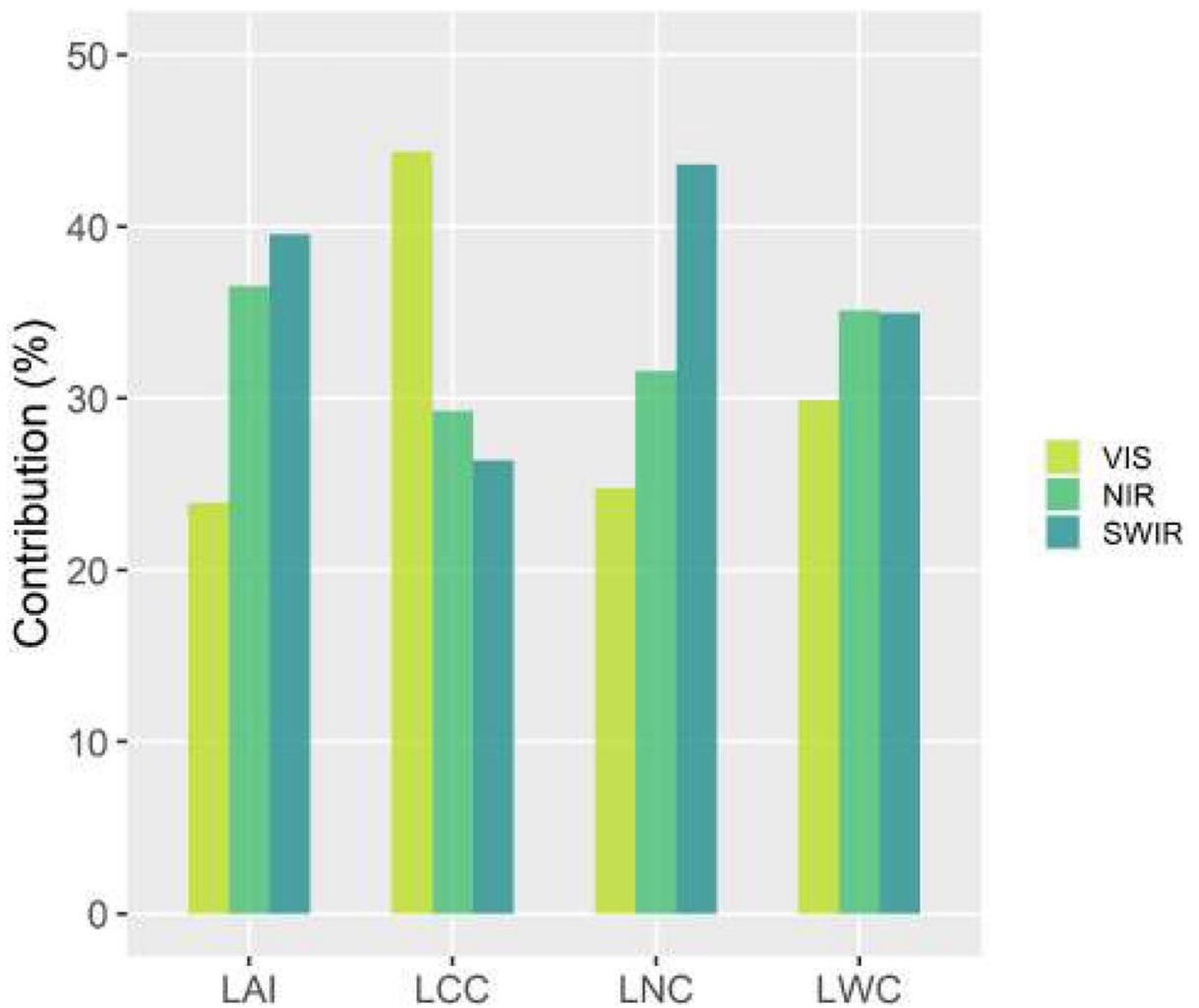
Fig. 10.
Bar plot showing the contribution (%) of the visible (VIS), near-infrared (NIR) and shortwave infrared (SWIR) spectral regions in the retrieval of Leaf Area Index (LAI), Leaf Chlorophyll Content (LCC), Leaf Nitrogen Content (LNC) and Leaf Water Content (LWC).

Fig. 11.

Contribution of the spectral bands in the retrieval of Leaf Chlorophyll Content (LCC) and Leaf Nitrogen Content (LNC). The vertical black dotted lines located at 750 nm and 1300 nm indicate the boundaries between the visible, near-infrared and shortwave infrared spectral regions.

Table 1

Descriptive statistics of the data collected in the field in 2020. For each variable (LCC=Leaf Chlorophyll Content, LNC=Leaf Nitrogen Content, LWC=Leaf Water Content, LAI=Leaf Area Index, CCC=Canopy Chlorophyll Content, CNC=Canopy Nitrogen Content, CWC=Canopy Water Content), the unit, minimum (Min), maximum (Max), average (Avg) and standard deviation (Sd) are reported.

Variable	Unit	Min	Max	Avg	Sd
LCC	$\mu g\ cm^{-2}$	14.47	70.34	43.77	10.96
LNC	$g\ cm^{-2}$	0.000110	0.000290	0.000187	0.000037
LWC	$g\ cm^{-2}$	0.005835	0.025361	0.011559	0.005475
LAI	$m^2\ m^{-2}$	1.03	6.60	3.47	1.33
CCC	$g\ m^{-2}$	0.42	3.99	1.59	0.87
CNC	$g\ m^{-2}$	1.74	19.12	6.64	3.43
CWC	g^{-2}	79.18	1086.10	409.01	256.39

Table 2

Parameterisation of the coupled PROSPECT-PRO-4SAIL model for the generation of the simulated dataset. Within the specified range (min–max), the values were generated randomly with a uniform distribution.

Model	Variable	Unit		Range
PROSPECT-PRO	N	Leaf structural parameter	-	1.2-1.8
	LCC	Leaf chlorophyll content	$\mu g\ cm^{-2}$	10-70
	C_{cx}	Leaf carotenoid content	$\mu g\ cm^-$	0-20
	C_{anth}	Leaf anthocyanin content	$\mu g\ cm^-$	0
	C_{bp}	Brown pigment content	$\mu g\ cm^-$	0
	c_p	Leaf protein content	$g\ cm^-$	0.0001-0.0015
	CBC	Carbon-based constituents	$g\ cm^{-2}$	0.001-0.006
	LWC	Leaf water content	$g\ cm^{-2}$	0.001–0.05
4SAIL	LAI	Leaf area index	$m^2\ m^{-2}$	0–7
	ALA	Average leaf angle	deg	30–60
	hots	Hot spot parameter	$m\ m^{-1}$	0.01
	a_{soil}	Soil scaling factor	-	0.1–0.5
	skyl	Fraction of diffuse radiation	-	0.1
	SZA	Sun zenith angle	deg	27
	OZA	Observer zenith angle	deg	0
	rAA	Relative azimuth angle	deg	0

Table 3

Summary of the goodness-of-fit statistics calculated on the 2020 dataset between measured and estimated values for each trait. Variable (LCC=Leaf Chlorophyll Content, LNC=Leaf Nitrogen Content, LWC=Leaf Water Content, LAI=Leaf Area Index, CCC=Canopy Chlorophyll Content, CNC=Canopy Nitrogen Content, CWC=Canopy Water Content); Active Learning algorithm (AL) (ABD=Angle-based diversity, EBD=Euclidean distance-based diversity, CBD=Clustering-based diversity, PAL=Variance-based pool of regressors); number of samples (n); coefficient of determination (r^2); p-value; root mean square error (RMSE); normalised RMSE (nRMSE); bias and relative bias (rbias)

Variable	AL	n	r^2	p-value	RMSE	nRMSE	bias	rbias
LCC	ABD	47	0.67	<0.001	5.88 µg cm^{-2}	11.7%	-1.08 µg cm^{-2}	-2.43%
LNC	EBD	47	0.87	<0.001	0.000013 g cm^{-2}	7.5%	-0.000001 g cm^{-2}	-0.50%
LWC	CBD	47	0.63	<0.001	0.003337 g cm^{-2}	17.1%	0.000189 g cm^{-2}	1.63%
LAI	CBD	47	0.82	<0.001	0.57 m^2 m^{-2}	10.3%	-0.01 m^2 m^{-2}	-0.23%
CCC	PAL	47	0.82	<0.001	0.36 g m^{-2}	10.2%	-0.02 g m^{-2}	-1.29%
CNC	EBD	47	0.92	<0.001	0.96 g m^{-2}	5.5%	-0.03 g m^{-2}	-0.40%
CWC	PAL	47	0.61	<0.001	160.76 g m^{-2}	16%	8.28 g m^{-2}	2.03%

Table 4

Summary of the goodness-of-fit statistics calculated on the 2021 dataset between measured and estimated values for each trait. Variable (LCC=Leaf Chlorophyll Content, LNC=Leaf Nitrogen Content, LWC=Leaf Water Content, LAI=Leaf Area Index, CCC=Canopy Chlorophyll Content, CNC=Canopy Nitrogen Content, CWC=Canopy Water Content); number of samples (n); coefficient of determination (r^2); p-value; root mean square error (RMSE); normalised RMSE (nRMSE); bias and relative bias (rbias).

Variable	n	r^2	p-value	RMSE	nRMSE	bias	rbias
LCC	36	0.62	<0.001	10.78 µg cm^{-2}	27.9%	8.66 µg cm^{-2}	22.1%
LNC	41	0.35	<0.001	0.000039 g cm^{-2}	28.4%	0.000021 g cm^{-2}	12.7%
LWC	41	0.74	<0.001	0.0041 g cm^{-2}	20.4%	-0.0032 g cm^{-2}	−26.7%
LAI	41	0.84	<0.001	0.95 m^2 m^{-2}	14.5%	0.06 m^2 m^{-2}	1.6%
CCC	41	0.79	<0.001	0.66 g m^{-2}	18.5%	0.41 g m^{-2}	26.4%
CNC	41	0.62	<0.001	3.5 g $^{-2}$	23.7%	1.54 g m^{-2}	21.8%
CWC	41	0.92	<0.001	276.92 g $^{-2}$	16.6%	-179.01 g $^{-2}$	−33.8%

Table 5

Coefficients of variation (CV) of the estimates on each PRISMA date. The values are expressed in percentage as average ± standard deviation. DOY=Day Of the Year, LCC=Leaf Chlorophyll Content, LNC=Leaf Nitrogen Content, LWC=Leaf Water Content, LAI=Leaf Area Index.

DOY	CV_{LCC}	CV_{LNc}	CV_{LWc}	CV_{LAI}
098	23.8±1.3%	40±2.8%	38.4±6.7%	19.7±3.2%
132	22.5± 3.9%	38.2±5.8%	30.5±6.4%	22±5.3%
138	27.6± 4.7%	44.7±5.5%	48.4±13.3%	30.9±7.8%
144	20.7± 5.9%	38.9±7.4%	41.8±10.3%	25.5±8.1%
178	28± 3.5%	44.1 ±3.9%	52.9±8.0%	27.8±7.1%
196	19± 1.9%	38.9±3.5%	46.5±7.5%	17±2.8%
213	16.2± 1.6%	37.8±2.8%	23.9±2.9%	16±1.9%
260	32.6± 3.4%	53±6.1%	29.4±7.1%	36.3±5.3%