Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis

Mariana Belgiu[a,*], Ovidiu Csillik[b,c]

[a] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[b] Department of Geoinformatics – Z_GIS, University of Salzburg, Schillerstrasse 30, 5020 Salzburg, Austria
[c] Department of Environmental Sciences, Policy and Management, University of California, Berkeley, Berkeley, CA 94720-3114, United States

ARTICLE INFO

ABSTRACT

Efficient methodologies for mapping croplands are an essential condition for the implementation of sustainable agricultural practices and for monitoring crops periodically. The increasing spatial and temporal resolution of globally available satellite images, such as those provided by Sentinel-2, creates new possibilities for generating accurate datasets on available crop types, in ready-to-use vector data format. Existing solutions dedicated to cropland mapping, based on high resolution remote sensing data, are mainly focused on pixel-based analysis of time series data. This paper evaluates how a time-weighted dynamic time warping (TWDTW) method that uses Sentinel-2 time series performs when applied to pixel-based and object-based classifications of various crop types in three different study areas (in Romania, Italy and the USA). The classification outputs were compared to those produced by Random Forest (RF) for both pixel- and object-based image analysis units. The sensitivity of these two methods to the training samples was also evaluated. Object-based TWDTW outperformed pixel-based TWDTW in all three study areas, with overall accuracies ranging between 78.05% and 96.19%; it also proved to be more efficient in terms of computational time. TWDTW achieved comparable classification results to RF in Romania and Italy, but RF achieved better results in the USA, where the classified crops present high intra-class spectral variability. Additionally, TWDTW proved to be less sensitive in relation to the training samples. This is an important asset in areas where inputs for training samples are limited.

1. Introduction

World population is expected to increase from 7.3 billion to 8.7 billion by 2030, 9.7 billion by 2050, and 11.2 billion by 2100 (United-Nations, 2015a). This population growth impacts food supply systems worldwide (Waldner et al., 2015), making urgent the development of sustainable natural resources management programs. The increasing role of agriculture in the management of sustainable natural resources calls for the development of operational cropland mapping and monitoring methodologies (Matton et al., 2015). The availability of such methodologies represents a prerequisite for realising the United Nations (UN) Sustainable Development Goals, including no poverty and zero hunger (United-Nations, 2015b). The Agricultural Monitoring Community of Practice of the Group on Earth Observations (GEO), with its Integrated Global Observing Strategy (IGOL), also calls for an operational system for monitoring global agriculture using remote sensing images.

There are various studies, using supervised or unsupervised algorithms, dedicated to cropland mapping from time series or single-date remote sensing images (Petitjean et al., 2012b; Xiong et al., 2017; Yan and Roy, 2015). The cropland mapping methods applied to time series images have proven to perform better than single-date mapping methods (Gómez et al., 2016; Long et al., 2013). For example, the phenological patterns identified using 250 m MODIS-Terra/Enhanced Vegetation Index (EVI) time series have been successfully used to classify soybean, maize, cotton and non-commercial crops in Brazil (Arvor et al., 2011). Patterns of vegetation dynamics identified from MODIS EVI data were used by Maus et al. (2016) to map double cropping, single cropping, forest and pasture. Senf et al. (2015) used multi-seasonal MODIS and Landsat imagery to differentiate crops from savannah, and Müller et al. (2015) successfully classified cropland and pasture fields from Landsat time series. Jia et al. (2014) investigated the efficacy of phenological features computed from the MODIS Normalized Difference Vegetation Index (NDVI) time series fused with NDVI data derived from Landsat 8 for land-cover mapping. This study confirmed that phenological features, including the maximum, minimum, mean and standard deviation values computed from the fused NDVI data, are relevant for classifying various vegetation categories such as
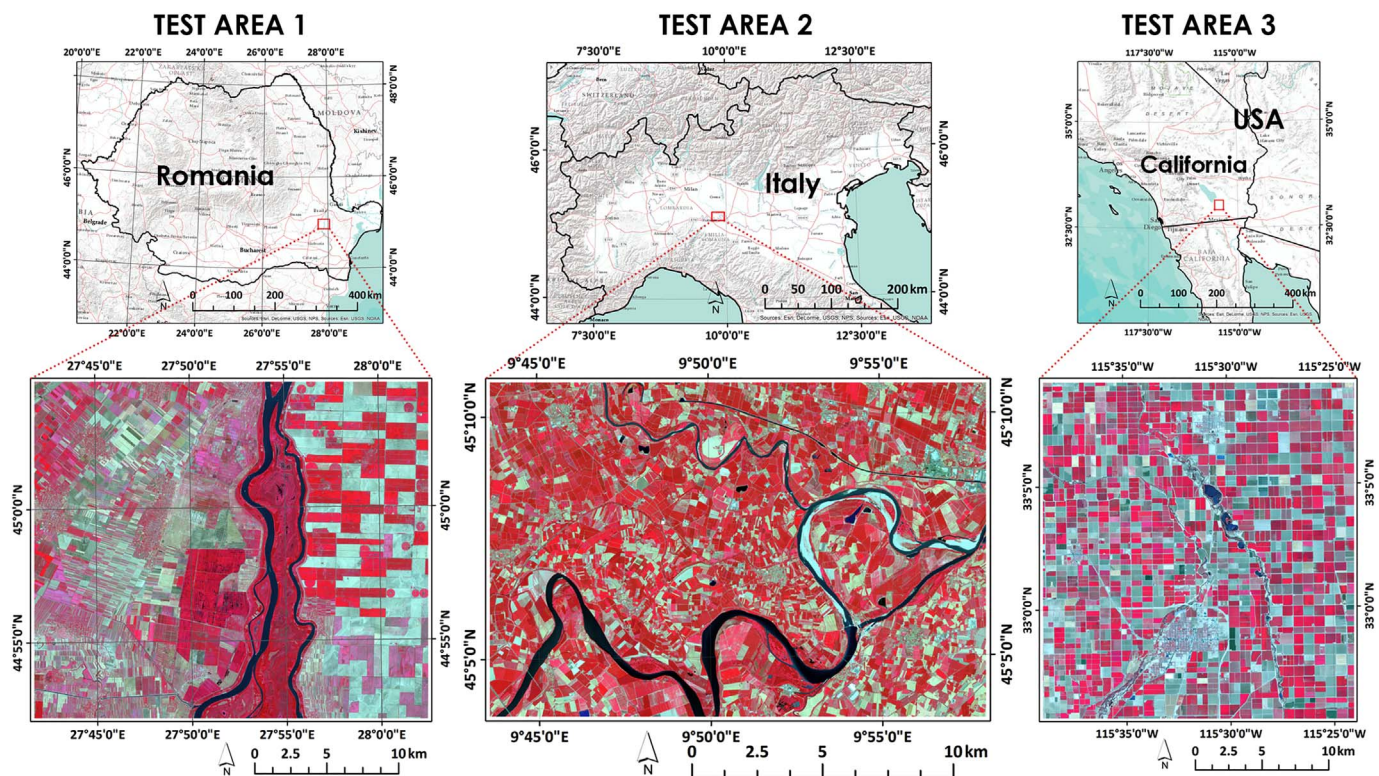
Fig. 1. Study areas located in Romania (Test Area 1 – TA1), Italy (Test Area 2 – TA2) and California, USA (Test Area 3 – TA3). On the lower part, the three subsets are illustrated using a false color combination (near-infrared, red, green bands) for the dates of August 6th (TA1), July 18th (TA2) and June 15th 2016 (TA3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

forest, grass and crop classes.

The above-mentioned studies focused on the classification of multi-temporal images at pixel level. Petitjean et al. (2012b) argue that the increasing spatial resolution of available space-borne sensors, including the Multispectral Imager sensor (MSI) carried on Sentinel-2, creates the possibility of applying object-based image analysis (OBIA) to extract crop types from multi-series data. OBIA is an iterative method that starts with the segmentation of satellite imagery into homogeneous and contiguous image segments (also called image objects) (Blaschke, 2010). The resulting image objects are then assigned to the target classes using supervised or unsupervised classification strategies. Lebourgeois et al. (2017) used OBIA for the mapping of smallholder agriculture in Madagascar, using pan-sharpened Pléiades images (0.5 m spatial resolution) for delineating the agricultural fields. The fields were then further classified using reflectance and spectral indices computed from Sentinel-2 (artificially created images from SPOT-5 images and Landsat 8 OLI/TIRS images) and from Digital Elevation Model (DEM)-based ancillary information, such as altitude or slope. Novelli et al. (2016) used single-date Sentinel-2 and Landsat 8 data to classify greenhouse segments from WorldView-2 images. Long et al. (2013) described a multi-temporal object-based approach for mapping cereal crops from Landsat SLC-off ETM + imagery (without using gap-filling schemes); they concluded that this approach allowed the accurate classification of crops located partially within data gaps and avoided the salt-and-pepper effect specific to a pixel-based approach, which usually leads to mixed-crop classifications of fields. Castillejo-González et al. (2009) showed that object-based analysis outperformed pixel-based analysis for cropland mapping that uses QuickBird images. Matton et al. (2015) and Li et al. (2015) also reported the advantages of using object-based methods to classify various crop types from SPOT-Landsat 8 time series and Landsat-MODIS data, respectively.

Despite the reported advantages of object-based methods to delineate agricultural parcels (Castillejo-González et al., 2009; Matton et al., 2015; Petitjean et al., 2012b), the implementation of object-based

frameworks for the spatio-temporal analysis of high-resolution satellite imagery such as Sentinel-2 time series in order to delineate and classify agricultural fields is very limited (Yan and Roy, 2016).

According to Petitjean et al. (2012a), cropland mapping based on time series analysis is challenged by (i) the lack of samples used to train the supervised algorithm; (ii) missing temporal data caused by clouds obscuration; (iii) annual changes of phenological cycles caused by weather or by variations in the agricultural practices. Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978) proved to be an efficient solution to handle these challenges (Baumann et al., 2017; Petitjean et al., 2012a). Maus et al. (2016) proposed a time-weighted version of the DTW method (TWDTW) able to classify crops with various vegetation dynamics. The TWDTW analysis performed well in identifying single cropping, double cropping, forest and pasture from EVI derived from MODIS data (Maus et al., 2016). The method has, however, been applied only at pixel level. Given the fact that agricultural management decisions are generally made at the level of agricultural parcels (Forster et al., 2010; Long et al., 2013), and given the increasing spatial resolution of Sentinel-2 data, it is worth investigating how this method performs when applied within an OBIA framework.

The overall objective of this study is to evaluate the performance of the TWDTW method for cropland mapping, based on freely available Sentinel-2 time series and using pixels and objects as spatial analysis units. Specifically, the performance of the TWDTW method is evaluated in three different agrosystem regions, in Romania, Italy and the USA. Subsequently, the classification results produced by TWDTW are compared with those achieved by a Random Forest (RF) classifier (Breiman, 2001). RF was successfully used in previous remote sensing studies dedicated to land use/land cover mapping from time series (Pelletier et al., 2016).

This paper is structured as follows: Section 2 describes the study areas and the data; Section 3 presents the proposed methodology; Section 4 is dedicated to the results; Section 5 highlights the main findings and the implications of this study, and is followed by our

Table 1
Characteristics of the three test areas: location, extent, climate and soils.

Test area	TA1	TA2	TA3
Country	Romania	Italy	California, USA
Lat/Long	45°00′N	45°10′N	33.05′N
	27°90′E	9°85′E	− 115°50′W
Extent (pixels)	2738 × 2333	1898 × 1278	2336 × 2521
Extent (ha)	63,877	24,256	58,890
Climate	Temperate continental	Mediterranean	Hot desert
Average annual temperature (°C) – 2016[a]	max. 19.5 mean 16.9 min. 13.2	max. 20.8 mean 17.7 min. 13.1	max. 32.0 mean 27.5 min. 19.4
Precipitation (mm) – 2016[a]	615.4	1078.2	49.3
Soils[b]	Calcic Chernozems and Calcaric Fluvisols	Eutric Cambisols and Orthic Luvisols	Calcaric Fluvisols with associated Solonchaks

[a] Data provided by WorldWeatherOnline.com

[b] Soil information extracted from FAO-UNESCO (1981) (TA1 and TA2) and FAO-UNESCO (1975) (TA3).

Conclusion.

2. Study areas and data

2.1. Study areas

The TWDTW method was applied to three study areas from diverse climate regions and with different crops, field geometries, cropping calendars and background soils (Fig. 1).

The first test area (TA1) is situated in south-eastern Brăila County, Romania, at 45°00′N, 27°90′E (Fig. 1). TA1 covers 63,877 ha (Table 1) and is situated in the Bărăgan Plain, one of the most fertile and productive agricultural areas in the country. The agricultural landscape is very fragmented in the western part of TA1, while in the eastern part the parcels are larger and more compact, making them suitable for intensive agricultural production. Situated in a steppe area with a temperate continental climate, this region is well known for cold winters and hot summers, which make it one of the most inhospitable areas in Romania (for the 2016, the annual mean temperature was 16.9 °C, with great variations in temperature, and precipitation of around 615 mm/year) (Table 1). However, the fertile soils represented by Calcic Chernozems and Calcaric Fluvisols (FAO-UNESCO, 1981) make this region attractive for various crops, including wheat (common wheat and durum wheat), maize, rice and sunflowers.

The second test area (TA2) is situated in southern Lombardy and northern Emilia-Romagna (Italy), at 45°10′N, 9°85′E (Fig. 1); it covers 24,256 ha (Table 1), and the major soils are Eutric Cambisols and Orthic Luvisols (FAO-UNESCO, 1981). The climate is classified as Mediterranean, with precipitation of 1078.2 mm/year, an average annual maximum temperature of 20.8 °C, and average annual minimum temperature of 13.1 °C for 2016. The major crops in the region are maize, permanent and temporary meadow for forage, and winter cereals (e.g. winter wheat, winter barley); double cropping is practised (i.e. winter crops followed by maize, or winter crops followed by forage) (Azar et al., 2016).

Test area 3 (TA3) is located in central Imperial County, southern California, at 33°05′N and − 115°50′W (Fig. 1). Covering 58,890 ha with low altitudes in the Colorado Desert, the area is characterized by high agricultural productivity and a hot desert climate (Table 1). Within California, Imperial County is one of the highest-producing counties for hay (alfalfa, over 15%), onions and lettuce (CDFA, 2016). The agriculture is heavily reliant on irrigation: for the year 2016, the average precipitation was below 50 mm, and the annual mean temperature was above 27 °C (Table 1). The predominant soils are Calcaric Fluvisols with associated Solonchaks, with a hyperthermic soil temperature regime; the area is intensively managed for the production of irrigated crops (FAO-UNESCO, 1975). Seven land use/land cover classes were selected for analysis, each of them occupying at least 2% of the study area: durum wheat, alfalfa, other hay/non-alfalfa, sugar beets, onions, fallow/idle cropland and lettuce.

2.2. Data collection and pre-processing

The available Sentinel-2 satellite images (Level-1C S2) for the three test areas were downloaded from the European Space Agency's (ESA) Sentinel Scientific Data Hub (Fig. 2). Using sen2cor plugin v2.3.1 (Muller-Wilm et al., 2013), available on the Sentinel Application Platform (SNAP) v5.0 and distributed under the GNU GPL license, we processed reflectance images from Top-Of-Atmosphere (TOA) Level 1C S2, to Bottom-Of-Atmosphere (BOA) Level 2A.

We selected temporal images with zero or close to zero (< 10%) cloud coverage, taken between February and September for TA1 (13 images), January and October for TA2 (13 images), and January and December 2016 for TA3 (21 images) (Fig. 2). Only visible bands 2, 3 and 4 and the near-infrared band (band 8) of Sentinel-2 data were used. The images from the three test areas were projected in WGS 1984 UTM Zone 35 N (TA1), Zone 32 N (TA2), and Zone 11 N (TA3).

3. Methodology

The methodological workflow consists of the following steps: (1) data pre-processing to create BOA-corrected reflectance images (Level 2A) from TOA (Level 1C) input data (see section 2.2 for more details); (2) data processing, which involves the creation of the NDVI time series, the generation of training and validation samples, and multi-temporal image segmentation; (3) data classification, where both TWDTW and RF are used to map the target classes; (4) evaluation, for assessing the classification accuracies obtained by pixel-based TWDTW (PB-TWDTW) and object-based TWDTW (OB-TWDTW), and for comparison of the results with those obtained by RF (Fig. 3).
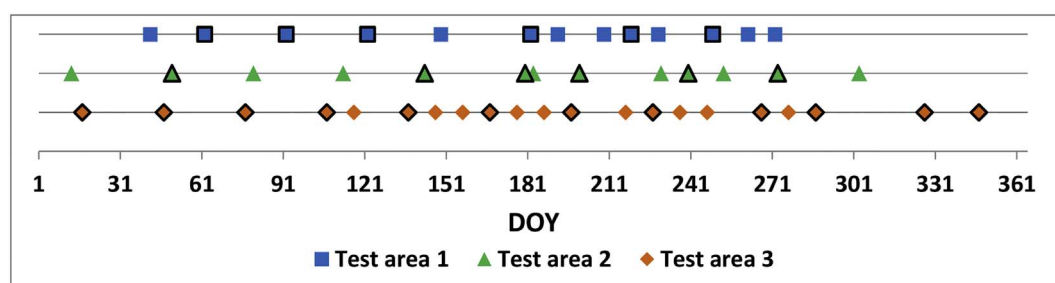
Fig. 2. Temporal coverage of the three Sentinel-2 time series data used in this study: TA1 (Romania), TA2 (Italy) and TA3 (California). The black-outlined symbols indicate the images used in the segmentation process.
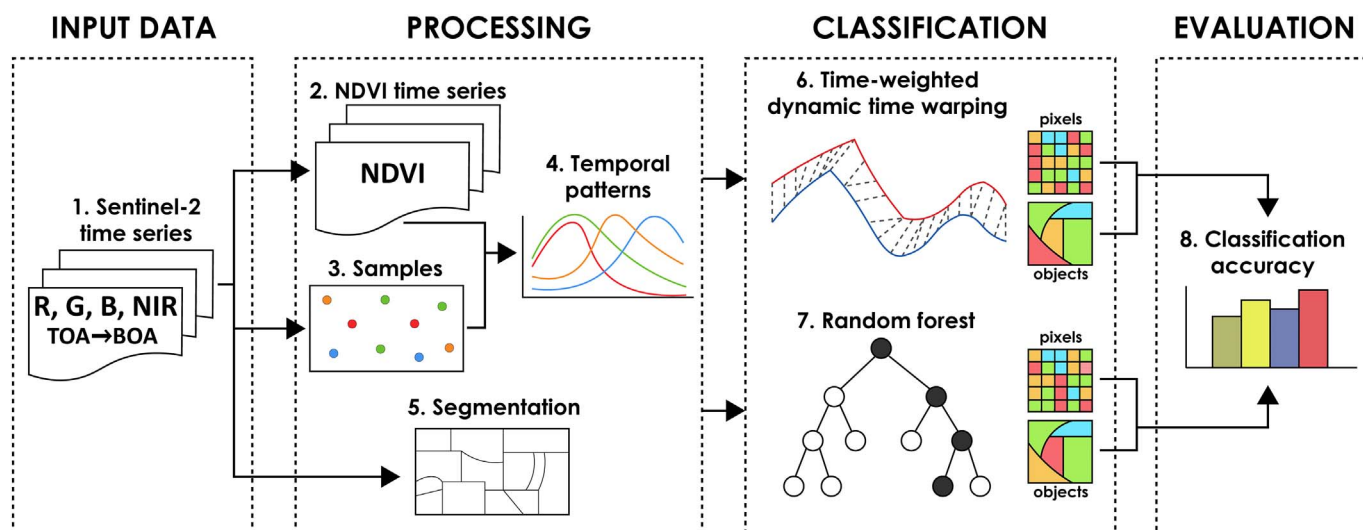
Fig. 3. The workflow of pixel-based and object-based TWDTW and RF approaches.

3.1. Training and validation samples

For TA1 and TA2, we randomly generated samples across the sites and classified them using visual interpretation keys based on expert knowledge and information about crop types from the European Land Use and Coverage Area Frame Survey (LUCAS). LUCAS is an in-situ data survey carried out by EUROSTAT to identify changes in land use and cover across the European Union. For TA3, we made use of CropScape – Cropland Data Layer (CDL) for 2016, provided by the United States Department of Agriculture, National Agricultural Statistics Service (Boryan et al., 2011; Han et al., 2012), in order to classify randomly generated samples. In all three test areas, the reference samples were split into two sets of disjoint training and validation samples. This was done at the polygon level to guarantee that training samples and validation samples were located in different agricultural parcels (Table 2).

3.2. Temporal phenological patterns

The temporal phenological patterns of the target classes were computed using the NDVI (Tucker, 1979) generated from 10 m resolution red and near-infrared spectral bands:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Table 2
Number of training and validation samples, for each class investigated.

Test area 1			Test area 2			Test area 3		
Class	Train.	Valid.	Class	Train.	Valid.	Class	Train.	Valid.
Wheat	38	89	Double cropping	59	124	Durum wheat	26	60
Maize	32	73	Forage	45	80	Alfalfa	69	160
Rice	24	57	Forest	41	102	Other hay/ Non alfalfa	50	116
Sunflower	30	71	Maize	42	147	Sugar beets	38	89
Forest	32	73	Water	33	68	Onions	21	49
Water	25	57	Winter cereals	31	84	Fallow/ Idle cropland	33	76
						Lettuce	24	56
Total	181	420	Total	251	605	Total	261	606

NDVI is one of the most-used vegetation indices for studying vegetation phenology (Yan and Roy, 2014); it reduces the spectral noise caused by certain illumination conditions, topographic variations or cloud shadows (Huete et al., 2002). Additional vegetation indices or spectral bands (such as red-edge spectral bands) are not considered in this study.

The crops available in the three test areas have different temporal patterns because of the diverse climate conditions and agricultural practices (Fig. 4). For TA1, the wheat crop has its period of maximum growth in May through June, and is harvested at the end of June (Croitoru et al., 2012; Sima et al., 2015). Maize emerges in May, has the highest values of NDVI from late June to late August, and is harvested in September. Rice is partially covered by water in June, thus lowering the NDVI values, and reaches peak NDVI values in late August/early September. Sunflower reaches its peak growing period in July and is harvested from early August. The last class investigated, namely the (deciduous) forest class, has green vegetation from late April through to late September. The decrease in NDVI values for maize and forest in July is due to this being a very dry month, with precipitation below 5 mm (Fig. 4).

In TA2, the winter cereals reach the peak of their vegetative phase in April–May, and is harvested in May–June; the maize reaches its highest vegetative phase in July, and is harvested between the end of August and late September; the forage crops (rye grasses, alfalfa or clovers) are harvested several times per year (Azar et al., 2016). The double-cropping class consists of the winter crops and maize, and of winter crops and the 'other crops' class. As the phenological dynamics of these two variations on double cropping are similar (Fig. 4), we decided to merge them in our classification approach.

Because TA3 comprises intensively managed agricultural parcels for the production of irrigated cash crops, the crops in TA3 show complex temporal patterns. Durum wheat and onions have similar temporal patterns, with the highest NDVI values in March, followed by reduced greenness for the rest of the year. Alfalfa and other hay (non-alfalfa) have the most irregular temporal patterns, since these crops are produced and harvested periodically throughout the whole year. However, increased greenness is seen in the first half of the year. Lettuce have two greenness peaks, in April and June–July, followed by bare soil afterwards. Sugar beet reaches its peak greenness in the first three months of the year, with low NDVI values after that, until the end of October, when the values start to increase again. The most individualized class is the fallow/idle cropland, with values of NDVI below 0.2 throughout the entire year. Fig. 4 illustrates the phenological patterns of the target crops identified in the Sentinel-2 time series for 2016.
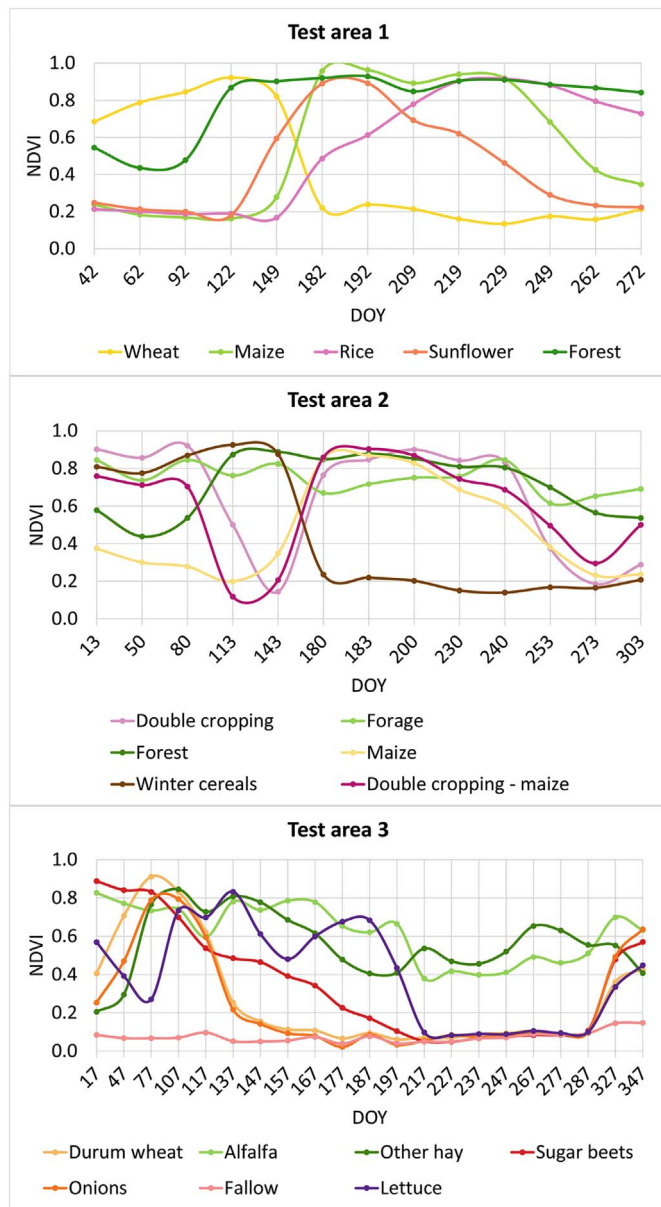
Fig. 4. Temporal patterns of NDVI for all classes mapped in the three test areas.

To avoid time-consuming trial-and-error and subjective selection of the SP (Drăguţ et al., 2010), we used an automated tool for assisting the segmentation procedure, namely Estimation of Scale Parameter 2 (ESP2) (Drăguţ et al., 2014). ESP2 relies on the concept of local variance across scales to automatically identify three suitable SPs for MRS. We used a hierarchical segmentation approach, meaning that each new level of segmentation is based on the previous level. In the end, the finest level of the hierarchy was chosen because it better reflected the boundaries within the site, avoiding an undesirable high degree of under-segmentation. For segmentation purposes, we used the red, green, blue and near-infrared spectral bands for 6 images of TA1, 6 images of TA2, and 12 images of TA3 (Fig. 2). The images selected for segmentation are distributed across the entire agricultural calendar, resulting in stacks of 24, 24 and 48 layers for TAs 1, 2 and 3 respectively, to be used as input in the segmentation process for the test areas. Finally, a number of raster datasets were exported, representing the mean NDVI values for each object, for each temporal image used (13 for TA1 and TA2; 21 for TA3). The resulting raster datasets were then stacked for each test area, ready to be analysed using the TWDTW method.

3.4. Time-weighted dynamic time warping

We used the TWDTW method as implemented in the R (v3.3.3) package *dtwSat* v0.2.1 (Maus et al., 2016). The method consists of three main steps: (1) generating the temporal patterns of the ground truth samples based on the NDVI time series; (2) applying the TWDTW analysis; (3) classifying the raster time series.

3.4.1. Retrieval of phenological stages

We used the pixel-based stack of NDVI datasets extracted from Sentinel-2 imagery and the training samples (Table 2) to extract their average NDVI signal for the complete agricultural calendars studied. We defined the temporal patterns using the function dtwSat::createPatterns (Maus, 2016), which uses a Generalized Additive Model (GAM) (Wood, 2011) to create a smoothed temporal pattern for the NDVI. The sampling frequency of the output patterns was set to 8 days, in order to create a smoothed line. The resulting temporal patterns were further used in the PB- and OB-TWDTW classifications.

3.4.2. TWDTW classifications

The Dynamic Time Warping (DTW) method was originally developed for speech recognition (Sakoe and Chiba, 1978) and later introduced in the analysis of time series images (Maus et al., 2016; Petitjean et al., 2012a; Petitjean and Weber, 2014). DTW works by comparing the similarity between two temporal sequences and finds their optimal alignment, resulting in a dissimilarity measure (Rabiner and Juang, 1993; Sakoe and Chiba, 1978). In the case of remote sensing data, DTW can deal with temporal distortions, and can compare shifted evolution profiles and irregular sampling thanks to its ability to align radiometric profiles in an optimal manner (Petitjean et al., 2012a).

To distinguish between different types of land use and land cover, Maus et al. (2016) introduced a time constraint to DTW that accounts for seasonality of land cover types, improving the classification accuracy when compared to traditional DTW. They implemented their method, Time-Weighted Dynamic Time Warping (TWDTW), into an open-source R package, *dtwSat* (Maus, 2016). We applied a logistic TWDTW, since it has been shown that this gives more accurate results than a linear TWDTW: logistic TWDTW has a low penalty for small time warps and significant costs for large time warps, which is different from the linear approach, which has large costs for small time warps, reducing the sensitivity of classification (Maus et al., 2016). As recommended in Maus et al. (2016), we used the function dtwSat::twdtwApply with α = − 0.1 and β = 50, which means that we added a time-weight to the DTW, with a low penalty for time warps smaller than 50 days and a higher penalty for larger time warps. Different values of α

Note that these patterns might vary from year to year because of changes in agricultural practices and/or varying meteorological conditions (Petitjean et al., 2012a); these changes might impact the "canonical temporal profiles" (Petitjean et al., 2012a) of the target crops, challenging most of the existing image classification methods, which are usually unable to deal with irregular temporal phenological signatures (Maus et al., 2016).

3.3. Segmentation of multi-temporal images

Sentinel-2 images were segmented into homogeneous objects using one of the most popular segmentation algorithms in OBIA, namely multi-resolution segmentation (MRS) (Baatz and Schäpe, 2000), implemented in the eCognition Developer (v9.2.1, Trimble Geospatial). MRS is a region-growing algorithm that starts from the pixel level and iteratively aggregates pixels into objects until some conditions of homogeneity imposed by the user are met. MRS relies on several parameters, which need to be tuned. These include the scale parameter (SP), which dictates the size and homogeneity of the resultant objects.
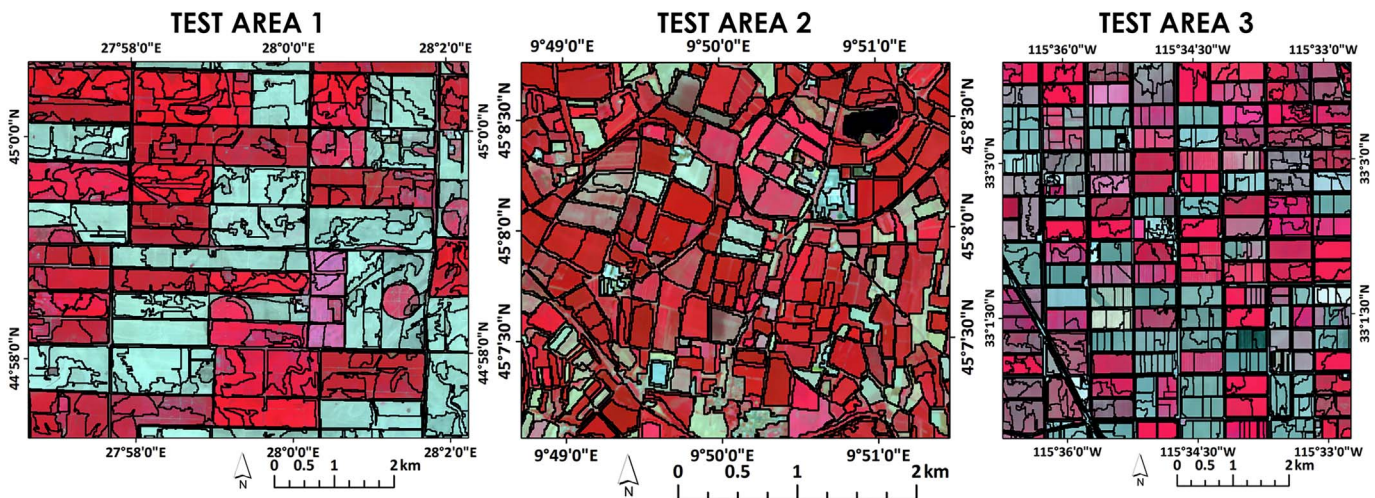
Fig. 5. Subsets of the three test areas segmented into 5198 (TA1), 6044 (TA2) and 12,543 (TA3) objects using MRS and ESP2. The objects boundaries are depicted in black, overlaid on a false color combination (NIR, R, G). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(α = − 0.2, − 0.1, 0.1 or 0.2) and β (β = 50 or 100) were tested for smaller subsets of the test areas, but the best classification results were obtained with the α = − 0.1 and β = 50 parameters. For the logistic weights, α and β represent the steepness and the midpoint, respectively. The α and β values are highly important when analysing time series from different years and when the phenological cycles of crops differ from one season to another. The dtwSat::twdtwApply function takes each pixel location in the NDVI time series and analyses it in conjunction with the temporal patterns identified for training samples. The output is a dtwSat::twdtwRaster with layers containing the dissimilarity measure for each of the temporal patterns (Maus, 2016). In the final step, the function dtwSat::twdtwClassify generates the categorical land cover map, based on the previously obtained dissimilarities, by assigning each pixel to the class with the lowest dissimilarity value.

3.5. Random Forest

RF is an ensemble learning classifier (Breiman, 2001) that has achieved efficient classification results in various remote sensing studies, including cropland mapping (Hao et al., 2015; Li et al., 2015; Novelli et al., 2016; Pelletier et al., 2016). A detailed review of RF and its efficiency in remote sensing can be found in Belgiu and Drăguţ (2016) and Gislason et al. (2006). In our study, RF was run using the *randomForest* (v.4.6–12) R package-based script published by Millard and Richardson (2015). Two parameters need to be tuned for RF, namely the number of trees which will be created by randomly selecting samples from the training samples (*ntree* parameter), and the number of variables used for tree nodes splitting (*mtry* parameter). In all three study areas, the *ntree* parameter was set up at 1000, because previous studies proved that there is no increase in the number of errors beyond the creation of 1000 classification trees from randomly selected samples (Lawrence et al., 2006). Following the RF classification results reported in previous studies and reviewed in Belgiu and Drăguţ (2016) and Gislason et al. (2006), the *mtry* parameter was established as the square root of the number of the available layers. Only the NDVI values computed from multi-temporal images were used as input variables in the RF classifier.

3.6. Classification accuracy assessment

The accuracies of the pixel- and object-based classifications obtained were evaluated in terms of overall accuracy, producer's accuracy, user's accuracy metrics (Congalton, 1991), and kappa coefficient (Cohen, 1960). The validation samples available for all three study

areas are shown in Table 2. The differences between the classification results obtained by TWDTW and RF were evaluated using McNemar's test (Bradley, 1968).

4. Results

4.1. Computational resources for applying TWDTW analysis

One of the biggest challenges in applying TWDTW is the computational time required. For a speedy application of the algorithm, the NDVI stacks were split into 9 tiles for TA1, 6 for TA2, and 9 for TA3, which were processed in a parallelized approach. NDVI stack-splitting did not influence the classification results, because each pixel is treated independently and additional information such as topological relations is not required. TWDTW analysis was run on a configuration using 3 out of 4 cores, with 3.20 GHz and 16-GB memory. The processing time varied between 25 h for TA1, 9 h for TA2, and 30 h for TA3 for the PB-TWDTW, and 3 min for the OB-TWDTW. The computational time depends on the number of layers in the NDVI stack and the number of classes being investigated. RF classifications required about 1 h for 25 iterations (for both pixel- and object-based RF). For segmentation purposes, we used the same configuration used for the TWDTW analysis (see above). The processing time varied between 44 min for TA1 (using 6 temporal images with 4 spectral bands each), 16 min for TA2 (using also 6 temporal images with 4 spectral bands each) and 2h34m for TA3 (using 12 temporal images with 4 spectral bands each).

4.2. Segmentation results

For the segmentation task, the finest level of a hierarchical MRS segmentation using ESP2 was chosen. The MRS algorithm was run using a shape parameter of 0.1 and a compactness parameter of 0.5 for all three areas. For TA1, ESP2 identified a SP of 205, resulting in 5198 objects with a mean area of 1229 pixels of 10 m resolution (Fig. 5). For TA2, the final SP was 123, resulting in 6044 objects with a mean area of 401 pixels (Fig. 5). For TA3, we obtained 12,543 objects with a SP of 129 (Fig. 5). In TA1 and TA3, agricultural fields are over-segmented because of the high spectral variability of individual crop parcels, whereas the segmentation results obtained in TA2 follow the field boundaries. Multiple temporal images well distributed across the agricultural calendar need to be used as input in the segmentation process in order to extract field boundaries that may appear on a certain date, due to crop management, irrigation practices or weather influences (Fig. 6). For example, using 6 out of 13 temporal images for TA1 in the
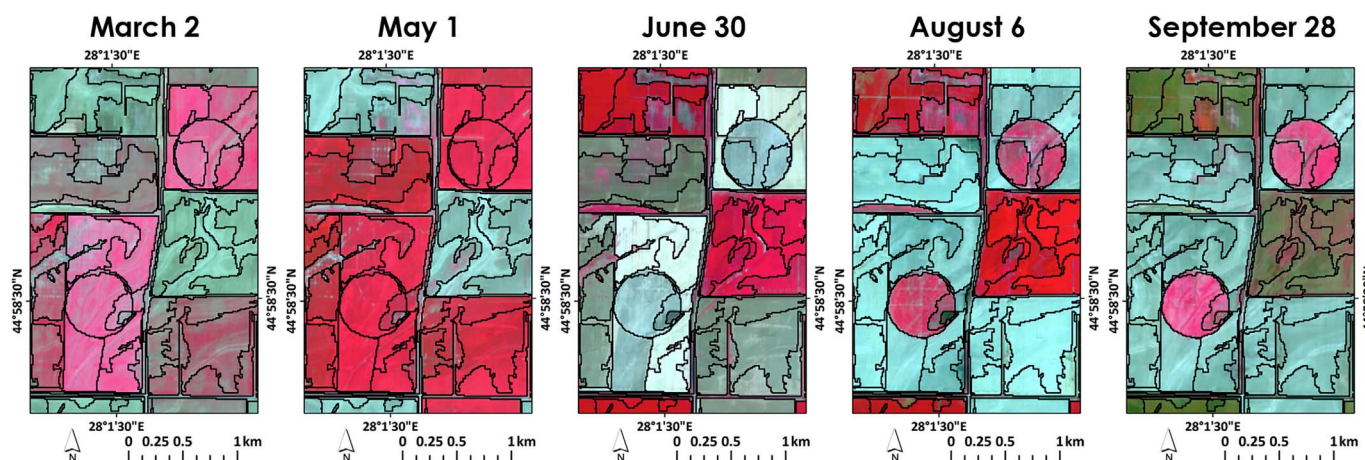
March 2   May 1   June 30   August 6   September 28

Fig. 6. Using multiple temporal images in the same segmentation process is useful in depicting the boundaries that may appear along the agricultural calendar. In this example, the irrigated circle parcels are extracted in the final segmentation (black lines), although they appear in the landscape at the end of June. The images are shown in false color composite (NIR, R, G). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

segmentation process ensured the extraction of the irrigated circle parcels, which appeared in the landscape at the end of June (Fig. 6).

4.3. Classification results

4.3.1. Overall accuracies of TWDTW

TWDTW achieved good classification results in all three test areas (Table 3 and Fig.7), the highest overall accuracy being obtained in TA1 (96.19% for OB-TWDTW; 94.76% for PB-TWDTW), followed by TA2 (89.75% for OB-TWDTW; 87.11% for PB-TWDTW). In TA3, OB-TWDTW achieved an accuracy of 78.05%, whereas PB-TWDTW obtained 74.92%. OB-TWDTW performed better than PB-TWDTW in all test areas (Table 3).

4.3.2. TWDTW and RF classification comparisons using McNemar's test

The differences between the classification outcomes achieved by TWDTW and RF were assessed using the McNemar's Chi-squared test (Table 4). In TA1, PB-RF yielded comparable results to OB-TWDTW and OB-RF, while PB-TWDTW achieved statistically different results. The performance of PB-TWDTW was similar to that of OB-RF and PB-RF in TA2, while the classification outputs obtained by OB-TWDTW were statistically different from those produced by PB-TWDTW and OB-RF. In TA3, all classifications are statistically different, except for the pair PB-RF and OB-RF.

4.3.3. User's and Producer's accuracies

In the case of TA1, the UA and PA for wheat, water and forest achieved the highest values (100%, or slightly lower (98.65% for forest)), for both PB and OB-TWDTW. For the sunflower class, the highest UA was reached by PB-TWDTW (98.41%). However, the same approach produced the lowest PA for sunflower, 87.32%, due to confusion with the maize class. For the rice class, both OB approaches achieved higher UA than PB classifications (a difference of 5.35% for

TWDTW and of 6.01% for RF). The lower accuracy of the PB classifications was due to confusion with the maize class. Although the temporal patterns of the two classes are distinct (the rice class having a temporally-shifted high peak in NDVI), the internal variability of the temporal pattern for the rice class is high. Some rice parcels have an earlier growing pattern, showing a closer resemblance to the maize pattern. OB classifications efficiently addressed the intra-class variability problem specific to rice crops in TA1.

The PA for the rice class was higher for both RF approaches than for TWDTW, with a maximum difference of 12.28% between PB-RF and PB-TWDTW, as a result of reduced confusion with the maize class. The lowest UA value was obtained by the PB-TWDTW approach for maize (82.72%), because of confusion with the rice and sunflower classes. The temporal patterns of these three classes have overlapping phases, with sunflower and maize sharing a similar pattern for the first half of the period investigated, while maize and rice are similar at the beginning and in the late-middle part of the agricultural calendar (Fig. 4). However, in the case of the PA for maize, both PB- and OB-TWDTW have the same value (91.78%), which is superior to the PB-RF approach (89.04%), but lower than the OB-RF classification (94.52%) (Fig. 8).

In TA2, TWDTW achieved the highest UA for the water class, for both PB and OB classifications (100%), followed by winter cereals with a UA of 89.36% for PB-TWDTW and 95.45% for OB-TWDTW (Fig. 9). Thus, the OB-TWDTW reduced the confusion between winter cereals on the one hand and cereals found in the double-cropping and forage classes, which occurred in the case of PB-TWDTW.

The maize class was also well classified: UA of 87.01% for PB-TWDTW, and 89.26% for OB-TWDTW. Similar results were obtained for the double-cropping class (UA of 87.5% for PB-TWDTW; 88.28% for OB-TWDTW). The forage class achieved the lowest UA, namely 75.86% for PB-TWDTW and 83.08% for OB-TWDTW, because of confusion with the forest class.

Generally, OB-TWDTW performed better than PB-TWDTW for all classes in TA2, the highest difference being for the forage class, where OB-TWDTW performed much better than PB-TWDTW (about 7% difference obtained for UA, and 12.5% difference for PA). OB-TWDTW classified the winter cereals class with a higher confidence than PB-TWDTW (about 5% higher value for the UA metric) (Fig. 9).

TWDTW outperformed RF in identifying the double-cropping class (for both PB and OB approaches), achieving a PA of 90.32% for PB-TWDTW and 91.13% for OB-TWDTW, whereas RF yielded PAs of 80.65% and 76.61% for PB and OB respectively. Furthermore, TWDTW also achieved better classification results for the winter cereals. RF on the other hand performed better in classifying the forage class. RF also obtained the highest PA for the maize class (95.24% for PB-RF).

Table 3
Overall accuracy (OA) and kappa coefficient of TWDTW and RF applied to all three test areas using both pixels and objects as the image analysis units.

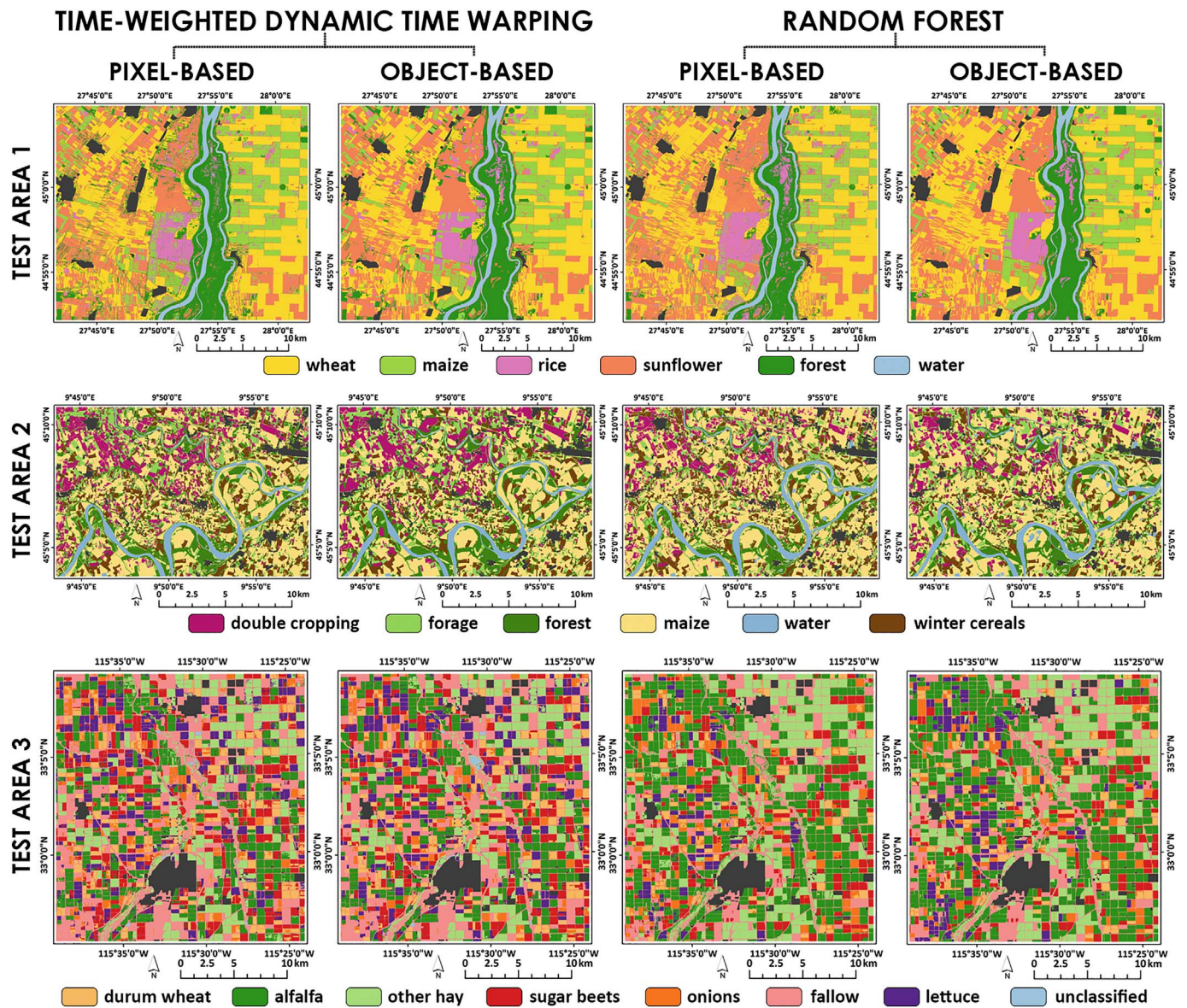|	TA1		TA2		TA3	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
PB-TWDTW	94.76	0.93	87.11	0.84	74.92	0.70
OB-TWDTW	96.19	0.95	89.75	0.87	78.05	0.74
PB-RF	97.14	0.97	87.44	0.84	88.78	0.87
OB-RF	97.62	0.97	86.28	0.84	88.28	0.86

Fig. 7. Pixel-based and object-based categorical maps of crops using TWDTW and RF. The grey values represent the settlement areas.

However, the UA obtained for this class by RF is much lower than that obtained by TWDTW.

In TA3, alfalfa and the other hay class achieved better UA for both TWDTW approaches, when compared to RF. In the case of alfalfa, the biggest UA difference, 13.26%, was between OB-TWDTW and OB-RF, as greater confusion occurred with sugar beet, onion and lettuce in the case of the latter approach (Fig. 10). On the other hand, durum wheat and lettuce achieved the lowest UA for TWDTW, of 51.35% for lettuce

using PB-TWDTW, and 57.45% for durum wheat using OB-TWDTW. In the case of wheat, there were confusions with sugar beets and onions, while lettuce was confused with alfalfa and the other hay class. In the case of PB- and OB-TWDTW, the PA values for durum wheat were higher than in the corresponding RF approach. For alfalfa and other hay, the PA values for the TWDTW classification are lower than for RF, because of confusions with sugar beets and lettuce. The PB-RF classification obtained the highest PA for sugar beets (88.76%) and onions

Table 4
Summary of the classification comparisons using McNemar's Chi-squared test.

		TA1		TA2		TA3	
		χ^2	p	χ^2	p	χ^2	p
PB-TWDTW	OB-TWDTW	4.167	0.041	11.25	< 0.001	17.053	< 0.001
	PB-RF	5.786	0.0162	0.023	0.880	71.76	< 0.001
	OB-RF	10.083	0.002	0.34	0.56	79.012	< 0.001
OB-TWDTW	PB-RF	1.125	0.289	3.841	0.05	47.08	< 0.001
	OB-RF	4.167	0.041	8.511	0.004	58.141	< 0.001
PB-RF	OB-RF	0.167	0.683	1.565	0.211	0.129	0.719

Fig. 8. User's and Producer's accuracy for TA1 represented together with the confidence interval (95% confidence level).
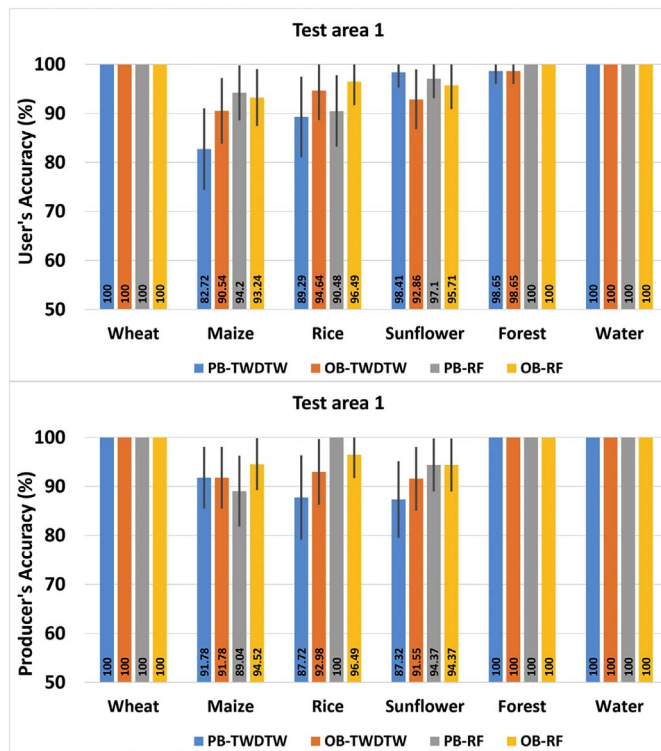

Fig. 10. User's and Producer's accuracy for TA3 represented together with the confidence interval (95% confidence level).
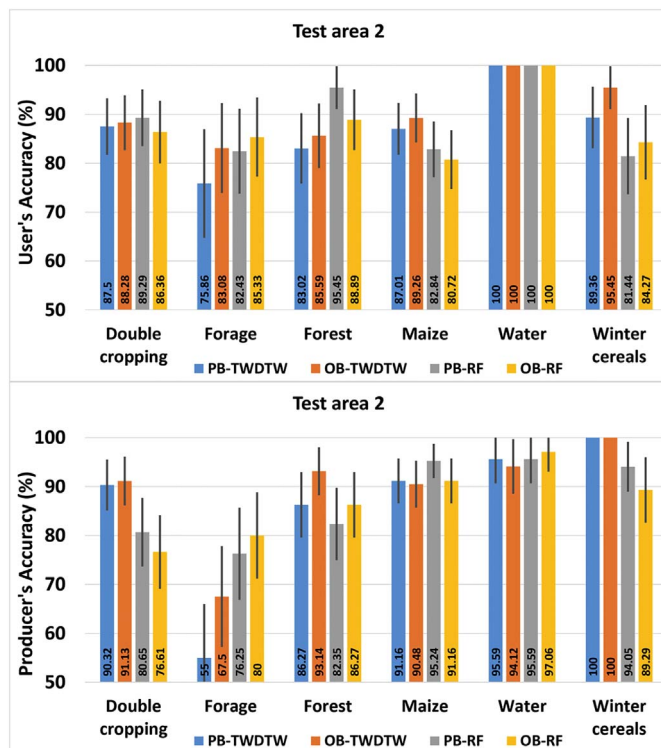

Fig. 9. User's and Producer's accuracy for TA2 represented together with the confidence interval (95% confidence level).

(87.76%); there was a small amount of confusion of sugar beets and onions with wheat and alfalfa. For all classes, the OB-TWDTW obtained higher or equal UAs and PAs when compared to PB-TWDTW, with the highest difference, 8.17%, for PA for onions (Fig. 10).

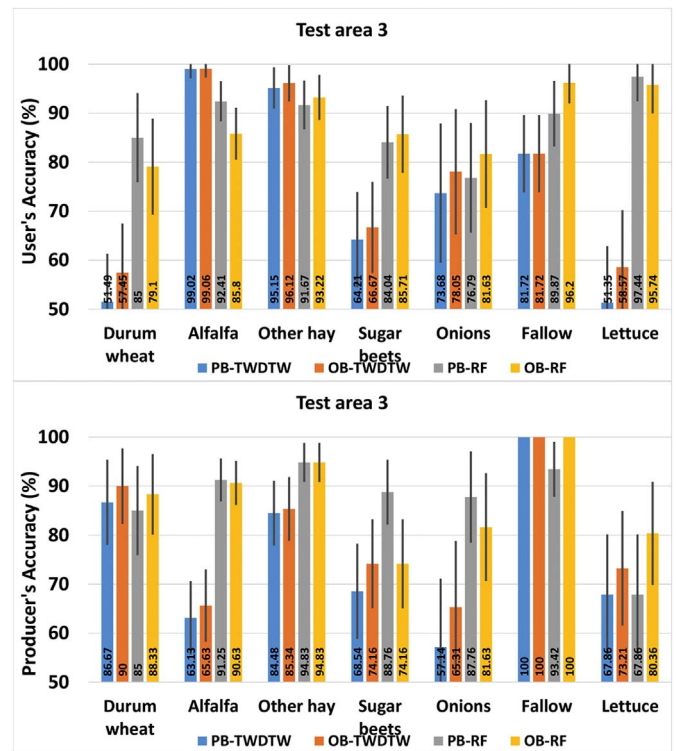PB- and OB-TWDTW obtained lower UA than PB- and OB-RF for

durum wheat, but the PA was higher for this class. This means that TWDTW correctly identified more ground truth pixels/objects as durum wheat, but the commission error for this class is much higher. In the case of alfalfa, the confidence for pixels and objects classified as alfalfa by TWDTW was greater than that obtained by RF, but the PA was lower for TWDTW than RF.

4.3.4. TWDTW dissimilarity measures

Complementary to a classification map, the TWDTW dissimilarity map shows how similar a classified pixel or object was to the temporal pattern of their assigned class (the closer to 0, the better) (Fig. 11). The crops mapped in TA3, for example, present the highest dissimilarity values due to the high intra-class spectral variance, which proved to have a great impact on the classification outputs yielded by TWDTW. The dissimilarity measure values of PB-TWDTW ranged between 0.45 and 12.37 for TA1, 0.69 and 26.81 for TA2, and 0.36 and 13.20 for TA3, while for OB-TWDTW the values ranged between 0.53 and 9.25 for TA1, 0.71 and 16.21 for TA2, and 0.37 and 8.09 for TA3 (Fig. 11).

The difference between the upper bounds of the PB- and OB-TWDTW approaches is an indication that grouping the pixels into objects helps reduce the salt-and-pepper effect characteristic of PB classification, by averaging the NDVI values inside each object. This is also apparent from Fig. 12, which shows that OB classification delivers a spatially homogeneous agricultural mapping product, avoiding the noise introduced by pixels inside a crop parcel that do not share the same properties as the parcel as a whole, which can be due to variations in the crop growth or different amounts of water retained by the soil, for example.

4.3.5. Area of the crop fields mapped

The estimated areas of the crop fields mapped were adjusted to the magnitude of the classification errors. We used the approach described by Olofsson et al. (2013) to compute the error-adjusted area estimates (at 95% confidence interval). In TA1, there are no significant differences between adjusted and mapped areas for the wheat, forest and
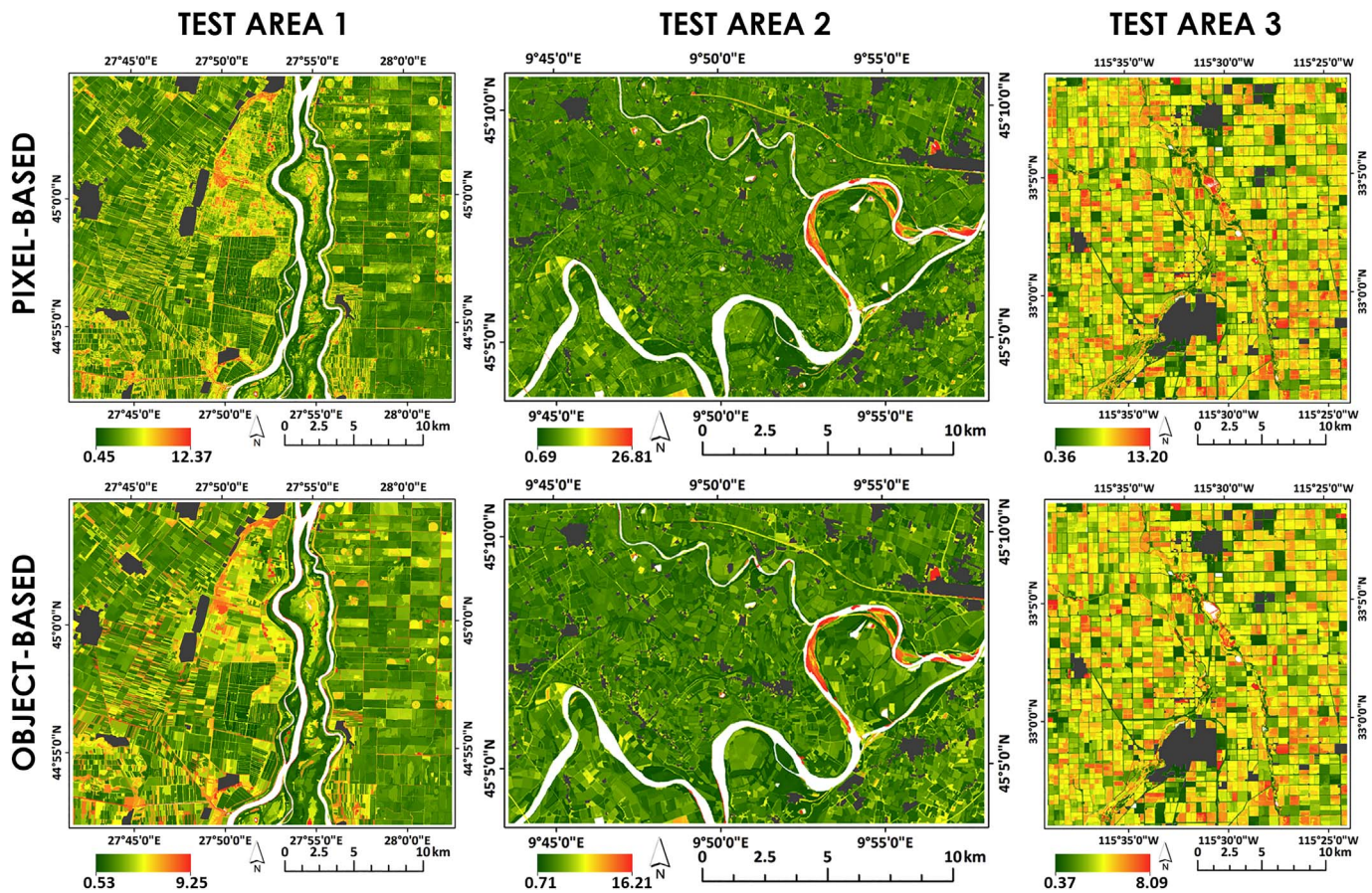
Fig. 11. TWDTW dissimilarity measure maps of the final PB and OB classifications, for all three test areas. Grey represents the built-up areas, while white areas are waterbodies (lakes, rivers) which remained unclassified.

water classes, due to their distinctive temporal patterns, which assure good classification results (Fig. 13). The RF approaches mapped larger areas of sunflower, with the biggest difference, of 3353 ha, for the adjusted area for both RF classifications, when compared to the PB-TWDTW results. By contrast, the TWDTW approaches mapped larger areas of maize than RF. For rice, the adjusted areas are greater than the mapped areas (except for the PB-RF), the greatest difference being of 1473 ha for the OB-TWDTW classification. The margin of error is greater for the TWDTW classification compared to RF, due to the higher UA obtained by the latter approach for rice (Fig. 13).

In TA2, the greatest difference between adjusted and mapped areas is produced by RF for the maize class, where the adjusted area is with 1655 ha smaller for PB-RF, and 1574 ha smaller for OB-RF. A high discrepancy occurred also for the double-cropping class, where the adjusted areas were higher, with 862 ha for PB-RF and 1113 ha for the OB-RF. TWDTW produced the highest discrepancy between mapped and adjusted areas for the maize and forage classes. In the case of maize, the adjusted area was 849 ha smaller for PB-TWDTW, and 551 ha smaller for the OB-TWDTW. In the case of the forage class, the adjusted area was with 681 ha greater for the PB-TWDTW, and 639 ha greater for the OB-TWDTW (Fig. 13).

In TA3, the complexity of the temporal patterns created many discrepancies between the adjusted and mapped areas. For durum wheat, sugar beets, fallow and lettuce, both TWDTW approaches mapped larger areas than the values of the adjusted areas, the biggest differences being recorded for the fallow/idle cropland class (3026 ha for PB-TWDTW results). Larger adjusted areas than mapped areas were obtained using TWDTW for alfalfa and the other hay class, with a 4783 ha difference for PB-TWDTW results. The area of alfalfa was underestimated because of the confusion between it and sugar beets, lettuce

and the other hay class. This is also reflected in the lower PA for alfalfa, namely 63.13% for PB-TWDTW and 65.63% for OB-TWDTW. The fallow area was overestimated because pixels/objects belonging to almost all other crops types (except for the other hay class) were misclassified as fallow; UA for this class is about 81.72% for both PB-TWDTW and OB-TWDTW. For RF classifications, the biggest difference between adjusted and mapped areas is for lettuce: adjusted areas measure 1932 ha more for PB-RF, and 966 ha for OB-RF. The largest error margins were obtained for alfalfa, fallow and lettuce, with a maximum of 1397 ha for PB-TWDTW in the case of lettuce, and 1083 ha for OB-RF for alfalfa (Fig. 13).

5. Discussion

This study evaluated the performance of the TWDTW method (Maus et al., 2016) in identifying various cropland classes from Sentinel-2 NDVI time series, using both pixels and objects as spatial analysis units. The method was applied in three different study areas situated in Romania, Italy and the USA. OB-TWDTW outperformed PB-TWDTW in all three study areas. OB-TWDTW proved to be computationally less intensive, since it was applied only on the centroids of the objects, thus reducing the number of analysis units from, e.g., 6,387,754 pixels to 5198 objects for TA1. Furthermore, the crops mapped using the object-based approach are spatially more consistent than those mapped using the pixel as the smallest analysis unit (Fig. 12). Our results are similar to those obtained by Castillejo-González et al. (2009), who reported that OBIA outperformed pixel-based approaches for cropland mapping based on QuickBird images.
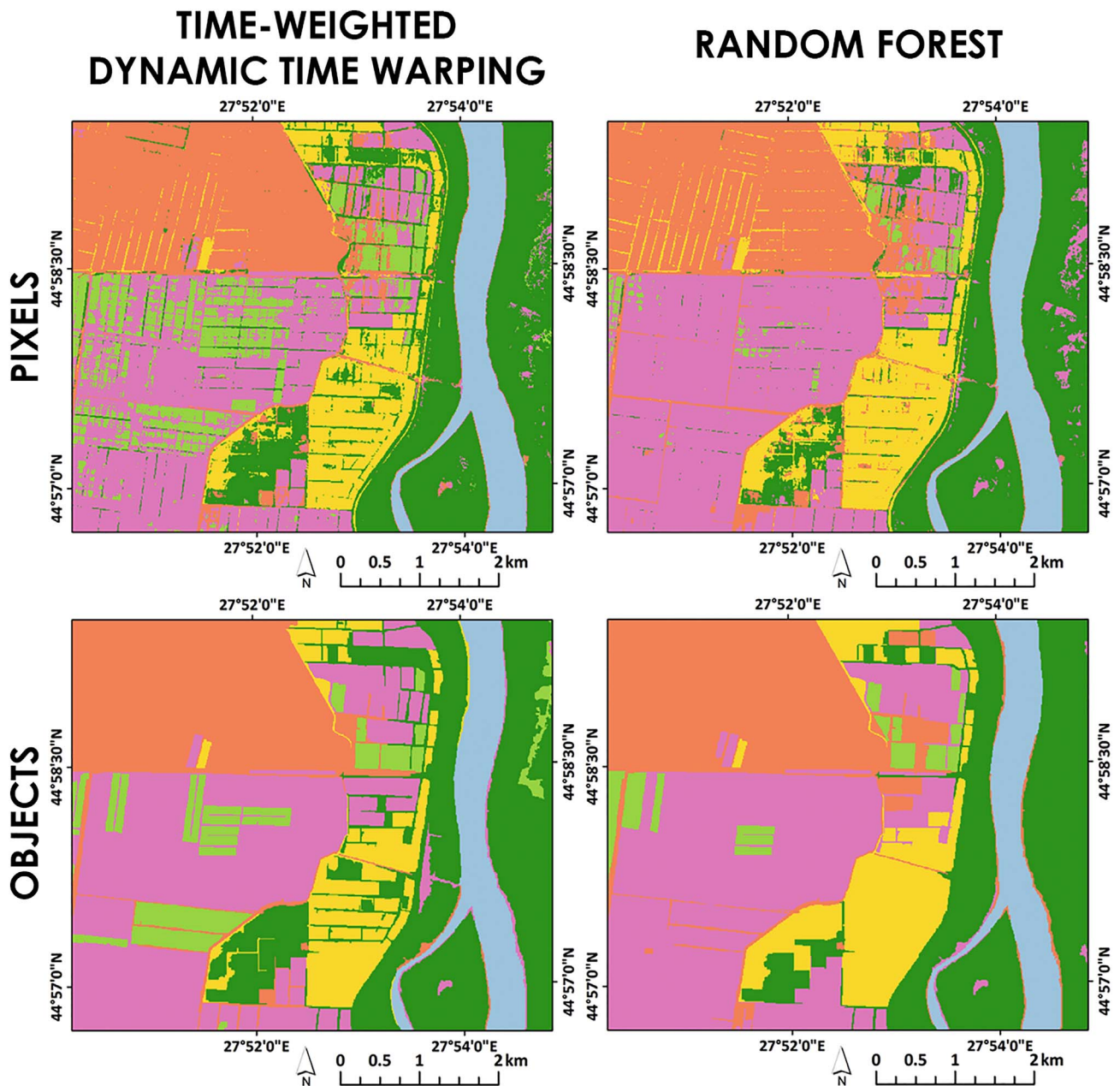
Fig. 12. A subset of TA1 showing the removal of salt-and-pepper effect from PB classification using objects as spatial analysis units in TWDTW and RF, respectively.

5.1. Segmentation of multi-temporal remote sensing images

Image segmentation is a challenging task in OBIA as it requires the user's intervention to define the optimal segmentation parameters for delineating the objects of interest (Belgiu and Drăguţ, 2014; Drăguţ et al., 2010). To deal with this challenge, we used the ESP2 tool (Drăguţ et al., 2014) to automatically identify the scale parameter of the MRS algorithm which would be applied to the Sentinel-2 time series. ESP2 was used as it is a data-driven, unsupervised segmentation evaluation approach that relies exclusively on image statistics to identify the optimal SP. Therefore, it works independently of any reference data required when supervised segmentation evaluation approaches are used (Li et al., 2015). The quality of the segmentation results is influenced by the field geometry (Yan and Roy, 2016). For example, larger parcels and rectangular parcels situated in the eastern part of TA1, in TA2 and TA3 were well delineated in all three study areas (Fig. 5). However, the fields with a high internal variation were over-segmented (especially in TA1 and TA3), and the narrower adjacent fields with more complex geometries were under-segmented (in TA1) (Fig. 5). The classification errors caused by over-segmentation can be overcome by applying post-classification rules for merging objects belonging to the same class (Johnson and Xie, 2011). Under-segmentation occurred especially in TA1, because of the high spatial fragmentation of the fields in the western part of the study area, where in addition the boundaries between these and neighbouring fields are not very distinct. Smaller fields which are merged into larger heterogeneous segments can be decomposed using morphological decomposition algorithms (Yan and Roy, 2016). As the three test areas do not include the full range of possible field geometries, we recommend further investigation on how image segmentation and TWDTW perform in agricultural areas with irregular geometries, with less distinct boundaries between neighbouring fields (Yan and Roy, 2016), with high confusion between agricultural fields and natural vegetation, and where large trees are present within fields (Debats et al., 2016).

Image segmentation has traditionally been applied to single-date satellite images (Desclée et al., 2006). Several studies report the
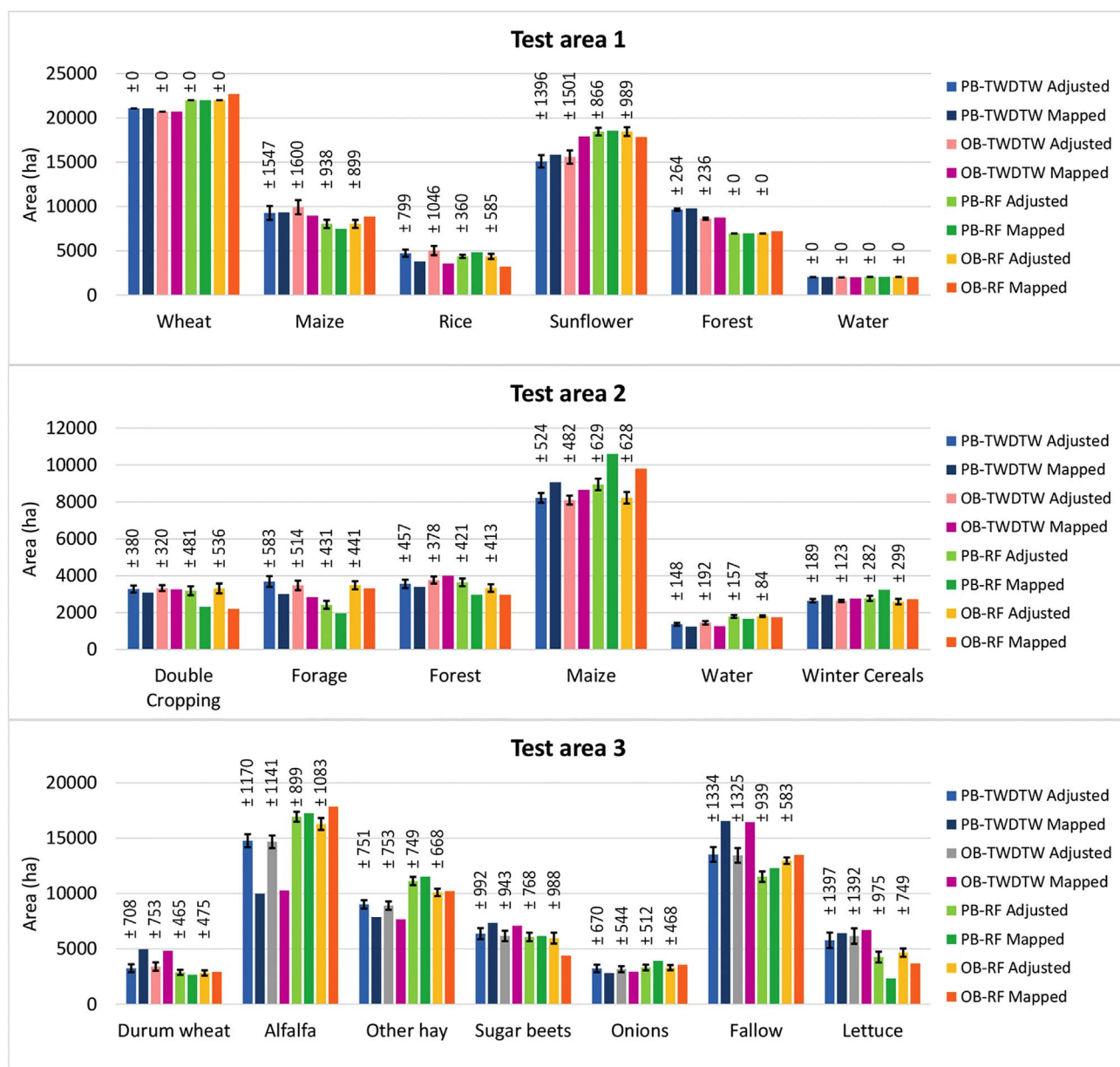
Fig. 13. Area uncertainty: mapped area and adjusted area using the information from error matrix (at 95% confidence interval) of classified crops for the three test areas and the four approaches: pixels-based (PB) and object-based (OB) TWDTW and RF, respectively.

advantages of satellite time series segmentation, such as consistent delineation of agricultural fields from Web Enabled Landsat Data (WELD) (Yan and Roy, 2014), better and faster forest change analysis from SPOT images (Desclée et al., 2006), robustness against shadowing and registration errors (Desclée et al., 2006; Mäkelä and Pekkarinen, 2001), and reduced salt-and-pepper effect apparent in per-pixel classifications (Matton et al., 2015). In our study, segmentation of multitemporal images proved to be an efficient approach for delineating crop fields, including fields whose boundaries are influenced by irrigation systems (Fig. 6). Furthermore, within-field heterogeneity (i.e. the presence of several crops within an individual field) specific to several crops present in the investigated areas was reduced by aggregating pixels into homogeneous spatial units through segmentation. Consequently, OB-TWDTW performed better than PB-TWDTW in identifying crops with high within-field heterogeneity such as forage in TA2, and lettuce, onions or sugar beets present in TA3.

5.2. TWDTW and Random Forest

The highest differences between the TWDTW and RF classification results were obtained for TA3. In this study area, TWDTW produced lower PA and UA for crop classes such as alfalfa, sugar beet or lettuce (see Table 3). The high intra-class spectral heterogeneity, coupled with the fact that the time series for TA3 do not cover the phenological cycle for some crops (e.g. sugar beets and lettuce) may contribute to these low values. Therefore, to ensure the best results for TWDTW, the time series should correspond with the phenological cycles of the crop types investigated, and further research is required to improve the efficiency of TWDTW under such conditions. On the other hand, TWDTW achieved better results than RF in TA2 for double-cropping and winter cereals. Despite the differences between TWDTW and RF presented above, RF achieved good classification results in all three study areas. Therefore, our study confirmed the efficiency of this classifier for cropland mapping already reported in the literature (Hao et al., 2015; Li et al., 2015; Novelli et al., 2016; Pelletier et al., 2016).

Table 5
Overall accuracy (OA) and kappa coefficient of TWDTW and RF applied to all three test areas using just three training samples per class.

	TA1		TA2		TA3	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
PB-TWDTW	92.14	0.91	86.78	0.84	66.34	0.60
OB-TWDTW	92.62	0.91	87.11	0.84	66.56	0.61
PB-RF	86.67	0.84	78.35	0.73	63.86	0.56
OB-RF	88.57	0.86	77.85	0.73	60.23	0.53

Given the similar classification results obtained by RF and TWDTW (except for TA3, where RF outperformed TWDTW), we decided to evaluate the sensitivity of these methods to the number of training samples. These samples, three per class of crop, were taken from all three test areas, randomly selected from among those presented in Table 2. TWDTW produced good classification outputs also with a small number of training samples in TA1 and TA2 (Table 5). In TA3, TWDTW obtained lower classification accuracies. For RF, however, the overall accuracy of the classification results was lower in all three study areas. Previous studies have also reported that RF achieves less accurate results when a small number of training samples is used (Millard and Richardson, 2015; Valero et al., 2016). The ability of TWDTW to achieve good classification results with a small number of training samples is a huge advantage which should be taken into account when regional and global cropland mapping projects are planned, especially in countries which lack input training samples (King et al., 2017; Matton et al., 2015; Petitjean et al., 2012a). Petitjean et al. (2012a) stated that TWDTW has also the advantage of being robust to irregular temporal sampling and to the annual changes of vegetation phenological cycles. These are important assets, especially in regions with high meteorological variability and where agricultural practices vary from one year to another. However, further research is required to test the sensitivity of TWDTW to the year-to-year variability of vegetation phenology. Additionally, the remote sensing community interested in developing automated methods for crop fields mapping would greatly benefit from the availability of an online, readily accessible crop calendar similar to those developed by FAO (goo.gl/zo4jpY). This crop calendar could guide the selection of time series that correspond with the phenological cycles of the target crop types.

5.3. Potential of Sentinel-2 images for cropland mapping

Our work took advantage of the increasing spatial and temporal resolution of Sentinel-2 imagery for cropland mapping. The increasing spatial resolution of the MSI sensor allowed us to successfully delineate crop field boundaries in all test areas, including where the agricultural landscape is very fragmented and hence the fields are smaller (see TA1). Other studies have also reported the added-value of the improved Sentinel-2 spatial resolution for mapping built-up areas (Pesaresi et al., 2016) and smallholdings (Lebourgeois et al., 2017), and for identifying greenhouses (Novelli et al., 2016). Novelli et al. (2016) showed that Sentinel-2 in combination with WorldView-2 images (used for segmentation) facilitate the accurate identification of greenhouses, whereas Lebourgeois et al. (2017) used very-high resolution Pléiades images to segment fields, which were further classified using spectral features and indices computed from Sentinel-2 data.

The higher temporal resolution of the freely available satellite imagery (with revisiting times of 16 days for Landsat, 10 days for Sentinel-2 in 2016, and 5 days for Sentinel-2B, launched in March 2017) is expected to increase the chance of finding cloud-free data (Gómez et al., 2016; Yan and Roy, 2015). In this study, a reduced number of cloud-free images were available, especially in TA1 and TA2 (both areas situated in Europe). Nevertheless, the images available proved to be sufficient for describing the temporal behaviour of the

target crops and for achieving satisfactory classification results. In regions with few Sentinel-2 images, RADARSAT-2 (Fisette et al., 2013) or Sentinel-1 data (Satalino et al., 2012) could be combined with multispectral Sentinel-2 images.

We used NDVI in our study because this vegetation index proved to be "sufficiently stable to permit meaningful comparisons of seasonal and inter-annual changes in vegetation growth and activity" (Huete et al., 2002). However, the soil background might cause variations in the computed phenological profile (Montandon and Small, 2008). Future research to investigate the soil spectra interactions with the NDVI index is therefore required. A potential solution to this problem is the utilization of alternative indices, such as the Soil Adjusted Vegetation Index (SAVI) (Huete, 1988), transformed SAVI (Baret and Guyot, 1991), or the generalized SAVI (Gilabert et al., 2002).

The potential of the Sentinel-2 spectral configuration has been evaluated in various studies dedicated to identifying water bodies (Du et al., 2016); mapping agricultural fields from single-date Sentinel-2 images (Immitzer et al., 2016); predicting leaf nitrogen concentration in vegetation (Ramoelo et al., 2015); estimating canopy chlorophyll content, leaf area index (LAI) and leaf chlorophyll concentration (LCC) (Frampton et al., 2013); evaluating burn severity (Fernández-Manso et al., 2016), or identifying *Prosopis* and *Vachellia* tree species (Ng et al., 2017). We limited this study to the utilization of 10 m resolution red and near-infrared spectral bands in the classification procedure. Future work to consider spectral indices such as the Enhanced Vegetation Index (EVI) or Normalized Difference Water Index (NDWI) (McFeeters, 1996), and the vegetation red-edge bands (bands 5, 6, 7 and 8a) for cropland mapping is recommended.

6. Conclusion

In this study, Sentinel-2 time series served as data input for cropland mapping using the TWDTW method. The analysis was applied on pixels and objects, in three test areas with different agricultural calendars, geometry of parcels and climate conditions. OB-TWDTW outperformed PB-TWDTW in all test areas in terms of the quality of the classification outputs, as measured by the overall accuracy metric and in terms of computational time. TWDTW performed less accurately, especially when the mapped crops presented a high intra-class variability or the yearly-based time series did not overlap with the phenological cycles of the crops. When compared to the RF classifier, TWDTW obtained similar classification outputs in two test areas. RF performed better in the third, where the within-field heterogeneity was very high. TWDTW proved to be more robust than RF when applied to the number of training samples. Automated segmentation of multi-temporal images using ESP2 and the MRS algorithm generated a satisfactory delineation of the agricultural parcels.

Given the high classification accuracy obtained without human intervention and the reduced computational time, the TWDTW method applied to objects could be integrated into operational programs dedicated to cropland mapping and monitoring based on satellite image time series.

References

Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V., Durieux, L., 2011. Classification

of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. Int. J. Remote Sens. 32, 7847–7871.

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio, P.A., 2016. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. Eur. J. Remote. Sens. 49, 361–380.

Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation-an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informationsverarbeitung. Wichmann-Verlag, Heidelberg, pp. 12–23.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35, 161–173.

Baumann, M., Ozdogan, M., Richardson, A.D., Radeloff, V.C., 2017. Phenology from Landsat when data is scarce: using MODIS and dynamic time-warping to combine multi-year Landsat imagery to derive annual phenology curves. Int. J. Appl. Earth Obs. Geoinf. 54, 72–83.

Belgiu, M., Drăguţ, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. ISPRS J. Photogramm. Remote Sens. 96, 67–75.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto Int. 26, 341–358.

Bradley, J.V., 1968. Distribution-free Statistical Tests. Prentice-Hall, New Jersey.

Breiman, L., 2001. Random Forest. Mach. Learn. 45.

Castillejo-González, I.L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J.M., Jurado-Expósito, M., de la Orden, M.S., González-Audicana, M., 2009. Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. Comput. Electron. Agric. 68, 207–215.

CDFA, 2016. California Agriculture Statistics Review 2015–2016. California Department of Food and Agricultural.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Croitoru, A.-E., Holobaca, I.-H., Lazar, C., Moldovan, F., Imbroane, A., 2012. Air temperature trend and the impact on winter wheat phenology in Romania. Clim. Chang. 111, 393–410.

Debats, S.R., Luo, D., Estes, L.D., Fuchs, T.J., Caylor, K.K., 2016. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens. Environ. 179, 210–221.

Desclée, B., Bogaert, P., Defourny, P., 2006. Forest change detection by statistical object-based method. Remote Sens. Environ. 102, 1–11.

Drăguţ, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int. J. Geogr. Inf. Sci. 24, 859–871.

Drăguţ, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogramm. Remote Sens. 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 Sens. 8, 354.

FAO-UNESCO, 1975. Soil Map of the World 1:5000000. Vol. 2 UNESCO, Paris North America.

FAO-UNESCO, 1981. Soil Map of the World 1:5000000. Vol. 5 UNESCO, Paris Europe.

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.

Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., Smith, A., Davidson, A., Shang, J., Jarvis, I., 2013. AAFC annual crop inventory. In: 2013 Second International Conference on Agro-geoinformatics (Agro-Geoinformatics). IEEE, Fairfax, VA USA, pp. 270–274.

Forster, D., Kellenberger, T.W., Buehler, Y., Lennartz, B., 2010. Mapping diversified peri-urban agriculture–potential of object-based versus per-field land cover/land use classification. Geocarto Int. 25, 171–186.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92.

Gilabert, M., González-Piqueras, J., Garcıa-Haro, F., Meliá, J., 2002. A generalized soil-adjusted vegetation index. Remote Sens. Environ. 82, 303–310.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recogn. Lett. 27, 294–300.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.

Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: a web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84, 111–123.

Hao, P., Zhan, Y., Wang, L., Niu, Z., Shakir, M., 2015. Feature selection of time series MODIS data for early crop classification using random forest: a case study in Kansas, USA. Remote Sens. 7, 5347.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices.

Remote Sens. Environ. 83, 195–213.

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.

Jia, K., Liang, S., Zhang, N., Wei, X., Gu, X., Zhao, X., Yao, Y., Xie, X., 2014. Land cover classification of finer resolution remote sensing data integrating temporal features from time series coarser resolution data. ISPRS J. Photogramm. Remote Sens. 93, 49–55.

Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J. Photogramm. Remote Sens. 66, 473–483.

King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.-P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M., Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. Remote Sens. Environ. 195, 13–29.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362.

Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 2017. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). Remote Sens. 9, 259.

Li, Q., Wang, C., Zhang, B., Lu, L., 2015. Object-based crop classification with Landsat-MODIS enhanced time-series data. Remote Sens. 7, 15820.

Long, J.A., Lawrence, R.L., Greenwood, M.C., Marshall, L., Miller, P.R., 2013. Object-oriented crop classification using multitemporal ETM + SLC-off imagery and random forest. GISci. Remote Sens. 50, 418–436.

Mäkelä, H., Pekkarinen, A., 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. Remote Sens. Environ. 77, 66–75.

Matton, N., Canto, G.S., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. Remote Sens. 7, 13208–13232.

Maus, V., 2016. dtwSat: time-weighted dynamic time warping for satellite image time series analysis. In: R Package Version 0.2.0.

Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F.M., Queiroz, G.R.D., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE J. Select. Top. Appl. Earth Observ. 1–11.

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Millard, K., Richardson, M., 2015. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sens. 7, 8489.

Montandon, L.M., Small, E.E., 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. Remote Sens. Environ. 112, 1835–1845.

Müller, H., Rufin, P., Griffiths, P., Siqueira, A.J.B., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sens. Environ. 156, 490–499.

Muller-Wilm, U., Louis, J., Richter, R., Gascon, F., Niezette, M., 2013. Sentinel-2 level 2A prototype processor: architecture, algorithms and first results. In: Proceedings of the ESA Living Planet Symposium, Edinburgh, UK, pp. 9–13.

Ng, W.-T., Rima, P., Einzmann, K., Immitzer, M., Atzberger, C., Eckert, S., 2017. Assessing the potential of Sentinel-2 and Pléiades data for the detection of Prosopis and Vachellia Spp. in Kenya. In: Remote Sensing. Vol. 9. pp. 74.

Novelli, A., Aguilar, M.A., Nemmaoui, A., Aguilar, F.J., Tarantino, E., 2016. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: a case study from Almería (Spain). Int. J. Appl. Earth Obs. Geoinf. 52, 403–411.

Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens. Environ. 129, 122–131.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. Remote Sens. Environ. 187, 156–168.

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.

Petitjean, F., Weber, J., 2014. Efficient satellite image time series analysis under time warping. IEEE Geosci. Remote Sens. Lett. 11, 1143–1147.

Petitjean, F., Inglada, J., Gançarski, P., 2012a. Satellite image time series analysis under time warping. IEEE Trans. Geosci. Remote Sens. 50, 3081–3095.

Petitjean, F., Kurtz, C., Passat, N., Gançarski, P., 2012b. Spatio-temporal reasoning for the classification of satellite image time series. Pattern Recogn. Lett. 33, 1805–1815.

Rabiner, L.R., Juang, B.H., 1993. Fundamentals of Speech Recognition. PTR Prentice Hall, Englewood Cliffs, N.J.

Ramoelo, A., Cho, M., Mathieu, R., Skidmore, A.K., 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. J. Appl. Remote. Sens. 9 094096–094096.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. 26, 43–49.

Satalino, G., Balenzano, A., Mattia, F., Davidson, M., 2012. Sentinel-1 SAR data for mapping agricultural crops not dominated by volume scattering. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 6801–6804.

Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., Hostert, P., 2015. Mapping land cover in complex Mediterranean landscapes using Landsat: improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sens. Environ. 156, 527–536.

Sima, M., Popovici, E.-A., Bălteanu, D., Micu, D.M., Kucsicsa, G., Dragotă, C., Grigorescu, I., 2015. A farmer-based analysis of climate change adaptation options of agriculture

in the Bărăgan Plain, Romania. Earth Perspect. 2, 5.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.

United-Nations, 2015a. World population prospects: the 2015 revision. In: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat. Department of Economic and Social Affairs, New York.

United-Nations, 2015b. Resolution adopted by the General Assembly on 25 September 2015. In: Washington: United Nations: UN General Assembly.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sens. 8, 55.

Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. ISPRS J. Photogramm. Remote Sens. 110, 1–13.

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. Ser. B Stat Methodol. 73, 3–36.

Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google earth engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244.

Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal web en-abled Landsat data. Remote Sens. Environ. 144, 42–64.

Yan, L., Roy, D.P., 2015. Improved time series land cover classification by missing-ob-servation-adaptive nonlinear dimensionality reduction. Remote Sens. Environ. 158, 478–491.

Yan, L., Roy, D., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86.