Response to Reviewer 2

We would like to thank the reviewer once again for his/her exhaustive and constructive comments. We sincerely appreciate the time and effort the reviewer dedicated to reviewing our manuscript.
In this response, we will provide more detailed explanations addressing the reviewer's comments, as well as presenting our plans for further analysis.
Our responses are in blue font.

Main comments

Before I list specific points in the manuscript, I wanted to raise the issue to the authors (it is not a must, more a proposition): I personally found Fig. 2 and 3 very interesting but under-utilised (and to some extend Fig. 6 too). Clearly, some models are not very skilful at capturing the synoptic climate during historical droughts and this raises in itself an important issue:

1) how can we confidently understand drivers using longer simulations? What setup does "work better" to capture what part of the signal (temporal vs. spatial)? I know that expending on this aspect would modify the paper substantially but it is of crucial importance to build on. It would also give more confidence to the method applied later on to determine the dominant climatic modes (I must say, as is, I am wondering how much trust we can have – which might also be due to the fact that I am not savvy with the methodology used here).

Previously, we gave a short first response, but here we add some more of our thoughts on this question and how we plan to tackle it.
To understand which set-up works better, it is necessary to perform sensitivity tests changing model configurations which would not be possible using all these CMIP5 models and is out of the scope of this study. Our objective is to identify extra-tropical patterns related to past droughts in CMIP5 model simulations and to understand the difference between these model simulations.
Importantly, we cannot expect the same responses from all models due to their internal physical differences, initial conditions, and model-dependent internal variability (which is what our result also shows at the end).

As we responded previously, in general, if the models are able to represent well the present-day climate modes in an acceptable way (as performed in several studies, e.g. Fasullo et al., 2020; Deser et al., 2018), we assume that the representation of the modes would be still valid in longer simulations.
To test whether the simulations work well, we compared the variability of soil moisture (Fig. 2.a) and soil moisture-geopotential height correlations (Fig. 3) between the CMIP5 simulations and the observation-based data in the manuscript. To assess which one is still better than others at representing the variability of soil moisture, we will add the ensemble spreads of each individual CMIP5 model and compare them against NOAH-LSM in Fig. 2.a in the revised manuscript. In addition, we will add more discussion regarding these methods and complications of model-observation comparisons in past and present simulations.

2) Another (more minor) issue: Perhaps I am not familiar with the term, but it would be useful to clarify what you mean by "internal variability" (and that might be done by just explaining what is considered an external forcing in your study). If not used to the terminology, one might wonder for instance, whether teleconnections are part of the internal variability or not?

With internal variability, we refer to different modes of climate variability involved in droughts. By external forcing, we mean volcanic and solar forcings. However, volcanic forcing is an internal climate forcing and not an external forcing. We will correct and clarify better what internal variability and forcings mean in our study in the revised manuscript.

3) Abstract: lines 4-5: "The focus is [...] during 850-2005, this excluding the anthropogenic trends from 1850 CE onwards". This is very confusing... It can only be understood if the methodology has been explained, otherwise one starts to wonder which time range was actually used.

We will correct the sentence.

4) Introduction: Lines 66-71, the definition of the knowledge gap and the leading questions is a little weak. "necessary to understand which modes of climate variability or atmospheric circulation patterns are involved in each climate model": this sounds very ambitious and is actually not what this study is tackling. Instead, it might be useful to explain why studying the "patterns in the mid-latitudes" and their effect on Mediterranean climates is important.

Thanks very much for the point. We will include more details on this in the revised manuscript.

5) S3.1: this section of the methods is at times quite confusing. I have read the comments of Rev. 1 and agree with their comments so I will just emphasize two points that I found particularly confusing:

As both reviewers commented, we will go through and revise the section for the next phase.

6) Line 112: "We use annual mean anomalies in order to include winter conditions in the analyses [...]": I don't understand this sentence.

We will correct the sentence as reviewer 1 also commented about it.

7) Lines 121-126: This part is at the same time important and confusing: why not normalising the whole record instead of just the Hist part? Would a z-scoring be useful?

The aim of using the method in line 121 is to remove the trends introduced by the anthropogenic increase in GHG after 1850 to focus on studying the natural variability of drought during the last millennium. We will write this paragraph better for the next phase.

8) S3.5: section very hard to follow, please provide additional information and background on the choices of the methods and their limitation.

We will go through and revise the section.

- Results

9) Fig.2: if I understand right, this is your control run for time series.

All time series are from forced last millennium and historical simulations. We will include this detail in the caption.

10) a) the modelled soil data are quite different from the observation time series (range and trend are similar but changes are often not synchronous). How does that affect your analysis? What does the intermodal comparison look like? Could you have a look at the cross-correlation between NOAH LSM and CMIP5 mean? Just so that we understand the limitations of the study.

We do not expect the models' soil data to be exactly synchronous with the observation-based time series. That is basically not possible because of the difference in the initial conditions between the models and the observation, and internal variability of the climate system. As long as the variability of the observed variable lies within a model spread, we assume that the model is able to represent the temporal variability of the observations (e.g., Liu et al., 2012). For the same reason, a cross-correlation analysis between NOAH and CMIP5 would not provide additional useful information. However, we can examine the spread of each individual model utilized here (as we responded to comment 1) to verify each model's soil variability and possible limitation compared to the observed data. We will include more discussion on this in the revised manuscript.

11) Line 238: "SOIL and Z500 from NOAH-LSM" isn't Z500 from ERA?

That is true. We will correct it.

12) Fig.3: control run for spatial analyses. I think that it would be good to indicate that the upper left panel shows the observation (SOIL vs. Z500). How do you deal with models not capturing the synoptic climate very well (are they still used for further analyses, and if so, why?)? Can you quantify the skillfulness of the models? To me this exercise seems to be providing an evaluation of the models to track spatiotemporal climate conditions associated to droughts. It could be a goal in itself (as proposed earlier).

We did not have control simulations in our analysis (see our response to 9). Regarding the question of how to deal with models that do not capture well the observed synoptic climate, as one of the objectives of our study is to understand the model differences in representing drought-related circulation over the Mediterranean region, we do not apply a selection criterion to choose the best or worst model (Also see our response to comment 1). Instead, we address and visualize the differences between the models in representing past drought-associated extra-tropical patterns. We already discussed briefly the implications of the model discrepancies in the discussion section (for instance, about MIROC in lines

370-373), but we will extend the discussion more and clarify our objective in the revised version.

Regarding the skillfulness of the models, it would be hard to assess the skillfulness of all variables associated with droughts, and there is no universal metric for it. We somehow asses the skillfulness of the models in representing soil moisture-associated circulation patterns by comparing the correlation fields between the soil moisture and geopotential heights of the observation-based data and model simulations in Fig 1. We can add, for instance, root mean square errors calculated between the spatial correlation coefficients in Fig 1 to summarize the figure and come up with some numbers.

13) Lines 286-288: I struggle with these sentences. Either too little or too much is said here. What is the role of model's setup and skilfulness in this observation?

We will correct the sentences for clarification. About the models' setup and ability, see our response to comment 1, and about their skillfulness, see our response to comment 12.

14) I also do not understand what is meant by "counterfactual"?

By "counterfactual", we refer to the climate conditions without the anthropogenic trends from 1850 onward, so the detrended Historical simulations. We will modify this word and also rephrase the sentences for clarification.

15) Fig.4: Would it be possible to quantify the antiphase? (bi-plot, cross- correlation)

Thanks for the suggestion. We will include more analysis of the time series (for instance, as the reviewer suggests, a cross-correlation analysis between the west and east regions) in the revised manuscript.

16) How do these time series compare to actual data (e.g., OWDA, HYDRO2K), even if the variables are different, there should still be some similarities?

We provide here a more extended response than the previous one:
Regarding actual data (for instance, OWDA), Kim and Raible (2021) have performed the comparison between summer droughts in a CESM simulation and OWDA, and found differences in the temporal occurrence of droughts between the model and the proxy-based reconstruction. If we include OWDA, we expect to see a similar result here (not temporarily synchronous). Also, OWDA provides only the summer hydroclimate variability of the region and not all annual means. We will include more discussions on the finding of Kim and Raible (2021, Clim. Past) in the revised manuscript.

17) A minor issue but might avoid confusion: I would recommend using years CE/AD on the scale (I am more often dealing with yrs BP, so I tend to read that directly... It's a matter of community standards)

We will correct this in the revised version.

18) Line 319: it might be me not being familiar with the jargon(and perhaps I missed it earlier, sorry): what are EA-WR and Eastern Atlantic pattern (perhaps need some description earlier in the manuscript?). I would generally recommend refraining from using abbreviations if not needed, it is much nicer to the reader in plain words.

They are the circulation patterns in the mid-latitude that mainly affect the European and Eurasian domains. We will include a brief description of these patterns or at least include references for them and not use abbreviations to avoid confusion.

19) Fig.5: how reliable are these results with regard to Fig.3?

In the manuscript in lines 371-373 (Discussion section), we briefly discuss the representation of circulation patterns in MIROC which is different from others relating to what is observed in Fig 3. We will extend this discussion to the result of Fig. 5 to all the models in the revised version.
Also, see our response to comment 12.

20) Fig.6: Very hard to read and capture the essence of these results. But once again, it seems to me here that the authors are showing an evaluation of the different models, so there is a tension between showing actual climatic results (fig 5) and showing how models perform and the inter-model spread, which is hard to reconcile.

Same as the response to comment 19.

21) Line 338: "the mean occurrence of pattern groups for each model": what is exactly meant here?

We refer to a mean frequency (in terms of the count of patterns every century). We will rephrase the word.

- Discussion:
22) On the NAO/EA-Mediterranean climate relationship: this is very interesting but I am a little puzzled to see that both NAO configurations can be associated to droughts. I am wondering if that could be discussed more?

We consider droughts that last at least two years (lines 155). During these periods, there could be fluctuation in climate patterns, although it looks like positive NAO-like conditions occur much more frequently (patterns 1 and 3 in Fig 5). We will add more discussion on this in the revised manuscript.

23) Another point of curiosity: can you detect in the time series (not averaged in 100-yr windows) any fingerprint of NAO (e.g., 7 yrs periodicity)? The identification of a "multi-decadal scale anti-phase" is very interesting and could also be further explored (is there an oscillatory component? Has it been observed before?). I had to think of an article by Mann et al. (Nat. Comms 2020) "Absence of internal multidecadal and interdecadal oscillations in climate model simulations", not sure if it is relevant here.

To detect the periodicity, we need to perform extra analysis like a spectral analysis (e.g., Stevenson et al., 2018). We will perform the analysis of the time series of soil moisture anomalies and the time series of occurrence of droughts in Fig. 4 to also quantify the periodicity of the occurrence of droughts. We will be able to add more discussions to the reviewer's questions and address the "absence of internal multidecadal oscillation in climate models" with the result from the new analysis.

24) Is there any way possible to test the observed seasonality pattern of climatic associations in the OWDA?

We did not include the time series of OWDA, as our focus is on how some climate models represent droughts and associated extra-tropical circulation on an annual time scale. OWDA is the reconstructed time series of summer (growing season) hydroclimate. We do not think we can see similar temporal and spatial patterns to OWDA (see our response to comment 16). Regarding the spatial circulation patterns, summer circulations over the region are in general much weaker than those during the wet period, which is shown in Kim and Raible (2021) Fig 9. We can provide more details on this in the revised manuscript.

25) Do you also observe a N-S antiphase (also mentioned in Markonis et al. 2018)?

About the N-S antiphase in Markonis et al. (2018), in my understanding, the N-S antiphase occurs between northern Europe and southern Europe. As our study covers only the Mediterranean region (southern Europe), we would not be able to observe a similar result to Markonis et al. (2018).

26) The E-W antiphase: is it stable on all timescales (Indeed the results you obtain are contradictory to Cook et al. 2016, this might be further discussed)?

To answer this question, we need to perform further analysis of the time series, for instance, a wavelength coherence analysis between the west and east time series to compare their frequencies. We will add more details and discussion about the antiphase when the analysis is performed for the next phase.

27) I am also surprised to read that climate background (e.g.,global temperatures) had no effect on droughts (or did I understand wrong?). See for instance Dermody et al. (Clim Past 2012).

We did not particularly examine how the global temperature (or SST as used by Dermody et al., 2012) is related to the Mediterranean droughts. However, the changes in the background climate condition caused by volcanic eruptions can influence the Mediterranean hydroclimate on an annual time scale by bringing more wetter conditions (Gao et al., 2019; Kim and Raible, 2021), as we mentioned in lines 53-55. This change seems to be more dynamically driven than thermodynamically (Rao et al., 2017).

28) Minor comment: perhaps cite Douville et al. (2021,Water cycle change IPCC report) instead of Masson-Delmotte et al. (2021)?

We will change it in the revised version.

29) Finally: I liked the approach of Hanel et al (Sci. Rep.2018) to distinguish meteorological (rainfall), agricultural (soil moisture) and hydrological (runoff) droughts. Perhaps an idea here to frame the research (most previous papers are looking at meteorological droughts)?

Thanks for the suggestions. We will be clear that our focus is on soil moisture droughts in the revised manuscript.

References

Gao, Y., & Gao, C. (2017). European hydroclimate response to volcanic eruptions over the past nine centuries. International Journal of Climatology, 37(11), 4146-4157.

Kim, W. M., & Raible, C. C. (2021). Dynamics of the Mediterranean droughts from 850 to 2099 CE in the Community Earth System Model. Climate of the Past, 17(2), 887-911.

Liu, C., Allan, R. P., & Huffman, G. J. (2012). Co‐variation of temperature and precipitation in CMIP5 models and satellite observations. Geophysical Research Letters, 39(13).

Rao, M. P., Cook, B. I., Cook, E. R., D'Arrigo, R. D., Krusic, P. J., Anchukaitis, K. J., ... & Griffin, K. L. (2017). European and Mediterranean hydroclimate responses to tropical volcanic forcing over the last millennium. Geophysical research letters, 44(10), 5104-5112.

Stevenson, S., Overpeck, J. T., Fasullo, J., Coats, S., Parsons, L., Otto-Bliesner, B., ... & Cole, J. (2018). Climate variability, volcanic forcing, and last millennium hydroclimate extremes. Journal of Climate, 31(11), 4309-4327.