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. This study develops a historical climate field reconstruction framework based on a Partial Convolutional (PConv) neural network, using climate model simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6) as training samples. Global surface temperature observations were constructed by merging the China Global Land Surface Air Temperature dataset (C-LSAT2.1) with the 200-member sea surface temperature ensemble from HadSST4.2.0.0, provided by the UK Met Office Hadley Centre, thereby forming an input ensemble. Model uncertainty was further quantified using the Deep Ensemble approach, generating multiple reconstruction members. The final monthly global surface temperature reconstruction was obtained by averaging all reconstruction ensemble members, while the corresponding uncertainty was estimated from the ensemble. Furthermore, validation against station observations indicates that the reconstructions perform well over Antarctica after 1961, where observational coverage is extremely sparse. Based on this framework, we developed the China global Artificial Intelligence Reconstructed Surface Temperature$_{\mathrm{CMIP6}}$ (C-AIRST$_{\mathrm{M}}$) dataset, 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. This dataset provides 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$_{\mathrm{M}}$ datasets can be downloaded at https://doi.org/10.6084/m9.figshare.33122444 (Ouyang et al., 2026a). It is also available from http://www.gwpu.net/en/h-col-103.html (last access: 30 July 2026).

1 Introduction

Global surface temperature (ST) is one of the most fundamental variables in the climate system, directly reflecting the state of the 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). 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 spatial coverage as complete as possible (Morice et al., 2021; Vose et al., 2021; Lenssen et al., 2019; Rohde, 2020; Sun et al., 2021). 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. 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.

Antarctica holds an irreplaceable position in the global climate system, and its enormous glacier masses store a substantial portion of global freshwater and play a major role in determining future sea level change (IPCC, 2013). Consequently, accurately characterizing the spatiotemporal variations of ST over Antarctica is essential for assessing Antarctic climate change. 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 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. 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. 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.

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. Qian et al. (2026) improved the accuracy of global historical climate reconstruction by integrating a Video Diffusion Model with a U-Net framework. Li et al. (2026a) applied a generative diffusion model to perform probabilistic reconstruction of SST fields, providing a new approach for quantifying uncertainty in historical climate data.

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 upon previous climate field reconstruction studies based on PConv, this study develops an AI-driven reconstruction framework that integrates Deep Ensemble with observational uncertainty propagation for uncertainty quantification of reconstructed global historical surface temperature fields. Furthermore, by incorporating more comprehensive observational information, the proposed framework improves the accuracy of temperature reconstruction over Antarctica after 1961.

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, uncertainty quantification and the data validation methods. Section 3 presents the global reconstruction results, including the ENSO events, 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 global 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., 2026b).

In this study, C-LSAT2.1 was used as the land surface air temperature (LSAT) component, while the 200-member HadSST4.2.0.0 sea surface temperature (SST) ensemble provided by the UK Met Office Hadley Centre (Kennedy et al., 2019) was used as the SST component. These datasets have a monthly temporal resolution and span the period from 1850 to 2024. A total of 105 ensemble members from Phase 6 of the Coupled Model Intercomparison Project (CMIP6; Eyring et al., 2016) historical simulations (Table S1) were used as training samples for the PConv model.

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). Specifically, the Byrd station data were obtained from the OSU Polar Meteorology Group (Bromwich et al., 2025b), the Dumont d'Urville station data from Météo-France (Météo-France, 2025), the Amundsen-Scott station data from the University of Wisconsin-

Madison (South Pole Meteorology Office, 2025), and the Scott Base station data from the National Institute of Water and Atmospheric Research Ltd (NIWA, 2025). Necessary gap filling was then performed for the Antarctic station records following the method described by Bromwich et al. (2025a). Stations with more than 25 valid years during 1961–1990 are selected, yielding a temporal span of 1957–2024. 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 reconstruction.

Data	Type	Resolution	Time	Reference
CMIP6	Grid	Monthly/-	1850–2014	Eyring et al. (2016)
C-LSAT2.1	Grid	Monthly/5°×5°	1850–2024	Wei et al. (2025)
HadSST4.2.0.0	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. The Antarctic land mask used in this study at a spatial resolution of 5° × 2.5° is shown in Fig. S2.

2.3 AI training and reconstruction process

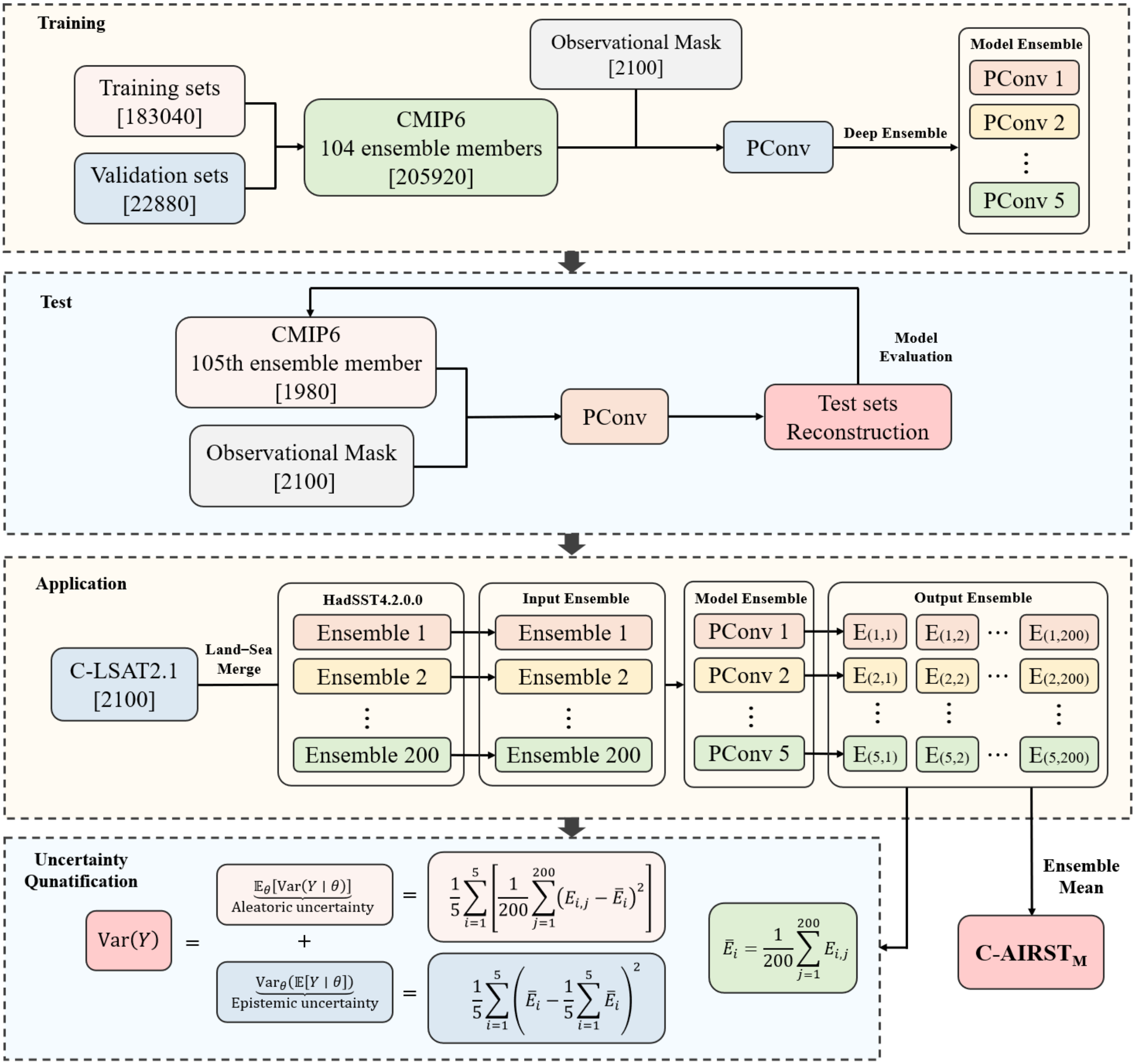


Figure 1: AI training and reconstruction process. The number in square brackets (e.g., [183040]) denotes the total number of the monthly temperature anomaly fields. In the output ensemble, E(i, j) denotes the (j)-th ensemble member generated by the (i)-th PConv model.

The PConv method has been successfully applied to the reconstruction of 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). 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. (2020) 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 SST 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.

The main contribution of this study is the integration of input observational uncertainty, represented by the HadSST4 ensemble members, and model uncertainty, represented by the Deep Ensemble approach, into a unified reconstruction framework. The law of total variance is then used to decompose and quantify these two sources of uncertainty, thereby providing predictive uncertainty estimates for reconstructed result. In addition to producing reconstructed global historical surface temperature fields, the proposed framework also provides associated uncertainty estimates, offering valuable confidence information for subsequent climate change analyses and the interpretation of reconstruction results.

The reconstruction workflow is illustrated in Fig. 1, and consists of the following steps:

(1) Resolution standardization: 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: The monthly temperature anomaly fields (relative to the 1961–1990 climatology) from 105 CMIP6 historical simulation ensemble members over the period 1850–2014 were used to construct the sample dataset. One ensemble member was randomly selected and reserved as the test set, while the remaining 104 CMIP6 ensemble members were used as the training set for the PConv model. The training samples were randomly partitioned, with 8/9 used for model training and 1/9 for validation. This resulted in 183,040 training samples and 22,880 validation samples for the AI model.

(3) Mask configuration: Observational data obtained by merging C-LSAT2.1 and HadSST4.2.0.0, with monthly data spanning 1850–2024, were used as masks for missing values for the reconstruction. Observed grid points were marked as “1” and missing grid points as “0”, serving as inputs for model training and validation.

(4) Model training and fine-tuning: During the PConv model training stage, five independent PConv models were trained using different random seeds (Seed = 1–5) for network initialization. Each model was 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 each training run required approximately 8 hours of computation on an NVIDIA GeForce RTX 4060 GPU.

(5) Model evaluation: The CMIP6 test ensemble member was input into the single trained PConv model, and the reconstruction was compared with the original test ensemble member to evaluate model performance (see Text S1).

(6) Reconstruction: C-LSAT2.1 was merged with each of the 200 ensemble members from HadSST4.2.0.0 to generate 200 input ensemble members. Each input ensemble member was then input into the five trained PConv models, producing a total of 1,000 globally complete monthly temperature reconstructions at a spatial resolution of 5° × 2.5° for the period 1850–2024. The ensemble mean of all reconstructions was taken as the final reconstructed dataset, referred to as China global Artificial Intelligence Reconstructed Surface Temperature$_{\mathrm{CMIP6}}$ (C-AIRST$_{\mathrm{M}}$).

2.4 Uncertainty Quantification

In deep learning, predictive uncertainty can generally be decomposed into aleatoric uncertainty and epistemic uncertainty (Kendall and Gal, 2017). Aleatoric uncertainty arises from the inherent randomness of the observation process and measurement noise, whereas epistemic uncertainty reflects uncertainty in model parameters due to limited knowledge of the underlying data-generating process (Abdar et al., 2021).

In this study, the 200 ensemble members provided by HadSST4.2.0.0 were used to approximate the observational uncertainty of sea surface temperature (SST). The HadSST4 ensemble members are generated using an observational error model that accounts for multiple sources of uncertainty, including Bucket measurement errors, ~~sampling~~ Engine-room intake measurement errors, and bias-adjustment uncertainties. Each ensemble member represents a physically plausible realization of the SST field constrained by the available observations (Kennedy et al., 2019). The C-LSAT2.1 dataset was merged separately with each HadSST4 ensemble member to construct 200 model input ensembles. Because these ensemble members collectively represent the irreducible uncertainty associated with the observation process, they were regarded as the aleatoric uncertainty of the input data, which was propagated through the reconstruction model via ensemble propagation (Kendall and Gal, 2017).

Epistemic uncertainty, also referred to as model uncertainty, was estimated using the Deep Ensemble approach, which is one of the most effective methods for quantifying epistemic uncertainty in deep learning (Lakshminarayanan et al., 2017). Specifically, five PConv models with identical network architectures, training datasets, and hyperparameters were independently trained using different random seeds for parameter initialization. The model with the best validation performance from each training run was retained to form the ensemble. Owing to different random initializations and stochastic optimization trajectories, the independently trained models converge to different parameter configurations, thereby reflecting uncertainty in model parameter estimation (Lakshminarayanan et al., 2017).

The 200 input ensembles were subsequently fed into each of the five trained PConv models, resulting in a total of 1,000 reconstructed temperature fields. This reconstruction ensemble simultaneously incorporates the propagated observational uncertainty from the input data and the prediction variability arising from different model parameterizations, thereby jointly representing both aleatoric and epistemic uncertainty. The ensemble mean of the 1,000 reconstructions was adopted as the final reconstructed temperature product. The total predictive uncertainty was decomposed according to the Law of Total Variance (Valdenegro-Toro and Mori, 2022):

$$\mathrm{Var}(Y) = \underbrace{\mathbb{E}_\theta[\mathrm{Var}(Y \mid \theta)]}_{\text{Aleatoric uncertainty}} + \underbrace{\mathrm{Var}_\theta(\mathbb{E}[Y \mid \theta])}_{\text{Epistemic uncertainty}}$$

Here, θ denotes the parameters of each independently trained PConv model in the Deep Ensemble, and Y represents the reconstructed temperature field. The first term, $\mathbb{E}_\theta[\mathrm{Var}(Y \mid \theta)]$, represents the propagated observational (aleatoric) uncertainty. For each fixed model θ, the variance of the predicted outputs was first calculated from the 200 ensemble predictions, yielding $\mathrm{Var}(Y \mid \theta)$. The average of these variances across the five models was then taken to obtain $\mathbb{E}_\theta[\mathrm{Var}(Y \mid \theta)]$. The second term, $\mathrm{Var}_\theta(\mathbb{E}[Y \mid \theta])$, represents the model (epistemic) uncertainty. Specifically, for each model θ, the ensemble mean prediction $\mathbb{E}[Y \mid \theta]$ was first computed from the 200 reconstructed outputs. The variance among the mean predictions of the five independently trained models was then calculated to quantify the uncertainty arising from differences in model parameters.

The sum of these two components yields the total predictive variance. Its square root corresponds to the predictive standard deviation σ, and the 95% predictive uncertainty interval is reported as 1.96σ.

It should be noted that the total predictive uncertainty estimated in this study represents an approximation of the true predictive uncertainty and is conditional on the adopted ensemble representation. Although the observational uncertainty of the SST input data is propagated through the reconstruction model, the 200-member HadSST4 ensemble primarily represents the uncertainty associated with SST bias corrections and does not explicitly incorporate the spatially correlated error covariance information. Therefore, these spatially correlated covariance errors are not explicitly propagated in our uncertainty quantification framework. The observational errors from C-LSAT2.1 are not explicitly incorporated into the uncertainty quantification framework. According to the uncertainty assessment of the C-LSAT2.1 dataset by Li et al. (2026b), the maximum 1σ station, sampling, and bias uncertainties are approximately 0.065 °C, 0.025 °C, and 0.005 °C, respectively, in 1850, and all decrease to below approximately 0.02 °C after the early twentieth century. Their omission is expected to have only a minor influence on the total predictive uncertainty estimated in this study. Consequently, the estimated uncertainty mainly reflects the contributions from SST input uncertainty and model uncertainty. The resulting total predictive uncertainty can therefore be regarded as a lower-bound estimate of the true predictive uncertainty, and the actual uncertainty over land regions is likely to be larger than that reported in this study.

2.5 Data validation methods

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 (Text S1). The primary evaluation metrics included the Bias, Correlation Coefficient (CC), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which were used to assess the consistency and bias of the reconstructed fields relative to reference datasets (Figs. S5–S10). Based on the model validation results (Text S1, Figs. S5–S10), this study ultimately adopted the merged product of C-LSAT2.1 and HadSST4.2.0.0 as the observational dataset, trained the AI model using CMIP6 samples, and estimated reconstruction uncertainty by incorporating the 200-member HadSST4.2.0.0 ensemble.

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. 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, DCENT-I Kadow et al.(2020), and the reconstruction by Bromwich et al. (2025a). 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 ENSO events

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 Fig. 2, the ocean coverage of NOAAGlobalTempv6 is higher than that of the unreconstructed C-$AIRST_M$, which is attributed to the ocean component included in the former (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).

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 b1). 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.

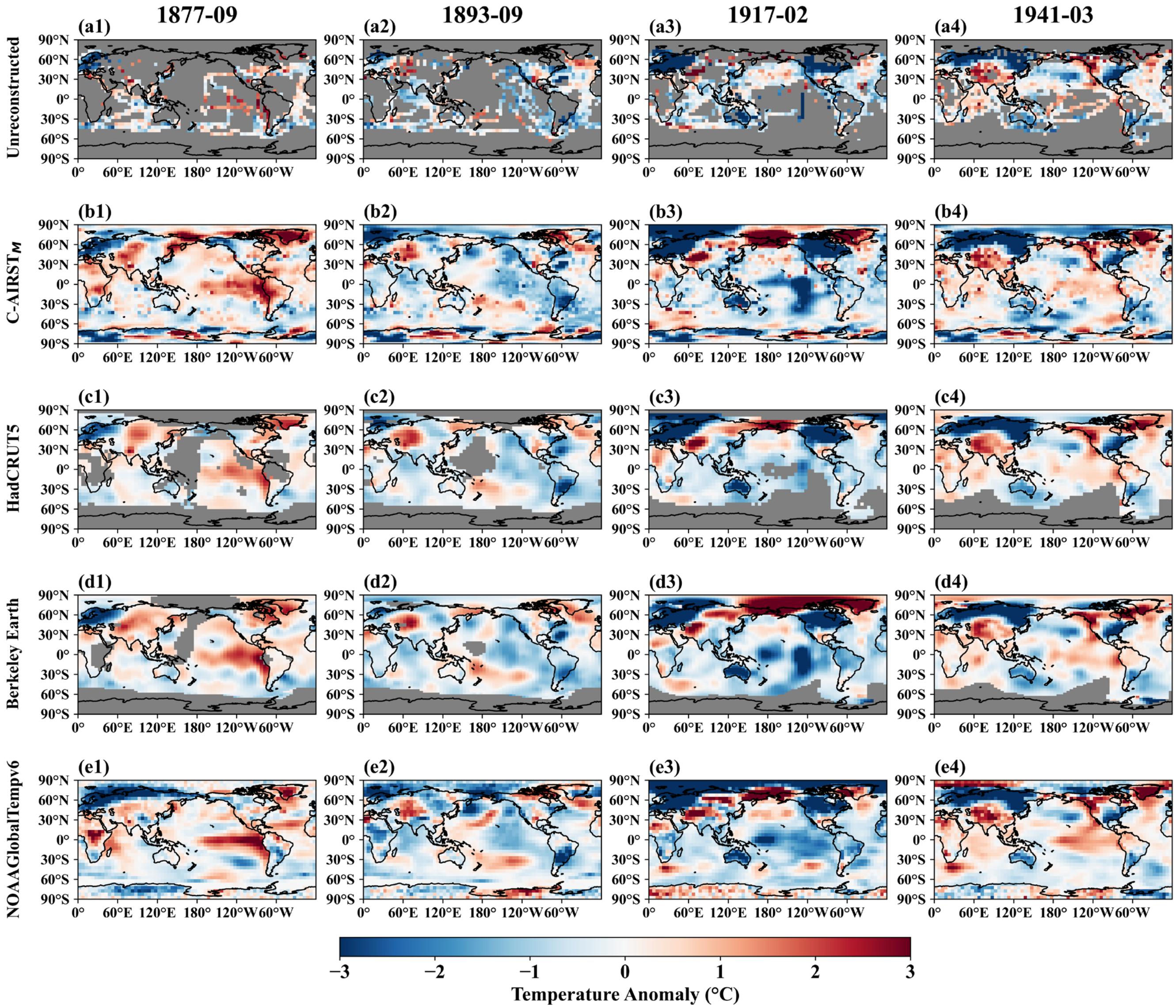


Figure 2: Global temperature anomaly (relative to the 1961–1990 climatology) fields before and after reconstruction for four typical ENSO events. (a1–a4) Unreconstructed C-AIRST$_M$; (b1–b4) C-AIRST$_M$; (c1–c4) HadCRUT5; (d1–d4) Berkeley Earth; (e1–e4) NOAAGlobalTempv6. From left to right, the columns show the temperature anomaly fields for September 1877, September 1893, February 1917, and March 1941, respectively.

For the ENSO reconstructions in September 1893, February 1917, and March 1941 (Fig. 2 b2–b4), the spatial patterns produced by the AI model are generally consistent with those from HadCRUT5 (Fig. 2 c2–c4) and Berkeley Earth (Fig. 2 d2–d4), 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 e2–e4). This discrepancy primarily arises from inherent differences between HadSST and ERSSTv6.

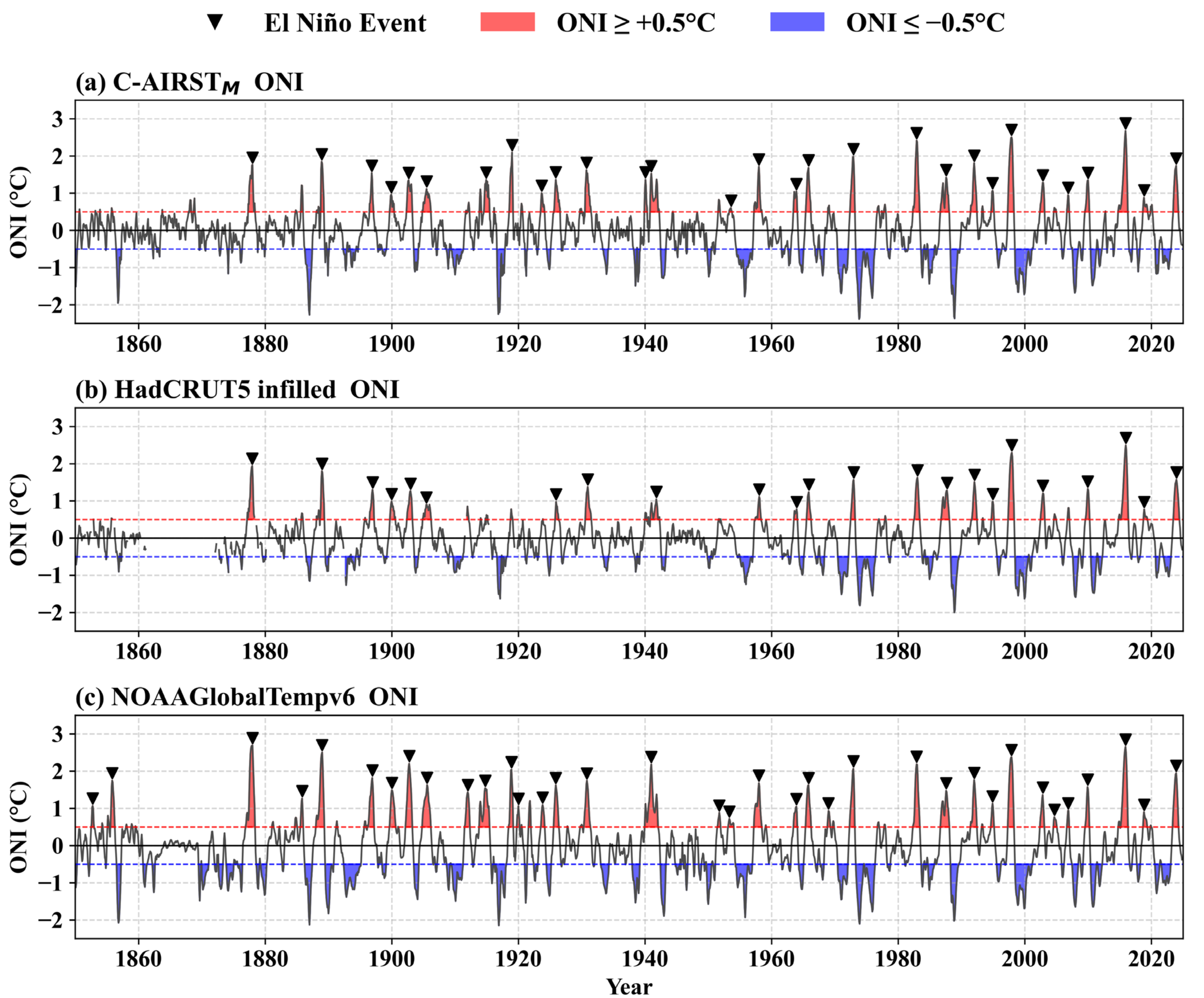


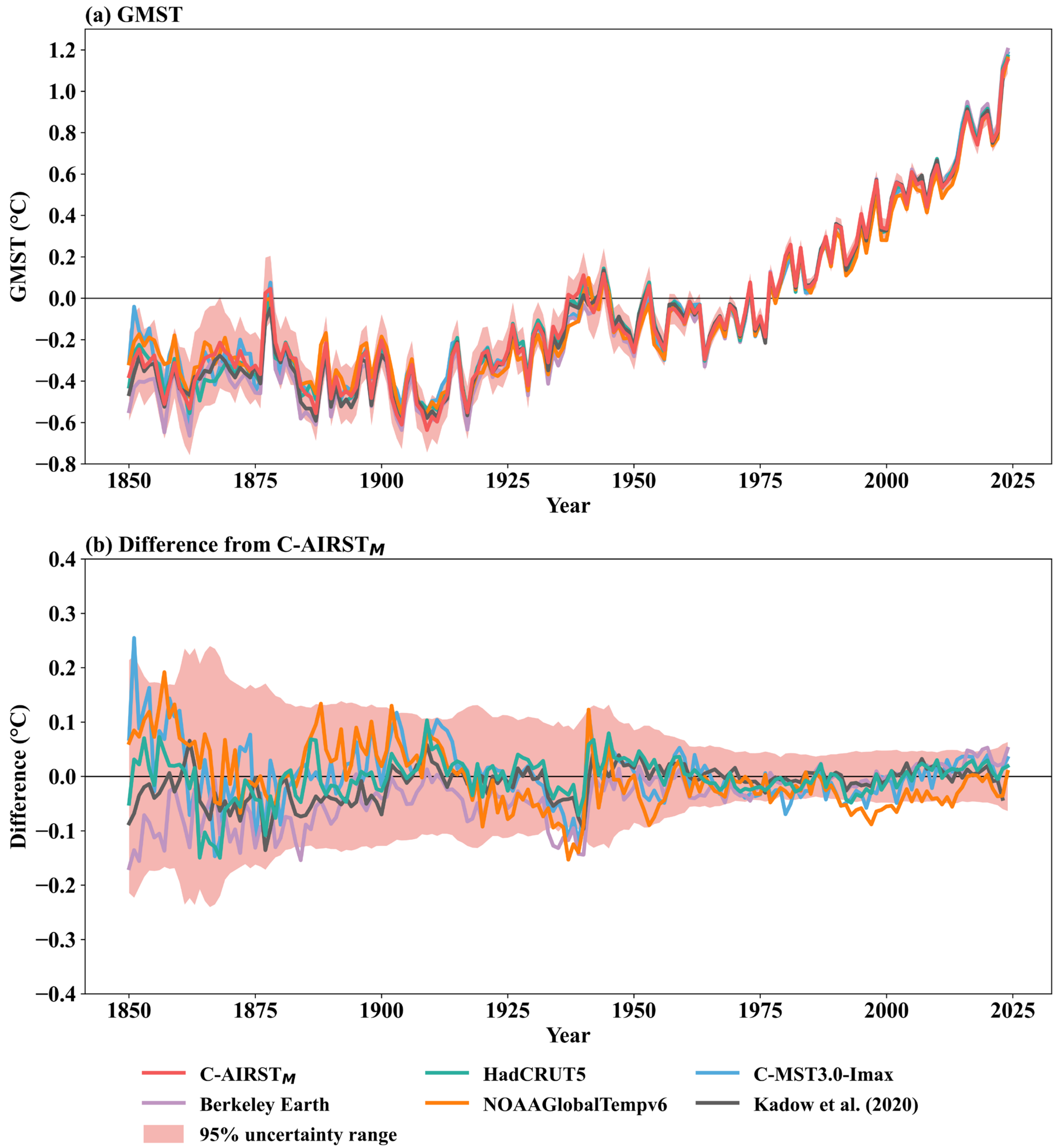
Figure 3: Ocean Niño Index (ONI) time series indicating ENSO events. (a) Time series of ONI from C-AIRST$_M$ over 1850–2024, with black triangles indicating El Niño events defined as ONI exceeding 0.5 °C for at least five consecutive months; (b–c) same as (a) but for HadCRUT5 infilled, and NOAAGlobalTempv6,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

AI model is able to effectively capture the spatial patterns of ENSO, while also reproducing historical ENSO events. As shown in Fig. 3, 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 1914–1915, and 1918–1919 (Yu et al., 2013), which are also evident in NOAAGlobalTempv6 (Fig. 3c). These results also corroborate the findings of Qian et al. (2026), who investigated historical ENSO variability and reconstruction using long-term Niño indices and their associated spatial patterns.

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 C-AIRST$_M$ 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)

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. 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).



325 **Figure 4: Global mean surface temperature (GMST) time series from 1850 to 2024 (relative to the 1961–1990 climatology). The red shading indicates the 95% uncertainty range estimated for C-AIRST$_M$. (a) GMST; (b) GMST differences from C-AIRST$_M$.**

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
C-AIRST$_M$	0.064±0.006	0.101±0.008	0.155±0.013	0.194±0.023	1.49
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

The C-AIRST$_M$ 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 0.15 °C. After 1960, differences between the AI reconstructions and other datasets further decrease to within 0.05 °C. During 1930–1940, C-AIRST$_M$ was approximately 0.1 °C warmer than the other datasets, whereas it was slightly cooler during the 1940–1945 World War II period (Fig. 4b). These differences mainly arise from the revised historical SST bias corrections implemented in HadSST4.2.0.0 (Sandford and Rayner, 2026). In particular, the systematic warm bias introduced by the widespread use of Engine Room Intake (ERI) measurements during World War II was corrected (Chan and Huybers, 2021; Kent and Kennedy, 2021). In addition, a truncation error in the Kobe SST records resulted in an approximately 0.4 °C cold bias during 1930–1940, which was corrected in HadSST4.2.0.0 (Sandford and Rayner, 2026). This SST correction also contributed to a higher estimate of the GMST in C-AIRST$_M$ during the same period. The long-term warming trends reconstructed by the AI model for different periods during 1850–2024 (Table 1) are generally consistent with the estimates from other global temperature datasets, including Berkeley Earth, HadCRUT5, and NOAAGlobalTempv6. The AI reconstruction also yields a Global Warming Level (GWL) of 1.49 °C, with only small differences relative to the other datasets, indicating that global warming in 2024 had approached the 1.5 °C 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.

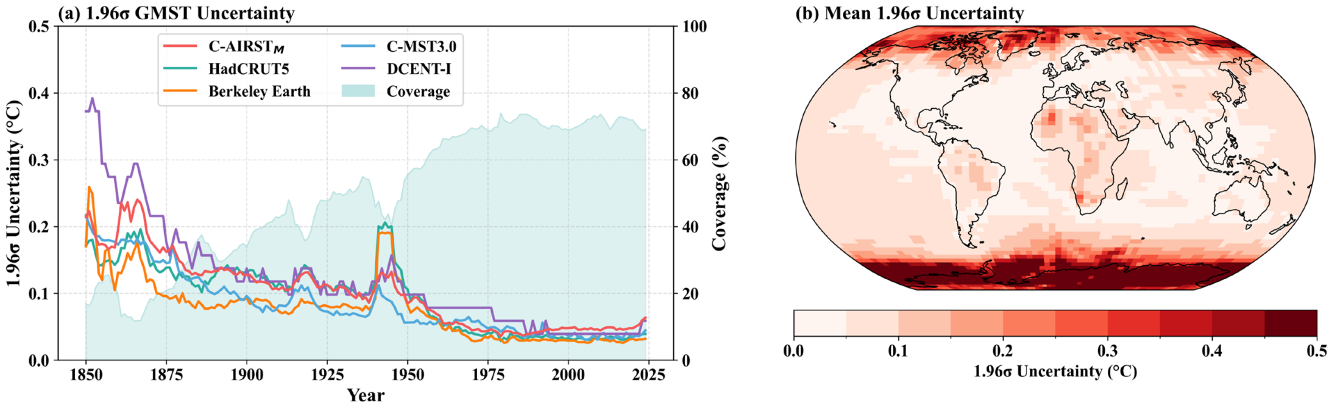


Figure 5: Uncertainty estimation. (a) Time series of the 1.96σ uncertainty in GMST from different datasets; (b) spatial distribution of the mean 1.96σ uncertainty estimated for C-AIRST$_M$.

Regarding the uncertainty estimation of GMST (Fig. 5), the uncertainty quantified in this study is not completely comparable to that of mainstream observational datasets, which generally consider multiple sources of uncertainty, including measurement errors, sampling errors, bias adjustments, and incomplete spatial coverage. The predictive uncertainty estimated here incorporates the propagation of input uncertainty represented by the 200 ensemble members of HadSST4.2.0.0, as well as the model uncertainty arising from the deep learning framework itself (see Section 2.4 for details on uncertainty quantification). Despite the differences in uncertainty sources, the estimated uncertainty in GMST is generally comparable to that of existing mainstream datasets. The 1.96σ uncertainty gradually decreases from approximately 0.2 °C in the 1850s to approximately 0.05 °C after the 1960s (Fig. 5a). In addition, uncertainty exhibits temporary increases during the two World War periods (the 1910s and 1940s), likely associated with reduced global observational coverage.

The spatial distribution of uncertainty (Fig. 5b) shows that the polar regions have substantially higher uncertainty than the low- and mid-latitude regions due to the long-term scarcity of observational records. The uncertainty in the polar regions generally exceeds 0.3 °C, whereas that in most low- and mid-latitude regions remains below 0.2 °C. Notably, Antarctica exhibits substantially larger uncertainty than the Arctic, both in terms of the magnitude of uncertainty and the spatial extent of high-uncertainty regions. This difference is primarily related to the much sparser observational network and insufficient historical observations in Antarctica compared with the Arctic.

3.3 Comparison of zonal temperature

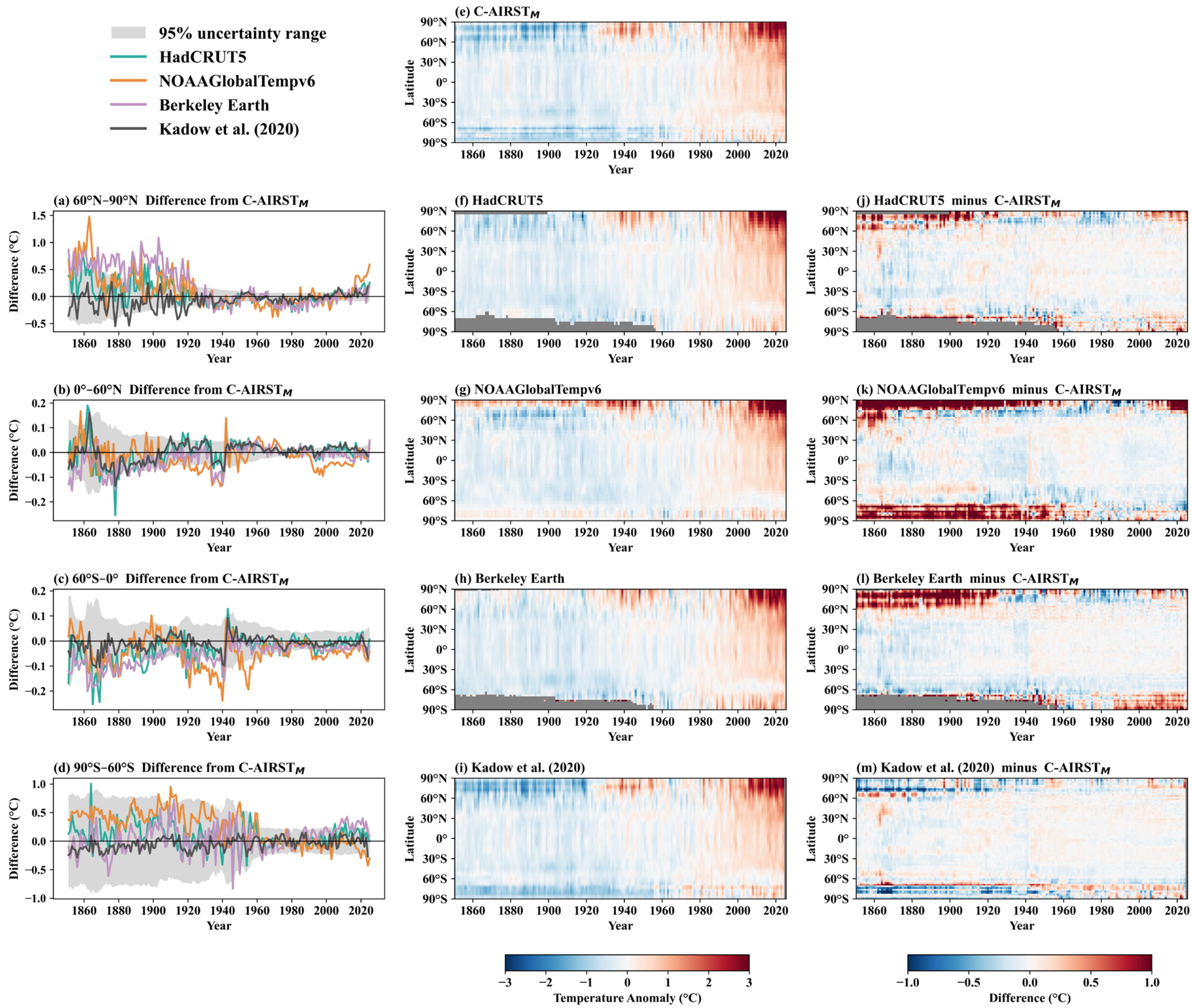


Figure 6 Zonal temperature comparison. The gray shading indicates the 95% uncertainty range estimated for C-AIRST$_M$. (a–d) time series of zonal-mean temperature anomaly differences between other datasets and C-AIRST$_M$ across different latitude bands; (e–i) Zonal-mean temperature anomalies from C-AIRST$_M$, HadCRUT5, NOAAGlobalTempv6, Berkeley Earth, and Kadow et al. (2020), respectively; (j–m) zonal-mean temperature anomaly differences between other datasets and C-AIRST$_M$.

In the zonal temperature comparison across different datasets (Fig. 6), it can be seen that within the 60°S–60°N band, where observational sampling is relatively sufficient, the differences between each dataset and C-AIRST$_M$ generally remain within ±0.1 °C to ±0.2 °C prior to the early 20th century (Figs. 6b, 6c). The Northern Hemisphere (NH), characterized by a larger land fraction and denser observational networks, is substantially better constrained by observations than the predominantly ocean-covered Southern Hemisphere (SH). Consequently, after 1900, the inter-dataset differences within the equator–60°N region 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 1960s. The 95% uncertainty intervals estimated by C-AIRST$_{M}$ in both regions remain within ± 0.1 °C, indicating relatively low uncertainty. 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. 6a), prior to 1920, both C-AIRST$_{M}$ and Kadow et al. (2020) exhibit a cold bias of approximately 0.6 °C relative to HadCRUT5, NOAAGlobalTempv6 and Berkeley Earth. It is also observed that the AI reconstructions (Figs. 6j–6l) 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. 6d), during 1850–1960, the estimates from different datasets are not entirely consistent. The Kadow et al. (2020) product and C-AIRST$_{M}$ 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.

The above analyses suggest that the reconstructions exhibit some systematic cold bias in the early Arctic (before 1920) and Antarctica (before 1960), where observational constraints are relatively weak. Correspondingly, the 95% uncertainty ranges in these periods are also relatively large, reaching approximately ±0.5 °C in the Arctic (Fig. 6a) and ±0.8 °C in Antarctica (Fig. 6d). Consequently, the reconstructed temperatures for these periods and regions should be interpreted as exploratory and treated with appropriate caution.

3.4 Global warming pattern

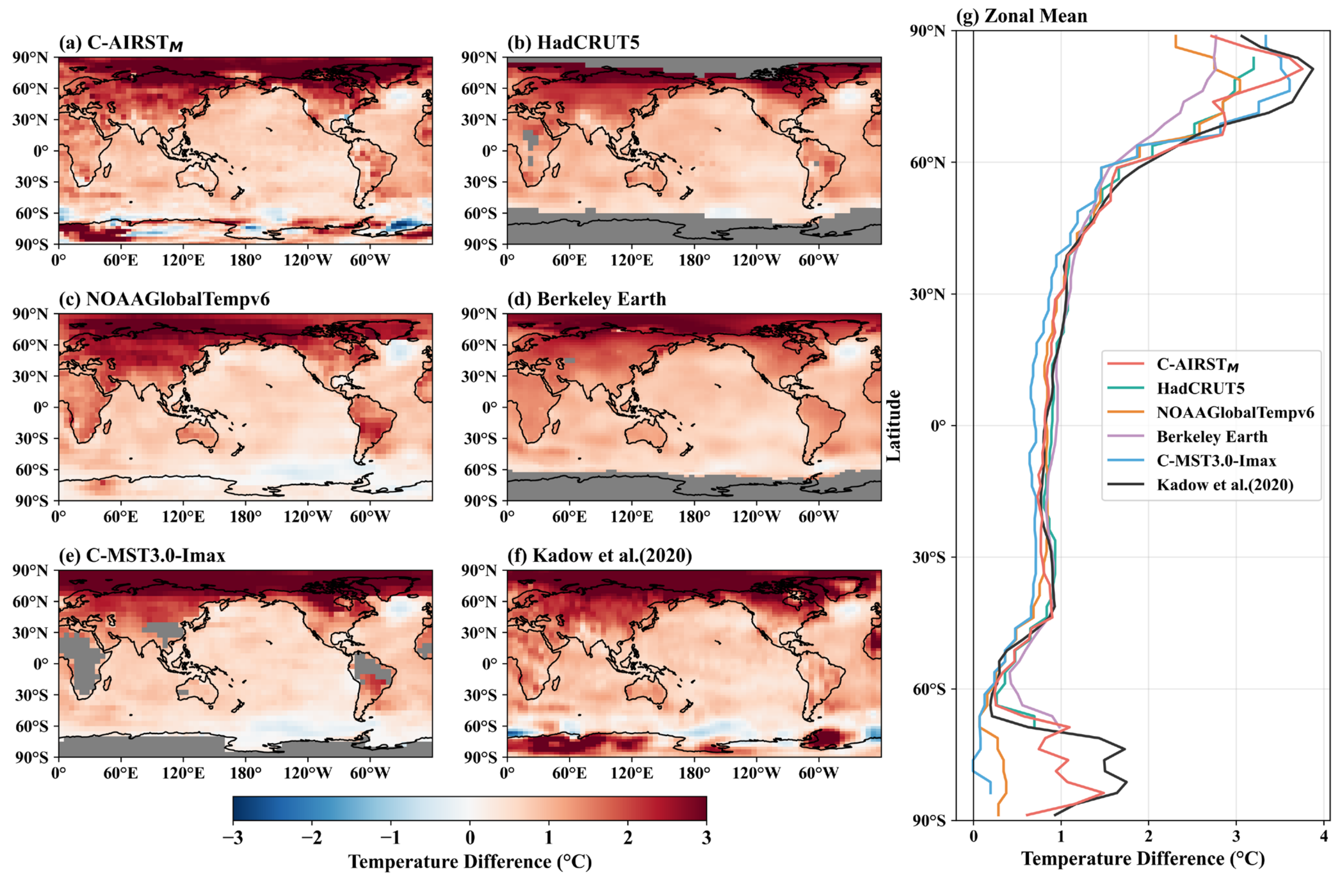


Figure 7 Global warming pattern. (a–f) Spatial distribution of mean global temperature warming over 2005–2020 relative to the 1850–1900 baseline period, (g) zonal-mean profile of warming magnitude.

Under the context of global warming, the spatial pattern of warming level from C-AIRST$_M$ is generally consistent with those of other datasets (Fig. 7). The main discrepancies are concentrated south of 60°S, where the PConv-based reconstruction (Figs. 7a, 7f) 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 (Figs. 2 b1, b2).

The C-AIRST$_M$ and reconstruction from Kadow et al. (2020) 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.

The latitudinal warming amplitudes further reveal the consistency and differences among the reconstructed products in representing polar warming (Fig. 7g). C-AIRST$_M$ and the reconstruction from Kadow et al. (2020) exhibit slightly stronger warming amplitudes near 80°N than the other datasets. In Antarctica, the warming amplitude estimated by C-AIRST$_M$ is higher than that of NOAAGlobalTempv6 by approximately 0.7 °C in the polar regions, respectively. These differences mainly arise from variations in the estimated baseline climatology for 1850–1900 among different datasets.

Furthermore, the zonal distribution of warming (Fig. 7g) shows a pronounced Arctic amplification effect, with warming in the Arctic substantially exceeding the global mean level. During 2005–2020, relative to the pre-industrial baseline period of 1850–1900, the cumulative warming in the Arctic region reached approximately 2.6–4.0 °C, markedly higher than the corresponding global mean warming over the same period.

Fig. 8 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 C-AIRST$_M$ is generally consistent with the reconstruction by Kadow et al. (2020). The spatial gradient of LSAT trends reconstructed by these two datasets is less uniform compared to other datasets, particularly across the continents during 1911–1970. 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.

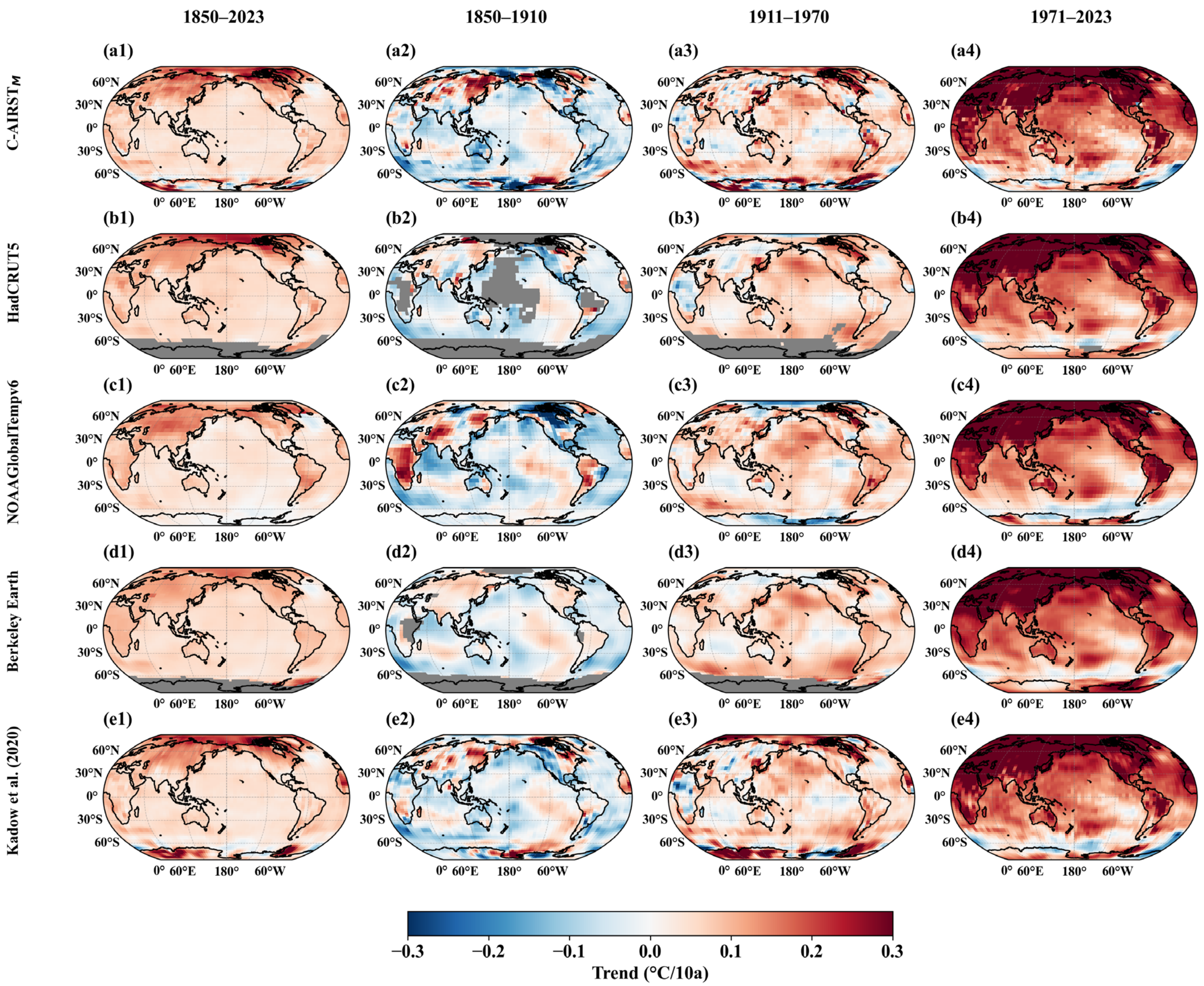


Figure 8 Spatial distribution of global surface temperature trends for 1850–2023, 1850–1910, 1911–1970, and 1971–2023. (a1–a4) C-AIRST$_M$, (b1–b4) HadCRUT5, (c1–c4) NOAAGlobalTempv6, (d1–d4) Berkeley Earth, (e1–e4) reconstruction from Kadow et al. (2020).

3.5 Comparison of regional land surface temperature

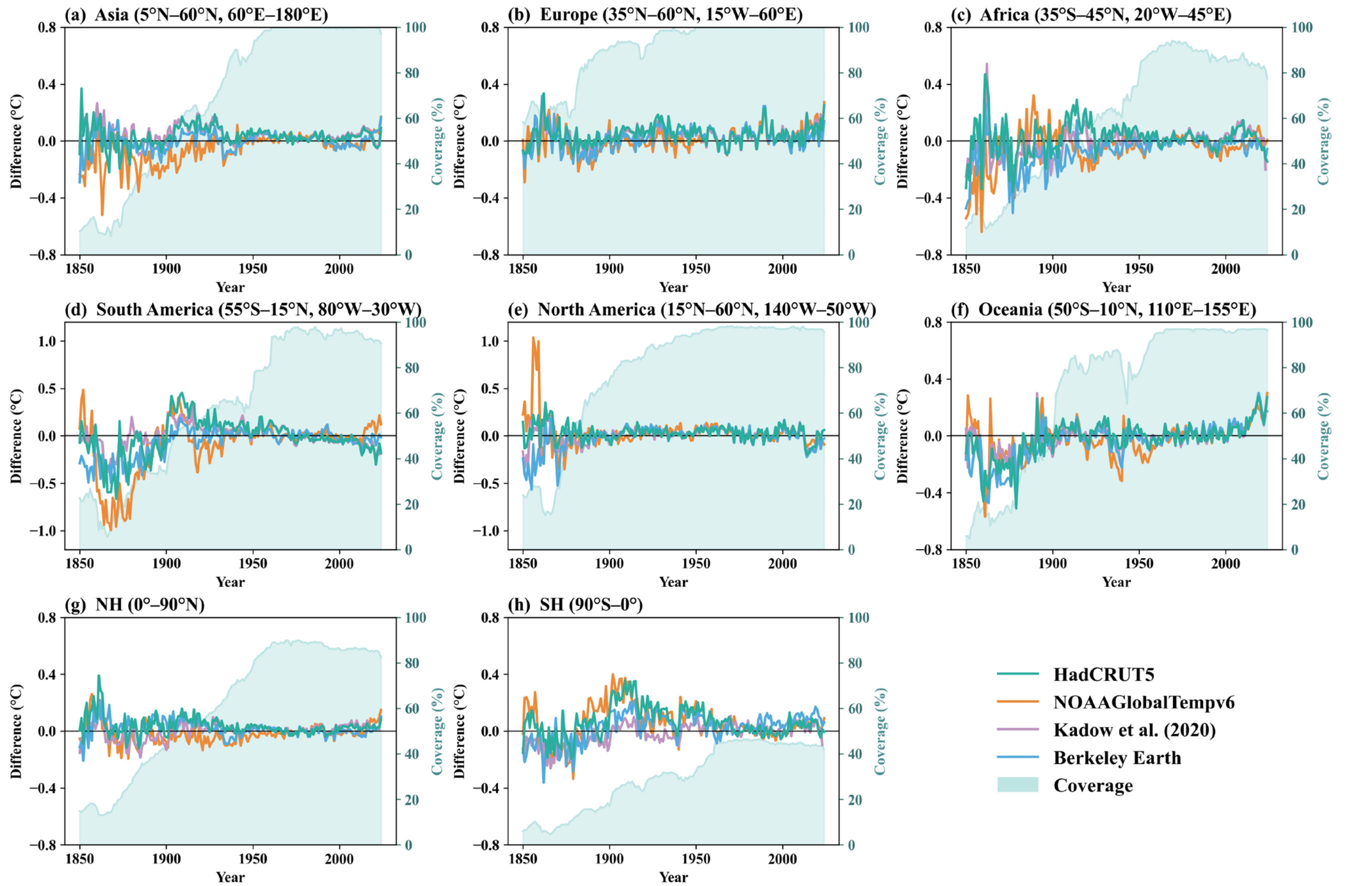


Figure 9: Annual mean surface temperature anomaly differences relative to C-AIRST$_M$ from 1850 to 2024 in different land regions. (a–h) Asia, Europe, Africa, South America, North America, Oceania, the Northern Hemisphere (NH), and the Southern Hemisphere (SH). Coverage indicates the unreconstructed data availability in each region.

One of the key focuses of this study is the reconstruction of global land surface data (Fig. 9, Table 3). To evaluate the reconstruction performance, we also present the effective data coverage of the original datasets across different regions (Coverage in Fig. 9). 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."

Fig. 9 shows that across different land regions, the differences between the C-AIRST$_M$ and other datasets gradually decrease to within 0.1 °C as observational coverage exceeds approximately 50%. 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, C-AIRST$_M$ exhibits differences of varying magnitudes relative to other datasets, ranging from 0.4 to 1.0 °C (Figs. 9c, 9d, 9h), which also leads to lower long-term trends in these regions for the former in Table 3.

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

Region/Dataset	C-AIRST$_M$	HadCRUT5	NOAA GlobalTempv6	Berkeley Earth	Kadow et al. (2020)
Asia	0.088±0.010	0.088±0.009	0.101±0.010	0.090±0.010	0.084±0.010
Europe	0.090±0.014	0.092±0.014	0.097±0.010	0.094±0.013	0.092±0.014
Africa	0.072±0.009	0.073±0.008	0.082±0.008	0.082±0.008	0.079±0.008
South America	0.061±0.010	0.067±0.007	0.090±0.009	0.079±0.007	0.058±0.008
North America	0.079±0.012	0.077±0.012	0.073±0.013	0.087±0.011	0.079±0.012
Oceania	0.039±0.008	0.054±0.008	0.050±0.009	0.058±0.008	0.050±0.008
NH	0.095±0.009	0.091±0.009	0.096±0.010	0.091±0.009	0.097±0.009
SH	0.060±0.007	0.059±0.006	0.059±0.006	0.070±0.007	0.068±0.006

Furthermore, during 1860–1870, when effective coverage in Africa, South America, and Oceania (Figs. 9c, 9d, 9f) drops below a critical threshold (~20%), and considering the strong spatial sampling heterogeneity and lack of inland temperature information in these continents (Wei et al., 2025), the C-AIRST$_M$ shows varying degrees of deviation from NOAAGlobalTempv6 and HadCRUT5. The most pronounced discrepancy occurs in South America (Fig. 9d), 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 trends of C-AIRST$_M$ across different regions are broadly consistent with those of HadCRUT5. However, C-AIRST$_M$ exhibits a weaker warming trend than other datasets over Oceania, which can be attributed to inherent temperature biases in early land surface products prior to 1900.

As shown in Table 3, C-AIRST$_M$ indicates the strongest warming over Europe during 1850–2024, with trend of 0.090 ± 0.014 °C/decade, while Oceania shows the weakest warming, at 0.039 ± 0.008 °C/decade. The warming rate in the SH is significantly lower than that in the NH, 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 the early 20th century, the AI-based reconstruction results show improved consistency and stability across these regions.

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	C-AIRST$_M$	
					r	RMSE
Dome A	−80.4	77.4	4084	2006–2019	0.63	1.18
Drescher	−79.2	−19.0	27	1993–2003	0.56	0.77
Ferrel	−77.9	170.8	45	2007–2023	0.88	0.56
G3	−70.9	69.9	84	2002–2020	0.40	0.84
GC41	−71.6	111.3	2763	1984–2005	0.96	0.44
General Belgrano	−78.0	−38.8	32	1956–1978	0.69	0.77
GF08	−68.5	102.1	2125	1987–2007	0.87	0.46
LG10	−71.3	59.2	2619	1993–2005	0.65	0.44
LG35	−76.0	65.0	2345	1994–2007	0.62	0.55
LG59	−73.5	76.9	2565	1994–2003	0.48	0.51
Mount Siple	−73.2	−127.1	230	1993–2006	0.84	0.34
Nordenskiold	−73.1	−13.4	497	1995–2019	0.82	0.31
Russkaya	−74.7	−136.9	100	1981–1989	0.68	0.70
Thurston Island	−72.5	−97.6	225	2011–2021	0.83	0.45
Average (14 stations)					0.71	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 10a). During model validation for Antarctic data (Fig. S8), 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 model 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 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 0.84 °C except for Dome A (Table 4). Dome A, located at the highest point of the East Antarctic Plateau (Table S2, Fig. S11), 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 C-AIRST$_M$ is 0.71, with corresponding mean RMSE value of 0.59 °C.

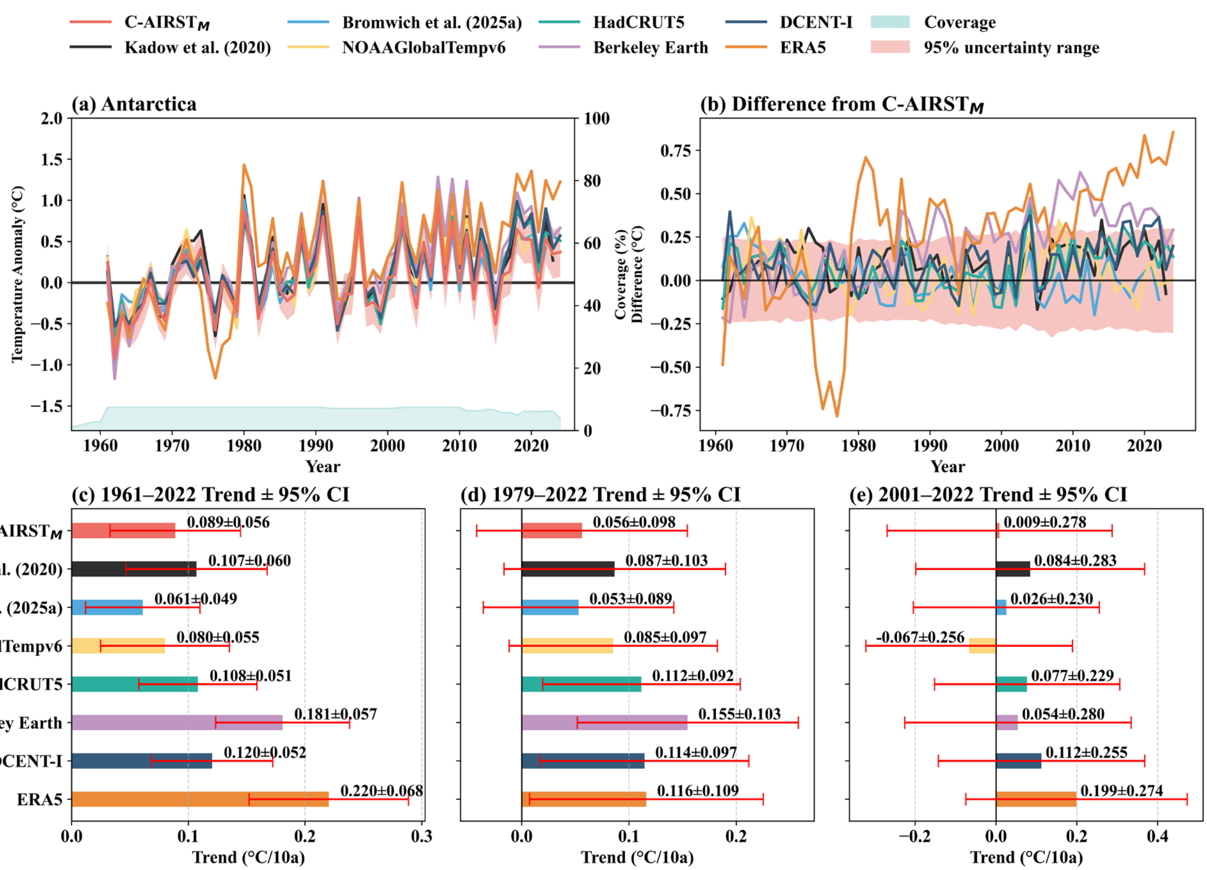


Figure 10: (a) Annual mean surface temperature anomaly time series over Antarctica from 1961 to 2024, and coverage refers to the proportion of Antarctic grid cells that are occupied by observational grid points before reconstruction; (b) Temperature difference from C-AIRST$_M$; (c–e) Linear trend of annual mean temperature and 95% confidence interval (°C per decade) during 1961–2022, 1979–2022 and 2001–2022.

Fig. 10 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 periods. To comprehensively evaluate the reliability and performance of the AI reconstructions, the results were systematically compared with the reconstructions from Kadow et al. (2020) and Bromwich et al. (2025a), HadCRUT5, NOAAGlobalTempv6, Berkeley Earth, DCENT-I and the ERA5 reanalysis product. The comparisons include a quantitative assessment of Antarctic annual mean temperature time series (Fig. 10) and Antarctic subregional temperature series (Fig. S13), as well as a qualitative analysis of the spatial distribution of temperature trends (Figs. 11 and S14).

Overall, C-AIRST$_M$ shows high consistency with multi-source observations and reconstruction products 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. 10a). C-AIRST$_M$ shows good agreement with the reconstruction from Bromwich et al. (2025a), NOAAGlobalTempv6, and HadCRUT5, with differences generally within 0.25 °C (Fig. 10b), and its 95% uncertainty interval remained around 0.25 °C.

During 1961–2022, the linear warming trends derived from C-AIRST$_M$ is 0.089 ± 0.057 °C/decade. The magnitude (Fig. 10c) is close to those reported by Bromwich et al. (2020) and NOAAGlobalTempv6, and are lower than those from HadCRUT5, Berkeley Earth, DCENT-I, and ERA5. During 1979–2022, the trend from C-AIRST$_M$ remains statistically insignificant, consistent with Bromwich et al. (2025a) and NOAAGlobalTempv6, whereas HadCRUT5, Berkeley Earth, and ERA5 still exhibit statistically significant warming trends (Fig. 10d). 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. S13, this bias mainly originates from warming contributions over East Antarctica and West 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 (see Fig. S12 for Antarctic geographical references), 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. C-AIRST$_M$ and all datasets indicate that Antarctic warming is not statistically significant from 2001 to 2022 (Fig. 10d). 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.

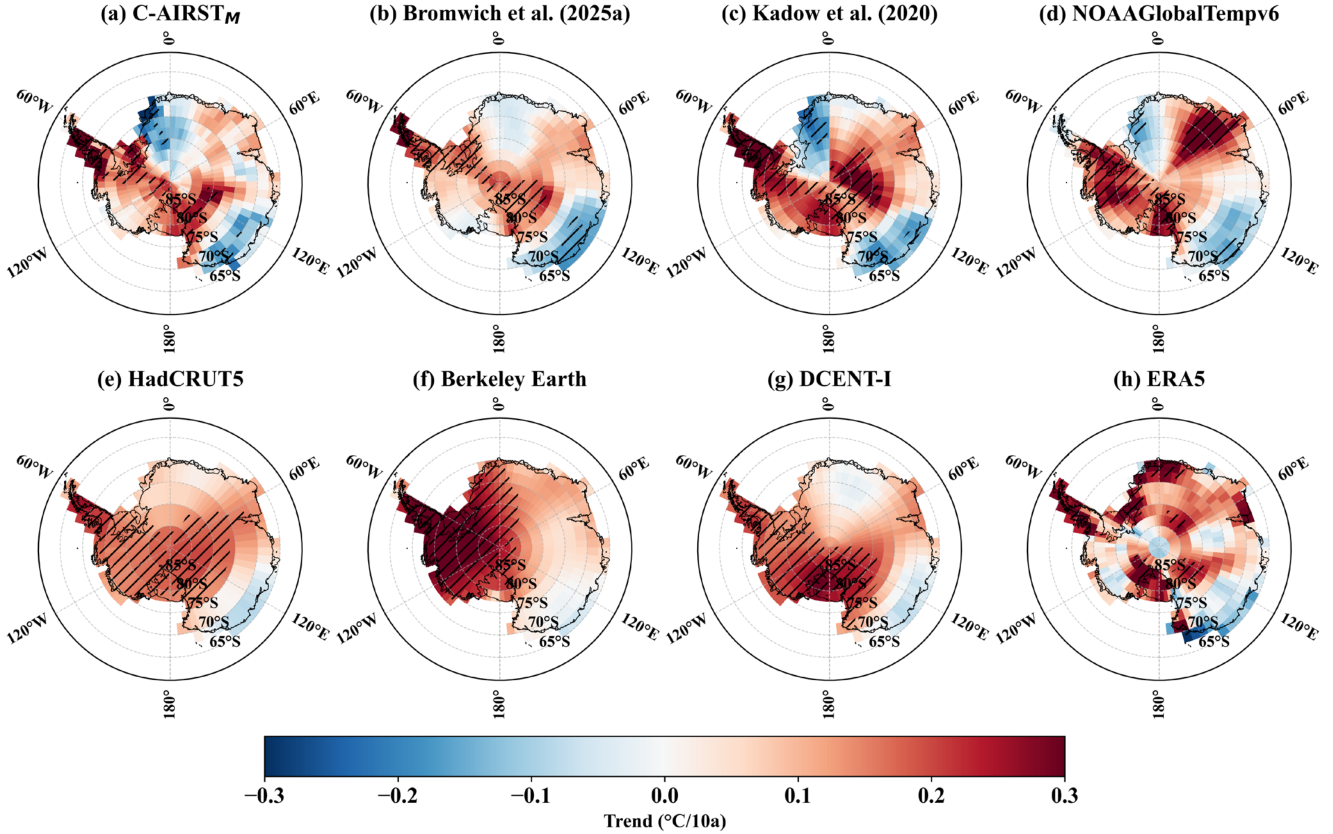


Figure 11: Spatial distribution of Antarctic annual mean surface temperature trends for 1979–2022 ("///" indicates regions where the temperature trend is statistically significant at the 0.05 level).

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). Fig. 11 presents the spatial distribution of ST trends and their statistical significance for 1979–2022 derived from C-AIRST$_M$, together with the results from other representative datasets.

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, while significant cooling is observed along the coast of Wilkes Land. In particular, the Antarctic Peninsula exhibits pronounced warming, with C-AIRST$_M$ yielding trend of 0.276 ± 0.109 °C/decade (Fig. S13). 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 C-AIRST$_M$ 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. 11c) 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 C-AIRST$_M$ (Fig. 11a), 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, Berkeley Earth and DCENT-I 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. 10d are higher in the former three datasets than in the latter two. It is worth noting that NOAAGlobalTempv6 also shows significant warming in eastern Queen Maud Land, while C-AIRST$_M$, Bromwich et al. (2025a), 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 have the potential to provide a reliable representation of the 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 11c). 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. Furthermore, with the rapid development of AI, generative diffusion models have demonstrated the potential to further improve the accuracy of historical global climate field reconstruction by employing a Video U-Net architecture to jointly model the temporal evolution of multiple spatial fields (Qian et al., 2026). By learning the probability distribution of SST fields under observational constraints and generating multiple physically plausible reconstructions, these models provide an alternative framework for probabilistic temperature field reconstruction and uncertainty quantification (Li et al., 2026a). Therefore, future developments in historical climate field reconstruction should not only account for the spherical geometry of the Earth but also preserve the spatiotemporal physical consistency of climate fields, thereby enabling more accurate and reliable historical temperature reconstructions together with probabilistic uncertainty quantification.

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, sensitivity experiments based on the PConv model indicate that the reconstructions may still exhibit a certain degree of systematic bias in the polar regions during the early period when observational records are sparse (Text S1; Figs. S5, S9). In particular, during periods with weak observational constraints (before the 1920s in the Arctic and before the 1960s in Antarctica), the model trained using 20CR and CMIP6 may introduce varying degrees of cold bias, which can reach approximately 0.5 °C under the Merge-H configuration (Fig. S9). This phenomenon is likely attributable to two main factors. First, the extremely limited availability of early polar observations provides insufficient spatial constraints for the model, leading to substantially increased uncertainty when extrapolating over large data-sparse regions (Fig. 6). Second, systematic biases inherent in the training datasets may also propagate into the reconstruction through the deep learning process. For example, the limited observational constraints in the early polar regions within the 20CR reanalysis, together with biases in CMIP6 simulations of the polar climate state and its variability, may affect the spatial relationships learned by the model. Consequently, similar early cold biases are not unique to C-AIRST$_M$; they are also evident in the reconstruction of Kadow et al. (2020), which exhibits an approximately 0.5 °C cold bias relative to other datasets (Fig. 6). Another limitation concerns the potential mismatch between the physical variables used for model training and those represented in the final reconstruction. The AI model is trained primarily on near-surface air temperature, whereas, the final reconstruction is constructed as a blended land-SAT/ocean-SST product. This introduces a potential physicalvariable mismatch between SAT and SST, which may be particularly relevant in coastal regions and near the seasonal sea-ice edge, where the distinction between land, ocean, and sea-ice surfaces can be important. Future work should investigate the effects of this SAT–SST mismatch and develop a more physically consistent treatment of land, ocean, and sea-ice-covered grid cells.

A further limitation concerns the representation of predictive uncertainty. The uncertainty estimated in this study is conditional on the adopted ensemble representation and does not fully account for all sources of observational uncertainty. In particular, the spatially correlated error covariance of the HadSST4 ensemble and the observational uncertainties of C-LSAT2.1 are not explicitly propagated through the reconstruction framework. Therefore, the estimated total predictive uncertainty should be regarded as an approximate, potentially lower-bound estimate of the true uncertainty.

Therefore, the reconstruction results for these regions and periods with weak observational constraints should be regarded as exploratory estimates with considerable uncertainty and interpreted with caution when analyzing climate trends and temperature anomalies. Future improvements could be achieved by incorporating additional early proxy records (e.g., ice cores, tree rings, and marine sediment records) and by adopting higher-resolution and more physically consistent reanalysis datasets as training constraints, thereby further improving the reliability of early climate field reconstructions in polar regions.

6 Data availability

The China global Artificial Intelligence Reconstructed Surface Temperature$_{CMIP6}$ (C-AIRST$_M$) datasets are publicly available at https://doi.org/10.6084/m9.figshare.33122444 (Ouyang et al., 2026a). They can also be accessed at http://www.gwpu.net/en/h-col-103.html (last access: 30 July 2026) for free.

7 Code availability

The code utilized in this project can be downloaded here at https://doi.org/10.5281/zenodo.19428292 (Ouyang et al., 2026b).

8 Conclusions

This study employed an AI model (PConv) to reconstruct global ST fields from observational datasets and performed a corresponding uncertainty estimation. Based on this framework, we developed a monthly global surface temperature dataset spanning 1850–2024, named the China global Artificial Intelligence Reconstructed Surface Temperature$_{CMIP6}$ (C-AIRST$_M$). The results demonstrate that C-AIRST$_M$ is consistent with multiple representative climate datasets. It exhibits good temporal and spatial continuity in low- and mid-latitude regions and over Antarctica after 1961, suggesting that it can serve as a useful dataset for extending long-term climate records, assessing polar climate change, and supporting climate monitoring, detection, and attribution. The main conclusions of this study are as follows:

(1) The PConv image inpainting framework effectively fills missing grid values and reproduces large-scale ENSO spatial structures under sparse early observational coverage. For pre-1900 periods with limited terrestrial station data, C-AIRST$_M$ reasonably recovers historical El Niño and La Niña signals consistent with HadCRUT5 and Berkeley Earth. C-AIRST$_M$ fills the 1910–1920 data gap in the Niño 3.4 index shown by HadCRUT5 and accurately captures well-documented historical El Niños events. Zonal comparisons indicate inter-dataset discrepancies are controlled within ±0.1 °C after 1900 across the 60°S–60N band; large divergences (0.4–1.0 °C) persist over Africa and South America before 1900 when observational coverage is below 40%, which requires cautious interpretation of early continental results.

(2) Over 1850–2024, C-AIRST$_M$ yields a GMST trend of 0.064 ± 0.006 °C per decade, nearly identical to HadCRUT5 (0.065 ± 0.006 °C) and DCENT-I (0.065 ± 0.006 °C), slightly lower than Berkeley Earth (0.070 ± 0.006 °C) and marginally higher than NOAAGlobalTempv6 (0.058 ± 0.006 °C). The 2024 pre-industrial (1850–1900) warming level of C-AIRST$_M$ reaches 1.49 °C, falling within the 1.43–1.62 °C range of all benchmark datasets. C-AIRST$_M$ accurately reproduces the four-stage GMST evolution: stable temperatures pre-1910, rapid warming 1910–1940, mild cooling mid-20th century, and accelerated warming post-1970. A small ~0.1 °C warm bias of C-AIRST$_M$ during 1930–1940 originates from revised ERI measurement bias corrections in HadSST4.2.0.0. The 95% GMST predictive uncertainty gradually declines from ~0.2 °C in the 1850s to ~0.05 °C after 1960, with temporary uncertainty spikes during two World Wars driven by reduced global observational sampling. Spatially, polar regions feature uncertainty exceeding 0.3 °C, and Antarctic uncertainty is universally higher than the Arctic due to far sparser historical stations.

(3) Before 1920, Arctic reconstructions from C-AIRST$_M$ show a systematic cold bias of ~0.6 °C relative to HadCRUT5, NOAAGlobalTempv6 and Berkeley Earth; for 1850–1960, Antarctic fields are ~0.5 °C colder than NOAAGlobalTempv6, which delivers full Antarctic spatial extrapolation. The 95% uncertainty bands hit ±0.5 °C for pre-1920 Arctic and ±0.8 °C for pre-1960 Antarctica, meaning these early polar outputs are exploratory rather than definitive. For continental land areas with sufficient post-1900 observations, inter-dataset differences shrink below 0.1 °C. Among all continents, Europe exhibits the strongest 1850–2024 warming trend (0.090 ± 0.014 °C/decade), while Oceania has the weakest trend (0.039 ± 0.008 °C). Consistent with all reference datasets, Southern Hemisphere warming rates are markedly lower than Northern Hemisphere counterparts, driven by vast ocean heat uptake.

(4) Although the effective observational coverage of Antarctica remains only ~8% after 1961, C-AIRST$_M$ achieves an average correlation coefficient of 0.71 and mean RMSE of 0.59 °C against 14 unused independent Antarctic stations. Unlike Kadow et al. (2020), which shows artificial abrupt longitudinal trend shifts near 0° longitude, C-AIRST$_M$ produces smooth, physically reasonable Antarctic spatial patterns consistent with Bromwich et al. (2025a) and NOAAGlobalTempv6. Over 1961–2022, the continental mean Antarctic warming trend of C-AIRST$_M$ is 0.089 ± 0.057 °C per decade, statistically significant, matching Bromwich et al. (2025a) and NOAAGlobalTempv6. C-AIRST$_M$

successfully captures the robust warming signal over the Antarctic Peninsula (0.276 ± 0.109 °C/decade) and significant cooling along Wilkes Land coast. For 2001–2022, no statistically significant Antarctic warming is detected by C-AIRST$_M$, highlighting the dominant role of internal climate variability over external anthropogenic forcing in recent Antarctic change.

A minor limitation is that C-LSAT2.1 land temperature is treated as a deterministic input without ensemble error propagation. According to Li et al. (2026b), the maximum 1σ land observational uncertainty reaches only 0.065 °C in 1850 and falls below 0.02 °C after the early 20th century, exerting only a minor influence on total predictive uncertainty. As such, the uncertainty values provided by C-AIRST$_M$ represent a lower bound of the full true uncertainty across all land grids. The PConv network adopts planar two-dimensional convolution, which may introduce minor spatial artifacts near polar regions and global map boundaries due to the spherical-to-planar projection. All AI-based reconstructions including C-AIRST$_M$ display non-negligible cold biases over sparsely observed pre-1920 Arctic and pre-1960 Antarctica, originating from insufficient spatial constraints and inherent biases within 20CR/CMIP6 training data. Future work will incorporate spherical convolution architectures, physical-informed loss functions, and paleoclimate proxy records (ice cores, tree rings) to mitigate high-latitude distortion and early-period biases. Additionally, an ensemble version of C-LSAT will be developed to propagate land observational uncertainty and achieve complete full-field uncertainty quantification for global temperature fields.

Overall, this study verifies the great potential of ensemble deep learning for long-term climate field reconstruction. The C-AIRST$_M$ dataset delivers spatially complete monthly global surface temperature anomalies at 5°×2.5° resolution spanning 1850–2024, with uncertainty estimates in gridded temperature products. C-AIRST$_M$ demonstrates reasonable temporal continuity and generally consistent spatial structures across mid/low latitudes and post-1961 Antarctica, providing valuable support for extending long climate records, polar climate diagnosis, and global climate monitoring, detection and attribution research.

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

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Inform. Fusion, 76, 243–297, https://doi.org/10.1016/j.inffus.2021.05.008, 2021.

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.

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.

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.

Kendall, A. and Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? in: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017, 30, 5574–5584, 2017.

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.

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.

Lakshminarayanan, B., Pritzel, A., and Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles, in: Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017, 30, 6402–6413, 2017.

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

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, H., Wang, Y., Yang, K., Huang, G., Xia, X., Chen, Z., Tao, W., Lyu, C., Chen, L., Zhang, M., Hu, K., Gong, H., Fu, D., and Lin W.: Probabilistic reconstruction of global sea surface temperature using generative diffusion models, arXiv preprint arXiv:2603.16272, https://doi.org/10.48550/arXiv.2603.16272, 2026a.

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., 131, e2025JD044732. https://doi.org/10.1029/2025JD044732, 2026b.

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$_{\mathrm{CMIP6}}$ (C-AIRST$_{\mathrm{M}}$), figshare [data set], https://doi.org/10.6084/m9.figshare.33122444, 2026a.

Ouyang, C., Li, Q., Li, Z., and Wei, S: An AI-Driven Reconstruction of Global Surface Temperature with Emphasis on Refining the Antarctic Record, Zenodo [code], https://doi.org/10.5281/zenodo.19428292, 2026b.

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.

Qian, Z., Liu, T., Bathiany, S., Yang, S., Hess, P., Bochow, N., Burmester, C., Gelbrecht, M., Groenke, B., and Boers, N.: Generative deep learning improves reconstruction of global historical climate records, arXiv preprint arXiv:2602.16515, https://doi.org/10.48550/arXiv.2602.16515, 2026.

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

Sandford, C., and Rayner, N.: Addressing the World War 2 Warm Anomaly in HadSST.4.2.0.0, Int. J. Climatol., e70388, https://doi.org/10.1002/joc.70388, 2026.

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.

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 J.: 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.

Valdenegro-Toro, M. and Mori, D. S.: A deeper look into aleatoric and epistemic uncertainty disentanglement, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022, 1508–1516, https://doi.org/10.1109/CVPRW56347.2022.00173, 2022.

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.

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.