Detail of the processing flow of the methodology for the tree cover classification

This section describes the processing flow relating to the tree cover classification methodology developed by using the Sentinel-2 L1C images and Sentinel-1 GRD images.

Sentinel-2 processing

The developed tree cover classification methodology is based on the following steps for Sentinel-2 images, applied separately to all the tiles covering Italy:

1. Selection of Sentinel-2 images acquired during summer 2017 (from June 1st to August 31st).
2. Calculation of the NDVI index for every Sentinel-2 image; NDVI was calculated using the bands 8 and 4 that have 10-m spatial resolution:

$$NDVI = \frac{Band\,8 - Band\,4}{Band\,8 + Band\,4}$$

3. Calculation of the raster of maximum NDVI value per pixel (named Maximum NDVI Summer), processing the summer NDVI rasters; this process is based on the assumption that, during summer, tree covered pixels reach higher NDVI values than non-tree covered pixels (nevertheless, other vegetation types such as grassland or crops can reach very high NDVI values).
4. Calculation of a value (named Threshold Maximum NDVI Summer) to be used as threshold for the raster Maximum NDVI Summer, with the assumption that for tree covered pixels the following condition is true:

Maximum NDVI Summer >= Threshold Maximum NDVI Summer

The value Threshold Maximum NDVI Summer is the 10th percentile of Maximum NDVI Summer values calculated on tree training areas (created for various tree cover types to account for variability of NDVI values). The 10th percentile was selected in order to exclude

possible outliers due to cloud shadows or topographic effects that could be included in tree training areas.

A Threshold Maximum NDVI Summer was calculated for each Sentinel-2 tile separately.

5. For each NDVI raster, calculation of a binary raster, applying the following conditional statement:
    - Set 0 where NDVI < 0.5;
    - Set 1 where NDVI ≥ 0.5;

    The threshold 0.5 has been chosen after several empirical tests, as optimal value for detecting tree covered pixels, in order to exclude non-tree covered pixels (such as grassland or crops) that usually have lower NDVI values when the soil is exposed because of harvesting or dry grass.
6. Processing of the previous NDVI binary rasters to get the percentage of occurrence of NDVI ≥ 0.5 (raster named High NDVI Summer), calculated as:

$$High\ NDVI\ Summer = \sum_{i=1}^{n} \frac{NDVI\ Binary_i}{n} \times 100$$

Where:

- n is the total number of acquisitions (excluding masked pixels);
- $NDVI\ Binary_i$ is the raster calculated in step 5, for the i image.

The assumption of this process is that, during summer, tree cover has a higher percentage of occurrence of NDVI values ≥ 0.5 than other vegetation cover (e.g. when crops are harvested NDVI values drop; during summer grassland tends to become dry and have lower NDVI values).

Sentinel-1 processing

The developed tree cover classification methodology also relies on the processing of Sentinel-1 images, in order to detect low vegetation pixels and exclude them from tree cover, according to the following steps:

1. Selection of Sentinel-1 images acquired during the whole year of 2017.
2. Selection of Sentinel-1 VH Polarization Ascending (Descending) orbit. Ascending and descending orbit are processed separately in order to account for geometric and topographic effects that can affect backscatter values.

3. For each Sentinel-1 image (separately for ascending and descending orbit), calculation of binary rasters (named VH Ascending Binary Raster and VH Descending Binary Raster), applying the following conditional statement:
    - Set 0 where backscatter > -20 dB;
    - Set 1 where backscatter ≤ -20 dB;

    The threshold -20 dB has been chosen after several empirical tests, based on the assumption that tree covered pixels tend to have higher backscatter values than other low vegetation pixels (such as grassland or crops).
4. Processing of the previous backscatter binary rasters (ascending and descending separately) to get the percentage of occurrence of VH backscatter ≤ -20 (two rasters named Percentage Ascending and Percentage Descending), calculated as:

$$Percentage\ Ascending = \sum_{i=1}^{n} \frac{VH\ Binary\ Ascending_i}{n} \times 100$$

$$Percentage\ Descending = \sum_{i=1}^{m} \frac{VH\ Binary\ Descending_i}{m} \times 100$$

    Where:
    - n is the total number of acquisitions in ascending orbit;
    - $VH\ Binary\ Ascending_i$ is the raster calculated in step 3, for the i image in ascending orbit;
    - m is the total number of acquisitions in descending orbit;
    - $VH\ Binary\ Descending_i$ is the raster calculated in step 3, for the i image in descending orbit.
5. Calculation of the VH Binary Raster with the following concurrent conditions on the percentage of occurrence:
    - Set 0 where Percentage Ascending ≥ 2% and Percentage Descending ≥ 2%;
    - Set 1 where Percentage Ascending < 2% and Percentage Descending < 2%.

    This calculation assumes that tree cover has backscatter values lower than -20 dB for less than 2% of the yearly acquisition, both in ascending and descending order.

Tree cover classification

The tree cover classification (woody vegetation) is calculated from the rasters resulting from the previous processing, with the following concurrent conditions:

- *High NDVI Summer > 70%*; this raster is derived from Sentinel-2 images; the threshold 70% has been empirically estimated as optimal value, assuming that tree covered pixels have NDVI >= 0.5 for at least 70% of the summer period.
- *Maximum NDVI Summer ≥ Threshold Maximum NDVI Summer*; this condition was calculated for each Sentinel-2 tile separately, in order to adapt the threshold to the different climatic conditions and vegetation characteristics.
- *VH Binary raster = 1*; this condition derived from Sentinel-1 allows for excluding low vegetation (especially cropland) from tree cover classification.

The methodology has been applied to produce a tree cover classification of 2017 images. The processed tiles over Italy have been mosaicked in order to produce a seamless tree cover classification

with 10-m spatial resolution (**Figure 1**).

A preliminary accuracy assessment was performed on a few study areas, showing an overall accuracy above 80%, although in a few regions (especially in southern Italy) classification issues have been observed for shrubland, because of the similarities with trees regarding spectral signatures, phenology, and backscatter values. In particular, in some areas of the south of Italy, such as Apulia and Calabria, the accuracy was lower than 80% due to the presence of olive trees not classified as tree cover (**Figure 2**). This is mainly due to the very low NDVI values of tree covered pixels, that barely reach 0.4 values during the summer months, probably because of the aridity of the area, the particular pattern of olive cultivations (e.g. distance between trees) that expose bare soil causing the mix of soil and tree spectral signatures (i.e. mixed pixels). Moreover, the Sentinel-2 spatial resolution (i.e. 10m) is often too broad to detect sparse olive trees. Further study is required to account for these low NDVI values contradicting the assumptions described in step 5 of the methodology for Sentinel-2 processing. Manual corrections were applied to these areas in order to improve the classification.

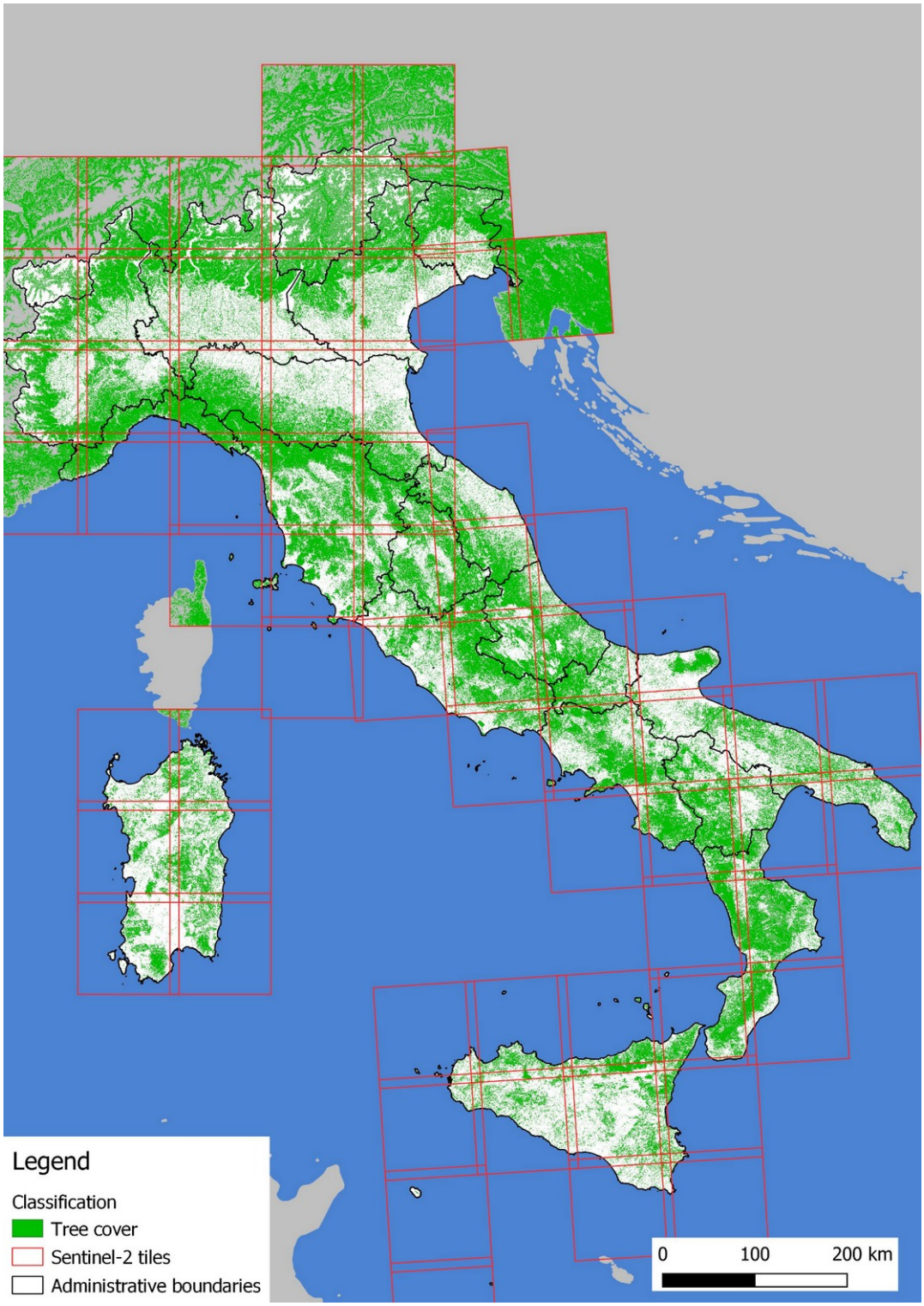

Figure 1: Tree cover classification (green) over Italy for 2017; all the Sentinel-2 tiles were classified according to the developed methodology.

Figure 2: Example of tree cover classification (green) in Apulia (Google Earth image, 2017); olive trees are partially omitted.