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

Chenxi Ouyang[1,2,3], Qingxiang Li[1,2,3], Zichen Li[1,2,3], Sihao Wei[1,2,3]

[1] School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, China
[2] Key Laboratory of Tropical Atmosphere–Ocean System, Ministry of Education, Zhuhai, China
[3] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China

Correspondence to: Qingxiang Li (liqingx5@mail.sysu.edu.cn)

Abstract. Accurate estimates of long-term surface temperature (ST) changes are fundamental not only for assessing observed warming, but also for improving the reliability of future climate projections. However, substantial missing information in global ST datasets, remains a major source of uncertainty in estimating global or regional temperature changes. Recent advances in artificial intelligence (AI) have promoted the effective application of deep learning approaches, such as image inpainting and transfer learning, in reconstructing incomplete geophysical datasets. In this study, partial convolutional neural network (PConv) models were trained using the 20CR reanalysis data and CMIP6 climate model outputs as training samples, with the aim of achieving a proper reconstruction of the global surface temperature dataset. To address differences among existing sea surface temperature (SST) datasets, we reconstruct global monthly ST fields since 1850 by merging the China global Land Surface Air Temperature (C-LSAT2.1) dataset with Extended Reconstructed Sea Surface Temperature (ERSSTv6) dataset and Met Office Hadley Centre's sea surface temperature (HadSST4) dataset, respectively. Although both reconstructions reliably reproduce large-scale spatial patterns and long-term variations, the merge of C-LSAT2.1 with HadSST4 exhibits greater physical consistency and is therefore adopted as our preferred reconstruction. In particular, validation against station observations indicates that the reconstructions perform well over the Antarctica after 1961, where observational coverage is extremely sparse. Based on this framework, we developed the China global Artificial Intelligence Reconstructed Surface Temperature$_{20CR/CMIP6}$ (C-AIRST$_{R/M}$) datasets, providing spatially complete global monthly ST anomaly reconstructions since 1850 with a spatial resolution of 5°×2.5°, with anomalies referenced to the 1961–1990 climatology. These datasets offer improved support for extending long-term climate records and for applications in polar climate assessment, as well as in climate monitoring, detection, and attribution studies. The C-AIRST$_{R/M}$ datasets can be downloaded at https://doi.org/10.6084/m9.figshare.30663797.v1 (Ouyang et al., 2025). They are also available from http://www.gwpu.net/en/h-col-103.html (last access: 21 November 2025).

1 Introduction

Global surface temperature (ST) is one of the most fundamental variables in the climate system, directly reflecting the state of the ~~Earth's~~ global energy balance, and it plays a central role in the monitoring and assessment of climate change (IPCC, 2013, 2021). Although sporadic ST observations have been available since the late seventeenth century, continuous observational datasets capable of representing global-scale ST variability did not emerge until the mid-nineteenth century (Cowtan and Way, 2014; Jones, 2016). The sparse distribution of observation sites introduces uncertainties in estimates of both global and regional climate changes (Katz et al., 2013; Karl et al., 2015; Huang et al., 2017; Li et al., 2021, 2022; Sun et al., 2021, 2022). Because the number of meteorological sensors and stations is limited, it is difficult to derive globally representative conclusions directly from observational records. Therefore, developing effective approaches to reconstruct global climate information has become particularly necessary (Vose et al., 2021; Morice et al., 2021). The evolution of the atmosphere and ocean follows the fundamental physical laws of mass, momentum, and energy conservation; therefore, these constraints imply that the climate field exhibits a certain degree of spatial and temporal continuity and predictability, and based on these properties, it is possible to infer missing information through statistical or dynamical relationships even in regions with sparse observations (Lorenz, 1963; Trenberth et al., 2003). Consequently, researchers have employed various methods to reconstruct missing climate information, including smoothing and interpolation techniques (Rayner et al., 2003; Vose et al., 2012; Lenssen et al., 2019; Li et al., 2021), principal component analysis (PCA) and its variants such as empirical orthogonal teleconnection (EOT) (Huang et al., 2017; Sun et al., 2021, 2022), and data-interpolating empirical orthogonal functions (DINEOF) (Beckers et al., 2003; Huang et al., 2017). These methods have played an important role in filling missing values and extracting climate signals from noisy data (Beckers et al., 2003; Wang and Clow, 2020). ~~However, missing information inevitably introduces uncertainties and structural biases. The propagation of observational errors, inconsistencies among reanalysis products, and the inherent limitations of interpolation assumptions can all affect the reliability of reconstruction results (Huang et al., 2017; Morice et al., 2021). These issues are particularly pronounced in regions such as Africa, South America, and Antarctica, where sparse observations and harsh environmental conditions pose greater challenges for traditional reconstruction methods.~~ Nevertheless, due to the inherent incompleteness of the observational system, climate reconstruction inevitably introduces uncertainties and potential biases (Huang et al., 2017; Morice et al., 2021). These issues are particularly pronounced in regions with sparse observations or complex environmental conditions, such as Africa, South America, and the polar regions. It should be emphasized that these challenges are not specific to any single method, but rather represent common limitations shared by current climate reconstruction approaches. These methods have been widely applied at the global scale, their performance exhibits regional variability, with uncertainties becoming more pronounced in areas characterized by extremely limited observations and complex environmental conditions. Among such regions, Antarctica, owing to its unique climatic background and observational constraints, serves as a critical testbed for evaluating the effectiveness and reliability of climate reconstruction methods.

~~The~~ Antarctica holds an irreplaceable position in the global climate system, and its enormous glacier masses store a substantial portion of ~~the world's~~ global freshwater and play a major role in determining future sea level change (IPCC, 2013). At the same time, this region's radiation budget and the complex interactions among ice, atmosphere, and ocean exert profound influences on the atmospheric circulation of the Southern Hemisphere (SH) and the global energy balance (Kennicutt et al., 2019). Consequently, accurately characterizing the spatiotemporal variations of ST over Antarctica is essential for assessing Antarctic climate change, diagnosing changes in ice sheet mass balance, and improving the performance of global climate models. ~~However, due to its extreme geographical conditions, harsh climate, and logistical and communication constraints, observational data from Antarctica remain extremely scarce and temporally uneven.~~ However, the widespread high-elevation ice sheets and complex terrain across Antarctica pose substantial challenges to station deployment. In addition, persistently low temperatures (often below −40 °C) and strong winds hinder the stable operation of observational instruments, while limited transportation and communication infrastructure further complicate station maintenance and data transmission (Wang et al., 2023). To alleviate the problem of insufficient observations, since the International Geophysical Year (1957/1958), national research programs have gradually established an automatic weather station (AWS) network across the Antarctica, providing valuable data for long-term climate monitoring (Jones et al., 2019; Wang et al., 2023). Nevertheless, most of these stations are concentrated along the coastal regions and near research bases, while observations over the interior plateau remain sparse. As a result, observation-based regional temperature fields often ~~hard~~ struggle to capture the overall spatial structure of ST variability (Bromwich et al. 2025a). Against this background, Antarctic temperature reconstruction has become a major focus of polar climate research. Researchers have commonly adopted multi-source data fusion strategies that integrate limited in situ observations with reanalysis products (It refers to the use of data assimilation techniques to integrate multi-source observations with numerical models, producing meteorological fields that are temporally and spatially continuous and consistent) and satellite remote sensing products, using statistical and machine learning methods for spatial and temporal interpolation or extrapolation. For instance, ~~Nicolas et al. (2014) and Bromwich et al. (2025a) applied ordinary kriging with different datasets as weights to spatially extrapolate Antarctic ST, effectively mitigating the problem of sparse observational coverage.~~ Nicolas et al. (2014) and Bromwich et al. (2025a) reconstructed Antarctic surface temperature by using reanalysis data to characterize the spatial covariance structure, and then applying an ordinary kriging framework to derive optimal interpolation weights, thereby enabling spatial extrapolation under sparse observational coverage. Ma et al. (2025) reconstructed the Antarctic temperature field using a deep learning model. Nielsen et al. (2023) improved the spatiotemporal consistency of polar temperature estimates by calibrating MODIS land/ice ST against AWS observations through linear regression. Moreover, Xie et al. (2026) combined limited in situ station data with MODIS ST retrievals within a Bayesian framework, reproducing the evolution of Antarctic climate variability since the beginning of this century. Despite these advances, Antarctic temperature reconstruction still faces considerable challenges. Systematic biases exist among different data sources, and reanalysis products often show substantial uncertainties at high latitudes. Satellite observations are significantly affected by cloud cover and complex topography, while traditional statistical interpolation methods struggle to fully capture nonlinear spatial structures (Wang et al., 2023; Wang et al., 2025; Bromwich et al., 2025; Ma et al., 2025). Therefore, developing reliable methods to reconstruct Antarctic surface air temperature under limited observational constraints remains one of the central scientific challenges in polar climate reconstruction research.

At present, a relatively mature technical framework for global ST reconstruction has been developed. The Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2021) includes five observational ST products: HadCRUT5, NOAAGlobalTemp-Interim, GISTEMPv4, Berkeley Earth, and China-MST-Interim. These datasets apply various reconstruction or interpolation methods to generate homogenized global ST records with ~~as complete a spatial coverage as possible~~ spatial coverage as complete as possible (Morice et al., 2021; Vose et al., 2021; Lenssen et al., 2019; Rohde, 2020; Sun et al., 2021). With the rapid development of artificial intelligence (AI), deep learning has created new opportunities in atmospheric science (Liu et al., 2018; Ham et al., 2019; Kadow et al., 2020; Irrgang et al., 2021). For example, NOAAGlobalTempv6 incorporates an artificial neural network (ANN) to extend data coverage and update the dataset to a globally complete product (Yin et al., 2024). Bochow et al. (2025) applied fast Fourier convolution to fill missing values in HadCRUT4, while Plésiat et al. (2024) examined the ability of partial convolution-based networks to reproduce historical spatial patterns of climate extremes. Partial convolutional networks (PConv), originally proposed by Liu et al. (2018) for image inpainting, perform convolution operations using only valid (non-missing) pixels and dynamically update the validity mask during training. This design greatly enhances reconstruction accuracy, particularly for fields with extensive missing regions. The underlying concept has since been extended to climate reconstruction tasks, in which PConv-based models learn spatial structures and nonlinear dependencies from large climate datasets, thereby enabling "intelligent" completion of incomplete climate fields (Kadow et al., 2020; Zhou et al., 2022; Jiao et al., 2023; Bochow et al., 2025; Ma et al., 2025). Building on these previous efforts, the present study applies PConv reconstruction framework for global ST anomaly (relative to the 1961–1990 climatology) fields. We construct training samples from long-term reanalysis products and climate model simulations, and use in situ Antarctic station observations as the primary reference.~~ By combining statistical approaches with convolutional neural networks, the methodology seeks to balance model interpretability with nonlinear representation capability. The reconstructed temperature anomaly fields are then systematically compared with existing observational and reconstructed products to assess the strengths and limitations of AI-based methods for this application.~~ By integrating statistical methods with convolutional neural networks and leveraging their nonlinear fitting capabilities, global and Antarctic surface temperatures are reconstructed. The results are then systematically evaluated against existing observational and reconstructed datasets to assess the strengths and limitations of this AI-based approach for this problem.

The remainder of this paper is organized as follows. Section 2 describes the data and methods, including data resources, the land–sea merging method, the AI training and reconstruction process, and the data validation methods. Section 3 presents the global reconstruction results, including the spatial pattern of ENSO and the Oceanic Niño Index (ONI), global mean surface temperature (GMST), zonal temperature comparisons, global warming patterns, and regional land surface air temperature comparisons. Section 4 presents the Antarctic temperature reconstruction results. Section 5 discusses the limitations and future perspectives. Sections 6 and 7 provide information on data and code availability, respectively. Section 8 concludes the paper.

2 Data and methods

2.1 Data resources

China land surface air temperature dataset (C-LSAT) is a global land surface air temperature (LSAT) product that provides both station-based and gridded data. It integrates a total of 14 data sources, including three global datasets (CRUTEM4, GHCNm, and BEST), three regional datasets, and eight national datasets. A major advancement of this dataset lies in its substantially improved station coverage across most Asian countries, particularly in China and its surrounding regions (Xu et al., 2018; Li et al., 2021; Wei et al., 2025). The dataset was assessed in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6), released in August 2021, as "fully meeting the IPCC requirements," and was accordingly incorporated and utilized in the report (IPCC, 2021). C-LSAT2.1 represents its latest version (Wei et al., 2025).

Building upon this foundation, the China global Merged Surface Temperature (C-MST3.0, Yun et al., 2019; Sun et al., 2021, 2022; Li et al., 2020, 2021) is constructed by merging the land surface air temperature dataset C-LSAT2.1 with the Extended Reconstructed Sea Surface Temperature dataset ERSSTv6 (Huang et al., 2025a, 2025b). According to the spatial extent of the reconstructed Arctic sea ice surface air temperature regions, three variants are defined: C-MST3.0-Nrec, C-MST3.0-Imin, and C-MST3.0-Imax (Li et al., accepted). ~~The China global Merged Surface Temperature (China-MST/C-MST) dataset is an established global ST product (Yun et al., 2019; Sun et al., 2021, 2022; Li et al., 2020, 2021). The latest version, C-MST3.0, is classified into three variants (C-MST3.0-Nrec, C-MST3.0-Imin and C-MST3.0-Imax) based on the spatial coverage of reconstructed Arctic sea ice ST (Li et al., under review). C-MST3.0 is constructed by combining ST from the China land surface air temperature (LSAT) dataset C-LSAT2.1 (Wei et al., 2025) with the Extended Reconstructed Sea Surface Temperature version 6 (ERSSTv6), released by NOAA/NCEI (Huang et al., 2025a, 2025b).~~

In this study, LSAT from C-LSAT2.1 are separately merged with the sea surface temperature (SST) products ERSSTv6 and the Met Office Hadley Centre's sea surface temperature (HadSST4.1.1.0) dataset (Kennedy et al., 2019), and the resulting merged datasets, referred to as "Merge-E" and "Merge-H", are used as the original inputs for the AI-based reconstruction. Both datasets provide monthly mean ST from 1850 to 2024.

In addition, two ~~historical~~ monthly ST anomaly (relative to the 1961–1990 climatology) datasets are employed to construct training sets for the AI reconstructions. The Twentieth Century Reanalysis version 3 (20CR; Slivinski et al., 2019), which provides monthly ST fields for 1850–2015, is used to train the "20CR-AI" model, whereas the ~~"Historical"~~ historical simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016), offering monthly ST fields for 1850–2014, are used to train the "CMIP6-AI" model. The 20CR dataset contains 80 ensemble members, whereas the CMIP6 dataset contains 105 ensemble members (Table S1).

To better reconstruct realistic climate conditions over Antarctica, we further employ monthly Antarctic station data. Most stations originate from SCAR READER (Turner et al., 2004), GHCNm v4 QCF and QFE (Menne et al., 2018)~~, AntAWS (Wang et al., 2023), and GSOD (NCEI, 1999)~~. Additional stations are obtained from the OSU Polar Meteorology Group (Bromwich et al., 2025b), Météo-France (Météo-France, 2025), the University of Wisconsin-Madison (South Pole

Meteorology Office, 2025), and the National Institute of Water and Atmospheric Research Ltd (NIWA, 2025). Stations with more than 25 valid years during 1961–1990 are selected, yielding a temporal span of 1957–2024. Following Bromwich et al. (2025a), necessary gap-filling procedures are applied to the Antarctic station data. Subsequently, homogenization tests are performed to eliminate discontinuities caused by station relocations, sensor changes, or other non-climatic shifts. Station observations are mapped to the corresponding 5°×2.5° grid cells (based on station longitude and latitude) and treated as valid data constraints for the reconstruction. Details of the station metadata and temperature anomaly time series are shown in Table S2 and Figure S1 in the Supplement. Table 1 presents information on the different types of data products used in this study.

Table 1 List of information on the various data used in this paper.

Data	Type	Resolution	Time	Reference
20CR	Grid	Monthly/0.7°×0.7°	1850–2015	Slivinski et al. (2019)
CMIP6	Grid	Monthly/-	1850–2014	Eyring et al. (2016)
C-LSAT2.1	Grid	Monthly/5°×5°	1850–2024	Wei et al. (2025)
ERSSTv6	Grid	Monthly/5°×5°	1850–2024	Huang et al. (2025a)
HadSST4	Grid	Monthly/5°×5°	1850–2024	Kennedy et al. (2019)
SCAR READER	Station	Monthly	-	Turner et al. (2004)
GHCNv4 QCF and QFE	Station	Monthly	-	Menne et al. (2018)
OSU Polar Meteorology Group	Station	Monthly	1958–2022	Bromwich et al. (2025b)
Météo-France	Station	Monthly	2014–2016	Météo-France (2025)
University of Wisconsin-Madison	Station	Monthly	1958–2022	South Pole Meteorology Office (2025)
NIWA	Station	Monthly	2016–2022	NIWA (2025)

2.2 LSAT and SST merging method

In this study, we adopted the data merging approach of Yun et al. (2019) to merge LSAT from C-LSAT2.1 separately with SST from ERSSTv6 and HadSST4. The land–ocean mask used in this process was obtained from the NCAR Command Language (NCL) "landsea.nc" mask file (download at: http://www.ncl.ucar.edu/Applications/Data/cdf/landsea.nc, last access: 23 July 2025). The mask file contains five categories: 0 for ocean, 1 for land, 2 for lake, 3 for small island, and 4 for ice shelf. In this study, land, lake, small island, and ice shelf were treated as land, while the other category was considered ocean. Due to the relatively low resolution of the reconstructed datasets (5°×2.5°), this merging approach results in some grid cells along the land–ocean boundaries, particularly along the Antarctica where observational data are sparse, containing only ocean information. AI-based reconstruction is influenced by features at the edges of missing data (Liu et al., 2018). Specifically, in Antarctic coastal transition zones where partial observations are available over the ocean but land observations are nearly absent, these grid cells in PConv can only rely on information from adjacent oceanic grids. The convolutional operation in PConv generates predictions based on the surrounding available data, which leads to a bias toward oceanic characteristics and makes it difficult to accurately represent the true local surface air temperature over land or ice sheets. In simple terms, "land-edge grid cells are biased by surrounding heterogeneous ocean grids." This effect is particularly pronounced in the merged dataset (Merge-E), which is constructed by combining with the fully ocean-covered ERSSTv6, and in regions with severe observational gaps (see Section 3.2 for a detailed discussion of GMST). To minimize the impact of ocean-dominated grid cells along land–ocean boundaries, we reasonably expanded the Antarctic land mask relative to the original land–sea mask at the 5°×2.5° scale (Fig. S2). This expansion has negligible effect on the merging process in regions with extremely sparse Antarctic observations but allows the AI reconstruction to better capture ST patterns in the Antarctic coastal and peripheral areas.

2.3 AI training and reconstruction process

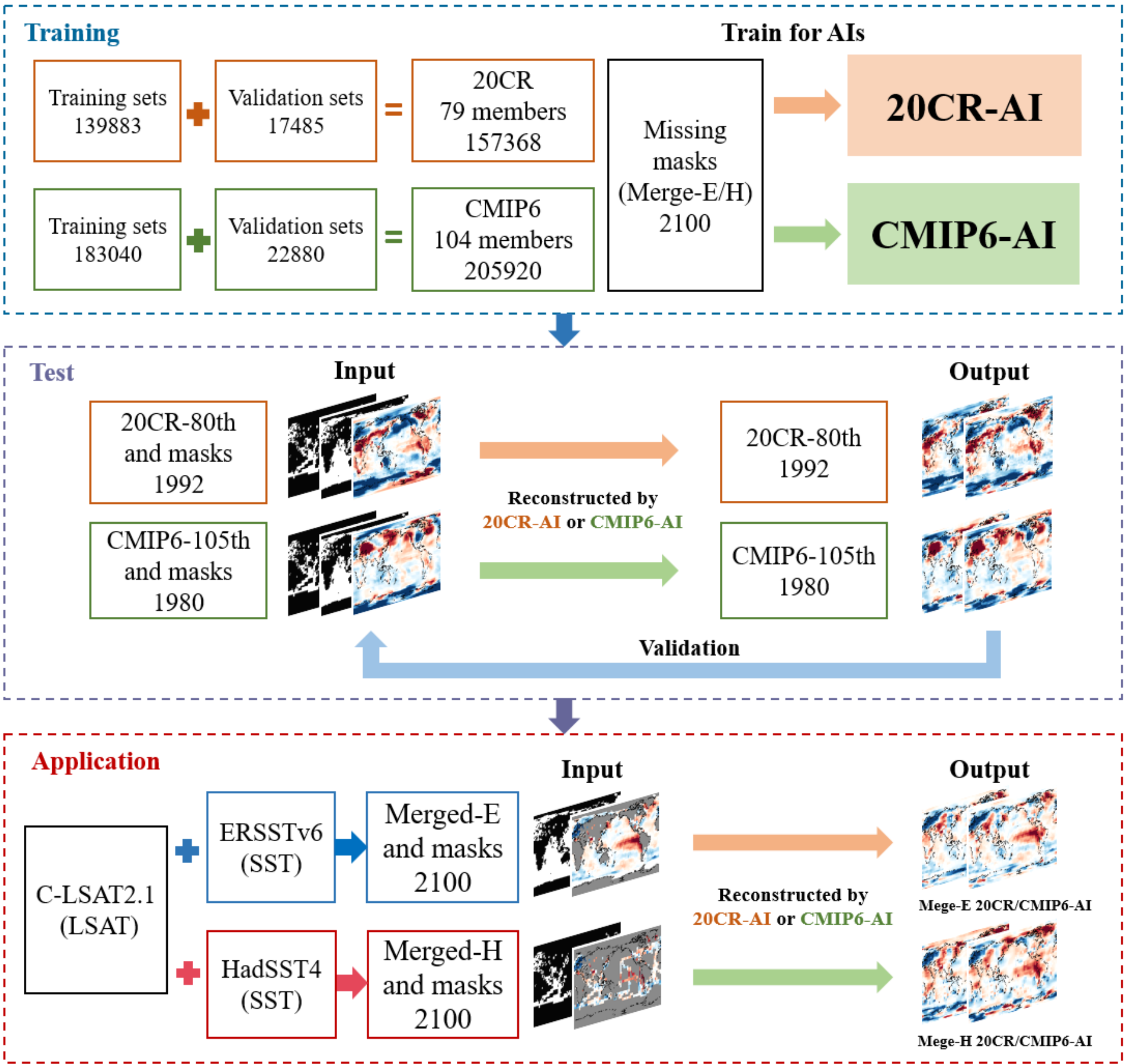


Figure 1: AI training and reconstruction (The numbers shown in the data boxes represent the sample size of the monthly anomaly fields, corresponding to the total number of months).

The PConv method has demonstrated excellent performance in reconstructing globally missing climate data, including ST anomaly fields, wind fields, and surface solar radiation (Kadow et al., 2020; Zhou et al., 2022; Jiao et al., 2023). The underlying principle is to treat the climate field as a two-dimensional image, with missing regions regarded as irregular holes to be inpainted. PConv is then trained to learn spatial features from a large set of climate data samples, enabling the inference of plausible spatial patterns in the missing regions (Kadow et al., 2020). ~~In this study, the AI reconstruction follows a similar approach.~~ This study builds upon the work of Kadow et al. (2020), whose primary contribution was the introduction of a PConv-based framework for climate data reconstruction. Kadow et al. mainly demonstrated the feasibility and accuracy of PConv for climate field reconstruction at the global scale with a resolution of 5° × 2.5°, and showed that the method can effectively capture the spatial patterns of ENSO-related sea surface temperature anomalies. The PConv approach performs well for missing grid points over the low- and mid-latitude areas, but provides limited assessment for polar regions or areas with severe observational deficiencies.

In this study, we introduce targeted improvements for the Antarctic region after 1961. On the one hand, additional Antarctic station data are incorporated to construct the input samples, ensuring sufficient valid data support after 1961. On the other hand, considering the coarse spatial resolution (5° × 2.5°) and the pronounced land–sea contrasts in Antarctic grid cells, the Antarctic land mask is adjusted. Based on these modifications, more detailed validation and analysis are conducted for both global and Antarctic domains to assess the performance and limitations of the model under different training datasets and different SST products.

The reconstruction workflow is illustrated in Figure 1 (adapted from Extended Data Fig. 1 of Kadow et al., 2020), and consists of the following steps:

(1) Resolution standardization: ~~All datasets used for training and reconstruction were regridded to a common spatial resolution of 5°×2.5°, resulting in a 72×72 grid.~~ The input datasets used for training are uniformly remapped from higher resolution to a spatial resolution of 5° × 2.5° using conservative interpolation, resulting in a regular grid composed of 72 × 72 grid cells.

(2) Sample partitioning: Monthly temperature anomaly fields from 20CR (80 ensemble members) for 1850–2015 and CMIP6 (105 ensemble members) for 1850–2014 ~~(2014)~~ relative to the 1961–1990 climatology were divided into training and testing sets. One ensemble member was randomly selected as the test set, while the remaining 79 members from 20CR were used to train the 20CR-AI model, and 104 CMIP6 members were used to train the CMIP6-AI model.

(3) Training sample construction: The training samples were randomly partitioned, with 8/9 used for model training and 1/9 for validation. This resulted in 139,883 training samples and 17,485 validation samples for the 20CR-AI model, and 183,040 training samples and 22,880 validation samples for the CMIP6-AI model.

(4) Mask configuration: Observational data from Merge-E and Merge-H, covering monthly values from 1850 to 2024, were used as masks for missing values in the respective reconstruction schemes. Observed grid points were marked as “1” and missing grid points as “0”, serving as inputs for model training and validation.

(5) Model training and fine-tuning: Both AI models were trained for 500,000 iterations with a learning rate of 0.0002, followed by an additional 500,000 iterations with a reduced learning rate of 0.00006. The batch size was set to 16, and computations were performed on an NVIDIA GeForce RTX 4060 GPU for approximately 8 hours.

(6) Reconstruction: The datasets Merge-E and Merge-H, which contain missing values, are both input into the trained 20CR-AI and CMIP6-AI models for inference, with the output being globally complete monthly surface temperature datasets at a spatial resolution of 5° × 2.5° for the period 1850–2024.

2.4 ~~Model Post-Processing~~ Data validation methods

~~To ensure comparability across different data sources, all external datasets, including observations, reanalysis products, and reconstructed data, were first regridded to a common 72×72 regular grid consistent with the AI reconstruction output, with a uniform spatial resolution of 5°×2.5°.~~ To ensure comparability across different data sources, all external benchmark ST datasets are first remapped to a spatial resolution of 5° × 2.5° consistent with the AI reconstruction inputs and outputs, using bilinear interpolation, resulting in a 72 × 72 regular grid. All data were converted to ST anomalies relative to the 1961–1990 climatology to eliminate differences in the climate reference among datasets. Subsequently, the performance of the AI reconstruction was systematically evaluated. The primary evaluation metrics included the spatial correlation coefficient and root mean square error (RMSE), which were used to assess the consistency and bias of the reconstructed fields relative to reference datasets in the spatial domain. Temporal stability and reconstruction accuracy were further examined using annual mean correlation coefficients and RMSE time series (Fig. S4, S5 and S6).

To comprehensively assess the global applicability and reliability of the AI reconstruction, it was compared against multiple representative global temperature datasets, including Berkeley Earth, HadCRUT5, NOAAGlobalTempv6, C-MST3.0-Imax, and the latest HadCRUT5 reconstruction results of Kadow et al. (2020) trained on the 20CR dataset, as well as four reconstruction scenarios in this study (Merge-E 20CR-AI, Merge-E CMIP6-AI, Merge-H 20CR-AI, Merge-H CMIP6-AI). The comparison included global annual mean temperature anomaly time series, linear trend estimates, and their statistical significance. In the Antarctica, where observational coverage is sparse, targeted evaluation of the AI reconstruction was performed. Reference datasets included the ERA5 reanalysis, Berkeley Earth, HadCRUT5, ~~GISTEMPv4,~~ Kadow et al.(2020), the reconstruction by Bromwich et al. (2025a), and Antarctic station data used in this study. Cross-validation among these multiple data sources was conducted to assess the reliability and consistency of the Antarctic reconstructions after 1961.

The GMST series of the benchmark datasets used in this study follow the standard products released by their respective official sources. All regional annual mean temperature series are first computed using area-weighted spatial averaging, followed by temporal averaging on an annual basis, in accordance with the World Meteorological Organization (WMO) methodology. Linear trends are estimated using ordinary least squares, and their statistical significance is assessed using a two-sided t-test. The Ocean Niño Index (ONI) is defined following the NOAA Climate Prediction Center standard, as the 3-month running mean of temperature anomalies over the Niño 3.4 region (5°S–5°N, 170°W–120°W), calculated relative to a centered 30-year base periods that is updated every 5 years.

3 Global Reconstruction Results

3.1 ~~Characteristics of the Global Reconstruction Results~~ ENSO spatial patterns and the Ocean Niño Index (ONI)

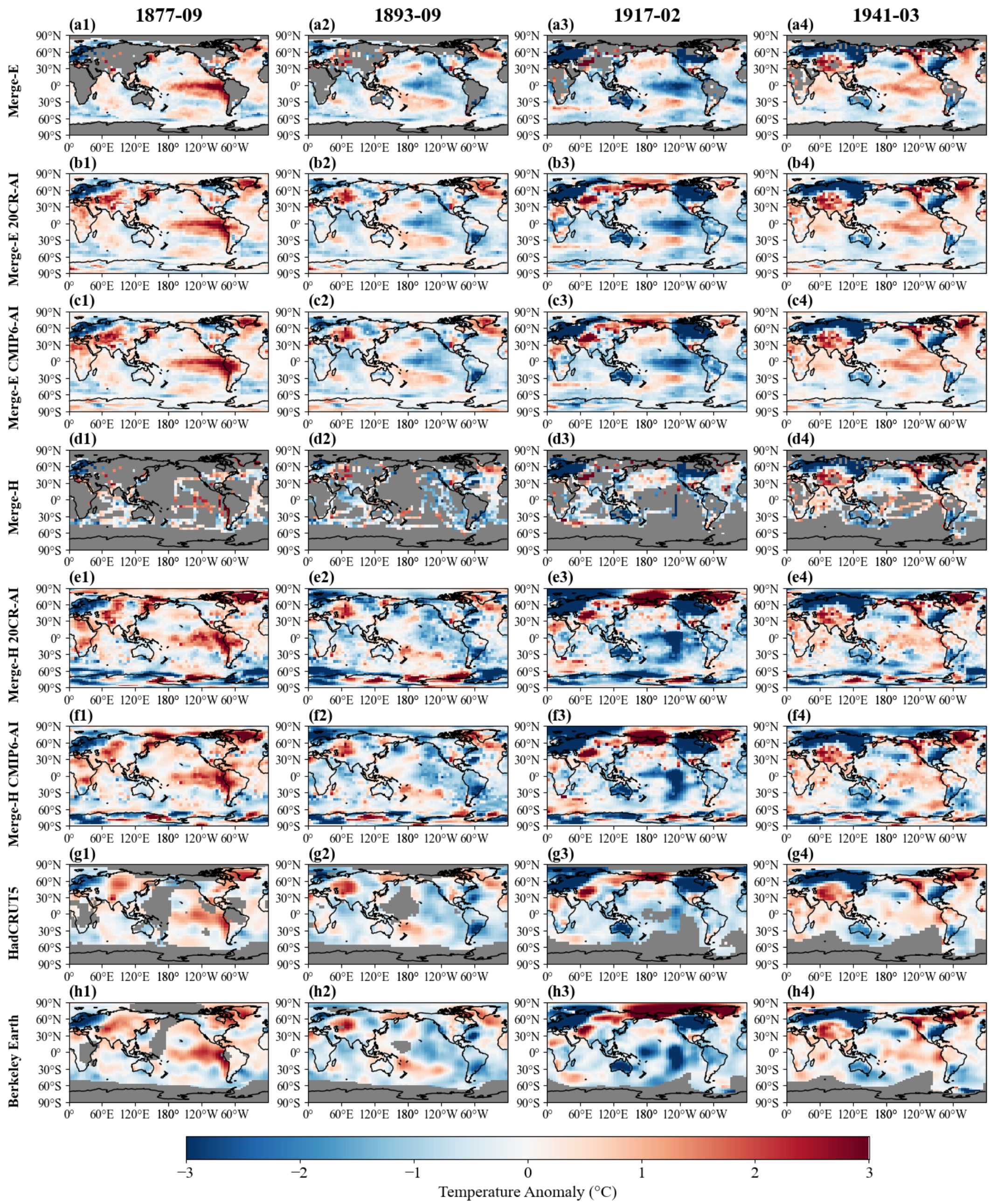


Figure 2: Global temperature anomaly (relative to the 1961–1990 climatology) fields before and after reconstruction for four typical ~~months~~ ENSO events. (a1–a4) Merge-E original; (b1–b4) Merge-E 20CR-AI; (c1–c4) Merge-E CMIP6-AI; (d1–d4) Merge-H original; (e1–e4) Merge-H 20CR-AI; (f1–f4) Merge-H CMIP6-AI; (g1–g4) HadCRUT5; (h1–h4) Berkeley Earth.

Prior to the 19th century, the C-LSAT2.1 dataset contained substantial missing data in Asia, South America, Africa, and Antarctica due to the sparse distribution of land-based stations, and this situation was improved considerably after the 19th century (Wei et al., 2025). In addition, Antarctic observational stations began recording data progressively after the 1957/1958 International Geophysical Year; however, observational coverage in Antarctica remains extremely sparse (Wang et al., 2023). As shown in Figure S3, the original ERSSTv6 dataset exhibits higher ocean data coverage compared to HadSST4. This is because ERSSTv6 employs an ANN reconstruction method, which allows for more accurate and stable reconstruction of SST in sparsely observed regions, reduces excessive smoothing, and better preserves spatial variability (Huang et al., 2025a, 2025b). In contrast, HadSST4 primarily relies on statistical interpolation and ensemble-based methods, which aim to correct biases and estimate uncertainties within grid cells, without performing large-scale spatial reconstruction (Kennedy et al., 2019).

Considering the differences between the two SST datasets, this study performed AI-based reconstruction using two distinct merged datasets to investigate potential discrepancies under different conditions. Figures 2 (b1–b4, c1–c4, e1–e4, f1–f4) show the complete temperature anomaly fields reconstructed from Merge-E and Merge-H using the 20CR-AI and CMIP6-AI models. ~~The global spatial patterns of the four reconstructed fields are largely consistent across most regions; however, in high-latitude areas with extremely sparse early observations, the spatial distributions of Merge-E and Merge-H reconstructions show some differences. The reconstruction performance of the model improves as the coverage of original valid data in the model validations (Fig. S4, S5 and S6). Compared to Merge-H, Merge-E exhibits higher spatial coverage and fewer missing gaps in polar regions, particularly over Antarctica and its surrounding seas. This leads to better AI reconstruction performance under the Merge-E mask than under Merge-H (Fig. S4 and S6). Visually, the reconstructed fields in Antarctica and adjacent regions are smoother in Merge-E than in Merge-H (Fig. 2), indicating that during image inpainting, the AI reconstruction is influenced to some extent by the colour, texture, and style features at the edges of missing regions in the two different datasets (Liu et al., 2018; Nazeri et al., 2019), thereby leading to distinct reconstructed features in the early periods when large areas of missing data occur in Merge-E and Merge-H.~~ In Fig. 2 (b1–b4, c1–c4), Merge-E uses the complete ocean component from ERSSTv6, which is not reconstructed by the AI model; therefore, the ENSO spatial patterns in these panels are directly derived from ERSSTv6 and are shown here for reference. During the strong El Niño event in September 1877, the AI model is able to reasonably reconstruct a coherent warming anomaly pattern over the equatorial western Pacific under data-sparse conditions (Fig. 2 d1, e1, f1). This pattern is characterized by pronounced warm anomalies in the tropical central and eastern Pacific, with alternating warm and cold anomalies distributed zonally along the equator, reflecting a typical ENSO signal.

For the ENSO reconstructions in September 1893, February 1917, and March 1941 (Fig. 2 d2–d4, e2–e4, f2–f4), the spatial patterns produced by the AI model are generally consistent with those from HadCRUT5 (Fig. 2 g2–g4) and Berkeley Earth (Fig. 2 h2–h4), both of which also use HadSST-based ocean components. However, the spatial distribution of warm and cold anomalies along the equatorial Pacific, namely the classic "warm–cold tongue" structure, is less pronounced than in ERSSTv6 (Fig. 2 a2–a4). This discrepancy primarily arises from inherent differences between HadSST and ERSSTv6.

~~Moreover, the AI reconstruction effectively reproduces characteristic climate events in key years. In particular, the results shown in Figures 2 (e1, f1) clearly capture the strong El Niño event of 1877, characterized by significantly positive SST anomalies in the central and eastern tropical Pacific and a distinct east–west dipole pattern of warm and cold anomalies along the equator, reflecting the typical signal of the El Niño Southern Oscillation (ENSO). This demonstrates that the model is capable of reconstructing large-scale spatial patterns and temporal evolution of the temperature field even under extremely sparse observational coverage, indicating robust performance and spatial consistency.~~

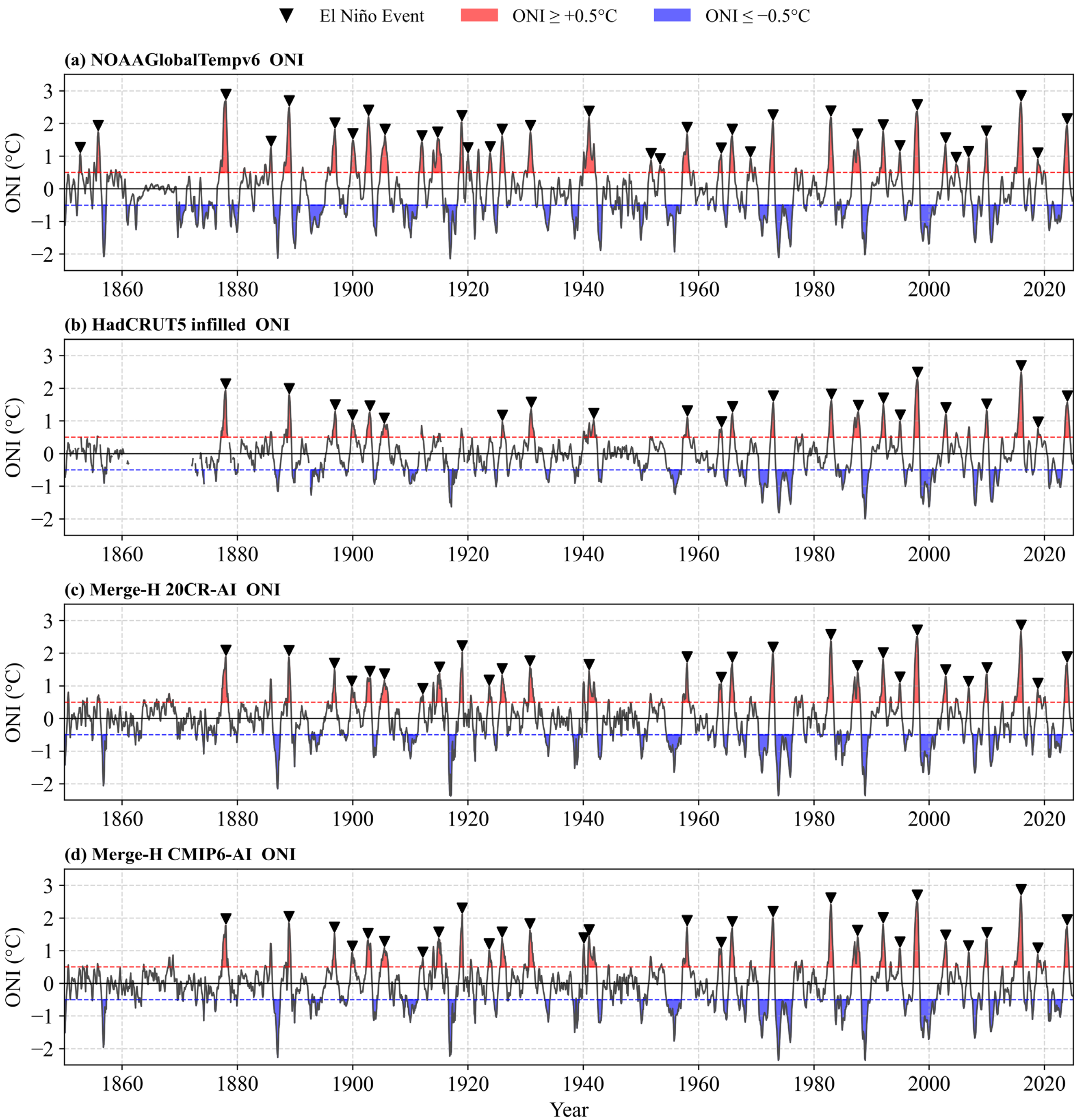


Figure 3: Ocean Niño Index (ONI) time series indicating ENSO events. (a) Time series of ONI from NOAAGlobalTempv6 over 1850–2024, with black triangles indicating El Niño events defined as ONI exceeding 0.5 °C for at least five consecutive months; (b–d) same as (a) but for HadCRUT5 infilled, Merge-H 20CR-AI,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

The two AI models are able to effectively capture the spatial patterns of ENSO, while also reproducing historical ENSO events. As shown in Fig. 3(a–d), the positive and negative phases of the ONI, which serves as an indicator of ENSO variability, are generally consistent with those from NOAAGlobalTempv6 and HadCRUT5. The AI-based reconstruction also fills the gaps in the ONI during 1910–1920 in HadCRUT5 infilled, where incomplete spatial coverage over the Niño 3.4 region led to biases or missing values. In addition, the reconstruction identifies the documented El Niño events of 1911–1912, 1914–1915, and 1918–1919 (Yu et al., 2013), which are also evident in NOAAGlobalTempv6 (Fig. 3a).

A few weak El Niño events show slight discrepancies compared with HadCRUT5 due to small differences in the amplitude of the Niño 3.4 index, indicating minor estimation biases in the AI-reconstructed index. It is also noteworthy that the ENSO amplitude in NOAAGlobalTempv6 is generally stronger than that in HadCRUT5 prior to 1950, and this discrepancy is primarily attributable to differences in the SST bias correction methods and spatial infilling approaches employed by the two datasets (Morice et al., 2021; Huang et al., 2025a; 2025b). Consequently, the Merge-H results based on HadSST are overall more consistent with HadCRUT5. In summary, the AI models perform well in representing the spatial structure of ENSO, they are also capable of reproducing historical ENSO events through the ONI.

3.2 Global Mean Surface Temperature (GMST)

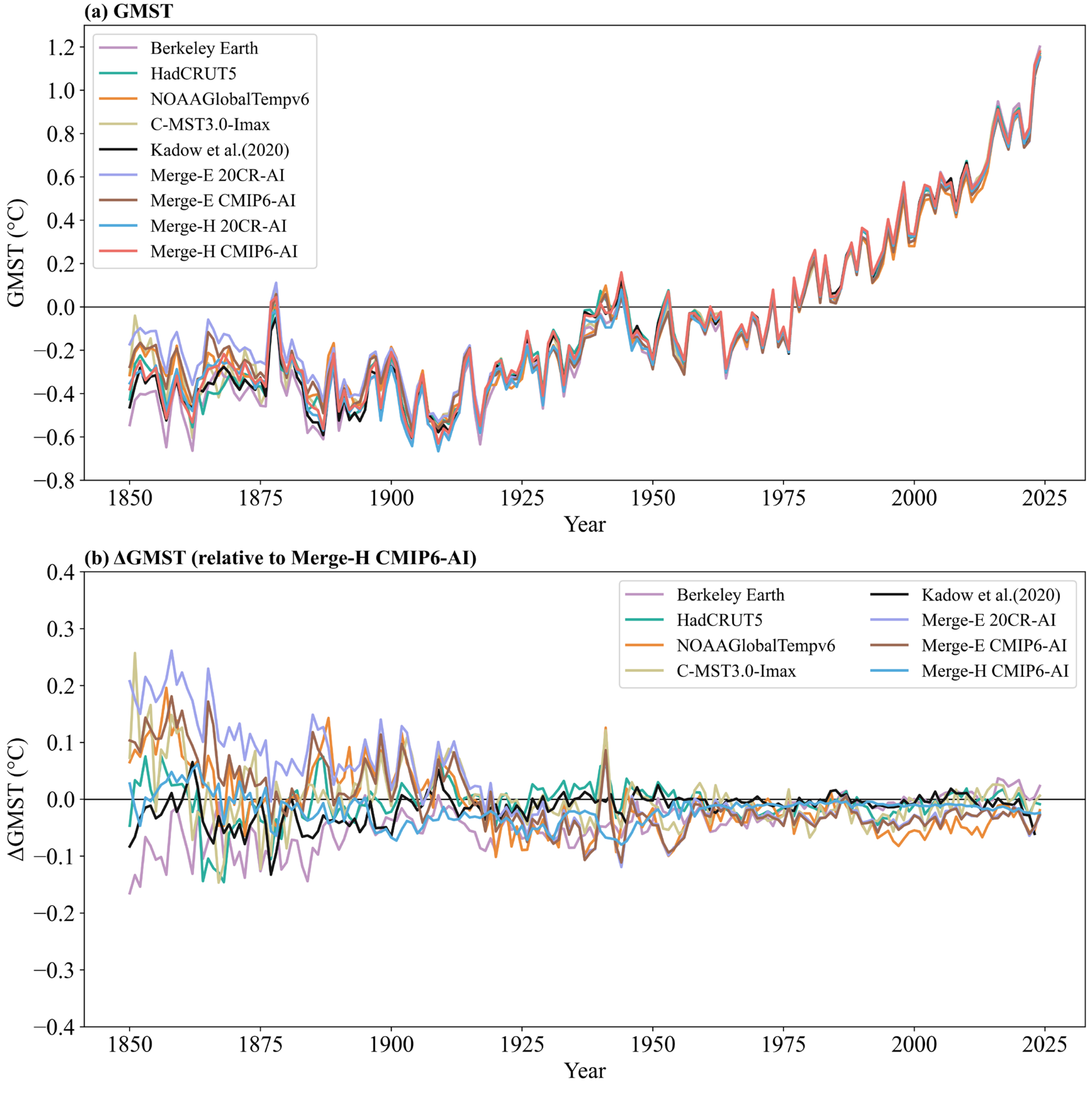


Figure 4: Global mean surface temperature (GMST) time series from 1850 to 2024 (relative to the 1961–1990 climatology). (a) GMST; (b) GMST differences from Merge-H CMIP6-AI.

Table 2: 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 Global mean surface temperature (GMST) time series reconstructed by AI is shown in Figure 4, reflecting the long-term variations of global mean temperature from 1850 to 2024. ~~The AI reconstruction results from Merge-E and Merge-H exhibit overall consistent trends; however, certain differences exist during periods with sparse early observations. Prior to the 1890s, the global ST anomalies in the Merge-E reconstruction are generally higher than those in the Merge-H reconstruction. Given the significant contribution of SST to global annual mean temperature variability, this difference is mainly attributable to the differing SST datasets used in the two reconstructions. PConv rely heavily on the spatial structures present in regions with valid observations when reconstructing climate fields containing extensive missing data (Liu et al., 2018; Reichstein et al., 2019; Toms et al., 2020). In the Merge-E scenario, when early land observations are extremely sparse but oceanic grid points exhibit relatively complete spatial coverage (Fig. S3), the dominant spatial gradients, covariance structures, and anomaly patterns learned by the model during training are inherently governed by the ocean. According to the general properties of deep learning, convolutional neural networks preferentially learn the most frequent and statistically stable features in the input data (LeCun et al., 2015). Consequently, during 1850–1890, when land observations are limited, the PConv model is inevitably dominated by oceanic signals during feature extraction. This causes the model to "fill" missing regions with spatial structures resembling those of the ocean, which are typically smoother and warmer, thereby producing systematic biases in the land driven by ocean-dominated features. Around the 1890s, however, the substantial increase in land-based observation stations (Wei et al., 2025) enriches the spatial information over land. As the spatial patterns and variability of land anomalies begin to occupy sufficient weight in the training data, the PConv model becomes capable of simultaneously learning both land and ocean features. This reduces the early-period dominance of oceanic coverage and allows the reconstruction to gradually converge toward a more realistic combined land-ocean spatial structure. In the~~

~~experiments described above, the SST product ERSSTv6 in Merge-E already incorporates ANN-based spatial infilling, resulting in smoother and more homogeneous oceanic features (Huang et al., 2025a, 2025b). This further amplifies the dominance of oceanic characteristics during feature learning. In contrast, in the Merge-H scenario, both land and ocean fields contain extensive missing data in the early period (Wei et al., 2025; Kennedy et al., 2019), meaning that no single domain provides a strong dominant signal. As a result, the PConv model is more likely to learn spatial structures jointly constrained by both the land and the ocean domains, thereby reducing the risk of domain-specific biases, particularly biases associated with the ocean. Consequently, the Merge-E reconstruction exhibits pronounced overfitting to oceanic structures before the late nineteenth century, producing warmer biases and ultimately yielding a lower long term trend and global warming level (GWL) than that of Merge-H (Table 1).~~ The AI reconstruction results from Merge-E and Merge-H exhibit certain differences during periods with sparse early observations (Fig. 4b). Prior to 1920, the GMST reconstructed by the Merge-E configuration is generally 0.1–0.2 °C higher than that of Merge-H, which is primarily attributed to differences in the underlying SST datasets used in the two approaches.

When dealing with climate fields containing large amounts of missing data, partial convolutional neural networks (PConv) rely heavily on the spatial structures of the regions covered by valid observations during the learning process (Liu et al., 2018). When early land-based observations are extremely sparse, while ocean grid cells have relatively complete spatial coverage, the model predominantly learns temperature gradients, covariance structures, and anomaly patterns from oceanic data during training (LeCun et al., 2015; Reichstein et al., 2019; Toms et al., 2020). Consequently, under conditions of scarce land observations during 1850–1920, the Merge-E reconstruction is consistently warmer than other datasets by 0.1–0.3 °C over the low- and mid-latitudes (Fig. S8), leading to lower long-term trends and GWL estimates compared with Merge-H (Table 2). This bias reflects an over-reliance on ocean-dominated features under early sparse observational conditions, resulting in systematic inconsistencies relative to other reconstruction schemes.

After 1920, however, as the number of land-based observation stations increases substantially (Wei et al., 2025), the PConv model is able to jointly learn both land and ocean characteristics, thereby reducing the bias introduced by early ocean-dominated training and allowing the reconstruction to gradually converge toward a more realistic coupled land–ocean structure. At the same time, since ERSSTv6 already incorporates ANN-based spatial infilling (Huang et al., 2025a, 2025b), the additional information introduced by the AI reconstruction in Merge-E is limited. Based on the above results, we adopt the more physically consistent merging scheme, the Merge-H product generated by combining C-LSAT2.1 with HadSST4, as the primary focus of our analysis.

~~It is noteworthy that the two AI models show small differences in their long-term reconstructions and trends under the Merge-H scenario (Fig. 3, Table 1). For the Merge-H case, the 1850–2024 warming trend produced by both AI reconstructions is 0.064 ± 0.006 °C per decade, which is broadly consistent with the trend in HadCRUT5. The GWL reconstructed by CMIP6-AI reaches 1.52 °C, which is slightly higher than the 1.5 °C estimated from 20CR-AI. This indicates that both AI reconstructions from Merge-H suggest that global warming in 2024 has reached 1.5 °C above the baseline of 1850–1900.~~

~~At the decadal scale, all reconstruction sequences clearly reproduce the pronounced global warming associated with the extreme 1876–1877 El Niño event, which led to the warmest years prior to the 1940s (Huang et al., 2020). From 1850 to the early 20th century, global mean temperatures remained relatively stable, occasionally interrupted by short-term cooling events associated with enhanced volcanic activity and slight decreases in solar radiation (Fang et al., 2022). Between the 1910s and 1940s, global mean temperature experienced the first rapid and sustained warming phase of the 20th century, with a marked increase in magnitude. From the mid-20th century to the mid-1970s, the warming trend slowed, followed by a renewed and significant warming phase starting in the late 1970s, which continues to the present day, reflecting the dominant influence of anthropogenic greenhouse gas emissions (IPCC, 2013, 2021). It is noteworthy that none of the AI reconstruction sequences show a significant "warming slow-down" during 1998–2012, consistent with the findings of Li et al. (2021). Overall, the AI reconstruction sequences shown in Figure 3 reasonably reproduce the phased characteristics of global climate changes over the past 175 years at the annual scale, further demonstrating the robustness and reliability of the AI framework for long-term climate reconstruction.~~

From 1850 to the early 20th century, global mean temperature remained generally stable (Fig. 4a), with occasional short-term cooling signals associated with enhanced volcanic activity and slight decreases in solar radiation (Fang et al., 2022). From the 1910s to the 1940s, the global mean temperature experienced the first rapid and sustained warming phase of the 20th century, with a pronounced increase in magnitude. From the mid-20th century to the mid-1970s, the warming trend weakened, followed by a renewed and more pronounced warming phase from the late 1970s onward, which continues to the present, reflecting the dominant influence of anthropogenic greenhouse gas emissions (IPCC, 2021). Notably, none of the AI-based reconstruction series exhibit a significant “warming slow-down” during 1998–2012, consistent with the findings of Li et al. (2021).

The Merge-H 20CR-AI and Merge-H CMIP6-AI show only minor differences in long-term reconstructed series and trends, and are generally consistent with the HadCRUT5 record (Fig. 4b, Table 2), with overall biases within 0.1 °C. After 1960, differences between the AI reconstructions and other datasets further decrease to within 0.05 °C. Around 1940, several datasets exhibit a warm bias of approximately 0.12 °C relative to the Merge-H CMIP6-AI reconstruction (Fig. 4b), which is attributed to non-standard temperature measurement practices during World War II, including changes in sea surface temperature observation methods (Chan and Huybers, 2021; Kent and Kennedy, 2021).

Both Merge-H AI reconstructions yield a long-term trend of 0.064 ± 0.006 °C/decade over 1850–2024 (Table 2). The global warming level (GWL) for 2024 derived from the Merge-H CMIP6-AI reconstruction reaches 1.52 °C, slightly higher than the 1.50 °C estimated from the Merge-H 20CR-AI reconstruction. This indicates that both AI-based reconstructions suggest that the year 2024 is approaching or has already reached the 1.5 °C global warming threshold relative to the pre-industrial period (1850–1900). Overall, the AI reconstruction series shown in Fig. 4 reasonably reproduce the phase-dependent characteristics of global climate evolution over the past 175 years at interannual scales.

3.3 Comparison of zonal temperature

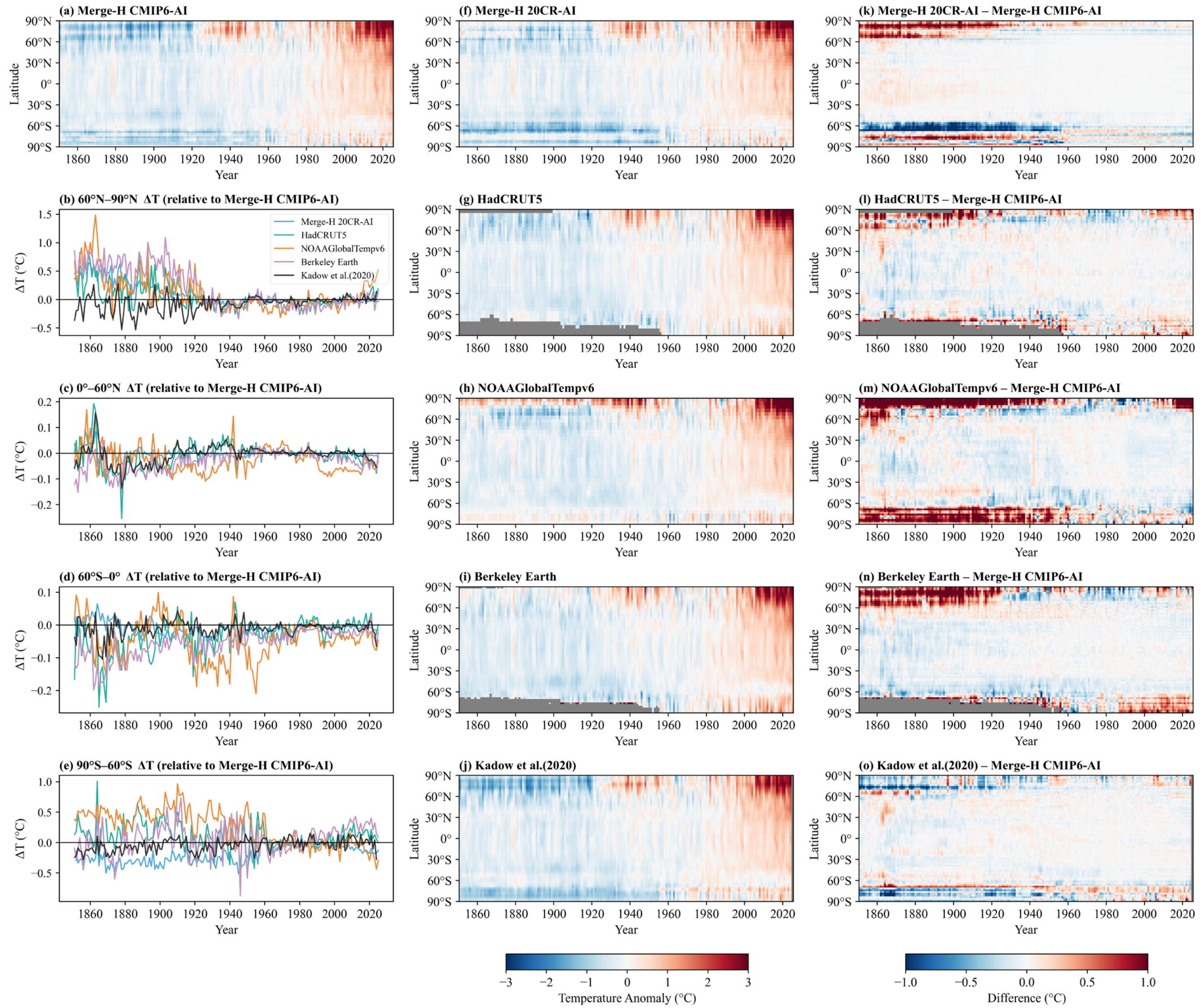


Figure 5 Zonal temperature comparison. (a) Zonal-mean temperature anomalies from Merge-H CMIP6-AI; (b–e) time series of zonal-mean temperature anomaly differences between other datasets and Merge-H CMIP6-AI across different latitude bands; (f–j) same as (a) for Merge-H 20CR-AI, HadCRUT5, NOAAGlobalTempv6,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

In the zonal temperature comparison across different datasets (Fig. 5), it can be seen that within the 60°S–60°N band, where observational sampling is relatively sufficient, the differences between each dataset and Merge-H CMIP6-AI generally remain within ±0.1°C to ±0.15 °C prior to the early 20th century (Fig. 5c, 5d). The Northern Hemisphere, characterized by a larger land fraction and denser observational networks, is substantially better constrained by observations than the predominantly ocean-covered Southern Hemisphere. Consequently, after 1900, the inter-dataset differences within the equator–60°N region rapidly decrease to within 0.1 °C. In contrast, within the equator–60°S region, this convergence is more delayed, with differences only gradually reducing to within 0.1 °C after 1950. Notably, only NOAAGlobalTempv6 exhibits a persistent cold bias of approximately 0.2 °C relative to the other benchmark datasets during 1920–1960 in this region, likely reflecting weaker constraints over the Southern Ocean.

In the Arctic region (Fig. 5b), prior to 1920, both Merge-H CMIP6-AI and Kadow et al. (2020) exhibit a cold bias of approximately 0.6 °C relative to NOAAGlobalTempv6 and Berkeley Earth, while this bias is smaller in Merge-H 20CR-AI. It is also observed that the AI reconstructions (Fig. 5l–5n) show varying degrees of cold bias in the Arctic compared with other datasets before 1920. After 1930, as Arctic observations increase, the differences among datasets rapidly decrease to within 0.2 °C.

In the Antarctic region (Fig. 5e), during 1850–1960, Merge-H 20CR-AI exhibits a cold bias of approximately 0.7 °C relative to other datasets, with a pronounced cold anomaly near 60°S (Fig. 5f, 5k). In the early period of extremely sparse observations over the Southern Ocean, the Merge-H 20CR-AI reconstruction shows a relatively strong systematic cold bias. The Kadow et al. (2020) product and Merge-H CMIP6-AI also remain approximately 0.5 °C colder than NOAAGlobalTempv6, which provides full Antarctic coverage, while the larger variability in HadCRUT5 and Berkeley Earth is likely due to the lack of full spatial extrapolation south of 60°S prior to 1960. As Antarctic observations increase around 1961, differences between the two AI reconstructions rapidly converge to within 0.2 °C. Combined with the validation results in Fig. S4f and S4h, it can be further seen that Merge-H 20CR-AI exhibits a larger number of high-RMSE regions in Antarctica than Merge-H CMIP6-AI, indicating greater instability in its polar reconstruction capability prior to 1961.

3.4 Global warming pattern

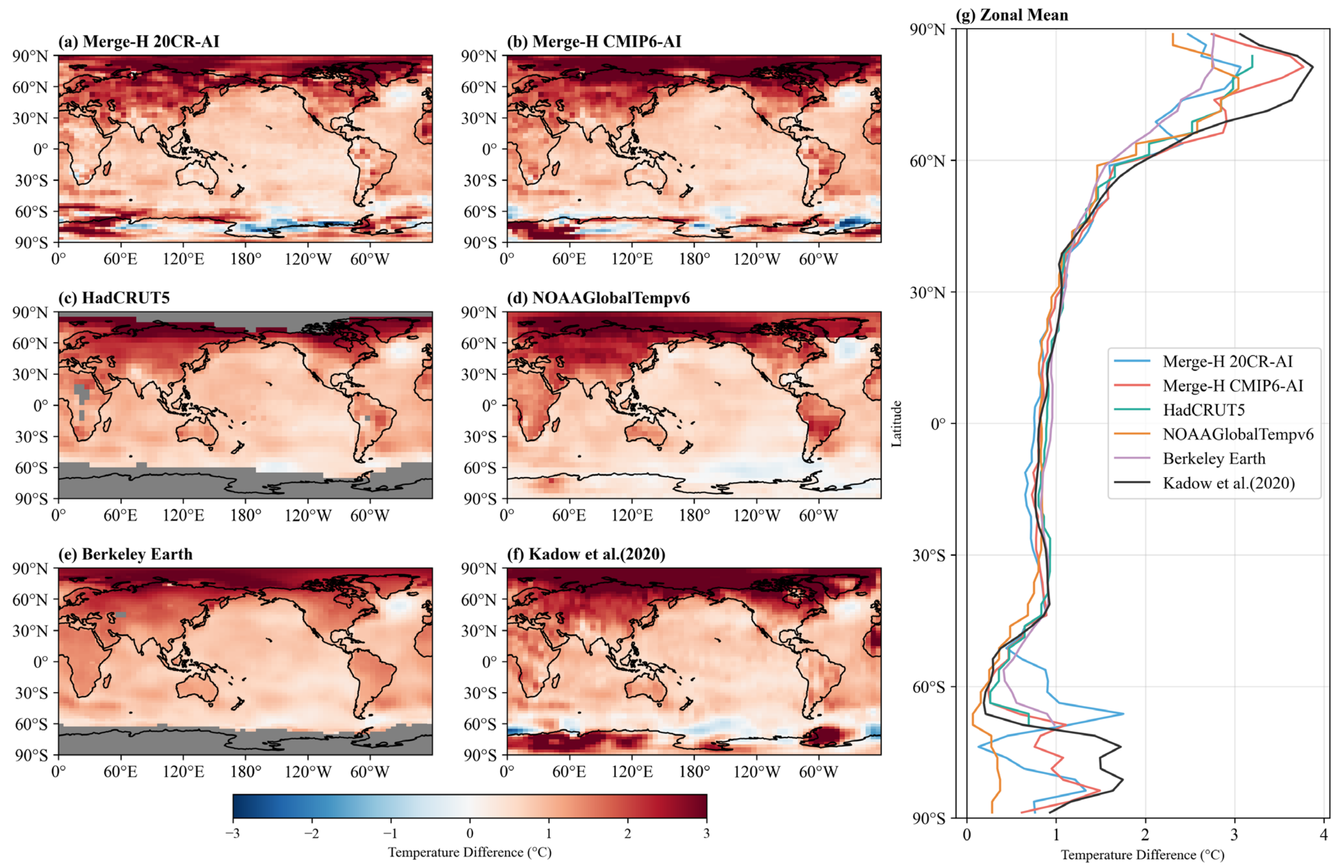


Figure 6 Global warming pattern. (a–f) Spatial distribution of mean global temperature warming over 2005–2020 relative to the 1850–1900 baseline period, (g) zonal-mean profile of warming magnitude.

Under the context of global warming, the spatial patterns of warming reconstructed by the two AI models are generally consistent with those of other datasets (Fig. 6). The main discrepancies are concentrated south of 60°S, where the AI-based reconstructions (Fig. 6a, 6b, 6f) exhibit spatial discontinuities in warming relative to NOAAGlobalTempv6. This may be attributed to the inherently less smooth spatial temperature fields produced by the AI reconstruction during 1850–1900 (Fig. 1 e1, e2, f1, f2).

As shown in Fig. 6g, between 60°S and 80°S, Merge-H 20CR-AI displays zonal variations of ±1 °C relative to Merge-H CMIP6-AI, while the Kadow et al. (2020) reconstruction also shows a pronounced warming increase of up to 1.5 °C near 80°S, which is higher than all other datasets. This indicates that Merge-H 20CR-AI is less stable in this region, consistent with the findings in the previous zonal temperature comparison section. The representation of Arctic amplification (In the context of global warming, temperature changes in the Arctic are significantly greater than the global average) by the AI models also differs among datasets (Fig. 6g), with inter-dataset differences of approximately ±1 °C near 80°S.

The two AI reconstructions based on HadSST data generally show higher consistency in warming patterns with HadCRUT5 and Berkeley Earth, which also use HadSST, compared to NOAAGlobalTempv6, which is based on ERSSTv6. In particular, near 110°W in the Southern Ocean adjacent to Antarctica, NOAAGlobalTempv6 exhibits a larger cooling magnitude than the other datasets. Although all datasets capture the North Atlantic "cold blob," the stronger cooling north of this region in NOAAGlobalTempv6 may represent an artifact of its underlying analysis procedure (Chan et al., 2025), a feature not observed in the other datasets.

Figure 7 presents the spatial distribution of long-term warming trends from 1850 to 2023 across different datasets, as well as the warming trend patterns for three sub-periods: 1850–1910, 1911–1970, and 1971–2023. The results from Merge-H 20CR-AI and Merge-H CMIP6-AI are generally consistent with the reconstruction by Kadow et al. (2020). The spatial gradient of LSAT trends reconstructed by these three datasets is less uniform compared to other datasets, particularly across the continents during 1911–1970. Such differences are attributed to the nonlinear reconstruction capability of PConv. A notable difference between the AI-based reconstructions in this study and NOAAGlobalTempv6 lies in the 1850–1910 period. Specifically, NOAAGlobalTempv6 shows a widespread warming trend over the African continent, whereas the warming magnitude and spatial extent in other datasets. All datasets consistently indicate that, under the background of global warming, the magnitude of the warming trend during 1971–2023 is clearly higher than those in the earlier periods of 1850–1910 and 1911–1970.

Overall, the AI-based reconstructions are able to reasonably reproduce the magnitude and spatial distribution of global warming. However, caution is still warranted in regions covered by sea ice and subject to extremely sparse observational constraints, particularly in the polar regions.

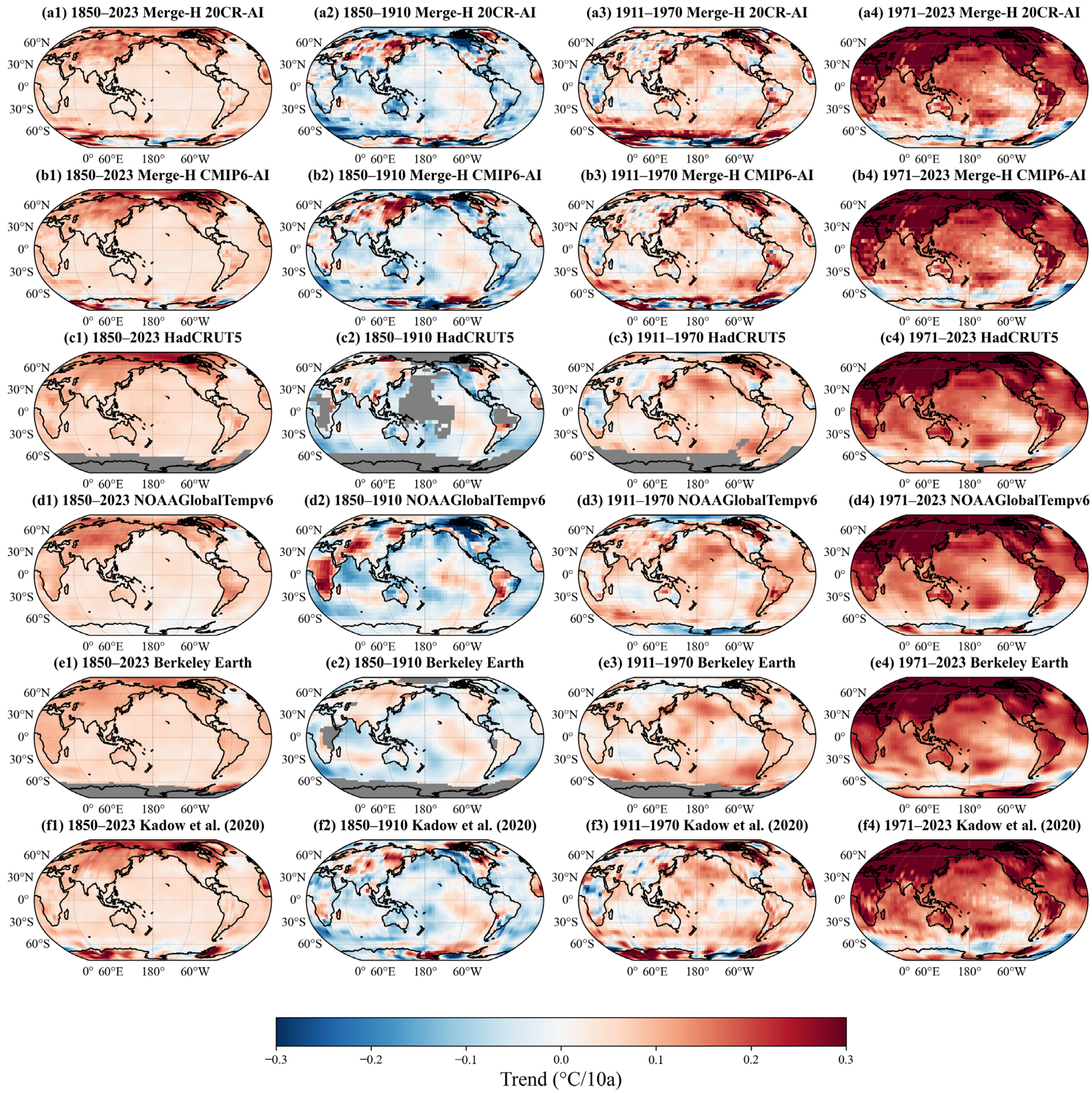


Figure 7 Spatial distribution of global surface temperature trends for 1850–2023, 1850–1910, 1911–1970, and 1971–2023. (a1–a4) Merge-H 20CR-AI, (b1–b4) Merge-H CMIP6-AI, (c1–c4) HadCRUT5, (d1–d4) NOAAGlobalTempv6, (e1–e4) Berkeley Earth, (f1–f4) reconstruction from Kadow et al. (2020).

3.5 Comparison of regional land surface temperature

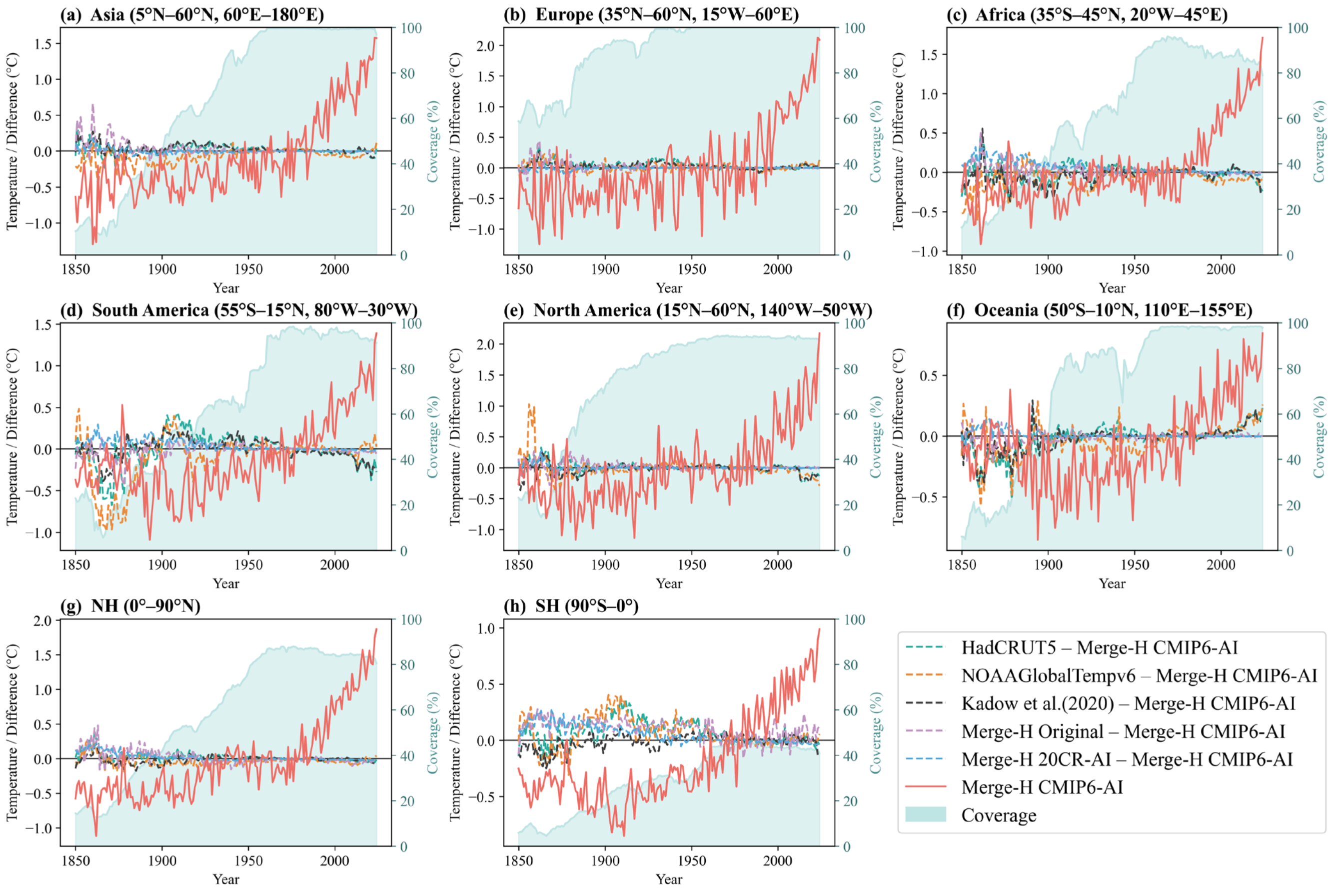


Figure 8: Annual mean surface temperature anomaly time series from 1850 to 2024 in different regions under the Merge-H Scenario. (a–h) Asia, Europe, Africa, South America, North America, Oceania, the Northern Hemisphere, and the Southern Hemisphere. Coverage indicates the original data availability in each region.

~~Global reconstruction of land regions is one of the main focuses of this study. We first analyse the long-term variations of ST in different land regions before and after reconstruction (Fig. 4). In this analysis, the original Merge-E dataset is compared with its two AI-based reconstructions. Considering the extremely sparse observational coverage over Antarctica, this subsection does not discuss the Antarctic region, which will be addressed in detail in Section 4 (Antarctic Reconstruction Results). As a reference for reconstruction performance, the effective data coverage of the original datasets in each region is also presented. Table 2 provides the linear trends of annual mean ST and their 95% CI for different regions from 1850 to 2024.~~ One of the key focuses of this study is the reconstruction of global land surface data (Fig. 8, Table 3). To evaluate the reconstruction performance, we also present the effective data coverage of the original datasets across different regions (Coverage in Fig. 8). In addition, Table 3 summarizes the linear trends of annual mean surface temperature over 1850–2024 for different regions, along with their 95% confidence intervals. The performance of the Antarctic land reconstruction is further evaluated in detail in Section 4, “Antarctic Reconstruction Results.”

Table 3: Temperature trends and 95% confidence intervals (CI) in different regions from 1850 to 2024 (°C per decade).

Region/Dataset	HadCRUT5	NOAA GlobalTempv6	Kadow et al. (2020)	Merge-H Original	Merge-H 20CR-AI	Merge-H CMIP6-AI
Asia	0.088±0.009	0.101±0.010	0.084±0.010	0.083±0.010	0.090±0.010	0.091±0.010
Europe	0.092±0.014	0.097±0.010	0.092±0.014	0.094±0.014	0.099±0.014	0.097±0.014
Africa	0.073±0.008	0.082±0.008	0.079±0.008	0.069±0.009	0.060±0.009	0.074±0.009
South America	0.067±0.007	0.090±0.009	0.058±0.008	0.068±0.009	0.055±0.010	0.064±0.010
North America	0.077±0.012	0.073±0.013	0.079±0.012	0.080±0.012	0.080±0.013	0.083±0.012
Oceania	0.054±0.008	0.050±0.009	0.050±0.008	0.041±0.008	0.035±0.009	0.038±0.008
NH	0.091±0.009	0.096±0.010	0.097±0.009	0.087±0.010	0.089±0.010	0.097±0.010
SH	0.059±0.006	0.059±0.006	0.068±0.006	0.054±0.007	0.050±0.008	0.063±0.007

~~Figure 4 shows that as observational coverage increases, the differences between different reconstruction results decrease markedly. This feature is particularly evident in regions with increasingly abundant observational data, where the AI-reconstructed annual variations become more stable. In contrast, in periods or regions with sparse early observations and low data coverage, discrepancies among reconstructions are more pronounced. The differences between the two AI reconstructions are mainly concentrated in regions such as Africa and South America, where early observations were scarce. Due to the large extent of missing data in these areas, which corresponds to large holes in the input images that need to be filled, the accuracy of the AI reconstructions is significantly affected (Liu et al., 2018). Considering Table 2, the differences between the two AI reconstructions are unavoidable in some regions during periods of low early data coverage. For the Merge-H reconstruction, the two AI models exhibit slight differences in long term trends, with 20CR-AI showing a lower temperature trend in all regions except Europe compared to CMIP6-AI. However, the trends remain within reasonable ranges relative to the pre-reconstruction data. Except for regions with substantial early data gaps, such as Africa, South America, and Oceania, most AI-reconstructed temperature trends exhibit stronger warming compared with pre-reconstruction trends.~~

~~On long timescales, all regions show significant warming trends, particularly since the 1970s, when the warming rate accelerated. Nevertheless, warming is uneven across continents due to differences in response scales and persistence (Li et al., 2022). According to Table 2, in the Merge-H reconstruction scenario, Europe experienced the most pronounced warming between 1850 and 2024, with trends of 0.099 ± 0.014 and 0.097 ± 0.014 °C per decade, whereas Oceania experienced the smallest warming, with corresponding trends of 0.035 ± 0.009 and 0.038 ± 0.008 °C per decade. The warming rate in the SH is noticeably lower than in the NH. This north–south asymmetry under global warming results from significant differences in~~

~~land–ocean distribution, with the Southern Ocean absorbing the majority of heat, leading to greater climate inertia in the SH (Hansen et al., 2010).~~

~~A comprehensive analysis of regional ST anomaly time series indicates that NH land areas contribute most significantly to global land warming. Within the NH, Europe, Asia, and North America are the primary contributors. In conditions of extremely low data coverage, the both reconstructions inevitably exhibit biases. However, as coverage increases, the AI reconstruction experiments show good consistency and stability in these regions.~~

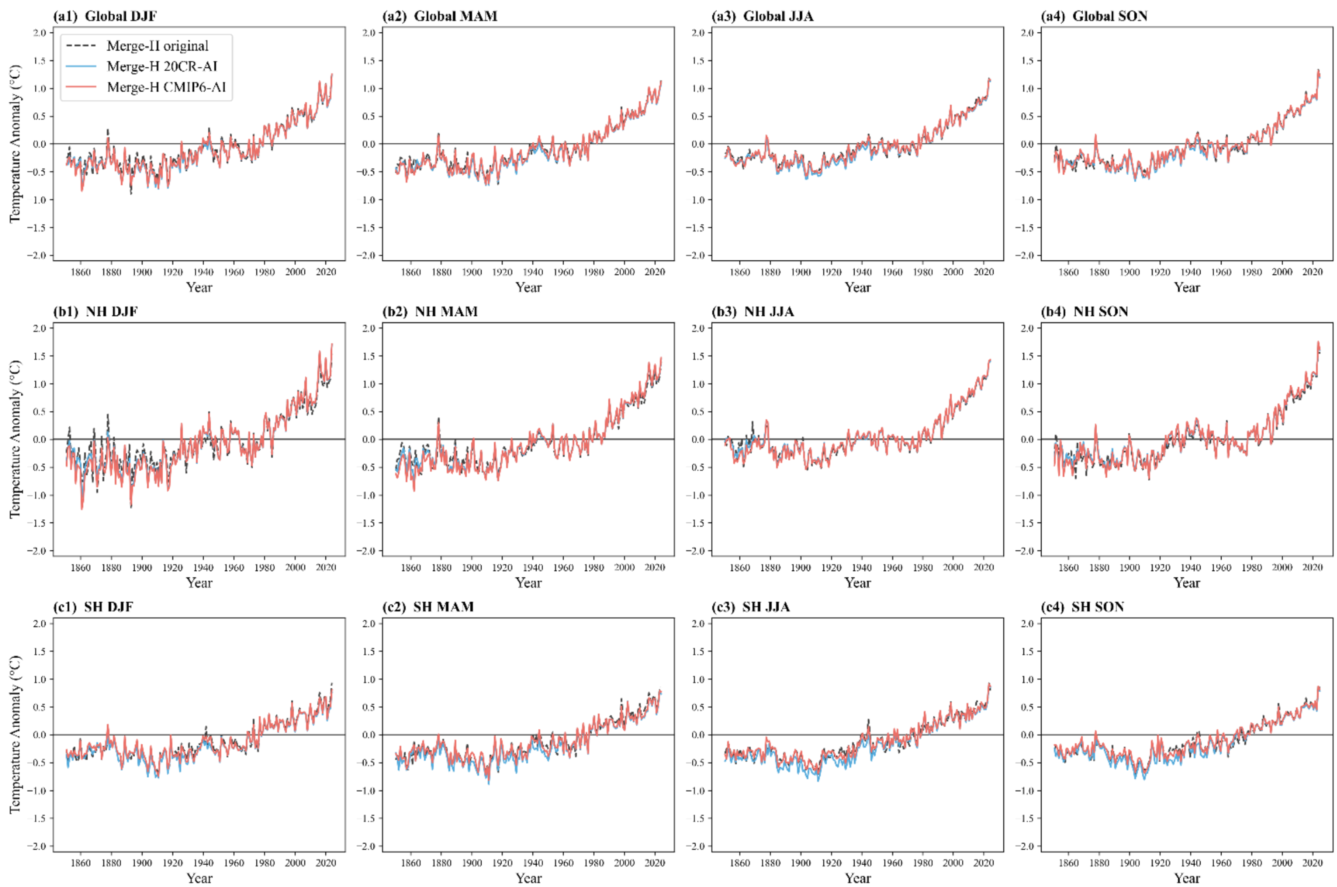
Figure 8 shows that as data coverage increases, the differences between the two reconstruction products and other datasets gradually decrease to within 0.1 °C after 1900. This feature indicates that in regions where observational data become progressively denser, the interannual differences among AI-based reconstructions tend to stabilize. In contrast, during earlier periods or in regions with sparse observations and low coverage, discrepancies among reconstruction products are mainly concentrated in Africa and South America, where early observational records are limited. When effective coverage is below 40% before 1900, Merge-H 20CR-AI is overall about 0.2 °C higher than Merge-H CMIP6-AI (Fig. 8c, 8d, 8h), which also leads to lower long-term trends in these regions for the former in Table 3.

Furthermore, during 1860–1870, when effective coverage in Africa, South America, and Oceania (Fig. 8c, 8d, 8f) drops below a critical threshold (~10%), and considering the strong spatial sampling heterogeneity and lack of inland temperature information in these continents (Wei et al., 2025), the AI reconstructions show varying degrees of deviation from NOAAGlobalTempv6 and HadCRUT5. The most pronounced discrepancy occurs in South America (Fig. 8d), where values are approximately 0.8 °C higher than NOAAGlobalTempv6 and 0.5 °C higher than HadCRUT5. However, around 1920, a systematic negative bias of about 0.3 °C appears. Over the full 1850–2024 period (Table 3), the long-term trend of the AI reconstruction in South America is broadly consistent with HadCRUT5. In Oceania, both Merge-H Original and the two AI-based reconstructions show consistently higher values than other datasets, resulting in lower long-term trends (Table 3), which can be attributed to inherent temperature biases in early land surface products prior to 1900.

On long timescales, regional surface temperature anomalies exhibit significant warming trends, particularly since the 1970s, when warming rates accelerate markedly. However, land regions across continents show heterogeneous warming patterns due to differing response timescales and persistence characteristics (Li et al., 2022). As shown in Table 3, the Merge-H 20CR-AI and Merge-H CMIP6-AI indicate the strongest warming over Europe during 1850–2024, with trends of 0.097 ± 0.014 °C/decade and 0.099 ± 0.014 °C/decade, respectively, while Oceania shows the weakest warming, at 0.035 ± 0.009 °C/decade and 0.038 ± 0.008 °C/decade. The warming rate in the Southern Hemisphere is significantly lower than that in the Northern Hemisphere, reflecting hemispheric asymmetry driven by the large land–ocean contrast and the strong heat uptake by the Southern Ocean, which enhances climate inertia in the Southern Hemisphere (Hansen et al., 2010).

From the integrated analysis of regional surface temperature anomaly series, it is evident that Northern Hemisphere land areas contribute most significantly to global land warming. Within the Northern Hemisphere, Europe, Asia, and North America are the dominant contributors. Under conditions of extremely low data coverage, both reconstructions inevitably exhibit biases; however, as coverage increases after 1900, the AI-based reconstruction results show improved consistency and stability across these regions.

~~3.6 Characteristics of seasonal surface temperature changes~~



~~Figure 5: Seasonal surface temperature anomaly time series from 1850 to 2024 under the Merge-H scenario. Global winter, spring, summer and autumn(a1–a4), Northern Hemisphere winter, spring, summer and autumn (b1–b4), Southern Hemisphere winter, spring, summer and autumn (c1–c4).~~

~~The seasonal ST anomaly time series from 1850 to 2024 are shown in Figure 5. It can be seen that all seasons exhibited a slight cooling trend prior to the 20th century. The period from the 1910s to 1940s corresponds to the first notable warming phase, followed by a relatively stable trend in global land temperatures during the 1940s–1970s. Since the late 1970s, global land temperatures have shown a pronounced and rapid warming trend, consistent with the overall global temperature changes.~~

~~From the perspective of interannual variability, winter (DJF) temperatures in the NH exhibit the largest interannual fluctuations, whereas summer (JJA) shows the smallest variability. This contrast is mainly attributed to the dominance of the NH. The larger interannual variability in winter land temperatures is associated with a relatively unstable atmospheric~~

circulation, enhanced planetary wave activity, and strong snow and sea ice feedbacks, which make land temperatures more sensitive to external perturbations. In contrast, the smaller interannual variability in summer is due to the dominance of solar radiation, a more stable boundary layer, and land processes such as evapotranspiration and soil moisture that provide thermal buffering, thereby reducing the impact of atmospheric disturbances on temperature (Jones et al., 2014). The four seasons in the SH do not exhibit significant differences in interannual variability, reflecting stronger thermal inertia and lower year-to-year fluctuations.

Table 3: Seasonal temperature trends and 95% CI from 1850 to 2024 (°C per decade).

Season/Dataset		Merge-H Original	Merge-H 20CR-AI	Merge-H CMIP6-AI
Global	DJF	0.062±0.007	0.067±0.007	0.068±0.007
	MAM	0.064±0.007	0.067±0.007	0.068±0.007
	JJA	0.057±0.006	0.059±0.007	0.058±0.006
	SON	0.062±0.007	0.063±0.007	0.062±0.006
NH	DJF	0.073±0.010	0.084±0.010	0.090±0.010
	MAM	0.071±0.008	0.078±0.008	0.084±0.008
	JJA	0.056±0.008	0.057±0.008	0.062±0.008
	SON	0.070±0.008	0.071±0.009	0.075±0.009
SH	DJF	0.051±0.006	0.050±0.006	0.046±0.006
	MAM	0.056±0.006	0.056±0.006	0.051±0.006
	JJA	0.056±0.005	0.061±0.006	0.055±0.006
	SON	0.052±0.006	0.054±0.006	0.049±0.006

As shown in Table 3, the reconstructed global temperature trends are fastest in DJF, with trends of 0.067 ± 0.007 and 0.068 ± 0.007 °C per decade, followed by MAM (0.067 ± 0.007 and 0.068 ± 0.007 °C per decade), while JJA shows the slowest warming rate (0.059 ± 0.007 and 0.058 ± 0.006 °C per decade), with the NH contributing the most to the long-term warming trend. A further comparison of different AI reconstruction scenarios indicates that CMIP6-AI reconstructions generally exhibit slightly higher seasonal temperature trends than the 20CR-AI reconstructions, whereas in the SH the former shows lower trends than the latter. Overall, the reconstructed seasonal temperature series capture the major climate change features revealed by both observations and models. The slight differences between CMIP6-AI and 20CR-AI highlight the importance of the coverage, spatial characteristics, and training samples of the original datasets for AI reconstruction performance, providing a reference for further improving multi-source climate data fusion and AI-based reconstruction methods.

4 Antarctic Reconstruction Results

Table 4: Validation between Antarctic station observations and monthly temperature anomalies from AI-reconstructed grid cells under two reconstruction schemes (r: correlation coefficient; RMSE: root mean square error, °C)

Station	Latitude	Longitude	Elevation(m)	Coverage	Merge-H 20CR-AI		Merge-H CMIP6-AI	
					r	RMSE	r	RMSE
Dome A	−80.4	77.4	4084	2006–2019	0.52	1.28	0.64	1.17
Drescher	−79.2	−19.0	27	1993–2003	0.60	0.72	0.56	0.77
Ferrel	−77.9	170.8	45	2007–2023	0.84	0.65	0.88	0.58
G3	−70.9	69.9	84	2002–2020	0.49	0.79	0.44	0.81
GC41	−71.6	111.3	2763	1984–2005	0.80	0.57	0.96	0.43
General Belgrano	−78.0	−38.8	32	1956–1978	0.70	0.76	0.69	0.77
GF08	−68.5	102.1	2125	1987–2007	0.87	0.46	0.87	0.45
LG10	−71.3	59.2	2619	1993–2005	0.73	0.43	0.71	0.43
LG35	−76.0	65.0	2345	1994–2007	0.70	0.51	0.64	0.53
LG59	−73.5	76.9	2565	1994–2003	0.63	0.45	0.54	0.48
Mount Siple	−73.2	−127.1	230	1993–2006	0.84	0.34	0.84	0.34
Nordenskiold	−73.1	−13.4	497	1995–2019	0.82	0.31	0.81	0.32
Russkaya	−74.7	−136.9	100	1981–1989	0.73	0.66	0.69	0.70
Thurston Island	−72.5	−97.6	225	2011–2021	0.74	0.56	0.85	0.43
Average (14 stations)					0.72	0.61	0.72	0.59

Among the 81 Antarctic stations used in this study (Table S2 and Fig. S1), the vast majority began recording Antarctic climate data after the 1957/1958 International Geophysical Year. Even so, the effective data coverage in Antarctica remained only around 8% after 1961 (Figure 9a). During model validation for Antarctic data (Fig. S6), the reconstructed results from both AI models showed high performance, with correlation coefficients with the test dataset reaching approximately 0.9 and RMSE below 1.3 °C after 1961. Due to the scarcity of observational data prior to 1961, validation and assessment of Antarctic data for earlier years are not feasible. Therefore, this study focuses on the systematic validation and analysis of Antarctic reconstructed data from 1961 onward.

To better evaluate the performance of the AI models in reconstructing ST over the Antarctic continent since 1961 and to ensure the validity of the verification, we selected 14 independent Antarctic stations that were not used in the reconstruction and were located outside the ~~effective grid cells~~ spatial coverage of the original observational grid before reconstruction. The monthly observed anomalies at these stations were compared with the corresponding reconstructed anomalies to assess the consistency and statistically test whether the reconstructions reliably reproduced temperature variability (Monaghan et al., 2008; Nicolas et al., 2014; Bromwich et al., 2025a). The method involves extracting the gridded data from the reconstructed product over the same time period as the station observations, and then performing a consistency check after removing the mean from each respective time series. The results indicate that most stations exhibit high correlation with the reconstructed grid cell temperature series, with RMSE below 0.9°C except for Dome A (Table 4). Dome A, located at the highest point of the East Antarctic Plateau (Table S2, Fig. S9), is influenced by surface winds controlled by synoptic-scale circulation, and its temperature records can reach extreme lows, making it more challenging for the reconstruction method to capture its variability (Scambos et al., 2018). Overall, the average correlation between the 14 stations and the two reconstruction schemes is 0.72, with corresponding mean RMSE values of 0.61°C and 0.59°C. Although there are slight differences among the reconstruction schemes, the CMIP6-AI reconstruction performs slightly better than the 20CR-AI reconstruction, and both schemes show good agreement with the observational data.

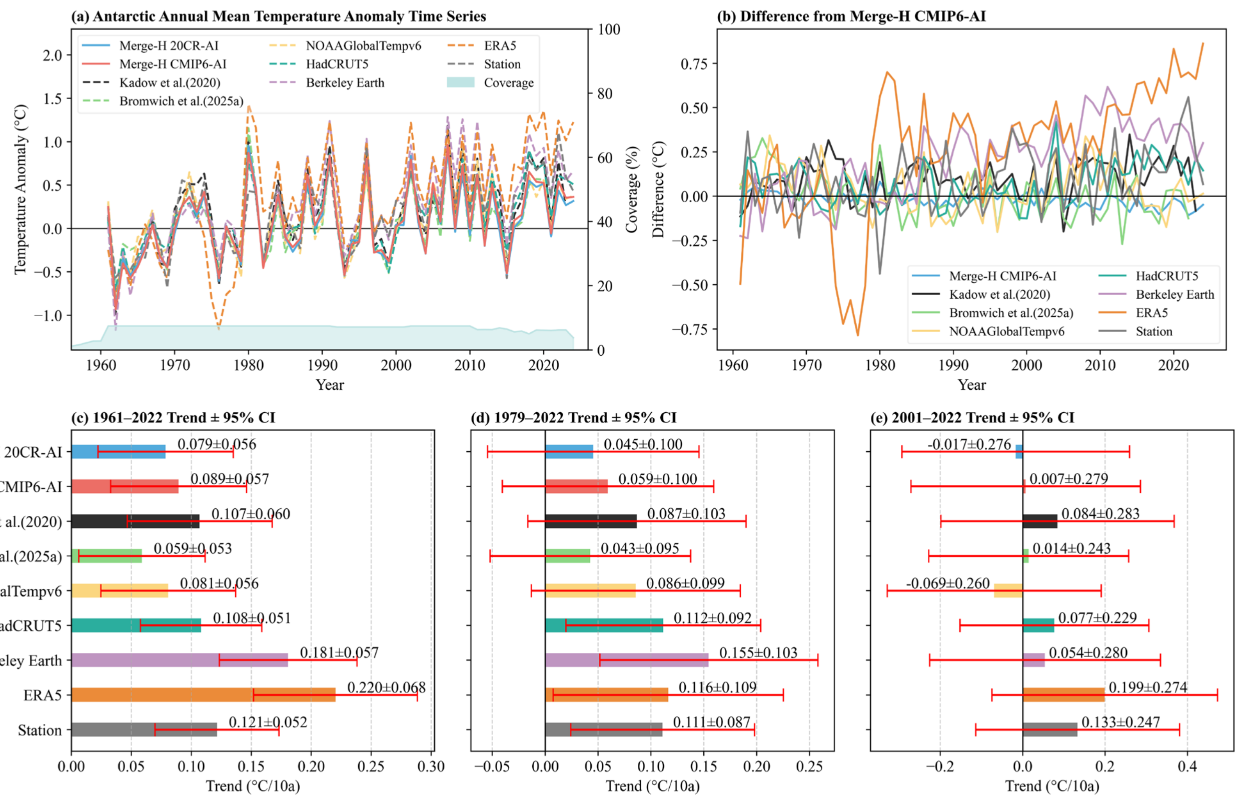


Figure 9: (a) Annual mean surface temperature anomaly time series over Antarctica from 1961 to 2024, and coverage refers to the proportion of Antarctic grid cells that are occupied by observational grid points before reconstruction; (b) Temperature difference from Merge-H CMIP6-AI; (c–e) Linear trend of annual mean temperature and 95% confidence interval (°C per decade) during 1961–2022, 1979–2022 and 2001–2022.

Figure 9 presents the annual mean ST anomalies over the Antarctic continent for 1961–2024, along with the linear temperature trends and their 95% CI for three ~~representative~~ periods. To comprehensively evaluate the reliability and performance of the AI reconstructions, the results were systematically compared with the ~~arithmetic~~ spatially weighted mean of 81 Antarctic observational stations used in this study (after mapping the station data onto grid cells; Station in Figure 9), the reconstruction from Kadow et al. (2020) and Bromwich et al. (2025a), HadCRUT5, NOAAGlobalTempv6, Berkeley Earth, and the ERA5 reanalysis product. The comparisons include a quantitative assessment of Antarctic annual mean temperature time series (Fig. 9) and Antarctic subregional temperature series (Fig. S11), as well as a qualitative analysis of the spatial distribution of temperature trends (Figs. 10 and S12).

~~Overall, the AI reconstructions exhibit high consistency with multiple observational and reconstruction products in terms of interannual variability (Fig. 8a). In particular, the reconstructed anomalies closely track the phases of strong cold and warm years, indicating that the AI method reliably captures interannual climate signals in the Antarctic region. During 1961–2024, the linear warming trends derived from the two AI reconstruction schemes are 0.078 ± 0.053 and 0.089 ± 0.053°C per decade, respectively (Fig. 8c). The trends are similar in magnitude, show minor differences, and are both statistically significant at the 0.05 level, demonstrating a pronounced warming trend over the Antarctic continent since the 1960s.~~

~~Notably, since 1979, the warming trends in the AI reconstructions remain non-significant, consistent with Bromwich and NOAAGlobalTempv6, whereas HadCRUT5, Berkeley Earth, and ERA5 still show significant warming trends, in agreement with the arithmetic mean of the station data (Station) (Fig. 6c). The higher trend estimates in these products may stem from biases introduced by statistical and reanalysis methods under sparse observational conditions in polar regions. Previous studies indicate that the ERA5 reanalysis system exhibits substantial seasonal biases in near-surface Antarctic temperatures, likely due to weak constraints on surface turbulent processes and increased assimilation errors under strong inversions and low-friction conditions (Garza-Girón et al., 2024; Yang et al., 2025). Additionally, Berkeley Earth's spatial statistical extrapolation may produce systematic overestimation in high-latitude Antarctica (Rohde et al., 2020), although the exact contribution of statistical and assimilation methods to these biases remains to be quantified. These results highlight that systematic differences among data sources must be carefully considered when assessing long-term climate trends in polar regions.~~

Overall, the AI reconstructions show high consistency with multi-source observations and reconstruction products (Bromwich et al., 2025a) in capturing interannual variability, particularly in the phase evolution of temperature anomalies during strong warm and cold events, which is largely synchronized with observations and representative reanalysis datasets (Fig. 9a). The differences between the two AI reconstruction approaches and Bromwich et al., NOAAGlobalTempv6, and HadCRUT5 are generally within 0.25 °C (Fig. 9b).

During 1961–2022, the linear warming trends derived from Merge-H 20CR-AI and Merge-H CMIP6-AI are 0.079 ± 0.056 °C/decade and 0.089 ± 0.057 °C/decade, respectively. These magnitudes (Figs. 9c–9e) are close to those reported by Bromwich et al. (2020) and NOAAGlobalTempv6, and are lower than those from HadCRUT5, Berkeley Earth, ERA5, and station-based datasets. However, it should be noted that the station data used to construct the Station-based series exhibit a highly uneven spatial distribution, with observations primarily concentrated in regions experiencing stronger warming (Fig. S3, Fig. 10). This sampling structure causes the Station-based estimates in Fig. 9 to better represent localized signals rather than the pan-Antarctic mean state, thereby potentially leading to an overestimation of the warming trend at the continental scale of Antarctica.

Since 1979, the trends from the AI reconstructions remain statistically insignificant, consistent with Bromwich et al. and NOAAGlobalTempv6, whereas HadCRUT5, Berkeley Earth, and ERA5 still exhibit statistically significant warming trends (Fig. 9d). This apparent overestimation in these products may arise from limited observational sampling over Antarctica and spatial smoothing or extrapolation procedures, resulting in reduced regional representativeness (Morice et al., 2021; Rohde et al., 2020). As shown in Fig. S11e and S11h, this bias mainly originates from warming contributions over East Antarctica. In addition, Bromwich et al. (2024) show that the pronounced Antarctic warming trend in ERA5 prior to 1979 may be overestimated, primarily due to a cold bias in the ECMWF model. Given the extremely sparse observations over the Southern Ocean and limited assimilation capability of early satellite records, this bias cannot be effectively constrained. After 1979, ERA5 exhibits further enhanced warming trends near the 0° longitude coastal sector, the Ross Ice Shelf, and Marie Byrd Land, which further increases the overall temperature trend. These results highlight that systematic differences among datasets must be treated with caution when assessing long-term climate change trends in polar regions.

Since 2001, both the AI reconstructions and all datasets except ERA5 indicate that Antarctic warming is not statistically significant (Fig. 9d). During this period of accelerated global mean temperature increase (Fig. 4a), no significant warming is detected over Antarctica, suggesting that internal climate variability continues to exert a stronger influence on Antarctic climate than external forcing, warranting further attention.

~~Overall, the AI reconstruction results demonstrate reasonable temporal consistency, trend significance, and magnitude of variability, validating the feasibility and potential utility of AI-based climate reconstruction for the Antarctic region.~~

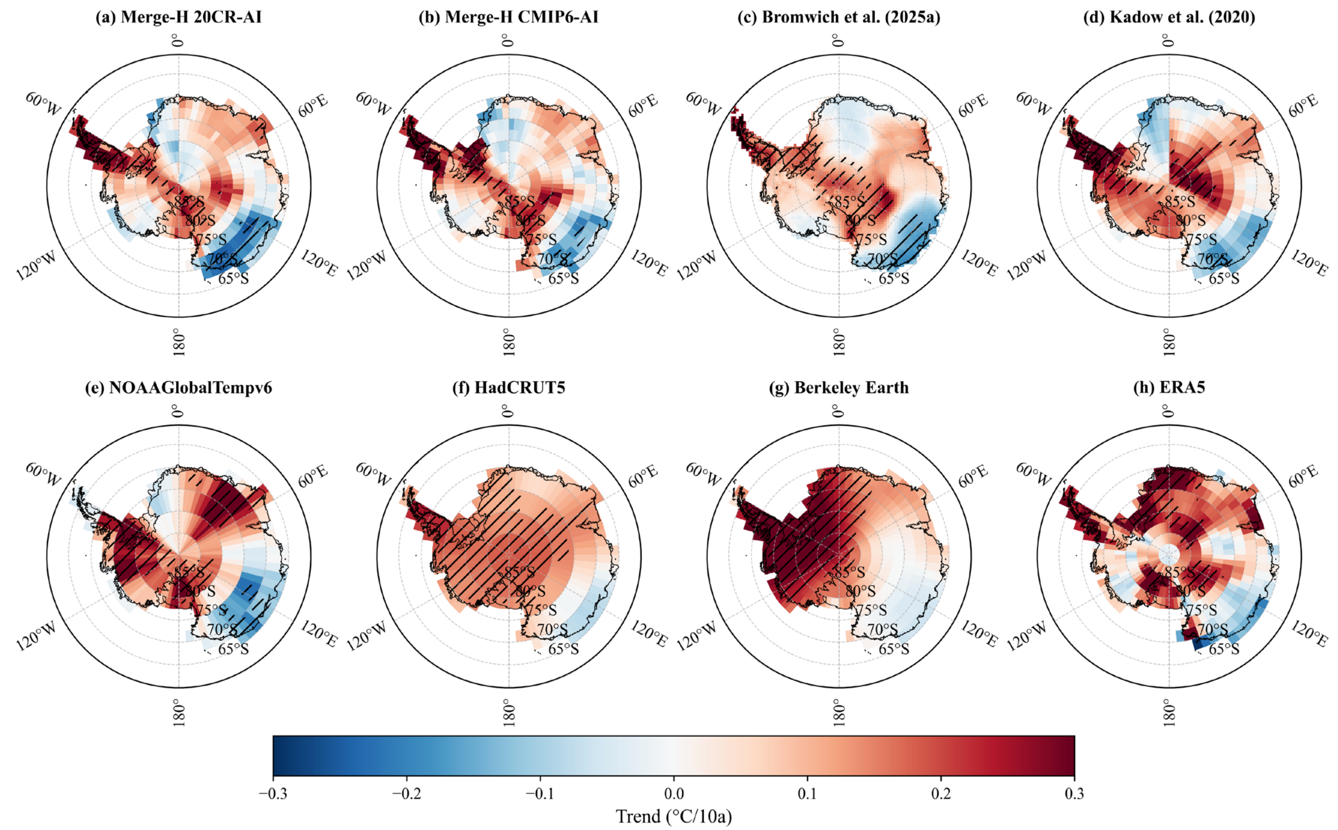


Figure 10: Spatial distribution of Antarctic annual mean surface temperature trends for 1979–2024 ("///" indicates regions where the temperature trend is statistically significant at the 0.05 level).

The spatial pattern of Antarctic temperature change is of critical scientific importance for understanding regional climate variability and its potential driving mechanisms. Due to substantial differences in topography, sea ice coverage, circulation features, and ocean–atmosphere interactions across the Antarctic continent, temperature changes are not uniformly distributed but exhibit complex regional responses (Turner et al., 2005, 2019; Marshall, 2003, 2006). Figure 10 presents the spatial distribution of ST trends and their statistical significance for 1979–2024 derived from the two AI reconstruction schemes, together with the results from other representative datasets. ~~Overall, both AI reconstruction schemes reveal significant warming over the Antarctic Peninsula, near the Ronne Ice Shelf, and the northeastern Ross Ice Shelf, while the Wilkes Land coast shows significant cooling. These spatial patterns are broadly consistent with the reconstructions of Bromwich, the NOAAGlobalTempv6 dataset, and reported observations or model simulations in these regions (Vaughan et al., 2003; Wang et al., 2025; Darelius et al., 2016; Clem et al., 2020; Sheehan et al., 2024), demonstrating the reliability of AI reconstructions in capturing key climate signals in Antarctica.~~

~~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). As a result, their temperature trend fields are smoother and have weaker spatial gradients. All three datasets indicate significant warming over the Antarctic Peninsula, with HadCRUT5 and Berkeley Earth also showing extensive warming near the South Pole, while GISTEMPv4 indicates significant warming east of the Ross Ice Shelf and in eastern Queen Maud Land. Notably, NOAAGlobalTempv6 also shows warming in eastern Queen Maud Land, whereas the two AI reconstruction schemes, reconstruction from Bromwich, and ERA5 reveal positive but statistically insignificant trends in this region, which lies in the transitional zone of Antarctic continental temperature variability.~~

~~In summary, the AI reconstructions successfully capture the main regional temperature trends and their statistical significance across Antarctica, reflecting climate signals consistent with observations and reanalysis data while exhibiting plausible spatial variability in data-sparse areas. This demonstrates that AI-based climate reconstruction not only provides high temporal consistency but also possesses strong potential for spatial applications, offering a novel tool for understanding Antarctic climate change processes and their underlying drivers.~~

Overall, all datasets consistently reveal significant warming over the Antarctic Peninsula, the vicinity of the Ross Ice Shelf, and the northeastern sector of the Ross Ice Shelf (see Fig. S10 for Antarctic geographical references), while significant cooling is observed along the coast of Wilkes Land. In particular, the Antarctic Peninsula exhibits pronounced warming, with the AI reconstructions yielding trends of 0.301 ± 0.115 °C/decade and 0.311 ± 0.113 °C/decade, respectively (Fig. S11). For the annual mean time series, all datasets show biases of approximately ±1 °C. In the relatively large regions of West Antarctica and East Antarctica, the differences between the AI reconstructions and other datasets are within 0.5 °C and 0.25 °C, respectively, and the regional temperature trends are generally consistent with the Station-based estimates.

The spatial pattern of Antarctic temperature trends is broadly consistent with the reconstruction results of Bromwich et al. and NOAAGlobalTempv6, and is also in agreement with observational evidence or model-based studies in these regions (Vaughan et al., 2003; Wang et al., 2025; Darelius et al., 2016; Clem et al., 2020; Sheehan et al., 2024). Notably, the reconstruction of Kadow et al. (2020) (Fig. 10d) exhibits a clear artifact near 0° longitude, where the temperature trend shifts abruptly from significant warming east of 0° to cooling west of it, indicating a pronounced spatial discontinuity. This issue is largely mitigated in the present reconstruction (Figs. 10a, 10b), which, consistent with Bromwich et al. and NOAAGlobalTempv6, does not show widespread significant warming near 0° longitude. This highlights that AI-based reconstructions of Antarctic climate signals require adequate observational sampling support.

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. This is also one of the reasons why the temperature trends in Fig. 8d are higher in the former two datasets than in the latter two. It is worth noting that NOAAGlobalTempv6 also shows significant warming in eastern Queen Maud Land, while the AI reconstructions, Bromwich et al., and ERA5 all indicate a certain warming signal in this region. However, since this area lies within a transitional zone of cold and warm variability over the Antarctic interior, the warming signal does not reach statistical significance.

Overall, the AI reconstructions successfully capture the spatial distribution of temperature trends and their statistical significance across major Antarctic regions. They not only reproduce climate signals broadly consistent with observational and reanalysis datasets, but also exhibit reasonable spatial variability in data-sparse regions. This suggests that AI-based climate reconstruction methods provide robust and reliable temporal evolution of Antarctic climate after 1961, when observational constraints become available, and also demonstrate good skill in representing spatial structures. These results provide a new approach for better understanding Antarctic climate variability and its underlying driving mechanisms.

5 Limitations and future perspectives

AI reconstruction results exhibit high consistency with existing observational and reconstructed datasets in terms of spatial structure, long-term trends, and interannual variability, demonstrating the feasibility and reliability of this AI approach for climate reconstruction. However, compared with traditional statistical interpolation or reanalysis models, AI methods still present challenges regarding model dependency, physical interpretability, and long-term consistency that warrant further investigation.

The AI reconstruction approach employed in this study is based on image inpainting techniques, which treat climate fields as two-dimensional images and use PConv to infer missing data from learned spatial features in local neighborhoods. This method performs well in mid-latitude and low-latitude regions where spatial continuity is relatively strong, but because the computation relies on planar convolution operations, it cannot fully account for the spherical geometry of the Earth. Consequently, systematic limitations remain in high-latitude regions. Similarly, Kadow et al. (2020) and Bochow et al. (2025) highlighted that whether using partial convolution or Fourier convolution, while historical climate fields can be reasonably reconstructed, projecting the spherical Earth onto a two-dimensional equidistant grid introduces geometric distortion at high latitudes, affecting continuity on the sphere and potentially generating "edge discontinuities" or "artifacts" at the poles or at the image seams (at the 0°/360° boundary of the global map). The "artifact" features observed in the edge reconstruction of our images support this observation (Fig. 2 and 10d). ~~Esteves et al. (2023) reported that scaling spherical convolutional neural networks (Scaling Spherical CNNs) can directly model data on the sphere, achieving competitive results in weather forecasting tasks, demonstrating the potential of spherical convolutional frameworks in atmospheric sciences. However, high-resolution global climate reconstruction with such methods remains significantly limited by current GPU memory capacity and the high cost of high-performance computing. Overall, balancing spherical geometric accuracy with computational feasibility while maintaining global spatial continuity is a key direction for future AI climate reconstruction.~~

~~Integrating spherical deep learning structures with physical constraints may enable more realistic and accurate reconstructions of polar climate variability.~~ To address this issue, Esteves et al. (2023) reported that scaling spherical convolutional neural networks (Scaling Spherical CNNs) can be used to model spherical data; however, their application remains constrained by current GPU memory limitations and the high computational cost of high-performance computing resources.

The stability of AI reconstruction results largely depends on the spatial patterns of the training samples and the availability of original valid data. ~~In this study, models trained on CMIP6 outputs generally captured higher long-term trends than those trained on 20CR, particularly, systematic biases also exist between the Merge-E and Merge-H reconstruction schemes.~~ In this study, the AI-based reconstruction approach inevitably exhibits systematic biases in polar regions during the early period without observational records. This indicates that while deep learning methods possess strong nonlinear fitting capabilities, they are constrained by the spatial structure and statistical characteristics of the input data. Moreover, we found that when missing regions are excessively large and contain no valid data, the AI model's prior correlation and RMSE deteriorate substantially, as observed in the pre-1961 polar regions. After introducing station observations post-1961 to increase effective coverage, the prior correlation improved and RMSE decreased, particularly for Merge-H, which had larger missing regions than Merge-E (Fig. S6). This finding emphasizes that while AI reconstruction can generate plausible climate fields under sparse observational conditions, sufficient image information or spatial constraints are required. Because the AI model does not explicitly include energy conservation or atmospheric dynamics constraints, extrapolations to data-sparse regions, such as early periods and polar regions, heavily rely on surrounding temperature spatial patterns, that is, the edge features of the missing regions (Liu et al., 2018). Thus, land–ocean boundaries, complex terrain, and extensive missing-data areas are major sources of AI reconstruction errors. Future studies may improve reconstructions by incorporating terrain corrections, adjusting land–ocean weighting, introducing local dynamical parameterizations, or using physics-informed loss functions. ~~Furthermore, combining AI's ability to capture nonlinear features and local spatial patterns with the interpretability, computational transparency, and uncertainty quantification advantages of statistical methods, forming "AI plus physics-driven" or "AI plus statistical fusion" frameworks, may represent a promising direction for climate reconstruction.~~

~~AI method demonstrates strong potential and scalability in climate reconstruction, yet its application in modeling and reconstructing climate systems remains an evolving area. The physical consistency and long-term stability of current AI reconstruction outputs still require further validation under more stringent constraints. Future research should incorporate climate dynamics and energy balance constraints while preserving the nonlinear fitting advantages of deep learning to ensure interpretability and generalizability in complex climate contexts. With continued advancement in high-performance computing and climate big data, AI reconstruction methods may achieve higher resolution and stronger constraint global climate field reconstructions. By integrating spherical neural networks, physical information, and multi-source data fusion frameworks, future AI reconstruction systems could more accurately reproduce climate evolution in data-sparse regions, supporting studies on polar amplification, local climate change, and data monitoring, detection, and evaluation. Overall, AI~~

~~is expected to play an increasingly important role in climate science, providing a robust technological foundation for understanding past and projecting future global and regional climate change.~~

6 Data availability

The C-AIRST$_{R/M}$ datasets are publicly available on the website at https://doi.org/10.6084/m9.figshare.30663797.v1 (Ouyang et al., 2025). They can also be accessed at http://www.gwpu.net/en/h-col-103.html (last access: 21 November 2025) for free.

7 Code availability

The code utilized in this project can be downloaded here at https://doi.org/10.5281/zenodo.19428292.

8 Conclusions

This study employed an AI model (PConv) to reconstruct global ST fields from two fused observational datasets. The results demonstrate that the AI reconstructions ~~shows high consistency~~ are broadly consistent with multiple representative climate datasets. ~~in terms of interannual variability, long-term temperature trends, and spatial patterns, validating the effectiveness and reliability of deep learning-based image inpainting approaches for climate reconstruction tasks.~~ Both Merge-H 20CR-AI and Merge-H CMIP6-AI datasets exhibit high temporal and spatial continuity in low- and mid-latitude regions and over Antarctica after 1961, providing a solid foundation for extending long-term climate records, assessing polar climate change, and supporting climate monitoring, detection, and attribution. ~~A comprehensive comparison of the global mean ST time series derived from different reconstruction schemes in this study indicates that the datasets exhibit largely consistent long-term trends. In particular, for Antarctica after 1961, the AI reconstruction aligns well with both observational and reanalysis data, indicating that the AI-based approach provides reliable support for reconstructing continuous global ST fields since the mid-19th century.~~ The main conclusions of this study are as follows:

~~(1) Convolutional neural network-based image inpainting can effectively fill global temperature data gaps. Both the 20CR-AI and CMIP6-AI models, which are trained on large sample datasets, produced satisfactory reconstruction results. The differences between the two models in reconstruction quality and trend estimation are minor, demonstrating good generalization and stability. Reconstruction performance improves significantly with increased data coverage, highlighting the importance of data completeness for model performance.~~

~~(2) AI reconstructions provide a novel tool for climate system research. Compared with traditional statistical interpolation and reanalysis products, the AI method has advantages in capturing complex spatial structures and nonlinear changes. It can restore large-scale temperature anomaly patterns such as those associated with ENSO, and generate physically reasonable, fully covered climate fields even in data-sparse regions.~~

~~(3) High reliability of Antarctic AI reconstructions after 1961. Although the effective data coverage in Antarctica is only around 10% after 1961, the AI models achieve correlation coefficients of approximately 0.9 and RMSE values generally below 1.3°C in this region. The reconstructed sequences show strong consistency with previous studies in both temporal evolution and spatial distribution, indicating that deep learning models can effectively recover ST variations in high-latitude regions with sparse observations when sufficient climate information is present.~~

~~(4) Consistent long-term trends across different AI models in global and Antarctic regions. The 20CR-AI reconstructions show slightly lower global temperature trends than CMIP6-AI, but both reveal that 19th-century temperature anomalies were generally lower, with a slow cooling trend from the mid-19th century to the early 20th century. Since the mid-20th century, particularly during the satellite era, global temperatures have increased significantly, and no pronounced "warming hiatus" has been detected since 1998. In Antarctica, surface temperatures have shown a gradual warming trend since 1961, statistically significant at the 0.05 level, with strong warming particularly evident over the Antarctic Peninsula, Ronne Ice Shelf, Ross Ice Shelf, and surrounding regions.~~

(1) The image reconstruction method based on Partial Convolution (PConv) neural networks can effectively fill in missing global surface air temperature information in the mid-to-low latitudes and capture large-scale spatial temperature anomaly structures such as ENSO. During periods of sparse early observations, both 20CR-AI and CMIP6-AI, which are large-sample-based AI reconstruction models, are able to reconstruct ENSO spatial patterns and reproduce historical ENSO events reasonably well. However, the estimation of ENSO variability shows some biases prior to 1960 and is slightly overestimated after 1979.

(2) The coverage of observational data is a key factor influencing the reliability of AI reconstruction results. During periods of sparse observations over land, different reconstruction schemes exhibit certain systematic discrepancies. As observational coverage significantly improves after 1900, the different reconstructions gradually converge and show better consistency. Overall, AI-based reconstructions can well characterize the long-term variability of global land surface temperature; however, results prior to 1900 in regions with low observational coverage should be interpreted with caution.

(3) The AI reconstruction results for Antarctica since 1961 are highly reliable. Despite an effective observational coverage of only ~8% in Antarctica after 1961, the models achieve correlation coefficients of around 0.9 in this region, with RMSE generally below 1.3 °C. The AI-reconstructed temperature series are consistent with previous studies in both temporal evolution and spatial distribution, indicating that deep learning models can effectively reconstruct surface air temperature variability in high-latitude regions when sufficient climatic information is available.

(4) The Merge-H 20CR-AI and Merge-H CMIP6-AI products show relatively unstable performance in the Arctic prior to 1930 and in Antarctica prior to 1960. Although 20CR-AI performs slightly better than CMIP6-AI on the test set overall, the two models exhibit different strengths and weaknesses in data-sparse polar regions. Moreover, 20CR-AI shows larger variability than CMIP6-AI. Therefore, considering overall robustness, it is recommended to preferentially use the more balanced dataset, Merge-H CMIP6-AI.

Based on the Merge-H AI reconstruction schemes, this study developed spatially complete global monthly ST anomaly datasets for 1850–2024 with a spatial resolution of 5°×2.5°, termed the China global Artificial Intelligence Reconstructed Surface Temperature$_{20CR/CMIP6}$ (C-AIRST$_{R/M}$) datasets, which are reconstructed independently using the 20CR-AI and CMIP6-AI schemes based on the merged C-LSAT2.1 and HadSST4. Overall, this study demonstrates the potential and application value of AI in climate data reconstruction. With further advancements in deep learning, physics-informed learning, and high-performance computing, future AI-based climate reconstruction frameworks are expected to achieve breakthroughs in global continuity and high resolution, offering more robust scientific support for understanding the evolution of the Earth's climate system.

Author contributions

CO: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing (original draft preparation; review and editing). QL: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing (review and editing). ZL: resources, validation. SW: resources, validation.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Earth System Science Data.

Financial support

This research has been supported by the National Natural Science Foundation of China (grant no. 42375022) and the National Key Research and Development Program of China (grant no. 2023YFC3008002).

References

Beckers, J. M. and Rixen, M.: EOF calculations and data filling from incomplete oceanographic datasets, J. Atmos. Oceanic Technol., 20, 1839–1856, https://doi.org/10.1175/1520-0426(2003)020<1839:ECADEF>2.0.CO:2, 2003.

Bochow, N., Poltronieri, A., Rypdal, M., and Boers, N.: Reconstructing historical climate fields with deep learning, Sci. Adv., 11, eadp0558, https://doi.org/10.1126/sciadv.adp0558, 2025.

Bromwich, D., Ensign, A., Wang, S. and Zou X.: Major Artifacts in ERA5 2-m Air Temperature Trends Over Antarctica Prior to and During the Modern Satellite Era. Geophys. Res. Lett., 51(21), https://doi.org/10.1029/2024GL111907, 2024.

Bromwich, D., Wang, S. H., Zou, X., and Ensign, A.: An updated reconstruction of Antarctic near-surface air temperatures at monthly intervals since 1958, Earth Syst. Sci. Data, 17, 2953–2962, https://doi.org/10.5194/essd-17-2953-2025, 2025a.

Bromwich, D., Wang, S. H., and Nicolas, J. P.: Reconstructed Byrd temperature record (1957–2022), Polar Meteorology Group [data set], https://polarmet.osu.edu/datasets/Byrd_recon/, last access: 23 July 2025, 2025b.

Chan, D., Chan, S. C., Siddons, J. T., Cable, A., Faulkner, A., Kent, E. C., Gebbie, G., and Huybers, P.: DCENT- I: A Globally Infilled Extension of the Dynamically Consistent ENsemble of Temperature Dataset. Geosci. Data J., 13:e70054. https://doi.org/10.1002/gdj3.70054, 2026.

Chan, D., and Huybers, P.: Correcting Observational Biases in Sea Surface Temperature Observations Removes Anomalous Warmth During World War II. J. Clim., 34(11): 4585–4602. https://doi.org/10.1175/JCLI-D-20-0907.1, 2021.

Clem, K. R., Fogt, R. L., Turner, J., et al.: Record warming at the South Pole during the past three decades, Nat. Clim. Change, 10, 762–770, https://doi.org/10.1038/s41558-020-0815-z, 2020.

Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Q. J. Roy. Meteor. Soc., 140, 1935–1944, https://doi.org/10.1002/qj.2297, 2014.

Cowtan, K., Hausfather, Z., Hawkins, E., Jacobs, P., Mann, M. E., Miller, S. K., Steinman, B. A., Stolpe, M. B., and Way, R. G.: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, Geophys. Res. Lett., 42, 6526–6534, https://doi.org/10.1002/2015GL064888, 2015.

Darelius, E., Fer, I., and Nicholls, K.: Observed vulnerability of the Filchner–Ronne Ice Shelf to wind-driven inflow of warm deep water, Nat. Commun., 7, 12300, https://doi.org/10.1038/ncomms12300, 2016.

Esteves, C., Slotine, J. J., and Makadia, A.: Scaling spherical CNNs, Proc. 40th Int. Conf. Mach. Learn., 202, 9396–9411, https://doi.org/10.48550/arXiv.2306.05420, 2023.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fang, S. W., Timmreck, C., Jungclaus, J., and Krüger, K.: On the additivity of climate responses to the volcanic and solar forcing in the early 19th century, Earth Syst. Dynam., 13, 1535–1555, https://doi.org/10.5194/esd-13-1535-2022, 2022.

~~Garza-Girón, R. and Tulaczyk, S. M.: Brief communication: Significant biases in ERA5 output for the McMurdo Dry Valleys region, Antarctica, The Cryosphere, 18, 1207–1213, https://doi.org/10.5194/tc-18-1207-2024, 2024.~~

Ham, Y., Kim, J., and Luo, J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, https://doi.org/10.1038/s41586-019-1559-7, 2019.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, 4, https://doi.org/10.1029/2010RG000345, 2010.

Huang, B., Thorne, P. W., Smith, T. M., Liu W., Lawrimore J., Banzon, V. F., Zhang, H. M., Peterson, T. C., and Menne M.: Further exploring and quantifying uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (v4), J. Clim., 29, 3119–3142, https://doi.org/10.1175/JCLI-D-15-0430.1, 2015.

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., Liu, C., Rennie, J. J., Schreck, C. J., Sun, F., Vose, R., Williams, C. N., Yin, X., and Zhang, H. M.: Uncertainty estimates for sea surface temperature and land surface air temperature in NOAA GlobalTemp version 5, J. Clim., 33, 1351–1379, https://doi.org/10.1175/JCLI-D-19-0395.1, 2020.

Huang, B., Yin, X., Boyer, T., Liu, C., Menne, M., Rao, Y. D., Smith, T., Vose, R., and Zhang, H.: Extended Reconstructed Sea Surface Temperature, Version 6 (ERSSTv6). Part I: An artificial neural network approach, J. Clim., 38, 1105–1121, https://doi.org/10.1175/JCLI-D-23-0707.1, 2025a.

Huang, B., Yin, X., Boyer, T., Liu, C., Menne, M., Rao, Y. D., Smith, T., Vose, R., and Zhang, H.: Extended Reconstructed Sea Surface Temperature, Version 6 (ERSSTv6). Part II: Upgrades on quality control and large-scale filter, J. Clim., 38, 1123–1136, https://doi.org/10.1175/JCLI-D-24-0185.1, 2025b.

Huang, J., Zhang, X., Zhang, Q., Lin, Y., Hao, M., Luo, Y., Zhao, Z., Yao, Y., Chen, X., Wang, L., Nie, S., Yin, Y., Xu, Y., and Zhang, J.: Recently amplified Arctic warming has contributed to a continual global warming trend, Nat. Clim. Change, 7, 875–879, https://doi.org/10.1038/s41558-017-0009-5, 2017.

IPCC: Climate Change 2013: The Physical Science Basis, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., https://doi.org/10.1017/CBO9781107415324, 2013.

IPCC: Climate Change 2021: The Physical Science Basis, Cambridge Univ. Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp., https://doi.org/10.1017/9781009157896, 2021.

Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., and Saynisch-Wagner, J.: Towards neural Earth system modelling by integrating artificial intelligence in Earth system science, Nat. Mach. Intell., 3, 667–674, https://doi.org/10.1038/s42256-021-00374-3, 2021.

Jiao, B., Su, Y., Li, Q., Manara, V., and Wild, M.: An integrated and homogenized global surface solar radiation dataset and its reconstruction based on a convolutional neural network approach, Earth Syst. Sci. Data, 15, 4519–4535, https://doi.org/10.5194/essd-15-4519-2023, 2023.

Jones, P. D.: The reliability of global and hemispheric surface temperature records, Adv. Atmos. Sci., 33, 269–282, https://doi.org/10.1007/s00376-015-5194-4, 2016.

~~Jones, P. D., Harpham, C., and Vinther, B. M.: Winter-responding proxy temperature reconstructions and the North Atlantic Oscillation, J. Geophys. Res. Atmos., 119, 6279–7013, https://doi.org/10.1002/2014JD021561, 2014.~~

Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nat. Geosci., 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, 2020.

Karl, T. R., Arguez, A., Huang, B., Lawrimore, J. H., McMahon, J. R., Menne, M. J., Peterson, T. C., Vose R. S., and Zhang H.: Possible artifacts of data biases in the recent global surface warming hiatus, Science, 348, 1469–1472, https://doi.org/10.1126/science.aaa5632, 2015.

Katz, R. W., Craigmile, P. F., Guttorp, P., Haran, M., Sansó, B., and Stein, M. L.: Uncertainty analysis in climate change assessments, Nat. Clim. Chang., 3, 769–771, https://doi.org/10.1038/nclimate1980, 2013.

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., and Killick, R. E.: An ensemble data set of sea surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set, J. Geophys. Res. Atmos., 124, 7719–7763, https://doi.org/10.1029/2018JD029867, 2019.

Kennicutt II, M. C., Bromwich, D., Liggett, D., et al.: Sustained Antarctic research: a 21st century imperative, One Earth, 1, 95–113, https://doi.org/10.1016/j.oneear.2019.08.014, 2019.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015

Kent, E. C., and Kennedy, J. J.: Historical Estimates of Surface Marine Temperatures. Annu. Rev. Mar. Sci., 13: 283–311. https://doi.org/10.1146/annurev-marine-042120-111807, 2021.

~~Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP uncertainty model, J. Geophys. Res. Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.~~

Li, Q., Sheng, B., Huang, J., Chao, L., Song, Z., Chao, L., Sun, W., Yang, Y., Jiao, B., Guo, Z., Liao, L., Li, X., Sun, C., Li, W., Huang, B., Dong, W., and Jones, P.: Different climate response persistence causes warming trend unevenness at continental scales, Nat. Clim. Chang., 12, 343–349, https://doi.org/10.1038/s41558-022-01313-9, 2022.

Li, Q., Sun, W., Huang, B., Dong, W., Wang, X., Zhai, P., and Jones, P.: Consistency of global warming trends strengthened since 1880s, Sci. Bull., 65, 1709–1712, https://doi.org/10.1016/j.scib.2020.06.009, 2020.

Li, Q., Sun, W., Yun, X., Huang, B., Dong, W., Wang, X., Zhai, P., and Jones, P.: An updated evaluation of the global mean land surface air temperature and surface temperature trends based on CLSAT and CMST, Clim. Dynam., 56, 635–650, https://doi.org/10.1007/s00382-020-05502-0, 2021.

Li, Z., et al.: Arctic warming trends and their uncertainties based on surface temperature reconstruction under different sea ice extent scenarios, Adv. Clim. Chang. Res., 14, 335–346, https://doi.org/10.1016/j.accre.2023.06.003, 2022.

Li, Z., Li, Q., Jiao, B., Xu, Q., Wei S., Ru, X., Si, P., Chao, L., Zhang, H., Lin, J., Liao, L., Zhang, H., Huang, B., and Jones, P.: An integrated uncertainty framework for the China-MST 3.0 global surface temperature dataset. J. Geophys. Res. Atmos., accepted.

Liu, G., Reda, F. A., Shih, K. J., et al.: Image inpainting for irregular holes using partial convolutions, Springer International Publishing, https://doi.org/10.48550/arXiv.1804.07723, 2018.

Lorenz, E. N.: Deterministic nonperiodic flow, in: The Theory of Chaotic Attractors, edited by: Hunt, B. R., Li, T. Y., Kennedy, J. A., and Nusse, H. E., Springer, New York, https://doi.org/10.1007/978-0-387-21830-4_2, 1963.

Ma, Z., Huang, J., Zhang, X., Luo, Y., Dou, T., and Ding, M.: Deep learning-based reconstruction of monthly Antarctic surface air temperatures from 1979 to 2023, Sci. Data, 12, 847, https://doi.org/10.1038/s41597-025-05175-6, 2025.

Marshall, G. J.: Trends in the Southern Annular Mode from observations and reanalyses, J. Climate, 16, 4134–4143, https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2, 2003.

Marshall, G. J.: Half-century seasonal relationships between the Southern Annular Mode and Antarctic temperatures, Int. J. Climatol., 27, 373–383, https://doi.org/10.1002/joc.1407, 2006.

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., and Lawrimore, J. H.: The Global Historical Climatology Network Monthly Temperature Dataset, Version 4, J. Climate, 31, 9835–9854, https://doi.org/10.1175/JCLI-D-18-0094.1, 2018.

Météo-France: Dumont d'Urville station from Météo-France [data set], https://meteofrance.com/, last access: 23 July 2025.

Monaghan, A. J., Bromwich, D. H., Chapman, W., and Comiso, J. C.: Recent variability and trends of Antarctic near-surface temperature, J. Geophys. Res. Atmos., 113, D04105, https://doi.org/10.1029/2007JD009094, 2008.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset, J. Geophys. Res. Atmos., 126, e2019JD032361, https://doi.org/10.1029/2019JD032361, 2021.

Nazeri, K., Ng, E., Joseph, T., Qureshi, F. Z., and Ebrahimi, M.: EdgeConnect: Structure guided image inpainting using edge prediction, IEEE/CVF Int. Conf. Comput. Vis. Workshop, https://doi.org/10.1109/ICCVW.2019.00408, 2019.

Nicolas, J. P. and Bromwich, D. H.: New reconstruction of Antarctic near-surface temperatures: multidecadal trends and reliability of global reanalyses, J. Climate, 27, 8070–8093, https://doi.org/10.1175/JCLI-D-13-00733.1, 2014.

Nielsen, E. B., Katurji, M., Zawar-Reza, P., and Meyer, H.: Antarctic daily mesoscale air temperature dataset derived from MODIS land and ice surface temperature, Sci. Data, 10, 833, https://doi.org/10.1038/s41597-023-02720-z, 2023.

NIWA (National Institute for Water and Atmospheric Research): Scott Base station, NIWA [data set], https://cliflo.niwa.co.nz, last access: 23 July 2025.

Ouyang, C., Li, Q., Li, Z., and Wei, S: China global Artificial Intelligence Reconstructed Surface Temperature$_{20CR/CMIP6}$ (C-AIRST$_{R/M}$), figshare [data set], https://doi.org/10.6084/m9.figshare.30663797.v1, 2025.

Plésiat, É., Dunn, R. J. H., Donat, M. G., and Kadow, C.: Artificial intelligence reveals past climate extremes by reconstructing historical records, Nat. Commun., 15, 9191, https://doi.org/10.1038/s41467-024-53464-2, 2024.

Rayner, N. A., Parker, D. E., Horton, E. B., et al.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res. Atmos., 108, D14, https://doi.org/10.1029/2002JD002670, 2003.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzier, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data, 12, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020, 2020.

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C.: A new estimate of the average Earth surface land temperature spanning 1753 to 2011, Geoinfor. Geostat.: An Overview, 1, 1–7, https://doi.org/10.4172/2327-4581.1000101, 2013.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015, Lecture Notes in Computer Science, 9351, Springer Cham, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Rüttgers, M., Lee, S., Jeon, S., and You, D.: Prediction of a typhoon track using a generative adversarial network and satellite images, Sci. Rep., 9, 6057, https://doi.org/10.1038/s41598-019-42339-y, 2019

Scambos, T. A., Campbell, G. G., Pope, A., Haran, T., Muto, A., Lazzara, M., Reijmer, C. H., and van den Broeke, M. R.: Ultralow surface temperatures in East Antarctica from satellite thermal infrared mapping: the coldest places on Earth, Geophys. Res. Lett., 45, 6124–6133, https://doi.org/10.1029/2018GL078133, 2018.

Sheehan, P. M. F., and Heywood, K. J.: Ross Ice Shelf frontal zone subjected to increasing melting by ocean surface waters, Sci. Adv., 10, eado6429, https://doi.org/10.1126/sciadv.ado6429, 2024.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., and et al.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. R. Meteorol Soc., 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

South Pole Meteorology Office: Amundsen-Scott South Pole Station climatology data, 1957–present (ongoing), AMRDC Data Repository [data set], https://doi.org/10.48567/szgp-6h49, last access: 23 July 2025.

Sun, W., Li, Q., Huang, B., Dong, W., Wang, X., Zhai, P., and Phil Jones: The assessment of global surface temperature change from 1850s: the C-LSAT2.0 ensemble and the CMST-Interim datasets, Adv. Atmos. Sci., 38, 875–888, https://doi.org/10.1007/s00376-021-1012-3, 2021.

Sun, W., Yang, Y., Chao, L., Dong, W., Huang, B., Jones, P., and Li, Q.: Description of the China global Merged Surface Temperature version 2.0, Earth Syst. Sci. Data, 14, 1677–1693, https://doi.org/10.5194/essd-14-1677-2022, 2022.

Toms, B. A., Barnes, E. A., and Ebert-Uphoff, I.: Physically Interpretable Neural Networks for the Geosciences: Applications to Earth System Variability, J. Adv. Model. Earth Syst., 12, e2019MS002002, https://doi.org/10.1029/2019MS002002, 2020.

Trenberth, K. E., and Stepaniak, D. P.: Covariability of components of poleward atmospheric energy transports on seasonal and interannual timescales, J. Climate, 16, 3691–3705, https://doi.org/10.1175/1520-0442(2003)016<3691:COCOPA>2.0.CO;2, 2003.

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., and Iagovkina, S.: The SCAR READER Project: Toward a high-quality database of mean Antarctic meteorological observations, J. Climate, 17, 2890–2898, https://doi.org/10.1175/1520-0442(2004)017<2890:TSRPTA>2.0.CO;2, 2004.

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., and Svetlana, L.: Antarctic climate change during the last 50 years, Int. J. Climatol., 25, 279–294, https://doi.org/10.1002/joc.1130, 2005.

Turner, J., Marshall, G. J., Clem, K., et al.: Antarctic temperature variability and change from station data, Int. J. Climatol., 40, 2986–3007, https://doi.org/10.1002/joc.6378, 2019.

Vaughan, D. G., Marshall, G. J., Connolley, W. M., et al.: Recent rapid regional climate warming on the Antarctic Peninsula, Clim. Change, 60, 243–274, https://doi.org/10.1023/A:1026021217991, 2003.

Vose, R. S., Arndt, D., Banzon, V. F., et al.: NOAA's Merged Land Ocean Surface Temperature Analysis, Bull. Amer. Meteor. Soc., 93, 1677–1685, https://doi.org/10.1175/BAMS-D-11-00241.1, 2012.

Vose, R. S., Huang, B., Yin, X., Arndt, D., Easterling, D. R., Lawrimore, J. H., Menne, M. J., Sanchez Lugo, A., and Zhang, H. M.: Implementing full spatial coverage in NOAA’s global temperature analysis, Geophys. Res. Lett., 48, e2020GL090873, https://doi.org/10.1029/2020GL090873, 2021.

Wang, K., and Clow, G. D.: Reconstructed global monthly land air temperature dataset (1880–2017), Geosci. Data J., 7, 4–12, https://doi.org/10.1002/gdj3.84, 2020.

Wang, R., Zheng, F., and Wang, H.: Influence of Winter Tasman Sea SST on the Antarctic Peninsula: A perspective from historical simulations, Adv. Atmos. Sci., 42, 1533–1547, https://doi.org/10.1007/s00376-024-4350-0, 2025.

Wang, S., Li, G., Zhang, Z., Zhang, W., Wang, X., Chen, D., Chen, W., and Ding, M.: Recent warming trends in Antarctica revealed by multiple reanalysis, Adv. Clim. Chang. Res., 16, 447–459, https://doi.org/10.1016/j.accre.2025.03.003, 2025.

Wang, Y., Zhang, X., Ning, W., Lazzara, M. A., Ding, M., Reijmer, C. H., et al.: The AntAWS dataset: A compilation of Antarctic automatic weather station observations, Earth Syst. Sci. Data, 15, 411–429, https://doi.org/10.5194/essd-15-411-2023, 2023.

Wei, S., Li, Q., Xu, Q., Li, Z., Zhang, H., and Lin, J.: Updates to C-LSAT2.1 and the development of high-resolution land surface air temperature and diurnal temperature range datasets, Earth Syst. Sci. Data, 17, 4985–5005, https://doi.org/10.5194/essd-17-4985-2025, 2025.

Xie, S., Wei, S., Li, Z., and Li, Q.: Recent Changes in Antarctic Surface Air Temperature Based on the Fusion of Satellite and In-situ Measurements. Theor. Appl. Climatol., https://doi.org/10.1007/s00704-026-06130-0, in press.

Xu, W., Li, Q., Jones, P. D., Wang, X. L., et al.: A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900, Clim. Dyn., 50, 2513–2536, https://doi.org/10.1007/s00382-017-3755-1, 2018.

~~Yang, Q., Zhao, C., Feng, J., Li, G., Gu, J., Xia, Z., Xu, M., and Yang, Z.: Comprehensive evaluation of iAMAS (v1.0) in simulating Antarctic meteorological fields with observations and reanalysis, Geosci. Model Dev., 18, 5373–5396, https://doi.org/10.5194/gmd-18-5373-2025, 2025.~~

Yin, X. G., Huang, B. Y., Menne, M., et al.: NOAAGlobalTemp Version 6: An AI-based global surface temperature dataset, Bull. Amer. Meteor. Soc., 105, E2184–E2193, https://doi.org/10.1175/BAMS-D-24-0012.1, 2024.

Yu, J. Y. and Kim, S. T.: Identifying the types of major El Niño events since 1870. Int. J. Climatol. 33, 2105–2112. https://doi.org/10.1002/joc.3575, 2013.

Yun, X., Huang, B., Cheng, J., Xu, W., Qiao, S., and Li, Q.: A new merge of global surface temperature datasets since the start of the 20th Century, Earth Syst. Sci. Data, 11, 1629–1643, https://doi.org/10.5194/essd-11-1629-2019, 2019.

Zhou, L., Liu, H., Jiang, X., Ziegler, A., Azorin-Molina, C., Liu, J., and Zeng, Z.: An artificial intelligence reconstruction of global gridded surface winds, Sci. Bull., 67, 2060–2063, https://doi.org/10.1016/j.scib.2022.09.022, 2022.