Evaluating Combinations of Temporally Aggregated Sentinel-1, Sentinel-2 and Landsat 8 for Land Cover Mapping with Google Earth Engine

Luis Carrasco [1,2,3,*], Aneurin W. O'Neil [1], R. Daniel Morton [1] and Clare S. Rowland [1]

[1]   NERC Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK; aneurin@ceh.ac.uk (A.W.O.); danm@ceh.ac.uk (R.D.M.); clro@ceh.ac.uk (C.S.R.)
[2]   National Institute for Mathematical and Biological Synthesis, 1122 Volunteer Boulevard, University of Tennessee, Knoxville, TN 37996, USA
[3]   Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville, TN 37996, USA
*   Correspondence: lcarrasc@utk.edu; Tel.: +1-865-438-9959

check for updates

Abstract: Land cover mapping of large areas is challenging due to the significant volume of satellite data to acquire and process, as well as the lack of spatial continuity due to cloud cover. Temporal aggregation—the use of metrics (i.e., mean or median) derived from satellite data over a period of time—is an approach that benefits from recent increases in the frequency of free satellite data acquisition and cloud-computing power. This enables the efficient use of multi-temporal data and the exploitation of cloud-gap filling techniques for land cover mapping. Here, we provide the first formal comparison of the accuracy between land cover maps created with temporal aggregation of Sentinel-1 (S1), Sentinel-2 (S2), and Landsat-8 (L8) data from one-year and test whether this method matches the accuracy of traditional approaches. hirty-two datasets were created for Wales by applying automated cloud-masking and temporally aggregating data over different time intervals, using Google Earth Engine. Manually processed S2 data was used for comparison using a traditional two-date composite approach. Supervised classifications were created, and their accuracy was assessed using field-based data. Temporal aggregation only matched the accuracy of the traditional two-date composite approach (77.9%) when an optimal combination of optical and radar data was used (76.5%). Combined datasets (S1, S2 or S1, S2, and L8) outperformed single-sensor datasets, while datasets based on spectral indices obtained the lowest levels of accuracy. The analysis of cloud cover showed that to ensure at least one cloud-free pixel per time interval, a maximum of two intervals per year for temporal aggregation were possible with L8, while three or four intervals could be used for S2. This study demonstrates that temporal aggregation is a promising tool for integrating large amounts of data in an efficient way and that it can compensate for the lower quality of automatic image selection and cloud masking. It also shows that combining data from different sensors can improve classification accuracy. However, this study highlights the need for identifying optimal combinations of satellite data and aggregation parameters in order to match the accuracy of manually selected and processed image composites.

Keywords: cloud computing; cloud masking; data fusion; gap filling; radar; supervised classifications

1. Introduction

Land cover maps help us better understand environmental processes, such as water and biochemical cycles, energy exchanges, or biodiversity alterations [1]. Land cover has been recognised as an Essential Climate Variable (ECV) [2] and has also been proposed as a Satellite Remote Sensing

Essential Biodiversity Variable [3]. The characterisation of land cover at a high thematic and spatial resolution (i.e., 30 m) enables monitoring of the Earth's surface at a scale comparable to human activity [4]. During the last few years, there has been an explosion in the availability of medium–high resolution satellite data, and many visualisation and processing platforms have emerged, resulting in a new generation of land cover maps and methods [5].

Land cover mapping typically involves a number of stages, including data selection, data processing, and land cover classification (supervised or unsupervised). When mapping large regions (i.e., national or regional scale), the selection and processing of satellite images poses several challenges. One of the first challenges is the large amount of data that must be extracted and processed. For example, the UK Land Cover Map 2015 (LCM2015), a national land cover map which classified two-date image composites over the course of a year, used over 100 Landsat images, each of which required manual cloud-masking. This volume of data requires a large amount of storage, together with significant computing power and time. Another critical challenge is the lack of spatial continuity due to differential cloud covers [6]. Cloud cover affects the spatial frequency of cloud-free images, resulting in spatial inconsistencies in classification. Additionally, acquiring complete spatial coverage for cloudy areas of the world requires the use of numerous partially cloudy images. This increases the number of images that must be obtained, stored, and processed.

The launch of new medium–high resolution satellites, such as those of the Copernicus Programme, is increasing the frequency of free image acquisitions. The combination of data from Landsat 8 (L8) and Sentinel-2A/B should provide an average minimum of one image every three days [7]. The Sentinel-1 radar imaging satellite (S1) is particularly promising in terms of dealing with data continuity issues, as the radar signal is not affected by clouds. The increase in the number of sensors and free satellite data has led to the emergence of new computing platforms that help users select and process large volumes of geospatial data. A good example is Google Earth Engine (GEE), a cloud-based computing platform which provides easy access to satellite datasets on a planetary scale [8]. Platforms such as GEE allow users to avoid storing images locally and provide greater computing power for image processing and analysis. This combination of new sensors and platforms is allowing researchers to find new approaches to large-scale land cover mapping by changing the way data are selected and processed.

The computing power of these new platforms, enables the application of complex algorithms to large amounts of data. New land cover mapping methods exploit this to fill the gaps generated by the lack of medium–high resolution data for certain dates or by automated cloud-masking procedures. Gap-filling techniques such as data fusion [9,10], pixel unmixing [11], data interpolation [12], or best-pixel selection [13,14] have all been applied to time series data for the land cover mapping of large areas. Temporal aggregation, which is the use of metrics (i.e., mean, median, max/min, etc.) derived from measurements such as reflectance or the Normalized Difference Vegetation Index (NDVI) that are calculated over a period of time [15,16], is also becoming a popular approach to dealing with data gaps and inconsistent numbers of available satellite images [17–19]. These techniques differ from approaches that use all of the available data [20], as they significantly reduce the volume of data to produce smaller, more manageable data sets. Temporal aggregation is simpler than other gap filling algorithms, and platforms such as GEE enable it to be applied over hundreds of images rapidly [5]. This means that the cost and time needed to produce large-scale land cover maps with no data gaps can be significantly reduced. The major concern with this methodology is that it relies on there being enough images available to create reliable aggregated measurements over a period of time. Insufficient data or poor quality automated pre-processing methods, such as cloud-detection algorithms, may lead to low classification accuracy [17]. However, it has not yet been tested whether land cover classification using temporal aggregation techniques is more accurate than manually pre-processed data. The reliability of temporal aggregations will depend on the number of good images over the temporal aggregation period. This will vary spatially and temporally, as well as depending on the repeat frequency of the satellite sensor. Therefore, an assessment of the optimal type, or combination, of satellite input data to obtain satisfactory land cover classifications is needed.

Here, we assessed the accuracy of land cover maps created using temporal aggregation of one-year of data derived from L8, S1, and Sentinel-2 (S2) satellites. The two main objectives were to (1) test whether land cover classifications, created using temporal aggregation of large amounts of automatically-processed satellite data, outperform those created with traditional non-aggregated two-date composites and (2) compare the classification accuracy of temporally aggregated data from different sensors. To do this, multiple datasets were created using the temporal aggregation of satellite-derived measurements taken over different time intervals within one year. GEE was used to apply automated cloud-masking and to temporally aggregate the data. All data were resampled to 30 metres to match the spatial resolution of the most coarsely resolved sensor (L8). This was done to isolate the effects of temporally aggregated measurements on the classification accuracy. A wide variety of datasets were created, including combined multi-sensor data, and supervised classifications were used to create land cover maps. The accuracy of the land cover maps was then assessed using a field-based land cover dataset. The methods were tested in the UK which is a relatively cloudy area of the world [21].

2. Materials and Methods

2.1. Study Area

The study area is in the UK and covers most of Wales (82.7%) and some bordering regions of England (Figure 1). The area covers 200×100 km^2 and corresponds to tiles T30UVC and T30UVD of S2. The region is geographically diverse, with mountainous areas in the north and centre. It is characterised by a maritime climate, with mild temperatures and rain on more than 200 days a year (https://www.metoffice.gov.uk/), so is often cloudy. The lowlands are dominated by grasslands dedicated to animal grazing, with some cropland in the east and near the north coast. The main urban areas are located in the south. Forested patches are abundant and distributed throughout the region. Semi-natural lowland habitats, formed by heathlands, natural grasslands, and saltmarshes are dispersed and fragmented, and they are mainly found in the coastal areas. Uplands are formed by a mosaic of coniferous woodlands, bogs, heath, and semi-natural grasslands [22].
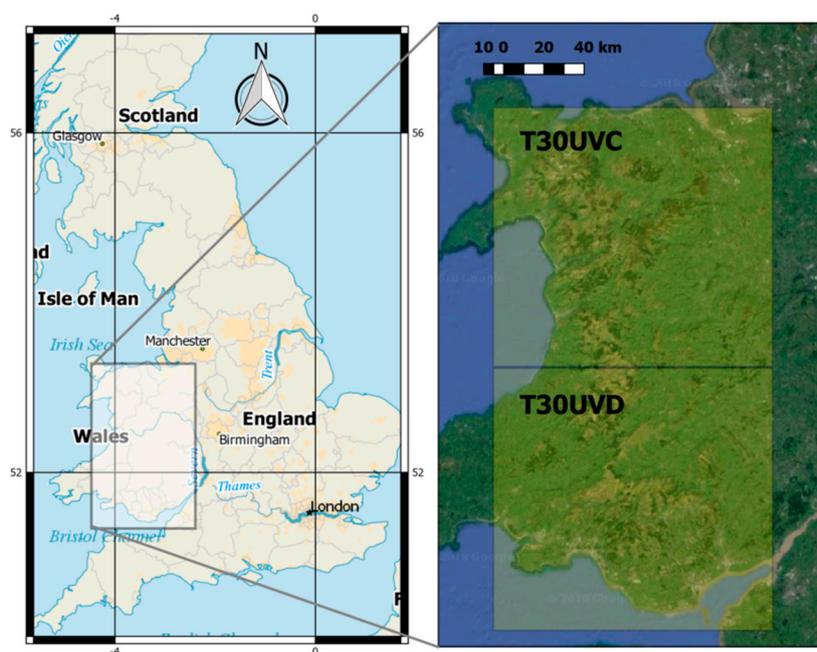

Figure 1. The study area: The area covered by tiles T30UVC and T30UVD of Sentinel-2, corresponding to the study area, are shown in the map on the right.

2.2. Datasets

The datasets were created using images from three satellites: L8, S2, and S1. L8 and S2 carry optical sensors and are the most common current data sources for medium–high resolution land cover mapping, while S1 carries a radar instrument (Table 1).

Table 1. The main features of the sensors used in this study.

	Landsat 8	Sentinel-2	Sentinel-1
Sensor (type)	OLI (optical)	MSI (optical)	C-SAR (radar)
Spatial resolution (m)	15 */30/100 *	10/20/60 *	5 **
Number of bands (used)	11 (7)	12 (9)	1
Spectral bands (μm)	0.435–0.451, 0.452−0.512, 0.533–0.590, 0.636–0.673, 0.851−0.879, 1.566−1.651, 10.60–11.19 *, 11.50–12.51, 2.107–2.294, 0.503–0.676 *, 1.363–1.384 *	0.449–0.545 *, 0.458–0.523, 0.543–0.578, 0.650−0.680, 0.698–0.713, 0.733−0.748, 0.773−0.793, 0.785−0.899, 0.855−0.875 *, 0.932−0.958 *, 1.338−1.414 *, 1.565−1.655, 2.100−2.280	
Repeat Frequency (days)	16	10	12
Swath (km)	180	290	80 **
Polarization	Not Applicable	Not Applicable	Dual (HH + HV, VV + VH) **

* Not used in this study; ** Strip Map Mode (used in this study).

Data from October 2016 to September 2017 were used. Data spanning the course of one year is ideal for the identification of land cover types that have seasonality patterns, such as agricultural crops, while also allowing for the analysis of medium-term land cover change [23]. All datasets, with the exception of the "two-date composite" datasets, were created by the temporal aggregation of optical or radar data across different time intervals. For example, a one-interval dataset aggregated the data taken over the whole year into one image, while a two-interval dataset aggregated the data into two images, representing the six-month winter/summer season split, and so on (Figure 2). Datasets with a higher number of intervals aggregated the data across smaller time intervals of equal length.
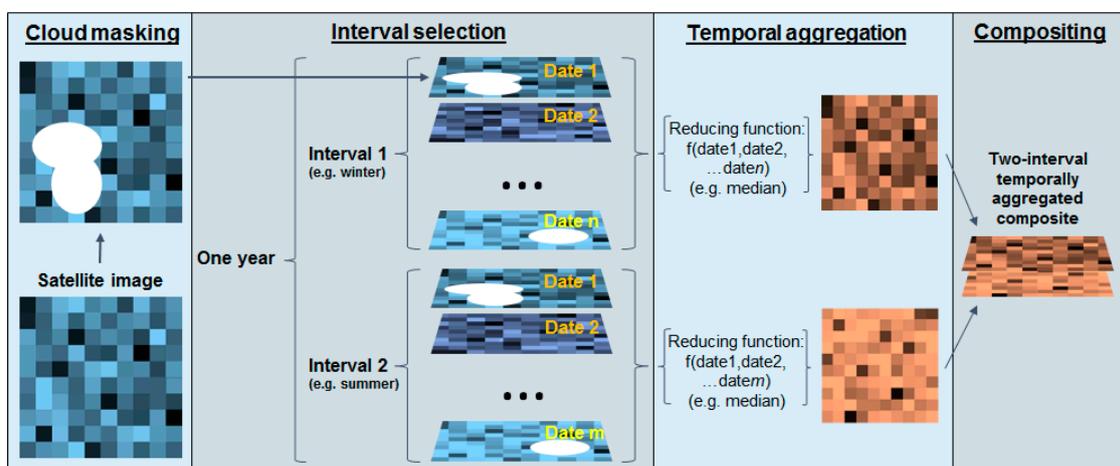

Figure 2. The workflow diagram for the dataset creation using temporal aggregation described in Section 2.2.

The image pre-processing (see the details for each dataset in the following sections) and the aggregation of the data across time intervals was performed using GEE [17], except in the case of the "two-date composite" datasets. For optical data (L8 and S2), automatic cloud and cloud shadow masking methods were used. Cloud-masking creates gaps with no-data in different locations for different image dates. This means that the temporal aggregation will be performed using a varying number of images (dates) for different pixels. In order to assess the effect of cloud gaps on temporal aggregations, maps and histograms were created representing the amount of cloud-free pixels for different time intervals.

All images were resampled to 30 metres to match the original spatial resolution of L8, which offers the coarsest resolution of the sensors used. Doing so minimises the effect of spatial resolution on the final classification accuracies, thus simplifying the comparison between datasets. Additionally, elevation, aspect, and slope bands were added to the datasets to increase the classification accuracy [24]. The datasets were divided according to the sensor or processing procedure. A summary of the datasets main features can be seen in Table 2. The acquisition dates of every image used for the temporal aggregations can be found in Supplementary Material S1.

Table 2. The list of datasets showing data inputs, the temporal aggregation method, and the number of intervals through the year. [1] denotes automatic cloud mask.

Type	Name	Sensor	Bands	Metric	Intervals
Landsat 8	*l8_med2*	L8	B1-7	median	2
	l8_mea2	L8	B1-7	mean	2
	l8_var2	L8	B1-7	median and variance	2
	l8_med1	L8	B1-7	median	1
	l8_mea1	L8	B1-7	mean	1
	l8_var1	L8	B1-7	median and variance	1
Indices	*ndvi_med2*	L8	NDVI	median	2
	ndvi_med1	L8	NDVI	median	1
	ndvi_var1	L8	NDVI	median and variance	1
	ndvi_mea1	L8	NDVI	mean	1
	ndmi_var1	L8	NDVI/NDMI	median and variance	1
	ndmi_med2	L8	NDVI/NDMI	median	2
	ndwi_var1	L8	NDVI/NDWI	median and variance	1
	ndwi_med2	L8	NDVI/NDWI	median	2
Sentinel 2	*s2_med1*	S2	B2-8,11,12	median	1
	s2_var1	S2	B2-8,11,12	median and variance	1
	s2_med2	S2	B2-8,11,12	median	2
	s2_var2	S2	B2-8,11,12	median and variance	2
	s2_mea2	S2	B2-8,11,12	mean	2
	s2_med3	S2	B2-8,11,12	median	3
	s2_med4	S2	B2-8,11,12	median	4
Sentinel 1	*s1_med1*	S1	VV,VH,(VV-VH)	median	1
	s1_med2	S1	VV,VH,(VV-VH)	median	2
	s1_med3	S1	VV,VH,(VV-VH)	median	3
	s1_med4	S1	VV,VH,(VV-VH)	median	4
	s1_med6	S1	VV,VH,(VV-VH)	median	6
	s1_med12	S1	VV,VH,(VV-VH)	median	12
Combined	*s1_s2*	S1; S2	VV,VH,(VV-VH); B2-8,11,12	median	12; 2
	s2_l8	S2; L8	B2-8,11,12; B1-7	median	2; 2
	s1_s2_l8	S1; S2; L8	VV,VH,(VV-VH); B2-8,11,12; B1-7	median	12; 2; 2
Two-date composite	*traditional*	S2	B2-8,11,12	reflectance	2
	auto_cm [1]	S2	B2-8,11,12	reflectance	2

2.2.1. Landsat 8

L8 data were extracted from the USGS Landsat 8 Surface Reflectance Tier 1 dataset provided by GEE. These data are derived from L8's OLI/TIRS sensors and have been orthorectified and atmospherically corrected to obtain surface reflectance. Bands one to seven, with an original spatial resolution of 30 metres, were used in this study. An automatic cloud masking procedure was applied using the C Function of Mask (CFMask; [25]) band included with the Landsat data.

Temporal aggregation was applied by calculating the mean, median, and variance of the reflectance values across all the available images for a specified time interval. A single seven-band image is, therefore, obtained for every interval for each aggregation function. One or two time intervals (with lengths of 12 and six months, respectively) were used for the Landsat 8 datasets. Temporal aggregation methods work best when there are sufficient cloud-free values for each pixel in the time interval to get a representative value from the aggregation function. Intervals that are too short should, therefore, be avoided for the aggregation function to work correctly. Based on the preliminary cloud-cover analysis for Landsat, shorter intervals would have resulted in images with a large number of no-data values or pixels with just a one cloud-free date. This would have increased the risk of using cloud-contaminated data, as the cloud-masking methods do not always work perfectly [26]. In total, six datasets were created using the mean, median, or a combination of median and variance for one or two intervals within the space of a year (Table 2).

2.2.2. Indices

Three spectral indices, the NDVI [27], the Normalized Difference Moisture Index (NDMI; [28]), and the Normalized Difference Water Index (NDWI; [29]), were used to create the "indices" datasets. The NDVI is calculated using the red and near-infrared (NIR) bands, $(\rho_{NIR} - \rho_{RED})/(\rho_{NIR} + \rho_{RED})$, where ρ represents spectral reflectance and characterises the "greenness" of the surface. The NDMI is calculated using the NIR and the mid-infrared (MIR) bands, $(\rho_{NIR} - \rho_{MIR})/(\rho_{NIR} + \rho_{MIR})$ and identifies the moisture content of soil and vegetation. The NDWI is calculated using the NIR and the green bands, $(\rho_{GREEN} - \rho_{NIR})/(\rho_{GREEN} + \rho_{NIR})$, and identifies water bodies. The first two indices are related to the structure and cover of vegetation, and all three indices have been widely used in land cover characterisation [23].

L8 data, atmospherically corrected and cloud masked as described in Section 2.2.1, were used to derive the indices. The mean, median, and variance of the indices were calculated for either one or two intervals. The temporally aggregated images using NDVI, or a combination of NDVI and NDMI/NDWI, were used to create the datasets (Table 2).

2.2.3. Sentinel-2

Sentinel-2 data with level 1C processing [30] provided by GEE were used. These data have been orthorectified and radio-corrected providing top-of-atmosphere reflectance values. Bands 2 to 8, 11, and 12 were used with original spatial resolutions of 10 or 20 metres. An automatic cloud masking procedure was applied using band QA60 of the S2 1C product, masking both opaque clouds and cirrus clouds.

Following the same methodology as the two previous sections, the mean, median, and variance of the S2 reflectance values were calculated for one, two, three, and four time intervals. Datasets using three and four time intervals were included because despite not being comparable to Landsat 8 datasets (with a maximum of two intervals), they may give some insights into the effects of adding additional time intervals onto the classification accuracies. These temporally aggregated images were then combined and resampled to 30 metres to create several "Sentinel-2" datasets (Table 2).

2.2.4. Sentinel-1

Radar data were analysed using the dual-polarised C-band data from the Synthetic Aperture Radar (SAR) instrument carried by the S1A and S1B satellites. The level-1 Ground Range Detected product (GRD, [31]) provided by GEE was used. The GRD images have been radiometrically calibrated and orthorectified, and the terrain correction has been applied using SRTM30 [32]. Two different polarisation modes were used: single co-polarisation with vertical transmit/receive (VV) and dual-band co-polarisation with vertical transmit and horizontal receive (VH). An extra pre-processing procedure, consisting of spatial filtering using a 7 × 7 Refined Lee speckle filter [33], was applied in order to eliminate the "speckle noise" characteristic of radar images and to make them functional for land cover detection at the spatial resolution used in this study. An extra band, VV-VH, was also created using the difference between the two polarisation modes. A three-band composite image was then created for each date, combining the three polarisation modes (VV, VH, and VV-VH), as this combination has been reported as optimal for land cover characterisation [34].

Radar data is not affected by clouds, so a considerable number of gap-free images can be obtained every month. However, radar data is affected by weather conditions (i.e., recent rainfall or wind) and with S1A and S1B can produce large data sets, so temporal aggregation may still be a valuable tool. For the Sentinel-1 datasets, the main focus was on analysing different numbers of time intervals for the temporal aggregation. Temporally aggregated images using the median values were obtained for each of the bands (VV, VH, and VV-VH), thus creating composites for one, two, four, six, and 12 intervals across the year. All datasets were resampled to 30 metres (Table 2).

2.2.5. Multi-Sensor

A fourth type of dataset was created by combining the best single-sensor type datasets. Three datasets were created by merging S2 with L8 data, S1 with S2 data, and S1 with S2 and L8 data. The selection of the best dataset for each of the single-sensor types was done *a posteriori*, once the classification accuracy had been estimated for the datasets in the previous sections (Table 2). Misalignments between satellite images from different satellites were examined visually by a comparison with easily identifiable ground locations and showed that spatial mismatches were always less than one 30 metre pixel.

2.2.6. Two-Date Composite

One final type of dataset was created using single-date images but without applying temporal aggregation techniques. Two relatively cloud-free S2 images were downloaded from the ESA Sentinel Hub (https://scihub.copernicus.eu/): one winter image (5 January 2017) and one summer image (17 June 2017). The images were obtained with a Level-1C pre-processing (orthorectified and radio-corrected). Only bands one to eight, 11, and 12 were used. An atmospheric correction was applied using the sen2cor algorithm [35] in the SNAP Toolbox. A terrain correction was then applied using a Minnaert algorithm with a slope correction [36]. In this case, cloud-masking was manually applied using visual interpretation. The two pre-processed images were combined into a two-date composite and resampled to 30 metres. This use of summer and winter images to create a two-date composite is the "traditional" method used by the UK Land Cover Maps [37,38]. It is used in this paper to provide a baseline accuracy against which to compare the other classifications and to determine whether the temporal aggregations match the accuracy of existing "traditional" methods.

A second "two-date composite" dataset was created using the same two images but without any extra pre-processing (Level-1C processing only). By doing so, the effects of the terrain correction and manual cloud-masking on classification could also be evaluated.

2.3. Land Cover Classification

Land cover classifications were carried out for each dataset by applying a supervised classification. Thirteen land cover classes were selected based on the UK Broad Habitat (BH; [39]) classification (Table 3). Training areas were based on polygons that had been identified with the same BH in the UK Land Cover Map 2000 (LCM2000; [37]) and 2007 (LCM2007; [38]), thus mimicking the classification methodology used for the UK LCM2015 (Rowland et al., in prep.). In this way, only stable areas or areas with a high probability of belonging to the assigned class can be selected. These areas were complemented with manually added polygons for rarer classes or classes for which the LCM2000 and LCM2007 were more inconsistent, such as coastal or semi-natural classes (e.g., fen). From the final set of training polygons, 10,000 points were randomly selected for each land cover class and were used to train a classification algorithm. A Random Forest (RF; [40]) classifier with 200 trees was trained and applied to each dataset to create land cover classifications. Because the validation of the classifications was performed by using an independent dataset, all training polygons were fed to the RF classifier, leaving out one-third of the training data for each bootstrap sample. The RF training and classification was implemented using the WEKA Data Mining Software [41].

Table 3. The definitions of land cover classes.

Land Cover Class	Description	Validation Data Classes *
Broadleaf woodland	Broadleaved tree species and mixed woodland	"Broadleaved, Mixed and Yew Woodland"
Coniferous woodland	Coniferous tree species where they exceed 80% of the total cover	"Coniferous Woodland"
Arable	Arable, horticultural and ploughed land; annual leys, rotational set-aside and fallow	"Arable and Horticulture"
Grassland	Managed grasslands and other semi-natural grasslands (grasses and herbs) on non acidic soils	"Improved Grassland" "Calcareous Grassland" "Neutral Grassland"
Acid grassland	Grasses and herbs on soils derived from acidic bedrock	"Acid Grassland"
Bog and fen	Wetlands with peat-forming vegetation such as bog, fen, fen meadows, rush pasture, swamp, flushes and springs	"Bog" "Fen, Marsh and Swamp"
Heather	Vegetation that has more than a 25% cover of species from the heath family	"Dwarf Shrub Heath"
Inland rock	Natural and artificial exposed rock surfaces	"Inland Rock"
Saltwater	Sea waters	"Saltwater"
Freshwater	Lakes, pools, rivers and man-made waters	"Freshwater"
Coastal	Beaches, sand dunes, ledges, pools and exposed rock in the maritime zone	"Supralittoral Rock" "Supralittoral Sediment" "Littoral Rock" "Littoral Sediment"
Saltmarsh	Vegetated portions of intertidal mudflats; species adapted to immersion by tides	"Littoral Sediment" **
Built-up areas	Urban and rural settlements	"Built-up Areas and Gardens"

* These classes are based on the UK Broad Habitats or Priority Habitats [39]. ** The Saltmarsh class is based on the "Saltmarsh" UK Priority Habitat, which is included in the "Littoral Sediment" Broad Habitat.

2.4. Accuracy Assessment

The Glastir Monitoring and Evaluation Programme (GMEP; [42]) dataset was used for an independent map validation. The GMEP is a field survey-based habitat and vegetation monitoring programme which provides high-resolution georeferenced data of the Welsh countryside. Polygon data for UK Broad Habitats, collected by the GMEP in 2013, 2014, 2015, and 2016 were used. Some BH

were merged to match the 13 classes used in this study. National Forest Inventory (NFI; [43]) data from 2016 were used to filter broadleaved and coniferous woodland patches, which had been harvested between the GMEP monitoring campaign and the acquisition of the satellite data. The GMEP data show a general lack of saltmarsh and inland rock polygons. In order to fill this gap between the classifications and the validation data, information on these two classes was extracted from the Natural Resources Wales (NRW) Terrestrial Phase 1 Habitat Survey (hereafter referred to as the Phase 1 Survey; [44]). The Phase 1 Survey is a semi-natural habitat map of Wales and is based on field surveys conducted over the course of several decades. Despite the potentially large time gap between the creation of Phase 1 Survey data and the dates of the satellite images used in this study, mismatches between the classifications and the validation data in terms of the presence of saltmarsh and inland rock polygons are unlikely to be due to changes in the land cover class, as these two classes are very stable over time. To avoid areas of change, the additional validation polygons were also manually reviewed against aerial photography.

The final set of validation polygons was buffered inwards by 30 metres to avoid selecting mixed pixels at the object boundaries. A total of ten thousand points were randomly selected from the validation set, and the differences between the validation data and the classifications were studied using confusion matrices. The overall accuracy (OA) estimation and the kappa coefficient [45] were used to compare the classification accuracy between datasets.

Finally, to understand the effects of the number of bands on the land cover predictions, the classification accuracies were plotted against the number of satellite data bands. The datasets were grouped by type, and a linear model with an adjusted R^2 was calculated.

3. Results

3.1. Cloud Cover

For temporal aggregations, sufficient observations are required per interval to ensure a good estimate. To evaluate this, the availability of cloud-free pixels during the year studied was estimated for L8 and S2. In general, the S2 data showed a very low number of pixels with zero or one cloud-free image(s) when using one or two time intervals, while L8 data showed more than 10% of pixels with zero or one cloud-free image(s) when using two intervals.

The temporal aggregation over one year led to high numbers of cloud-free pixels throughout the whole study area for the L8 and S2 satellites (Figure 3). Low values were observed for L8 for the summer and winter time intervals. Values lower than four are observed in coastal regions during summer and in central regions (corresponding to upland areas) during winter. Large areas with values of two or less were found for the winter interval in the centre of the map. No clear geographical patterns for low numbers of cloud-free images were found for S2. However, a triangular-shaped area in the southeast presented lower values for both the single-interval map and the summer/winter maps. This is because the study area is divided into two orbit paths of the S2 satellite, meaning that the date and number of image acquisitions for these two regions differ. In this case, the eastern orbit shows fewer cloud-free images, leading to the creation of this pattern.

All pixels had at least two cloud-free instances over the one-year interval for L8, with seven being the most common value (Figure 4a). However, almost 10% of pixels had just one cloud-free instance for the winter interval and about 10% of them had just two cloud-free values for summer. For S2, there were no pixels with less than seven cloud-free values over the one-year interval, with 26 being the most common value (Figure 4b). The winter interval did not present any pixels with less than four cloud-free values, while less than 5% of pixels in the summer interval had less than three. A bimodal shape was observed for the S2 histograms, which was due to the differences between the regions falling into the different S2 orbit trajectories [7].
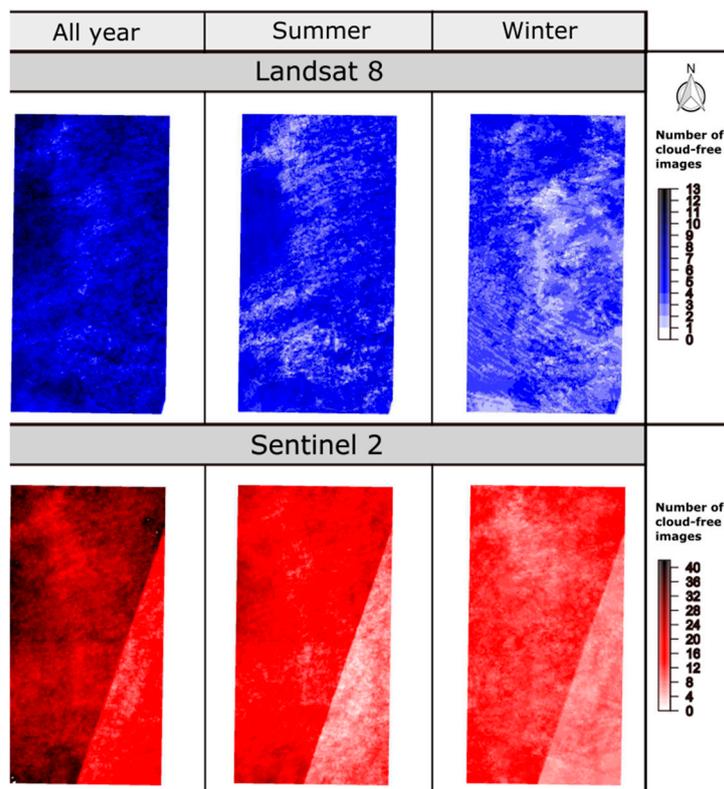
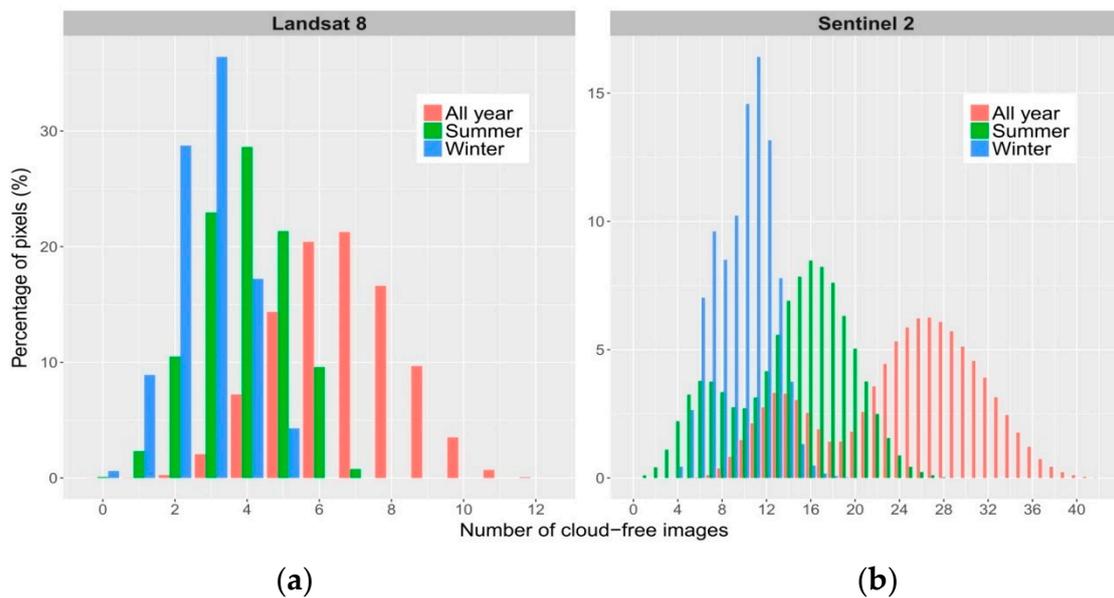**Figure 3.** Maps of the number of cloud-free images per pixel.

Figure 4. Histograms of the number of cloud-free images available per pixel for three different time intervals for (**a**) Landsat 8 and (**b**) Sentinel-2.

3.2. Classification Accuracy

The classification accuracy of each dataset scenario was assessed using confusion matrices and their derived accuracy indices. The confusion matrices can be found in Supplementary Material S2. The two traditional-style two-date composites, which were not temporally aggregated, obtained the highest classification accuracy, followed by the combined sensors datasets. The kappa index was very

closely related to the overall accuracy (OA), meaning that datasets with a higher OA also obtained higher kappa values. A summary of the accuracy results can be found in Figure 5.
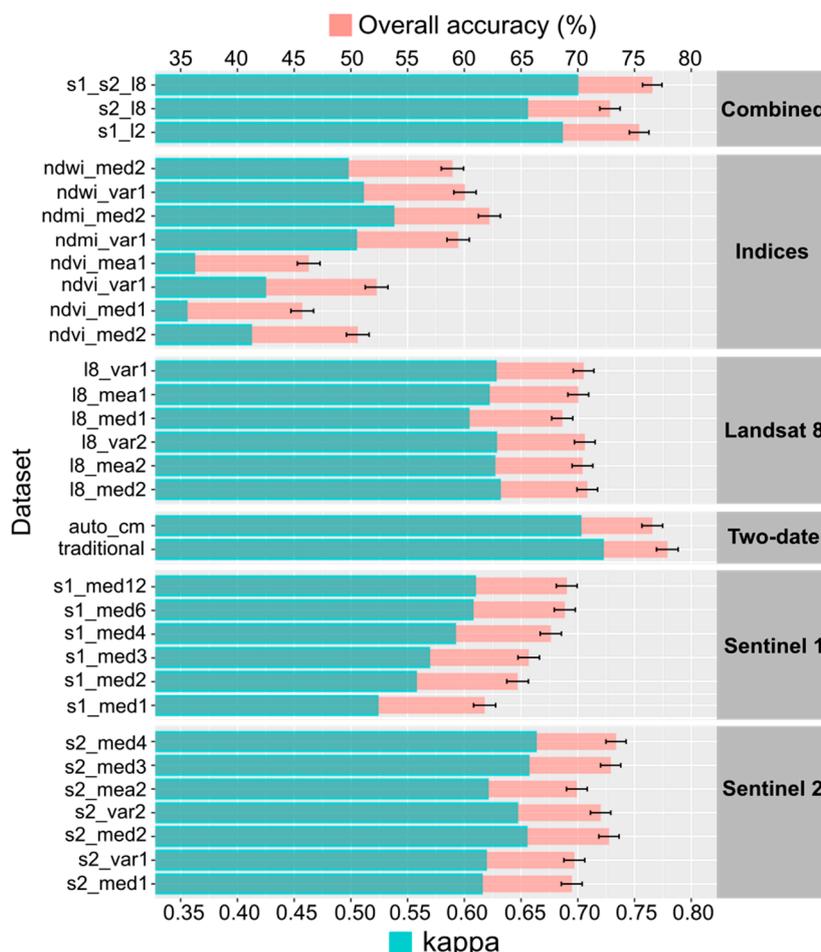
Figure 5. The classification accuracies: the error bars represent a 95 percent confidence interval for the overall accuracy, calculated using an exact binomial test.

All of the "Landsat 8" datasets obtained very similar levels of accuracy, with the OA ranging from 68.6% to 70.8%. The most accurate "Landsat 8" dataset used the median and variance values taken over two intervals. The datasets based on just the L8-derived indices obtained the lowest levels of accuracy. The temporal aggregation over the mean or median values of NDVI obtained very similar levels of accuracy, with the datasets using two intervals or a combination of median and variance over one interval being the most accurate. The combination of NDVI with NDMI or NDWI obtained higher levels of accuracy than the temporal aggregations using only NDVI, with the datasets using median values over two time intervals being the most accurate (OA = 62.2%).

The OA of the "Sentinel-2" scenarios ranged between 69.4% and 73.3%. The most accurate classification used median values calculated over three (OA = 72.9%) or four (OA = 73.3%) time intervals. These were followed by the dataset created using median values over two intervals (OA = 72.7%). Figure 6a shows this dataset and its classification for a region that presented cloudy pixels on several image dates. This dataset was slightly more accurate than the equivalent dataset created using the L8 data (OA = 70.8%). The OA of the "Sentinel-1" datasets increased with the number of time intervals. The scenario that used mean values over 12 time intervals obtained the maximum level of accuracy, with an OA of 69.0%, which was slightly below the most accurate scenarios of the "Landsat 8" or "Sentinel-2" types.
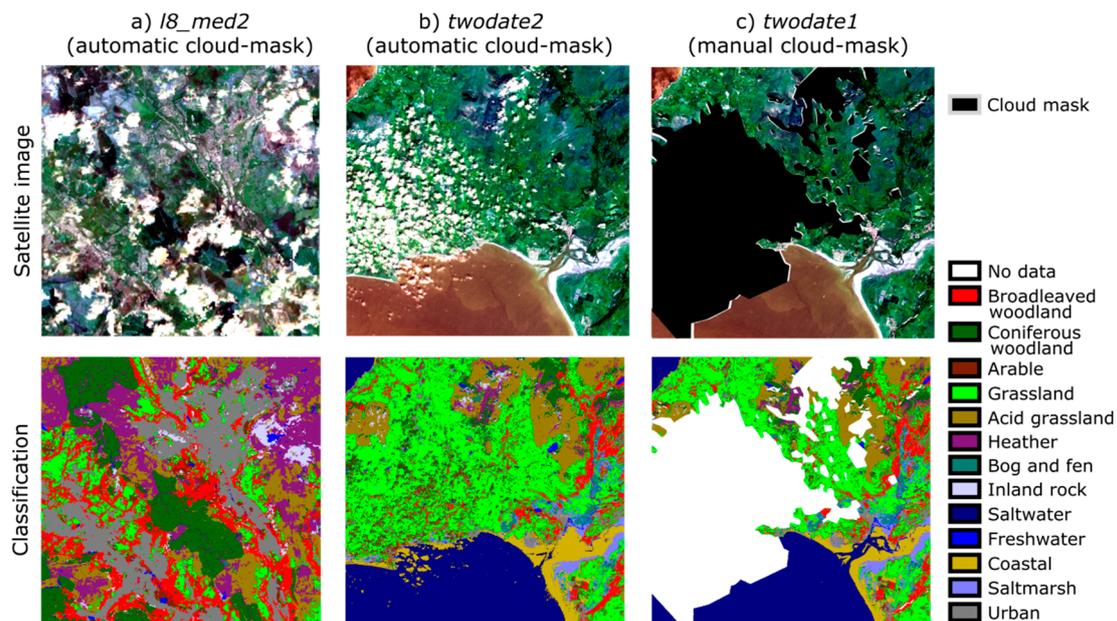
Figure 6. Details of the detected/undetected clouds for three datasets, (**a**) *l8_med2*, (**b**) *twodate2* and (**c**) *twodate1* and their classification results. The satellite images are displayed using the RGB bands for the summer interval.

The classification accuracy of individual land cover classes was compared for the most accurate Sentinel-1, Sentinel-2, and Landsat 8 classifications (Figure 7). Differences in the accuracy between Sentinel-2 and Landsat 8 were less than 7% for all classes, except for inland rock. Arable, bog and fen, inland rock, and sea water were slightly more accurately classified with the Landsat 8 dataset. Sentinel-2 datasets obtained better accuracy than Sentinel-1 for all classes except for arable and fen and bog. The higher frequency of the Sentinel-1 data may help to capture the phenology of arable lands, while its capability to detect soil moisture may help with the detection of fen and bogs. Inland rock was validated using only 12 points, due to its scarcity in the landscape, so slight changes in classifications produce a high variability in accuracy between datasets for this class. Saltwater's accurate detection and its confusion with coastal classes is strongly dependent on the tidal state at the time of image acquisition. However, inland rock and saltwater have relatively few validation points, so their accuracy will not have a significant impact on the overall classification accuracies.

The "Combined" datasets obtained better accuracy results than any of the single-sensor datasets using temporally aggregated data. The dataset that combined S1, S2, and L8 data obtained the highest accuracy with a OA of 76.5%, followed by the scenario combining S1 and S2 data (OA = 75.4%). The scenarios combining the two optical sensors, i.e., L8 and S2, obtained an appreciably lower level of accuracy (OA = 72.8%) than the other two combined scenarios. The land cover classification map using the S1/S2/L8 combined data—the most accurate temporally-aggregated scenario—can be seen in Figure 8. Figure 9 shows this classification, together with the classification of the least accurate scenario, as well as its discrepancies with the validation data.

Finally, the two traditional "two-date composite" datasets, which were based on non-temporally aggregated data, obtained the highest accuracy. The area of these classifications were, however, smaller than the rest of the classifications, as one of the satellite images covered only about 80% of the study area. The dataset that used manual cloud-masking obtained the highest accuracy, with an OA of 77.9%, although the dataset that used automatic cloud-masking was almost equal at 76.5%. Despite the fact that both datasets obtained similar classification accuracies, the automatic cloud-masking failed to identify large cloud patches in some regions (Figure 6b,c). This affected the classification of these cloudy areas but, due to the quantity and distribution of the validation areas, were not enough to substantially decrease the estimated overall classification accuracy.
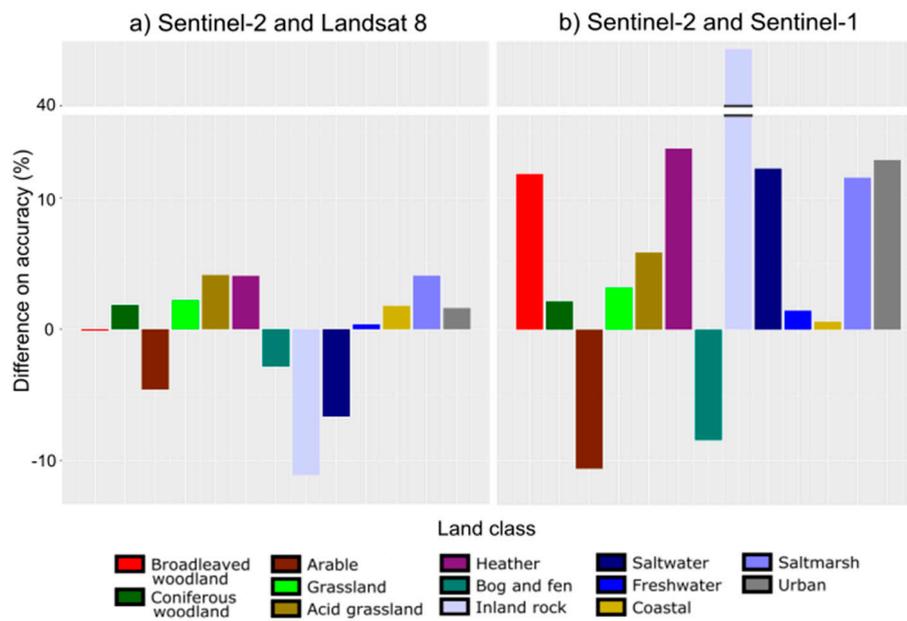
Figure 7. The difference in classification accuracy for each land cover class between Landsat 8, Sentinel-2, and Sentinel-1 datasets: The most accurate datasets for each satellite—*l8_med2*, *s2_med4*, and *s1_med12*—were used for the analysis. The accuracy was calculated by averaging the user and producer's accuracy for each land cover class. (**a**) The difference in accuracy between the Sentinel-2 and Landsat 8 datasets: the negative values represent a higher accuracy for the Landsat 8 dataset. (**b** The difference in accuracy between the Sentinel-2 and Sentinel-1 datasets: the negative values represent a higher accuracy for the Sentinel-1 dataset.
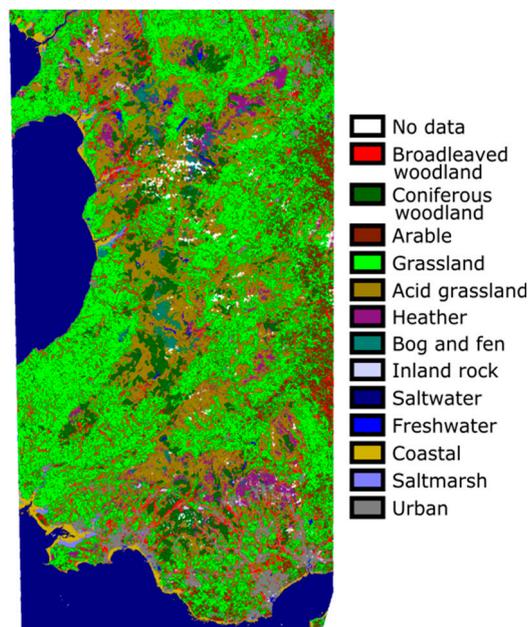
Figure 8. The land cover classification of the *s1_s2_l8* dataset. This classification showed the highest accuracy of all the aggregated datasets.
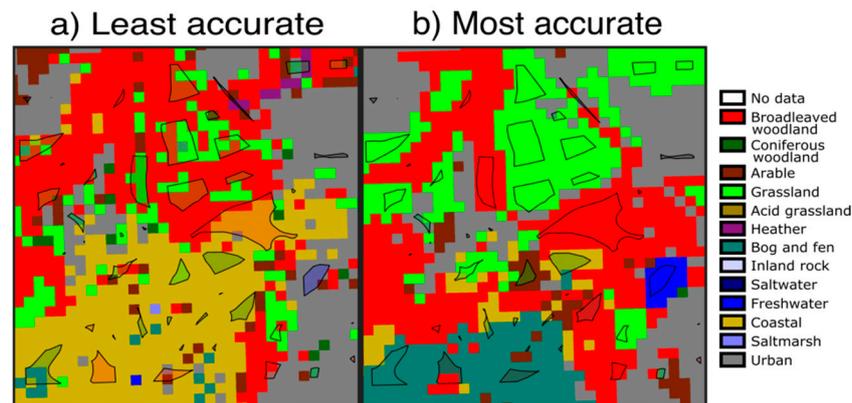
Figure 9. Details of the classifications for the least (*ndvi_mea1*) and most (*s1_l8_s2*) accurate temporally aggregated datasets: the polygons from the validation dataset are plotted on top of the classifications. The shaded portions of the polygons represent the inconsistencies between the classification and validation data.

The effect on the overall accuracy of increasing the number of bands varied between the different dataset types (Figure 10). The "Combined" and "Indices" datasets showed very high correlations between the number of bands and the OA. The increase in OA for the "Indices" datasets was very acute. However, the number of bands only ranged between one and four for this dataset type, while the range for the rest of the datasets was at least 20 bands. The correlation with the "Sentinel-1" type dataset was moderate, while correlations with the "Sentinel-2" and "Landsat 8" types were weak.
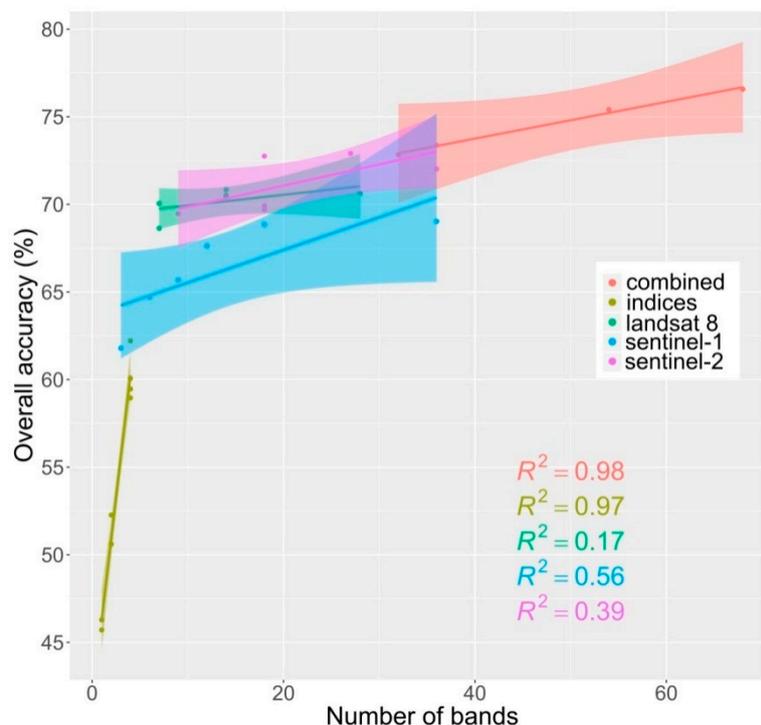

Figure 10. The relationships between classification accuracy and the number of bands.

4. Discussion

One of the challenges in producing land cover maps of large areas is the inconsistency in the number of cloud-free pixels available over a period of time. In this study, the differences in the number of cloud-free images per pixel over a year were analysed to evaluate the effects on the temporal aggregation of L8 and S2 data. For the data assessed in this study, to ensure at least one cloud-free

pixel per time interval, a maximum of two intervals per year were possible with L8, while more than two intervals could be used for Sentinel-2. This is due to Sentinel-2's higher revisit frequency during 2016 (10 days against 16 days for L8). With the recent launch of Sentinel-2B, the revisit frequency of both Sentinel-2A and 2B is expected to increase to an average of five days, with that number becoming larger for higher latitudes [7]. Such high revisit frequencies could enable the creation of temporal aggregations over at least five or six intervals per year. This will allow researchers to better characterise the seasonality of certain land types and to use time-series analysis approaches [23]. Differences in cloud-free coverage are also affected by the differences in satellite coverage, an issue that should be considered for large areas. For the study area, Sentinel-2 data showed a different level of cloud-free coverage for a large triangular region in the southeast, which corresponded to a different orbit path. This could result in inconsistencies in accuracy [12], and it should be acknowledged when deciding the length of the interval to be used for temporal aggregation or when evaluating the appropriateness of gap-filling methodologies.

The highest OA for a temporally aggregated dataset was 76.5% (the dataset combining L8, S1, and S2 data). Issues that may have the potential to decrease classification accuracy were detected with the training data; they include the overestimation of certain widespread classes (i.e., acid grassland) and the scarcity of training polygons for common upland classes (i.e., bog). The quality of the training data is key to obtaining satisfactory classification results [46]. A large area of this image was covered with uplands that are a complicated mix of habitats occurring in a complex mix of mosaics and gradual transitions between different habitats. These habitats can be difficult for surveyors in the field to map reliably and do not really conform to the idea of discrete patches of land cover implicit in mapping the dominant land cover in the pixel [22].

Some studies that have applied innovative input-data approaches for land cover mapping have obtained higher classification accuracies [47,48]. In general, most of those studies classified fewer classes than our study or they used classes that have traditionally been easier to characterise from remotely sensed data (i.e., forests or water). Inglada et al. [12], on the other hand, used interpolation methods to create a land cover classification of France, characterising a wide variety of land cover classes and obtaining kappa values ranging between 0.82 and 0.86. The main difficulty of comparing the accuracies of these studies with those obtained in our study is that their validation data was mainly based on the visual interpretation of remotely sensed data (satellite or aerial photography). Gebhardt et al. [17] applied temporal aggregation methodologies to create a land cover map of Mexico and obtained overall accuracies of up to 76% using ground-based validation data. Although the use of stratified-sampled ground reference data for validation, as we do here, is not exempt from error, it is considered a more accurate representation of the land surface and it can avoid the biases inherent in validating against remotely sensed data in terms of spatial or thematic resolution [49]. In any case, the aim of this study was to compare the accuracies of different input datasets. To do this, the training samples and validation samples were kept constant for every dataset. Using stratified-sampled ground data for validation was key for our purposes, as reliable data for rare land cover classes such as different types of natural grasslands can be difficult to obtain accurately from remote sensed datasets.

The classification accuracies varied from 46% to 76.5%. The "Indices" datasets, which were based on NDVI/NDMI/NDWI estimates, obtained the lowest levels of accuracy. The dataset that combined NDVI and NDMI over two time intervals was the most accurate; however, its accuracy was much lower than for the datasets using reflectance measurements. Although spectral indices have been shown to be a simple and effective way of analysing time series of vegetation data [50] and may improve the characterisation of the land cover when combined with other reflectance measures [51], our indices alone did not obtain good results in terms of predicting a wide variety of land cover classes. This is probably due to the fact that these indices considerably reduce the amount of spectral information used to characterize land cover and that indices are mostly useful when combined with reflectance bands. The "Sentinel-1" datasets obtained, on average, worse accuracies than the other two optical sensors. However, the dataset that used 12 time intervals obtained almost the same level of

accuracy as the most accurate L8-based classification. Previous studies have shown the potential of S1 to characterise certain land types, such as crops [52] or forests [53]. S1 has also shown good results for the land cover mapping of heterogeneous landscapes [34]. However, this study showed how S1's capabilities for characterising land cover are principally based on its image frequency, as it is able to use a considerable number of dates per year whilst maintaining a relatively low number of bands. The temporal aggregation of S1 data might provide a good strategy for including frequent radar data in classifications but without needing to include every image that can be acquired in a year.

The "Sentinel-2" datasets obtained higher accuracy than other single-sensor datasets with a maximum OA of 73%, which was significantly higher than the most accurate "Landsat 8" dataset. However, this S2-based dataset used four time intervals, while the L8 datasets used a maximum of two intervals due to the lower numbers of available cloud-free pixels. The accuracy of the datasets using two intervals did not differ between S2 and L8. It has to be noted that some authors have pointed out the poor performance of the cloud mask used for S2 [54]. Additionally, the S2 data currently available via GEE is not atmospherically corrected, whereas the L8 data is. These two limitations might have affected the potential of S2 and its additional bands to outperform L8. We used the L8 and S2 data sets from GEE despite their differences in processing because one of our aims was to determine to what extent pre-processing could be simplified by adopting workflows in GEE.

The "Combined" datasets, which used multi-sensor derived data, obtained the most accurate classifications of all the temporally aggregated datasets. The combination of L8 and S2 data produced lower levels of accuracy than those that combined optical and radar data (L8/S2 or L8/S2/S1). Known co-registration problems between Landsat 8 and Sentinel-2 [55,56] could have lowered the accuracy of the combined datasets. However, our methodology, avoiding boundaries between land cover types for classification training and validation pixels, should reduce the effects of spatial misalignment between images. The launch of the Harmonized Landsat and Sentinel Datasets [57] should minimize this issue in the near future. Our results support findings from previous studies, which have suggested the potential of combining optical and radar data to characterise and detect changes in land cover [58–60].

The effect of the number of selected time intervals on classification accuracy was studied. In general, a higher number of intervals resulted in a higher level of classification accuracy, and two-interval datasets outperformed one-interval datasets that included variance measurements. Increasing the number of intervals has a similar effect to increasing the number of single-date images per year and helps to characterise the land cover, as certain classes have distinctive phenological signatures that can be detected by the classification algorithm [61]. The accuracy obtained by the "Sentinel-1" datasets increased linearly according to the number of time intervals. The temporal aggregation of S1 data over a sufficient number of time intervals can help to improve the accuracy of land cover classifications, especially when combined with optical data, while avoiding the necessity of adding massive amounts of data.

A linear correlation between the number of bands and the classification accuracy was found for some dataset types. Yu et al. [62] showed in their review that accuracy increased according to the number of bands, especially if the datasets combined data from different sensors. This was also true for this study, except in the case of the "Sentinel-2" and "Landsat 8" datasets, probably because some of these datasets included variance data, which increased the number of bands without significantly increasing the accuracy. Whether additional bands increase accuracy will depend on whether they contain additional information, for example, by capturing a different point in time. The growth in the availability of remote sensed data and processing platforms over the last few years is facilitating the usage of larger volumes of data [5] and is, therefore, providing the potential for improving the accuracy of land cover classifications. However, the selection of optimal input data is still important, as the accuracy/data size trade-off has not disappeared. For example, GEE proved to be a great tool for selecting, cloud-masking, stacking, and extracting big satellite datasets. However, GEE memory limits on the size of arrays prevented the training of the RF algorithm with the large number of training

samples and input bands assessed here. To counter this we created the input data stacks in GEE and then exported them and processed them locally using established RF classification work flows. Processing in GEE is also currently affected by the very different levels of pre-processing applied to the data collections by the agencies providing the data. So the L8 data was atmospherically corrected and had an established cloud-mask, whereas the S2 data was not atmospherically corrected and had a cloud-mask that still has significant issues [54]. This complicates analyses between the capabilities of different sensors and will make GEE unsuitable for some types of processing.

The highest levels of accuracy were obtained by datasets that used traditional non-temporally aggregated two-date composites. The accuracy of the two-date composite with automatic cloud-masking was not significantly lower than the manually cloud-masked one. The images for these datasets were selected based on the low amount of cloud cover, and thus, the effects of poor cloud-masking for the automatically cloud-masked composite did not significantly affect the accuracy, as the quality of the classification was only affected in a relatively small area. Only the most accurate of the temporally aggregated datasets, which combined optical and radar data, matched the accuracy of the two-date composites. This highlights the quality of the methods that have underpinned the UK's Land Cover Map series to date. Some manual pre-processing tasks, such as manual cloud-masking or the manual selection of cloud-free areas for gap-filling, can require large amounts of labour [63]. Adopting the temporal aggregation methods and automatic cloud-masking could help to reduce the time needed to produce land cover maps of large areas, especially for regions for which the acquisition of cloud-free images is particularly difficult. However, increased accuracy should not be presupposed if these methods are used. For example, Senf et al. [10] found that data fusion matched the accuracy of manually processed Landsat images, but a lower level of accuracy was obtained when there were not enough cloud-free images available per year.

Approaches that use all available satellite data are also capable of dealing with the gaps and capable of reducing cloud-masking inaccuracies and have produced promising results for land cover mapping and the land cover change detection of large areas [20,64]. Non-aggregated datasets can be, however, very difficult to manage, and most classification algorithms cannot handle the combination of a large volume of training data, a large number of classes, and a high number of input bands, even when using cloud-computing platforms. On the other hand, temporal aggregation maintains reasonably small-sized data to feed the classification algorithms, while having the potential to reduce the processing time needed for land cover mapping of large areas. However, an optimal choice of satellite data and aggregation parameters are crucial to maintain the accuracy levels of the more traditional, manually intensive approaches.

5. Conclusions

The recent availability of frequent satellite data and cloud computing platforms are stimulating the emergence of new approaches for the land cover mapping of large areas. This paper has analysed one of these approaches, i.e., the temporal aggregation of automatically pre-processed satellite data, comparing it with traditional methods and studying the classification accuracy of different temporally aggregated datasets. Temporal aggregation of all the available images, over the course of one year, only matched the manually selected and processed two-date composite when an optimal combination of optical and radar data was used. Combined datasets (S1/S2 or S1/S2/L8) outperformed single-sensor datasets, while datasets based on spectral indices obtained the lowest levels of accuracy. This study provides, to the best of our knowledge, the first formal comparison of the accuracy between multi-sensor data and single-sensor data using S1, S2, and L8 data for land cover mapping, as well as providing a wide range of combinations and parameters for the classification input data.

Cloud computing platforms such as GEE allow researchers to compensate for the lower quality of automatic pre-processing methods by using larger amounts of satellite data. However, identifying the optimal input datasets, in terms of the best combination of satellite data, temporal interval, and other data aggregation parameters, are needed in order to match the accuracy of manually selected and

processed image composites. While optimal satellite datasets might differ slightly for other regions and land classes, our findings offer a framework for the comparison and selection of suitable temporally aggregated data. Further work will be needed to compare temporal aggregation with other gap-filling techniques, such as data interpolation or data fusion.

References

1. Lambin, E.F.; Geist, H.J.; Lepers, E. Dynamics of land-use and land-cover change in tropical regions. *Annu. Rev. Environ. Resour.* **2003**, *28*, 205–241. [CrossRef]
2. Global Climate Observing System. Essential Climate Variables. 2010. Available online: https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables (accessed on 1 August 2018).
3. Pettorelli, N.; Wegmann, M.; Skidmore, A.; Mucher, S.; Dawson, T.P.; Fernandez, M.; Lucas, R.; Schaepman, M.E.; Wang, T.; O'Connor, B.; et al. Framing the concept of satellite remote sensing essential biodiversity variables: Challenges and future directions. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 122–131. [CrossRef]
4. Lambin, E.F.; Geist, H.J. *Land-Use and Land-Cover Change: Local Processes and Global Impacts*; Springer Science & Business Media: Berlin, Germany, 2008.
5. Wulder, M.A.; Coops, N.C.; Roy, D.P.; White, J.C.; Hermosilla, T. Land cover 2.0. *Int. J. Remote Sens.* **2018**, *39*, 4254–4284. [CrossRef]
6. Kovalskyy, V.; Roy, D. The global availability of Landsat 5 tm and Landsat 7 etm+ land surface observations and implications for global 30 m Landsat data product generation. *Remote Sens. Environ.* **2013**, *130*, 280–293. [CrossRef]
7. Li, J.; Roy, D.P. A global analysis of Sentinel-2a, Sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* **2017**, *9*, 902.
8. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
9. Gevaert, C.M.; García-Haro, F.J. A comparison of STARFM and an unmixing-based algorithm for Landsat and Modis data fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [CrossRef]
10. Senf, C.; Leitao, P.J.; Pflugmacher, D.; van der Linden, S.; Hostert, P. Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sens. Environ.* **2015**, *156*, 527–536. [CrossRef]
11. Zhang, Y.; Atkinson, P.M.; Li, X.; Ling, F.; Wang, Q.; Du, Y. Learning-based spatial–temporal superresolution mapping of forest cover with MODIS images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 600–614. [CrossRef]
12. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sens.* **2017**, *9*, 95. [CrossRef]
13. Griffiths, P.; van der Linden, S.; Kuemmerle, T.; Hostert, P. A pixel-based Landsat compositing algorithm for large area land cover mapping. *IEEE J.-STARS* **2013**, *6*, 2088–2101. [CrossRef]

14. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W. Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year Landsat time series. *Can. J. Remote Sens.* **2018**, *44*, 67–87. [CrossRef]

15. DeFries, R.; Hansen, M.; Townshend, J. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* **1995**, *54*, 209–222. [CrossRef]

16. Loveland, T.R.; Reed, B.C.; Brown, J.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [CrossRef]

17. Gebhardt, S.; Wehrmann, T.; Ruiz, M.A.M.; Maeda, P.; Bishop, J.; Schramm, M.; Kopeinig, R.; Cartus, O.; Kellndorfer, J.; Ressl, R.; et al. Mad-mex: Automatic wall-to-wall land cover monitoring for the mexicanredd-mrv program using all Landsat data. *Remote Sens.* **2014**, *6*, 3923–3943. [CrossRef]

18. Winsvold, S.H.; Kaab, A.; Nuth, C. Regional glacier mapping using optical satellite data time series. *IEEE J.-STARS* **2016**, *9*, 3698–3711. [CrossRef]

19. Verhegghen, A.; Eva, H.; Ceccherini, G.; Achard, F.; Gond, V.; Gourlet-Fleury, S.; Cerutti, P.O. The potential of sentinel satellites for burnt area mapping and monitoring in the Congo basin forests. *Remote Sens.* **2016**, *8*, 986. [CrossRef]

20. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171. [CrossRef]

21. Armitage, R.P.; Ramirez, F.A.; Danson, F.M.; Ogunbadewa, E.Y. Probability of cloud-free observation conditions across Great Britain estimated using MODIS cloud mask. *Remote Sens. Lett.* **2013**, *4*, 427–435. [CrossRef]

22. Blackstock, T.; Burrows, C.; Howe, E.; Stevens, D.; Stevens, J. Habitat inventory at a regional scale: A comparison of estimates of terrestrial broad habitat cover from stratified sample field survey and full census field survey for wales, UK. *J. Environ. Manag.* **2007**, *85*, 224–231. [CrossRef]

23. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm.* **2016**, *116*, 55–72. [CrossRef]

24. Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [CrossRef]

25. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [CrossRef]

26. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]

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

28. Wilson, E.H.; Sader, S.A. Detection of forest harvest type using multiple dates of Landsat tm imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [CrossRef]

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

30. Gatti, A.; Bertolini, A. Sentinel-2 Products Specification Document, 2015. Available online: https://sentinel.esa.int/documents/247904/349490/S2_MSI_Product_Specification.pdf (accessed on 3 August 2018).

31. Bourbigot, M.; Piantanida, R. *Sentinel-1 User Handbook*; European Space Agency (ESA): Paris, France, 2016.

32. Farr, T.G.; Kobrick, M. Shuttle Radar Topography Mission produced a wealth of data. *EOS Trans. Am. Geophys. Union* **2000**, *81*, 583–585. [CrossRef]

33. Yommy, A.S.; Liu, R.; Wu, S. SAR image despeckling using refined lee filter. In Proceedings of the 2015 7th International Conference on Intelligent HumanMachine Systems and Cybernetics (IHMSC), Hangzhou, China, 26–27 August 2015; Volume 2, pp. 260–265.

34. Abdikan, S.; Sanli, F.; Ustuner, M.; Caló, F. Land cover mapping using sentinel-1 SAR data. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B7*, 757–761. [CrossRef]

35. 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. 1–8.

36. Gao, M.; Gong, H.; Zhao, W.; Chen, B.; Chen, Z.; Shi, M. An improved topographic correction model based on minnaert. *GISci. Remote Sens.* **2016**, *53*, 247–264. [CrossRef]

37. Fuller, R.; Smith, G.; Sanderson, J.; Hill, R.; Thomson, A.; Cox, R.; Brown, N.; Clarke, R.; Rothery, P.; Gerard, F. *Countryside Survey 2000 Module 7. Land Cover Map 2000*; Final Report; Centre for Ecology & Hydrology: Lancaster, UK, 2002.

38. Morton, D.; Rowland, C.; Wood, C.; Meek, L.; Marston, C.; Smith, G.; Wadsworth, R.; Simpson, I. *Final Report for LCM2007: The New UK Land Cover Map*; Centre for Ecology & Hydrology: Lancaster, UK, 2011.

39. Jackson, D. *Guidance on the Interpretation of the Biodiversity Broad Habitat Classification (Terrestrial and Freshwater Types): Definitions and the Relationship with Other Habitat Classifications*; Joint Nature Conservation Committee: London, UK, 2000.

40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

41. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The Weka data mining software: An update. *SIGKDD Explor.* **2009**, *11*, 10–18. [CrossRef]

42. Emmett, B.; Abdalla, M.; Anthony, S.; Astbury, S.; August, T.; Barrett, G.; Beckmann, B.; Biggs, J.; Botham, M.; Bradley, D.; et al. *Glastir Monitoring & Evaluation Programme. Second Year Annual Report*; Centre for Ecology & Hydrology: Lancaster, UK, 2015.

43. UK Forestry Commission. National Forest Inventory Woodland England 2015. 2016. Available online: https://data.gov.uk/dataset/ae33371a-e4da-4178-a1df-350ccfcc6cee/national-forest-inventory-woodland-england-2015 (accessed on 25 October 2018).

44. Blackstock, T.; Stevens, J.; Howe, L.; Jones, P. *Habitats of Wales: A Comprehensive Field Survey*; University of Wales Press: Cardiff, UK, 2010; pp. 1979–1997.

45. Wan, T.; Jun, H.; Hui Zhang, P.W.; Hua, H. Kappa coefficient: A popular measure of rater agreement. *Shanghai Arch. Psychiatry* **2015**, *27*, 62.

46. 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]

47. Chen, B.; Huang, B.; Xu, B. Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS J. Photogramm.* **2017**, *124*, 27–39. [CrossRef]

48. Qadri, S.; Khan, D.M.; Qadri, S.F.; Razzaq, A.; Ahmad, N.; Jamil, M.; Nawaz Shah, A.; Shah Muhammad, S.; Saleem, K.; Awan, S.A. Multisource data fusion framework for land use/land cover classification using machine vision. *J. Sens.* **2017**, *2017*, 3515418. [CrossRef]

49. Foody, G.M. The impact of imperfect ground reference data on the accuracy of land cover change estimation. *Int. J. Remote Sens.* **2009**, *30*, 3275–3281. [CrossRef]

50. Fan, X.; Liu, Y. A global study of NDVI difference among moderate-resolution satellite sensors. *ISPRS J. Photogramm.* **2016**, *121*, 177–191. [CrossRef]

51. Feng, D.; Yu, L.; Zhao, Y.; Cheng, Y.; Xu, Y.; Li, C.; Gong, P. A multiple dataset approach for 30-m resolution land cover mapping: A case study of continental Africa. *Int. J. Remote Sens.* **2018**, *39*, 3926–3938. [CrossRef]

52. Bargiel, D. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sens. Environ.* **2017**, *198*, 369–383. [CrossRef]

53. Haarpaintner, J.; Davids, C.; Storvold, R.; Johansen, K.; Arnason, K.; Rauste, Y.; Mutanen, T. Boreal forest land cover mapping in Iceland and Finland using Sentinel-1A. In Proceedings of the Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; Volume 740, p. 197.

54. Coluzzi, R.; Imbrenda, V.; Lanfredi, M.; Simoniello, T. A first assessment of the Sentinel 2 Level 1-C cloud mask product to support informed surface analyses. *Remote Sens. Environ.* **2018**, *217*, 426–443. [CrossRef]

55. Storey, J.; Roy, D.P.; Masek, J.; Gascon, F.; Dwyer, J.; Choate, M. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sens. Environ.* **2016**, *186*, 121–122. [CrossRef]

56. Yan, L.; Roy, D.P.; Zhang, H.K.; Li, J.; Huang, H. An automated approach for sub-pixel registration of Landsat-8 of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sens.* **2016**, *8*, 520. [CrossRef]

57. Claverie, M.; Ju, J.; Masek, J.; Dungan, J.L.; Vermonte, E.F.; Roger, J.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]

58. Solberg, A.H.S.; Jain, A.K.; Taxt, T. Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 768–778. [CrossRef]

59. Haack, B.; Bechdol, M. Integrating multisensor data and radar texture measures for land cover mapping. *Comput. Geosci.* **2000**, *26*, 411–421. [CrossRef]

60. Dusseux, P.; Corpetti, T.; Hubert-Moy, L.; Corgne, S. Combined use of multi-temporal optical and radar satellite images for grassland monitoring. *Remote Sens.* **2014**, *6*, 6163–6182. [CrossRef]

61. Tatsumi, K.; Yamashiki, Y.; Torres, M.A.C.; Taipe, C.L.R. Crop classification of upland fields using random forest of time-series Landsat 7 etm+ data. *Comput. Electron. Agric.* **2015**, *115*, 171–179. [CrossRef]

62. Yu, L.; Liang, L.; Wang, J.; Zhao, Y.; Cheng, Q.; Hu, L.; Liu, S.; Yu, L.; Wang, X.; Zhu, P.; et al. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int. J. Remote Sens.* **2014**, *35*, 4573–4588. [CrossRef]

63. Gong, P.; Yu, L.; Li, C.; Wang, J.; Liang, L.; Li, X.; Ji, L.; Bai, Y.; Cheng, Y.; Zhu, Z. A new research paradigm for global land cover mapping. *Ann. GIS* **2016**, *22*, 87–102. [CrossRef]

64. Hansen, M.C.; Potapov, R.; Moore, R.; Hancher, M.; Turubanova, S.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-resolution global maps of 21st-century forest cover change. *Science* **2013**, *6160*, 850–853. [CrossRef] [PubMed]