Response to comments:

Investigating hydroclimatic impacts of the 168-158 BCE volcanic quartet and their relevance to the Nile River basin and Egyptian history

Singh et al. 2022.

We are grateful to the reviewers for their comments and suggestions for improving the manuscript. We believe these have helped us further improve the manuscript. We present below the respective reviewer comments in bold, our responses in red, and use italic font when quoting text added to the revised version of the manuscript.

Reviewer 2: -

Comment 1.0. **The manuscript in its first version was already in a quite mature state – the revised version includes all the comments raised in the first review in a very comprehensive way. A big asset is the re-structuring of different parts of the manuscript including some modifications of the presentation and interpretation of the statistics related to the ensemble simulation approach used in the study. Therefore I congratulate the authors for the manuscript and suggest to publish it after addressing some minor comments listed below.**

Response 1.0. We thank the reviewer for this assessment.

C1.1. The first comment relates to the former comment of l. 514:

Conceptually, all ensemble members are equally accurate, given they are forced with the same set of external forcings. A thought experiment might be if one could estimate the outbreak of a volcano based on the single ensemble members and reconstruct volcanic activity. As one can see in the additional rainfall anomaly plots ensemble member 01, the event E01 shows a comparatively large response, whereas member 02 shows little or no response at all (even in the presence of the very strong E01 eruption). This might be also the case for a future volcanic eruption that despite a (clear) simulated ensemble mean response there is a chance on the regional scale for little or no response at all. Maybe the authors could just add some words to motivate a more nuanced view on the response of hydrological changes on even large volcanic eruptions, especially on the local-to-regional scale and the according disadvantages [of] only investigating ensemble mean statistics.

R1.1. In our first round of revisions, we inserted several additional sentences based on these results, and relevant to the reviewer's comment, in the discussion sections. We have amended these sentences for clarity. See line 515, page 24 of the revised manuscript:

"Any individual ensemble member might best represent the historical reality, but it is impossible to select the most accurate member absent supporting observational data from the period. Also, added noise due to natural variability can be greater at the regional scale, even to the extent of

altering the sign of observed changes among the individual ensemble runs. Thus, we mainly focused upon the mean from across the ensemble when examining the response to the eruptions for the various climate variables considered"

We certainly agree with the substance of the reviewer's comment, and have additionally added the following text in our conclusion section (Page 38; Line 811).

"However, we note that particularly on smaller spatial scales, as examined here, the variability in the modelled response as observed across our individual ensemble members may reduce the representiveness of the mean. The notable variability on display across our individual ensemble members, even to the quite substantial forcing represented by E1, also suggests that hydroclimatic responses on local to regional scales may depart from broader regional or hemispheric averages even after quite large volcanic forcings.

C1.2. The second comment is related to Line 743-785 In the Revised Manuscript:

The 1 sigma ctrl level is a very poor statistical threshold to argue for the statistical significance/confidence. Therefore the 2 sigma level, representing approximately the 95% confidence interval should be used. It is always advisable to rigorously test a hypothesis based on an a-priori set threshold of statistical significance instead of re-defining thresholds after the analysis is carried out to fit results to the according hypothesis and/or lines of argumentation. Therefore the presentation of results is now better, but implicitly also reveals the insight that the smaller northern hemispheric eruptions E02-E04 show a remarkably different response compared to the larger tropical eruption E01.

R1.2. We have now modified the discussion starting at line 699 to remove the discussion pertaining to the 1σ level, keeping the discussion for the 95% (±2σ) only. We have also modified plot 12 to remove the line representing the 1σ, as shown below. We have additionally added the following sentence added in the conclusion at line number 699.

"It is evident that the mean surface temperature response in the northern hemisphere is significant at the control period's $2\sigma_{ctrl}$ level (95% significance). However, while rainfall and river discharge responses are not significant at the $2\sigma_{ctrl}$ levels, several individual members do show significance at $2\sigma_{ctrl}$ as well. This signifies the important influence of the model's internal variability in representing the regional hydrological response to volcanic eruptions."

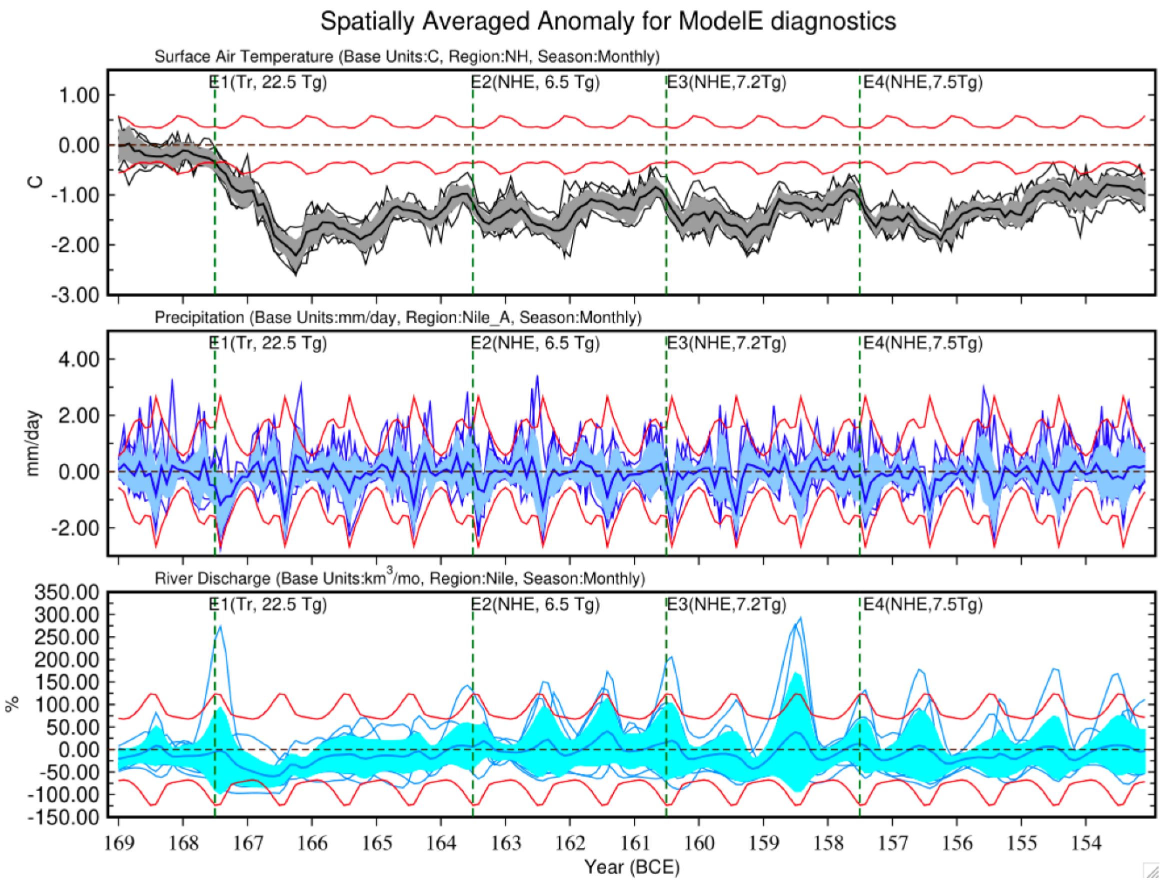

Fig 12: Monthly time series of individual ensemble and mean of surface temperature response (˚C) averaged over northern hemisphere (NH) (top panel), rainfall change (mm/day) for the model's spatial box representing the Nile River watershed (Latitude: (5N, 18N), Longitude: (30E, 42E)) (middle panel) and Nile River discharge anomaly (%) at the delta region (grid box centered at 29.0N, 31.25E). For each panel, the darker solid (thick) line shows the multi-ensemble mean, individual member (thin line), and the color envelope shows the associated variability (±σ; Standard deviation). The annual cycle of climate variability of the control run is shown as $2\sigma_{ctrl}$ lines (red solid line) along the x-axis for all three variables. The vertical dotted green line shows when each eruption happened.

Reviewer 3: -

Comment 1.0. The authors' changes have addressed most of my concerns in the first draft. The additions starting at line 275, in particular, help clarify the scope of the article.

Response 1.0. We thank the reviewer for their assessment.

C1.1. Nevertheless, the discussion of historical causation could still use improvements. In particular, the authors face a central problem that the correlation between the timing of eruptions and the timing of uprisings in Ptolemaic Egypt fits two scenarios equally well: (1) the occurrence (rather than non-occurrence) of eruption-induced droughts was necessary

(and/or sufficient) for the occurrence (rather than non-occurrence) of uprisings, or (2) the timing of eruption-induced droughts in some years (rather than droughts occurring in other years) was necessary (and/or sufficient) for the timing of uprisings in some years (rather than uprisings that would have taken place anyway in other years). This is not a fatal problem with the article. However, it needs to be openly admitted and addressed. The authors do not need to determine which scenario was the case, and they may clarify how the findings of this study add some credibility to scenario (1). However, they need to acknowledge that both scenarios remain possible and that some historical evidence points to (1) while other historical evidence and circumstances point to (2).

R1.1. We appreciate the reviewer's efforts in outlining and explaining their position on the importance of (and approaches to) assessing historical causality. This is valuable (and again we are happy that *Climate of the Past* has an open peer review system whereby the reviewer's commentary can be accessed and cited). We are similarly happy to amend our Introduction to explicitly cite the model of necessary and sufficient condition (described simply) as useful for further interrogating and thinking about causality, and something that readers should be aware of.

See now Lines 143-145 in the Introduction (of the clean manuscript; line numbers will differ in the Track Changes version), which state:

White and Pei (2020) argue that such questions represent a key challenge for climate historians and related scholars, recommending a framework wherein potential causes are assessed using a framework of necessary and sufficient conditions (put simply; see also Ludlow et al. (2023)).

See also Lines 857-860, which close the paper:

Relatedly, open questions remain as to where along the spectrum from proximate to ultimate causality (as per Gao et al., 2021) or necessary and sufficient conditions (as per White and Pei, 2020) hydroclimatic shocks lay in contributing to the revolts and other societal stresses that feature so prominently in Ptolemaic history.

We argue that to go much further in discussing the intricacies of causality is, however, beyond the scope of the present study, which was never intended to engage in a full discussion of how historical causality should be assessed, much less to attempt to determine the precise nature of that causality (as, to be fair, the reviewer notes). To do so would unfairly divert the reader from the considerable effort and time that expended in the modeling exercise, presented as the core of the paper. It is true that this modeling has been conducted in pursuit of a greater insight into the likely hydroclimatic consequences of the eruptions between 168-158 BCE, which were chosen not only for their scientific interest in terms of apparent magnitude and close temporal spacing, but because they coincided with a period of pronounced political turmoil in Ptolemaic Egypt. But it is for other work (already in progress) to make use of the modeling insights presented in this paper to deliver a dedicated analysis of the role of these volcanically induced hydroclimatic shocks in the incidence of any given social phenomenon (like revolt) of the period.

We do, however, explicitly note (as recommended by the reviewer) that the nature of any underlying causation is not settled: See Lines 139-141:

Much work remains to further characterize this causality, how direct or indirect it may have been, and whether this changed meaningfully through time (and between revolts that varied in geography and scale) according to (or in interaction with) other coincident potential causes (from longer term developments promoting chronic vulnerabilities, to more acute political and socioeconomic stresses).

C1.2. I would particularly encourage revisions in two sections:

The authors' discussion of causation and correlation starting at line 183 is confusing. It is true that historians frequently resort to the platitude "correlation isn't causation." That platitude often applies to situations in conventional history, where a correlation between phenomena A and B might be explained away by some set of factors (C, D, E, etc.) that influence both A and B, rather than any connection between A and B themselves. However, in much climate history, where there are no hidden variables influencing, e.g., both social unrest and volcanic eruptions, that problem doesn't apply and we need to drop the platitude altogether.

R1.2. We appreciate the reviewer's reasoning here, and have amended our relevant discussion to better frame and justify the inclusion of correlation vs. causality. We note that because the topic is familiar to many readers it thereby acts as an effective window into a brief discussion of causality, as well to introduce the work done to date by Ludlow and Manning (2016) and Ludlow et al. (2017) in establishing the statistical significance of the eruption-revolt association.

In Lines 124-135, we now state:

It is a truism that correlation does not establish causation. Genuine causality is, however, implied where significance testing suggests an observed correlation is unlikely to have arisen randomly, though this does not determine the direction or character of causality (Izdebski et al., 2023). Statistical significance may, moreover, be sensitive to many factors. These include here (1) the choice of statistical test, (2) the choice of revolt and eruption dates (if uncertainties exist), (3) judgements as to what constitutes "revolt" (vs. phenomena like food riots motivated more by desperation than politics), and (4) judgements concerning which eruptions to include in testing (e.g., seeking those with a meaningful impact vs. non-climatically effective eruptions introducing "noise" into the analysis), assessed by estimated volumes and heights of atmospheric SO2 injections, eruption locations, and more. Notably, thus, testing by Ludlow and Manning (2016) was followed by Manning et al. (2017) who also observed a statistically significant coincidence between eruptions and Ptolemaic-era revolts using different methods and variant dates.

C1.3. In this study [it] would be more accurate to say that this correlation does imply causation—yet the nature of that causal linkage remains unclear. Absent further research and clarification, we cannot say, for example, whether these uprisings might have occurred without the eruptions but perhaps in a different manner or in a different year. Moreover, even if the eruptions and Nile failures were a necessary condition for the uprisings, we would need further information and arguments to determine whether the eruptions should be deemed "the cause of" the uprisings. For example, if the Ptolemaic regime were especially vulnerable to Nile failures at this time, while another regime would have endured similar natural shocks without popular unrest, then it may be more appropriate to label those societal vulnerabilities "the cause of" the uprisings instead.

The discussion of causal "pathways" does not necessarily address this problem and could be even more misleading. After all, almost every causal pathway could be broken down into additional causal mechanisms ad infinitum. The length or complexity of the "pathway" per se doesn't actually change how we determine causation. I can't run over someone in a car and then claim that my actions didn't cause their death because really there was this whole complex pathway between pressing my foot on the accelerator, the motor running, the car moving, the impact, injury, blood loss, etc. What matters, as discussed in the review of the first version, are determinations of the appropriate contrast set in cause and effect and the strength of causal necessity and sufficiency.

R1.3. We appreciate the reviewer's continued argumentation and welcome the opportunity to engage with it in more depth here (given that we feel this would be too much of a diversion with the focus on modeling in the main text of the paper itself).

To begin, we agree that based upon the statistical testing of Manning et al. (2017) that the repeated coincidence in timing between historically dated revolts and ice-core-based eruption dates suggests causality. As we would describe it, the "direction" of the causal "pathway" cannot credibly flow in reverse here (i.e., revolts do not credibly cause eruptions). We have amended the text of our Introduction to emphasize this more clearly.

Lines 136-139 thus note (following on directly from the excerpt cited in our previous answer):

Logic dictates that the direction of any genuine causality must flow here from eruption to revolt (Izdebski et al., 2023). Further confidence in its reality arises from the existence of plausible mechanisms connecting volcanic hydroclimatic variability with revolt (i.e., via reduction of the Nile summer flood and consequent societal impacts).

Even in this case, it is not actually so straightforward to comprehensively reject concerns over correlation vs. causality. Doing so depends upon accepting the central finding of Manning et al. (2017), in which the coincidence in timing between revolts and eruptions is held to be real (i.e., causal) on the basis that it appears non-random (statistically significant), with the margin for error in terms of this statistical significance being deemed acceptably small (itself a value judgment that may vary by individual). Given that an individual's judgment as to whether the observed coincidence in timing (correlation) is likely to arise from actual causation will depend in large measure upon the level of statistical significance observed, it is critical to note that this might change meaningfully depending upon (1) the type of statistical testing chosen (there are always alternative tests possible), (2) the set of revolt dates chosen in cases where dating uncertainty exists, (3) whether it is accepted that all events considered by Manning et al. (2017) should be taken to be the same type of phenomenon for the purpose of testing (after all, each revolt will have been to varying degrees unique in circumstances, severity, geography, duration, motives, outcomes, etc., so that other revolt groupings may be legitimately proposed), and (4) what eruption dates should be included as relevant for the purpose of this testing (e.g., which eruptions were likely, in principle, to have a meaningful hydroclimatic impact as based upon their apparent magnitudes, locations, etc., and not simply introduce "noise" to any analysis). We have noted these issues more briefly in the main article text, as per the excerpt provided earlier (Lines 124-135).

One critical aid to any assessment of correlation vs. causality (in light of the above) is the delineation of causal pathways that can credibly link proposed/potential causal forces or factors (like explosive volcanism) to a complex societal phenomenon (like revolt). As Manning et al. (2017) propose, one of several (non-exclusive and simplified) possible causal pathways linking revolt to explosive volcanism involves the demonstrable dependence (for its agricultural fortunes and even political stability) of Ptolemaic Egypt upon a sufficient Nile summer flood (noting that what was "sufficient" will certainly have varied in time – a point the reviewer touches upon in a related context - and which we emphasize in Lines 140-144), and the ability of large explosive eruptions to diminish the level of summer flooding. It is the intention of the present paper to provide insight into the extent to which historical eruptions like those observed between 168-158 BCE might have impacted the flood, and hence contribute to more detailed assessments of Egyptian environmental history (including causality). We emphasize this on Lines 186-189:

Here, we intend to advance our understanding of the likely hydroclimatic impact of the 168-158 BCE eruption quartet as a foundation for ongoing efforts to more securely establish and qualify the causality underlying the observed association between eruptions, Ptolemaic-era revolts and other political and socioeconomic phenomena and developments.

As is evident from our response, we hold that referencing casual or contributory pathways is a useful way to conceptualize causality (quite common in the contemporary climate-conflict literature and thereby a useful reference point for readers coming from this field) and for thinking about the dependencies between different historical phenomena, including in our own case. Whilst we accept the reviewer's point that one could add additional contributory mechanisms "ad infinitum", we would argue for the utility in an exercise and framework that attempts to identify and order multiple potential causal / contributory forces and factors in trying to explain human history. Moreover, we do not see this as mutually exclusive to the framing of necessary and sufficient conditions favored by the reviewer. Thus, on Lines 151-156, we now state:

An alternative framing in many climate-conflict studies (not incompatible with that proposed by White and Pei (2020) or employed by Gao et al. (2021)) is to delineate multiple identifiable "pathways" that may enable or lead (through material (economic), political, cultural or psychological channels) to links between hydroclimatic variability and various forms of conflict (see Hsiang and Burke (2014) and Ide (2017) for reviews).

For the record, we have discussed the importance of many such considerations as they pertain to causality (in a specific case study on Ptolemaic Egypt) in a paper already in-press: Izdebski, A., Bloomfield, K., Eastwood, W. J., Fernandes, R., Fleitmann, D., Guzowski, P., Haldon, J., Ludlow, F., Luterbacher, J., Manning, J. G., Masi, A., Mordechai, L., Newfield, T., Stine, A. R., Senkul, C. and Xoplaki, E. (In Press, 2023), "The Emergence of Interdisciplinary Environmental History: Bridging the Gap between the Humanistic and Scientific Approaches to the Late Holocene," *Annales*, 77 (2), 1-48. We have also applied different tests using different lists of eruptions and revolts and found a continued statistical significance in the observed coincidence between revolt dates and eruption dates. See Ludlow, F. and Manning, J. G. (2016) "Revolts under the Ptolemies: A Paleoclimatic Perspective", In: Collins, J. J. and Manning, J. G. (eds.), *Revolt and Resistance in the Ancient Classical World and the Near East: The Crucible of Empire. Culture and History of the Ancient Near East Series*. Leiden: Brill, 154-171.

C1.4. I would also discourage the use [of] "proximate" vs. "ultimate" causation in lines 1303-1305 for the following reasons. First, the terms themselves are confusing and ambiguous. Without further explanation, many readers would assume that an "ultimate" cause should be somehow more fundamental than a "proximate" cause—exactly the opposite of how they are used in Gao et al. Second, the long time-series and abundant data for disasters and political events in imperial China enable inferences and analysis that just aren't possible (yet) for a case like Ptolemaic Egypt. Third, and most important, the spectrum between "ultimate" and proximate" fails to capture the central problems regarding causation in the case of Nile failures and uprisings in Ptolemaic Egypt. As explained in the review of the first draft, these problems are essentially two-fold. On the one hand, we cannot say (yet) whether these Nile failures were necessary or sufficient for the uprisings to occur at all, or only whether they were necessary or sufficient for the timing or perhaps character of social and political turmoil that was going to occur sooner or later anyway. On the other hand, even if eruption-induced Nile failures were necessary or sufficient for these events, we cannot say (yet) which condition was more exceptional and relevant and therefore appropriately labeled "the cause of" the uprisings: the degree of Nile failure or the particular vulnerabilities of Ptolemaic regime.

I would strongly encourage the authors to spell out these issues of historical causal explanation plainly and clearly.

R1.4. As noted earlier, following the reviewer's guidance, we have added text highlighting the framework of necessary and sufficient conditions to the Introduction, where it is presented as one useful framework for further interrogating the issue of causality. But given that the matter of how to assess causality is hardly settled in or between various disciplines, we argue (similar to the “pathway” framing) that we should keep the reference to “ultimate” versus “proximate” causality. Beyond Gao et al. (2021), it is finding increased use as a framework to further interrogate or characterize historical causality (see, e.g., Brian Villmoare, *The Evolution of Everything: The Patterns and Causes of Big History* (Cambridge University Press, 2022), in which ultimate and proximate causality are employed as central concepts - we have now cited this in the paper). It may well be that one approach or framing can ultimately be demonstrably shown as better than another, but the present paper is not the place to discuss this in depth. Rather, in our revisions, we now offer a range of such framings, for the reader to further examine.

We do take the reviewer’s point that some readers may assume that “ultimate” should be taken as the more important form of causality and have clarified that in the existing text. See now Lines 145-150:

Gao et al. (2021) employ a framework wherein the role of volcanically induced hydroclimatic “shocks” in the collapse of Chinese dynasties is characterized along a spectrum from “ultimate” to “proximate” causation (see also Villmoare (2022) for this framework). Here, smaller hydroclimatic shocks could act as the ultimate cause of collapse when enabled by high pre-existing stress, while larger shocks could act with greater independence as proximate causes without substantial pre-existing stress.

1.4. Finally, I would encourage the authors to tighten the language to improve readability, particularly in sections that have been added since the first draft (e.g., lines 180-183). The article is still unusually long. While the complexity of the topic does merit extra space, I believe it could be at least a page shorter simply by improving the style and removing ambiguities and redundancies.

We agree that greater concision can be achieved and have made many further excisions and tweaks to streamline the text for clarity and size (please see the Track Changes version of the manuscript). This includes the deletion of several paragraphs outright (e.g., see the Discussion and Conclusions section).

Regarding length, we note that addressing the issue of causality has itself contributed to a longer article, but feel the additions and overall length are justified here because (1) the interdisciplinary subject matter requires more context to guide its different potential audiences, (2) there are no set word limits (that we are aware of) in this journal, and (3) this paper is certainly not the longest of those already published in this special issue of *Climate of the Past*.