Research Article

Received: 16 March 2022 Revised: 24 August 2022 Accepted article published: 12 October 2022 Published online in Wiley Online Library: 27 October 2022

(wileyonlinelibrary.com) DOI 10.1002/ps.7233

Detection of *Ecballium elaterium* in hedgerow olive orchards using a low-cost uncrewed aerial vehicle and open-source algorithms

Jorge Torres-Sánchez,[a*] Francisco Javier Mesas-Carrascosa,[b] Fernando Pérez-Porras[b] and Francisca López-Granados[a]

Abstract

BACKGROUND: *Ecballium elaterium* (common name: squirting cucumber) is an emerging weed problem in hedgerow or superintensive olive groves under no tillage. It colonizes the inter-row area infesting the natural or sown cover crops, and is considered a hard-to-control weed. Research in other woody crops has shown *E. elaterium* has a patchy distribution, which makes this weed susceptible to design a site-specific control strategy only addressed to *E. elaterium* patches. Therefore, the aim of this work was to develop a methodology based on the analysis of imagery acquired with an uncrewed aerial vehicle (UAV) to detect and map *E. elaterium* infestations in hedgerow olive orchards.

RESULTS: The study was conducted in two superintensive olive orchards, and the images were taken using a UAV equipped with an RGB sensor. Flights were conducted on two dates: in May, when there were various weeds infesting the orchard, and in September, when *E. elaterium* was the only infesting weed. UAV-orthomosaics in the first scenario were classified using random forest models, and the orthomosaics from September with *E. elaterium* as the only weed, were analyzed using an unsupervised algorithm. In both cases, the overall accuracies were over 0.85, and the producer's accuracies for *E. elaterium* ranged between 0.74 and 1.00.

CONCLUSION: These results allow the design of a site-specific and efficient herbicide control protocol which would represent a step forward in sustainable weed management. The development of these algorithms in free and open-source software fosters their application in small and medium farms.

Keywords: remote sensing; site-specific weed management; digital transformation; machine learning; random forest; squirting cucumber

1 INTRODUCTION

Ecballium elaterium (L.) Richard, popularly known as squirting cucumber, belongs to the Cucurbitaceae family and is an herbaceous, creeping, perennial plant present in the Mediterranean basin. In recent years, it has caused serious problems as a weed in olive groves and other woody crops because it is a hard-to-control weed and thus can easily expand its populations.[1,2] The control of this weed is complicated by the fact that, being a perennial plant that resprouts from rhizomes, it has continuous emergence, such that at certain times of the year stands with seedlings, flowering plants and fruiting plants can be found simultaneously in the same field. It colonizes the inter-row area of orchards infesting the natural or sown cover crops. This is relevant because hedgerow or superintensive olive groves are mainly under a no-tillage system with natural covers, and only clearing work is carried out, which is incapable of controlling this weed. Blank *et al.*[1] have shown that the spatial distribution of *E. elaterium* in almond orchards shows an aggregate pattern in stands that are also stable over time by using geostatistical analyses. These characteristics make this weed susceptible to being managed by site-specific treatments, thus avoiding the application of homogeneous treatments in the entire plot and allows for an improvement in the efficacy of its control.

Remote sensing based on uncrewed aerial vehicles (UAVs) has shown suitability for the generation of weed infestation maps for both herbaceous,[3,4] and woody crops.[5,6] The availability of infestation maps allows the design of site-specific weed treatments, which represents a step forward in sustainable agriculture. In the detection of weeds with UAV imagery, object based image

* *Correspondence to: J Torres-Sánchez, imaPing Group, Department of Crop Protection, Institute for Sustainable Agriculture (IAS), Spanish National Research Council (CSIC), Córdoba, Spain. E-mail: jtorres@ias.csic.es*

a *imaPing Group, Department of Crop Protection, Institute for Sustainable Agriculture (IAS), Spanish National Research Council (CSIC), Córdoba, Spain*

b *Department of Graphic Engineering and Geomatics, Campus de Rabanales, University of Cordoba, Córdoba, Spain*

analysis (OBIA) has become almost standard in recent years, with a plethora of papers using this image analysis paradigm for weed detection.[5,7–11] In OBIA, the minimal unit for classification is the object instead of the pixel. Objects are groups of pixels with homogeneous values which are characterized by their spectral, geometrical and topological values, adding more variables to the classification process than in traditional pixel-based image analysis (PBIA). The creation of the objects is called segmentation, and although there are several image segmentation algorithms, all of them need the definition of a series of parameters. An optimal selection of the segmentation parameters is crucial because of its influence on the classification accuracy.[12,13]

In OBIA, as in PBIA, supervised and unsupervised classification algorithms, and the selection of one of these approaches depends on the classification problem to be addressed. Unsupervised OBIA approaches have demonstrated accurate results in weed detection problems with no need to identify different weed species. In these cases, the aim of the classification process is to discriminate a limited number of classes (soil, crop, weed) having features that allow its discrimination based on a reduced set of thresholds. For example, soil can be distinguished from vegetation using vegetation indices, and weeds can be differentiated from the crop using attributes such as height or their position among the crop rows.[6,8,14] However, in the cases when it is necessary to detect different weed species, supervised approaches are needed since they can more accurately deal with the discrimination of classes with similar appearance such as weeds belonging to the same family or genus.[4,15] Among the supervised classification models, the random forest (RF) classifier is a machine learning algorithm widely applied in remote sensing because of its robustness and resistance to outliers.[16,17]

Image classification processes can be performed on proprietary or open-source software. The use of free and open-source software may be an important advantage in relation to the implementation of precision agriculture and digitizing techniques, as economic costs have been identified as one of the barriers to the spread of these strategies.[18,19] Furthermore, the use of open-source software is gaining interest in the geospatial community,[20] and it has been successfully applied in previous works on weed detection using UAV imagery.[3,21]

As a result of the above discussion, the aim of the present work was to develop a methodology for the detection of *E. elaterium* infestations in hedgerow olive orchards oriented to the implementation of site-specific control strategies. This methodology consisted of combining images from a low-cost RGB sensor onboard a UAV platform with OBIA algorithms developed on free and open-source software. To the best of the authors' knowledge, this is the first time that remote sensing has been used for weed mapping in an olive orchard, and for detection of *E. elaterium*.

2 MATERIALS AND METHODS

2.1 Study fields and UAV flights

The studies were conducted in two plots from a hedgerow olive orchard located in the province of Córdoba, southern Spain (Fig. 1). This farm is located in an area with warm inland Mediterranean climate, characterized by mild winters, and summers with high temperatures and low precipitations. The private company ELAIA, owner of the fields, authorized the UAV flights and field work. Field 1 (central coordinates 37°48′21″N, 4°45′40″ W, WGS-84, planted in 2015) and Field 2 (central coordinates 37°48′15″N, 4°46′30″ W, WGS-84, planted in 2016) had an area of 10 615 m^2 and 11 758 m^2, respectively. Both fields were planted with the 'Arbosana' variety at a density of 3.75 × 1.35 m (1,975 trees ha^{-1}), under no-tillage and natural cover crops in the inter-row area. The selection of the fields' location and limits was based on the information about *E. elaterium* infestations provided by the company owning the olive farm.

UAV flights and field work were conducted on 05/19/2021, when the orchards were naturally infested by *E. elaterium* and other weeds such as *Convolvulus* spp. or *Amaranthus* spp., and on 09/27/2021 when *E. elaterium* was the only infesting weed. Coordinates of the *E. elaterium* plants and patches were registered using a RTK GNSS. A GNSS Leica Viva GS15 antenna was used for measuring the plants position. It is capable of receiving signals from GPS, GLONASS, Galileo and Beidou systems. The antenna was mounted to the top of a monopole at 1.80 m, setting as mask angle 15°. Ground control points were measured using incoming corrections from RAP network with a baseline distance lower than 10 km (https://www.juntadeandalucia.es/institutodeestadisticaycartografia/rap/nodos). Corrections were received *via* NTRIP (Network Transport of Radio Technical Commission for Maritime Service (RTCM) *via* Internet Protocol) on a RTCM3-iMAX mount point. Those, the RTCM-iMAX uses a real reference base GNSS station to send the network corrections. Under these conditions, the manufacturer reported an accuracy of the system with a RTK network of 8 mm + 0.5 ppm and 15 mm + 0.5 ppm taking into account horizontal and vertical accuracy. Figure 2 shows images of the *E. elaterium* plant disposition in the center of the inter-row areas on both dates. *E. elaterium* grows in spring, summer and autumn, and germinates in both spring and autumn, being possible to find flowering plants at any of these times.[2] Therefore, on both sampling dates coexisted plants in a wide variety of growth stages: from seedlings, to adult plants with flowers and fruits.

Aerial images were acquired with a quadcopter model Mavic Pro 2 (DJI, Shenzen, China) equipped with an RGB camera Hasselblad LID-20c with 20 Mp. The UAV was configured for flights at 50 m over the terrain following flight lines parallel to the hedgerows. The aerial images had forward and side overlaps of 90% and 85%, respectively. UAV flights were carried out around noon on sunny days with no wind.

UAV images were processed with Agisoft Metashape Professional Edition software (Agisoft LLC, St. Petersburg, Russia) for the generation of the geomatic products used in the classification workflows: (i) the orthomosaic containing the spectral information, and (ii) the Digital Surface Model (DSM) which provided height information. The configuration parameters for the generation of geomatic products were the same as in de Castro *et al.*[22] The orthomosaics from both fields and dates had a spatial resolution of 1.1 cm, and the DSMs had a spatial resolution of 2.2 cm. The production of the geomatic products was automatic, with the exception of the manual measurement of five ground control points (GCPs) in the images (one in each corner plus one in the field center). GCP coordinates were measured in the field with a RTK GNSS receiver at the day of flights.

2.2 *E. elaterium* detection algorithms

Different approaches were tested for *E. elaterium* detection depending on the agronomic scenario (Fig. 3). When *E. elaterium* was the dominant weed and there were also various weed species in the field (May), a supervised approach was applied consisting of the training of an RF model. When *E. elaterium* was the only infesting weed (September), an unsupervised algorithm with no need

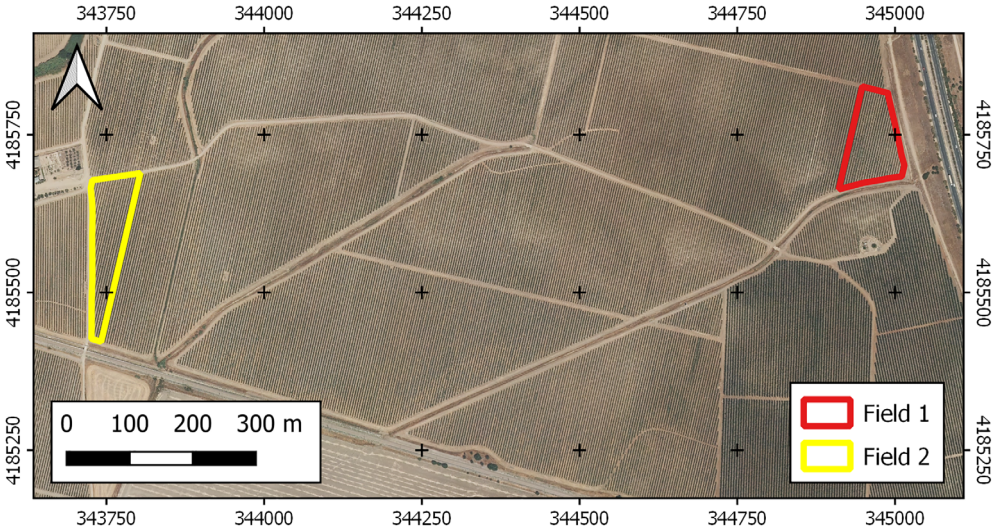

Figure 1. Map showing the location of the study fields: orthoimage from the Spanish National Plan for Aerial Orthophotography (PNOA). Coordinate system: WGS84 UTM 30 N.

of training dataset or user intervention was used. All the processes involved in the classification workflows were carried out using open-source software.

2.2.1 *Image segmentation and object feature creation*

In the supervised and the automatic weed detection algorithms there were some common steps related to the segmentation of the orthomosaic and the generation of the variables used in the classification processes. All these previous steps and the classification algorithms were programmed in the R computer language.[23]

In any OBIA classification algorithm, the first step is image segmentation. Multiresolution segmentation was applied as implemented in Terralib 5.2.1 (National Institute for Space Research (INPE), Brazil),[24] and executed in R through the package SegOptim.[13] This segmentation algorithm was used because it has been successfully applied in previous weed detection algorithms.[4,25,26] It requires the setting of a set of parameters (compactness weight, color weight, minimum segment size, and merging threshold) that influence the shape and size of the created objects, which has paramount importance due to their effect on the performance of OBIA classification algorithms.[12,13,27] The selection of the segmentation parameters was carried out by applying the SegOptim package, which uses an approach based on genetic algorithms (GA) to optimize these parameters. The parameter optimizer needs as input a subset of the image and a vector file with labeled samples of the searched classes. The aim of the GA is to find the set of segmentation parameters able to achieve the best results in the RF classification. According to that goal, the GA segments the image iteratively with different parameter sets and evaluates the accuracy of an RF model created with the labeled samples in every segmentation created in each iteration. The optimization process was applied in a 400 m^2 tile from the orthomosaic from field 1 in May. The optimization process was not repeated in the images from the other field and dates because they had similar characteristics and, consequently, the same segmentation parameters could be used in all of them. As the segmentation of a whole orthomosaic is a time-consuming task, the orthomosaics were divided into 20 × 20 m tiles that were individually segmented to speed-up the process.

Once the image is segmented, the next step in an OBIA classification algorithm is the calculation of the object features that will be used for discriminating among the different classes in the image. Table 1 shows the object variables used in the present work. Among the set of variables included in Table 1, there were

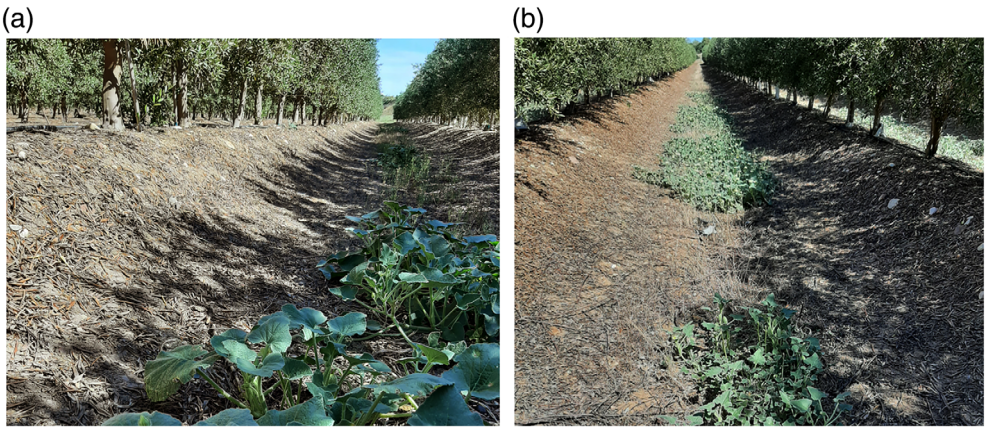

Figure 2. Field view of the *E. elaterium* infestations in May (a) and September (b).

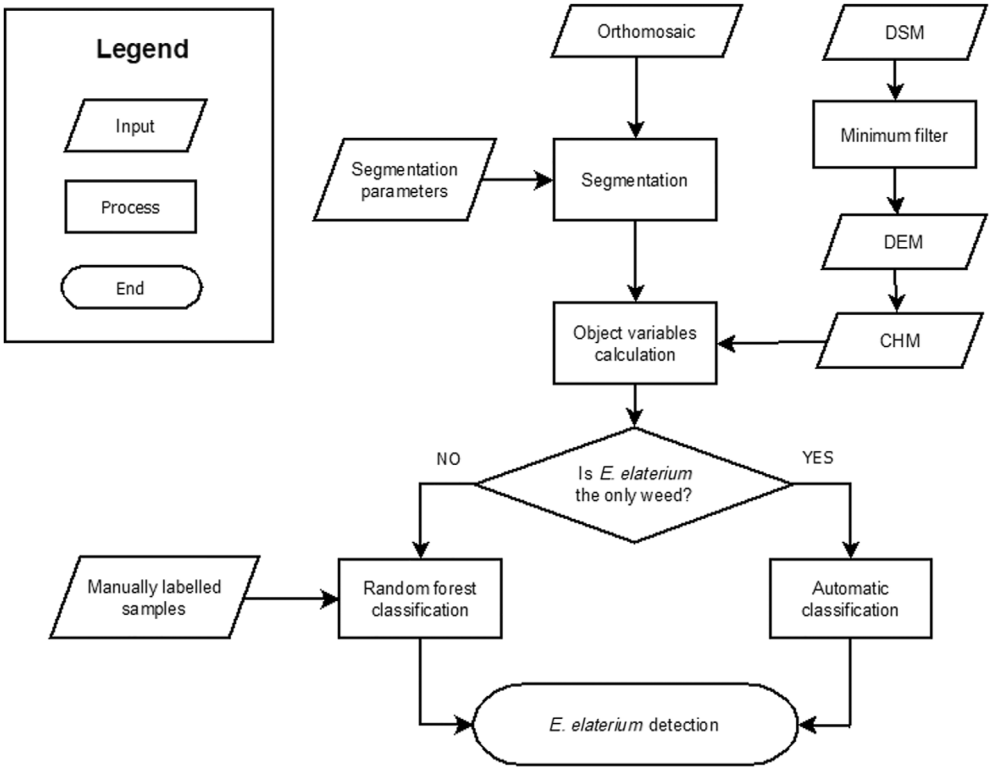

Figure 3. Flowchart of the workflow for *E. elaterium* detection in different scenarios.

four geometric, one topologic, and six spectral variables including two commonly used vegetation indices: excess green (ExG) and normalized green-red difference index (NGRDI).[28,29] Two of the variables are more complex features that require explanation about their calculation: height and distance to the olive row axis.

The height of the plants was included in the set of objects variables because it has been successfully used in previous research works for weed detection.[11,25] The height of the objects was calculated using a canopy height model (CHM) generated by subtracting the height of the ground from the DSM values. The height of the ground was estimated using a digital elevation model (DEM) which was created by applying to the DSM a minimum filter with a radius of 3 m using the process *wbt_minimum_filter* from the Whitebox Tools R package.[30] This filter assigns each cell in the DEM the minimum value in a moving window centered on each grid cell from the DSM. In the present work, a window with a side of 1.5 m was applied. It was assumed that the points with a minimum height corresponded to soil pixels. This simple approach was selected considering the smooth flat topography of the olive orchards.

The first step to calculate the distance to the olive row axis was to detect the olive rows. This was done by classifying all objects with a height over 0.5 m as belonging to the olive rows. Then, all these objects were merged to create a vector layer storing polygons corresponding to the olive hedgerows. A similar approach for crop row detection has been previously tested and validated in olive,[31] and other woody row crops such as vineyards.[22] After the olive row classification, their axes were extracted using the process *wbt_polygon_long_axis* from Whitebox Tools R package, and the distance to the closest olive row axis was calculated for all the objects in the orthomosaics.

2.2.2 *Random Forest model*

RF models were created for fields 1 and 2 in May, when there were other weeds apart from *E. elaterium*. These models were created to classify the following classes: 'Olive', 'Soil', 'Shadow', '*E. elaterium*', and 'Other weeds'. The first step in the creation of the RF model was the manual labeling of a set of objects belonging to the five classes. To create the training dataset, 90 circular plots with a radius of 1 m were randomly created over the orthomosaics of the fields. Image objects inside these circular plots were manually classified (Fig. 4) by only one expert to avoid

Table 1. Object variables used in the classification algorithms

Variable type	Variable name and formula
Spectral	$r=\frac{R}{R+G+B}$†
	$g=\frac{G}{R+G+B}$
	$b=\frac{B}{R+G+B}$
	$ExG=2g-r-b$
	$NGRDI=\frac{G-R}{G+R}$
	$Brightness=\frac{R+G+B}{3}$
Geometric	Area of the object
	Perimeter of the object
	Area/Perimeter
	Height (mean height in the canopy height model)
Topologic	Distance to olive row axis

† R, G, and B indicate the mean values for the red, green, and blue bands of the image object, respectively.

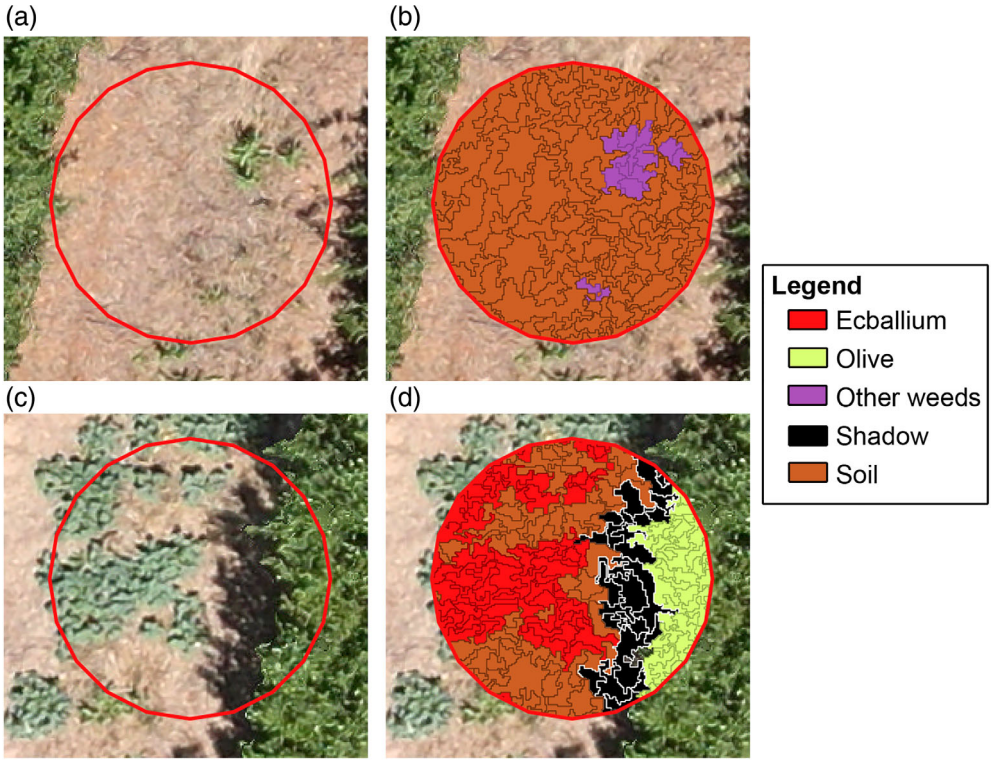

Figure 4. Detailed view of some training points in field 1 in May. Original images with plots containing other weeds (a) and *E. elaterium* (c), and results of the manual object labeling of the same plots (b and d).

discrepancies. The coordinates of the *E. elaterium* plants recorded during the field work were used when there was doubt when labeling weed objects.

RF models were created using the *randomForest* package in R.[32] The models were fed with the object variables included in Table 1 and were created using 500 trees, which is the recommended number of trees from a detailed review on the use of RF in remote sensing.[16] The number of variables randomly sampled as candidates at each split in the creation of the forest of decision trees was three, the closest number to the square root of the total number of variables. A study about the correlation among the input variables was not performed because RF can successfully handle multicollinearity,[16] and because the number of variables was not high enough to lead to excessive processing times. The training dataset was not class-balanced, with '*E. elaterium*' and 'Other weeds' as the less represented classes in all fields. As the training dataset needs to be class balanced,[16] a random undersampling process was implemented. Nonetheless, the RF models were also created using the whole training dataset to compare the results between models with balanced and unbalanced training datasets.

RF models for Septembers' data were also created compare their results with those from the automatic classification algorithm. These RF models aimed to classify four classes: 'Olive', 'Soil', 'Shadow', and '*E. elaterium*' since there were no other weeds.

2.2.3 *Unsupervised classification algorithm*

Starting after the segmentation of the image and the calculation of the object variables, an automatic and unsupervised classification algorithm was applied for the date with *E. elaterium* as the only infesting weed. It did not need any user intervention during the classification process, nor the previous creation of a training dataset, and it can be divided into the following steps:

(1) All objects whose height was higher than 0.5 m were classified as 'Olive'. This height threshold was selected because the only elements higher than 0.5 m in the field were the olive trees.
(2) The Otsu thresholding method[33] was applied to the brightness values of all the unclassified objects, and all the objects whose brightness was below the calculated threshold were classified as 'Shadow'.
(3) After the first steps, the remaining unclassified objects belonged to 'Soil' and '*E. elaterium*'. They were discriminated by applying a threshold to the ExG (Table 1) values. This threshold was calculated using the Otsu method for the object values of this spectral index. After this step, all the objects in the image were classified as 'Olive', 'Soil', 'Shadow', or '*E. elaterium*'.
(4) To avoid misclassification due to the presence of some objects with anomalous values, some enhancements were applied to the object classification:
    4.1 There were segments belonging to the 'Olive' class that could be classified as *E. elaterium* due to having low height values associated with small mistakes in the DSM creation. Since *E. elaterium* infestations grow in the center of the inter-row area (Fig. 1), all the objects classified as *E. elaterium* having a distance to the row axis lower than 1.2 m were classified as 'Olive'.
    4.2 Due to the growing habits of *E. elaterium*, some parts of the plants projected shadows over other parts. Consequently, there could have objects classified as 'Shadow' that belonged to '*E. elaterium*' class. To solve this possible misclassification,

all the objects classified as 'Shadow' that were located more than 1.2 m away from the row axis and had an ExG value higher than the ExG threshold calculated in step 3 were reclassified as '*E. elaterium*'. The 1.2 m distance threshold for this and the previous step (4.1) was based on the width of the olive hedgerow and of its projected shadow.

2.3 Validation

Confusion matrices were calculated using the training datasets manually labeled for the creation of the RF models as reference. The amount of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), were extracted from the confusion matrices. To assess the accuracy of the classifying algorithms, overall accuracy (OA) (Eqn (1)) was calculated from the confusion matrix for every date and field, together with the producer's and user's accuracy (PA and UA) (Eqns (2) and (3), respectively) for the '*E. elaterium*' class. PA represents the percentage of objects belonging to a class that is correctly classified, while UA represents the percentage of objects assigned to a class that actually belong to this class.

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$PA = \frac{TP}{TP + FN} \quad (2)$$

$$UA = \frac{TP}{TP + FP} \quad (3)$$

In the RF models, out-of-bag (OOB) samples were used to create the confusion matrix. The RF classifier uses a set of decision trees to make a prediction. In the creation of the trees, approximately two-thirds of the training dataset is used to train them, and the remaining third (the OOB samples) is used in an internal cross-validation process.[34] The OOB error is claimed to be an unbiased estimate of the generalization error,[35] although OOB error is also said to overestimate the true error in some situations,[36] which would imply that the RF performs better than indicated by OOB error.

For a more complete assessment of the RF model accuracy, apart from the use of the OOB estimations, each model was evaluated by applying it to the training dataset from the other field on the same date. For example, the RF model from Field 1 in May was validated by applying it to the training dataset from Field 2 in May. Using this approach, the accuracy of the models was evaluated twice, and the transferability of the models was also tested.

The unsupervised classification algorithm was validated in Fields 1 and 2 using the training datasets created for the RF models generated in September, when *E. elaterium* was the only weed in the olive orchards.

3 RESULTS

3.1 Image segmentation optimization

The optimization process of the segmentation parameters using GA yielded values of 0.78, 0.93, 0.21, and 62.55 for the compactness weight, spectral weight, merging threshold, and minimum segment size, respectively. These segmentation parameters were applied to the orthomosaics from both fields and dates.

3.2 Random Forest model

Due to a small area that presented some artifacts in the orthomosaic and DSM, one out of the 90 training plots from Field 1 in September was not used. Table 2 shows the number of objects manually labeled in the creation of the training datasets, detailing the number of objects for the classes associated with weeds, which were the less represented classes. In the balanced datasets created from the complete training datasets, the number of samples by class was equal to the number of objects for the less represented class. Consequently, in Field 1, there were 113 and 137 objects by class in May and September, respectively, and in Field 2, there were 111 objects for each class in May and 43 in September.

Table 3 shows the results achieved from the RF models created for analyzing the olive orchards in May, the month when there were different weed species in the study fields. The OA was equal to or higher than 0.89 in all cases. PA for *E. elaterium* ranged from 0.75 to 0.93, while UA was lower, yielding values between 0.45 and 0.86. In all the models, PA values for *E. elaterium* were higher when evaluated in the OOB data than when the models were validated in other fields. UA followed the same trend except in the model for the unbalanced training dataset in Field 2 where the UA for *E. elaterium* was higher when the model was tested in other fields.

Regarding the balance of the dataset, OA was always equal to or higher for the complete unbalanced dataset. There was not a clear trend for *E. elaterium* PA and UA, which were sometimes higher for the unbalanced dataset and sometimes higher for the balanced dataset.

Although in September, when *E. elaterium* was the only infesting weed, an automatic classification algorithm with no need of training data could be applied, RF models were also created and validated to compare the automatic approach with a supervised methodology. RF models for the classification of 'Soil', 'Olive', 'Shadow' and '*E. elaterium*' were created using data from May and September. Apart from the OOB validation, the RF models created from the May dataset in one field were validated using data from September in both fields, and the RF models created with data from September in one field were validated in September data from the other field. Table 4 shows the resulting accuracy metrics achieved in all these cases.

Table 2. Number of objects manually labeled for the training datasets, with detailed data for the objects belonging to the less represented classes in the complete training dataset

Date	Field	Complete dataset	*E. elaterium*	Other weeds	Balanced dataset
May	**1**	14 068	271	113	565
	2	15 392	111	267	555
September	**1**	17 406	137	*NA*	548
	2	14 909	43	*NA*	172

Table 3. Accuracy metrics for the random forest models in the scenarios with *E. elaterium* and other weeds (May)

Training field	Balanced training dataset	Validation field	Overall accuracy	Producer's accuracy *E. elaterium*	User's accuracy *E. elaterium*
1	No	OOB*	0.97	0.89	0.84
		2	0.96	0.75	0.61
	Yes	OOB	0.89	0.91	0.82
		2	0.96	0.75	0.69
2	No	OOB	0.98	0.93	0.73
		1	0.95	0.79	0.80
	Yes	OOB	0.92	0.91	0.86
		1	0.89	0.86	0.45

*OOB: based on out-of-bag data from the same field.

Table 4. Accuracy metrics for the random forest models in the scenarios with *E. elaterium* as the only infesting weed (September)

Training field	Balanced training dataset	Training date	Validation field	Overall accuracy	Producer's accuracy *E. elaterium*	User's accuracy *E. elaterium*
1	No	May	OOB*	0.97	0.94	0.86
			1	0.95	0.75	0.94
			2	0.97	0.95	0.89
		September	OOB	0.97	0.97	0.82
			2	0.98	0.98	0.91
	Yes	May	OOB	0.96	0.98	0.97
			1	0.95	0.87	0.44
			2	0.97	0.98	0.51
		September	OOB	0.95	0.96	0.95
			2	0.97	1.00	0.22
2	No	May	OOB	0.99	0.92	0.82
			1	0.99	0.77	0.77
			2	0.98	0.95	0.84
		September	OOB	0.98	0.95	0.91
			1	0.96	0.66	0.94
	Yes	May	OOB	0.97	0.98	0.97
			1	0.90	0.95	0.16
			2	0.96	0.98	0.22
		September	OOB	0.93	0.98	0.95
			1	0.95	0.80	0.38

*OOB: based on out-of-bag data from the same field.

In the OOB validation, the OA values were always equal to or higher than 0.93, achieving better results in the complete unbalanced dataset than in the balanced dataset. PA and UA for *E. elaterium* were over 0.80 in all cases and were higher in the balanced dataset, with the only exception of PA in the model created for Field 1 using data from September.

Table 5. Accuracy metrics for the automatic classification algorithm in the scenarios with *E. elaterium* as the only infesting weed (September)

Field	Overall accuracy	Producer's accuracy *E. elaterium*	User's accuracy *E. elaterium*
1	0.93	0.74	0.76
2	0.95	0.91	0.91

When applying the RF models created from the September dataset (*E. elaterium* as the only weed) in other fields and dates, OA values were equal to or higher than 0.9, PA for *E. elaterium* ranged from 0.66 to 1.00 with most of the values over 0.75, and UA for *E. elaterium* was between 0.16 and 0.94. OA values were very similar for the models created with the balanced and unbalanced datasets, although slightly higher for the latter one. Regarding the accuracy metrics for *E. elaterium*, although PA was higher for the models created with balanced data, the models created with the complete unbalanced dataset yielded better UA values.

3.3 Unsupervised classification algorithm

Table 5 shows the validation results of the unsupervised algorithm in both fields in September, when there were no other weeds apart from *E. elaterium*. All the accuracy metrics were over

0.70, and although OA values were similar in both fields, PA and UA for *E. elaterium* were higher in Field 2.

4 DISCUSSION

The optimization of segmentation parameters is of crucial importance in OBIA. However, according to a semisystematic review conducted by Gonçalves *et al.*,[13] 44% of scientific studies using OBIA do not explain how they selected these parameters. Among some recent papers using RF and OBIA for weed detection,[7,25,26] the present work is the only one that has applied a methodology for the optimization of all the segmentation parameters, while de Castro *et al.*[25] optimized the scale using ESP[12] but not the other segmentation parameters, and Michez *et al.*[26] only tested three scale parameter values.

The RF models created in this work were very accurate, with OA values ranging from 0.89 to 0.99. Furthermore, this accuracy was achieved regardless of whether they were validated on OOB samples or in other fields and dates. This fact implies that the RF models were robust and transferable. RF models created with training data from one field and date yielded high OA values when tested in other fields and dates. As an example, the RF model created for the detection of *E. elaterium* as the only infesting weed using data from Field 1 in May achieved an OA of 0.97 when validated in data from Field 2 in September. This is important because it means that it would not be necessary to create a model with manually labeled samples for each field and date. For example, in a practical use for *E. elaterium* management in an olive farm, it would be rather sufficient to create one model and it could be applied to the whole farm at any phenological moment if the imagery is acquired under similar light conditions, and flight and camera configurations. Although it has not been tested, the use of another UAV platforms or sensors should not affect the accuracy and transferability of the proposed methodology as long as parameters such as spatial resolution and spectral range of the images remain similar to the ones used in the present work.

The OA values yielded by the RF models created in the present work are in line with those obtained in UAV-based weed detection works by other authors. Michez *et al.*[26] and Gao *et al.*[7] reached OA values of 0.97 and 0.945 for weed mapping using OBIA and RF and using proprietary software, respectively. Other works that applied OBIA in open-source environments for weed detection reached OA values of 0.92 and 0.99, respectively.[3,21]

Moving the point of view from the OA to the accuracy achieved in the classification of *E. elaterium*, which is the agronomical leitmotif of this work, the RF models also achieved high accuracy values. In the most complicated scenario, *i.e.*, when there were different weed species in the olive orchard, the RF models achieved PA values for *E. elaterium* ranging from 0.75 to 0.93. This fact implies that the vast majority of the *E. elaterium* plants was detected. Most of the UA values for this class were over 0.70 when there were different weed species, which means that most of the objects classified as *E. elaterium* actually belonged to this class. Considering the high PA values, even lower UA values implying an overestimation of the weed presence could be accepted by olive growers, which would prefer this overestimation rather than missing the detection of *E. elaterium* plants that could escape from the weed treatment and act as a source of seeds.

In the analysis of the images from September (Table 4), when *E. elaterium* was the only infesting weed, the RF models also achieved high PA values for the detection of this weed. On average, these RF models allowed the detection of 91% of the objects labeled as *E. elaterium*. UA values for this weed were over 0.80 in most cases, although there were UA values below 0.25 in three RF models created with balanced data. The lower UA values were accompanied by high PA values, which imply that there was an overestimation of the *E. elaterium* infestation. According to the confusion matrices (data not shown), this overestimation was caused by shadow and soil objects being classified as *E. elaterium*. Maybe in this balanced dataset from field 2 in September (including less samples), the heterogeneity of objects belonging to *E. elaterium*, shadow and soil was underrepresented and caused the confusion among these classes. In any case, even in this situation, an adequate control of *E. elaterium* could be achieved since the RF models presented high PA values and consequently missed a small amount of weed plants.

Regarding the influence of balanced training datasets on the PA and UA values for *E. elaterium* in the RF models, there was not a clear influence of the balance of the samples in the models created for the multiweed scenario. These accuracy metrics were sometimes higher for the unbalanced dataset and sometimes higher for the balanced dataset. However, when *E. elaterium* was the only infesting weed, the UA values for this class were always

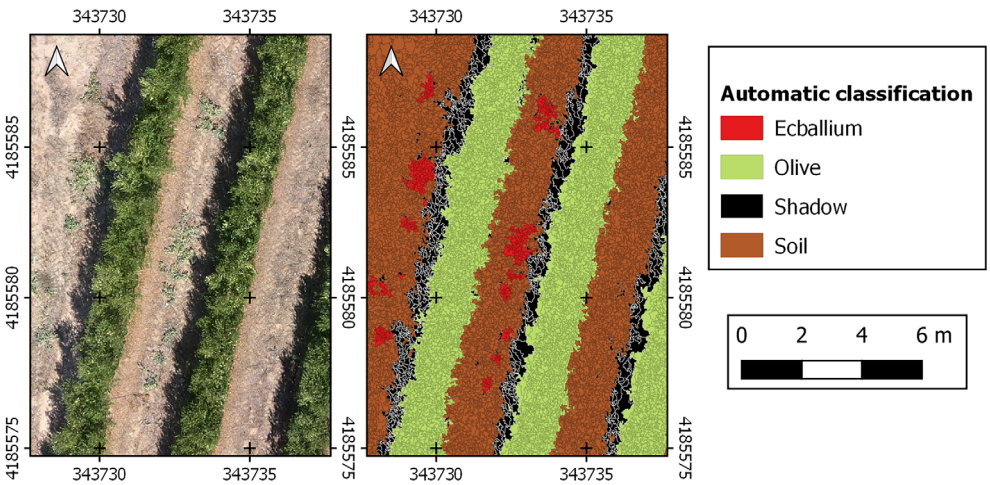

Figure 5. Detailed view of the orthomosaic (left) and the infestation map (right) created applying the automatic algorithm to Field 2 in September. Coordinate system: WGS84 UTM 30 N.

higher for the models created with the full dataset without class balancing, while OA and PA were similar regardless of the balance of the training data. The better results achieved for UA with the unbalanced dataset contrast with the general recommendation of using balanced training datasets for the use of RF in remote sensing,[16] and this fact should be more deeply studied in future research. Previous works devoted to weed detection using RF and OBIA. de Castro *et al*.[22] and Michez *et al*.[26] used balanced training data, while Gao *et al*.[7] did not balance the training dataset. However, none of these works compared the results of RF models created with balanced and unbalanced data.

The unsupervised classification algorithm created for the detection of *E. elaterium* when there were no other weeds reached high accuracy metrics, similar to those achieved by the supervised RF models, but with the added value of not needing the manual creation of a training dataset. The only user intervention required for using the automatic algorithm is the creation of a small training dataset for the optimization of the segmentation parameters using SegOptim package. However, this previous step is only required the first time that the algorithm is going to be used. Once the segmentation parameters are defined, they can be recursively used in a potential commercial application as long as the UAV imagery acquisition parameters (sensor, flying height, overlaps) remain stable. Furthermore, this previous step is also needed for the supervised classification approach. Consequently, and due to its lower requirement of user intervention, the unsupervised classification algorithm is recommended in olive orchards affected only by *E. elaterium*. As this is a hard-to-control weed,[1,2] it could also be used in fields where the rest of the weeds have been eradicated by previous weed treatments and the farmer is looking for the detection and treatment of the remaining *E. elaterium* plants. This would be the case for the studied olive orchards, where there were different weed species in May that were effectively controlled by the corresponding treatments applied on the farm, but *E. elaterium* tolerated these treatments and was still present in the fields in September.

The results achieved by the RF models and the unsupervised algorithm developed in the present work allow the generation of accurate infestation maps, such as the one shown in Fig. 5. These maps representing the location and extension of the *E. elaterium* patches open the door to the creation of site-specific weed treatments specifically designed for eradicating this hard-to-control weed. Furthermore, as stated by Blank *et al*.,[1] the knowledge generated about the spatial distribution in patches and patch temporal stability of *E. elaterium* could be useful not only for the understanding of the species ecology, which is vital for the design of appropriate weed management strategies, but also because this tactic would not require new yearly mapping but a relatively low frequency mapping (*e.g*. every few years) which is relevant in woody crops.

The use of both free and open-source software for the development of the *E. elaterium* classification algorithms together with inexpensive UAVs and embedded sensors is on line with the recent trend of developing affordable weed mapping systems reported by other works, such as Lam *et al*.[21] and Mattivi *et al*.[3] The use of low-cost technologies and methodologies can represent a boost for the adoption of precision agriculture techniques, as the economic costs have been identified as a barrier to the implementation of these technologies both in developed regions such as Europe and in developing countries such as India.[18,19]

The present work is, to the authors' knowledge, the first one to address weed mapping in olive orchards by remotely sensed imagery. Most other works on remote sensing for SSWM have focused on horticultural or arable crops, with few exceptions dealing with weed detection in woody crops,[5,6] although none in olive orchards. Therefore, the present paves the way to the implementation of SSWM in olive orchards under superintensive system, which has a very high added value due to its high productivity and precocity, with the consequent boost to crop sustainability by rationalizing the use of herbicide treatments in line with the EU Green Deal. In order to achieve an effective implementation of SSWM in olive orchards, together with weed detection between olive hedgerows, it is necessary to carry out research on weed detection under the canopy of olive trees (*i.e*. within the row). In this line of research, some preliminary work has been published on the use of ground sensors for the detection of *Conyza* spp. growing under olive hedgerows,[37] which cannot be detected by aerial remote sensing.

5 CONCLUSIONS

The present work presents a robust methodology for the detection of *E. elaterium*, a problematic weed that is expanding its infestations in the interrow area of olive and other woody crops. To the best of the authors' knowledge, this is the first time that remote sensing has been used for weed mapping in olive orchards. The developed workflow was based on low-cost UAV-RBG imagery and OBIA algorithms developed in open-source software. The classification of *E. elaterium* relied on RF models in scenarios with multiple weed species and on an unsupervised process when *E. elaterium* was the only infesting weed in hedgerow olive orchards. The proposed workflow was created and validated in commercial orchards with natural infestations, which hardened the achievement of the proposed objective in comparison with its development under controlled conditions in experimental fields. The accuracy of the classification algorithms allows the creation of site-specific treatment maps for increasing the sustainability of weed management. Therefore, the development of these algorithms in free and open-source software fosters their application in small and medium farms where economic cost is a barrier to the implementation of precision agriculture and digitizing techniques.

ACKNOWLEDGEMENTS

The authors thank ELAIA S.A. for allowing developing the field work and UAV flights in its olive orchards. We also thank Juan Carlos Cañasveras, Rubén Bujaldón and Antonio J. Cuevas for their technical assistance and help in the field work. This research was funded by MCIN/AEI/10.13039/501100011033 (project: PID2020-113229RB-C44).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1 Blank L, Birger N and Eizenberg H, Spatial and temporal distribution of *Ecballium elaterium* in almond orchards. *Agronomy* **9**:751 (2019).
2 Saavedra M, Características y control del pepinillo del diablo: *Ecballium elaterium* (L.) Richard, Junta de Andalucía, Consejería de Agricultura y Pesca, Sevilla (2000).
3 Mattivi P, Pappalardo SE, Nikolić N, Mandolesi L, Persichetti A, De Marchi M *et al*., Can commercial low-cost drones and open-source

GIS technologies be suitable for semi-automatic weed mapping for smart farming? A case study in NE Italy. *Remote Sens* **13**:1869 (2021).
4 Torres-Sánchez J, Mesas-Carrascosa FJ, Jiménez-Brenes FM, de Castro AI and López-Granados F, Early detection of broad-leaved and grass weeds in wide row crops using artificial neural networks and UAV imagery. *Agronomy* **11**:749, Multidisciplinary Digital Publishing Institute (2021).
5 de Castro AI, Peña JM, Torres-Sánchez J, Jiménez-Brenes FM, Valencia-Gredilla F, Recasens J *et al.*, Mapping *Cynodon dactylon* infesting cover crops with an automatic decision tree-OBIA procedure and UAV imagery for precision viticulture. *Remote Sens* **12**:56 (2020).
6 Jiménez-Brenes FM, López-Granados F, Torres-Sánchez J, Peña JM, Ramírez P, Castillejo-González IL *et al.*, Automatic UAV-based detection of *Cynodon dactylon* for site-specific vineyard management. *PLoS One* **14**:e0218132 (2019).
7 Gao J, Liao W, Nuyttens D, Lootens P, Vangeyte J, Pižurica A *et al.*, Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *Int J Appl Earth Obs Geoinformation* **67**:43–53 (2018).
8 López-Granados F, Torres-Sánchez J, Castro A-ID, Serrano-Pérez A, Mesas-Carrascosa F-J and Peña J-M, Object-based early monitoring of a grass weed in a grass crop using high resolution UAV imagery. *Agron Sustainable Dev* **36**:67 (2016).
9 Peña JM, Torres-Sánchez J, de Castro AI, Kelly M and López-Granados F, Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. *PLoS One* **8**: e77151 (2013).
10 Pérez-Ortiz M, Peña JM, Gutiérrez PA, Torres-Sánchez J, Hervás-Martínez C and López-Granados F, Selecting patterns and features for between- and within- crop-row weed mapping using UAV-imagery. *Expert Syst Appl: Int J* **47**:85–94 (2016).
11 Zisi T, Alexandridis TK, Kaplanis S, Navrozidis I, Tamouridou A-A, Lagopodi A *et al.*, Incorporating surface elevation information in UAV multispectral images for mapping weed patches. *J Imaging* **4**: 132 (2018).
12 Drăguț L, Tiede D and Levick SR, ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int J Geogr Inf Sci* **24**:859–871 (2010).
13 Gonçalves J, Pôças I, Marcos B, Mücher CA and Honrado JP, SegOptim—A new R package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data. *Int J Appl Earth Obs Geoinformation* **76**:218–230 (2019).
14 Peña JM, Torres-Sánchez J, Serrano-Pérez A, de Castro AI and López-Granados F, Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors* **15**:5609–5626 (2015).
15 Lottes P, Khanna R, Pfeifer J, Siegwart R, and Stachniss C, UAV-based crop and weed classification for smart farming, presented at the IEEE international conference on robotics and automation (ICRA), 2017, Singapore.
16 Belgiu M and Drăguț L, Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* **114**:24–31 (2016).
17 Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M and Rigol-Sanchez JP, An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogramm Remote Sens* **67**:93–104 (2012).
18 Barnes AP, Soto I, Eory V, Beck B, Balafoutis A, Sánchez B *et al.*, Exploring the adoption of precision agricultural technologies: a cross regional study of EU farmers. *Land Use Policy* **80**:163–174 (2019).
19 Mondal P and Basu M, Adoption of precision agriculture technologies in India and in some developing countries: scope, present status and strategies. *Prog Nat Sci* **19**:659–666 (2009).
20 Moreno-Sanchez R, Free and open source software for geospatial applications (FOSS4G): a mature alternative in the geospatial technologies arena. *Trans GIS* **16**:81–88 (2012).
21 Lam OHY, Dogotari M, Prüm M, Vithlani HN, Roers C, Melville B *et al.*, An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using *Rumex obtusifolius* as a case study. *Eur J Remote Sens* **54**:1–18 (2020).
22 de Castro AI, Jiménez-Brenes FM, Torres-Sánchez J, Peña JM, Borra-Serrano I and López-Granados F, 3-D characterization of vineyards using a novel UAV imagery-based OBIA procedure for precision viticulture applications. *Remote Sens* **10**:584 (2018).
23 R Core Team, R: A language and environment for statistical computing, Vienna, Austria (2019).
24 Baatz M and Schäpe A, Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation, in *Angewandte Geographische Informations-Verarbeitung XII*, ed. by Strobl J, Blaschke T and Griesbner G. Wichmann Verlag, Karlsruhe, Germany, pp. 12–23 (2000).
25 de Castro AI, Torres-Sánchez J, Peña JM, Jiménez-Brenes FM, Csillik O and López-Granados F, An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. *Remote Sens* **10**:285 (2018).
26 Michez A, Piégay H, Jonathan L, Claessens H and Lejeune P, Mapping of riparian invasive species with supervised classification of unmanned aerial system (UAS) imagery. *Int J Appl Earth Obs Geoinformation* **44**: 88–94 (2016).
27 Karl JW and Maurer BA, Spatial dependence of predictions from image segmentation: a variogram-based method to determine appropriate scales for producing land-management information. *Ecol Inf* **5**: 194–202 (2010).
28 Woebbecke DM, Meyer GE, Von Bargen K and Mortensen DA, Color indices for weed identification under various soil, residue, and lighting conditions. *Trans ASAE* **38**:259–269 (1995).
29 Gitelson AA, Kaufman YJ, Stark R and Rundquist D, Novel algorithms for remote estimation of vegetation fraction. *Remote Sens Environ* **80**: 76–87 (2002).
30 Lindsay JB and Whitebox GAT, A case study in geomorphometric analysis. *Comput Geosci* **95**:75–84 (2016).
31 Torres-Sánchez J, López-Granados F, Serrano N, Arquero O and Peña JM, High-throughput 3-D monitoring of agricultural-tree plantations with unmanned aerial vehicle (UAV) technology. *PLoS One* **10**:e0130479 (2015).
32 Liaw A and Wiener M, Classification and regression by randomForest. *R News* **2**:18–22 (2002).
33 Otsu N, A threshold selection method from gray-level histograms. *IEE Trans Syst Man Cybern* **9**:62–66 (1979).
34 Breiman L, Random forests. *Mach Learn* **45**:5–32 (2001).
35 Peters J, Baets BD, Verhoest NEC, Samson R, Degroeve S, Becker PD *et al.*, Random forests as a tool for ecohydrological distribution modelling. *Ecol Model* **207**:304–318 (2007).
36 Mitchell MW, Bias of the random forest out-of-bag (OOB) error for certain input parameters. *Open J Stat* **01**:205–211 (2011).
37 Pérez-Porras F, Torres-Sánchez J, López-Granados F, Cantón S, and Mesas-Carrascosa FJ, Detección de *Conyza* spp. en la hilera de cultivos leñosos: caso de estudio en olivar en seto, presented at the XVIII Congreso de la Sociedad Española de Malherbología, 26 April 2022, Mérida, Spain, pp 427–432.

