Review of Kim et al., 2023 submitted to Climate of the Past
By Cécile Blanchet

Disclaimer: not being a modeller, I will not be able to judge the technicalities of the study. I will therefore provide general comments on Mediterranean climates, the comparison to proxy data and (hopefully) help to improve the readability and accessibility of the manuscript for non-specialist audiences.

Scope and relevance: The manuscript shows modelling results from the CMIP5-PMIP3 ensemble to elucidate the main climatic regimes associated to droughts in the Mediterranean during the last millennium. This is an important topic that fully deserves our attention. I find the paper well-written and very clear, with scope and relevance suited for CP. However, the goals of the paper are at times not very clear with regard to knowledge gaps (see below) and some clarifications need to be provided about the methodology and the evaluation of the models selected to conduct the analysis. I therefore request a minor to moderate revision before the paper can be accepted and hope that my comments will be helpful.

Main comments

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

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?

- 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.
- 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.
- 2. Data: $2.1: “All simulations were run with the volcanic, solar, and greenhouse has (GHG) focing
- §3.1: 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.

  - 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?
- §3.5: section very hard to follow, please provide additional information and background on the choices of the methods and their limitation.
- Results
  - Fig. 2: if I understand right, this is your control run for time series. 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.
  - Line 238: "SOIL and Z500 from NOAH-LSM" isn't Z500 from ERA?
  - 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 skilfulness 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).
  - Lines 286-288: I struggle with these sentences. Either too little or too much is said here. What is the role of models setup and skilfulness in this observation? I also do not understand what is meant by "counterfactual"?
  - Fig. 4: Would it be possible to quantify the antiphase? (bi-plot, cross-correlation) 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? 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)
  - 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.
  - Fig. 5: how reliable are these results with regard to Fig. 3?
  - Fig. 6: Very hard to read and to 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.
  - Line 338: "the mean occurrence of pattern groups for each model": what is exactly meant here?
- Discussion:
  - 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? 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.
  - Is there any way possible to test the observed seasonality pattern of climatic associations in the OWDA? Do you also observe a N-S antiphase (also

mentioned in Markonis et al. 2018)? 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)?

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