Response to Reviewer 1

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.

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 for the point. We will correct the paragraph and include more details about the extended seasonality character of 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 will include more details about this particular article (Xoplaki et al., 2018) in the introduction.

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 will incorporate the reviewer's suggestion in the next phase.

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 point. We agree with the reviewer's comment regarding the need for an additional description of soil moisture for each model and a discussion on inter-model differences in our study. We will include this detail in the revised manuscript.

One point that we want to comment on is that 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. However, vertical soil layers (up to two-meter depth) could be a more important factor contributing to the observed model differences. We briefly discussed this issue in the discussion section relating to the discrepancy between MIROC-ESM and other models (lines 366–373). We will extend this discussion.

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.

As we responded to comment 4, we will include more details on soil moisture and soil components of the models in the revised manuscript.

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.

We used the soil moisture from NOAH-LSM instead of ERA5 because NOAH-LSM is forced with the observation-based dataset and the reanalysis data that the biases were corrected with respect to the observations (https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/GLDAS_NOAH025_3H.2.1/doc/README_GLDAS2.pdf). ERA5 does not directly assimilate any rain gauge data except for the United States (Lavers et al., 2018). Therefore, we assume that NOAH-LSM could be more appropriate to show more realistic present-day soil moisture variability, which is mainly influenced by precipitation variability. We will add this detail to the manuscript.

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...

We will include this detail in the revised version (see our responses to comments 4 and 5.).

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. Hence, this member was not taken for the analysis. We will include this explanation in the new version of the manuscript.

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.

In our research, we focus on investigating the natural variability of drought and extratropical circulation associated with the events during the last millennium. Hence, the effect of the post-1850 increase in GHG is excluded, but not the volcanic forcing as is a natural internal forcing. We do not assume from the beginning that internal variability is the main driver of drought (although our result, in the end, shows this point). We are aware that our text is misleading on that point, therefore, we will correct the manuscript for clarification.
As we mentioned already, we use the last millennium simulations instead of long control runs as we examine drought variability during the last millennium.

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 the soil moisture at 70 cm level was used. The sentence was to describe the soil variable of NOAH-LSM. We will include the same information for other climate models in the revised manuscript.

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.

Here we wanted to emphasize that we use the annual mean time series to include hydroclimate conditions of all seasons, including the wet seasons (referred to as only winter in our study, but we will correct this according to the reviewer's first comment.), unlike other studies focusing on the region that usually consider only the summer mean time series. We will clarify our point better in the revised manuscript.

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 will include this change.

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, therefore, to avoid redundancy, we will remove the sentences about the annual cycle in the revised version and move other sentences in the middle to the next paragraph.

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.

It would not affect the volcanic events during the pre-industrial period (850–1849) as the ensemble averages were only subtracted from the historical simulations. We will clarify the paragraph in the revised manuscript.

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

The filtering is applied only to the post-1850 period mainly to remove the anthropogenic GHG forcing which becomes apparent after 1850. We did not apply the filtering before 1850, as there is no apparent long-term up- or downward trend in soil moisture before 1850. In addition, except for CESM and GISS, other models have only one run during the last millennium which inhibits applying the method to all simulations. We will try to clarify this paragraph better for the next phase.

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.

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 (These means are shown in Fig 2.c). Then, the ensemble mean is subtracted from each of the historical simulations of the corresponding model. We may not have been clear on this point, therefore, will try to clarify this paragraph better.

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.

See our response to comment 16. The ensemble mean is calculated for each of the models and later, subtracted from each of the historical simulations of the corresponding model. Therefore, we filtered out the external response of that specific model, which takes into account the model's climate sensitivity.

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. 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 the 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 was fed separately to the algorithm. We will clarify this point better in the revised manuscript. About control runs, see our response for comment 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.

See our response to comment 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. Indeed, 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 the 60% of the total grid cells correspond to about 56.20% (if covering only the southern grid areas) – 61.54% (if covering only the northern grid areas) of spatial extent in the western Mediterranean, and 56.31% – 61.65% in the eastern Mediterranean (see the table below). These values are close to the 60% level that we use with the number of grid cells.

The maximum difference in the spatial coverage between the models is around 4%, which for all models except CCSM4, is the spatial extension of one grid cell. Therefore, using a threshold proposed by the reviewer that is based on the spatial extent may exclude a few drought events in our analysis, but it would not change our results much. We will add more details on our choice of threshold in the revised manuscript.

In addition, we noticed that we made a mistake in the number of grid cells for CESM and GISS-E2-R in Table 1 in the manuscript. We will also correct it.

Table 1. The number of 60% of land grid cells and the corresponding spatial coverage (spatial extension divided by the total area of the region) for the southmost (minimum) and northmost (maximum) regions.

	% of spatial coverage		
	Minimum coverage (southern area)	**Maximum coverage (northern area)**	**Difference between the minimum and maximum (%)**
WEST			
CESM	58.49	61.54	3.05
GISS-E2-R	58.01	61.05	3.04
CCSM4	57.81	61.59	3.78
bcc-csm1-1	56.20	59.58	3.38
MIROC-ESM	57.56	60.71	3.15
EAST			
CESM	56.31	58.97	2.66
GISS-E2-R	56.89	59.61	2.72
CCSM4	58.15	60.98	2.83
bcc-csm1-1	58.20	61.65	3.45
MIROC-ESM	57.13	59.66	2.53

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 will include this detail in the revised manuscript.
We added a spatial constraint on droughts by taking into account the percentage of grid points over the region to make sure that we take droughts that have a considerable spatial extension and are regional events and not local events (lines 159-160).
We could also have used directly the spatially weighted time series and taken a certain threshold for droughts, but in this case, a few strong negative anomalies in a few grid points can cause strong negative anomalies, not ensuring any 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 will move the sentence to the result section.

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 will adapt the notation so that everything is consistent throughout the entire section.

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).

Thanks for the comment. We will go through the section to clarify the text.

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 will correct these issues in the revised manuscript.

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 will include these details in the paragraph and correct the sentences for clarification.

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.

The principal component analysis (PCA) is mainly applied to reduce the spatial dimension of the dataset and increase the performance of the clustering method. With the PCA, the new field of PC(t)s are obtained from the whole spatio-temporal (*time x latitude x longitude*) Z500 with *t* drought years. The first *N* PCs, *N* in our study ranging between 5 and 6, are taken depending on the silhouette coefficients and explained PC variances. The dimension of space of PC is now reduced to *t x N*, instead of *t x latitude x longitude* of the original dataset. The K-mean clustering is applied to group similar Z500 in this *t x 5* PC space, then assigning the label to Z500 patterns that belong to the same cluster.
It is true that the PCs are uncorrelated, but the clustering method considers the geometrical distance between the points (euclidean or Mahalanobis), and not the covariance between them. It is common to apply PCA before grouping the data for a better performance. We will revise the section for clarification.

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.

We will correct these issues in the revised manuscript.

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.

Also, see our response to 27. The K-mean clustering groups the circulations (Z500) of drought years. After grouping the clusters as explained in 27, the mean values of each cluster are calculated using the Z500 anomalies that correspond to the drought years of the cluster (instead of using the normalized projected values of the PCs). Hence the clusters are able to be correlated based on their Z500 anomalies. We notice we did not include this detail in the manuscript. As we responded in comments 24 – 28, we will revise the section, and if we find it necessary, also we will add this detail to the diagram (Fig 1) for a better understanding.

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.

See our response to 24.

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.

See our response to comment 16.

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.

See our responses to comment 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.

We will try to accommodate the full 850-2005 in the Fig 2 area. Also as reviewer 2 commented that Fig 2 is under-utilized, we will add more details about these time series and the NOAH-LSM – CMIP5 comparison in the revised version.

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 will correct it in the revised manuscript.

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?.

Yes, that is correct. We will correct the sentence.

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.

We do not expect to see much difference between different reanalysis products as the present-day reanalyses are assimilated with similar observational-based data (although with different models). Here we provide the maps of correlation between the Z500 from the NCEP/NCAR reanalysis 1 (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html; Kalnay et al., 1996) and the NOAH-LSM soil moisture anomalies (Fig 1.b and d), along with the correlations between the ERA5 Z500 and the NOAH-LSM soil moisture (Fig. 1.a and c) which is in Fig. 3 of the manuscript.

The two datasets basically show similar structures. A slight difference is because of the horizontal spatial resolution (NCEP-NCAR has a coarser resolution of 2.5 x 2.5 degrees compared to ERA5 of 0.25 x 0.25 degrees).

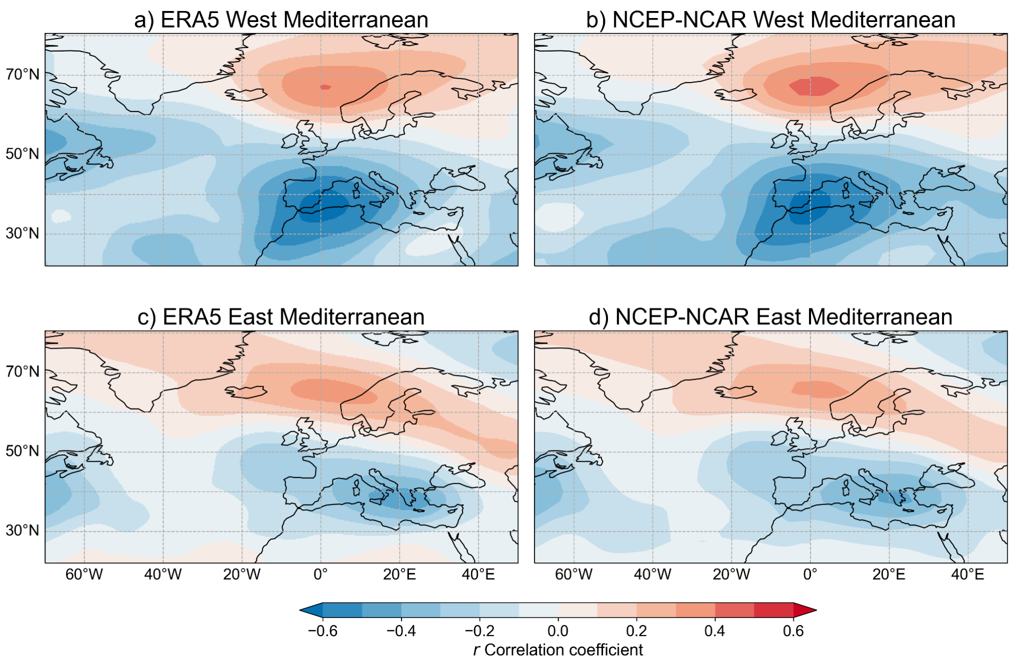

Fig 1. Correlation coefficients between the ERA5 Z500 and the NOAH-LSM (a) west, (c) east Mediterranean soil moisture anomalies, and between the NCEP-NCAR reanalysis and the NOAH-LSM (d) west, (d) east Mediterranean soil moisture anomalies.

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.

There could be some variability between the ensemble members, although we do not expect the centers of correlation would differ much between them. See the response to comment 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. As the correlation analysis using the ensemble means may smooth out the fluctuations, we provide here the mean correlation maps of three ensemble members of CCSM4 (Fig 2 below). Compared to the correlation fields of only one member (what is shown in Fig 3 in the manuscript), Fig 2 (below) shows more smoothed correlation coefficients. However, the locations and signs of patterns do not significantly change.

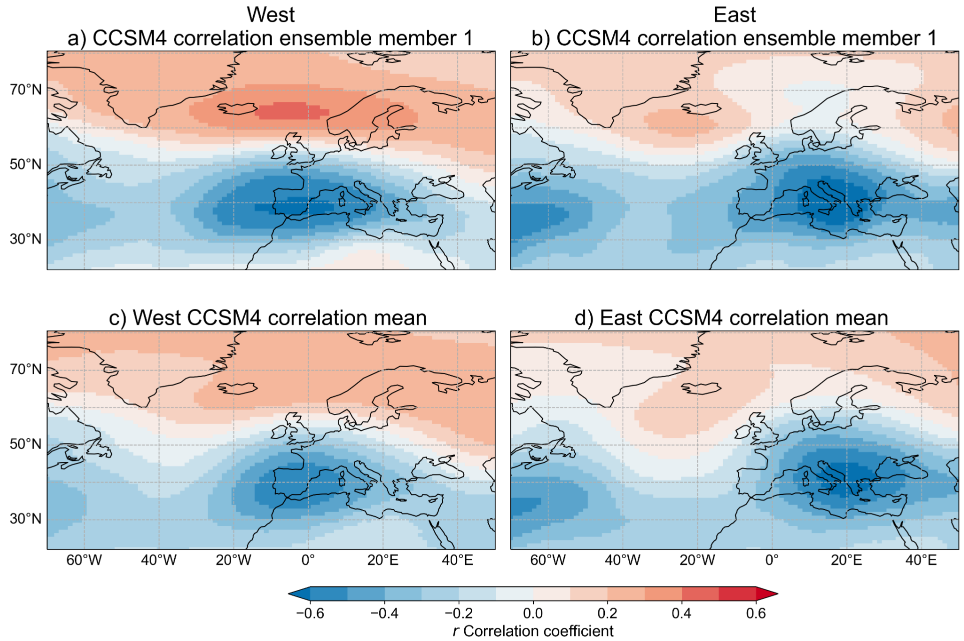

Fig 2. Correlation coefficients between Z500 and soil moisture anomalies in CCSM4 for (a) and (b) one ensemble member, and (c) and (d) the means of three ensemble members over west and east.

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 running count of droughts every 100 years which is calculated for each member, and the ensemble spread is a unit standard deviation of the time series of counts of all members. We did not include this detail, therefore, we will add it in the revised version.

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.

We guess this detail can be better shown when we provide an analysis of the time series for each ensemble member. Below, we include the time series of some ensemble members in CESM and GISS plotted separately in Fig 3.

Notice that in the figure, we do not see anymore a synchronous temporal pattern between the east and west or between the two members of the same model in the same region. Hence, the synchronous temporal pattern observed in Fig 4 in the manuscript mainly comes from averaging all ensemble members. We do not think we can add all the plots of individual ensemble members of all models, but we will include more discussion on this inter-member difference in the revised manuscript.

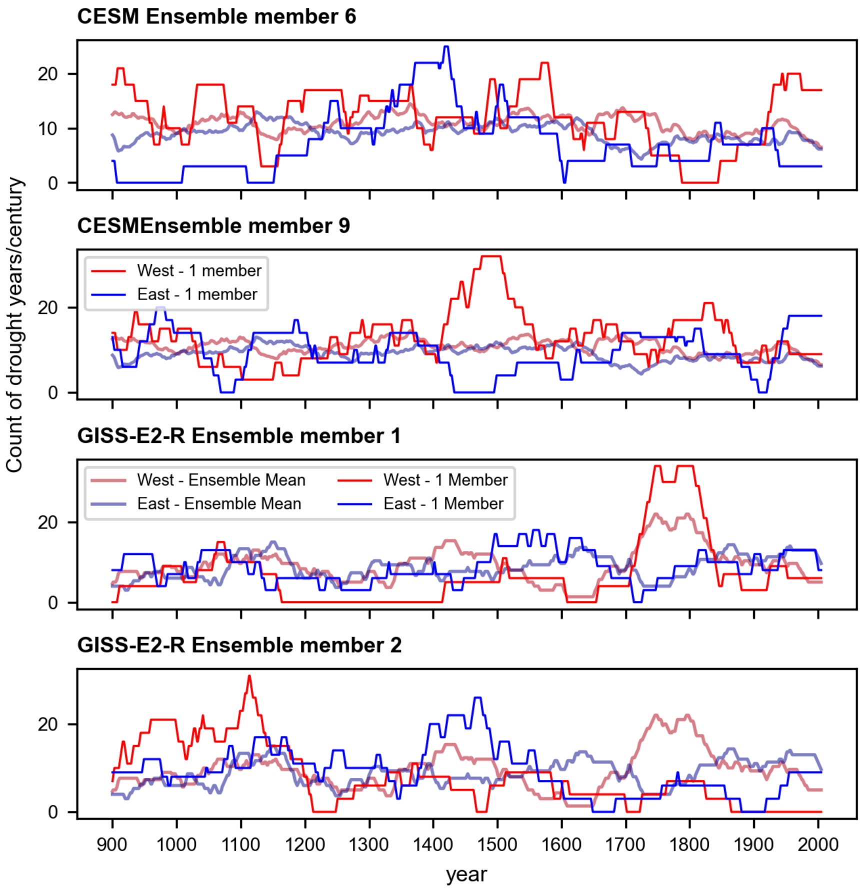

Fig 3. Occurrence of drought years in a moving window of a century in the western (red) and eastern Mediterranean (blue), for some members of CESM and GISS-E2-R.

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.

See also our response to comment 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?

We included the temperature anomalies as the variable can increase the intensity of the events under the presence of high-pressure systems (Zhou et al., 2019). We will include more details on the effects of temperature and geopotential height anomalies on droughts in the revised version.

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.

See our response to comments 23 to 30.

We will correct the minor comments in the revised manuscript.

Reference

Kalnay et al. (1996), The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470.

Lavers, D. A., Simmons, A., Vamborg, F., & Rodwell, M. J. (2022). An evaluation of ERA5 precipitation for climate monitoring. Quarterly Journal of the Royal Meteorological Society, 148(748), 3152-3165.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., ... & Gentine, P. (2019). Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. Proceedings of the National Academy of Sciences, 116(38), 18848-18853.