scientific reports

OPEN

Provenancing ancient materials with lead isotopes: overlap uncovered

Sarah De Ceuster[1], Jurian Hoogewerff[2] & Patrick Degryse[1]✉

Lead isotope analysis has been used to provenance metals such as lead, silver and bronze for many decades. Different approaches to interpret lead isotope ratios were proposed, and various limitations of the method have been discussed and addressed. Overlap in composition between different possible ore sources is always mentioned as a major limitation in lead isotope studies. However, it has never been comprehensively studied using a multivariate statistical approach. In this paper, the kernel density estimation (KDE) approach previously proposed by the authors is applied to calculate overlap between possible ore source regions. Firstly, the copper and lead ores of the same regions are compared, to assess if they are consistent and thus can be combined to increase sample size for provenance studies. Secondly, the pair-wise overlap between all the mining regions is calculated to determine if the distinction can actually be made between those ore fields. The use of one-dimensional KDE's is very effective for calculation and assessment of the overlap between ore sources. This study argues that merging the lead and copper ore data might increase the reliability of a region's KDE's in most cases, but the overlap should be assessed beforehand. Furthermore, the study provides useful tools to verify for every pair of possible ore sources if it is theoretically possible to discriminate between them, and to what extent.

Keywords Provenancing, Lead isotope analysis, Kernel density estimate, Ore field, Distribution overlap

Since lead isotopes were first applied to provenance archaeological artefacts[1], this field has been in continuous debate and controversy over the complexity of the statistical analysis due to the non-normal and overlapping distibutions of the different ore sources. A wide range of materials have been provenanced using lead Isotope analysis (LIA). First copper, lead, silver and bronze artefacts, and later also glass and glazes have been investigated with lead isotopes. An extensive chronological overview can be found in a special issue of the journal Archaeometry[2]. Multiple ore datasets providing lead isotope ratios have been compiled and published[e.g.3–5]. The OXALID database[3], created after a series of lead isotope data of the Isotrace Laboratory in Oxford were published between 1995 and 1998 in the journal Archaeometry, was the first attempt at open database that was later digitized (last entries in 2014). Brett Scaife has compiled another open-access online database[4], which also focuses on the Mediterranean region. The data it contains largely correspond to the areas covered in OXALID. Vogl and his colleagues[5] expanded the data published by Bret Scaife to more than 2700 lead isotopic compositions of ores extracted from 80 publications covering Europe and the Mediterranean area. Ongoing efforts have been made to combine the different datasets into one[e.g.6–11], by various researchers and groups, for example initiatives like IBERLID[7] and GlobaLID[6]/TerraLID. Multiple artefact datasets have also been investigated and published[e.g.12–15]. Numerous statistical approaches alternative to the biplot method were developed[e.g.16,17]. Albarède et al.[18] have advocated for the complementary use of geological parameters (model age, κ, μ) to determine the geological context of the ore sources. Moving away from merely provenance studies, the mixing of ores from different sources and the recycling of metal artefacts has been addressed as an analytical question to answer instead of an insoluble problem[19–22]. The research angle has further broadened from asking exclusively geographical origin questions to questions about the use of materials and artefacts, object biography, assemblages, material 'flow' and the causes of technological, social, … change[20]. This evolution in research questions, however, does not diminish the considerable need for solid basic provenancing techniques.

In previous papers De Ceuster and co-authors[23,24] have presented a novel approach using kernel density estimation (KDE) to describe the distribution of ore lead isotope ratios across mining regions to answer to the limitations of LIA for provenancing. Firstly, using KDE's effectively tackles the usually non-normal distribution

[1]Department of Earth and Environmental Science, KU Leuven, Leuven, Belgium. [2]Faculty of Sci & Tech, University of Canberra, Canberra, Australia. ✉email: patrick.degryse@kuleuven.be

of the lead isotope ratios for each ore deposit. Secondly, unlike the conventional biplots, using KDE's provides a statistically sound basis that effectively combines statistical and visual assessment.

In addition, KDE's provide the means to calculate and visualize the overlap between the distributions of the deposits or regions. In this paper two applications of distribution overlap are explored to further investigate the KDE approach's value. For this purpose, datasets[9,10] on copper and lead ore lead isotope ratios were utilized, encompassing Europe and the Middle East.

Firstly, often ore datasets specific to the artefacts in question (e.g. galena for lead or Cu ore for bronze) are used for provenancing[e.g.25] although the use of mixed ore datasets would increase the significance of the analysis. Likewise, combined ore datasets are used without assessing the compatibility of their (density) distributions[e.g.26]. We investigate if lead isotope values from different datasets, originally produced for different types of artefacts are isotopically consistent. If so, a combined dataset will theoretically improve the statistical sensitivity of LIA, not only due to the increased number of observations for each mining site but it might also improve discrimination between mining sites that may have overlapping lead isotope data. As often the same mineralization event produces the cogenetic ores, it can be expected that in many cases their lead isotope compositions are indistinguishable within a mining site. Although different isotope ratios might be the result of different mineralization fluids, separate geological events in the same region etc.

Secondly, we explore the possibility of distinguishing between mining sites with overlapping isotope data. The provenance postulate[27] assumes that the measurable variation of one or more variables within an ore source should be smaller than their variation between ore sources[17,20,28]. A major limitation of LIA for provenancing, however, is that many ore deposits show overlap in isotopic composition between geographically and/or geologically distinct mining regions[8]. The aim is to present the extent of the overlap between the distributions of all mining sites in the datasets.

Methodology

The lead and copper ore datasets that were used in this study were compiled from the literature and are available online[9,10]. These datasets include lead isotope ratio data for lead and copper ores respectively, the samples are categorized by mining region. Additionally, the original publication for each sample is referenced within the datasets.

As shown in previous papers[23,24], unidimensional KDE's show the non-normal distribution for each lead isotope ratio per mining region. For each region in the datasets the 'ore field' is represented by the density distribution, estimated for each isotope ratio, by summing the individual kernels over the data points and rescaling the area under the curve to 1 (Fig. 1). The distribution of the lead isotopic values of the ore samples acts as an estimation of the distribution of the population. These kernel density estimates are deliberately chosen unidimensional to increase readability and to minimize the theoretical error of order $O\left(n^{-4/(d+4)}\right)$[29]. This means the theoretical error increases with decreasing sample size n and increasing number of dimensions d. Since the datasets consist of rather small ore sample sizes per region, it is expedient to keep d small. A summary of the full KDE methodology for provenancing[23,24] can be found in Supplementary Materials S1.

All analyses were performed using R Statistical Software[30]. The Rscript for this study can be consulted in Supplementary Materials S2.

To assess consistency between the lead isotope data of the lead and copper ore datasets, their KDE's were overlain in one plot. To make a reliable visual comparison, the area under the density estimates is plotted in proportion to the number of samples available and the values are indicated on the x-axis. To assess overlap between deposits, the overlap function from the overlapping package for R[31] is used to calculate the overlap coefficient which is an estimate of the percentage of overlap between the estimated densities. This coefficient is calculated by integrating the minimum between two densities. It essentially calculates the relative probability that a value occurs in the overlapping interval(s) and thus the probability that a sample can be a member of either population.

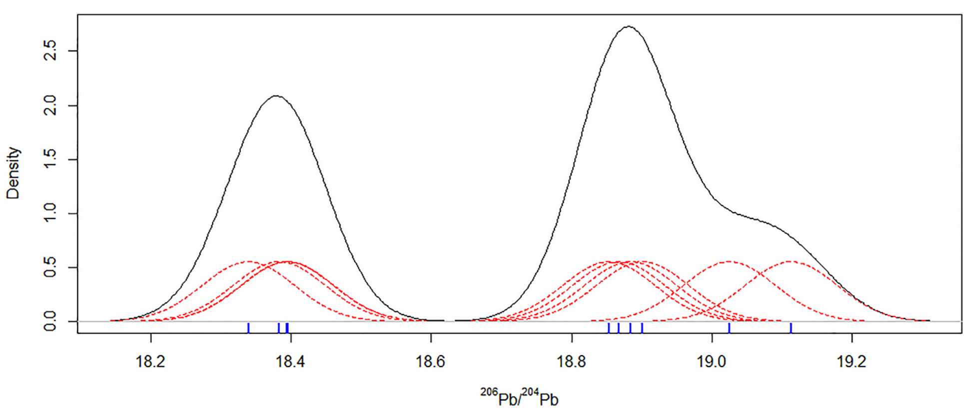


Fig. 1. Theoretical example of the kernel density estimation for 10 $^{206}Pb/^{204}Pb$ values. The blue vertical lines represent the values, the red dashed distributions are the so called kernels. Summing the kernels results in the black kernel density distribution estimate for the ore field.

Results and discussion

Cu versus Pb minerals

For the 45 mining regions that appear in both datasets, and thus have both lead and copper ores, the KDE's for the lead isotope data of lead ores were compared to the copper ore data to assess if and to what extent they overlap. The results (Supplementary Materials S3) can be divided into 3 categories: for 28 mining regions the KDE's largely overlap, for 13 regions they do overlap, but the range of the distributions is very different, resulting in a different resolution and for 4 regions they don't overlap (completely). An example per category is shown in (Fig. 2a–c).

There appear to be two possible reasons for the difference in resolution. On the one hand, as is the case for Italy-Sardinia-Iglesiente (the example in Fig. 2b), one of the ores shows bimodal estimates, while the other one does not. On the other hand, for some of the regions there are too few samples in the dataset. The copper ore dataset of France Massif Central, for example, contains merely two copper ore samples, which is not enough for a meaningful comparison. Also in the third category there are examples of regions that have a shortage of ore samples for a meaningful comparison. Another reason for no (complete) overlap could be a problem with the demarcation of the relevant mining regions or a separate mineralization event. In any of these cases the assessment of overlap between both datasets gives an indication of the applicability as a joined dataset before use. The regions where the resolution differs or there is no (complete) overlap should be treated separately to maintain the highest resolution possible in the results. Extra ore samples are required for the regions with insufficient samples, before a significant assessment can be made. When integrating them, a higher degree of uncertainty for the obtained results should be taken into account.

Distribution overlap between possible ore sources

For all regions with at least 20 ore samples in the dataset, the overlap is calculated using a paired comparison (PC) technique, calculating $n(n-1)/2$ pair-wise overlap comparisons. This is done for the lead ore dataset and the copper ore dataset separately. One might consider combining the datasets while separating regions with distinct signatures. This decision ultimately depends on the specific context of the study. In order to evaluate if the distinction between each pair can be made and to what extent, the results were divided into 4 colour-coded groups by degree of overlap according to the guidelines in (Table 1).

The colour codes were then implemented to create practical and easily customizable tools to verify if the distinction can be made between every two regions and to what degree. An easy-to-read confusion type overlap matrix for copper ores (Fig. 3) and lead ores (Fig. 4) was made. From Fig. 3 it can be deduced that the Armenian region and the Turkish regions, with exception of western Türkiye, have 5 or more regions with which they largely overlap for the copper ores. For the lead ores (Fig. 4) it is mainly northwestern Bulgaria, Portugal and the British, French and German regions that show high overlap.

To make a better overview, an interactive slicer dashboard where any mining regions can be selected, renders a pair-wise table of the selected regions with the overlap information and the colour code. A sample screenshot using the copper ore data is given in (Fig. 5). It shows the amount of overlap of the Turkish regions with the Cypriot regions in a colour-coded table. There appears to be little overlap for most combinations, only the Turkish Black Sea region shows large overlap with the Cypriot Larnaca Axis. The latter also shows average overlap with some other Turkish regions.

Conclusion

This paper shows the usefulness of KDE's to describe ore fields, not only as a provenancing method but also as a way to study overlap between them. In the first application, it emphasizes the need for caution when combining ore data without prior assessment and the increased reliability if they are combined advisedly. Additionally, it provides insights to assist researchers in making this assessment. In the second application, it provides some ready to use, customizable tools for researchers to verify if and to what extent mining regions are distinguishable when using LIA.

In order to increase the reliability of any analytical provenancing technique, more ore samples are needed. However, although the significance of the results would increase, larger datasets might not necessarily result in better differentiability, they might cause more overlap. To increase the differentiability, analytical techniques should be combined, and complemented with other contextual information.

The authors recognize that more detailed datasets would significantly enhance the understanding of ore fields. Such datasets would ideally include precise sample locations, elemental and geological data, the operational dates of specific mines, and details about measurement techniques. However, the datasets used in this study were designed to maximize the number of samples available for statistical analysis, which required sacrificing certain details that are often (mostly) absent in the literature.

The regional groupings employed here reflect how they are typically presented in existing literature and commonly used by archaeologists in conventional biplots. While information on deposit types and their effects on distributions would undoubtedly be valuable, it is frequently unavailable. One advantage of kernel density estimates (KDEs) is their ability to represent multimodal distributions, which can reveal distinct lead isotope signatures within ores from the same mine or region.

Furthermore, it is important to note that geological formations, while scientifically interesting, are not directly relevant to archaeological inquiries. Instead, the geographical region holds greater significance for understanding provenance. The aim of this study is to present an intuitive, visually accessible approach that can be applied broadly across the field, allowing researchers to utilize historical datasets responsibly and effectively.

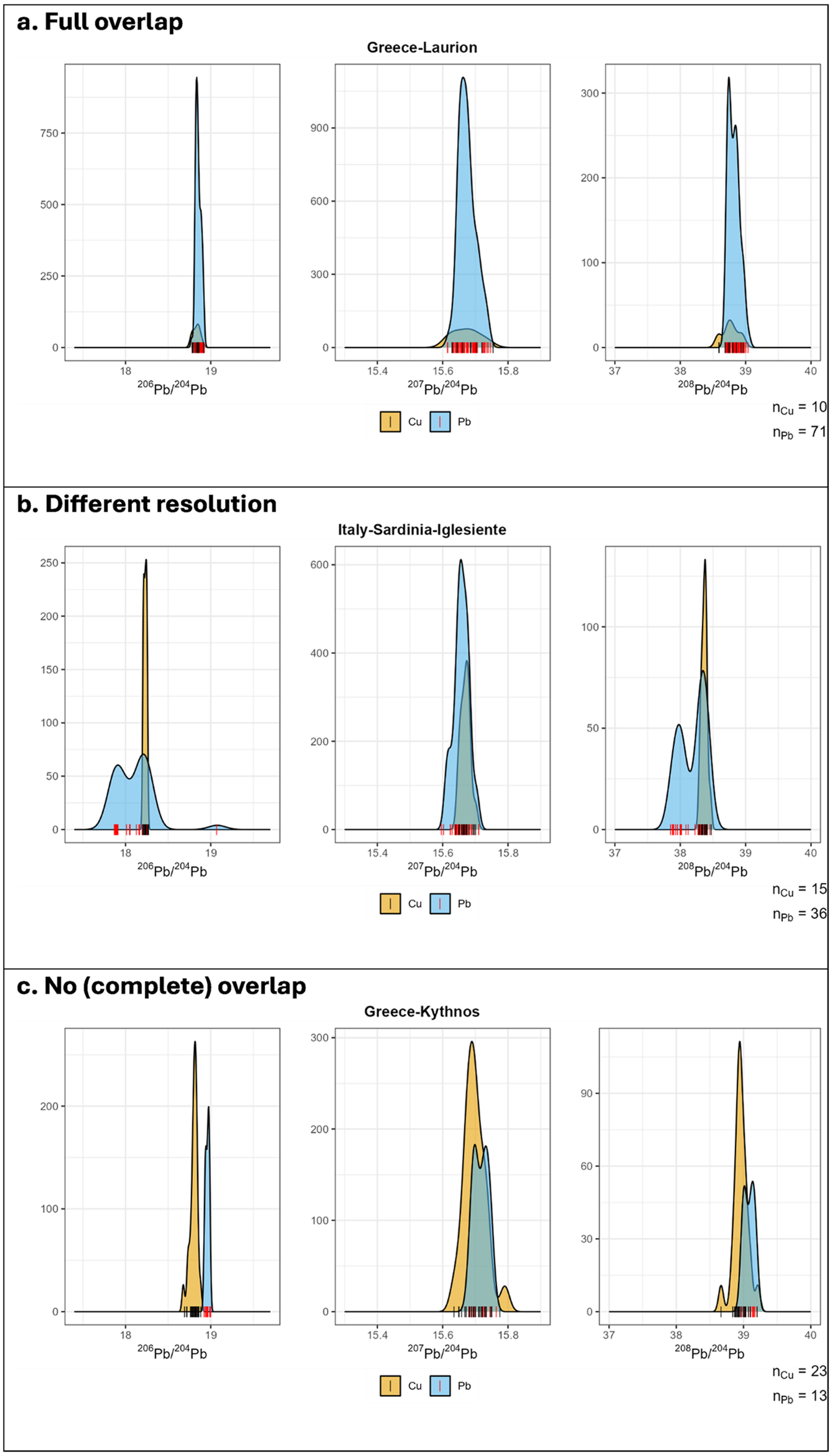


Fig. 2. Examples of the overlap comparison between lead ores and copper ores within one mining region to assess if and to what extent they overlap. (**a**) shows an example of full overlap for Laurion, (**b**) shows an example of difference in resolution for Sardinia Iglesiente and (**c**) shows an example of no (complete) overlap for Kythnos.

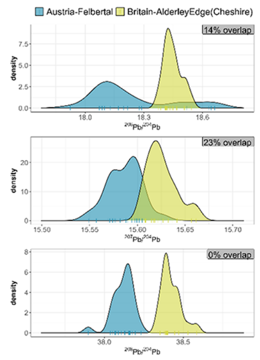
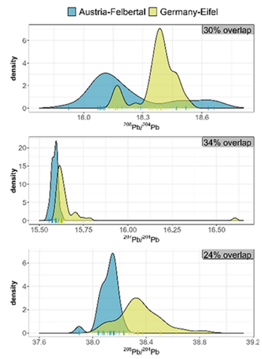
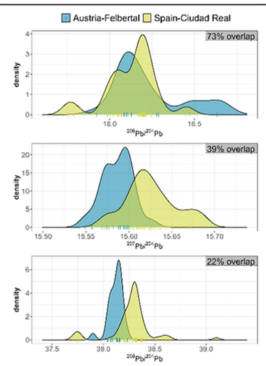
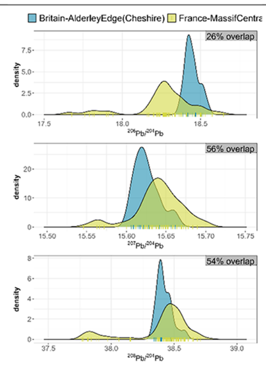
Description	Example	Colour code
At least 1 ratio under 20% overlap. Very distinguishable. No or only little overlap.		Blue
At least 2 ratios under 33% overlap. Distinguishable.		Green
At least 2 ratios under 50% overlap. Possibly distinguishable.		Yellow
At least 2 ratios higher than 50% overlap. Almost indistinguishable.		Red

Table 1. Guidelines for the division of the overlap results into 4 colour-coded categories.

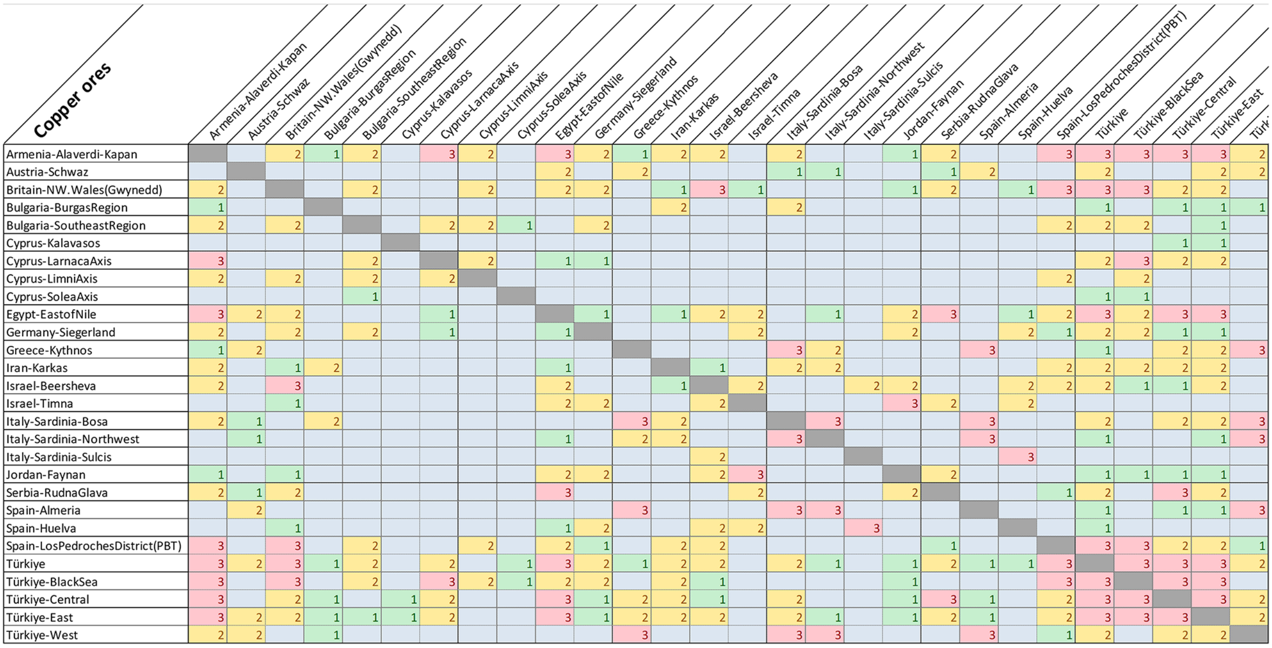


Fig. 3. Overlap matrix for copper ores. The first column and top row show a list of all mining regions with at least 20 samples in the copper ore dataset. The corresponding field for each pair of regions shows the differentiability between them as defined in the guidelines in (Table 1).

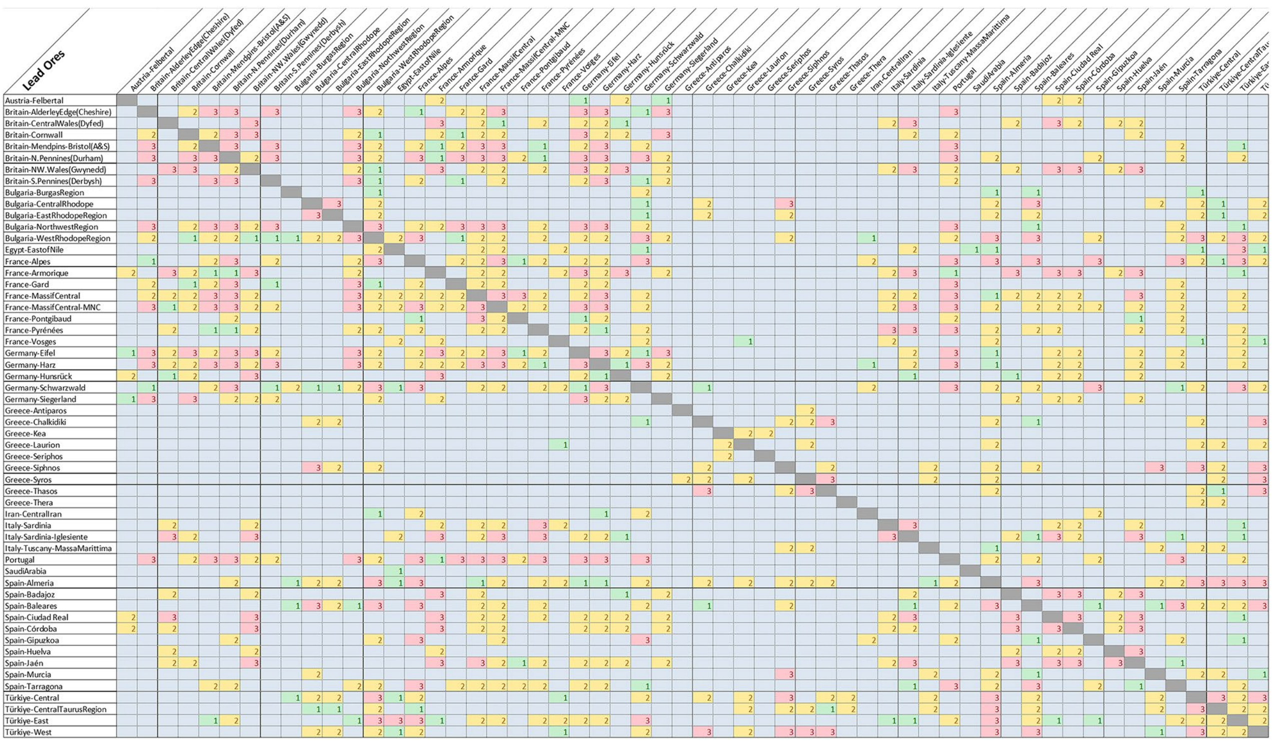
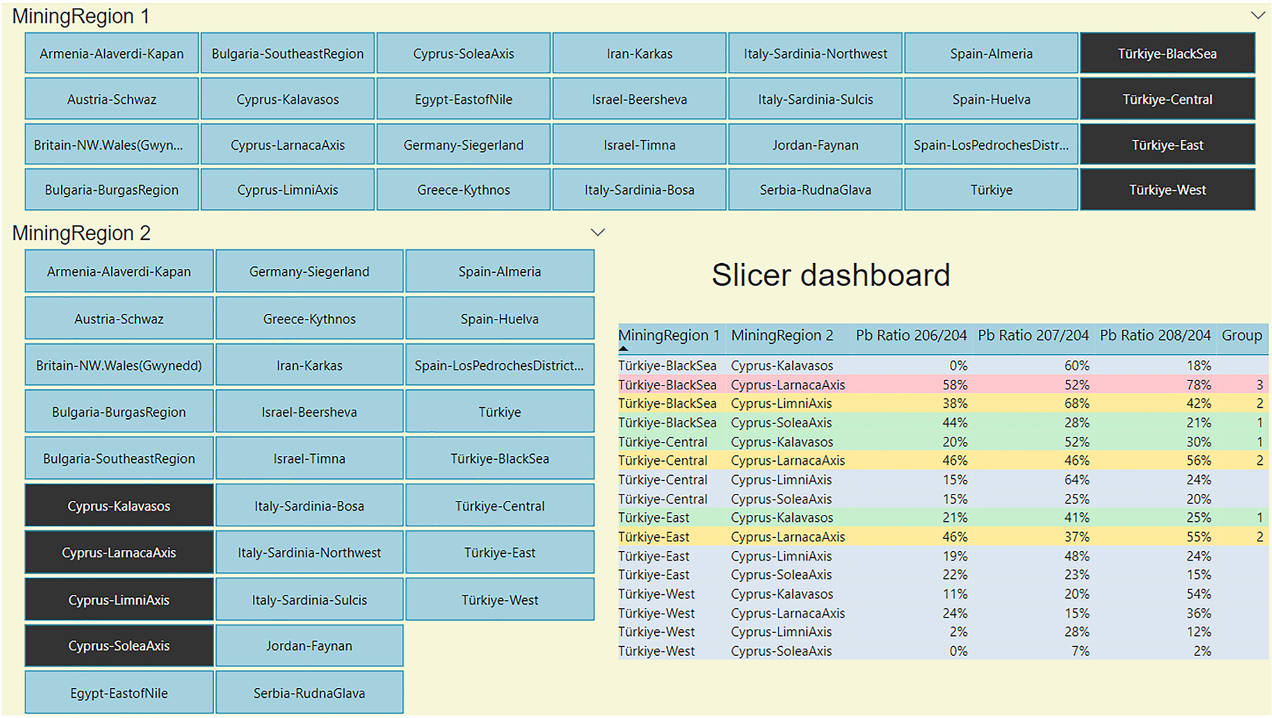


Fig. 4. Overlap matrix for lead ores. The first column and top row show a list of all mining regions with at least 20 samples in the lead ore dataset. The corresponding field for each pair of regions shows the differentiability between them as defined in the guidelines in (Table 1).



MiningRegion 1	MiningRegion 2	Pb Ratio 206/204	Pb Ratio 207/204	Pb Ratio 208/204	Group
Türkiye-BlackSea	Cyprus-Kalavasos	0%	60%	18%	
Türkiye-BlackSea	Cyprus-LarnacaAxis	58%	52%	78%	3
Türkiye-BlackSea	Cyprus-LimniAxis	38%	68%	42%	2
Türkiye-BlackSea	Cyprus-SoleaAxis	44%	28%	21%	1
Türkiye-Central	Cyprus-Kalavasos	20%	52%	30%	1
Türkiye-Central	Cyprus-LarnacaAxis	46%	46%	56%	2
Türkiye-Central	Cyprus-LimniAxis	15%	64%	24%	
Türkiye-Central	Cyprus-SoleaAxis	15%	25%	20%	
Türkiye-East	Cyprus-Kalavasos	21%	41%	25%	1
Türkiye-East	Cyprus-LarnacaAxis	46%	37%	55%	2
Türkiye-East	Cyprus-LimniAxis	19%	48%	24%	
Türkiye-East	Cyprus-SoleaAxis	22%	23%	15%	
Türkiye-West	Cyprus-Kalavasos	11%	20%	54%	
Türkiye-West	Cyprus-LarnacaAxis	24%	15%	36%	
Türkiye-West	Cyprus-LimniAxis	2%	28%	12%	
Türkiye-West	Cyprus-SoleaAxis	0%	7%	2%	

Fig. 5. Sample screenshot from an interactive slicer dashboard using the copper ore data. It contains two lists of all mining regions with at least 20 samples in the copper ore dataset. The user selects one or more regions on both lists to render a table that gives pair-wise information. The corresponding row for each pair of regions shows the differentiability between them as defined in the guidelines in (Table 1).

Data availability

The data that support the findings of this study are openly available in KU Leuven RDR. The lead ore data is available at https://doi.org/10.48804/D4DPLJ, the copper ore dataset at https://doi.org/10.48804/ZS1Q4U.

Received: 14 October 2024; Accepted: 31 January 2025
Published online: 07 February 2025

References

1. Brill, R. H. & Wampler, J. M. Isotope studies of ancient lead. *Am. J. Archaeol.* **71**, 63–77 (1965).
2. Degryse, P. & Editorial Isotopes in archaeological science. *Archaeometry* **62**, 1–2 (2020).
3. OXALID. Oxford archaeological lead isotope database. https://oxalid.arch.ox.ac.uk (2010).
4. Scaife, B. Lead isotope ratios for mediterranean ores. http://www.brettscaife.net/lead/data/index.html (1997).
5. Vogl, J., Rosner, M., Curbera, J., Peltz, U. & Peplinski, B. Lead isotope analysis in magic artefacts from the Berlin museums. *Archaeol. Anthropol. Sci.* **10**, 1111–1127 https://doi.org/10.1007/s12520-016-0445-6 (2018).
6. Klein, S., Rose, T., Westner, K. J. & Hsu, Y. K. From OXALID to GlobaLID: Introducing a modern and FAIR lead isotope database with an interactive application. *Archaeometry* **64** (4), 935–950 https://doi.org/10.1111/arcm.12762 (2022).
7. García de Madinabeitia, S., Ibarguchi, G., Santos Zalduegui, J. F., IBERLID & J.I. and A lead isotope database and tool for metal provenance and ore deposits research. *Ore Geol. Rev.* **137**, 0169–1368 https://doi.org/10.1016/j.oregeorev.2021.104279 (2021).
8. Killick, D. J., Stephens, J. A. & Fenn, T. R. Geological constraints on the use of lead isotopes for provenance in archaeometallurgy. *Archaeometry* **62**, 86–105. https://doi.org/10.1111/arcm.12573 (2020).
9. De Ceuster, S. & Degryse, P. Dataset of lead isotope ratios of lead ores. *KU Leuven RDR* **V1** https://doi.org/10.48804/D4DPLJ (2023).
10. De Ceuster, S. & Degryse, P. Dataset of lead isotope ratios of copper ores. *KU Leuven RDR* **V1** https://doi.org/10.48804/ZS1Q4U (2024).
11. Tomczyk, C. A database of lead isotopic signatures of copper and lead ores for Europe and the near East. *JAS* **46**, 105657 https://doi.org/10.1016/j.jas.2022.105657 (2022).
12. Blichert-Toft, J., de Callataÿ, F., Telouk, P. & Albarède, F. Origin and fate of the greatest accumulation of silver in ancient history. *Archaeol. Anthropol. Sci.* **14**, 64 https://doi.org/10.1007/s12520-022-01537-y (2022).
13. Westner, K. J. et al. Rome's rise to power. Geochemical analysis of silver coinage from the Western Mediterranean (fourth to second centuries BCE). *Archaeometry* **62**, 577–592 https://doi.org/10.1111/arcm.12547 (2020).
14. Vogl, J., Paz, B. & Völling, E. On the ore provenance of the Trojan silver artefacts. *Archaeol. Anthropol. Sci.* **11**, 3267–3277 https://doi.org/10.1007/s12520-018-0756-x (2019).
15. Artioli, G., Canovaro, C., Nimis, P. & Angelini, I. LIA of prehistoric metals in the central Mediterranean area: a review. *Archaeometry* **62** (1), 53–85 https://doi.org/10.1111/arcm.12542 (2020).
16. Rodríguez, J., Sinner, A. G., Martínez-Chico, D., Zalduegui, S. & AMALIA, J. F. A matching algorithm for lead isotope analyses: Formulation and proof of concept at the roman foundry of Fuente Spitz (Jaén, Spain). *J. Archaeol. Science: Rep.* **51**, 104192. https://doi.org/10.1016/j.jasrep.2023.104192 (2023).

17. Tomczyk, C. & Żabiński, G. A PCA-AHC approach to provenance studies of non-ferrous metals with combined pb isotope and chemistry data. *J. Archaeol. Method Theory* **31** (1), 93–143 (2024).
18. Albarede, F. et al. A miner's perspective on pb isotope provenances in the western and central Mediterranean. *J. Archaeol. Sci.* **121**, 105194 (2020).
19. Radivojević, M. et al. The provenance, use, and circulation of metals in the European bronze age: the state of debate. *J. Archaeol. Res.* **27**, 131–185 (2019).
20. Pollard, A. M. et al. *Beyond Provenance: New Approaches to Interpreting the Chemistry of Archaeological Copper Alloys* https://doi.org/10.2307/j.ctv7xbs5r (Leuven University, 2018).
21. Albarède, F. et al. From commodity to money: the rise of silver coinage around the ancient Mediterranean (sixth–first centuries BCE). *Archaeometry* **63** (1), 142–155 (2021).
22. Berger, D., Brügmann, G., Bunnefeld, J. H. & Pernicka, E. Identifying mixtures of metals by multi-isotope analysis: disentangling the relationships of the early bronze age swords of the Apa–Hajdúsámson type and associated objects. *Archaeometry* **64** (S1), 44–74. https://doi.org/10.1111/arcm.12714 (2022).
23. De Ceuster, S. & Degryse, P. A 'match–no match 'numerical and graphical kernel density approach to interpreting lead isotope signatures of ancient artefacts. *Archaeometry* **62**, 107–116 (2020).
24. De Ceuster, S., Machaira, D. & Degryse, P. Lead isotope analysis for provenancing ancient materials: a comparison of approaches. *RSC Adv.* **13** (28), 19595–19606 (2023).
25. Yagel, O. & Ben-Yosef, E. Lead in the levant during the late bronze and early iron ages. *J. Archaeol. Sci. Rep.* **46**, 103649 (2022).
26. Brozio, J. P. et al. The origin of neolithic copper on the central northern European plain and in Southern Scandinavia: connectivities on a European scale. *PLoS One* **18** (5), e0283007 (2023).
27. Weigland, P. C., Harbottle, G. & Sayre, E. V. Turquoise sources and source analysis: Mesoamerica and the Southwestern U.S.A. In *Exchange Systems in Prehistory* (eds Earle, T. K. & Ericson, J. E.) 15–34 https://doi.org/10.1016/B978-0-12-227650-7.50008-0 (Academic Press, 1977).
28. Pollard, A. M. What a long strange trip it's been: lead isotopes and archaeology. *Mine Microscope Adv. Study Anc. Technol.* 181–189 (2009).
29. Wand, M. P. & Jones, M. C. *Kernel Smoothing* (CRC, 1994).
30. R Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* https://www.R-project.org/ (2022).
31. Pastore, M., Di Loro, P. A., Mingione, M., Calcagni, A. *Overlapping Estimation of overlapping in empirical distributions*. https://CRAN.R-project.org/package=overlapping (2022).

Acknowledgements
This work was supported by the Special Research Fund (BOF-KUL) C1 (Grant No. C14/19/060 and C16/24/004).

Author contributions
S.D.C.: conceptualization, methodology, coding, formal analysis, investigation, data curation, writing—original draft, writing—review & editing, visualization. J.H.: coding, writing—review & editing, supervision. P.D.: conceptualization, writing—review & editing, supervision, project administration, funding acquisition.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-88909-1.

Correspondence and requests for materials should be addressed to P.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.