RC1:

Hennekam et al. provide a new calibration of XRF-CS derived Ti/Al measurements from Mediterranean core ODP 967 – a key site for the study of Plio-Pleistocene Saharan climatic variability. This is an important record which provides additional evidences for the timing and intensity of wetter/drier periods in the Sahara and the potential global/orbital controls of these fluctuations. The article is well written, with little grammatical revision required to the main body of the text. I believe this article asks two key questions: 1) how best can non-destructive and destructive geochemical methods be combined to provide an accurate record of past climatic variability? And 2) what can this new record inform about the long-term orbital influences on Saharan climatic variability throughout the Early Pleistocene to Mid Pleistocene?

The major strength of this manuscript is that it offers a valuable method to mitigate loss of material though WD-XRF analysis by instead selecting fewer (1060) samples to calibrate a non-destructive XRF-CS record (8497 samples). This permits a higher resolution Ti/Al record to be produced. However, I have a few concerns with this section.

Reply: We thank the reviewer for their positive and constructive comments on our work. Below, we reply in detail to the comments of this reviewer.

I believe the authors would benefit from emphasising the novelty of their study more clearly. Currently, on the basis of the text, it does not seem entirely clear how this calibration and XRF-CS record differs from that of Grant et al. (2022). Did the authors obtain new Ti/Al measurements? Or did they use those of Grant et al. (2022)? Similarly, did Grant et al. (2022) use the same WD-XRF dataset (Konijnendijk et al. 2014, 2015) to calibrate their record? Is this study using the same data and method as Grant et al. (2022), and simply testing how many samples are needed for accurate calibration? The authors must make the last two paragraphs of the introduction (and the materials and methods section) much clearer so that readers can establish the data output of this study.

Reply: This study and the study of Grant et al. 2022 were executed in parallel and hence there is indeed overlap between the studies (i.e., similar calibration approaches) but also important differences that merit a separate publication. First, there is a misconception that proper calibration of XRF-scanning data is only necessary to quantify the geochemical data. We show that appropriate calibration also allows to significantly improve (i.e., making it consistent with other established geochemical methods) the capturing of down-core geochemical variability. This will also make it possible to revisit old intensity datasets that were deemed unusable and extract useful paleoenvironmental data. We think it is important to explicitly describe our methods, including appropriate statistical testing, in a paper (and not just a supplement), as it is pivotal knowledge for many studies to come. We will highlight (in the last two paragraphs of the Introduction, but also in the Abstract and Conclusions) more clearly that the misconception about quantified XRF-scanning exists and that accurate calibration can much improve the capturing of geochemical variability. Moreover, we will clarify, shortly, in the Introduction the differences with Grant et al. (2022) and we will add a paragraph within the Methods section that will describe similarities and differences between Grant et al. (2022) and this study in detail. The calibration set of Grant et al. (2022) also used the WD-XRF dataset of Konijnendijk et al. (2014, 2015). Second, Grant et al. (2022) consider the full 5 Myr scanning XRF records from ODP967 in conjunction with other new proxy records from the same samples (stable isotopes and environmental magnetism), with a particular focus on the geochemical and lithological shift at 3.2 Ma, while we here focus on the 2.3-1.2 Myr interval. The available Ti/Al records (De Boer et al., 2021; Konijnendijk et al., 2014; 2015; Lourens et al., 2001) showed that the North African climate system seemed to behave differently after 1.2 Ma compared to before 2.3 Ma. Yet, that left a knowledge gap on the operation of the African climate system between 2.3-1.2 Ma, which we here for the first time address in detail. We will amend the Introduction to more clearly highlight this novelty.

The results table (Table 1). Instead of a Y or N value to indicate whether the null-hypotheses have been rejected, the authors should provide the P-value and test specific values. This could be included in supplementary material rather than the main text, but they must be accessible for researchers. Additionally, the authors need to account for the “multiple comparison problem” by adjusting the ð• ‘ Ž value

Reply: We will add relevant test results to Table 1, with a focus on the *p* values. To correct *p*-values for multiple comparisons, we will use the Bonferroni method (i.e., $p <= 0.05$ / #tests). We already applied this correction to our dataset and this will lead to one adjustment in Table 1 (see adjusted Table below). Specifically, the *p* value of the one-way ANOVA for Fig. 2e (10% calibration samples) is 0.0412 and hence this indicates that statistically, with the Bonferroni correction, the means of the data obtained with this XRF-scanning calibration and the WD-XRF data are not significantly different (= similar; a “Y” in the table). We will adjust the Methods section accordingly (i.e., the part that describes the statistical approach).

	Correlation r	Equality of variance (F-test)	One-way ANOVA	Student t-test	Non-parametric Mann-Whitney
Fig. 2a: Ti/Al intensities and Ti/Al WD-XRF	0.34	N (<0.0001)	N (<0.0001)	N (<0.0001)	N (<0.0001)
Fig. 2b: Ln(Ti/Al) intensities and Ln(Ti/Al) WD-XRF	0.30	N (<0.0001)	N (<0.0001)	N (<0.0001)	N (<0.0001)
Fig. 2c: Ti/Al Grant et al. 2017 calibration (n=45) and Ti/Al WD-XRF	0.39	N (<0.0001)	N (<0.0001)	N (<0.0001)	N (<0.0001)
Fig. 2d: Ti/Al all calibration samples (n=1060) and Ti/Al WD-XRF	0.74	N (<0.0001)	Y (0.4383)	Y (0.4817)	Y (0.9299)
Fig. 2e: Ti/Al 10% calibration samples (n=106) and Ti/Al WD-XRF	0.68	N (<0.0001)	Y (0.0412)	Y (0.0701)	Y (0.4184)
Fig. 2f: Ti/Al 5% calibration samples (n=53) and Ti/Al WD-XRF	0.68	N (<0.0001)	Y (0.2573)	Y (0.2813)	Y (0.6464)
Fig. 2g: Ti/Al 2% calibration samples (n=22) and Ti/Al WD-XRF	0.60	N (<0.0001)	N (0.0077)	N (0.0101)	N (<0.0001)
Fig. 2h: Ti/Al AvaaXelerate calibration samples (n=53) and Ti/Al WD-XRF	0.62	N (<0.0001)	N (0.0006)	N (0.0010)	N (<0.0001)
Fig. 2i: Ti/Al AvaaXelerate calibration samples (n=22) and Ti/Al WD-XRF	0.61	N (<0.0001)	N (<0.0001)	N (<0.0001)	N (<0.0001)

It is necessary for the authors to better explain why 53 samples are required for accurate calibration, and why, if this is sufficient, the 1060 sample calibration record is favoured for the subsequent discussion. I understand that it is necessary to reduce the number of samples to achieve the authors aims. However, I believe the justification for this amount is unclear as the test specific results have not been made available.

Reply: Indeed, 53 samples seem to be appropriate statistically, as the tests indicate similarity of the means between the XRF-scanning data and the XRF-bead data (we will provide this data in a new Table 1, see above) and a relatively high correlation coefficient r. However, this also indicates that the 1060 sample calibration performs best (i.e., highest *p* values of the tests and highest correlation r), which is why this calibration is favored for further use here. When such an extensive sample set would not be available already a 53 sample calibration set would be enough. We will clarify this in the Discussion.

For the high-resolution XRF-CS Ti/Al analysis and correlation to orbital records, I would like to first say that I am generally supportive of this analysis. The authors provide a detailed insight into the varying controls of orbital parameters on African wetter/drier periods. Unlike hematite dust transport, Ti/Al ratios provide a method to study the intensity of wetter/drier periods. Their statistical analysis and interpretation, that high-latitude forcing played an increasingly dominant role after the Mid Pleistocene Transition, appears reasonable and well argued. However, I believe this section needs further work and clarification/justification.

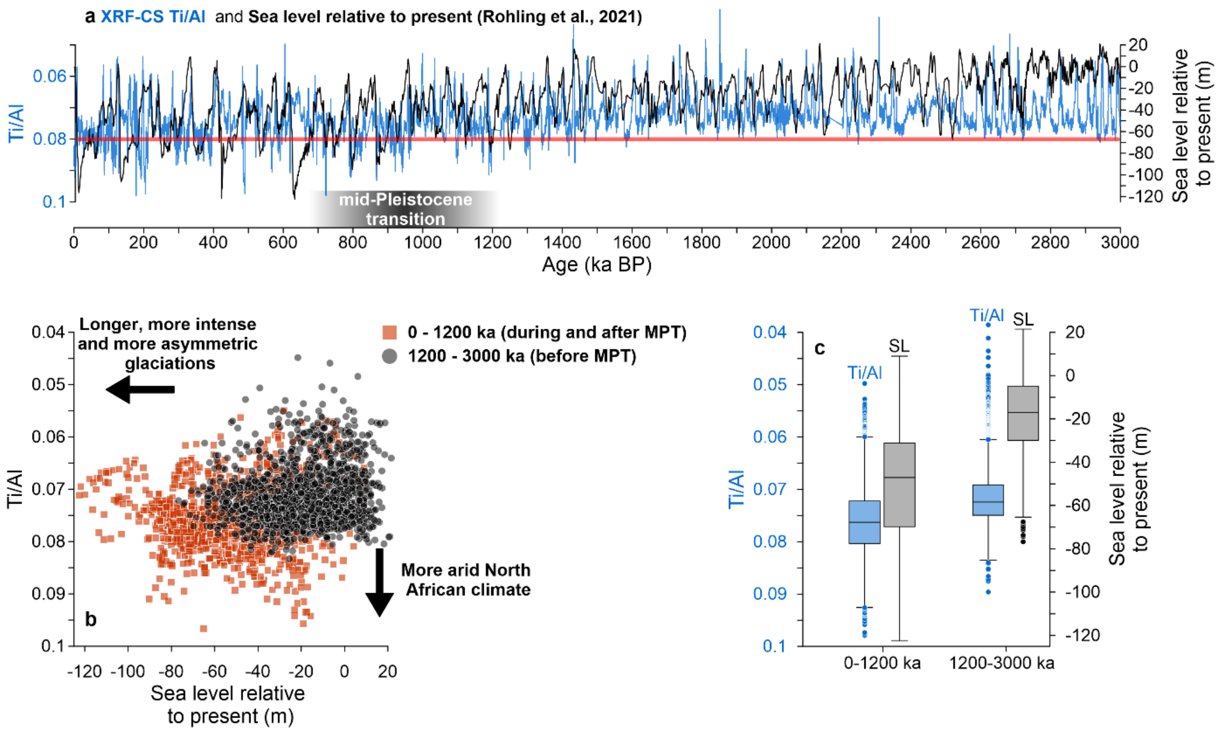
Firstly, the application of a 401 kyr window running correlation (long eccentricity band), based on the text, does not seem justified to the reader. Why was this running correlation window selected? The authors must explain why such a large window is necessary and crucial to their analysis and interpretations.

Reply: The 401-kyr window was chosen to obtain a relatively smooth running correlation record that focuses on long time-scale changes in variability. If the window is set to much shorter values (e.g., 100 kyr), then the smaller discrepancies between Ti/Al and insolation become more apparent, as fewer datapoints are involved. The ChangePoint statistics (see below), for instance, would probably indicate many more change points. With the 401-kyr window these statistics only focus on the largest changes in the record. We will justify and explain this in more detail in the Methods section of the revised version of the paper.

Secondly, as can be seen from the very well-made figures, the 95% confidence intervals (while they do represent extremes) are large and, considering this, there is some uncertainty when distinguishing the shift from low to high running correlation between >1.2 and <1.2 Ma. This is more of an issue for the correlation with sea-levels. Additionally, the claim for constant high correlation with sea-level after the MPT is not so clear; it appears that higher correlations exist from about 1.7 Ma, with an abrupt dip at 1.1 Ma, after which the high correlation returns. Perhaps the authors could perform a t-test of running correlation values between these two periods to test for significant differences?

Reply: Based on the comments of both reviewers (RC1 and RC2) and recent updates on sea-level proxy records, we have removed the running correlation (Fig. 5c in current manuscript) between Ti/Al and sea-level change at Gibraltar (RSL_{Gib}) from the revised version of the paper. The latest sea-level review produced by several co-authors involved in our study (Rohling et al., Submitted) shows that, compared to other sea-level proxy records, the RSL_{Gib} deviates quite considerably before ~1.5 Ma. We therefore investigated the ODP967 Ti/Al running correlation with the recent Rohling et al. (2021) sea-level record based on deconvolution of deep-sea benthic foraminiferal $\delta^{18}O$ records, even though the latter age models might differ somewhat with our ODP967 record. However, we found that the resultant running correlation remained close to 0 within uncertainties. In light of this, and the concerns of both reviewers about the weak/variable correlation between Ti/Al and sea-level, we will now present a straightforward cross-correlation between sea-level and ODP967 Ti/Al values older/younger than 1.2 Ma (new Figure 5b – see below) and box-whisker plots of the same values (new Figure 5c).

Considering these new plots with our change-point and wavelet analyses, we believe that the evidence suggests a high latitudinal impact on North African climate around the MPT, despite a weak running correlation of Ti/Al and sea-level. For example, (1) wavelet analysis of the Ti/Al record (Fig. 4a) shows a strengthening of wavelengths >100 kyr at the MPT, similar to high-latitude records; (2) Change point analysis highlights that indeed a statistical change occurs in Ti/Al at the MPT; (3) new Fig. 5b shows that large amplitude changes in both SL and Ti/Al share at least their timing, albeit nonlinearly; (4) The mean (t-test) and variance (F-test) is significantly different for both Ti/Al and sea level before and after MPT (new Fig. 5c; we will add these statistical results to the caption and text; all *p* values are <0.0001). We will slightly adjust the end of the discussion to include these points, and we will remove the text about the running correlation.



Furthermore, both the correlation with insolation (is this SITIG, 65N, 35N or 15N? Please clarify on figures) and sea-levels timing may benefit from further investigation using ChangePoint analyses. If using the R statistical software package, this can be achieved with packages such as BCP or ChangePoint. This may result in slightly different ages identified for these changes, but combined with the current analysis, would add an additional line of support to the authors argument. In either case, I believe that, while there is a deal of statistical uncertainty, the authors analysis provides important information for understanding the orbital controls on Saharan wetter/drier periods throughout the Pleistocene.

Reply: We will clarify in the figures and captions that insolation means SITIG in this case. We thank the reviewer for this comment, because the ChangePoint analysis is indeed a great addition to the paper, which we will implement. Our new ChangePoint analysis on the correlation of Ti/Al and insolation identifies changes at: 317 ka, 1081 ka, 1404 ka for the mean and 286 ka, 1114 ka, and 1404 ka for the standard deviation. Indeed, most of these ages fall within or just prior to the MPT.

While I am supportive of their analysis, the authors may benefit from additional reference to various studies which describe the suppressive effects of glacial termination melt-water

discharge on low-latitude forcing during the Middle and Late Pleistocene, causing monsoon intensification to lag insolation (e.g., Marino et al., 2015; Menviel et al., 2021; Häuselmann et al., 2015; Böhme et al., 2015). While most of these studies are limited to the LIG or Holocene, this may provide an additional line of support for some of the authors arguments.
Reply: We will add the suggested literature (Böhm et al., 2015; Häuselmann et al., 2015, Marino et al., 2015; Menviel et al., 2021) – and we will slightly expand the text – at the end of the Discussion to further support the suppressive effects of glacial termination meltwater discharge on the North African monsoon system. This will aid in explaining the change in the phase relationship between (low-latitude) insolation and monsoon intensity during the late Pleistocene.

I recommend that this paper be published in Climate of the Past subject to the authors addressing the concerns and the few grammatical/technical notes below. I suggest minor revisions as 1) results of the statistical testing and consideration of the "multiple comparison problem" (this may have some impact on the results, but is difficult to estimate without seeing the test specific results); and 2) the interpretation/discussion needs further analysis and justification to support these arguments, and currently the novelty is not well emphasised. However, I believe that this work will make a valuable contribution once these concerns are addressed.
Reply: We again thank the reviewer for the constructive comments and hope that our proposed changes will take away any concerns.

Technical/grammatical notes:
Line 37-39: References. The authors may benefit from adding a few references to palaeoanthropological/archaeological outputs and discussions, that are not necessarily climatic research initiatives, to highlight the broader relevance of their work. (E.g., Potts et al. 2020; Groucutt et al. 2015)
Reply: We will add Potts et al. (2020) and Groucutt et al. (2015) to the literature already cited within these lines.

Line 58-86: The last two paragraphs of the introduction. I believe these paragraphs are, in short, saying "As WD-XRF is destructive, how many samples are required to accurately calibrate a non-destructive XRF-CS record?". The authors may benefit from revising these paragraphs to emphasise the aims of the manuscript more concisely (or perhaps directly). Maybe this is due to my unfamiliarity with the methods, but it took me a few attempts to work-out the novelty of this article, as Grant et al. (2022) is described as having conducted a very similar WD-XRF calibration of an XRF-CS record for the 5 Ma period of the core. The paragraphs must emphasise the novelty of this study more clearly.
Reply: As explained in our above responses, we will highlight the novelty more clearly within the Introduction, and will discuss the differences/similarities with Grant et al. (2022) within the Introduction and in more detail in the Methods section.

Line 172: There have been various comments that WD-XRF analysis is more precise/better established than other methods. Can the authors provide further quantification of this?

Reply: The statement about a more precise/accurate measurement of WD-XRF is especially focused on its comparison to ED-XRF. The WD-XRF technique reaches simply a much higher spectral resolution than ED-XRF, which results in better results. We will clarify this in the text.

Line 223-224. The authors may wish to add a comment on the work of Tzedakis et al. (2017). Nature, 542: 427-432.
Reply: We will add a comment on the work of Tzedakis et al. (2017), stating that at this time (around the MPT) an increase in the deglaciation energy threshold likely resulted in glacial cycles with lower frequency and higher amplitude.

Table 1. Please include the results of the statistical tests either here or in supplementary material.
Reply: As discussed above, we will.

Fig. 2 and caption. "XRF-bead". Perhaps change this to WD-XRF-bead for clarity?
Reply: We will adopt this change.

Fig. 3 may benefit from the addition of correlation coefficients of the XRF-CS Ti/Al record and the respective humidity/aridity records from ODP 967.
Reply: We will add this to the figure and shortly discuss it in the Discussion.

Fig. 4g. Please clarify if insolation is the SITIG, 65N, 35N or 15N.
Reply: We will clarify in caption and figure that this is SITIG.

Fig. 5c. The figure may benefit from a dashed line running horizontally from 0. This would allow the reader to track changes more easily in the correlation.
Reply: This comment is no longer relevant as we will adjust Fig. 5.

Figures. (not necessary). The cyan text may benefit from being a few shades darker.
Reply: We will adjust the cyan text to a darker blue (including the lines), as in the figure above.

Literature

Böhm, E., Lippold, J., Gutjahr, M., Frank, M., Blaser, P., Antz, B., Fohlmeister, J., Frank, N., Andersen, M., Deininger, M., 2015. Strong and deep Atlantic meridional overturning circulation during the last glacial cycle. Nature 517, 73-76.

De Boer, B., Peters, M., Lourens, L.J., 2021. The transient impact of the African monsoon on Plio-Pleistocene Mediterranean sediments. Clim Past, 331-344.

Grant, K.M., Amarathunga, U., Amies, J.D., Hu, P., Qian, Y., Penny, T., Rodriguez-Sanz, L., Zhao, X., Heslop, D., Liebrand, D., Hennekam, R., Westerhold, T., Gilmore, S., Lourens, L.J., Roberts, A.J., Rohling, E.J., 2022. Organic carbon burial in Mediterranean sapropels intensified during Green Sahara Periods since 3.2 Myr ago. Communications Earth & Environment 3, 1-9.

Groucutt, H.S., Petraglia, M.D., Bailey, G., Scerri, E.M., Parton, A., Clark-Balzan, L., Jennings, R.P., Lewis, L., Blinkhorn, J., Drake, N.A., 2015. Rethinking the dispersal of Homo sapiens out of Africa. Evolutionary Anthropology: Issues, News, and Reviews 24, 149-164.

Häuselmann, A.D., Fleitmann, D., Cheng, H., Tabersky, D., Günther, D., Edwards, R.L., 2015. Timing and nature of the penultimate deglaciation in a high alpine stalagmite from Switzerland. Quaternary Science Reviews 126, 264-275.

Konijnendijk, T., Ziegler, M., Lourens, L., 2015. On the timing and forcing mechanisms of late Pleistocene glacial terminations: insights from a new high-resolution benthic stable oxygen isotope record of the eastern Mediterranean. Quaternary Science Reviews 129, 308-320.

Konijnendijk, T.Y.M., Ziegler, M., Lourens, L.J., 2014. Chronological constraints on Pleistocene sapropel depositions from high-resolution geochemical records of ODP Sites 967 and 968. Newsletters on Stratigraphy 47, 263-282.

Lourens, L.J., Wehausen, R., Brumsack, H.J., 2001. Geological constraints on tidal dissipation and dynamical ellipticity of the Earth over the past three million years. Nature 409, 1029-1033.

Marino, G., Rohling, E.J., Rodriguez-Sanz, L., Grant, K.M., Heslop, D., Roberts, A.P., Stanford, J.D., Yu, J., 2015. Bipolar seesaw control on last interglacial sea level. Nature 522, 197-201.

Menviel, L., Govin, A., Avenas, A., Meissner, K.J., Grant, K.M., Tzedakis, P.C., 2021. Drivers of the evolution and amplitude of African Humid Periods. Communications Earth & Environment 2, 1-11.

Potts, R., Dommain, R., Moerman, J.W., Behrensmeyer, A.K., Deino, A.L., Riedl, S., Beverly, E.J., Brown, E.T., Deocampo, D., Kinyanjui, R., 2020. Increased ecological resource variability during a critical transition in hominin evolution. Science advances 6, eabc8975.

Rohling, E.J., Yu, J., Heslop, D., Foster, G.L., Opdyke, B., Roberts, A.P., 2021. Sea level and deep-sea temperature reconstructions suggest quasi-stable states and critical transitions over the past 40 million years. Science Advances 7, eabf5326.

Rohling, E.J., Foster, G.L., Gernon, T.M., Grant, K.M., Heslop, D., Hibbert, F.D., Roberts, A.P., Yu, J., Submitted. Comparison and synthesis of sea-level and deep-sea temperature variations over the past 40 million years. Reviews of Geophysics. pre-publication version: https://www.essoar.org/doi/abs/10.1002/essoar.10510904.1

Tzedakis, P., Crucifix, M., Mitsui, T., Wolff, E.W., 2017. A simple rule to determine which insolation cycles lead to interglacials. Nature 542, 427-432.