Agricultural Systems 171 (2019) 36–50

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: Case study of the Duero river basin

Laura Piedelobo[a], David Hernández-López[b], Rocío Ballesteros[a], Amal Chakhar[b], Susana Del Pozo[a], Diego González-Aguilera[a], Miguel A. Moreno[c,*]

[a] Department of Cartographic and Land Engineering, University of Salamanca, Hornos Caleros 50, 05003 Ávila, Spain
[b] Institute for Regional Development (IDR), University of Castilla-La Mancha, Campus Universitario s/n, 02071 Albacete, Spain
[c] Regional Centre of Water Research (CREA), University of Castilla-La Mancha, Carretera de las Peñas km 3,200, 02071 Albacete, Spain

ARTICLE INFO

Keywords:
Crop classification
Remote sensing
Sentinel-2 and Landsat-8
Interoperability
Precision agriculture
Machine learning

Abstract: Satellite imagery is the foremost source of information to analyze and monitor land covers in several time ranges, especially over large areas. However, it is not always either freely available or easily compatible for the final users due to the different resolutions offered by sensors onboard the satellite platforms. Crop classification is an important task to control and make decisions related to the agricultural practice and its regulation. However, it is not trivial, especially for extensive areas. Thus, this paper proposes a new approach for crop classification in large areas by a combined use of multi-temporal open-source remote sensing data from Sentinel-2 (S2) and Landsat-8 (L8) satellite platforms. Having to deal with different spatial and temporal resolutions, special spatial regions (called Tuplekeys) were created within a local nested grid to allow a proper integration between the data of both sensors. Temporal variation of the Normalized Difference Vegetation Index (NDVI) was the chosen input to classify crops. Moreover, due to the massive quantity of data collected, filters considering some agronomic and edaphic criteria were applied with the dual goal of decreasing redundancies and increasing the process efficiency. Out of three different machine learning classifiers analyzed, a plot-based approach was considered for the algorithms calibration while a pixel-based approach was used for the final classification process. The methodology was both tested and validated in the Duero river basin (Spain), 78,859 km^2, for the 2017 spring and summer seasons. Finally, classification outputs were analyzed throughout their overall accuracy (OA), not only for the whole basin but also for each of the Tuplekeys so that the OA spatial distribution was evaluated as well. The Ensemble Bagged Trees (EBT) algorithm showed the maximum OA, 87% and 92%, when classifying crops individually (15 classes) and grouped (7 classes), respectively, proving both the accuracy and efficiency of the developed approach.

1. Introduction

The current increase in the demand for agricultural products due to the rising worldwide population requires the proper management and planning of the available resources. A well-grounded and detailed knowledge of agricultural areas and crop types is essential to perform economical, ecological and sustainable strategies (Cai et al., 2018; Davis et al., 2016; Zaks and Kucharik, 2011). Given the dynamic character of agriculture, crop monitoring turned out to be crucial to control both its evolution and trend. Among other parameters, cultivated surface, crops production and distribution and their water demand can be accurately estimated using alternative technologies (Burke and Lobell, 2017; Cai et al., 2018; Durgun et al., 2016; Murmu and Biswas, 2015).

Satellite remote sensing (RS) has been proven as a very valuable technique to provide large scale multi-temporal data. Thus, both land cover and natural resources maps are currently performed using this technique (Hasmadi et al., 2017; Piedelobo et al., 2018; Shelestov et al., 2017). However, for crop monitoring, proper spatial and temporal resolutions are required to assess crops at their different phenological stages (Ballesteros et al., 2014).

Nowadays, several open-source satellite platforms improve Earth

* Corresponding author.
E-mail addresses: lau_pm@usal.es (L. Piedelobo), david.hernandez@uclm.es (D. Hernández-López), rballesteros@usal.es (R. Ballesteros), amal.chakhar@alu.uclm.es (A. Chakhar), s.p.aguilera@usal.es (S. Del Pozo), daguilera@usal.es (D. González-Aguilera), miguelangel.moreno@uclm.es (M.A. Moreno).

https://doi.org/10.1016/j.agsy.2019.01.005
Received 18 October 2018; Received in revised form 8 January 2019; Accepted 16 January 2019

Observation (EO) capabilities for both agricultural management and crop monitoring. Specifically, Copernicus Open Access Hub and Earth Explorer are open data providers offering Sentinel-2 (S2) A and B and Landsat-8 (L8) images, respectively (Colkesen and Kavzoglu, 2017). Both satellite platforms have been successfully used in the agriculture field (Azar et al., 2017; Hansen and Loveland, 2012; Immitzer et al., 2016; Kussul et al., 2016, 2017; Piedelobo et al., 2018; Sexton et al., 2013; Skakun et al., 2016; Townshend et al., 2012). However, its individual use may entail significant limitations: (i) lack of spatial, temporal or spectral resolutions and (ii) clouds contaminating the images (Cai et al., 2018; Kussul et al., 2017; Skakun et al., 2017). Overcoming these limitations is conceivable using several platforms jointly, so that denser temporal, spatial and spectral input data can be collected (Hao et al., 2014; Roy et al., 2014).

Integrating RS data from several sources is not a trivial issue. Firstly, a proper interoperability between the satellite-based data is required, which entails efficient information management (Shelestov et al., 2017). As this fact has been proven for the freely available L8 and S2 images (Mandanici and Bitelli, 2016), authors decided to use both remotely-sensed data jointly for crop classification.

Specifically, multi-temporal series of the Normalized Difference Vegetation Index (NDVI), which has been the most widely used vegetation index for decades, were analyzed in this study. NDVI is highly sensitive to vegetation cover variations (Rouse et al., 1973) and hence can estimate crop productivity and monitor vegetation health and growth (Immitzer et al., 2016; Inglada et al., 2015; Peña et al., 2014; Piedelobo et al., 2018; Schultz et al., 2015; Sesnie et al., 2008; Wardlow and Egbert, 2008). However, using jointly multi-temporal NDVI data from both satellite platforms led to an extremely large amount of input information. This requires a proper data management, especially for a large agricultural area, which cannot be achieved by a single calibrated classification model.

This paper proposes a new accurate and efficient classification approach for crop mapping. Its development has considered: (i) a well-organized management of input data through special spatial regions, (ii) image pre-processing and feature extraction using NDVI statistics and considering special filters based on agronomic and edaphic criteria, (iii) the selection of proper ground-truth samples, (iv) the use of the most accurate classifiers and (v) accuracy assessment (Lu and Weng, 2007). A deep review has been made along several experiences regarding crop classification results worldwide, checking different algorithms, sensors, number of crops classified, ground-truth samples and study areas, among other parameters (Table 1).

As shown in Table 1, crop type mapping at very large scales and with high resolution imagery does not exist in the literature. The main challenge when performing a crop classification in large study areas remain in the heterogeneous climates, landscapes and phenology. Therefore, testing and validating the developed crop classification approach in the Spanish part of the Duero basin (78,859 km^2) required zoning the study area. Thanks to the creation of special spatial regions (called Tuplekeys) and filters based on agronomic and edaphic criteria, input data was efficiently organized and preprocessed to take the classification process. All this management of the input data improved the efficiency of the classification process without decreasing its accuracy.

Concerning the reference data, both field inspections made by the fluvial guards of the river basin and farmers' statements to the Common Agricultural Policy (n.d.) (CAP) were used, so enough samples calibrated the model and assessed its accuracy. 2017 crop classification in Duero basin was performed using three machine learning classifiers: (i) Decision Trees (DT), (ii) Ensemble Bagged Trees (EBT) and (iii) Weighted Nearest Neighbor (WNN). These algorithms were selected because of their proved accuracy and efficiency (Table 1).

Regarding the number of crop types classified, 15 individual classes and 7 grouped classes were labeled and hence classification algorithms were applied twice. Finally, both salt and pepper algorithm and masks for filtering artificial raw elements and forest areas were applied.

Finally, overall accuracy (OA) was calculated for both the whole basin and each of the Tuplekeys. Thereby, areas with worse classification accuracy, either due to less availability of satellite data or ground-truth data, were also evaluated. Among the three classifiers, EBT showed the best OA, 87% and 92%, when classifying crops individually and grouped, respectively. In addition, very competitive processing times were obtained considering the huge amount of satellite-based data for such a long time and large area analyzed. 16 h were enough for analyzing the whole basin using a commercial computer.

To describe the developed crop classification approach, the paper is organized as follows: after the Introduction, Section 2 describes input data requirements, proposed methodology and the study area where it has been tested and validated. Section 3 shows the results and accuracies obtained for each of the classifiers; Section 4 discusses the results and, finally, Section 5 summarizes all conclusions derived after testing the proposed crop classification methodology in the Duero river basin.

2. Materials and methods

2.1. Case study: the Duero river basin

The study area selected for testing, evaluating and validating the developed classification approach was the Spanish part of the Duero Hydrographic Basin (Fig. 1). This case study was chosen not only due to its large area and significant agricultural activity, but also to its noteworthy spatial and temporal hydroclimatic variability, which also defines its wide diversity of crops, landscapes and phenologies (ITACyL and AEMET, 2013; Fernández Pereira et al., 2015; Herrero Lizano, 2017). Thus, the crop classification approach could be entirely evaluated, including its efficiency, accuracy and required processing time when applied to large areas.

The Spanish part of the Duero river basin is mainly located in the Community of Castile-Leon, in the central-North area of the Iberian Peninsula (Ministry of Environment, 2007), covers 78,859 km^2 and has a continental climate, mostly considered semi-humid-semi-arid. Average annual rainfall is 612 mm/year with a high spatial and seasonal variability ranging from 350 to 2000 mm/year (ITACyL and AEMET, 2013; Ceballos et al., 2004; Fernández Pereira et al., 2015; Herrero Lizano, 2017).

According to the last hydrological plan of the basin, 3.7 million (Fernández Pereira et al., 2015) and 488,491 ha (2017 CAP statements) corresponded to rain-fed and irrigated crops in 2017, respectively. Most crops were wheat (21%), corn (19%) and barley (15%), with almost 200,000 declared plots. Other representative crops were sunflower (7%), fallow (6%), alfalfa (6%), beet (4%) and potato (3%).

2.2. Materials

2.2.1. S2 and L8 multispectral imagery

The presented work used jointly Landsat-8 (L8) and Sentinel-2 (S2) satellite data for the following reasons: (i) they are freely available, (ii) the interoperability between both platforms has been proven (Mandanici and Bitelli, 2016) and (iii) they are able to monitor large surfaces with high spatial, temporal and radiometric resolutions. This may explain their worldwide use as input data for land cover / land use monitoring and decision-making applications (Azar et al., 2017; Cai et al., 2018; Hansen and Loveland, 2012; Immitzer et al., 2016; Kussul et al., 2017; Piedelobo et al., 2018; Schultz et al., 2015; Sexton et al., 2013; Skakun et al., 2016; Townshend et al., 2012).

Specifically, multi-temporal variation of the well-known Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) was used to perform the crop classification in the study area. NDVI profiles have a specific shape depending on the crop type and describe the three most important phenological stages: onset of greenness,

Table 1
Review of crop classification experiences worldwide.

Research	Azar et al., 2017	Ustuner et al., 2017	Ok et al., 2017	Immitzer et al., 2016
Study area	Italy, > 10,000 km^2	Turkey, ~100 km^2	Turkey, ~95 km^2	Austria, ~1,000 km^2 and Germany, < 100 km^2
Number of crops classified	7	2 crops (3 stages for each) and 6 features	6 main classes 13 subclasses	7
Sensor / number of bands used	L8 / 5 (2,3,4,5,7)	RapidEye / 5	SPOT 5 / 4	S2 / 10
Image analysis	PB[b]	PB[b]	OB[a] and PB[b]	OB[a] and PB[b]
Index	EVI[d], NDFI[e], RGRI[f]	Statistical learning	Spectral signatures, statistics and Gini index	Band-metrics for OB[a] and reflectance for PB[b]
Ground-truth data	561 training fields and 287 validation fields (CUAA[g] 2013)	11,585 training pixels. 60 in-situ GPS and 828 Google Earth validation points (AD[i] and LCLU[m])	1,021 field samples (field surveys)	452 field samples (field surveys)
Number of images	26	1	1	1
Classification algorithms	MLC[h], EMD[i], SAM[j], NN[k]	SVM[n] and MLC[h]	MLC[h] and RF[c]	RF[c]
Overall accuracy	92%	81–85%	77 and 85%	76%

Research	Inglada et al., 2015	Tatsumi et al., 2015	Schultz et al., 2015	Peña et al., 2014
Study area	12 test sites worldwide 1,000–17,000 km^2	Peru, ~230 km^2	Brazil, ~7,150 km^2	USA, ~2,650 km^2
Number of crops classified	2–20	8	5	9
Sensor / number of bands used	SPOT 4 / 4, L8 / 6 (2,3,4,5,6,7) and RapidEye / 5	L7 ETM + / 3 (1,3,4)	L8 / 3 (4,5,6)	ASTER / 9
Image analysis	OB[a] and PB[b]	PB[b]	OB[a]	OB[a]
Index	Surface reflectance, NDVI[o], NDWI[p] and brightness	EVI[d]	Statistics, NDVI[o] and brightness	10 vegetation indices (i.e. NDVI[o], VIgreen[s], GNDVI[t])
Ground-truth data	Number not mentioned (field surveys and LCLU[13])	11,781 training pixels and 105,994 validation pixels (field surveys and photointerpretation)	677 training fields and 2,502 validation fields (field visits and photointerpretation)	1,007 field samples (field visits and LCLU[13])
Number of images	9–41	53	2	2
Classification algorithms	SVM[n], DT[q], GBT[r] and RF[c]	RF[c]	RF[c]	DT[q], LR[u], SVM[n] and NN[k]
Overall accuracy	> 80%	81%	80%	79–89%

[a] Object-based.
[b] Pixel-based.
[c] Random Forest.
[d] Enhanced Vegetation Index.
[e] Normalized Difference Flood Index.
[f] Red Green Ratio Index.
[g] Annual Agriculture Land Use Map.
[h] Maximum Likelihood Classification.
[i] Euclidean Minimum Distance.
[j] Spectral Angle Mapper.
[k] Neural Networks.
[l] Ancillary Data.
[m] Land Cover / Land Use.
[n] Support Vector Machines.
[o] Normalized Difference Vegetation Index.
[p] Normalized Difference Water Index.
[q] Decision Trees.
[r] Gradient Boosted Trees.
[s] Green Vegetation Index.
[t] Green Normalized Difference Vegetation Index.
[u] Logistic Regression.

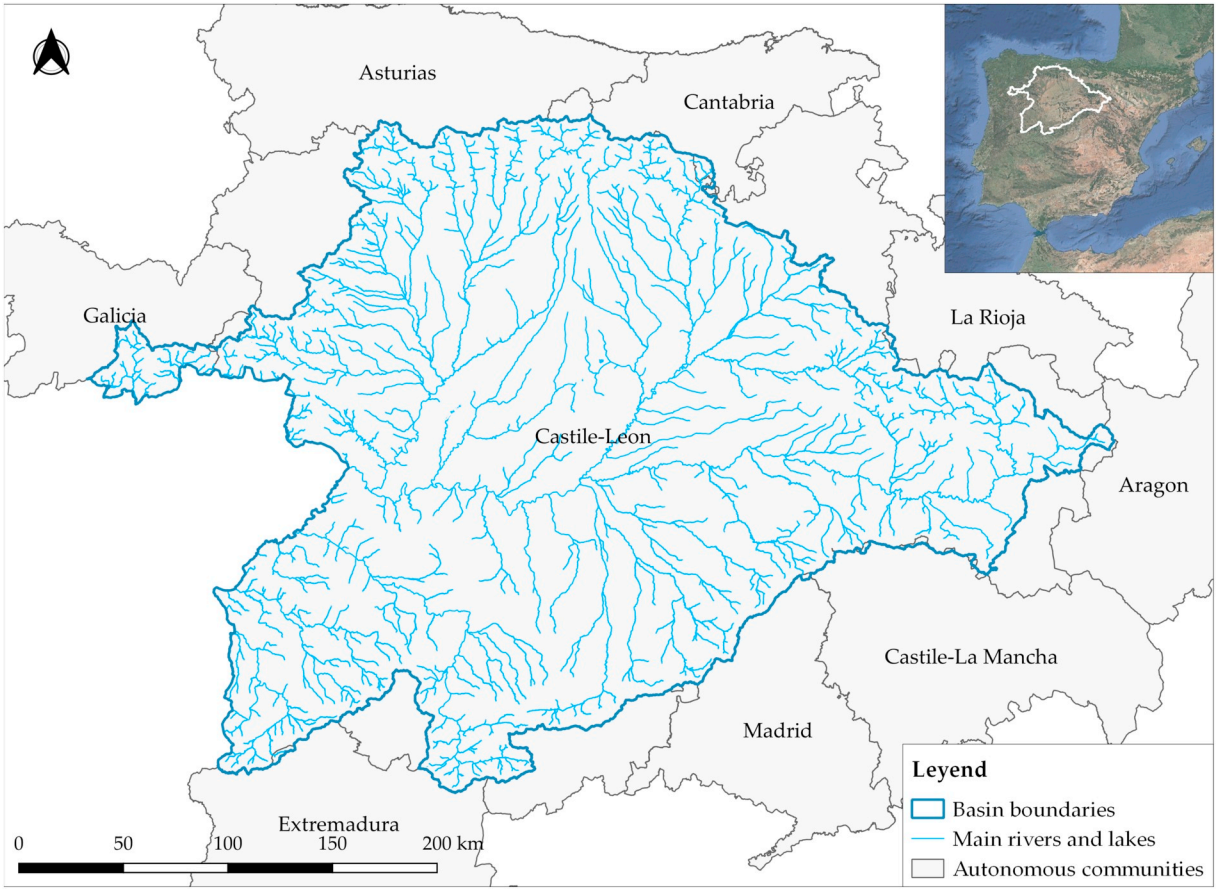

Fig. 1. The Duero Hydrographic Basin located in Spain.

flowering and onset of yellowing. To exploit this information, crop features extracted from multi-temporal NDVI metrics have been extensively used for crop mapping, achieving up to 90% OA values (Bontemps et al., 2015; Inglada et al., 2015; Lu and Weng, 2007; Matton et al., 2015; Peña et al., 2014; Pittman et al., 2010; Schultz et al., 2015; Valero et al., 2016; Wardlow and Egbert, 2008). However, the study of a very large area implies high spatial and temporal variability regarding climate and phenology, so crop characteristics occur in different time periods. In such cases, both zoning the analysis along the study area and gathering enough and representative ground-truth samples are needed.

In this work, 20,836 scenes from L8 (Level-1T) and S2 (Level-1C) were automatically downloaded and preprocessed (Section 2.3.1.) for the period between March 1st and October 31st 2017 with the aim of classifying the crops of the Duero river basin in 2017 attending multi-temporal variation of $NDVI_{TOA}$ values. To obtain the same spatial coherence between L8 and S2 NDVI values, input data was divided in 278 spatial regions (Tuplekeys) created within a local nested grid. 192 out of the 278 Tuplekeys covered the river basin in the storage Level Of Detail (LOD). As further explained in Section 2.3.2, this step was implemented due to L8 and S2 different spatial, temporal and radiometric resolutions (Table 2). Thus, completely integrating all input data in Tuplekeys solved the issue of pixels misalignment. Furthermore, several filtering criteria was applied so just the most representative NDVI scenes (8,671 out of 20,836) were used to perform the final uncalibrated pixel-based crop classification.

Also, as it is not the specific NDVI value in a pixel but the variation of this value through time and its comparison with several crop patterns in a study area what defines a crop classification, NDVI at the top of the atmosphere ($NDVI_{TOA}$) was analyzed, instead of at the bottom of the atmosphere ($NDVI_{BOA}$), avoiding atmospheric corrections and optimizing processing times (Eq. (1)).

$$NDVI_{TOA} = (NIR_{TOA} - R_{TOA})/(NIR_{TOA} + R_{TOA}) \quad (1)$$

Ranging $NDVI_{TOA}$ from -1 to $+1$, with higher values being representative of healthier crops (Rouse et al., 1973), and NIR_{TOA} and R_{TOA} the near-infrared and red reflectances at the top of the atmosphere, respectively.

2.2.2. Ground-truth reference samples

To ensure an accurate crop classification using remotely-sensed (RS) satellite data, a proper amount of reference samples must be provided for calibrating the algorithms and assessing their accuracy (Congalton, 1991). The quality and representativeness of ground-truth data can significantly influence the performance of the algorithms used and, consequently, the classification results (Chen et al., 2002). Field surveys, farmer's statements, aerial photographs, satellite images or land cover / land use (LCLU) maps are widely-used suppliers of ground-truth samples (Azar et al., 2017; Immitzer et al., 2016; Inglada et al., 2015; Lu and Weng, 2007; Ok et al., 2017; Peña et al., 2014; Schultz et al., 2015; Tatsumi et al., 2015; Ustuner et al., 2017).

The two main uses of the reference samples are: (i) calibrating the classifiers and (ii) evaluating the classification accuracy. Therefore, the calibration dataset supports the RS data and the classification algorithms to perform the final crop map (Delincé, 2017).

In this study, both summer field visits provided by fluvial guards of the river basin (from June to September 2017) and farmers' statements to the Common Agricultural Policy (n.d.) (CAP) of such year were used as reference data. 249 field visits determined the crop type, phenological stage and irrigation system established in the visit date. CAP information represented only irrigated plots and provided all farmers' declarations regarding crop type, surface and plot delimitation.

Table 2
Main characteristics of satellites and bands used. Adapted from Landsat Science Portal (n.d.) and The European Space Agency portal (n.d.).

Satellite platform	Landsat-8	Sentinel-2 (A & B)
Sensor	Operational Land Imager (OLI)	Multispectral Instrument (MSI)
Area covered by each scene	170 × 185 km	100 × 100 km
Spatial resolution of the bands used	30 m	10 m
Temporal resolution	16 days	5 days
Radiometric resolution	12 bits	12 bits
Bands used	Band 4 (Red: 630–680 nm)	Band 4 (Red: 645–683 nm)
	Band 5 (Near Infrared: 845–885 nm)	Band 8 (Near Infrared: 762–907 nm)

Table 3
Number of reference samples used for the 2017 crop classification in the Duero river basin, both from field surveys made by the fluvial guards and from the farmer's statements to the Common Agricultural Policy (CAP).

Aggregated crop class	Individual crop class	Number of CAP statements	Number of plots visited
Spring crops	Wheat	8,141	3
	Barley	6,109	8
	Rye	275	–
	Rapeseed	791	2
	Green peas	802	1
	Oat	1,008	–
Summer crops	Sugar beet	2,055	54
	Potato	1,142	60
	Corn	7,567	57
Alfalfa and ray-grass	Alfalfa	2,821	38
	Ray-grass	279	–
Sunflower	Sunflower	2,730	22
Fallow	Fallow	2,780	1
Vineyard	Vineyard	249	3
Pastures	Pasture	94	–
Total		36,843	249

352,403 irrigated plots had been declared to the CAP in 2017 but, after a spatial analysis, those with an area of < 1 ha were deleted to avoid the border effect in small plots. Lastly, considering the labeled classified crop types, 36,843 CAP statements were used (Table 3).

As a novelty, the accuracy of the final crop classification was evaluated per each of the algorithms not only for the whole basin but also per each of the Tuplekeys. Thus, the spatial distribution of the overall accuracy (OA) along the basin was analyzed. Therefore, all available ground-truth data was exploited by using different reference samples for either calibration or validation. All field visits were used for determining the classification OA at the level of the whole basin due to their higher accuracy stating the type and stage of the crops in the visit date. On the other hand, 70% of the CAP statements were used for calibrating the algorithms while the other 30% were used for assessing OA per Tuplekey, compensating the low number of field surveys.

Regarding the crop types classified, 15 individual and 7 grouped classes were labeled, applying the classification algorithms twice. Table 3 shows the number of agricultural plots visited and the CAP statements for the selected individual and gathered crop classes.

2.3. Classification methodology

Obtaining a proper crop classification through the proposed methodology requires to follow 4 main consecutive steps to use jointly multi-temporal RS data from S2 and L8, as outlined in Fig. 2.

2.3.1. Data collection and pre-processing

This step involves the execution of two different processes: (i) downloading and pre-processing RS satellite data and (ii) gathering ground-truth data from the fluvial guards of the river basin and the farmers' statements to the CAP. After downloading L8 and S2 data, $NDVI_{TOA}$ images were generated for both RS sensors. Both United States Geological Survey (USGS) and European Space Agency (ESA) provide services for querying and downloading L8 Level 1T and S2 Level 1C products. These services could be used through Application Programming Interfaces (API). Thus, the developed approach made use of these interfaces, automatically querying the available products (NIR_{TOA} and R_{TOA} images) for the study area and time range of interest (March 1st – October 31st 2017), downloading and keeping them in a system catalogue and calculating $NDVI_{TOA}$ images using Eq. (1) (Piedelobo et al., 2018).

2.3.2. Satellite data integration

Once $NDVI_{TOA}$ images were generated from both satellite sensors, a proper data integration was required to perform a precise crop classification. Therefore, special spatial regions were automatically created for the study area. These spatial regions, called Tuplekeys, had a grid structure compatible with both L8 and S2 data (ANZLIC, 2012; Purss et al., 2015; Stumpf et al., 2018; Villa et al., 2016). The goal was processing multiple $NDVI_{TOA}$ layers from both satellite platforms optimizing the spatial coherence between them. Thus, the following issues regarding Earth Observation (EO) multi-temporality and multi-resolution are solved:

- Pixel misalignment between L8 and S2 images since pixel sizes of both imagery sets are not multiples. This leads to the impossibility of directly comparing radiometric values from different dates.
- Pixel non-alignment in the upper Level of Detail (LOD) of the resolution pyramid. This issue is due to the different spatial resolution of L8 (30 m) and S2 (10 m) images. The size of a given image was N by N where $N = 3^n$, giving the image pyramid a hierarchical structure composed of n levels of the same image of different resolutions. The bottom of the pyramid corresponded to the full resolution of the given image. Each set of 3 × 3 neighbor pixels was replaced by their average as the pixel value of the image at the next level. This process, that divided the image size by 3 on each dimension, was repeated n times until finally an image of only 1 pixel (average of the entire image) was generated as the top of the pyramid with a dimension of 10 × 10 m (finest spatial resolution of S2). Thus, appropriate alignment of the pixels in all LODs was mandatory to perform multi-resolution analysis.

The unique local nested grid solves the storing, processing and comparison between the raster input data from both satellite platforms. This methodology was applied to generate the Tuplekey regions, organizing different footprints, pixel sizes and pixel positions at all pyramid levels (ANZLIC, 2012; Purss et al., 2015; Stumpf et al., 2018; Villa et al., 2016).

To obtain a local nested grid for the Duero river basin, the first step was the creation of a nested grid that covered all the Iberian Peninsula and had the coarsest spatial resolution. The method consisted of placing a fine grid inside the coarse grid. Thus, the coarse grid covered the entire Iberian Peninsula while the fine grid preserved the detailed features of the plots of the study area. The following parameters defined the Tuplekey generation process:

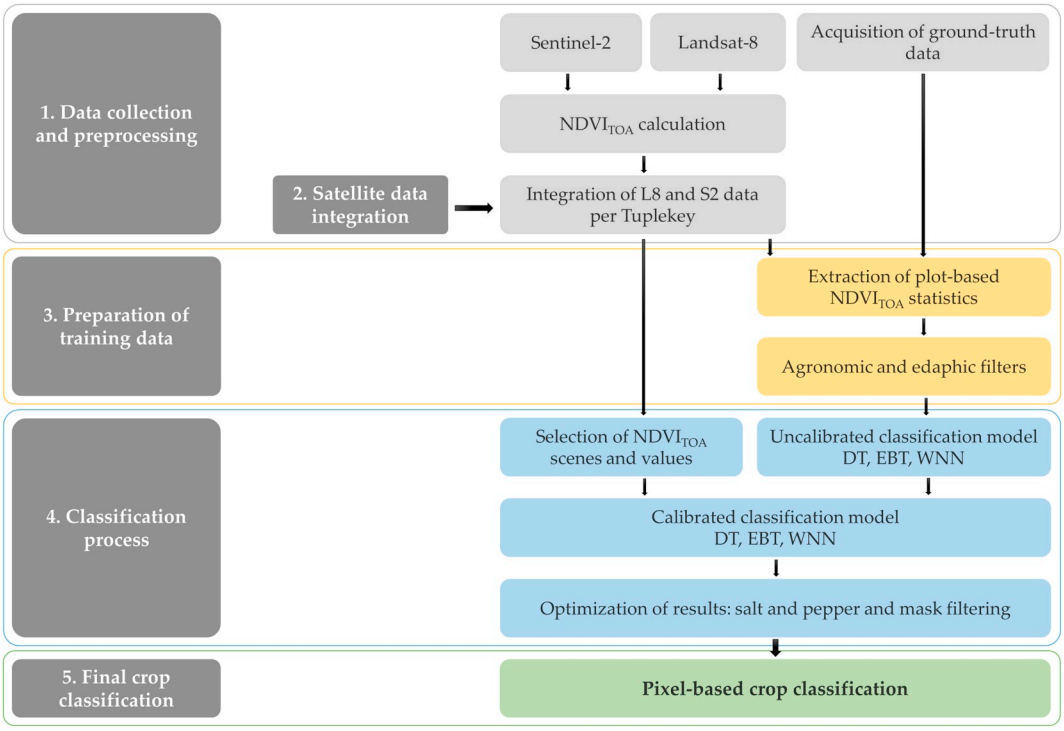

Fig. 2. Flowchart of the proposed methodology for crop classification.

- The coordinate reference system (CRS). In this case, UTM 30 WGS84.
- The coordinates of the upper left corner of the grid: Initial NW (North West) Origin longitude (DEG) and Initial NW (North West) Origin latitude (DEG). Coordinates of the grid cells at all pyramid levels remained as whole integer numbers, not decimal numbers.
- Region of Interest (ROI): the area of interest that covered the entire Iberian Peninsula was equal to 1,500,000.0 m in width.
- Ground Sample Distance (GSD) for the maximum LOD. It was defined as 10 m, which corresponded to the spatial resolution of S2.
- Interval of LODs. LOD 0 corresponded to the coarsest spatial resolution, which was defined as 7,290 m while LOD 6 corresponded to the finest spatial resolution, defined as 10 m. LOD 4 was chosen as the storage level to keep appropriate sized files.
- Recursive ratio factor in tiles. It was defined as 3 because L8 spatial resolution was 30 m and S2 was 10 m, so there was a direct proportion equal to 3 between them. Consequently, the nested grid in each higher resolution level must be divisible in 3 × 3 sequentially.
- The maximum raster file size. It corresponded to a L8 scene of 183,000.0 m of width.
- The final ROG (Region Of Grid). It was obtained by resizing the initial ROI, defined to cover the maximum initial distance entered (1,500,000.0 m), considering the inclusion of a scene of maximum size corresponding to a L8 scene (183,000.0 m).

Fig. 3 shows the developed local nested grid and Tuplekeys from LOD 0 (coarsest nested grid), which covered the entire Iberian Peninsula to LOD 4 (storage level), displaying just the 192 Tuplekeys that covered the river basin.

2.3.3. Preparation of training data

A plot-based approach was used to train the classification algorithms using NDVI$_{TOA}$ metrics (Azar et al., 2017; Immitzer et al., 2016; Inglada et al., 2015; Ok et al., 2017; Peña et al., 2014; Schultz et al., 2015; Tatsumi et al., 2015; Ustuner et al., 2017). Thus, both NDVI$_{TOA}$ average and standard deviation were calculated at the level of each plot of the reference samples. 25,790 CAP statements (70%) were used for this purpose. Since the approach generated a massive volume of information, several filters were applied to use just the most representative training data. This step ensured both a precise and efficient classification model.

1. Extraction of plot-based NDVI$_{TOA}$ statistics.
   Using both CAP data (plots geographic delimitation) and NDVI$_{TOA}$ images organized in the Tuplekey regions, the mean and standard deviation of NDVI$_{TOA}$ at a plot-based level was calculated. Considering the border effect when using RS data at this level, a 30 m buffer was applied inwardly before performing the metrics. Results were kept in two different files, one for L8 and one for S2 with the following information: identification of each plot, crop type and NDVI$_{TOA}$ statistics.
2. Agronomic and edaphic filters.
   The training dataset was filtered by agronomic and edaphic criteria so that only the most useful and representative data trained the classification algorithms. Therefore, when multiple images were available for the same region and week, just the maximum NDVI$_{TOA}$ value remained per plot since no significant change in the phenological stage of a crop occurs in less than a week. This filter's goal is decreasing the high volume of input data in the process. Moreover, picking the maximum NDVI$_{TOA}$ value ensured a less presence of clouds in the image since clouds have very low NDVI$_{TOA}$ values.
   In addition, to ensure both the agro-climatic and edaphic coherence of the training data, just 200 plots of each crop were selected for taking part in the calibration process. Selected plots were the closest to the centroid of each Tuplekey. Nevertheless, if there was a crop class represented by < 200 plots, the total amount would be considered. In this way, both the quantity and representativeness of the training data were assured throughout the basin.

2.3.4. Classification process

During this step, the training data prepared was used to calibrate the classification algorithms by a plot-based approach and per Tuplekey. The fact of both filtering the training dataset and calibrating the model per Tuplekey instead of for the entire basin made it possible

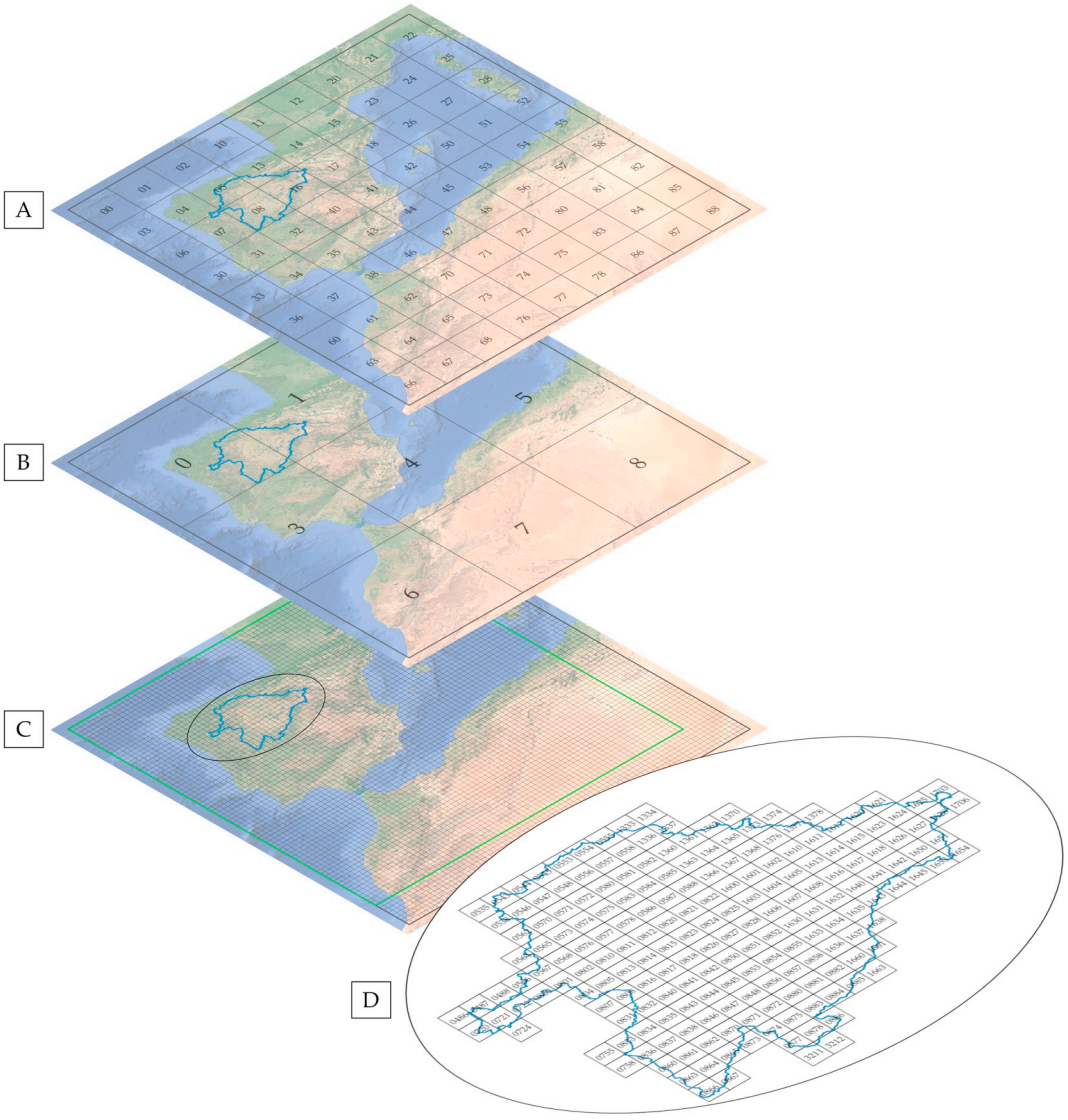

Fig. 3. Local nested grid in several LODs. **(a)** Local nested grid in LOD 2: each tile of the previous grid in LOD 1 was divided in 3 × 3. **(b)** Local nested grid in LOD 1: coarsest grid in LOD 0 with tiles divided in 3 × 3. **(c)** The coarsest nested grid that had LOD 0 and covered the entire Iberian Peninsula (in grey), the ROI (in green) and LOD 4 (little tiles in grey). **(d)** Zoom to the Duero basin to show the Tuplekeys in LOD 4 that cover the study area.

to differentiate the unique edaphic and agro-climate conditions of each region.

The classifiers chosen to be evaluated were Decision Trees (DT), Ensemble Bagged Trees (EBT) and Weighted Nearest Neighbor (WNN) since they show major guarantees in terms of speed and accuracy according to literature (Table 1). Table 4 describes the main characteristics of each method. The *ClassificationLearner* application of Matbal® was used with the aim of calibrating and validating the different algorithms.

The classification methodology goes through the following steps:

1 Selection of $NDVI_{TOA}$ scenes

To simplify the final classification process, S2 over L8 $NDVI_{TOA}$ scenes were selected when both were available for the same Tuplekey and date due to its better resolution.

2 Cloud cover filter

As clouds show very low $NDVI_{TOA}$, only scenes with an average value > 0.20 were used.

3 Classifiers calibration

The selected classifiers must learn from the prepared training data (70% of CAP plots). The plot-based calibration process was performed

Table 4
Algorithms evaluated for crop classification.

Method	Main characteristics
Decision Trees (DT)	Different number of leaves and maximum number of splits, up to 100
Ensemble Bagged Trees (EBT)	Random forest ensemble method with decision trees
Weighted Nearest Neighbor (WNN)	Distance weight. The number of neighbors was set to 10

per Tuplekey and classification algorithms learnt from the $NDVI_{TOA}$ signatures of the training data, considering all the plot-based statistics extracted during the previous step (Section 2.3.3).

4 Crops classified

Regarding the number of crops classified, 15 individual classes (alfalfa, corn, sugar beet, potato, sunflower, barley, wheat, fallow, rapeseed, green peas, vineyard, ray-grass, rye, oats and pastures) and 7 grouped classes (spring crop, summer crop, alfalfa and ray-grass, sunflower, fallow, vineyard and pastures) were labeled. Gathered classes were grouped attending the seasonality of the individual crops and their growing stages as their NDVI signatures are quite similar. Therefore, classification algorithms were applied two times: one considering all individual crop classes and the second one considering just the gathered classes.

5 Optimization of results

Finally, both the salt and pepper denoising approach and masks filtering artificial raw elements and forest areas were applied. Regarding the filtering masks, SIGPAC (Geographic Information System for Agricultural Plots, n.d.) land uses per cadastral parcel were the source of information. Finally, the salt and pepper denoising was applied to avoid the mixed-crop classification of plots (Basukala et al., 2017; Belgiu and Csillik, 2018).

2.3.5. Final crop classification: accuracy assessment

In this case study, classification accuracy was determined per classifier: (i) globally for the whole basin and (ii) per Tuplekey, so a spatial analysis of the OA could also be performed. Thus, an evaluation on how the number of images used per Tuplekey affects the final pixel-based classification was also carried out. Moreover, OA was determined considering: (i) individualized crop classes and (ii) gathered crop classes.

The confusion matrix allowed the comparison between reference data and final crop classes. Field surveys performed by the Duero fluvial guards in 2017 summer period were considered for assessing the OA for the whole basin. However, due to the lack of field samples, 30% of CAP statements were considered for assessing the classification accuracy per Tuplekey.

Overall, user's and producer's accuracies (OA, UA and PA) were calculated to assess the accuracy per classifier. In addition, per-class accuracy was evaluated regarding misclassified crops, both for individual and gathered crop classes, calculating omission and commission errors (OE and CE) (Congalton, 1991).

3. Results

3.1. Classification map for individual crop classes

This classification map offered a high overall accuracy (OA) when considering the 15 individual crop classes (alfalfa, corn, sugar beet, potato, sunflower, barley, wheat, fallow, rapeseed, green peas, vineyard, ray-grass, rye, oats and pastures) for all three classifiers and considering the whole basin. Ensemble Bagged Trees (EBT) classifier provided the best result with and OA of 87%, being Decision Trees (DT) and Weighted Nearest Neighbor (WNN) classifiers very close with OAs of 81% and 80%, respectively. Fig. 4 shows the final pixel-based crop classification map in the Duero river basin for individual crop classes using EBT and Table 5 its confusion matrix, as this classifier obtained the highest accuracy.

The following significant misclassified cases were obtained with WNN: (i) potato as corn 6 times and as sunflower 4 times, (ii) alfalfa as sugar beet 4 times and (iii) sugar beet as corn 4 times. On the other hand, with DT: (i) corn was confused with potato 4 times, (ii) sugar beet with alfalfa 3 times, (iii) alfalfa with pasture 3 times and (iv) sunflower with corn 4 times. However, other 25 misclassified crops were obtained when using DT, even not exceeding 2 times. As for EBT, only one significant misclassification occurred, potato confused with corn 4 times, among other 20 confusion cases, but not exceeding 2 times. Thus, EBT got less significantly confused crops than WNN and DT, which explains its better OA. These confusion cases are explained by the high similarity between their growing cycles. As Table 6 shows, potato and corn were the majorly confused classes due to their likeliness initial and late growing stages.

As for OAs per Tuplekey, WNN ranged 54–58% to 74–78% while global OA was 80% and DT presented the lowest values, ranging 42–46% to 62–66%, while global OA was 81%. The highest OAs per Tuplekey were again obtained when using EBT classifier, ranging mainly 70–74% and 74–78%, but still not reaching its global OA of 87%. Lower OAs were obtained per Tuplekey than for the whole basin due to using different validation data in each case. 30% of CAP statements were considered for estimating OA per Tuplekey instead of the field surveys due to the lack of these ones.

In addition, similar OAs were observed between adjacent Tuplekeys, remaining three spatial regions in the center of the basin as the lowest OA for all three classifiers (Fig. 5). The total number of $NDVI_{TOA}$ images used for the final crop classification was also calculated per Tuplekey, finding out that the lower number of images used, the lower OA was obtained.

Regarding the execution time needed for the classification process, EBT had the longest processing time, lasting 16 h and 8 min. On the other hand, DT classifier, slightly more accurate than WNN, needed only 2 h and a half, while WNN lasted almost twice that time, 5 h and 22 min. The developed approach was performed in a commercial computer with an Intel®CoreTM i9 7900× CPU (3.30 GHz), 64 bits and 64 Mb of RAM.

3.1.1. Per-class accuracy

Table 7 shows commission and omission errors (CE and OE) per class and classifier considering individual crop classes. Ray-grass, rye, oat and pasture showed both a CE and OE equal to 100% due to the lack of field surveys representing these crop classes. Moreover, other crops, i.e. alfalfa and barley, were misclassified as ray-grass and rye. Apart from them, the worst classified crop class was wheat, with a CE of 67% and OE ranging 75–86% for all three classifiers.

On the other hand, green pea was the only individual crop class with both a CE and OE equal to 0% for the three classifiers. Fallow also showed a CE of 0% for all the classifiers, but OE was just 0% when using DT while significantly high, 50 and 67% when using EBT and WNN, respectively.

EBT showed the lowest CE and OE among all three classifiers for alfalfa, corn, sugar beet and potato. Regarding DT, sunflower presented the lowest CE and OE when compared to the other classifiers. On the other hand, WNN showed higher CE and OE per crop class than the other algorithms. Therefore, it could be concluded that EBT performed a better classification for alfalfa, corn, sugar beet and potato and DT for fallow and sunflower while green pea is well classified using either of the classifiers.

3.2. Classification map for grouped crop classes

This classification map also offered high OAs when considering 7 grouped crop classes (spring crop, summer crop, alfalfa and ray-grass, sunflower, fallow, vineyard and pastures) for all three classifiers. EBT provided again the highest OA for the whole basin, 92%, followed in this case by WNN and DT, with OAs of 89% and 88%, respectively. Thus, an improvement of the OAs could be noticed for all three classifiers when considering gathered instead of individual crop classes (92% instead of 87% for EBT, 89% instead of 80% for WNN and 88% instead of 81% for DT). This fact was due to gather some individual

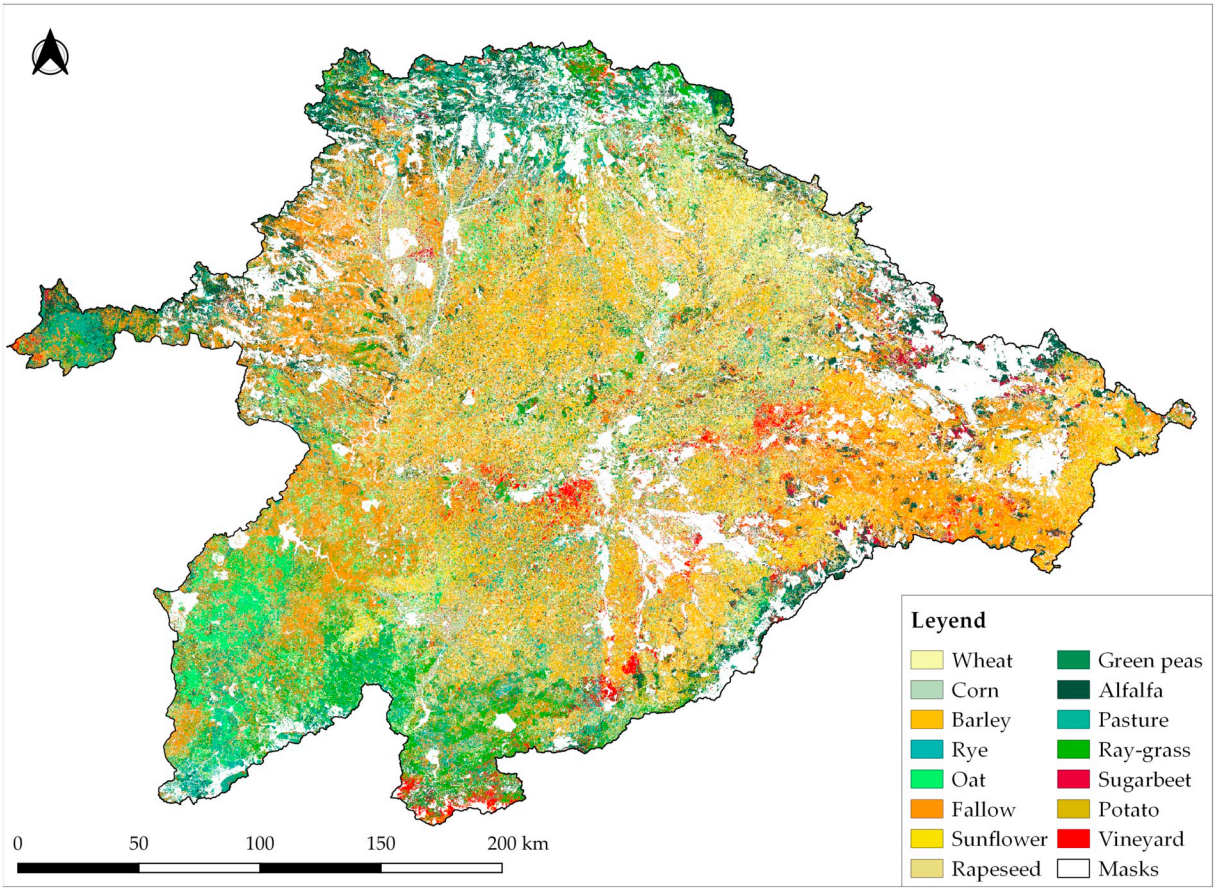

Fig. 4. 2017 final per-pixel crop classification performed with EBT classifier in the Duero river basin considering individual crop classes.

crop classes with similar growing cycles in the same class, decreasing the amount of misclassification cases. Final pixel-based crop classification in the Duero river basin and confusion matrix when using EBT and considering grouped crop classes are shown in Fig. 6 and Table 8, as the most accurate results.

WNN confusion matrix showed the following misclassified crops: (i) alfalfa and ray-grass classified as summer crop 6 times and as spring crop 3 times, (ii) summer crop confused with spring crop 4 times and with sunflower 3 times and (iii) sunflower misclassified as summer crop 3 times. Regarding DT classifier, the most significant confusion classes were: (i) sunflower classified as vineyard 8 times and as summer crop 4 times, (ii) summer crop classified as alfalfa and ray-grass 6 times and (iii) alfalfa and ray-grass confused with spring crop 3 times. Finally, the most noteworthy misclassification classes obtained with EBT were: (i) sunflower classified as summer crop 4 times and (ii) summer crop confused with spring crop 3 times and with alfalfa and ray-grass other 3 times. The growing stages showed again a high likeliness between the confused grouped classes (Table 9). Nevertheless, the most important remark was the decreased number of misclassified crops when classifying grouped instead of individual crop classes.

When getting OA per Tuplekey and hence its spatial distribution along the basin, higher and more homogeneous OA values could be noticed when crops were gathered instead of individualized. However, the three Tuplekeys with the lowest OA remained in the middle of the basin whether considering grouped or separated crop classes and it was directly related with the number of $NDVI_{TOA}$ images used. Spatial distribution of OAs per Tuplekey showed a significance likeliness between EBT and WNN classifiers, ranging mainly 86–90% to 90–94%, like their global OAs of 92 and 89% when assessing the accuracy for the whole basin. DT showed again the lowest values, ranging 42–46% to 62–66%, while its global OA was 88% (Fig. 7).

As for processing times when considering gathered crop classes, EBT, the most accurate classifier, showed again the longest processing time, lasting 16 h. On the other hand, DT needed only 2 h and a half, while WNN lasted 5 h. Therefore, executing processes were slightly quicker when crop classes were grouped instead of separated, but quite similar.

3.2.1. Per-class accuracy

Table 10 shows commission and omission errors (CE and OE) per class and classifier considering grouped crop classes. Pasture was the only class showing both a CE and OE equal to 100% due to the lack of field surveys representing it. As for the other crop classes, errors varied significantly, reaching 73%.

Comparing CE per classifier, fallow was the best classified crop, with a CE of 0% using either of the algorithms. EBT showed better results classifying summer crop and alfalfa and ray-grass (4 and 11%) and DT for vineyard (0%). As for OE, EBT generally presented the lowest values, especially for both fallow and vineyard, with an OE of 0%.

Therefore, for grouped crop classification, EBT showed the best results for classifying sunflower, summer crop and alfalfa and ray-grass; both EBT and DT classified perfectly fallow; while WNN showed the worst results.

3.3. Mixed pixel-based classification in the same plot

Mixed pixels' issue consisted in pixels that, even belonged to the same plot, were classified as different crop types. This problem took place even after applying the salt and pepper denoising approach and for either of the three classifiers. However, it significance depended on the classification algorithm used.

Table 5
Confusion matrix obtained with EBT classifier for the 2017 crop classification in the Duero river basin considering individual crop classes and field surveys performed by fluvial guards in 2017 summer period. Both UA and PA are shown in %.

	A[a]	C[b]	SB[c]	PO[d]	S[e]	B[f]	W[g]	F[h]	RA[i]	GP[j]	V[k]	RG[l]	R[m]	O[n]	P[o]	Total	UA[q]
A[a]	**35**	0	1	0	0	0	0	0	0	0	0	0	0	0	0	**36**	97
C[b]	0	**53**	2	4	2	0	0	0	0	0	0	0	0	0	0	**61**	87
SB[c]	0	0	**48**	2	0	0	0	0	0	0	0	0	0	0	0	**50**	96
PO[d]	0	2	2	**53**	1	1	0	0	0	0	0	0	0	0	0	**59**	90
S[e]	1	2	0	0	**18**	0	1	0	0	0	0	0	0	0	0	**22**	82
B[f]	0	0	1	1	0	**5**	1	0	0	0	0	0	0	0	0	**8**	63
W[g]	1	0	0	0	0	2	**1**	0	1	0	0	0	0	0	0	**5**	20
F[h]	0	0	0	0	0	0	0	**1**	0	0	1	0	0	0	0	**2**	50
RA[i]	0	0	0	0	1	0	0	0	**1**	0	0	0	0	0	0	**2**	50
GP[j]	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	**1**	100
V[k]	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	**2**	100
RG[l]	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	**0**	0
R[m]	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	**0**	0
O[n]	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	**0**	0
P[o]	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	**1**	0
Total	**38**	**57**	**54**	**60**	**22**	**8**	**3**	**1**	**2**	**1**	**3**	**0**	**0**	**0**	**0**	**249**	
PA[p]	92	93	89	88	82	63	33	100	50	100	67	0	0	0	0		

Bold indicates (1) the main diagonal, (2) total number of parcels identified with the same crop type as for each line and column, (3) individual crop classes in the first column and in the first line.

[a] Alfalfa.
[b] Corn.
[c] Sugar beet.
[d] Potato.
[e] Sunflower.
[f] Barley.
[g] Wheat.
[h] Fallow.
[i] Rapeseed.
[j] Green peas.
[k] Vineyard.
[l] Ray-grass.
[m] Rye.
[n] Oat.
[o] Pasture.
[p] Producer's accuracy.
[q] User's accuracy.

Table 6
Misclassified individual crop classes with similar phenological cycles. Numbers 1–12 mean the month of the year, light green "Initial and development stage" and darker green "Mid and late stage". Growing stages of the different crop classes were adapted from FAO (Food and Agriculture Organization of the United Nations). Alfalfa and pasture are not shown as they have pluriannual cycles.

	1	2	3	4	5	6	7	8	9	10	11	12
Sunflower												
Corn												
Potato												
Beet												

DT showed the quickest execution time but also the most noteworthy appearance of mixed pixels, followed by WNN. On the other hand, EBT, which presented the highest OA for both individual and grouped crop classes and for the whole basin or per Tuplekey, showed the lowest rate of mixed pixels at a plot level.

Fig. 8 shows the mixed pixels' issue per classifier for the 210th plot of the field visits performed by the Duero basin fluvial guards. According to the field survey, the crop was a sugar beet. However, it was labeled as sugar beet and potato, both as majoritarian classes, using WNN, and as sugar beet, potato, wheat and pasture using DT classifier. Thus, DT did not only classify two extra classes but also obtained scattered pixels, which made it difficult to distinguish the majoritarian crop class in the plot.

On the other hand, EBT classifier labeled the plot as sugar beet and alfalfa, but it was easier to distinct the majoritarian crop class. Therefore, a plot-based estimation of the majoritarian crop class must be done to solve this issue.

4. Discussion

The developed pixel-based classification approach was tested for performing the 2017 crop classification at the large area occupied by the Duero river basin in Spain, 78,859 km^2. Crop mapping at this kind of scale and using high resolution imagery does not exist in the literature, finding 17,000 km^2 as the largest area studied in previous research experiences (Inglada et al., 2015). The main challenge when performing a crop classification in large study areas remain in the highly heterogeneous climates, landscapes and phenology, so just one calibrated classification model was not enough (Eggen et al., 2016).

The presented methodology achieved almost perfect classification

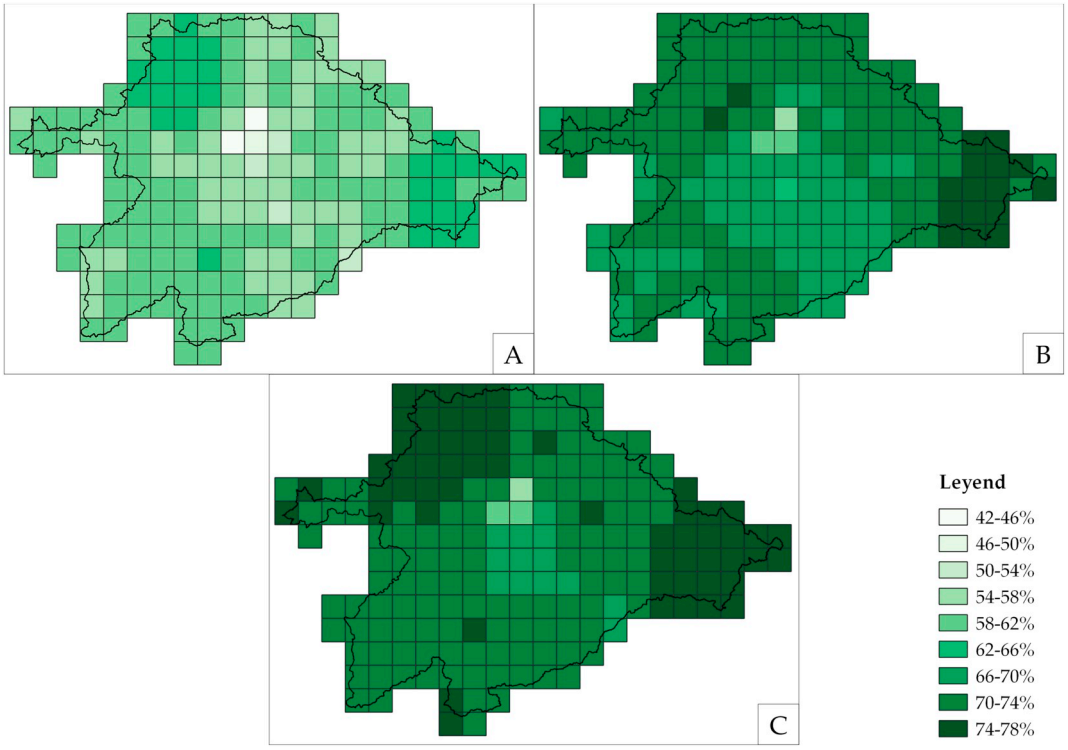

Fig. 5. Overall accuracy (OA) per each of the Tuplekeys when classifying individual crop classes and considering 30% of 2017 farmers' statements to the CAP as validation data. **(a)** Using Decision Trees (DT). **(b)** Using Weighted Nearest Neighbor (WNN). **(c)** Using Ensemble Bagged Trees (EBT).

Table 7
Commission and omission errors per class and classifier for the 2017 crop classification in the Duero river basin considering individual crop classes. Results are organized from the lowest to the highest CE obtained by EBT as the most accurate classifier. Both CE and OE are shown in % for each of the classifiers.

Individual crop classes	Commission error			Omission error		
	DT	WNN	EBT	DT	WNN	EBT
Green peas	0	0	0	0	0	0
Fallow	0	0	0	0	67	50
Corn	16	9	7	11	20	13
Alfalfa	18	18	8	11	6	3
Sugar beet	19	20	11	8	14	4
Potato	14	20	12	32	11	10
Sunflower	15	18	18	12	31	18
Vineyard	0	67	33	25	0	0
Barley	63	63	38	50	25	38
Rapeseed	50	50	50	0	67	50
Wheat	67	67	67	75	86	80
Ray-grass	100	100	100	100	100	100
Rye	100	100	100	100	100	100
Oat	100	100	100	100	100	100
Pasture	100	100	100	100	100	100

results by the effective integration of remotely-sensed data from two different open-source satellite platforms, Landsat-8 (L8) and Sentinel-2 (S2) A and B. These satellites offered two datasets with different spatial and temporal resolutions (30 and 10 m and 16 and 5 days, respectively) and hence its combined use increased both the amount and quality of the input data.

20,836 images from L8 (Level-1T) and S2 (Level-1C) were automatically downloaded and preprocessed in the period of March 1st – October 31st 2017 with the aim of obtaining a set of multi-temporal $NDVI_{TOA}$ images. The developed approach was based on integrating S2 and L8 $NDVI_{TOA}$ raster layers in special spatial regions (Tuplekeys) within a local nested grid that covered the whole basin. This novelty allowed to feed the classification process with a completely coherent and consistent database, organizing image footprints, pixel sizes and pixel positions at all pyramid levels (ANZLIC, 2012; Purss et al., 2015; Stumpf et al., 2018; Villa et al., 2016). Therefore, input data was divided in 278 Tuplekeys and 192 covered the river basin in the storage Level Of Detail (LOD).

However, the main goal was creating not only an accurate but efficient approach for crop classification of large areas. Thus, several filters based on agronomic and edaphic criteria were applied so just the most representative $NDVI_{TOA}$ images (8,671 out of 20,836) were used to perform the final calibrated pixel-based crop classification.

The three most accurate and efficient machine learning classifiers according to the literature (Table 1) were both tested and evaluated. According to the results, Ensemble Bagged Trees (EBT) showed the best performance, both when considering individual and gathered crop classes, reaching overall accuracies (OAs) of 87% and 92%, respectively, which reaffirm previous experiences. However, it also showed the longest executing time, 16 h.

Decision Trees (DT) was slightly more accurate than Weighted Nearest Neighbor (WNN), 81% and 80%, classifying individual crop classes, but less, 88% and 89%, considering aggregated crop classes. Nevertheless, DT has been widely used for crop classification, even not being the most accurate classifier since it is easily trainable and hence provides the quickest processing (Choodarathnakara et al., 2012).

OA was also assessed per Tuplekey using 30% of 2017 farmers' statements to the CAP, as well as the total number of $NDVI_{TOA}$ images used, finding out that low OA values were directly related with the quantity of worth classifying images filtered per Tuplekey. Generally, OA rates per Tuplekey were lower than the generic OA calculated for the whole basin for either of the classifiers due to the quantity of reference data, CAP declarations instead of field surveys.

Regarding per-class commission and omission errors (CE and OE), high rates were mainly associated with a lack of reference data from the field surveys. Thus, ray-grass, rye, oat and pasture for individual crop classification and just pasture for aggregated crop classification

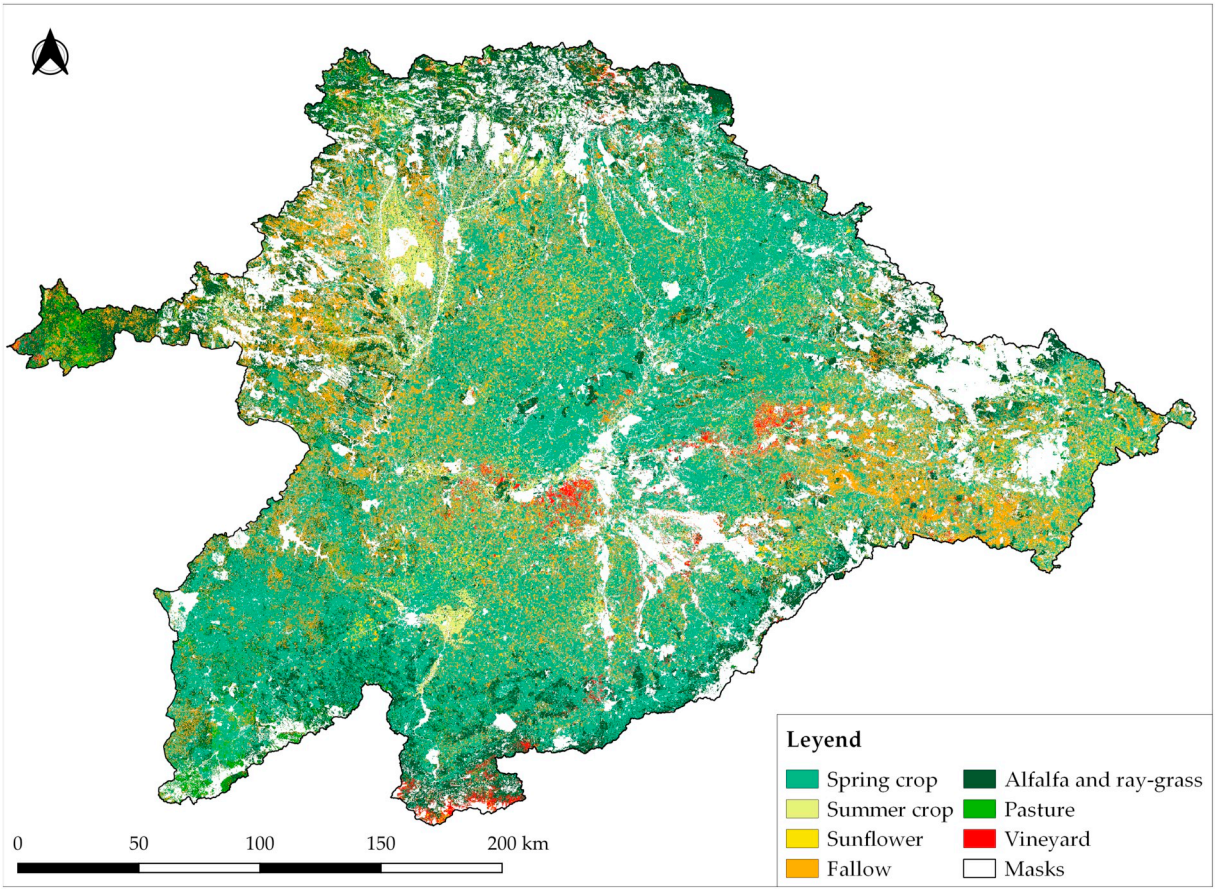

Fig. 6. 2017 final per-pixel crop classification performed with EBT classifier in the Duero river basin considering grouped crop classes.

Table 8
Confusion matrix obtained with EBT classifier for the 2017 crop classification in the Duero river basin considering grouped crop classes and field surveys performed by fluvial guards in 2017 summer period. Both UA and PA are shown in %.

	Spring crop	Summer crop	Alfalfa and ray-grass	Sunflower	Fallow	Vineyard	Pastures	Total	UA[b]
Spring crop	**12**	3	2	1	0	0	0	**18**	67
Summer crop	2	**164**	2	4	0	0	0	**172**	95
Alfalfa and ray-grass	0	3	**34**	0	0	0	0	**37**	92
Sunflower	0	1	0	**17**	0	2	0	**20**	85
Fallow	0	0	0	0	**1**	0	0	**1**	100
Vineyard	0	0	0	0	0	**1**	0	**1**	100
Pasture	0	0	0	0	0	0	**0**	**0**	0
Total	**14**	**171**	**38**	**22**	**1**	**3**	**0**	**249**	
PA[a]	86	96	89	77	100	33	0		

Bold indicates (1) the main diagonal, (2) total number of parcels identified with the same crop type as for each line and column, (3) gathered crop classes in the first column and first line.

[a] Producer's accuracy.

[b] User's accuracy.

Table 9
Misclassified grouped crop classes with similar phenological cycles. Numbers 1–12 mean the month of the year, light green "Initial and development stage" and darker green "Mid and late stage". Growing stages of the different crop classes were adapted from FAO (Food and Agriculture Organization of the United Nations).

	1	2	3	4	5	6	7	8	9	10	11	12
Sunflower												
Vineyard												
Summer crop												
Alfalfa and ray-grass												
Spring crop												

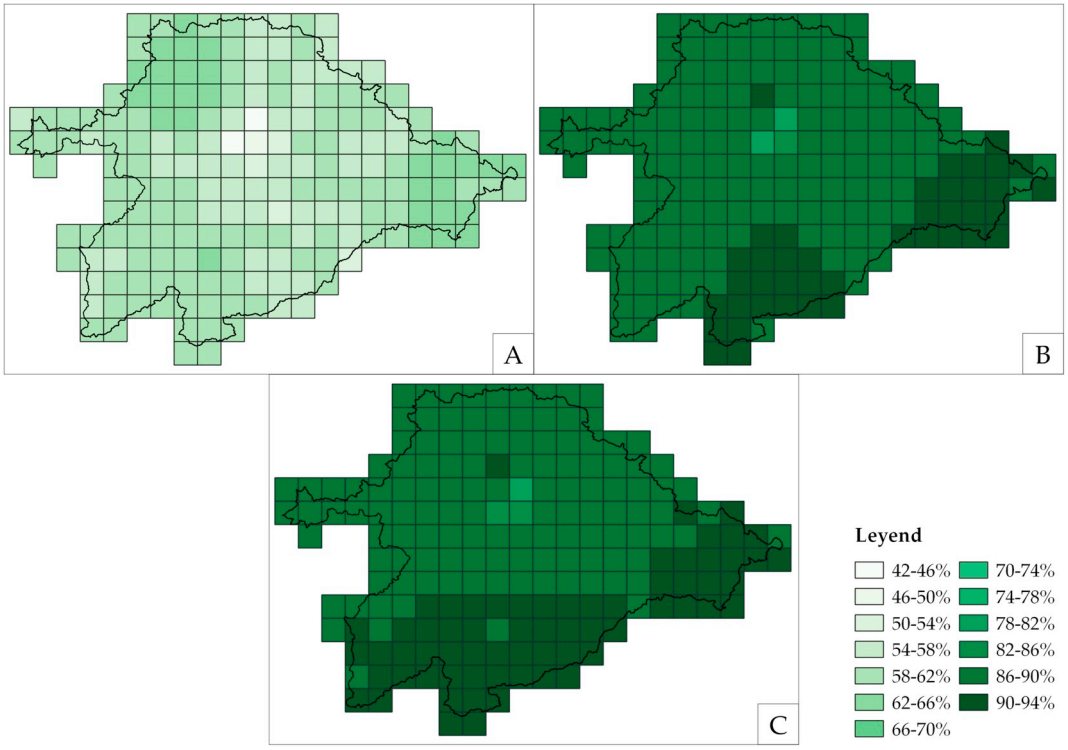

Fig. 7. Overall accuracy (OA) per each of the Tuplekeys when classifying grouped crop classes and considering 30% of 2017 farmers' statements to the CAP as validation data. **(a)** Using Decision Trees (DT). **(b)** Using Weighted Nearest Neighbor (WNN). **(c)** Using Ensemble Bagged Trees (EBT).

Table 10
Commission and omission errors per class and classifier for the 2017 crop classification in the Duero river basin considering grouped crop classes. Results are organized from the lowest to the highest CE obtained by EBT as the most accurate classifier. Both CE and OE are shown in % for each of the classifiers.

Grouped crop classes	Commission error			Omission error		
	DT	WNN	EBT	DT	WNN	EBT
Fallow	0	0	0	0	50	0
Summer crop	6	5	4	4	6	5
Alfalfa and ray-grass	16	24	11	16	6	8
Spring crop	14	14	14	29	43	33
Sunflower	55	23	23	17	23	15
Vineyard	0	67	67	73	0	0
Pasture	100	100	100	100	100	100

obtained both CE and OE of 100%, followed by wheat and vineyard, respectively. Therefore, using a larger and representative ground-truth dataset would allow to improve the crop classification (Lu and Weng, 2007). Defining several monitoring plots that represent the most significant crop types in the basin, as well as developing a methodology for assessing the crop type and stage, could be the solution to this issue.

The presence of mixed pixels was also recognized as an issue that affected the discrimination of the majoritarian crop class in a plot and hence OA per classifier. This problem was most significant using DT classifier, which also showed lower OAs, but the fastest processing. Mixed pixels are a very common issue when performing crop mapping, regardless the sensor resolution and the processing time needed (Basukala et al., 2017; Belgiu and Csillik, 2018; Choodarathnakara et al., 2012; Hsieh et al., 2012). Different approaches have been developed in this matter, i.e. using higher spatial resolution images to enhance the input datasets (Lu and Weng, 2007), but depending on non-freely available data. Another solution could be using soft classification methods based on fuzzy logic, so each pixel would belong to more than one crop class, having membership grades for each class (Choodarathnakara et al., 2012; Murmu and Biswas, 2015).

5. Conclusions

The main purposes of the presented study were: (i) to assess the

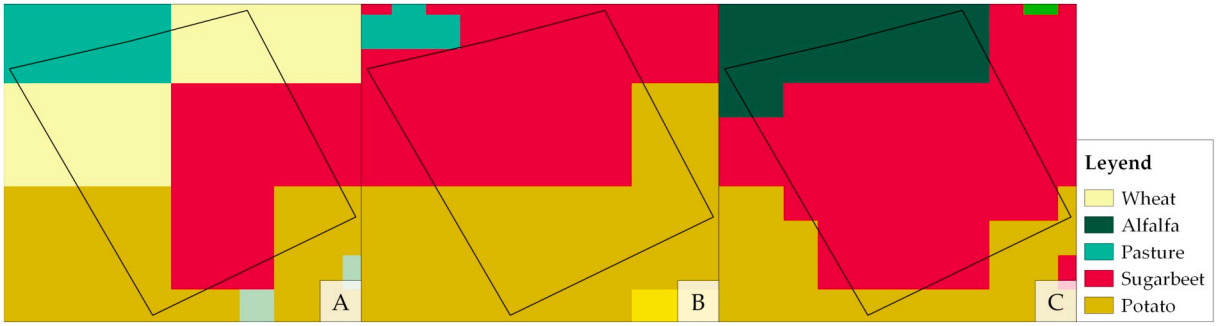

Fig. 8. Pixel-based crop classification for the 210th plot of the field visits performed by the Duero basin fluvial guards in 2017 summer period from the quickest to the slowest classifier. **(a)** Using Decision Trees (DT). **(b)** Using Weighted Nearest Neighbor (WNN). **(c)** Using Ensemble Bagged Trees (EBT).

accuracy when performing a cropland map in the Duero river basin using the three most accurate supervised classification algorithms according to the literature, (ii) to evaluate the integrate use of freely available multi-temporal remote sensing data from S2 and L8 satellite platforms since they have different spatial and temporal resolutions and (iii) to evaluate the approach outputs and efficiency considering both a long period of time and a large study area, and hence a huge quantity of information to be processed. Therefore, the developed classification methodology was capable to:

- Filter the available NDVI dataset attending both agronomic and edaphic criteria, improving the efficiency and therefore computation timing of the classification process since it did not consider redundant input data.
- Integrate the use of L8 and S2 NDVI data in a large study area by automatically creating special spatial regions called Tuplekeys that permitted its jointly used on a coherent way, organizing footprints and pixel sizes and positions at all levels of detail (LOD), increasing the available spatial and temporal resolution of the input data per Tuplekey.
- Test and evaluate the three most accurate supervised classification algorithms: Ensemble Bagged Trees (EBT), Decision Trees (DT) and Weighted Nearest Neighbor (WNN) both for individual and gathered crop classes.
- Obtain not only a generic OA value for the whole basin, but also its spatial distribution along the basin, calculating the accuracy per Tuplekey. Field surveys taken by the fluvial guards of the river basin during the summer of 2017 were used when calculating the classification accuracy for the whole basin. On the other hand, 30% of 2017 CAP statements were used when performing OA per Tuplekey due to the lack of field visits.
- Conclude that EBT classifier provided the best classification results both when considering individual and grouped crop classes, achieving the highest OA both globally and spatially distributed along the basin.
- Obtain 2017 crop classification in the Duero basin, covering almost 80,000 ha, with an OA of 87% and 92% with EBT classifier for individual and grouped crop classes, respectively, in just 16 h of computing time with a commercial computer, which proved the high efficiency and accuracy of the proposed methodology.

Acknowledgments

This research has been supported by the Duero Hydrographic Confederation. Therefore, authors want to thank Javier Fernandez Pereira for his help in achieving the goals of this project. Authors also want to thank the Spanish Ministry of Education, Culture and Sports for providing a FPU grant (Training Program for Academic Staff) to the corresponding author of this paper.

Author contributions

Conceptualization, M.A.M. and D.G.A.; Methodology, D.H.L., R.B., M.A.M., L.P. and S.d.P.; Software, D.H.L and M.A.M.; Validation, L.P., S.d.P. and A.C.; Formal Analysis, L.P., S.d.P. and R.B.; Investigation, L.P., S.d.P., D.H.L., R.B., M.A.M., A.C. and D.G.A.; Resources, D.H.L., M.A.M. and D.G.A.; Data Curation, L.P., S.d.P., D.H.L., R.B. and M.A.M.; Writing-Original Draft Preparation, L.P., S.d.P., A.C., M.A.M. and D.G.A.; Writing-Review & Editing, L.P., S.d.P. and M.A.M.; Visualization, M.A.M. and D.G.A.; Supervision, M.A.M. and D.G.A.; Project Administration, D.G.A.; Funding Acquisition, D.G.A.

Conflicts of interest

The authors declare no conflicts of interest.

References

ANZLIC, 2012. National Nested Grid (NNG) Specification Guideline. Australian and New Zealand Land Information Council, Canberra (31 pp).

Azar, R., Villa, P., Stroppiana, D., Crema, A., Boschetti, M., Brivio, P.A., 2017. Assessing in-season crop classification performance using satellite data: a test case in Northern Italy. Eur. J. Remote Sens. 49 (1), 361–380.

Ballesteros, R., Ortega, J.F., Hernández, D., Moreno, M.A., 2014. Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part II: application to maize and onion crops of a semi-arid region in Spain. Precis. Agric. 15 (6), 593–614. https://doi.org/10.1007/s11119-014-9357-6.

Basukala, A.K., Oldenburg, C., Schellberg, J., Sultanov, M., Dubovyk, O., 2017. Towards improved land use mapping of irrigated croplands: performance assessment of different image classification algorithms and approaches. Eur. J. Remote Sens. 50 (1), 187–201. https://doi.org/10.1080/22797254.2017.1308235.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523. https://doi.org/10.1016/j.rse.2017.10.005.

Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Matton, N., Morin, D., Popescu, R., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Ahmad, I., et al., 2015. Building a Data Set over 12 Globally distributed Sites to support the Development of Agriculture monitoring applications with Sentinel-2. Remote Sens. 7 (12), 16062–16090. https://doi.org/10.3390/rs71215815.

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. U. S. A. 114 (9), 2189–2194. https://doi.org/10.1073/pnas.1616919114.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47. https://doi.org/10.1016/j.rse.2018.02.045.

Ceballos, A., Martinez Fernandez, J., Luengo Ugidos, M.A., 2004. Analysis of rainfall trends and dry periods on a pluviometric gradient representative of Mediterranean climate in the Duero Basin, Spain. J. Arid Environ. 58, 215–233. https://doi.org/10.1016/j.jaridenv.2003.07.002.

Chen, J.M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S.G., White, H.P., Hall, R.J., Peddle, D.R., King, D.J., Trofymow, J.A., Swift, E., Van der Sanden, J., Pellikka, P.K.E., 2002. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. Remote Sens. Environ. 80 (1), 165–184. https://doi.org/10.1016/S0034-4257(01)00300-5.

Choodarathnakara, A.L., Kumar, T.A., Koliwad, S., Patil, C.G., 2012. Mixed pixels: a challenge in remote sensing data classification for improving performance. Int. J. Adv. Res. Comput. Eng. Technol. 1 (9), 261–271.

Colkesen, I., Kavzoglu, T., 2017. Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using sentinel-2 and Landsat OLI imagery. Remote Sens. Lett. 8 (11), 1082–1091. https://doi.org/10.1080/2150704X.2017.1354262.

Common Agricultural Policy Available online. https://pac.jcyl.es/web/jcyl/PAC/es/Plantilla100/1284146916532/_/_, Accessed date: 25 May 2018.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37 (1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B.

Davis, K.F., Gephart, J.A., Emery, K.A., Leach, A.M., Galloway, J.N., D'Odorico, P., 2016. Meeting future food demand with current agricultural resources. Global Environ. Chang. 39, 125–132. https://doi.org/10.1016/j.gloenvcha.2016.05.004.

Delincé, J., 2017. Technical Report on the Master Sampling Frame: The Field Experiments Conducted in Nepal. Technical Report No. 30. Global Strategy to Improve Agricultural and Rural Statistics (GSARS). Technical Report: Rome, December 2017. vol. 138.

Durgun, Y.Ö., Gobin, A., Van De Kerchove, R., Tychon, B., 2016. Crop area mapping using 100-m Proba-V time series. Remote Sens. 8 (7), 585–608. https://doi.org/10.3390/rs8070585.

Eggen, M., Ozdogan, M., Zaitchik, B.F., Simane, B., 2016. Land Cover Classification in complex and fragmented agricultural landscapes of the Ethiopian highlands. Remote Sens. 8 (12), 1020–1043. https://doi.org/10.3390/rs8121020.

Fernández Pereira, J., Rodríguez Arroyo, J., del Barrio, V., Ramos, M.A., Castrillón, M., Vaquerizo, E., Trujillo, H., Hernández, V., Gómez, S., Seisdedos, P., et al., 2015. Plan Hidrológico de la Parte Española de la Demarcación Hidrográfica del Duero 2015–2021; Confederación Hidrográfica del Duero: Valladolid, España; pp. 486 and Annexes.

Geographic Information System for Agricultural Plots Available online. http://www.mapama.gob.es/es/agricultura/temas/sistema-de-informacion-geografica-de-parcelas-agricolas-sigpac, Accessed date: 15 February 2018.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024.

Hao, P., Wang, L., Niu, Z., Aablikim, A., Huang, N., Xu, S., Chen, F., 2014. The potential of Time Series Merged from Landsat-5 TM and HJ-1 CCD for Crop Classification: a Case Study for Bole and Manas Counties in Xinjiang, China. Remote Sens. 6 (8), 7610–7631. https://doi.org/10.3390/rs6087610.

Hasmadi, M., Pakhriazad, H.Z., Shahrin, M.F., 2017. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. Geogr. - Malays. J. Soc. Sp. 5 (1), 1–10.

Herrero Lizano, J., 2017. Informe de Seguimiento del Plan Hidrológico de la Parte Española de la Demarcación Hidrográfica del Duero. Año 2017; Confederación Hidrográfica del Duero: Valladolid, España; pp. 45 and Annexes.

Hsieh, P.-F., Lee, L.C., Chen, N.-Y., 2012. Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. IEEE Trans. Geosci. Remote Sens. 39 (12), 2657–2663. https://doi.org/10.1109/36.975000.

Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 Data for Crop and tree Species Classifications in Central Europe. Remote Sens. 8 (3), 166–193. https://doi.org/10.3390/rs8030166.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an Operational System for Crop Type Map Production using High Temporal and Spatial Resolution Satellite Optical Imagery. Remote Sens. 7 (9), 12356–12379. https://doi.org/10.3390/rs70912356.

ITACyL and AEMET, 2013. Agroclimatic Atlas of Castilla y León. Available online. http://atlas.itacyl.es, Accessed date: 1 February 2018.

Kussul, N., Lemoine, G., Gallego Pinilla, F.J., Skakun, S.V., Lavreniuk, M., Shelestov, A.Y., 2016. Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A Data. IEEE J. Sel. Topics Appl. Earth Observ. 9 (6), 2500–2508. https://doi.org/10.1109/JSTARS.2016.2560141.

Kussul, N., Lavreniuk, M., Skakun, S.V., Shelestov, A.Y., 2017. Deep Learning Classification of Land Cover and Crop Types using Remote Sensing Data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Landsat Science portal Landsat 8 Bands. Available online. https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands/, Accessed date: 20 September 2018.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28 (5), 823–870. https://doi.org/10.1080/01431160600746456.

Mandanici, E., Bitelli, G., 2016. Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use. Remote Sens. 8 (12), 1014–1024. https://doi.org/10.3390/rs8121014.

Matton, N., Sepulcre Canto, G., Waldner, F., Valero, S., Morin, D., Inglada, J., Arias, M., Bontemps, S., Koetz, B., Defourny, P., 2015. An automated method for annual cropland mapping along season for various globally-distributed agrosystems using high spatial and temporal resolution time series. Remote Sens. 7 (10), 13208–13232. https://doi.org/10.3390/rs71013208.

Ministry of Environment, 2007 Feb 3. Real Decreto 125/2007, de 2 de Febrero, Por El Que Se Fija El Ámbito Territorial de las Demarcaciones Hidrográficas; Boletín Oficial del Estado N. 30: Madrid. pp. 5118–5120.

Murmu, S., Biswas, S., 2015. Application of Fuzzy Logic and Neural Network in Crop Classification: a Review. Aquat. Procedia 4, 1203–1210. https://doi.org/10.1016/j.aqpro.2015.02.153.

Ok, A.O., Akar, O., Gungor, O., 2017. Evaluation of random forest method for agricultural crop classification. Eur. J. Remote Sens. 45 (1), 421–432. https://doi.org/10.5721/EuJRS20124535.

Peña, J.M., Gutiérrez, P.A., Hervás-Martínez, C., Six, J., Plant, R.E., López-Granados, F., 2014. Object-based image classification of summer crops with machine learning methods. Remote Sens. 6 (6), 5019–5041. https://doi.org/10.3390/rs6065019.

Piedelobo, L., Ortega-Terol, D., del Pozo, S., Hernández-López, D., Ballesteros, R., Moreno, M.A., Molina, J.-L., González-Aguilera, D., 2018. HidroMap: a new tool for irrigation monitoring and management using free satellite imagery. ISPRS Int. J. Geo-Inf. 7 (6), 220. https://doi.org/10.3390/ijgi7060220.

Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens. 2, 1844–1863. https://doi.org/10.3390/rs2071844.

Purss, M.B.J., Lewis, A., Oliver, S., Ip, A., Sixsmith, J., Evans, B., Edberg, R., Frankish, G., Hurst, L., Chan, T., 2015. Unlocking the Australian Landsat Archive – from dark data to High Performance Data infrastructures. Geo. Res. J. 6, 135–140. https://doi.org/10.1016/j.grj.2015.02.010.

Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations; NASA: Washington, DC, USA, 1973; Pp. 309–317, NASA SP-351.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., et al., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001.

Schultz, B., Immitzer, M., Formaggio, A.R., Del' Arco Sanches, I., Barreto Luiz, A.J., Atzberger, C., 2015. Self-guided segmentation and classification of multi-temporal Landsat 8 images for crop type mapping in Southeastern Brazil. Remote Sens. 7 (11), 14482–14508. https://doi.org/10.3390/rs71114482.

Sesnie, S.E., Gessler, P.E., Finegan, B., Thessler, S., 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. Remote Sens. Environ. 112 (5), 2145–2159. https://doi.org/10.1016/j.rse.2007.08.025.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. Int. J. Digit. Earth 6 (5), 427–448. https://doi.org/10.1080/17538947.2013.786146.

Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017. Exploring Google Earth engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. Front. Earth Sci. 5 (17), 1–10. https://doi.org/10.3389/feart.2017.00017.

Skakun, S.V., Kussul, N., Shelestov, A.Y., Lavreniuk, M., Kussul, O., 2016. Efficiency assessment of multitemporal c-band radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine. IEEE J. Sel. Topics Appl. Earth Observ. 9 (8), 3712–3719. https://doi.org/10.1109/JSTARS.2015.2454297.

Skakun, S., Vermote, E., Roger, J.-C., Franch, B., 2017. Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. AIMS Geosci. 3 (2), 163–186. https://doi.org/10.3934/geosci.2017.2.163.

Stumpf, A., Michéa, D., Malet, J.-P., 2018. Improved co-registration of Sentinel-2 and Landsat-8 imagery for earth surface motion measurements. Remote Sens. 10 (2), 160–180. https://doi.org/10.3390/rs10020160.

Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Ramos Taipe, C.L., 2015. Crop classification of upland fields using random forest of time-series Landsat 7 ETM+ data. Comput. Electron. Agric. 115, 171–179. https://doi.org/10.1016/j.compag.2015.05.001.

The European Space Agency portal Sentinel-2 MSI Resolutions. Available online. https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions, Accessed date: 20 September 2018.

Townshend, J.R., Masek, J.G., Huang, C., Vermote, E.F., Gao, F., Channan, S., Sexton, J.O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M.C., Li, M., Wolfe, R.E., 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. Int. J. Digit. Earth 5 (5), 373–397. https://doi.org/10.1080/17538947.2012.713190.

Ustuner, M., Sanli, F.B., Dixon, B., 2017. Application of support vector machines for Landuse classification using high-resolution RapidEye images: a sensitivity analysis. Eur. J. Remote Sen. 48 (1), 403–422.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sens. 8 (1), 55–76. https://doi.org/10.3390/rs8010055.

Villa, G., Mas, S., Fernández-Villarino, X., Martínez-Luceño, J., Ojeda, J.C., Pérez-Martín, B., Tejeiro, J.A., García-González, C., López-Romero, E., Soteres, C., 2016. The need of nested grids for aerial and satellite images and digital elevation models. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B2, 2016. XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic, https://doi.org/10.5194/isprsarchives-XLI-B2-131-2016.

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250m NDVI data: an assessment for the U.S. Central Great Plains. Remote Sens. Environ. 112 (3), 1096–1116. https://doi.org/10.1016/j.rse.2007.07.019.

Zaks, D.P.M., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. Environ. Res. Lett. 6 (1), 014017. https://doi.org/10.1088/1748-9326/6/1/014017.