Preliminary Tests and Results Concerning Integration of Sentinel-2 and Landsat-8 OLI for Crop Monitoring

Andrea Lessio *, Vanina Fissore and Enrico Borgogno-Mondino

Department of Agricultural, Forest and Food Sciences, University of Turin, Largo Braccini 2, 10095 Grugliasco (TO), Italy; vanina.fissore@unito.it (V.F.); enrico.borgogno@unito.it (E.B.-M.)

Received: 12 September 2017; Accepted: 3 November 2017; Published: 5 November 2017

Abstract: The Sentinel-2 data by European Space Agency were recently made available for free. Their technical features suggest synergies with Landsat-8 dataset by NASA (National Aeronautics and Space Administration), especially in the agriculture context were observations should be as dense as possible to give a rather complete description of macro-phenology of crops. In this work some preliminary results are presented concerning geometric and spectral consistency of the two compared datasets. Tests were performed specifically focusing on the agriculture-devoted part of Piemonte Region (NW Italy). Geometric consistencies of Sentinel-2 and Landsat-8 datasets were tested “absolutely” (in respect of a selected reference frame) and “relatively” (one in respect of the other) by selecting, respectively, 160 and 100 well distributed check points. Spectral differences affecting at-the-ground reflectance were tested after images calibration performed by dark object subtraction approach. A special focus was on differences affecting derivable NDVI and NDWI spectral indices, being the most widely used in the agriculture remote sensing application context. Results are encouraging and suggest that this approach can successfully enter the ordinary remote sensing-supported precision farming workflow.

Keywords: data integration; sensitivity analysis; Sentinel-2; Landsat-8 OLI

1. Introduction

In the context of Copernicus Programme for Earth Observation [1], on 7 March 2017 the second of two twin satellite systems composing the Sentinel-2 (S2) mission was launched. S2 system is intended to acquire mid resolution multispectral imagery by Multi-Spectral Instrument (MSI), specifically designed for vegetation/agriculture monitoring [2–5]. S2 mission was planned to ensure continuity with the SPOT (Satellites Pour l’Observation de la Terre) mission and, possibly, to be also jointly used with Landsat data. MSI design integrates experiences from the ESA MERIS (Medium Resolution Imaging Spectrometer) and NASA MODIS (MODerate-resolution Imaging Spectro-radiometer) instruments, providing more spectral bands than Landsat-8 Operational Land Imager (L8) sensor [6].

At the moment, geometric and spectral characteristics and free data access of S2 data suggest synergies with L8 OLI ones. Some studies testing their spectral consistency have already been done [7,8], but with no particular focus on agricultural applications based on simplified and easy-to-use data processing workflows intended for an at-farm-level operational context. Agronomic interventions, in fact, are expected to be better addressed in space and time when crop properties can be continuously (i.e., with a proper frequency of observation) mapped by remote sensing [9–11]. It is worth reminding that the majority of plants physiologic processes can be described and correctly interpreted only dynamically [12]. Spotty observations (even if regular in time) cannot be enough for these purposes; differently, a daily acquisition is desirable. Many experiences have demonstrated that continuity of acquisitions improves analysis and comprehension of biophysical processes involving vegetated...
surfaces and their phenology [13–16]. Unfortunately, medium/high geometric resolution satellite missions, if singularly considered, are not able to guarantee daily acquisitions. Moreover, along the year, the number of “good” acquisitions can be further reduced by cloud cover occurrence [17–21]. The possibility of integrating similar data from different missions, especially if it is available for free, is therefore desirable and remains the most economically convenient approach for crop monitoring by remote sensing, especially in agriculture where incomes are small. A combination of remotely sensed data along common time series from multiple source can be only achieved if their consistency is demonstrated, or a proper data integration strategy is adopted [22,23]. Many works in literature report experiences concerning data integration of S2 and L8 datasets for different issues like geological [24], forest, environmental applications [25,26], and urban sprawl monitoring [27]. In general, these works do not refer to the consistency of spectral and geometric features of datasets, giving it as a fact. Differently, this research was specifically intended to investigate and quantify geometric and spectral consistency between S2 and L8 datasets, with special regard to agricultural applications where positioning and spectral accuracy are mandatory to ensure effectiveness of deductions. While testing consistency between different datasets from remote sensing sensors, two aspects have to be considered: the geometrical and the spectral one. The first is required since datasets are, in general, supplied already orthoprojected with different declared accuracies and, possibly, in respect of different coordinate systems. The second, instead, concerns the spectral features of bands acquired by sensors, that, in general, do not perfectly fit the same spectral ranges, even if belonging to the same spectral region. This can determine a not negligible inhomogeneity between data, making unreliable any their integration. According to these general considerations, data consistency analysis can in general respond to different requirements, depending on the application data integration is intended for. Consequently, the conditions under which tests are achieved have to be precisely defined. In the operative framework of this research, data consistency was tested under the following conditions: (a) images cover a specific geographic area (Piemonte region, NW Italy); (b) information is required for crop monitoring purposes; and, (c) image processing workflow is a simplified one to be easily transferred to the operational agronomic context.

The first condition basically determines the selection of images making results not general, but specific for the study area; the second and third ones need to be more properly discussed. Since the analysis is intended to generate responses specifically aimed at the crop monitoring context, it can be limited to some aspects of the problem. In the precision farming context, from the geometric consistency point of view, not only the relative positioning between datasets is important, but also the absolute one (in respect of the assumed coordinate systems). This fact, better discussed in the Materials and Methods section, is related to the expected utilization of the maps from satellite on board of GNSS (Global Navigation Satellite System) equipped agriculture machineries at a variable rate.

From the spectral consistency point of view things are more complicated. Firstly, comparison concerns an indirect measure (at-the-ground reflectance), whose computation from the raw data depends on the selected radiative transfer model (RTM). A different model brings different results. A strictly forcing condition in the analysis is therefore the type of adopted RTM. A wide literature refers about RTM ranging from the so called physically-based ones (ATCOR3 [28,29], FLAASH, MODTRAN [30], etc.), up to the simplified empirical ones, like the dark object subtraction (DOS) method. The formers are, in general, more difficult to be managed in operational contexts requiring a lot of information defining atmosphere quality at the moment of the acquisition. Less scientifically smart, but reliable enough for many applications, is the DOS approach, which mostly relies on the intrinsic radiometric features of the same image that has to be calibrated (see Material and Methods section). DOS, therefore, better responds to the operative requirements that are imagined to be transferred into the ordinary agronomic workflow. In this work, presented results refer to images that were calibrated using a DOS RTM; the expectation is that every user is able to manage it by ordinary free software.
Moreover, since crop monitoring by remote sensing is mostly performed using time series of spectral indices, this work looked at the effects of image calibration onto Normalized Differencing Vegetation Index (NDVI) [31] and the Normalized Differencing Water Index (NDWI) [32]. NDVI and NDWI were selected as references, among a huge list of proposed spectral indices, since they are largely the most used ones for agriculture applications, being assumed as proxies of, respectively, vegetation vigour and water content. Spectral comparison is intended to point out if measure differences are consistent or not with the intrinsic uncertainty of the single dataset. From an operational point of view, this is mandatory to determine if inter-dataset spectral differences are significant, or not, in respect of the expected intra-dataset uncertainty. Some works can be found in literature that tries to define and map uncertainty of reflectances and spectral indices using a sensibility analysis concerning the adopted RTM [33,34].

2. Materials and Methods

2.1. Test Area

Test area for S2 and L8 data consistency analysis is represented by the Piemonte region (Northern West Italy), and, in particular, that part of the regional territory where crop fields are mostly present. Two different subsets of the regional area were selected to test, respectively, the geometric and spectral consistency of datasets, mostly in dependence of the compared image footprints. Geometric assessment was performed using images covering the whole Piemonte region area, while spectral comparison concerned a smaller area corresponding to a single tile of S2 dataset (110 km × 110 km) centered at 454864 E, 4945139 N (WGS-84 UTM 32N reference system). Footprints of test images are reported in Figure 1 for both the situations.

![Figure 1](image_url)

*Figure 1. Test areas used for geometrical (a) and spectral (b) validation. Footprint of S2 and L8 data are shown, respectively, in dark and light grey.*

2.2. Dataset

Two different image datasets were used to test geometric and spectral consistency between the images. As far as geometric assessment was concerned two L8 images, four S2 tiles and twelve very high resolution aerial orthoimages were used. Differently, five quasi-coeval pairs of L8 and S2 images were used for spectral comparison. All of the satellite images were obtained for free from the EarthExplorer distribution system [35]. L8 images were obtained as Level 1 products: orthoprojected (WGS-84 UTM 32N) and not radiometrically processed (pixel values are 12 bits digital numbers). S2 images were obtained as Level 1C products: orthoprojected (WGS-84 UTM 32N) and calibrated as reflectance at the top-of-atmosphere. Acquisition dates and orbit information of images are reported in Table 1; technical features (from official sensors handbooks [36,37]) are reported in Table 2. Aerial

...
orthoimages (hereafter called ICE), used as reference for absolute geometric assessment were obtained from the Piemonte Region Map Office for free. Metadata declare an horizontal accuracy of 1 m, therefore suitable for a 1:5000 map scale. Reference system is WGS-84 UTM 32N.

Table 1. List of images used for geometric and spectral assessments.

<table>
<thead>
<tr>
<th>Type of Assessment</th>
<th>Images</th>
<th>Date (DD/MM/YY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric</td>
<td>Sentinel-2 MSI</td>
<td>T32TLQ, T32TMQ, T32TLR, T32TMR</td>
</tr>
<tr>
<td></td>
<td>Landsat 8OLI</td>
<td>Path Row</td>
</tr>
<tr>
<td></td>
<td>ICE aerial</td>
<td>orthoimages</td>
</tr>
</tbody>
</table>

Table 2. Technical features of S2 and L8 datasets.

<table>
<thead>
<tr>
<th></th>
<th>Sentinel-2 MSI</th>
<th>Landsat 8OLI</th>
<th>ICE Aerial Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bands (µm)</td>
<td>Geometric Resolution (m)</td>
<td>Bands (µm)</td>
<td>Geometric Resolution (m)</td>
</tr>
<tr>
<td>B1: 0.433–0.453</td>
<td>60</td>
<td>B1: 0.435–0.451</td>
<td>30</td>
</tr>
<tr>
<td>B2: 0.458–0.523</td>
<td>10</td>
<td>B2: 0.452–0.512</td>
<td>30</td>
</tr>
<tr>
<td>B3: 0.543–0.578</td>
<td>10</td>
<td>B3: 0.533–0.590</td>
<td>30</td>
</tr>
<tr>
<td>B4: 0.650–0.680</td>
<td>10</td>
<td>B4: 0.636–0.673</td>
<td>30</td>
</tr>
<tr>
<td>B5: 0.698–0.713</td>
<td>20</td>
<td>B5: 0.851–0.879</td>
<td>30</td>
</tr>
<tr>
<td>B6: 0.733–0.748</td>
<td>20</td>
<td>B6: 1.566–1.651</td>
<td>30</td>
</tr>
<tr>
<td>B7: 0.773–0.793</td>
<td>20</td>
<td>B7: 2.107–2.294</td>
<td>30</td>
</tr>
<tr>
<td>B8: 0.785–0.900</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B8a: 0.855–0.875</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B9: 0.935–0.955</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B10: 1.360–1.390</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B11: 1.565–1.655</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B12: 2.100–2.280</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Radiometric resolution: 12 bit Radiometric resolution: 16 bit Radiometric resolution: 8 bit

Since height information was needed to operate topographic correction during radiometric image pre-processing (area is not completely included within the regional boundary), the global SRTM DEM (Shuttle Radar Topography Mission Digital Elevation Model [38]) was obtained for the area [39]. SRTM DEM (pixel size = 90 m) was oversampled by bilinear interpolation up to 10 and 30 m, to be consistent respectively with S2 and L8 imagery.
All of the datasets were maintained at their nominal geometric resolution to ensure that all of the measures concerned the original spectral and geometric properties of data. It is worth reminding that image resampling necessarily move both geometry and radiometry of original data introducing some further uncertainties that could make results less reliable. All compared measures from images were therefore extracted from the not-resampled datasets according to some reference ground points.

2.3. Geometric Assessment

Geometric quality evaluation concerned two aspects: (a) absolute positioning accuracy of S2 L1C data in respect of the UTM 32N WGS84 reference frame; and, (b) relative positioning between S2 and L8 datasets (image shift).

2.3.1. Absolute Positioning Test

S2 images were geometrically compared with the available ICE aerial orthoimages, which were assumed as a reference map. Check Points (CPAs) collection followed a hierarchical and driven approach. Since the focus of the analysis was on agricultural areas, CPAs were distributed as more homogeneously as possible over them. Farming areas identification was based on the Corine Land Cover map [40] 2012 (CLC2012). Regularly distributed macro areas were preventively identified according to CLC2012 and the correspondent ICE orthoimages selected for CPAs collection. CPAs were then collected as more homogeneously distributed as possible within each selected orthoimage. The absolute positioning of images in farming areas is a basic issue, whereas satellite-derived crops maps are intended to be interpreted and used by farming machines equipped with GNSS technology [41]. Statistics concerning the distribution of CPAs in respect of CLC2012 Level I classes and of territory steepness were finally calculated to demonstrate properness and representativeness of CPAs in respect of the territorial features of regional agricultural areas.

2.3.2. Relative Positioning Test

A second analysis was achieved comparing a single S2 tile (orbit: 65, ID: T32TMQ), acquired on 6/8/2015 with the correspondent L8 scene (Path: 195, Row: 29) acquired on 7/8/2015. A hundred (100) of check points (CPRs), different from the previous ones, were collected regularly over images, mainly within the regional farming areas. The spatial distribution of both CPAs and CPRs is shown in Figure 2.

Position accuracy (absolute and relative) was tested according to the following strategy:

(a) differences (ΔE, ΔN) between measured (tested) and expected (reference) East (E) and North (N) coordinates of CPAs/CPRs were computed; (b) mean (µ_{ΔE}, µ_{ΔN}) and standard deviation values (σ_{ΔE}, σ_{ΔN}) were computed separately for E and N differences. µ_{ΔE}, µ_{ΔN} were assumed as measure of bias, σ_{ΔE}, σ_{ΔN} as measure of precision. It is worth to point out that in this work bias is intended as that part of the relative distance between the two datasets that can be somehow modelled. Bias is assumed negligible if µ<σ. Local total error of the generic ith point is:

$$\Delta_i = \sqrt{\Delta E_i^2 + \Delta N_i^2}$$

(1)

Statistic distributions of Δ_i (cumulative frequency) for both absolute and relative positioning assessment were computed. (c) RMSE (Root Mean Squared Error) [42], was assumed as synthetic measure of accuracy:

$$RMSE = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \Delta_i^2}$$

(2)

where n is the number of CPAs/CPRs (respectively, 160 for the absolute and 100 for the relative positioning assessment) and Δ_i the local total error according to (1).
2.4. Spectral Assessment

Radiometric assessment was performed considering 5 couples of (quasi-) coeval S2 and L8 images (see Table 1). Images were preventively masked to remove cloudy pixel. “Cloudy” from “not-cloudy” pixels at the single date were separated using the “cirrus” band (band 10) and the quality assessment (QA) layer, respectively, for S2 and L8 images (Figure 3).

Figure 2. Spatial distribution of check points (CPAs and CPRs) used for geometric assessment. (a) spatial distribution of aerial orthoimages (rectangles) used for the absolute positioning assessment; (b) example of spatial distribution of CPAs (yellow triangles) within one of the selected orthoimages; (c) spatial distribution of CPRs (yellow dots). Light green areas in (a–c) are the agriculture devoted one according to Corine Land Cover 2012 map.

Figure 3. Mask image used to exclude cloudy pixels (black) from the analysis. Check points used for spectral assessment are in yellow. Image extent correspond to a whole S2 tile (Orbit:65, Tile: T32TMQ).

Mapping of not-cloudy pixels was achieved by thresholding the above mentioned band 10 and QA layers. Threshold values were defined by photointerpretation and the following values used, equal for all of the images. For S2 images, a unique threshold for the available images was determined by photo interpretation of clouds on screen using band 10 (cirrus). This spectral band corresponds to a strong absorption band of water vapour. Its absorption is so strong that a photon emitted by the sun in this wavelength has nearly no chance to reach the earth surface, and even less to reach the satellite after that without being absorbed. The consequence is therefore that the surface is usually not visible on the images taken for the 1.38 µm channel. We defined the threshold of 0.0025 by a conservative approach, on the basis of the minimum value registered on images. For L8 images we used the QA layers which coherently reports the reliability of the data at each pixel, identifying the areas that were
affected by cloud coverage always of the same value (code 20480 in QA). All of the masked images were then calibrated into at-the-ground reflectance according to the RTM suggested by Moran et al. [43] (3) and based on a DOS approach [44].

\[
\rho_\lambda(x, y) = \frac{\pi d^2 [L_\lambda(x, y) - \hat{L}_{\lambda atm}]}{\tau_\lambda \cdot \sin[\beta(x, y)] \cdot I_\lambda + E_{down}}
\]

where \( \rho_\lambda \) is the at-the-ground reflectance, \( L_\lambda \) is the at-sensor-radiance (W·sr\(^{-1}\)·m\(^{-2}\)·µm\(^{-1}\)) obtained by applying gain and offset values (if needed, i.e., L8 images), \( \hat{L}_{\lambda atm} \) is the upwelling atmospheric scattered radiance (W·sr\(^{-1}\)·m\(^{-2}\)·µm\(^{-1}\)), \( d \) the Sun-Earth distance coefficient (astronomical units), \( E_{down} \) the downwelling atmospheric scattered radiance (assumed equal to \( \pi \cdot \hat{L}_{\lambda atm} \)), \( \tau_\lambda \) the atmospheric transmittance, \( \beta \) the local Sun incidence angle (rad), and \( I_\lambda \) the sun irradiance (W·m\(^{-2}\)·µm\(^{-1}\)).

The Sun-Earth distance coefficient \( d \) needed for S2 images calibration was calculated as proposed by Wahid and Akiyama [45]:

\[
d = 1 - 0.01672 \cdot \cos[0.9856 \cdot (\text{Julian Day} - 4)]
\]

Differently, \( d \) was directly obtained from the metadata file (MTL file) while calibrating L8 images. Moran RTM was applied to L8 dataset in the standard mode, being L8 Level 1 imagery provided in the ordinary 12 bit digital numbers. Differently, S2 Level 1C images, ordinarily supplied as TOA reflectance, were preventively recovered back to radiance values, \( L_\lambda(x, y) \), by inverting (5), and then the Moran’s RTM applied.

\[
\rho_\lambda^{TOA} = \frac{\pi d^2 \cdot L_\lambda(x, y)}{I_\lambda \cdot \sin \varepsilon}
\]

where \( \varepsilon \) is the Sun elevation angle at the date of image acquisition. Other parameters are the same of (3).

The values of RTM parameters values adopted during the calibration step are reported in Tables 3 and 4.

| Table 3. Sun irradiances and atmospheric transmittance values used for images calibration. Atmospheric transmittance values are the ones proposed in literature for a reference mid-latitude low-hazy atmosphere [46]. |
|-----------------|-----------------|-----------------|
| Band Name | Sun Irradiance (W·m\(^{-2}\)·µm\(^{-1}\)) | Atmospheric Transmittance | Wavelength (µm) |
| B1 | 1995 | 0.50 | 0.435–0.451 |
| B2 | 1998 | 0.60 | 0.452–0.512 |
| B3 | 1865 | 0.65 | 0.533–0.590 |
| B4 | 1594 | 0.65 | 0.636–0.673 |
| B5 | 987 | 0.80 | 0.851–0.879 |
| B6 | 248 | 0.89 | 1.566–1.651 |
| B7 | 77 | 0.92 | 2.107–2.294 |

Moran’s RTM is retained compliant with an operational workflow (largely used for agricultural applications), being quite robust, but simple enough to be properly managed by most of the users. It is
well known that this model is hybrid between the physically based and empirical ones, thus involving a high degree of approximations, mainly concerning atmosphere modelling. We expect that a further step for minimizing residual differences between images have to be considered after image calibration, possibly based on a statistical regression approach. In this way, the differences could be modeled globally with no regard about which and how involved model parameters condition image calibration performance.

\[ NDVI(x, y) = \frac{\rho_{NIR}(x, y) - \rho_{RED}(x, y)}{\rho_{NIR}(x, y) + \rho_{RED}(x, y)} \]  
\[ NDWI(x, y) = \frac{\rho_{NIR}(x, y) - \rho_{MIR1}(x, y)}{\rho_{NIR}(x, y) + \rho_{MIR1}(x, y)} \]

where \( \rho_{RED} \) is the at-the-ground reflectance in the red spectral range (band 4 for both for S2 and L8 dataset); \( \rho_{NIR} \) is the at-the-ground reflectance in the near infrared spectral range (band 8 and band 5, respectively, for S2 and L8); and, \( \rho_{MIR1} \) is the at-the-ground reflectance in the medium infrared spectral range (band 11 and 6, respectively, for S2 and L8). Pairs of S2 and L8 calibrated bands of (quasi-) coeval images were compared by calculating the Pearson’s \( R \) correlation coefficient. Pairs of S2 and L8 \( NDVI \) and \( NDWI \) images were compared as well. Since \( R \) only measures the degree of similarity between reflectance pattern over images, it cannot be used to derive information concerning the strength of differences: high \( R \) values could be associated with high differences (low consistency) if a bias is present. It is worth to point out that in this work bias is intended as that part of the relative spectral differences between the two datasets that can be somehow modelled. To take care of this issue the following statistics of reflectance differences (S2 minus L8) were calculated: mean (\( \mu \Delta \)) and standard deviation (\( \sigma \Delta \)). A deeper analysis was also performed to describe and model \( NDVI \) and \( NDWI \) differences, being the main players of agronomic applications of remote sensing. Scatterplots relating pairs of S2 and L8 \( NDVI/NDWI \) correspondent values were therefore generated. A bias was found and modeled by linear regression, (8) and (9).

\[ NDVI_{S2} = a_0 \cdot NDVI_{L8} + a_1 \]  
\[ NDWI_{S2} = b_0 \cdot NDWI_{L8} + b_1 \]

where \( a_0, a_1, b_0, b_1 \) are regression coefficients; \( NDVI_{S2}, NDVI_{L8}, NDWI_{S2}, NDWI_{L8} \) the local spectral index value respectively computed from S2 and L8 datasets. Once opportunely calibrated, (8) and (9) can be used to remove bias from L8 \( NDVI/NDWI \), thus making them more consistent with the
correspondent S2 ones. RMSE of spectral differences was used as synthetic statistic measure of spectral consistency of datasets.

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\Delta SI_i^2)}{n-1}}
\]  

(10)

where \(n\) is the number of CPs (250) and \(\Delta SI_i\) the difference between S2- and L8-computed spectral indices (\(NDVI\) e \(NDWI\)). The RMSE reduction rate (RRR) was used to quantify the degree of the consistency improvement according to (11):

\[
RRR = \frac{(RMSE - RMSE')}{RMSE} \times 100
\]

(11)

where \(RMSE\) and \(RMSE'\) are referred, respectively, to spectral index differences before and after bias removal.

3. Result and Discussion

3.1. Geometric Assessment

Concerning absolute positioning, CPAs were located around the regional farming areas for the above mentioned reasons. In Figure 4a, the CPAs distribution in respect of land use classes is reported showing a strong dominance of the agricultural one. Consequently, the majority of CPAs falls in flat areas mostly representing agriculture or built-up classes, generally located close to farming sites. Distributions of CPAs and of regional farming areas in respect of slope classes were compared to further demonstrate that CPAs well fit the regional farming context (Figure 4b,c).

![Figure 4. Histograms showing CPAs distribution in respect of land use classes and area steepness. (a) Occurrences (%) of land cover classes (Corine Land Cover 2012—Level 1) where CPAs fall in. Class label are: 1 = artificial areas, 2 = agricultural areas, 3 = forest and seminatural areas, 4 = Wetlands, 5 = Water bodies; (b,c) occurrences (%) of slope classes for both CPAs and the whole regional agricultural areas.](image)

Concerning CPRs, no similar statistics were computed since the selection criterion was different, simply following a regularly distribution of points over the compared scenes. Absolute and relative consistency were tested by computing the following statistics: a) mean (\(\mu_{AE}, \mu_{AN}\)), standard deviation (\(\sigma_{AE}, \sigma_{AN}\)) and \(RMSE\) values of the East and North coordinates differences (Table 5). A complete statistical description of error distribution was obtained by computing the cumulative frequency distribution of local total error \(\Delta_i\) (Figure 5).

Statistics concerning accuracy revealed that no systematic error was present, since \(\mu_{AE}\) and \(\mu_{AN}\) remained lower than the correspondent standard deviations (\(\sigma_{AE}, \sigma_{AN}\)). An exception came from the N coordinate differences in the S2 vs. L8 comparison, where an opposite situation is given. It is not a goal of this research to investigate the possible reason of this anomaly; nevertheless, it is worth to
point out this finding as a possible critical point. A practical suggestion from this information is that is much more desirable, while jointly using S2 and L8 datasets, operating at the L8 spatial resolution. Concerning the value of differences S2 RMSEs was 5.9 and 16.2 m, respectively, for the absolute and relative positioning quality tests. When considering the geometric resolution of both S2 and L8 images (10 and 30 m, respectively), RMSE reveal that error is about half a pixel in both of the cases, ensuring that spectral measures from S2 and L8 datasets refer to the same ground portion and, therefore, represent the same type of surface.

**Table 5.** Errors statistics: mean and standard deviation of E and N differences, Root Mean Squared Error (RMSE). (left) Absolute positioning statistics (S2 vs. ICE). (right) Relative positioning statistics (S2 vs. L8).

<table>
<thead>
<tr>
<th></th>
<th>Orthoimage ICE 2009 vs. Sentinel-2</th>
<th>S2 vs. L8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\mu_E) (m)</td>
<td>(\sigma_E) (m)</td>
</tr>
<tr>
<td>Orthoimage ICE 2009 vs. Sentinel-2</td>
<td>-0.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Figure 5. Distribution of cumulative frequency of positioning errors. (a) Absolute positioning errors; (b) relative positioning errors.

3.2. **Spectral Assessment**

Spectral assessment was performed with reference to all the selected pairs of (quasi-) coeval images (see Tables 1, 6 and 7) using the randomly selected not-cloudy pixels (250). Pearson’s correlation coefficient \(R\) was computed by comparing the corresponding (spectrally) L8 and S2 calibrated bands (Table 6). As expected, \(R\) increases while time distance between S2 and L8 compared images decreases. Highest correlations (\(R > 0.75\) in Table 6) were obtained for bands belonging to the infrared spectral region; \(R\) values, surprisingly, decrease in the shortest wavelengths of the visible region. It can be supposed that this fact can be related to the adoption of the Moran’s RTM, which is not able to completely model atmospheric scattering effects: it is worth to remind that it mostly acts in the visible part of the spectrum and turns to negligible in the infrared one. Mean (\(\mu\Delta\)) and standard deviation (\(\sigma\Delta\)) values of differences were computed too, to give a value to spectral consistency of the compared bands. Results of Table 6 show that reflectance differences are acceptable in the most of cases, i.e., measures made by the two compared sensors are consistent. Again, time distance between the compared images determines the highest differences in reflectance values. Differently, \(\mu\Delta\) and \(\sigma\Delta\) do not depend on band wavelength like \(R\) does. Since spectral indices are widely used as final measures to be compared with ground observations in agriculture/forest applications, the effect that the previously measured reflectance differences generate onto indices has to be explored. Scatterplots relating (quasi-) coeval pairs of S2- and L8-derived NDVI and NDWI values were generated (Figure 6). They show that S2 and L8 NDVI and NDWI values are generally consistent (Table 8), but a significant linear bias is present. Bias was modeled by linear regression. Values of gain \((a_0, b_0)\), offset \((a_1, b_1)\) and \(R\) Pearson’s correlation coefficient for all of the tested combinations are reported in Table 7.
Table 6. Statistics concerning band per band reflectance differences (S2 minus L8). Selected combinations are those relating bands that reasonably can be supposed a-priori correlated (i.e., bands belonging to the same region of the spectrum). When compared S2 and L8 scenes are the closest ones within the time series. Reported values show that reflectance differences are acceptable. R is the Pearson’s correlation coefficient. * Correlations are significant at $p = 0.01$.

<table>
<thead>
<tr>
<th>Bands</th>
<th>Compared S2 and L8 Image Pairs (Date S2–Date L8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L8</td>
<td>S2</td>
</tr>
<tr>
<td>-------</td>
<td>----</td>
</tr>
<tr>
<td>B1</td>
<td>B1</td>
</tr>
<tr>
<td>B2</td>
<td>B2</td>
</tr>
<tr>
<td>B3</td>
<td>B3</td>
</tr>
<tr>
<td>B4</td>
<td>B4</td>
</tr>
<tr>
<td>B5</td>
<td>B8</td>
</tr>
<tr>
<td>B6</td>
<td>B8a</td>
</tr>
<tr>
<td>B7</td>
<td>B12</td>
</tr>
<tr>
<td>B7</td>
<td>B13</td>
</tr>
</tbody>
</table>
They show that also NDVI and NDWI differences depend on time lag between the compared measures. Higher is the time distance between compared images, weaker is the correlation between S2 and L8 spectral indices. It can also be noted that (a) S2 tends to over-estimate NDVI values in respect of the correspondent L8 one; (b) S2 tends to under-estimate NDWI values in respect of the correspondent L8 one; and, (c) S2 NDVI values tend to saturate for high values more rapidly than L8 derived ones. Once bias affecting NDVI/NDWI was modeled, L8-derived NDVI/NDWI were corrected and the new differences tested (Table 8). Results are encouraging and demonstrate that, in spite of still resisting residual biases, probably related to the adopted RTM, spectral index measures from S2 and L8 dataset can be finally assumed, after correction, as belonging to the same standard statistic population. Results clearly pointed out that, while approaching data integration, looking for a joint use of different datasets (S2 and L8) along the same statistical (multi-temporal) distribution, a normalization process (possibly “relative”) has to be applied.

Table 7. Correlation coefficient (R Pearson) and parameters of linear regression (see (8) and (9)) relating S2 and L8 NDVI/NDWI. Linear regressions were used to remove bias error between S2 and L8 derived spectral indices. All of the correlations are significant at $p = 0.01$.

<table>
<thead>
<tr>
<th>Compared S2 and L8 Image Pairs (Date S2–Date L8)</th>
<th>NDVI</th>
<th>NDWI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_0$</td>
<td>$a_1$</td>
</tr>
<tr>
<td>06/08/2015–07/08/2015</td>
<td>0.9305</td>
<td>0.1811</td>
</tr>
<tr>
<td>02/03/2016–03/03/2016</td>
<td>0.7491</td>
<td>0.3214</td>
</tr>
<tr>
<td>10/08/2016–25/08/2016</td>
<td>0.3756</td>
<td>0.5032</td>
</tr>
<tr>
<td>26/09/2016–29/09/2016</td>
<td>0.9279</td>
<td>0.2492</td>
</tr>
<tr>
<td>12/10/2016–29/10/2016</td>
<td>0.6193</td>
<td>0.4195</td>
</tr>
</tbody>
</table>

Table 8. RMSEs of NDVI/NDWI differences from S2 and L8 imagery before ($RMSE$) and after ($RMSE'$) bias removal. RMSE reduction rate (RRR) is the amount (%) of improvement determined by bias modeling.

<table>
<thead>
<tr>
<th>Image Pair (S2–L8)</th>
<th>Time Delay (d)</th>
<th>NDVI</th>
<th>NDWI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>$RMSE'$</td>
</tr>
<tr>
<td>06/08/2015–07/08/2015</td>
<td>1</td>
<td>0.18</td>
<td>0.10</td>
</tr>
<tr>
<td>02/03/2016–03/03/2016</td>
<td>1</td>
<td>0.27</td>
<td>0.10</td>
</tr>
<tr>
<td>10/08/2016–25/08/2016</td>
<td>15</td>
<td>0.43</td>
<td>0.21</td>
</tr>
<tr>
<td>26/09/2016–29/09/2016</td>
<td>3</td>
<td>0.25</td>
<td>0.12</td>
</tr>
<tr>
<td>12/10/2016–29/10/2016</td>
<td>17</td>
<td>0.30</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Values of Table 8 demonstrate that bias removal determine an appreciable improvement in spectral consistence of S2 and L8 NDVI/NDWI values. Nevertheless, $RMSE'$ values are not still completely satisfying, being about three times higher than the expected theoretical NDVI/NDWI uncertainty values ($\sigma_{NDVI/NDWI}$) that are related to sensor performance. Some works in literature [31] suggest reference values of $\sigma_{NDVI/NDWI}$ around 0.025. If an index difference is considered in place of the index value, its uncertainty can be estimated by the variance propagation law as $\sigma_{\Delta} = \sqrt{2\sigma_{NDVI/NDWI}} = 0.035$. It is worth reminding that, $\sigma$ and RMSE can be considered comparable (coincident) if the mean value of the differences is zero. Since $RMSE'$ is computed after an Ordinary Least Squares estimation of parameters of the linear regression, this condition is certainly verified (mean of differences is about 0). $RMSE'$ and $\sigma_{\Delta}$ can be therefore compared; assuming $\sigma_{\Delta} = 0.035$ as the expected intrinsic (theoretical) NDVI/NDWI measure uncertainty and $RMSE'$ like the actual one. This demonstrates that the differences affecting spectral indices from S2 and L8 after bias removal are still about three times higher than the expected value. These results are very useful to inform operators that, if “mixed” time series of NDVI/NDWI are generated from DOS-calibrated data, significant index variations (in space and time) are those whose absolute value is higher than 0.08 for NDWI and 0.1 for NDVI. It is worth pointing out that,
in an operative scenario, where the expected product is a sequence of observations along a common time series, users will not be interested in comparing (quasi-) coeval S2 and L8 images to model spectral bias: L8 and S2 images of different days are preferable. In these conditions, bias modeling cannot be achieved looking at whatever pixels, since one has to admit that probably their spectral properties have changed (especially for vegetation). Consequently, a standardized procedure for data integration, once more, will have to be based on the identification of spectrally invariant pixels (structures, rocks, roads, etc.) to be used for calibrating bias removal model.

**Figure 6.** Scatterplots relating pairs (S2 vs. L8) of Normalized Differencing Vegetation Index (NDVI) (left) and Normalized Differencing Water Index (NDWI) (right). Black lines represent the $y = x$ expected trend; on $x$ axis are reported the L8 values, on $y$ axis the S2 values. They clearly show that a good correlation exists between the two datasets, but different from the expected trend. A linear bias can be modeled to improve measures consistency. It is evident that correlation decreases while time distance between compared image pairs increases (images are not perfectly coeval).
4. Conclusions

The consistency of S2 and L8 dataset was tested to get a preliminary evaluation in view of a joint use of spectral indices along common time series that are able to respond effectively to crop monitoring purposes. Tests concerned both geometric and spectral features of datasets. Results showed that geometric quality of S2 data is high both for absolute and relative (in respect of L8 images) positioning. Computed position errors are consistent with the geometric resolution of data (about half a pixel). No systematic error was found affecting planimetric coordinates. An anomaly was found affecting the north coordinate differences were a moderate bias was observed. This is however negligible at the geometric resolution of L8.

Concerning spectral assessment, the focus was on both at-the-ground reflectance and spectral indices (NDVI and NDWI) as obtainable applying a simplified RTM (Moran’s DOS). We found that reflectance differences between correspondent (or almost correspondent) spectral bands, in the most of the cases, is lower than 0.1. We also found that differences suffer from some systematic error especially affecting bands belonging to the visible spectral region. Concerning the effects of computed reflectance differences onto spectral indices, we found that S2 and L8 derived NDVI and NDWI values are generally consistent, but not negligible biases can be observed. They were modeled by linear regression. Depending on time distance between compared images, NDVI differences range between 0.18 and 0.30 if no bias modeling is performed. They can be reduced around 0.1 by linear bias correction. Depending on time distance between compared images NDWI differences range between 0.13 and 0.25 if no bias modeling is performed. They can be reduced around 0.08 by linear bias correction too. These values are however high in respect of the expected ones (about three times higher). It can be finally pointed out that: (a) spatial information from the S2 and the L8 dataset is consistent and released dataset well georeferenced; (b) spectral information, specifically related to NDVI and NDWI, if managed through a simplified DOS-based RTM is not satisfyingly consistent, making unreliable those time series obtained fusing the two datasets; (c) a bias affecting NDVI/NDWI differences was found and modeled. L8 unbiased NDVI/NDWI are much more consistent with the S2 ones, but significant differences persist; and, (d) NDVI from S2, probably depending on the bands used in the formula (B8 and B4) that do not perfectly fit the correspondent L8 ones (B5and B4), tends to be less effective in mapping NDVI differences when NDVI values are high, showing an unwanted saturation effect.

Finally, if S2 and L8 NDVI/NDWI are computed from DOS-based RTM calibrated images, their integration is reliable enough for quantitative purposes only if a further inter-calibration (bias removal) step is performed. Further investigations have therefore still to be done exploring eventual improvements to data integration given by calibrating images by more rigorous RTMs, which are possibly driven by spectrally invariant pixels.

Author Contributions: Enrico Borgogno Mondino conceived and designed the experiments; Andrea Lessio and Vanina Fissore performed the experiments; Andrea Lessio and Enrico Borgogno Mondino analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest concerning this work.

References


27. Lefebvre, A.; Sannier, C.; Corpetti, T. Monitoring urban areas with Sentinel-2A data: Application to the update of the Copernicus high resolution layer imperviousness degree. Remote Sens. 2016, 8, 606. [CrossRef]

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).