

Article

Not peer-reviewed version

---

# Combining Sentinel 2A Timeseries, Lidar Data and Archaeological Prospection for the Study of Ancient Roads and Poblational Pattern in Right Side of Guadalén River, (Vilches, Spain)

---

[Antonio J. Ortiz Villarejo](#) \* and [Luis M. Gutiérrez](#)

Posted Date: 13 February 2024

doi: 10.20944/preprints202402.0727.v1

Keywords: remote sensing; lidar; satellite; least cost path; roman empire; GIS



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Combining Sentinel 2A Timeseries, Lidar Data and Archaeological Prospection for the Study of Ancient Roads and Poblational Pattern in Right Side of Guadalén River, (Vilches, Spain)

Antonio J. Ortiz-Villarejo <sup>1,\*</sup> and Luís M. Gutiérrez-Soler <sup>2</sup>

<sup>1</sup> Anthropology, Geography and History Department, Modern History Area, Universidad de Jaén, Paraje "Las Lagunillas" s/n. C.P. 23071, Jaén, Spain

<sup>2</sup> Instituto Universitario de Investigación en Arqueología Ibérica, Universidad de Jaén, Paraje "Las Lagunillas" s/n. C.P. 23071, Jaén, Spain; lmsoler@ujaen.es

\* Correspondence: ajvillar@ujaen.es

**Abstract:** This paper shows the firsts results of a methodological approach in which Remote Sensing technologies (Satellite imagery and LiDAR) combined with Least Cost Path algorithm on GIS are being tested like tools to overcome locally-based interpretations about populational patterns and communication routes during Flavian municipality in the Comarca de El Condado region. The identification of fourteen hypothetical communication routes between already known archaeological sites thanks to a recent archaeological prospection, and the confirmation of three sections of them and a new archaeological site demonstrate that the combined use of the aforementioned technics has a very useful to delve more deeply, more widely and at a variety of different scales into the relation between the twelve known/selected Early Roman Empire sites and the surrounding landscape in the right bank of Guadalén River. It is expected that, in next phases, this methodological approach will allow us to go deeper in analysing and understanding spatial distribution, exploitation, and interrelation between different Early Roman Empire sites, helping us to fill the aforementioned gaps in our understanding of the archaeological *continuum* in the region reducing time and costs.

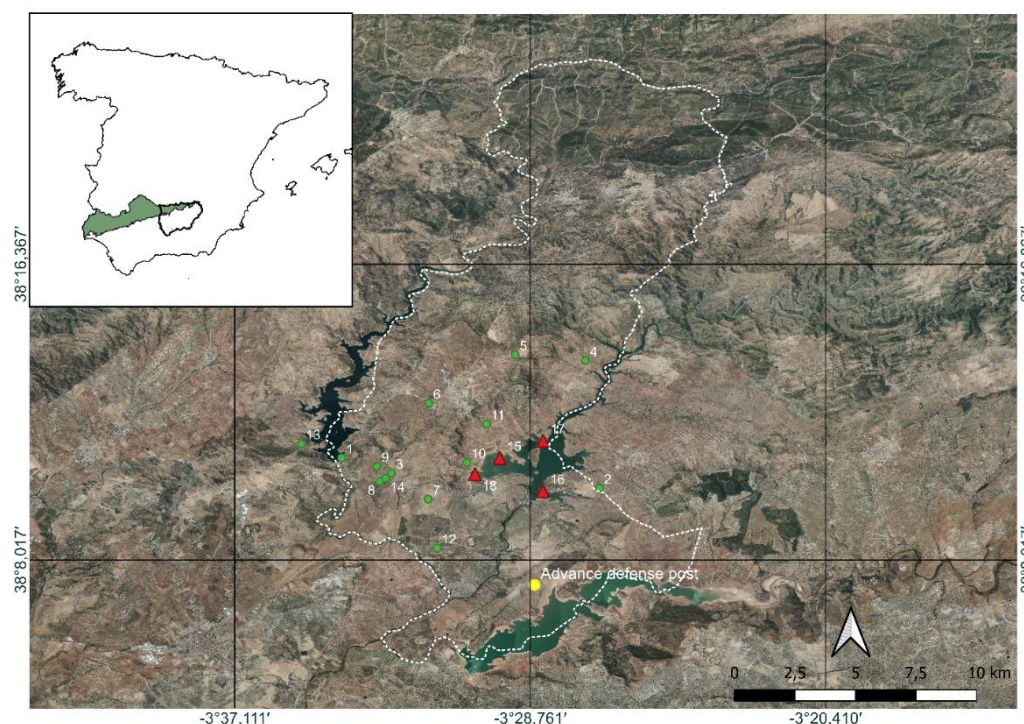
**Keywords:** remote sensing; lidar; satellite; least cost path; roman empire; GIS

## 1. Introduction

Sierra Morena is located north of the Guadalquivir River and is a mountainous region rather than a mountain range in the strict sense. It is the transition between the Meseta Central and the Guadalquivir depression that runs from east to west from the Portuguese Algarve region for more than 500 km (Alaminos Tererno, 2001) to Sierra Morena region, and it is formed by slates and granites and calcium-poor acidic soils.

The region of Sierra Morena has been and continues to be a land of frontier and passage between the south and the north of the Iberian Peninsula. Its occupation dates back to prehistoric times, a period documented in natural shelters in Aldeaquemada, and continues to the present day, encompassing all historical periods such as the Iberian period in the Cueva de los Muñecos, the Roman era in nearby regions like Cástulo and the Battle of Baecula, the Battle of Las Navas de Tolosa in the heart of the Middle Ages within the current Natural Park of Despeñaperros, the founding of the New Settlements in Sierra Morena by Charles III, which encompasses this project, and finally the construction of the A4 Highway that connects the center of the peninsula with the south, among other events.

The area under study is within the municipality of present-day Vilches in the Sierra Morena region (area with white dash in Figure 1)



**Figure 1.** Location of the area under study and sites studied inside actual administrative division of Vilches municipality (white dash) and Sierra Morena limits (green area). Source: PNOA 2016.

In spite of the intense historical activity the chronological period chosen for this study was the High Imperial Roman period, since, as we will see later, it is a time when there is a change in the population pattern of the area. To understand this pattern in this part of the Sierra Morena, it is essential to take into account the controversy that arose around the identification of the *municipium Flavium Baesucci*. As we mentioned, historiography has traditionally associated its location with the nucleus of the current population on which Vilches sits due to the epigraphic documentation found there (González Roman and Mangas, 1991, although today Santagón, a large archaeological site located in the valley, has been alternatively proposed.

The population distribution during the protohistory documented in the middle stretch of the Guadalimar valley is the starting point for the study of the occupancy model of the first moments of Roman times in the Upper Guadalquivir, from the end of the Second Punic War until the Flavian period. (Gutiérrez Soler et al., 1999, 1995; Royo et al., 1995)

The initially dispersed occupancy model for the Guadalimar valley evolved in the first half of the first century BC towards a type of concentrated settlement, highlighting the town of La Monaria as the main centre. The moments immediately after the Roman conquest (2nd century BC) represent the first permanent occupation of this sector of Sierra Morena, a period of generalized instability in the Upper Guadalquivir during the late-republican period, with an interest in economic exploitation centred on continuity and expansion of mining operations in Carthaginian hands (Alejo Armijo, 2019).

Until the Flavian era, the process of occupation of the Upper Guadalquivir maintained the traditional structured and hierarchical patronage systems (Gutiérrez Soler, 2002), reaffirming in this case the new administrative base of Roman Hispania in the region. It will not be until the Flavian period that the settlement pattern changes, intensifying the exploitation of the territory and increasing the number of rural settlements in the Linares-Bailén depression, an aspect already studied in the Subbética of Córdoba and in the Campiña Oriental in Jaén (Alejo Armijo, 2019).



It is undoubtedly the implantation of Latin law by Vespasian that triggered this change, since it turned the settlements into city centres, generalizing a dispersed rural settlement of small and medium population centres, which would consolidate the initial commitment to implementing colonial centres such as *Colonia Augusta Gemella Tucci* and *Colonia Salaria*, which promoted the distribution of Italian products (Castro Lopez, 1999).

In turn, throughout the territory, fortified enclosures are documented that control the entire Guadalimar valley and, possibly, the main roads (Arboledas Martinez, 2008). For that reason, we selected the site located on the access road to the city and interpreted as an advanced defence post of the city of Giribaile as a starting point for our analysis. It will be the moment in which the first *villae* are installed, buildings in the field that follow Roman architectural models, of 2 ha maximum (Pérez Bareas et al., 1992, p. 93), which highlights the appearance of *Terra Sigillata Hispanica* (TSH) and *Terra Sigillata Gallica* (TSG) and stucco coatings that clearly show a new evolution in the urban phase (Alejo Armijo, 2019).

Altogether, the new historical situation in which the Upper Guadalquivir is situated during the period of Roman domination is marked by some interests that leave their trace on the landscape: the control of the roads through different *nuclei* that articulate communication networks from Cástulo (Cástulo-Libisosa, Cástulo-Malaca, Cástulo-Corduba), through the exploitation of resources, with the mines of Sierra Morena, olive and cereal productions, the arboreal wealth of *saltus tugensis*, livestock resources... collected from the *villae*, organized following a cadastral model and the main territorial planning axes formed from the colonies of Salaria and Tucci (Alejo Armijo, 2019).

For a proper understanding of the historical moment and landscape at hand and also spatial distribution, we must not overlook, among other aspects, the intrinsic relationship that occurs between the society of that specific time and its immediate environment, which entails both direct and indirect modifications to it (Campana, 2018). This is why it is important to have knowledge of the geophysical characteristics of a region before conducting a study, as these characteristics will, in turn, influence the design of a methodology adapted to the objectives pursued.

Sierra Morena region, when viewed from the Extremadura or Castilian plains, the mountain range appears as a series of gentle undulations, but from the Andalusian side, it rises in large cliffs, steeply dropping to overcome a 700 m step.

The area has always been characterized by its demographic void, due in part to the characteristics of the soil that are responsible for its poor agricultural potential. Instead the people of the area practice traditional livestock husbandry and hunting, with little modification of the local vegetation cover (*Sierra Morena Andaluza*, n.d.).

Sierra Morena has a semiarid Mediterranean-type climate with some degree of continentality, which becomes more marked when moving eastwards. It manifests itself in a very pronounced thermal amplitude and in an abundance of days with frost. Summers are very hot and dry, with average temperatures of approximately 33 °C; winters are cold with very irregular rains. Rainfall is closely linked to altitude, increasing from 600–800 litres/m<sup>2</sup> (Alaminos Terreno, 2001) annual average in the peripheral areas, and with up to more than 1,500 l/m<sup>2</sup> recorded in the highest peaks of the western region, although up to 9 climatic types can be identified. The climate affecting Sierra Morena is type 8, with the greatest rainfall deficits due to the rain shadow effect and greater continental influence of the peninsular plateau that affects the most low-elevation and eastern parts of Sierra Morena (*Sierra Morena Andaluza*, n.d., p. 141).

Sierra Morena is partially covered by a typical Mediterranean forest rich in shrubs and herbaceous vegetation (C. M. Martín, n.d.) and has an irregular orography that determines accessibility in public open spaces, especially for walk-over surveys.

The rivers crossing the area under study are the Yeguas, Jándula, Rumblar, Guarrizas, Montizón and Guadalmena rivers.

2. Objectives

Based on the previously described characteristics, with the present research, we aim to shed light on the spatial distribution and communication networks during the Flavian period in the Upper Guadalquivir region, with a specific focus on the Giribaile area.

For this study, we build upon the previous archaeological prospecting work conducted by Alejo and Gutiérrez (2019), which has provided us with the locations of the sites used as the basis for our research. Our goal is to advance our understanding of the territorial organization of this environment, and for this purpose, we have outlined the following objectives:

1. Develop a methodology in which Remote Sensing technologies (satellite imagery and LiDAR) combined with the Least Cost Path algorithm in GIS are tested as tools to go beyond locally-based interpretations regarding population patterns and communication routes during the Flavian municipality in the Comarca de El Condado region.
2. Identify potential communication routes.
3. Establish the foundations for future research projects that will allow us to validate the methodology used and further advance our knowledge of the region.

3. Materials

For the development of this research, and taking into account the aforementioned objective and geophysical characteristics of the studied region, we consulted information from different sources. The comprehensive interpretation of these sources has allowed us to provide new data on the existence of archaeological material vestiges that, in turn, have served to confirm some working hypotheses and enabled us to propose new research. The sources consulted are as follows:

3.1. Geographic Documentation of High Imperial Sites

Regarding the sites under study, these come from two different sources, on the one hand from the Digital Guide of the Cultural Heritage of Andalusia databases and on the other from the prospecting carried out by the Alejo & Gutiérrez team (2019) in which a total of 21 archaeological sites were found. In total, 18 archaeological sites were represented (**Error! Reference source not found.**) with a UTM coordinate, each marking the centre of the identified site. The datum used has been ETRS89/ZONE 30N.

Table 1. Archaeological sites included in this article.

ID	SITIO	ID	SITIO
1	VALDEINFIERNO I	10	GALEOTE 2
2	CORTIJO DE ARQUILLOS EL VIEJO	11	LLANOS DE VICHI - CERRO MANZANO
3	CORTIJO EL RASO	12	NECROPOLIS DE LA LEONA
4	CORTIJO HORTALANCA	13	PALAZUELOS
5	CORTIJO JUAN CLAVERO	14	RECINTO SAN JULIAN
6	CORTIJO LA FLORINA	15	CERRILLO DEL CUCO
7	CORTIJO LOS GRAJOS	16	DEHESA DEL VIZCONDE
8	SAN JULIAN - EL RASO	17	SAN ALEJO
9	FUNDACION SAN JULIAN	18	SANTAGON

3.2. Lidar Data

The main aim of using Lidar data was the generation of a DEM with a spatial resolution as high as possible. In this case the generated DEM had a resolution of 5 m/pix, which was sufficient since in our case, we needed to generate a cost map in which the resolution was not determinant. In addition, we must consider both the large area covered and that a lower resolution would increase the weight

of the model to be analysed and therefore both the data processing time and the limits of the GIS itself, also requiring hardware with greater power.

The Lidar dataset used in this paper was downloaded from Spanish National Geographic Information Centre (Organismo Autónomo Centro Nacional de Información Geográfica, 1988) and are freely available. The used dataset covered the entire Spanish territory in several flights carried out between 2008 and 2015. The generated point clouds had a density of 0.5 points/m<sup>2</sup> and are organised in digital files distributed in 2 x 2 km files with altimetric information from LiDAR point cloud. All the files are coloured with infrared and RGB from orthophotos and are referenced to SRC ETRS 89, as established by the current Spanish standards (BOE, 2007).

### 3.3. Shapefile Reservoirs

As seen in the attached orthophoto (Figure 1) in the area under study, there are three reservoirs, of recent construction: Guadalén (1946-1954), Fernandina (1989-1991) and Giribaile (1992-1997) whose boundaries in shapefile (SHP) format were downloaded from the Institute of Statistics and Cartography of Andalusia (Junta de Andalucía, n.d.). This file provided us with a precise delimitation of the three reservoirs, whose boundaries were included both when calculating slope and optimal routes, since the present-day anthropic flat surface (which was non-existent in the historical moment of interest) would have distorted the results in the calculations of the routes.

These files have been basic for this research due to they were used to cut out the reservoir areas as we will see in methodology part.

### 3.4. Satellite Images

The analysis of satellite imagery data, is based on the study of the behaviour of radar waves in the different bands of the electromagnetic spectrum (Lasaponara and Masini, 2013) such as the thermal region composed of the near infrared (NIR: 700 nm to 1100 nm) and short wavelengths (SWIR: 1.1 nm to 3 nm) (Agapiou et al., 2013), or the L 1–2 GHz, (5–30 cm) (Caspari et al., 2020; Cisneros and Santos, 2003); X 8–12.5 GHz, (2.5–3.75 cm) bands (Bachagha et al., 2022; Monterroso Checa and Martínez Reche, 2018a), which have proven to be the most suitable for the detection of archaeological trace, especially in dry and fine earth territories (Tapete and Cigna, 2017).

The images processed in this article come from the Sentinel 2A satellite of ESA's Copernicus mission (B. P. Martín et al., 2020). The use of satellite images for the study of archaeological and heritage remains has been demonstrated in numerous publications (Casana, 2014; Derooin et al., 2012; Stewart, 2017, 2020; Tapete & Cigna, 2019) and, more specifically, in the field that is of interest to us, there are some examples, such as Monterroso et al. (2018), for the study of the Roman Mellaria or Zanni et al. (2019) in the Srem region of Serbia.

For our case study, we analysed a time series of images from 01/01/2017 to 12/31/2021, with a cloud cover of less than 30% (which meant the analysis of 192 images of the area under study) to determine the trends of the region to establish, if possible, climatic patterns such as periods of rain or droughts. We also sought the most suitable moment for analysing and optimizing the potential of these images. The dataset used were harmonized and the processing level of the product was the highest, level 2, that means that the surface of the images is Atmospherically corrected and also reflectance's are shown in cartographic geometry. All images were UTM georeferenced and has a spatial resolution of 10 m/pixels on the bands used in this study. Thus, with this analysis, it is possible to understand the phenology of the territory in a generally speaking way to later focus on the analysis of the images whose capture date is as close as possible to the moment of interest.

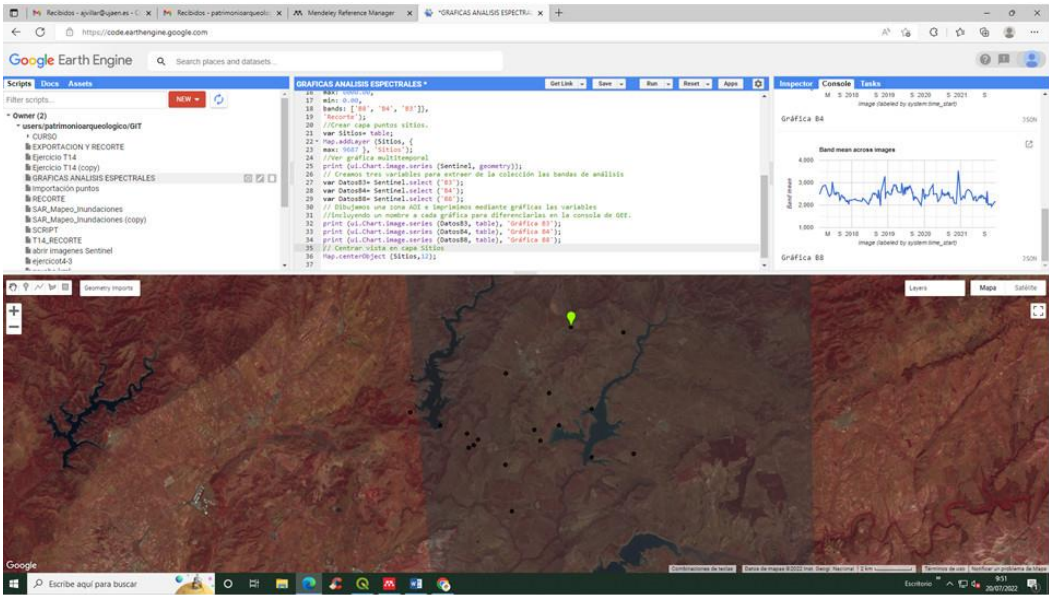
## 4. Methodology

In this section, we use two fundamental tools due to their potential: Google Earth Engine (GEE) (Google Earth Engine, n.d.) and GIS software, QGIS, version 3.22.3 (QGIS, n.d.).

#### 4.1. Google Earth Engine

We decided to use this platform to select the most suitable data for archaeological analysis of the studied region. The main advantage of this platform is that allows the processing of large quantities of images in the cloud, so it is not necessary to have a high-powered PC to perform these calculations or to download the entire time series of data that would require very high storage capacities. It is free, but it requires knowledge of the JavaScript programming language to perform the processing (**Error! Reference source not found.**). The script used was:

```
//multitemporal analisis
var Sentinel= ee.ImageCollection("COPERNICUS/S2_HARMONIZED")
.filterDate ('2017-01-01','2021-12-31')
.filterMetadata ('CLOUD_COVERAGE_ASSESSMENT','Less_than',30)
.filterBounds (geometry);
//Mean Value of the selected area.
var Mosaico= Sentinel.median ();
Map.addLayer (Mosaico, {
max: 6000.00,
min: 0.00,
bands: ['B8', 'B4', 'B3']},
'Composicion Sentinel');
//Cut area by geometry.
var Recorte= Mosaico.clip (geometry);
Map.addLayer (Recorte, {
max: 6000.00,
min: 0.00,
bands: ['B8', 'B4', 'B3']},
'Recorte');
//Create layer with location using points
var Sitios= table;
Map.addLayer (Sitios, {
max: 9687 }, 'Sitios');
//See multitemporal graphics
print (ui.Chart.image.series (Sentinel, geometry));
// Creation of three variables to extract bands from collection to analyse
var DatosB3= Sentinel.select ('B3');
var DatosB4= Sentinel.select ('B4');
var DatosB8= Sentinel.select ('B8');
//Draw an AOI and print graphics from variables including
//the name of each graph to differentiate in GEE console
print (ui.Chart.image.series (DatosB3, table), 'Gráfica B3');
print (ui.Chart.image.series (DatosB4, table), 'Gráfica B4');
print (ui.Chart.image.series (DatosB8, table), 'Gráfica B8');
// Centre view on layer Sitios
Map.centerObject (Sitios,12);
```



**Figure 2.** Screenshot of Google Earth Engine on which the used script as well as part of the results obtained can be seen. Source: The authors.

The data provided by the constellation of satellites of the Copernicus mission, specifically those of the Sentinel 2A satellite, are made up of a total of 12 bands of the electromagnetic spectrum. The focus of this study are bands 3, 4 and 8, whose main characteristics are shown (**Error! Reference source not found.**):

**Table 2.** Spectral Bands used and main characteristics.

	WAVELENGTH (µm)	SPATIAL RESOLUTION (m)
BAND 3 - Green	0.54 – 0.57	10
BAND 4 - Red	0.65 – 0.68	10
BAND 8 – Near Infrared (NIR)1	0.78 – 0.90	10

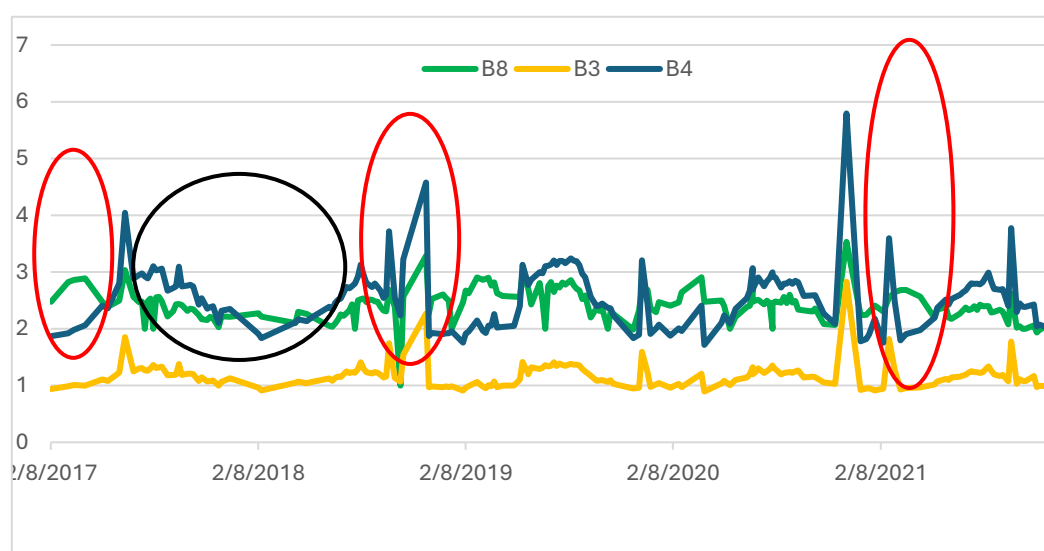
These bands were chosen because they reflect the phenology of the plants studied and absorb part of the sun’s energy in the visible spectrum, making the visible red and blue bands an interesting resource for study. As we move through the NIR, the reflectance increases. Thanks to the different infrared bands, we can recognize multiple aspects related to vegetation that, in addition to indicating the existence of this biological element, can also offer us information on the type of vegetation, humidity levels or growth phase. Crop marks are linked to the presence of buried walls and/or filled ditches in vegetated areas: the first constrains, whereas the second favours, vegetation growth. In both cases, crop marks are named positive and negative. In the case of bare ground, the commonly used proxy indicators are damp and soil marks. The first are variations in soil drainage caused by buried walls, moats, and ditches. The second appears as changes in soil colour due to ploughing that can bring to light stony materials of shallow walls or wet and organic clods of moats and/or earthworks (Belmonte et al., 2018).

The scientific literature and the available satellite dataset indicate that in the Mediterranean area, crop marks are generally evident during spring when the vegetation matures, especially in wheat fields (Belmonte et al., 2018).

As seen in **Error! Reference source not found.**, the behaviour of the selected bands is cyclical. During the period analysed, there was an alteration in the values of the bands studied periodically during the summer period, specifically in the month of August, a factor that is related to the period of summer drought that is characteristic of this area (red ellipses). However, if we account for the trends of the documented vegetation on the studied surface (olive grove) during this period, there is



a halting in the progression of the vegetation, so these variations should not be accounted for in the analysis we are conducting. On the other hand, large “plains” can be seen in the values where the bands appear uniform (black ellipse); in this case, they broadly coincide with the spring period, confirming what is established by Masini et al. for the Mediterranean (Belmonte et al., 2018). Thanks to these long uniform periods, these moments are optimal for the analysis of the growth status of the vegetation and water content since they will provide a clear image of the entire surface studied, allowing us to identify variations in the crop surface that could be linked with the existence of archaeological remains as indicated above, as numerous authors have shown (Agapiou et al., 2014, 2010; Reid, 2016; Verhoeven & Vermeulen, 2016).



Error! Reference source not found.. Behaviour of spectral bands. Red ellipses mark significant alterations. Black ellipse marks the studied chronological period.

Considering these characteristics, we proceeded to analyse the trends of bands B3, B4 and B8 (NIR) and select images with the date of capture as close as possible to the periods identified that met the following characteristics:

- The percentage of cloud cover in the image did not obstruct the studied area.
- The more recent the study date is, the better.

Thus, after selecting the images with the aforementioned characteristics, the images that met the established requirements were as follows:

1. S2A\_MSIL2A\_20180419 T105651\_N0207\_R094\_T30SVH\_20180424T162732
2. S2A\_MSIL2A\_20180504 T105621\_N0207\_R094\_T30SVH\_20180504T131626
3. S2B\_MSIL2A\_20180708 T105619\_N0208\_R094\_T30SVH\_20180708T141851
4. S2A\_MSIL2A\_20210905 T105621\_N0301\_R094\_T30SVH\_20210905T135834

The dates to which these images belonged were (Graphic 1, black ellipse):

April 19, 2018.

May 4, 2018.

July 8, 2018.

September 5, 2021.

Once the series of images on which to calculate the referred indices were selected, they were downloaded from the Copernicus Open Access Hub for subsequent processing with the GIS software mentioned above. From the group of selected images, we performed calculations on the image captured on April 19, 2018, since it provided the greatest amount of information related to the object of this article.

## 4.2. GIS

The software used was QGIS version 3.22.3-Białowieża, in which the results of the analyses carried out on the multispectral image captured with the Sentinel 2A satellite were merged with those obtained from the digital elevation model (DEM) analysis using the LASTool plug-in. For this, we use the least cost path (LCP) algorithms contained in SAGA GIS 2.12.99 included in QGIS 3.28.1-Firenze to calculate optimal routes and the raster tool to calculate the different indices analysed.

### 4.2.1. DEM Generation from Lidar Point Clouds: LASTools

To generate the DEM that will be the base for LCP analysis we used the LASTool plug-in from Rapidlasso (*LASTools*, 2020) is specifically meant for working with lidar data and contains many algorithms necessary for this context. Specifically, the algorithms used were lasmergePro, lasclip and blast2dem (Figure 3).



Figure 3. Workflow carried out for analyse the DEM.

The spatial resolution of the DEM obtained was 5 m/pix, which, as mentioned above, we consider optimal because of the relationship between mesh size and the ability of our team to analyse it in search of the optimal routes using the LCP algorithm. The use of a higher resolution would require more powerful computer equipment (Lugo & Alatrisme-Contreras, 2019). Thus, based on the results obtained that will be presented later, we feel that we have reached a good balance between physical requirements and results.

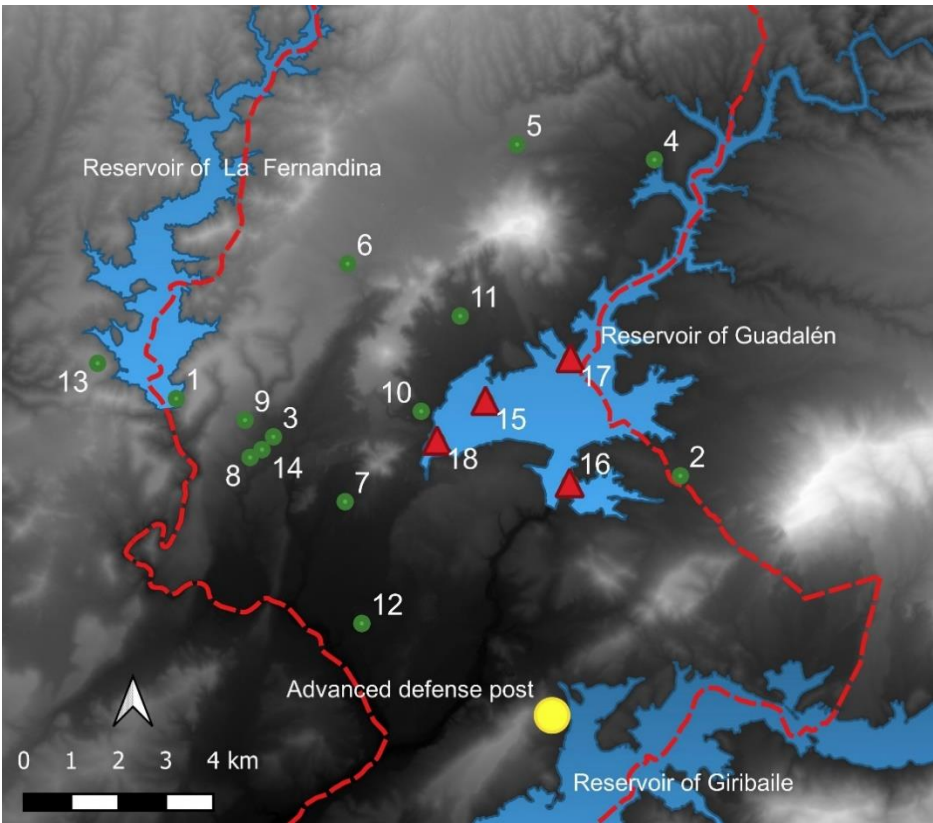
### 4.2.2. Least Cost Path (LCP)

In the absence of other criteria, it is normally assumed that travel routes are optimized with use by the population over time, which has led to the development of the LCP algorithm (Güimil-Fariña & Parcero-Oubiña, 2015) that is of interest to us.

With the application of this methodology, theoretical connections were established between the different archaeological sites and the one identified in this text as an access control post next to the road, which would serve as a guide on which to focus the search for archaeological evidence in the satellite imagery analysis because the spatial resolution provided by the data of the analysed satellite images is 10 m (**Error! Reference source not found.**). On the other hand, the Semi-Automatic Classification Plug-in (SCP) was considered for the semiautomated classification of surface vegetation cover, but in this case, it would not provide clear results because the surface of the terrain was very homogeneous, and consequently, the variation in band values analysed when correlating with the possible existence of archaeological remains, could be easily confused and thus provide a high rate of false positives that would force us to perform a generalized supervised classification of the surface visually. Therefore, with awareness of this limitation and to increase the chances of success in the identification of archaeological remains, we use the optimal tracing generated by the use of LCP as a guide that, despite not offering optimal or perfect results, is very practical (Lugo & Alatrisme-Contreras, 2019), helping and accelerating the identification of possible archaeological evidence (Seaman & Thomas, 2020).

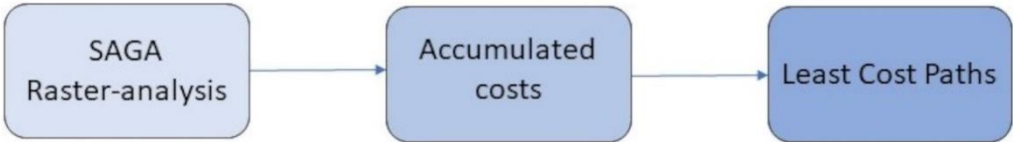
Prior to the analysis of this DEM, we proceeded to eliminate both the reservoirs within the study area and the archaeological sites documented in its interior by using the SHP layer referred to above (**Error! Reference source not found.**). The flat surface created by the dammed water would have caused the route to pass over the surface of the reservoirs since this would represent a lower cost and would therefore distort the result. In turn, there are two moments of data collection that must be considered: first is the date of the survey, in which, due to the long periods of drought, the water reserve of the reservoirs is very low, making it possible to exploit being able to identify archaeological sites that would otherwise be impossible (sites 15 to 18) and, on the other hand, the moment of Lidar

data capture, which have served as the basis for the generation of the DEM at hand, which reflect a higher level in the water reserves of these reservoirs and therefore cover the existing remnants.



**Figure 4.** DEM in which the limits of the reservoirs extracted for their calculations can be seen, as well as the initial point of route calculation (yellow), sites excluded from the analysis (red triangles) and destinations (green dots). In red discontinuous line the limits of the municipal term of Vilches. Source: DEM generated from IGN lidar point cloud by the authors.

Once the DEM was configured and prior to using the aforementioned LCP algorithm, a slope map and an accumulated cost map using the function accumulated cost included in the aforementioned SAGA GIS version, were calculated (Lugo & Alatrliste-Contreras, 2019) (**Error! Reference source not found.**), since it is the base cartography on which this algorithm works to calculate optimal routes between various points.



**Figure 5.** Workflow carried out using Least Cost Path.

The steps referred to here are basic for LCP since it performs its calculations on a raster model of costs between two points. Prior to this, the DEM was reclassified, depending on the slopes documented in its interior, into 4 categories establishing up to a maximum of 15% slope (**Error! Reference source not found.**).

- 0% - 4%
- 4% - 8%
- 8% - 11%
- 11% - 15%



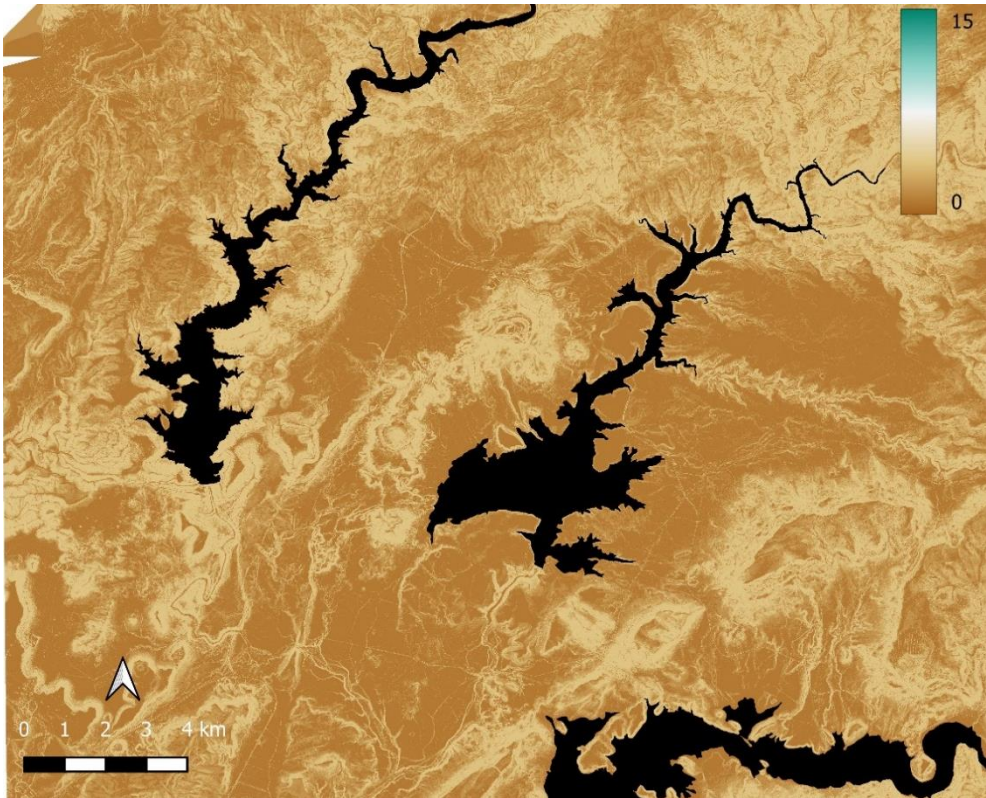


Figure 6. Reclassified slope map. Source: the Author.

Next, we proceeded to the cost calculation base map starting from the advance defense post next to the road to each of the identified sites, identifying a route for each of the points analysed (**Error! Reference source not found.**).

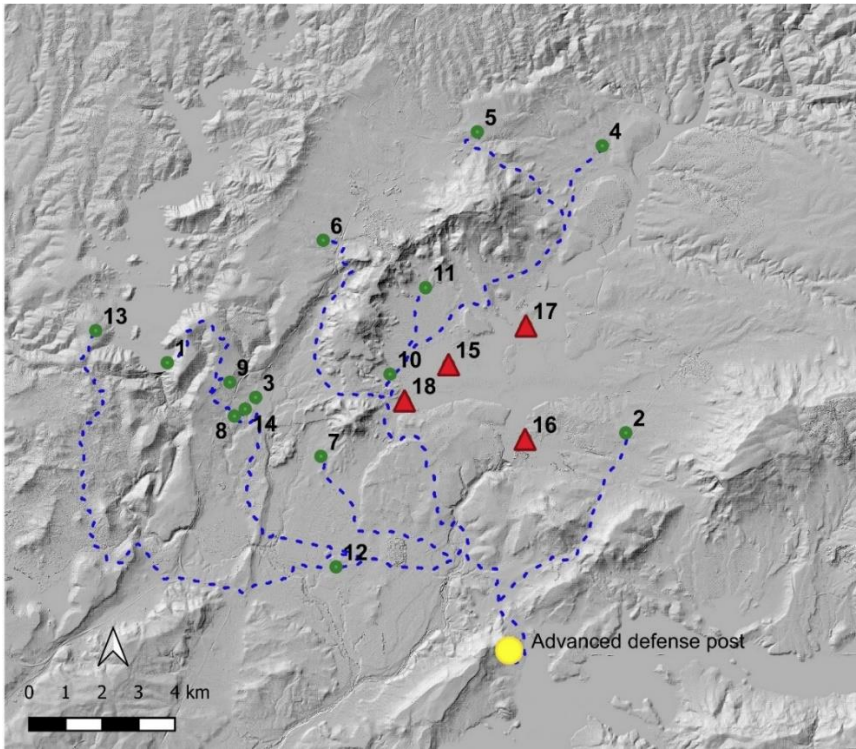
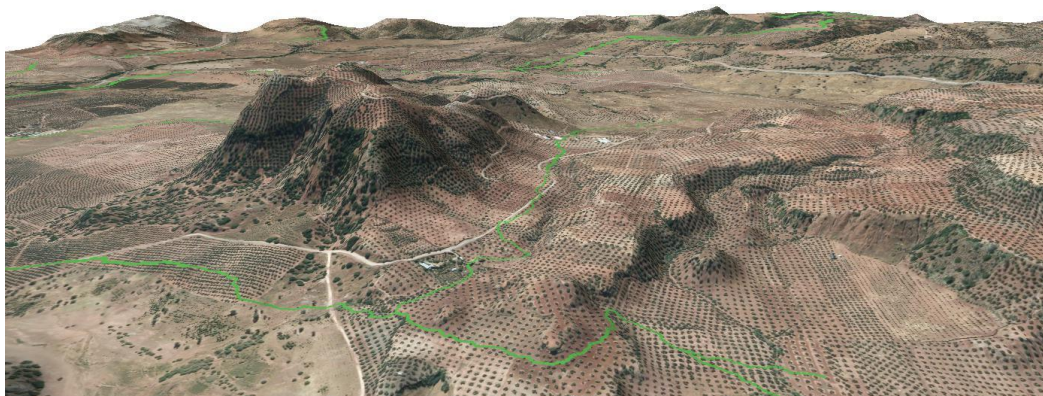


Figure 7. Results after applying Least Cost Path algorithm with possible communication routes between different archaeological sites. Source: Hillshade generated from DEM generated using IGN lidar point cloud by the authors DEM.



Finally, once the routes were identified and to test the reliability of the results obtained, we represented them on a 3D cartography obtained from merging the digital orthophotography of Andalusia and the DEM used (**Error! Reference source not found.**). As seen in the image, the layout aligns well to the orography of the terrain, so this suggests that the precision, within the limitations of the data, is good and that it therefore can be a good guide for the analysis with multispectral data.



**Figure 8.** Superposition of the routes obtained on a 3D model of the terrain made by fusing DEM data and colour Orthophotography (PNOA 2016). Source: the authors.

#### 4.3. Multispectral Satellite Imagery

In our case, as mentioned above, the focus was on bands B3, B4 and B8 (**Error! Reference source not found.**) those are the bands that provide a higher spatial resolution (10 m/pix) and of the Sentinel mission, and for its processing, we used the tools for the map algebra of the QGIS software to provide quantitative data with which we could interpret the information and monitor it in numerical terms. This is why we decided to use the indices contained in this study, such as NDVI, NDWI, and SAVI (Bannari et al., 1995; Dafalla, 2022), since the state of the vegetation reflected in these indices can be indirect indicators of the presence of archaeological remains in the subsoil, also there the most suitable one according the whit the selected bands.

##### Normalized Difference Vegetation Index (NDVI)

The NDVI is one of the most commonly used indicators in vegetation studies and precision agriculture (Bannari et al., 1995; Crippen, 1990). Its foundation lies in the analysis of the wavelength values emitted by the vegetation in the visible red band (B4) and in the near infrared band (B8)<sup>1</sup>. Its result allows us to quickly identify the distribution and type of vegetation throughout the territory, enabling the differentiation of various territorial elements or their qualities.

The formula used for its calculation is:

$$\text{VALUE} = (B08 - B04) / (B08 + B04)$$

The range of possible values during the input of bands oscillates between the values of -1 and 1:

- Negative values are associated with areas of water and snow.
- Positive values close to 0 represent rocky and bare areas that can acquire some vegetation until reaching values close to 0.3, corresponding to bare soils or with early-growth vegetation with soil exposure.

<sup>1</sup> The nomenclature used in these bands is that used in Sentinel satellites, but may vary depending on the satellite used. For example, band 8 corresponds to Band 5 on Landsat 8 satellites.

- Values greater than 0.3 indicate the presence of vegetation in a more obvious way. The closer to values of 1 our data are, the more leafy and the greater vegetation coverage we will find in our study area.

Normalized Difference Water Index (NDWI)

Through the calculation of the normalized differential water index (NDWI), we can identify water masses and areas of high humidity saturation. Thanks to this index, it is possible to identify potential wetlands and delimit wet environments.

As in the NDVI case mentioned above, the potential values obtained from the NDWI range between -1 and 1, with the values describing water surfaces and vegetation with water content or land areas and with the absence of moisture.

Among the diversity of existing methods for calculating this index, we opted for the McFeeters method (2007), which is based on the substitution of the SWIR band for the visible green band, highlighting the water masses.

The formula used for its calculation is as follows:

$$\text{VALUE} = (B03-B8)/B03 + B04)$$

Soil-Adjusted Vegetation Index (SAVI)

This soil-adjusted vegetation index shows a slight variance with respect to the traditional NDVI formula to avoid distortions in the analysis values when the vegetation is on exposed soils.

The SAVI index is more adapted to studies of vegetation analysis in early growth stages or sparse vegetation. In general, SAVI can be a good alternative for any soil where there is a low plant density and the exposure of the soil surface is relevant.

In this case, the exposure of the soil will be considered through the L factor, incorporating it into the formula. This factor buffers the presence of the soil by assigning values between 0 (for areas with high plant density) and 1 (for areas with low plant density). In this way, for soils with a high presence of vegetation, the L factor goes to a value of 0, leaving the equation unchanged and making it equivalent to the usual NDVI equation.

$$\text{VALUE} = ((B08 - B04)/(B08 + B04 + L)) * (1 + L)$$

Due to the characteristics of the soil in the studied area, in which the analysed soil can be classified as moderately exposed, we considered the L factor to be 0.5, and the resulting equation is as follows:

$$\text{VALUE} = ((B08-B04)/(B08 + B04 + 0.5)) * (1 + 0.5)$$

5. Results

After conducting the LCP analysis, we identified a total of 14 routes, one for each of the archaeological sites analysed outside the reservoirs.

As shown in **Error! Reference source not found.**, the average length of these routes was 8,702.21 m.

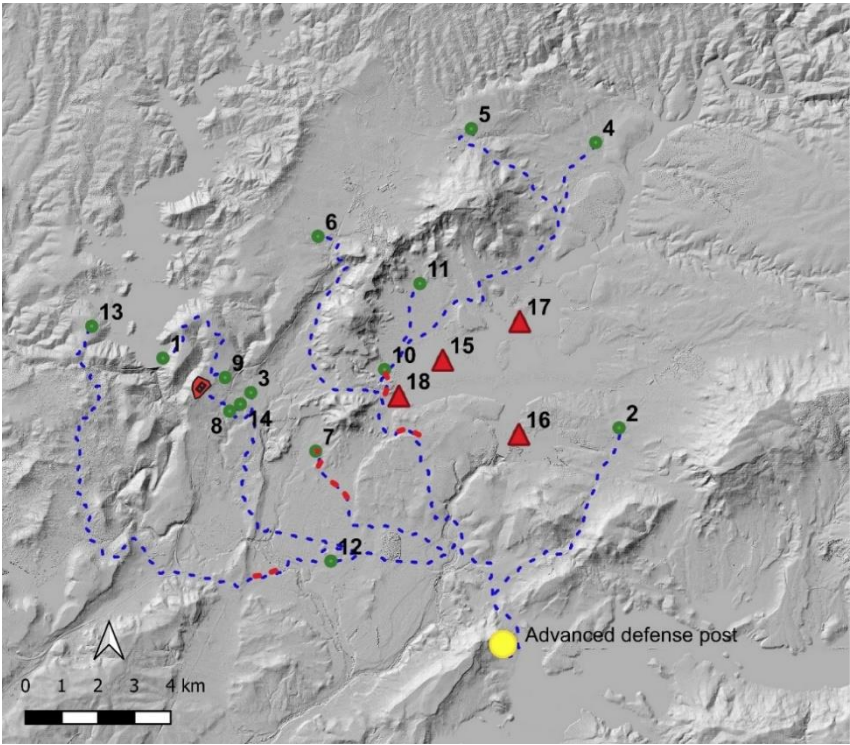
**Table 3.** Archaeological sites on identified routes ends and length.

	END POINT	Distance (m)		END POINT	Distance (m)
1	VALDEINFIERNO I	5.844	8	SAN JULIAN - EL RASO	7.348
2	CORTIJO DE ARQUILLOS EL VIEJO	8.321	9	FUNDICION SAN JULIAN	7.959
3	CORTIJO EL RASO	9.641	10	GALEOTE	7.348

4	CORTIJO HORTALANCA	10.229	11	LLANOS DE VICHI - CERRO MANZANO	6.322
5	CORTIJO JUAN CLAVERO	10.106	12	NECROPOLIS DE LA LEONA	6.321
6	CORTIJO LA FLORINA	6.754	13	PALAZUELOS	22.988
7	CORTIJO LOS GRAJOS	9.408	17	CERRILLO DEL CUCO	8.394

On the other hand, their overlap on the DEM allows us to visually see that the proposed layout is sufficient for our purposes, that is, to establish a guide concerning where to focus the analysis of the Sentinel 2 images.

Thus, thanks to the study we carried out, we were able to document possible remnants of communication channels on the routes. These remnants included an access control post next to the road - Cortijo de los Grajos (7), Necrópolis de la Leona (12) - Palazuelos (13) and an access control post next to the road - Galley 2 (10) **Error! Reference source not found..** As seen in Figures 10–12, the sections are short, but they seem to confirm the reliability of the results of the LCP algorithm.



**Figure 9.** The dashed red line and white arrows show the identified road sections, as well as the discovered archaeological site. Source: Hillshade from DEM generated using IGN lidar point cloud by the authors DEM.

**Error! Reference source not found.** shows the documented lengths, as well as the index that has allowed their identification.

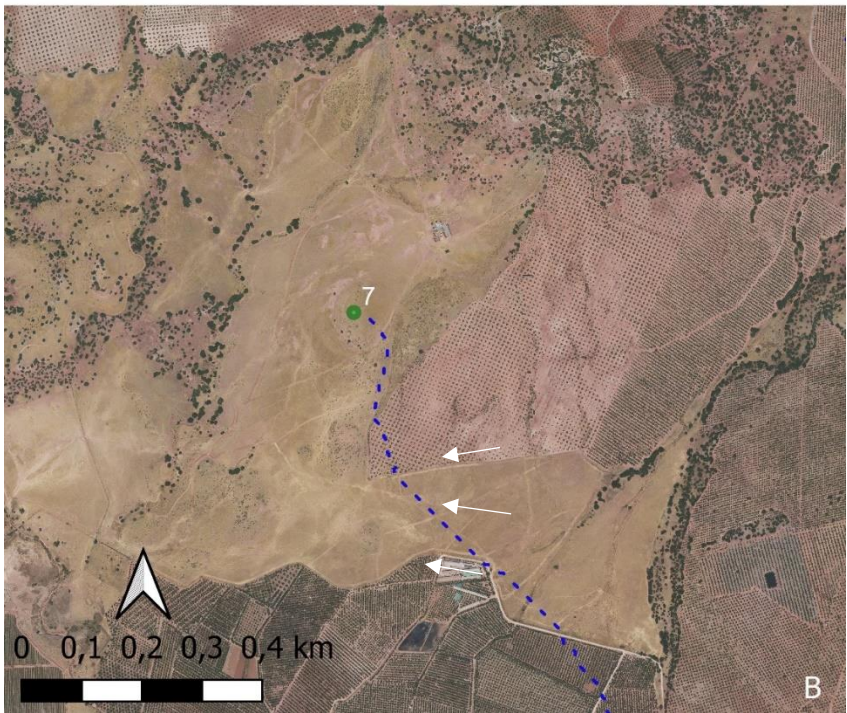
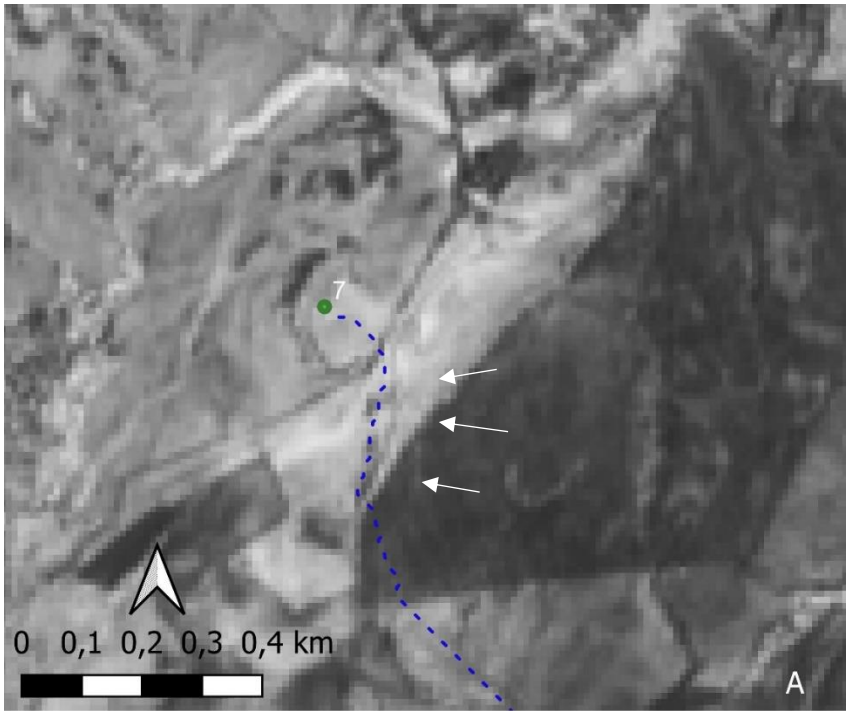
**Table 4.** Sections of routes identified as a percentage of the total route after carrying out the analysis of the multispectral images.

CODE	SITIO	LENGHT (m)	%	INDEX	FIGURE
------	-------	---------------	---	-------	--------



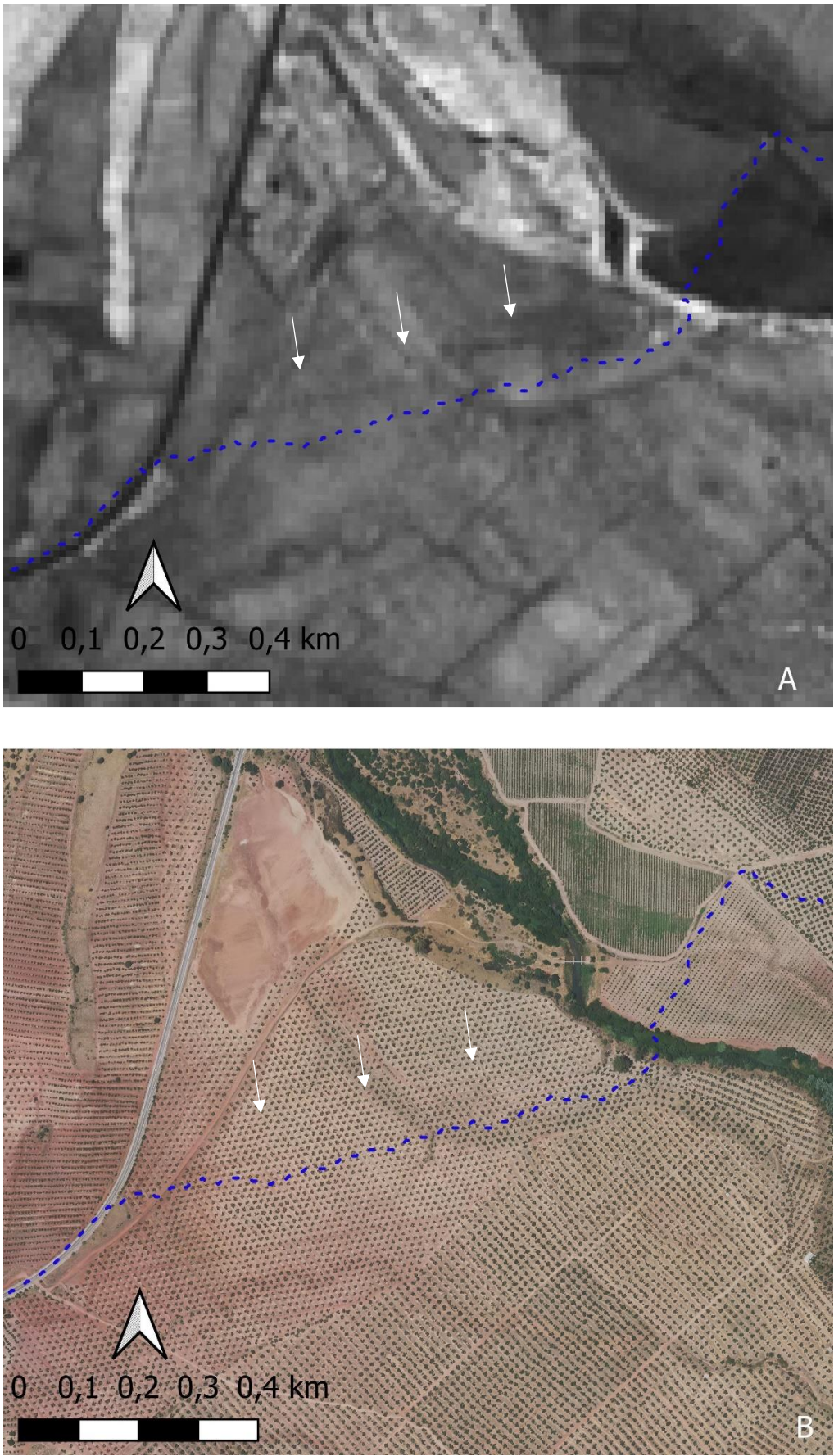
1-7	Advanced defence post-Cortijo de los Grajos	557	5.92	NDVI	10
1-10	Advanced defence post-Galeote	556	7.56	NDWI	12
12-13	La Leona-Palazuelos	667	4.88	SAVI	11

As we can see in the aforementioned images, the existing separation between the remnants of the identified route and the one proposed by LCP is approximately 50 m, and it even reproduces the route marked by it. On the other hand, considering the lengths identified, we consider them to be important sections, since in no case are the documented lengths less than 500 m, which meant the average theoretical identification of approximately 5% of the total route.



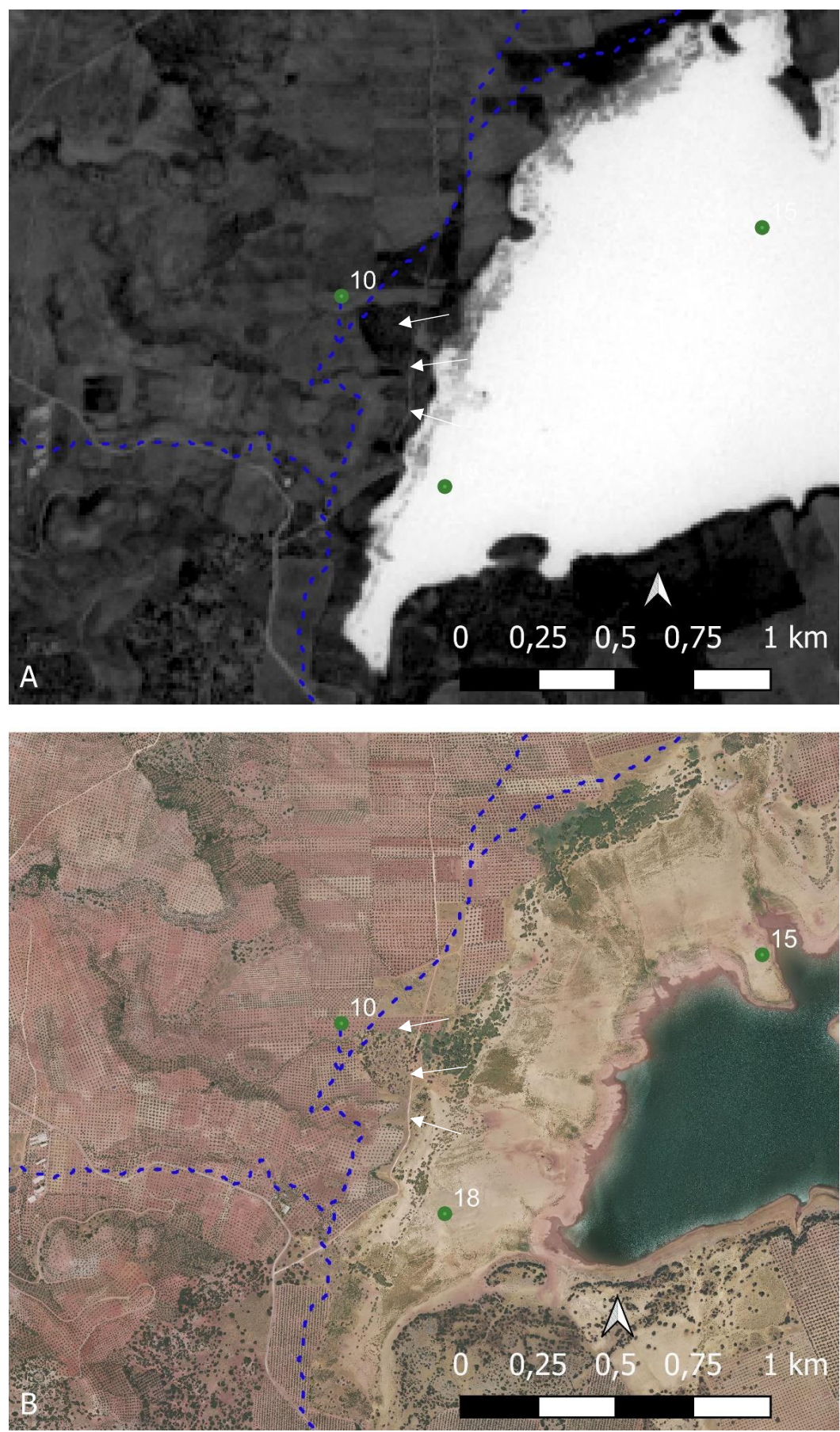


**Figure 10.** A. NDVI 19-04-18. Dashed blue dots LCP results. B. Orthophotography from PNOA 2019. The white arrows mark the identified section.





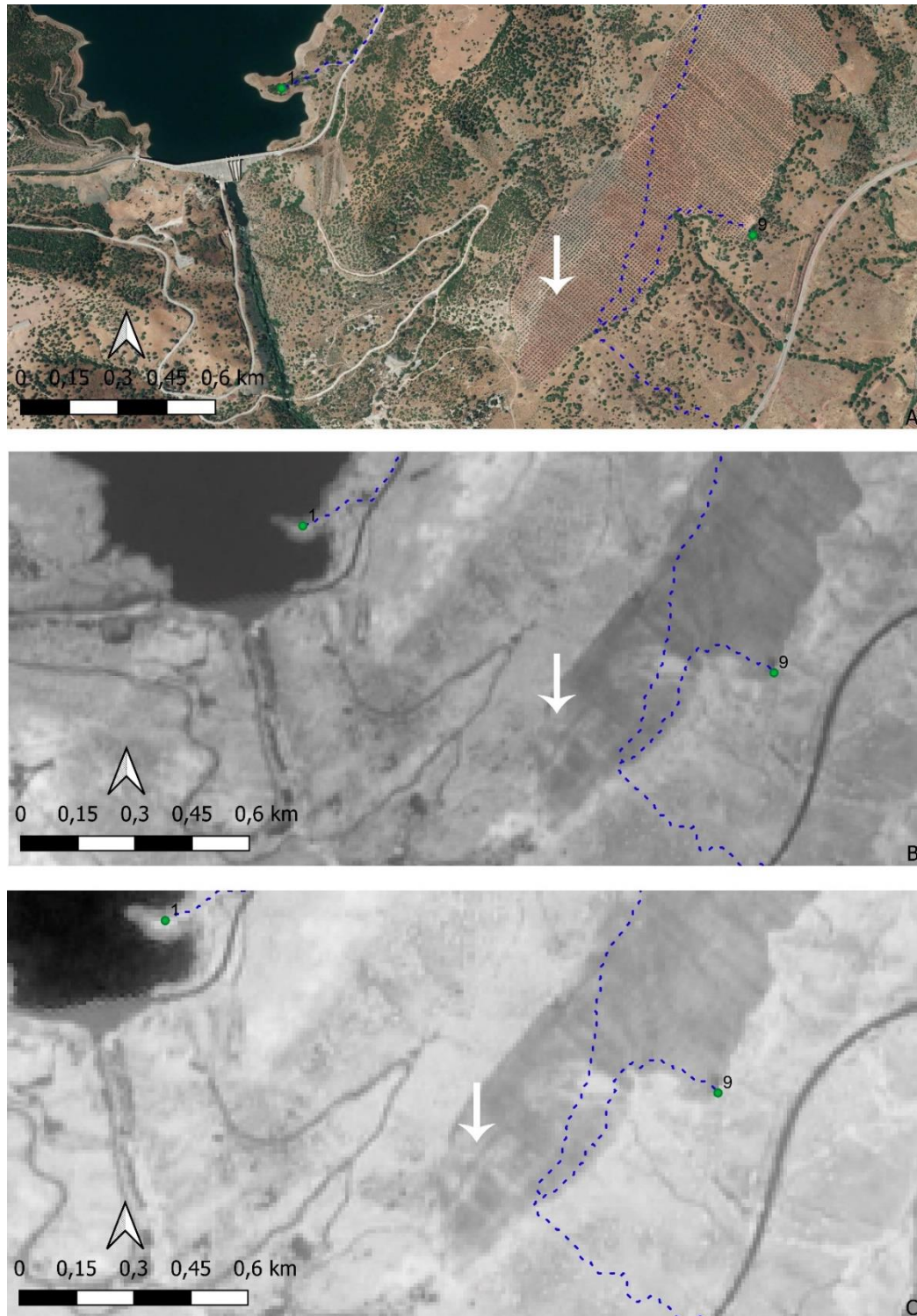
**Figure 11.** A. SAVI 19-04-18 Dashed blue dots LCP results. B. Orthophotography from PNOA 2019. The white arrows mark the identified section.





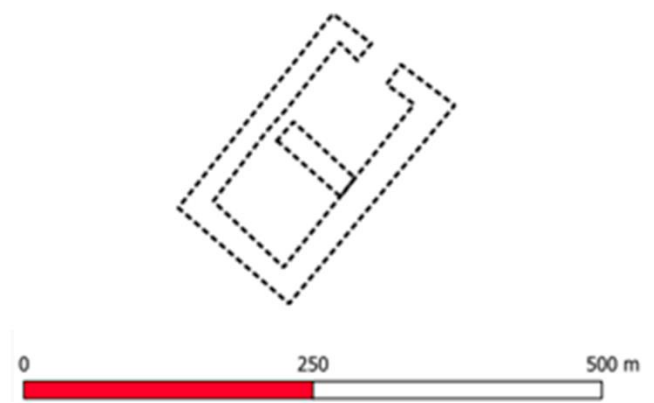
**Figure 12.** A. NDWI 19-04-18 Dashed blue dots LCP results. B. Orthophotography from PNOA 2019. The white arrows mark the identified section.

Finally, we were able to document the presence of a new archaeological site that had not yet been discovered. It is near the San Julián district between the Valdeinfierno mine (1) and the San Julián Foundry (9). Its presence has been identified fundamentally in the NDVI and SAVI analyses, and it went completely unnoticed in the colour orthophotography of Andalusia in 2016 (**Error! Reference source not found.**).



**Figure 13.** Identification of the aforementioned roman archaeological site. A) Coloured Orthophotography (PNOA 2016). B) SAVI, C) NDVI.

The approximate documented dimensions of the detected anomaly were 216.17 x 132.24 m, which represents an approximate area of 2.82 ha. As seen in **Error! Reference source not found.**, its layout is rectangular and oriented NE–SW. At its northern end, it has an opening of 31.37 m that could be identified as the access door to the building, which, in turn, is divided into two rooms separated by a dividing wall that seems to have an access of 12 m close to wall N.



**Figure 14.** Plan of the identified structure.

Regarding the chronology of the site, we still cannot specify anything, since in the surveys carried out in the environment, it has not been possible to document archaeological evidence on the surface that would allow us, on the one hand, to infer its presence and, on the other hand, its chronology. However, based exclusively on the comparison of the dimensions with other Roman villas in the area whose surface is approximately 2 ha (Alejo Armijo, 2019) to which we referred above, as well as the environment in which it is located, close to other well-dated archaeological sites, we propose a chronology for this of the Flavian era, although we are aware that these data must be confirmed or rejected in future works.

## 6. Discussion

In our case study, the results obtained through the use of the LCP algorithm have not been made by accounting for the various factors that can affect the displacement (density of vegetation, distance to natural resources, possibility of flooding of a land, etc.) of an individual through a region and therefore to the choice of one path or another for a certain route. Therefore, the cost calculation has been made based on the slopes of the environment generated from the Lidar. Despite these limitations, its use has proven to be effective for our objectives because it has allowed us to focus the search on specific areas of a wide area with “errors” of approximately 50 m with respect to the identification of sections that otherwise would have gone unnoticed. This approach, despite its simplicity, has proven to be valid for the proposal of expected archaeological evidences related to both archaeological remains of communication routes and buildings that must be confirmed or don’t in future archaeological campaigns.

Satellite images present various drawbacks:

1. Size of the analysed images (approximately 1 GB per image): their high weight makes it advisable that when analysing a time series such as the one analysed in this article (composed of a total of 192 images), it is recommended to process them in the cloud, but as has been seen above, for this you need programming knowledge.
2. Resolution: It is low for some archaeological purposes, so we must account for the entity of the archaeological remains that we are looking for because if these are smaller than 10 m, they will not be able to be identified due to the lack of spatial resolution. In this case, when we proceed to use these images, we must account for the purpose of their use and their limitations, since only in this way can we extract their maximum potential. I.E a road can be identified by its width and by its length, that expand over a territory. What is more, we have to consider that the road width



can be represented in several pixels (depending on the orientation of the road as well as the image) and consequently the mean value of the pixel on which the road is can vary and consequently the representation of the identified road can be seen wider than it is. This doesn't mean that we can measure the width or the it layout but at least we are identifying it presence, or at least, something similar to, that must be reviewed in future archaeological campaigns.

In contrast, its advantages are numerous because they allow us to analyse many images over a long period (in our case, 5 years), calculate several spectral indices that could provide direct or indirect information on the presence of archaeological remains in the subsoil and analyse large areas of the land in a short time.

Regarding the chronology and dimensions provided for the identified findings, we are aware that it is a mere approximation accounting for the environment in which we find ourselves and that the hypotheses established in this regard must be confirmed or not, either with a new archaeological survey or with an archaeological excavation since thus far, we are moving in the field of assumptions.

## 7. Conclusions

The results obtained are the result of the combined analysis of data from different sources and media, all of them from public funds. It is in their origin that the strengths and weaknesses of these sources reside. The strengths include free and open access, while its resolution is low since these products are not oriented to the needs of archaeology. However, along these lines, we have been able to appreciate how their combined use has great potential for the identification of archaeological remains that allow the understanding of the population dynamics of a region.

We must not forget that although the beginning of these works was the prospecting carried out by Alejo et al. and allowed the identification of 18 archaeological sites from the Flavian period, in this case, no sections of communication routes were identified; at the same time, the aforementioned archaeological site was not identified and therefore obliged us to carry out a new survey on the identified areas to confirm their existence. We must bear in mind that these data may experience variations or may reflect other elements that do not necessarily have to be of archaeological interest, which is why a new campaign of focused prospecting is necessary and in future steps either geophysical prospecting or archaeological excavation for their confirmation.

Finally, the identification of 3 possible sections of communication routes as well as a new archaeological site allows us to affirm that the methodology used is effective for the study of large areas of land, reducing time and costs when planning a prospecting campaign, which, in our opinion, should be complemented either with geophysical surveys or with archaeological excavations that allow the established hypotheses to be confirmed or not.

**Funding:** The translation of this work was supported by the Research Group PAIDI-HUM357 (GIPAJ).

## References

1. Agapiou, A., Alexakis, D. D., Sarris, A., & Hadjimitsis, D. G. (2014). Evaluating the potentials of sentinel-2 for archaeological perspective. *Remote Sensing*. <https://doi.org/10.3390/rs6032176>
2. Agapiou, A., Hadjimitsis, D. G., Themistocleous, K., Papadavid, G., & Toullos, L. (2010). Detection of archaeological crop marks in Cyprus using vegetation indices from Landsat TM/ETM+ satellite images and field spectroscopy measurements. *Earth Resources and Environmental Remote Sensing/GIS Applications*, 7831, 78310V. <https://doi.org/10.1117/12.864935>
3. Alaminos Tererno, F. J. (2001). *El medio ambiente en Andalucía en el umbral del siglo XXI*. Junta de Andalucía. Consjería de Medio Ambiente.
4. Alejo Armijo, M. (2019). *Poder y empoderamiento de la arqueología en Giribaile. Arquitectura social y representativa de la cultura ibérica e impacto territorial a través de la romanización*. Universidad de Jaén.
5. Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2), 95–120. <https://doi.org/10.1080/02757259509532298>
6. Belmonte, A., Masini, N., Lasaponara, R., Manzari, P., Marzo, C., & Sabia, C. (2018). On the characterization of temporal and spatial patterns of archaeological crop-marks. *Journal of Cultural Heritage*. <https://doi.org/10.1016/j.culher.2017.12.009>

7. Campana, S. R. L. (2018). *Mapping the Archaeological Continuum. Filling "empty" Mediterranean Landscapes*. Springer Briefs in Archaeology. Springer. <https://doi.org/https://doi.org/10.1007/978-3-319-89572-7>
8. Casana, J. (2014). Regional-Scale Archaeological Remote Sensing in the Age of Big Data. *Advances in Archaeological Practice: A Journal of the Society for American Archaeology*, 222–223.
9. Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, 34(1), 71–73. [https://doi.org/https://doi.org/10.1016/0034-4257\(90\)90085-Z](https://doi.org/https://doi.org/10.1016/0034-4257(90)90085-Z)
10. Dafalla, M. S. (2022). New Remote Sensing Vegetation Index for Recognition of Sparse Vegetation in Arid land. Case Study: Algezira State, Sudan. *Academia Letters, January*. <https://doi.org/10.20935/al4807>
11. Derooin, J. P., Téreygeol, F., & Heckes, J. (2012). Remote sensing study of the ancient Jabali silver mines (Yemen): From past to present. In *Remote Sensing and Digital Image Processing*. [https://doi.org/10.1007/978-90-481-8801-7\\_10](https://doi.org/10.1007/978-90-481-8801-7_10)
12. Google Earth Engine. (n.d.). <https://Earthengine.Google.Com/>.
13. Güimil-Fariña, A., & Parcero-Oubiña, C. (2015). "Dotting the joins": a non-reconstructive use of Least Cost Paths to approach ancient roads. The case of the Roman roads in the NW Iberian Peninsula. *Journal of Archaeological Science*, 54, 31–44. <https://doi.org/10.1016/J.JAS.2014.11.030>
14. Junta de Andalucía. (n.d.). Instituto de Estadística y Cartografía de Andalucía. <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/ProdCartografia/Ortofotografias/Orto16.Htm>.
15. Lasaponara, R., & Masini, N. (Eds.). (2013). *Satellite Remote Sensing. A new tool for archaeology*. Springer. <https://doi.org/10.1007/978-90-481-8801-7>
16. LASTools. (2020). <http://lastools.org/>
17. Lugo, I., & Alatrste-Contreras, M. G. (2019). Nonlinearity and distance of ancient routes in the Aztec Empire. *PLoS ONE*, 14(7). <https://doi.org/10.1371/journal.pone.0218593>
18. Martín, B. P., Martínez, A. R. S., Hernández, J. D., García, M. E. C., & Alcázar, G. V. (2020). El programa copernicus para la monitorización del territorio y los objetivos del desarrollo. *El Programa Copernicus Para La Monitorización Del Territorio Y Los Objetivos Del Desarrollo Sostenible*, 53(9), 120–125. <https://doi.org/10.7419/162.13.2020>
19. Martín, C. M. (n.d.). LOS ORETANOS. UNA VISIÓN DESDE EL TERRITORIO, LA SOCIEDAD Y LA IDEOLOGÍA.
20. McFeeters, S. K. (2007). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. <https://doi.org/10.1080/01431169608948714>, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
21. Monterroso Checa, A., & Martínez Reche, T. (2018). COSMO SkyMed X-Band SAR application – combined with thermal and RGB images – in the archaeological landscape of Roman Mellaria (Fuente Obejuna-Córdoba, Spain). *Archaeological Prospection*, 25(4), 301–314. <https://doi.org/10.1002/arp.1709>
22. QGIS. (n.d.). QGIS Un Sistema de Información Geográfica libre y de Código Abierto. Retrieved February 2, 2019, from <https://qgis.org/es/site/>
23. Reid, S. H. (2016). Satellite Remote Sensing of Archaeological Vegetation Signatures in Coastal West Africa. *African Archaeological Review*. <https://doi.org/10.1007/s10437-016-9222-2>
24. Seaman, A., & Thomas, L. S. (2020). Hillforts and Power in the British Post-Roman West: A GIS Analysis of Dinas Powys. *European Journal of Archaeology*, 23(4), 547–566. <https://doi.org/10.1017/eea.2020.19>
25. Sierra Morena Andaluza. (n.d.). 131–187.
26. Stewart, C. (2017). Detection of archaeological residues in vegetated areas using satellite synthetic aperture radar. *Remote Sensing*, 9(2). <https://doi.org/10.3390/rs90201018>
27. Stewart, C. (2020). SAR for Archaeological Prospection in Europe and in the Middle East (pp. 59–84). [https://doi.org/10.1007/978-3-030-10979-0\\_5](https://doi.org/10.1007/978-3-030-10979-0_5)
28. Tapete, D., Banks, V., Jones, L., Kirkham, M., & Garton, D. (2017). Contextualising archaeological models with geological, airborne and terrestrial LiDAR data: The Ice Age landscape in Farndon Fields, Nottinghamshire, UK. *Journal of Archaeological Science*, 81, 31–48. <https://doi.org/10.1016/J.JAS.2017.03.007>
29. Tapete, D., & Cigna, F. (2017). Trends and perspectives of space-borne SAR remote sensing for archaeological landscape and cultural heritage applications. *Journal of Archaeological Science: Reports*, 14, 716–726. <https://doi.org/10.1016/j.jasrep.2016.07.017>
30. Tapete, D., & Cigna, F. (2019). Detection of archaeological looting from space: Methods, achievements and challenges. In *Remote Sensing* (Vol. 11, Issue 20). MDPI AG. <https://doi.org/10.3390/rs11202389>
31. Verhoeven, G., & Vermeulen, F. (2016). Engaging with the canopy-multi-dimensional vegetation mark visualisation using archived aerial images. *Remote Sensing*, 8(9). <https://doi.org/10.3390/rs8090752>
32. Zanni, S., & De Rosa, A. (2019). Remote sensing analyses on sentinel-2 images: Looking for Roman roads in Srem region (Serbia). *Geosciences (Switzerland)*, 9(1). <https://doi.org/10.3390/geosciences9010025>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.