Response to Referee 2

The manuscript of Hollstein et al presents a synthesis of sea surface temperature (SST) reconstructions from the low-latitude Indo-Pacific for the mid-Holocene, MIS 5e, and MIS 11c and a comparison with the CESM1.2.2 time-slice simulations for 6 ka, 127 ka and 409 ka. The main finding is that the proxy compilation indicates generally warmer-than-preindustrial SSTs during MIS 5e and MIS 11c and a smaller zonal SSTgradie nt across the equatorial Pacific, whereas the CESM simulations tend to produce cooler SST anomalies and a larger zonal gradient. Further sensitivity experiments show that the implementation of paleo-vegetation in CESM reduces the model-proxy discrepancies but does not improve the zonal SST gradient. Based on these results, the authors suggest that refined model physics and improved paleo-proxy integration are needed to better simulate tropical climate behavior during interglacials. It is an interesting study and addresses a challenging topic on proxy-model comparison over the tropical oceans. I hope the following comments and suggestions will be helpful for strengthening the manuscript and clarifying the interpretation of the main results.

We thank the referee for the positive evaluation of our manuscript and for the constructive comments, which will help to improve the manuscript. We will address the comments as outlined below and include the changes in the revised version of our manuscript.

1. The proxy SST anomalies are mainly calculated relative to 1850–1900 COBEv2 SST rather than to late-Holocene or preindustrial values derived from the same proxy records. This approach may introduce site- and proxy-specific offsets related to calibration, ecology, habitat depth, and seasonality. This issue may be important because the proxy–COBE offsets reported in Tables S2 and S3 are comparable in magnitude to the reconstructed MIS 5e and MIS 11c warming. The CE-based sensitivity test is useful, although it is necessarily limited to a subset of the records. It would therefore be helpful if the authors could further quantify how the main results change when only records with a same-proxy CE reference are considered. In particular, it would be useful to examine whether the regional SST anomalies and the WPWP–EEP gradient retain similar signs and magnitudes.

To control the robustness of our results, we additionally calculated SST anomalies relative to the Common Era (CE) using all records covering the CE (total of 36 Mg/Ca and 24 UK′37 records; Fig. S4 of the manuscript). Considering the reduced data coverage, we find that the overall outcome of this study does not depend on the choice of the PI dataset. Specifically, the Mid-Holocene shows a heterogenous pattern of SST anomalies across the low-latitude Indo-Pacific, while MIS 5e and MIS 11c are marked by positive SST anomalies. The available data also indicate a smaller zonal SST gradient across the equatorial Pacific Ocean suggesting a weakening of the Pacific Walker circulation. Thus, using the CE as reference period would not change the overall results or conclusions of this study. We will include this information in the revised version of the manuscript. However, due to the small number of records that cover the CE and the mid-Holocene (58 records), MIS 5e (14 records) or MIS 11c (13 records), it does not appear useful to calculate regional SST anomalies for the WPWP, SCS and EEP areas. For instance, none of the SCS records, only one WPWP record and two EEP records cover the CE and MIS 5e. We therefore refrain from calculating regional SST anomalies and zonal gradients with the CE reference.

2. The proxy SSTs are averaged over relatively broad intervals (4–8, 123–131, and 405–413 ka), whereas the CESM simulations represent fixed equilibrium states at 6, 127, and 409 ka. The proxy averages and model simulations therefore do not necessarily represent exactly the same climatic state. I wonder

whether the proxy averages can straightforwardly be interpreted as "conservative estimates of peak warmth". Temporal averaging may damp local SST maxima, but it may also combine different orbital states and different timings of local SST maxima among sites. It would therefore be useful to test the sensitivity of the reconstructed regional SST anomalies and zonal SST gradient to the width of the selected time windows, using narrower windows where the data resolution allows. In addition, the possible influence of chronological uncertainty in the proxy records may also be considered.

We did as suggested and repeated the reconstruction of regional SST anomalies and zonal SST gradients with halved widths of the selected time slices: 6 ka (5–7 ka), 127 ka (125–139 ka) and 409 ka (407–411) (see Figs. 4 and 5). We find that this does not affect the overall results or outcome of the study. We will include this information in the revised version of the manuscript. The referee is correct that our phrasing that the reconstructed SST anomalies represent conservative estimates of peak warmth is at least misleading and we will correct this statement.

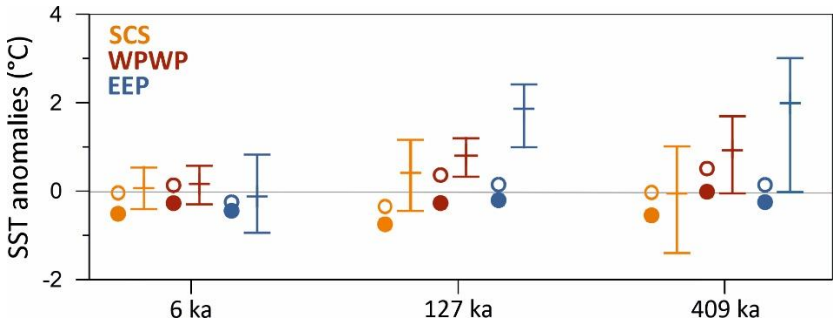


Figure 1. Same as Figure 3 of the main text, but with halving the time windows for the time slices. Comparison between interglacial proxy- and model-based SST anomalies in the SCS (orange), the WPWP (red) and the EEP (blue) during (a) the Holocene, (b) MIS 5e and (c) MIS 11c time slices. Bars indicate proxy-based assessments of computed mean SST anomalies and uncertainties (shown as median and 95 % confidence intervals), dots indicate SST anomalies derived from CESM runs without (filled) and with (open) expanded vegetation cover.

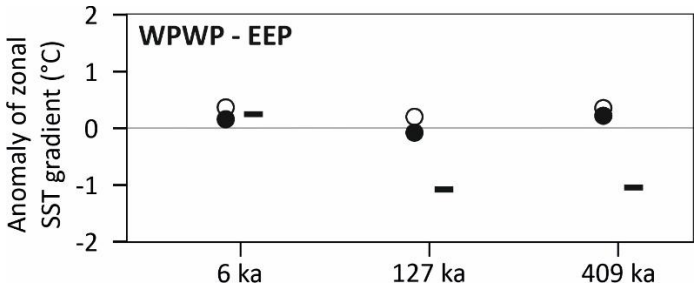


Figure 2. Same as Fig. 4 in the main text of the article, but with halving the time windows of the time slices. Anomalies of zonal equatorial Pacific SST gradients during mid-Holocene, MIS 5e and MIS 11c time slices relative to the preindustrial period. Bars indicate reconstructed anomalies, dots indicate SST anomalies derived from CESM runs without (filled) and with (open) expanded vegetation cover.

3. The regional SST means are weighted by the square root of the number of observations within each time window. It would be helpful to provide more justification for this weighting approach. A higher-resolution core does not necessarily provide more independent information about a regional spatial mean, particularly because adjacent measurements within an individual core are likely to be temporally autocorrelated. It would therefore be helpful to compare the current results with calculations based on

equal weighting among sites. A leave-one-site-out sensitivity test could also be informative, particularly for MIS 11c, where the number of available records is relatively small. Such tests would be particularly useful for evaluating the robustness of the WPWP–EEP SST-gradient anomalies, as this gradient represents one of the central results of the manuscript.

We tried different methods to calculate the regional means and associated uncertainties (Fig. 6).

As described in the manuscript, we finally used a Markov Chain Monte Carlo sampling to calculate weighted mean SSTs (weighted by the square root of the number of observations within the individual cores) of n randomly selected realizations of the time series within each area and time slice, where n equals the total number of SST records within the area and time slice. We weighed towards the number of observations within individual cores, because we assume that a weighted average better reflects the temporal mean of a time series. Autocorrelation would in theory be a problem, but non-climatic noise on the observations means that in practice the effect is negligible. Taking the square root of the number of observations reflects that the precision of the mean improves proportionally to the square root of the sample size due to the central limit theorem

By randomly sampling the records with replacement, we aim to consider spatial uncertainty. While we do not perform a leave-one-site-out sensitivity test as the referee suggests, the random selection with replacement of n realizations of the time series (with n equaling the total number of the SST records within the area and time slice) effectively achieves the same goal of not including all the time series in the calculation of the regional mean. This way, we test the sensitivity to the number of sites within each region and we find that this does not substantially influence the calculated area means, but the uncertainty increases when we randomly sample the records instead of considering all sites once.

The reviewer's point about whether a time series with more datapoints is more representative of the regional mean is valid. We also performed the uncertainty calculations using an equal weighing approach and find that reconstructed area mean SSTs are insensitive to the weighing approach (Fig. 6).

We will add this information to the manuscript supplementary information.

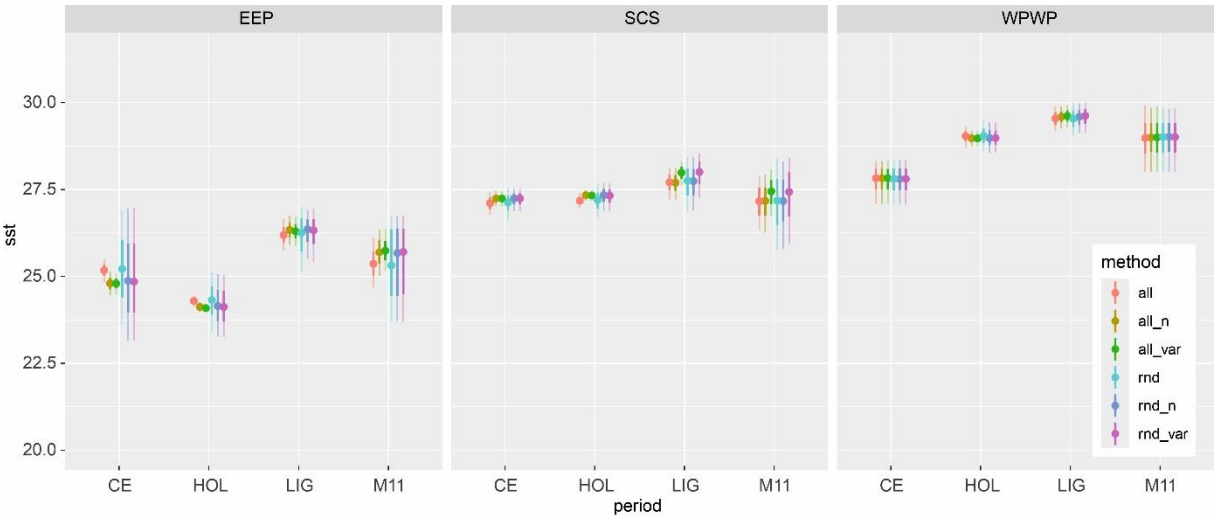


Figure 3. Comparison of reconstructed SSTs using different approaches. Regional mean SSTs in the EEP (left column), SCS (middle) and WPWP (right column) during the Common Era (CE) Holocene (HOL), MIS 5e (LIG) and MIS 11c (M11) time slices. Bars indicate of computed mean SSTs and uncertainties (shown as median and 95 % confidence intervals). We either included each core once (all_x), or randomly selected n cores (n equaling the number of cores included in the region (rnd_x) and used different weighing approaches, including equal weighing (all and rnd), weighted by the square root of

the number of data points of the individual time series (all_n, rnd_n) and weighted by 1/sd of the time series mean (across all ensembles) (all_var, rnd_var).

4. The spatial distribution of the proxy records is relatively uneven, with many sites located near continental margins or coastal regions, whereas large parts of the central and open Indo-Pacific are less well represented. Coastal SSTs may be more strongly influenced by regional processes such as upwelling, river discharge, monsoon variability, and changes in local ocean circulation, and may therefore not always reflect basin-scale SST variability. I would suggest to discuss more explicitly to what extent the available proxy network can be considered representative of the low-latitude Indo-Pacific as a whole. In particular, the limited coverage of the central ocean basins should be taken into account when interpreting the regional mean SST anomalies and when making broader statements about Indo-Pacific-wide warming.

Our compilation includes 89 sites from the deep sea (>1000 m water depth) and we defined the SCS, WPWP and EEP rectangles so that they are well represented by the proxy records.

Note that in general, palaeoceanographic sites tend to be located relatively close to the continents, where higher sedimentation rates facilitate the recovery of records with higher temporal resolution. Proxy records from these sites also have the advantage that they are less affected by secondary effects (such as dissolution) than records from greater water depths. From a very practical point of view, the logistics to obtain cores from far offshore are also more challenging. We will include a cautionary note that points to this limitation in the method section 2.1 (lines 87-90).

5. Given that many of the proxy records are located near continental margins or in coastal regions, changes in sea level may have a strong influence on local oceanographic conditions during past interglacials. In the CESM simulations, present-day topography and sea level are prescribed. It would be useful to discuss whether this could influence the proxy-model comparison, particularly for sites located close to continental margins or shallow shelves. This issue may also be relevant when interpreting the reconstructed anomalies in terms of large-scale Indo-Pacific temperature changes.

As outlined by the referee, sea level variations can exert a significant influence on the oceanographic and climatic conditions. This is particularly relevant on glacial-interglacial time scales, when large shelf areas, such as the Sunda or Sahul shelves become exposed / flooded. We submit that the precise sea level at 127 and 409 ka is hitherto unknown. This is especially true as local relative sea level variations and global mean sea level are difficult to disentangle at this scale. However, the time slices were deliberately selected, as the sea level was assumable comparable to the PI (with deviations <10 m). This was also done to eliminate one level of uncertainty from the model simulations, because large-scale effects that arise from sea level changes do not play a dominant role and differences between these interglacials are almost entirely attributed to orbital forcing and varying GHG concentrations. Sea level variations in the order of a few meters would also be smaller than the grid cells of CESM and therefore, the effects could not be modelled.

6. The manuscript discusses a possible mechanism linking extratropical warming and sea-ice loss to changes in the AMOC, reduced cold-water supply and upwelling, EEP warming, and ultimately a weaker Walker circulation. This is an interesting and physically plausible interpretation. However, the proposed mechanism is not directly evaluated in the analyses currently presented. Since the CESM simulations

are available, the manuscript could be further strengthened by examining relevant model diagnostics associated with these processes.

See our response to General Comment #1 by reviewer David Fastovich.

7. The conclusion on the proxy-model discrepancies, including the proposed "temperature conundrum" during past interglacials, is based on comparison with time-slice simulations of one specific model. It would be more informative to compare with the results from more models, especially to test to what extent the current conclusion is model dependent, and in addition to vegetation, whether the model results are also affected by simulation type. For example, in a transient simulation (Fig. 5a of Wu et al., 2025, https://doi.org/10.1038/s41467-025-57076-2), the simulated peak warmth of MIS 5 and MIS 7 in the western equatorial Pacific SST is similar to PI, and MIS 11c and MIS 9 are warmer than PI.

We agree that a comparison to additional models including transient simulations could be informative. The referee cites a study by Wu et al. (2025), which includes a transient simulation of LOVECLIM1.3. In this study, Wu et al. focus on the area between 120-180 °E and 10 °N – 10°S. For this area, the overall SST response matches our results. Beyond that, our results are in overall agreement with previous publications as well. For instance, Otto-Bliesner et al. (2021) suggests that the CMIP6-PMIP4 time slice simulations (individual models and multi-model ensembles) suggests negative anomalies (or cooler mean annual SSTs as compared to the PI) at 127 ka and would thus underestimate proxy-based SSTs. This information will be included in our discussion section 3.2.

Otto-Bliesner, B. L., Brady, E. C., Zhao, A., Brierley, C., Axford, Y. L., Capron, E., Govin, A., Hoffman, J. S., Isaacs, E., & Kageyama, M. (2021). Large-scale features of Last Interglacial climate: Results from evaluating the lig127k simulations for the Coupled Model Intercomparison Project (CMIP6)-Paleoclimate Modeling Intercomparison Project (PMIP4). Climate of the Past, 17, 63–94. https://doi.org/https://doi.org/10.5194/cp-17-63-2021

Wu, Z., Yin, Q., Berger, A., & Guo, Z. (2025). Forcing mechanisms of the half-precession cycle in the western equatorial Pacific temperature. Nat Commun, 16(1), 1841. https://doi.org/10.1038/s41467-025-57076-2