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.

Thank you for your insightful comments and suggestions on the manuscript. We have thoroughly reviewed each of your points and are committed to addressing them effectively.

In response to your feedback regarding the objectives of the paper, we recognize the need for modification. The primary aim of the paper is to provide a comprehensive characterization of the ERP climate. To achieve this, we, in the first part of the manuscript, focusing on the evaluation of the CCLM output. Additionally, we acknowledge the necessity of refining the employed methodology to offer a more accurate description of the applied analysis. It's essential to clarify that the primary focus of the paper is to compare the BCE climate to the pre-industrial climate and further characterize the climate over EM and NR during the BCE period. We want to highlight that covering the entire period of the transient simulation is beyond the scope of this specific study. Furthermore, we share your interest in comparing proxy data to our CCLM transient simulation output. We are currently in the process of preparing a dedicated manuscript specifically addressing the aspects related to proxy data. For a more detailed information, please see the following responses to the major comments.

Major comments

- 1. 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.

  We agree with this opinion, the data is currently under processing and being compared with the proxy record. Nevertheless, the purpose of this paper is composed by two parts, one is to provide an evaluation of the performance of the model for the present-day climate, and the other is to compare the ERP and PI periods, further provided a characterization of the ERP climate. We will modify the purpose of the manuscript and revised the introduction by explaining why there is a focus on the ERP period and why it is important to obtain a more detailed climatological understanding of the ERP. And for further comparing the simulated output with proxy is not within the study scope of this manuscript but another direction that we are currently working on.

- 2. 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.

  We are currently preparing a paper which is mainly discussing the set up and the changes in the different external forcing in detail (Hartmann, et al. 2024). We can share the manuscript confidentially to further illustrate the above-mentioned information. The external forcings are based on the recommendation for the PMIP4 past1000 contribution to CMIP6 (Jungclaus et al., 2017). We will add some more information in Section 2.1. Here is some more detailed information:

  - The orbital forcing is represented by the eccentricity, the obliquity and the longitude of perihelion.
  - The total solar irradiance represents the solar forcing.
  - The volcanic by the aerosol optical depth.
  - The changes of greenhouse gas concentrations are given in equivalent CO_2 concentrations which consists of CO_2, CH_4 and N_2O.
  - The land use changes are introduced as LAI, maximum and minimum plant coverage derived from the MPI-ESM transient simulation.

- 3. 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?

  We acknowledge the suggestion to conduct EOF analyses across the entire period for seasonal anomalies. However, in the specific context of characterizing the climate of the ERP periods, we find it necessary to perform EOF analyses over the selected periods. This is essential for achieving the main objective of comparing the ERP climate to the pre-industrial (PI) period.

  We have analyzed the trend of the entire period, based on General Circulation Model (GCM) output, which helps us to identify the BCE period for further investigation. Given the established connection between climate conditions and societal events (McCormick et al., 2012) and notably, this specifically focuses on climate change and the Roman Empire back to 100 BCE, leaving a gap concerning the earlier Roman Kingdom period, which dates back to around 500 BCE.

To address this gap and explore associations between climate change and the entire Roman period, we have in this study chosen the period 400-362 BCE. This period predates the established Roman Empire, capturing the rise from a kingdom. By examining the characteristics BCE period temperature and precipitation in this study, we can provide knowledge of the climate back to 2000 years ago in a more detailed spatial and temporal resolution.

- 4. 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.

  To address this comment, we have made substantial revisions to the method section. Additionally, we have added discussions on different cluster methods in both the introduction and discussion sections to provide a comprehensive overview of our approach.

  Also, we will provide additional details that support the specific procedures employed in this study. For a thorough understanding of these enhancements, please refer to our response in comment 6. We believe these adjustments significantly contribute to the clarity and robustness of our methodology, ensuring a more detailed presentation of our research approach.

- 5. 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?

  We chose the CRU dataset as our reference data for assessing the simulation output in the first part of the manuscript. We chose the first six principal components because with these first six components, we cover more than 75% of the dataset's variability. Given that the purpose of principal component analysis captures the major data patterns, using beyond the sixth component will not give valuable information for analysis. This choice guarantees that we are keeping the most important information of the dataset, helping us to better identify the most important precipitation and temperature in the subsequent analysis.

- 6. 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.

  The overlapped grid boxes of the different subregion should be taking into consideration. Homogeneous region identification is based on the calculated EOF pairs with reference to the CRU dataset (refer to Tables 1 and 2 in the manuscript). Subsequently, we establish homogeneous regions across the three datasets by employing a threshold criterion. Specifically, grid points with values equal to or exceeding the chosen threshold are designated as part of the subregion. For this study, we set the threshold at the 80th percentile for precipitation and the 75th percentile for temperature, determined from the respective loading distributions. This percentile-based thresholding considers both strong negative and positive values, accommodating dipole patterns in the EOF loadings.

  Therefore, for better illustrating the choice of the threshold in order to have less grid boxes being retained in more than one subregion, we will therefore present results using different percentiles as an example, focusing on the EM region. Figure R1 demonstrates that different thresholds influence subregion overlap, with lower thresholds causing more overlap and higher thresholds resulting in less coverage. Additionally, Table R1 and R2 provide insights into the standard deviation and the percentage of retained grid points for each subregion. These details contribute to a comprehensive understanding of the subregion identification process.

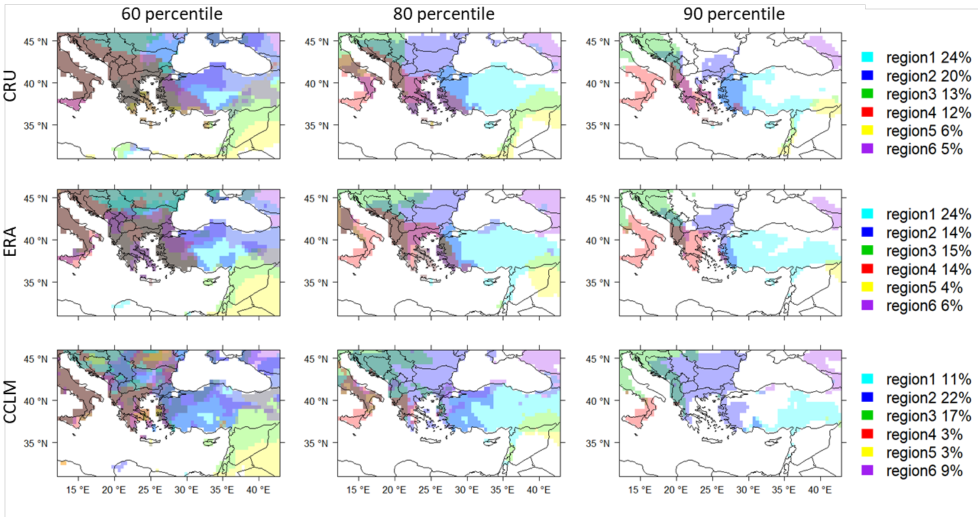


Figure R1. Subregions when using different thresholds for the EM region.

Table R1. Standard deviation of the subregions when using different thresholds.

std	Region-1	Region-2	Region-3	Region-4	Region-5	Region-6
CRU-60	11.27	8.30	11.16	9.53	6.88	8.73
CRU-80	10.06	6.68	10.58	7.08	4.98	8.30
CRU-90	9.09	5.18	9.21	4.92	2.95	6.86
ERA-60	10.30	6.80	10.29	9.31	4.58	8.79
ERA-80	8,24	4.74	9.82	7.05	3.69	10.33
ERA-90	7.37	2.97	8.47	5.05	1.94	9.40
CCLM-60	6.14	8.47	12.10	6.07	2.66	10.29
CCLM-80	4.74	7.13	12.69	6.56	1.97	12.88
CCLM-90	3.81	5.60	12.23	5.54	0.32	14.03

Table R2. Covered grid box percentage of the whole region of the subregions when using different thresholds.

	Region-1	Region-2	Region-3	Region-4	Region-5	Region-6	sum
CRU-60	0.32	0.33	0.20	0.18	0.16	0.13	1.32
CRU-80	0.19	0.21	0.08	0.10	0.07	0.03	0.68
CRU-90	0.09	0.09	0.05	0.06	0.02	0.02	0.33

ERA-60	0.34	0.30	0.23	0.19	0.12	0.14	1.32
ERA-80	0.19	0.17	0.10	0.12	0.04	0.04	0.66
ERA-90	0.12	0.07	0.05	0.07	0.009	0.02	0.33
CCLM-60	0.29	0.35	0.22	0.13	0.17	0.18	1.34
CCLM-80	0.17	0.23	0.11	0.05	0.04	0.07	0.67
CCLM-90	0.06	0.15	0.07	0.02	0.001	0.03	0.33

- 7. 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?

  As we mentioned in the manuscript. The classic patterns, EOFs, are orthogonal, hence, the eigenvectors are uncorrelated. For some applications, this is a useful characteristic (i.e., setting up multiple regression models with predictors that are not collinear). In meteorological applications, however, the orthogonality constraint may be disadvantageous, because most processes in the real world are not orthogonal (Storch and Zwiers, 1984). We thus applied the VARIMAX rotation to obtain rotated EOFs (REOFs) that are physically more consistent than the non-rotated patterns. REOFs are thus used for the validation of the model set-up in the present period (1980 - 2018 CE), while non-rotated EOFs are implemented for the comparison of the mean climate conditions in PI and ERP times, ensuring the preservation of meteorological characteristics in temperature and precipitation. The application of non-rotated EOFs guarantees the accurate interpretation of the regression involving the non-rotated principal components of precipitation and temperature with respect to large-scale circulation patterns.

- 8. 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.

  Thank you very much for pointing this out, we agree that it will be very interesting to investigate the temperature, precipitation, and sea level pressure for the entire simulation period. Nevertheless, the purpose of this paper is 1: to assess the transient simulation and 2: describe the characteristics of the ERP climate. Thus, for the EOF analysis, if we select the whole simulation period, we cannot characterize the ERP climate specifically but an overall trend over the last 2500 year. Nevertheless, for choosing the current two interesting period, we have analyzed the trend of the entire period, based on General Circulation Model (GCM) output, for more details regarding why we have chosen this two specific period, please refer to the response of comment 3. Given the substantial computational and time resources required for this ongoing process, the Regional Climate Model (RCM) paleo transient simulation is currently in progress. Consequently, the simulation results spanning the entire 2500-year period will not be available in the near term.

- 9. 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.

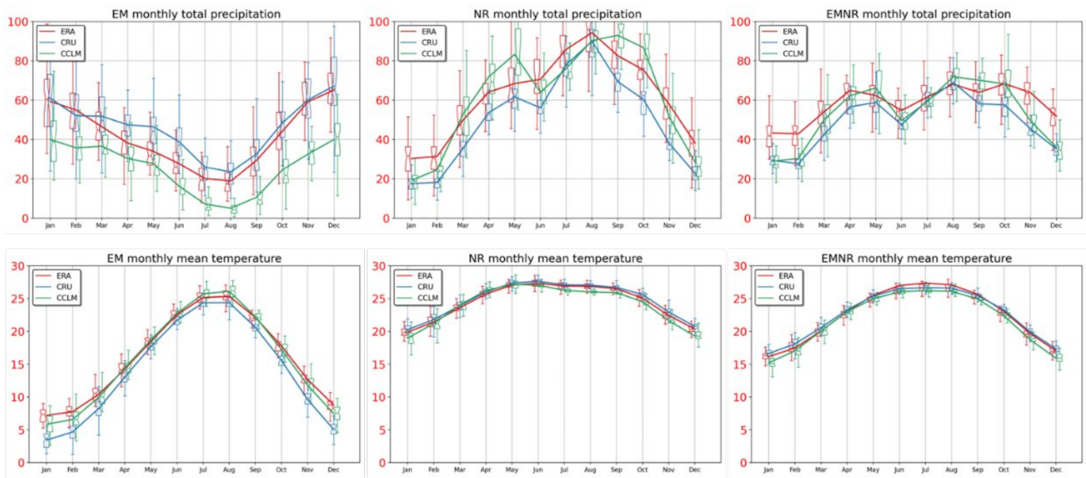
  We are aware of the different clustering methods like K-means and Hierarchical clustering. To enhance clarity on our choice of employing Empirical Orthogonal Function (EOF) analysis, we plan to include a paragraph discussing these alternative methods. Unlike machine learning-based clustering, regionalization based on EOF analysis retains a strong physical interpretation. Furthermore, the principal components derived from EOF analysis provide valuable insights into temporal variations and their connections to large-scale circulation patterns. This approach aligns with our goal of not only capturing patterns but also understanding the underlying physical processes driving regionalization.

- 10. 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.

  We understand that comparing mean regional climatological values first and then calculating anomalies with respect to the corresponding mean value for each dataset is indeed a valid approach. In the first part of the manuscript, we want to validate our simulation results. By calculating seasonal anomalies with reference to CRU data, we can not only obtain a more accurate assessment of differences in both the mean and temporal variability across datasets but also if there is a systematic bias of our simulation results against the observational/reanalysis data sets (such as overestimation or underestimation).

  This methodology ensures a proper evaluation of how each dataset deviates from the observational data (CRU), providing insights into the characteristics and variability within each dataset. In addition to this,

  We compared the absolute monthly total precipitation and mean temperature between the GCM, observational data and CCLM for period 1980-2018 to further demonstrate the mean state of the different data sets, see Figure R2, where the CRU data is the observational data, ERA is the global forcing data of regional climate model CCLM.



- 11. 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?

In Figure 7, we present the mean values across EM, NR, and the combined EMNR as a comprehensive comparison with the pre-industrial period. Additionally, we employ Empirical Orthogonal Function (EOF) analysis to explore the connection between major precipitation/temperature patterns and sea level pressure. The rationale behind assessing the CCLM output through clustering via EOF analysis is to scrutinize its efficacy in capturing temporal variance within specific subregions.

Initially, we attempted to perform EOF analysis over the entire region. However, due to distinct precipitation characteristics in EM and NR regions, the analysis struggled to identify variabilities specific to each. Consequently, we choose to separate the region, recognizing the unique patterns within EM and NR. Moreover, considering the different precipitation regime of the two regions, the application of EOF methods might yield different results. By clustering the region based on EOF analysis, we could evaluate the simulation results by looking at that if the model is reasonably simulation the climate of the different variabilities.

12. 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.

In examining large-scale circulation patterns associated with principal components, there are several methods available, including canonical correlation analysis and linear regression. In our specific approach, we utilized Empirical Orthogonal Function (EOF) analysis to derive the primary principal components of precipitation and temperature data. Given this, we have opted for linear regression in this study to investigate the potential relationship between the major precipitation/temperature EOF pattern and sea level pressure. This choice aligns with our aim to understand the specific associations between these climatic variables and provides a focused exploration of their interconnections through linear regression analysis.

Minor comments

All the minor comments will be addressed as suggested. Specially, response to some of the minor comments are described here.

- 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.

  Sentence reformulated as: We modified a high-resolution regional model, COSMO-CLM, for paleoclimate applications, by integrating all external forcings and conducted a transient simulation from 500 BCE to 1850 CE.

- 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.

  While the proxy- based climate reconstructions are location based and spread over several locations within the interested region. A RCM could help to close this gap and provide a better representation of the spatial patterns.

- 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?

  Researchers looked for the relationship between the Nile flood anomaly from climate modelling output with the volcanic eruption in 1902, 1912, 1963 and 1982. And found out that the Nile floods matched with the effects of volcanic eruptions, i.e., large-scale volcanic eruptions that cool the Earth's atmosphere and thus disrupt the normal flow of the Nile.

- p4, l120: the increase in energy in the climate system is continuous only for a continuous increase in GHGs. Please modify accordingly.

  Modified: Continuous increase of Greenhouse gases (GHGs) trap in turn the longwave radiation that is emitted by the Earth surface and lead to a continuous increase of energy in the climate system and effects on the global climate (Ramanathan and Feng, 2009)

- 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.

  Exactly, therefore here we are using "fully" forced. As in the introduction mentioned above, there is some study specific focusing on one of the external forcings, such as orbital or GHG.

- p4, l135-136: why this association should not be possible simply using a GCM? please better clarify

  The association between the regional climate with large scle circulations is possible but lack of detailed spatial information, therefore the sentence has been modified to: "Further, it enables the exploration of the association between the regional climate patterns at higher horizontal resolution and the large-scale atmospheric circulation patterns from the GCM world."

- 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.

  The comparison of the simulated results against proxy record is under processing and planned in another frame of work focus. In this paper we mainly want to compare the climate of the ERP to PI and characterize the ERP climate. So one have pre knowledge about the climate change of the two period and further discussing the trend or variability of the region by linking or comparing the simulation output with proxy records.

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

  The representation of albedo and aerosols are found to be the most important parameters in the studied region (see also Hartmann et al., submitted) and are set the same as in Bucchignani et al., (2016) in which the simulations are performed over MENA (Middle East and North Africa). In the manuscript we have modifies this part as "MENA region covers similar study region in this study which across the equatorial region, in our previous test simulations, we have found out that the parametrization of albedo and aerosols are the most important ones as described in Bucchignani et al., 2016. Therefore, we have chosen the general model setup from Bucchignani et al., 2016."

- 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.

  As mentioned in the previous response, before setting up our transient CCLM simulations, we have performed several test simulations to find the best configuration of the transient simulation. As our simulation domain covers similar region as MENA described in Bucchignani et al., 2016, we have therebefore ran several test simulations based on the simulation's setup of Bucchignani et al., 2016. The test simulations include several different setups, details in the table shown here:

Simulations	Specific setting
CCLM	itype_aerosol=1, soil layer=9 (0.005 - 11.5 m), pstbga= 0.045_ireals / 19330.0_ireals
CCLM-MENA	itype_aerosol=2, soil layer=7 (0.005 - 14.58 m), pstbga= 0.045_ireals / 19330.0_ireals