Article

Object-Based Time-Constrained Dynamic Time Warping Classification of Crops Using Sentinel-2

Ovidiu Csillik [1,]*[ID], Mariana Belgiu [2][ID], Gregory P. Asner [1] and Maggi Kelly [3,4][ID]

[1] Center for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ 85287, USA; gregasner@asu.edu

[2] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands; m.belgiu@utwente.nl

[3] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA; maggi@berkeley.edu

[4] Division of Agriculture and Natural Resources, University of California, Davis, CA 95616, USA

* Correspondence: ocsillik@asu.edu; Tel.: +1-650-943-3411

check for updates

Abstract: The increasing volume of remote sensing data with improved spatial and temporal resolutions generates unique opportunities for monitoring and mapping of crops. We compared multiple single-band and multi-band object-based time-constrained Dynamic Time Warping (DTW) classifications for crop mapping based on Sentinel-2 time series of vegetation indices. We tested it on two complex and intensively managed agricultural areas in California and Texas. DTW is a time-flexible method for comparing two temporal patterns by considering their temporal distortions in their alignment. For crop mapping, using time constraints in computing DTW is recommended in order to consider the seasonality of crops. We tested different time constraints in DTW (15, 30, 45, and 60 days) and compared the results with those obtained by using Euclidean distance or a DTW without time constraint. Best classification results were for time delays of both 30 and 45 days in California: 79.5% for single-band DTWs and 85.6% for multi-band DTWs. In Texas, 45 days was best for single-band DTW (89.1%), while 30 days yielded best results for multi-band DTW (87.6%). Using temporal information from five vegetation indices instead of one increased the overall accuracy in California with 6.1%. We discuss the implications of DTW dissimilarity values in understanding the classification errors. Considering the possible sources of errors and their propagation throughout our analysis, we had combined errors of 22.2% and 16.8% for California and 24.6% and 25.4% for Texas study areas. The proposed workflow is the first implementation of DTW in an object-based image analysis (OBIA) environment and represents a promising step towards generating fast, accurate, and ready-to-use agricultural data products.

Keywords: agricultural mapping; DTW; eCognition; land use/land cover mapping; object-based image analysis; satellite image time series

1. Introduction

Humankind has impacted almost every place on Earth causing important environmental changes [1,2]. One of the most widespread landscape changes is related to the cultivation of crops, which has been spatially extended and intensified to face the food supply needs of a constantly growing global population [3,4]. Efficient monitoring of crops is important in providing food security, understanding productivity, and supporting agricultural policies. In this regard, satellite image analysis can provide timely and relevant spatial and temporal information products and tools [5,6].

In recent decades, satellite images with global coverage and different spectral and temporal characteristics have become freely available, such as MODIS [7] and Landsat [8]. The recent Sentinel-2

satellite, a high resolution, multi-spectral spaceborne Earth imaging mission, is providing imagery with high temporal resolution (five days), spectral (13 bands), and spatial resolution (10–60 m). These assets are making the Sentinel-2 product suitable for satellite image time series (SITS) analysis [6]. SITS offers advantages over single-image approaches, particularly by allowing the incorporation of the phenological characteristics of vegetation throughout a year or multiple years in the classification task [9]. Given the large amounts of SITS acquired by diverse spaceborne platforms, we need efficient methods to turn this data into information [10].

Numerous methods and software packages for analyzing SITS exist, each with advantages and disadvantages [11], including the popular BFAST (Breaks For Additive Seasonal and Trend) [12], TIMESAT [13], SPIRITS (Software for the Processing and Interpretation of Remotely Sensed Image Time Series) [14], DTW (Dynamic Time Warping) [11], or WELD (Web Enabled Landsat Data) [9]. These methods are used in a wide range of applications including those focused on seasonality information extraction [15], phenological change detection [16], reconstructing cloud free time series [17], land cover mapping [11,18], automated mapping of crops [19,20], agricultural dynamics mapping during different cropping years [21], and detecting trends in forest disturbance and recovery [22].

SITS analysis aims at extracting meaningful information from pixels for a certain goal (e.g., crop mapping), but this comes with several issues that need to be addressed: (1) the shortcoming of sample availability, which can be overcome by using reference data from the past; (2) the irregular temporal sampling of images and the missing information in the datasets (e.g., cloud contamination) that can obscure the analysis; and (3) the variability of pseudo-periodic phenomena (e.g., vegetation cycles, which are influenced by weather conditions) [11]. All of these challenges require the development of methods capable of coping with the distortions of the temporal profile of physical variables and artifacts of SITS datasets [11]. One method that addresses these issues is Dynamic Time Warping (DTW), initially developed for speech recognition [23] and recently applied in the analysis of SITS [11,24–26].

DTW compares the similarity between two temporal sequences by finding their optimal alignment and providing a dissimilarity measure [23,27]. Previous research demonstrated that DTW performs successfully in various data mining scenarios [28] such as data mining of sequential data [29,30] or spoken word recognition [23]. DTW has recently become a focus for the remote sensing community as well. Petitjean et al. [11] introduced DTW from a theoretical point of view and demonstrated the utility of DTW in analyzing SITS. Baumann et al. [31] used DTW to re-align multi-year Landsat time series to a common phenology in order to eliminate the yearly phenological discrepancies. The authors demonstrated their approach for different forest types in eight different locations distributed across the United States. Gómez et al. [32] identified anomalous values in Landsat time series using DTW, replacing them by spatiotemporal interpolation for improved forest monitoring. Xue et al. [33] made use of DTW as a similarity measure in sample selection, linking a few reliable samples with other interpreted samples, thus increasing the number of available training samples. Guan et al. [34] used DTW similarity measure of the MODIS Normalized Difference Vegetation Index (NDVI) time series to map large-area rice cropping systems. To distinguish between different types of land use and land cover, Maus et al. [18] introduced a time-weighted DTW (TWDTW) that accounts for seasonality of land cover types, improving the classification accuracy when compared to traditional DTW. The authors implemented the TWDTW into an open-source R package, *dtwSat* [25].

Most Earth Observation data processing and SITS applications using DTW have applied the analysis at the scale of a pixel [35]. This is understandable, since until the recent advent of Sentinel-2, only medium to coarse resolution SITS were available at no cost (e.g., Landsat or MODIS). The increasing spatial resolution of Sentinel-2 images makes them suitable for object-based image analysis (OBIA). Through OBIA, similar pixels are grouped together into objects through image segmentation [36]. In a recent study dedicated to the classification of Sentinel-2 SITS using DTW, Belgiu and Csillik [26] demonstrated that object-based DTW outperformed pixel-based DTW for crops mapping in three study areas with different agricultural landscapes. At present, the remote sensing community lacks a dedicated workflow to perform an object-based DTW on SITS [18]. The availability of easy-to-use tools

to analyze and understand SITS is seen as a great demand for the successful application of this data to real-world problems [37,38].

This study builds on the previous work by Belgiu and Csillik [26] who compared pixel-based and object-based dynamic time warping with random forest classification for crop mapping. First, we extend the single-band analysis to multi-band time-constraint DTW for crop mapping using different study areas, time periods, more crops, and more images in the time series. The performance was assessed on two Sentinel-2 SITS from California and Texas, geographic areas characterized by highly complex agricultural landscape and practices. Secondly, we tested different time constraints in DTW classification and compared the results with those obtained by means of Euclidean distance measurement and a DTW without time constraint. Thirdly, we discuss the implications of DTW dissimilarity values in understanding the classification errors and evaluate the error propagations throughout our analysis. The approach is implemented as a ready-to-use algorithm in eCognition Developer software [39] available for the remote sensing community (see Supplementary Material). The developed workflow is able to adapt the DTW to perform on objects and solve one of the main challenges of DTW, namely a reduction in computational time [18].

The paper is structured as follows: Section 2 describes the Sentinel-2 datasets used, the design of the experiments, the background of DTW algorithm and the proposed single-band and multi-band object-based time-constraint DTW implementation. Section 3 is dedicated to the results related to the segmentation and classification accuracies. Section 4 discusses the main results, good practices in applying the object-based DTW analysis and states the advantages and limitations of the proposed workflow. Section 5 concludes the study and underlines the importance of such developments.

2. Materials and Methods

2.1. Study Areas and Datasets

2.1.1. Study Areas

The two study areas were chosen due to their complexity of agricultural practices, land parcel geometry, and the large number of crop types and cropping cycles, which together with the irregular temporal sampling of images would challenge most of the existing image classification methods.

The first test area (TA1) is in Southern California, in central Imperial County (33°00′N and 115°30′W) (Figure 1). The area covers 58,890 ha, with an extent of 23.36 × 25.21 km. Situated at low altitudes with a hot desert climate, the area is characterized by high agricultural productivity, especially for hay (alfalfa), onions, and lettuce [40]. The agricultural landscape is intensively managed to produce irrigated crops, with one or more cropping cycles. The crops occupying more than 2% of the study area (according to CropScape—Cropland Data Layer which is provided by the United States Department of Agriculture, National Agricultural Statistics Service [41,42]) were considered for classification, and include durum and winter wheat, alfalfa, other hay/non-alfalfa, sugarbeets, onions, sod/grass seed, fallow/idle cropland, vegetables (mostly lettuce, plus carrots, broccoli and greens), together with the water class.

The second test area (TA2) is in Northwestern Texas, near the border with New Mexico and Oklahoma, in the southern part of Dallam County (36°10′N and 102°35′W) (Figure 1). The area covers 59,400 ha, with an extent of 27 × 22 km. The semi-arid climate with hot summers and cool and dry winters highly influences the agricultural practices, characterized by center-pivot irrigation. The main crops cultivated according to CropScape [41,42] and occupying more than 2% of the study area are corn, cotton, winter wheat, alfalfa, fallow/idle cropland, grass/pasture, and double crop (winter wheat, followed by corn).

2.1.2. Sentinel-2 Satellite Image Time Series

Copernicus Sentinel-2 Level-1C (S2-L1C) products provide orthorectified Top-Of-Atmosphere (TOA) reflectance, including radiometric and geometric corrections. The Sentinel-2 SITS used in this study were composed of 28 Sentinel-2A and 2B images for TA1 (subset of S2-L1C T11SPS tile) and 20 images for TA2 (subset of S2-L1C T13SGA) (Figure 2). When analyzing time series, atmospheric correction is an important preprocessing step and, therefore, we processed the reflectance images from TOA Level 1C to Bottom-Of-Atmosphere (BOA) Level 2A using the *sen2cor* plugin [43,44]. The images were selected manually to minimize the cloud coverage percentage. The clouds and cloud shadow contamination occurred in two images for each SITS, covering less than 20%. The temporal distribution of the images covers well one agricultural year, from 23 September 2016 to 23 September 2017 for TA1 and from 18 October 2016 to 2 November 2017 for TA2. The biggest time differences between two consecutive images are 40 days for TA1 and 50 days for TA2, both in autumn and winter months when cloud covered images were unusable. The smallest time difference is 5 days in both test areas since the advent of Sentinel-2B images in July 2017 (Figure 2).
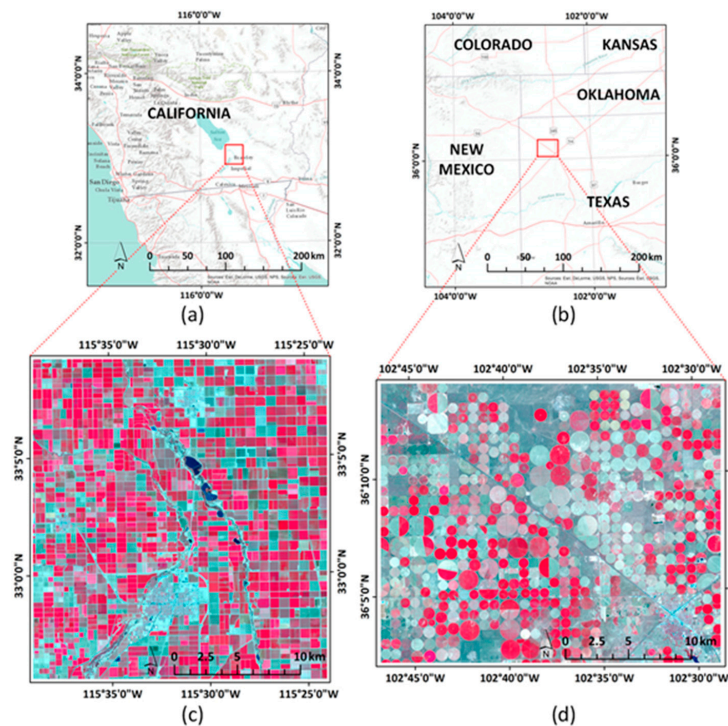

Figure 1. Test area locations in California (first test area (TA1)) (**a**) and Texas (second test area (TA2)) (**b**). The false color composites for TA1 (1 May 2017) and TA2 (20 July 2017) are depicted in (**c**) and (**d**), respectively.
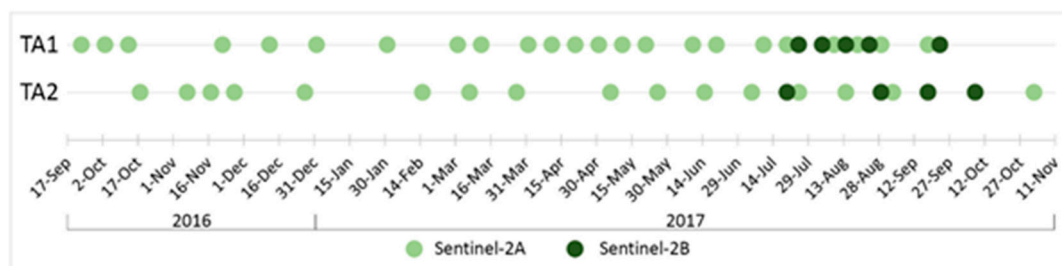

Figure 2. Timeline of the two satellite image time series (SITS) used, composed of Sentinel-2A and 2B images with an irregular temporal distribution.

2.1.3. Vegetation Indices Extracted from Sentinel-2 Images

Using most of the Sentinel-2 bands, we derived 5 widely used vegetation indices to construct time series for DTW (Table 1). We resampled the 20 m resolution bands (red-edge 1 and 2, narrow near-infrared and short-wave infrared) to perfectly align with the 10 m spatial resolution bands (blue, green, red, and near-infrared). The Normalized Difference Vegetation Index (NDVI) [45] is a chlorophyll sensitive index sufficiently stable to compare seasonal changes in vegetation growth [46]. The Green Normalized Difference Vegetation Index (GNDVI) was shown to be more sensitive to chlorophyll concentration than NDVI [47] and useful in differentiation in stressed and senescent vegetation [48]. The Normalized Difference Red-Edge (NDRE) is highly sensitive to pigment changes [49] and is correlated with drought induced seasonal variation in leaf photosynthetic rates [50]. The Soil Adjusted Vegetation Index (SAVI) minimizes soil brightness influences accounting for uncertainties resulting from variation in background condition by incorporating a correction factor L in the NDVI formula [48,51]. The Normalized Difference Water Index (NDWI) is useful to measure vegetation liquid water status and can be a good indicator of irrigation if the time series is sufficiently dense [52,53] (Table 1).

For single-band DTW, we used the NDVI time series. For multi-band DTW, we combined the NDVI with GNDVI, NDRE, SAVI, and NDWI to evaluate if integrating multiple time series of vegetation indices into DTW will yield more accurate results for crop classification. The entire workflow of our analysis is shown in Figure 3.

Table 1. Vegetation indices we used for crop mapping, their equation and range of values. Sentinel-2 bands used are the green (B3, 0.560 μm), red (B4, 0.665 μm), red-edge 1 (B5, 0.705 μm), red-edge 2 (B6, 0.740 μm), near-infrared (B8, 0.842 μm), narrow near-infrared (B8A, 0.865 μm), and SWIR (B11, 1.610 μm).

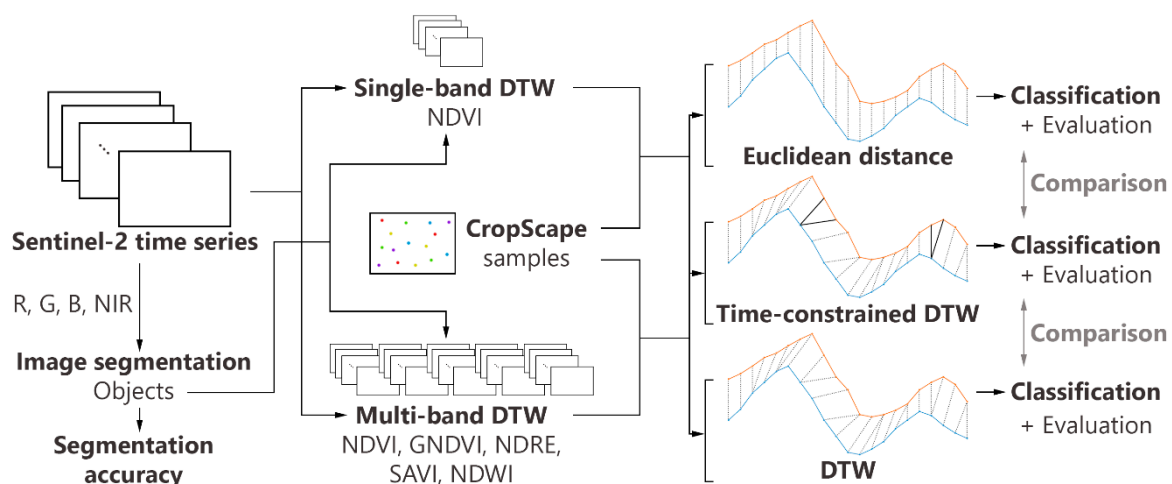| Name	Description	Equation	Range
NDVI	Normalized Difference Vegetation Index	(B8 − B4) / (B8 + B4)	[−1, 1]
GNDVI	Green Normalized Difference Vegetation Index	(B8 − B3) / (B8 + B3)	[−1, 1]
NDRE	Normalized Difference Red-Edge	(B6 − B5) / (B6 + B5)	[−1, 1]
SAVI	Soil Adjusted Vegetation Index	((B8 − B4) / (B8 + B4 + 0.5)) × (1 + 0.5)	[−1, 1]
NDWI	Normalized Difference Water Index	(B8A − B11) / (B8A + B11)	[−1, 1]

Figure 3. Workflow of our object-based dynamic time warping (DTW) classifications using multiple vegetation indices extracted from Sentinel-2 SITS, as shown in Table 1.

2.1.4. Reference and Validation Samples

For both test areas, we randomly generated points that were well spatially distributed across the study areas, covering most parcels with an identifiable land cover type. These were further classified according to the CropScape layer for 2017, which is temporally overlapping with our Sentinel-2 time

series. A temporal pattern for each class was obtained, but additional refinement was necessary to remove the inconsistent patterns and was done through visual interpretation and by analyzing the temporal pattern of NDVI. The good performance of a DTW classification comes from good quality reference samples [18]. DTW performs well also with a low number of samples, as demonstrated in [26], as long as their temporal patterns are representative for each class. Therefore, to avoid smoothing a reference temporal pattern by selecting multiple heterogeneous samples for the same class, we selected only 3 reference samples per class to generate the temporal patterns needed as input for the DTW classification, avoiding selecting samples that were contaminated with clouds or clouds shadow. For single-band DTW, we used the NDVI to generate the patterns, while for multi-band DTW we used the temporal patterns from all five vegetation indices (Table 1). By using a clean reference pattern for a class, DTW will be able to deal with the intra-class heterogeneity of patterns and match them well. The rest of the samples were used for validation and their number varied in accordance with class representativeness within each test area, i.e., the area of the class occupied in the CropScape (Table 2).

Table 2. The number of reference (ref.) and validation (valid.) samples for TA1 and TA2, respectively. For alfalfa in TA1 and winter wheat in TA2, two distinct temporal patterns were identified, as detailed in Section 2.1.5.

California (TA1)			Texas (TA2)		
Class	**Ref.**	**Valid.**	**Class**	**Ref.**	**Valid.**
Wheat	3	48	Corn	3	287
Alfalfa	3, 3	569	Cotton	3	122
Other hay/non-alfalfa	3	116	Winter wheat	3, 3	171
Sugarbeets	3	70	Alfalfa	3	34
Onions	3	98	Fallow/idle cropland	3	41
Sod/grass seed	3	82	Grass/pasture	3	129
Fallow/idle cropland	3	76	Double crop	3	60
Vegetables	3	47			
Water	3	30			
Total	**30**	**1136**	**Total**	**24**	**844**

2.1.5. Temporal Patterns of Reference Samples

The two test areas present complex phenological patterns of the crops. For simplicity, we will detail here the temporal characteristics of a crop based on the NDVI time series. All five vegetation indices have values between −1 and 1, with NDVI and GNDVI having predominantly higher values than NDRE, SAVI, and NDVI for both test areas, while the lowest values are mainly for NDRE and NDWI. Although the five vegetation indices follow similar patterns, there are variations dictated by their sensitivity to vegetation characteristics, i.e., biophysical properties of the classified crops. Temporal patterns of crops for all five vegetation indices are shown in Figure 4 for TA1 and Figure 5 for TA2.

For TA1, winter wheat, sugarbeets, and onions have a single cropping cycle with different temporal length, with onions having the shortest (January to May 2017) and sugarbeets having the longest cropping cycle (October 2016 to June 2017). Vegetables have 2–3 cropping cycles, similar to the grass seed. Fallow/idle cropland have values of NDVI between 0 and 0.1 throughout the whole year, while water class presents negative values of NDVI, with shorter peaks of positive values during the summer. Alfalfa have the highest diversity of temporal patterns for TA1, with two main types of patterns. In order to account for this, we used two reference temporal patterns for alfalfa, one with 5 cropping cycles and one with 7–8 cropping cycles over the timeframe analyzed (Figure 4). The other hay/non-alfalfa class presents 3 cropping cycles per year, shorter during the summer period.
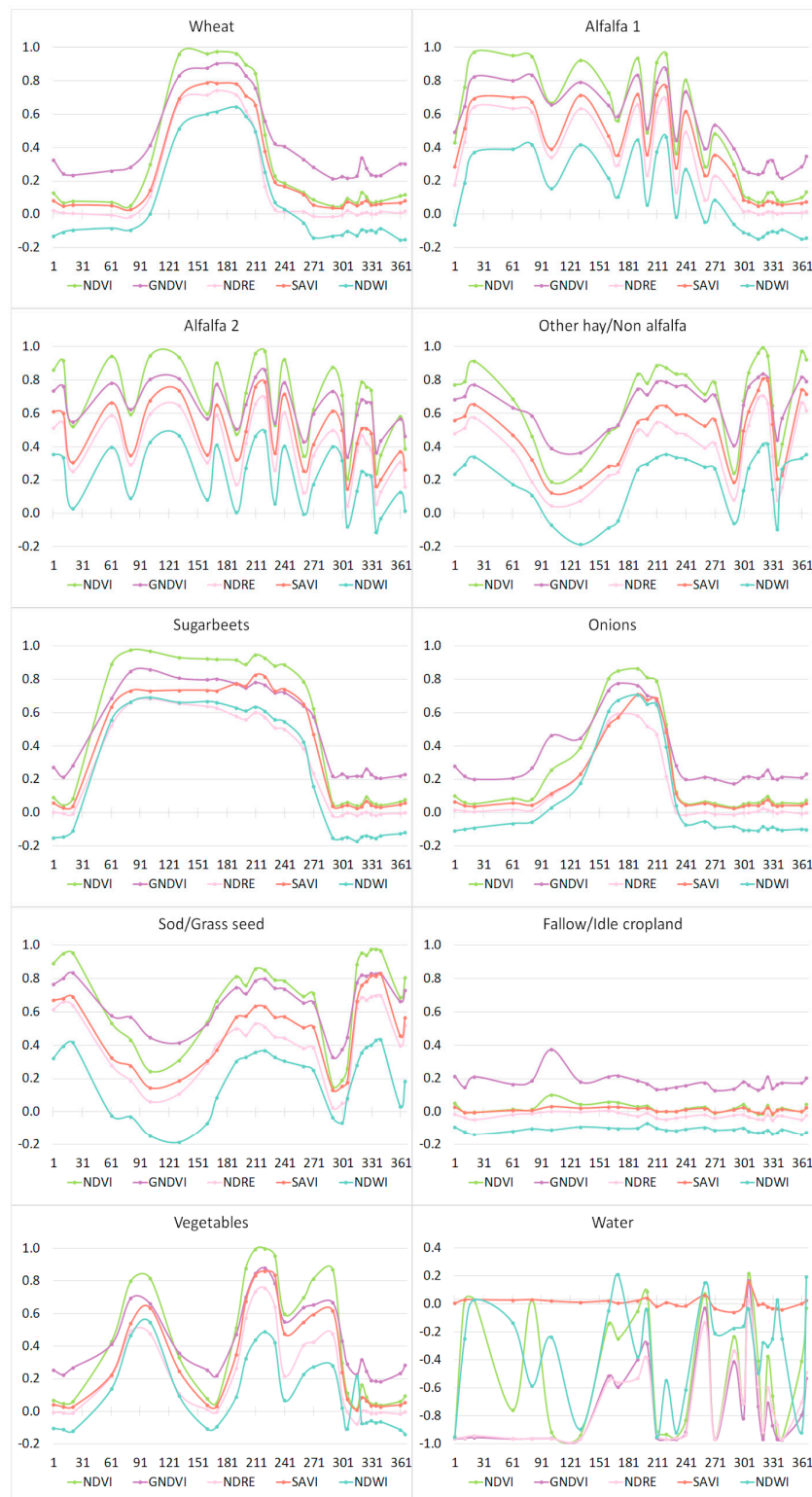
Figure 4. The temporal patterns of the classes analyzed for TA1, with values of five vegetation indices shown on the vertical axis. We analyzed a single agricultural year, the horizontal axis shows the day-of-the-time-series, namely from 23 September 2016 to 23 September 2017 (365 days). Two temporal patterns were identified for alfalfa (1 and 2). The vegetation indices are the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Red-Edge (NDRE), Soil Adjusted Vegetation Index (SAVI), and Normalized Difference Water Index (NDWI), derived as shown in Table 1.

Figure 5. The temporal patterns of the classes analyzed for TA2, with values of five vegetation indices shown on the vertical axis (NDVI, GNDVI, NDRE, SAVI, and NDWI). We analyzed a single agricultural year, the horizontal axis shows the day-of-the-time-series, namely from 18 October 2016 to 2 November 2017 (380 days). Two temporal patterns were identified for winter wheat (wheat 1 and 2).

For TA2, winter wheat presents two temporal patterns, with single crop in spring or double cropping cycles in late autumn and spring (Figure 5). Corn is a summer crop, similar with the cotton, but with longer cropping cycle and higher NDVI values than cotton. The temporal pattern of alfalfa shows 5 cropping cycles, longer during the cold season and shorter during the summer. The grass/pasture class shows two cropping cycles during spring and summer, but with values of

NDVI lower than 0.45. The last class analyzed, fallow/idle cropland, shows values of NDVI lover than 0.2 throughout the period analyzed, with an increase at the end (Figure 5).

2.2. Segmentation of Time Series

2.2.1. Multiresolution Segmentation of SITS

For the segmentation process, we created a time series of the 10 m spatial resolution bands (blue, green, red, and near-infrared) to take advantage of the best spatial resolution possible supported by Sentinel-2. We used an approach based on the multiresolution segmentation (MRS) [54], a region-growing algorithm that starts from the pixel level and aggregates pixels based on homogeneity conditions set by the user. The most important is the scale parameter (SP), which dictates the size of the object and, therefore, the internal homogeneity of the object. In order to objectively select the SP, we used the ESP2 tool (Estimation of Scale Parameter 2) [55], a method that automatically selects the appropriate SPs based on the local variance of objects across multiple scales [56,57]. We used a bottom-up hierarchical segmentation and selected the Level 2 of the results, which visually fitted best with the target crop parcels.

Segmenting multiple images from time series can be memory and time-consuming. Since the boundaries of the crop parcels can change their geometry over the timeframe analyzed, we selected all cloud free images available for each test area: 26 images for TA1 and 18 images for TA2. This resulted in 104 bands for TA1 and 72 bands for TA2 to be used in the segmentation process.

2.2.2. Segmentation Accuracy Assessment

The geometry of the agricultural fields delineated through segmentation has been evaluated using five segmentation accuracy metrics [58]. For TA1, we manually digitized 200 polygons as reference objects. For TA2 we digitized 681 polygons, representing all crop parcels with center-pivot irrigation, in order to have a complete overview of the segmentation performance in terms of accuracy. The manual delineation of the reference polygons considered every possible change of geometry through the periods analyzed, for example, a split of a larger parcel into smaller parcels for TA1 or a split of a center-pivot irrigation parcel into smaller slices in TA2. A minimum percent overlap of 50% between reference polygons and segmented objects needs to be achieved in order to match two corresponding objects [59]. The segmentation accuracy measures used are area fit index (AFI) [60], over-segmentation (OS), under-segmentation (US), root mean square (D) [61], and quality rate (QR) [62] (Table 3). Except for QR, the closer the values of the computed metrics are to 0, the better is the geometric accuracy of the resulting segments i.e., delineated agricultural fields. In case of QR, the closer its value is to 1, the better is the geometric correspondence between delineated objects and the reference objects (Table 3).

Table 3. Segmentation accuracy measures, their equations, ideal values and their range (C represents the total area of evaluated objects; R denotes the total area of reference objects). Area fit index (AFI), over-segmentation (OS), under-segmentation (US), root mean square (D), quality rate (QR).

Measure	Equation	Ideal Value	Range
Over-segmentation	$OS = 1 - (C \cap R) \div R$	0	[0, 1]
Under-segmentation	$US = 1 - (C \cap R) \div C$	0	[0, 1]
Area fit index	$AFI = (R - C) \div R$	0	OS: AFI > 0; US: AFI < 0
Root mean square	$D = \sqrt{(OS^2 + US^2) \div 2}$	0	[0, 1]
Quality rate	$QR = 1 - (C \cap R) \div (C \cup R)$	1	[0, 1]

2.3. Dynamic Time Warping

2.3.1. Theoretical Background

By combining time warping with distance measurements between elements of sequences, DTW is a nonlinear warping algorithm that decomposes a complex global optimization problem into local optimization problems using dynamic programming principle [23]. DTW has the advantage of time flexibility and was proven to achieve better results than the Euclidean distance measure for MODIS NDVI time series clustering [63]. Euclidean distance is a time-rigid method of comparing two sequences where each element of a sequence is compared with its associated element from the other sequence. DTW, on the other hand, is a time-flexible method that considers the temporal distortions of the time series (Figure 6), which can be associated with amplitude scaling or translation, time scaling or translation, shape changes, or noise in the data [64].
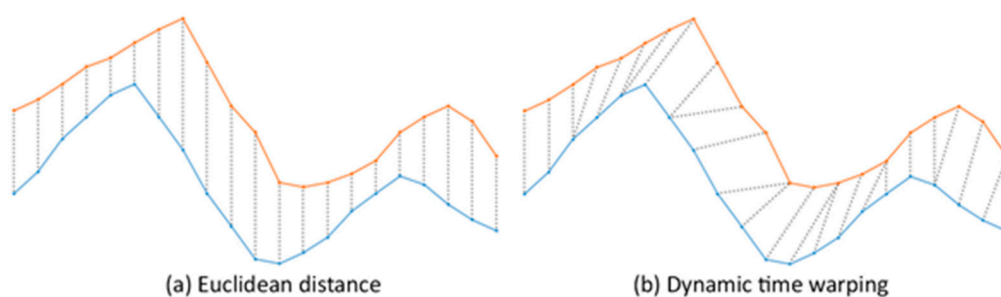

(a) Euclidean distance   (b) Dynamic time warping

Figure 6. Comparison between two sequences: (**a**) while Euclidean distance is time-rigid, (**b**) the dynamic time warping (DTW) is time-flexible in dealing with possible time distortion between the sequences. This flexibility is desirable for crop mapping, to deal with the intra-class phenological discrepancies caused by different environmental conditions.

The DTW algorithm works as follows. Suppose we have two sequences, $A = (a_1, \ldots, a_S)$ and $B = (b_1, \ldots, b_T)$ of length S and T respectively. Let δ be a distance between coordinates of sequences and m[S,T] a cost matrix. For single-band DTW, where each element of the sequence is a one-dimensional vector, we used δ as the squared Euclidean distance, in accordance with previous studies (Equation (1)) [11,65]. For multi-band DTW, where each element of the sequence is a multi-dimensional vector, we used a 5-dimensional δ in the computation of DTW. This is preferable to computing DTW individual for every dimension and then merge the results and gives a single alignment of the two multi-dimensional sequences compared. Using a multi-dimensional δ in computing DTW is recommended for remote sensing applications, where the dimensions are correlated phenomena [11]. For multi-band DTW we used the Euclidean distance to compare the elements of time series, as in Equation (2).

$$\delta\left(a_i,\ b_j\right) = \left(a_i - b_j\right)^2,\tag{1}$$

where a_i and b_j are elements of one-dimensional sequences A and B.

$$\delta\left(a_i,\ b_j\right) = \sqrt{\sum_{f=1}^{F}\left(a_{if} - b_{jf}\right)^2},\tag{2}$$

where a_{if} and b_{jf} are elements of the feature f of multi-dimensional sequences A and B, with F number of features.

Computing the alignment and comparison of the two sequences can be done recursively using Equation (3). Briefly, the computation of the entire DTW matrix starts by generating the first column and the first row, then the matrix is computed from left to right and from top to bottom, for each element being required the minimum between the upper, left, and the diagonal elements of the matrix [11]. In the end, the last element of the matrix from the bottom right indicates the degree of dissimilarity

between the two compared sequences, i.e., phenological temporal patterns. The smaller the DTW dissimilarity between two sequences, the similar are the two compared sequences.

$$D\left(A_i,\ B_j\right) = \delta\left(a_i,\ b_j\right) + min\begin{cases} \left(A_{i-1}, B_{j-1}\right) \\ \left(A_{i-1},\ B_j\right) \\ \left(A_i,\ B_{j-1}\right) \end{cases} \tag{3}$$

where $\left(A_{i-1}, B_{j-1}\right)$, $\left(A_{i-1},\ B_j\right)$, and $\left(A_i,\ B_{j-1}\right)$ are the upper left, upper and left neighboring elements of the $\left(A_i,\ B_j\right)$ element of the DTW matrix.

2.3.2. Time-Constrained DTW

Computing the entire DTW matrix can be a time-intensive procedure, especially when data rich sequences are compared. Furthermore, comparing an element of a sequence with all other elements of another sequence can lead to erroneous results, especially for crop mapping (e.g., aligning a winter crop with a summer crop). Therefore, applying global constraints on time warping have proven to speed up the processing, while providing meaningful results [65]. Two of the most-known global constraints are the Sakoe–Chiba band [23] and the Itakura parallelogram [66], which limits the warping path to a certain band around the diagonal of the matrix or allowing more distortions in the middle of the matrix, respectively.

For SITS data analysis, these global constraints might not be suitable mainly because the satellite images are irregularly sampled and consequently the defined constraint might lack a temporal meaning [11]. A more appropriate approach would be to limit the computation of the DTW matrix to a predefined maximum time delay, w. An element of the matrix will be computed only if the date difference ($date_{diff}$), which is a function returning the difference in time between the two dates of the compared elements of sequences, is smaller or equal to w. In computation of the DTW matrix, this condition must be evaluated beforehand. If it is true, the procedure remains unchanged, otherwise the element of the matrix is set to $+\infty$ [11]. The difference between a time-constrained and a time-weighted DTW relates to the fact that while for time-constrained DTW all elements found within w time delay are considered in computation, for time-weighted DTW a temporal cost is added using a linear or a logistic model [18].

Two special cases may arise when choosing the time-constraint on the warping path. First, when $w = 0$ the computation is limited to the diagonal elements, resembling the Euclidean distance. Second, when w is higher than the temporal length of the longest sequence the entire DTW matrix is computed. Therefore, the value of w needs to be adjusted to the characteristics of SITS analyzed to make sure that a warping path exists [11]. The procedure of computing a time-constrained DTW is described in Algorithm 1. Examples of two DTW matrices for TA1 and TA2 with 45-day time-constraints are depicted in Figure 7.

We compared different single-band and multi-band DTW classifications after applying time constraints of 15, 30, 45, and 60 days, respectively (for the remainder of the article, the terms DTW15, DTW30, DTW45, and DTW60 will be used). These values were selected to consider the diversity of crop cycles lengths, in accordance with the temporal sampling of the SITS used. Values higher than 60 days would already allow the alignment of winter crops with summer crops. Furthermore, we compared these results with those obtained by computing the entire DTW matrix (DTWFull) and computing the Euclidean distance ($w = 0$) (DTW0).

Algorithm 1. Computation of time-constrained DTW matrix.

|	
$date_{diff} = abs(date_i - date_j)$	#compute difference in time between the ith and j^{th} images
$w = maximum\ time\ delay$	#maximum time delay for the warping path
$m[1,1] = \delta(a_1, b_1)$	#compute first element of the matrixh
$for\ j\ in\ range\ (2, T)\ do$	#compute the rest of the first line of the matrix
$\quad if\ date_{diff} \leq w:$	#verify if the time difference falls within the allowed w
$\qquad m[i,j] = \delta(a_1, b_j) + m[1, j-1]$	
$\quad else: \quad m[i,j] = +\infty$	
$for\ i\ in\ range\ (2, S)\ do$	#compute the rest of the matrix, line by line, within w size
$\quad for\ j\ in\ range\ (1, T)\ do$	
$\qquad if\ date_{diff} \leq w:$	
$\qquad\quad m[i,j] = \delta(a_i, b_j) + min\{m[i-1, j],\ m[i,\ j-1],\ m[i-1,\ j-1]\}$	
$\qquad else: \quad m[i,j] = +\infty$	
$return\ m[S, T]$	#return the last element of the matrix, as DTW dissimilarity
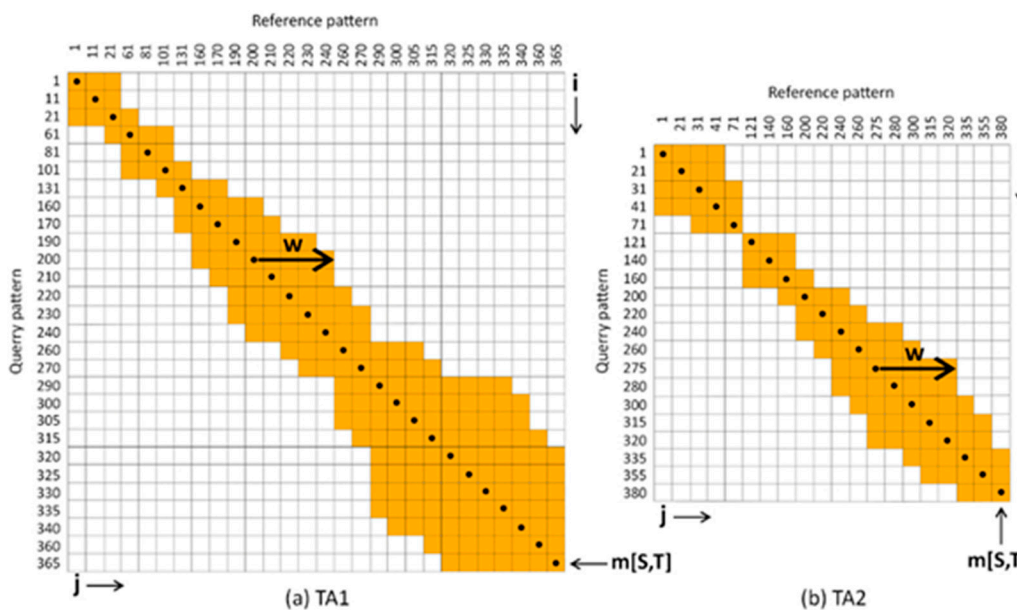
Figure 7. Computing the alignment between two sequences of TA1 (**a**) and TA2 (**b**). The vertical and horizontal values represent the date of an image from SITS, starting from 1 to 365 for TA1 and from 1 to 380 for TA2. Indices *i* and *j* are used to parse the matrix by line and by column, respectively. In these two examples, a maximum time delay, *w*, of 45 days is depicted, meaning that only the elements of the matrix who fall within this condition (orange) are computed. With black dots is represented the main diagonal of the DTW matrix (resembling Euclidean distance). After computing the matrix from upper left to lower right, the last element of the matrix, *m[S,T]*, is returned, as a measure of DTW dissimilarity between the two compared sequences.

2.3.3. Implementation of the Object-based DTW for SITS

The object-based DTW workflow is implemented into eCognition Developer software, version 9.4 (Trimble Geospatial), a commercial software package for OBIA applications. The object-based customized algorithm works on a stack of layers (e.g., vegetation indices), a complete segmentation of the scene into objects and reference samples for each class of interest. An object is classified into the class to which it had the minimum DTW dissimilarity value. Hence, the time complexity of the algorithm is $O(S \times T)$, where *S* and *T* are the sizes of the sequences compared.

2.3.4. Classification Accuracy

Various studies addressed the issues of evaluating the accuracy measures of maps, with descriptions, advantages, limitations, or extensions, offering suggestions for an efficient and

objective use within studies [67,68]. In this study, we use the error matrix and its related accuracy measures, namely the producer's (PA) and user's accuracy (UA) and overall accuracy (OA) [69]. Furthermore, we acknowledge the fact that using reference and validation samples from CropScape may introduce additional errors in the classifications. We compared our PAs, UAs, and OAs in conjunction with those available from CropScape for individual crops, at the state level. Although we heavily altered the samples obtained from CropScape to better match the reality of our study areas, thus increasing the accuracy of the samples, we combined the errors from our analysis with the errors from CropScape by computing the square root of the sum of the two squared errors, in order to see the maximum potential errors for PAs, UAs, and OAs.

3. Results

3.1. Segmentation Results

Executing ESP2 with a shape parameter of 0.1 and compactness of 0.5 resulted in an SP of 241 and 3,492 objects for TA1 and an SP of 307 and 1,585 objects for TA2 (Table 4). Good segmentation results were achieved for both test areas, with QR metric of 0.91 for TA1 and 0.82 for TA2, and very low US values (0.04 and 0.02, respectively). The positive values for AFI, as well as 0.05 and 0.16 for OS, denoted that the images were slightly over-segmented for both test areas (Table 4).

Table 4. Segmentation accuracy results depicting the scale parameter (SP), number of objects, over-segmentation (OS), under-segmentation (US), area fit index (AFI), root mean square (D), and quality rate (QR) metrics, as detailed in Table 3.

Test Area	SP	Objects	OS	US	AFI	D	QR
TA1	241	3492	0.05	0.04	0.01	0.05	0.91
TA2	307	1585	0.16	0.02	0.14	0.11	0.82

3.2. Overall Classification Accuracies

The computation time of the evaluated object-based DTW classifications was approximately 1 hour for TA1 and 15 minutes for TA2. For both single-band and multi-band DTW, the pattern of OAs values followed the same asymmetrical concave downward form, with peaks reached around 30 and 45 days of constraint (Table 5). For TA1, both single-band and multi-band DTW achieved highest OAs for time-constraints of 30 and 45 days, with 79.5% and 85.6% overall accuracy, respectively. Lowest OAs were achieved for DTW0 and DTWFull, of 75.2% and 72.3% for single-band approach, and for DTW0, DTW15, and DTWFull, of 76.1%, 80.3%, and 82.3% for multi-band approach (Table 5). For TA2, the highest OAs were achieved for 45 days constraints for single-band approach (89.1%) and for 30 days constraints for multi-band DTW (87.6%). The lowest OAs for TA2 were for DTW0, DTW15, and DTWFull for single-band approach (86.6%, 86.7%, and 82.3%) and DTW60 and DTWFull for multi-band approach (85.5% and 83.3%) (Table 5). The OAs pattern highlighted the importance of time-constraints in DTW for crop classification.

Table 5. Classification accuracies for single-band and multi-band object-based time-constrained dynamic time warping (DTW) using 15, 30, 45, and 60 days. For comparison, the accuracies obtained using no time constraint (DTWFull) and using Euclidean distance (DTW0) are shown. With bold fonts are shown the DTW classifications with the highest overall accuracies.

Test Area	DTW Type	DTW0 (%)	DTW15 (%)	DTW30 (%)	DTW45 (%)	DTW60 (%)	DTWFull (%)
TA1	single-band	75.2	77.7	79.5	79.5	76.6	72.3
	multi-band	76.1	80.3	85.6	85.6	84.3	82.3
TA2	single-band	86.8	86.7	87.8	89.1	88.6	82.3
	multi-band	86.1	86.1	87.6	86.0	85.5	83.3

3.3. User's and Producer's Accuracies

The classification results with the highest OAs for both test areas are shown in Figure 8. These are DTW30 for single-band and multi-band for TA1, DTW45 for single-band and DTW30 for multi-band approach for TA2 (Figure 8). For complete error matrices for these four classifications, see Supplementary Tables S1 to S4.
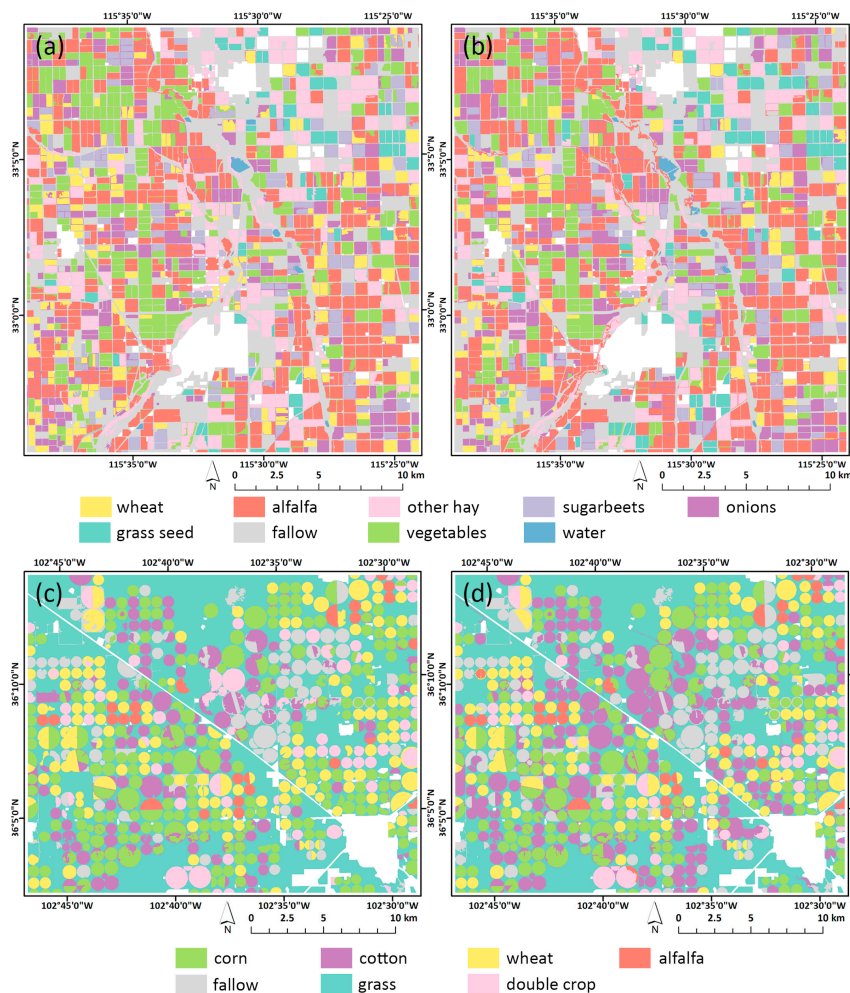

Figure 8. Best classification results for single-band (**a**) and multi-band DTW (**b**) for TA1 using in both cases a time constraint of 30 days (DTW30). Best classification results for single-band (**c**) and multi-band DTW (**d**) for TA2 using 45- and 30-day time constraints, respectively (DTW45 and DTW 30). For clarity, the developed/low to medium intensity areas are masked with white.

For TA1, single-band DTW30 yielded highest UAs for water (100%), alfalfa (98.2%), and onions (88.4%), while the lowest UAs were for vegetables (33.8%), wheat (56.5%), and other hay (65.9%) (Table 6). Vegetables were confused with alfalfa and this is understandable since the vegetables class contains multiple classes (lettuce, carrots, broccoli, or greens) with high intra-class diversity and number of cropping cycles that might resemble the temporal pattern of alfalfa (see Figure 4). For alfalfa, 98.2% of the validation samples were classified correctly, which is explained by the fact that alfalfa had no striking similarities with any other crop in terms of temporal pattern. Additionally, using two different reference patterns for alfalfa helped achieve the high UA value. Moreover, using 30 days flexibility in DTW allowed an accurate alignment of the high-internal diversity of temporal patterns for alfalfa. Regarding PAs, the highest values were obtained by fallow (100%), vegetables (97.8%), and sugarbeets (91.4%), while the lowest PAs were for sod/grass seed (51.2%), water (63.3%),

and alfalfa (77.7%) (Table 6). The sod/grass seed class was confused with the other hay/non-alfalfa class. What we classified partially as fallow was a water body and this can be explained by the presence of biomass in the water body which moved its temporal pattern towards that of fallow/idle cropland class (see Figure 4). For most crops, high intra-class heterogeneity of temporal patterns led to confusion with the complex patterns of alfalfa, which led to a PA of 77.7% for alfalfa. Vegetables had only one misclassified validation sample (97.8% PA), sign that its temporal pattern was well differentiated from the others (Figure 4).

For multi-band DTW30 for TA1, OA was higher with 6.1% than the single-band DTW30. This increase in value is attributable to a significant increase in UAs and PAs for some classes (Table 6). For example, using multiple vegetation indices increased the UA for wheat with 25.4% by reducing confusions with alfalfa, non-alfalfa, sugarbeets and onions. It also increased the UA for other hay with 8.4%, and sugarbeets with 7.7%. In case of PAs, the highest increase in PA was for water, with 23.3%, sod/grass seed with 13.4%, and alfalfa with 7.7% (Table 6). As expected, water was better classified when using multiple vegetation indices, of which some are more sensitive in detecting water content (e.g., NDWI).

Table 6. User's (UA) and producer's accuracy (PA) for most accurate single-band (DTW30) and multi-band (DTW30) object-based time-constrained DTWs for TA1, in California.

Crop	Single-Band DTW30		Multi-Band DTW30	
	UA	PA	UA	PA
Wheat	56.5	81.3	82.0	85.4
Alfalfa	98.2	77.7	98.8	85.4
Other hay/non-alfalfa	65.9	78.4	74.4	75.0
Sugarbeets	69.6	91.4	77.3	97.1
Onions	88.4	85.7	91.8	90.8
Sod/grass seed	82.4	51.2	77.9	64.6
Fallow/idle cropland	86.4	100	92.7	100
Vegetables	33.8	97.8	39.1	97.8
Water	100	63.3	100	86.7

For TA2, the single-band DTW45 achieved the highest UAs for wheat (98.1%), corn (96.0%), cotton (88.6%), and alfalfa (87.2%), while the lowest UAs were for fallow (66.7%), grass/pasture (77.3%), and double crop (79.7%) (Table 7). The fallow/idle cropland class was confused with cotton and wheat, while the grass/pasture class was confused with the cotton and fallow classes. For the latter case, this is because the temporal patterns of the three were mostly flat without prominent peaks (see Figure 5). The highest PAs were for alfalfa (100%), grass/pasture (97.7%), corn (92.3%), and wheat (90.6%), while the lowest PAs were for fallow (48.8%), cotton (82.8%), and double crop (85.0%) (Table 7). Notable class confusions occurred between the fallow and grass/pasture classes, and between the corn and cotton and double crop classes.

The multi-band DTW30 for TA2 had 1.5% lower OA than the single-band DTW45. Although it increased the UAs for double crop with 18.1% by reducing the confusion with corn and for grass/pasture with 6.7% by reducing the confusions with cotton and fallow, it decreased the UAs for cotton with 16.1% by increasing the confusion with corn and for fallow with 10.8% by increasing the confusion with wheat class (Table 7). Regarding PAs, multi-band DTW30 increased the PAs for cotton with 9.8% by reducing the confusion with grass/pasture and for fallow/idle cropland class with 9.7%, by reducing the confusion with grass/pasture and wheat. However, PAs decreased for double crop with 11.6% by slightly increasing confusion with multiple classes, for corn with 5.6% by increasing confusion with cotton crops, and for winter wheat with 3.5% by increasing the confusion with fallow class (Table 7). Overall, the lowest OA for multi-band DTW30 compared with single-band DTW45 for TA2 was attributable to two main increases in confusion between classes: between what we classified as cotton and was corn and what we classified as fallow that was winter wheat in the validation dataset.

Table 7. User's (UA) and producer's accuracy (PA) for most accurate single-band (DTW45) and multi-band (DTW30) object-based time-constrained DTWs for TA2, in Texas.

Crop	Single-Band DTW45		Multi-Band DTW30	
	UA	PA	UA	PA
Corn	96.0	92.3	96.5	86.8
Cotton	88.6	82.8	72.4	92.6
Wheat	98.1	90.6	98.0	87.1
Alfalfa	87.2	100	85.0	100
Fallow/idle cropland	66.7	48.8	55.8	58.5
Grass/pasture	77.3	97.7	84.0	97.7
Double crop	79.7	85.0	97.8	73.3

3.4. DTW Dissimilarity Results

The implemented workflow delivered two maps, a categorical classification of SITS (Figure 8) and a map of DTW dissimilarity values (Figure 9). The latter can be seen as a map of uncertainty for classification and depicts the DTW dissimilarity values of the classified objects, which represents the final values from the bottom right corner of the DTW matrix. This map is useful to understand which objects had a strong similarity with the assigned class (values closer to 0) or had a doubtful assignment to one of the target classes (the largest values). High values can be also a sign of high intra-class heterogeneity of temporal patterns. In our case, the DTW dissimilarity values for best single-band classifications ranged from 0 to 4 for DTW30 in TA1 and from 0 to 5.1 for DTW45 in TA2, with higher values corresponding to complex temporal patterns, like alfalfa in TA1 and double crop in TA2. Higher DTW dissimilarity values can also be attributed to the presence of other classes that were not considered in this study due to their unrepresentativeness and, therefore, had a distinct pattern than those of our target classes. For comparison purposes, Figure 9 shows the DTW dissimilarity values normalized using a min–max normalization approach.
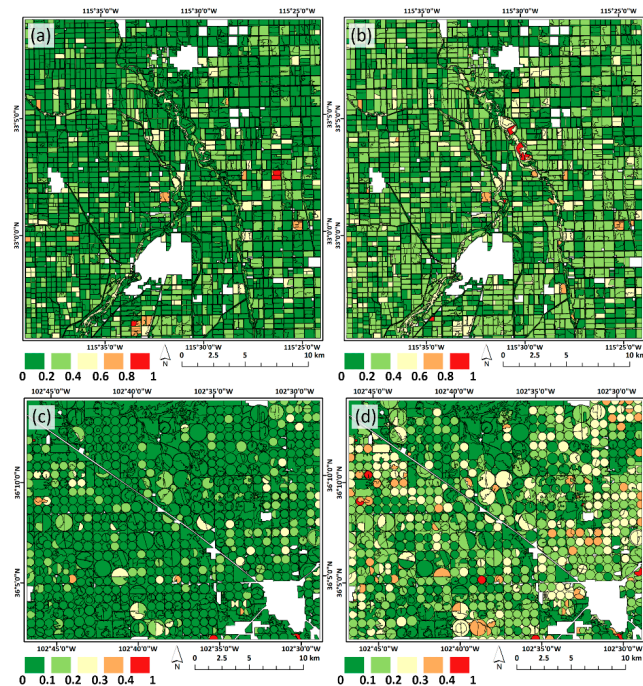

Figure 9. DTW dissimilarity values for single-band (**a**) and multi-band DTW (**b**) for TA1 using in both cases a time constraint of 30 days (DTW30). DTW dissimilarity values for single-band DTW45 (**c**) and multi-band DTW30 (**d**) for TA2 using 45- and 30-day time constraints, respectively. For clarity, the developed/low to medium intensity areas are masked with white.

Additionally, although we classified the objects into the class with the minimum DTW dissimilarity value, we stored for each object the DTW dissimilarity values resulted from comparing the objects' temporal pattern with all reference temporal patterns. In this way, we can better understand the uncertainty of classifications and confusions between classes. We plotted these values against each other, fitted a linear function and computed the R^2 for TA1 in Figure 10 and for TA2 in Figure 11. These results can be correlated with the corresponding UAs and PAs from Tables 6 and 7, respectively.

For single-band DTW30 in TA1, the highest R^2 values better explained the confusions between classes (Figure 10). For example, the low UA for vegetables (33.8%) presented an R^2 of 0.63 with alfalfa1 and 0.67 with sugarbeets. The low PA for water (63.3%) had an R^2 of 0.97 with the fallow class, while the low UA for sod/grass seed (51.2%) had a high R^2 of 0.96 with other hay/non alfalfa and 0.78 with alfalfa2. Other remarkable high R^2 values were for onions–wheat (0.86), sugarbeets–alfalfa1 (0.87), and other hay–alfalfa2 (0.84) (Figure 10). The 0.68 R^2 value between alfalfa1 and alfalfa2 suggested that it was useful to use two temporal patterns of alfalfa for classification.
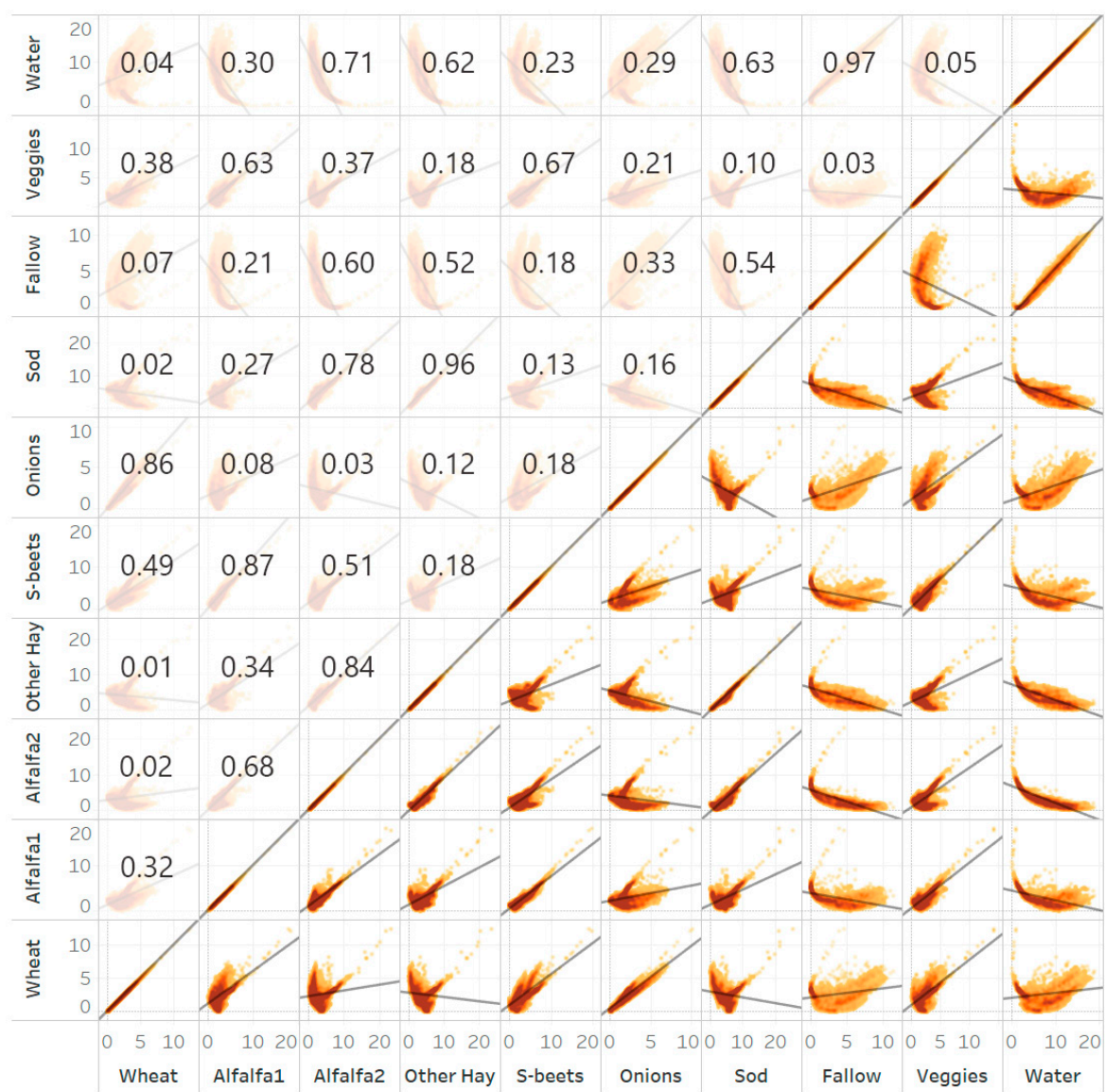

Figure 10. DTW dissimilarity values scatter plots computed for each class for the single-band DTW30 of California, with R^2 values in the upper left of the diagonal. Classes analyzed are wheat, alfalfa1, alfalfa2, other hay/non-alfalfa, sugarbeets, onions, sod/grass seed, fallow/idle cropland, vegetables, and water.

For single-band DTW45 in TA2, the low PA (48.8%) and UA (66.7%) for fallow/idle cropland can be explained by high R^2 values of fallow with grass/pasture (0.58) and wheat (0.46) (Figure 11). Part of the lower PA for double crop (85.0%) is due to an R^2 of 0.86 with alfalfa, while the DTW dissimilarity of cotton showed an R^2 of 0.86 and 0.67 for grass/pasture and corn, respectively (Figure 11). As for TA1, splitting winter wheat in two distinct patterns was useful, since the two patterns have an R^2 of 0.34.
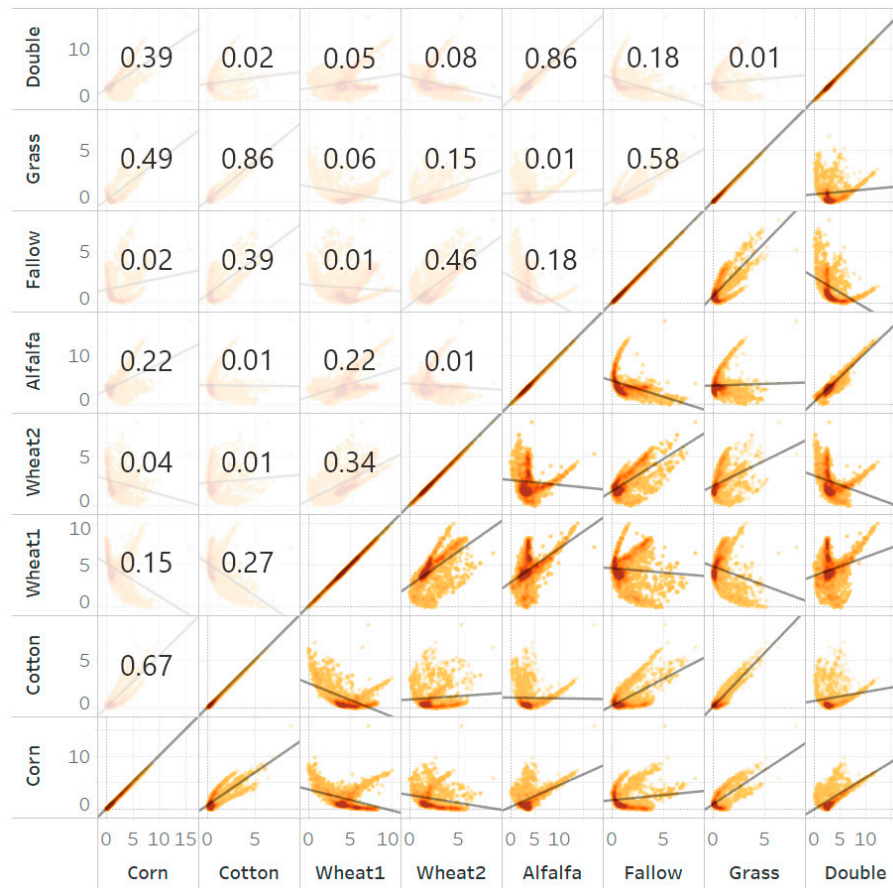

Figure 11. DTW dissimilarity values scatter plots computed for each class for the single-band DTW45 of Texas, with R^2 values in the upper left of the diagonal. Classes analyzed are corn, cotton, winter wheat1, winter wheat2, alfalfa, fallow/idle cropland, grass/pasture, and double crop.

3.5. Evaluation of Combined Uncertainties

Although we heavily altered the samples from CropScape and corrected them using multiple approaches, we acknowledged the propagation of errors into our analysis. Considering the CropScape errors at the state-level for California (8.7%) and Texas (22.1%) [70], and supposing we did not altered the samples, then the combined errors for OAs are 22.2% and 16.8% for single-band DTW30 and multi-band DTW30 for TA1, and 24.6% and 25.4% for single-band DTW45 and multi-band DTW30 for TA2.

For TA1 in California, most of our UAs and PAs errors were lower than those from state-level CropScape (Table 8). We had significant lower PAs errors for vegetables, fallow, sugarbeets, other hay, and wheat, while we also had lower UAs errors for alfalfa, onions, sod/grass seed, vegetables, and water. Our UAs errors for sugarbeets and PAs errors for alfalfa and water were higher than those from CropScape (Table 8).

Table 8. Errors for UAs and PAs for single-band DTW30 and multi-band DTW30 for California, compared with those from CropScape. With * are shown the CropScape errors at the scale of California and UA+ and PA+ represent the combined errors, shown in bold.

Crop	CropScape		Single-band DTW30				Multi-band DTW30			
	UA*	PA*	UA	UA+	PA	PA+	UA	UA+	PA	PA+
Wheat	27.6	31.9	43.5	**51.5**	18.7	**37.0**	18.0	**33.0**	14.6	**35.1**
Alfalfa	12.5	8.4	1.8	**12.6**	22.3	**23.8**	1.2	**12.6**	14.6	**16.8**
Other hay	34.5	46.4	34.1	**48.5**	21.6	**51.2**	25.6	**43.0**	25.0	**52.7**
Sugarbeets	14.0	53.3	30.4	**33.5**	8.6	**54.0**	22.7	**26.7**	2.9	**53.4**
Onions	28.6	21.4	11.6	**30.9**	14.3	**25.7**	8.2	**29.8**	9.2	**23.3**
Sod	34.7	52.8	17.6	**38.9**	48.8	**71.9**	22.1	**41.1**	35.4	**63.6**
Fallow	19.5	19.2	13.6	**23.8**	0	**19.2**	7.3	**20.8**	0	**19.2**
Vegetables	79.9	54.9	66.2	**103.8**	2.2	**54.9**	60.9	**100.5**	2.2	**54.9**
Water	7	7.8	0	**7.0**	36.7	**37.5**	0	**7.0**	13.3	**15.4**

For TA2 in Texas, we got lower UAs errors than CropScape state-level for corn, winter wheat, grass, and double crop and lower PAs errors for all classes, except for cotton from single-band approach (Table 9). We had slightly higher UAs errors for fallow and cotton from the multi-band approach and higher PAs errors for cotton from the single-band DTW approach (Table 9).

Table 9. Errors for UAs and PAs for single-band DTW45 and multi-band DTW30 for Texas (TA2), compared with those from CropScape. With * are shown the CropScape errors at the scale of Texas and UA+ and PA+ represent the combined errors, shown in bold.

Crop	CropScape		Single-band DTW45				Multi-band DTW30			
	UA*	PA*	UA	UA+	PA	PA+	UA	UA+	PA	PA+
Corn	9.7	16.2	4.0	**10.5**	7.7	**17.9**	3.5	**10.3**	13.2	**20.9**
Cotton	16.6	10.7	13.4	**21.3**	17.2	**20.3**	27.6	**32.2**	7.4	**13.0**
Wheat	16.9	13.5	1.9	**17.0**	9.4	**16.5**	2.0	**17.0**	12.9	**18.7**
Alfalfa	13.0	26.1	12.8	**18.2**	0	**26.1**	15.0	**19.8**	0	**26.1**
Fallow	30.5	57.0	33.3	**45.2**	51.2	**76.6**	44.2	**53.7**	41.5	**70.5**
Grass	30.7	17.1	22.7	**38.2**	2.3	**17.3**	16.0	**34.6**	2.3	**17.3**
Double	75.2	83.9	20.3	**77.9**	15	**85.2**	2.2	**75.2**	26.7	**88.0**

4. Discussion

Our study reported on the first operational implementation of DTW as a ready-to-use workflow in an OBIA environment focused on crop mapping applications that yielded high overall accuracies. We compared single-band DTW (using NDVI time series) with multi-band DTW (using NDVI, GNDVI, NDRE, SAVI, and NDWI) and found that multi-band significantly increase the overall accuracy for California, while slightly decreasing it for Texas. Furthermore, we evaluated the impact of time constraints in DTW over the final classification accuracy and had best results using time delays of 30 and 45 days. Poorest results were obtained by Euclidean distance or DTW without time constraints.

As reported in Belgiu and Csillik [26], an object-based DTW has the advantage of generating vector digital spatial data that can be further used in agricultural monitoring activities. Therefore, it can successfully support (if not replace) labor-intensive and error-prone manual digitization of crop fields from satellite imagery. Furthermore, the object-based DTW reported here reduced the computational time required by DTW to classify target classes from SITS, since it uses objects rather than pixels as input spatial analysis units. As a comparison, Belgiu and Csillik [26] analyzed a similar region in size as TA1 but with fewer images (21) using a pixel-based approach with *dtwSat* package [25] in R and the execution time was 30 hours without the segmentation step.

4.1. Segmentation of SITS

Segmentation of satellite images is one of the main challenges of an OBIA approach [71]. A very good indicator of the spatial match between delineated agricultural fields and reference objects is represented by the low value of the US metric (i.e., 0.04 for TA1 and 0.02 for TA2). The US is considered the most important indicator of the segmentation accuracy [55,72] since it is difficult to re-process and classify correctly the initially under-segmented objects [73]. OS value was higher for TA2 since the circle-pivot irrigation parcels had mainly two sizes, of 50 or 200 ha, the bigger ones being over-segmented. As expected, the Sentinel-2 time series provided satisfactory agricultural field delineation capabilities especially in regions with smaller fields [73].

The two test areas presented a high complexity of cropping cycles and, therefore, changes in the geometry of parcels throughout the entire time frame analyzed. Incorporation of multi-temporal images in the segmentation procedure allowed the delineation of homogeneous objects through time. We used all cloud free 10 m spatial resolution bands in the segmentation process. Other solutions may require either a knowledge-driven selection of images or a statistical selection (e.g., principal component analysis). Furthermore, combining multiresolution segmentation with superpixels can help in reducing the resources and time needed for the segmentation process [74].

4.2. The Importance of Time Constraints in DTW for Crop Mapping

In both test areas, the best classification accuracies were obtained with maximum time delays of 30 and 45 days in aligning the temporal sequences using DTW. These delays proved to be adequate, not being too rigid and not allowing too much freedom in computing the DTW matrix, thus avoiding inconsistent matching of crops. Moreover, achieving best results using time-constraints of 30 and 45 days showed that within the two test areas the temporal variations of cropping cycles are around this timeframe. Longer time-constraints (e.g., 100 days) were found best by Maus et al. [18] when analyzing multi-year time series of MODIS Enhanced Vegetation Index (EVI). Using time delays in DTW outperformed both classifications using Euclidean distance and DTW without time constraint. The time constraints in a DTW classification should be knowledge-based and set in accordance with the temporal sampling of the images from SITS (see also Figure 7) [11]. We achieved better results when using a time constraint and this is in line with the results obtained by Petitjean et al. [11] and Maus et al. [18], who applied DTW at the pixel-level for classifying FORMOSAT-2 SITS and MODIS SITS, respectively.

4.3. The Advantages of DTW

Selection of the appropriate classifier is very important in any image classification scenario [75]. The workflow described here used the DTW method due to its already demonstrated advantages to identifying classes of interest from SITS [11,18,26,76]. Firstly, DTW can deal with temporal variations of phenological patterns caused by differences in the agricultural practices or by different weather conditions. This is a very important asset when vegetation dynamics, i.e., phenology, are used for the classification purpose [31]. Secondly, DTW is not very sensitive to the number of reference samples, as already proven by Belgiu and Csillik [26], as long as their temporal pattern is characteristic for each target class. No training samples are needed if one performs clustering of temporal patterns into a predefined number of clusters and then use expert knowledge or other rules to assign meaningful crop classes. These are large advantages for mapping crops at larger extents or for countries with a lack of reference samples [77].

4.4. Single-band and Multi-band DTW

NDVI was used for the single-band DTW approach because of its power in identifying crop classes. For example, Peña-Barragán et al. [78] found NDVI to have a prediction power of around 50% amongst 20 ASTER vegetation indices for crop identification. For multi-band approach, we added

GNDVI, NDRE, SAVI, and NDWI, computed using most of the Sentinel-2 spectral bands. Red-edge bands proved to be very efficient for different vegetation mapping applications due to their potential to estimate vegetation chlorophyll content [48,79,80]. The suitability of red-edge and short-wave infrared (SWIR) bands of Sentinel-2 had been shown in different applications including chlorophyll content and leaf area index (LAI) estimation of various crops including corn, sugarbeet, onion, garlic, and wheat [81] and LAI estimation of wheat and potatoes crops [82], estimation of crops and grass chlorophyll and nitrogen content [83], and crops and tree species mapping [84]. SWIR spectral bands quantify canopy reflectance in relation to moisture content and their importance for different vegetation mapping applications was proven in previous studies [85].

Using a multi-band DTW increased the accuracy by 6.1% in California by reducing confusion between alfalfa and vegetables, alfalfa and sugarbeets, and other hay and sod/grass seed. In Texas, multi-band DTW had 1.5% lower accuracy than single-band DTW. In the latter case, although it reduced many instances of class confusion, it increased confusion between corn and cotton, two classes that shared a similar pattern and the usage of multiple vegetation indices might have smoothed their temporal profiles. Unlike other classifiers, DTW does not compute a variable importance measure for multi-band DTW so we cannot have a deeper knowledge of the importance of each vegetation index used. However, DTW computed a measure that can act as a measure of confidence for the classification, namely the DTW dissimilarity value [25].

4.5. DTW Dissimilarity and Classification Accuracies

We used the DTW dissimilarity values to understand class confusions and see what parcels had temporal patterns different to those we had as reference patterns. Furthermore, the DTW dissimilarity can be a valuable source in understanding spatio-temporal autocorrelation within time series of satellite images. Some possible applications are related to refinement of classification or segmentation results by finding outliers, clustering of seasonal crops or as a feature to be used in other machine learning classifiers. More research is needed on this topic in order to understand the implications of DTW in measuring spatio-temporal autocorrelation.

A variety of errors are encountered when dealing with land use/land cover classification [86,87], especially for global land cover mapping [88,89]. The majority of our UAs and PAs for both test areas were higher than those obtained by CropScape at the state-level [42,70], which make our DTW approach suitable to be applied at larger scales for land use/land cover mapping. Furthermore, we investigated the error propagations and combined the errors from the samples obtained from CropScape with our final DTW classification results. Considering the complexity of our study areas, we see the combined errors of 22.2% and 16.8% for California and 24.6% and 25.4% for Texas as very good. Immitzer et al. [84] obtained a similar error of 23.2% when using Sentinel-2 bands for mapping 7 crops in Southern Germany, with vegetables and sunflower classes having the lowest UAs and PAs. Belgiu and Csillik [26] had a classification error of 21.9% when analyzing a similar area as we did in California using a time-weighted DTW. Lambert et al. [90] mapped crop types at smallholder level with 20% classification errors and stated that red edge and near infrared bands of Sentinel 2 were the most important features in the classification.

4.6. Limitations and Future Developments

Besides advantages discussed above, it is worth mentioning some current limitations of the proposed methodology. The current approach allows the detection of crops from a single agricultural year. If multi-year crop classification is required, the workflow can be applied for each year separately. Different measures can be used inside DTW to assess the similarity between two temporal sequences. We implemented the default Euclidean distance measure. Yet, there are other studies that proposed the utilization of Canberra distance instead [91], which can be seen as a weighted Manhattan distance, performing better than the classic Euclidean DTW distance in Landsat time series clustering. Therefore, further distance measures need to be tested carefully in the future. To further improve the

computational speed, various temporal constraints, or bounds on warping window size proposed in the literature [28,30,92,93] will be evaluated for various SITS classifications scenarios.

Every crop modeling approach should have a well-defined domain of relevance, a mechanistic framework and an evaluation of the new model [94]. Besides achieving this, our proposed workflow is also meeting other criteria that guarantee a successful model, like simplicity, generality, robustness, accuracy, and fruitfulness [95].

5. Conclusions

This work reports on single-band and multi-band object-based time-constraint DTW workflow in an OBIA environment for crop mapping using Sentinel-2 time series data. Our findings represent a helpful step towards generating ready-to-use spatial data products that describe agricultural fields. For testing purposes, we used two complex agricultural landscapes from California and Texas, achieving satisfactory classification accuracy when time delays of 30 and 45 days were used for DTW. This outperformed the classifications using the Euclidean distance or DTW without time constraints. These kinds of approaches are increasingly necessary as we move into imagery-dense workflows with satellite constellations providing increasing spatial and temporal resolutions. This work demonstrated its efficiency for crop mapping. Nevertheless, other applications such as deforestation monitoring, disaster mapping or mapping of tree species can successfully benefit from using this approach.

References

1. Gibbs, H.K.; Ruesch, A.S.; Achard, F.; Clayton, M.K.; Holmgren, P.; Ramankutty, N.; Foley, J.A. Tropical Forests were the Primary Sources of New Agricultural Land in the 1980s and 1990s. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16732–16737. [CrossRef] [PubMed]
2. Deininger, K.; Byerlee, D. *The Rise of Large Farms in Land Abundant Countries: Do they have a Future?* Policy Research Working Papers; The World Bank: Washington, DC, USA, 2011.
3. Alexandratos, N.; Bruinsma, J. *World Agriculture towards 2030/2050: The 2012 Revision*; ESA Working Paper; FAO: Rome, Italy, 2012.
4. Foley, J.A.; Defries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K.; et al. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [CrossRef]
5. Atzberger, C. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sens.* **2013**, *5*, 949–981. [CrossRef]
6. Bégué, A.; Arvor, D.; Bellon, B.; Betbeder, J.; De Abelleyra, D.; P. D. Ferraz, R.; Lebourgeois, V.; Lelong, C.; Simões, M.; R. Verón, S. Remote Sensing and Cropping Practices: A Review. *Remote Sens.* **2018**, *10*, 99.
7. Justice, C.O.; Townshend, J.R.G.; Vermote, E.F.; Masuoka, E.; Wolfe, R.E.; Saleous, N.; Roy, D.P.; Morisette, J.T. An Overview of MODIS Land Data Processing and Product Status. *Remote Sens. Environ.* **2002**, *83*, 3–15. [CrossRef]
8. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The Global Landsat Archive: Status, Consolidation, and Direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [CrossRef]

9.   Yan, L.; Roy, D.P. Automated Crop Field Extraction from Multi-temporal Web Enabled Landsat Data. *Remote Sens. Environ.* **2014**, *144*, 42–64. [CrossRef]

10. Chen, J.; Dowman, I.; Li, S.; Li, Z.; Madden, M.; Mills, J.; Paparoditis, N.; Rottensteiner, F.; Sester, M.; Toth, C.; et al. Information from Imagery: ISPRS Scientific Vision and Research Agenda. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 3–21. [CrossRef]

11. Petitjean, F.; Inglada, J.; Gancarski, P. Satellite Image Time Series Analysis Under Time Warping. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3081–3095. [CrossRef]

12. Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting Trend and Seasonal Changes in Satellite Image Time Series. *Remote Sens. Environ.* **2010**, *114*, 106–115. [CrossRef]

13. Jönsson, P.; Eklundh, L. TIMESAT—A Program for Analyzing Time-Series of Satellite Sensor Data. *Comput. Geosci.* **2004**, *30*, 833–845. [CrossRef]

14. Eerens, H.; Haesen, D.; Rembold, F.; Urbano, F.; Tote, C.; Bydekerke, L. Image Time Series Processing for Agriculture Monitoring. *Environ. Model. Softw.* **2014**, *53*, 154–162. [CrossRef]

15. Jonsson, P.; Eklundh, L. Seasonality Extraction by Function Fitting to Time-Series of Satellite Sensor Data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1824–1832. [CrossRef]

16. Verbesselt, J.; Hyndman, R.; Zeileis, A.; Culvenor, D. Phenological Change Detection while Accounting for Abrupt and Gradual Trends in Satellite Image Time Series. *Remote Sens. Environ.* **2010**, *114*, 2970–2980. [CrossRef]

17. Roerink, G.J.; Menenti, M.; Verhoef, W. Reconstructing Cloudfree NDVI Composites using Fourier Analysis of Time Series. *Int. J. Remote Sens.* **2000**, *21*, 1911–1917. [CrossRef]

18. Maus, V.; Câmara, G.; Cartaxo, R.; Sanchez, A.; Ramos, F.M.; de Queiroz, G.R. A Time–Weighted Dynamic Time Warping Method for Land Use and Land Cover Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3729–3739. [CrossRef]

19. Zhong, L.; Hu, L.; Yu, L.; Gong, P.; Biging, G.S. Automated Mapping of Soybean and Corn using Phenology. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 151–164. [CrossRef]

20. Waldner, F.; Canto, G.S.; Defourny, P. Automated Annual Cropland Mapping using Knowledge based Temporal Features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 1–13. [CrossRef]

21. Arvor, D.; Meirelles, M.; Dubreuil, V.; Bégué, A.; Shimabukuro, Y.E. Analyzing the Agricultural Transition in Mato Grosso, Brazil, using Satellite derived Indices. *Appl. Geogr.* **2012**, *32*, 702–713. [CrossRef]

22. Kennedy, R.E.; Yang, Z.; Cohen, W.B. Detecting Trends in Forest Disturbance and Recovery using Yearly Landsat Time Series: 1. LandTrendr—Temporal Segmentation Algorithms. *Remote Sens. Environ.* **2010**, *114*, 2897–2910. [CrossRef]

23. Sakoe, H.; Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoust.* **1978**, *26*, 43–49. [CrossRef]

24. Petitjean, F.; Weber, J. Efficient Satellite Image Time Series Analysis Under Time Warping. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1143–1147. [CrossRef]

25. Maus, V.; Camara, G.; Appel, M.; Pebesma, E. dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. *J. Stat. Softw.* **2019**, *88*, 1–31. [CrossRef]

26. Belgiu, M.; Csillik, O. Sentinel-2 Cropland Mapping using Pixel-based and Object-based Time-Weighted Dynamic Time Warping Analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [CrossRef]

27. Rabiner, L.R.; Juang, B.-H.; Rutledge, J.C. *Fundamentals of Speech Recognition*; PTR Prentice Hall Englewood Cliffs: Upper Saddle River, NJ, USA, 1993; Volume 14.

28. Ratanamahatana, C.A.; Keogh, E. Everything You Know about Dynamic Time Warping is Wrong. In Proceedings of the Third Workshop on Mining Temporal and Sequential Data, Seattle, WA, USA, 22 August 2004; Citeseer: University Park, PA, USA, 2004; Volume 32.

29. Petitjean, F.; Ketterlin, A.; Gançarski, P. A Global Averaging Method for Dynamic Time Warping, with Applications to Clustering. *Pattern Recognit.* **2011**, *44*, 678–693. [CrossRef]

30. Esling, P.; Agon, C. Time-Series Data Mining. *ACM Comput. Surv.* **2012**, *45*, 12:1–12:34. [CrossRef]

31. Baumann, M.; Ozdogan, M.; Richardson, A.D.; Radeloff, V.C. 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.* **2017**, *54*, 72–83. [CrossRef]

32. Gómez, C.; White, J.C.; Wulder, M.A.; Alejandro, P. Integrated Object-Based Spatiotemporal Characterization of Forest Change from an Annual Time Series of Landsat Image Composites. *Can. J. Remote Sens.* **2015**, *41*, 271–292. [CrossRef]

33. Xue, Z.; Du, P.; Feng, L. Phenology-Driven Land Cover Classification and Trend Analysis Based on Long-term Remote Sensing Image Series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1142–1156. [CrossRef]

34. Guan, X.; Huang, C.; Liu, G.; Meng, X.; Liu, Q. Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance. *Remote Sens.* **2016**, *8*, 19. [CrossRef]

35. Camara, G.; Assis, L.F.; Ribeiro, G.; Ferreira, K.R.; Llapa, E.; Vinhas, L. Big Earth Observation Data Analytics: Matching Requirements to System Architectures. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, Burlingame, CA, USA, 31 October–3 November 2016; ACM: New York, NY, USA, 2016; pp. 1–6.

36. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; Van der Meer, F.; Van der Werff, H.; Van Coillie, F.; et al. Geographic Object-Based Image Analysis—Towards a New Paradigm. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 180–191. [CrossRef] [PubMed]

37. Bunting, P.; Clewley, D.; Lucas, R.M.; Gillingham, S. The Remote Sensing and GIS Software Library (RSGISLib). *Comput. Geosci.* **2014**, *62*, 216–226. [CrossRef]

38. Körting, T.S.; Garcia Fonseca, L.M.; Câmara, G. GeoDMA—Geographic Data Mining Analyst. *Comput. Geosci.* **2013**, *57*, 133–145. [CrossRef]

39. Trimble, T. *ECognition Developer 9.4 Reference Book*; Trimble Germany GmbH: Munich, Germany, 2018.

40. Messer, C.; Stock, C.; Averill, T. *California Agricultural Statistics Review 2015–2016*; California Department of Food and Agriculture: Sacramento, CA, USA, 2016.

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

42. Han, W.; Yang, Z.; Di, L.; Mueller, R. CropScape: A Web Service based Application for Exploring and Disseminating US Conterminous Geospatial Cropland Data Products for Decision Support. *Comput. Electron. Agric.* **2012**, *84*, 111–123. [CrossRef]

43. Muller-Wilm, U.; Louis, J.; Richter, R.; Gascon, F.; Niezette, M. Sentinel-2 Level 2A Prototype Processor: Architecture, Algorithms and First Results. In Proceedings of the ESA Living Planet Symposium, Edinburgh, UK, 9–13 September 2013; pp. 9–13.

44. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 Sen2Cor: L2A Processor for Users. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; pp. 9–13.

45. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]

46. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]

47. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

48. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the Capabilities of Sentinel-2 for Quantitative Estimation of Biophysical Variables in Vegetation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 83–92. [CrossRef]

49. Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [CrossRef]

50. Sims, D.A.; Gamon, J.A. Relationships between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sens. Environ.* **2002**, *81*, 337–354. [CrossRef]

51. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

52. Gao, B.-C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]

53. Ferrant, S.; Selles, A.; Le Page, M.; Herrault, P.-A.; Pelletier, C.; Al-Bitar, A.; Mermoz, S.; Gascoin, S.; Bouvet, A.; Saqalli, M.; et al. Detection of Irrigated Crops from Sentinel-1 and Sentinel-2 Data to Estimate Seasonal Groundwater Use in South India. *Remote Sens.* **2017**, *9*, 1119. [CrossRef]

54. Baatz, M.; Schäpe, A. Multiresolution Segmentation: An Optimization Approach for High Quality Multi-scale Image Segmentation. In *Angewandte Geographische Informationsverarbeitung*; Strobl, J., Blaschke, T., Griesebner, G., Eds.; Herbert Wichmann Verlag: Berlin, Germany, 2000; Volume 12, pp. 12–23.

55. Drăguţ, L.; Csillik, O.; Eisank, C.; Tiede, D. Automated Parameterisation for Multi-scale Image Segmentation on Multiple Layers. *ISPRS J. Photogramm. Remote Sens.* **2014**, *88*, 119–127. [CrossRef]

56. Drăguţ, L.; Tiede, D.; Levick, S.R. ESP: A Tool to Estimate Scale Parameter for Multiresolution Image Segmentation of Remotely Sensed Data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 859–871. [CrossRef]

57. Woodcock, C.E.; Strahler, A.H. The Factor of Scale in Remote Sensing. *Remote Sens. Environ.* **1987**, *21*, 311–332. [CrossRef]

58. Eisank, C.; Smith, M.; Hillier, J. Assessment of Multiresolution Segmentation for Delimiting Drumlins in Digital Elevation Models. *Geomorphology* **2014**, *214*, 452–464. [CrossRef] [PubMed]

59. Whiteside, T.G.; Maier, S.W.; Boggs, G.S. Area-based and Location-based Validation of Classified Image Objects. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 117–130. [CrossRef]

60. Lucieer, A.; Stein, A. Existential Uncertainty of Spatial Objects Segmented from Satellite Sensor Imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2518–2521. [CrossRef]

61. Clinton, N.; Holt, A.; Scarborough, J.; Yan, L.I.; Gong, P. Accuracy Assessment Measures for Object-Based Image Segmentation Goodness. *Photogramm. Eng. Remote Sens.* **2010**, *76*, 289–299. [CrossRef]

62. Winter, S. Location Similarity of Regions. *ISPRS J. Photogramm. Remote Sens.* **2000**, *55*, 189–200. [CrossRef]

63. Zhang, Z.; Tang, P.; Huo, L.; Zhou, Z. MODIS NDVI Time Series Clustering under Dynamic Time Warping. *Int. J. Wavelets Multiresolution Inf. Process.* **2014**, *12*, 1461011. [CrossRef]

64. Lhermitte, S.; Verbesselt, J.; Verstraeten, W.W.; Coppin, P. A Comparison of Time Series Similarity Measures for Classification and Change Detection of Ecosystem Dynamics. *Remote Sens. Environ.* **2011**, *115*, 3129–3152. [CrossRef]

65. Fu, A.W.-C.; Keogh, E.; Lau, L.Y.; Ratanamahatana, C.A.; Wong, R.C.-W. Scaling and Time Warping in Time Series Querying. *VLDB J.* **2008**, *17*, 899–921. [CrossRef]

66. Itakura, F. Minimum Prediction Residual Principle Applied to Speech Recognition. *IEEE Trans. Acoust.* **1975**, *23*, 67–72. [CrossRef]

67. Bennett, N.D.; Croke, B.F.W.; Guariso, G.; Guillaume, J.H.A.; Hamilton, S.H.; Jakeman, A.J.; Marsili-Libelli, S.; Newham, L.T.H.; Norton, J.P.; Perrin, C.; et al. Characterising Performance of Environmental Models. *Environ. Model. Softw.* **2013**, *40*, 1–20. [CrossRef]

68. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2008.

69. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]

70. USDA. *National Agricultural Statistics Service, Cropland Data Layer*; USDA: Washington, DC, USA, 2017.

71. Arvor, D.; Durieux, L.; Andrés, S.; Laporte, M.-A. Advances in Geographic Object-Based Image Analysis with Ontologies: A Review of Main Contributions and Limitations from a Remote Sensing Perspective. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 125–137. [CrossRef]

72. Neubert, M.; Herold, H.; Meinel, G. Assessing Image Segmentation Quality—Concepts, Methods and Application. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 769–784.

73. Yan, L.; Roy, D.P. Conterminous United States Crop Field Size Quantification from Multi-Temporal Landsat Data. *Remote Sens. Environ.* **2016**, *172*, 67–86. [CrossRef]

74. Csillik, O. Fast Segmentation and Classification of Very High Resolution Remote Sensing Data Using SLIC Superpixels. *Remote Sens.* **2017**, *9*, 243. [CrossRef]

75. Heydari, S.S.; Mountrakis, G. Effect of Classifier Selection, Reference Sample Size, Reference Class Distribution and Scene Heterogeneity in Per-pixel Classification Accuracy using 26 Landsat Sites. *Remote Sens. Environ.* **2018**, *204*, 648–658. [CrossRef]

76. Petitjean, F.; Kurtz, C.; Passat, N.; Gançarski, P. Spatio-Temporal Reasoning for the Classification of Satellite Image Time Series. *Pattern Recognit. Lett.* **2012**, *33*, 1805–1815. [CrossRef]

77. Matton, N.; Canto, G.S.; Waldner, F.; Valero, S.; Morin, D.; Inglada, J.; Arias, M.; Bontemps, S.; Koetz, B.; Defourny, P. 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.* **2015**, *7*, 13208–13232. [CrossRef]

78. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-Based Crop Identification using Multiple Vegetation Indices, Textural Features and Crop Phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [CrossRef]

79. Curran, P.J.; Dungan, J.L.; Gholz, H.L. Exploring the Relationship Between Reflectance Red Edge and Chlorophyll Content in Slash Pine. *Tree Physiol.* **1990**, *7*, 33–48. [CrossRef]

80. Horler, D.N.H.; Dockray, M.; Barber, J. The Red Edge of Plant Leaf Reflectance. *Int. J. Remote Sens.* **1983**, *4*, 273–288. [CrossRef]

81. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors* **2011**, *11*, 7063–7081. [CrossRef]

82. Herrmann, I.; Pimstein, A.; Karnieli, A.; Cohen, Y.; Alchanatis, V.; Bonfil, D.J. LAI Assessment of Wheat and Potato Crops by VENμS and Sentinel-2 Bands. *Remote Sens. Environ.* **2011**, *115*, 2141–2151. [CrossRef]

83. Clevers, J.G.P.W.; Gitelson, A.A. Remote Estimation of Crop and Grass Chlorophyll and Nitrogen Content using Red-Edge Bands on Sentinel-2 and -3. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 344–351. [CrossRef]

84. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]

85. Chrysafis, I.; Mallinis, G.; Siachalou, S.; Patias, P. Assessing the Relationships Between Growing Stock Volume and Sentinel-2 Imagery in a Mediterranean Forest Ecosystem. *Remote Sens. Lett.* **2017**, *8*, 508–517. [CrossRef]

86. Foody, G.M. Status of Land Cover Classification Accuracy Assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]

87. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making Better Use of Accuracy Data in Land Change Studies: Estimating Accuracy and Area and Quantifying Uncertainty using Stratified Estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131. [CrossRef]

88. Zhao, Y.; Gong, P.; Yu, L.; Hu, L.; Li, X.; Li, C.; Zhang, H.; Zheng, Y.; Wang, J.; Zhao, Y.; et al. Towards a Common Validation Sample Set for Global Land-Cover Mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814. [CrossRef]

89. Congalton, R.G.; Gu, J.; Yadav, K.; Thenkabail, P.; Ozdogan, M. Global Land Cover Mapping: A Review and Uncertainty Analysis. *Remote Sens.* **2014**, *6*, 12070–12093. [CrossRef]

90. Lambert, M.-J.; Traoré, P.C.S.; Blaes, X.; Baret, P.; Defourny, P. Estimating Smallholder Crops Production at Village Level from Sentinel-2 Time Series in Mali's Cotton Belt. *Remote Sens. Environ.* **2018**, *216*, 647–657. [CrossRef]

91. Zhao, Y.; Lin, L.; Lu, W.; Meng, Y. Landsat Time Series Clustering under Modified Dynamic Time Warping. In Proceedings of the 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Guangzhou, China, 4–6 July 2016; pp. 62–66.

92. Salvador, S.; Chan, P. Toward Accurate Dynamic Time Warping in Linear Time and Space. *Intell. Data Anal.* **2007**, *11*, 561–580. [CrossRef]

93. Keogh, E.; Ratanamahatana, C.A. Exact Indexing of Dynamic Time Warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [CrossRef]

94. Sinclair, T.R.; Seligman, N. Criteria for Publishing Papers on Crop Modeling. *Field Crops Res.* **2000**, *68*, 165–172. [CrossRef]

95. Li, Z.; Zhu, Q.; Gold, C. *Digital Terrain Modeling: Principles and Methodology*; CRC Press: Boca Raton, FL, USA, 2004.