Modeling 2020 regulatory changes in international shipping emissions helps explain 2023 anomalous warming

Ilaria Quaglia[1] and Daniele Visioni[2]

[1]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY
[2]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY

Correspondence: Daniele Visioni (dv224@cornell.edu)

Abstract. The summer of 2023 has seen an anomalous increase in temperatures even when considering the ongoing greenhouse-gases driven warming trend. Here we demonstrate that regulatory changes to sulfate emissions from international shipping routes, which resulted in a significant reduction in sulfate particulate released during international shipping starting on January 1 2020, have been a major contributing factor to the monthly surface temperature anomalies during the last year. We do this by including in Community Earth System Model (CESM2) simulations the appropriate changes to emission databases developed for the Climate Model Intercomparison Project version 6 (CMIP6). The aerosol termination effect simulated by the updated CESM2 simulations of $+0.14 \pm 0.07$ W/m^2 and $0.08K \pm 0.03K$ is consistent with observations of both radiative forcing and surface temperature, manifesting a similar delay as the one observed in observational datasets between the implementation of the emission changes and the anomalous increase in warming. Our findings highlight the importance of considering realistic near-future changes in short-lived climate forcers for future climate projections, such as for CMIP7, for an improved understanding and communication of short-term climatic changes.

1 Introduction

In 2020, the International Maritime Organization (IMO) prescribed changes in the sulfur content of shipping fuels which resulted in a strong reduction in the particulate emissions of the sector, especially over the North Atlantic corridor. Past research indicated that such change would lead to a minor increase in the global Earth Energy Imbalance (Partanen et al., 2013), mainly through a reduction in cloud formation (Jin et al., 2018). To date, no Coupled Model Intercomparison Project Phase 6 (CMIP6) Earth System Model simulations have directly tried to ascertain the magnitude and role that this abrupt change in forcing would have on the Earth System, which we try to do here in the Community Earth System Model (CESM). We leverage the CESM2 Large Ensemble (LENS2, Rodgers et al. (2021)), which post-2014 uses an emission pathway (the Shared Socio-economic Pathway (SSP) 3-7.0) that does not consider any change in shipping emissions post-2020, and consider a new ensemble of simulations with the only change starting in January 2020 with a sudden drop in sulfate shipping emissions of 90%, hereby named NOSHIP or CESM2-LENS2 without shipping emissions. This drop is consistent with changes reported in (Hoesly and Smith, 2024) (see Methods and Fig. A1). In fact, none of the existing scenarios used in CMIP6 (Meinshausen et al., 2020) include substantial changes in shipping emissions in the present or near future of the same magnitude as those planned by

the IMO. Our simulations can help bridge this gap and provide a useful reference point for future emission scenarios such as those for the Phase 7 of CMIP (Meinshausen et al., 2023). The relevance of air quality measures on regional climate has been highlighted before (Zheng et al., 2020; Takemura, 2020; Schumacher et al., 2024): however, the IMO change case may present a unique case study in which global radiative balance may have been measurably affected due to a sharp, almost-instantaneous, event, thereby providing a useful benchmark for climate models' sensitivity.

2   Results

2.1   Top of Atmosphere radiative forcing

Compared to the CESM2-LENS2 in the default emission scenario, our simulations show an increase in Absorbed Solar Radiation (ASR) and net radiation at the top of atmosphere (TOA, longwave+shortwave) that is more consistent with the increase as measured by CERES (Fig. 1); furthermore, this overall increase is consistently observed as coming from an increase in Cloud Radiative Forcing (CRF, see Fig. A2), and not from a direct aerosol forcing (Clear Sky) as in CERES (Fig. A3). This is also confirmed by the different magnitude of optical depth changes simulated by CESM2 (Fig. A2). The global NET RF effect as diagnosed by the change between LENS2 and NOSHIP ~~, is of the order of 0.2~~ (Fig. 1c, orange and green lines, respectively), is 0.14 W/m^2, ~~in agreement with previous studies and estimates Partanen et al. (2013); Diamond (2023); Hodnebrog et al. (2024)~~ within the range of previous estimates (with a minimum of +0.06 to a maximum of +0.37 W/m^2 across different studies (Partanen et al., 2013; Yuan et al., 2024; Yoshioka et al., 2024; Skeie et al., 2024; Forster et al., 2024)). This indicates that CESM can better capture short-term forcing changes if the proper emission changes are accounted for, and being more consistent with actual anthropogenic aerosol emissions would help reconcile CESM projected TOA energy imbalance with the one observed. It is worth highlighting that ERA5 reanalyses also present a considerable mismatch between their radiative fluxes and the observed ones, posing question about the reliability of this reanalyses product for this specific purpose.

Our results are also consistent with previous studies in which aerosol emissions have already been shown to present a source of bias for CESM2 compared to observations (Ramachandran et al., 2020; Zhang et al., 2019) in other regions.

2.2   Temperature impacts

It has already been highlighted that global surface temperatures in 2023 cannot be explained by natural variability alone Rantanen and Laaksonen (2024). In Fig. 2 (see also Fig. ~~A~~A5-~~A6~~A7) we try to understand if such observed temperature anomalies can be connected with the shipping emission changes. Even if a radiative forcing change at a specific time can be detected, its translation to a temperature signal can be complicated by natural variability and internal climatic oscillations, and by time-lags due to slow oceanic response and cloud adjustments: as an example, estimates of the exact magnitude of the cooling signal from Pinatubo have varied (0.14-0.5 K, Canty et al. (2013)) even if the forcing change has been robustly evaluated Soden et al. (2002); Schmidt et al. (2018).
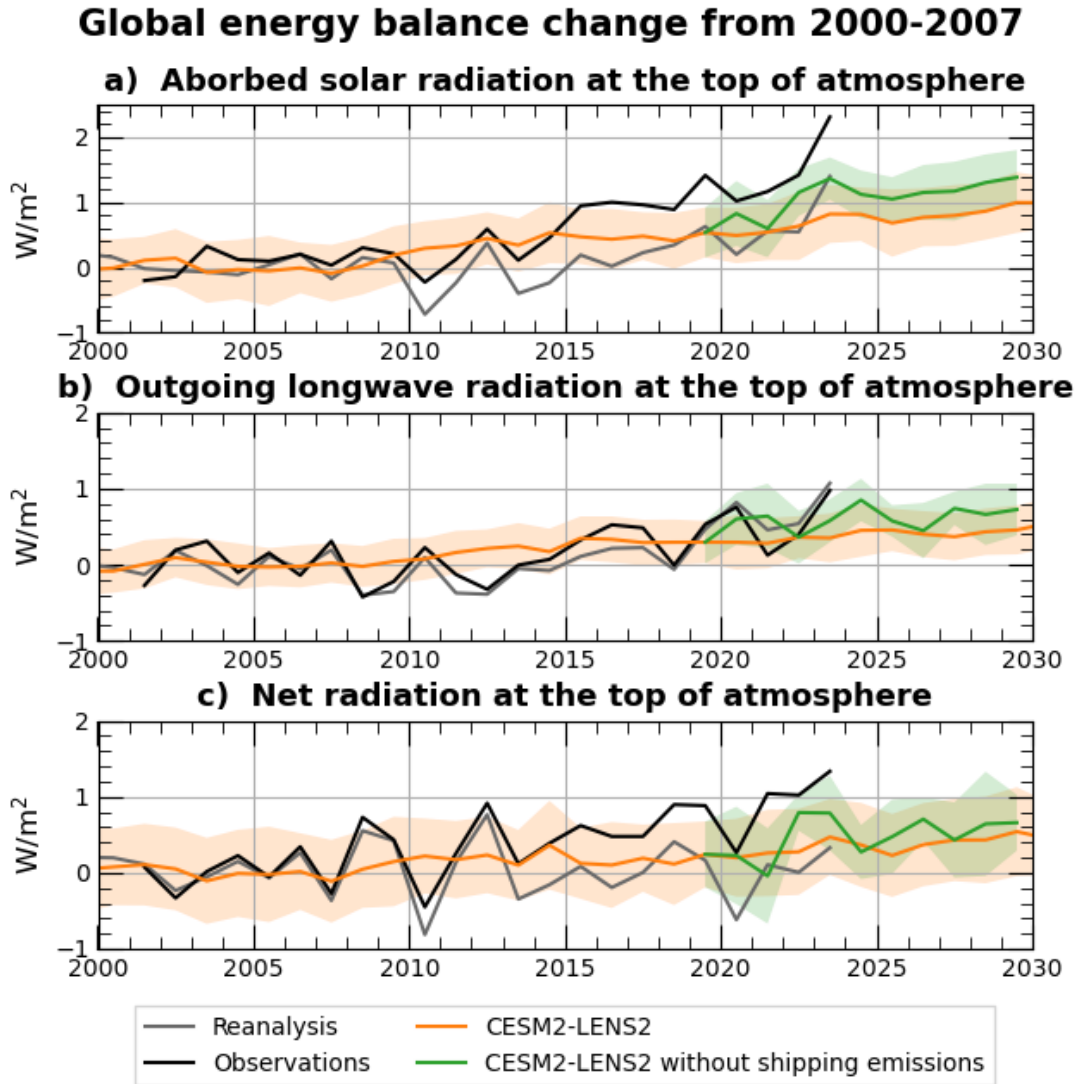
Figure 1. Time series of annual mean deviation from the 2000-2007 period for globally averaged (a) Absorbed Solar Radiation (ASR, defined as incoming minus outgoing shortwave), (b) Outgoing Longwave Radiation (OLR) and (c) ASR minus OLR (NET) radiative flux at the top of the atmosphere (TOA). The black line represents observational data from CERES(NASA/LARC/SD/ASDC, 2023), the gray line reanalysis from ERA5(Hersbach et al.), the orange and green lines the ensemble mean of CESM2-LENS2 simulations with and without shipping emissions, respectively. The shaded area for CESM2-LENS2 simulations represents one standard deviation calculated on ensemble members.

Here we use our two ensembles to perform an attribution of the 2023 temperature impact by comparing their results in the three years following the change in emission between the two scenarios. We ~~compare~~ compared our model's results with the ~~anomaly in one~~ anomalies in various surface air temperatures ~~dataset (Berkeley), looking~~ datasets (see Fig. A), but considering the small variations across them, here we only show the model's results compared to Berkley (see Methods). We looked at monthly de-trended global temperature anomalies over the period 2020-2023~~(~~, i.e. removing the, assumed linear, contribution from greenhouse gasses ~~).~~ and seasonality. Figure 2b shows the monthly de-trending method applied to August months which has been used for each month of the 3-year time series in Figure 2a. For the NOSHIP ensemble we considered an average and high estimates of shipping emission impact (light and dark green lines, respectively), calculated as the ensemble average and ensemble average plus one standard deviation in NOSHIP. The likelihood of occurring temperatures in Figure 2c and d is calculated for the measured temperature anomalies (black line) and for the measured temperature anomalies without the average and high contribution estimated from ship emissions (light and dark red lines corresponding to the light and dark green lines of panel 2a, respectively).

In LENS2, the mean ensemble response does not show anomalous increases in any of the years under analyses, whereas the NOSHIP ensemble demonstrates a striking agreement with observations in the manifestation of an anomaly compared to past years starting in June 2023, so 3 years after the change in shipping emission ~~. An analysis~~ (Fig. 2d). Analysis of the ENSO state (Fig. ~~A7) indicates that such an anomaly cannot be connected to an El-Niño-driven dynamical response, as the anomaly manifests even~~ A8) suggests that the anomaly cannot solely be attributed to a strong El Niño event starting in 2023, as it appears in ensemble members ~~in a El Niña state~~ even during El Niña states. This indicates that CESM responds to the global mean increase in incoming solar radiation in a way that is physically consistent with observations, manifesting in a delay in the surface temperature response with a time-lag of the order of three years.

Compared with the overall distribution of expected monthly anomalies in LENS2 (Fig. 2c), measured anomalies from Berkeley are statistically unlikely for the months from July to December (more than 2 σ), with September peaking with a likelihood of 0.02 %; considering the temperature contribution from shipping emissions, our results indicate that the same anomalies were roughly 9 times more likely to occur, and 27 times if we consider the higher-end estimate of the shipping contribution estimated by CESM (average + 1 σ of the NOSHIP ensemble) (Fig. 2d). For each month but September, this result indicates that, on its own, changes in shipping emission can be considered as a primary contributor capable of explaining the 2023 anomalies in light of internal variability.

The changes observed in the global mean also do not necessarily translate to an even change across the globe: in Fig. 3a we also show which regions, based on the difference between the two CESM ensembles, show significant changes in their annual mean temperatures in 2023. Our results indicate that most of the regions affected are ocean ones, especially in the North and South Atlantic and over the Pacific, with areas of significance over land in the Middle East and at high latitudes in the Northern Hemisphere. This is explainable by observing where most of the changes in absorbed solar radiation at the surface are, which we show in Fig. 3b considering the integrated change over the 2021-2023 period. In Fig. 3c and 3d we show that the changes in shipping emissions can also have some detectable effects on daily temperature extremes. Using the 90th percentile of daily

90 maximum temperature (see Methods), we show that, while the variability is quite high, a signal is distinguishable especially in late 2023 and in 2024 in the NOSHIP ensemble, with a higher percentage of very hot days especially in 2024.

3 Conclusions

Our results underscore the importance of considering realistic changes in aerosol emissions when discussing the evolution of surface temperatures at different timescales: we show both that, by 2030, the overall temperature increase is projected to be
95 ~~0.2~~ 0.14 K (Figure ~~A3~~A), and that in the 2020-2024 period there is clearly a heightened signal that contributes to the anomalous, sudden increase. Compared to previous research analyzing aerosol emission changes that mostly happened over land (i.e. Forster et al. (2020) in the case of COVID), in this case the complex interplay of cloud-mediated changes, absorbed solar radiation by the ocean, and ocean response may help explain why the emergence timescale is different. Our findings also points to the fact that future policy decisions around abrupt reductions in tropospheric aerosols might want to take into account their
100 surface temperature impact: while clear that policies that improve air qualities save lives Partanen et al. (2013), the presence of recent international agreements such as the Paris Agreement focused on avoiding future increases in temperature indicate that such ~~policies~~ air-quality focused policies, and more broadly other possible human and natural emission changes, may also need to be evaluated in the context of breaches of global temperature thresholds and related potential damages and risks. For instance, ~~would it~~ it would be legitimate to ask if such policies should ~~contribute to~~ more explicitly be framed in terms of
105 estimates of the remaining carbon budget ~~(Rogelj et al., 2019)?~~before they're enacted (Rogelj et al., 2019).

~~In future planned analyses, we will~~ Our forcing estimates of $+0.14$ W/m^2 $\pm$ 0.07 from CESM2 is located within the range of other recent works, which often used different methodologies to come to their conclusions. For instance, Yuan et al. (2024) found a forcing of 0.2 ± 0.11 W/m^2 (over the global oceans) indirectly estimating it from cloud changes as simulated by
110 NASA's Global Earth Observing System, which would result in a global forcing very close to ours overall. However, their temperature estimate of 0.16K is twice as large as our estimate of 0.08K $\pm$ 0.03. Other studies like Skeie et al. (2024) tried to estimate the effective radiative forcing by conducting fixed-SSTs similar to ours, using four models (CESM2-CAM6, NASA GISS ModelE, NorESM2, OsloCTM3), and finding a range of 0.06 to 0.09 W/m^2, similar to Yoshioka et al. (2024) which found 0.13 W/m^2 using HadGEM3-GC3.1, very close to the 0.14 W/m^2 $\pm$ 0.02 W/m^2 found in UKESM by Jordan and Henry (2024)
115 under similar experimental protocols. In this latter case, they also estimate the temperature response in UKESM to be 0.046K $\pm$ 0.010K. A future assessment of the different methodologies used will be necessary to reconcile these estimates, perhaps coupling it with a rigorous multi-model assessment in CMIP7. A cause for our estimate being towards the higher end of others might be our use of fully coupled simulations, which may result in a warming-driven feedback on cloud forcing, and which we pursued to try to reconcile our estimates of both forcing and temperature changes with available observations. It is also
120 possible that our results are overestimated due to an excessive sensitivity of CESM2 to cloud-aerosol interaction, or that they are driven by our ensemble size, and our specific ways of detecting 'significance' (for instance, see Watson-Parris et al. (2024) for a different interpretation of this while using a very similar modeling set-up as ours): if this is the case, however, it will be

necessary to find other explanations for the 2023 anomalous temperatures that currently don't seem to exist - especially as the persistent anomaly even after 12 months appears to rule out a statistical fluctuation.

125

Finally, we note that Forster et al. (2024) suggested that the 2024 global aerosol radiative forcing was made more negative due to the contribution from the Canadian wildfire. While our study is not suited to directly quantify such a potential contribution, due to prescribed biomass burning emissions that predate that specific event, we note that our analyses of CERES fluxes show a global positive increase in Clear sky fluxes between 2023 and 2024 (Fig. A2) that is hard to reconcile with such a hypothesis.

130 While it is certainly possible that increased warming resulted in higher wildfire risks at high latitudes (as suggested by our Fig. 3a results), satellite data of top-of-atmosphere radiative fluxes does not seem to support a forcing compensation between sulfate from shipping and wildfire aerosols. However, future studies including also realistic biomass burning could better clarify such matters.

135 Future analyses will shine further light on both the reliability of CESM's response in terms of changes in cloud cover and boundary layer sulfate and in terms of the response of such changes on the regional and global radiative forcing and temperature response. It is clear from this work already that in many cases the CESM ensemble is already underestimating the increasing radiative imbalance and, for instance, ~~on~~ the near-term projection of max daily temperatures. However, here we think we have successfully argued that further analyses of the current years' changes should not ignore real-world policy changes, both those

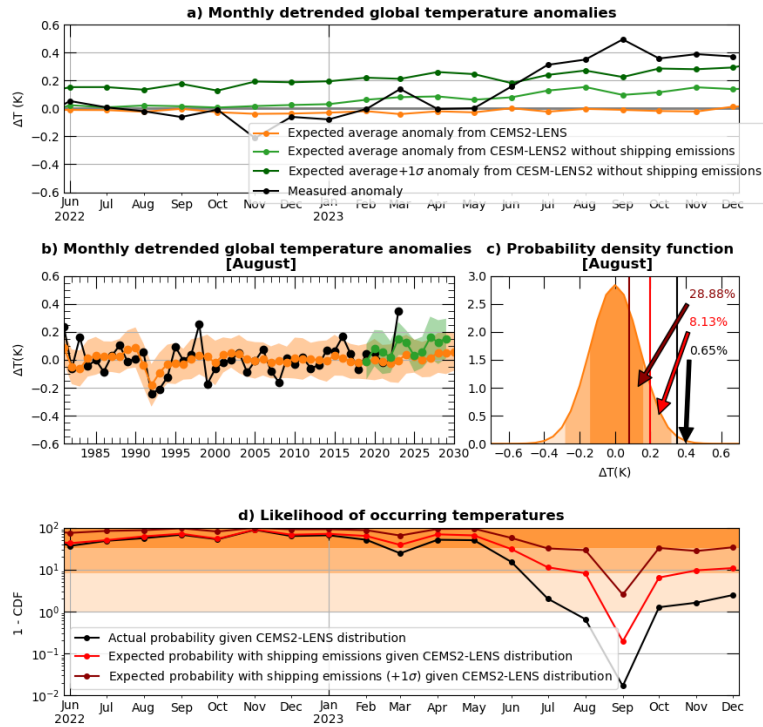140 already happened and future planned ones.

Figure 2. a) Detrended deseasonalized values of monthly mean global temperature changes from the 1981-2000 average. The black line represents observational data from Berkeley (in the legend: measured anomalies, ~~Rohde and Hausfather (2020)~~), the orange and light green line represent the ensemble mean of CESM2-LENS2 simulations with and without shipping emissions (in the legend: average anomaly from CESM2-LENS2 with and without shipping emissions, respectively), and the dark green line is the sum of the ensemble mean plus one standard deviation (1σ) of CESM2-LENS2 simulations without shipping emissions (in the legend: expected average + 1σ anomaly from ~~CESM-LENS2~~ CESM2-LENS2 without shipping emissions). b) Time series of globally averaged detrended temperature change from 1981-2000 as in (a) but only for the month of August, for observational data (black line), the ensemble mean of CESM2-LENS2 simulations with shipping emissions (orange line) and the ensemble mean of CESM2-LENS2 simulations without shipping emissions (light green line), as defined in the legend of panel (a). The shaded area represents one standard deviation on the ensemble members. c) Probability density function (PDF) of globally averaged detrended temperature change from 1981-2000 for the month of August in CESM2-LENS2 with shipping emissions, corresponding to the values of the orange line shown in panel (b). The PDF includes values from 1981 to 2020 for all ensemble members. Vertical lines represent the year 2023 for the measured anomalies (black line), anomalies due to shipping emissions (red line, calculated as difference between observed anomalies and average ensemble anomalies from CESM2-LENS2 without shipping emissions), and average ensemble anomalies plus one standard due to shipping emissions (dark red line, calculated as difference between observed anomalies and average ensemble anomalies plus one standard deviation from CESM2-LENS2 without shipping emissions). The values (in %) represent the right-tail values with respect to the CESM2-LENS2 PDF (one minus the cumulative density function, 1- CDF) for the three vertical lines. Legend is shown in the following panel. The three shaded area, from darkest to lightest orange, represent values within 1σ, 2σ and 3σ. d) Time series of right-tail values as defined in panel (c) for the month of August in 2023, here for each month of the year from June 2022 to the end of 2023.
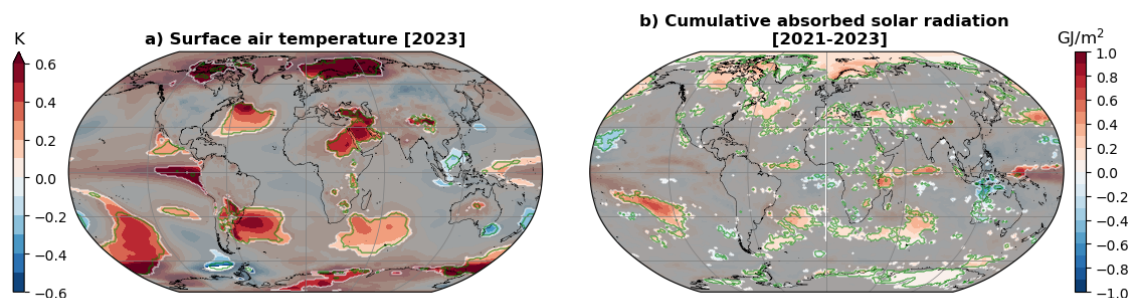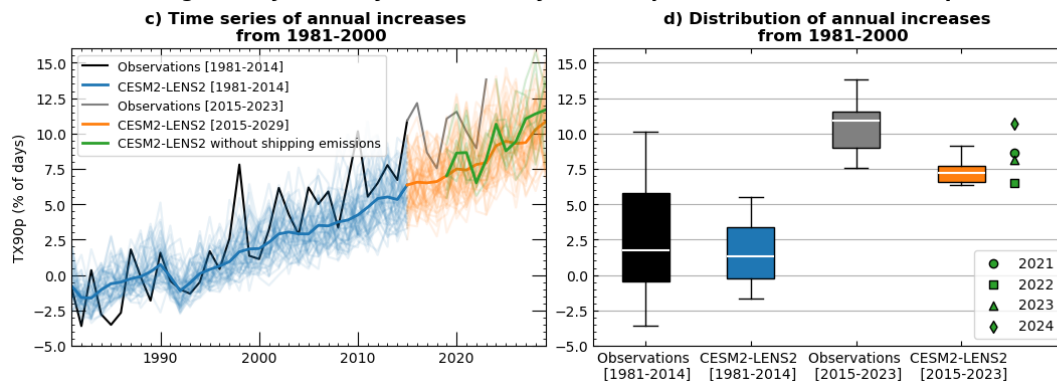
7

Figure 3. Maps of changes in surface air temperature in 2023 (a) and cumulative absorbed solar radiation over 2021-2023 (b) due to reduction in shipping emissions in CESM2-LENS2. Shaded areas indicate regions where the differences are not statistically significant at the 10% level, green contours indicate regions where the differences are not statistically significant at the 5% level. c) Time series of annual increases of percentage of days of the year when daily maximum temperature over land are greater than the 90th (TX90p) from 1981-2000 for observations from Berkeley and CESM2-LENS2, distinguished in pre- and post-2014, and CESM2-LENS2 without shipping emissions. d) Box plots of TX90p shown in panel c for Berkeley and ensemble-mean values in CESM2-LENS2, distinguished in pre- and post-2014. Green markers represent years from ~~2032~~ 2023 to 2024 for ensemble-mean values in CESM2-LENS2 without shipping emissions.

Data availability. Simulation data used in this work are available at Quaglia (2024).

Appendix A: Methods

A1   CEMS2-LENS2 simulations

The Large Ensemble Community Project (LENS2) consists of 100-member ensemble simulations that cover the historical period 1850-2014 (referred as Historical) and the period 2015-2100 under the SSP3-7.0 emission scenario (referred as SSP3-7.0). The simulations are performed Rodgers et al. (2021) with the Community Earth System Model version 2 run within the Community Atmosphere Model version 6 (CESM2-CAM6) Danabasoglu et al. (2020).

Ensemble members are initialized to capture to capture the internal variability of the climate system: ten members are initialized from different years of the pre-industrial control cycle, from 1001 to 1191 to minimize drift, and forty are initialized from four initial dates (1231, 1251, 1281, and 1301) that identify the phases of the Atlantic Meridional Overturning Circulation (AMOC), and for each starting date the ensemble members were created by randomly perturbing the atmospheric potential temperature field. We used 50 of the 100 ensemble members, i.e., members with smoothed biomass burning emissions of the CMIP6 protocol (11-year running means over the period 1990-2020 (Rodgers et al., 2021).

We additionally performed 10-member ensemble simulations under the SSP3-7.0 scenario (referred as NOSHIP or CESM2-LENS2 without shipping emissions), branched from the available ensemble members with start date 1011-1191 and smoothed biomass burning emissions of the CMIP6 protocol, reducing sulfur emissions from shipping Hoesly et al. (2018) by 90% everywhere over the oceans in accordance with IMO 2020 regulations from 2020 to 2030 (Fig. A1a) and keeping all other emissions the same. The drop in sulfur emissions results in an average reduction over this decade compared to the reference period 2000-2007 of 4.2 Tg-S/yr globally. Due to our computational constraints, we only performed 10 ensemble members compared to the original ensemble of 50. However, following Tebaldi et al. (2021) and Frankcombe et al. (2018), our ensemble size is representative of the perturbed state to allow us to estimate forcing changes.

A2   Observations and reanalysis

Sulfur shipping emissions are from the Community Emission Data System (CEDS) which provides estimates of emissions of anthropogenic greenhouse gases, reactive gases and aerosols, from 1750 to nowadays, based on existing emission inventories, emission factors, and activity/driver data (Hoesly and Smith, 2024).

To compare simulated radiative fluxes, we used satellite data from the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled Top-of-Atmosphere fluxes version 4.2 (CERES_EBAF_Edition4.2, NASA/LARC/SD/ASDC (2023)) and climate reanalysis data from the fifth generation European Centre for Medium-range Weather Forecasts (ECMWF) reanalysis (ERA5, Hersbach et al.).

Simulated surface temperatures are compared with ERA5, Berkeley Earth, Met Office Hadley Centre/Climatic Research Unit global surface temperature data set version 5.0.2.0 (HadCRUT5), and the NOAA Global Surface Temperature Dataset

version 6.0 (NOAAGlobalTempv6). For the results in our Figure 2 we tested all datasets but ultimately only showed Berkley, as our conclusions were largely independent of the dataset chosen (see Figure A5).

Berkeley Earth Land/Ocean Temperature Record (Rohde and Hausfather, 2020) combines the Berkeley Earth land-surface temperature field with an interpolated version of the Met Office Hadley Centre Sea Surface Temperature dataset version 4.0.0.0 (HadSST4). HadCRUT5 (Morice et al., 2021) uses a statistical infilling method to integrate sea-surface temperature data from the HadSST4 with land-surface air temperature data from the Climatic Research Unit temperature dataset version 5.0.0.0 (CRUTEM5). NOAAGlobalTempv6 (Huang et al.) combines the land-ocean surface temperature analysis from the Extended Reconstructed Sea Surface Temperature (ERSSTv5) with land surface air temperature analysis, which are is on the Global Historical Climatology Network-Monthly (GHCN-M) temperature database.

A3   Calculation of temperature anomalies

Temperature anomalies for both observations and modeling results are calculated by subtracting the average value on the 1981-2000 period. Each month is detrended considering the timeseries for each month separately, removing the trend calculate as a linear fit on the 1981-2020 period.

A4   Probability density function of detrended temperature anomalies

The probability density function (PDF) for each month of the year (Fig. A5A6) is calculated on the detrended temperature anomalies from 1981 to 2020 (Fig. AA5), as define before, in CESM2-LENS2 with shipping emissions simulation, including all ensemble members, henceforth referred as CESM2-LENS2 PDF. The right-tail values are calculated as one minus the cumulative density function (1 - CDF) given CEMS2-LENS PDF for observational data from Berkeley, and the so called "Expected probability with shipping emissions given CEMS2-LENS distribution" and "Expected probability with shipping emissions +1σ given CEMS2-LENS distribution" (see Fig. 2c and d).

The latter two are the right-tail values, given the CEMS2-LENS distribution, for the detrended temperature anomalies calculated as the difference in the temperature anomalies minus the linear trend in the period 1981-2020 between observed values and the ensemble average in CESM2-NOSHIP ("expected probability with shipping emissions given CEMS2-LENS distribution") and between observed values and the ensemble average plus one standard deviation in CESM2-NOSHIP ("expected probability with shipping emissions +1σ given CEMS2-LENS distribution"). This allows us to calculate the probability that the monthly temperature anomaly would have happened if the further warming contribution from shipping emissions had not been removed, treating the LENS2 as the counterfactual world where shipping emissions are maintained, and the difference between the observational dataset and the NOSHIP mean as the counterfactual observational dataset in a world that had maintained shipping emissions unaltered.

A5   Calculation of Extreme Temperature Index

TX90p is a climate change index defined as the percentage of days when daily maximum temperature greater than the 90th percentile of daily maximum temperature (TX90p), according to CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI, http://etccdi.pacificclimate.org). We use this as a measure of changes in extreme temperatures over land that might be relevant for future impacts assessment.

Thresholds for TX90p are calculate as described in Zhang et al. (2005) but without using a bootstrap procedure: the 90th percentile of daily maximum temperature is calculated for each day of the year over a base-period of 40 years, from 1961 to 2010 (and including all ensemble-members, in CESM case), and using data from 5 consecutive days centred on the day of interest. This means that in CESM the threshold for a given day of the year is calculated over a distribution of 40x50x5 data (years x ensemble-members x days). Each calendar day of the time series is compared with the respective threshold for that day of the year. We validated our calculations for CESM by applying the same method to the Berekley data set and comparing our results with those calculated by Climpact (https://www.climdex.org/, with results shown in Fig. A8 A9 where we demonstrate a close match between available results and ours).

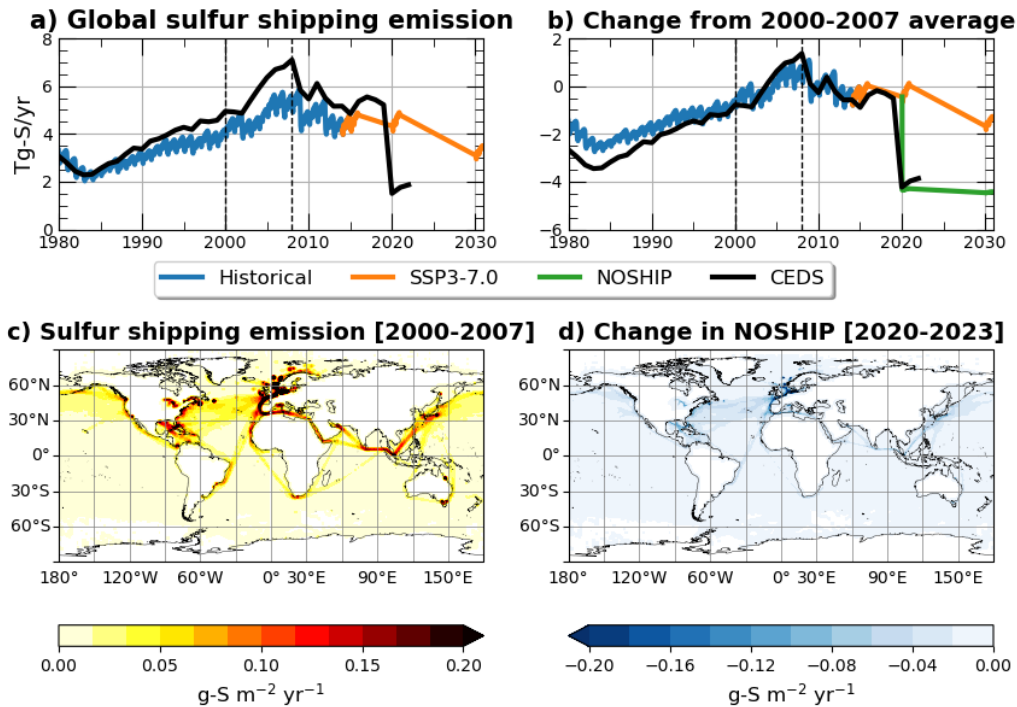**Figure A1.** a, b) Time series of the global mean of sulfur ($SO_2 + SO_4$, in Tg-S/yr) shipping emissions and its change from the from the 2000-2007 period in CESM2-LENS2 ~~from the~~ and Community Emission Data System (CEDS). Dotted vertical lines define the reference period 2000-2007. c, d) Maps of sulfur shipping emissions averaged over the reference period and the change in sulfur shipping emissions (g-S/yr/m^2) from the reference period averaged over the years 2020-2023, respectively.
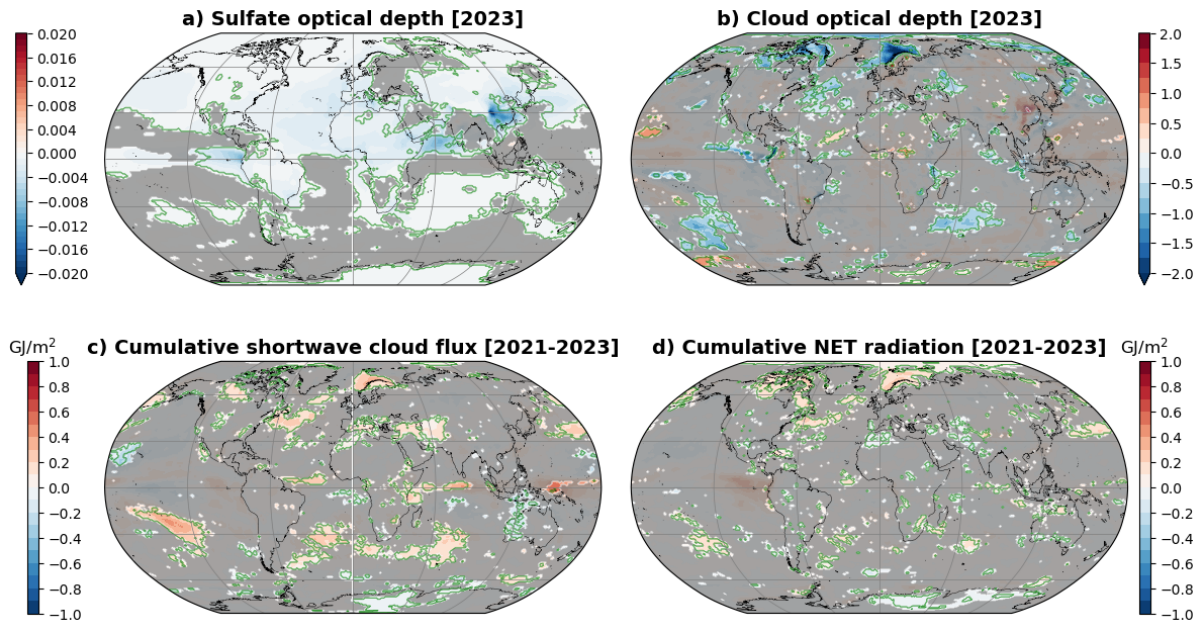
Figure A2. Maps of changes in sulfate and cloud visible optical depth in 2023 (a and b, respectively), and cumulative shortwave cloud flux and NET radiation over 2021-2023 (c and d, respectively) due to reduction in shipping emissions in CESM2-LENS2. Shaded areas indicate regions where the differences are not statistically significant at the 10% level, green contours indicate regions where the differences are not statistically significant at the 5% level.
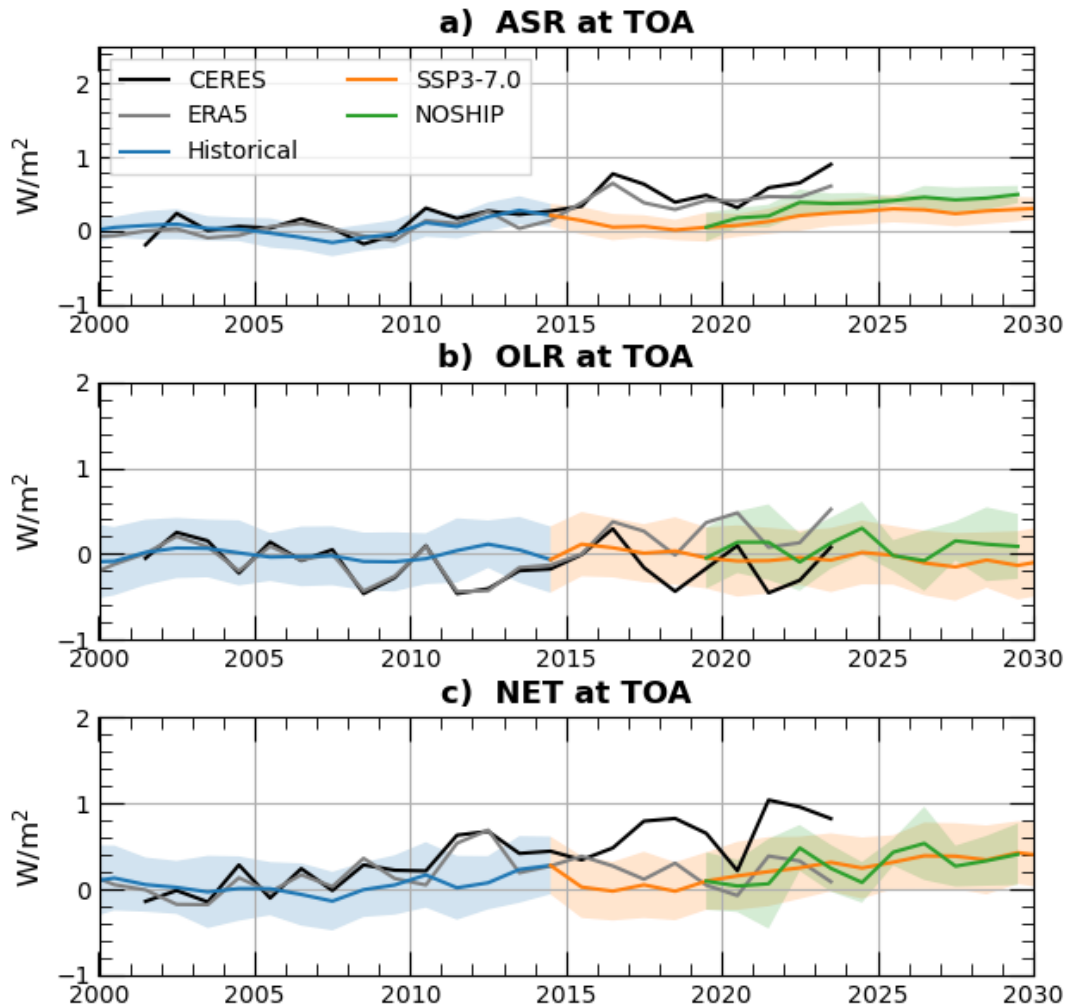
Figure A3. Time series of annual mean deviation from the 2000-2007 period for globally averaged (a) Absorbed Solar Radiation (ASR, defined as incoming minus outgoing shortwave), (b) Outgoing Longwave Radiation (OLR) and (c) ASR minus OLR (NET) radiative flux at the top of the atmosphere (TOA), in clear sky ~~condition~~conditions. The shaded area for CESM2-LENS2 simulations (Historical, SSP3-7.0, NOSHIP) represents one standard deviation calculated on ensemble members.

Global mean surface air temperature change from 1981-2000

~~Time series of detrended global mean temperature changes from the 1981-2000~~ and observations and reanalysis (Berkeley, ~~distinguished for each month of the year~~ ERA5, NOAAGlobalTempv6, HadCRUT5). The shaded area for CESM2-LENS2 simulations ~~(Historical, SSP3-7.0, NOSHIP)~~ represents one standard deviation calculated on ensemble members.

~~Time series of detrended global mean temperature changes from the 1981-2000~~ and observations and reanalysis (Berkeley, ~~distinguished for each month of the year~~ ERA5, NOAAGlobalTempv6, HadCRUT5). The shaded area for CESM2-LENS2 simulations ~~(Historical, SSP3-7.0, NOSHIP)~~ represents one standard deviation calculated on ensemble members.

Figure A4. Time series of 12-months rolling mean of global mean temperature changes from 1981-2000 ~~. The shaded area~~ for CESM2-LENS2 simulations (Historical, SSP3-7.0, NOSHIP) ~~represents one standard deviation calculated on ensemble members.~~
~~Time series of detrended global mean temperature changes from the 1981-2000~~ and observations and reanalysis (Berkeley, ~~distinguished for each month of the year~~ ERA5, NOAAGlobalTempv6, HadCRUT5). The shaded area for CESM2-LENS2 simulations ~~(Historical, SSP3-7.0, NOSHIP)~~ represents one standard deviation calculated on ensemble members.
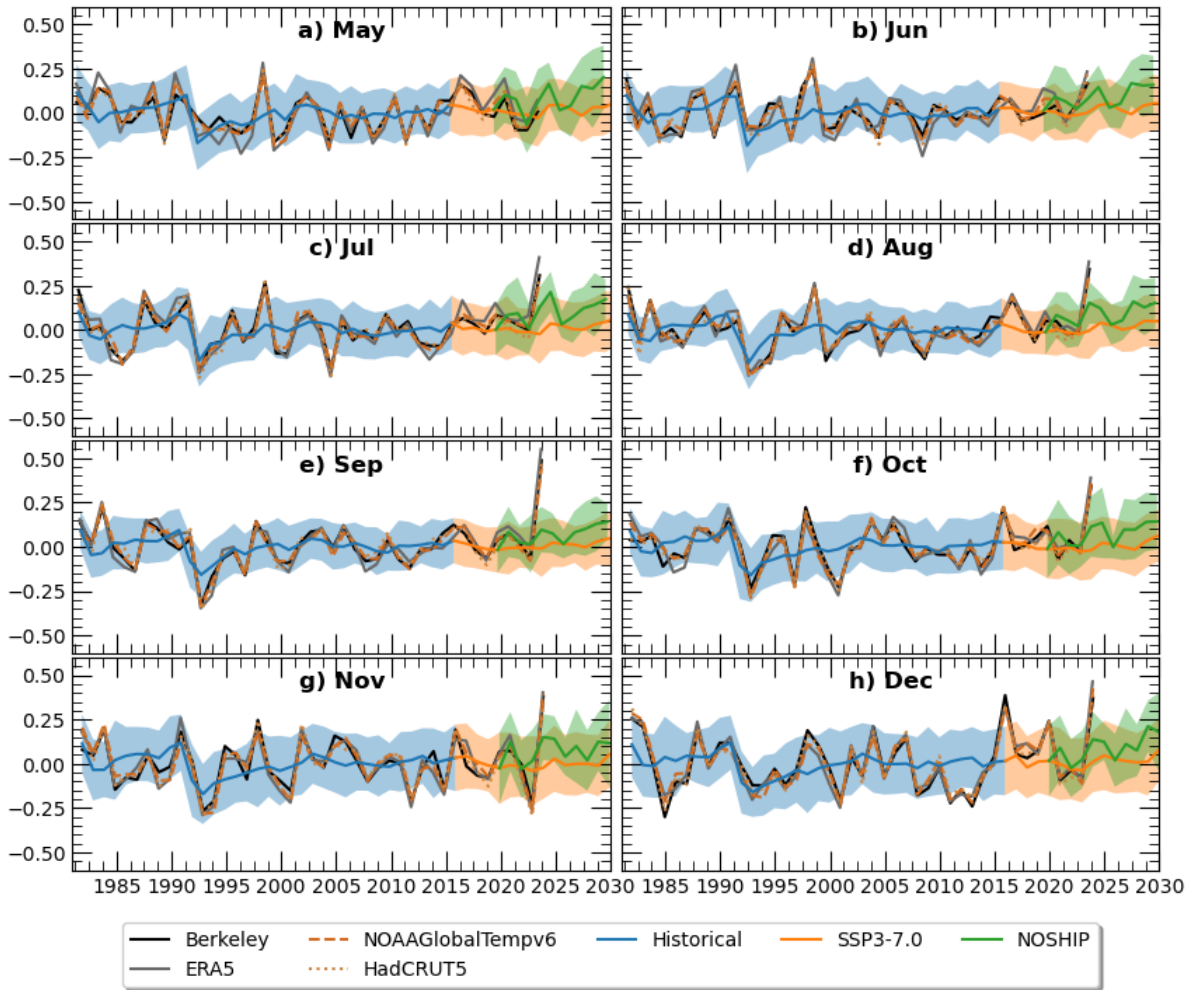
Figure A5. Time series of detrended global mean temperature changes from the 1981-2000, distinguished for each month of the year, for CESM2-LENS2 simulations (Historical, SSP3-7.0, NOSHIP) and observations and reanalysis (Berkeley, ERA5, NOAAGlobalTempv5, HadCRUT5). The shaded area for CESM2-LENS2 simulations represents one standard deviation calculated on ensemble members.
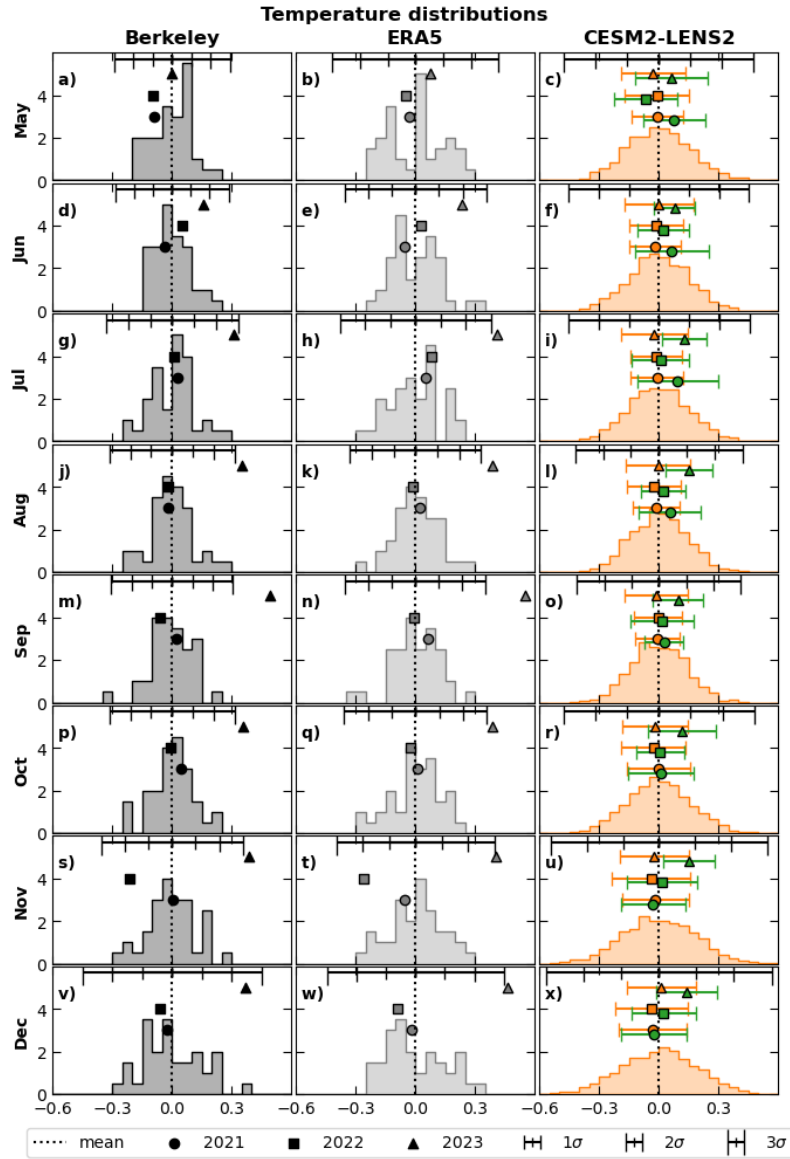
Figure A6. Probability density function of temperature changes from the 1981-2000 mean for values from 1981 to 2020 in Berkeley (first column), ERA5 (second column), and CESM2-LENS2 (Historical+SSP2-4.5, third column). Values from 2021-2023 are shown separately with different markers as defined in the legend. The dashed vertical line represents the mean value and the error bars 1σ, 2σ, and 3σ calculated from the respective PDF 1981-2020. CESM2-LENS2 panels includes all ensemble members, the 2021-2023 values for both SSP3-7.0 (orange) and NOSHIP (green) simulations, and the standard deviation calculated on the ensemble members.
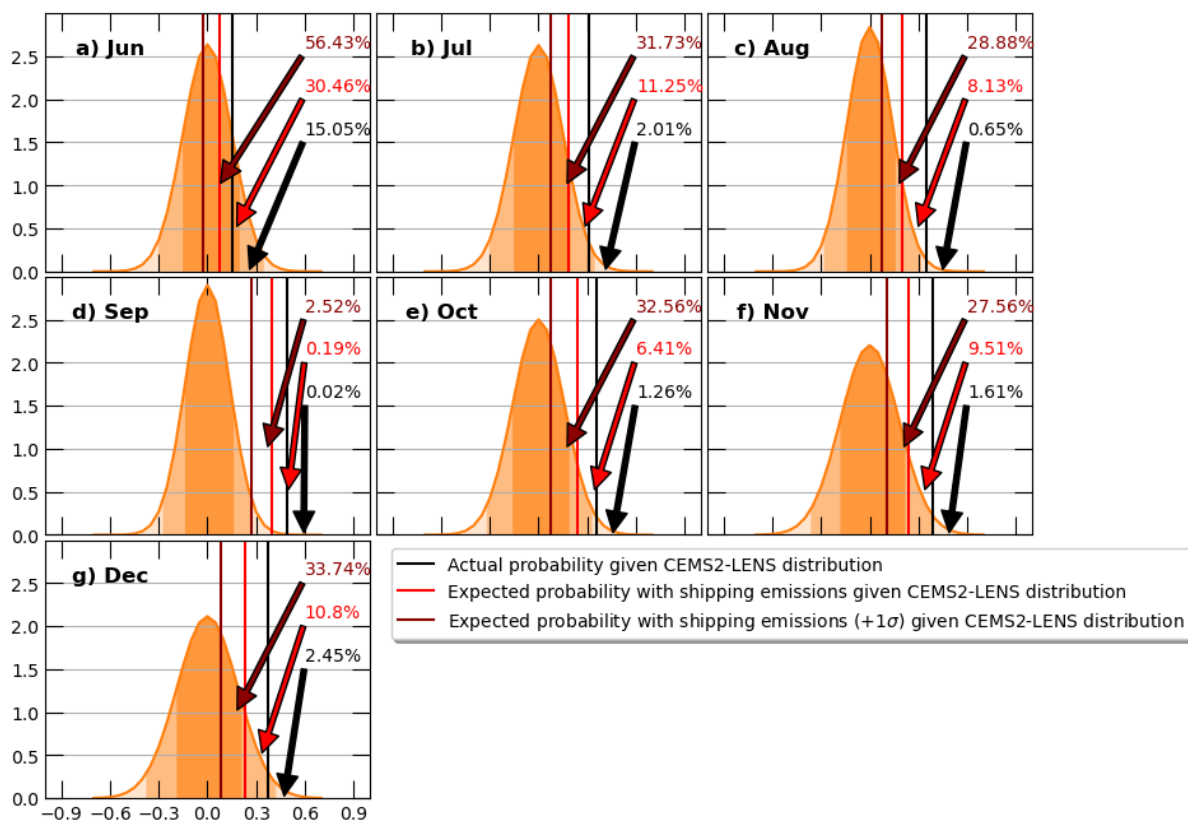
Figure A7. Probability density function (PDF) of globally averaged detrended temperature change from 1981-2000 in CESM2-LENS2 for the months of the year from May to December. The PDF includes values from 1981 to 2020 for all ensemble members. Vertical lines represent the year 2023 for the measured anomalies (Berkeley, black line), anomalies due to shipping emissions (red line, calculated as difference between Berkeley and average ensemble anomalies from NOSHIP), and average ensemble anomalies plus one standard due to shipping emissions (dark red line, calculated as difference between Berkeley and average ensemble anomalies plus one standard deviation from NOSHIP). The values (in %) represent the right-tail values with respect to the CESM2-LENS2 PDF (one minus the cumulative density function, 1- CDF) for the three vertical lines. The three shaded area, from darkest to lightest orange, represent values within 1σ, 2σ and 3σ.
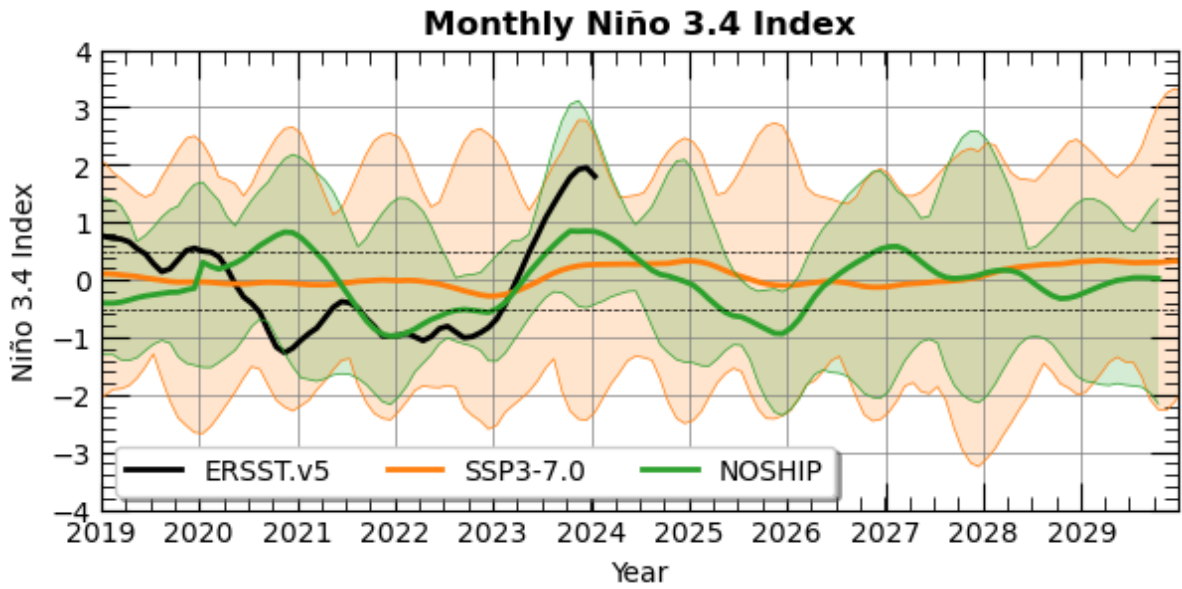
Figure A8. The El-Niño 3.4 anomalies. In CESM2-LENS2, the index is calculated using a 20-yr sliding climatology (1950-1999 from Historical) and smoothing the anomalies with a 5-month running mean. The shaded area for CESM2-LENS2 simulations (SSP3-7.0 and NOSHIP) represents one standard deviation calculated on ensemble members. ERSST.v5 data are taken from https://www.ncei.noaa.gov/access/monitoring/enso/sst.
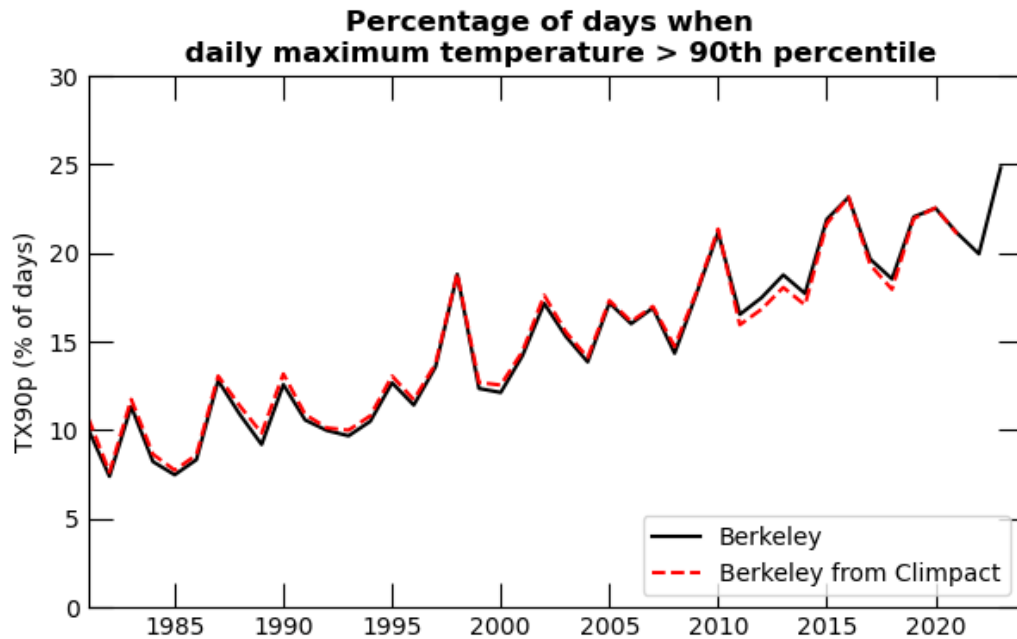
Figure A9. Percentage of days of the year when daily maximum temperature greater than the 90th percentile of daily maximum temperature (TX90p) from the Berkeley data set. The black line comes from our calculations (see Methods section), compared with Climpact data (dashed red line) taken from https://www.climdex.org/.

References

220 Canty, T., Mascioli, N. R., Smarte, M. D., and Salawitch, R. J.: An empirical model of global climate – Part 1: A critical evaluation of volcanic cooling, Atmospheric Chemistry and Physics, 13, 3997–4031, https://doi.org/10.5194/acp-13-3997-2013, 2013.

Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Van Kampenhout, L., Vertenstein, M., Bertini,
225 A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, https://doi.org/10.1029/2019MS001916, 2020.

Diamond, M. S.: Detection of large-scale cloud microphysical changes within a major shipping corridor after implementation of the International Maritime Organization 2020 fuel sulfur regulations, Atmospheric Chemistry and Physics, 23, 8259–8269,
230 https://doi.org/10.5194/acp-23-8259-2023, 2023.

Forster, P. M., Forster, H. I., Evans, M. J., Gidden, M. J., Jones, C. D., Keller, C. A., Lamboll, R. D., Quéré, C. L., Rogelj, J., Rosen, D., Schleussner, C.-F., Richardson, T. B., Smith, C. J., and Turnock, S. T.: Current and future global climate impacts resulting from COVID-19, Nature Climate Change, 10, 913–919, https://doi.org/10.1038/s41558-020-0883-0, 2020.

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Hall, B., Hauser, M., Ribes, A., Rosen, D., Gillett, N. P., Palmer, M. D., Rogelj,
235 J., Von Schuckmann, K., Trewin, B., Allen, M., Andrew, R., Betts, R. A., Borger, A., Boyer, T., Broersma, J. A., Buontempo, C., Burgess, S., Cagnazzo, C., Cheng, L., Friedlingstein, P., Gettelman, A., Gütschow, J., Ishii, M., Jenkins, S., Lan, X., Morice, C., Mühle, J., Kadow, C., Kennedy, J., Killick, R. E., Krummel, P. B., Minx, J. C., Myhre, G., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Schleussner, C.-F., Seneviratne, S. I., Szopa, S., Thorne, P., Kovilakam, M. V. M., Majamäki, E., Jalkanen, J.-P., Van Marle, M., Hoesly, R. M., Rohde, R., Schumacher, D., Van Der Werf, G., Vose, R., Zickfeld, K., Zhang, X., Masson-Delmotte, V., and Zhai, P.: Indicators of Global Climate
240 Change 2023: annual update of key indicators of the state of the climate system and human influence, Earth System Science Data, 16, 2625–2658, https://doi.org/10.5194/essd-16-2625-2024, 2024.

Frankcombe, L. M., England, M. H., Kajtar, J. B., Mann, M. E., and Steinman, B. A.: On the Choice of Ensemble Mean for Estimating the Forced Signal in the Presence of Internal Variability, Journal of Climate, 31, 5681 – 5693, https://doi.org/10.1175/JCLI-D-17-0662.1, 2018.

245 Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 monthly averaged data on single levels from 1940 to present, https://doi.org/10.24381/cds.f17050d7.

Hodnebrog, , Myhre, G., Jouan, C., Andrews, T., Forster, P. M., Jia, H., Loeb, N. G., Olivié, D. J. L., Paynter, D., Quaas, J., Raghuraman, S. P., and Schulz, M.: Recent reductions in aerosol emissions have increased Earth's energy imbalance, 2024.

250 Hoesly, R. and Smith, S.: CEDS v_2024_04_01 Release Emission Data, https://doi.org/10.5281/zenodo.10904361, 2024.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-i., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Huang, B., Yin, X., Menne, M. J., Vose, R. S., and Zhang, H.-M.: NOAA Global Surface Temperature Dataset (NOAAGlobalTemp), Version 6.0 [202406], https://doi.org/10.25921/rzxg-p717.

Jin, Q., Grandey, B. S., Rothenberg, D., Avramov, A., and Wang, C.: Impacts on cloud radiative effects induced by coexisting aerosols converted from international shipping and maritime DMS emissions, Atmospheric Chemistry and Physics, 18, 16 793–16 808, https://doi.org/10.5194/acp-18-16793-2018, 2018.

Jordan, G. and Henry, M.: IMO2020 Regulations Accelerate Global Warming by up to 3 Years in UKESM1, https://doi.org/10.22541/essoar.172046812.23872610/v1, 2024.

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M. K., and Wang, R. H. J.: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500, Geoscientific Model Development, 13, 3571–3605, https://doi.org/10.5194/gmd-13-3571-2020, 2020.

Meinshausen, M., Schleussner, C.-F., Beyer, K., Bodeker, G., Boucher, O., Canadell, J. G., Daniel, J. S., Diongue-Niang, A., Driouech, F., Fischer, E., Forster, P., Grose, M., Hansen, G., Hausfather, Z., Ilyina, T., Kikstra, J. S., Kimutai, J., King, A., Lee, J.-Y., Lennard, C., Lissner, T., Nauels, A., Peters, G. P., Pirani, A., Plattner, G.-K., Pörtner, H., Rogelj, J., Rojas, M., Roy, J., Samset, B. H., Sanderson, B. M., Séférian, R., Seneviratne, S., Smith, C. J., Szopa, S., Thomas, A., Urge-Vorsatz, D., Velders, G. J. M., Yokohata, T., Ziehn, T., and Nicholls, Z.: A perspective on the next generation of Earth system model scenarios: towards representative emission pathways (REPs), Geoscientific Model Development Discussions, 2023, 1–40, https://doi.org/10.5194/gmd-2023-176, 2023.

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 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/https://doi.org/10.1029/2019JD032361, e2019JD032361 2019JD032361, 2021.

NASA/LARC/SD/ASDC: CERES Energy Balanced and Filled (EBAF) TOA and Surface Monthly means data in netCDF Edition 4.2, https://doi.org/10.5067/TERRA-AQUA-NOAA20/CERES/EBAF_L3B004.2, 2023.

Partanen, A. I., Laakso, A., Schmidt, A., Kokkola, H., Kuokkanen, T., Pietikäinen, J.-P., Kerminen, V.-M., Lehtinen, K. E. J., Laakso, L., and Korhonen, H.: Climate and air quality trade-offs in altering ship fuel sulfur content, Atmospheric Chemistry and Physics, 13, 12 059–12 071, https://doi.org/10.5194/acp-13-12059-2013, 2013.

Quaglia, I.: Data from: Modeling 2020 regulatory changes in international shipping emissions helps explain 2023 anomalous warming, https://doi.org/10.5281/zenodo.11204379, 2024.

Ramachandran, S., Rupakheti, M., and Lawrence, M. G.: Aerosol-induced atmospheric heating rate decreases over South and East Asia as a result of changing content and composition, Scientific Reports, 10, 20 091, https://doi.org/10.1038/s41598-020-76936-z, 2020.

Rantanen, M. and Laaksonen, A.: The jump in global temperatures in September 2023 is extremely unlikely due to internal climate variability alone, npj Climate and Atmospheric Science, 7, 34, https://doi.org/10.1038/s41612-024-00582-9, 2024.

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J.-E., Simpson, I. R., Stein, K., Stuecker, M. F., Yamaguchi, R., Bódai, T., Chung, E.-S., Huang, L., Kim, W. M., Lamarque, J.-F., Lombardozzi, D. L., Wieder, W. R., and Yeager, S. G.: Ubiquity of human-induced changes in climate variability, Earth System Dynamics, 12, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021, 2021.

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., and Séférian, R.: Estimating and tracking the remaining carbon budget for stringent climate targets, Nature, 571, 335–342, https://doi.org/10.1038/s41586-019-1368-z, 2019.

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

295 Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., Bardeen, C. G., Conley, A., Forster, P. M., Gettelman, A., Portmann, R. W., Solomon, S., and Toon, O. B.: Volcanic Radiative Forcing From 1979 to 2015, Journal of Geophysical Research: Atmospheres, 123, 12 491–12 508, https://doi.org/10.1029/2018JD028776, 2018.

Schumacher, D. L., Singh, J., Hauser, M., Fischer, E. M., Wild, M., and Seneviratne, S. I.: Exacerbated summer European warming not captured by climate models neglecting long-term aerosol changes, Communications Earth & Environment, 5, 182, 300 https://doi.org/10.1038/s43247-024-01332-8, 2024.

Skeie, R. B., Byrom, R., Hodnebrog, Ø., Jouan, C., and Myhre, G.: Multi-model effective radiative forcing of the 2020 sulphur cap for shipping, https://doi.org/10.5194/egusphere-2024-1394, 2024.

Soden, B. J., Wetherald, R. T., Stenchikov, G. L., and Robock, A.: Global Cooling After the Eruption of Mount Pinatubo: A Test of Climate Feedback by Water Vapor, Science, 296, 727–730, https://doi.org/10.1126/science.296.5568.727, 2002.

305 Takemura, T.: Return to different climate states by reducing sulphate aerosols under future CO2 concentrations, Scientific Reports, 2020.

Tebaldi, C., Dorheim, K., Wehner, M., and Leung, R.: Extreme metrics from large ensembles: investigating the effects of ensemble size on their estimates, Earth System Dynamics, 12, 1427–1501, https://doi.org/10.5194/esd-12-1427-2021, 2021.

Watson-Parris, D., Wilcox, L. J., Stjern, C. W., Allen, R. J., Persad, G., Bollasina, M. A., Ekman, A. M. L., Iles, C. E., Joshi, M., Lund, M. T., McCoy, D., Westervelt, D., Williams, A., and Samset, B. H.: Weak surface temperature effects of recent reductions in shipping SO_2 310 emissions, with quantification confounded by internal variability, EGUsphere, 2024, 1–30, https://doi.org/10.5194/egusphere-2024-1946, 2024.

Yoshioka, M., Grosvenor, D. P., Booth, B. B. B., Morice, C. P., and Carslaw, K. S.: Warming effects of reduced sulfur emissions from shipping, https://doi.org/10.5194/egusphere-2024-1428, 2024.

Yuan, T., Song, H., Oreopoulos, L., Wood, R., Bian, H., Breen, K., Chin, M., Yu, H., Barahona, D., Meyer, K., and Platnick, S.: Abrupt reduc- 315 tion in shipping emission as an inadvertent geoengineering termination shock produces substantial radiative warming, Communications Earth & Environment, 5, 281, https://doi.org/10.1038/s43247-024-01442-3, 2024.

Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., and Hao, J.: Drivers of improved PM $_{2.5}$ air quality in China from 2013 to 2017, Proceedings of the 320 National Academy of Sciences, 116, 24 463–24 469, https://doi.org/10.1073/pnas.1907956116, 2019.

Zhang, X., Hegerl, G., Zwiers, F. W., and Kenyon, J.: Avoiding Inhomogeneity in Percentile-Based Indices of Temperature Extremes, Journal of Climate, 18, 1641–1651, https://doi.org/10.1175/JCLI3366.1, 2005.

Zheng, Y., Zhang, Q., Tong, D., Davis, S. J., and Caldeira, K.: Climate effects of China's efforts to improve its air quality, Environmental Research Letters, 15, 104 052, https://doi.org/10.1088/1748-9326/ab9e21, 2020.