In this new version of the manuscript we have carried out the changes and modifications to text and figures that were previously indicated in our reply to Reviewer 1 and to Reviewer 2.

Response to Reviewer 1

"Reconstructing land temperature changes of the past 2,500 years using speleothems from Pyrenean caves (NE Spain)"

We greatly appreciate this positive review and acknowledge all the technical comments provided by Rev.1 (see below). We have proceeded to correct them as most of them are grammar typos or refer to not clear sentences. Regarding the specific comments, we certainly agree about the need to better define the timing of the DA, RP, MCA, LIA periods since there are different opinions in the literature and also regional differences. We have included the time intervals we use in this manuscript for the five identified periods at the end of the method section. We use approximate ages, based on previous information (IPCC reports and previous papers). Besides, we have replaced AD or BC by CE and BCE in the text and figures, according to terminology using in PAGES2k and other networks focused on last 2000 years.

Technical corrections:
Lin 73 – I think a brief definition of the Great Acceleration would be useful here. **Done**
Line 77 – "record decadal temperature changes…" would be less awkward **Done**
Lines 88-97 – please use correct δ18O notation. **Done**
Line 98 – what certain periods are you referring to? **Done**
Line 105 – does this altitude impact the record? **In fact, yes. We observe that the variations of d18O in the stalagmites occur at different range for each cave and altitude can be one of the causes (see Fig. 3). To minimize local effects on the composite we normalized the stalagmite records, as explained in methods section.**
Line 120 – missing m for meters after 2-3 **Done**
Line 158 – slightly moderated by precip…are you indicating that temperature is the main relationship with d18O and that precipitation is impacting it to a small extent? **We modified the sentence to make it clear.**
Line 178 – Lacks instead of lack **Done**
Line 182 – One first batch – might be better phrased as "The first batch" **Done**
Line 228 – were instead of are **Done**
Line 235 – "were included in the iscam composite record" would be a better way to say this **Done**
Line 259 – I am not sure that F14C is the correct way to represent the 14C data? The NOSAMS website is a good reference if you need one:
https://www2.whoi.edu/site/nosams/calculations-and-reporting-of-results/
There is paper by Paula Reimer (2004) where F14C is the unit recommended for post-bomb samples dating. We include that reference in our manuscript.
Line 329 - please use correct δ18O notation. **Done**
Line 336 – driving instead of driven **Done**
Line 338 – "…due to the large dependence of temperature on d18Or in this region" might be clearer **Done**
Lines 345-346 – not a sentence – please re-work **Done**
Line 351 – significant instead of significance **Done**
Line 354 – no need for the work "what" **Done**
Line 355 – "Still" is not needed **Done**
Line 368 – phrases "varies at distinct range" and "it is really complicated" need to be re-phrased **Done**
Line 391 – check for an extra parenthesis **Done**

Line 401 – are you talking about speleothem d18O? It might be a good idea to specify this
Done
Line 409 – the phrase "in spite it may continue…" seems a bit awkward – I would recommend
re-phrasing **Done**
Line 414 - Roman Period has already been defined – can just use RP **Done**
Line 426 – "in spite of recent observations indicating that …" would be a better way to phrase
this **Done**
Line 432 – strange character shows up between our and speleothem **Done**
Lines 476-478 – this needs to be rephrased to be more clear to the reader
Lines 481-484 – this needs to be rephrased to be more clear to the reader
Line 485 – need "the" between isolate and last **Done**
Line 504 – I am not sure what is meant here by the Iberian-RP **Done**
Line 506 – either add "The" or "A" before the word record **Done**
Line 589 – Besides is not necessary here **Done**
Line 600 – NAO is not a driver of floods. A better way to phrase this is Atmospheric conditions
associated with low/negative NAO index" **Done**
Line 614 – should higher be highest? **Done**
Line 1090 – whatever you decide to do for the 14C, please change it here **Done**
Line 1101 – for Figure 4, it might be helpful to have warm/cool indicated for the Pyrenees
record presented in (a). **Good remark. Done**
Line 1138 – I cannot see the line with the red asterisk **The table was not in the right format.
Corrected.**
Line 1161 – please indicate (left) and (right) for the age models and proxy profiles,
respectively. **Done.**

Response to Reviewer 2

Reply to reviewer 2

"Reconstructing land temperature changes of the past 2,500 years using speleothems from
Pyrenean caves (NE Spain)"

We very much appreciate this detailed review, which will certainly help us to improve our
manuscript. Most of the comments and suggestions made in Rev2 can be easily incorporated
in a new version where we will make efforts on better integrating the role of other factors,
such as precipitation amount, on the isotopic variability. We agree that a mixture of processes
influences the d18O signal, with temperature being the dominant one but precipitation
(amount, source) also playing a role. We also think that the final message of this study would
not change substantially, but this new focus will simplify the discussion. Most of the changes
we can make in the manuscript are in that direction. Additionally, we also agree with Rev2 that
the last section of the discussion (5.2.3) needs to be enlarged by adding potential forcings and
including new reconstructions of NAO, EA, AO, etc in Fig. 6.

Major comments:

- Rev2 feels that the manuscript remains rather descriptive (and sometimes too long) for
  part of the discussion section, making it difficult to extract the value of the new
  speleothem record. We agree that probably too much effort was put into comparison with
  other regional records without focusing on the significance of the new d18O composite
  profile. We have therefore reinforced the importance of this new record and also
  streamlined sections 5.2.1 and 5.2.2. Additionally, Rev2 feels that section 5.2.3 (forcings)

should be more detailed. In the new version, we will emphasize in section 5.2.3 the differences from northern to southern Europe, taking into account the different influence of patterns such as NAO or EA on the decadal scale. Regarding the reflection of recent warming in the speleothem data, we have rewritten the paragraph dedicated to the Instrumental Era to make clearer our interpretation of the results obtained.

- Rev2 considers that we cannot present these data as a robust temperature reconstruction, due to different aspects related to both the statistics and the structure. We are aware of this, as other factors also play a role (precipitation source, amount…) as pointed out along the text. Therefore, we agree to modulate this message throughout the text (starting with the title where "temperature" could be changed by "hydroclimate").

  o Regarding the statistics, we have used different methods, which will be included in the methodology section and throughout the text.
    ▪ First, we note that most of the correlation analysis between our speleothems is included in the Iscam software. Iscam is employed to compare speleothems, providing an intra-site correlation age model. In contrast to the widely used but subjective 'wiggle matching' approach, the Iscam method uses Monte Carlo MC simulations and cross-correlation calculations (Fohlmeister, 2012). We now indicate more clearly in the text the cases where the data have been detrended and smoothed and how.
    ▪ Second, regarding the correlation between Seso Cave stalagmites and instrumental records, we agree that the statistical information is missing. In the new revised version, all this information will be included in the methods. We have mainly used the PAST application for correlation analyses (Hammer et al., 2001) but more detail can be given.
    ▪ Thirdly, the comparison with other regional or global datasets is not carried out using any statistical technique. Sometimes the use of statistics is precluded (or made difficult) by the different resolution between records (e.g. Pyrenean lakes and tree rings), and sometimes differences between records are important due to expected differences associated with their location, altitude or mixed signals.
  o Regarding the structure, we agree that section 5.1. could be shortened and streamlined. It is also good to read the advice from Rev2 to remove negative wording. It is true that the text will be improved by positive wording and by emphasizing the benefits of using a composite record rather than the limitations. In the new version we will also move here the comparison with the PAGES2k reconstruction to support the temperature interpretation.

Minor comments:

L88ff change d18O to δ18O (check whole manuscript, occurs several times). **Yes, it was a mistake. It is corrected.**

L97 delete "as". **Done**

L135 from which elevation is the speleothem? Does elevation have an influence on the data? Is there a gradient between the caves in terms of mean annual temperatures? Surprisingly no individual cave temperatures are given, only a range for all caves… **The elevation was already indicated, next to the coordinates. There is indeed a temperature gradient between the caves and possibly influenced the data. Due this one and other local factors, d18O data were**

normalized before going into Iscam. Cave temperatures will be indicated in the text in the revised version (except for Pot au Feu cave where no monitoring was carried out).

L149ff On which time-scales? Seasonal or interannual? **Interannual.**

L178 lacks **Done**

L235 is there any difference in the result if the order is changed? It might be worth checking if the main features (e.g., warmer RP than IE) remain when the input of iscam is changed. **We did that checking and the main features are maintained. But the correlation during the overlapping sections is maximized with this order.**

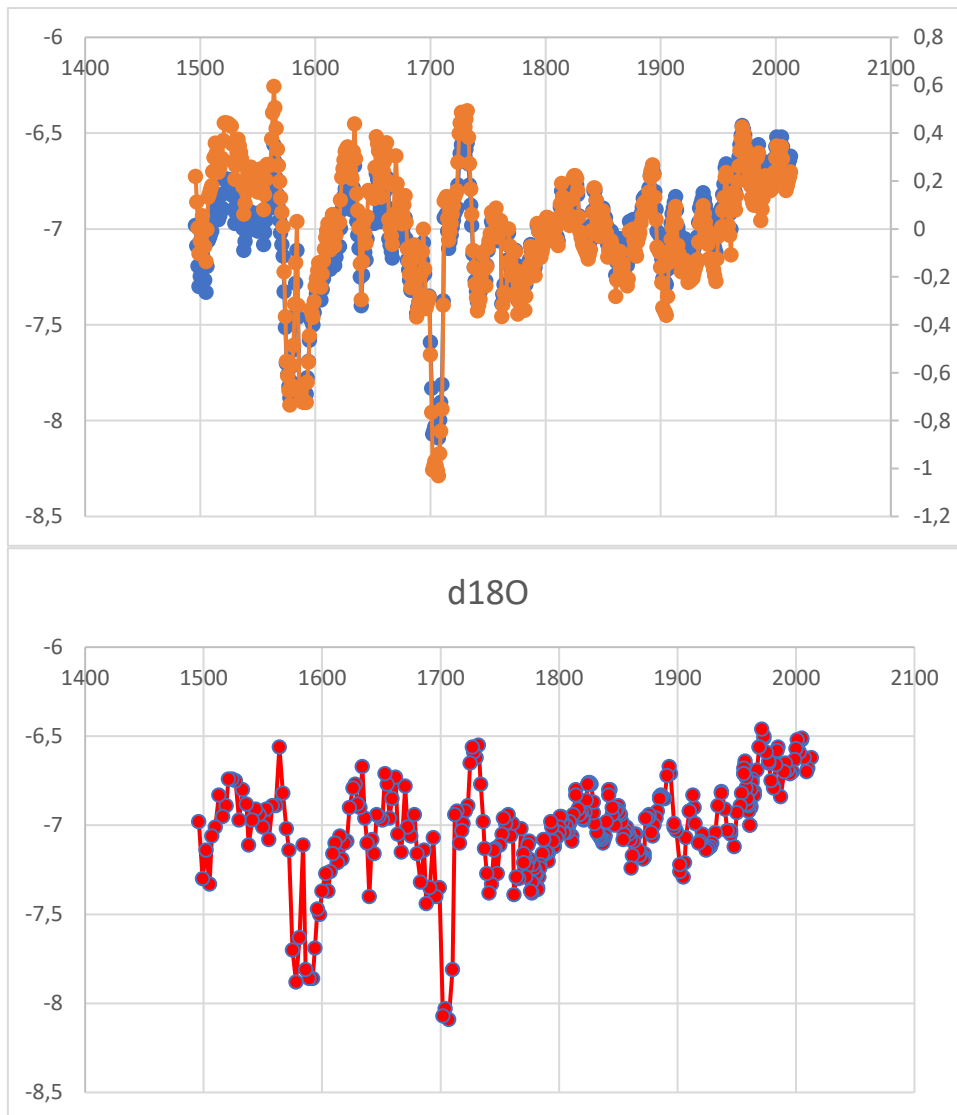
L297 Do you mean Figure A1? **No, Figure 3 was the correct one.**

L298: More like 0.3-0.4. Also Charlie has r = 0.5 and Isa 0.45. **Corrected.**

L308ff not sure if this is needed here. Seems to be all repeated within the discussion section later on. **The idea is to have that paragraph as Results and later those data should be discussed in comparison with other sites.**

L332 Again, which timescale? Seasonal? Interannual? **Interannual.**

L336 driving. **Done**

L346ff How is this correlation calculated? Number of degrees of freedom? Autocorrelation taken into account? how are T and speleothem data smoothed? Are they downsampled to the same resolution? What is the correlation of the detrended data**? The d18O series and the instrumental temperatures were first resampled (linear interpolation) to obtain the same regular spacing (annual). Then, correlation was computed using PAST software (Hammer et al., 2001). Spearman's rank correlation analysis, a nonparametric measure as an alternative to Pearson correlation analysis, was preferred to account for nonlinear relationships, with r indicating the correlation coefficient (PAST software; Hammer et al, 2001). The Bonferroni test was applied to prevent data from spuriously appearing as statistically significant by making an adjustment during comparison testing. P-value is also indicated for all the correlation analysis. All that information will be included in the text (Methods section). Autocorrelation was not taken into account. Below, we show in blue (interpolated) and orange (interpolated and detrended) isotopic data to indicate the small importance of this analyses. In red, below, is the original d18O series, without interpolation.**

There is a large discrepancy during the 1980s, so the "correlation" may be also driven just by the trend. This however may also reduce the validity of the speleothem T reconstruction on the decadal scale. **See graphs above. We don't think that the trend here is so important for correlation but, certainly, during the 1980s there is a large discrepancy among the Pyrenean and the global climatic series, that may be related to the effect of precipitation.**

L351 compare previous comment. **See our response above, it also applies here.**

L354 would move this rough "rule of thumb" to the end of this paragraph, if used later on. Or remove, if temperature control cannot be validated with more arguments. **Moved**

L355 Statistically, it is a large part! Assuming the r value of c. 0.5 is true also in the presence of autocorrelation and age uncertainties, this is still only 25% of the variance explained. So, 75% of the variance are related to other processes. **Included this argument**

L361ff For example, flow path changes can influence δ18O values without any clear correlation to precipitation amount as shown e.g., by the work of Treble et al 2022 or Priestley et al 2023 **References included to support this argument**.

L366 With this paragraph, the authors are weakening their own interpretation. Also they state, that "…comparing the warming magnitude of the RP with the MCA or the IE is not feasible…", but this is literally done in L414ff or L482ff! **The reviewer is partially right… and we have changed the text accordingly. The composite d18O record comes from different stalagmites from different cavities that were normalized and detrended so comparing RP with MCA or IE is not adequate. Still, comparing our record with previous ones is certainly possible and necessary. This is the focus of the discussion and not the comparison among periods.**

L397 also here, the link between precipitation and δ18O is highlighted. This is one of many examples where it is evident that precipitation amount is an important driver of speleothem δ18O. I am not sure why the authors do all the effort to demonstrate a temperature dependence, it would save a lot of text and discussion if it would be acknowledged that there are several influencing processes… **Yes, we agree. This idea represents the main change we are carrying out in the text of the manuscript.**

L414 contradiction to previous claim, that relative magnitudes of individual periods cannot be compared because they are recorded by different caves? Also the RP warming is mainly driven by the Gloces speleothem(s), which do not have a modern analogue. **Yes, the reviewer is right, the RP is a warm period but we cannot say is the warmest one. Previous sentence: "Considering the last 2500 years, the Roman Period (RP) stands out as the warmest period from the speleothem composite record (Fig. 4a)" is changed to "Considering the last 2500 years, the RP stands out as a clear warm period from the speleothem composite record (Fig. 4a)".**

L429 between? **Done**

L432 delete the square **Done**

L540ff this paragraph does not really add any value. Would delete completely… **Ok, we can remove it.**

L570 to 588: The comparison of speleothem data to PAGES reconstruction is not really a discussion of the drivers of this variability. I would bring this comparison much earlier to support the interpretation of the Pyrenees composite record as a temperature record. **We have changed this part of the discussion and moved it from 5.2.3 to 5.2.2. Thanks for the suggestion. We keep the PAGES2k temperature record in Fig. 6 since it is a representation of the temperature variable at a regional scale.**

L589ff I find the discussion of potential forcing mechanisms a bit superficial and short. Extending this discussion a little bit would make the whole paper less descriptive. **Agree. This section of the manuscript will be enlarged with more discussion on climate mechanisms.**

Why should the centennial scale be coherent across Europe, while the decadal NAO has a N-S dipole? **We were expecting that, if our record is dominated by temperature, the NAO influence may be lower than if the record is dominated by precipitation amount, a variable that controls precipitation patterns (more than temperature latitudinal patterns at centennial scale). In any case, compilations including records all over Europe, such as PAGES2k or treerings from Büngten et al. present a profile where all that variability among sites (sometimes in synchrony, sometimes in asynchrony) is already included. Part of this discussion is now included in 5.2.3.**

L610 again, contradiction to own argument, that discussing relative magnitudes is not meaningful. Please clarify… **Corrected.**

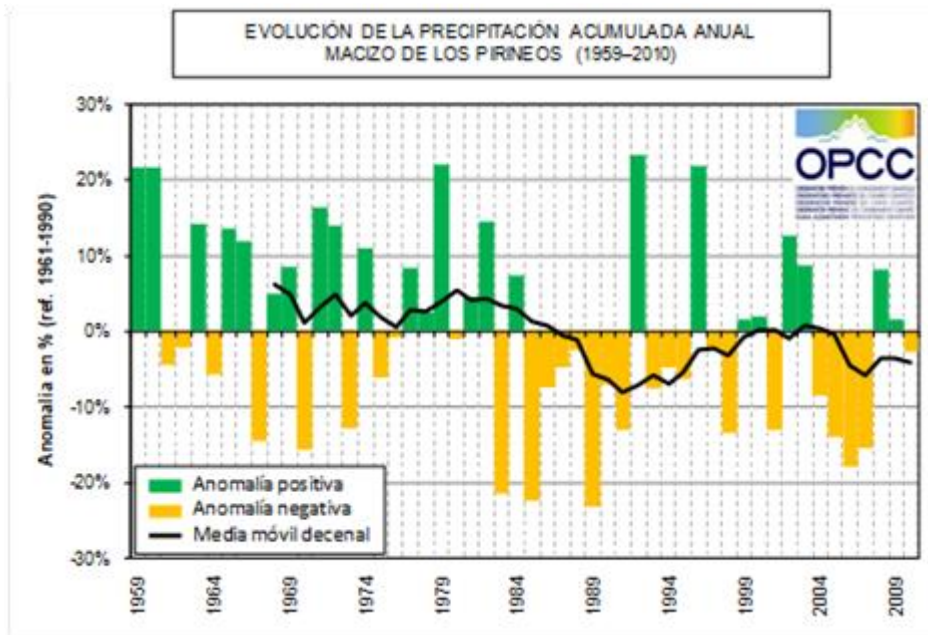
L613 any explanation/hypothesis why this is the case, also in the light that the alpine areas such as the Pyrenees should be more sensitive to the recent warming? Could the precipitation amount effect play a secondary role here as well? **It could be… but the decrease in precipitation for the last two decades is a very small one compared to the one in the 90s, according to last records compiled by the Pyrenean Observatory of Climate Change (see graph below). We have included a sentence about this hypothesis.**

L616 Again, why should Pyrenees, Central and Northern Europe be coherent? Is the nature of the forcing decisive if changes are coherent or not (as seen on decadal scale across Europe due to patterns like NAO, AO, EA/WR, …?, and if so what is the difference? **Yes, we think that the records will be more coherent if they are dominated by temperature than if they are dominated by precipitation since those atmospheric decadal-scale indexes (NAO, AO, EA…) are influencing precipitation in a very direct way. In any case, this is probably too speculative to be included in the text and it is only indicated as a hypothesis in section 5.2.3 about mechanisms.**

Figure 3: which age models are used, the preliminary StalAge models or the final iscam age model? **We used iscam age but not the final one. We generated Iscam model for Seso and Las Gloces caves separately, using 4 stalagmites in Seso and 2 in Gloces. Pot au Feu and B1 caves, that only have one stalagmite each, are here with the StalAge model. This is indicated in the figure caption.**

Figure 4: Not sure why the threshold values of ± 0.75 are highlighted when is has been previously claimed that relative magnitudes are not meaningful. Also, are colors colorblind-friendly? **We are not comparing warm periods among them but consider that the difference among warm and cold is significative. The threshold value is aleatory, we just wanted to highlight the most extreme ones.**

Figure 5: I find Fig 5 is a bit overloaded. Not sure if three regional rainfall and four temperature reconstructions are needed to be shown in this plot…? would also rather shift the NAO

reconstruction in the next figure (forcings). Also there are newer NAO/AMV reconstructions available with higher resolution (e.g., Hernandez et al. 2020, Lapointe et al., 2020, Becker et al., 2020, Ortega et al., 2015 ...). **Thanks for the suggestions. We prefer to keep the NAO reconstruction since Fig 5 is all speleothem records but will include another newer NAO record with higher resolution in Fig 6 (mechanisms)**

Table 2 is cut off and not complete. **Corrected.**

Figure A1: A: Spanish label in bottom left panel. Also: It might be more intuitive to show the r values in the proxy panels instead of age model plots. **Corrected**

Figure A2: The Seso composite seems to be rather uncorrelated or even anticorrelated to the overall composite record. Why is the Seso Cave record with the most speleothems not so dominant in the final composite? **Seso record is dominant in the last 200 years of the composite but not before (eg. LIA) when Las Gloces record is dominant. I think it is a question of Iscam software, not easy to correct or change.**

Figure A5: If the δ18O is detrended, could there be a correlation to precipitation amount for the 20th century? There is almost no effect when d18O is detrended. Visually, an **anticorrelation, at least for 1900 to 1980, is observed. Less negative values of d18O would correlate with higher precipitation, what does not make much sense in terms of precipitation amount influence. At the same time, we are aware that the chronological model is not so precise to resolve changes that are under decadal scale... At least it is not precise enough for that part of the record, where U-Th is not accurate and we just have 14C bomb peak data. We include this limitation in the manuscript.**

References

Ólafsdóttir, K. B., & Mudelsee, M. (2014). More accurate, calibrated bootstrap confidence intervals for estimating the correlation between two time series. Mathematical Geosciences, 46(4), 411-427.

Lu, J., Zhang, H., Li, H., Sha, L., Zhang, R., Zhao, J., ... & Cheng, H. (2022). Climatic and anthropogenic influence on vegetation in southeastern China during the past 120 years inferred from speleothem. Quaternary International, 625, 60-65.

Kathayat, G., Sinha, A., Tanoue, M., Yoshimura, K., Li, H., Zhang, H., & Cheng, H. (2021). Interannual oxygen isotope variability in Indian summer monsoon precipitation reflects changes in moisture sources. Communications Earth & Environment, 2(1), 96.

Warken, S. F., Fohlmeister, J., Schröder-Ritzrau, A., Constantin, S., Spötl, C., Gerdes, A., ... & Scholz, D. (2018). Reconstruction of late Holocene autumn/winter precipitation variability in SW Romania from a high-resolution speleothem trace element record. Earth and Planetary Science Letters, 499, 122-133.

Priestley, S. C., Treble, P. C., Griffiths, A. D., Baker, A., Abram, N. J., & Meredith, K. T. (2023). Caves demonstrate decrease in rainfall recharge of southwest Australian groundwater is unprecedented for the last 800 years. Communications Earth & Environment, 4(1), 206.

Treble, P. C., Baker, A., Abram, N. J., Hellstrom, J. C., Crawford, J., Gagan, M. K., ... & Paterson, D. (2022). Ubiquitous karst hydrological control on speleothem oxygen isotope variability in a global study. Communications Earth & Environment, 3(1), 29.

Hernández, A., Sánchez-López, G., Pla-Rabes, S. et al. A 2,000-year Bayesian NAO reconstruction from the Iberian Peninsula. Sci Rep 10, 14961 (2020). https://doi.org/10.1038/s41598-020-71372-5

Lapointe, F., Bradley, R. S., Francus, P., Balascio, N. L., Abbott, M. B., Stoner, J. S., ... & Labarre, T. (2020). Annually resolved Atlantic sea surface temperature variability over the past 2,900 y. Proceedings of the National Academy of Sciences, 117(44), 27171-27178.

Becker, L.W.M., Sejrup, H.P., Hjelstuen, B.O. et al. Palaeo-productivity record from Norwegian Sea enables North Atlantic Oscillation (NAO) reconstruction for the last 8000 years. npj Clim Atmos Sci 3, 42 (2020). https://doi.org/10.1038/s41612-020-00147-6

Ortega, P., Lehner, F., Swingedouw, D., Masson-Delmotte, V., Raible, C. C., Casado, M., & Yiou, P. (2015). A model-tested North Atlantic Oscillation