Response to Reviewer#2

Manuscript number： https://doi.org/10.5194/essd-2025-717

Title： An AI-Driven Reconstruction of Global Surface Temperature with Emphasis on Refining the Antarctic Record

Comments from reviewer：

This manuscript presents a deep learning approach based on partial convolution, trained using 20CR reanalysis data and CMIP6 historical simulations, to reconstruct global monthly gridded surface air temperature from 1850 to 2024, with a particular improvement over the Antarctic region. Overall, the manuscript is well-structured, the methodology is clearly described, and the design of the reconstruction framework as well as the selection of sea surface temperature datasets are appropriate. The reconstructed results are validated against multiple observational datasets and Antarctic station records, demonstrating good consistency. The resulting dataset is valuable for studies of global temperature variability, particularly for Antarctic climate analysis and global temperature change assessments, and is generally well aligned with the scope of ESSD as a data journal.

It is also noteworthy that Dr. David Bromwich, an expert in Antarctic temperature research, pointed out during the public discussion that the Antarctic temperature reconstruction presented in this study shows certain advantages compared to state-of-the-art reanalysis datasets (e.g., ERA5).

I have no objection to the publication of this manuscript after minor revisions. The following comments mainly concern aspects such as figure presentation and result interpretation, which should be further improved by the authors during revision:

Response: We sincerely thank the reviewer for the careful evaluation of our manuscript and for the detailed and constructive comments. These suggestions have provided valuable guidance for improving our work. We have thoroughly revised the original manuscript in accordance with the reviewer's recommendations and have addressed each comment point by point. Detailed responses and corresponding revisions are provided below.

In the following, black text indicates the reviewer's comments, blue text indicates our responses, *blue italic text* indicates the revised content, and *red italic text* with strikethrough indicates deleted content, *black italic text* indicates the original content.

In Figs. 2, 4, and 5, the font sizes of the axis labels and tick marks are relatively small and difficult to read. It is recommended to increase the font size to improve readability. In addition, the legend in Fig. 6 appears somewhat blurred; improving the resolution of this figure would help enhance clarity.

Response: We have adjusted the label sizes and improved the image clarity in this section.

Changes to the manuscript：

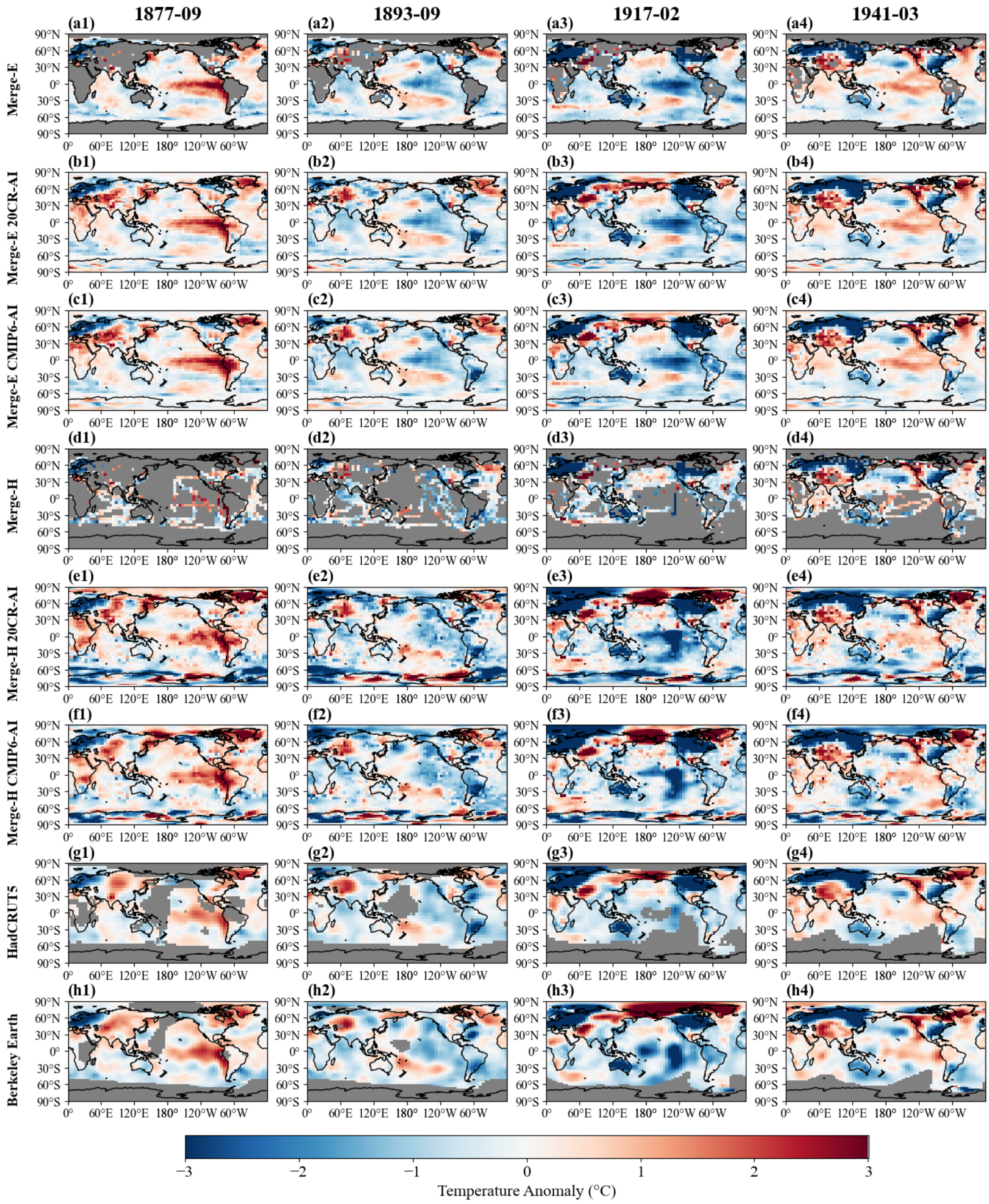

Figure 2: Global temperature anomaly fields before and after reconstruction for four typical ~~months~~ ENSO events. (a1–a4) Merge-E original; (b1–b4) Merge-E 20CR-AI; (c1–c4) Merge-E CMIP6-AI; (d1–d4) Merge-H original; (e1–e4) Merge-H 20CR-AI; (f1–f4) Merge-H CMIP6-AI; (g1–g4) HadCRUT5; (h1–h4) Berkeley Earth.

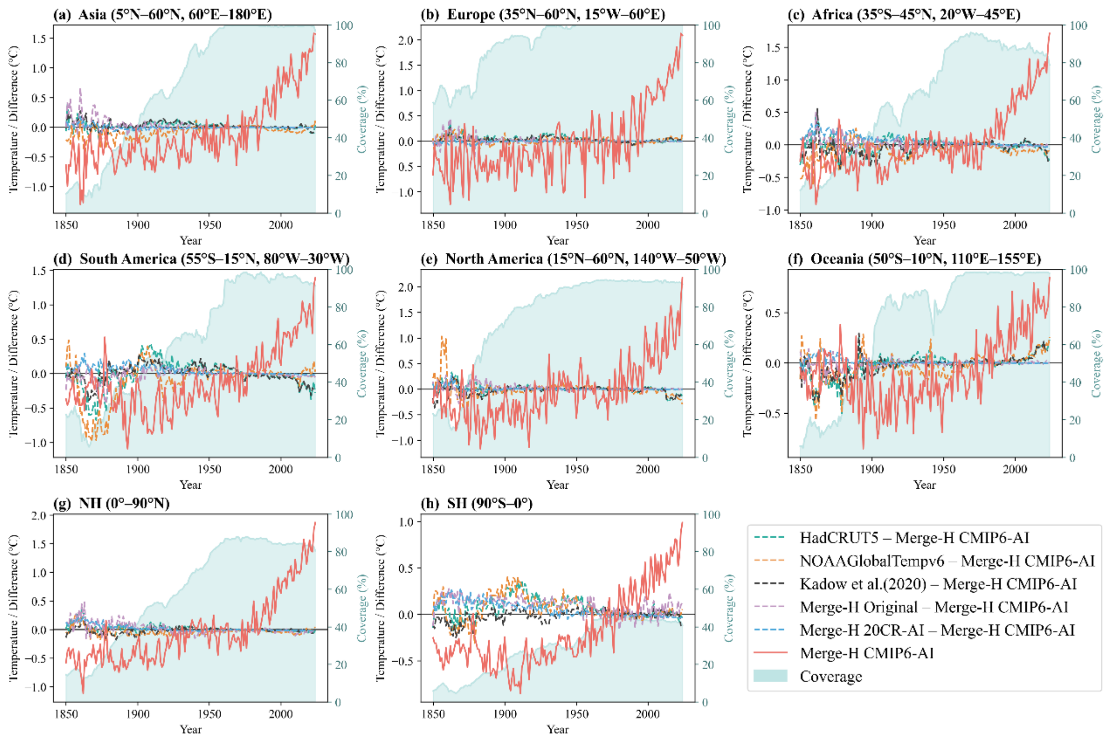

Figure 7: Annual mean surface temperature anomaly time series from 1850 to 2024 in different regions under the Merge-H Scenario. (a–h) Asia, Europe, Africa, South America, North America, Oceania, the Northern Hemisphere, and the Southern Hemisphere. Coverage indicates the original data availability in each region.

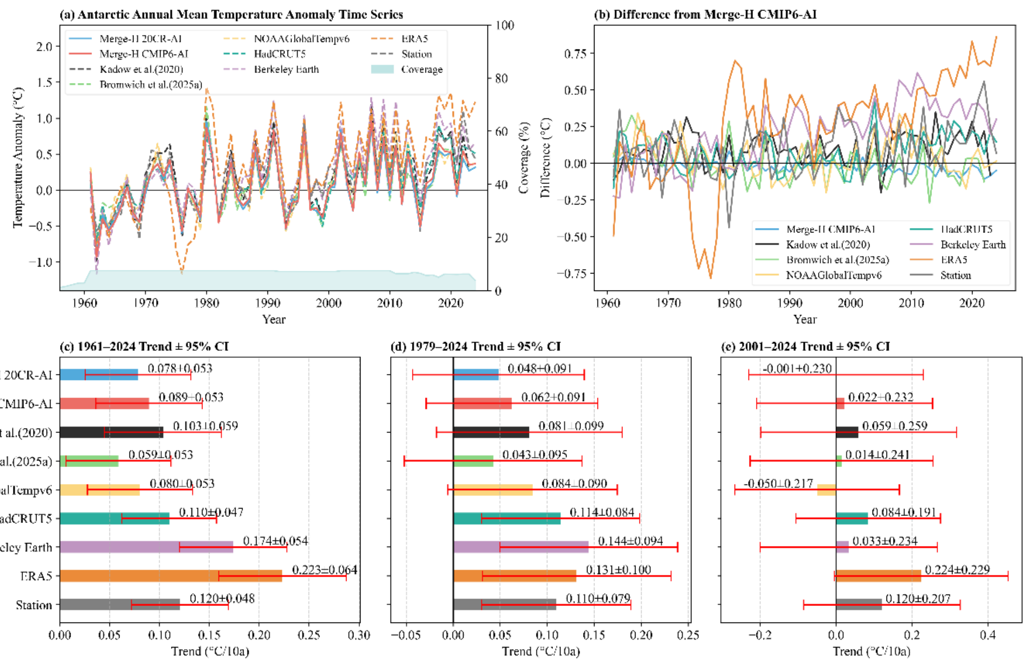

Figure 8: (a) Annual mean surface temperature anomaly time series over Antarctica from 1961 to 2024; (b) Temperature difference from Merge-H CMIP6-AI; (c–e) Linear trend of annual mean temperature and 95% confidence interval (°C per decade) during 1961–2024, 1979–2024 and 2001–2024.

Figure 2 presents reconstruction results for four representative months, but the manuscript does not explain the criteria used to select these months. It would be helpful if the authors could briefly clarify whether these months represent different observational coverage periods, climate conditions, or other considerations.

Response: Figure 2 shows the spatial temperature distribution map, which is intended to provide an intuitive comparison of the reconstruction performance for ENSO spatial patterns. Therefore, we added comparisons with HadCRUT5 and Berkeley Earth. In addition, to better evaluate the reconstruction performance of the entire dataset in representing ENSO events and variability, we included Figure 3 to enable a more comprehensive analysis and assessment.

Changes to the manuscript:

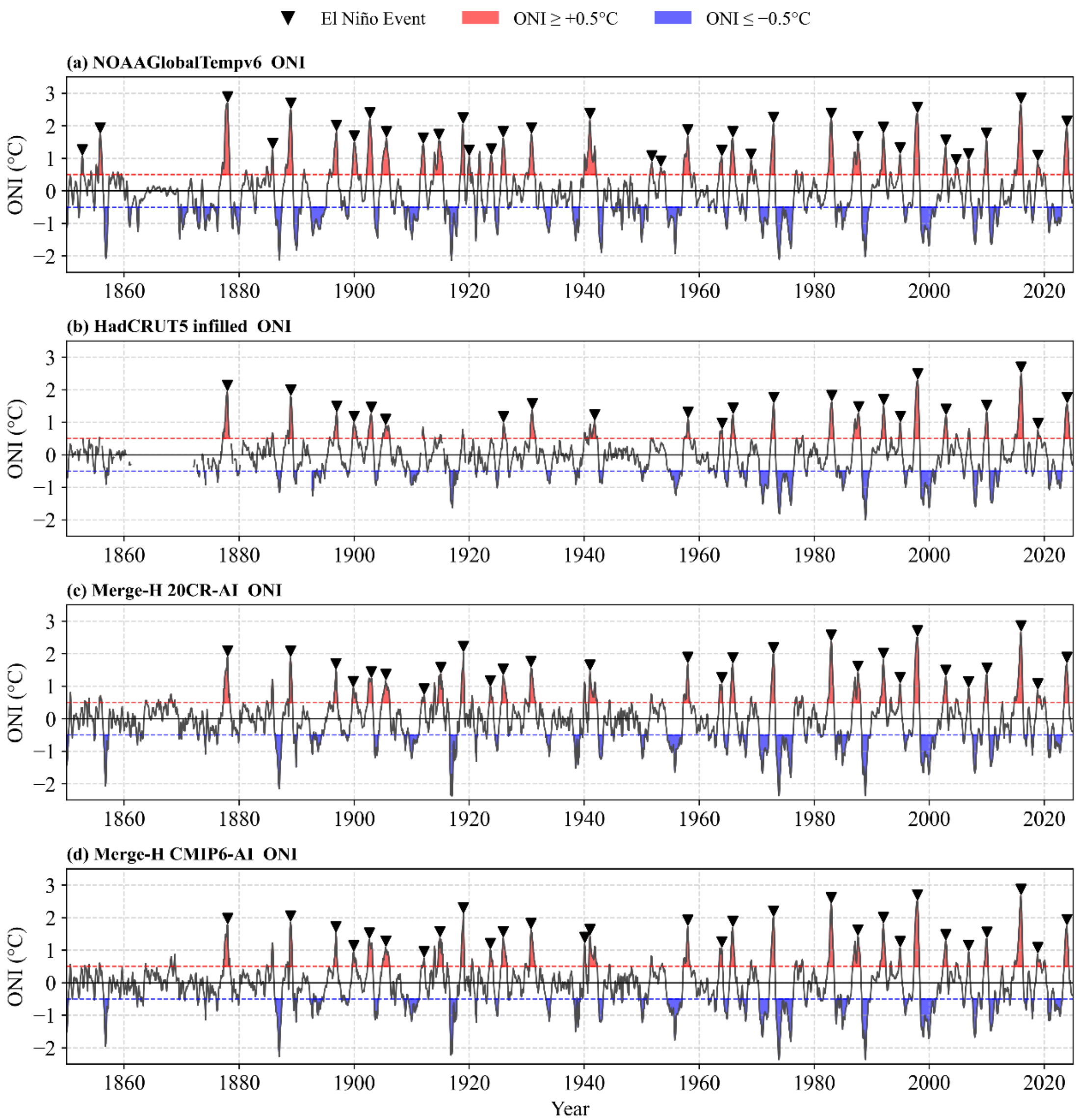

Figure 3: Ocean Niño Index (ONI) time series indicating ENSO events. (a) Time series of ONI from NOAAGlobalTempv6 over 1850–2024, with black triangles indicating El Niño events defined as ONI exceeding 0.5 °C for at least five consecutive months; (b–d) same as (a) but for HadCRUT5 infilled, Merge-H 20CR-AI, and Merge-H CMIP6-AI, respectively.

Figure 3 shows the global annual mean temperature time series. Since the reconstructed series is visually very similar to the CMST3.0-Imax dataset, the contrast between the curves could be improved. For example, the CMST3.0-Imax series could be plotted with a thicker solid line to make the comparison clearer.

Response: Due to the overlap of the time series in this figure, which makes them difficult to distinguish, we added difference plots relative to Merge-H CMIP6-AI. We also included the results from Kadow et al. (2020) for comparison.

Changes to the manuscript:

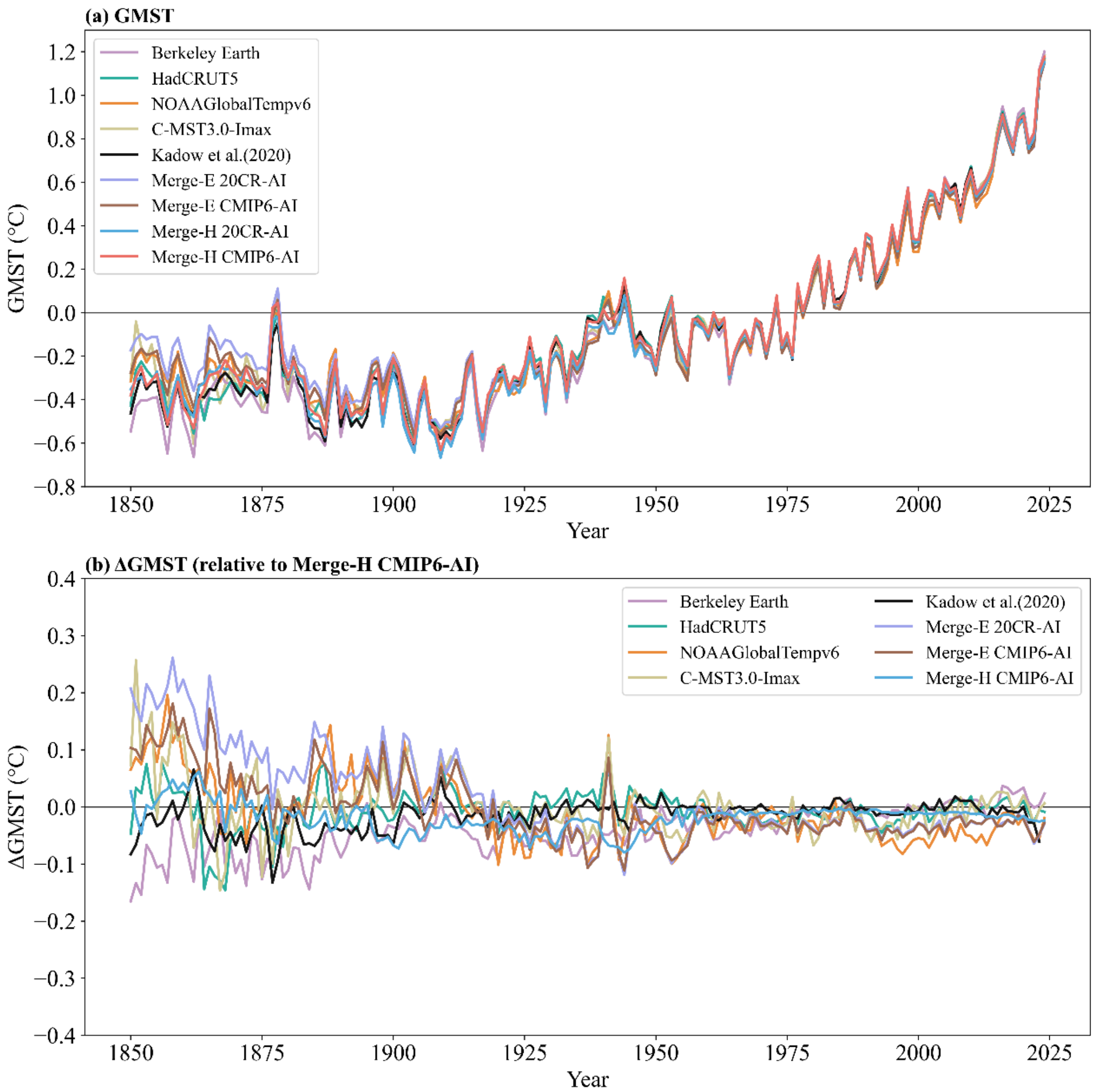

Figure 4: Global mean surface temperature (GMST) time series from 1850 to 2024 (relative to the 1961–1990 climatology). (a) GMST; (b) GMST differences from Merge-H CMIP6-AI.

In Table 1, the warming trends for Merge-H 20CR-AI and Merge-H CMIP6-AI during 1850–2024 are very similar (both 0.064 ± 0.006 °C per decade), yet the corresponding global warming levels (GWL) differ substantially (1.45 and52 °C). However, lines 255–256 state that "The GWL reconstructed by CMIP6-AI reaches 1.52 °C, which is slightly higher than the 1.5 °C estimated from 20CR-AI." The authors may wish to check whether the GWL value reported in Table 1 is correct or if there is a discrepancy between the table and the text.

Response: The GWL value for Merge-H 20CR-AI in Table 1 was a typographical error, and has been corrected to 1.50°C in the revised version.

Changes to the manuscript:

Table 1: Trends of GMST over different periods and 95% confidence intervals (°C per decade), the global warming level (GWL) denotes the increase of GMST (°C) in 2024 relative to the 1850–1900 reference period, the dataset of Kadow et al. (2020) covers the period up to 2023, and its GWL is GMST in 2023 relative to the 1850–1900 baseline period.

Dataset/Period	**1850–2024**	**1900–2024**	**1950–2024**	**1979–2024**	**GWL**
Berkeley Earth	0.070±0.006	0.107±0.008	0.163±0.104	0.206±0.024	1.62
HadCRUT5	0.065±0.006	0.100±0.008	0.156±0.015	0.203±0.023	1.53
NOAAGlobalTempv6	0.058±0.006	0.098±0.008	0.155±0.013	0.196±0.024	1.45
C-MST3.0-Imax	0.061±0.006	0.098±0.008	0.159±0.014	0.210±0.022	1.50
Kadow et al. (2020)	*0.065±0.006*	*0.100±0.008*	*0.151±0.013*	*0.188±0.022*	*1.43*
Merge-E 20CR-AI	0.052±0.007	0.097±0.008	0.157±0.013	0.199±0.023	1.37
Merge-E CMIP6-AI	0.057±0.007	0.099±0.008	0.157±0.013	0.199±0.023	1.43
Merge-H 20CR-AI	0.064±0.006	0.105±0.008	0.155±0.014	0.196±0.023	*1.50*
Merge-H CMIP6-AI	0.064±0.006	0.101±0.008	0.156±0.014	0.199±0.023	1.52

The manuscript notes that observational records in Antarctica are sparse prior to 1961, making direct validation difficult. It would be helpful if the authors could briefly discuss the limitations or uncertainties associated with the reconstructed temperature fields during this early period, so that potential users of the dataset are aware of these limitations.

Response: In the revised version, we have added a comparison of the zonal mean temperature series for each dataset, including Antarctica prior to 1961.

Changes to the manuscript:

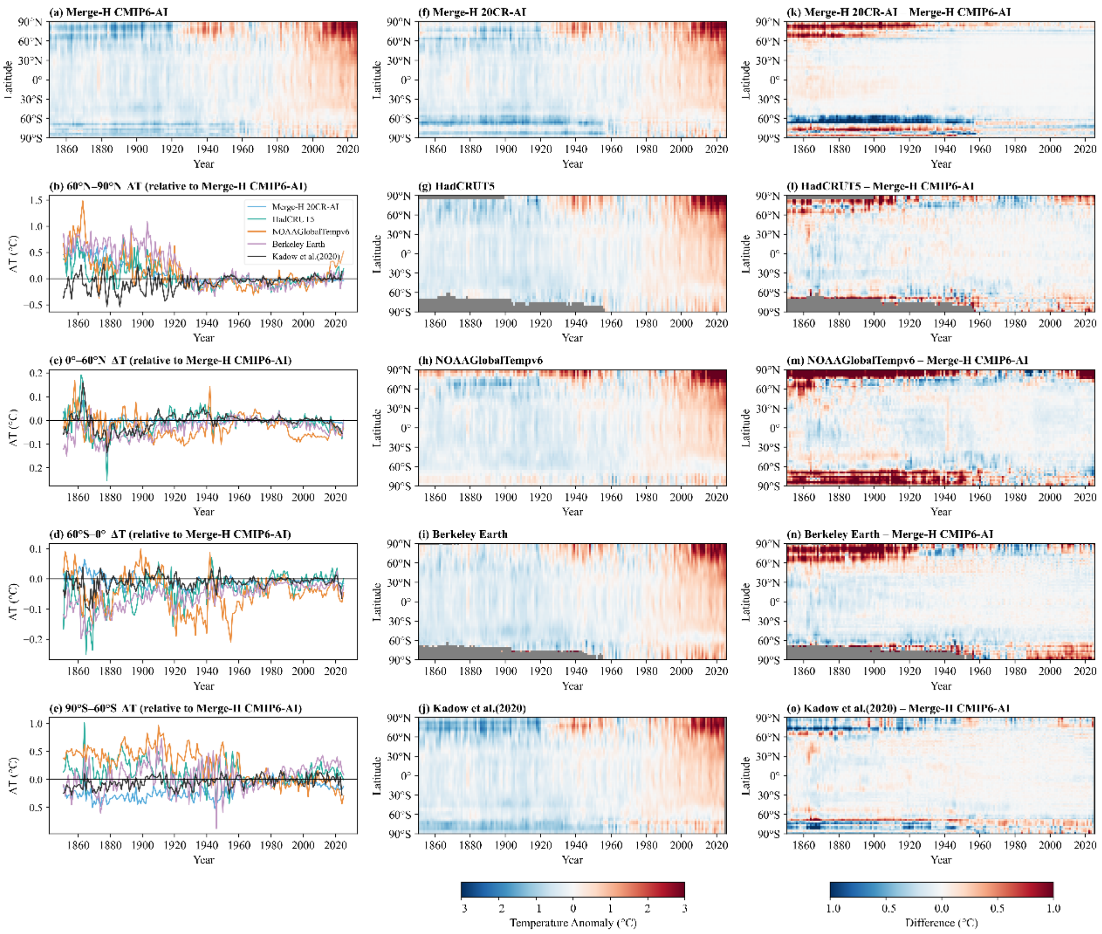

Figure 5 Zonal temperature comparison. (a) Zonal-mean temperature anomalies from Merge-H CMIP6-AI; (b–e) time series of zonal-mean temperature anomaly differences between other datasets and Merge-H CMIP6-AI across different latitude bands; (f–j) same as (a) for Merge-H 20CR-AI, HadCRUT5, NOAAGlobalTempv6, Berkeley Earth, and Kadow et al. (2020), respectively; (k–o) zonal-mean temperature anomaly differences between other datasets and Merge-H CMIP6-AI.

In line 412, the manuscript refers to “optimal interpolation, EOF, or spatially weighted averaging”. The term “EOF” in this context may need clarification. If the authors intend to refer to the empirical orthogonal teleconnection method, it may be more appropriate to use “empirical orthogonal teleconnection (EOT)”.

Response: We thank the reviewer for the comment and have revised this section accordingly in the revised manuscript.

Changes to the manuscript:

~~In contrast, the ERA5 product exhibits some spatial discrepancies: it shows significant warming near Queen Maud Land and west of the Ronne Ice Shelf, whereas cooling along the Wilkes Land coast is not statistically significant. The HadCRUT5, Berkeley Earth, and GISTEMPv4 global temperature reconstructions rely on limited ground-based observations in Antarctica and extend spatially using different statistical methods (optimal interpolation, EOF, or spatially weighted averaging) (Morice et al.,~~

~~2021; Rohde et al., 2020; Lenssen et al., 2019)~~

In contrast, the ERA5 product shows certain spatial differences. It exhibits significant warming over Queen Maud Land and west of the Ross Ice Shelf, while cooling along the Wilkes Land coast does not pass statistical significance testing. HadCRUT5 and Berkeley Earth display relatively smooth spatial patterns of temperature trends over Antarctica, with weaker spatial gradients, and fail to capture the cooling signal over Queen Maud Land and the significant cooling along Wilkes Land coastal regions identified in Bromwich et al. (2025a) and NOAAGlobalTempv6

There are a few minor language issues throughout the manuscript. For example, in line 10, the comma in "substantial missing information in global ST datasets, remains a major source of uncertainty" should be removed. In line 63, "observation-based regional temperature fields often hard to capture" could be revised to "often fail to capture". A careful language editing pass would further improve the readability of the manuscript.

Response: We thank the reviewer for the careful review of the language details and for the valuable suggestions. We have thoroughly proofread and polished the entire manuscript line by line, correcting the relevant grammatical and expression issues.