

01/02/2024

Dear Dr. Blasco López,

We thank you and the reviewers for peer-reviewing and editing our pre-print. The comments have brought to light weaknesses, inaccuracies, and ambiguities in our work, and afforded us a chance to correct them. We follow with a point-to-point reply to the reviewers.

The corrected manuscript has been uploaded to the repository, and can be found at <https://doi.org/10.1101/2023.11.12.566733>. The R script has not changed beyond adjustment of the line colors in Figures 1 and 2, following the request from Miriam Belmaker. This change has been updated in the repository at <https://zenodo.org/doi/10.5281/zenodo.10115388>

We hope that the revised pre-print will be found suitable for recommendation in PCI Archaeology.

Best wishes,

Nimrod & Uri

Reviewer 1: Jesús Rodríguez

Dr. Rodríguez raised concerns about the explanation of the methods and hypotheses, and also about the suitability of the datasets we used, Fairbanks and the Judean Desert, for applying the methodology. We thank him for his thorough and challenging critique. We follow with a point-to-point reply.

1. **Critique:** The reviewer points out a major flaw in our approach in that we assume that the probability of a specimen surviving to be dated is assumed to be proportional to the number of individuals of its taxon that existed in a specific region and time. While agreeing that the population density of a species in a limited area is one of the main factors affecting the probability of finding their remains in the paleontological record and obtaining radiocarbon dates for them, the reviewer notes that other interacting factors can drastically modify the probability of finding a certain taxon in a fossil assemblage. Among such factors he lists biotic and abiotic accumulation agents such as prey body-size selection; site characteristics, e.g., suitability for denning; or post-depositional processes that affect differentially the skeletal elements of different taxa.

Reply: This is an important critique, which highlights to us a fault in our explanation of the methods and hypotheses. It is important to emphasize that although we aim to reconstruct long-term ecological interactions between predator and prey taxa, our primary data are the distributions of observations of single taxa over time, aggregated across several find spots in each region. Therefore, biotic and abiotic biases in specimen frequencies should not affect our results unless we have reason to believe that these processes acted differently through time on specific taxa. For example, the hypothetical fact that predator tibiae preserve less well than herbivore tibiae does not matter to the distribution of predator remains over time. Conversely, if we have reason to believe that predator tibiae preserve less well during a particular time interval compared to other periods, this could bias our results. Here, we make the uniformitarian assumption that there are no changes over time in the biotic or abiotic factors affecting the deposition or post-depositional survivability of *specific taxa* (top predators/larger herbivores) and that the population density of a species in a region is therefore the main factor affecting the probability of finding their remains in the paleontological record and obtaining radiocarbon dates for them. This assumption relies on the comparability of the depositional environments for each dataset throughout time, which consist of dry desert caves or gravel deposits.

We added this explanation to the penultimate paragraph of the introduction.

2. **Critique:** The reviewer notes that the hypothesis tested by the authors is that the summed probability distribution (SPD) of radiocarbon dates of predators is different from those of prey. The authors state that if differences exist, they prove the existence of a

signal of predator-prey interactions in the analysed datasets. The reviewer found this assumption insufficiently justified in the preprint.

Reply: This point was clarified. The sixth paragraph of the introduction now reads: “We hypothesize that the summed probability distribution (SPD) of predator radiocarbon dates, insofar as it tracks changes in population size, will differ from those of prey, with either greater or lesser divergence than expected from a random sample of SPDs from the same time range. A non-random divergence would suggest that predator and prey populations covaried. This minimalist hypothesis assumes nothing about the wavelength, mechanism, or cause of predator-prey interaction, which we believe cannot be tested with the current data. If supported, this hypothesis would provide preliminary evidence that long-term regional radiocarbon data encode predator-prey interaction signals. This could justify constructing larger datasets to enable in-depth investigation of the structure of these signals.”

- 3. Critique:** Regarding the Fairbanks data, compiled by Leonard et al. (2007), the reviewer rightly notes that we omitted specimens with only minimum age estimates from the text without explaining the reason. Importantly, the reviewer raises an objection that the datasets we chose do not fulfill the criterion of randomly-selected specimens. This objection is based on three observations: **firstly**, that both the Judean Desert and Fairbanks datasets contain far more predator dates than is expected in an ecosystem, and therefore the datasets violate the assumption that the probability of a specimen surviving to be dated is proportional to the population density of the species. **Secondly**, the reviewer cites Lazagabaster’s (2022) paper, which states that large predators from the Judean Desert were always dated while herbivores were not, and to our decision to omit other mammals from the Judean Desert dataset. **Thirdly**, the reviewer notes that the 12 leopard bones from the Judean Desert represent an MNI of six, and were derived from three caves, raising concerns over pseudoreplication.

Reply: Thank you for catching the missing reference to Leonard et al., which has now been added to the text. We also explained in the first paragraph of the methods that only specimens with both minimum and maximum age estimates were included. As the reviewer noted, this is obligatory because otherwise the dates cannot be calibrated. The gist of the reviewer’s critique is, however, the selectivity of the samples. **First**, we agree that both the Judean Desert and Fairbanks datasets contain far more predator dates than is expected in an ecosystem, but disagree that this violates the assumption that the probability of a specimen surviving to be dated is proportional to the population density of the species. The estimate of predator and prey frequencies at any point in time is not obtained from their numerical ratio, which would then indeed have to reflect a reasonable predator/prey balance. Rather, it is derived from the independent calculation of the probability densities of the dates of each group. In this case, the frequencies are irrelevant. This reply is also relevant to the selective dating of

large carnivores by Lazagabaster et al. (2022). We have clarified this point in the last paragraph of the introduction. **Second**, the reviewer is right that we omitted taxa from the Judean Desert database, and focused on the leopard-hyrax-ibex triad only. The reason for that is that the leopards in the Judean Desert hunted almost exclusively ibex and hyrax (Perez, I., Geffen, E., & Mokady, O. (2006). Critically Endangered Arabian leopards *Panthera pardus nimr* in Israel: estimating population parameters using molecular scatology. *Oryx: The Journal of the Fauna Preservation Society*, 40(3), 295–301.). Gazelles are found in the open plains, away from the cliff habitats of the leopard; wild boar were introduced towards the end of the study period, at ~1000 BCE, and were not taken by the leopards — probably because their habitats do not overlap, but also because the small Arabian leopard could not take a wild boar. The striped hyena, a large carnivoran, was similarly omitted as a predator because it is a carrion eater that very rarely hunts. Therefore, we focused on the three species we are certain to have had trophic interactions. This is now explained in the first paragraph of the methods section. **Third**, the leopards of the Judean Desert clustered around specific wadis where large hyrax populations and water could be found (Davidovich et al, Leopard traps in the Judean Desert reveal long-term impact of humans on top predator populations. Revision submitted to *Quaternary Science Research*), and therefore the aggregation of observations in these caves does not necessarily mean they were likely from the same individual: They represent thousands of years of taphonomic “sinks” in the hotspot of leopard activity in the Judean Desert. Lazagabaster’s MNI estimation of six assumes that two specimens from the same cave represent different individuals only if the calibrated dates’ 95% confidence intervals do not overlap. Also, this type of ‘chronological minimum number of individuals’ calculation is conceptually equivalent to flattening the summed probability density curves we are comparing, and making them *ipso facto* similar to random noise. All in all, the MNI estimation is probably over-conservative, as the three leopard specimens from one of these caves, from which aDNA could be extracted, belonged to three different individuals (ibid.). Due to this spatial aggregation of leopards in the regions where the caves containing them were found and the genetic results, we do not think that pseudo-replication is a serious problem, but honestly — it cannot be controlled. We acknowledge this potential problem in the second paragraph of the methods section in the manuscript.

4. **Critique:** The reviewer notes that the interpretation of the results for the Judean Desert Dataset is problematic. The KL divergence value is said to be “smaller than 94% of the divergences measured for (random) predator – (real) prey distributions”. Again, the authors conclude that the “...predator and prey distributions are not random, and that the low divergence between them is, therefore, unlikely to be due to chance.” However, figure 2 shows that the observed KL value is not smaller but larger than most of the values generated at random. Thus, if the KL value observed is within 95% of the values obtained in the Monte Carlo approach, the conclusion should be that the KL value observed may be obtained by chance. Only if the value observed in the dataset is larger

than 95% of the randomly produced values it is possible to conclude that the divergence observed is too large to be a result of chance. In any case, a theoretical justification of the meaning of a similarity or divergence between the SPDs of predators and prey is required.

Reply: the hypothesis was not phrased properly in the introduction. It now reads: "We hypothesize that the summed probability distribution (SPD) of predator radiocarbon dates, insofar as it tracks changes in population size, will have either greater or lesser divergence than expected from a random sample of SPDs from the same time range. A non-random divergence would suggest that predator and prey populations covaried". We do not presume to know whether to expect a staggered pattern (predator increases/lag/prey decreases) or a synchronicity (predator increases/prey decreases) at this resolution.

Regarding the methodological aspects of the divergence metric (KL divergence) and its significance, we chose it for its ubiquitous use in information theory and its intuitive derivation from the Shannon entropy value of the distribution (we cite Deng, J., Wang, Y., Guo, J., Deng, Y., Gao, J., & Park, Y. (2019). A similarity measure based on Kullback–Leibler divergence for collaborative filtering in sparse data. *Journal of Information Science and Engineering*, 45(5), 656–675, but see also the nice explanation in StackExchange: (<https://stats.stackexchange.com/users/89649/skander-h>), S. H. (n.d.). *Why is Kullback-Leibler divergence a better metric for measuring distance between two probability distributions than squared error?* <https://stats.stackexchange.com/q/411116>).

Note that we avoided treating the bootstrap support values we obtained as p-values, as they cannot be treated as such (e.g., Hillis, D. M., & Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*, 42(2), 182–192.). These issues are now explained in the fourth and sixth paragraphs of the methods section, respectively.

Minor comments: corrected.

Reviewer 2

Critique: The study is based on the Summed Probability Distribution (SPD), which is first mentioned by the authors in line 68. I think it is necessary for the authors to explain what SPD consists of, what its applications are, and what its limitations are. Regarding the limitations, Williams (2012) proposed various recommendations for using the summed radiocarbon probability distributions,

01/02/2024

such as a minimum sample size to obtain reliable distributions.

Reply: We have added a short explanation on the SPD being a summed probability distribution of calibrated radiocarbon dates, and referenced the important publication by Williams (2012). We do not wish to digress into a longer exposition of this topic, which we believe is fairly well known. Williams (2012) has indeed recommended using a very large number of dates to construct an SPD (~500), but we agree (and cite) Crema's (2022) recent review of the topic in which he notes that the uniformity of the material (single species in a single local, as opposed to a hefty part of the archaeological record of Australia), the strength of the pattern, and our missing prior are important here (ibid., p.1392). A look at the very detailed study of the subject by Hinz (2020) reveals the importance of the signal strength and the sample uniformity (in our case, single taxa from a single context, as opposed to all archaeological dates in a region, sensu Williams 2012). Especially figure 6 in Hinz (2020) reveals well that a strong signal can be detected by a sample size on the order of magnitude that we employ (10^1 ...). On a final note, our ultimate goal is to look for covariance between pairs of SPDs, so the most important aspect of sample size is its equivalence between the compared SPDs. These considerations were explained in the fourth and fifth paragraphs of the text.

Critique: Although the authors state the hypothesis of the paper, I believe it should be explained in more detail why this hypothesis is proposed and what the implications would be if the results showed less or more divergence than a random sample of SPDs. My understanding is that if the results show less divergence than expected, it would indicate that the predator-prey species were coexisting, and if the divergence is greater, that they were coexisting to a lesser extent. As I mentioned above, I am not an expert on the techniques used by the authors of the paper, so I think it is important that the authors expand on this part and clarify possible misinterpretations.

Reply: See reply on points 2 and 4 to Reviewer #1. It is essentially the same comment.

Critique: Regarding the sample used, I was attracted by the fact that in Fairbanks there are more samples of predators (wolves) than of prey (horses and reindeer). Why is this? The Late Pleistocene Rancho La Brea tar seeps site came to mind, where there is a greater representation of carnivores than herbivores (Spencer et al., 2003). Readers may be interested in taphonomic issues and interpretations of the deposit-accumulating agents analyzed. In the analyzed assemblages, prey species appear that are not in the analysis. What is the reason for this? Are they non-preferred prey for the predators? Another aspect that may be interesting to explain is the preferred prey of the predator species in the regions analyzed.

Reply: See reply on point 3 to Reviewer #1. It is essentially the same comment.

Critique: The authors talk about calibrating radiocarbon dates before converting them to Summed Probability Distributions (SPDs) in lines 95-97. I find it interesting that

01/02/2024

the authors explain why this calibration is necessary and what it consists of.

Reply: Calibration is a necessary step in constructing summed probability distributions, and the subject is explained extensively in the cited literature (especially Williams, 2012). In a nutshell, uncalibrated dates do not have a probability distribution that can be summed. Again, we do not wish to digress more into the basic principles of radiocarbon dating in this short paper.

Critique: Regarding the Kullback-Leibler (KL) divergence analysis, I think it would be interesting if the authors could explain how the results should be interpreted. A summary of the KL divergence statistics between predator and prey SPDs in the Fairbanks and Judean Desert datasets is given in Table 1. In the case of the Fairbanks, the KL value (1.7174) is close to the minimum of what is expected between the random and prey distributions (1.559), and in the case of the Judean Desert, the KL value (5.0741) is close to the maximum of the divergence between the random and prey distributions (5.24). In both cases, however, the authors say that the results have the same trends, being less than 98 and 94% of the divergences measured for (random) predator-(real) prey distributions, respectively. Because of the interpretation problems that can arise, I think it would be useful to explain how to interpret the results of Kullback-Leibler divergence analysis, including figures 1 and 2, and when two distributions can be considered not different.

Reply: Following the 4th reply to Reviewer #1, we clarified our choice of use of the KL divergence value to quantify the difference between SPDs (fourth paragraph of the 'Methods' section), and also corrected the confusing way in which we presented our hypothesis.

Minor comments:

In the title, the authors could use the words predator and prey, separated by a hyphen: A note on predator-prey dynamics in radiocarbon datasets. **(accepted)**

Line 77: Change to *Panthera pardus nimr*. **(corrected)**

Line 78: Change to *Capra ibex nubiana* or *Capra nubiana*. **(corrected)**

Figures 1 and 2 show two parts, so they could be distinguished as A and B. **(we added 'top' and 'bottom' to the caption)**

In the caption of Figure 2, the Kullback-Leibler divergence analysis is missing. **(corrected)**

Table 2 could be organized by genus as Table 1 to better visualize the specimens of each genus and their dating. **(corrected)**

Line 192. Change to Evaluating Bayesian radiocarbon-dated event-count [REC] models for the study of long-term human and environmental processes. **(corrected)**

Lines 208-209: Change to Climate change and cyclic predator-prey population dynamics in the high Arctic. **(ok)**

Line 225: Remove [1974].

01/02/2024

Reviewer #3 (Miriam Belmaker)

Change the color scheme of the plots. (accepted).