Dear Reviewer #1,

Thank you very much for your very helpful and substantive suggestions. We have addressed them, and they will be included in the revised version. Detailed answers and how we will address them can be found below.

We very much appreciate your support and hope that our plan for the revised version meets your expectations and requirements

On behalf of all co-authors, yours sincerely

Eva Hartmann

RC1

Review 'Climatic Impacts of the 536 and 540 CE Volcanic Eruptions Simulated with COSMO-CLM over the Middle East and Africa'

The authors have carried out climate simulations for the Eastern Mediterranean region to study the regional impact of the mid-6th-century eruptions of 536 and 540 CE. They use both a global climate model (MPI-ESM) and a regional high-resolution model (COSMO-CLM). The authors find that the different regions in the Eastern Mediterranean show diverging hydroclimate responses and cooling duration after the eruptions.

Although the study could contribute to the understanding of the volcanic impact in the studied region, I have several major issues with this manuscript. These should be addressed before I would consider this manuscript to be ready for publication. Below are my main concerns in order of appearance.

Methods:

From the methods and throughout the manuscript it does not become clear which simulations are done for this study and which existing ones are used. In the table it says the simulations for this study cover 500-1850 CE, but the reference period used to calculate the anomalies is somehow covering 1-1850 CE. Past2k simulations for the MPI-ESM model exist (which are mentioned in the methods), as well as a 10-member ensemble for the LALIA. As the authors describe in the introduction, more ensemble members reduce the internal variability, so why don't the authors include these simulations in their study? And why are the individual simulations presented the figures? Showing them individually does not reduce the internal variability, which was the argument for using more than one ensemble member in the first place.

As shown in Table 1, our simulation period covers 500 BCE to 1850 CE, while the reference period is 1-1850 CE for comparability reasons with other studies. As the reviewer correctly states, other Past2k ESM simulations as well as the 10-member ensemble for LALIA do exist. However, neither of those is available with the required temporal resolution to force the regional model at its lateral boundaries. CCLM as an RCM requires at least a 6-hourly output ratio, and these Past2k ESM simulations are available only with a monthly resolution. We have decided not to increase the temporal resolution from monthly to 6-hourly statistically, in order to avoid introducing errors, and for this reason we have repeated the global simulations, writing out the output with the required high-resolution 6-hourly output intervals.

We performed a second realisation not only to reduce internal variability, which is an inherent feature of the climate system. The second realizations helps to assess the robustness of results concerning the changes and impacts of external forcings, specifically the volcanic eruptions. In the revised manuscript, we will clarify the simulation design and workflow and present results from the individual simulations, showing that the signal is not confined to a single run or only evident in the ensemble mean, but emerges within at least the two independent realisations.

Results:

It is also unclear which model simulations are presented in the different sections here. Which plots show the MPI-ESM simulations, and which the COSMO-CLM simulations?

All displayed figures and results show the regional COSMO-CLM simulations. We will correct and acknowledge this in the revised version.

The results sections are describing a lot of wetter/dryer/warmer/colder for different regions, which reads a bit chaotically. Having this type of results summarized in a table or map together with the climatology would help streamline the results a lot.

In the revised version, we will provide a selection of the results in a clearer and more understandable way.

Figures:

The figures in this manuscript are unreadable. The subfigures should be larger, especially in figures 4, 5, and 6, and the font size of figure titles and labels is way too small for all

figures. For figure 3, the variation in precipitation is not visible for some regions because the range on the y-axis is too broad. I suggest showing different ranges for the different regions. I am also missing significance testing in figures 2, 3, and 7, as well as a discussion on the significance of the findings.

We will improve the readability of the figures. In Figure 3, for instance, we use the same axes for consistency. However, to improve readability, we will adjust the y-axis scale as needed. A test on the statistical significance in different figures will also be added and discussed. We will also address the points mentioned with respect to the significance testing and the corresponding interpretations.

Discussion:

Perhaps the largest concern I have with the manuscript is the lack of discussion of the results. After the introduction, previous studies are completely ignored. The authors should do a better job at including the relevant papers in a discussion. The results should be compared with previous studies (include citations!), and the findings should be put in a larger context.

We agree with the reviewer and will add more relevant literature, a comparative discussion with previous studies, and extend this section. This will be achieved by placing results in a broader context, including both the citations from the introduction and referencing additional papers.

I am also missing a discussion on the uncertainties of the study. The authors argue that climate models give insight in the processes behind the anomalies after eruptions, but do not actually discuss in depth why they find the results that they do. I like the atmospheric circulation section, but it seems a bit of an afterthought. Throughout the paper, different regions are indicated with abbreviations, but for this section they use different names/regions. The atmospheric circulation connection back to the precipitation and temperature changes could be a bit more elaborate.

We agree with the reviewer that this part can be improved by including additional paragraphs on the uncertainties of the study. We have already discussed certain limitations in the submitted paper, including the limited number of ensemble simulations and the impact of internal variability on our findings. For the revised version, we will also elaborate on a stronger link between the atmospheric circulation findings and the other subsections, thematically and linguistically. In addition, we will also correct conventions and abbreviations to be consistent throughout the whole revised text.

***Conclusions**:*

The summary given in the conclusions reads a lot better than the results section but is rather lengthy. The flow of the paper would be much approved if this is moved to the results section and call it 'Results and Discussion', or if it is given its own 'Discussion' section. A comparison with other studies can then be added, as well as putting the results in context. This way a lot of repetition can be avoided, and the results can be discussed right away, which will make the paper more pleasant to read.

Thank you very much for this suggestion. We will merge the Discussion and Results sections into one coherent section, where results will be presented and discussed immediately afterward. That will reduce potential duplication and will also allow shortening the Conclusion section.

The importance of the findings should be strengthened in the conclusion. Mainly mentioning that they can be used for comparison with 'existing proxy reconstructions, historical documents and observations' is not sufficient. Why not include the comparison to available proxy reconstructions and historical documents in this study? It would strengthen the paper if a comparison is carried out.

For this study, we aimed to detail and emphasise the regional climate simulations with the paleo-RCM accounting for changes in external forcings, including the physical interpretation of the period under study. We agree that a follow-up new study, currently in preparation, should include a dedicated proxy/reconstruction and model comparison.

Please find the minor comments below.

We sincerely thank you for the additional detailed comments and suggestions. We will consider each of them carefully and revise the manuscript accordingly.

Where no specific response is provided below, the comment will be addressed in the revised manuscript.

***Introduction**:*

Line 25: 'Existing evidence…' add from tree ring reconstructions

We will include evidence from tree-ring reconstructions. However, the available studies sometimes reach contrasting conclusions following methodological corrections, revised interpretations, and shifts in the periods considered. Moreover, their spatial coverage is largely confined to the central European Alps and the Altai

region (e.g. Büntgen et al., 2016, 2022), rather than our study area. Tree-ring reconstructions therefore provide only a partial and insufficient representation of climatic conditions in the study area during the LALIA. In addition, precipitation signals are available from only two reconstructions, covering Greece and southern Türkiye.

Line 38-39: 'Together, these considerations highlight the need to better constrain the spatiotemporal climate response across the Mediterranean, the Middle East, and Africa,…'

This cannot be concluded from the text above. I suggest moving this farther down.

Following the short comment to the point above, we will build the justification on the uncertainty and unclear signal in the study area.

Line 58-61: Here you give previous literature on the topic. Use these studies in the discussion to compare your results to.

Thank you, we will do so.

Line 66: 'Simulations…' Buntgen did not carry out simulations. Be careful not to mix simulations and reconstructions.

We will clarify the distinctions between the individual studies and tools.

Line 73: van Dijk et al. (2022) used global ESM simulations that were analyzed for more local domains. This is not dynamically downscaling.

Line 77: '…here we make use of…' This implies you did not carry out the simulations for this study. But in the methods section, it sounds like you did. This should be clarified.

We will rewrite this section.

Line 89: '…lower, and the need for multiple realizations and explicit regional analysis is therefore even stronger.'

Do you think two simulations is a large enough ensemble (this should be added to the discussion)? There is a 10 member MPI-ESM ensemble available for the 6th century, including this would strengthen your argument for using multiple ensemble members.

As discussed above, we could not use the 10-member ensemble because it is available only at monthly resolution. Converting these monthly data artificially to the

6-hourly resolution required by CCLM would have introduced substantial biases. The added value of the RCM would be completely over-compensated by the pre-processing of the monthly mean data.

***Methods**:*

Line 116-117: 'T63). 'Two realizations are available with identical external forcings but different initial conditions, enabling an assessment of robustness against internal variability at the regional scale.'

Are these the past2k simulations? What period do they span? More information is needed here.

We will add this information, they cover the period 500 BC-1850 AD.

Line 138-139: 'simulation. The reconstructions selected for PMIP4 past1000 were officially extended to this period for past2k experiments (Jungclaus et al., 2017).'

This feels a bit out of place. In this section you are describing the COSMO experiments. I suggest moving this up to the MPI-ESM section above. Also clarify if these are the simulations used in this study or not. If not, why are they not included?

This is correct, we describe the COSMO experiments and the corresponding forcings that have been implemented.

Line 142-144: So you ran both the global model and the regional model for this study?

Yes, we ran both the global model and the regional model.

Line 146-147: Two ensemble members is not a lot. I am missing a discussion about this in the manuscript.

As already mentioned above, there is a good justification for 2 ensemble members. Considering as well the computational storage and time requirements 2,350-year-long ESM and RCM experiments require, we have not prepared additional ensemble members; we will expand this discussion. Besides these technical constraints, even a small number of two simulations allows testing the imprint of changes in external forcings, not only in a statistical sense, but based on physical reasoning using independent realizations of the two climatic trajectories. For instance, if the imprint of the volcanoes is hardly visible in both simulations, it can also hardly be argued

that it is caused by the stratospheric aerosols. Likewise, if the imprint is substantial and is evident in the two independent realizations (and the corresponding RCM simulations), there is a high chance that the imprint is based on physical mechanisms (i.e., direct changes in radiation and/or changes in atmospheric circulation) and not only due to internal variability in one of the two simulations.

Figure 1: Not all the domains described in the caption are visible in the figure. The figure could also be larger.

The reviewer is correct, thank you for pointing this out. An incorrect version of the map was inadvertently inserted, and therefore part of the actual domain is missing. The correct map is provided in our response to Referee 2. We will replace the figure, increase its size, and improve its readability in the revised manuscript.

Line 167: How well is the geographical variability captured in the different model grids?

We will add a similar map as Figure 1 for the ESM. Additionally, the ESM simulations will be added in the other figures, for example the climate diagrams (Figure 6). Further, we will show the climate classifications of the domain in both grids.

Line 171-173: 'Nicholson (2018) shows the general circulation for Africa in the boreal

summer (July/August) and in the boreal winter (January).'

This provides no information. Please explain what they found.

This entire section (2.4) reads a bit difficult. The explanation of the different circulation patterns should be combined with the temperature and precipitation patterns, so the reader gets an overview over the climatic situation for each subregion. This would also make it easier to put the anomalies from the eruptions in context in the results and discussion section.

We will improve the writing of this section.

***Results**:*

Line 195: The 30-year running mean is not described in the results. Why did you choose to show this time interval? I suggest choosing a different time interval that is more relevant, like the 15-year mean.

We used a 30-year period, following the standard climatological reference-period length. This also allows the variations across the 2,350-year time series to be presented more clearly.

Line 196: wet and dry is not shown in Figure 2.

Line 196: replace 'overlaps' with 'similarities'

Line 198: Showing the major eruptions described in the text in the figure as well would help the reader.

Figure 2: The figures are hard to read. The blue and red lines indicating the warm and cold year hide the actual temperature. I suggest making these lines transparent so the temperature can still be read from the figures.

Line 199: Also cite Sigl et al. (2015)

Line 207: 'On a seasonal scale, the results are comparable and are therefore presented in the Appendix.'

In the introduction, you argue for seasonal rather than annual results, yet here you mainly show annual means and have the seasonal means in the appendix. How is this making the study different from previous studies?

No previous study has presented regional climate simulations for this domain and its subdomains over the full period from 500 BCE to 1850 CE, using an RCM taking account of changes in external forcings explicitly, i.e., implementing these changes directly into the radiation routines of the source code. We therefore begin with annual means to provide an overview of the results across the entire 2,350-year period. At this broad scale, the seasonal patterns do not differ substantially from the annual results. We consequently present the seasonal means in the appendix and draw on the seasonal results in the subsequent sections, where specific periods and regional features are examined in greater detail.

Line 229-231: 'JJA is the wet season in SAH, ARP, CAF, and NEAF under the influence of the ITCZ and the dry season in the northern (MED, WCA) and the southern (SEAF) part of the domain.

Is this in the climatology, or a result? This needs clarification. As mentioned before, a map or table with the climatology and the anomalies for different regions would help organize the results.

We will clarify this.

Line 235: 'No clear pattern of distribution of extremes can be identified in either season.'

So, unlike what was stated in the introduction, seasonal patterns do not matter after all? This should be discussed.

Line 242: Figure 2 is referenced here, but it should be Figure 3.

Line 250-251: '…also the wettest or driest period, especially at the seasonal scale.'

Are these anomalies outside of the range of variability?

We will correct and discuss those points.

Line 258: I suggest using a different title for the section. You also described regional anomalies in the sections above.

We agree and suggest modifying the title to: Regional DJF and JJA climate anomalies

Line 262: '…that two realizations are…' Is this referring to the MPI simulations, or the COSMO simulations? This needs clarification.

Figure 3: The y-axis on the first few subplots is too broad, which makes it hard to read the variations in precipitation for these regions. I suggest using a different y-axis range for the first four regions. In addition, the suggestions for figure 2 apply here as well.

Figure 4: The subfigures are unreadable. Make sure the subfigures are large enough to easily read the titles, labels, tick marks, and legend. It would be more intuitive to plot the average of the simulations. Showing them individually does not reduce the internal variability, which was the argument for using an ensemble in the first place.

Line 314: 'It should be noted that this is the average value for both the northern and southern hemispheres.'

I don't see the added value of this sentence.

Line 321-324: '. In summary, CAF, NEAF, and SEAF have the smallest annual temperature amplitude but the highest variation in precipitation, which is also associated with high variability within the subdomains. SAH and ARP have a higher temperature amplitude but very low annual precipitation. MED and especially WCA show the largest temperature variation with medium amounts of precipitation, although fluctuations within the subdomains can also be large here.'

It would be nice to explain a bit more where the differences between the different regions come from. I suggest starting with this section before discussing the anomalies and connect the two sections a bit better. This gives an overview (and visualisation) of the climatology of the regions, which should be used to put the anomalies in context.

This is a very good suggestion. We will exchange the two sections and will give them a better connection.

Line 331: '…November, which can be up to twice the reference precipitation amount.'

I do not see this from the figure.

Figure 6: As with the other figures, this figure is hard to read.

Line 342: 'Overall, these results are consistent with findings from previous studies.'

This is highly insufficient. Explain which studies (use references!) and what exactly the similarities are between the results from those studies and your own.

Figure 7: Again, unreadable figures. Why are the individual simulations plotted?

Line 350-351: '…fields at the upper and lower levels of the troposphere.'

Why were these levels chosen?

Line 367-368: 'Despite the surface warming in some regions, temperatures at approximately 12 km height are cooler than the Common Era mean across the entire domain.'

It is very unsatisfying to end the section with this. Why is this?

We agree that the section currently ends too abruptly and does not adequately explain the results. We will rewrite the unclear passages, expand the discussion of all results presented in the section, and revise the figures to improve their readability.

***Conclusions**:*

Line 380-382: '…spatially extensive cooling in the years following the eruptions. The strongest temperature anomalies occur within the first one to five post-eruption years, while weaker but still discernible cooling persists for up to about 15 years.'

This is not a new finding and should be discussed in relation to previous studies.

Line 405: '…confidently confirm the extremely harsh climatic conditions in the 540s within our study area.'

Is a one-degree cooling in the study regions really 'extremely harsh'? The authors should put the findings in a wider context.

Line 408-411: '…and likely, a weakening of the Indian monsoon influence. Reduced incoming solar radiation after the eruptions cools the surface and thus weakens land-sea thermal contrasts. In summer, this likely reduces the influence of the Etesians over the Mediterranean and favors a greater influence of maritime air masses, thereby increasing humidity and precipitation in MED.'

The authors do not show this in this study. Is this theory based on previous studies? These should be cited!

Line 425-426: '…therefore offer an important basis for comparison with natural archives, documentary evidence, and historical observations.'

There are a lot of proxy reconstructions available. The authors could compare their study to these available archives to strengthen there findings.

Line 429: '…move beyond generalized narratives of volcanic cooling toward a more regionally differentiated understanding…'

Several previous studies have investigated the more regional impact of volcanic eruptions, including for the Near East and Mediterranean areas. These studies should not be ignored.

As a response to all comments regarding the Conclusions section: We will revise the complete section. To focus on the 'real' conclusions here, we will insert and enlarge the discussion in the new combined 'Results and Discussions' section, where we will also place our findings in the context of existing literature.