ESTIMATION OF BIOPHISICAL PARAMETERS IN RICE CROPPING SYSTEM FROM SENTINEL-2 DATA AND HYBRID APPROACH: PERSPECTIVE FOR PRECISION AGRICULTURE APPLICATION

Marta Rossi [1]*, Gabriele Candiani [2], Marco Gianinetto [3,2], Francesco Nutini [2], Mirco Boschetti [2]

[1] Università degli Studi Milano – Bicocca, DISAT, Piazza della Scienza 1, 20126 Milano, Italy
[2] IREA, National Research Council, Via Bassini 15, 20133 Milano, Italy (candiani.g, nutini.f, boschetti.m)@irea.cnr.it
[3] Department of Architecture, Built Environment and Construction Engineering (DABC), Politecnico di Milano, Via Ponzio 31, 20133 Milano, Italy – marco.gianinetto@polimi.it

KEY WORDS:
Rice crop monitoring, Hybrid modelling, Precision Agriculture, Sentinel-2, Multispectral Remote Sensing, Canopy Nitrogen Content, Leaf Area Index.

ABSTRACT:

The availability of information able to assess crop nutritional status in space and time is a crucial issue to support sustainable agriculture in smart farming framework. Remote sensing techniques have become popular methods to support precision farming activities by producing spatial variability maps of crop conditions. In this framework, an experimental activity has been conducted to estimate Leaf Area index (LAI) and test potentiality of Canopy Nitrogen Content (CNC) retrieval, due to the importance of these parameters for assessing crop nutritional status. This study focuses on rice and retrievals have been conducted using a hybrid approach based on Radiative Transfer Model (RTM) simulations and Machine Learning Regression Algorithms (MLRA). A Look Up Table of rice spectra, with a cardinality of 2000 samples, was generated ranging crop parameters as input to the PROSPECT-PRO RTM. Simulations were resampled to 8 bands Sentinel 2-like configuration and were then used to train Gaussian Process Regression (GPR) and Neural Network (NN) MLR algorithms, also testing a sample selection procedure based on Active Learning (AL). Cross-validation results showed good performance for LAI retrieval using both the standard hybrid model (GPR: R2 ~ 0.78 , NN: R2 ~ 0.72) and AL approach (GPR: R2 ~ 0.71, NN: R2 ~ 0.67). Preliminary tests conducted to estimate CNC revealed promising results for plant nutritional status assessment. The within-field spatial variation of estimated CNC from Sentinel-2 (S2) data in a precision farming experiment resulted coherent with the observed heterogeneity in the field and to corresponding prescription maps used to manage the fertilisation.

1. INTRODUCTION

If in the last century production and consumption of food happened parallel to each other, nowadays the global megatrends (climate change, population growth, technological change, etc) gradually cause the supply-demand balance to shift. Considering this scenario, farmers are forced to increase the yields, while at the same time protecting their most important production factor, the environment, natural resources like soil from degradation, water and air from pollution and climate from emissions of greenhouse gases (FAO 2016). Remote sensing can contribute for a continuous monitoring of plant development and relative Nitrogen (N%) concentration during all growth stages to improve farm management: in this context, precision farming reveals to be a current promising solution, since holding the potential to provide farmers the capacity to assess, understand and manage the within-field variability, as a prerequisite to define sustainable agro-practices able to reduce farming costs and environmental impact (Nutini et al., 2018).
Rice is considered the world's most important staple crop, being globally cultivated and representing a strategic crop for food security in many countries. Italy contributes with almost 50% (Eurostat 2019, https://agridata.ec.europa.eu/extensions/DashboardRice/RiceProduction.html) to the European rice production. The study area of Lomellina (Pavia province, Italy) produces almost 5 million quintals of rice per year. Nitrogen (N), which is applied through fertilisation in agro-practices, is fundamental for maximising crop production and minimizing environmental impact. Therefore, a sustainable fertilisation is a pre-requisite for modern agriculture, and it can be achieved assessing crop status in space and time.
From rice spectral signature it is possible to assess vegetation biophysical and biochemical variables. This study focuses on the estimation of Leaf Area Index (LAI), which is defined as the green leaf area per unit of ground surface area, representing a proxy of total crop biomass and Canopy Nitrogen Content (CNC), measured in grams of N in crop per m^2 of ground area, representing the quantity of N uptake by the plant leaves (i.e. the canopy). Indeed, according to recent model developments, it is possible to simulate directly, thanks to PROSPECT-PRO model (Féret et al., 2021), the effect of Leaf Nitrogen Content (LNC) $[g\ cm^{-2}]$ on leaf spectra. Once coupled with a canopy model (i.e., SAIL4) it is possible to assess effect at plant level. Through inversion of such leaf-canopy radiative transfer model (PROSAIL-PRO) it is possible to estimate CNC, as recently demonstrated for wheat and corn (Berger et al., 2020).
In this framework, this study aims at testing a hybrid retrieval approach, useful for rice monitoring as a support to precision farming activities. The first step consisted in assessing the potentiality of the PROSAIL-PRO RTM for rice spectra simulation taking into account the influence of the background. After, a hybrid retrieval approach is developed by i) generating a rice specific Look Up Table (LUT), setting different background

* Corresponding author

and vegetation parameters, and ii) training different ML algorithms. Finally, contribution of an Active Learning (AL) was tested as an optimisation process for LUT sampling selection. Results are then validated using independent datasets, in view of assessing the model exportability in real operational conditions. Finally, the obtained maps of rice parameters of interest, which have been generated from Sentinel-2 (S2) data for the year 2018 on SATURNO project study area (Nutini et al., 2021), are analysed in terms of temporal and spatial information provided for rice crop monitoring and discussed as a contribution in precision farming application.

2. MATERIAL AND METHODS

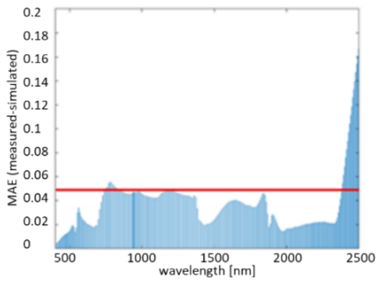
2.1 Study area and datasets exploited

The field measurements of the experimental datasets employed in this study used to test the PROSAIL-PRO RTM were collected by CNR-IREA and analysed in Stroppiana et al., 2009 and consist of two agronomic experiments carried out in 2004 and 2006 in Opera (south of Milano - Italy, Lat 45°23', Long 9°11'). Specifically, they included full resolution FieldSpec® spectral data and crop (rice) parameters from ground measurements. A 2018 dataset from SATURNO project of CNR-IREA was also used to validate the hybrid model when applied to real S2 imagery to generate biophysical parameters' maps. The ground measurements of the parameters of interest were acquired for the 2018 dataset in real farm condition, in Lomellina study area, which is one of the biggest rice districts in Italy and Europe (Nutini et al., 2021). The S2 dataset for 2018 crop season from SATURNO project included 9 clear sky images over the Lomellina study area (01/06/18, 16/06/18, 21/06/18, 26/06/18, 01/07/18, 06/07/18, 26/07/18, 31/07/18, 05/08/18).

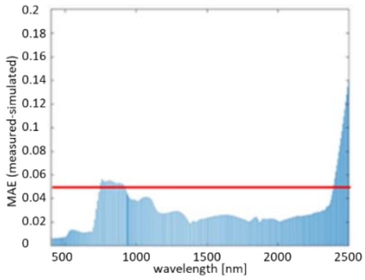
2.2 Hybrid method: RTM + MLRA

Hybrid methods for biophysical variables retrieval rely on the generation of simulated spectra using physically based RTMs for the training of ML Regression Algorithms (MLRAs), under the assumption of a more general applicability as compared to the training carried out using measured data, since RTMs allow to simulate a wide and crop-specific range of leaf and canopy parameters.

2.2.1 Simulated rice spectra and LUT generation: First, the PROSAIL-PRO model was tested to assess the reliability in rice spectral simulation performance, by providing the input parameters required (i.e., leaf and canopy level parameters, sun-view geometry and the background soil reflectance), using the data from the experimental datasets available (i.e., 2004 and 2006 datasets). Due to the variability of spectral backgrounds characterizing rice cropping systems (i.e., dry/wet soil and flooded condition)., the effect of different conditions has been tested. Error metric (Mean Absolute Error - MAE) between measured and simulated spectra were computed for the different conditions in the experiment (phenological stage, fertilisation level and background presence). Results showed that when the appropriate background types were provided to the model, a reliable simulation could be obtained, with an overall MAE less than 5% on the full spectral range, for both the tested datasets (Figure 1). On the contrary, when only one type of soil was provided, a higher MAE was obtained, ranging between 8%-12%, meaning a less reliable simulation of the real crop conditions.
After this test, the PROSAIL-PRO MATLAB script provided by CNR-IREA was run to simulate in a short time a large LUT training database (2000 samples), containing realistic ranges of variation of biophysical variables considering both measured values from 2006 dataset and reference data from literature (Campos-Taberner et al., 2016) and their corresponding simulated spectra, exploiting different backgrounds due to their ascertained importance in reproducing realistic spectra. For the LUT, a cardinality of 2000 random samples was chosen, since representing a reasonable numerosity compared to the state-of-the-art literature (Verrelst et al., 2020).



a)



b)

Figure 1. Results from spectral comparison in terms of MAE for 2004 dataset (panel a) and 2006 dataset (panel b). After 2300 nm spectra were atmospherically contaminated. Error estimated are therefore not considered. The red line represents the 5% MAE threshold.

2.2.2 ML algorithms: The MLRA model training was performed for Gaussian Process Regression (GPR) and Neural Network (NN) algorithms using ARTMO software (https://artmotoolbox.com/). GPR is considered one of the most promising kernel-based ML method for vegetation properties retrieval, having the advantage of providing uncertainty estimates on the predictions (Verrelst et al., 2020), while NN is potentially more accurate than other estimation techniques since it optimises directly over the variables of interest (Delloye et al., 2018). The Single Output model was performed, with 5% of Gaussian Noise added to both the parameters and spectra, as proposed for hybrid retrieval procedures to generalize the model and prevent overfitting on the RTM outputs (Verrelst et al., 2020). A cross-validation was performed with random partitioning of observations in 10 subsets (k-fold strategy). A data reduction (i.e., subset of LUT) procedure based on AL was also tested using ARTMO's AL module. AL is a subfield of ML seeking to optimize models to improve performance through intelligent sampling of training datasets (Verrelst et al., 2020). It is an iterative process which selected the best subsets of spectra from the original 2000 samples' LUT. 5 different diversity and uncertainty criteria where tested. Euclidean distance-based diversity (EBD), angle-based diversity (ABD) cluster-based diversity (CBD), variance-based pool of regressors (PAL), and residual regression AL (RSAL).
The reduced LUTs obtained as output, containing the optimal performing samples for the analysed LAI and CNC parameters (305 and 153 samples, respectively), were given as input to MLRA Toolbox to train the GPR and NN algorithms.

2.2.3 Sentinel 2 - Spectral configuration: Since this study focused on satellite S2 images, simulated spectra were resampled to a S2-like 8 bands spectral configuration (i.e., bands from 3 to 8, 11 and 12), this choice is conformed to European Space Agency (ESA) SENTINEL2 Toolbox, which excludes also band B2 (490 nm) according to ATBD (Algorithm Theoretical Based Document (Weiss and Baret, 2016) and findings from recent publication (Upreti et al., 2019).

3. RESULTS

3.1 Hybrid validation results

LAI estimation in cross-validation resulted well-correlated with independent data for both GPR and NN algorithms and both standard hybrid (R2 ~ 0.78 with GPR, R2 ~ 0.72 with NN) and AL optimization approach (R2 ~ 0.71 with GPR, R2 ~ 0.67 with NN) (Figure 2, panels a, b). CNC estimates were quite well correlated for low measured (independent) values (< 4 $[g\ m^{-2}]$) . Slightly better accuracy was obtained when using GPR algorithm (R2 ~ 0.37) with respect to NN (R2 ~ 0.33). An improvement is evidenced when using the AL method (passing from R2 ~ 0.37 to R2 ~ 0.45 with GPR and from R2 ~ 0.33 to R2 ~ 0.47 with NN) (Figure 2, panels c, d). AL results show a less evident saturation effect.

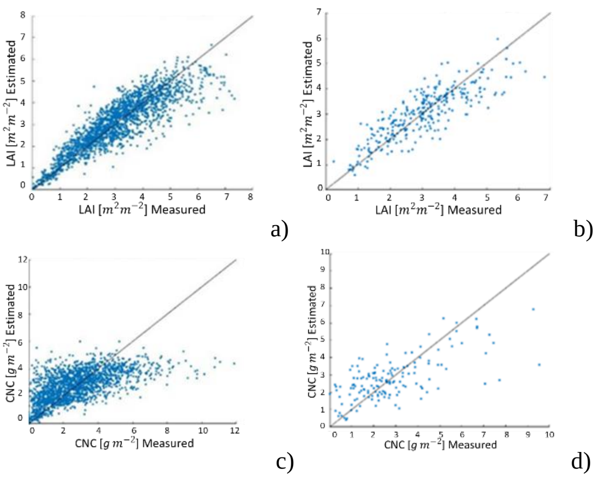


Figure 2. LAI $[m^2 m^{-2}]$ (panels a, b) and CNC $[g\ m^{-2}]$ (panels c, d) estimated values against measured (independent) in cross-validation phase, using the hybrid retrieval approach with GPR algorithm, considering the 2000 samples LUT (left panels) and the reduced LUTs from AL optimization (right panels)

The robustness and exportability of the model was assessed performing a validation using real farm condition datasets (2018 – SATURNO project). Results showed that LAI estimates were well correlated with measured values (data not showed), whereas CNC estimates showed the same saturation behaviour after 3-4 $[g\ m^{-2}]$, obtained in cross-validation phase, when compared to ground plant nitrogen uptake values (data not showed), although not representing a problem for model application to support fertilisation application at tillering and stem elongation phase, since for this purpose low values of CNC are expected in field. In general, the best accuracy results were obtained using the hybrid method with AL optimization and GPR algorithm, for both the parameters. The best method was consequently applied to the 9 selected level 2A S2 images that covered the Lomellina study area in the period from June to August of 2018 (from emergence to heading stage). Five of these images matched with ground measurements (Nutini et al., 2021).

From the generated LAI map, it was analysed the reliability of crop growth time series according to phenological development at both district (entire Lomellina) and farm level (single field) (Figure 3). The temporal maps from June to August resulted in agreement with the expected physical process of rice growth, with LAI estimates increasing in time from 0 to 6 $[m^2 m^{-2}]$.

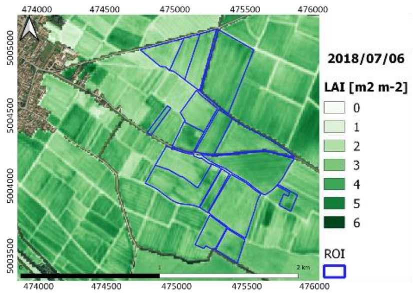


Figure 3. Example of retrieved LAI maps for the 6th of July. The blue polygons represent the Region Of Interest (ROI).

CNC map's reliability was further assessed by performing a within-field analysis. These maps were compared to the prescription maps generated in the SATURNO project for Variable Rate Technology (VRT) application (Figure 4).

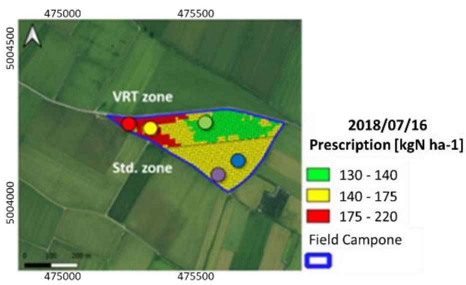


Figure 4. Prescription map from SATURNO project and selected samples, coloured circles, from 6th of July CNC map.

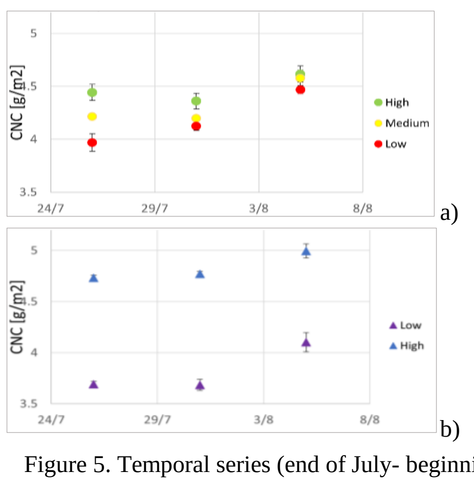


Figure 5. Temporal series (end of July- beginning of August 2018) of CNC estimates after 17th of July fertilisation, extracted in correspondence of the sample areas (coloured circles, Figure 4) for VRT (panel a) and Std. (panel b) zones

Figure 5 shows the prescription map used to fertilise the field with VRT on the 17th of July: upper part was managed with three target doses (140, 175 and 220 [$kg\ ha^{-1}$]) of "Urea" according to crop condition (below, average and above) as detected with a smart scouting approach while the lower part of the field was managed with a standard homogeneous fertilisation dose (175 [$kg\ ha^{-1}$]) (Nutini et al., 2021). Sample areas were selected in the map, corresponding to i) high, medium and low and ii) high and low CNC estimates in the upper and lower part of the field, respectively. These 5 sample areas were selected on the 6^{th} of July image, where the different ranges were well visible.

The CNC time series corresponding to the selected sample areas were extracted using the "Profile Tool" QGIS plugin, in order to investigate the effect of variable rate fertilisation against the standard dose application (Figure 5). The analysis showed the positive consequence of the VRT fertilisation applied in the upper part, on the 17^{th} of July, according to the prescription, resulting in a final homogeneous N plant content, although plants were having different growing conditions at the beginning. On the contrary, in the homogeneously managed half of the field, lower part, differences in rice crop development were kept after the fertilisation, reflecting a non-optimized N management.

4. DISCUSSION

LAI generated maps resulted reliable EO-derived products to describe the crop development in the study area, at both district and farm level. Those maps also provided the information needed to assess differences occurring in different farms or fields due to multiple factors such as rice variety, sowing date, water management and soil quality hence nutritional status. The analysis of LAI temporal profile can be useful to monitor different crop growing behaviours as a consequence of different management conditions. The within-field nutritional spatial variation highlighted by the CNC generated maps resulted coherent with the observed heterogeneity in the field in agreement with VRT fertilisation management. Comparison of CNC estimates in area from different fertilisation techniques (i.e., the VR and the standard), highlighted how the lower dose of fertiliser used in the VRT zone with high CNC values, saved up to 20% on N application while maintain proper crop growth. From these findings, it can be stated that such digital geo-products represent a promising decision-supporting spatial information contribution to support crop monitoring and fertilisation management purposes, in the precision farming framework.

5. CONCLUSIONS

From the experiments, it resulted that the rice simulated spectra using the PROSAIL-PRO model agreed well with measured data when a representative database of background reflectance is provided (overall MAE lower than 5%). The hybrid retrieval approach, provided accurate estimates, showing improvement in the performance when an AL method is used. In general, the GPR algorithm provided slightly more accurate results than the NN. LAI estimates were always robust, whereas CNC estimates showed a saturation behaviour for measured values greater than 4 [$g\ m^{-2}$], although not representing a problem for a model application to support fertilisation agro-practices. The source of variability in the estimation for high values need to be further investigated, since CNC estimation from RTM is a quite new approach in remote sensing. The application of the model to satellite images was able to provide crop monitoring information at district and farm level. Within-field detected variability is the foreseen information requested by user community for a quantitate support of sustainable fertilisation. In conclusion the proposed method demonstrated the feasibility of a direct estimation of biophysical variables from S2 products as a contribution for precision farming applications.

ACKNOWLEDGEMENTS

The research activities have been framed in the CNR-DIPARTIMENTO DI INGEGNERIA, ICT E TECNOLOGIE PER L'ENERGIA E I TRASPORTI project "DIT.AD022.180 Transizione industriale e resilienza delle Società post-Covid19 (FOE 2020)", sub task activity "Agro-Sensing".

REFERENCES

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

Campos-Taberner, M., García-Haro, F.J., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A., Boschetti, M., 2016. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. Remote Sens. Environ. 187, 102–118. https://doi.org/10.1016/j.rse.2016.10.009

Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. Remote Sens. Environ. 216, 245–261. https://doi.org/10.1016/j.rse.2018.06.037

Nutini, F., Confalonieri, R., Crema, A., Movedi, E., Paleari, L., Stavrakoudis, D., Boschetti, M., 2018. An operational workflow to assess rice nutritional status based on satellite imagery and smartphone apps. Comput. Electron. Agric. 154, 80–92. https://doi.org/10.1016/j.compag.2018.08.008

Nutini, F., Confalonieri, R., Paleari, L., Pepe, M., Criscuolo, L., Porta, F., Ranghetti, L., Busetto, L., Boschetti, M., 2021. Supporting operational site-specific fertilization in rice cropping systems with infield smartphone measurements and Sentinel-2 observations. Precis. Agric. https://doi.org/10.1007/s11119-021-09784-0

Stroppiana, D., Boschetti, M., Brivio, P.A., Bocchi, S., 2009. Plant nitrogen concentration in paddy rice from field canopy hyperspectral radiometry. F. Crop. Res. 111, 119–129. https://doi.org/10.1016/j.fcr.2008.11.004

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

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

Weiss, M., Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER - Version 1.1. Sentin. ToolBox Level2 Prod. 53.