**Introduction**

Southern and Western Iberia are still considered one of the last refugia for Middle Paleolithic populations. The claims for a late survival of Neanderthals in these areas are mostly based on radiocarbon results from the sites of Cueva Antón (Spain), Gorham’s Cave (Gibraltar) and Gruta da Oliveira (Portugal), all dated to ca. 37 ka cal BP or later. Recently, however, all these dates have been questioned due to possible age underestimations.

A consensus does exist on the fact that currently available archaeological data south of the Ebro is fragile, and that the Middle Paleolithic record along the western edge of Iberia, in particular, is still basically unknown.

Currently, more than 80 absolute dates establish that the Middle Paleolithic occupation of westernmost Iberia occurred from ca. 200 to 37 ka cal BP. Unfortunately, however, a significant number of the available results presents very large standard deviations, making them unreliable to build a detailed regional sequence. In other cases, problems related with sample collection, or the lack of resolution of dating methods in the past, might be behind significant underestimations of the results. This is the case of, for example, the site of Foz do Enxarrique, which until very recently was considered one of the latest Middle Paleolithic sites in Iberia (ca. 34 ka cal BP) but new OSL results, using post infra-red stimulated luminescence, place the occupation of the site at ca. 44 ka cal BP.

**Technology**

**Settlement**

**Regional setting**

The territory considered for this study is a rectangular strip with a maximum North-South length of c. 560 km and a maximum East-West width of c. 220 km, corresponding with the current political borders of Portugal. Altitudes range from 0 m above current sea level (a.c.s.l.) to 1993 m a.c.s.l. in the mountain peak of Serra da Estrela in the northeast.

From north to south, the primary rivers are the Minho, Douro, Mondego, Tagus and the Guadiana. Most of these rivers flow from east to west disgorging in the Atlantic Ocean. The coastline has c. 1794 km in length.

Three great structural geologic units can be distinguished (1) the Hesperic Massif, with crystalline metamorphic and igneous rocks, and Paleozoic formations; (2) the western and southern Meso-Cenozoic rims (3) and the Tagus and Sado Meso-Cenozoic basin.

To the north the landscape is mountainous in the interior areas with plateaus, while the south features mostly rolling plains with a climate somewhat warmer and drier than the cooler and rainier north.

**Materials and methods**

**Archaeological sites**

A dataset comprising all Middle Paleolithic sites reported in Portugal was created using the Endovélico – Archaeological Management and Information System, curated by the Direção Geral do Património Cultural (DGPC) (<http://arqueologia.patrimoniocultural.pt/>). This initial dataset included all entries with associated geographic information, in a total of 274 locations. Most of these sites, however, are dubiously attributed to the Middle Paleolithic, since in most cases there is no associated dating, nor reporting of any clear evidence of distinctive Middle Paleolithic stone tools within the assemblages. These inconsistencies are a result of the fact that many of the reported locations are surface findings, attributed to the Middle Paleolithic based on the rudimentary aspect of the lithic materials. To avoid, as much as possible, the inclusion of erroneous information in the database we adopted the filtering scheme presented in Figure XXX. Thus, using the initial list we cross-check information with data from publications and reports, subdividing the sample into two other sub-samples: the first, represented by 54 sites, integrating all contexts for which we were able to confirm their chronological attribution to the Middle Paleolithic, either by absolute or relative chronology; the second, comprising 46 sites, composed of all the contexts for which we were able to collect lithic raw materials information.

**GIS methods and considered variables**

The location of sites included in the short samples were checked and validated by visual inspection using satellite imagery in Google Earth. From there, a KML file was exported and imported into a Geographical Information System (ESRI ArcGIS v.10.6). Digital geographical information was added to the same database, including: a 30X30-m pixel Digital Elevation Model (DEM), obtained from the Shuttle Radar Topography Mission (Farr et al., 2007); a 1:25 000-scale water courses vector layer; and a 1:1 000 000 Lithological Map raster layer (obtained from <https://sniamb.apambiente.pt/content/geo-visualizador>).

Elevation, Slope, Aspect (see Burrough et al., 2015 for the definition and calculation details), and Background Lithology variables were extracted from these datasets (see Supplementary Information XXX), calculated as the mean values for a 20-m radius influence area for each site, using zonal algorithms in ArcGIS.

Distance to closest river was obtained using Euclidean calculation and using all hydrological classes in the dataset. Background lithological information was organized into two major categories: sites located in landscapes dominated by carbonate rocks and sites located in any other lithological background.

Additionally, sites were also characterized as open-air or cave/rockshelter sites, and the dominant raw material in each site, measured as the material representing more than 60% of all lithic artifacts in the assemblage, was recorded.

The results of attribute extraction were exported from ArcGIS as a series of tabular CSV files and are available, together with all the raster layers used for the analysis, site locations and the scripts used for the statistical analysis presented below, in our online research compendium (XXX).

**Statistical analysis**

All datasets concerning both spatial and archaeological variables were imported into the R Programming Environment for statistical analysis.

Two types of approach were taken in this regard. First, a Principal Component Analysis (PCA) was used to initially explore any latent patterns of association among the different spatial and archaeological variables. Second, to analyze each individual geographical variable and check if the location of sites could be attributed to a random choice, a series of Wilcoxon Rank Sum tests were performed. This test was used here as a nonparametric equivalent of the paired student's *T*-test because the population data did not follow a normal distribution. Following standard approaches in spatial pattern analysis comparisons were made between site variable distributions and random sample variable distributions. This was only applied in the case of open-air sites since this type of sites can presumably be located in any point across the landscape. For this, a total of 360 random points was generated in ArcGIS, using the Douro river valley as a northern limit for their spatial distribution. The total amount of points was based on the standard approach of using ten times more random points than archaeological sites, and the Douro valley was used as a limit because our database did not include any archaeological sites situated to the north of that river.

Finally, since the location of caves is limited in terms of lithological background, the comparison for this type of sites was not done using random points generation but using a set of 116 caves where archaeological deposits have been confirmed to exist.

**Results**

**General patterns**

Figure XXX presents the distribution of the 54 sites included in our short sample. A first evaluation of the distribution of the sites across the territory reveals a more dense concentration of sites in Central Portugal, a very weak presence of sites in northern regions, with only one site located in the Côa River valley (Aubry et al., 2020), and no sites located to the North of the Douro valley. Most of the sites present in this sample are open-air locations (n = 36), being noteworthy that both coastal (using current coastline as reference) and inland territories appear represented. Table XXX summarizes the median and interquartile ranges for each of the geographical variables used. Except for Slope, all remaining variables are not significantly different between open-air and cave sites location. In fact, cave and rockshelter sites tend to be located in areas with steeper slopes than open-air sites. In terms of elevation, the maximum altitude at which archaeological sites are found is 512 meters above current sea level (a.c.s.l.), corresponding this value to the cave site of Lapa do Picareiro, in central Portugal (Benedetti et al., 2019).

**Comparison with random samples**

**Principal Component Analysis**

Three key assumptions for PCA were considered before applying the method to our dataset: (1) the Bartlett’s Test for Sphericity p-value result of 0.002 rejects the null hypothesis that the intercorrelation matrix comes from a noncollinear population; (2) the Kaiser-Meyer-Olkin (KMO) value of 0.57 approves sampling adequacy; (3) and the determinant of the correlation matrix is positive at 0.578.

Table X presents the ‘loadings’ of each variable in the composition of the two first components, which represent, respectively c. 37.5% and c. 22% of the overall variability. Latitude and Slope were the variables that explained most variance in PC1, while Altitude and Distance to rivers explain most of the variance in PC2. Conversely, in PC1, the variables with less impact on the variability, *i.e.* Altitude, Aspect and Distance to water courses, can be considered the most important variables for settlement location, since their more restricted values represent the most sought after qualities for site occupation.

When plotted

**Discussion**

**Archaeological implications**

**Biogeographical/paleoecological implications**

**Conclusion**