

Article

Assessing Spatial Limits of Sentinel-2 Data on Arable Crops in the Context of Checks by Monitoring

Blanka Vajsová¹, Dominique Fasbender² , Csaba Wirnhardt^{3,*}, Slavko Lemajic⁴
and Wim Devos³

¹ Piksel Srl, Via Breda 176, 20126 Milan, Italy; blanka.vajsova@piksel.com

² Institut wallon de l'évaluation, de la prospective et de la statistique, Route de Louvain-la-Neuve, 2, 5001 Belgrade, Belgium; d.fasbender@iweeps.be

³ European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, I-21027 Ispra, VA, Italy; Wim.Devos@ec.europa.eu

⁴ Arhs Developments, 2b, rue Nicolas Bové, L-1253 Luxembourg, Luxembourg; Slavko.Lemajic@ext.ec.europa.eu

* Correspondence: Csaba.Wirnhardt@ec.europa.eu

Received: 11 May 2020; Accepted: 6 July 2020; Published: 9 July 2020



Abstract: The availability of large amounts of Sentinel-2 data has been a trigger for its increasing exploitation in various types of applications. It is, therefore, of importance to understand the limits above which these data still guarantee a meaningful outcome. This paper proposes a new method to quantify and specify restrictions of the Sentinel-2 imagery in the context of checks by monitoring, a newly introduced control approach within the European Common Agriculture Policy framework. The method consists of a comparison of normalized difference vegetation index (NDVI) time series constructed from data of different spatial resolution to estimate the performance and limits of the coarser one. Using similarity assessment of Sentinel-2 (10 m pixel size) and PlanetScope (3 m pixel size) NDVI time series, it was estimated that for 10% out of 867 fields less than 0.5 ha in size, Sentinel-2 data did not provide reliable evidence of the activity or state of the agriculture field over a given timeframe. Statistical analysis revealed that the number of clean or full pixels and the proportion of pixels lost after an application of a 5-m (1/2 pixel) negative buffer are the geospatial parameters of the field that have the highest influence on the ability of the Sentinel-2 data to qualify the field's state in time. We specified the following limiting criteria: at least 8 full pixels inside a border and less than 60% of pixels lost. It was concluded that compliance with the criteria still assures a high level of extracted information reliability. Our research proved the promising potential, which was higher than anticipated, of Sentinel-2 data for the continuous state assessment of small fields. The method could be applied to other sensors and indicators.

Keywords: NDVI time series; small agriculture parcels; fields; CAP; similarity; spatial limits; Sentinel-2; PlanetScope; checks by monitoring; CbM

1. Introduction

The European Copernicus program operates its own constellation of Earth observation satellites and produces an immense amount of satellite data. Together with the fast development of cloud processing platform services, it gave rise to an impulse to exploit these resources and technology advances within the EU's Common Agriculture Policy (CAP).

A substantial element of the CAP is a system of farmer subsidies granted to ensure income stability and to remunerate them for environmentally friendly farming and delivering public goods not normally paid for by the markets, such as taking care of the countryside (for instance flowering

buffer strips for pollination) [1]. However, this farmers' direct support is ruled by eligibility criteria, commitments, and other obligations under the CAP regulations. Compliance with these rules has so far been checked by administrative and on-the-spot checks (OTSC), a sample approach practice performed yearly by the EU Member States on at least 5% of aid applications [2]. The OTSC are performed by either classical field-based visits or checks via satellite imagery, called the control with remote sensing (CwRS) method. The CwRS uses a sequence of satellite images acquired during the crop season over the same selected sample of parcels. The sequence is usually composed of 2–4 data captures, from which one is of sub-meter resolution (for area measurement), complemented with images of slightly lower spatial resolution (for detecting field conditions over the season). A separate photointerpretation analysis of each image is the result at the end of the crop season as the conclusion for all sampled holdings. Satellite images have been used for the purpose of CAP control since 1993, and until now the number of acquired images has grown from 100 to 1400 captures for one control year. Both the amount of data and the quality (radiometric, spectral, and spatial resolution) of the satellite imagery have improved considerably. The list of various satellites used during the years includes: European Remote Sensing (ERS) satellite, Formosat 2, Radarsat, Landsat 5, Disaster Monitoring Constellation 2 (DMC 2), ResourceSat-1, RapidEye, THEOS, Cartosat 2, Deimos 1, CosmoSkymed, EROS A/B, Ikonos, QuickBird, Pleiades 1A/1B, SPOT 4/5, WorldView-4, PlanetScope, Kompsat-3/3A, SPOT 6/7, GeoEye-1, WorldView-1/2/3, last nine are currently used.

In 2018, the European Commission (EC) provided a legal and technical framework to exploit the advantages of Copernicus satellite data in the context of CAP controls. An amendment to Commission Implementing Regulation (EU) No. 809/2014 [3] gave the EU Member States an option to gradually replace the OTSC with checks by monitoring (CbM). This introduced a monitoring and prevention system that mainly relies on continuous provision of Sentinel satellite data. The free availability of global coverage and temporal resolution of the data enabled a shift from the sampling and sequence mode to the continuous assessment of the whole territory. Such constant assessment provides an individual time for the decision on the eligibility for each holding, which could make financial procedures more flexible in comparison to the OTSC approach. Furthermore, the checks mechanism is coupled with an alert component. Farmers receive early warnings and can correct their aid application to avoid noncompliant declarations, avoiding penalties at the end of the season. In practice, this involves automatic image analysis to determine the actual state of the field. This gives information either about the presence of a crop, its phenological stage, or about a specific activity conducted in a given time range. The accumulated information about the state of the land is continuously compared with the expected scenario of events implied by the obligations and commitments of the specific payment scheme [4].

As the CbM approach is based upon the use of Copernicus Sentinel data (Sentinel-1/2), its application scope is limited by the radiometric and geometric characteristics of these satellites. While the revisit time of Sentinel satellites provides an unprecedented temporal resolution, the ground sampling distance (GSD) and geometric positional accuracy could be too coarse for smaller spatial features. The CbM faces a tradeoff between the large amount of free data and its medium spatial resolution. Current OTSC image data have a spatial resolution of no more than 2.2 m for a panchromatic band. As a result, one of the major concerns of the EU Member States regarding the operational implementation of monitoring is the perceived inadequacy of the Sentinel sources for small fields. Cases where conclusion cannot be made regarding the eligibility of Sentinel imagery due to the size of the field (or other observed features) require alternative data sources, such as a higher resolution data stack, geotagged photographs, or field inspections. Inevitably, these alternative sources and processes have a higher cost than the automated processing of free Sentinel data, so managing them is essential.

This study addresses the concern of EU Member States regarding the monitoring of the small fields. It specifically focuses on Sentinel-2, i.e., optical data. The goals of the research are: (i) to quantify the capability of Sentinel-2 to provide information on the activity or state of a field; (ii) to examine the impacts of various geospatial parameters of the field on the suitability of Sentinel-2 data;

(iii) to estimate the limiting geospatial criteria above which the Sentinel-2 data stacks still provide reliable information on the state or activity of the field, all in the context of the CbM of small parcels. A similarity assessment of the two sets of NDVI time series (Sentinel-2 and PlanetScope) extracted over the same fields is used as the main tool to achieve the research goals.

2. Materials and Data Preparation

2.1. Study Area and Vector Data

The study area selected was in the Netherlands, as it is one of very few EU Member States that makes annual geospatial aid application (GSAA) information publicly available. The test areas in Figure 1 are in the southeast part of the Netherlands, crossing the provinces of North Brabant and Limburg. North Brabant is mostly flat, with the majority of the area still above the sea level, whereas Limburg has a hillier landscape, within which the highest point is 322 m.s.l, while the southern part of the province is slightly undulated.

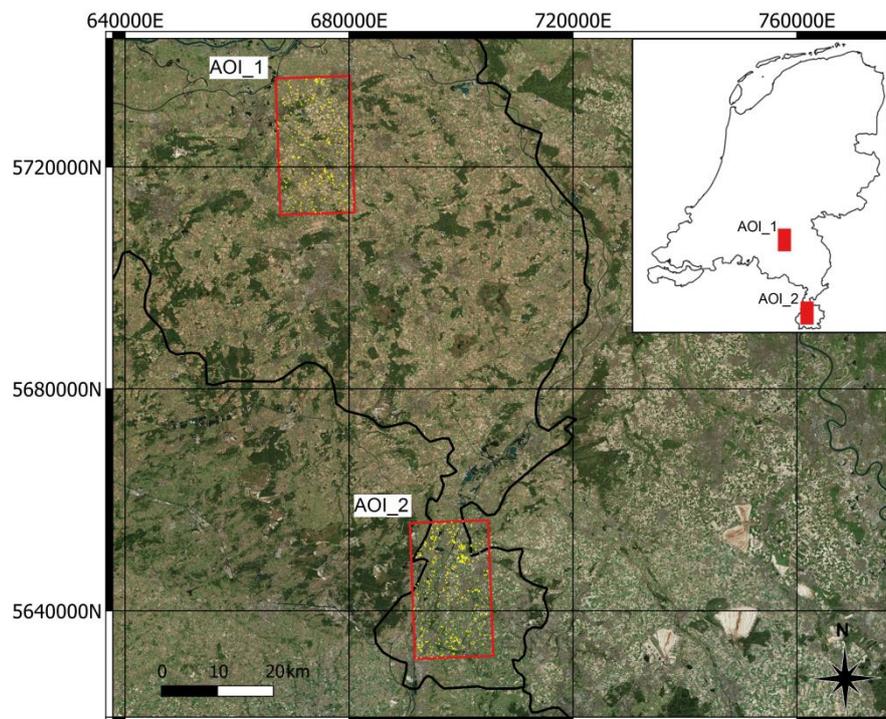


Figure 1. Position of areas of interests (AOIs) in the Netherlands.

For delineations of boundaries of agricultural fields, as well as the other relevant information assigned by farmers, we made use of the GSAA data declared for the 2018 claim year. The dataset is publicly available through The Netherlands open geodata infrastructure (<http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home>) [5].

We chose area of interest 1 and area of interest 2 (AOI_1 and AOI_2), characterized by a high accumulation of agriculture parcels with arable crops having a size below 0.5 ha. The magnified example of a few parcels within each AOI is displayed in Figure 2.

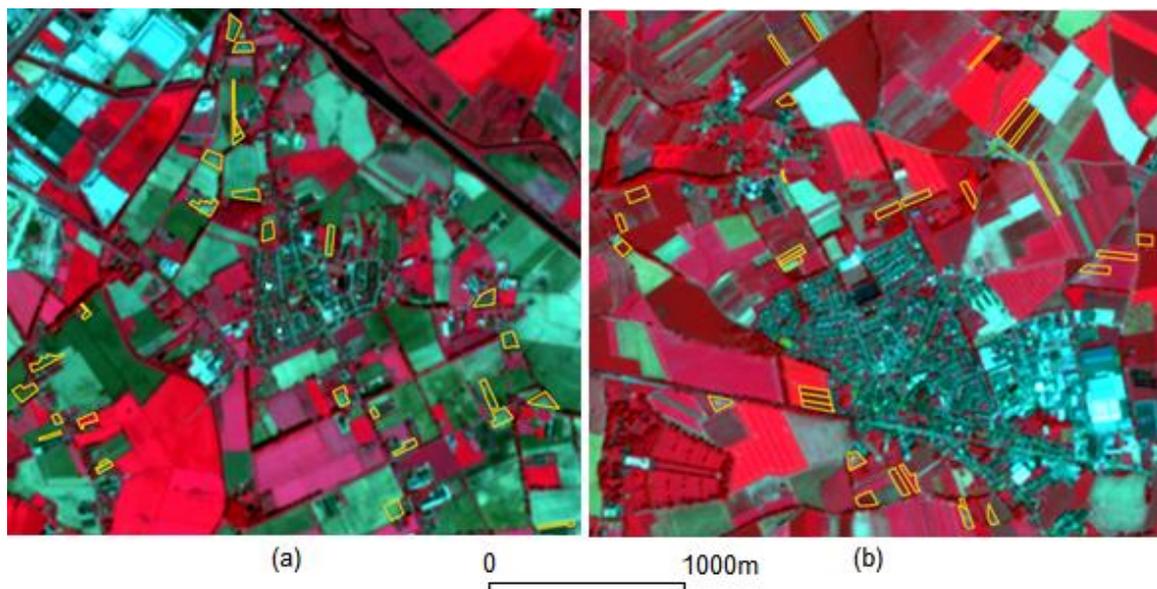


Figure 2. Magnified detail showing crop parcels (fields) within an area of interest (AOI): (a) example of AOI_1 in the background Sentinel-2 image from 06/05/2018; (b) example of AOI_2, in the background Sentinel-2 image from 27/06/2018. Sentinel-2 images are illustrated in false-color composite (R: near-infrared (NIR), G: red, B: green).

The size of the study area was approximately 67,340 ha, including 867 crop parcels (fields) below 0.5 ha in size, with 25% of the parcels being smaller than 0.2 ha, as shown in Figure 3. In the sample of parcels, the four main crops for the 2018 declaration year were maize (silage), winter wheat, potatoes for consumption, and sugar beet. The complete distribution is shown in Figure 4. The distribution of the geometry and the variety of arable crops demonstrate the representatives of the test site.

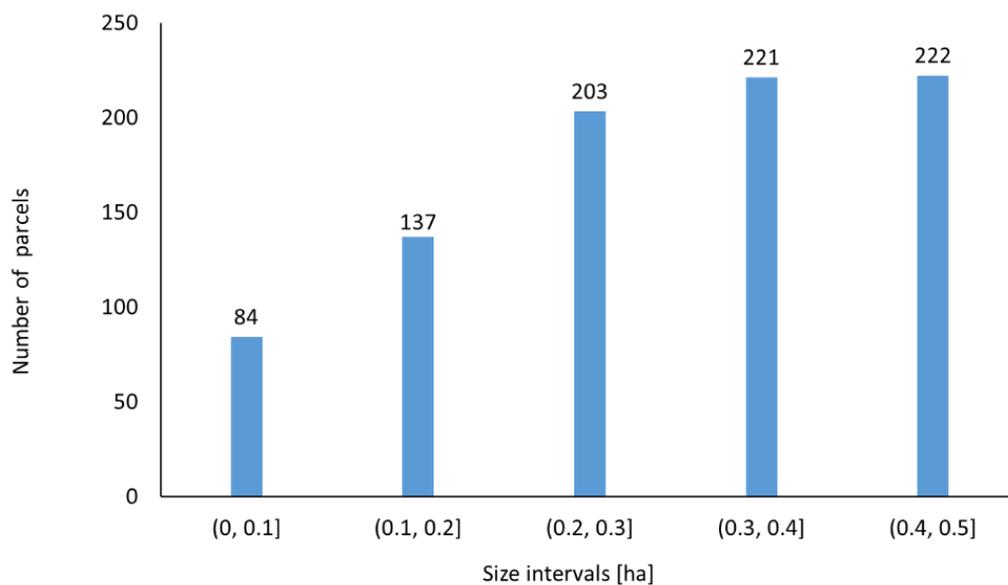


Figure 3. Distribution of parcels in a specific size category.

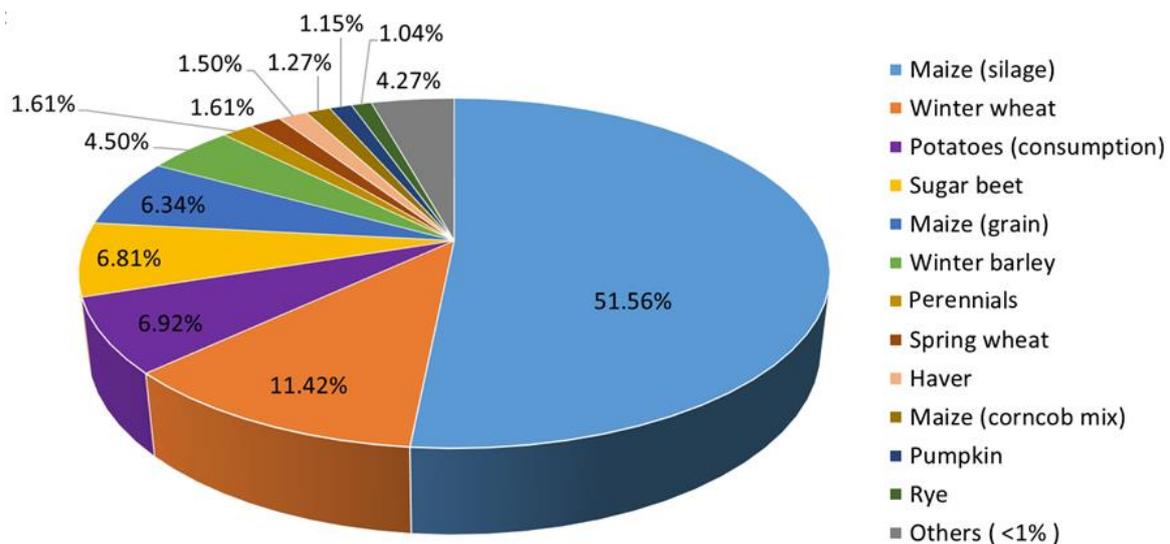


Figure 4. Distribution of arable crop types in the sample.

2.2. Satellite Data

Since the quality of cloud cover (CC) mask and geolocation accuracy parameters of image data have a direct impact on a number of satellite images included in the process and on the correctness and quality of NDVI values retrieved, details on these data characteristics are described in the following two sections.

2.2.1. Sentinel-2 Data

NDVI time series were extracted from the level 2A (L2A) products from both Sentinel-2A and Sentinel-2B satellites. These L2A products contain applied radiometric and geometric corrections—the bottom of atmosphere reflectance (BOA) in the Universal Transverse Mercator - World Geodetic System (UTM-WGS 84) cartographic coordinate reference system. Sentinel-2 offers 13 spectral bands ranging from 10 to 60 m, with a temporal resolution of 5 days. For the needs of the study, only band 8 (NIR) and band 4 (RED) were fully utilized. The Level-2A image metadata provide the scene classification (SCL) map. The quality of the SCL given by recognition of clear pixels over land and water reached overall accuracy (OA) of 91.5% [6]. The SCL information was used to filter out cloudy images. The absolute geolocation accuracy was reported as close to 11 m at 95% confidence level for both satellites. However, the geolocation accuracy for Sentinel-2B was not stabilized. The temporal co-registration error was within one pixel for 92% of Sentinel-2A and 87% of Sentinel-2B products. The co-registration of multispectral bands resulted in less than 0.3 pixels [7].

2.2.2. PlanetScope Data

A constellation of PlanetScope (PS) satellites collects multispectral (MS) imagery in 4 bands, with a ground sampling distance (GSD) of 3.7–3.9 m. The NDVI time series extraction was applied on Analytic Ortho Scene products (Level 3B), exploiting band 4 (NIR) and band 3 (RED). These products are radiometrically and geometrically corrected and contain surface reflectance values (BOA) in the UTM-WGS 84 cartographic coordinate reference system. The GSD is resampled to a pixel size of 3 m [8]. The quality of the CC mask was unknown at the time of processing. According to the PS Analytic Ortho Scene Product Specifications, the requirements for the geometric quality are expressed at positional accuracy of less than 10 m in two-dimensional root mean square error (RMSE_{2D}) [8]. The benchmarking test on the positional geometric accuracy for the purpose of CAP checks performed by the EC Joint Research Centre (JRC) resulted in circular error at the 90th percentile (CE(90)) of

9.93 m [9]. This outcome is in line with the horizontal accuracy assessment of one PS Ortho Scene [10], where the RMSE_{2D} resulted in 5.42 m (i.e., CE(90) 8.22 m).

2.3. NDVI Time Series

2.3.1. Quality of NDVI Time Series

Time series in our context are defined as a series of consecutive observations acquired at regular or irregular time intervals [11]. The quality of data entering the process has a crucial impact on the correctness and consequently the reliability of extracted information. Satellite-derived NDVI time series are omnipresent in the remote sensing of vegetation, but their application could be hindered by prevalent noise caused by varying atmospheric conditions and sun sensor–surface viewing geometries. It is important to consider the strength and character of any noise present in the NDVI dataset when selecting an approach to noise reduction. However, applying noise reduction to relatively clean time series will tend to degrade them [12]. Therefore, in our processing flow, we did not apply any noise reduction technique on the time series. After initial data filtering based on the metadata of the imagery (date, CC mask), we obtained relatively smooth time series that were deemed directly suitable for further analysis. Another motivation for our approach was to avoid adding additional bias into the similarity comparison.

While the noise caused by meteorological conditions at the time of an acquisition is a common issue regardless of the size of the parcels, the geometric positional accuracy of the image data is an influential factor on the reliability of time series, in particular for small parcel sizes. One could say that the sensitivity to any geometrical displacement increases with the decreasing size of the parcel. However, other factors, such as the geometry of parcels, also play important roles. Narrow elongated parcels are particularly vulnerable to this positional parameter. For the time series extraction, the positional factor is mainly a mixture of absolute geolocation accuracy of the source data and the multitemporal co-registration accuracy of raster images. The potential impacts of poor co-registration on various applications, such as change detection [13,14] and pixel- and object-based classifications [15], have already been pointed out by other researchers.

2.3.2. NDVI Time Series Extraction

The following steps were performed for the NDVI time series extraction:

1. Images were found within a given timeframe: 1/2/2018–30/11/2018 (L2A product available from 1/4/2018);
2. Images were filtered based on CC information;
3. NDVI layers were calculated using Equation (1) [16,17];

$$NDVI = \frac{NIR - red}{NIR + red} \quad (1)$$

where NIR and red are the surface reflectance of the NIR band and red band, respectively.

4. The mean value for each field is calculated, considering pixels within its boundaries (Section 2.4);
5. Visualization is performed.

The whole workflow of the Sentinel-2 NDVI time series extraction was carried out in the Joint Research Center's internal Earth Observation Data and Processing Platform (JEODPP). JEODPP is a versatile petabyte-scale platform providing a cluster environment for batch processing, web-based remote desktop access with a variety of software suites, and a web-based interactive visualization and analysis ecosystem [18]. Unlike PS images, the image filtering was done separately for each parcel using pixel-based assessment of SCL data. As a result, the number of NDVI values in the time series varies with each parcel.

The PS image data were downloaded through the Planet Explorer platform (<https://www.planet.com/explorer/>) [19] using predefined metadata filters. The CC filter was limited to 0% for the entire ortho scene, relying on the high temporal resolution of PS imagery to counteract this strict criterion. The Analytic Ortho Scene products were merged by an acquisition date into separate mosaics. The calculation of PS NDVI values and the extraction of mean NDVI time series were accomplished by in-house-developed python scripts, using the built-in raster calculator and zonal statistics functionalities of the open source software Quantum geographic information system (QGIS).

2.4. Parameters of Parcels

Since one of the goals of this research was to find and understand the dependency of Sentinel-2 suitability on parcels' geospatial and contextual characteristics, the following parameters were selected and their potential impacts were studied:

- Parcel size;
- Geometry of parcel (regular, irregular, various elongations);
- Crop type (study is limited to arable crops, reason explained in Section 5);
- Surrounding land use or land cover of the parcel (same or different within 5 m distance);
- Number of Sentinel-2 pixels in the parcel without any buffer;
- Number of Sentinel-2 pixels in the parcel with a 5 m negative buffer (i.e., full or clean pixels);
- Percentage of Sentinel-2 pixels lost after application of a 5 m negative buffer (i.e., the relative difference between the numbers of pixels before and after applying the 5 m negative buffer).

Regarding the size of parcels, rule-of-thumb considerations of Sentinel-2's ground sampling distance (10 m) and earlier experiences indicated that Sentinel-2 should operate well on agricultural parcels larger than 0.3–0.5 ha, depending on their geometry and neighborhood conditions [20]. We, therefore, considered only parcels less than or equal to 0.5 ha.

The number of pixels in the parcel expresses how many pixels participated in the mean NDVI value calculation for a given polygon. The counting procedure (i.e., deciding which pixel was inside the polygon, which would participate in further analysis) could have a direct impact on the resulting NDVI time series. The concept used in this study was based on the sum of all raster cells whose center was within the target polygon. A disadvantage of this approach compared to more complex methods is that its count is more affected by the relative position of the raster cells to the overlaid polygon layer. An example of the relationship between the number of counted pixels and the position of a given polygon in the raster is illustrated in Figure 5.

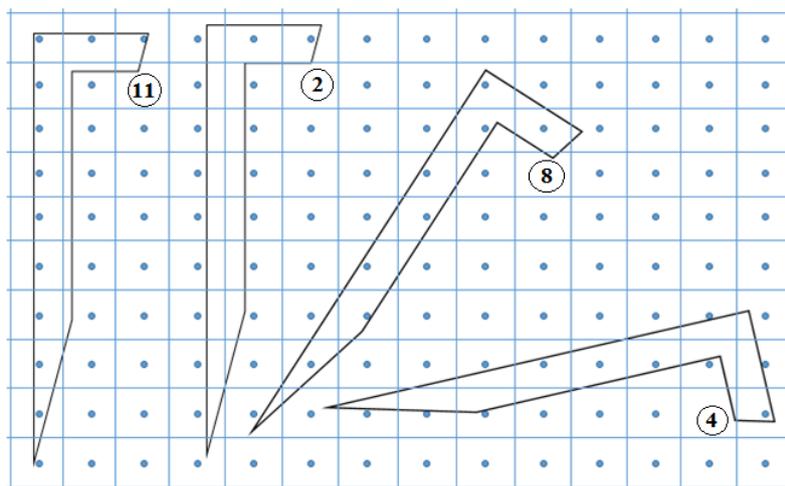


Figure 5. Different numbers of pixels linked to various positions of the parcel on the raster. Number of pixels from left to right: 11, 2, 8, 4. Adapted from [21].

The noise in the time series caused by heterogeneous pixels on the borders of parcels might be reduced by introducing a negative buffer, thus isolating clean pixels to represent the parcel. In some cases, it might also mitigate the negative effect of the imperfect positional geometry accuracy of the image. The distance of 5 m assures a high probability that the full pixel is inside the parcel. The 5-m threshold is derived from the Sentinel-2 pixel size. The NDVI values were calculated as an arithmetical mean, which to a certain extent also suppressed the undesirable influence of the border pixels. We opted not to use any buffer during the NDVI time series extraction. The reason was two-fold: firstly, we intended to further study the significance of the surrounding parameters; and secondly, we preferred to keep the testing sample as large as possible, without excluding very small parcels that would have disappeared after application of the negative buffer.

Although no negative buffer was applied for NDVI time series construction, the theoretical percentage of Sentinel-2 pixels lost after the application of the 5 m negative buffer was calculated and analyzed as the parcel's parameter (see list above). This was not considered at the beginning of the study but emerged during our analysis. This was found to be a powerful factor combining the impacts of the shape, size, and position of the polygon (field) on the raster image.

3. Methodology

3.1. General Approach

The test image data are captured by sensors of similar radiometric characteristics but of different spatial resolutions. One assumes that the higher the spatial resolution of the image data, the more detailed and reliable the information they provide about an imaged object becomes. Reflecting this simple assumption into the agriculture context, we first followed the evolution of a specific composite indicator (NDVI) to determine the state of the agriculture parcel over time. This temporal behavior of mean NDVI values of the parcel was portrayed as the NDVI time series. The temporal change of the state of each agriculture parcel was, therefore, represented by two NDVI time series retrieved from two data sources of different spatial resolution. Next, we carried out a trend behavior comparison and similarity assessment. We based our methodology on an assertion that the less similarities between the time series, the lower the reliability of information obtained from the time series that originated from coarser spatial resolution data. It is worth noting that we neither intend to assess the phenological phases of vegetation dynamics timing nor compare the absolute values of the NDVI indicator. The method relies on the time series' curve comparison, while considering the time series from the higher resolution data as a reference. In addition, the fact that we applied comparative analysis makes the used methodology insensitive to various environmental conditions (climate, soil type, etc.). Prior to the investigation on small fields, the validation test on the fields measuring above 3 ha was performed. Two parcels out of 400 (i.e., 0.5%) were just below the threshold of similarity (Section 3.2, point 2). The results from the validation proved that the time series from both sensors were very comparable.

In order to establish the working framework for our study, we constructed NDVI time series over 867 small agricultural parcels (below 0.5 ha) extracted from Sentinel-2 and PlanetScope image datasets. Next, the following steps were performed:

1. Visual assessment of the similarity of the NDVI time series pairs to identify a portion of the sample for which the time series deviated. This assessment had a complementary purpose and allowed for familiarization with the analyzed data (Section 3.3);
2. Statistical assessment of the NDVI time series pairs' similarity to identify a portion of the sample for which the time series deviated. These are the parcels (group 1) for which we assert that there is a higher probability of Sentinel-2 data providing insufficient information on the state or a change of the state of the land phenomenon (Section 3.4);
3. Statistical modeling to estimate the geospatial criteria of parcels limiting the suitability of the Sentinel-2 data use in CbM (Section 3.5)

- (a) Classification and regression analysis to assess the impact of various parcels' parameters (listed in Section 2.4) on the similarity of the NDVI time series;
 - (b) Modeling dependencies between the selected parcels' parameters and the probability that the correlation coefficient exceeds the predefined threshold. Identification of criteria for defining the term "small parcels" in the context of CbM with Sentinel-2 satellites;
4. The limiting geospatial criteria were applied on the tested sample of parcels and the result was compared with the outcome of the statistical similarity assessment method.

For the mathematical approach described in points 2 and 3 above, we set a concept of similarity (Section 3.2) defining the appropriate thresholds. Figure 6 illustrates the methodology and its steps.

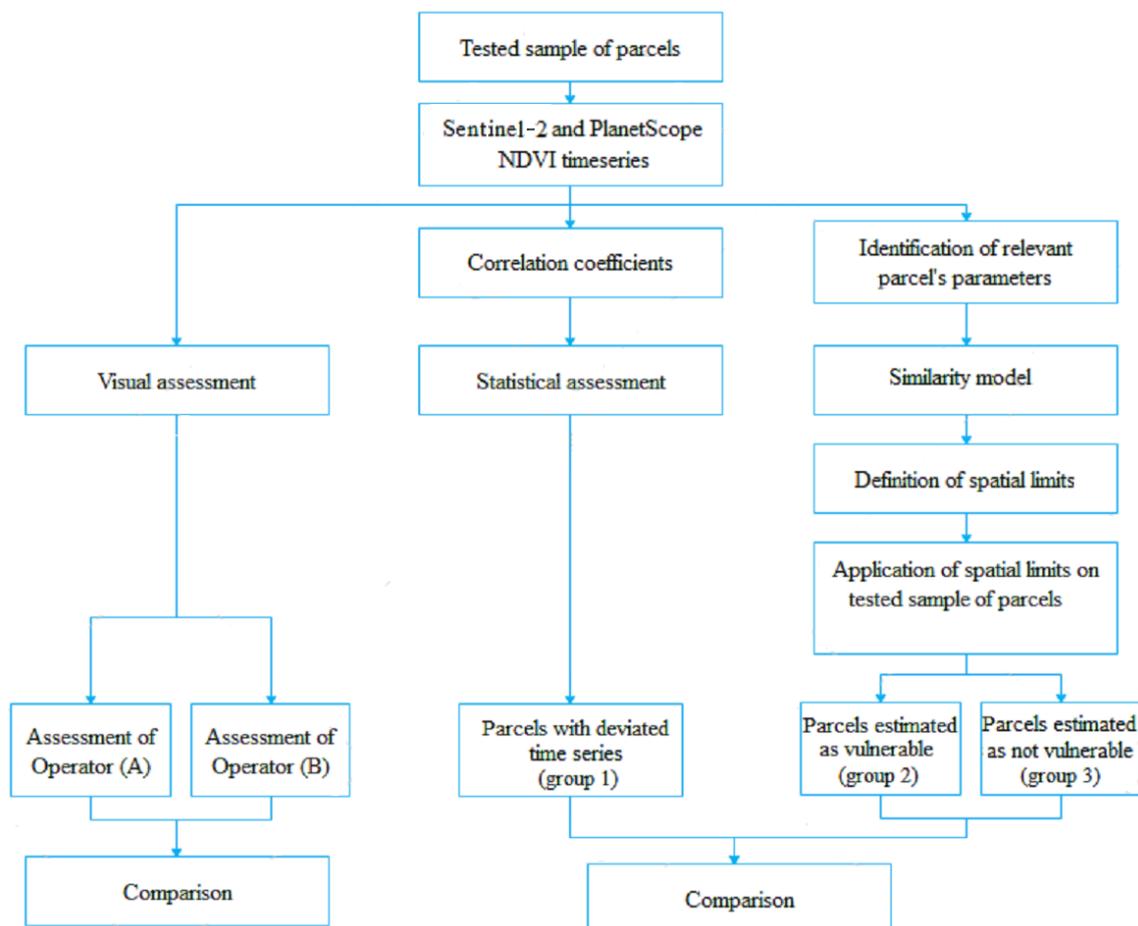


Figure 6. Workflow of the methodology.

3.2. Concept of Similarity

The evaluation of the similarity of the NDVI time series stems from shape-matching analysis. The comparison does not consider the absolute values of NDVI as such (i.e., is invariant to scale and vertical shift transformation), but instead assesses the trends and patterns of the curve, and thus the traceability of the specific changes in the behavior and agriculture activity on the field. To fulfill the requirements for comparisons, the correlation coefficient between the two series was selected as a sufficient quantitative index [22].

For the time series similarity assessment and the limiting criteria definition, we adopted the following three methodologies and respective thresholds:

1. Visual assessment based on expert judgment, without thresholds;

2. Correlation coefficient as an index of similarity. The time series were considered as similar (i.e., not deviating) if the correlation coefficient was higher than the value of 0.5 with 95% probability;
3. Similarity model quantifying the probability of exceeding the correlation coefficient value of 0.75 as a function of geospatial parameters of parcels. The time series were assumed to be similar if the probability that the correlation coefficient exceeded the value 0.75 was equal to or higher than 80%.

These thresholds on the correlation coefficient were defined for the context and purpose of the study. For the statistical test on the 95% confidence interval (point 2 above), we decided to apply the limit of 0.5, indicating a moderately positive correlation. In the similarity model for the derivation of limiting parameters, we pointed to the probability, which was set to 80%, of the exact value of the correlation coefficient, and hence we stuck to a stricter correlation value of 0.75, interpreted as a high (strong) positive correlation [23–25]. Even if expressed with different absolute values, these two criteria are similar.

3.3. Visual Assessment

We applied the visual evaluation of the similarity of the time series pairs as an independent control method to the statistical methodology described below. Since the statistical concept intended to simulate the genuine human perception of the profile's trend behavior, we considered the expert judgment as an adequate complementary approach. Since the visual assessment is impractical and work-intensive for large datasets, we maintained the reasonable size of the sample to enable this kind of manual evaluation. Two experts in remote sensing assessed the similarity of the NDVI time series independently from each other, portrayed by two curves on the same graph. The following categories were used for visual inspection: agreement, disagreement, doubt, and decision not made due to small number of observations. Typical examples of agreement, disagreement, and doubtful cases of the profiles are illustrated in Figure 7. In the assessment, the following aspects were considered: (i) the Sentinel-2 NDVI time series shadowed the trend of PlanetScope; (ii) distinct peaks and patterns visible in the PlanetScope profile were perceived as visible in the Sentinel-2 profile.

We used Cohen's kappa index [26] to measure the reliability of the evaluation. Cohen's kappa is a measure of an agreement between two operators who determine which category a finite number of subjects or decisions belong to, while the agreement due to chance is factored out. However, the index does not consider the degree of their disagreement. In the classification set, we could, however, observe different severities of disagreement. To address this issue, we opted for a modification to Cohen's kappa called the Cohen's kappa [27], as defined in Equation (2):

$$\kappa_w = 1 - \frac{\sum_{i=1}^k \sum_{j=1}^k w_{ij} p_{ij}}{\sum_{i=1}^k \sum_{j=1}^k w_{ij} e_{ij}} \quad (2)$$

where p_{ij} are the observed probabilities, e_{ij} are expected probabilities, and w_{ij} are the weights that express the degree of disagreement between the two operators. The higher the disagreement, the higher the weight. A prior study [28] gave a practical example of how to calculate observed (p_{ij}) and expected (e_{ij}) probabilities.

It is important to emphasize that PlanetScope NDVI time series were taken as a reference. We did not judge the correctness of its profile for identifying a specific crop. However, visual inspection of the NDVI time series was performed to understand the quality and uncertainty of the reference. It was concluded that we doubted the correctness of 2% of all PlanetScope NDVI profiles, mostly due to the weak trend.

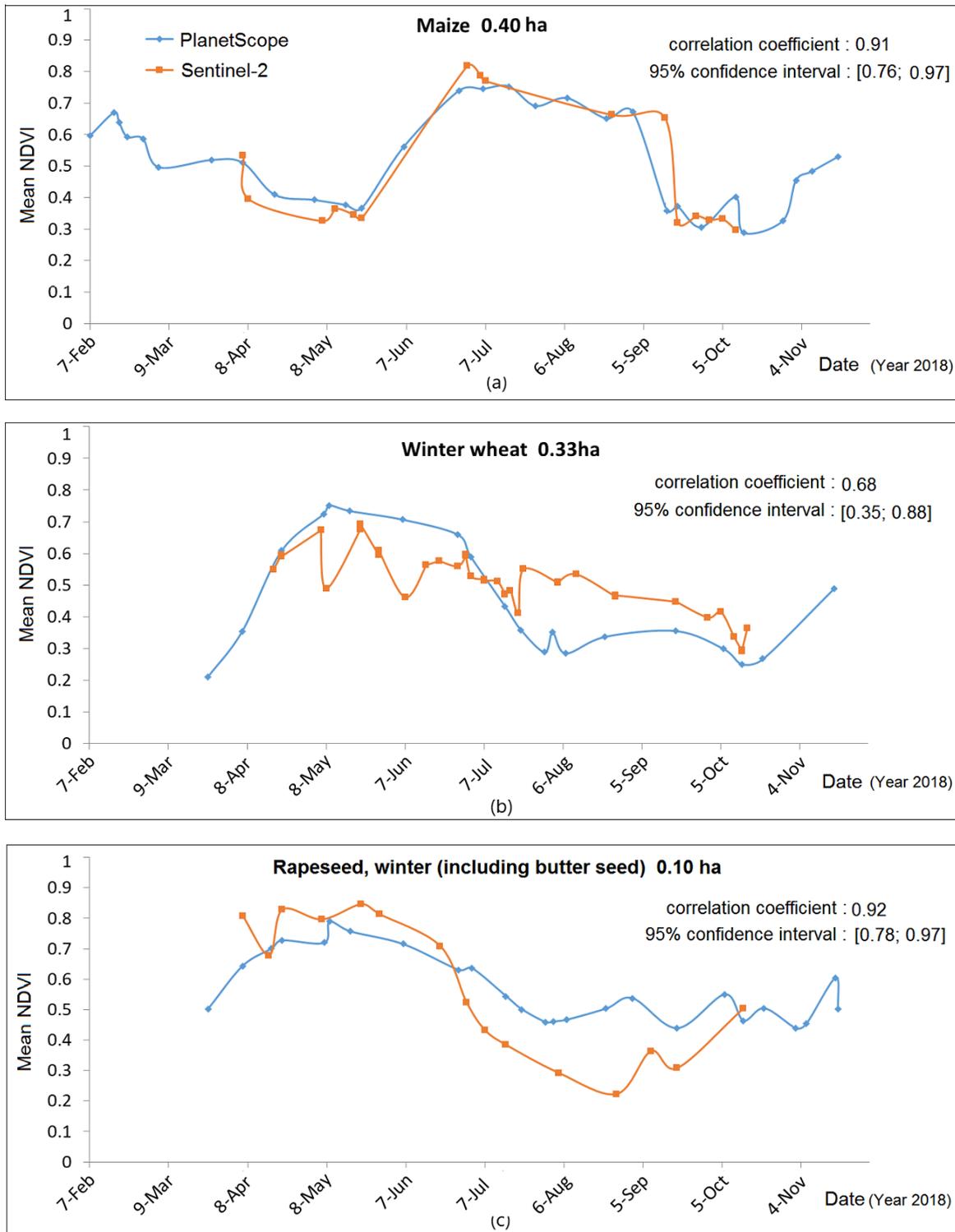


Figure 7. Typical examples of the visual assessment of NDVI time series pairs: (a) illustration of the agreement; (b) illustration of the disagreement; (c) illustration of the doubtful case.

3.4. Correlation Coefficient as the Index of Similarity

In this study, the Pearson’s correlation coefficient [22] was chosen to assess the similarity between the NDVI time series from PlanetScope and Sentinel-2. The correlation coefficient between two quantities is generally estimated using paired observations (i.e., each observation is a pair of both

quantities). With paired observations, one can simply rely on the following estimator for the estimated correlation coefficient:

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

where \bar{x} and \bar{y} are the observed averages of the time series and n is the number of observed pairs.

When comparing two time series from two different sensors, it is very unlikely that the data points from each time series are observed on the same dates. In order to circumvent this data issue, Ref. [29] proposed to fill the gaps by interpolating each of the time series at the observed dates of the other time series (i.e., the union of the observed dates). The authors concluded that relying on a linear interpolation was sufficient for translating the intuitive visual comparison of the time series pairs, while having the advantage that observations were temporally aligned, allowing estimation of the correlation coefficient. This method also proposes a solution for assessing the uncertainty of the estimated correlation coefficient.

3.5. Similarity Model

We built a model of the estimated correlation coefficients and selected geospatial parameters of the sampled parcels. The objective of this particular model was not necessarily to accurately predict if the parcel was suitable to be monitored with Sentinel-2 data, it was to better understand the conditions where Sentinel-2 and PlanetScope NDVI time series do not agree, as our analysis revealed that setting only a parcel's size as a threshold is not a sufficient criterion.

In the first step, we relied on a classification and regression tree (CART) algorithm in order to select the most relevant factors. This algorithm was based on an iterative definition of the dichotomous split (i.e., the nodes) of the input factors so that the separation between the parcels with or without agreement was maximized. Theoretically, the advantage of this approach is that the most relevant factors for the classification can be selected automatically.

The parameters that were present in the classification tree were:

- Percentage of Sentinel-2 pixels lost after application of a 5 m negative buffer;
- Surrounding land use and land cover of the parcel (same or different than tested parcel);
- Number of Sentinel-2 pixels in the parcel with a 5 m negative buffer (i.e., full or clean pixels).

Other parameters (listed in Section 2.4) were originally included in the classification tree but in lower nodes of the tree. Furthermore, we observed that they caused inconsistencies because of data over-fitting (e.g., the classification associates a high agreement to a node characterized by an elongated parcel simply because of one particular parcel). It was, thus, decided to discard those factors for further analyses.

As a second step, we decided to directly model the estimated correlation coefficient from the three retained parameters (see above). For this, we built the following linear regression model:

$$Y_i = \alpha_0 + \alpha_1 \text{SameCrop}_i + \alpha_2 \text{PurePixels}_i + \alpha_3 \text{PercLost}_i + \varepsilon_i, \quad (4)$$

where α_j are the parameters of the model and where for each i -th parcel:

- SameCrop_i is equal to 1 if the same crop is found in the surrounding and equal to 0 otherwise;
- PurePixels_i is the number of clean pixels;
- PercLost_i is the percentage of Sentinel-2 pixels lost after the 5 m negative buffer;
- ε_i is an error term;
- Y_i is the transformed correlation coefficient (i.e., using Fisher's transformation).

The use of Fisher's transformation is important in order to (i) meet the normality hypothesis of the linear regression model and (ii) avoid the modelled nonsensical correlation coefficient being larger than 1 or smaller than -1 [29].

Using this model, we could:

- Estimate the expected correlation coefficient for any parcel;
- Evaluate the uncertainty of an expected value;
- Evaluate the probability that the parcel's correlation coefficient would exceed a given threshold.

For each of these three potential applications, one must use the fitted model and rely on the inverse of the Fisher's transformation.

4. Results

In this chapter, we summarize the results for each of the three methods of the similarity assessment. Section 4.1 presents the results of the independent visual assessment of two experts and subsequent evaluation of their agreement via a weighted Cohen's kappa index. Section 4.2 shows the outcome from the statistical approach using the correlation coefficient as an index of similarity. Section 4.3 deals with the outcome of the similarity model and the estimation of limiting geospatial parameters of parcels, using criteria from Section 3.2. From the test sample, we can estimate the suitability of Sentinel-2 and compare the results with the statistical assessment of NDVI time series pairs in Section 4.4.

4.1. Visual Assessment

The similarity of the time series pairs was assessed by two independent operators (A, B). The assessment classification contained four categories: agreement, deviation, doubtful, and no decision taken due to lack of observations. The results are summarized in Table 1. Table 2 outlines the distribution of the operators' rates, while Table 3 shows how we assigned the weights to each category. The relationship "agreement–disagreement" was assigned a weight of 2, while "agreement/disagreement–doubtful" was assigned a weight of 1.

Table 1. Results from the visual assessment.

	Agreement	Deviation	Doubtful	No Decision Due to Lack of Observations
Operator (A)	87%	5.5%	5.5%	2%
Operator (B)	73%	19%	6%	2%

Table 2. The distributions of operators' rates.

Operator (B)	Operator (A)				Sum
	Agreement	Deviation	Doubtful	Sum	
Agreement	625	2	7	634	
Deviation	95	45	26	166	
Doubtful	36	2	15	53	
Sum	756	49	48	853	

Table 3. The assignment of weights.

Operator (B)	Operator (A)		
	Agreement	Deviation	Doubtful
Agreement	0	2	1
Deviation	2	0	1
Doubtful	1	1	0

Using the values from Tables 2 and 3, we obtained the weighted Cohen's kappa equal to 0.43 (95% confidence interval: 0.36–0.49). According to [26], the value could be interpreted as a moderate agreement.

4.2. Statistical Assessment using Correlation Coefficient

A scatter plot between PlanetScope and Sentinel-2 NDVI values confirmed that the assumptions of the Pearson's correlation coefficient were satisfied. After calculating the correlation coefficient for each pair, we performed a test on the confidence interval using the threshold described in Section 3.2 to assign the similarity status. The results are reflected in Table 4.

Table 4. Results from the statistical assessment.

Agreement	Deviation	Sum
784	83	867
90%	10%	100%

One can already observe that the rate of agreement is higher for the statistical approach than for the visual assessment. This is particularly noteworthy with operator B. For the present dataset, contrary to the visual evaluation, the statistical approach was able to provide a similarity assessment for each of the 867 parcels. However, there was an obvious inferior limit of a minimum of 3 observations when computing the correlation coefficient.

4.3. Fitting the Similarity Model with the Parcels' Characteristics

As described in Section 3.5, we relied on the linear model from Equation (4). The model's parameters were estimated using ordinary least squares. The estimated values and some statistics can be found in Table 5.

Table 5. Estimated parameters of the regression model and their significant tests.

Term	Estimate	Standard Error	t Ratio	Prob> t
Intercept	1.828	0.239	7.655	0.000
SameCrop	0.531	0.077	6.918	0.000
PurePixels	0.015	0.006	2.388	0.009
PercLost	−0.798	0.281	−2.834	0.003

The intercept of the model (i.e., α_0) was positive, indicating that the values of the correlation coefficients tended to be high for the whole dataset. The parameters for the “same crop” and for “pure pixels” were significantly positive. This means that having the same crop around the parcel or containing more pure pixels increases the chances of having a large correlation coefficient. On the other hand, the parameter for the percentage of pixels lost after the negative buffer was negative, indicating that the larger the loss, the smaller the correlation coefficient. Figure 8 illustrates (a) parcels having at least partly the same crop in the surroundings as the analyzed parcel; and (b) isolated parcels with different crops surrounding them compared to the tested parcel.

The worst case is, thus, met for a parcel that (i) is not surrounded by the same crop and (ii) does not have a single pure pixel. In these conditions, the estimated correlation coefficient is equal to 0.77 and the 95% confidence interval is [0.68,0.84] (the interval is not symmetric around the estimated correlation coefficient because of the inverse of Fisher's transformation).

Applying the threshold for the similarity model set described in Section 3.2, we can deduce the exact values of the parcel's geometric parameters, which we consider as spatial limits for Sentinel-2 suitability (i.e., causing difficulty in extracting meaningful information). Looking at Figure 8, we observe that for all parcels surrounded (fully or partially) with the same crop, the probability of exceeding the correlation coefficient of 0.75 is higher than 85%. On the other hand, for the crop isolated parcels, we might generalize the following assertion—the parcels containing less than eight full pixels and losing more than 60% of their pixels measured after the application of a 5 m buffer have a high probability (i.e., close to 80% or more) of having a correlation coefficient below the value of 0.75. Hence,

we consider the parcels to be compliant with these criteria (illustrated with the red box in Figure 8b) as potentially problematic, with a high risk that Sentinel-2 will not provide a reliable NDVI profile (i.e., group 2, Figure 6).

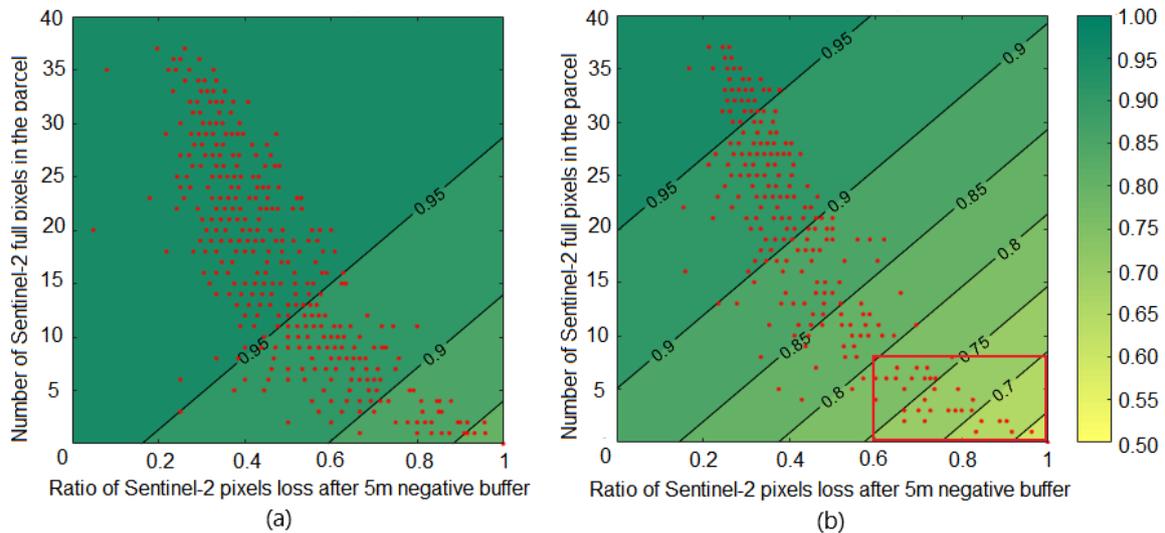


Figure 8. Model quantifying the probability of exceeding the correlation coefficient value of 0.75 as a function of the number of full pixels in the parcel and the percentage of pixels that are lost after application of the 5 m buffer. (a) Parcels with the same crop surroundings and (b) parcels with different crop surroundings. The background colors represent the variability of the probability. The red box illustrates the limiting criteria.

4.4. Comparison of the Results from the Statistical Assessment and Similarity Model

Applying the limiting criteria retroactively to the tested dataset of isolated parcels (325 parcels), we obtained 271 parcels above the limiting criteria (estimated group 3) and 54 parcels below the limiting criteria, i.e., group 2. These two groups of parcels were compared with the results from Section 4.2 to observe the distribution of the deviating parcels (group 1). Table 6 shows that 37% of parcels out of the vulnerable group 2 did not pass the similarity test, while only 10% of parcels from the group of parcels above the limiting criteria gave a negative similarity outcome in the statistical assessment.

Table 6. Application of the limits derived from the model on the samples of isolated parcels and the distribution of group 1.

	Parcels Below the Limiting Criteria (Estimated Group 2)	Parcels above the Limiting Criteria (Estimated Group 3)
Number of isolated parcels	54	271
Proportion of parcels that deviated according to the statistical assessment, i.e., group 1 (Section 4.2)	37%	10%

5. Discussion

For the purpose of this study, a set of normalized difference vegetation index time series from Sentinel-2 calculated over a number of fields with various geospatial parameters was compared with a set of NDVI time series from a higher resolution satellite (i.e., PlanetScope) extracted for the same group of fields. A similarity assessment of the two sets of NDVI time series was used as the main tool to reach the research goals. The latest studies dealing with image data time series in the agriculture domain mainly address the issue of understanding and monitoring of the crop phenology [30–32],

focusing on cropland mask construction [33,34] and land cover classification [11,35]. The use of time series to assess cropping intensity became an actual topic due to its importance as a critical input data variable for many climate land surface and crop models [36,37]. The most recent research studies in the context of the CAP exploit the Sentinel-2 data time series to detect agriculture land use anomalies [38] or map the grassland use intensity [39]. The paramount topic in the agriculture sector is, however, crop identification. Previously, [40] presented a first experience with crop type classification using one of the first acquired (preoperational) Sentinel-2 images. The benefits of the subsequently available time series were explored in [41], using a random forest (RF) classifier for crop mapping purposes. This study confirmed that multitemporal information increases the crop type classification accuracy. The benchmark for various classification methods on simulated Sentinel-2 time series was described in [42]. Further research on real multitemporal Sentinel-2 data [43] compared the performance of supervised machine learning techniques (support vector machines (SVMs) and RF) with respect to their accuracy and execution time, highlighting the superior performance of Sentinel-2 for this type of application when compared to Landsat-8 OLI. Additionally, [44] showed that the SVM and RF outperformed the traditional statistical maximum likelihood classification method. Since the temporal resolution of Sentinel-2 is constrained by cloud cover conditions, [45–48] explored multisource data (Sentinel-1–Landsat-8) use to complement Sentinel-2 time series and increase the crop type classification accuracy.

An operational solution for multitemporal crop classification approaches at a national scale was developed by European Space Agency's Sentinel-2 for Agriculture project (Sen2Agri). Sen2Agri delivered an open-source system that derives 4 basic products from Sentinel-2 and Landsat 8 time series: cloud-free composites, dynamic cropland masks, crop type maps, and vegetation status indicators. Additionally, [49] described the overall approach and key principles embedded in the processing chain and presented the results from an independent validation of the monthly cropland masks and crop type maps. However, the CbM approach, as described in [4], is primarily based on the detection of particular phenomena ("markers" and behavior) to continuously monitor a presumed agricultural business scenario, allowing early feedback to be issued to individual farmers. Crop mapping is not foreseen as a key component, but crop identification remains a valid and relevant ancillary instrument. Member States adopting CbM use the crop identification and classification algorithms for various reasons.

Regarding the topic of monitoring small fields, recent research [50] dealt with assessments on the impact of sensor pixel sizes for EU CAP. However, the study limited its consideration to a simple assumption that for fields with no pure pixels inside their boundaries, no CAP services can be provided, mainly pointing to crop determination. Our study presents the application of the comparative spatiotemporal time series analysis to assess the limits of satellite data regarding the smallest traced entity in the context of CbM, taking into account its full complexity.

The potential of Sentinel-2 data for the state assessment and the event detection on small parcels, as presented in Section 4.2, was found to be better than expected. The evaluation was performed over cloud-free pixels with a sufficient number of observations, and therefore the constraint for small parcels in the agricultural regions very frequently covered by clouds during the growing season, as mentioned for instance in [49], was not considered.

Concerning the visual assessment, a difference (0.43 Cohen's kappa) between the operators was observed. Even if according to [26] this might still be assessed as a moderate agreement, other researchers studies [51] considered this as a low degree of the agreement. While one operator (A) evaluated the NDVI time series similarly, and thus provided the same information content for 87% of parcels, the other operator (B) reached a level of 73% from the total amount. In comparison with the statistical approach, which resulted in 90% of agreements, there was visible conformity with operator A, whilst operator B most likely adopted stricter evaluation criteria. This confirms that in this case, the visual assessment was a subjective methodology strongly influenced by experience and personal perception. As such, we considered it as a supportive element for further analyses.

Regarding the PlanetScope NDVI time series (as a reference for our analysis), based on expert judgment it was concluded that for 2% there were doubts about the correctness due to the weak trend of the profiles. No dependency on the parcel size was found. For 80% of the parcels, the profiles of both sensors were assessed as not being similar. Additionally, 60% of these parcels were below our limiting criteria (<8 pixels and >60% pixels lost). After detailed analysis, we could observe that 65% of these parcels had elongated geometry, measuring 4–15 m wide. Parcels with such a shape are generally vulnerable, especially if the long side is perpendicular to the eventual shift. For the proper decision on the correctness, we would, however, need to have collected evidence during the crop season on the field.

From the investigation of the contextual parameters of the parcels, there is no clear evidence that the type of arable crop in the agriculture parcel would have an impact on the similarity of the NDVI time series pairs. The research was, however, restricted to arable crops only. This restriction is caused by the chosen statistical evaluation methodology of similarity using the correlation coefficient. The correlation coefficient uses averages as a pivot in order to measure the linearity between each NDVI time series pair. Thus, it assumes that the times series have marked trends over time. By contrast, parcels where slight or no trend is expected (e.g., permanent grazed grassland or permanent crops) exhibit time series with an almost constant value, and few fluctuations can be considered as noise. Hence, applying this methodology on these crops is not appropriate. Although the arable crop itself does not prove to have an impact on the similarity, the surface surrounding the evaluated parcel seems to be a significant factor. The results obtained from the study showed a higher rate of positive similarity status for parcels having the same crops in the surroundings. As a limitation of this result, we consider the fact that the algorithm used for the decision did not consider the amount of surrounding crops. In this respect, the resulting number of parcels with the same crop in the surrounding could be exaggerated (i.e., biased). We, therefore, assume that the positive influence of having the same crop in the surroundings could be even higher than the one shown in Figure 8.

Further analyses on the relationship between the correlation coefficient and quantitative parameters identified two parameters that were more relevant than the others; a number of clean or full pixels and a share of pixels lost after applying half of the negative pixel buffer. The possible explanation for that is the importance of the shape and position of the parcel on the raster. As outlined in Section 2.4, the parameter expressing the loss of pixels took into account both the shape and position, and emerged as a very suitable descriptor of the parcel. By contrast, the size of the parcel was revealed to be a vague description, with an apparent lack of a strong correlation.

Working further with the three selected parameters, i.e., surrounding the parcel, the ratio of Sentinel-2 pixels lost, and the number of full pixels after application of the buffer, we built a model displaying the possible impacts of their relationships, and combinations on the probability of the estimated correlation coefficient. From a priori defined thresholds, we derived the limiting parameters of the parcel geometry, according to which we could estimate and quantify the proportion of parcels (estimated group 2) with a higher risk of Sentinel-2 data failure to provide conclusive information about the parcel. Looking at Figure 8, one might theoretically question the stability of the model, as the tested sample does not cover all possible combinations of parameters. However, some of these combinations or dimensions are simply not realistic in the field.

The results in Table 6 show the proportions of parcels assessed as deviating in the similarity assessment exercise (group 1) for both groups of isolated parcels below (i.e., estimated group 2) and above the limiting parameters (estimated group 3). The proportions demonstrate the correctness and validity of the proposed limiting criteria.

As indicated earlier in Section 2.2, the geolocation accuracy of Sentinel-2B was described as unstable during the tested period. We, therefore, investigated a representation of Sentinel-2B in the NDVI time series sample and the proportion of Sentinel-2B in the NDVI profiles of parcels that deviated (group 1, Section 4.2). For both cases, the participation of Sentinel-2B amounted to approximately 40%. We did not observe any higher occurrence of Sentinel-2B values in those NDVI time series that did not

pass the test of similarity. We concluded that Sentinel-2B instability did not impact the results of this study. A positive aspect of the applied approach might be the fact that the curve matching method applied to compare the similarity of NDVI profiles was not highly sensitive to a single date outlier, unless it was a periodical or longer duration phenomenon.

As mentioned in the introduction section, the quality of the produced time series has a direct impact on the similarity assessment results. The cloud cover mask and the co-registration accuracy leave much room for improvement and we assume that the potential of Sentinel-2 imagery will be even higher in the near future, for instance after the introduction of a global reference image [7].

6. Conclusions

The introduction of the checks by monitoring approach in the framework of the European Common Agriculture Policy gave an impulse to explore the ability of Sentinel-2 data to detect a parcel's behavior in terms of the geometry of that parcel. The research demonstrated that the size of the parcel is not an appropriate measure for the evaluation of Sentinel-2 data suitability for CbM. The number of full pixels and the ratio of pixels lost after the application of a half-pixel-wide negative buffer emerged as more appropriate parameters. Our findings also suggest that the surrounding land use or land cover is a relevant factor that needs to be duly taken into account.

For the majority of tested parcels, the Sentinel-2 data provided the same or similar information about the state and the activity on the field as PlanetScope imagery. Considering the deliberate small size of the parcels in the studied sample, the results proved the high potential of Sentinel-2 imagery. Thus, for the checks by monitoring, the study provided evidence of a good design strategy involving the trade-off between the temporal and spatial resolution of the Sentinel-2 satellite.

The limiting criteria expressed by geospatial parcel parameters serve to identify a group of “vulnerable” agricultural parcels for which there is a high probability that Sentinel-2 might not be a suitable satellite to assess their state and related agricultural activity. Member states, therefore, might adopt the definition of these spatial limits when designing their monitoring approaches.

In our research, we applied the comparison technique on the time series pairs retrieved from two image datasets of different spatial resolution to estimate the potential and spatial limitations of the coarser one. This innovative approach, as well as the whole evaluation strategy with thresholds adapted to the specific needs, could be used for a similar assessment of any other satellite data and application purpose.

Author Contributions: Conceptualization, B.V. and D.F.; methodology, B.V. and D.F.; software, D.F., C.W., and B.V.; validation, B.V.; formal analysis, D.F., B.V., and S.L.; investigation, B.V. and D.F.; resources, S.L. and B.V.; writing—original draft preparation, B.V. and D.F.; writing—review and editing, W.D.; B.V.; D.F., C.W., S.L.; visualization, B.V. and D.F.; project administration, B.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the Planet Labs Company for providing the necessary image dataset and technical support regarding PlanetScope imagery and the Planet platform. We are also very grateful to Davide De Marchi for sharing scripts and technical support for the JEODPP platform.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CEMA. *CAP 2021–2027 Study, CAP Reform Highlights*; European Agricultural Machinery Association: Brussels, Belgium, 2019.
2. European Commission. *Commission Implementing Regulation (EU) No 809/2014 of 17 July 2014 Laying down Rules for the Application of Regulation (EU) No 1306/2013 of the European Parliament and of the Council with Regard to the Integrated Administration and Control System, Rural Development Measures and Cross Compliance*; European Commission: Brussels, Belgium, 2014.

3. European Commission. *Commission Implementing Regulation (EU) 2018/746 of 18 May 2018 Amending Implementing Regulation (EU) No 809/2014 as Regards Modification of Single Applications and Payment Claims and Checks*; European Commission: Brussels, Belgium, 2018.
4. Devos, W.; Lemoine, G.; Milenov, P.; Fasbender, D.; Loudjani, P.; Wirnhardt, C.; Sima, A.; Griffiths, P. *Second Discussion Document on the Introduction of Monitoring to Substitute OTSC: Rules for Processing Application in 2018–2019*; European Commission: Brussels, Belgium, 2018.
5. Nationaal Georegister. Available online: <http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home> (accessed on 14 May 2019).
6. European Space Agency. *S2 MPC Level 2A Data Quality Report*; European Space Agency: Paris, France, 2018.
7. European Space Agency. *S2 MPC Level 1C Data Quality Report*; European Space Agency: Paris, France, 2018.
8. Planet Labs Inc. Planet Imagery Product Specifications. 2020. Available online: https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (accessed on 30 May 2020).
9. Lemajic, S.; Vajsova, B.; Aastrand, P. *New Sensors Benchmark Report on PlanetScope: Geometric Benchmarking Test for Common Agricultural Policy (CAP) Purposes*; European Commission: Brussels, Belgium, 2018.
10. Dobrinic, D.; Gasparovic, M.; Župan, R. Horizontal Accuracy Assessment of PlanetScope, RapidEye and WorldView-2 Satellite Imagery. In Proceedings of the 18th International Multidisciplinary Scientific Geoconference SGEM 2018, Albena, Bulgaria, 2–8 July 2018.
11. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *Isprs J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [[CrossRef](#)]
12. Hird, J.N.; McDermid, G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sens. Environ.* **2009**, *113*, 248–258. [[CrossRef](#)]
13. Shi, W.; Hao, M. Analysis of spatial distribution pattern of change-detection error caused by misregistration. *Int. J. Remote Sens.* **2013**, *34*, 6883–6897. [[CrossRef](#)]
14. Xiaolong Dai; Khorram, S. The effects of image misregistration on the accuracy of remotely sensed change detection. *Ieee Trans. Geosci. Remote Sens.* **1998**, *36*, 1566–1577. [[CrossRef](#)]
15. Kukawska, E.; Lewiński, S.; Krupiński, M.; Malinowski, R.; Nowakowski, A.; Rybicki, M.; Kotarba, A. Multitemporal Sentinel-2 data-remarks and observations. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Brugge, Belgium, 27–29 June 2017; pp. 1–4.
16. Kriegl, F.J.; Malila, W.A.; Nalepka, R.F.; Richardson, W. Preprocessing Transformations and Their Effects on Multispectral Recognition. In Proceedings of the Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, USA, 13–16 October 1969; University of Michigan: Ann Arbor, MI, USA, 1969; pp. 97–131.
17. Rouse, W.; Haas, R.H.; Schell, J.A.; Deering, D.W. *Monitoring Vegetation Systems in the Great Plains with ERTS*; NASA SP-351: Washington, DC, USA, 1973; pp. 309–317.
18. Soille, P.; Burger, A.; Marchi, D.; Kempeneers, P.; Rodriguez Aseretto, D.; Syrris, V.; Vasilev, V. A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Gener. Comput. Syst.* **2017**, *81*, 30–40. [[CrossRef](#)]
19. Planet Explorer. Available online: <https://www.planet.com/explorer/> (accessed on 14 May 2019).
20. Devos, W.; Lemoine, G.; Milenov, P.; Fasbender, D. *Technical Guidance on the Decision to go for Substitution of OTSC by Monitoring*; European Commission: Brussels, Belgium, 2018.
21. Vajsova, B.; Fasbender, D.; Wirnhardt, C.; Lemajic, S. *Applicability Limits of Sentinel-2 Data Compared to Higher Resolution Imagery for CAP Checks by Monitoring*; European Commission: Brussels, Belgium, 2019.
22. Duveiller, G.; Fasbender, D.; Meroni, M. Revisiting the concept of a symmetric index of agreement for continuous datasets. *Sci. Rep.* **2016**, *6*, 1–14. [[CrossRef](#)]
23. Akoglu, H. User’s guide to correlation coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91–93. [[CrossRef](#)]
24. Schober, P.; Boer, C.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)]
25. Mukaka, M. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.
26. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]
27. Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213–220. [[CrossRef](#)] [[PubMed](#)]

28. Weighted Cohen's Kappa|Real Statistics Using Excel. Available online: <http://www.real-statistics.com/reliability/interrater-reliability/weighted-cohens-kappa/> (accessed on 24 June 2019).
29. Fasbender, D.; Vajsová, B.; Wirnhardt, C.; Lemajic, S. A Simple Similarity Index for the Comparison of Remotely Sensed Time Series with Scarce Simultaneous Acquisitions. *Remote Sens.* **2019**, *11*, 1527. [[CrossRef](#)]
30. Pan, Z.; Huang, J.; Qingbo, Z.; Wang, L.; Cheng, Y.; Zhang, H.; Alan Blackburn, G.; Yan, J.; Liu, J. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 188–197. [[CrossRef](#)]
31. Mercier, A.; Betbeder, J.; Baudry, J.; Denize, J.; Leroux, V.; Roger, J.-L.; Spicher, F.; Hubert-Moy, L. Evaluation of Sentinel-1 and -2 time series to derive crop phenology and biomass of wheat and rapeseed: Northern France and Brittany case studies. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI, International Society for Optics and Photonics, Strasbourg, France, 9–11 September 2019; Volume 11149, p. 1114903.
32. d'Andrimont, R.; Taymans, M.; Lemoine, G.; Ceglar, A.; Yordanov, M.; van der Velde, M. Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series. *Remote Sens. Environ.* **2020**, *239*, 111660. [[CrossRef](#)]
33. Valero, S.; Morin, D.; Inglada, J.; Sepulcre, G.; Arias, M.; Hagolle, O.; Dedieu, G.; Bontemps, S.; Defourny, P.; Koetz, B. Production of a Dynamic Cropland Mask by Processing Remote Sensing Image Series at High Temporal and Spatial Resolutions. *Remote Sens.* **2016**, *8*, 55. [[CrossRef](#)]
34. 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](#)]
35. Fauvel, M.; Lopes, M.; Dubo, T.; Rivers-Moore, J.; Frison, P.-L.; Gross, N.; Ouin, A. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sens. Environ.* **2020**, *237*, 111536. [[CrossRef](#)]
36. Hao, P.-Y.; Tang, H.-J.; Chen, Z.-X.; Yu, L.; Wu, M.-Q. High resolution crop intensity mapping using harmonized Landsat-8 and Sentinel-2 data. *J. Integr. Agric.* **2019**, *18*, 2883–2897. [[CrossRef](#)]
37. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624. [[CrossRef](#)]
38. Kanjir, U.; Đurić, N.; Veljanovski, T. Sentinel-2 Based Temporal Detection of Agricultural Land Use Anomalies in Support of Common Agricultural Policy Monitoring. *Isprs Int. J. Geo Inf.* **2018**, *7*, 405. [[CrossRef](#)]
39. Griffiths, P.; Nendel, C.; Pickert, J.; Hostert, P. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sens. Environ.* **2019**, *238*, 111124. [[CrossRef](#)]
40. 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](#)]
41. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [[CrossRef](#)]
42. Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* **2015**, *7*, 12356–12379. [[CrossRef](#)]
43. Sitokostantinou, V.; Papoutsis, I.; Kontoes, C.; Lafarga Arnal, A.; Armesto Andrés, A.P.; Garraza Zurbano, J.A. Scalable Parcel-Based Crop Identification Scheme Using Sentinel-2 Data Time-Series for the Monitoring of the Common Agricultural Policy. *Remote Sens.* **2018**, *10*, 911. [[CrossRef](#)]
44. Feng, S.; Zhao, J.; Liu, T.; Zhang, H.; Zhang, Z.; Guo, X. Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3295–3306. [[CrossRef](#)]
45. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved Early Crop Type Identification by Joint Use of High Temporal Resolution SAR and Optical Image Time Series. *Remote Sens.* **2016**, *8*, 362. [[CrossRef](#)]
46. Sonobe, R.; Yamaya, Y.; Tani, H.; Wang, X.; Kobayashi, N.; Mochizuki, K. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *Giscience Remote Sens.* **2017**, *54*, 918–938. [[CrossRef](#)]
47. Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium. *Remote Sens.* **2018**, *10*, 1642. [[CrossRef](#)]

48. Sun, C.; Bian, Y.; Zhou, T.; Pan, J. Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region. *Sensors* **2019**, *19*, 2401. [[CrossRef](#)] [[PubMed](#)]
49. Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* **2019**, *221*, 551–568. [[CrossRef](#)]
50. Meier, J.; Mauser, W.; Hank, T.; Bach, H. Assessments on the impact of high-resolution-sensor pixel sizes for common agricultural policy and smart farming services in European regions. *Comput. Electron. Agric.* **2020**, *169*, 105205. [[CrossRef](#)]
51. L McHugh, M. Interrater reliability: The kappa statistic. *Biochem. Med. Časopis Hrvat. Društva Med. Biokem. Hdm* **2012**, *22*, 276–282. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).