

Scalable parcel-based crop identification scheme using Sentinel-2 data time-series for the monitoring of the Common Agricultural Policy

Vasileios Sitokonstantinou ^{1,*}, Ioannis Papoutsis ¹, Charalampos C. Kontoes ¹, Alberto Lafarga Arnal ², Ana Pilar Armesto ² and Jose Angel Garraza ²

¹ Institute for Space Applications and Remote Sensing, National Observatory of Athens, I. Metaxa and Vas. Pavlou St, Penteli, 15236 Athens, Greece; vsito@noa.gr (V.S.); ipapoutsis@noa.gr (I.P.); kontoes@noa.gr (C.C.K.)

² INTIA Tecnologías e Infraestructuras Agroalimentarias, Av. Serapio Huici, 22, 31610 Villava, Navarra, Spain

Crop type classification using a dense Sentinel-2 image time-series

In order to assess the importance and impact of a very dense time-series of Sentinel-2 imagery in crop type classification accuracy, we have performed an SVM (quadratic kernel) classification scheme that incorporates almost all available imagery within the year of inspection, irrespective of the cloud percentage. Thus, for the study site in Navarra, Spain we have acquired, then pre-processed and used 39 Sentinel-2 images (January to October – Figure S1) to classify crop types based on the 2017 farmer declarations, as part of their Common Agricultural Policy (CAP) subsidy applications.

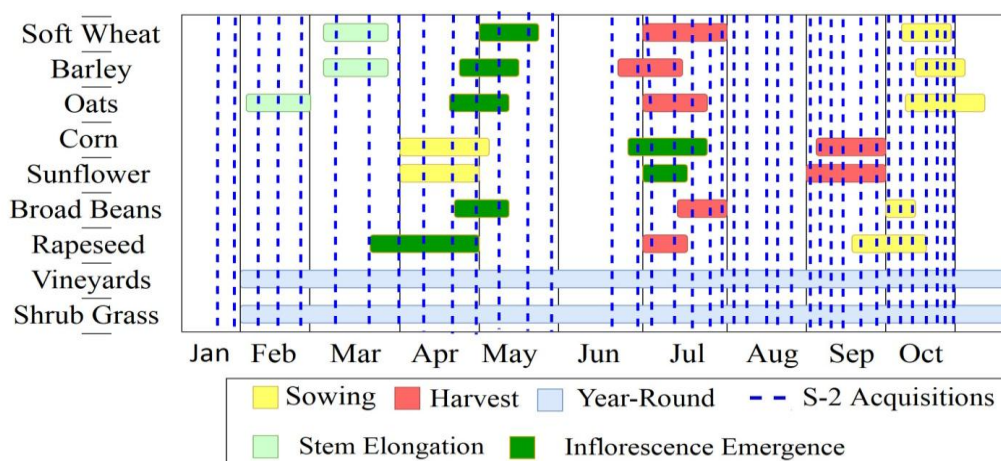


Figure S1. Timeline of phenology stages for key crops in Navarra, overlaid with the acquisition dates of Sentinel-2 images for 2017.

Methodology

The dataset comprises of all available Sentinel-2 imagery (including acquisitions from both S2A and S2B) and thus substantial cloud coverage is evident throughout the time-series. Therefore, cloud covered pixels within the time-series are first masked out and then missing values are filled in.

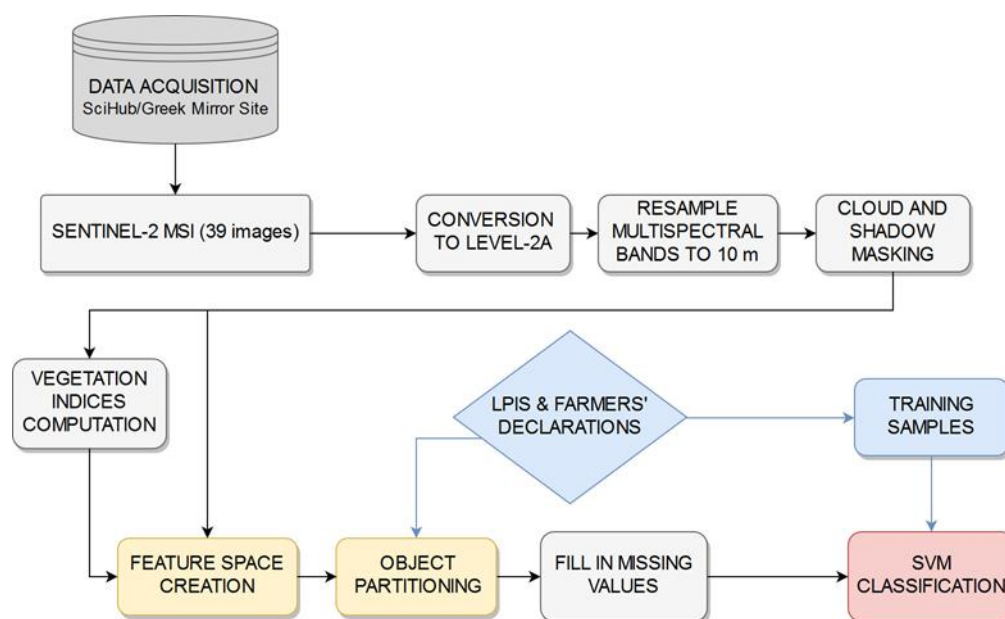


Figure S2. Flowchart of overall methodology: imagery acquisition, data pre-processing, cloud masking, feature space creation, fill in missing values, crop type classification

Figure S2 displays the processing workflow of the crop type identification scheme, from the acquisition of Sentinel-2 data to the SVM classification. The feature space, on which the supervised classification algorithm is applied to, comprises of the time-series of Sentinel-2 imagery and the vegetation indices Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Plant Senescence Reflectance Index (PSRI), computed for each image acquisition (Table S1).

Table S1. Sentinel-2 MSI feature space of multispectral band reflectances and vegetation indices

Type	Sentinel-2 MSI
Reflectances	10 Bands [B02, B03, B04 (VIS), B08, B8A (NIR) of 10 m spatial resolution, B05, B06, B07 (Red-Edge) and B11, B12 (SWIR) of 20 m spatial resolution] x 39 image acquisition (390 features)
Vegetation Indices	NDVI, PSRI, NDWI x 39 image acquisitions (117 features)

In order to mask out the shadowed and clouded pixels of each image, we used the Level-2A scene classification product (SC)¹ that is offered as a byproduct of the Sen2Cor atmospheric and terrain correction process. Each mask accounts for shadows, clouds of medium and high probability, cirrus clouds and snow (classes 3, 8, 9, 10, 11 of the SC product). Then for each image in the time-series, the pixels classified to the aforementioned categories are given “no data” values.

Prior to the application of the classifier, the feature space is transformed from the pixel level to the object level by averaging the pixel values within the boundaries of each Land Parcel Identification System (LPIS) polygon. During this object partitioning process, the parcels that have more than 50% of their pixels masked out are marked as “cloud covered”, while all else are marked as “cloud-free”. The feature values of the “cloud-free” parcels are computed using only the unmasked pixels of the parcel. On the other end, utilizing the a priori knowledge of the farmers’ declarations, and by assuming their validity, the “no data” feature values of the “cloud covered” parcels are filled in, as described in the following two steps (Figure S3).

¹ <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm>

1. For each Sentinel-2 acquisition used in the time-series stack, calculate the average value of the 9 crop type classes for each feature (spectral bands and vegetation indices). Only “cloud-free” parcels partake in the computation.
2. For each feature and for each “cloud covered” parcel, the “no data” values are replaced with the average remote sensing value, as calculated in step 1, considering their declared crop type class.

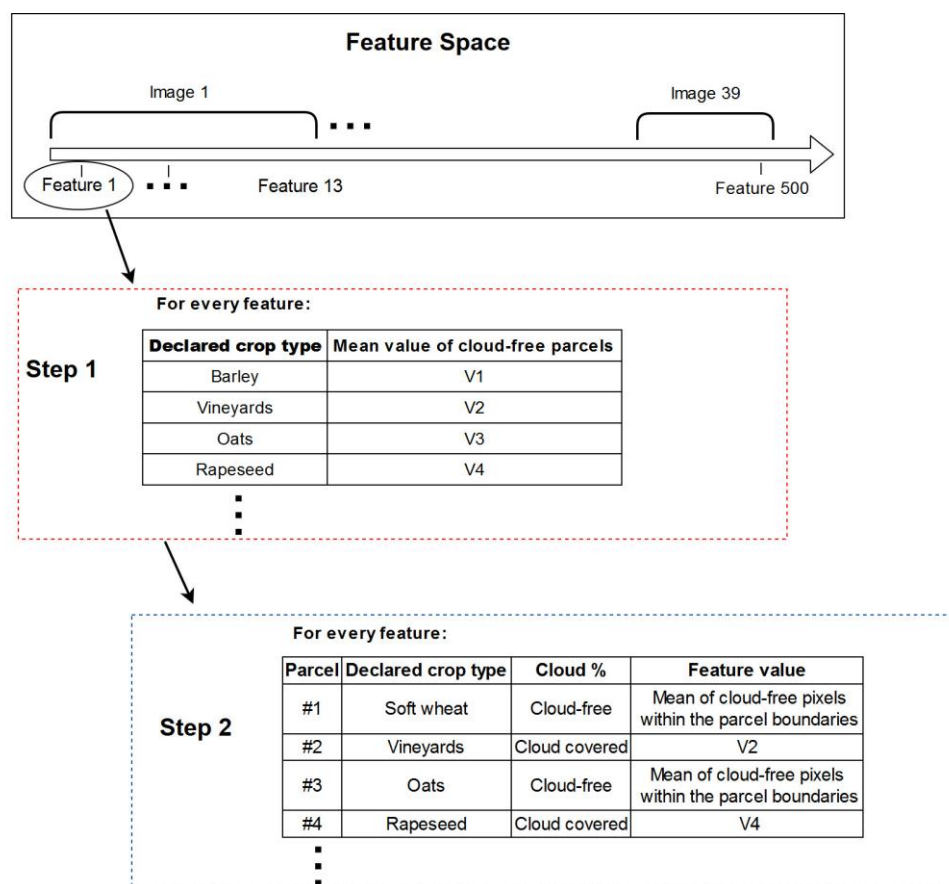


Figure S3. Filling missing values of “cloud covered” parcels. The average value for each crop type class is calculated from the corresponding “cloud-free” parcels for each feature. Then “cloud-covered” parcel values, for each feature, are substituted with the respective average class value, based on their declared crop type.

Images of more than 80% cloud coverage, over the area of interest, bypass the process of masking and missing value filling and are directly incorporated into the feature space after resampling (see Figure S2). This is done to avoid having features, constituted by parcel objects of forced values that have been computed based on a limited number of “cloud-free” instances.

In order to evaluate the influence of cloud coverage in crop type classification, the 39 Sentinel-2 images were then split into “clouded” and “cloudless” classes, with 19 and 20 images respectively. The division into the two classes was performed by calculating the number of all masked pixels falling within area occupied by the parcels. Images with cloud coverage less than 3% are marked as “cloudless”, while the “cloudy” class contains all remaining imagery. Prior to the crop type classification, features are ranked based on their importance that is computed using the *Relieff* algorithm in MATLAB.

Results

Figure S4 below displays the contribution of the “clouded” and “cloudless” image classes, in number of features, for different quintiles of the ranked feature space.

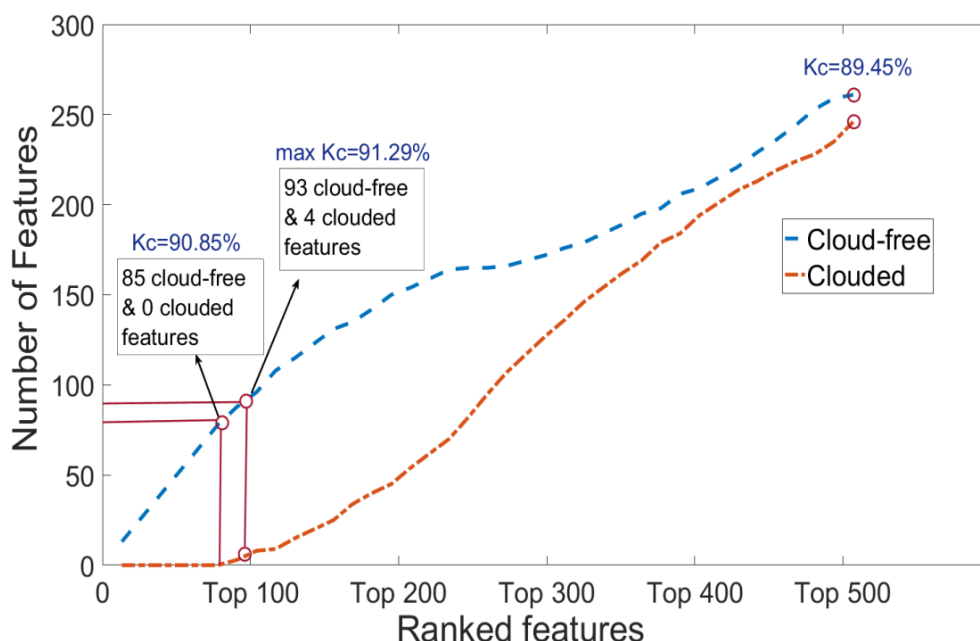


Figure S4. Contribution of “clouded” and “cloudless” images in the top ranked features of the variable space

The analysis of feature importance clearly shows that “cloudless” images are the most important in the classification process. Sorting the feature space according to the individual feature importance weights revealed that the 85 top ranked features came exclusively from the “cloudless” class. Utilizing solely these top 85 features resulted in a crop type classification of an overall Cohen’s kappa 90.85%. The features come only from images that were sensed in March (30%), April (30%), June (18%) and July (22%). Imagery that was sensed in May is absent, as all three acquisitions within the month had cloud coverage larger than 62%.

The highest accuracy, Cohen’s kappa of 91.29%, is achieved when classifying with the first 97 features of the ranked feature space; utilizing 93 cloud-free and 4 clouded features. On the other hand, using the entirety of the feature space (507 features) resulted in 89.45% accuracy. It becomes evident that images coming exclusively from the “cloudless” class are responsible for a near maximum overall accuracy, and therefore the incorporation “clouded” imagery is considered redundant.

It can be argued that this approach does not justify the extra processing effort that is demanded, for merely a marginal increase in thematic accuracy; particularly for operational applications, where computational efficiency is an important consideration. It can be also observed that the top ranked features stem from imagery that was sensed over several different months, thus covering the better part of the phenology timeline, as shown in Figure S1. All in all, cloud-free images that cover critical phenological stages make up the optimal feature space, for both accurate and computationally efficient crop type classification.

