

Extended Abstract

# Cropland Mapping Using Earth Observation Derived Phenological Metrics †

Federico Filippini ‡, Daniela Smiraglia \*‡, Stefania Mandrone and Antonella Tornato

Italian Institute for Environmental Protection and Research, ISPRA; federico.filippini@isprambiente.it (F.F.); stefania.mandrone@isprambiente.it (S.M.); antonella.tornato@isprambiente.it (A.T.)

\* Correspondence: daniela.smiraglia@isprambiente.it; Tel.: +39-06-5007-2261

† Presented at the 1st International Electronic Conference on Agronomy, 3–17 May 2021; Available online: <https://iecag2021.sciforum.net/>.

‡ These authors contributed equally to this work.

**Abstract:** Satellite Earth Observation provides timely and spatially explicit information on crop phenology that can support decision making and sustainable agricultural land management. Accurate classification and mapping of croplands is a primary information for agricultural assessment. This study presents a digital agriculture approach that integrates Earth Observation big data analytics based on machine learning technologies to classify and map main crop types. Two supervised machine learning models were calibrated using Random Forests algorithm from phenological metrics, estimated from NDVI and LAI vegetation indices time series calculated using Sentinel-2 MSI satellite acquisitions. Models were calibrated for the Toscana region in Italy. Results show a satisfactory overall accuracy (~78%) in croplands classification, and that model calibrated using LAI time series performs slightly better than the model calibrated using NDVI time series. The proposed approach offers a potential to accurately map crop types useful to support agricultural land management and monitoring systems for large areas over time.

**Keywords:** phenological metrics; random forests; NDVI; LAI; Sentinel-2

**Citation:** Filippini, F.; Smiraglia, D.; Mandrone, S.; Tornato, A. Cropland Mapping Using Earth Observation Derived Phenological Metrics. *Proceedings* **2021**, *68*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor:

Published: date

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cropland mapping is becoming increasingly important in environmental topics which deals with sustainable agriculture production and natural resources management [1]. Nowadays, a wide number of stakeholders is interested in this topic, such as national authorities, local environmental agencies, regional government and authorities, municipalities, universities and research centers, civil protection agencies, insurance companies, industries. Cropland mapping product answers the information need deeply felt by users in response to the growing interest shown by the European policies to climate change mitigation and adaptation and foster sustainable agricultural practices, especially today in the context of the European Green Deal strategy [2].

The information provided by increasing availability of Earth Observation (EO) data makes satellite images of paramount importance for identify, characterize and map crop typologies in both space and time dimensions by exploiting the radar backscatter and the optical response of vegetation [3,4]. The European Commission (EC) commitment to encourage the development of EO products, possibly taking advantage of Copernicus In situ Component, makes value-added information derived from satellites of primary importance for supporting agricultural land management. Indeed, the EC has finally sanctioned the use of Copernicus Sentinels data, integrated with EGNOS/Galileo, for the control and granting of Common Agricultural Policy (CAP) payments by local authorities, promoting the open data with a common data-sharing approach (Regulation (EU) 746/2018).

Multitemporal satellite images have proven to be successfully used to estimate vegetation biophysical parameters and identify phenological patterns [5–7]. Recently, Copernicus Sentinel-2 satellite constellation, equipped with MSI sensor, allows to sense Earth surface at high spatial, spectral and temporal resolutions, showing its potential for the estimation of vegetation parameters, such as phenological metrics (e.g., the start of season, the length of season, or the end of season) [7,8].

Many authors investigated the efficacy of spectral and biophysical time series indices to differentiate crop types [9,10]. Vegetation spectral indices have been and are still widely used to detect the status of vegetation (e.g., growth, health, and cover), the most popular of which is the Normalized Difference Vegetation Index (NDVI) [11]. However, NDVI has saturation as its limit at high values. On the other hand, vegetation biophysical characteristics, as the canopy structure and photosynthetic capacity, are well described by the Leaf Area Index (LAI) largely used in agricultural studies in heterogeneous smallholder and fragmented agroecosystems [12,13].

Furthermore, the advances in analytical techniques, such as the machine learning algorithms, enable dealing with fast and robust analysis applied to big data. Among these, the Random Forests (RF) is an ensemble learning classifier successfully used in vegetation classification applications, including crop mapping [7,10,14].

The aim of this study is to present a digital agriculture approach that integrates EO big data analytics, based on supervised machine learning model using temporal statistics and phenological metrics estimated from NDVI and LAI time series as predictors, to identify and map the main crops types. Performances of two supervised machine learning models, calibrated using RF algorithm for a study area in central Italy, are presented and discussed.

## 2. Materials and Methods

### 2.1. Study Area

Tuscany region is located in central Italy and covers about 23,000 square kilometers. The climate ranges from the Mediterranean dry climate along the coastline to the Temperate humid and wet climate in inland and northern areas of the region. Tuscany is mainly hilly (about 67%) and mountainous (about 25%), and it includes also some plains (about 8%). The cultivated areas represent about 39% of the region mainly characterized by arable land, vineyards and olive groves.

### 2.2. Satellite Images

Sentinel-2 (S2) satellites images, acquired from November 2015 to October 2019 with cloud cover lower than 90%, were acquired for the 4 granules corresponding to the study area. The Multi-Spectral Instrument (MSI) sensor onboard S2 is characterized by high spatial resolution (10 m, 20 m and 60 m), high revisit time (5 days with two satellites), and 13 spectral bands from the visible to shortwave infrared. The spectral bands of the images in the MUSCATE format, distributed by Theia as the bottom of the atmosphere (BOA) reflectance, orthorectified, terrain-flattened and atmospherically corrected with MACCS-ATCOR Joint Algorithm (MAJA) [15], were processed for spatial resampling at 10 m masked for invalid pixels (cloud, cloud\_cirrus, cloud\_shadow, topographic\_shadow, snow, edge, sun\_too\_low). A static mask, generated from Copernicus Land Monitoring Service datasets, has been applied to mask out pixel not corresponding to croplands.

### 2.3. Crop Types Maps

The reference crop types maps used in this study were made available by the Tuscany Regional Agency for Agriculture (<http://dati.toscana.it/organization/artea>) for the years from 2016 to 2019. This study focused only on the main crop types of the arable land, excluding the permanent crops such as vineyards and olive groves. Selected crop typologies were grouped into 8 classes taking into account the temporal pattern of the crops in

the study area: winter cereals, clover and alfalfa, maize, sorghum, sunflower, rape, horticultural crops, and soy. The centroid of each crop parcel polygon in the reference maps were used to query the raster predictors generated from satellite images.

#### 2.4. Time Series and Temporal Predictors

Two vegetation indices were selected to derive the main crop types in the study area: the NDVI and the LAI. The NDVI was calculated following the Equation (1):

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

where *RED* corresponds to S2 MSI spectral band *B4* and *NIR* corresponds to S2 MSI spectral band *B8*. The Leaf Area Index (LAI) is defined as half of the total green (i.e., photosynthetically active) leaf area per unit horizontal ground surface area. The biophysical processor [16] available in SNAP software was used to estimate the LAI from surface reflectance data.

The vegetation indices time series were first gap-filled and daily interpolated using Stinemann algorithm [17], and later temporally smoothed using the procedure based on second order weighted polynomial fitting and the Whittaker smoothing, described in [18]. From NDVI and LAI time series, temporal statistics and phenological metrics, derived following Gu et al. [19], were calculated and used as temporal predictors in the classification model (Table S1).

All predictors with a Pearson correlation coefficient higher than 0.9 and a variance inflation factor (VIF) higher than 2.0 [20] were removed to avoid multi-collinearity.

#### 2.5. Random Forests Classification

R package 'mlr' [21] was used to set the RF hyperparameters combination (i.e., *mtry*, *min.node.size*, *ntree*) through a 5-fold cross-validation with 20 repetitions and select those with the higher Cohen's kappa coefficient. Tuned hyperparameters were used to calibrate classification models from NDVI and LAI predictors using R-package 'ranger' [22]. Variables importance for the final set of selected predictors used in the models was calculated using the Gini index.

A stratified sampling method was applied to the crop type reference map of the year 2019 in order to select the pixels which represent all 8 classes of crop types, and that can be used as training samples for the classification and as test samples to verify the accuracy of the classification obtained. The 70% of the pixels were used as training samples and the remaining 30% as the test samples.

The results of the classifications obtained were evaluated by means of confusion matrices according to the test samples. Overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and Cohen's kappa coefficient (K) were assessed.

Finally, crop types map product for year 2019 was predicted using the calibrated supervised machine learning models.

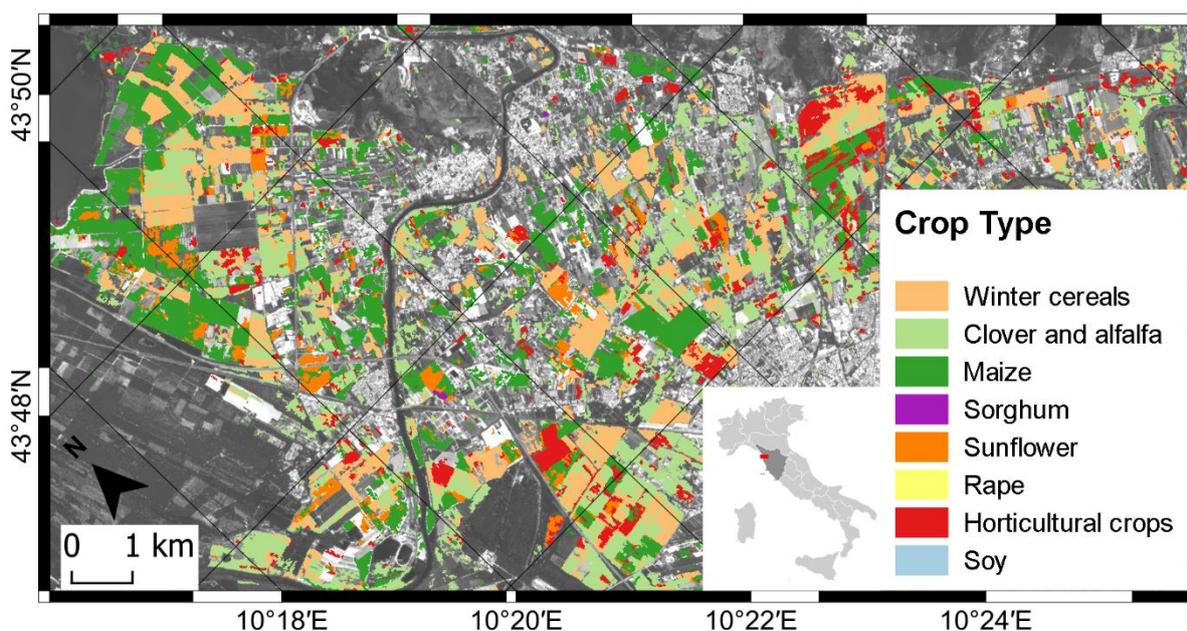
### 3. Results and Discussions

The RF hyperparameters tuning produced the following setting: *mtry* = 5, *min.node.size* = 2, *ntree* = 893 for NDVI, and *mtry* = 4, *min.node.size* = 3, *ntree* = 424 for LAI. The selected predictor variables reporting the higher Gini index were 13 for NDVI and 11 for LAI (Table S1).

Resulting spatial crop types map is showed in Figure 1. Regarding the classification obtained from the NDVI time series analysis, an overall accuracy of 78.6% was achieved with a Cohen's kappa coefficient of 0.54 (Table 2). Some classes were more accurately classified than others, such as clover and alfalfa (UA = 91.1%; PA = 82.8%), maize (UA = 69.5%; PA = 58%), and winter cereals (UA = 55.9%; PA = 69.2%). On the contrary, the sorghum

was classified worst (UA = 6%; PA = 26.8%). Rape and soy obtained low user’s accuracy (17.2% and 17.6% respectively).

As for the classified crop types resulting from the LAI time series analysis, an overall accuracy of 78.3% was attained with a Cohen’s kappa coefficient of 0.59 (Table 3). Differently from the NDVI model results, the LAI model generally showed high user’s and producer’s accuracies for all the classes, except for rape (UA = 10.7%) and the misclassification was with winter cereals principally.



**Figure 1.** Crop types map of year 2019 for the area of the city of Pisa (Tuscany, Italy).

Notwithstanding the overall accuracies and Cohen’s kappa coefficient are similar for both the NDVI and the LAI model, comparing the results for individual classes, the latter showed slightly higher performances.

High misclassifications of horticultural crops may be related to different seeding time of the horticultural species, that could increase the variability in terms of predictors values range. With respect to soy and rape, it should be noted that the small number of reference crops used for model calibration and validation could be the reason for such a low classes accuracy.

**Table 2.** Confusion matrix of RF result from the NDVI time series analysis. Producer (PA), user’s (UA), and overall (OA) accuracies in percentage as well as the Cohen’s kappa coefficient (K) are reported.

K = 0.54		Classification									
Reference map	Winter cereals	Clover and Alfalfa	Maize	Sorghum	Sunflower	Rape	Horticultural crops	Soy	Total	PA %	
Winter cereals	3804	1609	4	2	8	24	44	0	5495	69.2	
Clover and Alfalfa	2974	17084	53	45	36	29	410	0	20631	82.8	
Maize	2	21	528	62	215	0	71	11	910	58	
Sorghum	1	1	8	11	11	0	9	0	41	26.8	
Sunflower	6	20	139	34	230	0	62	3	494	46.6	
Rape	7	1	0	0	0	11	0	0	19	57.9	
Horticultural crops	10	16	28	29	54	0	540	0	677	79.8	
Soy	0	0	0	0	0	0	0	3	3	100	
<b>Total</b>	6804	18752	760	183	554	64	1136	17	<b>28270</b>	<b>OA%</b>	
UA %	55.9	91.1	69.5	6	41.5	17.2	47.5	17.6	<b>OA%</b>	78.6	

**Table 3.** Confusion matrix of RF result from the LAI time series analysis. Producer (PA), user’s (UA), and overall (OA) accuracies in percentage as well as the Cohen’s kappa coefficient (K) are reported.

K = 0.59		Classification									
Reference map	Winter cereals	Clover and Alfalfa	Maize	Sorghum	Sunflower	Rape	Horticultural crops	Soy	Total	PA %	
Winter cereals	9759	2510	0	0	0	62	40	0	12371	78,9	
Clover and Alfalfa	2170	7743	0	0	0	13	146	0	10072	76,9	
Maize	0	0	19	5	4	0	2	1	31	61,3	
Sorghum	0	0	4	20	1	0	2	0	27	74,1	
Sunflower	0	0	3	0	12	0	0	2	17	70,6	
Rape	13	0	0	0	0	9	0	0	22	40,9	
Horticultural crops	5	2	0	5	0	0	428	1	441	97,1	
Soy	0	0	0	1	1	0	0	3	5	60,0	
<b>Total</b>	11947	10255	26	31	18	84	618	7	<b>22986</b>	<b>OA %</b>	
UA %	81,7	75,5	73,1	64,5	66,7	10,7	69,3	42,9	<b>OA %</b>	78,3	

#### 4. Conclusions

The study demonstrates the EO big data analytics capacity to provide thematic products to support agricultural land management and fulfill users’ requirements. The phenological metrics estimated from high-resolution imagery sensed by Copernicus S2 satellites constellation, combined with thematic reference dataset related to crop types, together with the use of advanced computational analytic techniques (RF algorithm), allowed crop types mapping in heterogeneous, small, and fragmented agricultural systems. The calibrated NDVI and LAI supervised machine learning models show similar performances, with LAI model yielding better results.

The supervised machine learning model, applied to a wider spatial extent, could contribute to the sustainability measurement and assessment foreseen to the European Green Deal strategy, in terms of sustainable agricultural practices and environmental monitoring, climate change mitigation and adaptation, in accordance with the stakeholder requirements.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1).

**Author Contributions:** Conceptualization, F.F. and D.S.; methodology, F.F.; formal analysis, F.F.; investigation, D.S.; writing—original draft preparation, D.S. and F.F.; writing—review and editing S.M. and A.T.; supervision, A.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research was funded by Italian Space Agency (ASI) in the framework of agreement between ASI and the Italian Institute for Environmental Protection and Research (ISPRA) on “Air Quality” (Agreement number F82F17000000005).

**Acknowledgments:** This work contains modified Copernicus Sentinel data Copernicus Service information (2021).

**Institutional Review Board Statement:**

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:**

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Azar, R.; Villa, P.; Stroppiana, D.; Crema, A.; Boschetti, M.; Brivio, P.A. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. *Eur. J. Remote Sens.* **2016**, *49*, 1, 361–380, doi:10.5721/EuJRS20164920.
2. Taramelli, A.; Tornato, A.; Magliozzi, M.L.; Mariani, S.; Valentini, E.; Zavagli, M.; Costantini, M.; Nieke, J.; Adams, J.; Rast, M. An interaction methodology to collect and assess user-driven requirements to define potential opportunities of future hyperspectral imaging sentinel mission. *Remote Sens.* **2020**, *12*, 1286, doi:10.3390/RS12081286.
3. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* **2015**, *7*, 12356–12379, doi:10.3390/rs70912356.
4. Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium. *Remote Sens.* **2018**, *10*, 1642, doi:10.3390/rs10101642.
5. Weissteiner, C.J.; López-Lozano, R.; Manfron, G.; Duveiller, G.; Hooker, J.; van der Velde, M.; Baruth, B. A Crop Group-Specific Pure Pixel Time Series for Europe. *Remote Sens.* **2019**, *11*, 2668, doi:10.3390/rs11222668.
6. Gao, F.; Anderson, M.C.; Hively, W.D. Detecting Cover Crop End-Of-Season Using VEN $\mu$ S and Sentinel-2 Satellite Imagery. *Remote Sens.* **2020**, *12*, 3524, doi:10.3390/rs12213524.
7. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130, doi:10.1016/j.jag.2018.06.007.
8. Vrieling, A.; Meroni, M.; Darvishzadeh, R.; Skidmore, A.K.; Wang, T.; Zurita-Milla, R.; Oosterbeek, K.; O'Connor, B.; Paganini, M. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. *Remote Sens. Environ.* **2018**, *215*, 517–529, doi:10.1016/j.rse.2018.03.014.
9. Djamai, N.; Fernandes, R.; Weiss, M.; McNairn, H.; Goita, K. Validation of the Sentinel Simplified Level 2 Product Prototype Processor (SL2P) for mapping cropland biophysical variables using Sentinel-2/MSI and Landsat-8/OLI data. *Remote Sens. Environ.* **2019**, *225*, 416–430, doi:10.1016/j.rse.2019.03.020.
10. Belgiu, B.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523, doi:10.1016/j.rse.2017.10.005.
11. Rouse Jr, J.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Third ERTS-1 Symposium, Washington DC, USA, 10–14 December 1973; NASA: Washington, DC, USA, 1974; pp. 309–317.
12. Lambert, M.J.; Traoré, P.C.S.; Blaes, X.; Baret, P.; Defourny, P. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali’s cotton belt. *Remote Sens. Environ.* **2018**, *216*, 647–657, doi:10.1016/j.rse.2018.06.036.
13. De Peppo, M.; Dragoni, F.; Volpi, I.; Mantino, A.; Giannini, V.; Filipponi, F.; Tornato, A.; Valentini, E.; Nguyen Xuan, A.; Taramelli, A.; et al. Modelling the ground-LAI to satellite-NDVI (Sentinel-2) relationship considering variability sources due to crop type (*Triticum durum* L., *Zea mays* L., and *Medicago sativa* L.) and farm management. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI, Proceedings SPIE Remote Sensing, 2019, Strasbourg, France, 21 October 2019*; SPIE Press: 11149, 111490I.
14. Lebourgeois, V.; Dupuy, S.; Vintrou, É.; Ameline, M.; Butler, S.; Bégue, A. A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM). *Remote Sens.* **2017**, *9*, 259, doi:10.3390/rs9030259.
15. Hagolle, O.; Huc, M.; Desjardins, C.; Auer, S.; Richter, R. MAJA Algorithm Theoretical Basis Document. Available online: <https://doi.org/10.5281/zenodo.1209633> (accessed on 7 December 2017).
16. Weiss, M.; Baret, F. S2 ToolBox Level 2 Products: LAI, FAPAR, FCOVER. 2016. Available online: [https://step.esa.int/docs/ex-865tra/ATBD\\_S2ToolBox\\_L2B\\_V1.1.pdf](https://step.esa.int/docs/ex-865tra/ATBD_S2ToolBox_L2B_V1.1.pdf) (accessed on 31 March 2019).

17. Stineman, R.W. A consistently well behaved method of interpolation. *Creat. Comput.* **1980**, *6*, 54–57.
18. Filipponi, F.; Smiraglia, D.; Agrillo, E. Earth Observation for Phenological Metrics (EO4PM): Temporal discriminant to characterize forest ecosystems (manuscript in preparation).
19. Gu, L.; Post, W.; Baldocchi, D.; Black, T.; Suyker, A.; Verma, S.; Vesala, T.; Wofsy, S. Characterizing the seasonal dynamics of plant community photosynthesis across a range of vegetation types. In *Phenology of Ecosystem Processes*, Noormets, A., Ed.; Springer: New York, NY, USA, 2009; pp. 35–58, ISBN 978-1-4419-0026-5 2.
20. Zuur, A.F.; Ieno, E.N.; Elphick, C.S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **2010**, *1*, 3–14, doi: 10.1111/j.2041-210X.2009.00001.x.
21. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. mlr: Machine Learning in R. *J. Mach. Learn. Res.* **2016**, *17*, 1–5, doi: 10.5555/2946645.3053452
22. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17, doi:10.18637/jss.v077.i01.