Supplement of Earth Syst. Sci. Data, 18, 6503–6526, 2026
https://doi.org/10.5194/essd-18-6503-2026-supplement

Supplement of

An AI-driven reconstruction of global surface temperature with emphasis on refining the Antarctic record

Chenxi Ouyang et al.

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

S1. Sensitivity experiments and model evaluation.

As shown in Fig. S3, we designed a series of sensitivity experiments using different Sea Surface Temperature (SST) datasets and training datasets to evaluate the performance of the reconstruction framework. Specifically, C-LSAT2.1 (Wei et al., 2025) was merged separately with ERSSTv6 (Huang et al., 2025a; 2025b) and the median field of HadSST4.2.0.0 (Kennedy et al., 2019) following the land–ocean merging method described in Section 2.2 (LSAT and SST merging method). These two merged datasets were referred to as Merge-E and Merge-H, respectively. Furthermore, two independent sets of training samples were used to train the PConv models: 80 ensemble members from The Twentieth Century Reanalysis version 3 (20CRv3) dataset (Slivinski et al., 2019) and 105 ensemble members from Coupled Model Intercomparison Project Phase 6 (CMIP6) historical simulations (Table S1; Eyring et al., 2016). The corresponding trained models were denoted as 20CR-AI and CMIP6-AI, respectively. Under the masks provided by Merge-E and Merge-H, the 20CR-AI and CMIP6-AI models were independently used to reconstruct the corresponding test sets, followed by model evaluation. This resulted in four sensitivity experiment configurations: Merge-E 20CR-AI, Merge-E CMIP6-AI, Merge-H 20CR-AI, and Merge-H CMIP6-AI (Fig. S3).

During model training and evaluation, the leave-one-out strategy described in Section 2.3 (AI training and reconstruction process) was adopted, where one ensemble member was withheld as the test dataset while the remaining members were used for model training. Since these sensitivity experiments were designed to investigate the impacts of different observational configurations and training datasets on reconstruction performance, only one training and reconstruction experiment was conducted for each configuration. The evaluation metrics included Bias, Correlation Coefficient (CC), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

The corresponding results are presented in Figs. S4–S9. Overall, the test performance under the Merge-E configuration was generally superior to that under the Merge-H configuration (Fig. S4). This difference is mainly attributed to the fact that the ocean component of Merge-E, ERSSTv6, applies an Artificial Neural Network (ANN)-based spatial extrapolation method (Huang et al., 2025a; 2025b) and incorporates more extensive SST information (Fig. S4). Therefore, Merge-E and Merge-H are not completely comparable due to differences in their SST data coverage and reconstruction methods.

For the Merge-E configuration, the performances of Merge-E 20CR-AI and Merge-E CMIP6-AI were generally comparable, with Merge-E 20CR-AI showing slightly better performance. Specifically, Merge-E 20CR-AI achieved higher reconstruction accuracy for the global area-weighted monthly temperature anomalies than Merge-E CMIP6-AI, with improved CC, RMSE, and MAE values (Fig. S5). It also showed better spatial correlation and spatial RMSE performance (Figs. S7 and S8), whereas the two configurations exhibited comparable results in terms of the averaged temporal correlation and RMSE (Fig. S6). These results indicate that when observational coverage is relatively high, the 20CR-AI model performs slightly better than CMIP6-AI. However, analysis of the final reconstruction products revealed that Merge-E may have learned excessive oceanic characteristics during the early period (before the 20th century). Specifically, the spatial patterns of ocean temperature anomalies may have been propagated into poorly observed terrestrial regions, resulting in warm biases of approximately 0.1–0.2°C over low- and mid-latitude land areas compared with NOAAGlobalTempv6, despite both datasets using the same SST component (ERSSTv6). Therefore, these biases are primarily attributed to the land component, suggesting that the Merge-E configuration may introduce unrealistic land temperature signals during early periods (Fig. S10).

For the Merge-H configuration, Merge-H CMIP6-AI generally performed better than Merge-H 20CR-AI. The former showed improved reconstruction accuracy for global area-weighted monthly temperature anomalies (Fig. S5), as well as better averaged temporal correlation and RMSE metrics (Fig. S6). In contrast, Merge-H 20CR-AI exhibited slightly better performance in terms of averaged spatial correlation and spatial RMSE, although the differences between the two configurations were relatively small. Although both monthly spatial correlation coefficients and grid-point temporal correlation coefficients can be used to evaluate reconstruction performance, this study places greater emphasis on grid-point temporal correlation because it directly assesses the capability of the reconstruction to capture the temporal evolution of climate variability at each grid point.

Considering the overall performance of different sensitivity experiments, Merge-H CMIP6-AI was selected as the final configuration. Specifically, C-LSAT2.1 merged with HadSST4.2.0.0 was used as the observational input, and the PConv model trained with CMIP6 historical simulations was adopted for reconstruction. Reconstruction and uncertainty estimation presented in the main text was performed using this configuration, resulting in the China global Artificial Intelligence Reconstructed Surface Temperature$_{\mathrm{CMIP6}}$ (C-AIRST$_{\mathrm{M}}$).

Table S1: CMIP6 ensemble members for training AI model.

Model	Number	Ensemble
ACCESS-CM2	5	r1i1p1f1–r5i1p1f1
ACCESS-ESM1-5	3	r1i1p1f1–r3i1p1f1
BCC-CSM2-MR	3	r1i1p1f1–r3i1p1f1
BCC-ESM1	3	r1i1p1f1–r3i1p1f1
CanESM5	5	r1i1p1f1–r5i1p1f1
CESM2	1	r1i1p1f1
CESM2-WACCM	1	r1i1p1f1
CMCC-ESM2	1	r1i1p1f1
CNRM-CM6-1	10	r1i1p1f2–r10i1p1f2
EC-Earth3	4	r1i1p1f1, r2i1p1f1, r9i1p1f1, r12i1p1f1
FGOALS-g3	4	r1i1p1f1, r3i1p1f1, r4i1p1f1, r5i1p1f1
FIO-ESM-2-0	3	r1i1p1f1–r3i1p1f1
GISS-E2-1-G	12	r1i1p1f1–r6i1p1f1, r1i1p3f1–r6i1p3f1
GISS-E2-1-G-CC	1	r1i1p1f1
GISS-E2-1-H	11	r1i1p1f1–r6i1p1f1, r1i1p3f1–r5i1p3f1
GISS-E2-2-G	6	r1i1p1f1–r6i1p1f1
GISS-E2-2-H	5	r1i1p1f1–r5i1p1f1
INM-CM4-8	1	r1i1p1f1
INM-CM5-0	1	r1i1p1f1
IPSL-CM6A-LR	6	r1i1p1f1–r6i1p1f1
MIROC6	5	r1i1p1f1–r5i1p1f1
MPI-ESM1-2-LR	3	r1i1p1f1–r3i1p1f1
MRI-ESM2-0	5	r1i1p1f1–r5i1p1f1
NorESM2-LM	3	r1i1p1f1–r3i1p1f1
NorESM-MM	3	r1i1p1f1–r3i1p1f1

Table S2: Antarctic observational stations information.

Used Station	Latitude	Longitude	Elevation(m)	Source	Used Station	Latitude	Longitude	Elevation(m)	Source
Adelaide	-67.80	-68.90	26	2	Mawson	-67.60	62.90	8	1
Amundsen-Scott	-90.00	0.00	2830	5	Mawson Base	-67.60	62.88	14	2
Baldrick AWS	-82.77	-13.05	1968	2	Mcmurdo Sound Naf	-77.88	166.73	24.1	2
Base Almirante Brown	-64.88	-62.88	1	2	Mirny	-66.55	93.02	40	1, 2
Base Arturo Prat	-62.50	-59.68	5	2	Molodeznaja	-67.67	45.85	50	2
Base Baia Terra Nova	-74.70	164.10	92	2	Neumayer	-70.67	-8.25	50	2
Base Bernardo O'Higgins	-63.32	-56.68	10	2	Nico	-89.00	89.67	2935	2
Base Jubany	-62.23	-58.63	11	2	Novolazarevskaja	-70.77	11.83	119	1, 2
Base Marambio	-64.23	-56.72	198	2	O'Higgins	-63.30	-57.90	10	2
Base San Martin	-68.12	-67.13	7	2	Orcadas	-60.73	-44.73	8	1, 2
Belgrano II	-77.87	-34.62	256	1, 2	Palmer	-64.80	-64.10	8	2
Bellingshausen	-62.18	-58.88	14	2	Palmer Station	-64.77	-64.08	8	2
Bonaparte Point	-64.78	-64.07	8	2	Pegasus North	-77.95	166.50	10	2
Brianna	-83.89	-134.14	549	2	Petrel	-63.50	-57.30	18	2
Butler Isla	-72.20	-60.17	115	2	Possession Is	-71.88	171.20	30	2
Byrd Station	-80.01	-119.40	1530	2	Progress	-69.38	76.38	64	2
Cape Bird	-77.22	166.43	70	2	Rothera	-67.57	-68.12	33	2
Cape Phillips	-73.52	169.75	568	2	Sanae AWS	-71.70	-2.80	817	2
Cape Ross	-76.72	162.97	150	2	Schwerdtfeger	-79.90	169.98	60	2
Casey	-66.28	110.52	41	1	Scott	-77.80	166.80	16	2
Centro Met Antartico Pdte E	-62.18	-58.98	48	2	Scott Base	-77.84	166.76	16	1, 6
Centro Metan	-62.42	-58.88	10	2	Signy Island	-60.71	-45.60	6	2
Concordia	-75.10	123.40	3233	2	Siple Dome	-81.65	-148.78	620	2
D-47	-67.40	138.73	1560	2	Ski Blu	-74.80	-71.48	1589	2
Davis	-68.60	78.00	13	1, 2	Syowa	-69.00	39.58	21	1, 2
Dinamet Uruguay	-62.18	-58.85	17	2	Theresa	-84.60	-115.82	1463	2
Dome C	-74.50	123.00	3280	2	Vernadsky	-65.25	-64.27	11	2
Dome C II	-75.12	123.37	3250	2	Vostok	-78.45	106.87	3488	1, 2
Dome Fuji	-77.32	39.70	3810	2	Zhongshan	-69.37	76.38	18	2
Dome Plateau Eagle	-76.42	77.02	2830	2	**Validation Station**				
Dumont D'Urville	-66.67	140.02	43	1, 4	**Validation Station**	**Latitude**	**Longitude**	**Elevation(m)**	**Source**
Elaine	-83.13	174.17	60	2	Dome A	-80.40	77.40	4084	2
Elizabeth	-82.62	-137.08	549	2	Drescher	-72.87	-19.03	27	2
Esperanza	-63.20	-56.59	13	1, 2	Ferrell	-77.90	170.80	45	2
Faraday	-65.30	-64.30	11	1, 2	G3	-70.90	69.90	84	2
Ferraz	-62.10	-58.40	20	2	GC41	-71.60	111.30	2763	2
Fossil Bluff	-71.32	-68.28	69	2	General Belgrano	-78.00	-38.80	32	2
Gill	-79.98	-178.60	55	2	GF08	-68.50	102.10	2125	2
Great Wall	-62.22	-58.97	10	2	LG10	-71.30	59.20	2619	2
Halvfarryggen Ep11	-71.15	-6.68	694	2	LG35	-76.00	65.00	2345	2
Harry	-83.00	-121.38	945	2	LG59	-73.50	76.90	2565	2
Henry	-89.00	-1.02	2755	2	Mount Siple	-73.20	-127.05	230.0	2
King Sejong	-62.22	-58.75	11	2	Nordenskiold	-73.05	-13.40	497	2
Kohnenep9	-75.00	0.00	2892	2	Russkaya	-74.70	-136.90	100	2
Law Dome Summit	-66.73	112.83	1368	2	Thurston Island	-72.50	-97.60	225	2
Leningradskaya	-69.50	159.38	291	2	**Data Source**				
Lettau	-82.52	-174.45	55	2	1	SCAR READER			
Limbert	-75.87	-59.15	40	2	2	GHCNm			
Linda	-78.47	168.38	50	2	3	OSU Polor Meteorology Group			
Maitri	-70.77	11.75	50	2	4	Meteo France			
Manuela	-74.95	163.68	80	2	5	University of Wisconsin-Madison			
Marilyn	-79.95	165.12	75	2	6	NIWA			

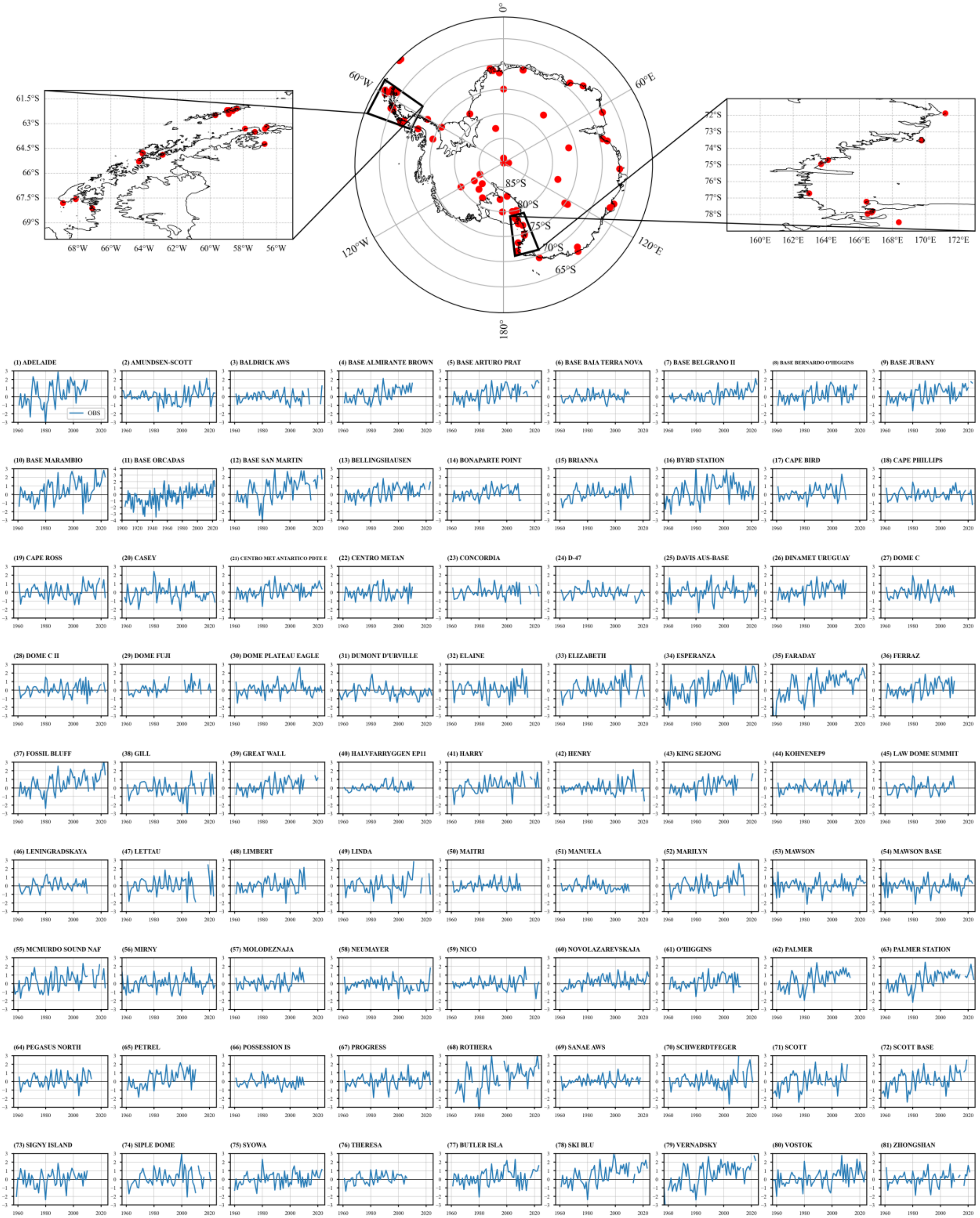


Figure S1: Antarctic stations location and annual mean temperature anomaly time series.

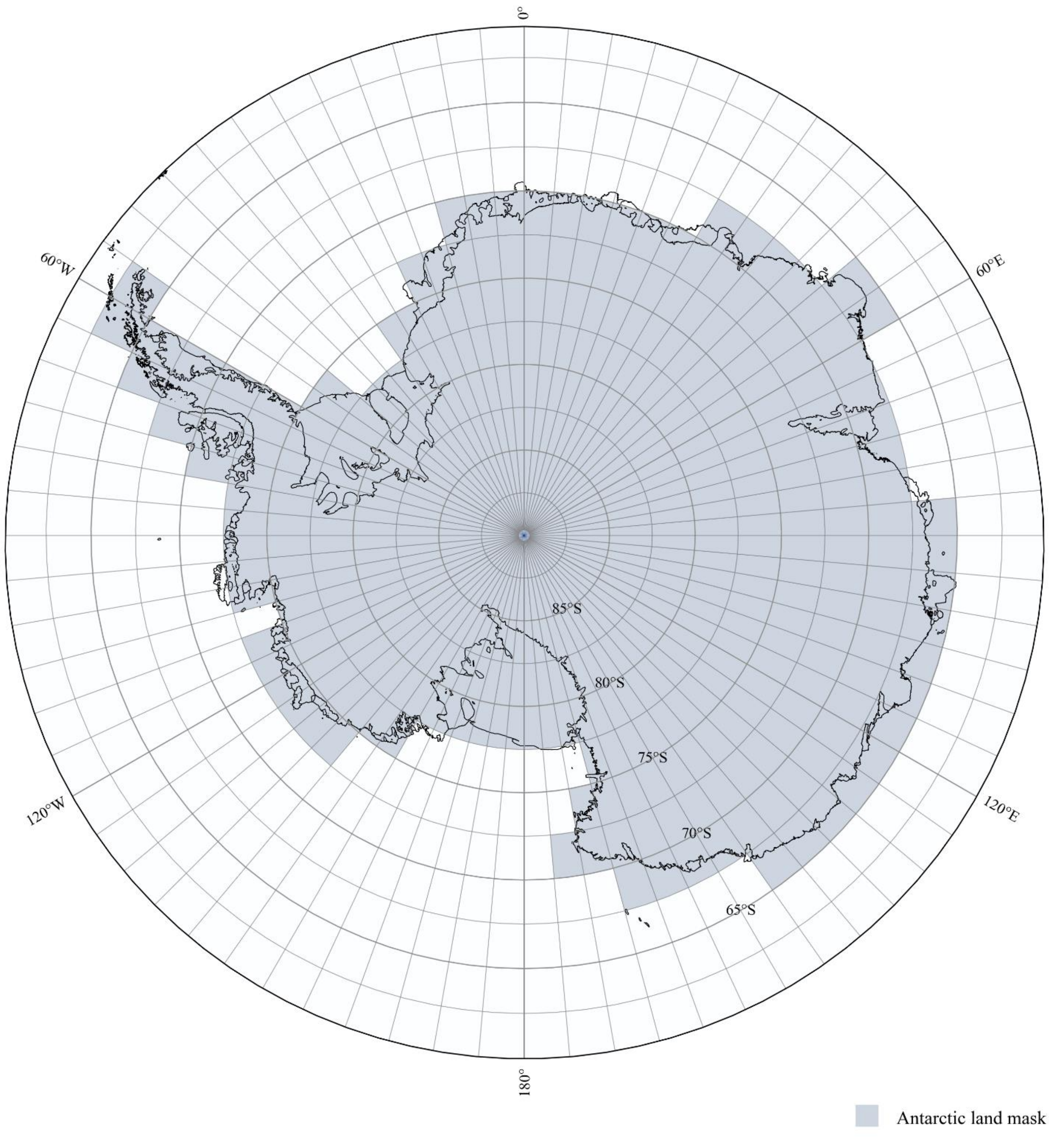


Figure S2: Antarctic land mask.

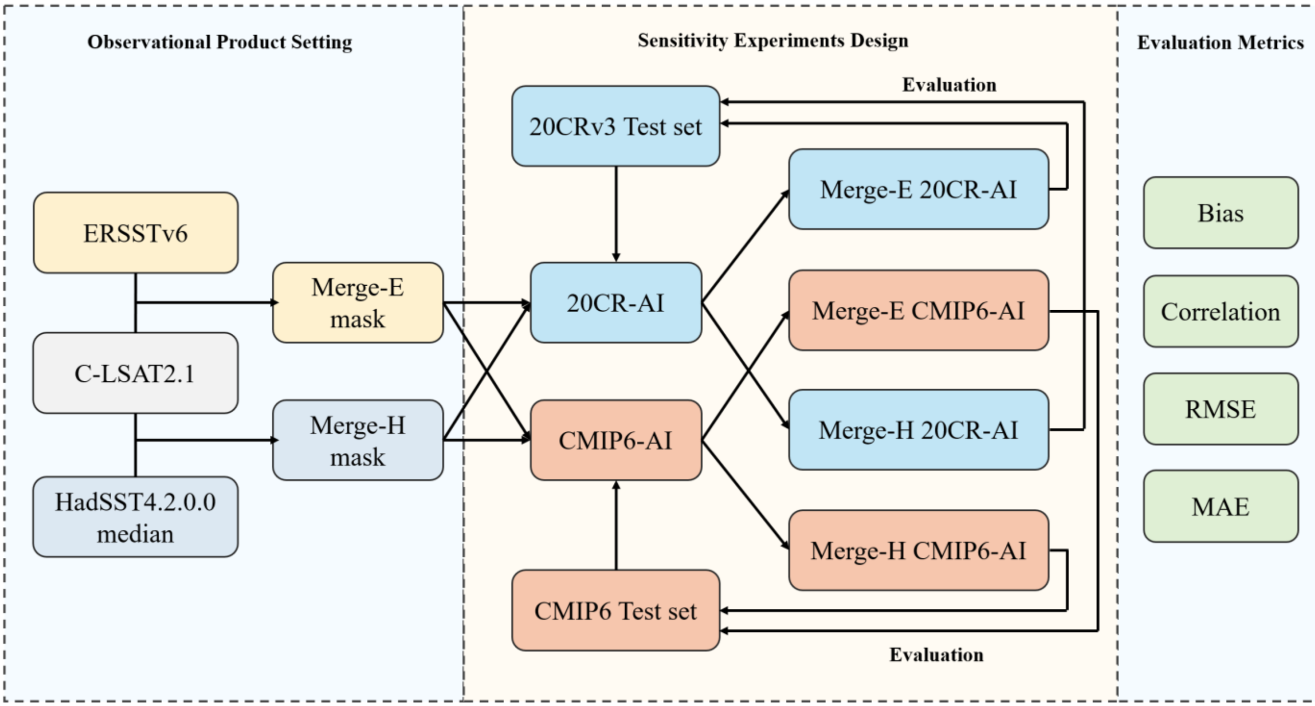


Figure S3: Sensitivity experiments and model evaluation.

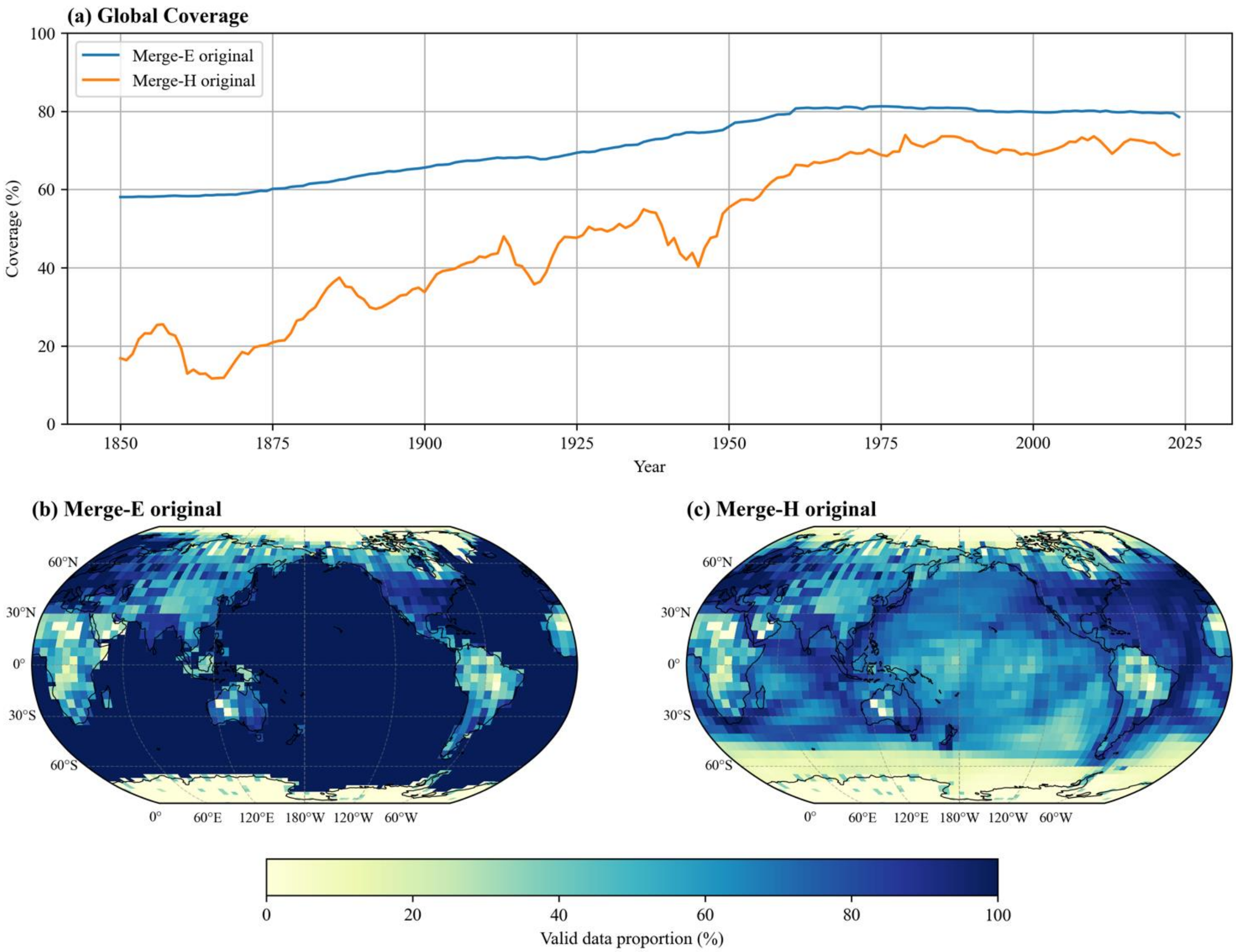


Figure S4: Global annual mean original effective data coverage for Merge-E and Merge-H during 1850–2024 (a). Spatial distribution of grid-level effective data coverage for Merge-E (b). Same as b but for Merge-H (c).

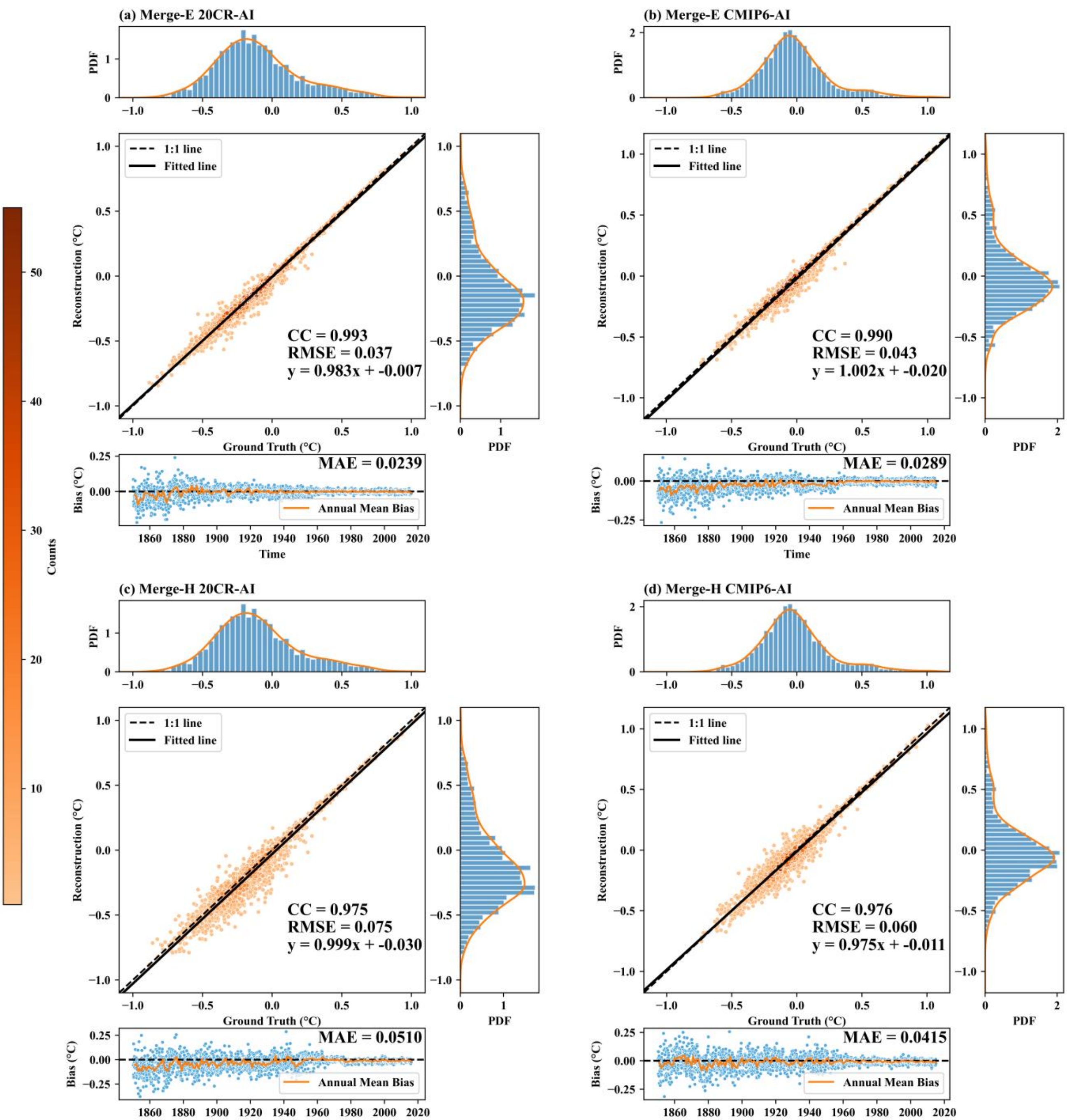


Figure S5: Model evaluation results for different sensitivity experiments. Each dot represents the monthly global area-weighted mean temperature from 1850 to 2014. The reconstruction performance was evaluated using the correlation coefficient (CC), root mean square error (RMSE), and mean absolute error (MAE). The histogram represents the probability density function (PDF) of the distribution. (a) Comparison between the 20CR test set reconstructed by Merge-E 20CR-AI and the its original test dataset. (b–d) Same as (a), but for Merge-E CMIP6-AI, Merge-H 20CR-AI, and Merge-H CMIP6-AI, respectively.

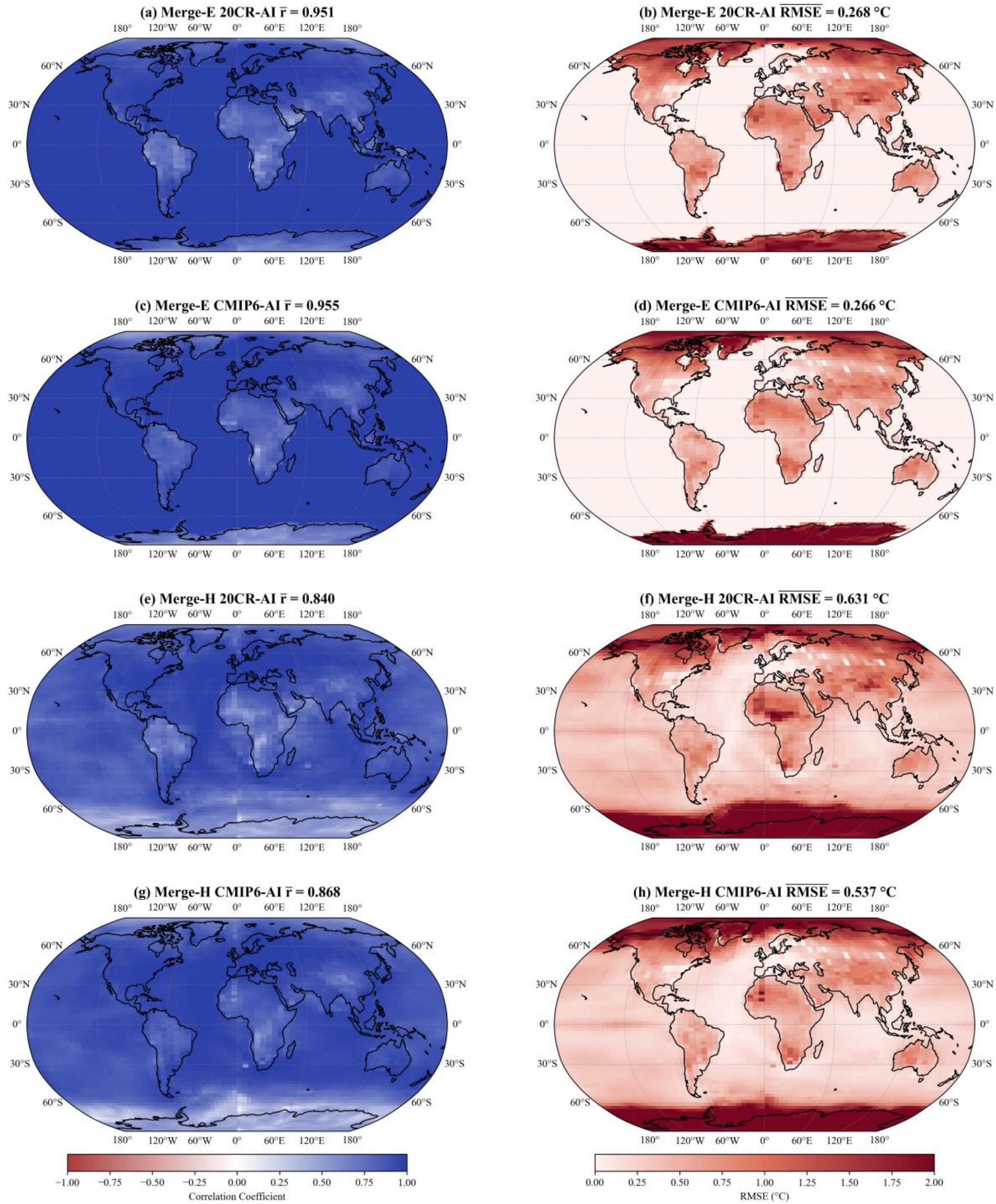


Figure S6: Monthly grid spatial validation results of the AI reconstructions for 1850–2024. Spatial correlation coefficients and RMSE between the 20CR-AI reconstruction and the original 20CR test set under the Merge-E mask (a–b). Same as a–b but for the CMIP6-AI reconstruction of the CMIP6 test set under the Merge-E mask (c–d). Same as a–b but for the 20CR-AI reconstruction of the 20CR test set under the Merge-H mask (e–f). Same as a–b but for the CMIP6-AI reconstruction of the CMIP6 test set under the Merge-H mask (g–h).

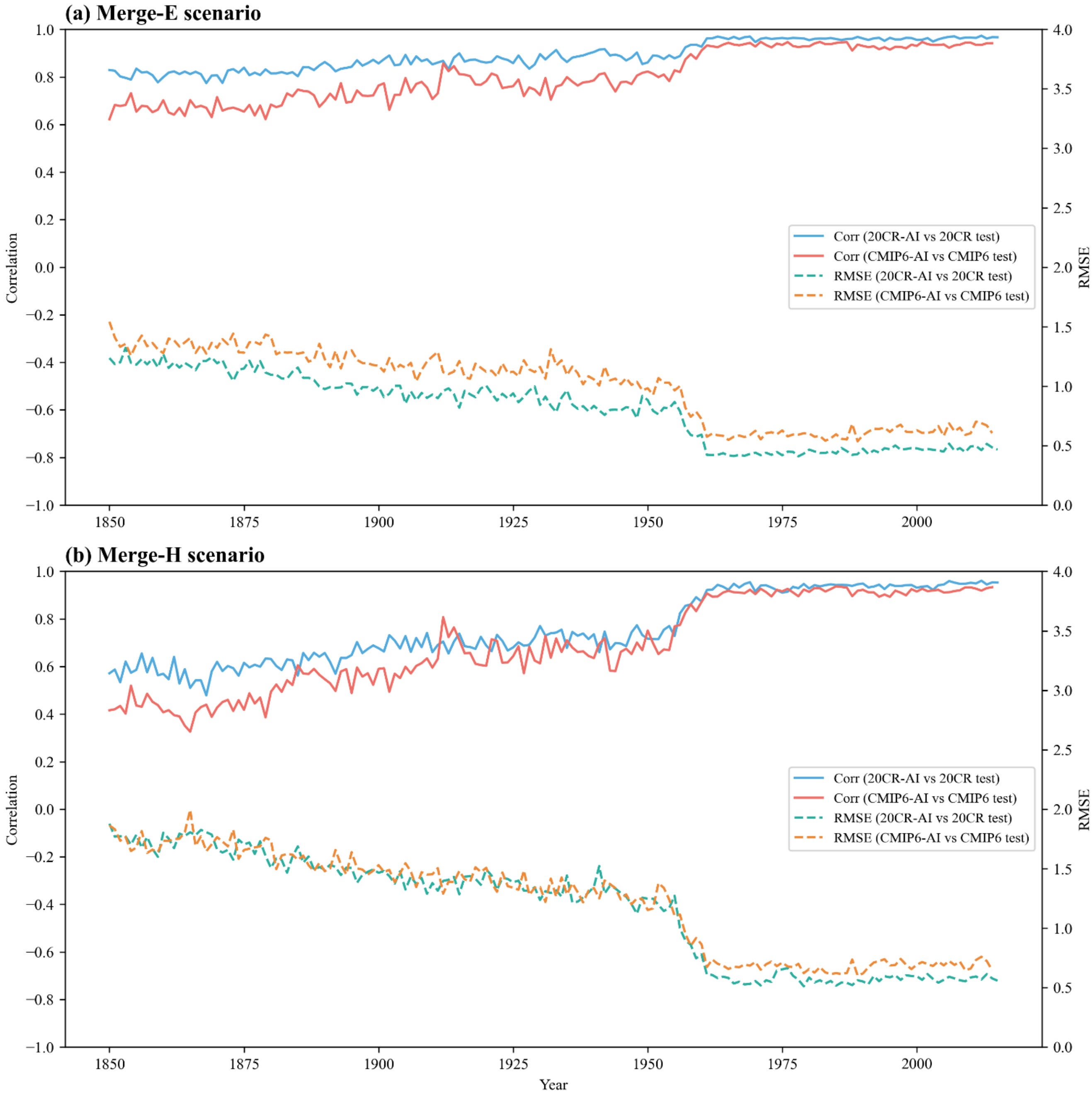


Figure S7: Global annual mean spatial grid correlation coefficients and RMSE of the 20CR and CMIP6 test sets. Annual mean spatial correlation coefficients and RMSE between the 20CR-AI reconstruction and the original 20CR test set, and between the CMIP6-AI reconstruction and the original CMIP6 test set, under the Merge-E mask (a). Same as a but under the Merge-H mask (b).

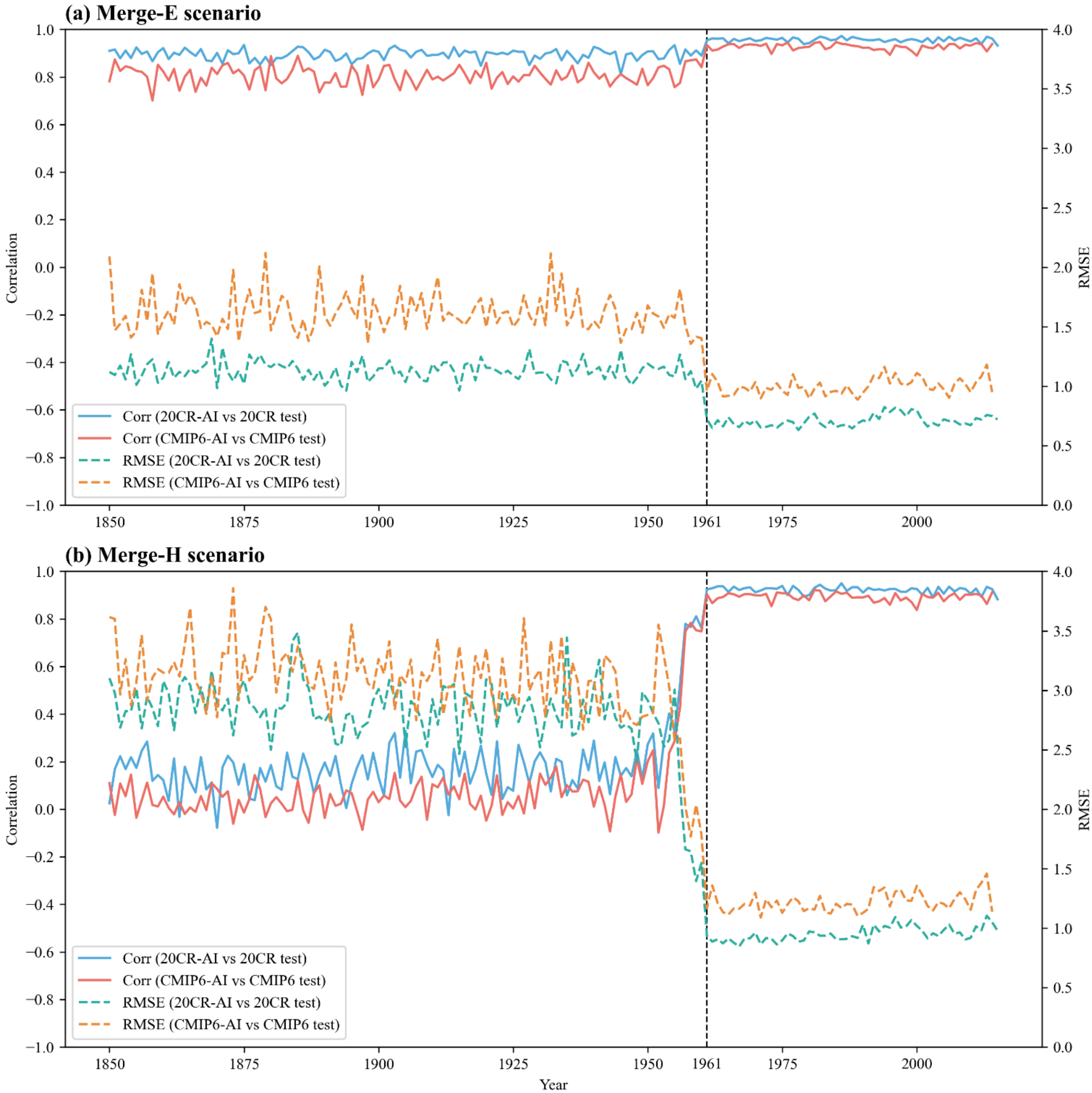


Figure S8: Antarctic annual mean spatial grid correlation coefficients and RMSE time series. Annual mean spatial correlation coefficients and RMSE between the 20CR-AI reconstruction and the original 20CR test set, and between the CMIP6-AI reconstruction and the original CMIP6 test set, under the Merge-E mask (a). Same as a but under the Merge-H mask (b).

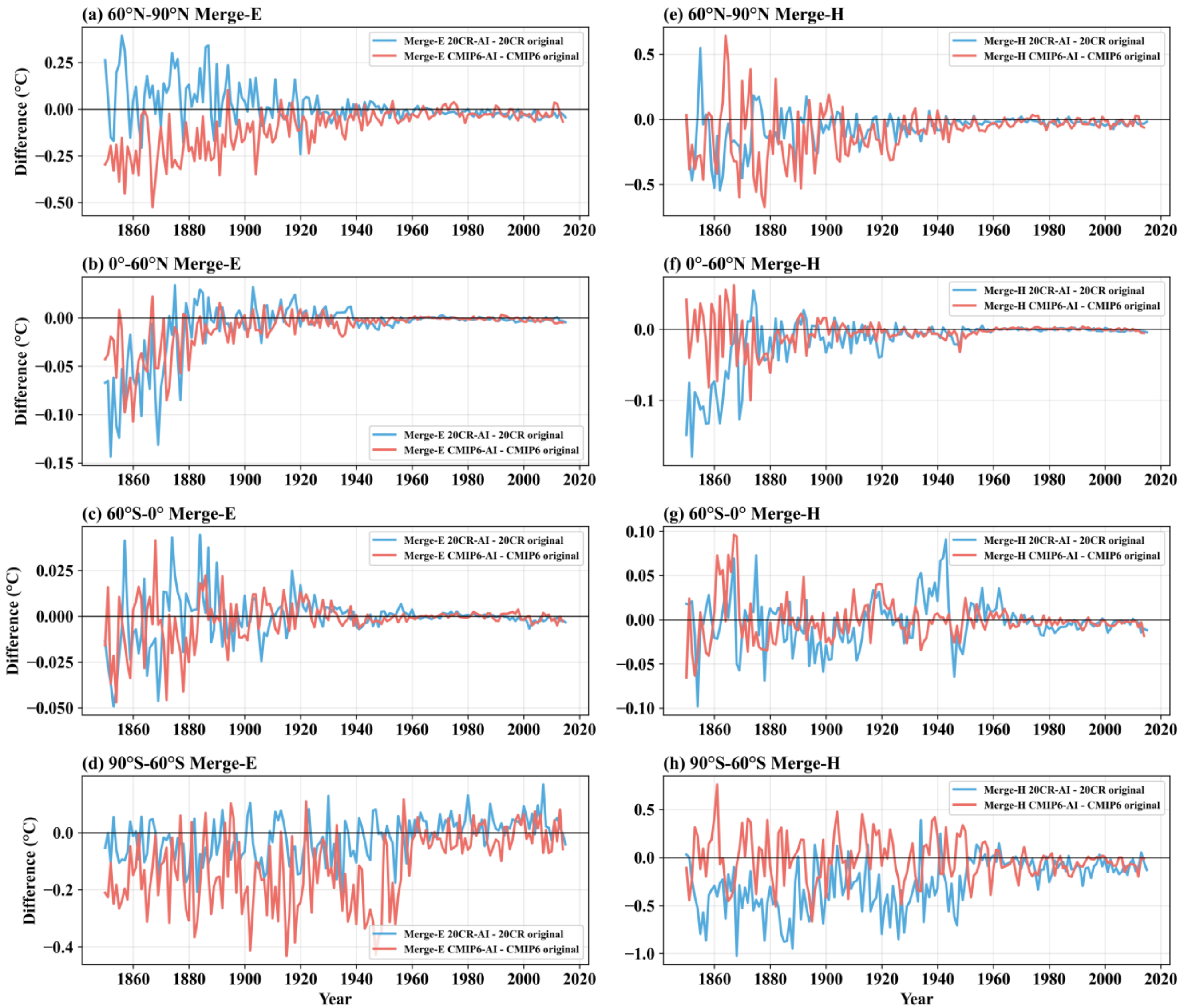


Figure S9: Annual mean difference time series across different latitude bands from 1850 to 2014. (a–d) Differences between the reconstructed test sets and original test sets under the Merge-E mask for the two models; (e–h) same as (a–d), but under the Merge-H mask.

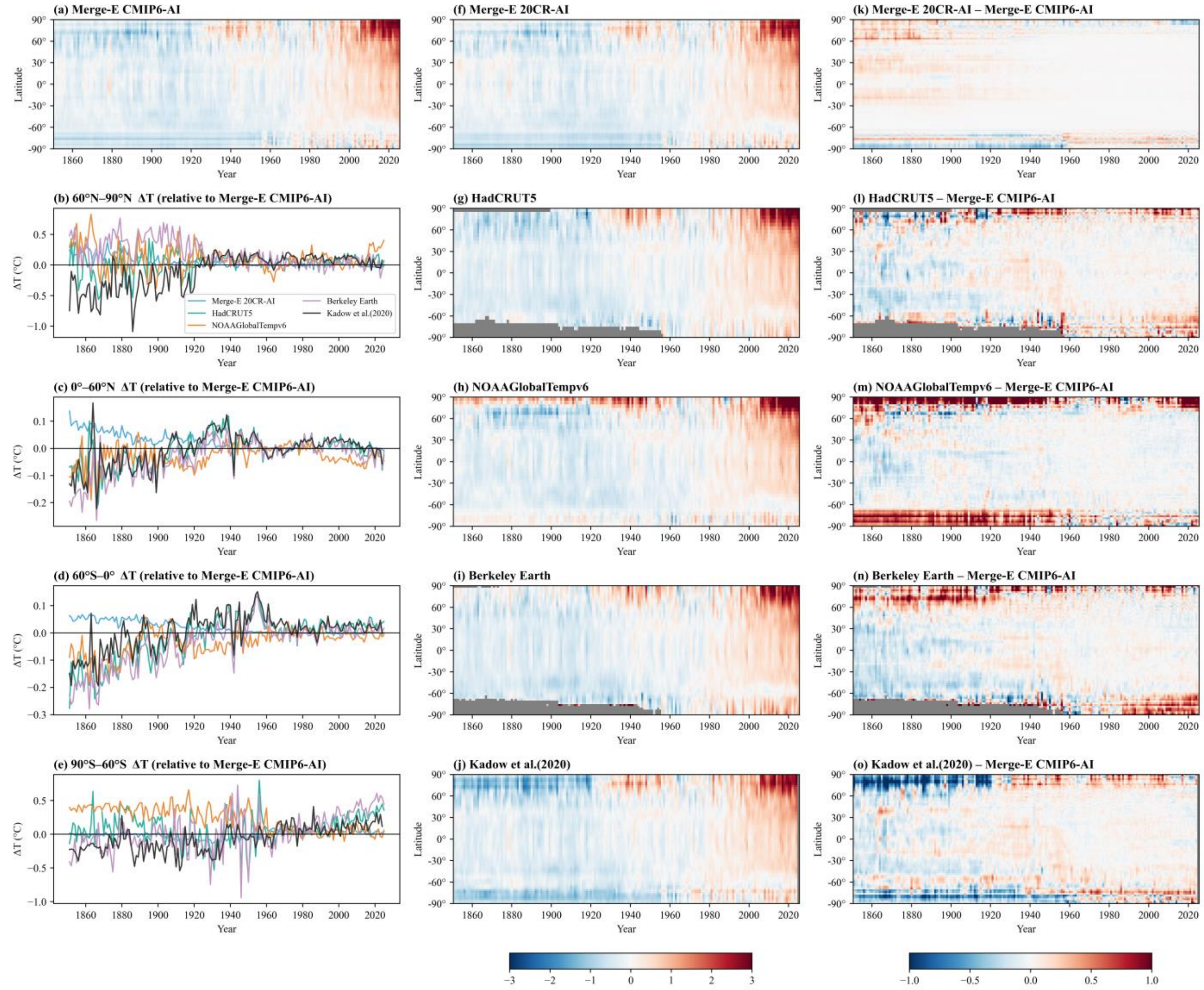
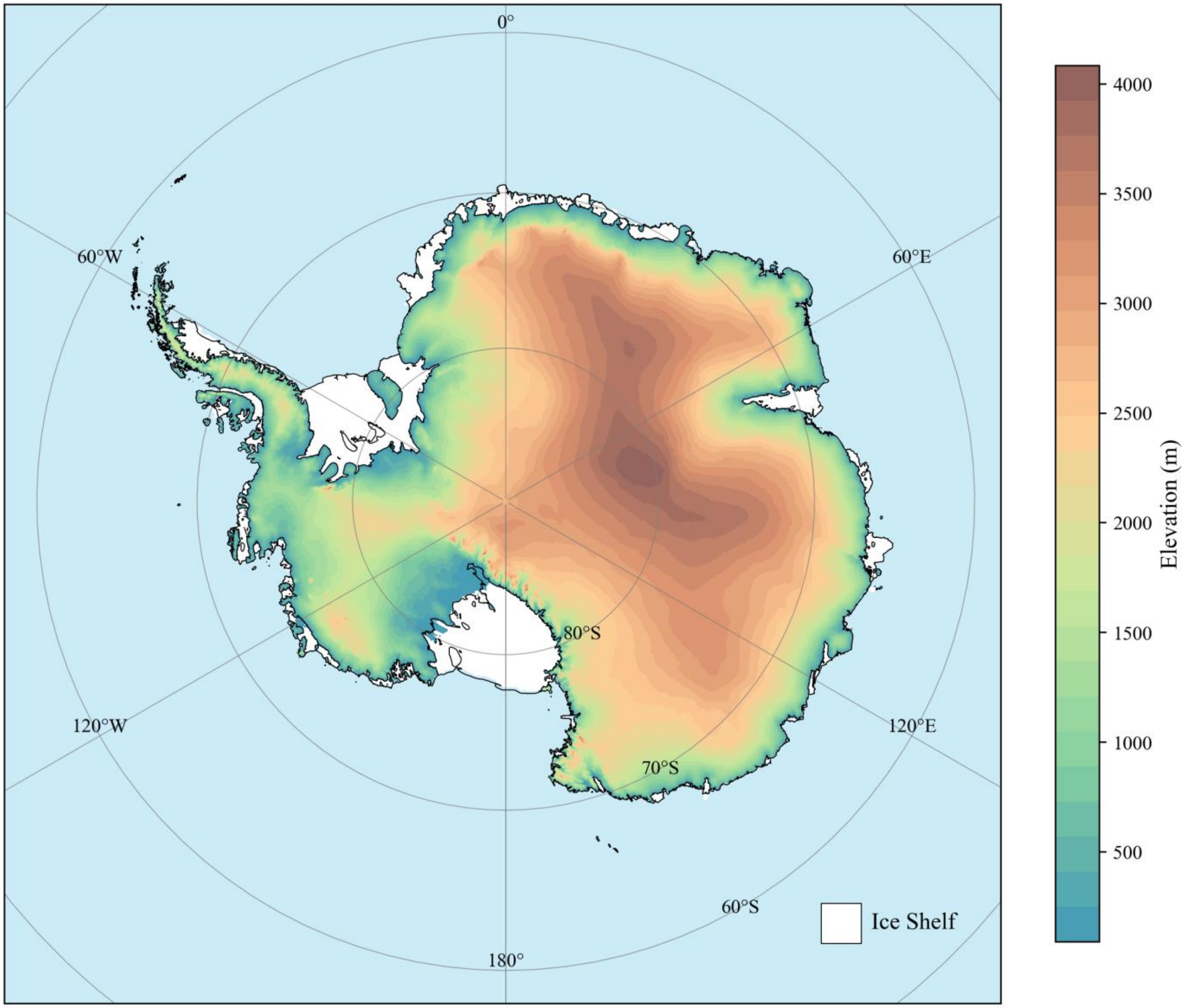


Figure S10: Zonal temperature comparison. (a) Zonal-mean temperature anomalies from Merge-E CMIP6-AI; (b–e) time series of zonal-mean temperature anomaly differences between other datasets and Merge-E CMIP6-AI across different latitude bands; (f–j) same as (a) for Merge-E 20CR-AI, HadCRUT5, NOAAGlobalTempv6, Berkeley Earth, and Kadow et al. (2020), respectively; (k–o) zonal-mean temperature anomaly differences between other datasets and Merge-E CMIP6-AI.



95

Figure S11: Topographic elevation map of the Antarctic continent.

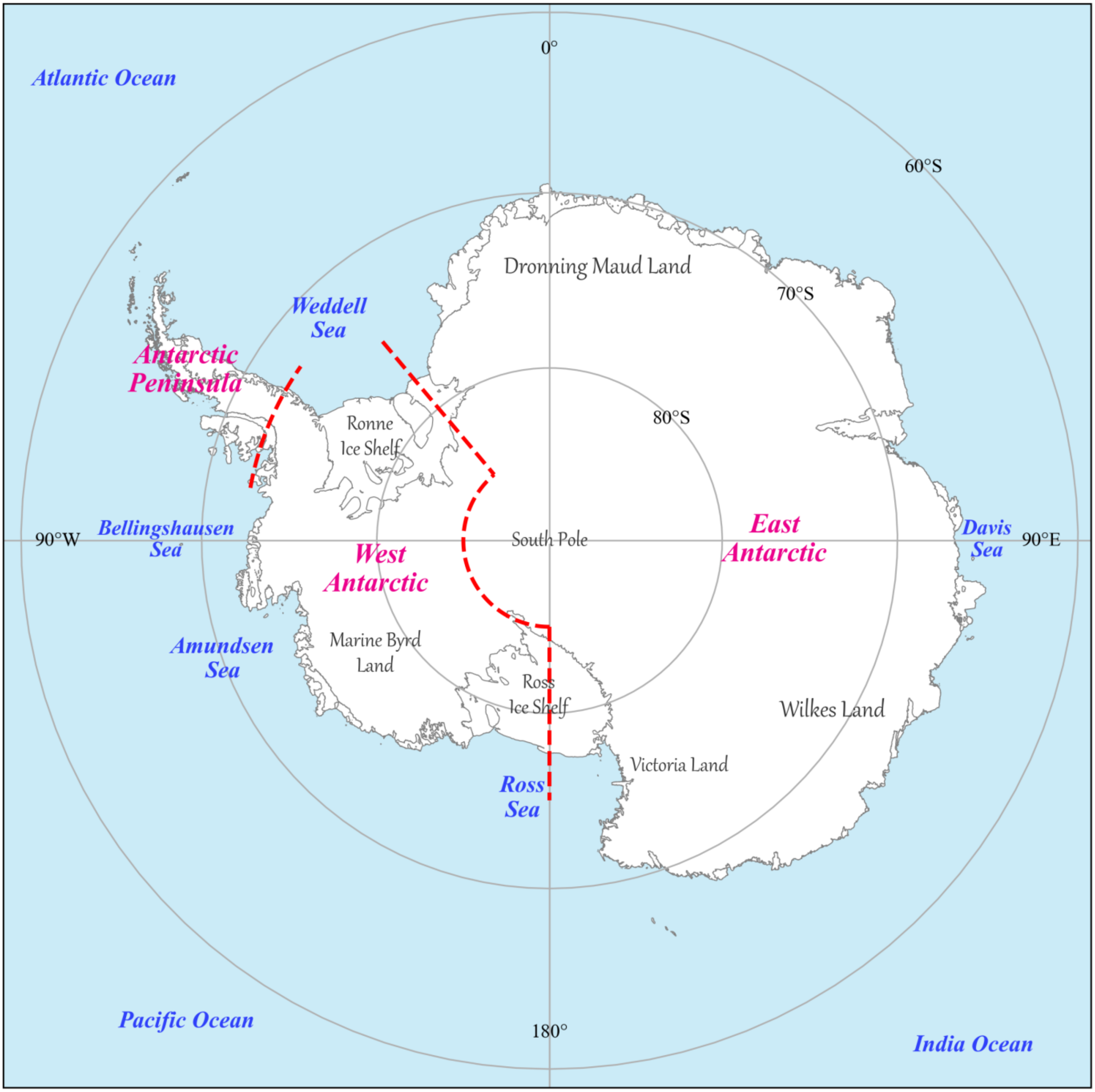


Figure S12: Geographic location map of the Antarctic continent (the red dashed lines indicate the boundaries dividing the Antarctic Peninsula, West Antarctica, and East Antarctica).

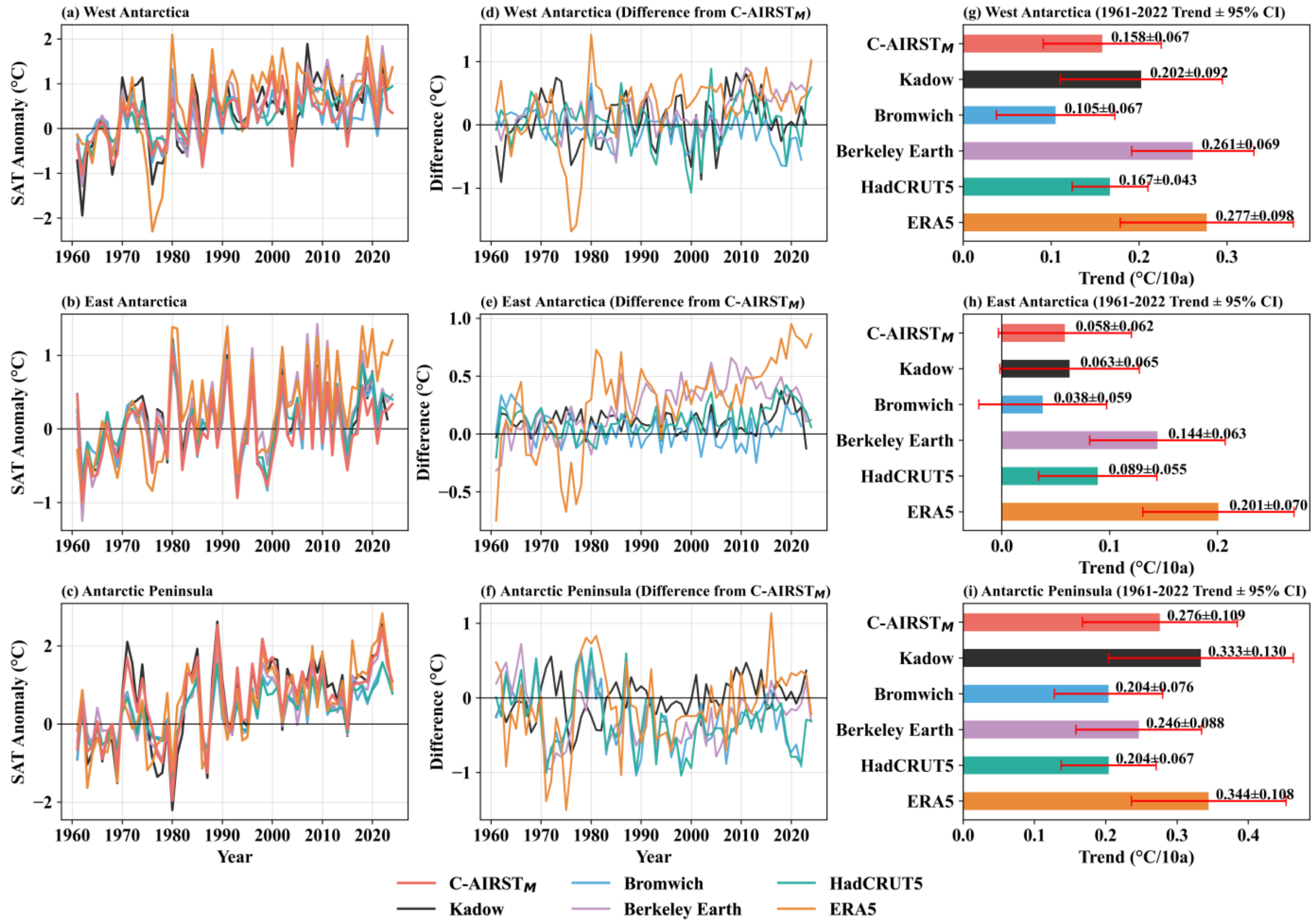


Figure S13: Area mean temperature time series and trends for Antarctic subregions.

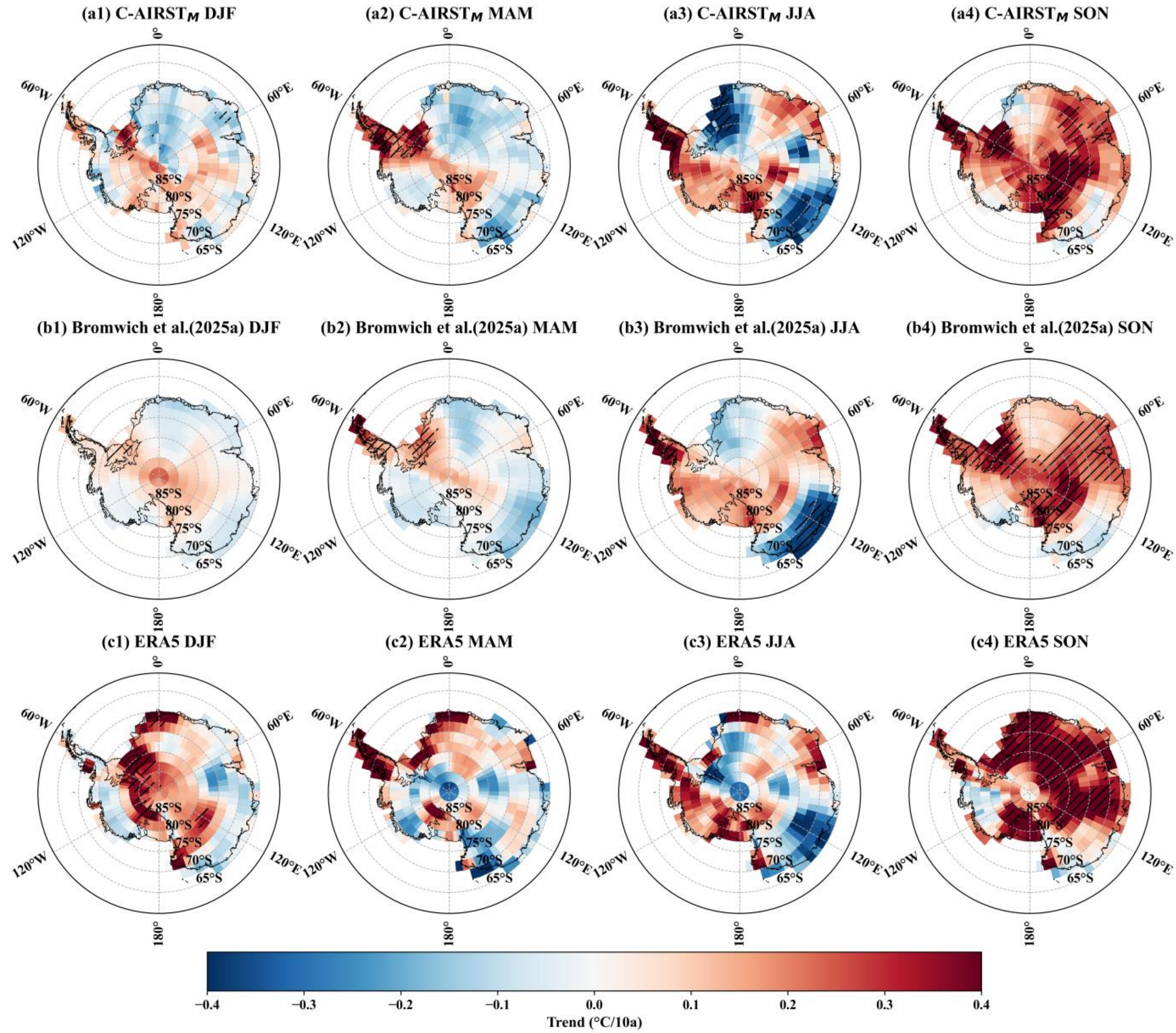


Figure S14: Spatial distribution of seasonal mean temperature trends over Antarctica during 1979–2022.

References

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.

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.

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.

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

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.