Performance evaluation of OBIA-based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: a case study from Almería (Spain)

Antonio Novelli*[a], Manuel A. Aguilar[b], Abderrahim Nemmaoui[b], Fernando J. Aguilar[b], Eufemia Tarantino[a]

*Corresponding Author: antonio.novelli@poliba.com
http://dx.doi.org/10.1016/j.jag.2016.07.011

[a] DICATECh, Politecnico di Bari, Bari, via Orabona 4 C.A.P. 70125, ITALY (antonio.novelli@poliba.it, eufemia.tarantino@poliba.it)

[b] Escuela Superior de Ingeniería de la Unversidad de Almeria (UAL) , Almeria Ctra. Sacramento s/n 04120 - La Cañana de San Urbano Almería, España (maguilar@ual.es, an932@ual.es , faguilar@ual.es)

Abstract

This paper shows the first comparison between data from Sentinel-2 (S2) Multi Spectral Instrument (MSI) and Landsat 8 (L8) Operational Land Imager (OLI) headed up to greenhouse detection. Two closely related in time scenes, one for each sensor, were classified by using Object Based Image Analysis and Random Forest (RF). The RF input consisted of several object-based features computed from spectral bands and including mean values, spectral indices and textural features. S2 and L8 comparisons were also extended through using a common segmentation dataset extracted form VHR World-View 2 (WV2) imagery to test differences only due to their specific features contribution. The best band combinations to perform segmentation were found through a modified version of the Euclidian Distance 2 index. Four different RF classifications schemes were considered: L8 features extracted from both L8-based segments WV2-based segments; S2 features extracted from both S2-based segments and WV2-based segments. The best overall accuracies, evaluated on the whole study area, were 89.1%, 91.3%, 90.9% and 93.4% respectively.

Keywords:Sentinel-2 MSI, Landsat8 OLI, WorldView-2, Greenhouses, object-based classification, segmentation quality.

1. Introduction

The agricultural practice of Plastic-Covered Greenhouse (PCG) is of strategic economic importance in semiarid regions, thus its development requires an adequate monitoring (Briassoulis et al., 2013). Due to its peculiar characteristics (i.e. different material, spectral signatures and local agricultural practice) remote sensing PCG mapping is far to be solved as demonstrated by the increasing number of scientific works produced in the last decade. Particularly, the topic is relevant with passive remotely sensed data with which PCG detection has been carried out through two main approaches: Pixel Based (PB) and Object Based Image Analysis (OBIA).

PB methods have been applied both on very high resolution and medium-low resolution data. Indeed, recent studies were conducted by using World-View 2 (WV2) (Koc-San, 2013; Pala et al., 2015), QuikBird and Ikonos (Agüera et al., 2008; Agüera et al., 2006; Arcidiacono and Porto, 2010; Carvajal et al., 2010), Landsat 8 (L8) Operational Land Imager (OLI)/Thermal Infrared Sensor (TIRS) (Novelli and Tarantino, 2015a, b) and Landsat Thematic Mapper data (TM) (Picuno et al., 2011) data. Also for plastic-mulched landcover, literature shows case study through Landsat TM (Lu et al., 2014), L8 OLI (Chen et al., 2016) and the joint use of Landsat ETM+, OLI and Modis sensor data (Lu et al., 2015).

OBIA approach is much more recent, starting with the use of true colour aerial data (Tarantino and Figorito, 2012) and continuing with both WV2 and GeoEye-1 stereo pairs (Aguilar et al., 2014) and WV2 and L8 data (Aguilar et al., 2015). The OBIA approach was also chosen in this paper since it has demonstrated to be more efficient in PCG detection than PB one (Wu et al., 2016).

As far as the authors' knowledge, this is the first paper in proposing both a comparison regarding PCG detection between satellite date provided by the novel Sentinel-2A (S2) MSI and L8 OLI, and the use of OBIA approach on an user produced Level 2A (L2A) S2 MSI image. S2 MSI L2A data are atmospherically corrected (bottom of atmosphere reflectance values) by means of the sen2cor algorithm (Muller-Wilm et al., 2013). Indeed, recent literature has showed many studies with simulated S2 MSI data but at the moment, only a few scientific papers have dealt with real S2 MSI data (Du et al., 2016; Fernández-Manso et al., 2016; Immitzer et al., 2016). However in these

2

studies were used S2 MSI Level 1C (L1C) data for vegetation classifications purposes (Immitzer et al., 2016), for burn severity discrimination (Fernández-Manso et al., 2016), to detect build-up areas (Pesaresi et al., 2016) and to map water bodies (Du et al., 2016).

The proposed comparisons were undertaken by applying an OBIA approach coupled with Random Forest (RF) classifier (Breiman). In the aforementioned OBIA studies, the segmentation quality (i.e. the first stage of OBIA process) was only assessed by visual inspection, or even manually digitised, instead of applying a quantitative approach. In this paper a further improvement was proposed for the segmentation quality assessment by means of discrepancy measurements. Indeed among several available methods, this approach is considered as one of the most useful and reliable (Clinton et al., 2010; Liu et al., 2012; Yang et al., 2015). Specifically, this was achieved from a modified version of the Euclidean Distance 2 index originally proposed by Liu et al. (2012). Besides the S2 and L8 based segmentations, it was also considered a more accurate segmentation obtained from a higher resolution WV2 image. In this sense, WV2 segmentation was employed as the basis to test the differences between S2 and L8 classifications results attributable to only their informative content. Very high Overall Accuracy (OA) values were obtained when testing four different RF classifications schemes from using a very small training set: L8 features with L8-based segmentation, L8 features with WV2-based segmentation, S2 features with S2-based segmentation and S2 features with WV2-based segmentation.

2. Study area and Data pre-processing

2.1 Study Area

The study area (Figure 1) is located in the so-called "Sea of Plastic", province of Almería (Southern Spain). The main economic activity is agriculture (tomato, pepper, zucchini, cucumber, aubergine, green bean, melon, watermelon and Chinese cabbage) under plastic covered greenhouses. Different typologies of plastic materials are used to cover greenhouse structures. The most common ones are

EVA (Ethylene-vinyl acetate) and polyethylene films both with different kind of colours and thickness. In this region the PCG detection is troublesome because of the variable transmissivity associated to the kind of plastic cover and its cleanliness. Furthermore, the variation of greenhouses spectral signatures over time is also due to reflectance variations experienced by cultivated plants under them.
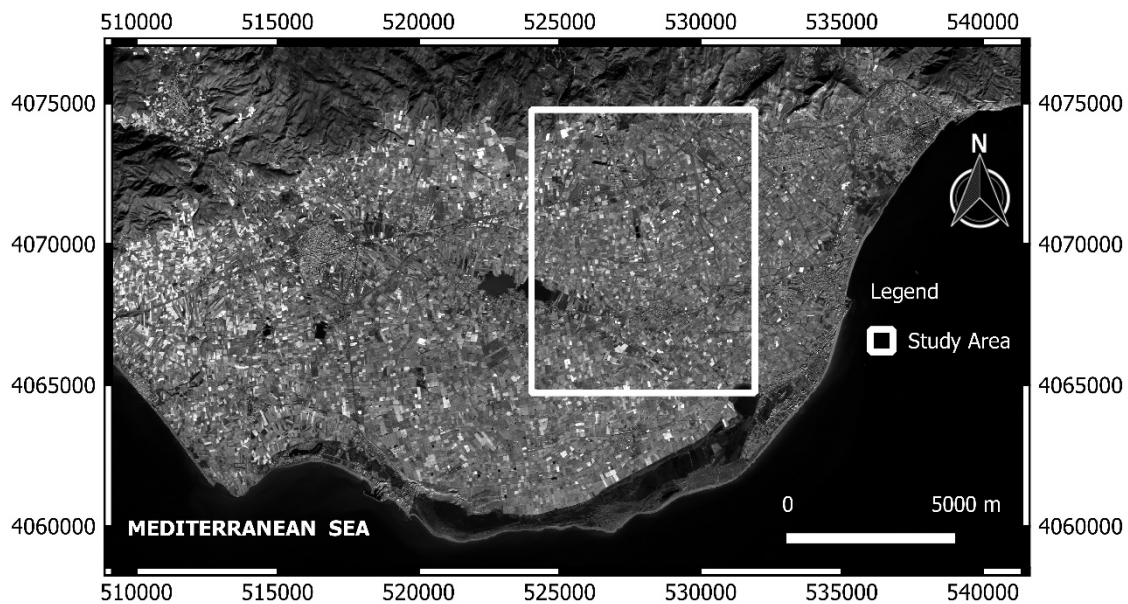

Figure1. Location of the study area depicted by means of the Red band of the Sentinel-2 image. Coordinate system: ETRS89 UTM Zone 30N

2.2 Satellite data pre-processing

Three optical satellite scenes from S2, L8 and WV2 were used in this case study. In each scene the study area was not affected by clouds presence. All the satellite data were atmospherically and geometrically corrected before the segmentation process. Moreover, L8 and S2 images were co-registered to the WV2 panchromatic (PAN) image since a very accurate spatial matching is required to perform multi-sensor comparisons (Townshend et al., 1992; Zhang et al., 2014)

The WV2 data (5 July 2015) consisted of a bundle combination of PAN and MultiSpectral (MS) images. The chosen WV2 images presented the following characteristics: Ortho Ready Standard Level-2A (ORS2A) format, dynamic range of 11-bit, lacking of DRA (dynamic range adjustment), geometric resolution of 2 m (MS) and 0.5m (PAN). WV2 was the first very high resolution

commercially available 8-band MS satellite with the following bands: coastal (C, 400–450 nm),

blue (B, 450–510 nm), green (G, 510–580 nm), yellow (Y, 585–625 nm), red (R, 630–690 nm), red

edge (RE, 705–745 nm), near infrared-1 (NIR1, 760–895 nm) and near infrared-2 (NIR2, 860–1040

nm). WV2 digital numbers were converted to ground reflectance values by using the ATCOR

algorithm implemented in Geomatica v. 2014 (PCI Geomatics, Richmond Hill, ON, Canada). In

particular, the core of the correction algorithm is based on the MODTRAN (MODerate resolution

atmospheric TRANsmission) radiative model (Berk et al., 1998). Seven Ground Control Points

(GCP), obtained by differential GPS, and a 10-m grid spacing DEM (courtesy of the Andalucía

Government) with 1.34m of vertical accuracy (root mean square error) were used to compute the

sensor model and to orthorectify the WV2 MS and PAN images. The sensor model was based on

rational functions refined by a zero order transformation in the image space (RPC0), being

computed through Geomatica v. 2014.

The L8 OLI scene (8 January 2016), Path 200 and Row 34, was downloaded at no cost from the

USGS EROS web site as Level 1 Terrain Corrected (L1T) product with 30m of geometric

resolution (Roy et al., 2014). The L8 image was composed by the following bands: coastal aerosol

(C, 430–450 nm), blue (B, 450–510 nm), green (G, 530–590 nm), red (R, 640–670 nm), near

infrared (NIR, 850–880 nm), shortwave infrared-1 (SWIR1, 1570–1650 nm), shortwave infrared-2

(SWIR2, 2110–2290 nm) and cirrus (CI, 1360–1380). The OLI panchromatic band was not used in

this study. The extracted subset was atmospherically corrected by applying the ATCOR algorithm

and co-registered with the WV2 PAN orthoimage through Geomatica v. 2014.

The S2 MSI image (12 January 2016, orbit R051) was downloaded at no cost from the Copernicus

Scientific Data Hub web site as a Level 1C (L1C) product. S2 MSI L1C product is characterized by

Top Of Atmosphere (TOA) reflectance values, cartographic projection, 12-bit dynamic range and

tiles/granules consisting of 100 km^2 ortho-images in UTM/WGS84 projection. The selected study

area is comprised within the T30SWF granule. The MSI sensor collects up to thirteen bands with

three different geometric resolutions: 60m, 20m and 10m. Costal (C, 443 nm), water vapour (WV,

1375 nm) and cirrus (CI, 1376) at 60 m resolution. Four red edge/NIR bands with central wavelength at 705 nm, 740 nm, 783 nm and 865 nm respectively, short wave infrared-1 (SWIR1, 1610 nm) and short wave infrared-2 (SWIR2, 2190 nm) at 20m resolution. Blue (B, 490 nm), Green (G, 560 nm), Red (R, 665 nm) and Near Infrared (NIR, 842 nm) at 10m resolution. The sen2core algorithm (Muller-Wilm et al., 2013) was used to produce a Level 2A (L2A) SE2 MSI product characterized by atmospherically corrected Bottom Of Atmosphere (BOA) reflectance values. Finally, the S2 study area was extracted from the selected granule and co-registered with the WV2 PAN image by using Geomatica v. 2014. 60m bands were not considered in the subsequent comparisons.

3. Methods

3.1 Segmentation algorithms

In this study was used the MultiResolution Segmentation (MRS) algorithm provided by eCognition v. 8.8. MRS is a bottom-up region merging object algorithm (see Baatz and Schäpe (2000) and Tian and Chen (2007) for a complete mathematical description). It takes into account each pixel as a separate object and subsequently pairs of image objects are merged to form bigger segments (Darwish et al., 2003). However, MRS requires user driven parameters and it is not easy to obtain a satisfactory segmentation for the required objects (Tian and Chen, 2007). Indeed, the MRS algorithm output depends on three main factors or parameters: scale, shape and compactness. The Scale Parameter (SP) or heterogeneity criteria is the most influent since it controls the size of segments and thus the over-segmentation and under-segmentation error (Frauman and Wolff, 2005). Moreover, other input information such as the considered band combination should also be fixed into MRS.

It is worth noting that the initial geometric resolution of the corrected L8 image (30m GSD) was increased to 1.875 m by simply halving four times the original pixel size. The same procedure was applied to the corrected S2 image, being the 10m and 20m GSD bands split, without any

6

resampling, in 2m GSD bands. This was necessary in order to enhance the fit between the S2-L8 images segmentations and WV2 image segmentation. In the end, and by applying the chessboard segmentation algorithm included in eCognition v.8.8., the higher resolution WV2-based segmentation (see 3.2 segmentation quality assessment) was applied to test the classification results attained from a common and accurately segmented dataset.

3.2 Segmentation quality assessment

The selection of the best three MRS parameters was carried out with a modified version of the supervised discrepancy measure named Euclidean Distance 2 (ED2) (Liu et al., 2012). As a supervised segmentation quality metric the modified ED2 works with a set of reference objects used to evaluate the goodness of the segmentation. Moreover, it can be assessed through the capabilities of GIS softwares. In this study 400 reference plastic covered greenhouse objects, manually delimitated on the corrected WV2 scene, were taken as reference objects. The 400 reference geometries were manually digitized considering only the plastic covered greenhouses common to the three satellite images. The references contours were drawn using the WV2 PAN orthoimage.

ED2, in its original formulation, starts the computations with the definition of the corresponding segment dataset. For each considered image segmentation output the dataset owned the segments that spatially overlap the reference polygons. A further constrain is imposed over the corresponding segment dataset (Clinton et al., 2010): a considered segment can be labelled as a corresponding segment if the area of intersection between a reference polygon and the candidate segment is more than half the area of either the reference polygon or the candidate segment (overlapping criteria). After defining the corresponding segments dataset, the ED2 index (1) evaluates the segmentation quality in a two dimensional Euclidean space by means of the Potential Segmentation Error (PSE) and the Number-of-Segments Ratio (NSR). The PSE (2) metric measures the geometric discrepancy as the ratio between the total area of under-segments and the total area of reference polygons. The under-segmentation error occurs when the contour of a reference polygon r_k divides the

7

corresponding segment s_i into at least two parts. Parts falling outside the reference polygon represent the under-segmented area (PSE = 0 means no under-segmentation). The NRS (3) measures the arithmetic discrepancy between the reference geometries and the candidate segments, being defined as the absolute difference between the number of reference polygons (m) and the number of corresponding segments (v) divided by the number of reference polygons.

$$ED2 = \sqrt{(PSE)^2 + (NSR)^2} \tag{1}$$

$$PSE = \frac{\sum |s_i - r_k|}{\sum |r_k|} \tag{2}$$

$$NSR = \frac{|m - v|}{m} \tag{3}$$

A high ED2 value indicates a significant geometric discrepancy, otherwise a significant arithmetic discrepancy, or both.

The implemented modification of the ED2 index was introduced to consider the side effect of the overlapping criteria that act as a filter both on candidate corresponding segments and on reference geometries. When the number of reference geometries rises, there are often reference geometries without any corresponding segments. In those cases the true number of employed reference geometries will be lower than the original one. Therefore, the ED2 index should take this into account to avoid bias when computing both PSE and NSR.

In this study the overlapping criteria side effect was corrected by increasing both the PSE and NSR values when not all reference geometries meet the overlapping criteria. Being n the number of excluded reference geometries, the new computed PSE (4) and NSR (5) will be:

$$PSE_{new} = \frac{\sum |s_i - r_k| + n \times \max(|s_i - r_k|)}{\sum |r_k|} \tag{4}$$

$$NSR_{new} = \frac{|m - v - n \times v_{max}|}{m - n} \tag{5}$$

Where $\max(|s_i - r_k|)$ is the maximum over segmented area found for a single reference geometry, v_{max} represents the maximum number of corresponding segments found for one single reference geometry and $\sum |r_k|$ computes the total area of the $m - n$ reference geometries. The modified ED2

index was then applied to obtain the optimized MRS parameters from which extracting the potentially best segmentation for each one of the three input satellite images used in this work.

3.3 Training set and Features extraction

Three training sets of 60, 90 and 120 segments were created from the three best estimated segmentations for L8, S2 and WV2. For each training set, one half of the geometries were related to the "Greenhouse" class and the other half to the class labelled as "Other". Special attention was given to the selection of each single training segment. They were manually selected to allow be considered as "pseudo-invariant" objects (similar geometry and same class) for the two classes and the three segmented satellite images.

S2 and L8 comparisons were obtained from the following classification schemes: L8 with L8-based segmentation (L8_SEG_L8), L8 with WV2-based segmentation (L8_SEG_WV2), S2 with S2-based segmentation (S2_SEG_S2) and S2 with WV2-based segmentation (S2_SEG_WV2). Notice that the geometric information provided by the WV2 segmentation (i.e., common input segmentation) was used to test differences only due to S2 and L8 informative content. The comparisons were repeated for the three different training sets.

Features included in the classification process were computed at object level, compiling for each considered object a vector containing spectral information, texture data and spectral indices. Texture data were obtained from the Haralick Grey Level Co-occurrence Matrix (GLCM) (Haralick et al., 1973). The best achieved segmentation, for each classification scheme, provided the geometric attributes for the classification input features computed by using eCognition v8.8. Table 1 summarizes the content of each i-th vector composed by 126 features for the S2 image and 87 features in the case of the L8 image. The larger number of S2 object features was mainly due to the iteration of 20m red edge/NIR bands in the place of 10m S2 NIR band (i.e., we also tested the 20m red edge/NIR bands instead of the 10m NIR one to compute some indices in Table 1).

Table 1. Sentinel-2 (S2) and Landsat 8 (L8) object based features (feat.)

Typology	Tested feat.	L8 No. of feat.	S2 No. of feat.	Description	Reference
Spectral information	Mean and Standard deviation (SD)	16	20	Mean and SD of each band	(Definiens, 2009)
Indices	NDVI (Normalized Vegetation Index)	1	5	(NIR−R) / (NIR+R)	(Rouse Jr et al., 1974)
	GNDVI (Green NDVI)	1	5	(NIR−G) / (NIR+G)	(Gitelson et al., 2002)
	PMLI (Plastic-mulched landcover index)	1	1	(SWIR1-R) / (SWIR1+R)	(Lu et al., 2014)
	SWIR1_NIR	1	5	(SWIR1-NIR) / (SWIR1+NIR)	This study
	SWIR2_NIR	1	5	(SWIR2-NIR) / (SWIR2+NIR)	This study
	SW1_SW2_NIR	1	5	(((SWIR1+SWIR2)/2)-NIR)/ (((SWIR1+SWIR2)/2)+NIR)	This study
	CIRRUS_NIR	1	-	(CIRRUS-NIR) / (CIRRUS+NIR)	This study
Texture	GLCM_h	8	10	GLCM homogeneity all directions	(Haralick et al., 1973)
	GLCM_d	8	10	GLCM dissimilarity all directions	(Haralick et al., 1973)
	GLCM_e	8	10	GLCM entropy all directions	(Haralick et al., 1973)
	GLCM_c	8	10	GLCM contrast all directions	(Haralick et al., 1973)
	GLCM_a	8	10	GLCM angular 2nd moment all directions	(Haralick et al., 1973)
	GLCM_cor	8	10	GLCM correlation all directions	(Haralick et al., 1973)
	GLCM_sd	8	10	GLCM standard deviation all directions	(Haralick et al., 1973)
	GLCM_m	8	10	GLCM mean all directions	(Haralick et al., 1973)

3.4 Random Forest classifier design and classification accuracy assessment

In this study the RF classifier was used as a tool to perform comparisons between S2 MSI and L8 OLI scene in PCG detection, since it has performed good classification results in several remote sensing studies demonstrating its robustness against a high number of variables (Breiman;

Rodriguez-Galiano et al., 2012; Smith, 2010). A detailed review of the RF classifier algorithm is beyond the scope of this paper. More information about the mathematical formulation and its parameters can be found in the literature (e.g. Breiman (2001), Dietterich (2000)).

The RF algorithm was applied by means of STATISTICA v10® (StatSoft Inc., Tulsa, OK, United States). If the number of input trees is large enough by default, the software will determine the best final model as the one (i.e., as the specific number of trees) that yields the smallest error estimate for the testing sample. In this case study, a precautionary values of 500 trees (always above the best solution found by the software) was chosen and the number of random predictive variables was computed from the expression $p = log_2(M + 1)$, being M the total number of predictor variables (features) (Hill and Lewicki, 2007). Lastly, the input training set was divided in two sub-sets for each classification: approximately 2/3 of the available data were used to train the classifier and the remaining ones to validate the training. The last one is usually called Out-Of-Bag (OOB) data. OOB accuracy is an unbiased estimator of the classification OA accuracy. However, this estimation would be based on objects more than pixels and thus the error due to an erroneous segmentation would not be considered. In this way, and to provide a more reliable and complete accuracy indicator, pixel based confusion matrices based on ground truths manually selected for the whole study area were computed (Figure 2). Ground truth was built up over the geometric base of the corrected WV2 PAN image, taking into account the land cover of the S2 and L8 images. Finally, the classification accuracy assessment in this work was based on the ground truth shown in Figure 2, using a pixel-based error matrix. Hence, the accuracy measures computed were user's accuracy (UA), producer's accuracy (PA), overall accuracy (OA) and kappa coefficient (kappa) (Congalton, 1991)

4. Results and discussion

4.1 Segmentation procedure

The best input segmentations were computed by using the MRS algorithm and the modified ED2 index. This task was carried out first by varying only the SP value and fixing shape and compactness to 0.5. Then, when a local optimum SP was found, the research continued considering: SP values within an interval of the local optimum, shape values from 0.1 to 0.9 and compactness fixed to 0.5. For each calculation, SP and shape parameters were incremented in steps of 1.0 and 0.1 respectively. Several band combinations were tested for the three satellite data, the visible and near infrared bands turning out to be the most important regarding the final quality of the segmentation. Table 2 summarizes the characteristics of the best estimate segmentation performed over the three atmospherically and geometrically corrected images. The compactness parameter was always fixed to 0.5 since in literature there are evidences of its negligible weight in the final output of the MRS algorithm (if compared to shape and, above all, SP parameter) (Drăguţ et al., 2014; Liu and Xia, 2010). Finally, for the S2 data the 10m GSD bands were considered the most valuable to produce a high quality segmentation results.
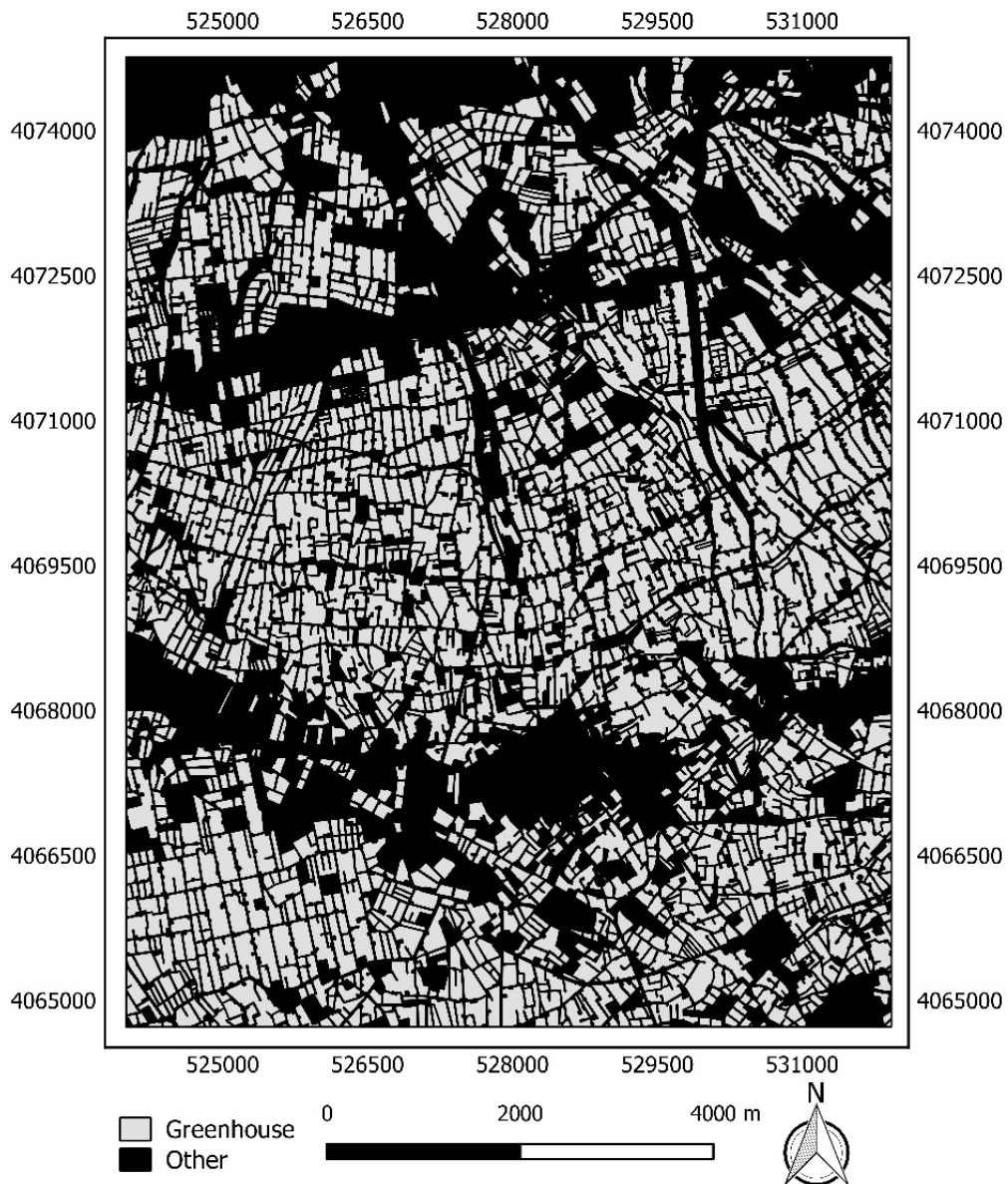
Figure 2. Manually digitized ground truth. Coordinate system: ETRS89 UTM Zone 30N

Table 2. Best estimate ED2 values and their associated input bands and Scale, Shape and Compactness parameters.

Satellite data	Band combination	No. objects	Scale	Shape	Compactness	Modified ED2
Landsat 8	Blue-Green-NIR	9596	43.0	0.3	0.5	0.424
Sentinel 2	Blue-Green-NIR	10561	39.0	0.2	0.5	0.319
WorldView2	Blue-Green-NIR2	10990	37.0	0.4	0.5	0.198

Table 2 shows that the best greenhouses segmentation results in the case of L8 and S2 satellite images were obtained from using the same bands combination. Moreover, a very similar ED2 result (0.199) was calculated with the Blue-Green-NIR1 band combination for the WV2 data. This turned out be a very important finding since it was proved the stability of the best bands combination in

13

order to retrieve the best segmentation on greenhouses. It is important to highlight that this result was obtained from atmospherically corrected images and using the same 400 reference geometries. Figure 3 shows a comparison of the three selected segmentation over the same area. Figure 3 shows a very high visual quality for the WV2 chosen PCG segmentation. Figure 3 also depicts that the segmentation based on L8 features performed the worst, while the S2-based segmentation still provided a good visual segmentation quality. This figure allows to appreciate the capability of the modified ED2 index to represent the segmentation quality of both VHR and medium resolution images.

4.2 Results from Random Forest classifier

The aim of the classification stage was to test differences in S2 and L8 PCG detection results. For this purpose their spectral content was coupled both with their respective geometric information (best L8 and S2 based segmentations respectively) and with the more accurate WV2-based segmentation. The four combinations (L8_SEG_L8, L8_SEG_WV2, S2_SEG_S2, S2_SEG_WV2) were applied to the three extracted training sets. For each classification, the input geometries for the training set were the ones corresponding to the chosen segmentation. Figure 4 shows a subset of the best classifications results according to each one of the considered four combinations. S2 classification featured a better visual quality than L8 classification for both S2 and WV2 based segmentations. In particular, S2

classification proved to be more adequate to discriminate narrow objects than L8 classification by using both segmentation approaches.
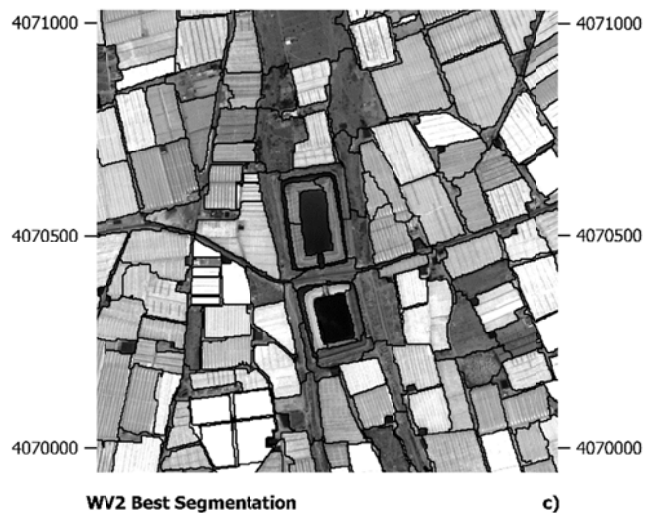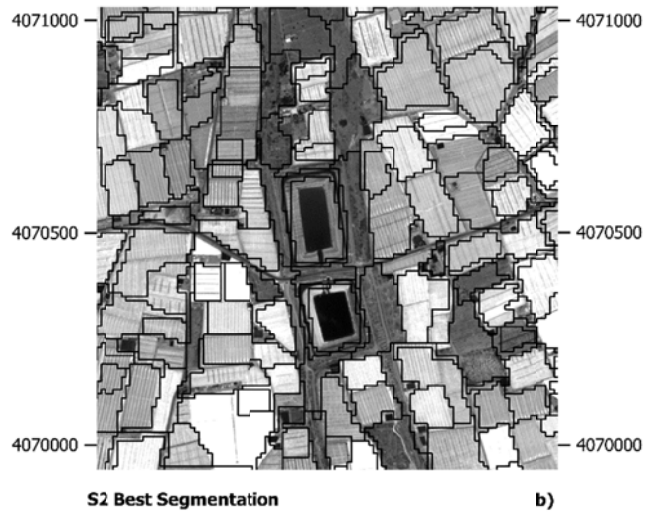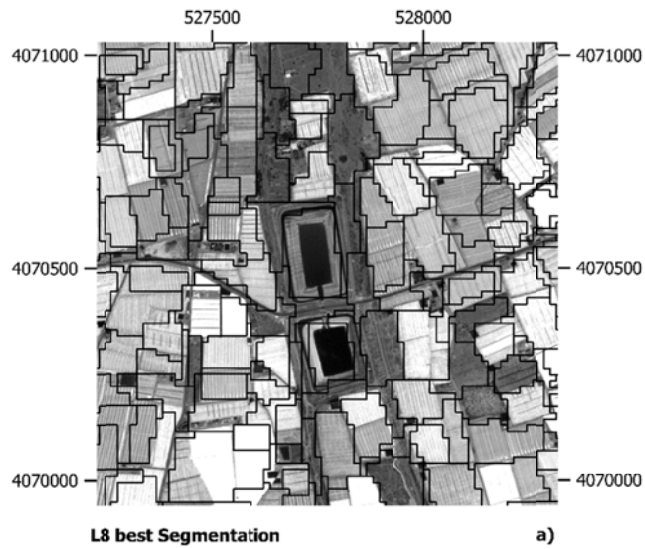
Figure3. Visual comparison over the WV2 orthoimage (RED band) of different best segmentations depending on the image data: a) best estimate L8-based segmentation; b) best estimate S2-based segmentation; c) best estimate WV2-based segmentation. Coordinate system: ETRS89 UTM Zone 30N

It might be due to mixed pixels are still commonly present in the medium resolution satellite

imagery as a result of the heterogeneity of the landscape and the limitations imposed by the spatial

15

resolution of the images. This fact was already reported by Wu et al. (2016) using an OBIA approach on a pan-sharpened Landsat 8 OLI image. In this regard, the reader should bear in mind the important difference between the best resolution for S2 (10m) and for L8 (30m).

Table 3 presents the achieved OA, Kappa Index of Agreement (KIA), Producer's Accuracy (PA) and User's Accuracy (UA) for the two considered classes.

Table 3. Achieved OA, KIA, PA and OA for the considered classifications. These results should be understood as true accuracies, and not as estimated ones, since they were computed using a ground truth comprising the whole working area.

Scene-Segments combination	training set	OA (%)	KIA	PA Greenhouse (%)	PA Other (%)	UA Greenhouse (%)	UA Other (%)
L8_SEG_L8	120	89.0	0.769	94.7	81.0	87.6	91.4
	90	89.1	0.773	93.1	83.4	88.8	89.6
	60	87.9	0.744	96.2	76.2	85.1	93.3
L8_SEG_WV2	120	89.8	0.791	90.6	88.8	92.0	86.9
	90	91.3	0.818	95.4	85.5	90.3	92.9
	60	90.8	0.806	95.1	84.8	89.9	92.5
S2_SEG_S2	120	90.9	0.864	95.1	84.8	89.9	92.5
	90	89.7	0.784	94.4	83.0	88.8	91.2
	60	89.8	0.786	94.9	82.4	88.5	92.0
S2_SEG_WV2	120	93.4	0.810	96.9	88.5	92.3	95.3
	90	92.6	0.844	97.0	86.2	90.9	95.3
	60	92.7	0.848	96.6	87.3	91.5	94.7

The obtained OA values, ranging from 87.9% (L8_SEG_L8 with 60 training geometries) to 93.4% (S2_SEG_WV2 with 120 training geometries), can be considered satisfactory taking into account the minimum value of 85% proposed by Congalton and Green (2008). Also KIA values showed a substantial and an almost perfect agreement (Landis and Koch, 1977). Both KIA and OA confirmed that S2 classifications performed always better than the corresponding L8 classifications. In particular, the difference in accuracy between S2 and L8 increased when the common WV2-based best segmentation was used. Since this result was achieved from the same segmentation (see 3.1 Segmentation algorithms), the attained differences can be attributed to the better performance of S2 features when undertaking the RF classification training process.

The PA reports about the number of pixels correctly classified in a particular category as a percentage of the total number of pixels actually belonging to that category in the image, being related to omission error. The PA for the "Greenhouse" class was always better than the PA of the "Other" class. These result confirmed the high classification quality achieved from the RF classifier for the "Greenhouse" class. Moreover, the lower "Other" class PA accuracy can be explained taking into account its high heterogeneity. Particularly, one only class was used to address all the spectral variability of the totality of land covers (i.e. water, vegetation, soil and build-up areas) different from the "Greenhouse" class. The UA is related to the probability that a pixel classified in a map actually represent that category on the ground. It is also related to commission error (error in field). UA values show that the reliability of the classification was very high for both classes, although the "Other" class featured slightly better UA values. Only in the case of the L8 with WV2-based segmentation (60 training segments), the UA of the "Greenhouse" class was greater than the corresponding UA for the "Other" class.

Although the training datasets were constituted of a very little number of geometries, the results were not very sensitive to the number of geometries. This was especially true with S2 data in which the best accuracies were always coupled with the 120 training geometries, whereas for the L8 data the best results were achieved from the training sets composed of 90 geometries.

RF classifier is also capable to estimate the features importance in the training process by means of the Gini index and the OOB estimation (Rodriguez-Galiano et al., 2012). Normalized features or ratios were always among the most important 10 features for both L8 and S2. However, the consistency of normalized features in S2 classification was stronger as compared to L8 outputs. In fact, among the top ten L8 most significant features, it was frequent to find mean-value spectral features and textural features.

This result seems to point out a better S2 stability in order to efficiently extract plastic-covered greenhouses regardless atmospheric conditions.
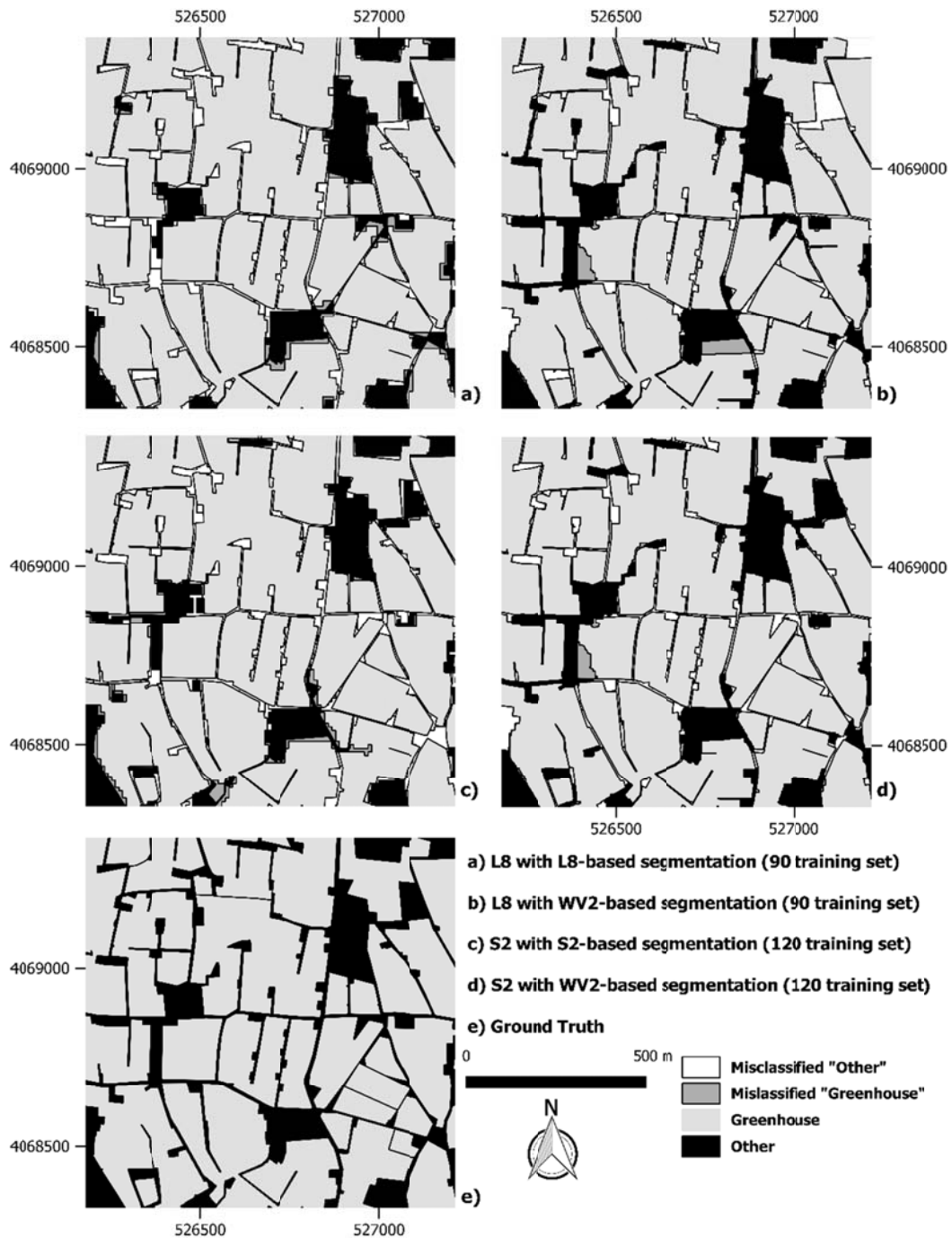
Figure 4. Comparisons of the best classification results for the four considered combinations. Coordinate system: ETRS89 UTM Zone 30N.

5.Conclusions

To the best knowledge of the authors, this study proposes a first comparison regarding PCG detection between S2 and L8 satellite data through applying an OBIA approach and RF classifier. The high geometric contribution of WV2 based segmentation was employed to test the effectiveness of S2 and L8 starting from a common and very good segmented image dataset.

The accuracy values achieved using very small training sets were very high for both sensors, also thanks to the modified ED2 index used to quantitatively assess the reliability of the best estimate segmentations. This can be considered as a further confirmation of the relevant contribution of the segmentation process in the final PCG detection. In this sense, another finding of this work was related to proving that Blue, Green and NIR bands are strongly related to the best segmentation of greenhouses for atmospherically corrected S2, L8 and WV2 data. Future research can benefit of this information to save computation time.

Overall, the final results showed that S2 performed better than L8, particularly when the best common segmentation from WV2 was used for both satellites.

The accuracy of the results obtained in this study makes this approach highly recommended for PCG mapping and detection. Further research will be focused both on the definition of new and more performant segmentation quality metrics and in the assessment of the best S2 and L8 features.

Acknowledgments

References

Agüera, F., Aguilar, F.J., Aguilar, M.A., 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. ISPRS Journal of Photogrammetry and Remote Sensing 63, 635-646.

Agüera, F., Aguilar, M., Aguilar, F., 2006. Detecting greenhouse changes from QB imagery on the Mediterranean Coast. Int. J. Remote Sens 27, 47514767.

Aguilar, M.A., Bianconi, F., Aguilar, F.J., Fernández, I., 2014. Object-based greenhouse classification from GeoEye-1 and WorldView-2 stereo imagery. Remote sensing 6, 3554-3582.

Aguilar, M.A., Vallario, A., Aguilar, F.J., Lorca, A.G., Parente, C., 2015. Object-Based greenhouse horticultural crop identification from multi-temporal satellite imagery: A case study in Almeria, Spain. Remote Sensing 7, 7378-7401.

Arcidiacono, C., Porto, S., 2010. Pixel-based classification of high-resolution satellite images for crop-shelter coverage recognition, XXVIII International Horticultural Congress on Science and Horticulture for People (IHC2010): International Symposium on 937, pp. 1003-1010.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. Herbert Wichmann Verlag: Berlin, Germany, pp. 12-23.

Berk, A., Bernstein, L., Anderson, G., Acharya, P., Robertson, D., Chetwynd, J., Adler-Golden, S., 1998. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. Remote Sensing of Environment 65, 367-375.

Breiman, L.,2001. Random Forests. Machine Learning 45, 5-32.

Briassoulis, D., Babou, E., Hiskakis, M., Scarascia, G., Picuno, P., Guarde, D., Dejean, C., 2013. Review, mapping and analysis of the agricultural plastic waste generation and consolidation in Europe. Waste Management & Research 31, 1262-1278.

Carvajal, F., Agüera, F., Aguilar, F., Aguilar, M., 2010. Relationship between atmospheric corrections and training-site strategy with respect to accuracy of greenhouse detection process from very high resolution imagery. International Journal of Remote Sensing 31, 2977-2994.

Chen, Z., Wang, L., Wu, W., Jiang, Z., Li, H., 2016. Monitoring Plastic-Mulched Farmland by Landsat-8 OLI Imagery Using Spectral and Textural Features. Remote Sensing 8, 353.

Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. Photogrammetric Engineering and remote sensing 76, 289-299.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment 37, 35-46.

Congalton, R.G., Green, K., 2008. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

Darwish, A., Leukert, K., Reinhardt, W., 2003. Image segmentation for the purpose of object-based classification, International Geoscience and Remote Sensing Symposium, pp. III: 2039-2041.

Definiens, A., 2009. Definiens eCognition Developer 8 Reference Book. Definiens AG, München.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning 40, 139-157.

Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS Journal of Photogrammetry and Remote Sensing 88, 119-127.

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 Sensing 8, 354.

Fernández-Manso, A., Fernández-Manso, O., Quintano, C., 2016. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. International Journal of Applied Earth Observation and Geoinformation 50, 170-175.

Frauman, E., Wolff, E., 2005. Segmentation of very high spatial resolution satellite images in urban areas for segments-based classification, Proceedings for 3rd International Symposium Remote Sensing and Data Fusion Over Urban Areas. Tempe, Arizona.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote sensing of Environment 80, 76-87.

Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on, 610-621.

Hill, T., Lewicki, P., 2007. STATISTICS Methods and Applications. StatSoft, Tulsa, USA.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with sentinel-2 data for crop and tree species classifications in Central Europe. Remote Sensing 8, 166.

Koc-San, D., 2013. Evaluation of different classification techniques for the detection of glass and plastic greenhouses from WorldView-2 satellite imagery. Journal of Applied Remote Sensing 7, 073553-073553.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. biometrics, 159-174.

Liu, D., Xia, F., 2010. Assessing object-based classification: advantages and limitations. Remote Sensing Letters 1, 187-194.

Liu, Y., Bian, L., Meng, Y., Wang, H., Zhang, S., Yang, Y., Shao, X., Wang, B., 2012. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. ISPRS Journal of Photogrammetry and Remote Sensing 68, 144-156.

Lu, L., Di, L., Ye, Y., 2014. A Decision-tree classifier for extracting transparent plastic-mulched landcover from Landsat-5 TM images. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 7, 4548-4558.

Lu, L., Hang, D., Di, L., 2015. Threshold model for detecting transparent plastic-mulched landcover using moderate-resolution imaging spectroradiometer time series data: a case study in southern Xinjiang, China. Journal of Applied Remote Sensing 9, 097094-097094.

Muller-Wilm, U., Louis, J., Richter, R., Gascon, F., Niezette, M., 2013. Sentinel-2 Level 2A Prototype Processor: Architecture, Algorithms And First Results, ESA Special Publication, p. 98.

Novelli, A., Tarantino, E., 2015a. Combining ad hoc spectral indices based on LANDSAT-8 OLI/TIRS sensor data for the detection of plastic cover vineyard. Remote Sensing Letters.

Novelli, A., Tarantino, E., 2015b. The contribution of Landsat 8 TIRS sensor data to the identification of plastic covered vineyards, pp. 95351E-95351E-95359.

Pala, E., Tasdemir, K., Koc-San, D., 2015. Unsupervised extraction of greenhouses using approximate spectral clustering ensemble, Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, pp. 4668-4671.

Pesaresi, M., Corbane, C., Julea, A., Florczyk, A.J., Syrris, V., Soille, P., 2016. Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. Remote Sensing 8, 299.

Picuno, P., Tortora, A., Capobianco, R.L., 2011. Analysis of plasticulture landscapes in Southern Italy through remote sensing and solid modelling techniques. Landscape and urban planning 100, 45-56.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67, 93-104.

Rouse Jr, J.W., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication 351, 309.

Roy, D.P., Wulder, M., Loveland, T., Woodcock, C., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment 145, 154-172.

Smith, A., 2010. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. Journal of Spatial Science 55, 69-79.

Tarantino, E., Figorito, B., 2012. Mapping rural areas with widespread plastic covered vineyards using true color aerial data. Remote Sensing 4, 1913-1928.

Tian, J., Chen, D.M., 2007. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. International Journal of Remote Sensing 28, 4625-4644.

Townshend, J.R., Justice, C.O., Gurney, C., McManus, J., 1992. The impact of misregistration on change detection. Geoscience and Remote Sensing, IEEE Transactions on 30, 1054-1060.

Wu, C.F., Deng, J.S., Wang, K., Ma, L.G., Tahmassebi, A.R.S., 2016. Object-based classification approach for greenhouse mapping using Landsat-8 imagery. International Journal of Agricultural and Biological Engineering 9, 79.

Yang, J., He, Y., Caspersen, J., Jones, T., 2015. A discrepancy measure for segmentation evaluation from the perspective of object recognition. ISPRS Journal of Photogrammetry and Remote Sensing 101, 186-192.

Zhang, J., Pu, R., Yuan, L., Wang, J., Huang, W., Yang, G., 2014. Monitoring powdery mildew of winter wheat by using moderate resolution multi-temporal satellite imagery. PloS one 9, e93107.