Manuscript egusphere-2023-119

Dear Editor Dr. Goose,

We would like to thank you again for this opportunity to submit our manuscript and for dealing with it. We also would like to thank the reviewer for his/her constructive comments. We sincerely appreciate the time and effort the reviewers and the editor dedicated to our manuscript. Note that because some new results were added, the manuscript is updated, and some parts were modified (these parts are marked in the marked-up file).

In this response, we provide more detailed explanations addressing the reviewers' comments. Our responses are in blue font, and the line numbers are within [the brackets].

The response to Reviewer 1 is on pages 2–16, and for Reviewer 2 is on pages 17 –27

Sincerely,

Woon Mi Kim, on behalf of all authors

Response to Reviewer 1

Abstract:
0) I think the abstract is in general well written and it emphasizes the findings of the manuscript.

Thanks very much for the comment. We'd like to inform you that the abstract is modified in the revised version. We included some more details as some results have been updated.

GC2. Introduction:
I think the introduction does its job in representing the frame of the problem and the case for this study.

Thanks very much for the comment.

There are a couple of issues I would like the authors to consider though:
GC2.
1) Page 1, lines 21-23. 'The climate of the...' This paragraph divides the seasons into wet and dry and refers to the wet one specifically as the winter season. There is considerable spatial variability in the rainy seasons in the Mediterranean lands and for many, the wet seasons are spring and autumn. I think this winter/summer separation may be misleading. Consider for instance Xoplaki et al (2004; DOI 10.1007/s00382-004-0422-0) or others, where an extended wet season is used. I think this discussion should incorporate better this seasonality character. Note page 2, line 24 where you mention 'highly variable seasonal hydroclimate...', although the end of this paragraph (line 30) will make the emphasis again on winter.

Thanks again for the suggestion. As the reviewer suggested, we corrected the corresponding sentences to:
[30-34] "While considerable spatial and temporal variability exists across the Mediterranean region, the climate is characterized by a wet cold season and a dry hot season (Xoplaki et al., 2004; Lionello et al., 2006). The wet cold season, typically extending from fall to spring, is a crucial period for moisture supply, where a large proportion of precipitation is provided by mid-latitude circulations and westerlies. Therefore the season dictates the intensity and impacts of droughts on the region."

2) Page 3, lines 64-65. 'They also found that ... is model-dependent'. This seems to relate specifically to the results of the current manuscript and perhaps should be discussed specifically either here, providing more specific background and describing in which way this can happen, or/and at the end of the manuscript discussing what this manuscript provides in
the context of previous evidence.

We included some more details on Xoplaki et al. (2018):
[74-80] "Using simulations from the CCSM4 and MPI-ESM climate models, Xoplaki et al. (2018) showed that multidecadal variations in the eastern Mediterranean hydroclimate are explained by internal climate dynamics. By comparing three historical periods with large

hydroclimate events over the region, they found notable differences in the climate patterns during the same periods between the two models. The observed discrepancies in climate patterns and timing of hydroclimate events between the models, and also between the models and the proxy records, indicate that exact temporal and spatial agreement of events between climate models and proxy records cannot be expected."

GC3.

3). Section 2.1 Page 3, lines 83-86.

The CMIP and PMIP and other model efforts could include a reference. I am aware they were just provided in the previous paragraph, but it seems pertinent to me that they get those also in the Data section. I leave this to the taste of the authors.

We included more references in the Data section, including the references for each of the models used in this study in Table 1.

4). Section 2.1 Page 3, lines 85+ 'We consider only simulations...'

- There are some issues that can be considered related to the use of soil moisture in different models: The models have very different depths and therefore they will possibly produce different soil moisture statistics, even if depths only down to 0.7 m are considered. Using this set up of models is fair and possibly it contributes to inter-model differences, but also it is arguable that models with shallower depths may be less realistic. Note that some models like GISS only have a depth of 3m. This will produce potentially a different vertical distribution of moisture and different temporal variability. Perhaps it is something the authors would like to comment on.

Thanks very much for the suggestion. We discuss more the difference in soil moisture from the models in the discussion section of the revised manuscript:

[468-480] "Particularly, MIROC-ESM has only half of the grids compared to bcc-csm1-1 over both the western and eastern regions and presents relatively coarse vertical soil layers (Table 1). The coarse horizontal resolutions over the land and atmosphere may affect the variables associated with soil moisture, such as precipitation, which is sensitive to the horizontal grid size (Haren et al., 2015). In addition, poor land vertical resolutions are probably insufficient to represent soil hydrology associated with vegetation and soil moisture memory effects that affect regional hydroclimate conditions (Hagemann and Stacke, 2015). This also can be the reason why MIROC-ESM shows spatial correlation patterns that differs distinctly from NOAH-LSM outside the focus region in Fig. 3.
Assessing which model represents soil moisture variability better is a complicated task since it is known that the magnitudes of soil moisture depend largely on the internal physics of land surface models (Fang et al., 2016; Berg and Sheffield, 2018). Moreover, soil moisture interactions with the atmosphere (Berg and Sheffield, 2018) and vegetation dynamics related to soil processes (Huang et al., 2016) vary across the CMIP5 models. Choosing the same vertical soil moisture level (70 cm) across all the models to represent ecosystem-related depth, i.e., root zone, over the entire Mediterranean region may be another influential factor since the root zone varies with the region (Kleidon, 2004)"

And as we responded before, a soil depth below two meters is generally considered less important for atmospheric processes. Hence, the fact that GISS has only a three-meter depth would not significantly affect the atmospheric processes we focus on.

5) In Table 1 it is included that a depth of 0.7 m is considered. I think this is relevant and should also be included in the text. It would be desirable to include a rationale for this decision and how it may influence the timescales of variability of soil moisture. What is the typical range of depths for the bedrock, that limits the presence of water, for the area of interest in general in these models? Including a shallow limit may be more representative for the whole area but also exclude lower frequency variability more typical of deeper levels. I think including a rationale for this in the text of Section 2.1 would be good.

We included that we used the 70 cm level and the reason for it in:
[170-172] "The 70 cm level is a deep soil level that can reflect the impacts of soil moisture change on vegetation and ecosystems, hence, better representing soil moisture droughts, including their persistent characteristics (Dirmeyer, 2011; Ghannam et al., 2016; Esit et al., 2021). "

Rather than the bedrock, we discuss briefly root-zone level as it is the level related to vegetation and is important for soil moisture droughts.

[478-480] "Choosing the same vertical soil moisture level (70 cm) across all the models to represent ecosystem-related depth, i.e., root zone, over the entire Mediterranean region may be another influential factor since the root zone varies with the region (Kleidon, 2004)"

Also, refer to our response in 4)

6) 3. Section 2.2. Why are NOAH-LSM data used for soil moisture and ERA5 only for circulation? I mean what does NOAH-LSM offer that you would not get in this manuscript from ERA5 soil moisture?. There is probably good reasons for it. I would just suggest that a motivation for the use of these data, beyond the fact that soil moisture observations are scarce, is provided.

As we responded during the first revision phase, we used the soil moisture from NOAH-LSM instead of ERA5 because NOAH-LSM is forced with the observation-based dataset. Therefore, we assume that NOAH-LSM could be more appropriate to show more realistic present-day soil moisture variability. We included this detail in:

[124-126] "We use the NOAH-LSM land variable instead of the one from ERA5, as NOAH-LSM is forced with the biases-corrected observation-based datasets. Therefore, NOAH-LSM could be a better choice that reflects more realistic present-day soil moisture variability."

7) Line 87 '... analysis: Giss-E2-R, CCSM4...'
References should be provided for these simulations. Considering previous points, I would suggest that a minimal background is at least provided in how different the soil moisture modelling can be among all models. This would help giving the reader an

idea of what to expect in terms of inter-model differences. For instance, vertical resolution can be very influential for soil moisture, as indicated in line 90...

Thanks for your suggestion. We included more information in Table 1, such as the names of the land component models and the references for each model.

8) Line 87 '... 12 ensemble members of CESM1'
There are 13 available if I am not wrong. Nothing critical but maybe you want to state why that selection. Here it can also be stated this refers to the all forcing simulations, although it is quite clear in the context (see next).

When we retrieved the CESM-LME dataset from https://www.earthsystemgrid.org, the first ensemble member (member 001) of the variable geopotential height (Z3) was missing for 850 – 1849 CE. Hence, this member was not taken for the analysis.
We included the number of CESM-LME members (2 to 13) we took from the data portal in

[104-105] "we use 12 ensemble members of CESM1 from members 2 to 13 available in https://www.earthsystemgrid.org."

9) Line 91 ' All simulations were run with the volcanic, solar...' It is clearly stated earlier in the introduction that this is all about internal variability. Perhaps some comment about focussing on forced simulations and not in long control runs (also available in cmip) is pertinent.

We primarily focus on investigating the temporal variability of droughts and the associated mid-latitude circulation patterns during the last millennium (850-2005 CE), and therefore, we use transient simulations instead of control runs. From the previous studies (e.g., Cook et al., 2017; Rao et al., 2017; Xoplaki et al., 2018), we know that droughts over the region are mainly driven by the internal climate dynamics, but we do not assume this from the beginning in our study. We were not clear about it in the former version, therefore, in the revised version, we clarified our motivation in the abstract and the introduction better, adding that the objective is to understand drought-related circulation patterns during 850-2005 CE in the coupled climate models:

[2-5] "The Mediterranean region is expected to experience significant changes in hydroclimate, reflected in increases in the duration and severity of soil moisture droughts. While numerous studies have explored Mediterranean droughts in coupled climate models under present and future scenarios, understanding droughts in past climate simulations remains relatively underexplored. Such simulations can offer insights into long-term drought variability that observational records cannot capture. Therefore, our study investigates circulation patterns in the Euro-Atlantic domain associated with multi-year soil moisture droughts over the Mediterranean region during the last millennium (850–2005 CE) in CMIP5-PMIP3 and CESM-LME simulations"

[83-85] "the objective of this study is to examine how different coupled climate models depict the temporal variability of droughts and associated circulation patterns during the last millennium. Additionally, we assess the discrepancies among the models in representing drought-related circulations."

10) Line 101 ' ...has four layers up (down) to two meters'
Ok, but I understand that only the soil moisture of the first 0.7 m is used, right? Explaining this in relation to the previous information in Section 2.1.

Yes, from NOAH-LSM, also soil moisture at a 70 cm level was used. Here, we wanted to provide some more detail about the soil variable from NOAH-LSM. To avoid confusion, we slightly modified the sentence to:
[117-118] "The soil moisture variable from NOAH-LSM has four layers that extend to a depth of two meters"

And in the method section, we mention that we take soil moisture at 70 cm from all data sets.
[130-131] "The soil moisture content of soil layers from NOAH-LSM and the climate models are vertically integrated to 70 cm."

GC4. 3. Methods
3.1 Calculation of anomalies.
I find the definition of annual anomalies and the explanations for it somewhat confusing. There is an emphasis in this explanation that I do not really understand. Perhaps this is my fault but, please, consider it and see if this influences the text as it stands now.

11) 1. Section 3.1 Page 5, lines 111+.
- 'These variables are transformed into the annual mean anomalies at each grid point'. At this point I would say it is not possible to understand what is meant if additional information is not provided first, and I would argue to simply write '...transformed into annual anomalies'. I think that will be enough also for the rest of it.
- 'We use annual mean anomalies in order to include winter conditions in the analysis, as it is an important season for the annual hydroclimate in the Mediterranean'
I think it is important to discuss this in the context of what is indicated in GC2.1. Specifically for this sentence, it is not possible at this stage to understand what is meant. Not a big thing because one can perhaps understand it from the following, but: a) I would suggest to include the description for the intra-annual anomalies before and then explain they are averaged; if these arguments stay in the text.

Thanks very much for your suggestion. We included the detail you mentioned and modified the paragraph to:
[133-137] "The variables are all transformed into annual mean anomalies. Wet seasons are critical periods for the moisture supply of the region, and therefore, when strong circulation patterns take place (Xoplaki et al., 2004; Lionello et al., (2006). Therefore, we use annual mean anomalies to capture the mean variability of hydroclimate throughout the entire year instead of focusing on some particular seasons, i.e., summer growing seasons. This is to include the influences of wet seasons, which is the crucial season for moisture in the region, in the analysis."

12) b) I would actually (respectfully) challenge that this is true. I think it is not, but maybe I understood things wrong. See below.
- 'Prior to the anomaly calculation ... to the 70 cm level'.
These two sentences are relevant and the reader could go with knowing this earlier in the data section (GC3.2).

We moved the sentence earlier and modified it to:
[130-131] "The soil moisture content of soil layers from NOAH-LSM and the climate models are vertically integrated to 70 cm."

13) 'Then, first, the monthly anomalies... monthly values' [...] 'Second, the necessary annual mean ... these monthly anomalies'.
I do not understand these two sentences. If you calculate monthly anomalies (differences between monthly data and the long-term monthly mean you are essentially subtracting the annual cycle. If you average out the resulting monthly anomalies what you get is the annual anomaly, equal to calculating anomalies from annual data by subtracting the long-term mean. Therefore, I do not understand the previous sentence about the importance of winter or the unnecessary emphasis on the monthly anomalies if you end up in annual anomalies. The sentences in the middle about reference periods read fine to me. Perhaps it fits better with the next paragraph and the issue of the trends since their definition is somewhat related.

We agree with the reviewer's comment. We removed all the redundant sentences about calculating the annual cycles and modified the pointed part as:
[137-138] "The annual mean anomalies are calculated by subtracting the multi-year means from the annually averaged time series."

14) 'Thereby, the ensemble means of the anomalies... Maher et al. (2018)'.
With the preceding sentences I am confused. I understand the purpose is subtracting the ensemble average to each member to get rid of the long-term trends; or better said of the forced response in general as it will also affect volcanic events for instance. This would be consistent with the following arguments in the paragraph. Consider explaining/rephrasing these paragraphs.

As we responded in the previous phase, the method would not affect the volcanic events occurring in the pre-industrial period (850–1849 CE) as the ensemble averages were only subtracted from the historical simulations. We have revised the paragraph related to this particular methodology for further clarification:

[141-147] "In the next step, we remove the strong unprecedented trends in the Hist simulations (1850--2005 CE). To achieve this, we calculate the ensemble means of the annual anomalies for each climate model during the period 1850--2005~CE. Then, for each model, we subtract their corresponding ensemble mean from each of the ensemble member anomalies. The approach follows a similar method to Maher et al. (2018), which aims to exclude a trend caused by increasing anthropogenic GHG concentration from a time series. Note that the method is not applied to the LM period, therefore, the anomalies during LM still contain forced signals such as those from volcanic eruptions. The number of ensemble members can have an influence on the final output of Hist."

15) 'This method guarantees that only internally driven variability remains in the time series of the variables'. ... mostly internally driven variability, but only in the post-1850 period. This should be perhaps indicated specifically. However, what I miss a bit is the rationale of why it is done like this. Why is the forced signal intended to be filtered out in the post-1850 (natural and anthropogenic) and not in the pre-1850 period? The reader has to guess that perhaps this is intended to get rid of the long-term trends, and the rest of the forced signal goes with it

As we responded in the previous phase, the filtering is applied only to the post-1850 period mainly to remove the anthropogenic GHG forcing, which becomes apparent after 1850.
Refer to our response in 14).

16) Note also that the ensemble is calculated over different models in CMIP5-PMIP3 if I understood well (?). Now this is an issue that can be relevant, as different models will show different levels of response to external forcing. When this is done over an ensemble with the same model you can reduce the external response in each model run as it is assumed that the external response is common to all runs. One aspect here is that depending on the number of runs you are also subtracting also internal variability that has not been averaged out in the ensemble average, and this is a limitation that can be discussed.

Refer to our response in 14).
The ensemble means are calculated for each of the five CMIP5-PMIP3-CESM and not over all CMIP5-PMIP3-CESM models. Therefore, each model has one ensemble mean for the period of 1850-2005. We have revised the paragraph about the method.

17) A more relevant aspect can be that if you are using different models in the ensemble as it is the case, the assumption that a common response to external forcing is filtered out does not hold and by subtracting from each model the ensemble average you do not actually filter out the external response of that specific model. Therefore, this would work for the CESM-LME ensemble but not if you mix in the ensemble different models. This can be a sensitive issue that should be at least clearly discussed here or indicated here and better discussed later in the text.

Refer to our responses in 14) and 16).

18) The two periods, LM and Hist, ... models and regions.
The pre-1850 and post-1850 are merged. However, they do not have the same nature, as in the pre-1850 the natural externally forced variability exists and in the post-1850 it is not intended to exist. Although if calculations have been done for some models out of a multi-model ensemble, some leftovers of externally forced variance should remain. Nevertheless, this should be in principle not important if externally forced signals are not pursued and are not expected to play a role either; actually, for that purpose, control runs could be as adequate or more adequate. I am trying to highlight with the previous sentences that some level of confusion can be transmitted to the readers.

Thanks very much for the point. As we responded in the previous phase, we are aware that the two simulations (pre-1850 and post-1850) do not have the same nature, as they were not run seamlessly, and external forced variability is only subtracted from the historical simulations. The two periods are merged only to calculate continuously the temporal variability of droughts in Fig. 4. But note that for the pattern detection, each period (also each model and each region) was fed separately into the algorithm. We mentioned it in:

[250-251] "The PC-KCA procedure is performed for each model – CESM1, GISS-E2-R, CCSM4, bcc-csm1-1, and MIROC-ESM –, experiment – LM or Hist –, and region – western or eastern Mediterranean – separately."

About control runs, see our response in 9).

19) 2. Section 3.3 Page 6.

I wonder why not considering the soil moisture from ERA5 or ERA5- land additionally or instead to NOAH-LSM. This relates to GC3.3. Since ERA5 produces soil moisture out of its assimilation system, this would be physically consistent with other ERA5 variables used here. Perhaps the authors want to make some comments in the text about the advantage of using the NOAH data here instead.

Refer to our response in 6).

20) 3. Section 3.4 Page 6, lines 161+

- 'When these temporal and spatial ...'

The regional arguments about drought incorporate the condition of having 60% of the horizontal gridpoints with negative SOIL. This faces some difficulties with the fact of using different horizontal resolutions. However, this could be overcome by considering the spatial size of anomalies by using grid box area and the total spatial extension that the anomalies represent, instead of the number of gridpoints.

Thanks for the point. As we responded in the previous phase, we used the number of grid cells as a spatial threshold because the models have different horizontal resolutions, and it would be difficult to set the same threshold based on the spatial extension. We checked that 60% of the total grid cells correspond to about 56.20% (when a drought encompasses only the southern areas) – 61.54% (when a drought encompasses only the northern areas) of spatial extent in the western Mediterranean. For the eastern Mediterranean, the percentages are from 56.31% to 61.65%. These values are close to the 60% threshold that we used with the number of grid cells. We included the approximated values of the areal extent (56%-61%) in the revised manuscript.

[181-183] "At least 60% of all horizontal grid points (which is approximately from 56% to 61% of the areal extent) within the region (the west or east Mediterranean) need to be under negative SOIL conditions during all consecutive drought years."

21) 'Next, the weighted spatial average of SOIL is ... Mediterranean'

How is this weighting done? Perhaps I missed it earlier in the text? If this is done, why not using this regional average to determine the occurrence of spatially large enough droughts? (??). I do not think the result will be very different from the

previous approach and it would overcome the use of a percentage of gridpoints with different model resolutions.

The weighting average was performed by weighting the soil moisture values considering the area of the corresponding grid cell. We included this detail in the revised manuscript.
[198-200] "The time series of SOIL is generated by applying spatial weights to the soil moisture anomalies, taking into account the spatial extent of each grid cell within the confined region."

We added a spatial constraint on droughts by taking into account the percentage of grid points under negative anomalies within the region to make sure that we take droughts that have a considerable spatial extension and are regional events and not local events as explained in **[181-183].** Refer to our response in 21).

We could also have used the spatially weighted time series directly without applying a spatial threshold, but in such cases, a few strong negative anomalies in a few grid points can cause strong negative anomalies, not ensuring a minimum spatial extension of droughts.

22) 4. Section 3.4 Page 7, lines 171+
'Only climate models with some ensemble members ... not detectable'
This is already a report of results. I have no major concern with anticipating it, but what is the reason for it?. The sentences are descriptive of the fact that pan Mediterranean droughts happen in some simulations with an specific feature and not in others. Can you at this stage argue about this feature?, or should this moved further down in the text and a discussion provided?

We removed the sentence as we mainly focus on analyzing east and west regions and not on the pan-Mediterranean events (as we included in the former version, they are rarely detected in the last millennium when the increase in GHG is not present)

23) 5. Section 3.5 Pages 7-9
I get the structure of the methodology in general, however, I suggest the authors revise this section for a more clear explanation, revising notation and perhaps the current state of the explanation of details of the methods that can be more important for the understanding of this section. I will not go to details in an exhaustive way, just provide some examples. The text is providing the sequence of a method. How the outcome of each of the steps feeds the following steps should be clear in explanation and, I would suggest, a homogeneous notation. I also recommend there is a rationale/justification for the conceptual use of each step. I will briefly try to highlight this with examples on the following, but please, go beyond those.

Thanks for the comment. We revised and modified the entire section 3.5, and corrected the notations.

24) References: the authors provide reference but please take care they are appropriate. For instance, there are many ways of applying PCA and the reference to the correct texts that describe the approach presented herein should be provided (see below).

We corrected the reference. Also refer to our response in 23).

25) Use of maps and series: This should be clear from the text. For instance, in the PCA, I understand T(t) are the principal components. Please, indicate that and also the range of the parameter t. The same with s, it is good to indicate the range so that the number of modes or the retained number of modes is well defined.

We modified the text as
[216-218] "PCA decomposes a spatio-temporal field *X(t,l)* using spatial functions *ui(l)*, where *l* is the spatial dimensions (*latitude x longitude* in our study), and their associated temporal functions *Ti(t),* where *t* is the time steps in years (*t* is the total drought years)"

Also, refer to our response in 23).

26) In equation (1), where are the eigenvalues?. I understand they are multiplying either the spatial or the temporal component. They should be indicated. I understand it is particularly relevant if they multiply the temporal component. The reason for it is that the temporal component will have a standard deviation 1 or different from it and this will impact the Kmeans procedure as it will affect the distance.
Explained variances are mentioned but it is not said that the 70% reflect accumulated variance accounted for by a number of pcs. Some sentences are confusing. E.g. line 191 'PCA is applied to the Z500 fields during droughts for each model'. I do not understand this.

We did not include the information on eigenvalues as we found it redundant. PCA is applied prior to a k-clustering analysis (KCA), with the objective of enhancing the effectiveness of the clustering technique. But we included a reference (Hannachi et al., 2007), which explains in detail the usage of PCA in atmospheric and climate sciences. We went through and modified the entire section 3.5. Refer to our response in 23).

27) How do the PCs enter the following analysis? They have unit standard deviation (perhaps not), and their inter-pc correlation is 0. How does this play a role in the Kmeans clustering, what does it mean physically because pcs should indicate different modes in time that are afterwards grouped, even if they are uncorrelated.

Refer to our response in 23).
We used the PC fields with the dimension of time *t* x truncated spatial *N* for the k-mean clustering analysis. We explained the transition from PCA to KCA in:
[226-228] " The resulting new PC field of Z500 has a spatio-temporal dimension of *N x t,* with *t* being the total droughts.
In Step 3, KCA is applied to the *N x t* PC field to detect similar patterns among *t* drought years and group them together."

28) The notation of how the T(t) go into equation (2) should be consistent with this equation. I suggest that the notation is blended for the various steps of the analysis. If it is not done, readers will have to accommodate how things fit from their knowledge and from the different steps of the method. If notation is blended, this section would

actually describe one thing, the approach followed in the manuscript, not several independent methods. Some features of notation are repeated for different things, e.g., s as parameter for points in space and the Silhouette coefficient for each point.

Refer to our response in 23).

29) There are 71 clusters. I learned this from the text but realized that this is the sum of all clusters from different models in Table 2. For each model 5 to 6 pcs are retained and from this, between (mostly) 3 and 6 clusters are formed. What does this mean? Are the 3 clusters gathering the information of the 5-6 pcs? In which way? Some rationale/explanation for what is conceptually happening is good for the reader.

Refer to our response in 23).

30) Line 220: at this level I do not know what the correlation between clusters means. All in all, section 3 needs, in my opinion, to be well revised do deliver a more clear and consistent text. This does not necessarily imply changes in the calculations, nor the results of the following sections. But it may impact (positively) the interpretation.

Refer to our response in 23).

GC5.
4. Results
1. Related to previous comments:
4.1 Observation-model comparison

31) Line 237: '... by subtracting from each of the ensemble members the anomalies at each grid point'. This relates to previous comments and could be explained better.

Refer to our response in 14).

32) Line 238: 'The spatial correlations between SOIL and Z500 of NOAH-LSM and each of the climate simulations are presented in Fig. 3'.
I do not think it is wrong at all but I wonder what is the gain of using NOAH instead of the soil moisture from ERA5 in this figure.

Refer to our response in 6).

33) 2. Figure 2 and related
2b and 2c are good in showing the impact of using different references. However, I would say that the two are not really needed. With one of them it would be enough to explain it. Perhaps the space could be saved to accommodate a time series of the full 850-2005 period. I leave it for the authors to decide.
The shading in Fig 2b seems to indicate consistency in the range of variability with the range of 'observed' NOAH-LSM variability, which is good to indicate.

Thanks very much for the suggestion. We extended Fig. 2 by including the time series for the entire 850-2005 CE, we added more details about these time series, and the NOAH-LSM – CMIP5 comparison in the revised version.

[291-295] "The ensemble means of standardized SOIL during 850--1849 CE (Fig. 2b) shows no apparent monotonic trend during LM. In general, the ensemble means of each model exhibits decreases in SOIL since 1850 CE. The decreases are noticeable even considering the ensemble spreads of each model, except for MIROC-ESM. In MIROC-ESM, a decline of SOIL is observed at the beginning of Hist, but then SOIL increases around 1950 CE. This trend of SOIL in MIROC-ESM is remarkably different from other models and may emphasize model-dependent response to forcing drivers, i.e., GHG forcing."

34) 3. Figure 3 and related
Watch some statements like ' ... negative correlations over southern Europe, but the correlations outside Europe are not significant'. They are for CCSM4 and bcc in the western low latitudes of the domain.

We corrected the sentences based to:
[314-315] "The rest of the models also show negative correlations over southern Europe. For GISS-E2-R and MIROC-ESM, the correlations are mostly significant in the European domain but not outside of the continent."

35) Line 255, '... all models present similarities to the NOAH-LSM, fed...'
What we see in Fig. 3 is the result of both NOAH-LSM and ERA5, right?.

We corrected it to NOAH-LSM-ERA5 in the text and in Fig. 3.

36) Actually, it is likely that the large-scale structure we see there is more dependent on the global model; one could actually test if it changes with other reanalysis products. I think this is likely out of the scope of the study, but it is not totally off the line of argumentation because this correlation field with ERA5 is what we consider 'truth', but it could change if we would have used a different reanalysis product.

Thanks for the point. As we already responded in the previous phase, we did not find much difference between different reanalysis products as the present-day reanalyses are assimilated with similar observation-based data.

37) Perhaps a more relevant issue: if you consider the variability among patterns in Figure 3, what is the variability among simulations of one single model ensemble?. I would assume that it is smaller, but it may be worth reporting.

As we already responded in the previous phase, there could be some variability between the ensemble members, although we do not expect that the centers of correlation would differ much among the ensemble members. Refer to the response in 38) (below).

38) In the actual figure 3, for those models with ensembles, is the pattern that is shown the result of one single experiment? If so, I think it should be clearly stated in

the caption and in the text. I would advise against including correlations using ensemble averages, but I don't think this is what is being shown.

Yes, the correlation fields in Fig. 3 are from the first member of each model. We included this information in the caption for Fig. 3, in the revised manuscript:
[Figure 3.] " For the climate models – CESM1, GISS-E2-R, CCSM4, bcc-csm1-1, and MIROC-ESM – the first ensemble member is used for the calculation."

As we already responded in the previous phase, the correlation analysis using the ensemble smooths out the fluctuations, however, the locations and signs of patterns do not significantly change.

39) 4.2 Mediterranean drought …
1. Figure 4 and related
In relationship to the role of external forcing, or the lack of it ('This fact emphasizes that external forcing signals do not play a role in droughts over the Mediterranean ...' line 286), there are several issues that may be worth commenting.
- ***How are droughts calculated over sub-ensembles of simulations (GISS, CESM1) in Fig. 4, and how is the ensemble spread provided for them? This should be explained (sorry if I missed it) in the text and figure caption.***

Drought years are calculated for each ensemble member using the definition in Section 3.4. Fig. 4 shows the number of drought years in a moving window of a century, and this counting is performed for each ensemble member. The ensemble spread is a unit standard deviation of the time series of the number of droughts across all ensemble members. We included this detail in the caption for Fig 4.

[Figure 4] "Number of drought years in a moving window of a century in the western (red) and eastern Mediterranean (blue). Thick lines for those models and periods that have more than one ensemble member correspond to the ensemble mean, and color-shading indicates the ensemble spread represented by a unit standard deviation."

40) Also, I expect it will justify the different temporal resolution of the curves in Fig 4 for GISS and CESM1 in comparison to the others. However, it can be misleading as the results of those two models may be read as if soil moisture for CESM1 would be consistently higher (smaller) in the western (eastern) Mediterranean during the late 17th and 18th century, or in GISS also during the 18th and early 19th centuries... or in GISS opposite to that during the 17th century. This would not be possible and would contradict the first statement of no role in external forcing as it is very unlikely that different model runs of a sub-ensemble with different initial conditions will coincide systematically in simulating relatively dry or wet periods unless forcing would play a major role. The only reason for that would be external forcing. Also, the finding of opposite phasing between west and east is interesting, and I would argue that it should be more perceptible for GISS and CESM1 if individual runs are considered, in relation to the statement.

As we responded in the previous phase, we did not find a synchronous temporal response between the east and west across all ensemble members of the same model. Hence, the synchronous temporal response observed in Fig. 4 in the manuscript mainly comes from averaging all ensemble members. Here, we add again the same figure that we included in our response during the initial phase, which compares the time series of the frequencies of droughts from some ensemble members in CESM1 and GISS-E2-R between the west and east regions.

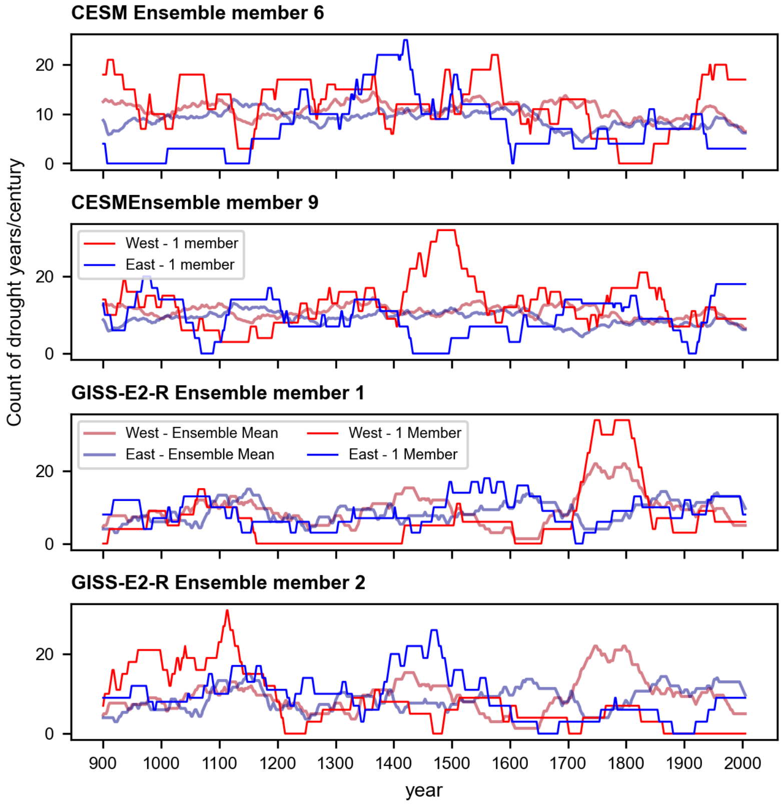

Fig 1. Number of drought years in a moving window of a century in the western (red) and eastern Mediterranean (blue), for some members of CESM and GISS-E2-R.

We included some more description about this finding in the revised manuscript in:
[354-357] " Increased drought events in the west during mid-1500~CE are only shown in CCSM4. The same is also observed across the ensemble members of the same model. The ensemble members of CESM1 and GISS-E2-R do not exhibit unanimous periods of low or high drought occurrence (figure not shown)."

41) '...observed, more clearly in those models and periods with one ensemble member' (line 296). Therefore, I suggest the authors revise how the ensemble behavior is presented for those two models. For instance, the last statement of Section 4.2 is sensitive: 'For those models and periods with more ensemble members... sometimes this association is blurred...', it should indeed be the expected behavior, even more than what is shown. If you resort to individual simulations, this should be more clearly evidenced. The ensemble spread should expectedly blur everything since a dry or wet century in one run should not be expected to be consistently dry or wet in most of the other sub- ensemble runs.

Refer to our response in 40).

42) 2. Figure 5 and related
I am not against showing temperature anomalies in association to the Z500 patterns, but why not showing precipitation and actual drought patterns. How do the geopotential anomalies account for drought occurrence?

Thanks very much for the suggestion. As we do not analyze in detail the effect of temperature or precipitation on droughts, we decided to leave only the Z500 patterns in Figure 5.

43) Figure 6: I need a better description of the methods section to better interpret results and figure out whether they can be dependent on methodological choices.

Refer to our responses for comments 23 to 30.

Minor comments:

MC1. Introduction, page 1 line 19. ' Ocean, teleconnections and large-scale modes of variabity.' Ocean, and teleconnections with other large-scale modes of variability
We corrected it to "high in the Atlantic Ocean and other teleconnection patterns" in **[29]** .

MC2. Page 2, line 40. '...change in climate boundary conditions...'
'... change in external forcing...' seems to me better in this context. The current sentence is maybe a bias from climate modelling,
Modified as suggested in **[51]**.

MC3. Page 2, line 42. 'Several natural proxy-based ...' Natural meaning?
We modified it to "proxy-based reconstruction" in **[53] .**

MC4. Page 2, line 44. '...summer dry and wetness...'. Dryness?
Modified as suggested in **[55].**

MC5. Page 2, line 42. '...of droughts variability...' drought variability
[Modified as suggested in **[53].**

MC6. Page 5, line 111. '...at vertical soil layers...' You mean perhaps, vertically integrated soil moisture content?, or something of the sort…
We modified it to " soil moisture content of vertical soil layers" in **[130]**.

MC7. Page 7, line 172. '...show few numbers of...'
Cases of?
We removed the sentence in the revised manuscript.

Response to Reviewer 2

Main comments

1) 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).

Thanks for the suggestions. We extended Fig. 2 by including the time series for the entire 850-2005 CE period and added more discussion about it **[291-295]**. We also extended the discussion on Fig. 6 (Fig. 5 in the old manuscript), which is related to Fig 7 (Fig. 6), focusing on the dominant frequencies of the detected patterns in **[424-429].**

[291-295] "The ensemble means of standardized SOIL during 850–1849 CE (Fig. 2b) shows no apparent monotonic trend during LM. In general, the ensemble means of each model exhibits decreases in SOIL since 1850 CE. The decreases are noticeable even considering the ensemble spreads of each model, except for MIROC-ESM. In MIROC-ESM, a decline of SOIL is observed at the beginning of Hist, but then SOIL increases around 1950 CE. This trend of SOIL in MIROC-ESM is remarkably different from other models and may emphasize model-dependent response to forcing drivers, i.e., GHG forcing."

[424-429] "The frequencies of occurrences of patterns presented in the lower panels of Fig. 6 indicate that the occurrence is not uniform over time, with certain patterns appearing only during specific periods. Nevertheless, the dominant frequency of each circulation pattern is estimated using a power spectra analysis (i.e., Cook et al., 2016), and the result reveals that the values noticeably vary across the models and ensemble members (Fig. A2). In general, the most dominant frequencies span a wide range of the multi-decadal (10-100 years) time scales, which partially seems to explain the time scales of the co-variability of SOIL and drought periods in Fig. 5."

2) Clearly, some models are not very skillful at capturing the synoptic climate during historical droughts and this raises in itself an important issue: 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).

As we responded in the previous phase, 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 in depicting these patterns. For this, we compare the patterns and their frequencies (Figs 6 and 7). However, we do not particularly measure which model is the best or worst, as there is no universal objective criterion for it, and 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).

Still, we assume that if the models are able to represent well the present-day drought-related circulation in an acceptable way (shown in the correlation patterns in Fig. 3), the representation of the modes would still be valid in longer simulations. Therefore, we performed correlation tests between the SOIL-Z500 correlation fields in NOAH-LSM-ERA5 and other models in Fig. 3. to compare the spatial patterns of soil-related circulation. The coefficients are included in Fig.3 and discussed in **[314-325]**.

[314-325] "The numbers on the top of the panels indicate the correlation between the correlation patterns of NOAH-LSM–ERA5 and each of the climate models presented in Fig. 3. The correlation analysis is performed on the entire regions in Fig. 3, which means that these values measure the overall closeness of the entire correlation fields to NOAH-LSM–ERA5 without considering the statistical significance of individual grid locations. We use this quantity to compare the model's representation of SOIL-related Z500 patterns between the two regions, the western and eastern Mediterranean. Overall, the correlation coefficients are higher in the western region than in the eastern Mediterranean. In the western region, the maximum coefficient is 0.82, shown by bcc-csm1-1, while the maximum in the eastern region is 0.62, also by the same model. The minimum value in the western region is 0.42 by GISS-E2-R, and in the eastern region, it is 0.14 by CCSM4. The overall comparison of spatial correlation coefficients implies that the variability of Z500 associated with SOIL in the climate models is better represented over the western region than over the eastern region."

For temporal variability, we assess whether the ranges and ensemble spreads of soil moisture anomalies from the models encompass the variability of the present-day observation-based soil moisture anomalies (NOAH-LSM), shown in Fig. 2a, and discussed in **[282-290].**

[282-290] " During 1950–2005 CE (Fig. 2a), the variability of SOIL from NOAH-LSM (the range of maximum and minimum over 1950–2005 CE is 18.29 mm/month) is within the range of variability of the CMIP5-PMIP3 model SOIL values (the range of maximum and minimum across the models of 22.75 mm/month). This observation indicates that the overall magnitudes of variability between the observation-based data and the models are comparable. However, the standard deviation (σ) of SOIL across the four models over the entire time (3.06 mm/month) is lesser than the σ of SOIL of NOAH-LSM (4.37 mm/month). This distinct σ indicates that there is some degree of discrepancies in SOIL variability among the models. GISS-E2-R, CESM1, and CCSM4 show σ of SOIL of 7.73, 5.44, and 4.43 mm/month, respectively, which are higher than that of NOAH-LSM, while bcc-csm1-1 and MIROC-ESM presents σ of 2.03, and 2.46 mm/month, respectively"

3) 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 that arise from the interaction of climate systems without the influence of external or internal forcings (volcanic eruptions). We rephrased it to "internal climate dynamics" when we want to distinguish internal variability (without eruptions) from external climate signals. For instance,

[352-360] "This large variability in drought occurrence across the models and the ensemble members, depicted by the ensemble spreads in Figs 4 and 2b, implies that internal climate dynamics is the primary driver of droughts in the region during LM and Hist with the anthropogenic forcing removed. In these time series, external or volcanic forcing signals are not visible in the variability of droughts."

4) 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 modified the sentence to:
[7-9] "Primarily, we examine the differences among the models in representing drought variability and related circulation patterns. For the analysis, we exclude the anthropogenic trends from 1850–2005 CE."

Also note that the abstract is modified in the revised manuscript, as the results are updated.

5) 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 suggestion. We modified the motivation in the introduction as:
[83-86] " the objective of this study is to examine how different coupled climate models depict the temporal variability of droughts and associated circulation patterns during the last millennium. Additionally, we evaluate the differences among the models in representing drought-related circulations. The focus is on the mid-latitude circulation patterns that have more impacts on the Mediterranean hydroclimate."

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

We corrected the sentence to:
[133-137] "The variables are all transformed into annual mean anomalies. Wet seasons are critical periods for the moisture supply of the region, and therefore, when strong circulation patterns take place (Xoplaki et al., 2004; Lionello et al., (2006). Therefore, we use annual mean anomalies to capture the mean variability of hydroclimate throughout the entire year

instead of focusing on some particular seasons, i.e., summer growing seasons. This is to include the influences of wet seasons, which is the crucial season for moisture in the region, in the analysis."

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 CE, as this trend can affect the overall analysis of the time series (continuous droughts are found after 1850-1900 when the trends is not removed.). Therefore, the GHG effect is not included in the analysis. We added this detail in the revised manuscript:

[140-141] "In the next step, we remove the strong unprecedented trends in the Hist simulations (1850–2005 CE). This means that the effects of increased GHG on droughts is not included in the analysis."

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

We revised the entire section 3.5, and the section is now modified from the previous version.

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

All time series are from the forced (transient) last millennium and historical simulations. We included this detail in the caption of Fig 2.
[Figure 2] "10-year running means (thick lines) and ensemble spreads (color-shaded) of standardized SOIL during 850–2005 CE with respect to 850–1849 CE, from the PMIP3-CESM1 transient LM and Hist simulations."

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.

This question is partially responded in 2). Also, as we responded in the initial response phase, 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 (an ensemble spread given by a unit standard deviation or maximum-minimum value range across the models), we assume that the model is able to represent the temporal variability of the observations. For the same reason, a cross-correlation analysis between NOAH-LSM and CMIP5 would not provide additional useful information. We included discussions about the implication of differences in soil data in:

[459-467] "It needs to be pointed out that MIROC-ESM presents many unique patterns that fail to join other circulation patterns, largely occurring with the eastern Mediterranean patterns. Several reasons can be put forth to explain this particular characteristic of MIROC-ESM. As depicted in Fig. 2, MIROC-ESM presents a different SOIL trend compared to the other models during Hist. This distinct trend could potentially be attributed to different circulation types compared to those of other models.
Another argument to consider is that, in general, SOIL-related circulation fields in the eastern region seem to exhibit lower statistical similarity with the observation-based circulation condition, based on correlation coefficients in 3b. While this is the case for all models, lower correlation coefficients in the eastern regions can be associated with a reduced number of horizontal grid points compared to the western regions. Fewer grid points may not reflect the entire temporal variability of SOIL, and, therefore, the associated circulation variability."

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

Yes, we modified it to NOAH-LSM–ERA5 in the text.

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 in 9).
As we responded in the initial phase, regarding the question of how to deal with models that do not capture well the observed synoptic climate, one of the objectives of our study is to understand the model differences in representing drought-related circulation over the Mediterranean region. Therefore, we do not apply a selection criterion to choose the best or worst model (Also refer to our response to comment 1). Instead, we attempt to address and visualize the differences between the models in representing drought-associated mid-latitude circulation patterns (For instance, in section 4.2). Regarding the skillfulness of the models, for the same reason, it would be hard to assess the skillfulness of all variables associated with droughts, and there is no universal metric for it.

Still, in the revised version, we performed a comparison of SOIL-related circulation (Z500) patterns between the models and NOAH-LSM–ERA5 by calculating the correlation coefficients of the correlation fields in Fig. 3. These coefficients measure the similarity of the correlation fields between the models and NOAH-LSM–ERA5, and we used them mainly to address the differences between the east and west regions. We included these new values in Fig. 3 top left of the panels, the explanation of the methods **[163-167]** , and the discussion of the result **[314-325].** Also, refer to our response in 2).

[163-167] "In addition, to quantify the spatial similarity of the correlation patterns, PCC is calculated between the correlation pattern NOAH-LSM--ERA5, for the western and eastern regions, separately, and the patterns of each climate model. For this spatial comparison of correlation patterns, the spatial resolutions are interpolated to those of the coarser models, which are bcc-csm1-1 and MIROC-ESM (Table 1). This PCC would provide a numerical value to evaluate the overall resemblance between the present-day NOAH-LSM–ERA5 field and the climate models."

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?

For clarification, we modified the part to:
[357-360] "This large variability in drought occurrence across the models and the ensemble members, depicted by the ensemble spreads in Figs. 4 and 2b, implies that internal climate dynamics is the primary driver of droughts in the region during LM and Hist with the anthropogenic forcing removed. In these time series, external or volcanic forcing signals are not visible in the variability of droughts."

About the models' setup and ability, refer to our response in 1), and their skillfulness, to our response in 12).

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

We excluded this word to avoid confusion and rephrased the sentence to:
[359] "Hist with the anthropogenic forcing removed."

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

Thanks for the suggestion. We included wavelet coherence analysis to assess the phase relationship between the two regions and their variations over time, in Fig 5.

[202-204] "We perform wavelet coherence (Grinsted et al., 2004) on the time series of the number of droughts and the soil moisture anomalies between the west and east Mediterranean. The purpose is to assess the phase relationship, dominant frequencies, and temporal variations of correlations in soil moisture anomalies and drought occurrences between these two regions."

The discussion of the result can be found in**:**
[372-390] "To evaluate closely the dominant frequency of the association between the two regions, and their temporal co-variability, the wavelet coherence analysis is performed on the time series of SOIL (Fig. 2b) and presented in Fig. 5a. At first glance, the wavelet coherence analysis suggests that timing and frequencies of co-variability are not the same across the models for both the soil moisture anomalies and the number of droughts. Also, the association is not uniform across the time-frequency space. The analysis performed on SOIL between the western and eastern regions (Fig. 5a) indicates co-variability that ranges from interannual to multi-decadal time scales, depending on the model. CCSM4 shows co-variability of higher frequencies (less than a 32-year period.). In general, the association between the two regions is in-phase. The co-variability between the western and eastern regions from all models is less pronounced and less significant in all time-frequency bands compared to the result presented by Cook et al. (2016) based on OWDA, which has shown significant in-phase co-variability of SOIL between east and west in diverse timescales.

The time series of SOIL does not necessarily indicate that dry periods are in phase. To compare only the dry periods, the wavelet coherence analysis is performed on the number of drought years (Fig. 4), which is presented in Fig. 5b. The co-variability between the two regions is significant in some time scales, for instance, on a multi-decadal time scale (around 32 years and higher) with an anti-phase relationship in CESM1, bcc-csm1-1, and MIROC-ESM. For GISS-E2-R, CCSM4, and MIROC-ESM, the anti-phase association is also significant on a high-frequency band (of less than 32 years). This result seems to agree with the observation from Fig. 4. It is observed that the anti-phase co-variability also depends on the time period, for instance, CCSM4 during 1400-1500 CE, the co-variability is in-phase. Again, the overall result points out the associations that are not uniform across time and that vary across the models. The occurrence of droughts and dry periods can be associated with dominant drought-driving circulation patterns of each region. More details on this are provided in the next sections."

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?

As we responded in the previous phase, 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 (but not temporarily synchronous). Also, OWDA provides only the summer hydroclimate variability of the region and not all annual means as we used.

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 included CE after the years.

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.

We corrected the text and wrote the full name of the patterns:

[151-155] "The separation is motivated by the suggested influences of the circulation patterns over the Mediterranean region (Dunkeloh et al., 2003; Lionello et al., 2006), such as the NAO, East Atlantic (EA), and Eastern Atlantic-West Russian (EA-WR). The western Mediterranean is more intensely influenced by NAO than the eastern region, and the eastern region is not only affected by NAO but also strongly linked with East Atlantic-type patterns."

We used abbreviations after this paragraph.

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

In the discussion section in lines **[459-467]** (refer to our response to the comment in 10), we discuss the representation of circulation patterns, mostly relating it to MIROC-ESM which is the model that shows distinct patterns from others in Fig 3 and Fig 5. We discuss the reasons for this possible difference (which is also valid for other models) based on different horizontal resolutions and soil moisture processes in the models.

[463-467] "Another argument to consider is that, in general, SOIL-related circulation fields in the eastern region seem to exhibit lower statistical similarity with the observation-based circulation condition, based on correlation coefficients in Fig. 3b. While this is the case for all models, lower correlation coefficients in the eastern regions can be associated with a reduced number of horizontal grid points compared to the western regions. Fewer grid points may not reflect the entire temporal variability of SOIL, and, therefore, the associated circulation variability."

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.

Thanks for your comments. In Fig. 7 (Fig 6 in the former version), we try to summarize the frequencies of the patterns in Fig 6 (Fig 5 before), and show the difference in frequencies of these patterns among the models, in which some patterns only appear in one model. We discuss briefly it in:

[505-509] "Our analysis also shows that the contribution of the patterns to droughts may greatly depend on the choice of the model. For example, the importance of EA-WR in the eastern droughts is apparent in CCSM4 and bcc-csm1-1 but not in other models (Fig. 7). The frequencies of a shared pattern between some models also vary greatly among them. This highlights the fact that climate models have their preferred circulation patterns associated with Mediterranean droughts."

The inter-model spread of these patterns in terms of their temporal variability is shown by the time series in Fig 6 (old Fig 5), explained in the caption:

[Figure 6] "Below each map, the time series of the number of occurrences of the corresponding patterns per century for the western (red) and eastern (blue) Mediterranean are plotted, together with the respective ensemble spread of occurrence (shaded)"

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

We refer to the mean frequency during LM and Hist of the circulation patterns. We rephrased the sentence to:
[430-431] "Fig. 7 summarizes the mean frequencies of the circulation patterns during the entire LM and Hist in Fig. 6, represented by the mean number of patterns per century for each mode"

- 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, therefore, the events are of multi-year duration. 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 6). We extended this explanation in the revised manuscript.

[403-406] "Positive NAO is known to be the dominant climate mode that drives a drier condition over the Mediterranean region (Lionello et al., 2006; Kim and Raible, 2021). This explains a large percentage of occurrence of P1 and P3 (42%) during droughts. Although it seems contradictory that P2 depicting a negative NAO condition also occupies a significant percentage of the occurrence (20%), the occurrence of P2 indicates the fluctuation of NAO patterns throughout multi-year droughts."

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.

The periodicity (mainly of co-variability between the two regions) from wavelet coherence analysis is provided in Fig 5. and lines **[372-390]**. The result shows no uniform and stable temporal frequency of co-variability. This may support the ""absence of internal multidecadal oscillation in climate models" of Mann et al., (2020).
Also, refer to our response in 15).

Additionally, we performed power spectra density analysis to find frequencies of the patterns in Fig 6, but we did not find consistent frequencies among the models (Fig. A2). Nevertheless, the patterns have frequencies of occurrence on multi-decadal time scales that are consistent with Fig. 4. We included the discussion in:

[424-429] "The frequencies of occurrences of patterns presented in the lower panels of Fig. 6 indicate that the occurrence is not uniform over time, with certain patterns appearing only during specific periods. Nevertheless, the dominant frequency of each circulation pattern is estimated using a power spectra analysis (i.e., Cook et al., 2016), and the result reveals that the values noticeably vary across the models and ensemble members (Fig. A2). In general, the most dominant frequencies span a wide range of the multi-decadal (10-100 years) time scales, which partially seems to explain the time scales of the co-variability of SOIL and drought periods in Fig. 5."

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

Refer to our response in 16).

As we responded before, we did not include the time series of OWDA, as our focus is on the representation of drought variability and associated mid-latitude patterns in climate models. OWDA is the reconstructed time series of only summer (growing season) hydroclimate. We do not think we can see similar temporal and spatial patterns to OWDA.

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

As we responded in the previous phase, the N-S antiphase occurs between northern Europe and southern Europe (Markonis et al., 2018). 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)?

Refer to our response in 15).

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).

As we responded during the initial phase, 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 **[64-66].** 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 changed it in the revised version.
[338-339] "a change that is mostly attributed to anthropogenic effects on global and regional climate (Douville et al., 2021; Seneviratne et al., 2021)."

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. Our focus is on soil moisture droughts, therefore taking the soil moisture anomalies from the subsurface layers to quantify droughts. We clarified it in the revised manuscript.

[100-101] "Our focus is on droughts that affect deep soil moisture conditions, known as soil moisture droughts (Dai, 2011)."

[170-171] " The 70 cm level is a deep soil level that can reflect the impacts of soil moisture change on vegetation and ecosystems, hence, better-representing soil moisture droughts including their persistent characteristics."

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.

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.