

PCI Archaeology

RE: Response to recommender - *A meta-analysis of Final Palaeolithic/earliest Mesolithic cultural taxonomy and evolution in Europe*

Dear recommender,

CAS – School of Culture &
Society

Department of Archaeology
and Heritage Studies

Felix Riede
Professor, Ph.D.

Date: 03/12/23

Tel.: +45 8716 2081
E-mail: f.riede@cas.au.dk

<http://person.au.dk/do/farkfr@hum>

Page 1/14

first, allow us to express our thanks and respect for so efficiently and swiftly handling this long and complex manuscript of ours. We truly appreciate the effort and time invested by yourself and the three reviewers. In the following, we reply to each substantive review comment in turn, with our responses marked by > and provided in *italics*. Associated with our reply letter is a revised manuscript in two versions, one with track changes and one without.

Please note in relation to your own thorough review of the code that we opted for a Docker-like packaging of all relevant components. This is now reflected in two ways:

1. The old/deprecated package versions can be downloaded and installed from the Posit Package Manager via the `install_packages.R` script provided in this repository.
2. Alternatively, users can download and employ the `CultTaxFinalPal.sif` Singularity/Apptainer container provided on Zenodo (<https://doi.org/10.5281/zenodo.10061126>).

To use the Singularity file interactively, use the following command: ``singularity exec CultTaxFinalPal.sif R``.

We hope very much that you find the revised manuscript in good order and worthy of recommendation.

Warmly and on behalf of all co-authors,

Felix

Decision: **minor revisions**.

The reviews are **overall very positive**, but point out some discussion aspects that are worth addressing in my opinion. They also suggested several modifications or additions that will surely serve to improve your paper.

> We are very happy to hear this as this has been a challenging paper to write. We are very happy to make the modifications requested.

I would like to add that I found the **methodology and approach well described and argued**. The authors are transparent and careful in their interpretation, which is very positive. Moreover, I could replicate the figures and analysis using the provided code and data. However, I would like to point out that many packages used in your analyses are deprecated, or even not available on CRAN anymore. I think it could help the replicability to either 1) change the code so that it uses packages that are still available, or 2) provide a file with all deprecated packages to be downloaded together with the supplementary.

> Please see our comment above. Ironically, the depreciation of packages turns out to be a real Achilles heel of reproducibility, especially over the slightly longer term.

Review #1

Reviewed by anonymous reviewer, 18 Oct 2023 00:21

The Manuscript is well written, and the description of the rationale are clear and comprehensive.

> Thanks!

The title is clear and reflect the content of the article. The abstract presents the findings concerned in the analysis. The introduction clearly explains the motivation for the study and the research question is well presented. According to the authors the aim of this research « has been to meta-analytically assess the validity of current cultural taxonomic schemes for the Final Palaeolithic and earliest Mesolithic (15-11 ka cal. BP) in Europe and on this basis to infer patterns and processes of material culture diversification, cultural transmission, and adaptation». I subscribe to the need of performing such analysis.

The methods and analysis are described with detail, however my lack of knowledge on data analysis, R or learning machine methods doesn't allow me to detect flaws in the design of the research or in the analysis. But the raw data are available and other researchers can replicate the analysis. The geographic units used seem logic, however, I would like to have more information on the archaeological sites chosen. We don't have information on the identity of the «regional experts» and despite Figure 3 captions says, «A detailed breakdown of key sites can be found in SI Table 2», I only had access to the type of site: open air, cave or rock shelter. The reader cannot check which sites were used for each region. I suggest that Figure 3 caption includes the list of the sites targeted in the analysis.

> We are happy to hear that R1 recognises the relevance of such analyses. The first thing to note here is that all sites are listed in some detail in the spreadsheets associated with the paper and which can be downloaded at the data link provided. There, they are shown with coordinates, references, and extensive metadata. This can easily be cross-checked with Fig. 3. Note also that, in our analysis no sites are really 'targeted' specifically. Naturally, all data should be linked back to actual sites but the sum of their total is also more than the individual sites.

We have further added on page 7: "Regional experts were sourced in the lead team's network and had to have extensive first-hand experience with the particular area they represent; most commonly, these regional experts are also based in their respective region."

Interpretation of the results is cautious and seemed to be adequately supported by the results. The mentioned «diversification of material culture in the realm of armature shapes towards the end of the Palaeolithic» was already observed but conventional analysis but it is interesting to prove it using these methodologies.

The references are appropriate and accurate. Regarding the Figures 5a, 5b and 5b and 7a and 7b are difficult to read. Unfortunately, I have no suggestions to improve them in order to become more easily read.

> Yeah, well, we have worked a great deal on these figures and considered many different forms of visualization. We appreciate the difficulty here but hope that high-resolution images in the eventual publication will suffice.

The paper presents valuable research that will be of significance to scholars working on this subject all over the Europe. Furthermore, the authors might consider raising a minor question regarding the influence of the French school during the during the 19th and first half of the 20th centuries. Some of the NACs were identified by French scholars abroad or by local archaeologists seeking for similarities with the classical Perigord sequence (mainly for the late Pleistocene phases). Some cultural taxonomic denominations can be explained by the influence of the French school?

> *This is an understandable comment but perhaps one that also needs some unpacking – and whose adequate treatment lies outside the scope of this paper. As Hussain (2019) has shown in his doctoral thesis, speaking of a 'French school' per se is a crass simplification. What this usually refers to is the Bordian tradition, but which now is largely superseded by a technological perspective, i.e. the 'French school' in contemporary terms. Yet nonetheless, we can refer to a certain Perigordian-centrism with longstanding attempts to correlated sequences and NACs with their supposed 'reference' sequences in SW-France. This reflects a deep-seated and ultimately paleontological logic of reasoning grounded in a good dose of assumed pan-European universalism. Much more can be said here but mindful of the length of our manuscript – and its aim to actually present novel analyses – we suggest not to expand on this topic here.*

Some minor corrections could be done to improve this paper, but this work will be an important contribution to the Pleistocene-Holocene transition.

Review #2

The manuscript "*A meta-analysis of Final Palaeolithic/earliest Mesolithic cultural taxonomy and evolution in Europe*" by Mr Riede and co-authors, which has been submitted to PCI Archaeology uses a meta-analysis approach to evaluate the efficacy and replicability of contemporary cultural classifications of prehistoric cultural taxonomies focusing on the Final Palaeolithic and the earliest Mesolithic in Europe (c. 15,000 to 11,000 BP). To achieve this aim, the authors used a high-level computational approach on a large spatiotemporal scale, combining a set of statistical tests designed to accumulate research results on novel integrated dataset including key sites, lithic toolkit composition, blade and bladelet production technology as well as lithic armatures.

Overall, the manuscript is well written. Data collection is adequately and openly presented in sufficient detail with additional information structured into chapters provided in the supplementary information. The literature cited is very informative and relevant to the topic of the current manuscript. All figures are appropriate and the statistical tests are displayed with accuracy. The argumentation is well stated as it is clearly indicated in the abstract.

> *Many thanks.*

The main point of this study is that the results of meta-analysis provide better estimates of the relation in the population than single studies, especially when integrating operational chain analysis to resolve cultural taxonomic questions. While I overall agree

with their findings and these data is potentially be of great interest for a broad readership, the presented manuscript would benefit from some clear information/discussion that I have resumed in two main points:

Selected studies and dataset

This study emphasizes to us the coherence of the various NACs groupings and the existence of a cultural diversification tracked over time. However, considering the large spatiotemporal scale, the inclusion of a limited set of studies biases estimates about the effect sizes in the population, since the results do not identify all possible studies on the phenomenon.

> In our defense, we mobilise the vital notion of the sample. Our approach does not rely on an analysis of 'all possible cases' to make its point. In fact, it is unclear what 'all possible cases' would mean and how to access such a complete population. Of course, a map at scale 1:1 is the most faithful representation of reality but is it feasible? Is it useful? Given the size of our database, we are actually quite confident in the rusticity of our results.

Data inconsistencies

The selection of reliable key sites identified as those that hold rich information on lithic typo-technology well published by regional experts in prestigious journals, might influence meta-analysis results and lead to misleading inferences about the issue of taxonomic designations. Furthermore, considering the fact that typical specimens are usually selected for drawing because they are representative of taxonomic entities, the inclusion of complete specimens rather than fragments might biases the estimates about the effect sizes in the population.

> These are good and valid points. We do address the potential biases introduced by resource limitations and variable drawing practices (see references to Lopes 2009 and Saville 2009). However, the bottom line is that the practice of basically always selecting the 'prettiest' pieces for drawing makes our analysis conservative in terms of capturing the total variability of forms.

Minor remarks:

Line 24: do you mean domain or module? Please correct accordingly in the text.

> module! This is corrected accordingly.

Line 55: unretouched components are excluded from the meta-analysis, though Belloisian and Laboraian unretouched knives, *identified as an important production goal,*

were included in the dataset. The exclusion of unretouched components, which might biases the meta-analysis results, needs to be explained.

> Unfortunately, these elements are not systematically recorded across regions, so any comparative analysis at this macro-scale would need future work. Because our workflow required – ideally – equal recording status across region, we needed to limit our attention to those elements most commonly recorded.

Line 286: the authors stressed that the Epigravettian macro-unit stands apart with a distinct laminar technological organization and suggested that this might be a bias of limited data from the long-lived Epigravettian. This very important point in this study seem to be mentioned as a side note here and needs to be more clarified in discussion.

> Given the paucity of relevant data on precisely this topic, we would in fact prefer not to go into more detail here. This issue is, we feel, clearly flagged up for future work in our paper but this is not the place, we feel, to expand on it.

Line 573: The discussion would have benefit from the integration of other aspects of material culture such as bone/ivory technology, genetic data, cave art, burials ... This would have strengthen the argumentation.

> We do include several references to other forms of material culture and to the patterns of arhythmical changes in lithic and organic technology as observed by Roux, Perles, Valentin, Naudinot, and others. It is obvious to us – and also to R2 – that an alignment of lithic analysis with other classes of material culture would be a strong next step to take. And we mention this in the Conclusion.

Review #3 (Dirk Leder)

Aim of the paper:

The manuscripts aims to tackle the traditional construct of archaeological cultural units in Europe at 15-11 ka cal BP by employing a meta-data analysis based on material culture (lithics). The study investigates the underpinnings and coherences of 86 named archaeological units (taxa) using multivariate statistical analyses. The objective is to clean up a taxonomic mess that evolved during more than a century of academic research and debate.

General observations and minor issues:

The abstract clearly lays out the topic of interest, whereas the results formulated therein lack clarity. It may be worthwhile considering to be a bit more straightforward stating, e.g. “Our data analysis suggests that 20 of the 86 analysed taxa are not distinct enough and therefore should undergo revision [or something in that area]”.

The introduction is well written, clearly states the geographic and temporal scope of the work as well as the issue at hand and the subjects (NACs) involved. Also, different research traditions/schools and political agendas are appreciated in brief and the motivation to write this text is clearly stated. However, after all the justified criticism of NACs in general, it seems advisable to state in brief why grouping individual assemblages might be beneficial still, if at all. Is there any benefit to it? Should we speak of different time slots instead, or are there other ways to communicate Palaeolithic cultural units, e.g. to students?

> We have added a passage on page 7 highlighting the inevitable need to classify in analyses that go beyond an individual assemblage. The question of communication to students or the public is moot here, however. Our aim here is not at all concerned with communication to students but analytical clarity. Note that biology has no issue communicating complex taxonomic issues to students, so should we really continue using obviously flawed concepts just for the sake of communicating assumed narratives? This is a really interesting debate that has raged a few times on Social Media when discussing the uses of typology, for instance, where some will continue to highlight its usefulness in communication. We think this should be taken up for consideration.

The materials used and the methodology are clearly described and the various statistical analyses iterated and referenced well. A few minor points should be addressed though.

- How many sites per region are included in this study?

*> This is presented in detail in the included data tables and also in our companion study: Hussain, Shumon T, Felix Riede, David N Matzig, Miguel Biard, Philippe Crombé, Javier Fernández-López de Pablo, Federica Fontana, et al. “A Pan-European Dataset Revealing Variability in Lithic Technology, Toolkits, and Artefact Shapes ~15-11 Kya.” *Scientific Data* 10, no. 1 (2023): 593. <https://doi.org/10.1038/s41597-023-02500-9>.*

- The usage of discrete data (presence/absence) in comparison to continuous data (e.g. percentage) might prove disadvantageous, but testing this would be a project for the future.

> Very true but I hope we can leave this for future work.

- Please explain the reasoning behind the usage of millennial-scale time slices over climate-driven time slices.

> *As we explain on page 10, the decision to sample the record in equidistant and equally sized time-steps is an attempt to not unduly bias the classification by preconceptions about human-climate relationships and the consequences of certain climate regimes on culture change dynamics. It is an attempt to circumvent what Bevan and Crema (2021) have recently labeled the ‘modifiable reporting unit problem’ – very much an issue with the taxonomic units of the Final Palaeolithic/earliest Mesolithic. The reasoning is quite similar to excavating in artificial layers or spits, which allow subsequent analyses of settlement intensity etc. precisely because the units are standardized. In our analysis it would be difficult, for instance, to quantify taxonomic diversity if our units were not organized in time bins.*

- I am not sure raw material economy is a relevant indicator of culture-taxonomic differences in general, e.g. Baltic flint has been used intensively at different times, and locally available raw materials likely were used more intensively across time-boundaries than exotic one in general.

> *We absolutely agree! In fact, we make that very same point specifically in relation to the Fürstein Facies.*

- Is there a reason SES was preferred over p-values?

> *Generally, one should be cautious of p-values! More specifically, the SES quantifies the relative difference beyond significance and is a value that’s comparable across different data types/sets.*

- In respect to armature outlines the authors mention, “The subsampling was conducted in a stratified way using the splitstackshape R package [146], where we chose two outlines per NAC whenever available.” How did you ensure, these two outlines are representative of all armatures in that NAC? Assuming, some assemblages may encompass e.g. rectangular, triangular and bipointed/lunate shaped armatures. Please clarify/specify.

> *Given the nature of our dataset, all armature outlines included should be considered equally representative of a given NAC – otherwise they would not be in the dataset. And the very logic of sub-sampling is also to ensure (statistical) representativity. The alternative would be to manually and hence arbitrarily choose specimens and that would definitely be against our approach as particular preconceptions about classification would guide these decisions.*

Discussion and Conclusion are clearly formulated and some interesting results of the data analyses are highlighted and contextualised well.

Given recent genetic evidence (Posth et al. 2023) of population movements around 14 ka cal BP from south to north, it would seem important to highlight potential connections between Late Epigravettian and Azilian/FMG based on the analysis.

> Very interesting remark, thank you for the suggestion! While this isn't our key research question here, we fully recognise the importance and relevance of this recent palaeogenomic work. To accommodate this, we have modified the closing passage of our paper thus:

“In this context, it is interesting to note the recently suggested connections between the Epigravettian and Azilian/Federmessergruppen as traditionally defined and the consistent clustering of artefacts associated with these cultural taxonomic groups in our analysis. This may be taken as tentative support for biocultural connections among these populations. Going forward, it becomes a critical matter to more precisely define how if at all traditional named archaeological cultures provide meaningful analytical units for the genetically, culturally, and ecologically fluid worlds of Pleistocene mobile foragers. Improved data-driven cultural taxonomic precision may allow us to better align and thereby mutually enrich the many emerging datasets – climatic, genetic, and archaeological – that allow us to address these dynamics.”

The authors comprise an international team that worked together according to a previously agreed upon standardized data frame. This step is crucial, as it allows for the accumulation of vast data facilitating inter-regional comparisons of archaeological material. The study is therefore a major improvement over previous studies wherein selected sites/regions have been analysed by just a few authors and results are based on a much smaller database. On the other hand, regionally differing research traditions might have introduced a bias into the study as the authors rightfully points out. By that token, the same tool might therefore be classified differently based on established research frameworks. This however, is a potential bias that will have to be checked in the future by researchers investigating ‘foreign’ assemblages from geographically distant regions.

Main concern:

As the authors rightfully point out the plurality of named archaeological cultures (NACs), rely on learned scientific traditions rather than data-driven inters-site correlations. The current paper was written in the spirit of breaking up these ‘classic’ cultural groupings and critically re-evaluate their coherence based on the analysed data. While I agree with the various analytical steps and selected statistical methods in general, my main concern lies in the base units selected in the various statistical analyses. The data was collected on the basis of 350 sites (with varying data qualities), which for the purpose of the analyses are then lumped back together into NACs and even macro-units. This to me seems like the very definition of a hermeneutic circle and naturally leads to redundancies with well-established expectations of the NACs defining criteria. It is

puzzling then why the authors did not choose sites as the base unit of their analysis instead. The latter approach would have permitted a less biased approach in detecting clusters defined by material culture attributes rather than by already established NACs and macro-units.

> The response to this concern requires a slightly longer explanation. When we initially began this project, our aim was in fact to conduct a similar study on a much smaller but highly resolved suite of key sites directly. That is, we intended to make this analysis anchored exclusively in sites. Along came the covid pandemic, shattering all plans to access collections directly; we therefore shifted to an approach – also very valid and, in our view quite novel – working with a combination of legacy data and expert assessment. That said, our outline data are at the site level and so are the analyses of these. Beyond the outlines, typological and especially technological data are not available at all at the necessary volume and standards to conduct the sort of analysis we do here. It was therefore a painful but necessary decision to integrate these data at the level of named archaeological cultures. Equally, as one of our aims was to assess the robusticity of groupings, this was not possible without in fact grouping or by analyses at the site level, in which case higher-order groupings could only be derived in a strictly exploratory manner. We have endeavoured to conduct all our analysis at the lowest (site-nearest) level wherever possible, yet without assuming their correctness. This should be evident from our results given that one key conclusion is that many of these units don't hold up as very coherent given the data.

With the risk of sounding defensive or lazy, we would put it to future work to use our or other data to seek coherent clusters of sites that can then be interpreted as 'cultures' or the like. Given the length of the current manuscript, this is not something we should do here, however.

Another issue arising from the lumping of sites arises when tool types and technological variables are present only in few cases among these groups. By lumping them with sites that lack such traits, these exceptional occurrences all of a sudden become a standard component (attribute) of this group in the analysis. For example, the majority of Late Magdalenian sites lacks bipoints and zinken that might however be present at few exceptional sites. By lumping all these sites into a single package, the entire macro-unit Magdalenian now has the named tools as a characteristic feature used in further analyses. Naturally, they would then overlap with the Azilian on the hand and the Hamburgian on the other.

> The Zinken are a curious and possibly unique example of such a qualitative trait that appears highly characteristic of a particular group. However, Zinken-like artefacts actually occur in other contexts, too (also much later Mesolithic ones!). Is the focus on this artefact class perhaps also a product of research historical biases? We consider it doubtful that a single artefact class can or should alone define an operational grouping. Instead, our approach assumes that similarly named groups should

be characterized by a consistent recurrence of features, however defined. Our present attempt is therefore to provide amalgamated information in that spirit. Note also that our approach is a macro-archaeological one where subtle differences, in principle, should not be significant. That said, please see our further comment on Zinken below.

Once this issue has been addressed, I think there is much potential in this paper that has the advantage over other publications of relying on solid database with relevance on a large geographic scale.

Specifics:

‘Low Countries’ might be an expression not familiar to most readers and I had to search for it too. Benelux might be an option instead.

> Fair point. However, the term is well represented in public dictionaries (Encyclopaedia Britannica, Wikipedia, etc.) and changing it across text, tables and all data sheets would be a large task that we do not consider worth the gain.

Table 1: The time slices should be equipped with ages cal BP for better orientation.

> Amended.

Table 2. The * in the table caption does not show up in the table itself.

> Amended.

Figure 3. The map displays country boundaries where one would expect to see boundaries of the defined regions instead. This should be adjusted.

> Here, we respectfully disagree. As argued extensively in the Introduction, modern political and cultural boundaries clearly have had and continue to have an effect on archaeological practice including classification. We therefore find it valuable to include contemporary national boundaries. Conversely, we quite deliberately have not drawn boundaries around our defined regions so as to not visually suggest prehistoric geographic units or territories. The discussion of mapping practices goes somewhat beyond this paper but a corollary of classification is its projection onto maps – and this is not something we want to do here. Instead, modern mapping tools allow point mapping, which we feel is the least assumption-laden way to show our site data.

Figures 5A-C. Text and figure captions should clearly state that NACs form the analytical base unit herein, not macro-units.

> This is amended for Figures 5A and B; for Figure 5C it is noted that the analysis is conducted at the level of specific sites.

I assume Zinken were subsumed under borers, and wonder whether this is a good idea, as it is one of the signature tools of the Hamburgian?

> Actually, Zinken were separated out as a separate artefact category; as it turns out, the discriminatory power of this single tool class is not as strong as one might have assumed. There is much confusion about this artefact category – more than many people would assume. Zinken-like tools occur in/are recognised (correctly or not) in many Late Palaeolithic as well as later prehistoric contexts.

Lines 262. Have all artefact outlines been used in this analysis or two per NAC as above? Please specify.

> Yes, this is clarified on line 263 and is also clear from the N provided in Figure 9.

Lines: 373-377. The authors suggest, “the place of Bromme in the Terminal Pleistocene and its relationship with ABP-associated complexes requires critical re-valuation...”. In my understanding, this perfectly fits expectations of the Bromme phase emerging from FMG, but with tanged points (TPC) while foreshadowing developments eminent in the following FBT/LBI. Please clarify.

> We have tried to clarify this in lines 377ff.; the sequence sketched out by R3 is the one we also agree with but which is not universally reflected as consensus in the literature where the Bromme and similar cultural taxa are very often lumped with the ‘tanged point cultures’, classically and influentially in Taute (1968) and Kozłowski (1999).

Lines 579-584. “Overall, the results of our macro-archaeological analyses are complex, surprising, and to some extent sobering”. “...our results confirm the broad heuristic utility of some traditional named cultural taxonomic groupings...” (e.g. the Magdalenian, the Ahrensburgian). As mentioned above, I suspect this is in part the result of the data-lumping into NACs and macro-units.

> We disagree with this suspicion given our analysis. We believe that we have in fact quite rigorously evaluated the utility of these units – and in fact we reject most of them leaving!

Typos

In the introduction, remove “(Ivanovaité et al., 2020; Sauer and Riede, 2019)” and leave [66, 81] to keep consistency in citing (I don’t have a line number)

> *Amended.*

L141 : replace (Matzig 2021a) by corresponding number

> *Amended.*

L238 : replace (Mantel 1967) and (Guillot and Rousset 2013) by corresponding numbers

> *Amended.*

L241 : replace (Sokal 1979) by corresponding number

> *Amended.*

Code

General

I can't find the Rproj file in the zip file downloaded on the supplementary link on Zenodo

> *This has been added.*

Otherwise, I was able to replicate all analysis and figures (after having installed the deprecated packages and dependencies)

I would advise for your next papers to try to choose packages that have fewer dependencies. Also, for GIS analysis, I would advise you switching to terra and sf instead of using raster, rgeos, rgdal etc. that are no longer being maintained.

> *Please see our comments on reproducibility in the beginning of this cover letter.*

Overall, we hope to have addressed all three reviewers' concerns with our replies and amendments. We further hope that you find our revised manuscript in good order and now worthy of recommendation.

Warm wishes on behalf of the author team,



Felix Riede