Review of "The climate of the Eastern Mediterranean and the Nile River basin 2000 years ago using the fully forced COSMO-CLM simulation" by Zhang et al. 2023

General Comments

In this manuscript the authors present the results of the first fully-forced, high resolution simulation of the climate of the Eastern Mediterranean and Nile river regions over the last 2500 years using the COSMO-CLM regional climate model. The work is mainly divided in two parts. The first one, where the authors evaluate the model with several modifications necessary for the simulation of past periods, for the present-day. And the second one where they assess climate changes between two past periods of time, namely the Early Roman Period (ERP) and the Pre-Industrial (PI) period. Here they assess differences in seasonal values of precipitation and temperature, as well as their connection to changes in the atmospheric circulation.

The paper presents some interesting results and its contents fit well within the scopes of Climate of the Past. Nonetheless, I do believe that the manuscript suffers from a series of major issues that need to be properly addressed before it could be considered for publication for the journal.

First of all, the objectives of the paper are in my opinion not very well defined. The employed methods are not always clearly described, both concerning the description of the experimental design of the presented simulations as well as for the statistical methods employed in the analysis of their results. This makes the understanding of the different analysis not always straightforward. Additionally, the presented analysis can sensibly be extended, making a full use of the transient simulation and of the driving GCM data, to understand discriminated model biases, as well as possible differences in the climate of different time frames of the simulation period. The beauty of your simulations is that you have so much data from which we could really learn a lot about past climate changes of the investigated area and their drivers. I would try to make a full use of them. Finally, it would be interesting to know how the model performs in past times, performing a comparison against proxy data for the study period.

Please, find below more detailed comments on which I based my judgement.

Major comments

- p3, l85-87: you mention that for this area a dense network of natural archives is available covering the last 2000 years. These data should be acknowledged and used for the comparison against your model results.
- p4., l133: you need to present a summary of the model setup, particularly concerning information on how you implemented changes in the model to take into account changes in the forcing.
- section 2.3.1: since you have data, wouldn't it be better to conduct EOF analyses of seasonal anomalies over the entire simulation time period and detect possible trends? In this way you could also compare changes across different periods. I think that this analysis, considering the fact that the study presents for the first time the results of a transient simulation for the area at high resolution, would be quite interesting. Also, are the presented results sensitive to the relatively short length of the two time periods considered?
- section 2.3.2: the method you use for the clustering of the different regions according to seasonal values of precipitation and temperature is not entirely clear. This part needs to be revised and possibly extended with additional details.
- section 2.3.2: Additionally, there are many choices that seem arbitrary in your method and that need further testing: for example, l219-220, why choosing only 6 EOFs for CRU and all for the other datasets?
- section 2.3.2: Another important point: do the different regions you derived from the different datasets contain different number of grid-boxes? This is a point that needs particular attention, in particular for the conclusions you draw from Fig. 3 and Fig. 5. When you quantify the match between datasets across regions, as performed in Fig. 5 and Fig. 3, you need to consider overlapping regions with the same number of points.
- section 2.3.2: Why for the present-day you use rotated EOFs and for the investigation of past periods you use non-rotated ones?
- section 2.3.3: since you have the results, why not showing the analysis in temperature, precipitation and mean sea level pressure for the entire simulation period? I think this would give some important and interesting insights on the simulated climate of the given period and area. In any case, whenever you show the differences between the two selected periods you must use the transient results for the entire simulation period to assess whether the obtained differences are comparable to the ones of other periods or if they particularly stand out? in the latter case, you could eventually try to assess why.

- section 2.3.2 and 2.3.3: In the paper the authors do not acknowledge in any way how the outcomes and conclusions of the manuscript are subject to the series of different arbitrary criteria they applied throughout their analysis. At least some discussion is needed here, to make readers aware that some changes might occur when changing some details of the method.

- Fig3 (same for Fig. 5): Why not comparing first the mean regional climatological values for a given region between the different datasets and then comparing the anomalies of each time series calculated with respect to the corresponding mean value of each dataset. Basically, instead of calculating all the anomalies with respect to the mean value of CRU in each region, it would be more appropriate to remove from the time series of each dataset the corresponding mean for the calculation of the seasonal anomalies. In this way you would have a proper assessment of the differences in the mean in each dataset as well as in their temporal variability.

- Fig. 7: Why are you now simply comparing spatial means over the entire region? in particular, what is the need for all the previously conducted analyses on sub-regions that you performed in previous sections in this context?

- section 3.3: Alternatively, you could also consider to conduct a canonical correlation analysis between SLP and precipitation and temperature over the entire period of time.

Minor comments

- p1, l14: at the regional scale
- p1, l15: atmospheric dynamics
- p1, l15: for present and future climate conditions
- p1, l17: please try to better describe in the manuscript what are the teleconnections relevant for the study domain
- p1, l17: you do not develop COSMO-CLM. You rather apply a high-resolution climate model modified for its application to paleoclimate studies. Make sure in the text that some studies already applied modified versions of COSMO-CLM to paleoclimate.
- p1, l23: comparable climatic conditions between the two considered periods
- p1, l23: variability of what? please specify
- p1, l27-29: period needs reformulation
- p1, l30: shed lights into
- p2, l17: involved in what? reformulate
- p2, l59: In summer
- p3, l78-81: reference needed
- p3, l90: in all PMIP phases (and not only PMIP4) model results were compared against proxies: please reformulate this period accordingly
- p3, l94-95: Can you specify the difference between proxy records and climate reconstructions? do not climate reconstructions rely on proxy records? I think that a better choice here would be simply using proxy-based climate reconstructions.
- p3, l97-104: I miss here some discussion, also based on previous literature, on why the application of RCMs to the study of the past is relevant.
- p3, l108-111: where can I see this? a proper discussion of available proxies for the region is needed
- p3,l112: I think that here you have a good chance to introduce the work conducted in this study and the simulations performed with COSMO-CLM and MPI-ESM.
- p3, l112: the listed forcing include both internal and external forcing: please correct.

- p4, l116: here it would be appropriate to also mention some of the works of Berger about the estimation of orbital parameters for the past.
- p4, l117-118: this sentence needs reformulation.
- p4, l119: I guess just some of the Nile flooding match volcanic eruptions and not all? maybe it might be interesting to report some example?
- p4, l120: the increase in energy in the climate system is continuous only for a continuous increase in GHGs. Please modify accordingly.
- p4, l124-126: you have to introduce before in the text the GCM simulation you are using in your study, as well as the fact that you are using this for running an RCM. This is not explicitly mentioned up to this point in the introduction. See also one of my comments before.
- p4, l124-126: Also, be aware that the land cover changes are specific to the target area and study period
- p4, l127: "but those forcings are not yet fully implemented in the RCM": please be aware that many other studies with modified forcing were already performed with COSMO-CLM.
- p4, l128-130: See comments above: so far it is not clear if you will be conducting a study with COSMO-CLM. Also, note that you use sometimes COSMO-CLM and some otehrs CCLM. Following the specifications of the CLM-community, I would recommend to always use the acronym COSMO-CLM for the model throughout the manuscript
- p4, l135-136: why this association should not be possible simply using a GCM? please better clarify
- p4, l136-138: This is not shown here. I would rather frame it as the possibility to use the results for the study of extreme events on societies. Still, in this case you must make clear that you need a proper comparison against proxy data before using the model results for past times.
- p4, l152: Armstrong et al. 2019 do not use COSMO-CLM. A more appropriate reference here would be the one of Soerland et al. 2019.
- p4, l. 152: ".. have been recently performed": COSMO-CLM activities and participation to CORDEX covers more than 15 years. Please, reformulate.
- p5, l155: pleas,e be more careful in the use of extreme terms: you actually do not revolutionise COSMO-CLM since other studies have already applied at least part of the changes necessary for the application of the model to paleoclimate studies that you are considering.

- p5, l156: As already stated before, it is not sufficient to include the reference to the paper of Hartmann et al. 2023 here. You need to provide a summary of the applied changes to the model and to its configuration.

- p5, l170: for which area they apply COSMO-CLM for, in the study of Bucchignani? why using their configuration? please specify

- section 2.1: As you mention later in the text, the configuration of the model is very important for an RCM cause it is region-dependent. What was the starting setup of your model? did you use the default setup for Europe provided by the CLM-community? you did not apply any additional changes beside the ones in accordance to Bucchignani et al. 2016? Eventually, provide more context on the reason for your choices in the model setup

- p5, l177-180: can you provide here more context on why selecting the two periods 1800-1850 CE and 400-362 CE in your study? I think that, also considering your performed analysis, a more appropriate choice would be the one of two periods with the same length. Also, you have so exciting results: why not performing the analyses over the entire simulation period?

- section 2.1: please also add here the horizontal resolution of your model as well as the extent of the domain over longitudes and latitudes

- Fig1: Is the outer box the entire domain of your simulation? please specify

- p6, l186: We evaluate: evaluate is in my opinion a more appropriate choice than "validate", since we have to acknowledge that also observations are not the absolute reality

- section 2.2: please be aware that the original resolution of ERAInterim is not 0.5°. Specify if you interpolated the data onto the target grid yourself or if you simply retrieved interpolated data from the ECMWF server.

- p7, l214: corresponding instead of according

- p7, l219: maybe it would allow a better understanding of your method if adding here (or eventually in the supplements) a figure with the EOFs obtained for the CRU dataset, for both considered variables

- p7, l221-222: can you provide more details on why the CRU REOFs accounf for the 75 % of variance in in each dataset?

- p7, l222: How do you calculate the 80th percentile of the loadings? also, you calculate the 80th percentile of the loadings with respect to REOFs of CRU or for each dataset separately? How are your results sensitive to the specific arbitrary choice of 80th and 75th percentile?

- p7, l223: Not clear what you mean by "paired" REOFs. Could you please provide additional details on the pairing methodology?

- p7, l226-228: specify that you produce the Taylor diagrams also for each season. In the Taylor diagrams, do you compare results over the same region? If yes, are these the regions derived from CRU? Eventually, please specify. In case you have calculated the correlation based on the different regions results are not appropriate, since for each dataset the regions are different.

- p7, l230: "between the first millennium ...": you actually do not consider the entire millennium, since your simulation starts only in 500 BCE, right? please reformulate

- p7, l234: why non-rotated EOFs in this case?

- p7, l234: again, specify that you conduct the analyses separately for the different seasons. Also, specify the variable that you use as input for the EOF analysis.

- p7, l238: "The index k represents the grid-point index covering the geographical domain": This is not clear. Do you mean that k indicates a given grid box of the domain of study? please reformulate

- p7, l240: "as the noise": is the noise? reformulate

- p7, l241: please justify why using only 3 non-rotated PCs in this case. Depending on the selected variable and season, it seems that the first 3 EOFs do not explain the same amount of variance. Wouldn't it be better to select the number of EOFs depending on a fixed threshold of total explained variance such as the 80%?

- section 3.1: Can you provide more details on how you calculate the spatial correlation between EOFs of each dataset for the different regions? Are you considering the same points in each region?

- p8, l254: Again, can you explain what you mean by paired EOFs? what is the CRU paired with in the first column?

- p8, l256: "For example, GPCC winter region 5 corresponds to GPCC REOF11 as the counterpart to CRU REOF5.": not clear, please reformulate

- p8, l261-262: Reference needed

- p8, l263: the strength of what?

- Fig. 2: What is the darker blue area in the different plots? In each panel it seems that you have more colours than in the legend. Why is that if you only select 6 regions in each case?

- p9, l268-274: are you comparing mean values calculated over different regions in each of the two datasets? if yes, I do not think this is appropriate. For such comparison you should calculate averages in each dataset over corresponding areas and then calculate the differences.
- p9, l273: "..higher than 0.7": It is not clear between what you calculate the correlation. Also, where can I see the values of the correlation and standard deviation in each case? It would be nice to see these values.
- caption of Fig3: what we see in the legend? please add to the caption
- p10, l279: can you better specify, possibly with the help of a figure, how do these precipitation regimes differ?
- p10, l282-284: But this should then be an issue of the GCM, if the issue is related to the large-scale circulation. Could you check if this is the case?
- Table2: why you select again 6 regions?
- p10, l293-294: "Precipitation in NR is mainly concentrated...": I cannot see this from Fig. 4. Please add some climatological characterisation of the different regions in terms of precipitation if you aim to discuss their features.
- p10, l295-297: In the method section it is not clear which model configuration was the base for your simulations. In case you selected the model setup suggested for Europe by the CLM community, you should make this clear in the text. Eventually, you should provide some reason why you did not tune the model properly for the study area, since you mention this as a possible cause of model bias
- p10, l300: also connected to previous questions: if the explained variance is 50% in this case, why not selecting more REOFs?
- Fig.4: why not considering all regions of the domain?
- p11, l308-311: Again, which regions are you considering in the taylor diagrams? the regions of CRU? in this case, please specify it. If you are instead considering the different regions of each dataset, this comparison would not be appropriate for the same reasons given before.
- p11, l307: "it is visible": where can I see this? table 2? please specify
- section 3.2: Why not conducting the analysis over the entire period? also, it would be more appropriate to compare statistics calculated over period of the same length (here 51 against 39 years), in particular when comparing interannual variability

- p13, l340-342: It would be interesting to see a plot of the volcanic forcing over the entire simulation period to better understand the relevance of the ERP with respect to other periods in terms of forcing p14, l348-354: Are these differences particularly different than differences arising from the interannual variability of the entire simulation? you could also have a map of the bias with a dot where the differences are above the standard deviation of the entire simulation for the considered variable and season
- Fig 7: connected to the comment from before: Could you add a map of the mean differences obtained for each variable and season.
- p14, l360-362: Again, you should do these analyses considering periods with the same length
- p14, l372: Reformulate:the NR region is highlighted in blue in Fig. 8. Also, specify to which regions correspond the red and green color boxes
- p15, l393-395: what about summer precipitation and winter temperatures? why not considering them? please specify
- p15, l396: EOFs applied to which variable now? seasonal anomalies of SLP?
- p15, l398: "..the three leading EOFs..": specify of winter precipitation
- section 3.3: as already suggested, EOF analysis of SLP and its projection onto precipitation and temperature should be calculated over the entire period of time. In this way you could assess how relevant changes are for the given periods compared to others, as well as possible trends over the entire period of time.
- p15, l401: which peninsulas? specify
- p15, l401: "...indicating the influence of the westerly circulation and the land-sea interaction": How? can you better specify how this indicates the influence of westerly circulation and land-sea interactions?
- p15, l403: "...likely connected with the direct impact of the large-scale circulation": again, could you better explain how is this plausible?
- p16, l418: did you check cyclones in your data or you have some reference for this statement?
- p18, l453: as for the precipitation, you could extend a bit on the dynamical drivers of the obtained temperature patterns
- p18, 448-458: It is not clear what is causing the differences in dynamics between two periods. Can you make some guesses? Again, here it would be very important to consider the variability of precipitation and temperature for each point over all the simulation period. Or maybe in terms of the EOF, to check how the PCs evolve over the entire simulation period.

- p19, l465: "...we presented the first fully forced CCLM adapted for paleoclimatic applications ...": You actually did not really present it, since you miss lot of crucial information about the model configuration in the methods section

- Conclusion section: you are still missing a comparison against proxies

- p19, l473: "...limitations in representing convective processes linked to the ITCZ.": You actually did not demonstrate this. Based on which ground can you affirm such statement? Eventually, you have all the tools and data for demonstrating this point. For example, you could check how is the ICTZ simulated in your model chain? In case the workload for conducting this analysis would be too much, I would try to be much more careful in explaining the possible reasons of model biases. For example, by extending the description of the paper of Adams et al. 2016, making it clear that this does not discuss an RCM, but the results of GCMs. Again, it would be interesting to see what happens in terms of the ITCZ in your driving GCM MPI-ESM.

- p19, l476477: The same applies to the conclusions of lines 476-477: you have all the data and tools to quickly check whether the drying bias is associated to the representation of clouds in the model.

- p20, l492: "... and is subject to various circulation mechanisms... " Reformulate. In particular, specify which ones are these mechanisms.

- Taylor diagrams in the appendix: The captions of figures in the supplements need some reformulation. Also in the Taylor diagrams it is not clear for which regions you are conducting the analyses. The CRU regions? please specify