

Time-varying global energy budget since 1880 from a new reconstruction of ocean warming

Article

Supplemental Material

Wu, Q. ORCID: <https://orcid.org/0000-0002-0686-8922>,
Gregory, J. M. ORCID: <https://orcid.org/0000-0003-1296-8644>,
Zanna, L. and Khatiwala, S. (2025) Time-varying global energy
budget since 1880 from a new reconstruction of ocean
warming. *Proceedings of the National Academy of Sciences of
the United States of America*, 122 (20). e2408839122. ISSN
1091-6490 doi: 10.1073/pnas.2408839122 Available at
<https://centaur.reading.ac.uk/122225/>

It is advisable to refer to the publisher's version if you intend to cite from the
work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1073/pnas.2408839122>

Publisher: National Academy of Sciences

All outputs in CentAUR are protected by Intellectual Property Rights law,
including copyright law. Copyright and IPR is retained by the creators or other
copyright holders. Terms and conditions for use of this material are defined in
the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



1

2 **Supporting Information for**

3 **Time-varying global energy budget since 1880 from a new reconstruction of ocean warming**

4 **Quran Wu, Jonathan M. Gregory, Laure Zanna and Samar Khatiwala**

5 **Quran Wu**

6 **E-mail: quran.wu@reading.ac.uk**

7 **This PDF file includes:**

8 Figs. S1 to S20

9 Table S1

10 SI References

1. Observational Constraints

To constrain the Green's function (GF) G at a given location requires: a) observations of tracers there and b) the tracers' surface histories or boundary conditions. Here we describe how to derive these two inputs for each of the four constraints used in this study. The procedure for deriving observational GFs is summarised in Fig. S1.

A. Transient Tracers. The CFC observations are derived from the gridded GLODAP data (1), which has a near-global coverage and a nominal year of 1994. We construct the CFC-11 surface history (CFC_{11}^s) by scaling its GLODAP value (assigned to year 1994) with its atmosphere history ($\text{CFC}_{11}^{\text{air}}$), following Khatiwala et al. (2).

$$\text{CFC}_{11}^s(\mathbf{r}_s, t_s) = \lambda(\mathbf{r}_s, t_s) |_{\text{CFC}_{11}} \text{CFC}_{11}^s(\mathbf{r}_s, 1994) \frac{\text{CFC}_{11}^{\text{air}}(t_s)}{\text{CFC}_{11}^{\text{air}}(1994)}. \quad [1]$$

$\text{CFC}_{11}^{\text{air}}$ is derived from Bullister (3). The GLODAP data has no value in the Arctic; we fill in this void using CFC-11 from CMIP6 historical simulations (4) (Table S1 2nd column), with a linear transition over 50°N – 60°N . The time evolution of the CFC-11 saturation is accounted for by $\lambda|_{\text{CFC}_{11}}$. We derive this quantity by fitting Eq. 1 to CFC-11 from CMIP6 simulations.

Simulated CFC_{11}^s has only a small spread across models in the polar regions (solid lines, Fig. S2 top row); it also agrees well with the GLODAP observation (cross sign) in the Antarctic. Inferred $\lambda|_{\text{CFC}_{11}}$ values are similar among CMIP6 historical simulations (Fig. S2 bottom row); they all imply that CFC-11 has become increasingly saturated at the surface, consistent with the stabilisation of its atmosphere mixing ratio. We average outputs from six CMIP6 models to evaluate Eq. 1; CanESM5 is excluded as it is an outlier of the ensemble. The CFC-12 surface history is constructed in a similar way.

Table S1. Climate model information. The first column shows model names, except for HadISST (last row), which is an observational product. The 2nd column indicates whether a model includes CFC-11 and CFC-12 in its historical simulation. The 3rd column indicates whether a model has a strong or weak response to anthropogenic aerosol (AA) forcing. The 4th and 5th columns show SST trends ($^\circ\text{C}$ per decade) averaged in the Southern Ocean (SO, 65°S – 30°S) and the North Atlantic (NA, 30°N – 65°N), respectively, for the 1950–1980 period. The 6th and 7th columns show the corresponding trends for the 1980–2020 period. The last column indicates whether a model's surface heat flux (hdfs) is used in SI Appendix 2.

Model name	CFC in simulation	response to AA	SO SST trend 1950–1980	NA SST trend 1950–1980	SO SST trend 1980–2020	NA SST trend 1980–2020	hdfs
CNRM-CM6-1	No	weak	0.01	0.17	0.13	0.24	used
CNRM-CM6-1-HR	No	weak	0.05	0.00	0.27	0.18	
GFDL-ESM4	Yes	weak	0.07	-0.17	-0.00	0.29	
GISS-E2-1-G	No	weak	0.05	0.18	0.03	-0.09	
IPSL-CM6A-LR	No	weak	-0.02	-0.06	0.09	0.17	used
MIROC6	No	weak	0.01	-0.06	0.03	0.18	
MPI-ESM1-2-LR	No	weak	0.04	-0.01	0.07	0.07	
MPI-ESM1-2-HR	No	weak	0.04	0.05	0.13	-0.04	
CMCC-CM2-SR5	No	weak	0.02	0.17	0.14	0.31	
CanESM5	Yes	strong	0.11	0.07	0.17	0.37	
ACCESS-CM2	No	strong	-0.04	-0.09	0.07	0.22	
CESM2	Yes	strong	-0.01	-0.00	0.15	0.27	
EC-Earth3-CC	Yes	strong	0.00	-0.16	0.08	0.41	
GFDL-CM4	Yes	strong	-0.06	0.03	0.12	0.26	
HadGEM3-GC31-LL	No	strong	-0.01	0.09	0.28	0.32	
MRI-ESM2-0	Yes	strong	0.00	-0.09	0.10	0.12	
NorESM2-LM	Yes	strong	-0.06	-0.17	0.05	0.27	
HadISST	N/A	N/A	0.08	-0.21	0.03	0.22	

B. Climatological Tracers. The climatological temperature ($\bar{\Theta}$) and salinity ($\bar{S}$) observations are derived from the 25-year (1993–2017) average of the ECCO ocean state estimate (version 4 release 4) (5). Synthetic surface histories of $\bar{\Theta}$ and $\bar{S}$ are constructed by repeating their climatological surface values in time indefinitely. We truncate the time integral at 8000 years when applying the GF convolution to $\bar{\Theta}$ and $\bar{S}$; extending it further makes no difference to the integral because the entire ocean volume is ventilated (comes into contact with the atmosphere) in less than 8000 years. The observation time of $\bar{\Theta}$ and $\bar{S}$ can be arbitrary value. Note that $\bar{\Theta}$ and $\bar{S}$ only constrain the spatial distribution of the G function at a given $\mathbf{r}$ (i.e. the water-mass distribution at $\mathbf{r}$), because they do not contain information on transit time from the surface to depths.

C. Deep Ocean Constraints. There are few observational constraints on G in the deep ocean (below 2000 m), because CFC-11 and CFC-12 in the GLODAP data have very low concentrations below 1000 m compared with their surface values (see Fig. 9 in Key et al. (1)). For this reason, we have not attempted to derive observational GFs for the deep ocean in this study. Natural

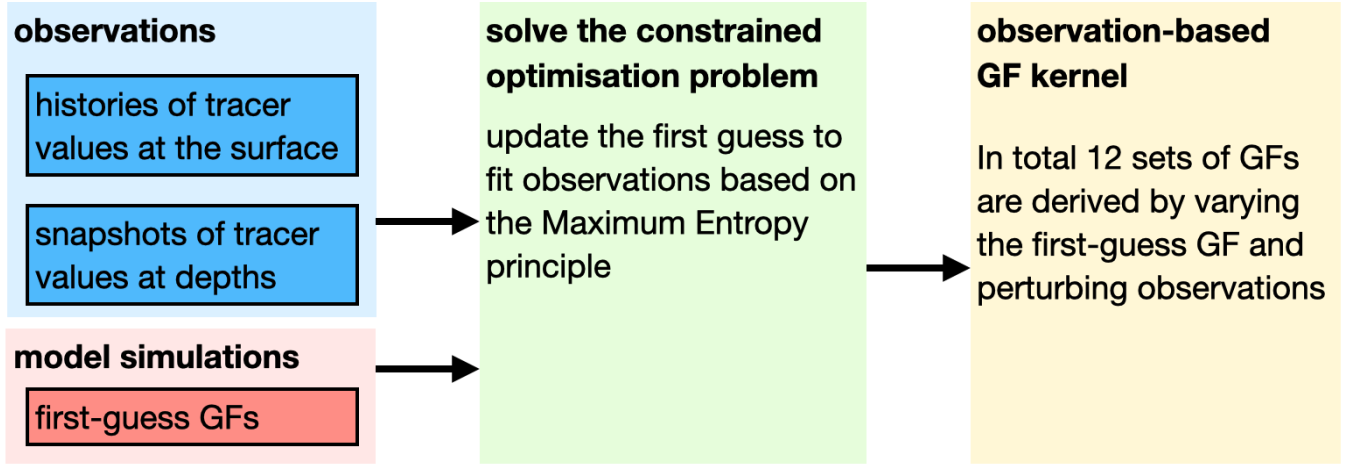


Fig. S1. The procedure of inferring Green's functions (GF) from ocean tracer observations.

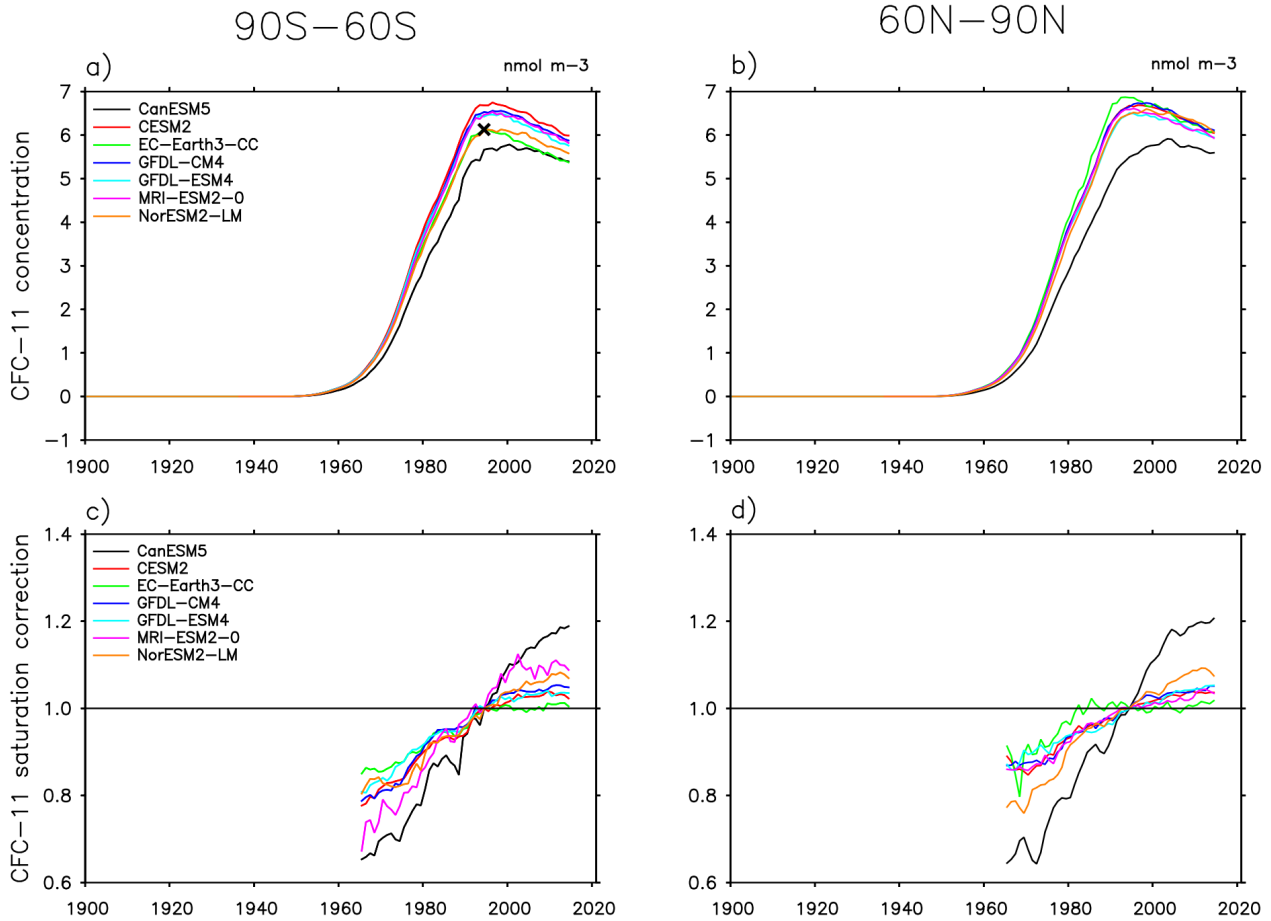


Fig. S2. CFC-11 concentration at the ocean surface (top row) and its saturation coefficient (i.e. $\lambda_{\text{CFC-11}}$ in Eq. 1) (bottom row). Both quantities are derived from CMIP6 historical simulations. They are zonally averaged first, and then further averaged over 90°S–60°S (left column) and 60°N–90°N (right column). In (a) the cross sign indicates the corresponding value from the GLODAP data.

radiocarbon, with its half-life of 5700 years, is useful to constrain G in the deep ocean on multi-century timescales (2, 6). We have not used natural radiocarbon because its half-life is much longer than the 150-year history of ocean warming studied here. Oxygen and phosphate would also help to constrain G in the deep ocean (6) but, like $\bar{\Theta}$ and $\bar{S}$, they do not contain information on transit time, which is critical for estimating transient ocean warming.

2. Ocean Heat Uptake Boundary Conditions

The procedure for deriving OHU boundary conditions is summarised in Fig. S3. The boundary conditions contain two components: 1) the global mean of SST anomalies (SSTa) from observations with respect to 1870–1880 and 2) spatial anomalies of simulated surface excess temperature (SSTe) with respect to the global mean. SSTa and SSTe are different because SSTa is affected by temperature redistribution, while SSTe is not.

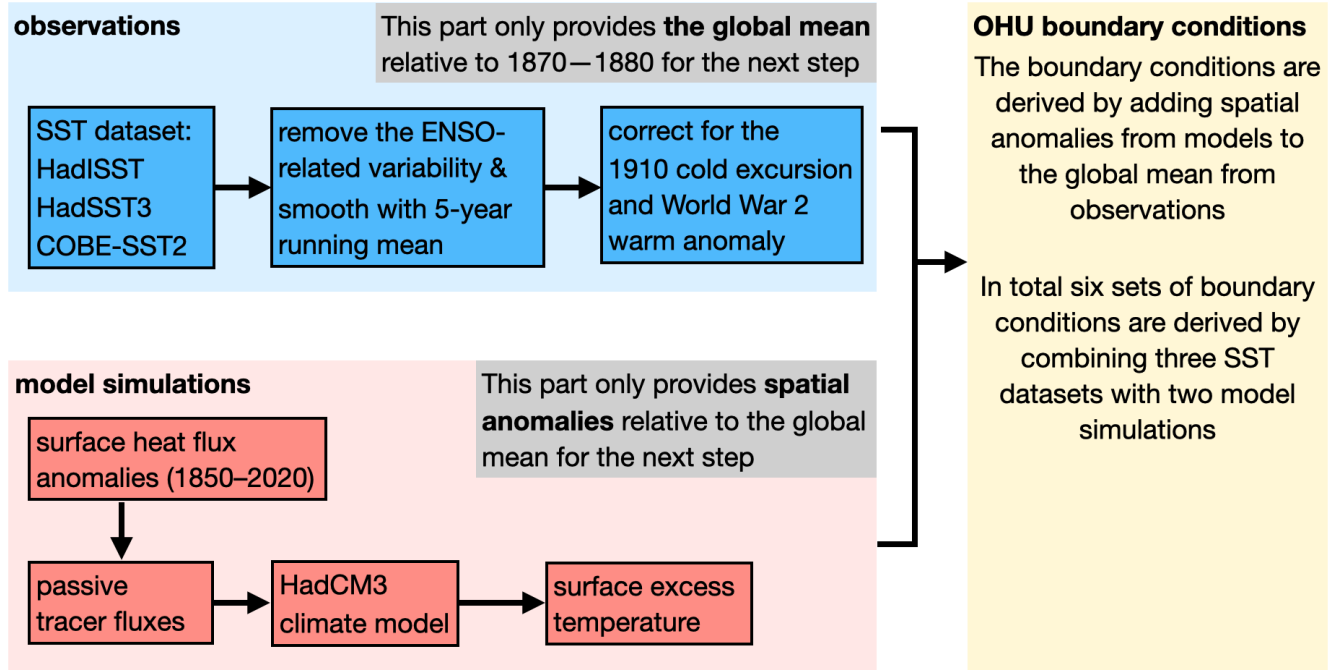


Fig. S3. The procedure for deriving OHU boundary conditions.

A. Global Mean. Three SST datasets are used to estimate the global mean of the OHU boundary conditions: a) HadISST (7), b) HadSST3 (8) and c) COBE-SST2 (9). We exclude the HadSST4 dataset because it suggests a much stronger global warming than the others. We exclude the ERSST5 dataset because it is much colder than the others during 1920–1960.

We apply two adjustments to the global mean of SSTa. First, we reduce the interannual variability in it, which is largely an imprint of vertical heat redistribution due to the El Niño–Southern Oscillation (10). To remove this redistribution signal, we remove the part of the global mean that is linearly related to the Niño-3.4 index. A 5-year running mean is subsequently applied to further smooth the time series. Second, we correct SSTa for potential biases in 1900–1920 (a cold excursion) and during World War 2 (a warm anomaly); the detail of this is described in the main text.

B. Spatial Anomalies. We simulate SSTe as a passive tracer in the HadCM3 climate model (11) following the governing equation of excess temperature Θ_e (recall that SSTe is Θ_e at the surface). The integration is conducted in a pre-industrial simulation, because the evolution of Θ_e is primarily controlled by the climatological ocean transport (12). The surface source of Θ_e is the surface heat flux anomaly relative to the climatology (denoted as Q_a). We derive Q_a from GFDL-ESM4 and MIROC6’s historical simulation (1850–2014) and the SSP2–4.5 (medium emission) extension (2015–2020). A comparison between SSTe and SSTa is shown in Fig. S4 (compare the top and bottom row). Note that SSTe shows stronger warming than SSTa in the Southern Ocean during 1980–2020; this is a typical feature in climate models under CO₂ forcing (13).

C. Model Selection. GFDL-ESM4 and MIROC6 are selected from 17 CMIP6 climate models (Table S1) to provide Q_a . The selection procedure is as follows. First, we exclude models that have a ‘strong’ response to anthropogenic aerosol forcing (Table S1, 3rd column), because they are likely inconsistent with observations (14). Next, we exclude models that have unrealistic SSTa trends in the North Atlantic and the Southern Ocean by a comparison with HadISST (observations) (Table S1, 4th–7th columns).

We focus on these two regions because they dominate the ocean’s uptake of heat and carbon (2, 15). The comparison is done for two periods: 1950–1980, when the aerosol forcing dominates, and 1980–2020, when the greenhouse gas forcing dominates (16). This process identifies GFDL-ESM4 and MIROC6 as the two best models in reproducing observed SSTa trends; their average matches observations well for SSTa patterns at high latitudes (Fig. S4, compare (b) and (c) with (e) and (f)). Prior to 1950, GFDL-ESM4 and MIROC6 do not agree well with observations for SSTa patterns (Fig. S4, compare (a) with (d)). It is unclear whether this discrepancy arises from errors in models or early observations.

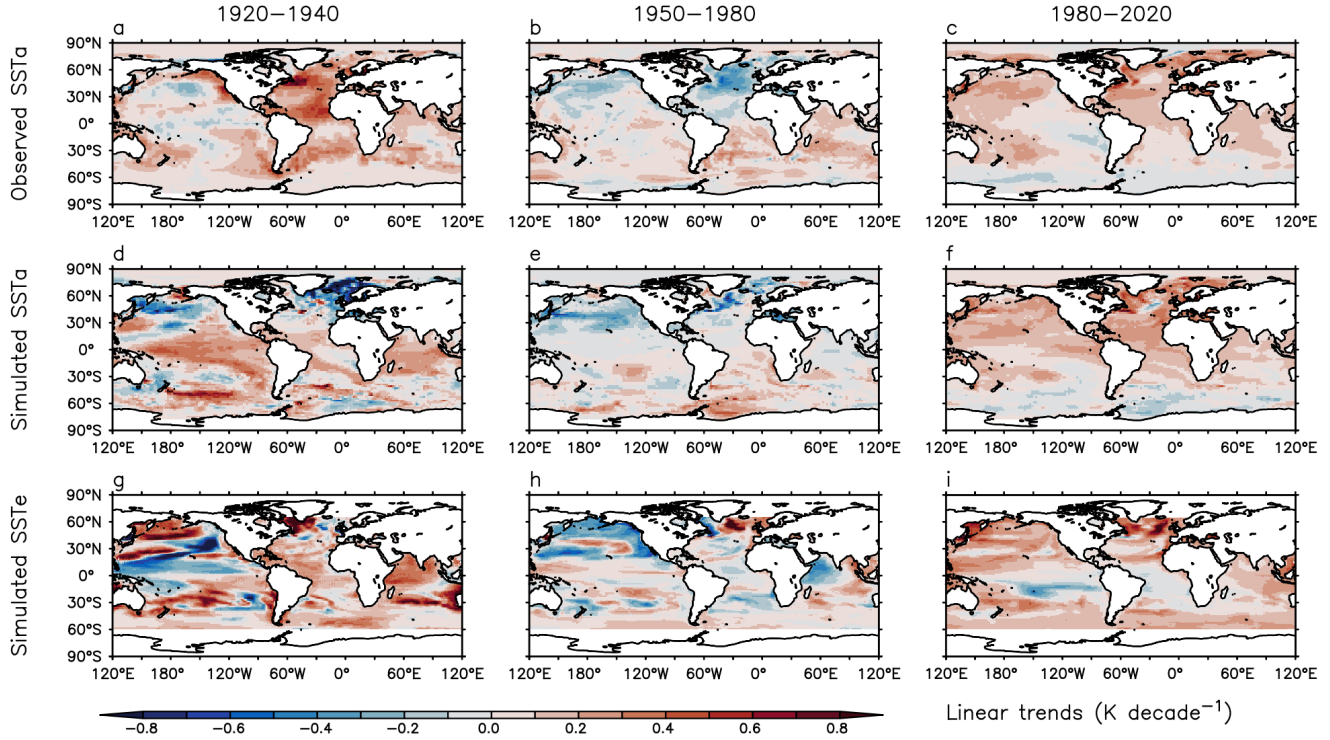


Fig. S4. SSTa and SSTE linear trends over 1920–1940 (1st column), 1950–1980 (2nd column), and 1980–2020 (3rd column). (a)–(c) are derived from HadISST. (d)–(f) are derived from GFDL-ESM4 and MIROC6 (shown as their average). (g)–(i) are the boundary condition for estimating OHU, i.e. global mean of SSTa + spatial anomalies of SSTE.

3. Propagation of Error

The uncertainty of the sum of two variables x and y is computed as

$$\text{var}(x + y) = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}, \quad [2]$$

where σ_x and σ_y are the standard deviation of x and y , respectively, and σ_{xy} is their covariance. The uncertainty of the product of x and y is approximated as

$$\text{var}(xy) \approx \mu_x^2 \mu_y^2 \left(\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} + \frac{2\sigma_{xy}}{\mu_x \mu_y} \right), \quad [3]$$

where μ_x and μ_y are the expectation values of x and y , respectively. Similarly, the uncertainty of the ratio between y and x is approximated as

$$\text{var}\left(\frac{y}{x}\right) \approx \frac{\mu_y^2}{\mu_x^2} \left(\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x \mu_y} \right). \quad [4]$$

We consider the correlations among the radiative forcing F , the radiative response R and the energy imbalance N when computing the uncertainty of functions based on them. We estimate their correlations across the seven models that have the F , R and N diagnostics in the **historical** experiment; the F , R and N are all averaged for 2000–2014 to highlight the forced response. The resulting correlations are $\text{corr}(F, N) = 0.57$, $\text{corr}(F, R) = -0.71$ and $\text{corr}(N, R) = 0.18$.

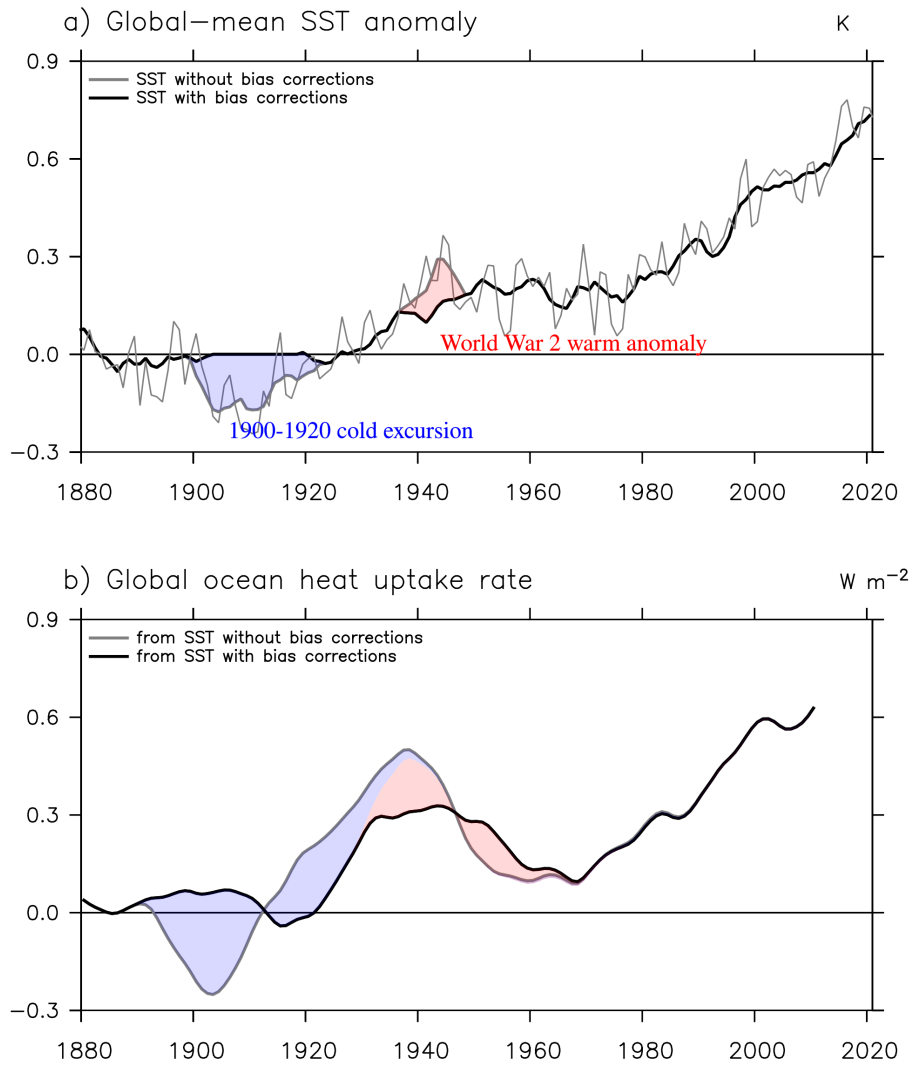


Fig. S5. Potential biases in the global-mean SST record and their impact on the global OHU estimate from the Green's function approach. a) Global-mean SST anomaly from three SST datasets: HadISST, HadSST3 and COBE-SST2, shown as the multi-data mean. The black and grey line show the SST anomaly with and without the bias corrections, respectively. b) Global OHU rate inferred from the two SST timeseries in (a) with corresponding line style. In both panels, changes due to the SST bias corrections are shown as shading. In (a), the thin grey line shows the annual SST anomaly; the thick grey line is derived by removing the variability associated with the Niño-3.4 index in the thin grey line, followed by a 5-year running mean.

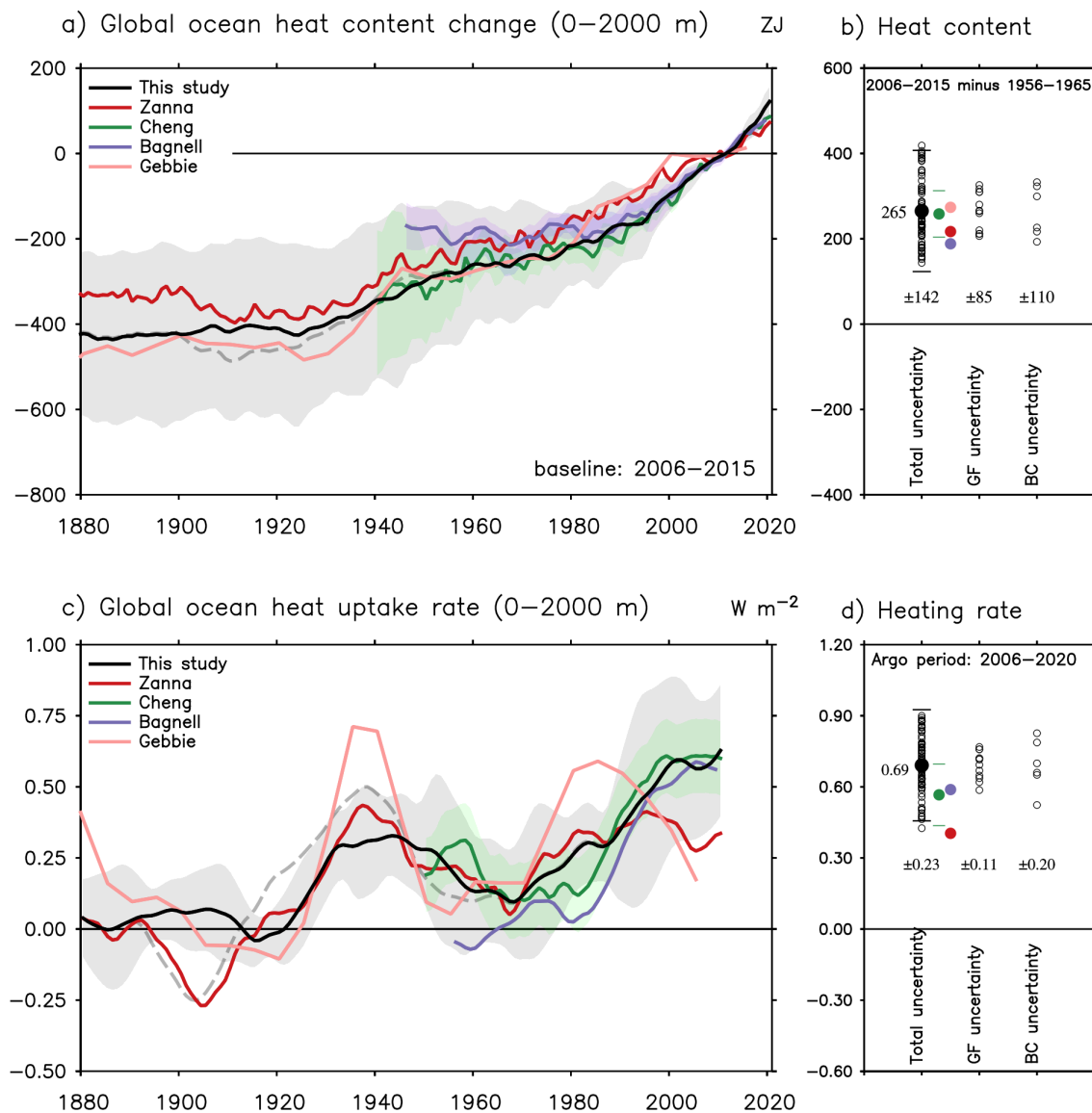


Fig. S6. Global ocean heat uptake during the historical period (0–2000 m). Different estimates are color coded. “This study”, “Zanna” and “Gebbie” are based on the Green’s function (GF) method; the other two are in-situ estimates. a) time evolution of ocean heat content change relative to the 2006–2015 baseline (1 ZJ = 10^{21} J). b) ocean heat content change between 2006–2015 and 1956–1965. c) time evolution of ocean heat uptake rate per unit area of Earth’s surface. d) ocean heat uptake rate during the Argo period (2006–2020). In (c), the rate of change is computed as linear trends of a 20-year running window. In (b) and (d), the spread of our GF estimate is decomposed into that due to the GF kernel G and that due to the boundary condition (BC) Θ_o^* ; individual members are shown as circles. Shading and error bars indicate the 2σ -error. In (a) and (c), the dashed black line is the same as the solid black line, except that it is computed from SST datasets without bias corrections.

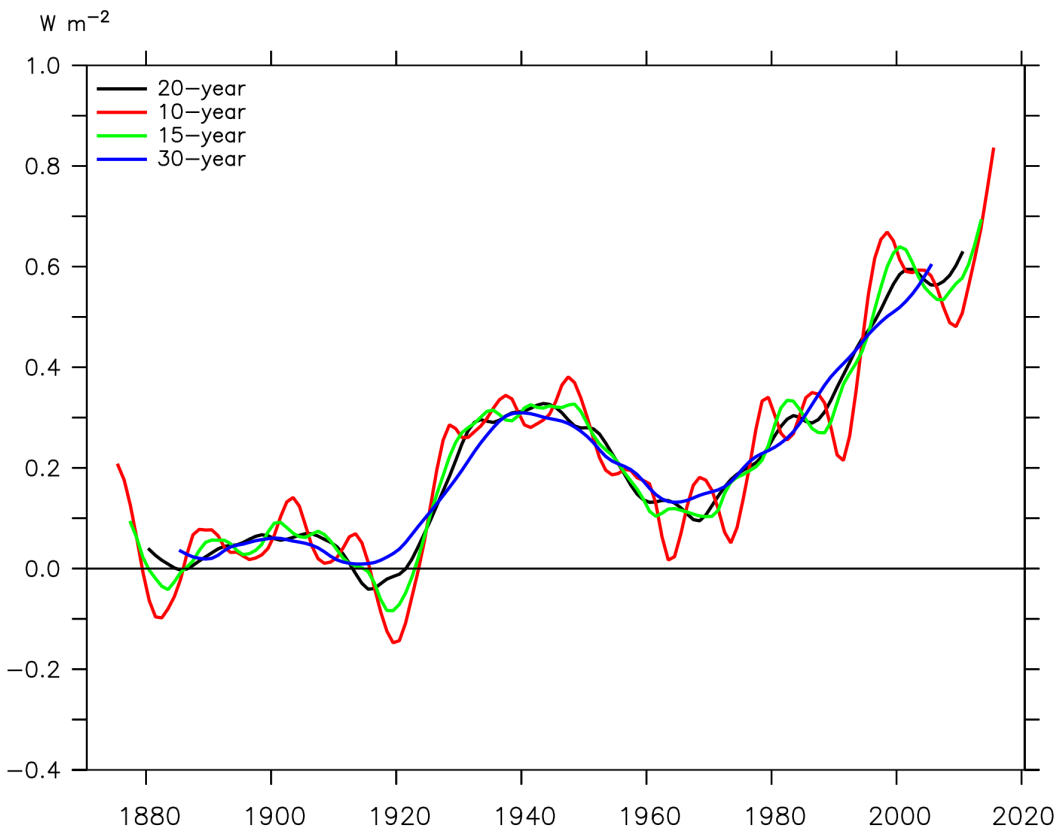


Fig. S7. Global OHU rate computed as linear trends of OHC change over a moving window. Results from different choices of the span of the moving window is shown in different colors; the black line is the same as the black line in Fig. 1c. The OHC data is derived from the ensemble-mean of the GF OHC estimate of this study, i.e. Fig. 1a black line.

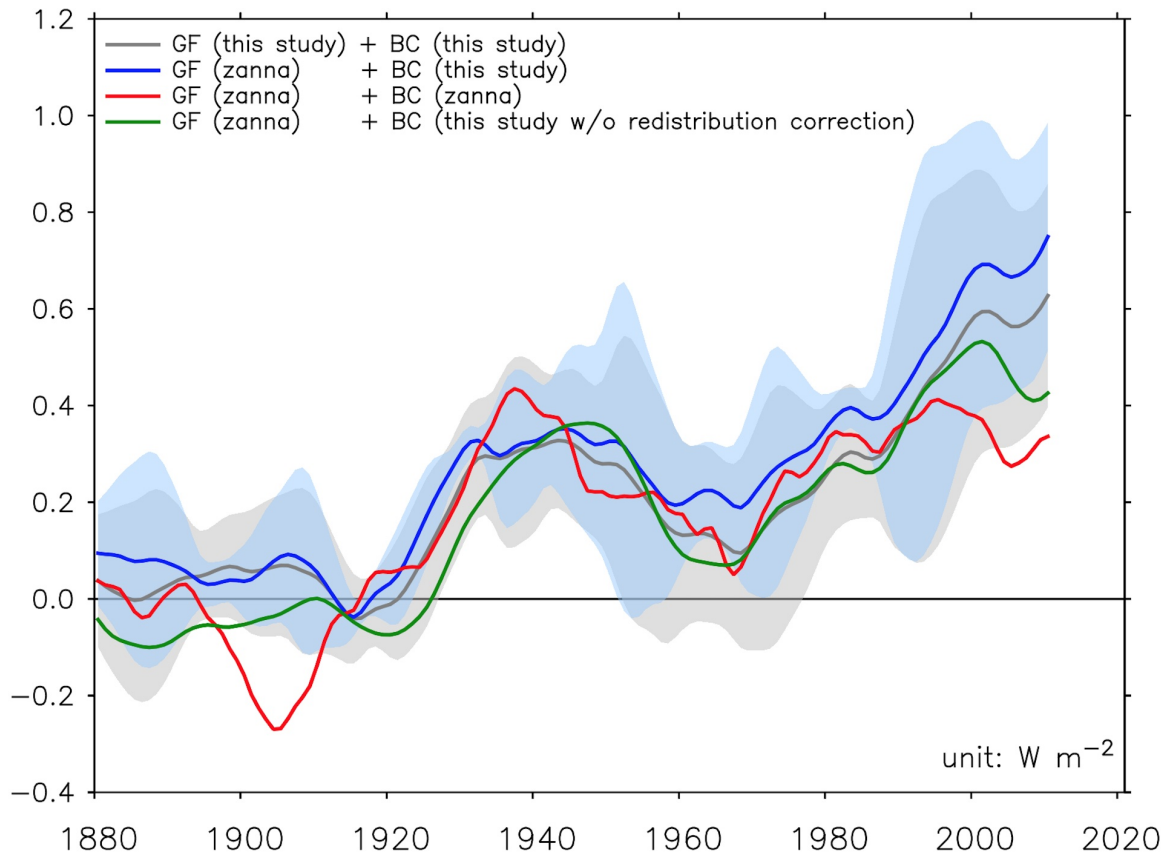


Fig. S8. Comparing the influences of the GF and the boundary condition (BC) in driving the difference between this study and Zanna et al. (17) for the OHU rate. The results of this study and Zanna et al. are shown as the gray and red line, respectively. Combining the GF of Zanna et al. and the BC of this study gives the blue estimate; removing the redistribution correction from the BC results in the green estimate.

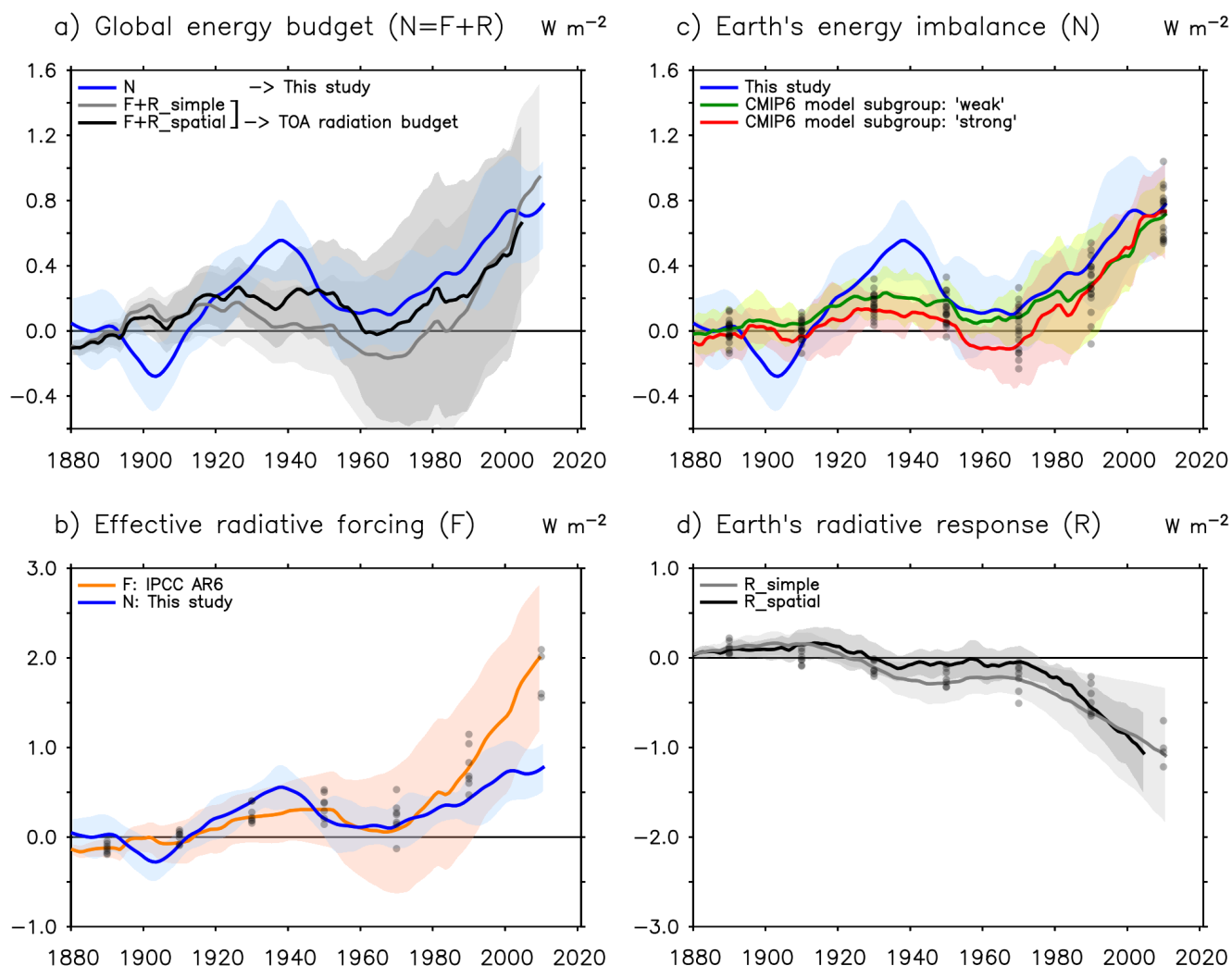


Fig. S9. Evaluating the global energy budget since 1880 using observation-based data. The three components examined here are: Earth's energy imbalance N , the radiative forcing F , and Earth's radiative response R . Everything is same as Figure 2 in the main text, except that the N estimate shown here is computed from SST datasets without bias corrections. In all panels, shading indicates the 2σ -error. F and R are both smoothed by a 20-year running mean. N is computed as linear trends of a 20-year running window. The N estimate of this study (without SST bias corrections) is shown as the blue line in (a-c), the same in each panel. In (b-d), simulations from 17 climate models are shown as pale dots, plotted every 20 years for clarity. Note that only seven models provide the F and R diagnostics and only four of them provide F and R up to 2020. In (b), the models are split into those with a "weak" and "strong" response to anthropogenic aerosol forcing, respectively.

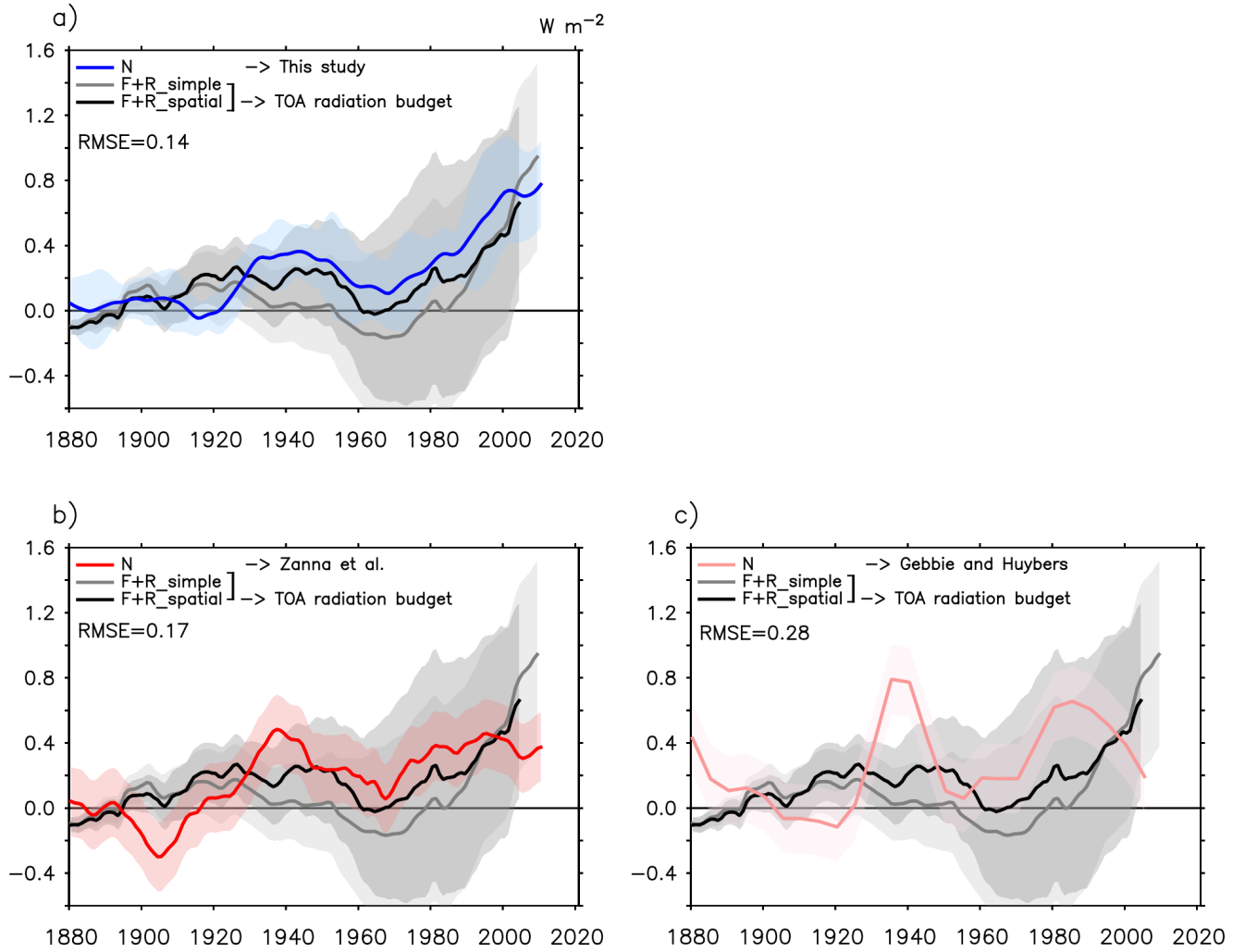


Fig. S10. Comparing the global energy budget based on the N estimate of: a) this study, b) Zanna et al. (17) and c) Gebbie and Huybers (18). (a) is the same as that in Fig. 2; $F + R_{\text{simple}}$ and $F + R_{\text{spatial}}$ in (a) are repeated in (b) and (c) for comparison. The RMSE between the best estimate of N and $F + R_{\text{spatial}}$ is listed in every panel. Shading indicates the 2σ -error. Because the Zanna estimate and the Gebbie estimate do not provide the 2σ -error, we use the time-mean error in (a) for (b) and (c).

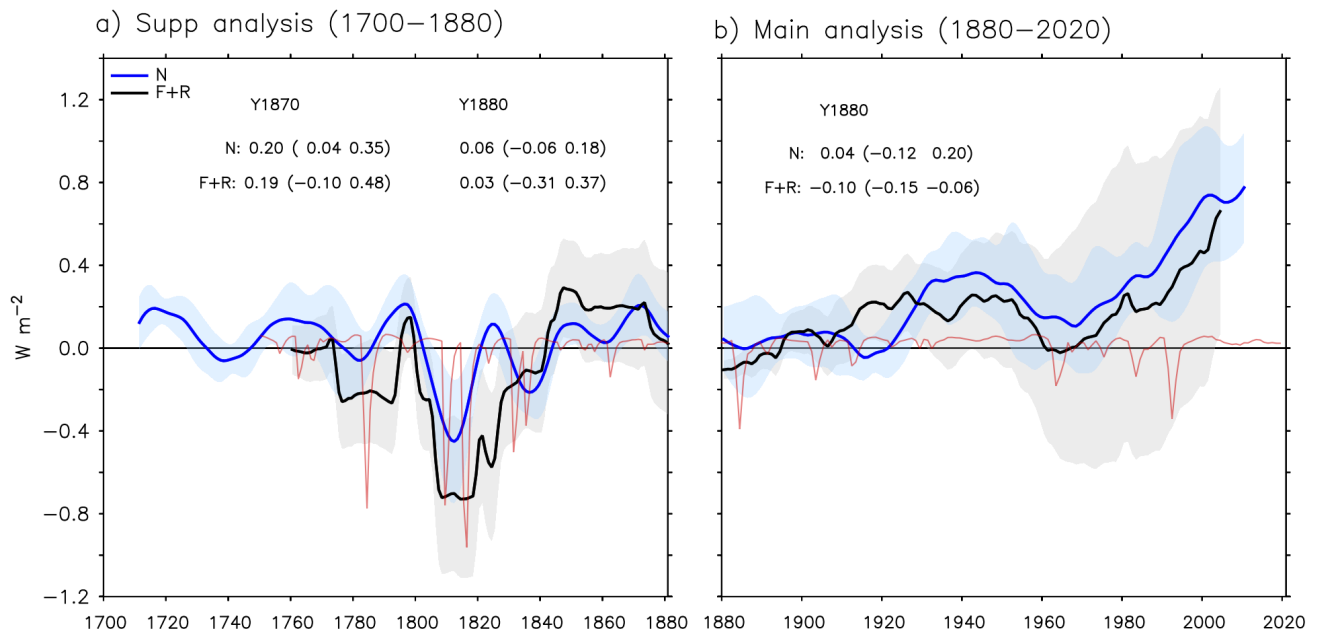


Fig. S11. Evaluating Earth's energy budget for (a) 1700–1880 and (b) 1880–2020. The two analyses are computed separately using different datasets and reference periods. Earth's energy imbalance N is shown in blue and the sum of radiative forcing F and radiative response R in black. The blue and black lines in (b) are the same as those in Figure 2a. Shading indicates the 2σ -error. The values of N and $F + R$ are listed in (a) for years 1870 and 1880 and in (b) for 1880. The AR6 natural forcing estimate is shown by the red line, which indicates major volcanic eruptions. Its magnitude is scaled down by a factor of 5 for better illustration.

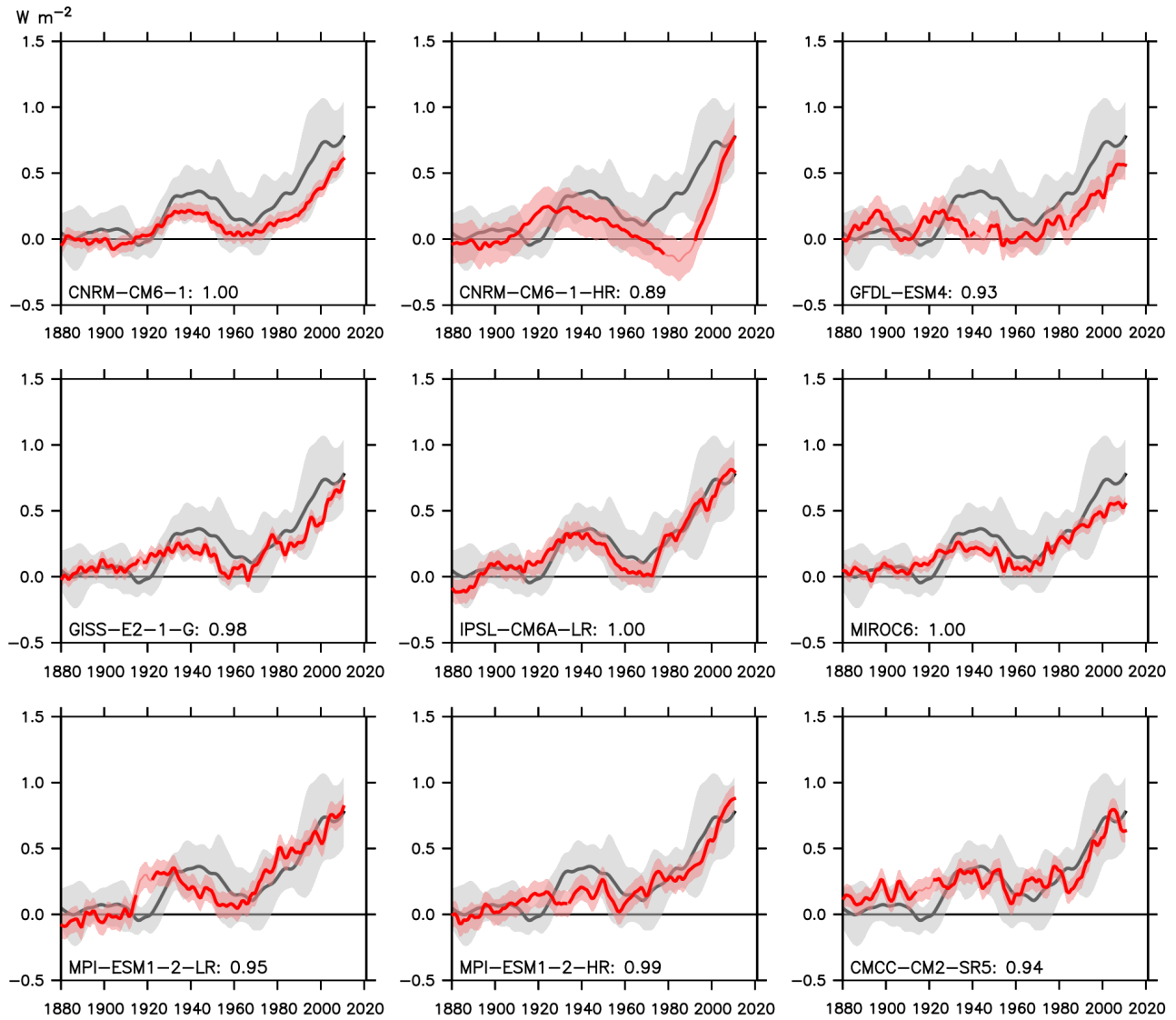


Fig. S12. Earth's energy imbalance N in nine CMIP6 models which have a 'weak' response to anthropogenic aerosol forcing. The model and observation-based result are shown as the red and black line, respectively. The observation is the same as the blue line in Figure 2b of the main text. The thick red line indicates that the model result is consistent with the observation within the 2σ -error (shading). The numbers denote the fraction of time the model result is within the observed range in 1880–2010.

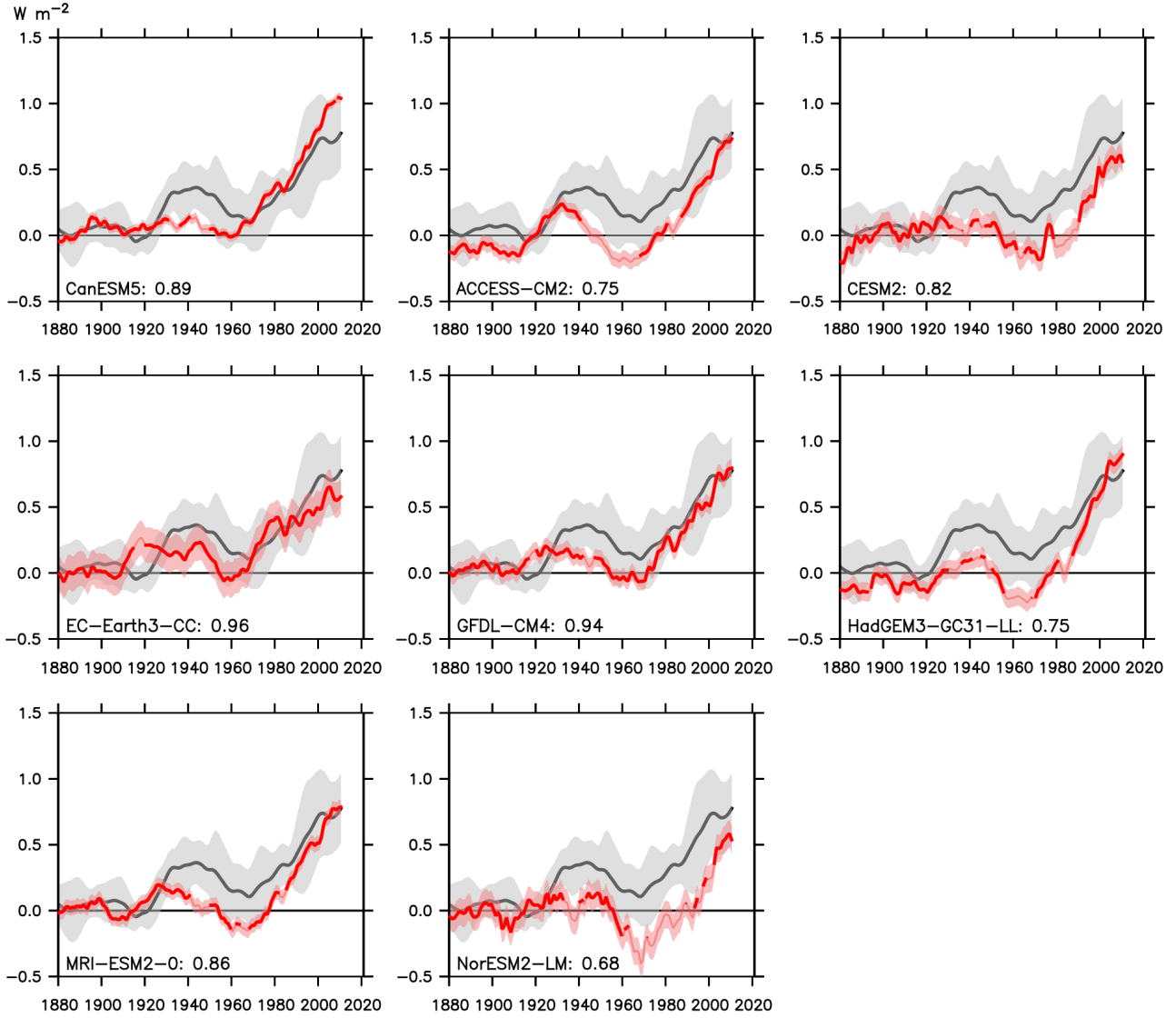


Fig. S13. Earth's energy imbalance N in eight CMIP6 models which have a 'strong' response to anthropogenic aerosol forcing. The model and observation-based result are shown as the red and black line, respectively. The observation is the same as the blue line in Figure 2b of the main text. The thick red line indicates that the model result is consistent with the observation within the 2σ -error (shading). The numbers denote the fraction of time the model result is within the observed range in 1880–2010.

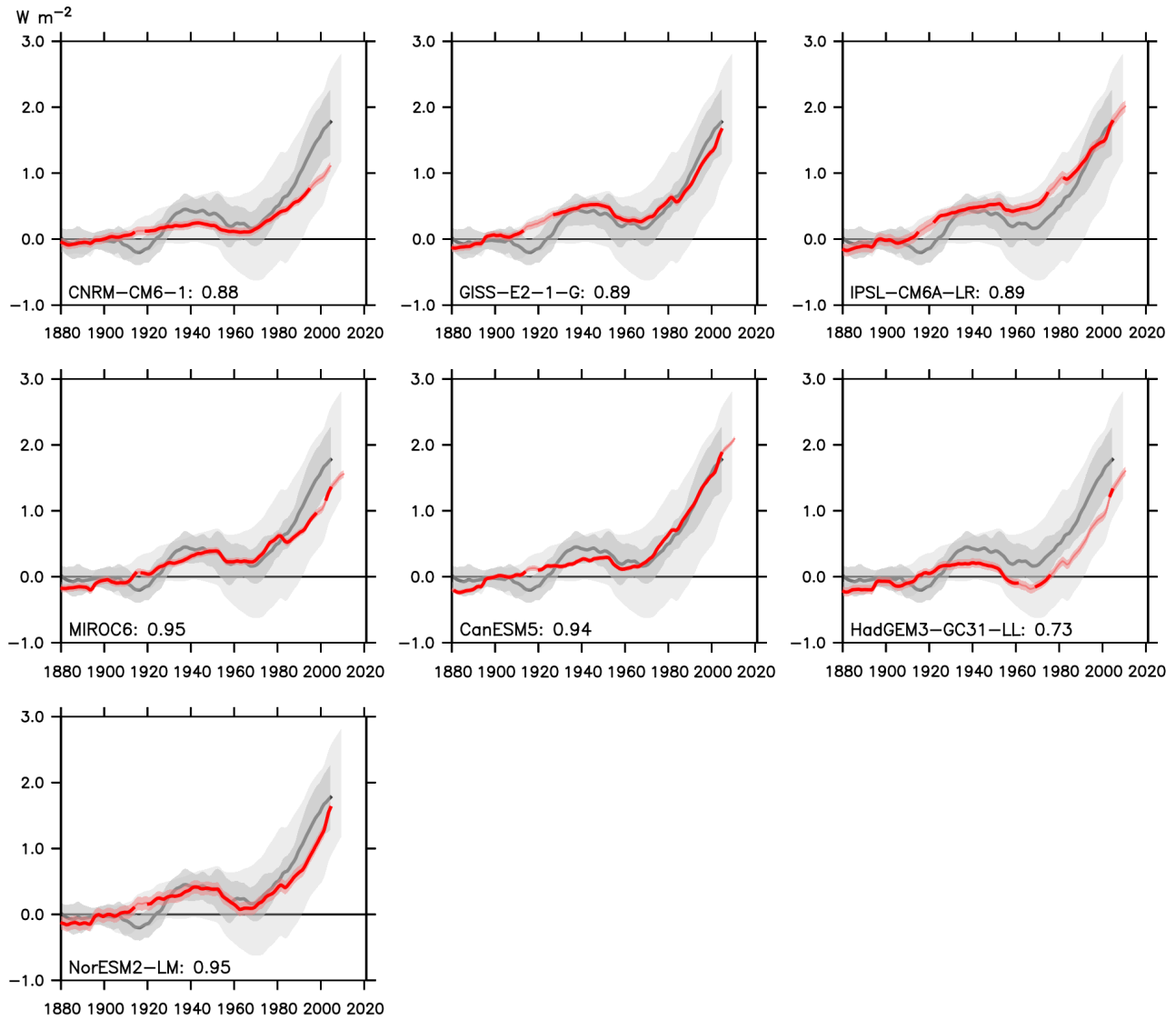


Fig. S14. Effective radiative forcing F in seven CMIP6 models. The model and observation-based result are shown as the red and black line, respectively. The observation is the inferred $F = N - R_{\text{spatial}}$ in Table 2 of the main text. The thick red line indicates that the model result is consistent with the observation within the 2σ -error (shading). The numbers denote the fraction of time the model result is within the observed range in 1880–2004. The light shading denotes the F range in IPCC AR6.

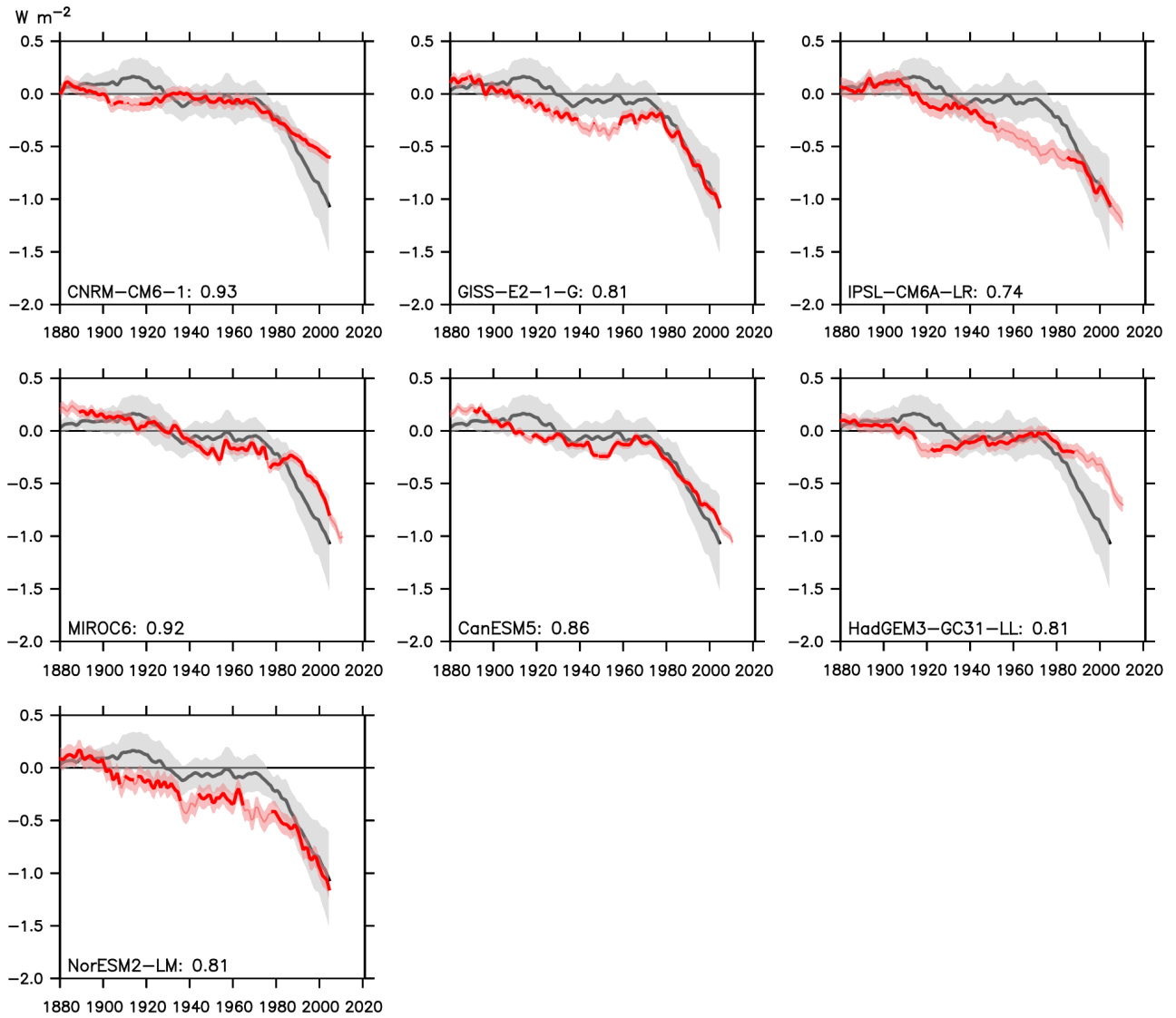


Fig. S15. Earth's radiative response R in seven CMIP6 models. The model and observation-based result are shown as the red and black line, respectively. The observation is the same as the black line (R_{spatial}) in Figure 2d of the main text. The thick red line indicates that the model result is consistent with the observation within the 2σ -error (shading). The numbers denote the fraction of time the model result is within the observed range in 1880–2004.

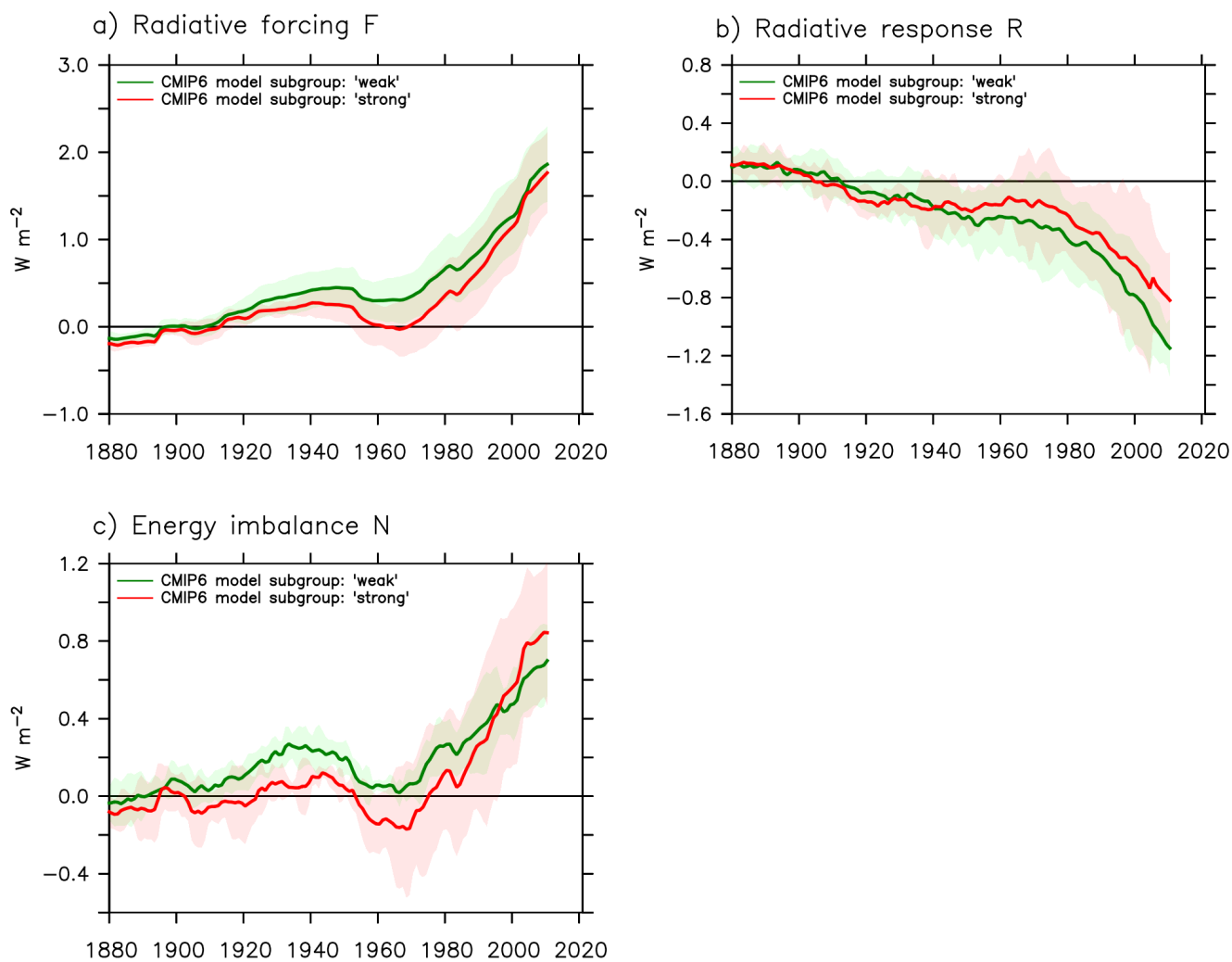


Fig. S16. Radiative forcing F , radiative response R and Earth's energy imbalance N from seven CMIP6 models. Results from the four “weak” models and the three “strong” models are shown separately. Solid lines denote ensemble means and shading denotes the 2σ -spread. Fig. 2c and panel (c) are not identical because the latter only uses a subset of models in the former.

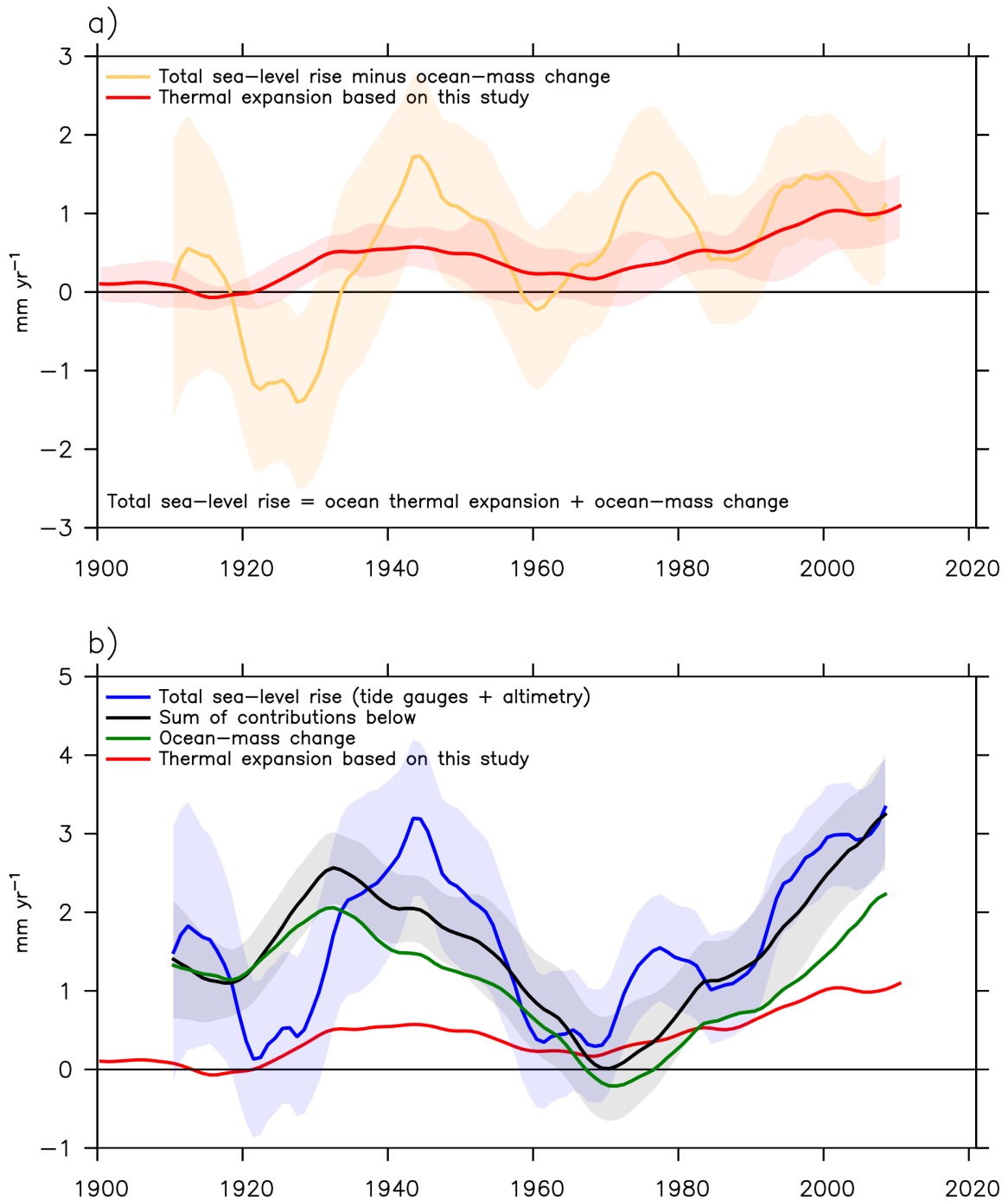


Fig. S17. Global-mean sea-level budget. The total sea-level rise is decomposed into contributions from: 1) ocean thermal expansion 2) ocean-mass change. The total sea-level rise and ocean-mass change are all derived from observation-based reconstructions in (19). The thermal expansion is computed from the ocean warming estimate of this study. The rate of change is computed as linear trends of a 20-year moving window.

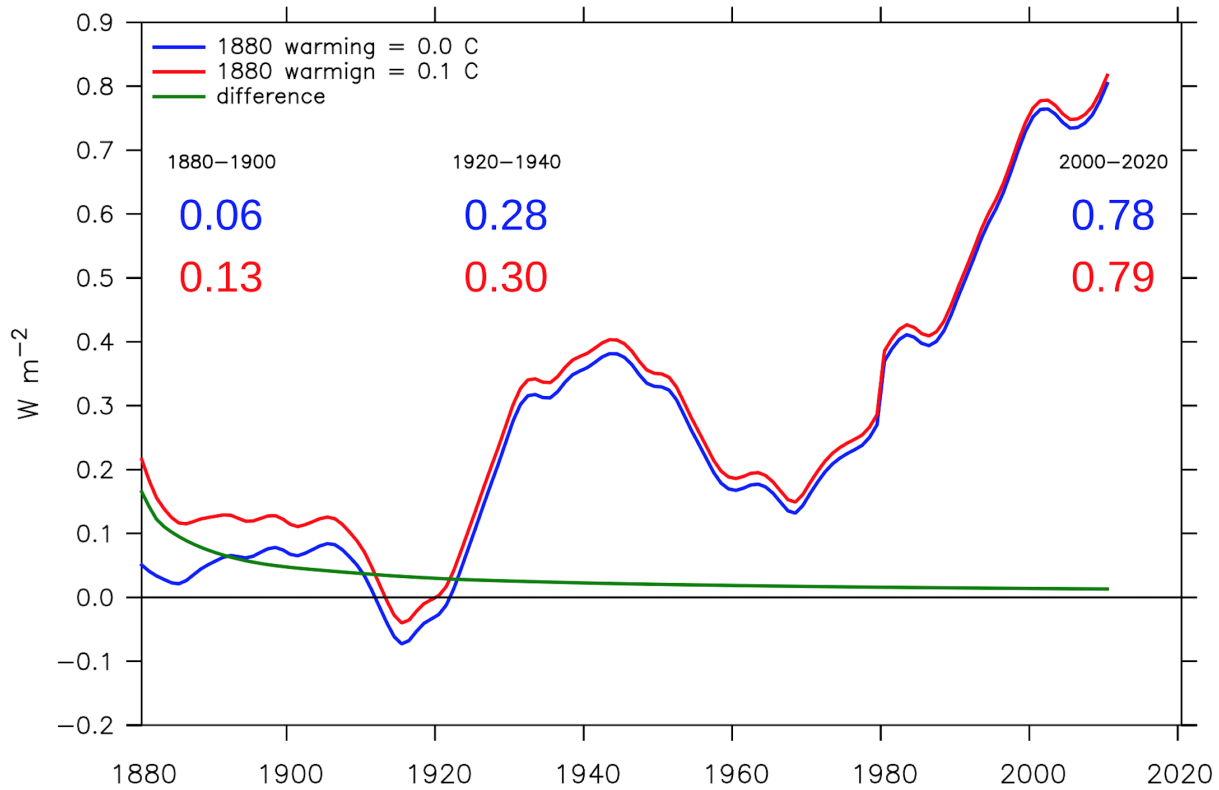


Fig. S18. Estimates of Earth's energy imbalance N based on two assumptions of global warming in year 1880. The blue estimate assumes that 1870–1880 is near equilibrium, i.e. no global warming. The red estimate assumes 0.1 C uniform warming in 1880. Their difference is shown as the green line. The blue line is the same as the blue line in Figure 2a of the manuscript. Time means of the blue and red estimates are listed with matching colours for 1880–1900, 1920–1940 and 2000–2020.

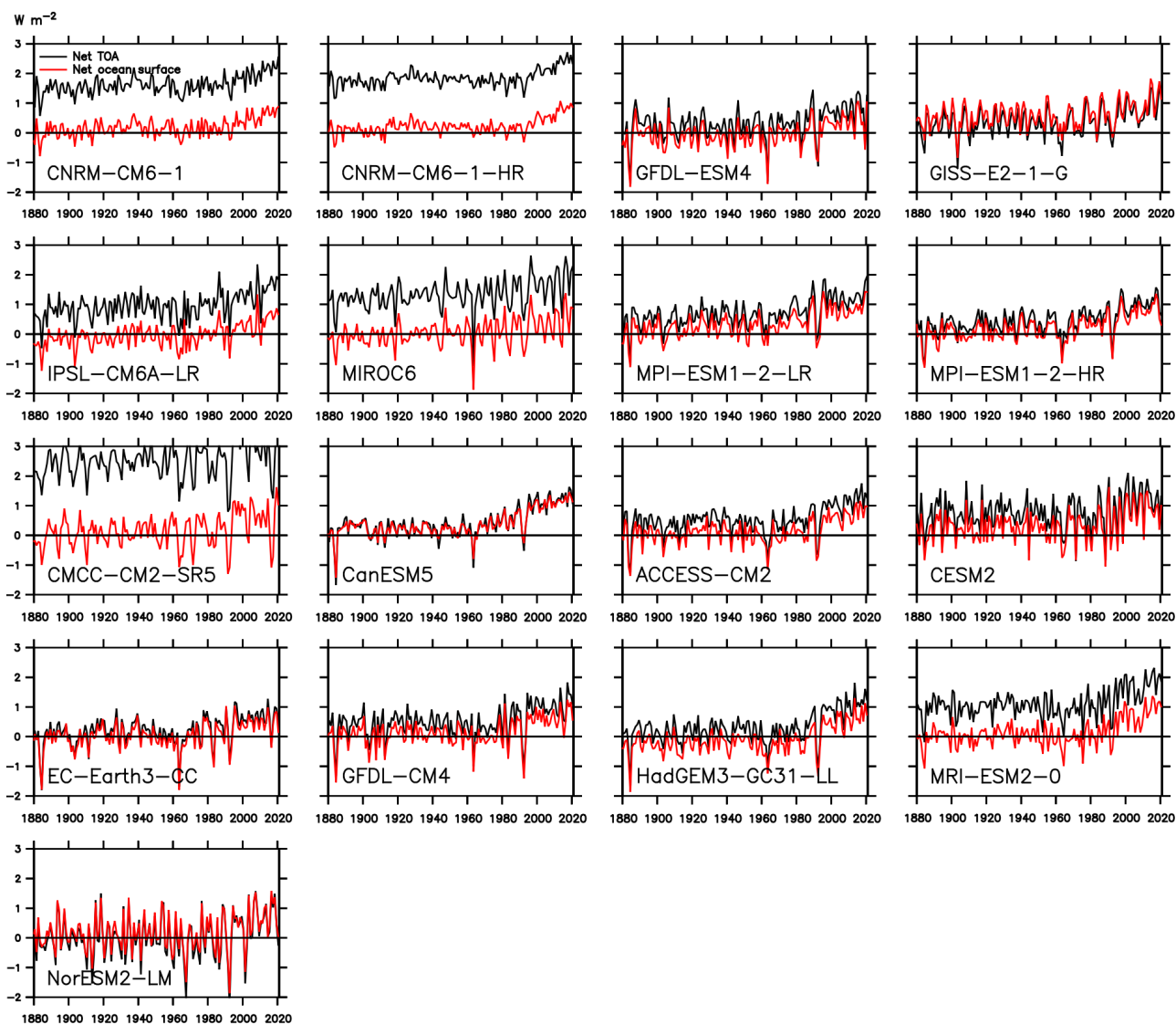


Fig. S19. Net top-of-atmosphere radiative flux (black) and net ocean surface heat flux (red) in the historical experiment. Both quantities are averaged over Earth's surface area and shown for individual climate models.

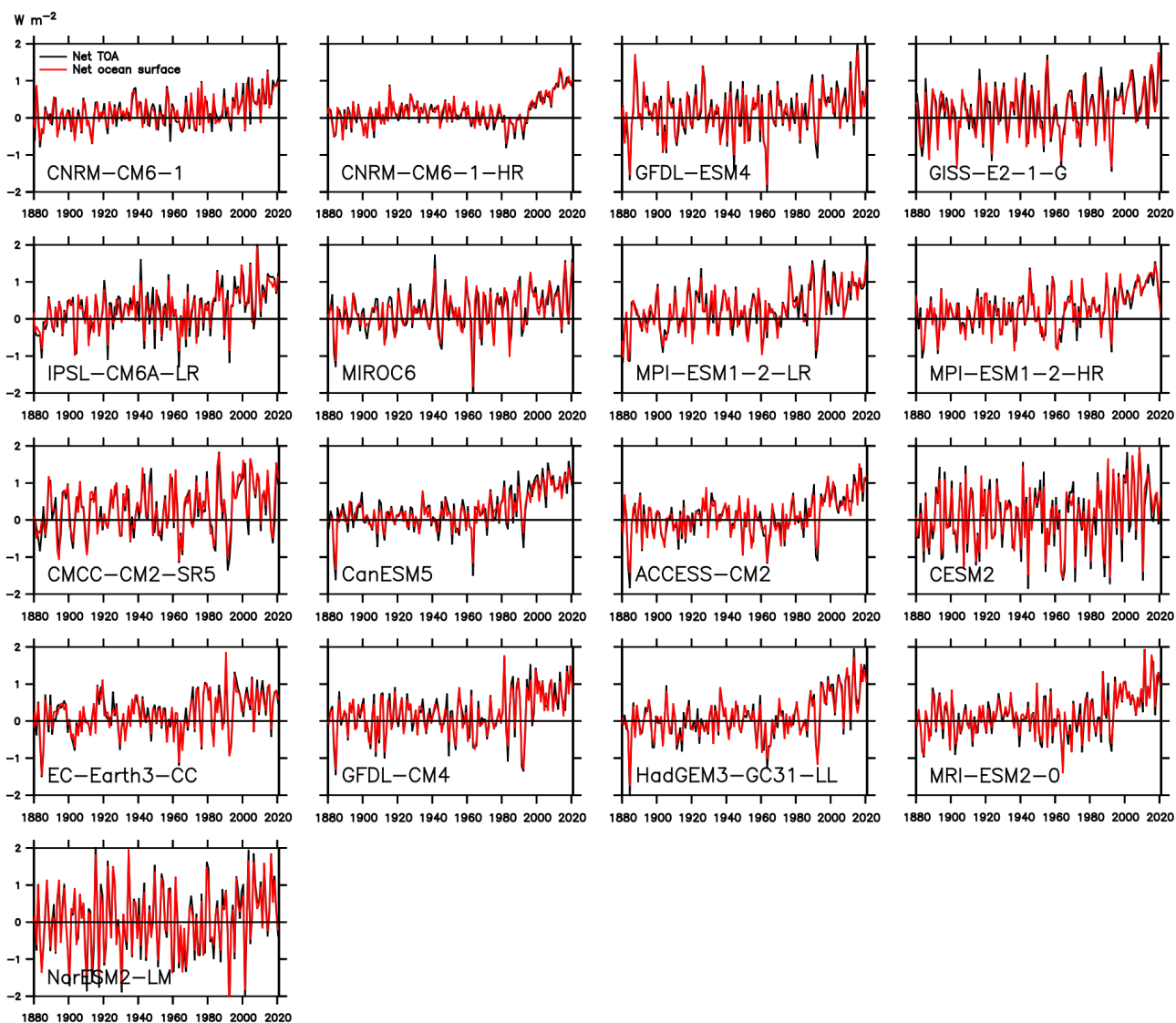


Fig. S20. Net top-of-atmosphere radiative flux (black) and net ocean surface heat flux (red) in the historical relative to the piControl experiment. Both quantities are averaged over Earth's surface area and shown for individual climate models.

References

1. RM Key, et al., A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031 (2004).
2. S Khatiwala, F Primeau, T Hall, Reconstruction of the history of anthropogenic CO₂ concentrations in the ocean. *Nature* **462**, 346–349 (2009).
3. JL Bullister, Atmospheric Histories (1765–2015) for CFC-11, CFC-12, CFC-113, CCl₄, SF₆ and N₂O (2015).
4. V Eyring, et al., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model. Dev.* **9**, 1937–1958 (2016).
5. G Forget, et al., ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model. Dev.* **8**, 3071–3104 (2015).
6. S Khatiwala, F Primeau, M Holzer, Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age. *Earth Planet. Sci. Lett.* **325–326**, 116–125 (2012).
7. NA Rayner, et al., Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
8. JJ Kennedy, NA Rayner, RO Smith, DE Parker, M Saunby, Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.* **116**, D14103 (2011).
9. S Hirahara, M Ishii, Y Fukuda, Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **27**, 57–75 (2014).
10. D Roemmich, J Gilson, The global ocean imprint of ENSO. *Geophys. Res. Lett.* **38** (2011).
11. C Gordon, et al., The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16**, 147–168 (2000).
12. Q Wu, JM Gregory, Estimating ocean heat uptake using boundary Green’s functions: a perfect-model test of the method. *J. Adv. Model. Earth Syst.* **14** (2022).
13. JM Gregory, et al., The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) contribution to CMIP6: investigation of sea-level and ocean climate change in response to CO₂ forcing. *Geosci. Model. Dev.* **9**, 3993–4017 (2016).
14. J Robson, et al., The role of anthropogenic aerosol forcing in the 1850–1985 strengthening of the AMOC in CMIP6 historical simulations. *J. Clim.* **35**, 3243–3263 (2022).
15. T Kuhlbrodt, JM Gregory, Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophys. Res. Lett.* **39**, 2012GL052952 (2012).
16. AJ Dittus, et al., Sensitivity of historical climate simulations to uncertain aerosol forcing. *Geophys. Res. Lett.* **47** (2020).
17. L Zanna, S Khatiwala, JM Gregory, J Ison, P Heimbach, Global reconstruction of historical ocean heat storage and transport. *Proc. Natl. Acad. Sci.* **116**, 1126–1131 (2019).
18. G Gebbie, P Huybers, The Little Ice Age and 20th-century deep Pacific cooling. *Science* **363**, 70–74 (2019).
19. T Frederikse, et al., The causes of sea-level rise since 1900. *Nature* **584**, 393–397 (2020).