A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses

Rosa Coluzzi[*], Vito Imbrenda, Maria Lanfredi, Tiziana Simoniello

IMAA-CNR (Institute of Methodologies for Environmental Analysis - National Research Council of Italy), C.da Santa Loja, 85050 Tito Scalo, PZ, Italy

A B S T R A C T

Cloud detection in optical remote sensing images is a crucial problem because undetected clouds can produce misleading results in the analyses of surface and atmospheric parameters. Sentinel-2 provides high spatial resolution satellite data distributed with associated cloud masks. In this paper, we evaluate the ability of Sentinel-2 Level-1C cloud mask products to discriminate clouds over a variety of biogeographic scenarios and in different cloudiness conditions. Reference cloud masks for the identification of misdetection were generated by applying a local thresholding method that analyses Sentinel-2 Band 2 (0.490 μm) and Band 10 (1.375 μm) separately; histogram-based thresholds were locally tuned by checking the single bands and the natural color composite (B4B3B2); in doubtful cases, NDVI and DEM were also analyzed to refine the masks; the B2B11B12 composite was used to separate snow.

The analysis of the cloud classification errors obtained for our test sites allowed us to get important inferences of general value. The L1C cloud mask generally underestimated the presence of clouds (average Omission Error, OE, 37.4%); this error increased (OE > 50%) for imagery containing opaque clouds with a large transitional zone (between the cloud core and clear areas) and cirrus clouds, fragmentation emerged as a major source of omission errors (R^2 0.73). Overestimation was prevalently found in the presence of holes inside the main cloud bodies. Two extreme environments were particularly critical for the L1C cloud mask product. Detection over Amazonian rainforests was highly inefficient (OE > 70%) due to the presence of complex cloudiness and high water vapor content. On the other hand, Alpine orography under dry atmosphere created false cirrus clouds. Altogether, cirrus detection was the most inefficient. According to our results, Sentinel-2 L1C users should take some simple precautions while waiting for ESA improved cloud detection products.

1. Introduction

Sentinel-2 is one of the core missions of the Copernicus Earth Observation programme of the European Union. With enhanced observation capabilities, it ensures continuity and complementarity with Landsat and SPOT (Satellite Pour l'Observation de la Terre) observations (Fletcher and European Space Agency, 2012). This mission aims to meet different user needs and to improve numerous Copernicus operational applications (ESA Sentinel-2 Team, 2007) such as:

- land monitoring service: land use and land cover state and changes; biogeophysical parameters estimation; forest monitoring; urban mapping; spatial planning; agro-environmental monitoring; natural resource monitoring; land carbon/carbon storage; global crop monitoring; coastal zone monitoring; soil sealing;
- risk management: floods and forest fires, subsidence and landslides,

volcano eruptions;
- food security/early warning systems;
- water management;
- soil protection;
- terrestrial mapping for humanitarian aid and development;
- global change issues.

Despite the short lifetime, many authors have already experienced the great potentialities of Sentinel-2 data to: classify crop and tree species (Immitzer et al., 2016); monitor natural and anthropic vegetation (Bontemps et al., 2015; Greco et al., 2018; Song et al., 2017); map glaciers (Paul et al., 2016) and water bodies (Du et al., 2016; Toming et al., 2016; Yesou et al., 2016); assess and monitor water constituents (Dörnhöfer et al., 2016); classify burn severity (Fernández-Manso et al., 2016; Huang et al., 2016); map built-up areas (Lefebvre et al., 2016; Pesaresi et al., 2016); and detect Sub-

Pixel Landscape Feature (Radoux et al., 2016). More in general, these data have proven to be useful in several geological studies (van der Werff and van der Meer, 2016).

As an optical remote sensing system operating in the wavelength range between 0.443 μm and 2.190 μm, Sentinel-2 data are sensitive to cloud cover. To correctly implement Copernicus applications and, in general, for retrieving accurate surface parameters, the first required step is the detection of clouds into the Sentinel-2 MSI (MultiSpectral Instrument) imagery because these can severely disturb the correct extraction of atmospheric or surface information using optical remote sensing satellite data (Greenhough et al., 2005; Huete et al., 2002; Kaufman, 1987; Nakajima et al., 2011; Woodcock et al., 2008; Gao and Li, 2017).

From an optical remote sensing point of view, clouds can be roughly classified in:

- opaque/dense clouds, mainly composed of droplets that are highly reflective in the VIS region and generally located at low-medium altitudes;
- cirrus, consisting of a large number of thin non-spherical ice crystals that are normally translucent in the VIS region, relatively highly reflective in the SWIR spectrum, and located at high altitude.

Clouds play an important role in climate variability and change as they reflect solar (albedo effect) and terrestrial (greenhouse effect) infrared radiation. The different spectral properties make cirrus the principal cloud type controlling the Earth's radiation budget (Dessler and Loeb, 2013; Zhou et al., 2013). Clouds usually cover large portions of the Earth's surface: globally cirrus clouds, the most difficult to be detected (Xu et al., 2014), are thought to cover on average about 17% of the surface with > 50% in the Intertropical Convergence Zone (Kovalskyy and Roy, 2015 and references therein). Recent estimates suggested that low altitude clouds, typically non-cirrus, may co-exist with spatially overlapping cirrus clouds and that, more in general, about 30% of low clouds are overlapped by high clouds (Kovalskyy and Roy, 2015 and reference therein).

Clouds largely affect the transmission of radiation between the satellite sensors and surface targets, reducing the ability to identify land covers, decreasing the accuracy of the retrieval of surface parameters, and providing misleading information on monitored surfaces (Kazantzidis et al., 2011; Lanfredi et al., 2015; Li et al., 2011; Sun et al., 2016). The multiple types of clouds and the complexity of land structures further complicate their detection due to the frequent difficulty in distinguishing clouds from the underlying landscape, even at high resolution (Hagolle et al., 2010; Jedlovec et al., 2008).

Sentinel-2 MSI, even if lacks thermal bands which are included in other sensors such as Landsat and ASTER, has spectral bands useful to provide better cloud screening and compensation of atmospheric effects: Band 1 (centered at 0.443 μm with 60 m of spatial resolution), Band 2 (centered at 0.490 μm with 10 m of spatial resolution), Band 9 (centered at 0.945 μm with 60 m of spatial resolution), Band 10 (centered at 1.375 μm with 60 m of spatial resolution), and Band 12 (centered at 2.190 μm with 20 m of spatial resolution). In particular, Band 10 (hereafter B10) at 1.375 μm is based on MODIS (Moderate Resolution Imaging Spectroradiometer) and Landsat 8 OLI (Operational Land Imager) sensors design heritage and provides the possibility to examine the incidence of cirrus clouds. The opportunity to dispose of all this information, contextually acquired by the same sensor, is rather new in satellite systems primarily intended for operational services. It represents a prerequisite for establishing fundamental environmental and climate data records to obtain consistent products and harmonized time series (Berger et al., 2012; Simoniello et al., 2004).

Sentinel-2 images are among the first high resolution data distributed with an associated cloud mask, followed by the VENμS (Vegetation and Environment monitoring on a New MicroSatellite) mission (French and Israeli cooperation) launched on 1 August 2017. Sentinel-2 Level-1C (L1C) products include cloud masks to simplify the user's work and to provide many useful products ready for delivery. Cloud masks identify cloudy pixels and separate them from those that are cloud-free. They include both dense clouds and cirrus clouds, by specifying the cloud type with an indicator. Moreover, statistical information about the percentage of dense cloud and cirrus pixels is enclosed.

During these two years of activity, many efforts have been made by ESA and related Sentinel-2 working groups to improve the performance of Sentinel-2 data and products. Corrections were implemented to obtain a high signal to noise ratio in L1C bands. In particular, for SWIR B10, scattered pixels labeled as "no data" because of the presence of instrument noise have been corrected. Sentinel-2 cloud masks are currently adjusted to minimize underdetections, which leads, on the other hand, to overdetections (Clerc et al., 2015).

By taking into account the relevance of cloud-free images for surface characterization and the peculiarities of the Sentinel-2 imagery, the aim of this study is to perform a first assessment of the cloud mask product included in the Sentinel-2 L1C data.

This study wants to evaluate if: 1) Level-1C cloud masks are enough accurate to select cloud-free pixels for land users' needs; 2) there are differences in performance depending on the environmental conditions; 3) the type and the configuration of clouds influence the detection accuracy.

To this aim, we selected scenes from different biogeographic scenarios and cloudiness conditions to infer the most common error sources as well as to identify environments that are particularly critical for the L1C cloud detection algorithm. Such an assessment can be very useful to support informed surface studies, based on high quality data.

2. Study areas and data

To assess the Sentinel-2 L1C cloud mask product in different environmental and sensing conditions, we selected scenes from different biogeographic regions, characterized by heterogeneous land cover and located at different altitudes (from the sea level to above 2000 m a.s.l.). Sentinel-2 tiling grid in KML format (available at the following link: https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products, last access 25/01/2018) was used to obtain a preview of the test-sites on Google Earth.

Finally, nine Sentinel-2 L1C images, acquired in different seasons (in the period August 2015 – March 2016), characterized by different cloud types (cirrus and opaque) and cloud coverage (0.37–50%), were downloaded (Table 1). Moreover, the selected images were processed in different ground-stations: Matera Core Ground Segment (MTI), Svalbard Core Ground Segment (SGS), and Spanish Processing Archive Centre located in Madrid (EPA).

Even if the images have different processing baselines (from N02.00 to N02.04), the corrections implemented in the processing chain do not involve modifications in the cloud mask processing steps.

Data are available at no cost at the Sentinel Scientific Data Hub, the public archive of Sentinel-2 products (https://scihub.copernicus.eu/dhus/#/home). The global distribution of the selected test sites is illustrated in Fig. 1.

Level-1C is composed of granules, also called tiles, of about 110 km × 110 km (10,980 × 10,980 pixels at 10 m of spatial resolution) ortho-images in UTM/WGS84 (Universal Transverse Mercator/World Geodetic System 1984) projection. Each tile consists of 13 compressed JPEG-2000 images (total spectral range between 0.443 μm and 2.190 μm); each image represents one single band (intensity range of 15 bits) with a spatial resolution depending on its native resolution: 10 m for the 4 VIS/NIR bands (the classical blue at 0.490 μm, green at 0.560 μm, red at 0.665 μm, and near-infrared at 0.842 μm); 20 m for the 6 red-edge/shortwave infrared bands (four narrow bands in the vegetation red-edge spectral domain respectively at 0.705 μm, 0.740 μm, 0.783 μm, and 0.865 μm, and two large SWIR bands at 1.610 μm and 2.190 μm); and 60 m for the 3 atmospheric correction and cloud

Table 1

List of the selected Sentinel-2 images for testing the L1C cloud mask product with the main acquisition characteristics: sensing date, location, cloud cover percentage, surface features, solar zenith angle (SZA), incidence/view zenith angle (IZA) for band 2 (B2) and band 10 (B10). The first part of the granule (tile) name is the same for all the identified images (S2A_OPER_MSI_L1C_TL); the granule ID reports the second part.

Granule ID	Sensing date	UTM zone	Clouds percentage[a]	Surface features[b]	SZA	IZA B2	IZA B10
SGS_20151225T153139_A002650N02.01	2015-12-25	33TVF	50.00	F/A/B/S	66.07	4.88	4.95
MTI_20160306T224020_A003683.N02.01	2016-03-06	18TYL	34.66	Se/B/F	49.61	8.79	8.76
SGS_20151211T155243_A002450.N02.00	2015-12-11	32TQR	32.00	F/A/B/Se	69.29	3.59	3.65
SGS_20151208T152443_A002407.N02.00	2015-12-08	32SNE	19.00	D	60.18	8.47	8.50
SGS_20160101T063159_A002744.N02.01	2016-01-01	55HGV	10.00	F/A	25.48	3.67	3.76
EPA_20160706T084949_A000877.N02.04	2015-08-23	34VCJ	11.36	F/A/B	46.20	6.06	6.09
EPA_20151202T202847_A002293.N02.00	2015-11-30	31TDF	1.00	B/Se	63.70	2.62	2.70
EPA_20160720T202438_A000648.N02.04	2015-08-07	33TXF	0.37	A/BS	27.71	9.04	9.07
SGS_20160217T201717_A003425.N02.01	2016-02-17	19NHC	8.01	F	29.73	5.47	5.52

[a] According to the Quality_Indicators_Info into the metadata file of the granule.

[b] A = agriculture, B = built-up, BS = bare soil, D = desert, F = forest, S = snow, Se = sea.

screening bands (0.443 μm for aerosol retrieval, 0.945 μm for water vapor retrieval, and 1.375 μm for cirrus cloud detection) (European Space Agency (ESA), 2015). For details about the Sentinel-2 spectral responses see Appendix A.

The L1C pixel values refer to the Top Of Atmosphere (TOA) reflectance. Cloud masks are associated with each tile. The product is also complemented by: general information regarding granule elements; geometric info providing the geolocation of the granule; quality indicators (radiometric quality, geometric quality, image content quality, quality control checks information); auxiliary data (Ground Image Processing Parameters – GIPP, Digital Elevation Model - DEM, and Global Reference Image - GRI, a set of unitary reference images covering one orbit); and meteorological ECMWF (European Center Medium

Weather Forecast) parameters (Total Column Ozone - TCO3, Total Column Water Vapor – TCWV, and Mean Sea Level Atmospheric Pressure - MSLP) (Thales Alenia Space, 2016).

2.1. Level-1C cloud mask product

Cloud masks associated with images are vectors in Geography Markup Language (GML) format, easy to be overlapped to raster band images. Level-1C cloud detection is based on a simple algorithm that identifies the following classes: opaque clouds, also called dense clouds, and cirrus (https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks, last access 25/01/2018).

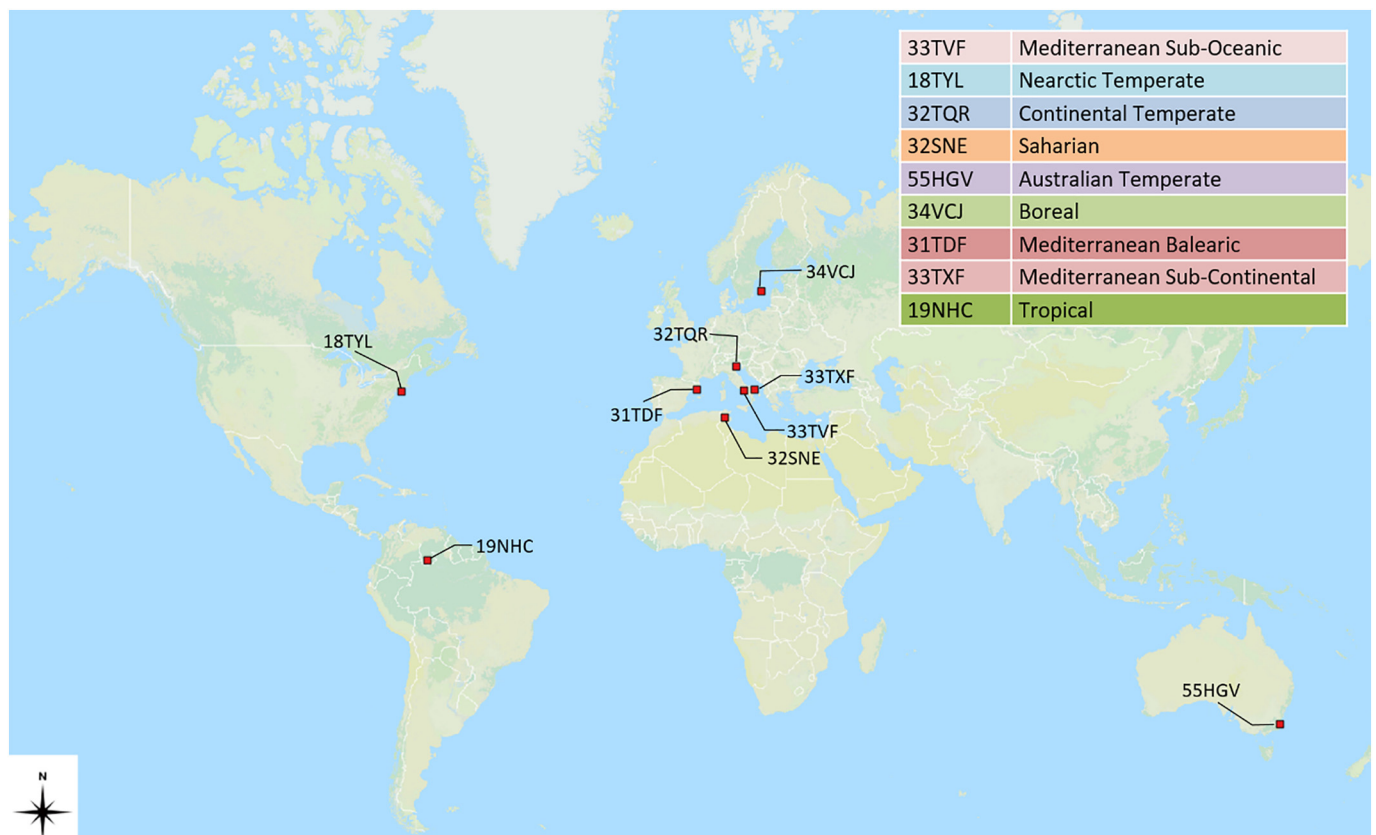LC1 algorithm is based on visible (Band 1 at 0.443 μm or Band 2 at

33TVF	Mediterranean Sub-Oceanic
18TYL	Nearctic Temperate
32TQR	Continental Temperate
32SNE	Saharian
55HGV	Australian Temperate
34VCJ	Boreal
31TDF	Mediterranean Balearic
33TXF	Mediterranean Sub-Continental
19NHC	Tropical

Fig. 1. Geographical distribution of selected evaluation sites.

0.490 μm) and SWIR spectral bands (Band 10 at 1.375 μm jointly with Band 11 at 1.610 μm or Band 12 at 2.190 μm). All the considered bands are resampled at the same spatial resolution (60 m) by a radiometric interpolation.

The L1C algorithm detects dense and cirrus clouds differently:

2.1.1. Dense cloud detection

A first threshold is applied on the blue spectral region B2 (0.490 μm) where dense clouds have a high reflectance. To avoid confusion between snow and clouds, SWIR channels B11 (1.610 μm) and B12 (2.190 μm) are also used because snow has a low reflectance at these wavelengths.

Since ice clouds and snows have both a low reflectance in the bands B11 and B12, an additional threshold on B10 (1.375 μm) is applied to include high altitude ice clouds eventually discarded by the previous SWIR thresholds (cirrus are not included at this step because they are transparent at Band 2 wavelengths).

2.1.2. Cirrus cloud detection

Band 2 and Band 10 are used because cirrus generally have low reflectance in the blue band (0.490 μm) and high reflectance in the SWIR band at 1.610 μm (Band 10 corresponds to a high atmospheric absorption band and only high altitude clouds are detected).

To limit false detections (due to high reflectance in the blue or due to the fact that clouds are not spectrally registered), the following filters using morphology-based operations are applied on both dense and cirrus masks:

- erosion, to remove isolated pixels;
- dilatation, to fill the gap and extend clouds.

The morphology operation is applied to the opaque and the cirrus cloud mask separately. The precise size of the neighboring pixels involved in dilation and erosion is taken from the associated GIPP. After morphology operations, if a pixel is classified both dense and cirrus, the dense cloud mask prevails.

Finally, the cloud mask is set to three values:

- 0 is a cloud-free pixel;
- 1 is a dense cloud pixel;
- 2 is a cirrus cloud pixel.

If measurements are not available in one or more bands needed to calculate the cloud mask, the value is set to NODATA. Statistical information on the percentage of opaque and cirrus cloud pixels is included in the metafile.

2.2. Auxiliary data

In order to evaluate the influence of the background (water/land) on the cloud detection performance, we used land/water masks derived from cloud-free Sentinel-2 tiles acquired in different dates and corresponding to the analyzed area (same UTM zone). The water mask was obtained by applying an upper threshold to B12 as the SWIR values for water are generally lower than those of land surfaces (Mishra and Prasad, 2015).

Thresholds were selected scene by scene. False positive results, i.e., pixels which were not from the water class such as some roofs, were manually removed. The land mask was obtained as complementary to the water mask.

Moreover, to evaluate the possible prevalence of specific land covers under misdetected clouds, we used the Climate Change Initiative Land Cover (CCI-LC) map by ESA (http://maps.elie.ucl.ac.be/CCI/viewer/download.php, last access 25/01/2018). This is a global LC map at 300 m derived from MERIS and SPOT-VGT data (ESA CCI LC project, 2016).

The map includes two different Levels: i) Level 1, containing information available at global scale; ii) Level 2, containing information at regional scale, where available. The map is also characterized by a set of quality flags.

We chose to adopt the CCI-LC 2010 map at Level 1 because it is globally consistent and available for all the case studies analyzed in this paper. For our purposes, we grouped the original 22 land cover classes into 7 macro-classes as follows: Cropland, Forest, Sparse Vegetation, Bare Soils, Water, Snow and Ice, Urban areas.

3. Method

To evaluate the reliability of the L1C cloud mask product, we firstly generated reference masks for each tile/granule and then compared them with the Sentinel-2 product.

To establish if a satellite image pixel is cloudy or free of clouds is not a simple matter. Usually the involvement of a trained expert is recognized as a key element to have confidence in the quality of cloud masks. Often experts manually perform cloud masks used for validating cloud cover algorithms. The U.S. Geological Survey currently distributes different cloud validation mask datasets manually derived from the Landsat data (Foga et al., 2017; Hughes and Hayes, 2014; M.J. Hughes, 2016; Scaramuzza et al., 2012, 2016; Scaramuzza and Dwyer, 2016. Data available at the following links: https://landsat.usgs.gov/landsat-7-cloud-cover-assessment-validation-data; https://landsat.usgs.gov/sparcs; https://landsat.usgs.gov/landsat-8-cloud-cover-assessment-validation-data. Last access 25/01/2018).

The reference cloud masks used in this work were generated by applying spectral techniques supervised by a trained expert backed up by support data. Such a supervised approach is expected to give a better performance than a truly automatic technique.

3.1. Reference cloud masks generation

The methodology designed for generating the reference cloud masks is based on an approach that considers cloud masking as a particular case of object-identification by band thresholding. Out of automatic processing chains, it represents one of the widely applied cloud detection method because of the simplicity of the algorithms, the fast operation, and the high reliability (see, e.g., Hagolle et al., 2010; Jedlovec et al., 2008; Zhu et al., 2015).

The adopted band thresholding roughly follows the approach of L1C product chain by considering Sentinel-2 Band 2 (0.490 μm) and Band 10 (1.375 μm) separately; for our purpose the thresholds are locally tuned to refine the cloud identification.

In particular, the blue spectral region of Band 2 is used to identify dense cloud pixels that show high brightness in that spectral region. This band is downsampled with an average filter at 60 m following L1C processing chain.

In Fig. 2 the spectral signature of typical opaque clouds (red stars) is compared with the signatures of different types of land cover (see the high reflectance values in VIS bands 1–4). Cirrus signature roughly follows the spectral profile of the background, water in this case. The high atmospheric adsorption of SWIR Band 10 is utilized to detect cirrus clouds, which have a comparable high reflectance at that wavelength (Hollstein et al., 2016), whereas only very low values are shown for the other classes (see orange stars at 1.375 μm - B10 in Fig. 2 and in particular the magnified panel 6n).

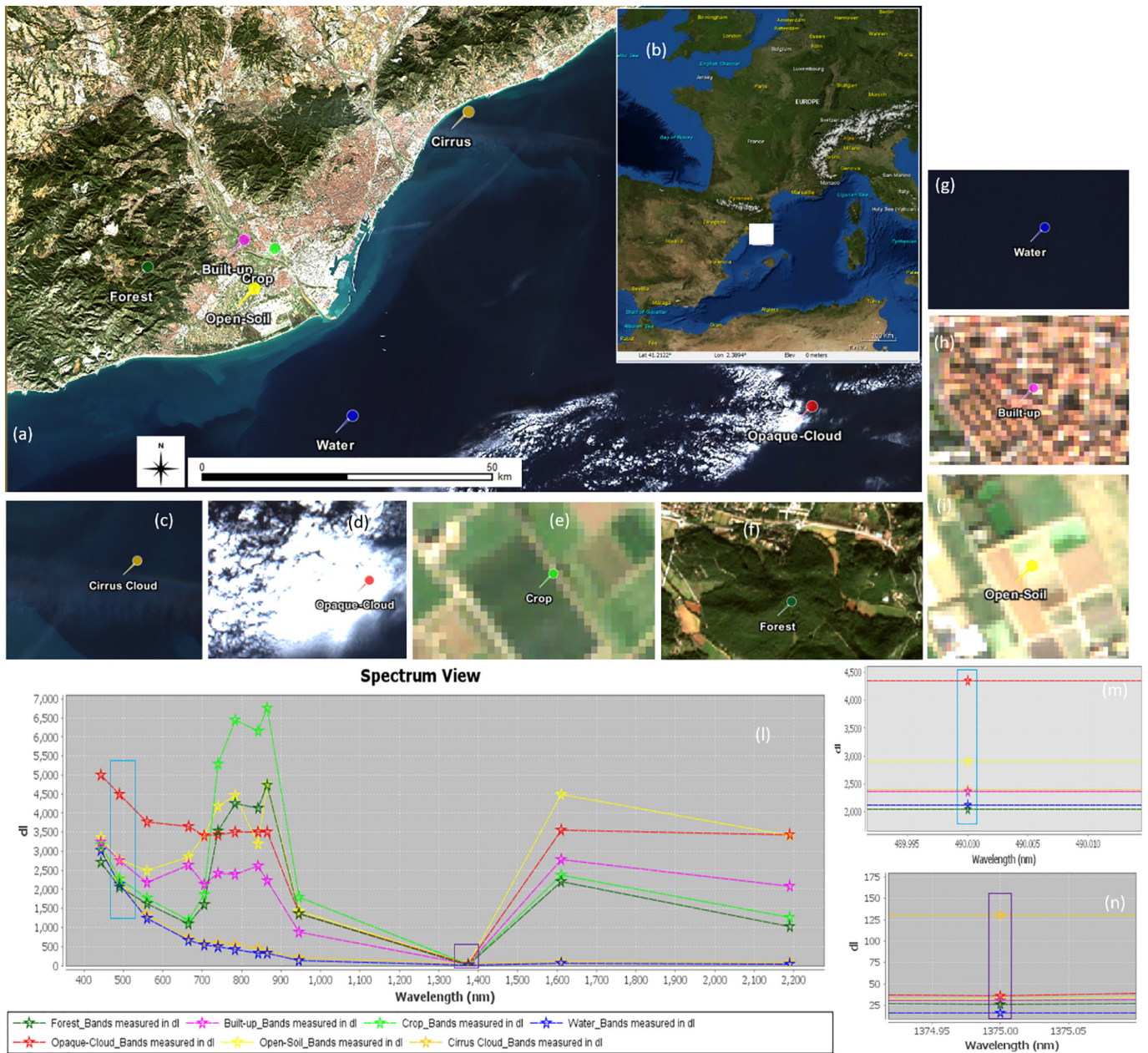The combination of the two bands (Band 2 and Band 10) was used

Fig. 2. Example of typical cloud signatures (dense/opaque red stars, cirrus orange stars) compared with the signatures of different land cover types selected on the Spanish sites (Granule 31TDF) (a-b). The profiles correspond to the pins located in the large image and magnified on the small panels: c) cirrus cloud; d) opaque cloud; e) crop; f) forest; g) water; h) build-up; i) open soil. As the spectrum view (l) shows, opaque clouds have the highest reflectance in VIS region, included in B2 utilized for their detection (m). Cirrus clouds have a signature similar to that of the water in background and the highest response in SWIR 1.375 μm - B10 (n). (dl in spectral profiles is for dimensionless TOA - Top Of Atmosphere - reflectances multiplied by 10,000). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for the detection of high-altitude, cold, and wispy opaque clouds that are transparent (with low reflectance) in the blue spectral region and show quite high reflectance at 1.375 μm (Acharya and Yang, 2015; Bo-Cai Gao et al., 2002; Gao, 2000; Gao et al., 1998; Gao and Li, 2017).

Cloud detection by thresholding a spectral band is primarily based on the evaluation of the grey-level histograms of each band to determine a reflectance threshold (in our application TOA reflectance value) which separates the histogram portion of the cloudy pixels from the cloud-free one (Bley and Deneke, 2013). Such a selection is generally scene-dependent due to the high variability of atmospheric conditions, clouds, and surface target characteristics. Fig. 3 shows the

variability of B10 signal distribution for the selected granules (tiles).

Cirrus clouds are usually found in the upper atmosphere and are composed of irregular ice crystals that can be less opaque than non-cirrus clouds at reflective wavelengths but can scatter and absorb radiation significantly (Kovalskyy and Roy, 2015 and references therein).

The signal received by the 1.375 μm band is absorbed by atmosphere, particularly by water vapor, therefore targets at surface appear dark (low B10 values) in a conventional greyscale image. On the contrary, since cirrus are located at higher altitude (the path-length radiation is shorter and the atmosphere is less absorbent), their signal is detectable and they appear bright in B10 (high values). Such a signal
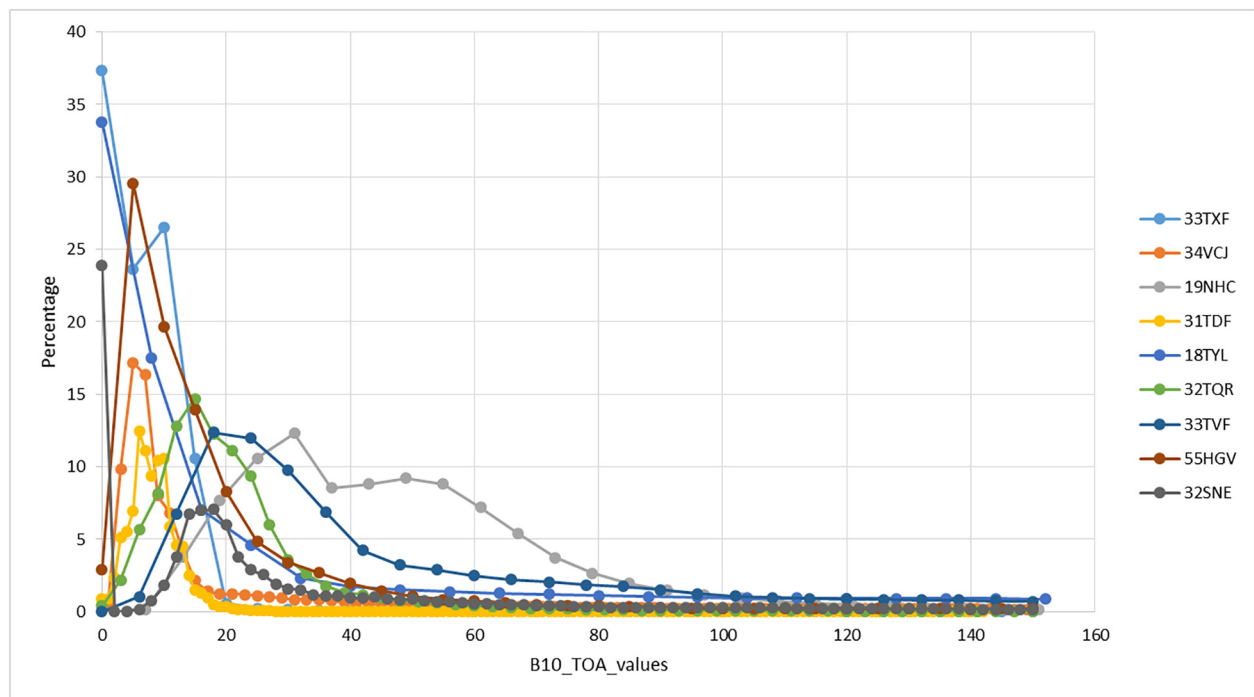
Fig. 3. Distribution of B10 signal (percentage of pixels) for the selected granules in Table 1 (B10 unit is dl = dimensionless Top Of Atmosphere (TOA) reflectances multiplied by 10,000).

strongly depends on the absorption and the scattering properties of cirrus clouds (optical depth, single scattering albedo, and phase function (Meyer et al., 2004)) as well as the water vapor absorption above, within, and below the cirrus clouds. Therefore, to a first approximation the threshold value for separating clouds from cloud-free pixels was defined scene by scene by selecting the TOA reflectance value that best marks the upper distribution tail corresponding to the areas visually identified as cloudy.

To verify and refine the histogram-based threshold, we adopted a visual interpretation method based on single band (B2, B10) and examination of others bands and image composites. We basically used the classical natural color composite (B4B3B2) as well as the B2B11B12 color composite that is useful to display snow and to distinguish it from clouds (Hollstein et al., 2016; Sun et al., 2016).

Radiometric enhancement filters (e.g., adaptive, gaussian, equalization) were applied to single channel or band composites to better visualize the cloud presence in the image. Consequently, the threshold on B2 or B10 was tuned to obtain the best overlap between the visualized clouds and the provisional cloud mask (optimal threshold identification - best mask).

When a suitable threshold is defined, each TOA reflectance band is converted in a binary map: background is set to zero (all values less than the threshold); foreground is set to one (all reflectance values higher than the threshold).

Then, in order to minimize the potential errors due to the inclusion of cloud-free pixels in the masks, we also considered:

- Normalized Difference Vegetation Index (NDVI) map to refine the mask of opaque clouds;
- Digital Elevation Model (DEM) to refine the mask of cirrus cloud.

To eliminate the presence of highly (and usually spatially isolated) reflective pixels in the VIS sometimes identified as clouds, such as bright bare soils and white buildings, the NDVI was calculated by using Sentinel-2 Band8 (NIR) and Band4 (RED) (at 10 m of spatial resolution

then resampled at 60 m).

NDVI values range from −1 to +1. It represents one of the most adopted index for surface analysis (Coluzzi et al., 2007; Pignatti et al., 2015; Simoniello et al., 2015). Negative values of NDVI generally correspond to water and dense clouds; values close to zero are related to barren areas of rock, sand, and snow; medium, positive values represent shrub and grassland; whereas high positive values indicate densely vegetated and forest areas. As the NDVI values of bare soil and build-up areas are higher than those of clouds, to eliminate highly reflective pixels in B2, we applied a threshold on the NDVI maps. In particular cases, where some doubtful bright patterns were retained, we observed the persistence of such patterns in multitemporal images acquired around the sensing date of the considered granule.

To remove false cirrus in B10 due to low water vapor content in correspondence of areas at high altitude, a DEM was used. B10 detects the radiation that is reflected by the high-altitude clouds and absorbed by the water vapor closer to the ground; so, as reported above, cirrus clouds appear largely brighter than other targets in presence of cloud-free atmospheres that contain water vapor. The detection of cirrus clouds using this channel critically depends on the total absorption of solar radiation by water vapor. If the atmosphere is relatively dry (low water vapor content), land surfaces at high altitude can appear bright as well due to the short radiation path-length (Acharya and Yang, 2015).

These false cirrus clouds usually appear with shapes similar to the orographic patterns and can be easily removed manually or by applying a spatial filter that compares the potential cloud shapes with the contourline extracted from the DEM.

Finally, we applied an erosion filter to remove isolated pixels from the cloud masks; we also used an adaptive average filter (3 × 3 box) to replace values altered by cloud presence with estimations obtained on the basis of the surrounding valid pixels.

At this point, two different cloud masks were generated: one starting from Band 2 (generally associated to opaque clouds) and the other one starting from Band 10 (generally associated to cirrus clouds), both processed at 60 m. As our procedure is principally focused on the
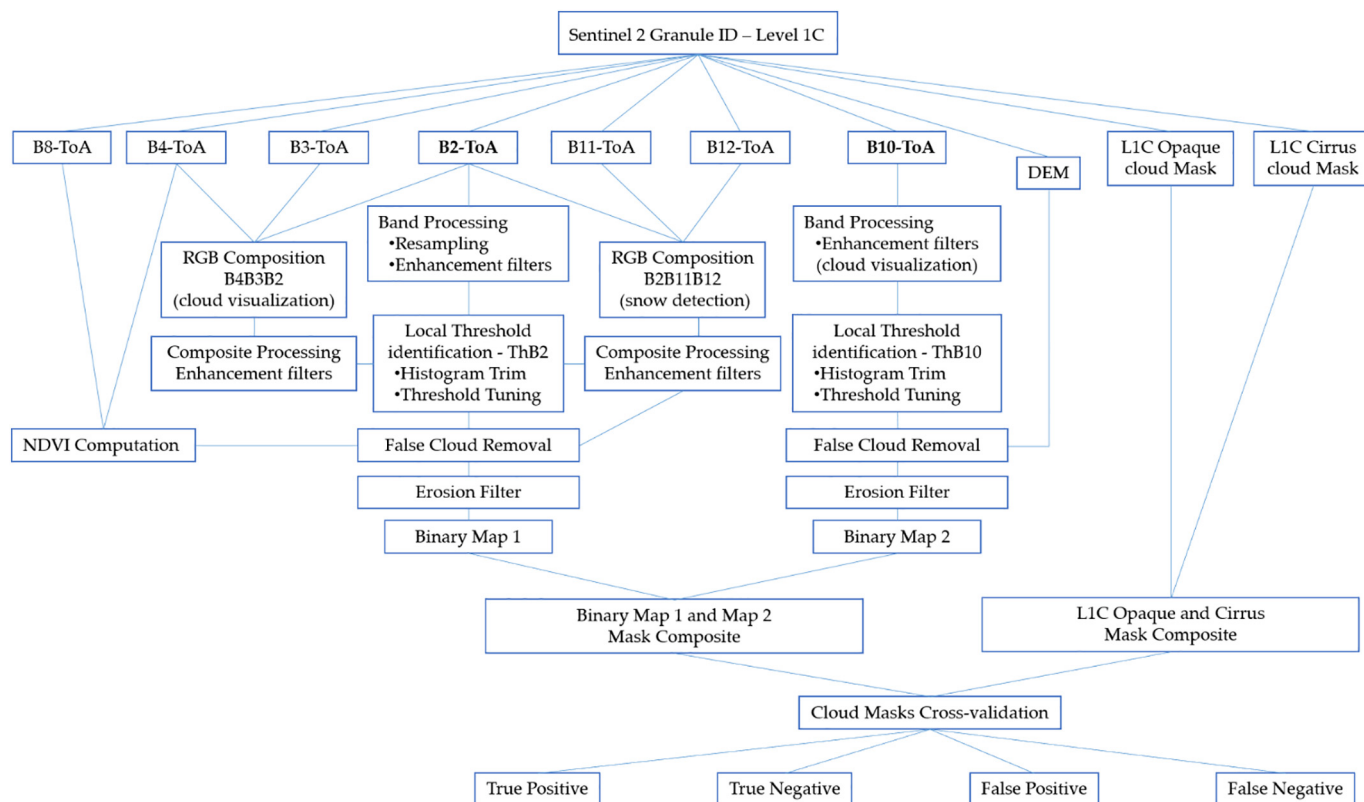
Fig. 4. Scheme of the procedure adopted to evaluate the cloud mask of Sentinel-2 L1C product.

generation of reliable cloud reference masks, more than on cloud classification (requiring further data and analyses (Hollstein et al., 2016; Zhu et al., 2015)), the two generated masks were jointed in one bitmap (cloud/no-cloud). To compare the elaborated masks with Sentinel-2 products, also the L1C masks were reclassified into a binary map.

Procedures applied to generate the reference masks are synthesized in Fig. 4.

3.2. Accuracy

According to a widely adopted practice (Foody, 2002; Liu et al., 2007), we evaluated the accuracy of the L1C cloud masks for each test-site by overlaying them with the reference masks, thereby obtaining validation maps and the corresponding confusion matrices (error matrices).

The validation map clearly emphasizes the areas of agreement and disagreement. In particular, we classified four categories: 1) True Positive (TP), i.e., agreements between cloudy pixels 2) True Negative (TN), i.e., agreements between no-cloudy pixels 3) False Positive (FP), i.e., disagreements No Clouds/Clouds (no-cloudy pixels in the reference mask, cloudy pixels in the L1C product) 4) False Negative (FN), i.e., disagreements Clouds/No Clouds (cloudy pixels in the reference mask, no-cloudy pixels in the L1C product).

From the confusion matrix, we estimated the overall accuracy, the omission and commission errors and the Cohen's Kappa coefficient.

The overall accuracy (A) indicates the ratio between the number of pixels correctly classified and the total number of pixels. It represents a basic estimation of similarity between the two masks and ranges from 0 (absence of agreement) to 100 (total agreement).

The user's accuracy (UA) is the ratio between TP pixels and the sum of TP and FP pixels. It is directly related to the commission error (CE), i.e., CE(%) = 100% − UA(%), which represents the percentage of cloud pixels in the L1C cloud mask which are falsely classified

(overestimation). The producer's accuracy (PA) is the ratio between TP pixels and the sum of TP and FN pixels. It is directly related to the omission error (OE), i.e., OE(%) = 100% − PA(%), which describes the percentage of the cloud pixels in the reference mask that have been omitted in the L1C cloud mask (underestimation).

The kappa coefficient (K) (Congalton, 1991) provides an additional statistical measure of the agreement between the two cloud masks. K includes off-diagonal elements also by taking into account the commission and omission errors. Therefore, by including also information on these errors, it represents a more realistic and reliable indication about the probability that a pixel is correctly labeled. K varies from 0 to 1; the higher the values of the kappa coefficient, the higher the agreement between the reference and L1C cloud mask.

3.3. Measure of cloud fragmentation

To measure the degree of cloud fragmentation, we adopted the effective mesh size index (m_{eff}). It is borrowed from landscape ecology concepts (Ingegnoli, 2003), i.e., the study of the pattern and interaction among ecosystems within a region of interest (Clark, 2010) and is generally applied to quantify fragmentation processes related to landscape configuration. The effective mesh size is based on the probability that two points randomly chosen within the examined area are connected, i.e., are located in the same patch (Jaeger et al., 2007). This index is widely adopted as a fragmentation measure in different contexts (natural and anthropogenic environments) and scales (from local to continental) see, e.g., (European Environment Agency and FOEN, 2011; Redon et al., 2014; Schmiedel and Culmsee, 2016).

The more fragmented the clouds, the lower the probability that the two randomly chosen cloudy pixels will be connected, and the lower the effective mesh size:

$$m_{eff} = \frac{1}{A_{total}}(A_1^2 + A_2^2 + ... + A_i^2 + ... + A_n^2)$$
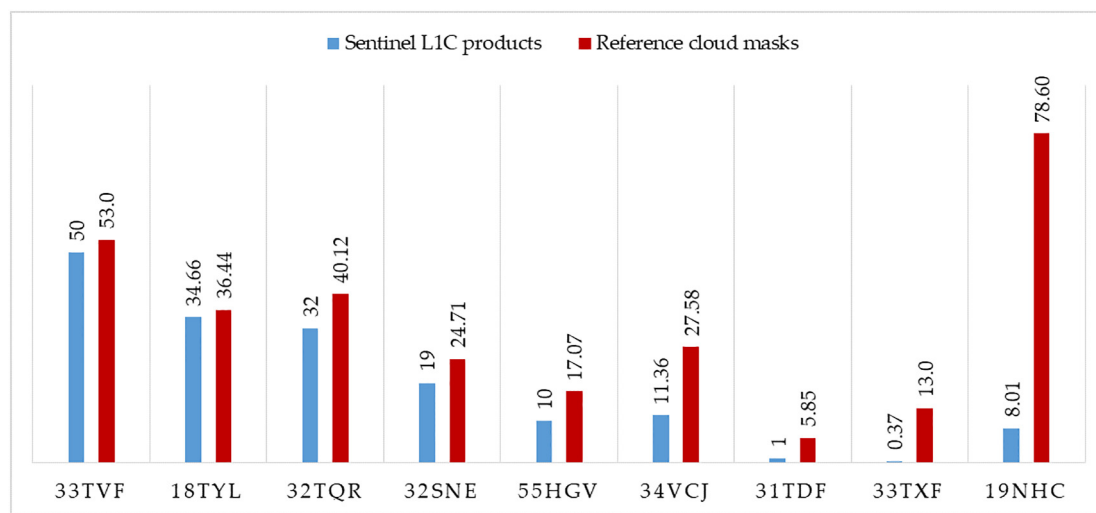
(1)

Fig. 5. Percentage of cloudy pixels in the Sentinel-2 L1C cloud product and the reference mask for the investigated granules (see Table 1).

where $A_1, ..., A_n$ represent the patch sizes of a cloud numbered from 1 to n, and A_{total} is the total area of the investigated region. The minimum m_{eff} value is the ratio between the pixel and tile dimensions (only one pixel is cloudy, absence of connection); whereas the upper limit is represented by the tile dimension (the image is totally covered by clouds, all the pixels are connected).

3.4. Utilized software platforms

Sentinel-2 data are analyzed using the Sentinel Application Platform (SNAP) v4.0 (available at http://step.esa.int/main/download/). Accuracy measures and complementary elaborations were implemented in QGIS 2.18.3 (QGIS Development Team, 2017) and Fragstats 4.2 free softwares (McGarigal et al., 2012).

4. Results and discussion

4.1. General performance of Level-1C cloud masks

A first comparison between the reference masks and the Sentinel-2 L1C cloud products shows a high variability of detected cloud coverage among the images (Fig. 5), hereafter identified by the UTM zone reported in Table 1. According to L1C products, the cloud cover percentage for the selected scenes ranges from 1.0% (granule 31TDF) to 50.0% (granule 33TVF) whereas for the reference masks it is comprised in the range 5.8% (granule 31TDF) - 78.6% (granule 19NHC).

The percentage of cloud cover provided by L1C is lower than that detected by reference masks in all the cases included in our sample, i.e., the Sentinel-2 mask underdetects the cloud coverage systematically. The highest difference (> 70%) is obtained for the granule 19NHC. In this image, corresponding to a portion of the Amazon rainforest, a high percentage of the contaminated pixels is occupied by thin clouds. As for the other tiles, this difference varies from 1,8% (granule 18TYL) up to 16,2% (granule 34VCJ).

The spatial distribution of misdetected pixels is shown in Fig. 6, whose rows report the natural color composite B4B3B2, the Band 10 image, and the validation map classified as indicated in Section 3.2 for each test-site. The validation maps (last column in Fig. 6) clearly show the location of over- (green pixels) and under-detected (blue pixels) cloud covers.

The image sequences highlight the relevance of multi-band analyses; for example, the test-site 32SNE appears almost clear in the RGB

composition (Fig. 6a) whereas the inspection of B10 (Fig. 6a′) shows several cirrus clouds spread across a large part of the image. Conversely, in 33TVF (Fig. 6c-c′) and 19NHC (Fig. 6i-i′), several cirrus clouds are present in the respective B10 images. No apparent opaque clouds can be noticed in this band whereas these can be easily observed in VIS composites: a large dense stratus in the center-right part of the image for 33TVF and a diffused number of small white patches of a spotted cumulus in the 19NHC image.

Moreover, in such images some product anomalies, already described in the Sentinel-2 Data Quality (Clerc et al., 2017a), can be noticed:

- the along-track striping due to the observation parallax effect between odd and even detectors, noticeable both in the natural composite (mainly over sea) and in B10 (Fig. 6b, e, and Fig. 6d′, e′, h′);
- the across-track intra-detector noise pattern induced by the compression noise on blind pixels used for dark signal correction (visible mainly in Fig. 6d′ and h′).

As referred in the more recent Data Quality Report (Clerc et al., 2017a, 2017b), a solution to filter out this noise has been identified and its operational implementation is currently under study.

4.2. Performance dependency on environmental conditions

The performance of the L1C cloud mask algorithm, used as a binary cloudy/non-cloudy classifier, was evaluated through the analysis of accuracy metrics (Table 2) obtained from confusion matrices. All the metrics identify the Amazonian rainforest 19NHC granule as the worst case (A = 29.44% and K = 0.05). Such a granule appears to be an outlier in the statistics of our test sample since it holds a large number of opaque clouds, which are sparse and different in size, jointly with diffused cirrus clouds, which appear largely underestimated in the Sentinel-2 mask (Fig. 6i-i″). Complexity and variability in cloud types jointly with a high content of water vapor (mean TCWV ~90 kg/m^2) seem to mislead the L1C detection in rainforest environments (Kovalskyy and Roy, 2015; Verhegghen et al., 2016).

For the other sites the overall accuracy (A) ranges from 84.25% to 97.44%. The lowest value, corresponding to a north-European granule (34VCJ, Fig. 6g″), is due to the massive presence of ragged and wispy clouds that are largely classified as free by the L1C product. The best overall accuracy is instead associated to a granule of Mediterranean environment (granule 33TXF, Fig. 6e″), combining portions of opaque and cirrus clouds. As it is evident in the validation map, for such a
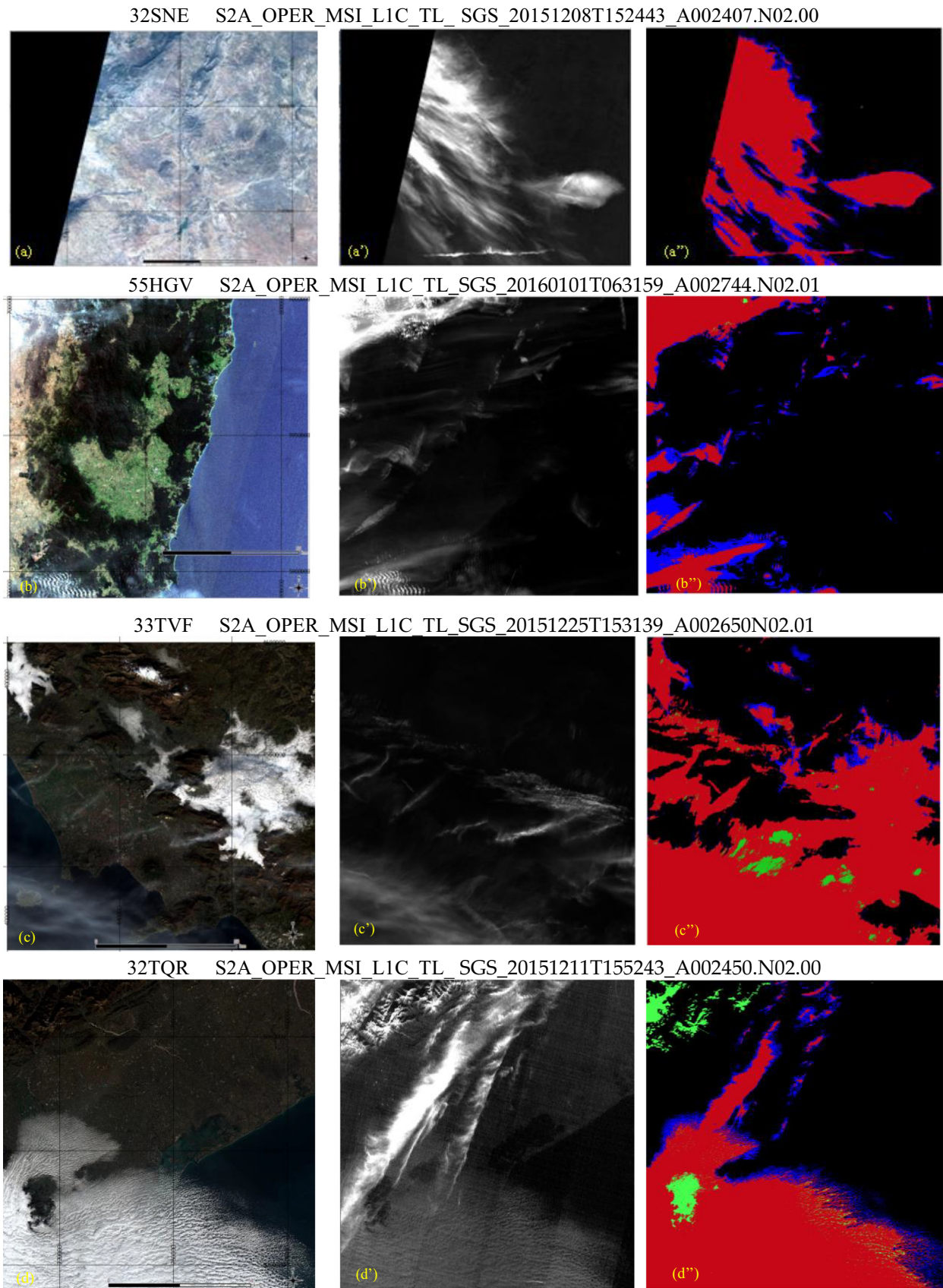
32SNE   S2A_OPER_MSI_L1C_TL_ SGS_20151208T152443_A002407.N02.00

55HGV   S2A_OPER_MSI_L1C_TL_SGS_20160101T063159_A002744.N02.01

33TVF   S2A_OPER_MSI_L1C_TL_SGS_20151225T153139_A002650N02.01

32TQR   S2A_OPER_MSI_L1C_TL_SGS_20151211T155243_A002450.N02.00

Fig. 6. Images of the investigated test-sites: natural color composite B4B3B2 (left panels), Band 10 (middle panels), and the validation maps (right panels). Natural composite and B10 show up dense and cirrus clouds. The validation map highlights pixels correctly classified by L1C cloud mask (red), undetected clouds (blue), and overestimated clouds (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
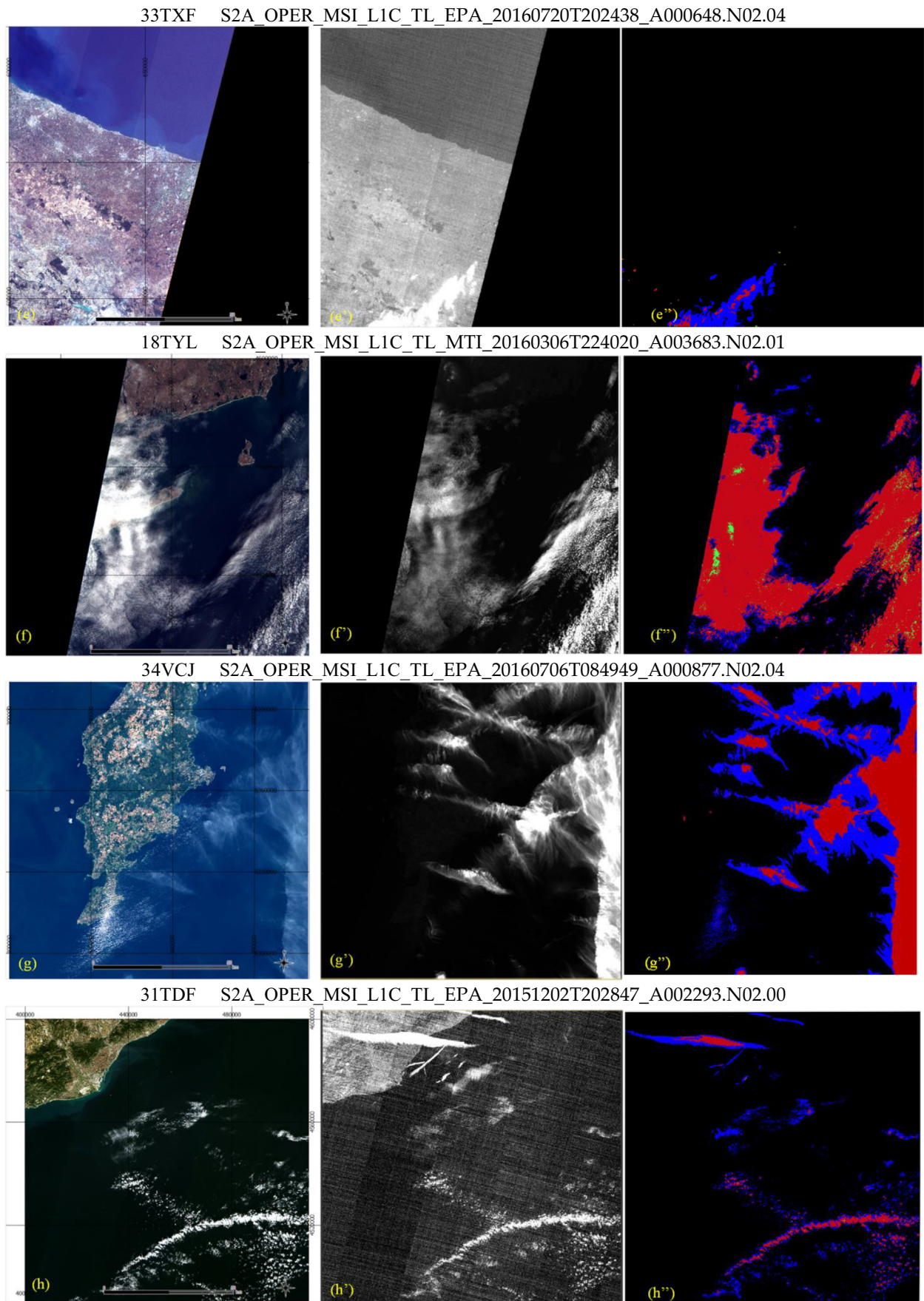
33TXF   S2A_OPER_MSI_L1C_TL_EPA_20160720T202438_A000648.N02.04

18TYL   S2A_OPER_MSI_L1C_TL_MTI_20160306T224020_A003683.N02.01

34VCJ   S2A_OPER_MSI_L1C_TL_EPA_20160706T084949_A000877.N02.04

31TDF   S2A_OPER_MSI_L1C_TL_EPA_20151202T202847_A002293.N02.00

Fig. 6.   (*continued*)

19NHC   S2A_OPER_MSI_L1C_TL_SGS_20160217T201717_A003425.N02.01

Color	Legend
Black	True Negative - Agreement between No-Clouds pixels
Red	True Positive - Agreement between Cloud pixels
Blue	False Negative - Disagreements Clouds/No Clouds
Green	False Positive - Disagreements No Clouds/Clouds
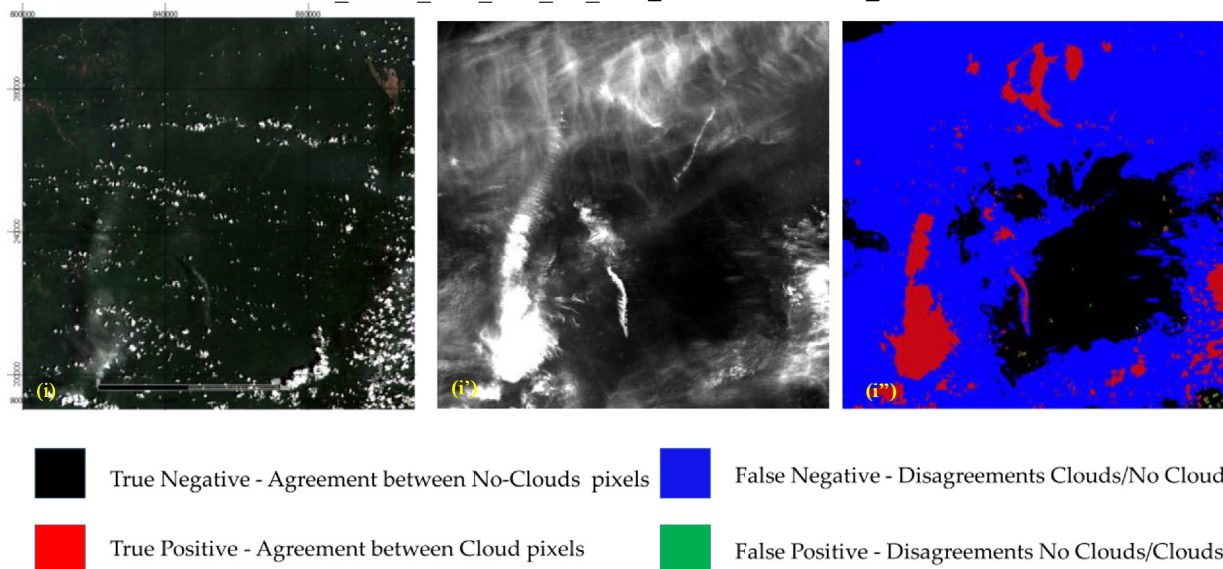
Fig. 6. (*continued*)

granule the cloudy area is relatively small and the good overall accuracy comes from the correct classification of non-cloudy pixels. If we consider just the fraction of correctly classified clouds (red), this is largely smaller than the undetected cloud portion (blue) and determines very high omission errors (87%) and low K (0.22).

According to the K values, the best L1C performance is obtained for a granule of the central Italian Apennines (33TVF, Fig. 6c″) where thin cirrus and large dense clouds coexist.

Only four of the nine test-sites show K values higher than 0.80; we found the lower the K values, the higher the omission errors ($R^2 = 0.95$), which range from about 4% (granule 33TVF, Fig. 6c″) to about 90% (granule 19NHC, Fig. 6i″). Conversely, commission errors have no significant influence on K coefficient ($R^2 = 0.04$ for linearity). Therefore, the performance of the L1C cloud mask is mostly determined by undetected clouds.

To evaluate potential dependency of the L1C underestimation on the surface background, we evaluated the omission errors over water and land (U-water and U-land in Table 2) by computing them under the land/water mask described in Section 2.2. The values estimated for the different tiles are significantly variable, however the values averaged over the tiles containing both the backgrounds are quite comparable (34,22% for water and 42.56% for land). Moreover, on the basis of the CCI-LC map, for each tile, we evaluated the distribution of land covers to determine the prevalent class in the background of misdetected clouds. No specific land cover appeared to dominate in the dataset, as a likely cause of misdetection, not even bright surfaces. We did not find

those misclassification errors due to white buildings reported in some S2 data quality reports (Clerc et al., 2015). Similarly, no relationship was found for illumination (SZA solar zenith angle) and observation geometry (SVA satellite view angle).

Among the test sites significantly affected by cloud overestimation, the 32TQR granule (Fig. 6d-d″) represents the most relevant case. This includes very heterogeneous land covers located within a wide elevation range (from the sea level to above 1800 m a.s.l.). For such a tile, the L1C mask shows the highest cloud overestimation (CE 8.5%) due to the appearance of false cirrus clouds located in the upper-left side of the granule in correspondence of mountain areas. The B2B11B12 color composite was used to exclude the presence of snow on the Italian Alps where snowfalls are frequent in the acquisition season (sensing date 11 December 2015); a DEM of the area showed that the false cloud formations were located at an elevation higher than 1000 m and their shape strictly followed orographic patterns. In such a case, the L1C overestimation was induced by high TOA reflectance values of surface targets in Band 10 due to particularly dry atmospheric conditions (TCWV between 2.5 and 3.5 kg/m^2) and to the consequent low absorption of reflected radiation.

4.3. Performance dependency on cloud type and structure

According to the results illustrated above, the L1C cloud mask product principally underestimates the actual presence of clouds within the selected scenes. A detailed visual assessment of the validation maps

Table 2
Accuracy metrics of L1C cloud mask products. A overall accuracy, CE commission errors (overestimation), OE omission errors (underestimation), K Kappa coefficient; U-Water and U-Land percentage of undetected clouds over sea and over land per total reference cloudy pixels over water and over land, respectively; m_{eff} effective mesh size index; OE$_{buffer}$ omission error for cloud masks with an added buffer (Section 4.4).

| UTM zone	A(%)	CE(%)	OE(%)	K	U-Water (%)	U-Land (%)	m_{eff} (ha)	OE(%)$_{buffer}$
33TVF	96.15	3.02	4.18	0.92	1.57	3.68	166,829	0.27
18TYL	94.54	2.98	12.28	0.88	10.44	47.22	114,575	0.31
32TQR	92.52	8.58	12.34	0.84	6.66	18.72	153,471	0.51
32SNE	94.64	0.04	18.98	0.86	–	18.98	90,440	0.00
55HGV	93.9	0.96	35.08	0.75	61.54	34.89	8453	3.59
34VCJ	84.25	0.5	56.88	0.52	52.43	81.64	68,481	9.84
31TDF	95.75	0.52	72.51	0.42	72.66	69.2	300	24.86
33TXF	97.44	4.82	87.00	0.22	–	87	833	0.00
19NHC	29.44	0.73	89.69	0.05	–	89.69	732,814	11.34
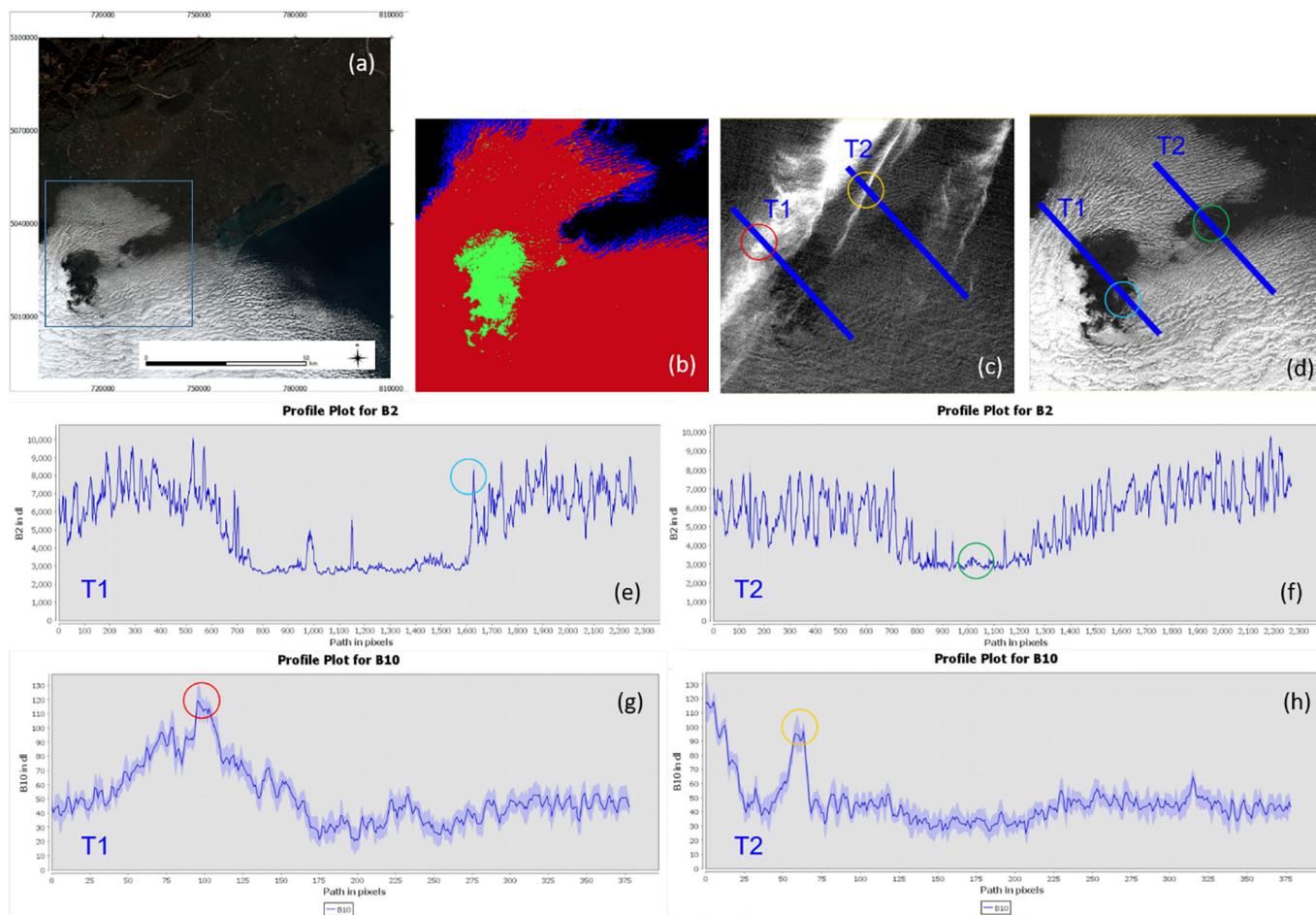
Fig. 7. Example of overdetected cloudy pixels in a hole inside the main cloud body. Natural color composite of the 32TQR granule (a), detail of validation map (b), transects on B10 (c), and B2 (d) and relative profiles for T1 (e, f) and T2 (g, h).

reveals that the majority of the undetected pixels are located around the core of the clouds, in the transition zone (buffer) between thick cloud pixels (completely cloudy) and pixels under clear atmosphere (completely cloud-free). This is the case of 6 of the 9 test-sites: 32SNE (Fig. 6a″), 55HGV (Fig. 6b″), 33TVF (Fig. 6c″), 32TQR (Fig. 6d″), 18TYL (Fig. 6f″), and 34VCJ granules (Fig. 6g″). The visual examination also shows that the lowest values of K (highest omission errors) mainly correspond to scenes with a large number of small and patched clouds, such as in 31TDF (Fig. 6h-h″) with K = 0.42 and 19NHC (Fig. 6i-i″) with K = 0.05.

The analysis of the fragmentation index (Table 2) shows that omission errors are inversely related to the cloud mesh size (R^2 = 0.73, standard deviation = 20); low accuracy values are associated to low values of the m$_{eff}$ index, i.e., to the presence of highly unconnected cloudy pixels. The complexity of error sources for the granule 19NHC is further confirmed since both m$_{eff}$ and OE are the highest. As for the other tiles, underdetection is likely due to the many small, generally thin, and spectrally fuzzy fragments usually observed in the transition zone between the cloud core and the cloud-free area. The OE therefore is low when the cloud core is large whereas the value of m$_{eff}$ is high, as it reveals the increase in connectivity due to the presence of such a core.

The differences between the L1C and the reference cloud masks are confirmed to be strongly dependent on cloud fragmentation (R^2 = 0.76, standard deviation = 17): the lower the clouds fragmentation, the better the L1C performance.

If we exclude the above-mentioned case due to dry atmosphere

conditions, also overestimation is related to cloud fragmentation for the presence of holes/gaps inside the main cloud body (see granules 33TVF, 18TYL, 32TQR). The profile of the transects in Fig. 7 clearly shows that the values of B2 and B10 recorded for pixels inside the hole (overdetected) are similar to those bordering the open hole, which is instead correctly detected as cloud-free. The configuration of such a false detection seems to suggest that the dilation filter adopted into the L1C processing chain overfills the observed gap.

For land surface applications, such a false detection is few significant because it limits the number of valid pixels useful for the analysis but does not alter the results obtained for cloud-free areas. Differently, underestimation can severely affect land surface analyses. To assure highly confident cloud-free pixels, also cloud buffers/edges should be labeled as cloud since they have altered reflectance values with respect to those of the neighboring pixels not contaminated by clouds. The examples in Fig. 8 clearly demonstrate the presence of undetected cirrus, plainly evident in the B10 profiles, and its impact on VNIR bands both over sea and over land. The alterations in TOA reflectance values due to cirrus clouds are particularly visible in B8 (Fig. 8m) and B4 (Fig. 8n) over sea.

Fig. 9 shows the spectral signatures of bright bare soil and vegetation in actual cloud-free areas and in areas covered by undetected clouds. The high incidence of undetected clouds found in our analyses is particularly relevant considering the efforts of the ESA Sentinel-2 Core Segments to minimize underdetections (Clerc et al., 2015). Omitted cirrus pixels in L1C masks can alter the estimation of land surface
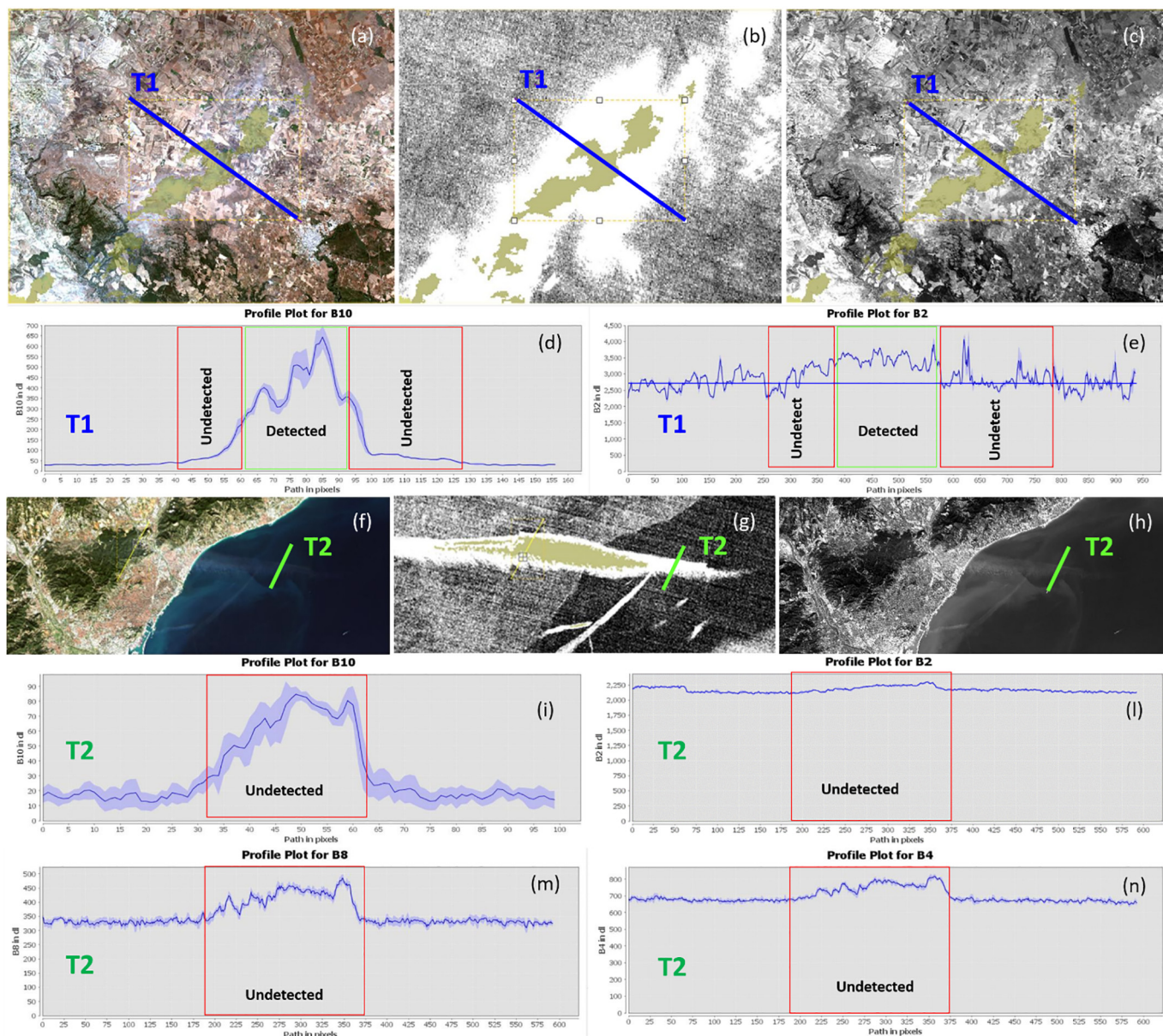
Fig. 8. Effects of cirrus on VNIR reflectances. Transects over land (T1) and sea (T2) are drawn on the following images: natural color composite (B4B3B2) (a, f), B10 (b, g), and B2 (c, h) images of the 33TXF and 31TDF granules. Reflectance profiles of T1 and T2 are shown in: (d, i) B10, (e, l) B2, (n) B4, and (m) B8. As evident in B10 image, L1C cloud mask (greenish brown) detected only the cirrus core corresponding to higher values (green box) in B10 profile (d). In the same transect portion also B2 values are high (e). For the portion of undetected cirrus (red box), B2 values are lower than the core values, but anyway they are higher than those of the cloud-free areas. Similar effects can be seen for the transect over sea. The impact of the cirrus is even higher in B4 and B8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

parameters, such as NDVI, chlorophyll *a* (Cha), dissolved organic matter (CDOM), NDWI (Normalized Difference Water Index), providing misleading results (Gao and Li, 2017; Du et al., 2016; Toming et al., 2016; Gao, 2000).

4.4. Suggestions/recommendations for the remote sensing community

This section provides some suggestions for users interested in the analysis of surface parameters, summarizing recommendations from our work and ESA quality assessments (Clerc et al., 2015, 2017a,

2017b; Gascon et al., 2017). They can be considered valid until corrections in the cloud mask processing chain will be implemented (at the current processing baselines, N02.06, no modifications involved the cloud mask processing steps (Clerc et al., 2018a)).

The Sentinel-2 Level 1C cloud mask product exhibits the worst performances in the following cases:

- in correspondence of snowy surfaces (wrongly identified as cirrus/opaque clouds);
- in correspondence of bright surfaces such as sand and white buildings (sometimes identified as opaque clouds);
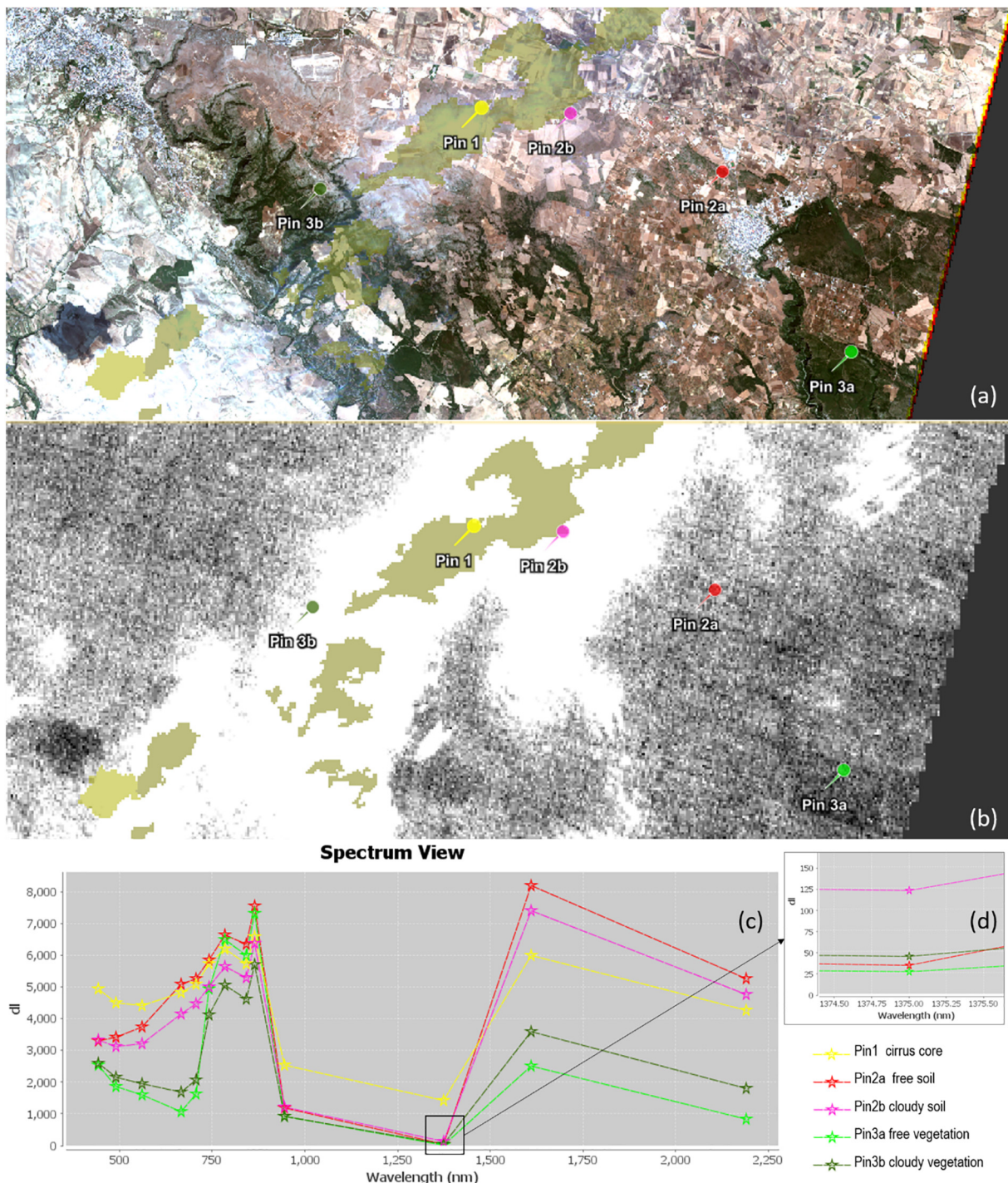
Fig. 9. Example of spectral signatures for cloud-free and undetected cloudy surfaces belonging to the same land cover types: bare soil Pin2a and Pin2b, and vegetation Pin 3a and Pin 3b. Pins over natural color composite (B4B3B2) (a) and Band 10 (b) of 33TXF granule (L1C cloud mask in greenish brown); Top Of Atmosphere (TOA) reflectance for the considered pins (c); magnified image in correspondence of Band 10 values (d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- over high elevation surfaces with particular dry atmospheric condition (sometimes confused as cirrus clouds);

- in correspondence of cloud boundaries/buffers and cirrus clouds.

To overpass these issues we suggest:

- to detect the presence of snow into the images by using, for example, the B2B11B12 as RGB composition;

- to exclude the presence in the cloud mask of highly reflective pixels in the VIS by using the NDVI and, if some doubtful bright patterns are retained, to observe the persistence of such patterns in

multitemporal images acquired around the sensing date of the considered granule;

- to verify the presence of false cirrus clouds in correspondence of high altitude lands by checking values of elevation and the water vapor content (Sentinel-2 product includes also the DEM and the TCWV of the area); in these cases DEM can be used to look at surface morphology because generally the L1C clouds follows orographic patterns, therefore false clouds can be easily removed manually or by applying a spatial filter that compares the shapes of potential clouds with the contourlines

extracted from the DEM;

- to label cloud buffers as cloudy because of the high probability of misclassification of these areas.

In case of large data sets, an automatic approach, such as the F-mask (Zhu et al., 2015) or the multitemporal cloud detection (MTCD) method (Hagolle et al., 2010), would be more helpful. If the application is possible, the simplified hybrid method based on dark pixel and MVC NDVI (Ramoino et al., 2017) requires a lower computation time. Anyway, a prudential buffer on opaque L1C cloud mask and/or a fixed threshold on B10 could limit misleading interpretations until a new operational implementation of Sentinel-2 Level 1C cloud mask product will be available.

The threshold on B10 could be fixed by selecting the TOA reflectance value that better discriminates cloud pixels from no-cloud pixels by simply trimming the B10 distributions. The extension of the buffer for a single image could be estimated by overlaying the L1C opaque cloud mask with the B4B3B2 RGB composition (evaluable in SNAP or other GIS environments). The cloudy area in the RGB composition which exceeds the L1C mask provides a reasonable visualization of the undetected margins; the maximum extent of these margins could be used to define the buffer width. This prudential masking strongly reduces the OE, possibly clearing the error, but also implies the removal of a large amount of cloud-free pixels.

A softer buffer masking could be implemented to improve the mask performance limiting the increase in CE. For our data set, a 1,5 km buffer (the median value of the maximum extents estimated for tile) and a threshold on B10 set to 35 dl (the minimum threshold over the database) allow a mean reduction in OE of about 90% (minimum reduction of 67%, see OE_{buffer} in Table 2) with an average CE = 20%. Since this strategy aims to obtain high quality surface data, a relatively high CE could be acceptable if this increase is associated to a net decrease of the number of underestimated pixels, which can severely affect land surface analyses.

5. Conclusions

This study provides a first assessment of the Sentinel-2 Level 1C cloud mask product in different biogeographic regions. The comparison with reference masks, elaborated by integrating visual interpretations of radiometric data and support information (NDVI, DEM), showed that

the L1C product is mostly working (average overall accuracy of 86.5%). The L1C algorithm showed comparable performances both over water and land, and our results do not point out any dependence of the error probability on the different land covers in the background.

Nevertheless, the performance of L1C cloud masks is limited by a non-negligible number of undetected clouds. The largest part of misclassified pixels was found on borders/buffers of cirrus and opaque clouds. Most omission errors (up to ~87%) were strictly dependent on the fragmentation level of the cloud structures, which was generally determined by the presence of many small, thin, and spectrally fuzzy fragments located in the transition zone between the cloud core and the cloud-free area. The commission errors (up to ~5%), markedly lower than the omission ones, were principally determined by the cloud configuration, particularly in the presence of holes inside the main cloud bodies. Altogether, cirrus clouds confirmed to be the most difficult to be correctly detected.

Two extreme environments in peculiar atmospheric conditions were particularly critical for a correct detection. Complex clouds over rainforests and in the presence of high atmospheric water vapor content determined the highest under-detection error (~89%); high mountain orography patterns in dry atmosphere determined the highest overdetection error (~8%).

In these two years, since the Sentinel-2 launch, great efforts were implemented by the ESA and the Sentinel-2 community to provide high level quality data, but further work is required to obtain consistent environmental and climate information and to fully exploit the great capability of the Sentinel-2 mission in operative chains.

In the meantime, we suggest carrying out preliminary analyses and prudential actions for minimizing effects of residual clouds in surface applications of Sentinel-2 Level 1-C data.

Acknowledgements

Appendix A. Sentinel-2 spectral response functions

Table A.1
Central wavelength, bandwidth, and spatial resolution for the 13 spectral bands of the Sentinel-2.

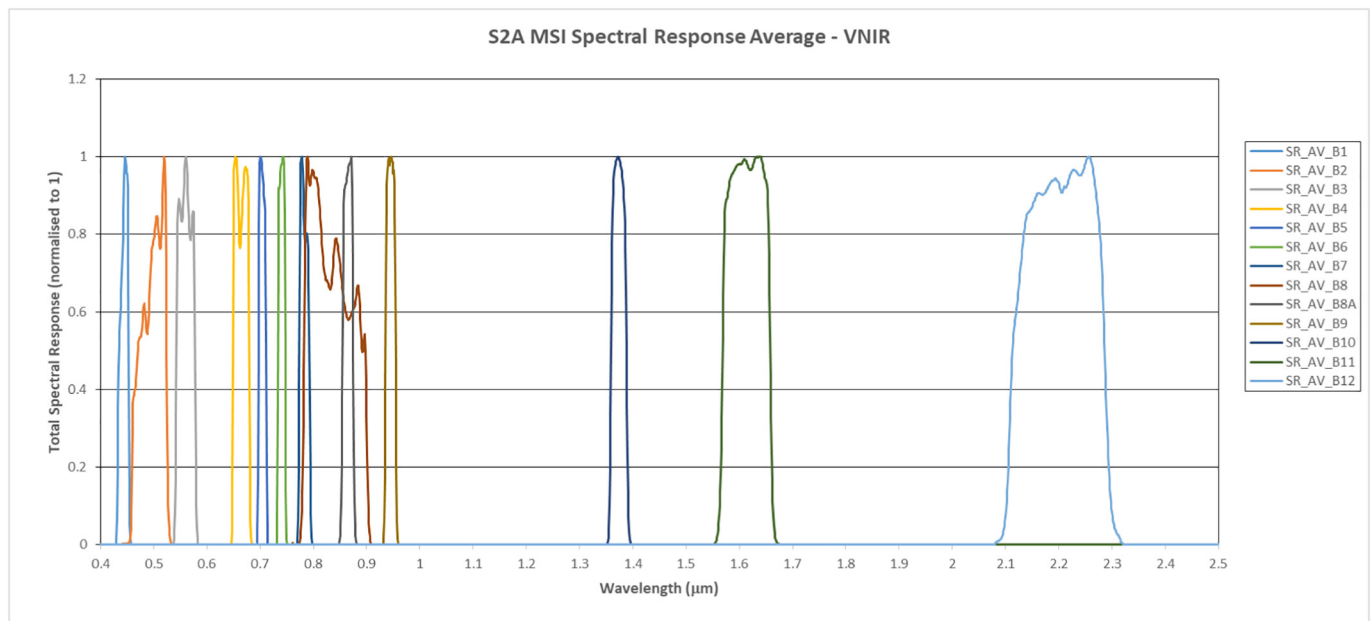| Band	Centre λ μm	Spectral width Δλ nm	Spatial resolution m
B1	0.443	20	60
B2	0.490	65	10
B3	0.560	35	10
B4	0.665	30	10
B5	0.705	15	20
B6	0.740	15	20
B7	0.783	20	20
B8	0.842	115	10
B8A	0.865	20	20
B9	0.945	20	60
B10	1.375	30	60
B11	1.610	90	20
B12	0.443	180	20

Fig. A1. Sentinel-2A Spectral Response Functions, available before 19/12/2017 and used by the Ground Segment until 15/01/2018. The y-axis values represent the average spectral response determined through ground-based testing of the Multi Spectral Instrument for each spectral band. Data are based on ground measurements performed in the frame of the Assembly, Integration and Test (AIT) campaign of the MSI of the ESA.
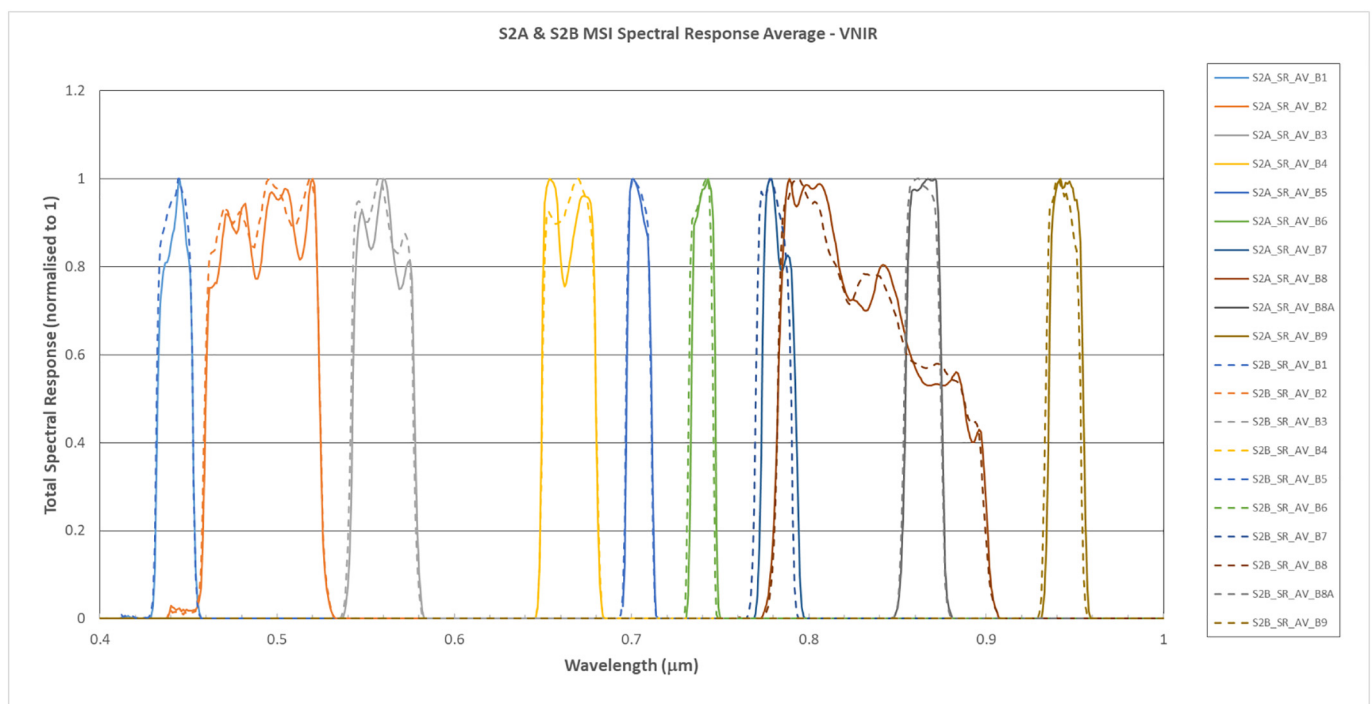

Fig. A2. Sentinel-2A and Sentinel-2B Spectral Response Functions for Visible and NIR bands, available after 19/12/2017 and used by the Ground Segment as from 15/01/2018. The y-axis values represent the average spectral response determined through ground-based testing of the Multi Spectral Instrument for each spectral band. Data are based on ground measurements performed in the frame of the Assembly, Integration and Test (AIT) campaign of the MSI of the ESA ((Clerc et al., 2018b); (https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses, last access: 24/01/2018).
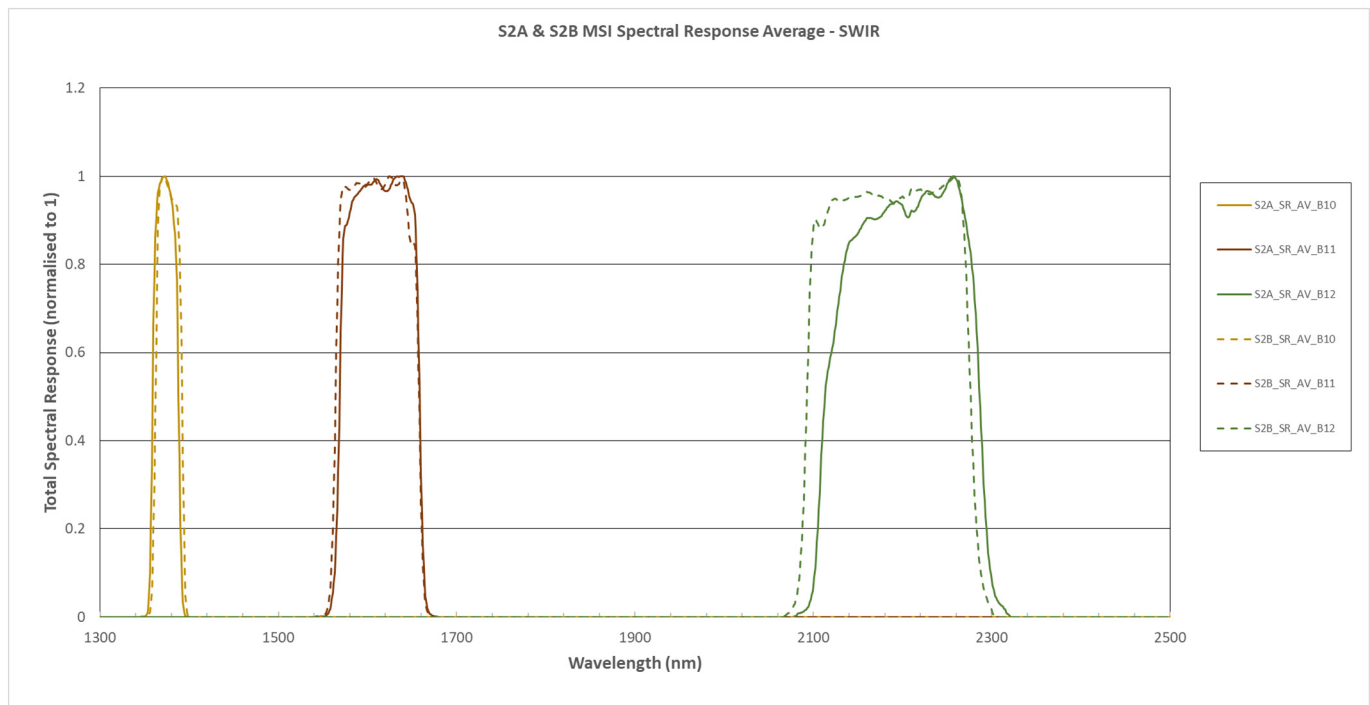
Fig. A3. Sentinel-2A and Sentinel-2B Spectral Response Functions for SWIR bands, available after 19/12/2017 and used by the Ground Segment as from 15/01/2018. The y-axis values represent the average spectral response determined through ground-based testing of the Multi Spectral Instrument for each spectral band. Data are based on ground measurements performed in the frame of the Assembly, Integration and Test (AIT) campaign of the MSI of the ESA ((Clerc et al., 2018b); https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses, last access: 24/01/2018).

References

Acharya, T.D., Yang, I., 2015. Exploring Landsat 8. Int. J. IT Eng. Appl. Sci. Res. 4 (4) (April 2015).

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's sentinel missions in support of Earth system science. Remote Sens. Environ. 120, 84–90. https://doi.org/10.1016/j.rse.2011.07.023.

Bley, S., Deneke, H., 2013. A threshold-based cloud mask for the high-resolution visible channel of Meteosat Second Generation SEVIRI. Atmos. Meas. Tech. 6, 2713–2723. https://doi.org/10.5194/amt-6-2713-2013.

Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Morin, D., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Defourny, P., Koetz, B., 2015. "Sentinel-2 for agriculture": supporting global agriculture monitoring. IEEE 4185–4188. https://doi.org/10.1109/IGARSS.2015.7326748.

Clark, W.R., 2010. Principles of landscape ecology. Nat. Educ. Knowl. 31034.

Clerc, S., Devignot, O., Pessiot, L., S2 MPC TEAM, 2015. S2 MPC Data Quality Report - 2015-11-30.

Clerc, S., Devignot, O., Pessiot, L., S2 MPC TEAM, 2017a. S2 MPC Data Quality Report - 2017-01-09.

Clerc, S., Devignot, O., Pessiot, L., S2 MPC TEAM, 2017b. S2 MPC Data Quality Report - 2017-04-06.

Clerc, S., Devignot, O., Pessiot, L., S2 MPC TEAM, 2018a. S2 MPC Data Quality Report - 2018-06-07.

Clerc, S., Devignot, O., Pessiot, L., S2 MPC TEAM, 2018b. S2 MPC Data Quality Report - 2018-08-01.

Coluzzi, R., di Donna, I., Lanorte, A., Lasaponara, R., 2007. Mapping forest fuel types by using satellite ASTER data and neural nets. In: Remote Sensing for Agriculture, Ecosystems, and Hydrology IX. Int. Soc. for Optics and Photonics 6742, 67420T. https://doi.org/10.1117/12.748033.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46. https://doi.org/10.1016/0034-4257(91)90048-B.

van der Werff, H., van der Meer, F., 2016. Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing. Remote Sens. 8, 883. https://doi.org/10.3390/rs8110883.

Dessler, A.E., Loeb, N.G., 2013. Impact of dataset choice on calculations of the short-term cloud feedback: calculation of the cloud feedback. J. Geophys. Res. Atmos. 118, 2821–2826. https://doi.org/10.1002/jgrd.50199.

Dörnhöfer, K., Göritz, A., Gege, P., Pflug, B., Oppelt, N., 2016. Water constituents and water depth retrieval from sentinel-2A—a first evaluation in an oligotrophic lake. Remote Sens. 8, 941. https://doi.org/10.3390/rs8110941.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from

sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. Remote Sens. 8, 354. https://doi.org/10.3390/rs8040354.

ESA CCI LC project, 2016. Land Cover CCI, Product User Guide - Version 2.5.

ESA Sentinel-2 Team, 2007. GMES Sentinel-2 Mission Requirements.

European Environment Agency, FOEN (Eds.), 2011. Landscape Fragmentation in Europe: Joint EEA-FOEN Report, EEA Report. Europ. Environment Agency [u.a.], Copenhagen.

European Space Agency (ESA), 2015. Sentinel-2 User Handbook.

Fernández-Manso, A., Fernández-Manso, O., Quintano, C., 2016. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. Int. J. Appl. Earth Obs. Geoinf. 50, 170–175. https://doi.org/10.1016/j.jag.2016.03.005.

Fletcher, K., European Space Agency (Eds.), 2012. Sentinel-2: ESA's Optical High-resolution Mission for GMES Operational Services, ESA SP. ESA Communications, Noordwijk.

Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Joseph Hughes, M., Laue, B., 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sens. Environ. 194, 379–390. https://doi.org/10.1016/j.rse.2017.03.026.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4.

Gao, B., 2000. Quantitative improvement in the estimates of NDVI values from remotely sensed data by correcting thin cirrus scattering effects. Remote Sens. Environ. 74, 494–502. https://doi.org/10.1016/S0034-4257(00)00141-3.

Gao, B.-C., Li, R.-R., 2017. Removal of thin cirrus scattering effects in Landsat 8 OLI images using the cirrus detecting channel. Remote Sens. 9, 834. https://doi.org/10.3390/rs9080834.

Gao, B.-C., Kaufman, Y.J., Han, W., Wiscombe, W.J., 1998. Correction of thin cirrus path radiances in the 0.4–1.0 μm spectral region using the sensitive 1.375 μm cirrus detecting channel. J. Geophys. Res. Atmos. 103, 32169–32176. https://doi.org/10.1029/98JD02006.

Gao, B., Yang, P., Han, W., Li, R., Wiscombe, W.J., 2002. An algorithm using visible and 1.38-μm channels to retrieve cirrus cloud reflectances from aircraft and satellite data. IEEE Trans. Geosci. Remote Sens. 40, 1659–1668. https://doi.org/10.1109/TGRS.2002.802454.

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus sentinel-2A calibration and products validation status. Remote Sens. 9, 584. https://doi.org/10.3390/rs9060584.

Greco, S., Infusino, M., De Donato, C., Coluzzi, R., Imbrenda, V., Lanfredi, M., Simoniello,

T., Scalercio, S., 2018. Late spring frost in Mediterranean beech forests: extended crown dieback and short-term effects on moth communities. Forests 9, 388. https://doi.org/10.3390/f9070388.

Greenhough, J., Remedios, J.J., Sembhi, H., Kramer, L.J., 2005. Towards cloud detection and cloud frequency distributions from MIPAS infra-red observations. Adv. Space Res. 36, 800–806. https://doi.org/10.1016/j.asr.2005.04.096.

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. Remote Sens. Environ. 114, 1747–1755. https://doi.org/10.1016/j.rse.2010.03.002.

Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., 2016. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. Remote Sens. 8, 666. https://doi.org/10.3390/rs8080666.

Huang, H., Roy, D., Boschetti, L., Zhang, H., Yan, L., Kumar, S., Gomez-Dans, J., Li, J., 2016. Separability analysis of sentinel-2A multi-spectral instrument (MSI) data for burned area discrimination. Remote Sens. 8, 873. https://doi.org/10.3390/rs8100873.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Hughes, M.J., 2016. L8 SPARCS Cloud Validation Masks. https://doi.org/10.5066/F7FB5146.

Hughes, M., Hayes, D., 2014. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. Remote Sens. 6, 4907–4926. https://doi.org/10.3390/rs6064907.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. Remote Sens. 8, 166. https://doi.org/10.3390/rs8030166.

Ingegnoli, V., 2003. Landscape Ecology: A Widening Foundation. Springer Verlag, New York, Berlin, Heidelberg.

Jaeger, J., Schwarz-Von Raumer, H.-G., Esswein, H., Müller, M., Schmidt-Lüttmann, M., 2007. Time series of landscape fragmentation caused by transportation infrastructure and urban development: a case study from Baden-Württemberg, Germany. Ecol. Soc. 12. https://doi.org/10.5751/ES-01983-120122.

Jedlovec, G.J., Haines, S.L., LaFontaine, F.J., 2008. Spatial and temporal varying thresholds for cloud detection in GOES imagery. IEEE Trans. Geosci. Remote Sens. 46, 1705–1717. https://doi.org/10.1109/TGRS.2008.916208.

Kaufman, Y.J., 1987. The effect of subpixel clouds on remote sensing. Int. J. Remote Sens. 8, 839–857. https://doi.org/10.1080/01431168708948693.

Kazantzidis, A., Eleftheratos, K., Zerefos, C.S., 2011. Effects of cirrus cloudiness on solar irradiance in four spectral bands. Atmos. Res. 102, 452–459. https://doi.org/10.1016/j.atmosres.2011.09.015.

Kovalskyy, V., Roy, D., 2015. A one year Landsat 8 conterminous United States study of cirrus and non-cirrus clouds. Remote Sens. 7, 564–578. https://doi.org/10.3390/rs70100564.

Lanfredi, M., Coppola, R., Simoniello, T., Coluzzi, R., D'Emilio, M., Imbrenda, V., Macchiato, M., 2015. Early identification of land degradation hotspots in complex bio-geographic regions. Remote Sens. 7. https://doi.org/10.3390/rs70608154.

Lefebvre, A., Sannier, C., Corpetti, T., 2016. Monitoring urban areas with sentinel-2A data: application to the update of the Copernicus high resolution layer imperviousness degree. Remote Sens. 8, 606. https://doi.org/10.3390/rs8070606.

Li, Q., Lu, W., Yang, J., 2011. A hybrid thresholding algorithm for cloud detection on ground-based color images. J. Atmos. Ocean. Technol. 28, 1286–1296. https://doi.org/10.1175/JTECH-D-11-00009.1.

Liu, C., Frazier, P., Kumar, L., 2007. Comparative assessment of the measures of thematic classification accuracy. Remote Sens. Environ. 107, 606–616. https://doi.org/10.1016/j.rse.2006.10.010.

McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer Software Program Produced by the Authors at the University of Massachusetts.

Meyer, K., Yang, Ping, Gao, Bo-Cai, 2004. Optical thickness of tropical cirrus clouds derived from the MODIS 0.66 and 1.375-/spl mu/m channels. IEEE Trans. Geosci. Remote Sens. 42, 833–841. https://doi.org/10.1109/TGRS.2003.818939.

Mishra, K., Prasad, P.R.C., 2015. Automatic extraction of water bodies from Landsat imagery using perceptron model. J. Comput. Environ. Sci. 2015, 1–9. https://doi.org/10.1155/2015/903465.

Nakajima, T.Y., Tsuchiya, T., Ishida, H., Matsui, T.N., Shimoda, H., 2011. Cloud detection performance of spaceborne visible-to-infrared multispectral imagers. Appl. Opt. 50, 2601. https://doi.org/10.1364/AO.50.002601.

Paul, F., Winsvold, S., Kääb, A., Nagler, T., Schwaizer, G., 2016. Glacier remote sensing using Sentinel-2. Part II: mapping glacier extents and surface facies, and comparison to Landsat 8. Remote Sens. 8, 575. https://doi.org/10.3390/rs8070575.

Pesaresi, M., Corbane, C., Julea, A., Florczyk, A., Syrris, V., Soille, P., 2016. Assessment of the added-value of Sentinel-2 for detecting built-up areas. Remote Sens. 8, 299. https://doi.org/10.3390/rs8040299.

Pignatti, S., Acito, N., Amato, U., Casa, R., Castaldi, F., Coluzzi, R., ... Matteoli, S., 2015. Environmental products overview of the Italian hyperspectral prisma mission: The SAP4PRISMA project. In: In Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, pp. 3997–4000.

QGIS Development Team, 2017. QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Radoux, J., Chomé, G., Jacques, D., Waldner, F., Bellemans, N., Matton, N., Lamarche, C., d'Andrimont, R., Defourny, P., 2016. Sentinel-2's potential for sub-pixel landscape feature detection. Remote Sens. 8, 488. https://doi.org/10.3390/rs8060488.

Ramoino, F., Tutunaru, F., Pera, F., Arino, O., 2017. Ten-meter sentinel-2A cloud-free composite—Southern Africa 2016. Remote Sens. 9, 652. https://doi.org/10.3390/rs9070652.

Redon, M., Bergès, L., Cordonnier, T., Luque, S., 2014. Effects of increasing landscape heterogeneity on local plant species richness: how much is enough? Landsc. Ecol. 29, 773–787. https://doi.org/10.1007/s10980-014-0027-x.

Scaramuzza, P., Dwyer, J., 2016. L7 Irish Cloud Validation Masks. https://doi.org/10.5066/F7XD0ZWC.

Scaramuzza, P.L., Bouchard, M.A., Dwyer, J.L., 2012. Development of the Landsat data continuity mission cloud-cover assessment algorithms. IEEE Trans. Geosci. Remote Sens. 50, 1140–1154. https://doi.org/10.1109/TGRS.2011.2164087.

Scaramuzza, P., Foga, S., Dwyer, J., 2016. L8 Biome Cloud Validation Masks. https://doi.org/10.5066/F7251GDH.

Schmiedel, I., Culmsee, H., 2016. The influence of landscape fragmentation, expressed by the 'Effective Mesh Size Index', on regional patterns of vascular plant species richness in Lower Saxony, Germany. Landsc. Urban Plan. 153, 209–220. https://doi.org/10.1016/j.landurbplan.2016.01.012.

Simoniello, T., Cuomo, V., Lanfredi, M., Lasaponara, R., Macchiato, M., 2004. On the relevance of accurate correction and validation procedures in the analysis of AVHRR-NDVI time series for long-term monitoring. J. Geophys. Res. 109. https://doi.org/10.1029/2004JD004819.

Simoniello, T., Coluzzi, R., Imbrenda, V., Lanfredi, M., 2015. Land cover changes and forest landscape evolution (1985–2009) in a typical Mediterranean agroforestry system (High Agri Valley). Nat. Haz. Earth Syst. Sci. 15 (6), 1201–1214. https://doi.org/10.5194/nhess.

Song, X., Yang, C., Wu, M., Zhao, C., Yang, G., Hoffmann, W., Huang, W., 2017. Evaluation of sentinel-2A satellite imagery for mapping cotton root rot. Remote Sens. 9, 906. https://doi.org/10.3390/rs9090906.

Sun, L., Wei, J., Wang, J., Mi, X., Guo, Y., Lv, Y., Yang, Y., Gan, P., Zhou, X., Jia, C., Tian, X., 2016. A Universal Dynamic Threshold Cloud Detection Algorithm (UDTCDA) supported by a prior surface reflectance database: UDTCDA, Surface Reflectance Database. J. Geophys. Res. Atmos. 121, 7172–7196. https://doi.org/10.1002/2015JD024722.

Thales Alenia Space, 2016. Sentinel-2 Products Specification Document (PSD) - S2-PDGS-TAS-DI-PSD. Sentin.-2 Prod. Specif. Doc. PSD 489.

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T., 2016. First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. Remote Sens. 8, 640. https://doi.org/10.3390/rs8080640.

Verhegghen, A., Eva, H., Ceccherini, G., Achard, F., Gond, V., Gourlet-Fleury, S., Cerutti, P.O., 2016. The potential of sentinel satellites for burnt area mapping and monitoring in the Congo Basin forests. Remote Sens. 8, 986. https://doi.org/10.3390/rs8120986.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free access to Landsat imagery. Science 320https://doi.org/10.1126/science.320.5879.1011a. (1011a-1011a).

Xu, Meng, Jia, Xiuping, Pickering, M., 2014. Automatic cloud removal for Landsat 8 OLI images using cirrus band. IEEE 2511–2514. https://doi.org/10.1109/IGARSS.2014.6946983.

Yesou, H., Pottier, E., Mercier, G., Grizonnet, M., Haouet, S., Giros, A., Faivre, R., Huber, C., Michel, J., 2016. Synergy of Sentinel-1 and Sentinel-2 imagery for wetland monitoring information extraction from continuous flow of sentinel images applied to water bodies and vegetation mapping and monitoring. IEEE 162–165. https://doi.org/10.1109/IGARSS.2016.7729033.

Zhou, C., Zelinka, M.D., Dessler, A.E., Yang, P., 2013. An analysis of the short-term cloud feedback using MODIS data. J. Clim. 26, 4803–4815. https://doi.org/10.1175/JCLI-D-12-00547.1.

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sens. Environ. 159, 269–277. https://doi.org/10.1016/j.rse.2014.12.014.