2nd Review of :
egusphere-2023-119
Extratropical circulation associated with Mediterranean droughts during the Last Millennium in CMIP5 simulations
W. M. Kim, S. J. González-Rojí and C. Raible

Evaluation:

The authors have included responses and changes to the comments. I am satisfied with most of them. I am not convinced about a couple of them that I mark with (*). I do suggest improving the description of the methodology. Additional to that there are several points which I think should be considered. I regard them as minor and easy to do, so I leave it to the Editor whether I should see the responses and the corrected manuscript again.

GC1. Introduction:
I think the introduction has improved with the changes implemented. A minor comment:

1. Page 2, line 51. '… no substantial change in external forcing…'
This statement can be missleading. I understand in the context of what the authors want to convey but I suggest rephrasing. Forcing changes are substantially smaller than say the glacial-interglacial transition, but still detectable in reconstructions and substantial enough to be climatically relevant. Consider rephrasing.
I think the relevant issue here is that the LM is the period in which the (natural) forcings are most comparable to nowadays and the one, if we think preindustrial, that offers the reference before the development of the large anthropogenic forcing.

2. As a reference for drought variability at global scales during the last millennium and potential responses to internal variability and external forcing consider https://cp.copernicus.org/articles/16/1285/2020/

GC2. Section 2: Data

1. Page 4, line 99: (LM; 850-1850)
See consistency with pate 5 line 140: 850-1849

Page 4, line 109: '…are transient ~~that~~ and were run…'

GC3. Section 3: Methods

1. Line 142: '… the effects of GHG on drought ~~is~~ are not…'

Line 167: '… present-day NOAH-LSM-ERA5 correlation field and that obtained from the climate models …' (?)

Line 175+ Drought definition. It is not clear if the drought is composed only of negative SOIL values (as it is literally stated) or if it contains the dates of the two positive values at the end ('… and continue until two consecutive years of positive anomalies…'
'.. of only negative values without being interrupted by a particularly wet year in between…' However, I understand from the definition that if there is a positive SOIL anomaly, the drought continues until two positive anomalies are found' – I am only trying to point out that I find the explanation confusing in this respect

Line 183 '… a substantial portion of the region experiences drought conditions …'

Line 186. 'Nevertheless, this approach avoids changes in the initial SOIL values, as …' I do not understand what is meant here.

Line 198+ 'Also, 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'.
This is not clear to me and I think it should be explained, what is done, how is it taken into account. It is pointed out that it is somehow taken into consideration, but I suggest to indicate how.

2. (*) Line 214+, PCA calculation
   - Line 222: '…PCs, represented as …Ti(t) * ui(l)
     In my understanding this is not correct. Please, indicate which variable are the PCs and which the eigenvectors, also indicate where the eigenvalues are included in eq (1).
     I think this part of the methods is not satisfactory. The explanation should state how the covariance (or correlation?) that is diagonalized is defined (in space or time) and then which ones are PCs, eigenvectors, EOFs, eigenvalues, etc, in a clear way. I think this influences the interpretation of the results and helps reproducibility.
   - Why 70%? Likely arbitrary but I would expect the text to say that the results are not very sensitive to this decision. Why should N be < or = to 7?...
   - I understand that the dates corresponding to the identified droughts are selected and the Z500 is considered only for those dates and then the PCA is applied. Perhaps I have skipped this explanation somewhere, but I have not found it. It should be clearly explained.
   - Do the xi in Eq (2) refer to the X(t,l) in eq (1)? This does not make sense to me. Can you refer the notation of Eq (2) to that of Eq (1) so that the reader can understand how the output of the PCA feeds the KCA? In my understanding these should be PC values because you end with N modes X t drought years, with (I assume) a PC being a time series with t time steps including drought years.
   - Therefore, each value in a cluster could be an array of N values corresponding to how a given date (Z500 anomaly map) is represented by those N values in the space of EOFs. However this

does not fit Eq (2). I may be wrong though and other approaches may be possible. What I am trying to highlight is that 3.5 needs a clear explanation of the methodological approach and its parts, with the notation of the different parts being consistent with each other.

- The final number of clusters is 71. However, I understand that since the analysis is performed on each model, many of those will be similar. Perhaps worth commenting this here?

- There is quite a number of typos, please revise them. Please, also in the rest of the text. I will avoid pointing at the grammar issues, but please take care of this.

GC4. (*) Section 4.1

I agree with the comments about Fig 2 and Fig 3 in general if we consider the details, maps, correlations, etc. However, perhaps I would have a different take on the actual interpretation of them.
Consider Fig. 2 first. Most models have a large low frequency variability, with large multi-decadal or multi-centennial departures from the long term mean. Some of them often longer than the reference period considered in Fig 2. They are to a large extent not consistent among the different models, which therefore indicates that they are more obviously related to internal variability than to the external forcing common to all experiments. If this is the behaviour of a real SOIL variable then the 1959-1979 interval considered as a reference is a very short interval of time and may correspond to a very specific state in the NOAH model, assuming that it also represents reasonably reality. However, a longer integration with NOAH or if we had more observations, would supposedly show a considerable level of low frequency variability; we do not know how much because we do not know how well the models in Fig2b represent reality. With this in mind, the fact that one simulation represents less variability during the selected period or another one represents more, does not mean that this or that model is doing better or worse, because it is not intended that these simulations represent the real 1950-79 variability, unless it would be clearly responding to external forcing or those simulations were driven by observations, which are not.
The previous reasoning extends to the other arguments related to Fig3. All arguments oriented to a better or worse representation (e.g. line 324, 325), I wouldn't agree with them, because Fig 2, indicates that except for the trends in the last decades SOIL responds to internal variability and thus the maps in Fig 3 are expected to show some level of similarity but not to represent faithfully what the NOAH model does within that comparably short period.
I think the authors should consider this argument and see how it impacts the orientation of the text and the interpretation of the figure.

Specific details:

- It would be good to indicate in the text or in the figure caption the time interval used for correlations.

- The maps shown in Fig 3 are indicating some relation to zonal circulation, NAO, which is mentioned in the text. If desired, this could be objectively calculated by indicating correlations with the NAO index in each model. But I would understand also that the authors would not want to go in that direction.

GC5. Section 4.2

1. Figure 4 caption and in the text. The mean percentages of total drought years and the mean duration of droughts are calculated from the ensemble means…
   I have reservations about the meaning of these numbers because of being calculated from the ensemble means. If the quantity that was being analysed would depend mostly on external forcing, I would agree, because the metrics based on the ensemble average would be meaningful as a filtered version obtaining after cancelling noise from internal variability. However, the behaviour of soil and drought occurrence here is shown and argued (e.g. Line 358) to be related to internal variability. Therefore, the statistics that would be comparable to the real world or representative for it according to each model are those of individual simulations, not the ensemble average.
   Arguing from a different angle: if we would have enough of a high number of runs, the ensemble average should tend to be flat and with no droughts.
   The previous arguments also would justify why this happens (L 368): '… more clearly in those models and periods with one ensemble member'

2. Line 371: ' This seems to agree with Cook et al (2016)'
   Do you mean the simultaneous occurrence or the period or both? Please indicate more specifically and if you think it is important elaborate…

   The comparison with Cook et al (2016) is interesting. For instance also in Lines 378-381.
   Do you think this comparison holds even if the reconstructions of Cook et al basically address the growing season (jja)? How can the statements of agreement or disagreement suffer from this? Some comments about this would be pertinent.

GC6. Line 405: 'Although it seems contradictory that P2 depicting a negative NAO condition also occupies a significant percentage of the occurrence…'

   I think this could perhaps be due to the size and definition of the windows used. I do not mean that it is wrong but could be an effect of that and if so, it may be worth commenting on it. The P2 pattern favours inflow from the SW into the Iberian Peninsula. The western side of it, over Portugal and southwestern Spain should not be dry with this pattern. However, dryness could affect the lands of northern Africa and central

Mediterranean Islands and over Italy. I think it is likely that the occurrence of drought with this pattern in the western box reflects the balance of wetness in the west/northwest region of the box and dryness in the rest. Perhaps it is worth assessing that and commenting.
This also takes me to suggest that it would be interesting to see the composites of soil for each group of patterns. One could actually show the composites over the whole Mediterranean, not only the boxes. This does not necessarily require an increase in the number of figures. The Z500 anomalies can be shown with lines using hatching for significance and the soil pattern with shading in the same map. It would help to understand how the different patterns influence drought in the region of interest.