Reply to Reviewer Comment C9

Santos et al try to present the consistence/inconsistency of Portuguese palaoclimatic data from 4 different sources including reconstructed European grid based seasonal temperature of Luterbacher et al (2004)/Xoplaki et al. (2005), local repeated borehole temperature observation from one site in Portugal, palaeoclimate simulation and their signature in these boreholes and finally documentary based indices from documentary sources from the late Maunder minimum. Noting that the gridded reconstructions are not consistent with the other results, they propose a new reconstruction based on a two-stage calibration procedure using information from these sources, and compare the obtained results with a new derived series. Although this is an interesting approach, conclusions and interpretations are not supported by scientific evidence, parts are misleading and need reconsideration.

Reply: We would like to thank the reviewer for his/her comments. They were truly helpful in improving our manuscript. Nevertheless, we disagree that our results are not supported by scientific evidence, as we wish to enlighten in our point-by-point replies below. The corresponding changes in the manuscript are tracked in blue.

1) The argue, that the Luterbacher et al. and Xoplaki et al. data are not in agreement with the authors data, is confusing and not supported by this kind of analysis.
First, it is striking that the 20th century grid point data close to Lisbon from New et al./instrumental series are in very good agreement with each other (not surprisingly), but not with the simulations and the other paleo evidence presented by the authors.
There is a clear bias, also in terms of variability and trend. What are the reasons for this different behaviour? For Luterbacher et al and Xoplaki et al. New et al. 2000 data have been use to calibrate the statistical model within the 20th century. Meteorological station information and LMM indexed data from Lisbon have been included in the reconstructions.
Luterbacher et al. and Xoplaki et al. reported on uncertainties based on unresolved variability in the statistical calibration. The uncertainties have been shown for the European annual average and are (30year filtered) larger than 0.5_C around
1600. Assuming the uncertainties are similar at European scale and for the gridpoint close to Lisbon (clearly smaller during the LMM as indexed series from Lisbon have been used as predictor) then the simulations and the other palaeo evidence would certainly be within the 2 standard errors of the filtered time series around the Luterbacher et al. 2004 reconstruction. Plotting in addition also the uncertainties of the authors data (see major point below), then there is no support of the conclusions and criticism related to the Luterbacher et al. reconstruction.
Reply: Regarding the bias referred by the reviewer, we totally understand his/her criticism. In fact, the plot in Fig. 2b was misleading, because the time series are represented as anomalies with respect to their own full periods (different baselines). To improve clarity, this figure and respective caption were changed so that the same baseline is considered for all the plotted time series, i.e. the anomalies are now with respect to their common period (1901-1989). The new Fig. 2b now clearly displays the limitation of the Lut2004 reconstruction when compared to its calibrated time series (CalT). We believe that this modified figure can help clarifying these reviewer's concerns.

Furthermore, we may have misunderstood the reviewer's comment, but we would like to refer that assuming that the errors in the local temperature for Lisbon are of the same order of magnitude than those for the European average is not correct at all. The error in the spatial average depends on spatial correlations between local errors. However, these local errors are not indicated in Luterbacher et al. (2004). Additionally, they do not provide errors at 30-year time scales, since their calibration period is much longer. The errors at 30-year time scales also depend on the temporal autocorrelation of the annual errors, which are also not indicated in Luterbacher et al. (2004). Thus, it is not possible for us to establish the local uncertainty at the 30-year time scale in Lut2004.

An additional point is that the disagreement between Lut2004 and the simulations is not random, but systematic. Therefore, even if the differences would lie within the uncertainty intervals, there would still exist a systematic bias between both series.

For further information on the uncertainties, please see our reply to the comment below.

2) Importantly, it is commonly practise to present uncertainties for new quantitative reconstructions. Can I please ask the authors to show results including those by clearly providing evidence on how the uncertainties have been calculated both statistically for the different types of data and if possibly also applying expert judgements?
Reply: We agree with the need for providing uncertainty measures in the reconstructions. However, we would like to stress that our study is not a reconstruction itself, but rather a post-reconstruction adjustment of the Lut2004 reconstruction. Therefore, the uncertainties in the Lut2004 reconstruction also remain in our CalT. The only difference is that the raw reconstructed time series for Lisbon was calibrated/corrected in terms of bias, scale and, more importantly, in terms of long-term trends that are not properly resolved in the raw series.

As we also think that some kind of uncertainty measure should be given, the non-linear trend is now accompanied by an uncertainty interval obtained from the two paleoclimatic simulations. Fig 4a was changed in order to depict this interval and the following sentence was added to the manuscript:

Lines 299-308: "*The uncertainties in the CalT series are the combination of the original uncertainties in the Lut2004 dataset plus additional uncertainties related to the non-linear trend used in the adjustment. The former are discussed in detail in Luterbacher et al. (2004). The latter can be estimated through the assessment of the consistence between Sim1 and Sim2. For this purpose, the SSA filtering was applied separately to Sim1 and Sim2. The mean absolute difference between the two non-linear trends obtained from Sim1 and Sim2 provides a measure of the uncertainty related to the simulations. It has an approximate value of 0.05ºC. However, this number provides just a lower bound, since it does not explicitly consider uncertainties related to the simulation itself, which are difficult to assess due to the limited number of available simulations with similar characteristics.*"

3) In order to show the full spread of paleo climate simulations I suggest to show also the new IPCC AR5 paleo GCM runs, selecting one or two gridpoints around the area of interest. Two simulations from the same model are not fully appropriate for this kind of analysis. Related to this point concerns the Fig 2b and corresponding text: The match between the profiles and the paleo simulations could be by chance, therefore more model outputs would be necessary to get an idea of the full spread from paleo simulations.
Reply: We understand this reviewer's concern and we thank his/her suggestions. However, we would like to note that there is a trade-off selecting climate simulations for this study. On the one hand, there are several available GCM simulations, developed within the last IPCC framework, which allow sampling the uncertainty space of climate simulations. On the other hand, GCMs present a number of limitations, being their coarse resolution (grid spacing of ca. 200 km) the most relevant for the sake of this regional study. From our viewpoint, GCM simulations are not suitable for such a regional study. Although not so abundant, RCM simulations largely overcome this limitation, by more accurately reproducing regional climates through dynamical downscaling. Since this is precisely the signal recorded by the reconstruction employed in our study (Lut2004), we decided to use such high-resolution data from two regional paleoclimatic simulations (Sim1 and Sim2). Further, the skill of these simulations has already been documented over Portugal in previous studies. Unfortunately, at the moment, these are the only two available regional simulations covering Portugal with the necessary

characteristics. Hence, it is not feasible to increase the ensemble size of model simulations, though it would be very useful for assessing the uncertainties referred by the reviewer. However, this is a valuable suggestion for future research, when more regional simulations will gradually become available over Portugal.

As we think this is a very pertinent issue, we have added the following discussion to the last section:

Lines 359-369: "*Due to the relatively coarse spatial resolution of data generated by state-of-the-art GCMs, they are not suitable for regional-scale assessments. Since such scales are precisely the focus of this study, temperature series from two high-resolution regional paleoclimatic simulations (Sim1 and Sim2) are employed instead of GCM runs. These two simulations were documented and validated in previous studies. Unfortunately, there are only two available simulations covering Portugal with such high-resolution characteristics. Hence, it is not possible to increase the ensemble size of model simulations, though it would be very useful for uncertainty assessments. In forthcoming research, new regional paleoclimatic simulations over Portugal, also using different models, should be used to enhance the robustness and evaluate the significance of the current adjustment*."

4) I agree with the other reviewer that the method for re-calibration process and the underlying assumptions are not clear and need more explanations.
Reply: Thank you for your advice. Following this comment and the other reviewer's comments, we have revised the whole manuscript so as to improve the description of the applied methodologies.

5) The 20th century gridded data presented in Luterbacher et al and Xoplaki et al. are not from Mitchell and Jones, but from New et al. 2000, J. Climate
Reply: Thank you for this correction. This reference was changed accordingly.

6) The summary and conclusions part are too long and include many repetitions. This part needs to be shortened and rewritten taking into account the new analysis from the reviews
Reply: This section was fully revised taking into account the two reviewer's suggestions, including further discussion related to the point 3 above. In addition, the former first two paragraphs in this section were shortened and rephrased in order to avoid repetitions.

7) The abstracts mentions twice 'attributed'. The study clearly does not deal with attribution. Therefore this should be reformulated.
Reply: We agree that this term might be misleading. The term 'attributed' was changed to 'commonly related' at the two instances.

8) The argumentation for this kind of paleo study in the introduction needs more convincing evidence
Reply: We understand this concern and thank the reviewer for his/her advice. We totally agree that the motivation of our study must be clearly stated. Therefore, we have added the following text to the introduction:

Lines 86-92: "*The identification of possible inconsistencies with the above-referred data sources enables a post-reconstruction adjustment of this time series. In effect, this calibrated time series may help understanding past climate variability in Portugal and its main driving mechanisms, namely the role of external vs. internal forcing mechanisms on temperature*

variability. This attribution analysis provides critical information for model validation and for assessing the reliability of regional climate change projections."

9) It would be nice if the time series could be updated to 2014
Reply: Unfortunately the ECA&D project does not provide information for Lisbon after 1999. As such, we cannot extend our analysis to the most recent years. Nonetheless, as the focus of our study is more on the reconstruction of past climate variability, we think this very recent data is of secondary relevance for the present study goals.

10) The whole paper needs a English check
Reply: A final English check was carried out.