# Neural Networks to Map Archaeological Lithic Artefacts

Joaquín Jiménez-Puerto[1*], Gianni Gallello[2)], Teresa Orozco[3)]

[1*)] Department of Prehistory, Archeology and Ancient History (PREMEDOC Research Group). University of Valencia. Av / Blasco Ibáñez 28, 46010, Valencia, Spain; email: joaquin.jimenez@uv.es; https://orcid.org/0000-0001-9760-9602

[2)] Department of Prehistory, Archeology and Ancient History (PREMEDOC Research Group). University of Valencia. Av / Blasco Ibáñez 28, 46010, Valencia, Spain

[3)] Department of Prehistory, Archeology and Ancient History (PREMEDOC Research Group). University of Valencia. Av / Blasco Ibáñez 28, 46010, Valencia, Spain

http://doi.org/10.29227/IM-2024-01-41

Submission date: *15.4.2023* | Review date: *28.5.2023*

## *Abstract*

*An analytical non-destructive strategy to chemically characterize lithic artefacts has been developed. Around 100 archaeological lithic materials found in Neolithic-Chalcolithic sites in the Mediterranean region of the Iberian Peninsula and nowadays stored in different museums of the Valencian Community (Spain), were studied. The materials belong to different typologies of rock (diabase, sillimanite, ophite and amphibolite) and were analysed employing portable energy dispersive X-ray fluorescence spectroscopy (pXRF) directly in the rock surface. The obtained data were processed through neural networks protocol, specifically the so-called Kohonen networks or Self Organised Maps (SOM), to map the geologic samples. This self-organized topological feature maps are suitable to deal with multidimensional representations and map them in a two-dimensional space of neurons, following an unsupervised learning protocol. SOM is used to reduce multidimensional data onto lower-dimensional spaces and clustering procedures. As a result, SOM create spatially organized representations, which enhance the discovery of correlations between data. In this case the method has enabled the evaluation of elemental markers related to each rock type behaving as a fine hidden pattern detector and so understand the possible advantages and disadvantages of the analytical method employed to define provenance issues. The attribution suggested by statistics is mainly driven by the composition of rocks essential minerals which are linked to the different petrogenetic conditions. The results showed that in most of the cases the distribution and dispersion of the chemical profile depend of the kind of rock, and clearly suggest that a good way to identify stone tools raw material procurement is to look for elemental markers, being the prior step to create an approximation to ancient exchange networks and their evolution in a diachronic axis.*

*Keywords: neural networks, archaeological lithic artefacts, non-destructive strategy, Kohonen networks, self organised maps (som), Dolerites*

**Introduction**

Polished stone tools are common archaeological artifacts allover the Mediterranean basin of the Iberian Peninsula. Previous works (1,2) highlighted the potential enclosed in the characterization and identification of raw material outcrops to enable the study of plausible exchange networks during the late Neolithic and Copper Age. This work, carried out in the frame of the project NEONETS[1], aims to develop a methodological approach for the classification of unknown lithic material, getting a deeper dive in the available analytical methods to perform this task in a non-invasive way. More specifically the use of a portable X-ray fluorescence (pXRF), which provides a reliable tool to determine the concentrations of a set of elements present in lithic artefacts through a semi-quantitative method. The benefits of using a non-destructive approach for characterising lithic artefacts include the preservation of the artefacts, as well as the possibility to obtain information without altering or damaging them. gaining insights into the provenance of the materials without destroying them (3,4).

Due to the complexity of the results obtained statistical tools must be employed for data processing. PCA is one of the most popular analyses which have the potential to provide good classification capabilities, nevertheless, due to the nature of our data, there is a more suitable tool to treat complex data: neural networks. They offer a flexible and powerful approach to classification that can handle a wide range of problems and applications (5–7). Between all the possible architectures, self-organized-maps (SOMs) stand out for their clustering, visualization, and classification capabilities. They can produce a low-dimensional representation of high-dimensional data while preserving similarity relations between the presented data items. Additionally, SOMs are unsupervised and use competitive learning to learn from the input data without explicit labels or targets. This makes them useful for exploratory data analysis and pattern recognition tasks where the underlying structure of the data is not well understood (8). It is the aim of this preliminary work to explore the classification possibilities of the SOM, to characterize properly each lithic item using the semi-quantitative pXRF elemental analysis, in order to create a calibration set. This calibration set will be employed in future works to perform a supervised learning architecture to correctly characterize unknown lithic items.

[1] NeoNetS. A Social Network Approach to Understanding the Evolutionary Dynamics of Neolithic Societies (C. 7600-4000 cal. BP). Prometeo/2021/007. GVA.

## Material And Methods

### The Sample

At this preliminary stage of the work, we have focused on the acquisition of a statistically significant sample by analysing many lithic assemblages. To do so the collections of many local museums have been analyzed by a pXRF device with a Vanta C Series Handheld XRF Analyzer including rhodium (Rh) anode 40 kV X-ray tube, SDD (Silicon Drift Detector). The measurement spot is characterised by Single-click 3 mm diameter collimation. Acquisition time was 60 s and calibration GeoChem 2-beam METHOD-G2-VCR was used. A wide area was selected that comprises the provinces of Castellón, Alicante and Valencia (see figure 1). Archaeological sites with levels from the Late Neolithic and Copper Age were selected due to their frequency and variability. Most of the selected items have already been studied and geologically characterized in previous works (9), fact that enables us to check the validity of further classifications. Nevertheless, some of them are not still determined and will be consequently used as unknown items. In addition, some non-archaeological outcrops of raw material have also been analyzed, to enable provenance studies in future works.

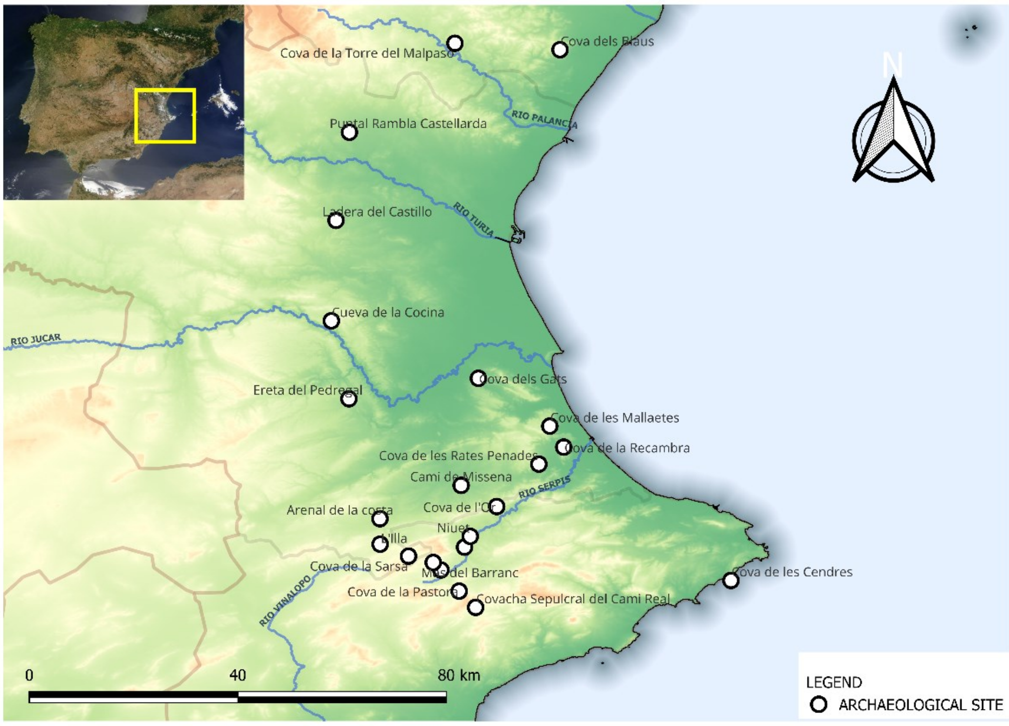


Fig. 1. Localization of the sites providing sampled artifacts.

The multielement quantification was conducted through the analysis of different points for each artifact, resulting in a database. The items were organizad attending to their mineralogical characterization in the following classes: dolerites, silimanites, schists, amphibolites, hornfels and unknown items, with more than 800 measurements. They provided quantifications for the following elements: Mg, Al ,Si ,P ,S ,K ,Ca ,Ti ,V ,Cr ,Mn ,Fe ,Co ,Ni ,Cu ,Zn ,As ,Rb ,Sr ,Y ,Zr ,Nb ,Mo ,Sn ,Sb ,Ba ,Pb ,Bi. In order to combine all the observations of a given item we have calculated the average of all of them, for each element. Moreover, we needed to address the problem of those readings that provide a result under the limit of detection from the pXRF. In order to do so we have separated the items by class, to look for the absolute minimum value corresponding to each class. Next, all the readings under the limit of detection have been substituted by the minimum value of their class. As a result of all the pre-processing a sample of 235 items is available, and is uploaded in Zenodo repository (10).

### Data analysis through SOM

Self-Organizing Maps, have several advantages over other methods of classification such as PCA. Neural networks can model complex nonlinear relationships between input and output variables, which may be difficult or impossible to capture using linear models or rule-based systems. They are also useful for prediction and decision-making tasks due to the ability to learn patterns and rules which are hidden behind the data structure. SOM can perform multiple computations simultaneously, which makes them suitable for processing large amounts of data quickly, and are very tolerant to the existence of outliers in the input data (11–13). Moreover, SOMs can be used to visualize high-dimensional data by mapping it onto a two-dimensional grid, which allows humans to better understand and interpret the data, by detecting clusters of similar patterns without supervision, which makes them useful for feature detection and extraction tasks, as well. We can't either forget that SOMs preserve the topological relationships between input data points in their feature maps, which can help identify patterns and structures that may not be apparent in the original data (8). Overall, neural networks offer a flexible and powerful approach to classification that can handle a wide range of problems and applications. This particular work focuses on the clustering and classification properties of the SOM which will be useful to build a calibration set of items. This set is being built by evaluating the composition of each lithic artifact, attending to their elemental features. In a first stage an exploration of the data will be conducted through SOM, using all the items. After that SOM will be employed to explore the possibilities of creation of a calibration set, trying to classify some unknown elements. All the data analysis were conducted using R (version: 4.0.2) (14). Furthermore, the following packages were employed: plotly (version: 4.10.0) (15) and kohonen (version: 3.0.11) (16).

In order to tackle this task SOMs will map high-dimensional data onto a low-dimensional grid allowing them to identify clusters of similar patterns without supervision. To use SOMs for clustering data packages, one would typically start by defining a set of input features that describe the relevant characteristics of the data. These features include in our case the concentration of the different elements provided by pXRF. Next, we will train a SOM using a set of training data. During training, the SOM adjusts its weights and biases to minimize the difference between the input data and the corresponding output neurons on the grid. This process results in a feature map that captures the underlying structure of the input data in a low-dimensional space. Once the SOM has been trained, one can use it to cluster new data packages by mapping them onto the feature map and identifying which output neurons they are closest to. Data packages that are mapped to nearby output neurons are likely to be similar in terms of their underlying characteristics and may belong to the same cluster.

**Results**

The SOM gave us an idea of the relevance of each element that can be seen in figure 2. It must be noted that the first five more relevant elements detected by SOM are Al, Nb, Fe, Ti, Bi and Sb, but as previous works have stated, the values of Al and Fe, as lighter elements are often disregarded since the analytical results are affected both by the low penetration depth, together with surface contamination and weathering (17–19). Ti also must be considered with caution due to low penetration depth (19). Therefore, Nb, Bi and Sb should be considered as good proxy elements for the classification and clustering (17), and can be seen in figure 3. The powerful visualizations that SOMs provide, enable a first-glance exploration that shows four main groupings: the silimanites, the schists (which are divided in those which contain Ca and those who don't), the amphibolites and the dolerites. But the detection shows a best-fit for 6 clusters. In addition, some small groups such as hornfels are also very close to each other, although they don't have a cluster on their own. Archaeological dolerites and non-archaeological raw materials outcrops conform a very well delimited cluster despite the existence of a couple of outliers. Moreover, many of the unidentified rocks are grouped together and couldn't be classified easily in the initial explorations. As a pivotal point of this work, we will try to examine to possibilities to classify the unknown items through a calibration set.

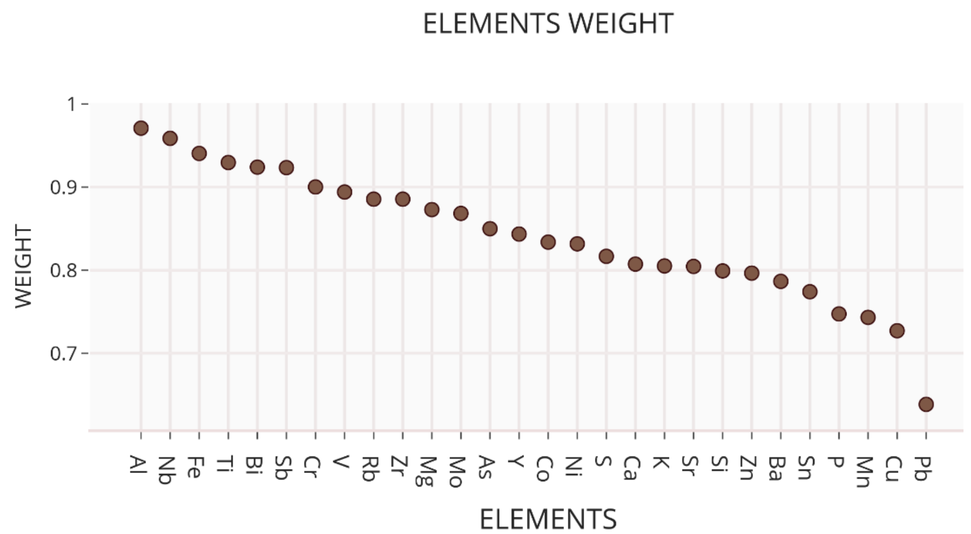


Fig. 2. Elements weights for the SOM with 6 clusters.

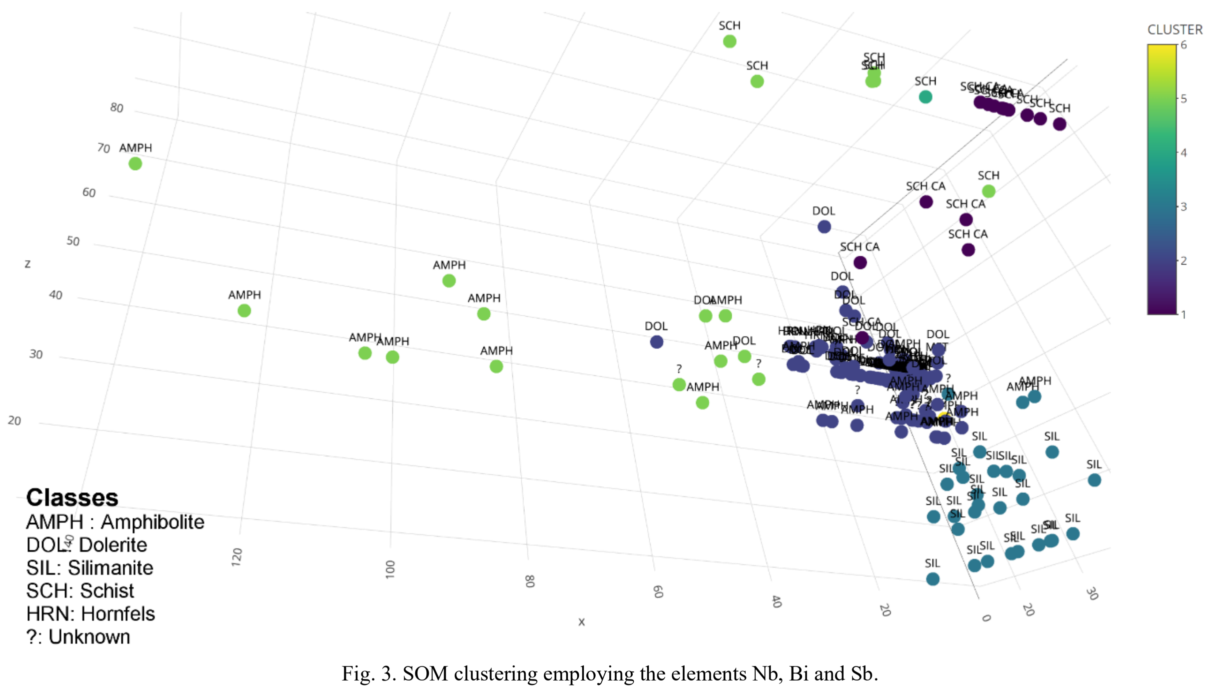


Fig. 3. SOM clustering employing the elements Nb, Bi and Sb.

As a first approximation, we tried to characterize some unknown items using the previously known ones. In order to do that, we have grouped the already characterized archaeological items, excluding the outliers to limit the noise. We have chosen those groups with a greater number of elements. Thus, 4 groups have been created, corresponding to dolerites, amphibolites, silimanites and schists (Table 1). The median of each element has been calculated for each group, in order to properly visualize the estimated situation coordinates for each group. The results can be seen in figure 4, showing that silimanites and schists are easily differentiated. Dolerites seem to be properly classified, but they are not easily distinguished from amphibolites, that despite being in the peripheral part of the cluster are not separated. Finally, some of the unknown samples are closer to the dolerite and other to the amphibolite while others cannot be related with any of the lithic types plotted in the model.

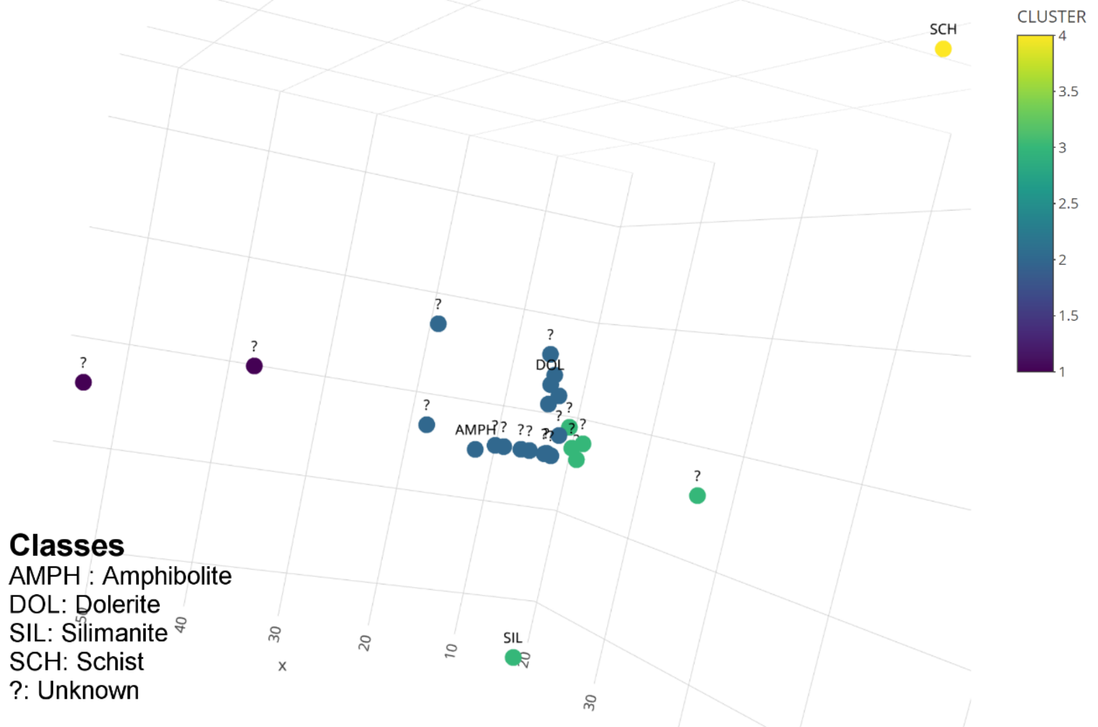


Fig. 4. SOM clustering for unknown elements, with an exploratory calibration set.

Tab. 1. Lithic artifacts and the most relevant elements mean concentrations employed for the calibration set exploration. The concentrations are expressed as mg/kg.

| ID | Class | Al | Nb | Fe | Ti | Bi | Sb |
|---|---|---|---|---|---|---|---|
| 56 | Dolerite | 84755 | 7 | 74721 | 8418 | 22 | 43 |
| 112 | Silimanite | 316228 | 10 | 2186 | 343 | 29 | 17 |
| 139 | Schist | 101809 | 21 | 68785 | 6855 | 64 | 64 |
| 192 | Amphibolite | 82463 | 14 | 87669 | 13481 | 23 | 35 |
| 207 | ? | 8035 | 3 | 1940 | 710 | 23 | 36 |
| 208 | ? | 92893 | 8 | 69174 | 9562 | 21 | 36 |
| 213 | ? | 8761 | 3 | 500 | 643 | 23 | 36 |
| 214 | ? | 9469 | 3 | 3884 | 774 | 21 | 39 |
| 215 | ? | 98486 | 19 | 57581 | 11563 | 21 | 37 |
| 216 | ? | 344185 | 3 | 1077 | 420 | 38 | 36 |
| 218 | ? | 28556 | 3 | 14253 | 1307 | 22 | 37 |
| 219 | ? | 91878 | 38 | 103228 | 14646 | 21 | 40 |
| 220 | ? | 70430 | 11 | 101396 | 12792 | 21 | 36 |
| 221 | ? | 85203 | 7 | 68556 | 6872 | 21 | 36 |
| 222 | ? | 92736 | 5 | 72160 | 7074 | 21 | 36 |
| 223 | ? | 90159 | 4 | 66265 | 7813 | 21 | 38 |
| ID | Class | Al | Nb | Fe | Ti | Bi | Sb |
| 225 | ? | 90275 | 6 | 85676 | 9285 | 21 | 44 |
| 226 | ? | 84835 | 54 | 110510 | 20275 | 21 | 36 |
| 227 | ? | 65078 | 6 | 65760 | 5992 | 22 | 36 |
| 228 | ? | 94276 | 5 | 76469 | 7083 | 21 | 42 |
| 229 | ? | 78968 | 20 | 60354 | 12295 | 21 | 47 |
| 230 | ? | 100097 | 6 | 65494 | 6854 | 23 | 36 |
| 231 | ? | 86463 | 8 | 74725 | 7576 | 21 | 36 |
| 232 | ? | 280529 | 6 | 24608 | 2258 | 27 | 38 |
| 233 | ? | 112547 | 6 | 38864 | 9033 | 23 | 36 |
| 234 | ? | 77519 | 10 | 93239 | 13768 | 21 | 36 |
| 235 | ? | 104949 | 6 | 62956 | 7826 | 21 | 41 |

## Discussion

The lithic samples were characterised in the first place by previous works **(9)** and subsequently by pXRF, from the chemical point of view. it is worth noting that XRF can be limited by its sensitivity to surface layers and its inability to distinguish between elements with similar atomic numbers. Additionally, XRF requires calibration against known standards and can be affected by matrix effects, which can lead to biases in the analysis. Only three elements (Nb, Bi, Sb) were chosen to discriminate among different lithic types, whose relationships were used for clustering. In addition, it must be noted that the reliability of classifications produced by SOMs depends on various factors, such as the quality and representativeness of the input data, the choice of parameters and hyperparameters, and the complexity of the model. In general, SOMs are considered to be a robust and effective tool for clustering, visualization, and classification tasks. However, like any machine learning algorithm, SOMs are not infallible and may produce errors or biases if the input data is noisy or biased. Therefore, it is important to validate the results of SOMs using appropriate metrics and to interpret them in their own context.

## Conclusion

Due to exploratory nature of this work the results provided should be considered preliminary, and further studies will surely provide newer nuances. Nevertheless, the data package developed thanks to pXRF non-invasive tool fully justifies the efforts. Moreover, neural networks appear as a promising tool whose potential has not yet been completely unravelled on archaeologic studies, and their skills on feature detection, clustering and prediction, can be of great interest. Furthermore, this work opens a door to the exploration of the possibility of further characterization through pXRF analysis, highlighting the versatility of this tool in combination with neural networks architecture

## Acknowledgments

We would like to thank to all the persons that made possible the analysis in the different musems all around the Comunidad Valenciana (MARQ, SIP, SIAP, Museo Arqueológico de Alcoi and MAOVA).

The authors acknowledge the Ministry of Education, Culture and Sport of the Valencian Government for funding the projects NeoNetS "A Social Network Approach to Understanding the Evolutionary Dynamics of Neolithic Societies (C. 7600-4000 cal. BP)" (Prometeo/2021/007). Gianni Gallello acknowledges the financial support of the Beatriz Galindo Fellowship (2018) funded by the Spanish Ministry of of Universities (Project BEA-GAL18/00110 "Development of analytical methods applied to archaeology" and the project founded by the Spanish Ministry of Science and Innovation EvolMED "Evolutionary cultural patterns in the contexts of the neolithisation process in the Western Mediterranean" (PID2021-127731NB-C21).

## Funding

This work was supported by Ministry of Education, Culture and Sport of the Valencian Government for funding the projects NeoNetS "A Social Network Approach to Understanding the Evolutionary Dynamics of Neolithic Societies (C. 7600-4000 cal. BP)" [Prometeo/2021/007]; and "Smartphone and Green Analytical Chemistry" [Prometeo/2019/056]

## Declaration of Competing Interest

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

## Data availability

Data will be made available on request.

**References**

1. Ramacciotti M, Gallello G, Jiménez-Puerto J, Bernabeu J, Orozco Köhler T, Rubio-Barberá S, et al. "Non-destructive characterisation of dolerite archaeological artefacts". Microchemical Journal. December 2022;183:108080.
2. Bernabeu J, Orozco T. 90: "Fuentes dc materias primas y circu· lación de materiales durante el final dei Neolítico en el Pais Valen· ciano. Resultados de! análisis peuológico dei utiJlaje pulimentado". Cuadernos de Prehistoria de la Universidad dc Granada. 1989;14-5.
3. Frahm E, Doonan RCP. "The technological versus methodological revolution of portable XRF in archaeology". Journal of Archaeological Science. February 2013;40(2):1425-34.
4. Jones MC, Williams-Thorpe O, Potts PJ, Webb PC. "Using Field-Portable XRF to Assess Geochemical Variations Within and Between Dolerite Outcrops of Preseli, South Wales". Geostand Geoanalyt Res. November 2005;29(3):251-69.
5. Honório KM, da Silva ABF. "Applications of artificial neural networks in chemical problems". Artificial neural networks-architectures and applications. 2013;203-23.
6. Zupan J, Gasteiger J. Neural networks in chemistry and drug design. John Wiley & Sons, Inc.; 1999.
7. Rojas R. Neural networks: a systematic introduction. Springer Science & Business Media; 2013.
8. Miljković D. "Brief review of self-organizing maps". In: 2017 40th international convention on information and communication technology, electronics and microelectronics (MIPRO). IEEE; 2017. p. 1061-6.
9. Orozco-Köhler T. Aprovisionamiento e Intercambio: Análisis petrológico del utillaje pulimentado en la Prehistoria Reciente del País Valenciano (España). Vol. 867. British Archaeological Reports Limited; 2000.
10. Jimenez-Puerto J, Gallello G. "pXRF Lithic tools database [Internet]". Zenodo; 2023 [citado 14 de abril de 2023]. Available at: https://zenodo.org/record/7829946
11. Fuchs C. "Self-Organizing System". In: Encyclopedia of Governance, ed by Mark Bevir. SAGE; 2007.
12. Ashby WR. "Principles of the self-organizing system." In: Systems Research for Behavioral Sciencesystems Research. Routledge; 2017. p. 108-18.
13. Kohonen T, Huang TS, Schroeder MR. Self-Organizing Maps. 3rd ed. Berlin, Heidelberg: Springer Berlin / Heidelberg; 2012.
14. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available at: https://www.R-project.org/
15. Sievert C. "Interactive web-based data visualization with R, plotly, and shiny.". CRC Press; 2020.
16. Wehrens R, Kruisselbrink J. "Flexible self-organizing maps in kohonen 3.0." Journal of Statistical Software. 2018;87:1-18.
17. Potts PJ, Bernardini F, Jones MC, Williams-Thorpe O, Webb PC. "Effects of weathering onin situ portable X-ray fluorescence analyses of geological outcrops: dolerite and rhyolite outcrops from the Preseli Mountains, South Wales". X-Ray Spectrom. January 2006;35(1):8-18.
18. Ogburn D, Sillar B, Sierra JC. "Evaluating effects of chemical weathering and surface contamination on the in situ provenance analysis of building stones in the Cuzco region of Peru with portable XRF". Journal of Archaeological Science. April 2013;40(4):1823-37.
19. Williams-Thorpe O, Potts PJ, Webb PC. Field-Portable "Non-Destructive Analysis of Lithic Archaeological Samples by X-Ray Fluorescence Instrumentation using a Mercury Iodide Detector: Comparison with Wavelength-Dispersive XRF and a Case Study in British Stone Axe Provenancing". Journal of Archaeological Science. February 1999;26(2):215-37.