A brave new world for archaeological survey: Automated machine learning-based potsherd detection using high-resolution drone imagery

H.A. Orengo[a,*], A. Garcia-Molsosa[b]

[a] *Ramón y Cajal Researcher, Catalan Institute of Classical Archaeology, Spain*
[b] *Marie Skłodowska-Curie Fellow, McDonald Institute for Archaeological Research, University of Cambridge, UK*

ABSTRACT

Archaeological pedestrian survey is one of the most popular techniques available for primary detection of archaeological sites and description of past landscape use. As such it is an essential tool not just for the understanding of past human distribution, economy, demography and so on but also for cultural heritage management and protection. The most common type of pedestrian surface survey consists of fieldwalking relatively large tracts of land, recording the dispersion of items of material culture, predominantly pottery fragments, by teams of archaeologists and students. This paper presents the first proof of concept for the automated recording of material culture dispersion across large areas using high resolution drone imagery, photogrammetry and a combination of machine learning and geospatial analysis that can be run using the Google Earth Engine geospatial cloud computing platform. The results show the potential of this technique, under appropriate field circumstances, to produce accurate distribution maps of individual potsherds opening a new horizon for the application of archaeological survey. The paper also discusses current limitations and future developments of this method.

1. Introduction

Pedestrian survey is one of the most important and fundamental data collection techniques in the archaeologist's toolbox. Together with remote sensing and geophysics, is one of the most popular and reliable approaches for the detection and characterisation of archaeological sites and, with different intensity and methodological approaches, it has been conducted from the beginning of the discipline in the 19th century. Archaeological survey is one of the few methods we possess to quantitatively evaluate human past occupation and dispersion at a large scale and, therefore, the relevance of the data it provides goes well-beyond archaeology. Today, "fieldwalking" usually involves systematic walking by teams of archaeologists to record surface-visible material culture, usually pottery fragments ("sherds"), and analysing the dispersion of these datable relics of human presence provide insights into changing landscape use and settlement intensity. Archaeological survey developed as a discipline since the 1950–60s and rapidly increased its application in important archaeological areas, such as Israel with 394 surveys from the period between 1989-98 (Kletter and De-Groot, 2001) and Greece with almost a hundred from 1975 to 1999 (Alcock and Cherry, 2004). Fish (1999), writing in 1997, documents a ten-fold increase in the percentage of articles related to settlement patterns

published in leading journals since the 1960s. Alcock and Cherry (2004) estimate that millions of hectares of the Mediterranean alone had been surveyed by the beginning of the 21st century with a marked increase in survey activity since then.

By the late 1970s intensive archaeological field survey had become a systematic and quantitative technique and the number and density of potsherds started to be recorded in detail (e.g. Cherry et al., 1978; Crowther, 1983; Shennan, 1985). GIS is nowadays routinely employed to plan the survey but also to represent and analyse the resulting data (e.g. Lock et al., 1999; Gillings and Sbonias, 1999; Bevan and Conolly, 2004). Obtaining well-distributed quantitative potsherd distribution data over a landscape usually requires large teams of surveyors or long periods of time. They usually walk following parallel lines or a grid to record and collect all or a selection of sherds from the field, which is a ploughsoil under ideal circumstances but it can include many types of surfaces and vegetation. All areas or a sample of them (usually in the form of transects) presenting good ground visibility are surveyed to offer an estimation of the past human occupation of the study area. Archaeological field survey, just as archaeological excavation, implies a strong investment as many people (who require accommodation and sustenance) are needed during several field seasons to cover a relatively large area. The need to store large quantities of pottery sherds (large

surveys can collect several millions of these) can also be expensive and require the use of large facilities for as many years as required for their study (Schiffer et al., 1978: 15, Alcock, 2000). Post-processing and study of the collected material and generation of ceramic dispersion maps also requires a strong investment in time and effort as the analysis and classification of such large quantities of pottery sherds can last for years.

Archaeological survey forms an important part of all commercial, public and research-based archaeology. In many countries it is a requirement for development projects to survey the affected area in order to identify the presence of possible archaeological sites and other cultural elements that might be in danger of being destroyed during the construction process (Schiffer et al., 1978). The pressures of salvage survey and the need for a speedy and accurate recording are strikingly described by Wendorf (1962: 54): "The archaeological teams follow as closely behind the surveyors and as far ahead of the right-of-way clearing machinery as possible. Even under ideal conditions the timing will still be close, and there may not be more than three to four weeks between the survey and dozer clearing the right-of way."

This paper aims to provide a proof of concept for a method to detect multi-period surface ceramic distributions, that augments traditional data-collection strategies and provides new ways to approach some of the shortcomings of walker-based recording. The proposed method is able to record individual potsherds across large survey areas using a semi-automated workflow. It is important to note here that this method does not aim to substitute archaeological fieldwalking but complement much of the non-specialist work conducted by groups of people for long periods of time in conductive environments so there is more time and resources available to dedicate to specialised work. The results of the test described here have the potential to revolutionise one of the most basic archaeological techniques when applied in favourable settings. In this way, it could reduce primary data collection costs while substantially increasing the survey intensity and the reliability of the results and provide important information to facilitate material collection and analysis.

The workflow relies on the combination of a series of recent independent technological developments:

1. In the first place drones or Unmanned Aerial Vehicles have become significantly cheaper while their technical capacities have increased considerably, in particular, flight time, remote control reach, flight planning, stability, image quality, sensor-based location and obstacle avoidance. These now allow enough flight time to record larger areas flying at very low altitude following a planned route while taking continuous overlapping pictures of enough resolution to clearly provide very high-resolution imagery of archaeological features (Stek, 2016; Orengo and Knappett, 2018).

2. Digital photogrammetry software has become much easier to use and accessible with the implementation of semi-automated co-registration, point cloud, surface and texture generation workflows and the widespread distribution of low-cost and open source photogrammetry software. It has also become increasingly powerful with the adoption of new techniques, such a structure-from-motion. In consequence its use has been incorporated to a large number of archaeological workflows (Grün et al., 2004; Verhoeven et al., 2012; Orengo, 2013; Orengo et al., 2015).

3. Machine learning is a branch of artificial intelligence that automates analytical model building and can learn from classified data improving its analytical capabilities with minimal human intervention. Machine Learning applications have also had an important development during the last years becoming a common option for data mining and analysis, pattern identification and decision making. Although, its application to archaeology is still limited (but see Menze and Ur 2012, Oonk and Spijker, 2015, Liss et al., 2017, Orengo et al., 2019 (submitted)) its potential has long been recognised (van der Maaten et al., 2007).

Fig. 1. Location of APAX survey area and plots 36 and 591 within it.

4. The last years have also seen an important development of widely-accessible cloud computing services including Amazon AWS and Google Cloud Platform. These offer the possibility to use parallel and distributed computing, large storage space and incorporate processing services, such as data analytics and machine learning. These allow the extremely intensive computation power necessary for the development of large scale and intensive analyses and are rapidly being integrated into archaeological research (e.g. Agapiou, 2017, Orengo and Petrie, 2017, 2018, Rayne et al., 2017, Garcia et al., 2019).

2. Case study

We selected two fields, plot 36 and plot 591 (Fig. 1), from the on-going survey developed by the Archaeological Project at Abdera and Xanthi (APAX) in which the authors are currently involved (Georgiadis et al. in preparation). APAX is a project directed by the Ephoria (archaeological service) of Xanthi (Thrace, north-eastern Greece) in collaboration with researchers of the National and Kapodistrian University of Athens, the Aristotle University of Thessaloniki and the Catalan Institute of Classical Archaeology. APAX aims to study the changing occupation patterns of the Archaic and Classical city of Abdera (ca. 7th to 3rd C BC) while achieving a good knowledge of the distribution of archaeological sites around it in order to: a) provide a better understanding of long-term settlement patterns and historical change, and b) facilitate the protection and promotion of Xanthi's heritage. Intensive survey was considered as the most adequate technique to achieve these objectives and, consequently, it was conducted during the summers of

the years 2016–18 for a duration of five weeks each. The survey is still under development but these three fieldwork seasons have resulted in the recording of around 580,000 fragments of pottery of which 41.000 have been collected and classified. Survey is conducted during summer months, which is the only period in which researchers and students, who participate in the survey campaign as part of their training in archaeology, can dedicate enough time for the development of fieldwork. This forced timetable is problematic because the highest visible for archaeological material is after plots have been ploughed, which in this area is variable according to cultivation type.

The plots selected for the pilot study were done so on the basis of being representative of field conditions during a typical fieldwork season at Abdera. These are ploughsoils that are cultivated periodically and correspond well to the common survey choice in Mediterranean field survey. Rather than plots with good ground visibility and ideal soil conditions for ceramic detection we selected those in which visibility was relatively low and the survey conditions were not the best available. The selected fields had not been recently ploughed (which would have increased the visibility of potsherds), show a high presence of vegetation, stones and shadows (all elements that result in a reduction of the efficacy of the detection) and the sediment present hues similar to those of ceramics. This selection intended to evaluate the success of the technique by testing it under less than perfect circumstances to ensure its application to the widest possible range of ploughsoil conditions.

The plots are located on each of the two areas in which the project is divided. In the urban sector (plot 36) the specific objective of the intensive survey was to map the relative density of pottery distribution across the city, in order to analyse changing uses of the urban space. Plot 591 is part of a transect located to the north of the ancient town. This area contains remains of both the cemeteries and the rural settlement linked to the town. Outside the urban area, the specific objectives were the detection of pottery concentrations, which could be differentiated from the ceramic background noise present thorough the area. The standard survey method adopted by APAX consisted in the laying of parallel lines separated by a distance of 5 m inside the urban area and 10 m outside. Surveyors walked these lines counting all visible ceramic fragments and collecting only those particular sherds that can provide chronological and typological information.

3. Methods

The pilot study presented here aimed to explore whether an automated pottery sherd recording methodology could be developed to respond to the research questions that motivated the intensive quantitative approach in a similar or more efficient manner (in terms of time and accuracy) than the standard fieldwalking currently employed within the project. To do so a workflow integrating drone high-resolution photography to record in detail the surveyed fields, photogrammetry to join all these photographs in a single orthophotomosaic, machine learning and other geospatial analyses to identify and isolate ceramic fragments in the photomosaic and cloud computing to efficiently process these analysis was developed and tested. In the following paragraphs this workflow (Fig. 2) is presented in detail:

3.1. Drone-based image acquisition

We selected a DJI Phantom 4 Pro v.2.0 to conduct the drone-based image acquisition of the fields under survey. The selection of this model was based on the fact that it falls well within the budget of most archaeological projects ($1499) and its overall quality. Phantom 4 allows for a flight time of 30 min (sufficient to cover a field of 70 m^2 at the required speed and height), is very stable thanks to its GPS and range of sensors, which renders it capable to accurately follow a planned route while avoiding obstacles, and its camera has good quality optics, mechanical shutter (which eliminates rolling shutter distortion of images acquired from moving platforms) and a resolution of 20 MP.

Litchi drone-planning software was employed to set the flight path over the areas of interest. After several tests a height of 3 m above ground was selected, which was considered sufficient to clearly identify ceramics on the field. In order to obtain enough overlap between photographs, which is necessary for the photogrammetry software to correlate them in a single orthomosaic, the parameters defined were: a) a flight speed of 0,61 m/s, b) photographs were automatically taken every 2 s, and c) the flight path was defined using waypoints that formed parallel straight lines separated by a distance of 3 m following the orientation of the fields.

3.2. Pedestrian survey for comparison purposes

Immediately after the flight the fields were intensively surveyed by a team of 4 surveyors (a very experienced team leader and three students with two weeks of field experience). The survey of the fields followed the standard intensive procedure within APAX described above for urban and rural areas: plot 36 was divided using 5 m-apart parallel lines while in plot 591 used 10 m-apart lines.

Subsequently both fields were divided in a 5 × 5 m grid and all pottery sherds in each grid were recorded. This type of total coverage goes well-beyond standard practice given the amount of time that needs to be invested and the little extra information that it could provide in terms of ceramic distribution and concentration. APAX has only developed total coverage of a field (including the collection of materials) in a few cases within small areas in order to check specific types of pottery and their distribution and/or to respond to specific questions related to the use of an area. In this case the objective was to compare the best possible results achievable by traditional survey with those of the automatic detection.

3.3. Photogrammetric processing of drone-acquired photographs

The photogrammetric process included all images obtained from the plots from a single flight to produce an orthophotomosaic, that is, a single image resulting from the merging of all drone-acquired photographs. Agisoft PhotoScan Professional v. 1.2.6 was employed to conduct the orthophotomosaic generation process, which consisted in the alignment of the photographs and the creation of a sparse point cloud. From the cloud point a mesh was generated and employed to produce the orthomosaic joining all photographs, which was exported as an uncompressed tiff file.

3.4. Computational processing of the images

The orthoimages resulting from the photogrammetric process were then uploaded into Google Earth Engine (EE), a web-based geospatial analysis platform with access to Google Cloud parallel computing resources (Gorelick et al., 2017). Earth Engine was considered an ideal environment to conduct the identification of pottery sherds in the orthophotomosaics as it implements machine learning algorithms that can be combined with other geospatial analyses that are useful to increase the reliability of the identification. More importantly it provides access to Google Could computing services which distribute the analysis between multiple computing cores making possible this computation, which would have been far too heavy for any personal computer. Lastly, EE is, upon registration, free to access and use. This allows the algorithm to be reproduced, modified and employed by anyone with their own orthoimages independently of their access to computing resources or coding skills.

EE, however has some limitations ingesting imagery that need to be considered when uploading data. The most evident is the limitation to a maximum size of 10 Gb for each uploaded image. However, many images can be uploaded up to the current storage limit of 250 Gb. This limitation can be partially overcome by uploading the three bands (red, green and blue) of the RGB orthoimage separately and then joining
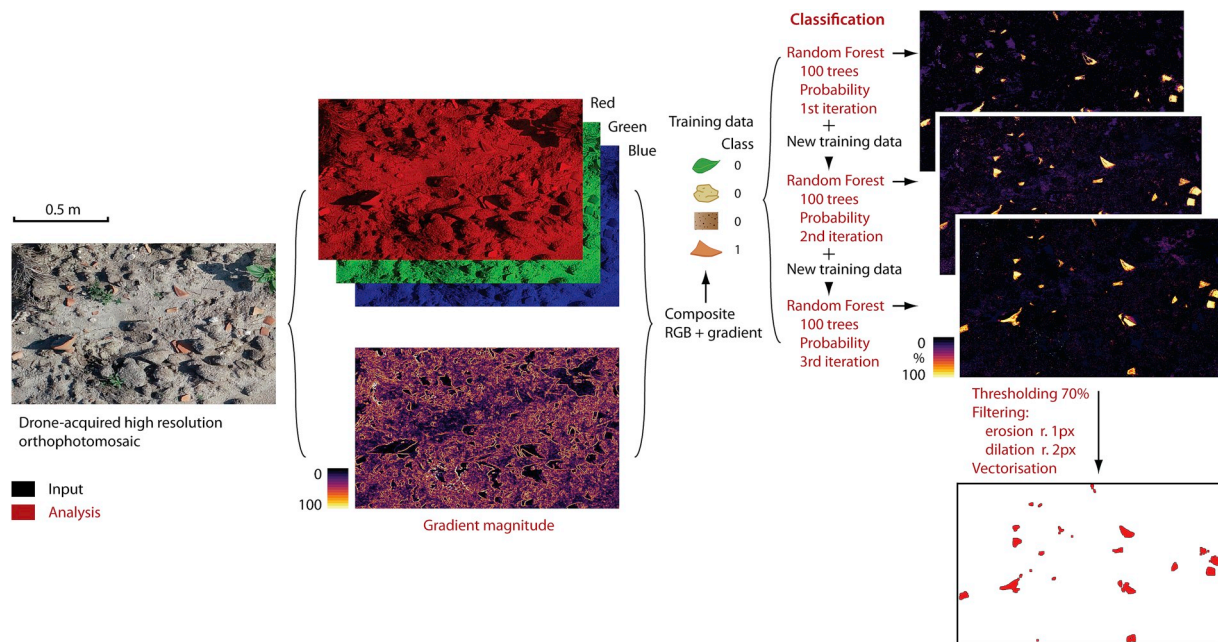
Fig. 2. Computational workflow for the detection and extraction of pottery sherd from drone acquired imagery.

them in a single multiband composite. This would allow uploading an image of a maximum size of 30 Gb, which would correspond to a very large plot of around 75 × 75 m (larger than a football field). The RGB orthoimages of our plots were usually well within EE allowance and they could be ingested without problems. The second limitation is that EE cannot currently ingest subcentimetric images. This was solved scaling up the images prior to their upload.

The algorithm developed for the automatic extraction of potsherds within EE is provided as Supplementary Material to this paper. It was written in EE's implementation of JavaScript as a single script consisting of several parts described below:

1. Import of the already ingested image, which could include the creation of a mosaic or a composite if the image has been divided into parts or the bands ingested separately.
2. Generation of texture and/or gradient analysis. This is an important step as it adds a new layer of information to the RGB composite for the identification of ceramics that is not based on colour but on spatial texture (not to be confounded with the physical texture of pottery, which is far too small to be identified using this pixel size). Pottery sherds usually have soft surfaces (similar values between contiguous pixels along their surfaces). In contrast to this, stone surfaces (although some particular types of geological backgrounds can produce distributions of surface flat stones), bare soil and vegetation generally present rugged surfaces with a much higher degree of variation between contiguous pixel values. Several texture analysis methods were tested (Fig. 3) and compared including entropy (Fig. 3, D), grey-level co-occurrence matrix (Haralick et al., 1973; Conners et al., 1984) and Geary's C (Anselin, 1995) (Fig. 3, E). The standard deviation of pixel values within a kernel was also tested (Fig. 3, F). After several tests we found that a kernel radius of 4 pixels, which equals a kernel size of 9 × 9 pixels (around a centimetre) was the most adequate as it would include a minimum number of pixels from small ceramic fragments while being large enough to identify the inconstant values of pixels in the rest of the field. The results of the different texture analyses were compared to a gradient magnitude analysis, which computes the variation of pixel values along the x and y axes. Best results were obtained by standard deviations (Fig. 3, F) entropy (Fig. 3, D), and gradient magnitude (Fig. 3, C). The last one was selected because a) it is not

dependent on a specific kernel size, which makes it applicable without the need to scale it to different pixel resolutions; b) in contrast to kernel-based methods it does not significantly reduce the area defining ceramic fragments and; c) its lower use of computational resources. The use of texture and gradient analysis could be complemented by the actual topography of the sherds by generating a Digital Surface Model (DSM) directly from the drone images using the same photogrammetric process employed for the generation of the orthophotomosaic. In Fig. 3 (G-L) we present a test of this method only for comparison purposes. The DSM has been processed to extract information about the morphological character and texture of the surface (including ceramics) independently of its colour or the presence of shadows as shown in Fig. 3 (G-H), where potsherds are distinguishable for presenting a flatter surface (less variation in height) than areas around them. While the standard deviation of the high resolution orthomosaic (Fig. 3, I) still provides better results than any of the tested topography-based methods (Fig. 3, J-L), these could potentially offer a complementary band that would help differentiate particularly types of ceramics, such as those presenting decorated surfaces, for which image-based texture analysis alone could not provide satisfactory results. However, the extremely large computational resources and processing time necessary to conduct this analysis prevented us to develop this avenue at this stage as we aim to create a reproducible workflow using standard and freely available tools.

3. Generation of a composite image, which joined the orthophotomosaic's RGB bands and the results of the gradient magnitude analysis. The composite image was employed both to extract training data values, to detect pottery sherds and to test the results of the automated identification.
4. Generation of training data. Training data engineering is nowadays considered to be a complex process requiring expertise and experience. Initial training data were selected by drawing polygons using EE's vector editing tools on top of a) visible ceramic fragments and b) all other elements of the image (bare soil, vegetation, stones, and so on). These two categories were associated to a table column 'class' in which the type of feature was identified, 1 for pottery fragments and 0 for all other elements. The creation of the training data consisted in assigning to each class the values of the pixels delimited by the polygons in each band of the composite.
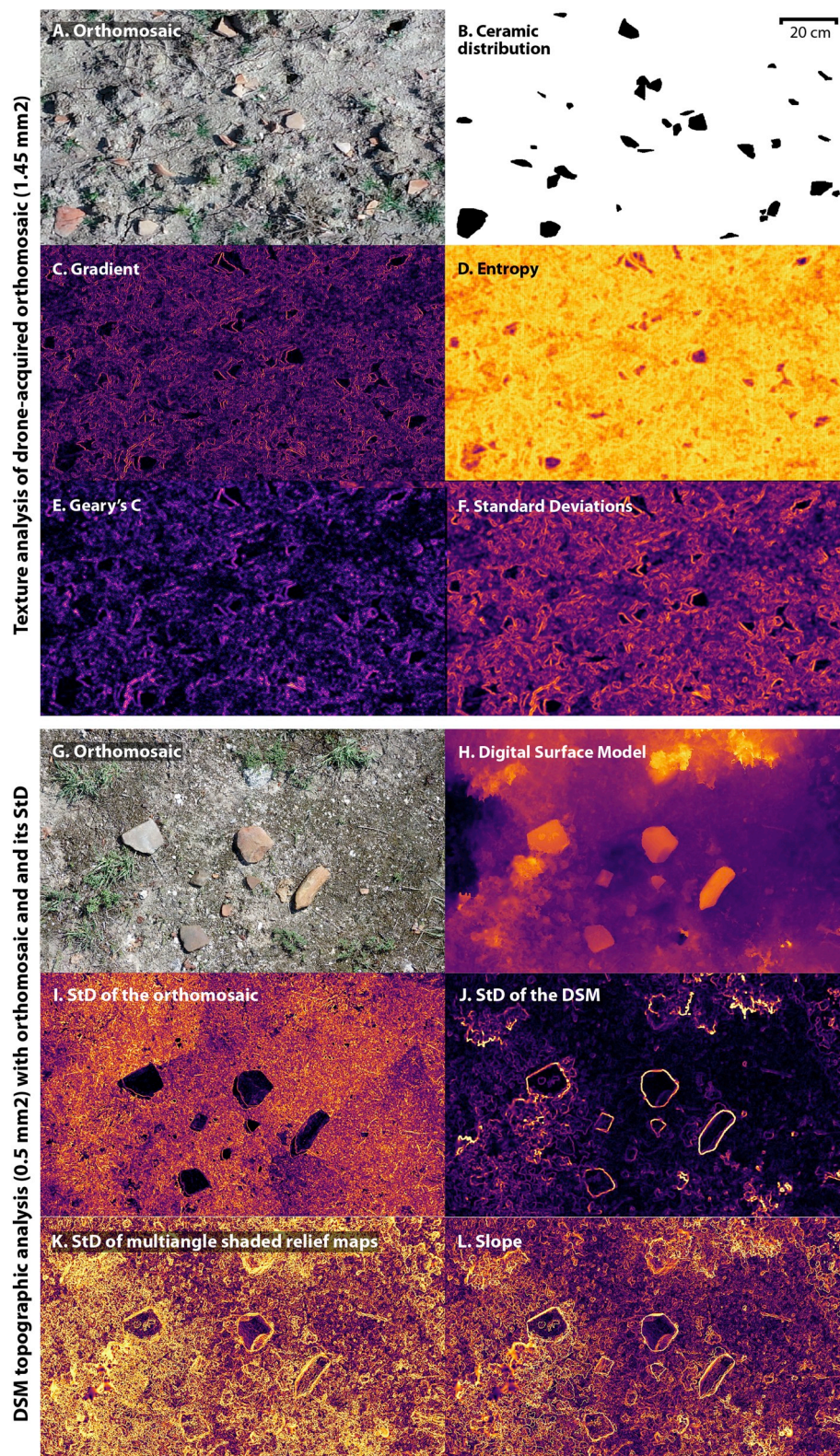
Fig. 3. Comparison of different texture analysis methods and gradient magnitude.

5. Training of the machine learning classifier using the training data. In this case, we selected a Random Forest (RF) classifier (Breiman, 2001). We tested other popular classifiers, such as Support Vector Machines and CART, also available in EE. RF provided superior results consistently, which explains its application to the few archaeological research making use of machine learning applications for image analysis (e.g. Menze and Ur, 2012; Liss et al., 2017). We chose 100 trees considering this is a high enough number to assure optimal results (Oshiro et al., 2012). Given the fact that we were only interested in differentiating pixels belonging to pottery sherds from all other pixels in the image we set the classifier to probability mode. That is, it would output the probability (from 0 to 1) of each

pixel to belong to a pottery sherd.

6. The composite image was then classified using the trained RF classifier. This produced a first classification output. The classification was compared to the orthomosaic to evaluate how it fit visible ceramic fragments. New training polygons were generated to: a) disregard groups of false-positive pixels erroneously identified as having a high probability of belonging to pottery sherds and b) include ceramic pixels that had not been recognised as such. A new iteration of the RF algorithm was then run. The second iteration of the algorithm produced a very high identification rate in both fields; however the third iteration provided only minor improvements.

7. The classification resulting from the third iteration still falsely identified many soil pixels as ceramic fragments. This result is very difficult to avoid as soils can have red-orange colour tonalities similar to those of ceramics. These, however, were disconnected and sparse without forming groups of high percentage pixels as it was common with ceramic sherds. In order to remove these a morphologic filter was developed that would eliminate isolated high-value pixels without affecting large groups of these belonging to sherds. The filter disregarded classification values below 70% and then applied to the rest of pixel values a focal minimum filter with a kernel radius of 1 pixel. This removed 3×3 (around 4–5 mm^2) or smaller pixel groups with values lower than the selected probability threshold. A focal maximum filter was then applied to the results of the first filter in order to: a) recover the pixels eliminated from the edge of grouped pottery pixels, an edge effect of kernel-based filters, such as the focal minimum and b) increase the size of pottery pixels groupings to reduce the number of separated clusters of high probability pixels belonging to a single sherd.

8. The last process in the algorithm was the vectorisation of the filter results, which generated a polygon vector layer in which each feature represented a pottery sherd.

3.5. GIS treatment and generation of outputs

The final vector layer could be exported and incorporated into any standard GIS software for visualisation and further analysis. Treatment in GIS can provide multiple outputs, from the standard density map (without traditional limitations based on sampling) to new measures that could incorporate not just number of ceramics but their size, orientation and the distance between them. With this method it can also be extrapolated ceramic pixel-based density maps to consider very small ceramic fragments that would have not been picked up on the field or the number of vegetation-related pixels in a field (or in grid units).

4. Results: assessing relative time cost of manual vs. automated technique

The results of the automatic recording of surface archaeological material can be best understood if compared with those from current intensive survey methods (Fig. 4 and Table 1). The comparison is offered so the reader can make her/his own approximate calculations. Depending on the type of soil and conditions of the plot, visibility, type and quantity of material culture and number and experience of surveyors these results can vary greatly.

For the automated data collection, in plot 36, 564 photographs were acquired. The photogrammetric process produced an orthoimage with pixel sizes of 1.34 mm^2. Plot 591 required 696 images and the resulting orthomosaic had a ground spatial resolution of 1.45 mm^2. Both resulting orthoimages were larger than the survey grid as a result of the need to include extra rows in all directions to achieve a perfect correlation of photographs and the variability of the drone GPS-based geolocation of the grid which could result in deviations of several meters for the flight, which need to be compensated by increasing the flight area.

For the manual data collection, the intensive survey grid with total recording took 5.5 h, for plot 36 and 4.5 h for plot 591 to record all the pot sherds visible in the surface of each field while the drone flight took around 20 min (see Table 1). In terms of total time spent on the field a survey team could provide faster recording depending on the survey strategy and number of surveyors. However, counting the accumulated time of all four surveyors, drone-based recording was 47–69 times faster than the total recording intensive survey developed in the same plots. This very intensive type of survey recorded 2080 fragments in plot 591, while plot 36 yielded 15779. The automatic recording identified 1597 (76.8% of the sherds counted by total recording) in plot 591 (with a minimum identified potsherd area of 1.3 cm^2) and 5189 (32.9% of the sherds counted by total recording) in plot 36 (with a minimum potsherd area of 1.1 cm^2). The presence of false positives was measured by checking visually 30 of the detected sherds in the orthoimage of each field. Field 591 produced 3 false positives and plot 36 only produced 1.

The standard intensive survey using 5 m-separated surveyors following parallel lines recorded 1071 pottery sherds in plot 36 and took a team of 4 surveyors 45 min. The automated recording method was therefore able to document almost five times more ceramic fragments 9.5 times faster (in accumulated surveyors times) than the standard method.

Plot 591, outside of the ancient town's area, used 10 m-separated lines. The four-people survey team recorded 221 potsherds during the 20 min they took to survey the field. The automatic recording method was therefore able to document 7.2 times more ceramic fragments 3.4 times faster than the standard method.

Photogrammetric processing of the photographs to generate an orthophotomosaic took a 8-processor 64 GB-RAM PC 3.2 h for field 36 and 3.5 for field 591. Although the upload of orthoimages was shorter than 10 min with a fiber optic Internet connection, the ingestion of the orthomosaics in EE (an internal EE transformation process that takes place once these have been uploaded) took around 4 h. In regards to the machine learning process the generation of the first training set took 20 min and the calculation of the first iteration of the whole algorithm a little less than an hour. The third iteration took almost 2 h. While the total processing time from photogrammetric treatment to the generation of ceramic distribution maps in vector format can be estimated to be around 13–15 h, the time in requiring human intervention would be around 2 h depending on the experience of the user.

5. Discussion

Fig. 4 in combination with Table 1 provides a comparison between the total recording survey and the automated recording for this specific test. Despite the fact that more ceramic fragments were recorded using total recording field survey the large amount of time necessary to conduct it makes it unpractical in terms of time-cost compared to representative results. The comparison of the results shows similar distributions, since the automatic recording was able to identify a statistically significant number of pottery sherds (32.9% of 5189 fragments in plot 36 and 76.8% of 1597 in plot 591). While plot 591 offers a clear concentration of ceramics, fragments in plot 36 are much more dispersed and some differences in density can be observed between those resulting from the total count survey and the automatic recording. It is interesting to note here that those grid cells surveyed by experienced surveyors systematically provided much larger numbers than those surveyed by students recently introduced to archaeological survey. This well-known bias (Plog et al., 1978; Schiffer et al., 1978) is clearly visible in Fig. 4 where, in the total recording carried out in plot 36, the experienced surveyor counted the cells of the southern central area of the grid (compare with the density map generated from the automatic recording below). This method produces continuous ceramic fragment dispersions without the need to use a grid and markedly improves potsherd distribution analysis and interpretation. In this regard, plot 36 density map derived from individually detected sherds shows a pattern
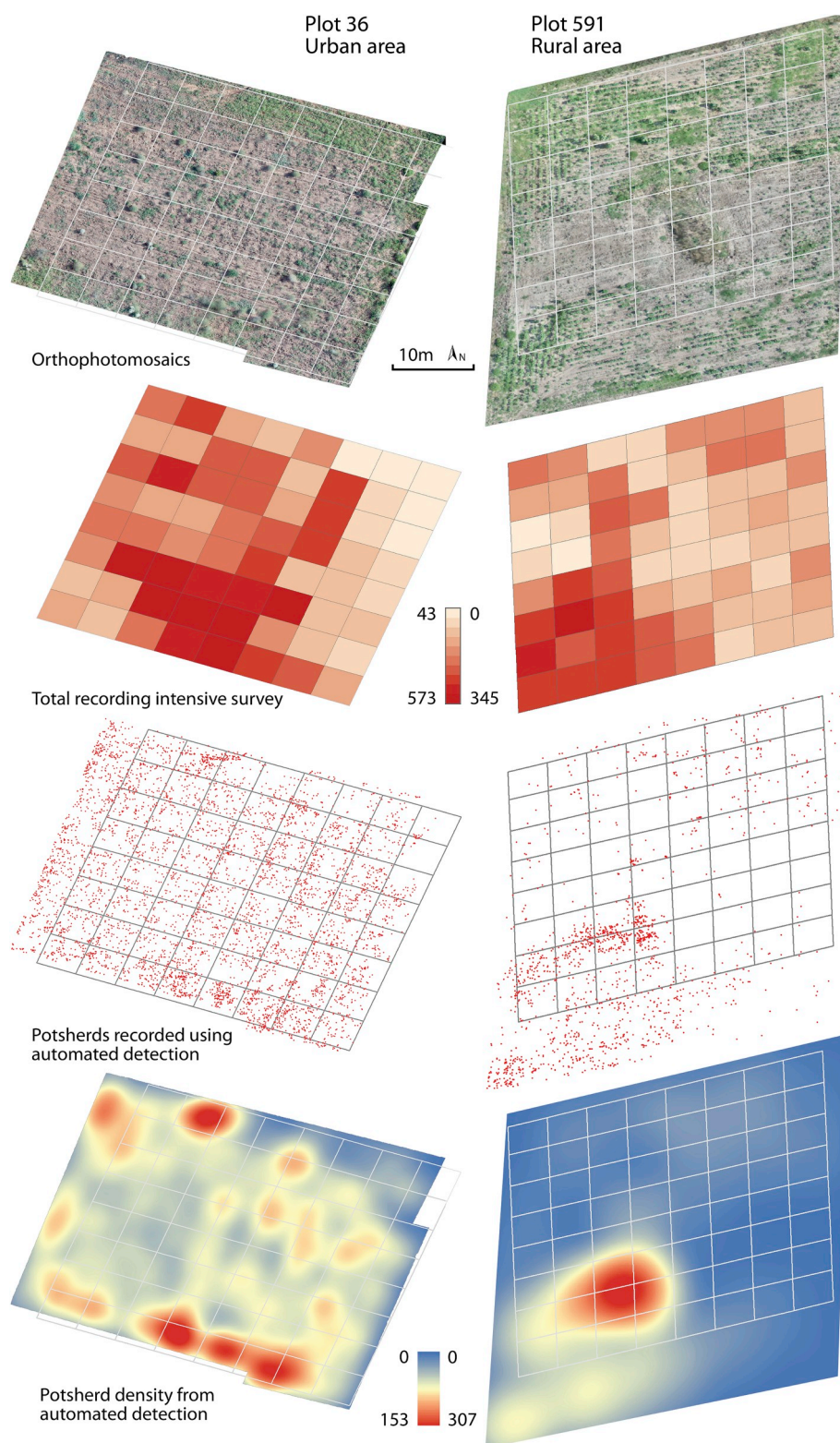
Fig. 4. Results from the intensive total count survey compared to those of the automatic detection.

that is probably related to current and past ploughing patterns in the field, which is not visible at all in the gridded data as the grid cells size is larger than the space between plough lines.

Also, the capacity to estimate the approximate area of the fragments allows a more thorough interpretation of their distribution. Fig. 5 presents the histograms of ceramic fragment sizes (area in cm^2) in plots 36 and 591. These clearly show divergent trends with plot 591

presenting a much more pronounced presence of larger ceramic fragments. This can be interpreted in terms of material recently surfaced or closer to the surface, while plot 36 pattern with thousands of very small fragments well-distributed through the field seems to present a case for a long-ploughed field were ceramics have been intensely fragmented and distributed beyond their original concentration area.

It is interesting to note that the automated method detected a higher

Table 1

Comparison of the costs and potsherds recorded with the different survey methods employed. Total time refers to the survey time multiplied for the number of surveyors. Urban plot 36 used 5 m-separated lines while rural plot 591 used 10 m-separated lines.

Plots	Pedestrian total record		Pedestrian lines		Automatic recording		
	Total time	Sherds	Total time	Sherds	Time	Photographs	Sherds
36 - 5 m lines	22 h	15779	3 h	1071	19 min	564	5189
591 - 10 m lines	18 h	2080	1:20 h	221	23 min	696	1597
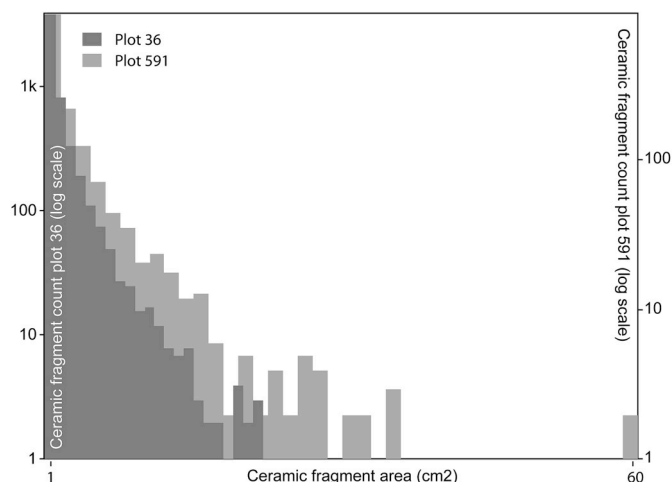
Fig. 5. Histograms of potsherd fragment sizes in both plots.

percentage of the ceramic fragments recorded during the field total recording survey in plot 591 than in plot 36. This is related to the redder hue of this field (see orthoimage in Fig. 4) that forced a more restrictive selection of the training data as there was less difference between RGB values for soil and for ceramic fragments than in field 591. This, is also in direct relation to the higher number of false positives detected in plot 591. The more inclusive the training data the more the false positives. In this regard, a fast measure of success can be obtained testing a selection of identified fragments against the orthophotomosaic, which in turn can be used to evaluate the potential percentage of missed fragments, which should keep an inverse relation to the lack of false positives when using the same ceramic training data in plots with different environmental conditions. In the same way, a fast inspection of the orthoimage can provide approximate percentages of undetected sherds that, in combination with the number of identified ceramic fragments, can be employed to provide estimates of total potsherds per field.

The recording using standard side-by-side survey following parallel lines with distances of 5 and 10 m provided less satisfactory results in terms of time and potsherd count than the automated recording. While the automated recording requires significant post-processing time per plot this can be done by a single person with minimal investment of time. Assuming the use of several computers or a powerful enough server the processing can be conducted in parallel for several fields thus greatly increasing the efficiency of the workflow.

The workflow presented here for the automated identification, mapping and quantification of pottery sherds in agricultural fields offers an important complementary technique to traditional fieldwalking employed in archaeological survey. The workflow can save large amounts of time and resources while increasing traditional survey accuracy and consistency when applied in optimal environments. In contrast with current survey practices, which provide density per areas, it can identify individual finds and even small fragments that would not usually been picked up and/or quantified in standard surveys. Not only it can estimate the number and distribution of fragments but it can also measure their approximate size. This can save much time during post-

processing, when sherds need to be measured, and provide new survey indices that can better describe the distribution of material culture and improve its interpretation. For example, instead of the ceramic density per area (based on the number of sherd collected in a given transect or survey line) the total number of pixels belonging to ceramic per area can be estimated resulting in much improved indices as pot sherds present highly variable sizes. By selecting pixels belonging to vegetation as training data this method would allow a quantifiable evaluation of visibility conditions in a field on a pixel basis or per unit of area, which are usually subjectively estimated during standard surveys for the whole field (Shennan, 1985: 44, Thomson, 2004). Just to provide an example, 23.5% of the cells in field 591 (Fig. 6) correspond to vegetation. These data can be used to correct the results of the ceramic count per grid unit according to vegetation density as shown in Fig. 6.

Using a commercial drone, such as the Phantom 4 with GPS, the resulting orthophotomosaic is already scaled and georeferenced saving the postprocessing time usually necessary for the geolocation of survey lines, sectors or transects. Although the scaling is lost when employing EE (as it cannot currently work with subcentimetric pixels) the original georeferenced orthoimage produced through the photogrammetric process can be easily employed to fit the resulting classification.

One of the big issues in archaeological survey is the comparability of survey data across regions and survey types. There is no standard survey methodology, and each project decides their own variables. Methods are usually designed to adapt to local environmental and cultural conditions and the survey objectives (Mattingly, 2000). A few surveys implement total collection of sherds while most others only collect a selection of the most representative types. Fieldwalking strategies also vary enormously with some surveys using walking lines (at variable distances and of variable length) or different grid sizes. This method can help tackle this problem by going down to the survey's atomic component, the potsherd, which can always be comparable between surveys following this same methodology. On a similar note, the method also improves the comparability of results within a single survey, which is usually highly subject to variability between the skills and efforts of individual field walkers and their relative tiredness (Plog et al., 1978; Banning, 2002: 224–225) as seen for the total count survey of field 36. By contrast the drone-derived results are mechanically consistent at the cost of more intuitive recognition of "special" finds (see below).

However, a series of issues need to be taken into account at this early stage of development:

1. Large computational resources are needed to conduct the photogrammetric and machine learning analysis. These have been solved here using freely available EE computing infrastructure but this is limited to a maximum field size of 75 m^2, which is slightly larger than current commercial drones can record in a single flight. Multiple or mosaicked fields can still be analysed. The continued development of cloud computing and the unremitting increase in computational power will likely increase the availability of computational resources for survey teams wanting to apply this technique during the next years.
2. A good knowledge of machine learning and experience in training data engineering is necessary for optimal results. This is comparable to the initial training necessary for novel surveyors, who require a
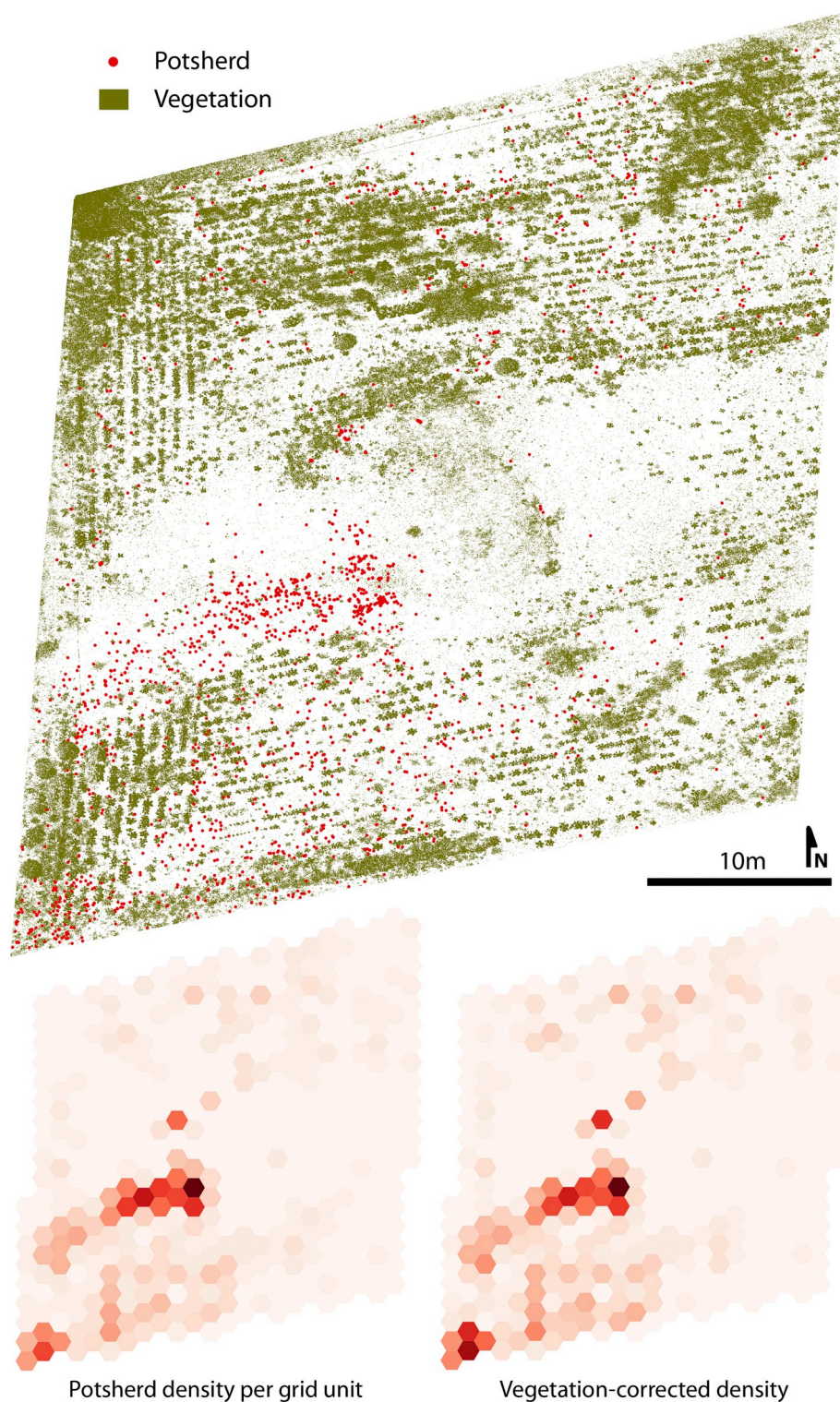
Fig. 6. Potsherd distribution plotted against vegetation density in plot 591.

few days or weeks of survey experience to be able to reach an optimum identification rate.

3. Best detection rates should be obtained in sedimentary plains with recently ploughed soils that offer a good colour contrast with pottery. In this regard the fields in Abdera required a more complex set up of the training data than it would have been necessary under ideal conditions but both plots provided more accurate results than those provided by traditional pedestrian line survey. Although some drones already incorporate sensors that can be employed to keep a constant height above ground in irregular terrain or avoid trees in forested areas these types of environment still represent a barrier for the application of this technique.

4. Light conditions and the presence of shadows can reduce the detection rate and will require more carefully selected training data. A particular drawback with dark shadows is the reduction of detection capability of the texture/gradient layer due to their low-contrast. Improvements in drone-based photogrammetry, pre-treatment of aerial images using photography software and a careful selection of

time of day and year to acquire the images should offer much improved results. Vertical light or uniform light conditions will provide the better results but the method works with any kind of light that allows the differentiation of ceramics. In fact the selected fields were recorded with oblique light. Drone survey, being not dependent on the availability of surveyors during summer months and holidays, offers the necessary freedom to conduct survey during most seasons and adapt to particular field conditions (not just those that are not under cultivation in summer). Also, being much faster than traditional fieldwalking the moment of the day offering best light conditions can be selected without the survey taking long enough for these conditions to change.

5. The method, based on spectral information and texture, will only detect one range of ceramics. While the more common type of oxidant ceramics or red-ware range from red to orange, the field might present dark reductive pottery types or light buff-ware that will not be detected by the algorithm. Other types of material culture with different colour and/or texture will go largely ignored. Also, material of little archaeological interest, such as modern building bricks (discarded dumps of which are frequently encountered by archaeologists in the Aegean countryside, but can be more easily ignored by a human data collector), will also be included if we aim at locating ceramics within a similar colour range. This can be addressed running different machine learning processes for the various ranges of pottery colour present in the field that can be combined in multitype distribution maps. Equally, other types of material culture, such as lithics or figurines, can also be identified if their colour is different enough from that of the field. In these cases it might be convenient to alter the script, for example not to use the texture/gradient band when the type material culture selected do not present flat surfaces or select different light or surface vegetation conditions during survey as dark pottery might not be easily differentiable from dark shadows.

6. The filtering process to eliminate small groups or single isolated false-positive pixels can also eliminate very small ceramic fragments. The kernel radius employed (see supplied code as Supplementary Material) would disregard any fragment smaller than 3×3 px (around $4–5 \, mm^2$), which, in any case are not collected during field survey. More problematic is the case of ceramic fragments positioned in such a way that only the section of the wall is visible forming a long but thin line of ceramic pixels. No matter how long the visible edge of the sherd is it will not be identified if it does not fit entirely within the filter kernel. The use of higher resolution imagery can address this problem.

7. During the data training process a trade-off between obtaining high values for ceramic pixels and disregarding those pixels that do not correspond to potsherds should be assumed. When pixels corresponding to ceramics present spectral information within the bounds of non-ceramic pixel values (as defined during the training process) then there will be an increase of false positives. To Illustrate this point the spectral and gradient signatures of the potsherds (as selected for the training data), in comparison to those presented by soil and vegetation (the other most common types of elements in the field) are presented in Fig. 7. The signatures show much overlap in the RGB bands between potsherd and soil classes with a higher value on the Red band for the ceramic fragments and lower Blue band values as expected. If very strict parameters for pottery are selected for the training data then fewer pottery sherds will be identified as such. Therefore, the user will need to choose between having false positives or achieving lower rates of detection. In our work we have preferred the latter. The better the quality of the orthomosaics, the conditions of the field and the training data the less significant this trade-off is on the results.

Fig. 7 also highlights the very discriminant values of vegetation that set it apart and ensures correct identification of field visibility values. More interesting, however, is the clearly discriminant values
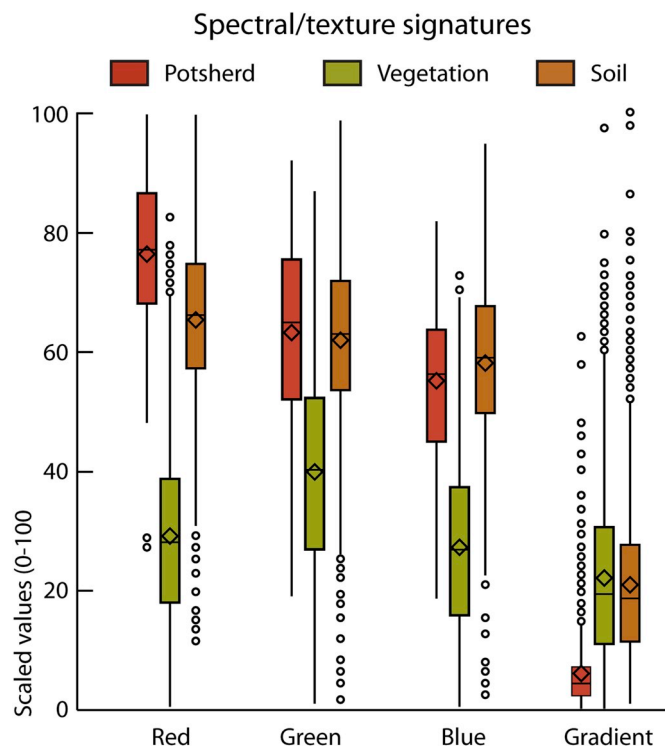

Fig. 7. Spectral/gradient signatures for potsherd training data, vegetation and soil.

provided by the gradient magnitude layer, with values for ceramic well below those of soil and vegetation in accordance to its uniform texture represented by a lower interpixel variability. This clearly illustrates the importance of texture analysis for the successful application of this method.

8. The postprocessing of finds usually involves a laboratory-based analysis of selected sherds in which the most significant ceramic fragments are measured and recorded and the type of pot and other important information, such as the pot date or function, are extracted by specialists. This is, of course, not possible with this method. However, the workflow does not aim to exclude sherd collection, which his still essential for interpretation. Standard survey practices quantify fragments after collection during postprocessing, when ceramic density maps are generated. The automated method generates distribution data with a delay of a few hours or a day. These data can then allow to conduct fast targeted small-scale collection at significant concentrations (as defined by potsherd densities but also sizes) by a few specialists and, in doing so, it can save large long-term storage space usually consumed by survey projects in which material collection is systematic and only after the analysis of the finds significant concentrations of material are mapped. Material collection will always be necessary for in-depth pottery analysis but this method can help reducing the amount of stored ceramics by defining significant potsherd sampling points instead of the traditional systematic collection. What this approach cannot do is to interpret material culture. For example, even if it is able to identify figurines or written tablets, these will just be identified as a group of pixels with high probability of being ceramic. There is a risk therefore that valuable particular types of finds will be recorded but not identified as such and therefore left uncollected.

9. When large areas are being surveyed and these incorporate different types of soils or environments (or they are surveyed under very different light conditions) it might not be adequate to apply the same training data systematically to all fields and it might need to be

updated with pixels selected from different orthomosaics.

Other concerns not directly related to the survey method and its results but of a more practical nature are the difficulty of obtaining permissions for drone photographic survey in different jurisdictions. A relatively new technology, many countries have only recently begun regulate drone usage but in many cases have done so in a way which may make application of this workflow bureaucratically difficult or impossible. Flights might also be constrained in restricted areas such as military bases or airport surroundings. On the plus side, the very low altitude of necessary flights (3 m) and the location of survey usually being unpopulated areas may make it easier to argue to the authorities that risks and security concerns are unwarranted.

Also important if the production of potsherd distribution and density maps while on the field is sought, is the need to have a reliable and relatively fast internet access, which is necessary for the upload of large plot orthoimages into EE. This is not always available in many survey areas but we are confident that the setting up of satellite-based global broadband internet services such as Starlink and OneWeb during the next few years will alleviate this need.

Despite these drawbacks the results obtained in the fields under analysis have yielded accurate results even when these were selected as complex testing areas that had not recently been ploughed and incorporated vegetation, stones, shadows, and red-orange soil tones. The workflow has been optimised for speed, ease of application and low-cost as images were taken in jpeg and had not been treated after their acquisition, an open to the public cloud computing platform has been employed and all information is derived from photographic visible RGB bands taken using an affordable popular drone.

The workflow is applicable to many survey situations. Extensive survey aiming to locate sites could be carried out, for example, by acquiring images following long parallel lines (at distances of, for example, the typical site's ceramic dispersion radius) across the study area. However, this workflow efficiency is proportional to the quantity of ceramics to be identified per unit of area. That is, the more ceramics in the field the more time it saves with respect to traditional survey. Also the smaller the area the more similar the environmental conditions (such as the type and colour of the soil) and the more applicable a single training set is to the whole study area. In this regard, intra-site surveys have the most to gain from it but any intensive survey in adequate areas with a minimum of pottery sherd visibility would benefit from complementing current survey methods with drone-based automated classification approaches.

6. Future developments

If the trends in technical developments documented in the introduction continue, it is possible to expect an important increase in the quality and applicability of this or similar workflows during the next years.

As drones incorporate differential GPS and increase flight time, planning and execution of survey missions will be faster and more efficient. As their cameras improve their quality and resolution, better spectral signatures can be obtained for ceramics and both kernel-based texture analyses and gradient magnitude will markedly increase their prediction capacity as shown in Fig. 3. An increase in computing capacity with the wider availability of cloud computing facilities at a lower price will allow reducing computing times in both photogrammetry-based processes and machine learning-based detection or maintain them when larger fields or higher resolutions are involved.

The incorporation of multispectral and, most importantly, thermal imagery will also imply a qualitative increase in ceramic detection given the difference in thermal signatures between pottery, stones, soil and vegetation. Many commercial drones allow payloads of up to 10 kg and multispectral and thermal cameras (and combinations of these) are available for many popular professional drone-models. Thermal

cameras resolution is still coarse in comparison with optical and multispectral cameras but important improvements can be expected in the future.

Improvements in computing will also result in deep-learning algorithms being employed for the detection of ceramic sherds. Potsherds' particular morphology, texture and straight edges in combination with their spectral signature will allow deep learning processes to detect them reliably and in combination with other machine learning processes even differentiate between ceramic types.

7. Conclusions: a new era for archaeological survey?

We have presented here a new workflow combining drone-based photogrammetry and machine learning for the automated recording of surface distributions of archaeological material. With a fraction of the cost than that of traditional pedestrian survey, this method has the potential to deliver faster results and higher analytical capabilities when applied under favourable conditions. This workflow uses free or low cost software and cloud computing services and the code employed and clear instructions on how to use it are available as Supplementary Material. It can be applied with minor adjustments to user-uploaded orthomosaics after the selection of appropriate training data. The only pre-requisite is a drone able to fly following a pre-programmed path carrying a camera with enough resolution to record potsherds or other items of material culture of interest. These characteristics are nowadays available in many professional and recreational drones at relatively low prices.

The workflow presented here addresses a specific part of an intensive survey: the quantification of surface material culture items. This is the most mechanical part of the work and requiring large amount of hours to obtain basic information, which is necessary for the interpretation of cultural data but, by itself, has limited value. Aspects, such as site chronology, the use of space and how these change through time will continue to be addressed through the analysis of collected relevant material by specialised archaeologists. Automated sherd detection can represent an important contribution, allowing to focus human resources on those aspects of a survey that can yield more relevant answers to research questions, and directing the efforts to the areas with more potential.

The results show that this method provides a reliable and detailed recording of material consistently surpassing those of traditional fieldwalking even in less than optimal conditions. However, there is still room to maximise detection of material culture and discrimination of false positives. This paper aimed to provide a strong methodological and conceptual basis on which future applications can build up. There is little doubt that technical advances will largely contribute to the widespread adoption of this and related approaches during the next years.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jas.2019.105013. For future updated versions of the code, please refer to https://github.com/horengo/Orengo_Garcia-Molsosa_2019_JAS.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Agapiou, A., 2017. Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. Int. J. Digit. Earth 10, 85–102.

Alcock, S.E., 2000. Extracting meaning from ploughsoil assemblages: assessments of the past, strategies for the future. In: Francovich, R., Patterson, H. (Eds.), Extracting Meaning from Ploughsoil Assemblages. Oxbow Books, Oxford, pp. 1–4.

Alcock, S.E., Cherry, J.F., 2004. Introduction. In: Alcock, S.E., Cherry, J.F. (Eds.), Side-by-Side Survey. Comparative Regional Studies in the Mediterranean World. Oxbow Books, Oxford, pp. 1–9.

Anselin, L., 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27, 93–115.

Banning, E.B., 2002. Archaeological Survey. Kluwer Academic, New York.

Bevan, A.H., Conolly, J.W., 2004. GIS, archaeological survey and landscape archaeology on the island of Kythera, Greece. J. Field Archaeol. 29 (1/2), 123–138.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Cherry, J.F., Gamble, C., Shennan, S., 1978. Sampling in Contemporary British Archaeology. BAR 50, Archaeopress, Oxford.

Conners, R.W., Trivedi, M.M., Harlow, C.A., 1984. Segmentation of a high-resolution urban scene using texture operators. Comput. Vis. Graph Image Process 25 (3), 273–310.

Crowther, D., 1983. Old land surfaces and modern ploughsoil: implications of recent work at Maxey, Cambridgeshire. Scott. Archaeol. Rev. 2, 31–44.

Fish, S.K., 1999. Conclusions: the settlement pattern concept from an Americanist perspective. In: Billman, B.R., Feinman, G.M. (Eds.), Settlement Pattern Studies in the Americas: Fifty Years since Virú. Smithsonian Institution Press, Washington, pp. 203–208.

Garcia, A., Orengo, H.A., Conesa, F., Green, A., Petrie, C., 2019. Remote sensing and historical morphodynamics of alluvial plains. The 1909 indus flood and the city of Dera Gazhi Khan (province of Punjab, Pakistan). Geosciences 9, 21.

Georgiadis, M.; Garcia-Molsosa, A.; Orengo, H.A.; Kefalidou, E. and Kallintzi, K. In Preparation. APAX Project 2015-2018: A Preliminary Report. (Hesperia).

Gillings, M., Sbonias, K., 1999. Regional survey and GIS: the boeotia project. In: Gillings, M., Mattingly, D., van Dalen, J. (Eds.), Geographical Information Systems and Landscape Archaeology. Oxbow, Oxford, pp. 35–54.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Grün, A., Remondino, F., Zhang, L.I., 2004. Photogrammetric reconstruction of the great buddha of Bamiyan, Afghanistan. Photogramm. Rec. 19 (107), 177–199.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst., Man, Cybernet., SMC-3 (6), 610–621.

Kletter, R., De-Groot, A., 2001. Excavating to excess? Implications of the last decade of archaeology in Israel. J. Mediterr. Archaeol. 14, 76–85.

Liss, B., Howland, M.D., Levy, T.E., 2017. Testing Google Earth Engine for the automatic identification and vectorization of archaeological features: a case study from Faynan, Jordan. J. Archaeol. Sci.: Report 15, 299–304.

Lock, G., Bell, T., Lloyd, J., 1999. Towards a methodology for modelling surface survey data: the sangro valley project. In: Gillings, M., Mattingly, D., van Dalen, J. (Eds.), Geographical Information Systems and Landscape Archaeology. Oxbow, Oxford, pp. 55–64.

Mattingly, D., 2000. Methods of collection recording and quantification. In: Francovich, R., Patterson, H. (Eds.), Extracting Meaning from Ploughsoil Assemblages. Oxbow Books, Oxford, pp. 5–15.

Menze, B.H., Ur, J.A., 2012. Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. Proc. Natl. Acad. Sci. 109 (14), E778–E787.

Oonk, S., Spijker, J., 2015. A supervised machine-learning approach towards geochemical predictive modelling in archaeology. J. Archaeol. Sci. 59, 80–88.

Orengo, H.A., 2013. Combining terrestrial stereophotogrammetry, DGPS and GIS-based 3D voxel modelling in the volumetric recording of archaeological features. ISPRS J. Photogrammetry Remote Sens. 76, 49–55.

Orengo, H.A., Krachtopoulou, A., Garcia-Molsosa, A., Palaiochoritis, K., Stamati, A., 2015. Photogrammetric re-discovery of the Eastern Thessalian hidden long-term landscapes. J. Archaeol. Sci. 64, 100–109.

Orengo, H.A., Knappett, C., 2018. Towards a definition of Minoan agro-pastoral landscapes: results of the survey at Palaikastro (Crete). Am. J. Archaeol. 122 (3), 479–507.

Orengo, H.A., Petrie, C.A., 2017. Large-scale, multi-temporal remote sensing of palaeo-river networks: a case study from Northwest India and its implications for the Indus civilisation. Remote Sens. 9 (7), 735.

Orengo, H.A., Petrie, C.A., 2018. Multi-scale relief model (MSRM): a new algorithm for the visualization of subtle topographic change of variable size in digital elevation models. Earth Surf. Process. Landforms 43, 1361–1369.

Orengo, H.A., Conesa, F.C., Garcia-Molsosa, A., Lobo, A., Green, A.S., Petrie, C.A., 2019. Living on the edge of the desert: automated detection of archaeological mounds in Cholistan (Pakistan) using machine learning classification of multi-sensor multi-temporal satellite data. In: Submitted to Proceedings of the National Academy of Sciences, submitted for publication.

Oshiro, T.M., Perez, P.S., Baranauskas, J.A., 2012. How many trees in a random forest? In: In: Perner, P. (Ed.), Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science, vol. 7376. Springer, Berlin, Heidelberg, pp. 154–168.

Plog, S., Plog, F., Wait, W., 1978. Decision-making in modern surveys. In: In: Schiffer, M.B. (Ed.), Advances in Archaeological Method and Theory ume 1 Academic Press, New York.

Rayne, L., Bradbury, J., Mattingly, D., Philip, G., Bewley, R., Wilson, A., 2017. From above and on the ground: geospatial methods for recording endangered archaeology in the Middle East and north africa. Geosciences 7 (4), 100.

Schiffer, M.B., Sullivan, A.P., Klinger, T.C., 1978. The design of archaeological surveys. World Archaeol. 10 (1), 1–28.

Shennan, S., 1985. Experiments in the Collection and Analysis of Archaeological Survey Data: the East Hampshire Survey. Department of Archaeology and Prehistory, University of Sheffield, Sheffield.

Stek, T.D., 2016. Drones over Mediterranean landscapes. The potential of small UAV's (drones) for site detection and heritage management in archaeological survey projects: a case study from Le Pianelle in the Tappino Valley, Molise (Italy). J. Cult. Herit. 22, 1066–1071.

Thomson, S., 2004. Side-by-side and back to front: exploring intra-regional latitudinal and longitudinal comparability in survey data. Three case studies from Metaponto, southern Italy. In: Alcock, S.E., Cherry, J.F. (Eds.), Side-by-Side Survey. Comparative Regional Studies in the Mediterranean World. Oxbow Books, Oxford, pp. 65–85.

van der Maaten, L.J.P., Boon, P.J., Paijmans, J.J., Lange, A.G., Postma, E.O., 2007. Computer vision and machine learning for archaeology. In: Clark, J.T., Hagemeister, M. (Eds.), Digital Discovery. Exploring New Frontiers in Human Heritage. Computer Applications and Quantitative Methods in Archaeology. Archaeolingua, Budapest.

Verhoeven, G., Taelman, D., Vermeulen, F., 2012. Computer vision-based orthophoto mapping of complex archaeological sites: the ancient quarry of Pitaranha (Portugal-Spain). Archaeometry 54, 1114–1129.

Wendorf, F., 1962. A Guide for Salvage Archeology. Museum of New Mexico Press, Santa Fe.