The CESM2 Single Forcing Large Ensemble and Comparison to CESM1: Implications for Experimental Design

Isla R. Simpson[a] , Nan Rosenbloom[a] , Gokhan Danabasoglu[a] , Clara Deser[a] , Stephen G. Yeager[a] , Christina S. McCluskey[a] , Ryohei Yamaguchi[b] , Jean-Francois Lamarque[a] , Simone Tilmes[c] , Michael J. Mills[c] , Keith B. Rodgers[d,e]

[a] *Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

[b] *Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan*

[c] *Atmospheric Chemistry Observations and Modelling Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

[d] *Center for Climate Physics, Insitute for Basic Science, Busan, South Korea*

[e] *Pusan National University, Busan, South Korea*

Corresponding author: Isla Simpson, islas@ucar.edu

ABSTRACT: Single Forcing Large Ensembles are a relatively new tool for quantifying the contributions of different anthropogenic and natural forcings to the historical and future projected evolution of the climate system. This study introduces a new single forcing large ensemble with the Community Earth system Model version 2 which can be used to separate the influences of greenhouse gases, anthropogenic aerosols, biomass burning aerosols, and all remaining forcings, on the evolution of the Earth System from 1850 to 2050. Here, the forced responses of global near surface temperature and associated drivers are examined in CESM2 and compared with those in a single forcing large ensemble with CESM2's predecessor, CESM1. The experimental design, the imposed forcing and the model physics all differ between the CESM1 and CESM2 ensembles. In CESM1 an "all-but-one" approach was used where everything except the forcing of interest is time evolving, while in CESM2 an "only" approach is used, where only the forcing of interest is time evolving. This experimental design choice is shown to matter considerably for anthropogenic aerosol-forced change in CESM2, due to state dependence of cryospheric albedo feedbacks and non-linearity in the Atlantic Meridional Overturning Circulation (AMOC) response to forcing. This impact of experimental design is, however, strongly dependent on the model physics and/or the imposed forcing as the same sensitivity to experimental design is not found in CESM1, which appears to be an inherently less non-linear model in both its AMOC behavior and cryospheric feedbacks.

1. Introduction

Historically, the Earth's climate system has evolved under a mixture of natural and anthropogenic forcings and it will continue to do so moving forward. A common approach that is used to disentangle and understand the relative contributions of such forcings to the evolution of the climate system is to perform Earth System Model (ESM) experiments in which only some forcings are evolving in time while others are held fixed. These experiments, which we will refer to as single forcing experiments even though they may be used to isolate the influence of multiple forcings at once, are most informative when a relatively large number of ensemble members are available, such that the forced signal can be isolated from the internal variability (e.g., Deser et al. 2020a). Many modelling centers have performed single forcing experiments under the coordinated framework of the "Detection and Attribution Model Intercomparison Project" (DAMIP, Gillett et al. 2016) as part of the Coupled Model Intercomparison Project (CMIP) and their utility has been recognized by the World Climate Research Program through emphasis on single forcing large ensembles as part of the lighthouse activity on "Explaining and Predicting Earth System Change" (Smith et al. 2022).

Single forcing experiments have been a core component of Intergovernmental Panel on Climate Change (IPCC) reports and have been used to conclude that human influence has unequivocally warmed the climate (IPCC 2021; Gillett et al. 2021). Beyond this, they have been used to investigate the wide ranging impacts of individual forcings on various aspects of the climate system. These include: the global patterns of surface temperature and precipitation anomalies induced by greenhouse gas versus aerosol forcing (Deser et al. 2020b; Shi et al. 2022); the influence of aerosol forcing on precipitation in the Sahel region (Dong et al. 2014; Giannini and Kaplan 2019; Hirasawa et al. 2020; Zhang et al. 2021) and other monsoon regions of the world (Li et al. 2018; Undorf et al. 2018; Monerie et al. 2022); the influence of ozone depleting substances and greenhouse gases on precipitation over Australia (Delworth and Zeng 2014); assessment of the aerosol-forced contribution to trends in the Pacific Ocean (Allen et al. 2014; Dittus et al. 2021) ; assessment of the counteracting influence of greenhouse gases and aerosols on Arctic sea ice (Mueller et al. 2018) and Arctic temperatures (England et al. 2021); the impacts of individual forcings on the North Atlantic Ocean circulation (Watanabe and Tatebe 2019; Dagan et al. 2020; Baek et al. 2022) and global sea level rise (Fasullo et al. 2020); the impacts of land cover change and irrigation on surface temperature and precipitation (Singh et al. 2018); and, the impact of individual forcings on various other aspects of the hydrological cycle and extreme weather (Chiang et al. 2021; Pendergrass et al. 2019; Bonfils et al. 2020; Seong et al. 2021; Touma et al. 2021). Single forcing experiments can also provide a useful testbed for exploring model sensitivity to differences in imposed forcings (e.g., Fyfe et al. 2021) or for inter-comparing the response to forcings among models (e.g., Menary et al. 2020; Dittus et al. 2021).

In the design of single forcing experiments choices must be made. Under the DAMIP protocol, the forcing of interest is evolving in time while all others are held fixed at pre-industrial values, referred to as the "only" method, hereafter. Another option is to evolve all forcings in time *except* the one of interest and determine that forcings influence by differencing this experiment from an all forcing simulation, referred to as the "all-but-one" method, herafter; this was the choice made for the single forcing large ensemble with the CESM1 model. There is also a choice as to what year forcings are held fixed at; DAMIP fixes them at 1850, while the CESM1 single forcing large ensemble fixed them at 1920. Whether these various design choices will produce the same answer as to a forcing's influence will depend on whether there are substantial non-linearities or state dependencies in the system, and prior results have been mixed as to whether this is the case. Meehl et al. (2004) found the global mean temperature response to forcings was approximately linearly additive while Feichter et al. (2004) and Ming and Ramaswamy (2009) found that it was not. A more recent study by Deng et al. (2020) assessed additivity of the response of greenhouse gases and aerosols in time slice experiments with CESM1. They found that for global mean temperature, the influence of these forcings were approximately linearly additive but for other features, such as autumn Arctic sea ice cover and East Asian precipitation, non-linearities did exist.

Here, we present a new single forcing large ensemble with the Community Earth System Model version 2 (CESM2). The aims of this study are two-fold: (1) to introduce this new dataset that researchers can use to further probe the impacts of individual forcings on the evolution of the Earth system according to this model and (2) to understand differences in the global mean temperature and radiative responses between this single forcing large ensemble and its predecessor (the CESM1 single forcing large ensemble (Deser et al. 2020b)). With regards to the second goal, we find substantial differences in the anthropogenic aerosol-forced global mean near surface air temperature evolution between the CESM1 and CESM2 ensembles. Three factors have the potential to contribute to this: differences in imposed aerosol emissions; differences in model physics; and, differences in the experimental design. Here, we make use of additional targeted experiments to attempt to isolate the relative role of the experimental design and, while this will be shown to have an influence, its impact is found to be sensitive to model physics and/or imposed forcings. The models, experimental design and methods are described in section 2. In section 3, we compare the global mean surface air temperature (GMST) response between the CESM2 and CESM1 single forcing large ensembles and reveal a substantial difference in the aerosol-forced response. In section 4, we then explore the influence of experimental design on the aerosol-forced GMST change in the CESM2 single forcing large ensemble and follow this with a comparison to CESM1 in section 5. Discussion and conclusions are provided in section 6.

2. Models, Experiments and Methods

a. CESM2 and its single forcing experiments

1) The Model

CESM2 is the latest generation Earth System Model developed by the U.S. National Center for Atmospheric Research in collaboration with others (Danabasoglu et al. 2020). The default configuration of CESM2, which was used to contribute experiments to CMIP6 (Eyring et al. 2016), simulates the global coupled Earth system at approximately 1° horizontal resolution. The atmospheric component is the Community Atmosphere Model version 6 (CAM6, Bogenschutz et al. 2018) with a model top at ~40 km and 32 layers in the vertical. It is coupled to the Parallel Ocean Program version 2 (POP2) ocean model (Smith et al. 2010; Danabasoglu et al. 2012), the Community Land Model version 5 (CLM5, Lawrence et al. 2019) and the Community Ice Code version 5 (CICE5, Hunke et al. 2015) and all the simulations in this study have fixed ice sheets. We refer readers to Danabasoglu et al. (2020) for more details and to the following studies for evaluation of various aspects of CESM2: Lawrence et al. (2019) for the representation of land surface processes; Simpson et al. (2020) for the large scale atmospheric circulation and its variability; Meehl et al. (2020) for the representation of monsoons; Capotondi et al. (2020) for the representation of Pacific sea surface temperature variability; and, DuVivier et al. (2020) for the representation of sea ice. In this description, we focus on the aspects of CESM2 that are of particular relevance to the single forcing large ensemble.

Within CESM2 with CAM6, atmospheric greenhouse gas concentrations are prescribed as monthly time evolving global concentrations. Aerosol forcing is introduced into the model via emissions of black carbon (BC), particulate organic matter (POM), sulfur dioxide (SO_2), sulfate (SO_4) and secondary organic aerosol precursor gas (SOAG). In CAM6, the aerosol scheme is the four-mode Modal Aerosol Module (MAM4, Liu et al. 2016). This consists of a very simple secondary organic aerosol scheme that does not include the oxidation of Volatile Organic Compounds and it is not interactively coupled to biogenic emissions (Tilmes et al. 2019). For carbonaceous aerosols, compared to its predecessor (MAM3, Liu et al. 2012), MAM4 contains an additional primary accumulation carbonaceous aerosol mode to allow for an explicit treatment of the microphysical aging of primary carbonaceous aerosols. Hydrophobic BC and POM are emitted into this fourth primary aerosol mode, where they do not activate cloud condensation nuclei (CCN) and cannot be removed by wet deposition. Over time (of the order 2-3 days) they move into the hydrophilic accumulation mode, where they are available for cloud droplet activation as CCN and also participate in wet deposition. This explicit treatment of the aging of carbonaceous aerosol in MAM4 compared to MAM3, where ageing was instantaneous, has the overall effect of increasing the lifetime and subsequent burdens of BC and POM (Liu et al. 2016, and also discussed in the Appendix). The cloud microphysics scheme is version 2 of the Morrison-Gettelman scheme (MG2, Gettelman and Morrison 2015) which, unlike its predecesor in CESM1, now includes dependence of mixed-phase immersion freezing ice nucleation on aerosols, i.e., dust aerosol acts as ice nucleating particles on which super cooled liquid water or vapor can freeze. CAM6 does not have a prognostic representation of stratospheric or tropospheric ozone or volcanic aerosol and, therefore, these forcings are prescribed.

In CLM5, each grid cell is composed of multiple land units (vegetated, lake, urban, glacier and crop) and each land unit has a specified number of columns which are then divided up into multiple patches. These patches contain a plant or crop functional type (PFT or CFT) which is prescribed through a land use time series file. Land use and land cover change can, therefore, be introduced through specified evolution of the PFTs and CFTs within each land unit and/or varying the fractional area covered by the land unit components, which allows transitions between natural vegetation, crop and glacier land units (a new feature within CLM5). The land use time series specification also determines the soil texture, wood harvest, industrial Nitrogen fertilizer application amounts and the area of the land surface equipped for irrigation. Irrigation is applied dynamically within the model to the irrigation-equiped area and is applied to achieve a target soil moisture level (Lombardozzi et al. 2020). When using biogeochemistry mode (as in the experiments here), leaf area index and canopy height are prognosed by the model. While CLM5 includes the simulation of fire internally, the current default is that this does not produce emissions that are seen by the atmosphere - biomass burning emissions to the atmosphere are prescribed from forcing datasets.

2) Forcings and Experiments

The CESM simulations used in the following analysis are summarized in Table 1. The baseline ensemble for the CESM2 single forcing large ensemble is the second set of 50 members of the CESM2 large ensemble (Rodgers et al. 2021). The CESM2 large ensemble, referred to as LENS2 hereafter, is a 100-member ensemble of simulations run under CMIP6 historical forcings between 1850 and 2014 and forcings of the Shared Socio-economic Pathway 3-7.0 (SSP3-7.0, Meinshausen et al. 2020) thereafter. A mixture of “macro” and “micro” initialization strategies were used to introduce ensemble spread where “macro” refers to initializing each model component from different years of the CESM2 pre-industrial control and “micro” refers to introducing ensemble spread through a round-off level perturbation applied to the initial atmospheric potential temperature field. The first and second set of 50 members of LENS2 are run with different biomass burning aerosol emissions over the period 1990 to 2020 (see Rodgers et al. (2021)). The first 50 members of LENS2 use the default CMIP6 biomass burning aerosol dataset, which contains higher levels of interannual variability during 1997-2014 compared to earlier and later periods due to the inclusion of satellite derived emissions. To avoid these artificial discontinuities in biomass burning aerosol variability, which has been shown to cause a rectified climate response in some regions (Fasullo et al. 2021; DeRepentigny et al. 2022), the second set of 50 members of LENS2 were run with a smoothed (11-year running mean) version of the biomass burning dataset which alters the emissions from 1990 to 2020. All of the CESM2 single forcing large ensemble experiments use the smoothed biomass burning emissions dataset, and are compared with the corresponding second 50-member set of LENS2 simulations (see supplemental Fig. 1 for a comparison of the smoothed and default CMIP6 biomass burning emissions).

Table 1. Summary of CESM experiments. The presence of a "1" or "2" in the experiment name indicates the simulation was performed with CESM1 or CESM2, respectively.

Name	Period	Model	# Members	Description
LENS2	1850-2100	CESM2	50	All forcings evolving. CMIP6 Historical and SSP3-7.0 but with smoothed BMB
AAER2	1850-2050	CESM2	15	Only anthropogenic aerosols evolving. Other forcings fixed at 1850
GHG2	1850-2050	CESM2	15	Only greenhouse gases evolving. Other forcings fixed at 1850
BMB2	1850-2050	CESM2	15	Only biomass burning emisssions evolving. Other forcings fixed at 1850
EE2	1850-2050	CESM2	15	Forcings other than AAER, GHG and BMB evolving. AAER, GHG and BMB fixed at 1850s
LENS1	1920-2100	CESM1	40	All forcings evolving. CMIP5 Historical and RCP8.5 forcings
XAAER1	1920-2080	CESM1	20	All forcings evolving except anthropogenic aerosols which are kept fixed at 1920s levels
XGHG1	1920-2080	CESM1	20	All forcings evolving except greenhouse gases which are kept fixed at 1920s levels
XBMB1	1920-2030	CESM1	15	All forcings evolving except biomass burning which is kept fixed at 1920s levels
XAAER2	1920-2050	CESM2	3	All forcings evolving except anthropogenic aerosols which are kept fixed at 1920s levels
AAER1	1850-2050	CESM1	3	Only anthropogenic aerosols evolving. Other forcings fixed at 1850

As summarized in the top portion of Table 1, the CESM2 single forcing large ensemble consists of four sub-ensembles of 15 members each that run from 1850 to 2050 following the “only” approach where *only* the forcing(s) of interest are evolving in time and others are held fixed at 1850s values: the greenhouse gas ensemble (GHG2); the anthropogenic aerosol ensemble (AAER2); the biomass burning aerosol ensemble (BMB2) and the “Everything Else” ensemble (EE2), where the number 2 in each acronym refers to CESM2. The ensemble members differ through a macro initialization, i.e., they are initialized from different years of the pre-industrial control simulation (the same years as members 1-10 and 91-95 of LENS2) to minimize any effects of ocean persistence on ensemble spread from the beginning of the run. In GHG2, only greenhouse gas concentrations are evolving in time and all other forcings are held fixed at their 1850s values. Fig. 1a shows the time-evolution of two of the important greenhouse gases (CO_2 and CH_4) and others can be seen in supplemental Fig. 2. In AAER2, only “anthropogenic aerosol” emissions are evolving in time. By “anthropogenic aerosols” here, we refer to industrial, agricultural, domestic and transport related emissions and acknowledge that this is not *all* anthropogenic emissions because it does not include anthropogenic influences on biomass burning (e.g., van Marle et al. 2017). The global emissions of three of the main aerosols or aerosol pre-cursors (SO_2, BC and SO_4) in the AAER2 ensemble are shown in Fig. 1b and others can be seen in supplemental Fig. 3. In BMB2 only biomass burning emissions are evolving in time, with three of the main emissions sources shown in Fig. 1c and others shown in supplemental Fig. 1. The transitions in variance of biomass burning emissions between the historical portion, the smoothed 1990-2020 period, and then the subsequent SSP3-7.0 projection period are quite apparent and this should be improved upon in any emissions datasets that are developed in the future. All other forcings, aside from those that are time evolving in GHG2, AAER2 and BMB2 are time evolving in the EE2 ensemble. Some of the main forcings that are evolving in this ensemble are the solar insolation (Fig. 1d black), the stratospheric volcanic aerosol (Fig. 1d red), stratospheric and tropospheric ozone (Fig. 1e for Southern Hemisphere (SH) stratospheric ozone) and land use and land cover change (Fig. 1f). The stratospheric volcanic aerosol and tropospheric and stratospheric ozone concentrations are derived from the average of a three-member ensemble of simulations with the Whole Atmosphere Community Climate Model version 6 (WACCM6, Gettelman et al. 2019), also run under historical and SSP3-7.0 forcings. This choice was made for all CESM2 experiments to limit forcing differences between CESM2 and CESM2-WACCM and, given the similarity between CESM2 and CESM2-WACCM in the troposphere and lower stratosphere, the CESM2-WACCM ozone and volcanic aerosol fields will be more consistent with the model dynamics and atmospheric structure than the CMIP6 forcing datasets. Overall, each forcing is time evolving in one of these sub-ensembles, allowing the additivity of forcing contributions to be tested.

The CESM2 pre-industrial control simulation is also used to examine the behavior of the Atlantic Meridional Overturning Circulation (AMOC). This simulation is run for 2000 years under forcings that are representative of 1850s conditions, following the CMIP6 protocol and we make use of simulation years 400 to 2000.

b. CESM1 and its single forcing experiments

1) The Model

CESM1 is the previous generation of CESM released in 2010 (Hurrell et al. 2013). It has been widely used, including through the CESM1 large ensemble (Kay et al. 2015) and the CESM1 single forcing large ensemble (Deser et al. 2020b). Between CESM1 and CESM2, major developments were undertaken in the atmosphere and land components, in particular. The atmospheric component of CESM1 is CAM5 and this contains the older aerosol scheme, MAM3, which, as discussed above, does not allow for the explicit treatment of the aging of primary carbonaceous aerosols although

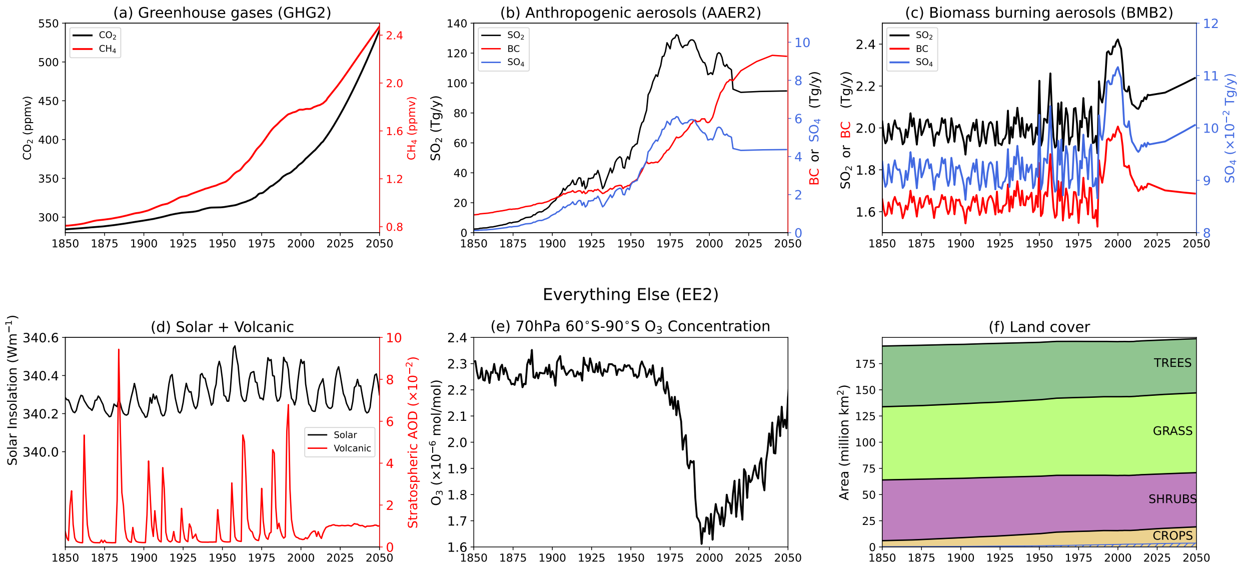

FIG. 1. Annual means of selected forcings and their evolution in the single forcing ensembles. (a) Global CO_2 (left axis) and CH_4 (right axis) concentrations as they evolve in the GHG2 ensemble. (b) Global emissions of SO_2 (left axis) and BC and SO_4 (right axis) as they evolve in the AAER2 ensemble. (c) Global emissions of SO_2 and BC (left axis) and SO_4 (right axis) as they evolve in the BMB2 ensemble. (d) - (f) show various forcings that evolve in the "Everything Else" ensemble: (d) Solar insolation (left axis) and stratospheric Aerosol Optical Depth (AOD) from the WACCM6 simulations that produced the volcanic aerosol forcing (right axis); (e) 70hPa ozone concentration averaged over 60°S to 90°S; and (f) surface area covered by various land cover types (blue hatching shows the irrigated land surface area) and note that bare ground is not shown.

the representation of secondary organic aerosols is the same as in MAM4. The cloud microphysics scheme is the first version of the Morrison-Gettelman scheme (MG1, Morrison and Gettelman 2008) which does not relate mixed-phase imersion freezing ice nucleation to aerosols. In CAM6, many other atmospheric parameterizations underwent development compared to CAM5. For example, CAM5 uses an older generation parameterization of shallow convection and boundary layer turbulence compared to CAM6 and it has a simpler representation of orographic drag.

The land component of CESM1 is CLM4 (Lawrence et al. 2011). As with the atmosphere, the CLM5 model component of CESM2 contains major developments compared to this older model version. Major updates were performed on the representation of soil and plant hydrology, snow density, river modeling, carbon and nitrogen cycling and crop modeling. These updates are, in general, found to lead to improvements in the representation of many land surface processes in the newer generation of the model (Lawrence et al. 2019) with impacts on the representation of climate variability (e.g., Simpson et al. 2022). Of relevance for the prescription of time evolving forcings, CLM4 does not allow for the time-evolution of the fractional area of each grid point covered by different land cover types, e.g., it does not allow for transitions in the weighting of natural vegetation versus crops within a grid cell, just time-evolution of the plant functional types within a given land cover type. It has a much simpler representation of crops and it does not have a representation of irrigation.

2) Forcings and Experiments

Forcings within CESM1 are prescribed in a similar manner to CESM2. Greenhouse gases are represented by prescribed global surface concentrations and aerosols are introduced through emission sources. Solar variability is introduced through variations in the total solar irradiance, time evolving volcanic aerosols and ozone concentrations are prescribed and land use and land cover change is introduced via time evolving plant functional types. CMIP5 era forcings are used in the CESM1 experiments. Historical forcings are used prior to 2005 and forcings of the Representative Concentration Pathway 8.5 (RCP8.5) are used thereafter (Meinshausen et al. 2011; Lamarque et al. 2011). Much like for CESM2, the ozone forcing in the CESM1 experiments is prescribed based on WACCM simulations (in this case using WACCM4 and a two member ensemble with a 10-year running mean applied to each month of ozone forcing separately). In CESM1 the volcanic aerosol forcing does not come from WACCM, but rather CMIP5 forcing is used.

As summarized in the second portion of Table 1, the baseline ensemble for CESM1 is the CESM1 large ensemble (Kay et al. 2015), referred to as LENS1 hereafter. This is a 40-member ensemble that runs from 1920 to 2100 in which all forcings are evolving, with each member differing through a round-off level perturbation introduced to the atmospheric potential temperature field at initialization (micro initialization). The CESM1 single forcing large ensemble (Deser et al. 2020b) consists of 3 sub-ensembles of 15 or 20 members that use the "all-but-one" method where all forcings are time evolving except the forcing of interest, which is held fixed at the values for 1920. Ensemble spread is also introduced through micro initialization and we refer to these sub-ensembles as XFORCING where X denotes that FORCING is held fixed. In the XGHG1 ensemble all forcings are evolving except greenhouse gases, in the XAAER1 simulation all forcings except anthropogenic aerosols are evolving (again, by anthropogenic aerosols here, we do not include anthropogenic influences on biomass burning) and, in the XBMB1 ensemble all forcings except biomass burning are evolving. The number 1 here denotes that these are CESM1 simulations. The time-evolution of the biomass burning emissions, greenhouse gas concentrations and anthropogenic aerosol emissions in CESM1 can be compared with those in CESM2 in supplementary Figs. 1 to 3. Note that a fourth ensemble was originally included in which all forcings except land use and land cover change were evolving but this dataset has since been retracted due to an error. The XGHG1 and XAAER1 ensembles extend from 1920 to 2080 and the XBMB1 ensemble extends from 1920 to 2030. Given that the "all-but-one"approach is used, the influence of a given forcing must be determined by taking the difference between LENS1 and the XFORCING ensemble and because not all forcings are represented within an XFORCING ensemble, a complete test of additivity cannot be performed.

Years 400 to 2200 of the CESM1 pre-industrial control simulation are also used. This simulation was run under forcings that are representative of 1850s conditions, following the CMIP5 protocol.

c. Experimental design sensitivity tests

As will be shown, substantial differences in the inferred response to aerosol forcing are found between CESM1 and CESM2. To test the influence of the method used, i.e., "only" versus "all-but-one", we perform 3-member ensembles of an "only" anthropogenic aerosol experiment with CESM1, referred to as AAER1 and an "all-but-one" anthropogenic aerosol experiment with CESM2, referred to as XAAER2, as summarized in the bottom portion of Table 1. In AAER1, CESM1 is run with time evolving anthropogenic aerosol forcing from 1850 to 2050 with all other forcings held fixed at 1850s values and members differ through micro initialization. In XAAER2, CESM2 is run from 1920 to 2050 with all forcings evolving except anthropogenic aerosols which are held fixed at those of 1920 and the members differ via macro initialization.

d. Methods

We focus on the period 1920 to 2050, which is common to the majority of CESM simulations (Table 1), and, unless otherwise stated, we consider ensemble mean anomalies from the 1920-1940 average, which is at the beginning of the CESM1 single forcing large ensemble. For CESM1 the influence of a given forcing at a given time period is therefore given by the difference between LENS1−XFORCING at that time period and LENS1−XFORCING averaged over 1920 to 1940. For CESM2, the influence is simply given by the difference between that time period and the average of 1920 to 1940 for the single forcing ensemble.

To quantify uncertainty and statistical significance a bootstrapping approach is used whereby, within each ensemble, members are randomly sampled with replacement and a new ensemble mean calculated. This is repeated 1000 times and the uncertainty range on anomalies for that ensemble is given by the 2.5th to 97.5th percentile range of those bootstrapped ensemble means. For the 3-member sensitivity tests, rather than performing bootstrapping on the 3-member ensembles, we either compare them to uncertainty ranges on the larger ensembles that are calculated by subsampling 3 members with replacement or we estimate uncertainty and significance levels by bootstrapping equivalent sample sizes from the respective pre-industrial control, under the assumption that the internal variability of the 1850s climate is representative of that throughout the simulation.

3. Global mean surface air temperature evolution: CESM2 versus CESM1

We begin by comparing the evolution of global mean near surface (2-m) air temperature ($[T_s]$, and we will use $[x]$ throughout to denote the global mean of variable x) between the CESM2 single forcing large ensemble and CESM1 single forcing large ensemble in Fig. 2. Recall that the CESM1 single forcing large ensemble does not have all forcing contributions represented, as the equivalent of the "Everything Else" ensemble was not performed. We, therefore, estimate the CESM1 "Everything Else" contribution as the residual LENS1−((LENS1−XGHG1) + (LENS1−XAAER1) + (LENS1−XBMB1)) for comparison with EE2. This can only be done out to 2030 when the XBMB1 simulation ends and assumes linearity, which may not be valid.

Figure 2a shows the time-evolution of $[T_s]$ for LENS2 (black) and the contributions that are inferred to be due to the different forcing components. This can be compared with the equivalent for CESM1 in Fig. 2b. In both CESM1 and CESM2 greenhouse gases (red) act to increase $[T_s]$, while anthropogenic aerosols (blue) act to decrease it. The role of biomass burning aerosols (brown) in $[T_s]$ evolution is fairly minimal, while "Everything Else" (green) acts to cool the planet during the major volcanic eruptions of the 20th Century (e.g., El Chichon in the early 1960s and Pinatubo in the early 1990s) and to warm the planet relative to 1920-1940 throughout the first half of the 21st Century (most apparent in CESM2 where the EE contribution can be examined beyond 2030). Exactly what is producing this warming warrants further investigation but it is potentially related to the lack of large volcanic eruptions in the projected future forcings (Fig. 1d, red).

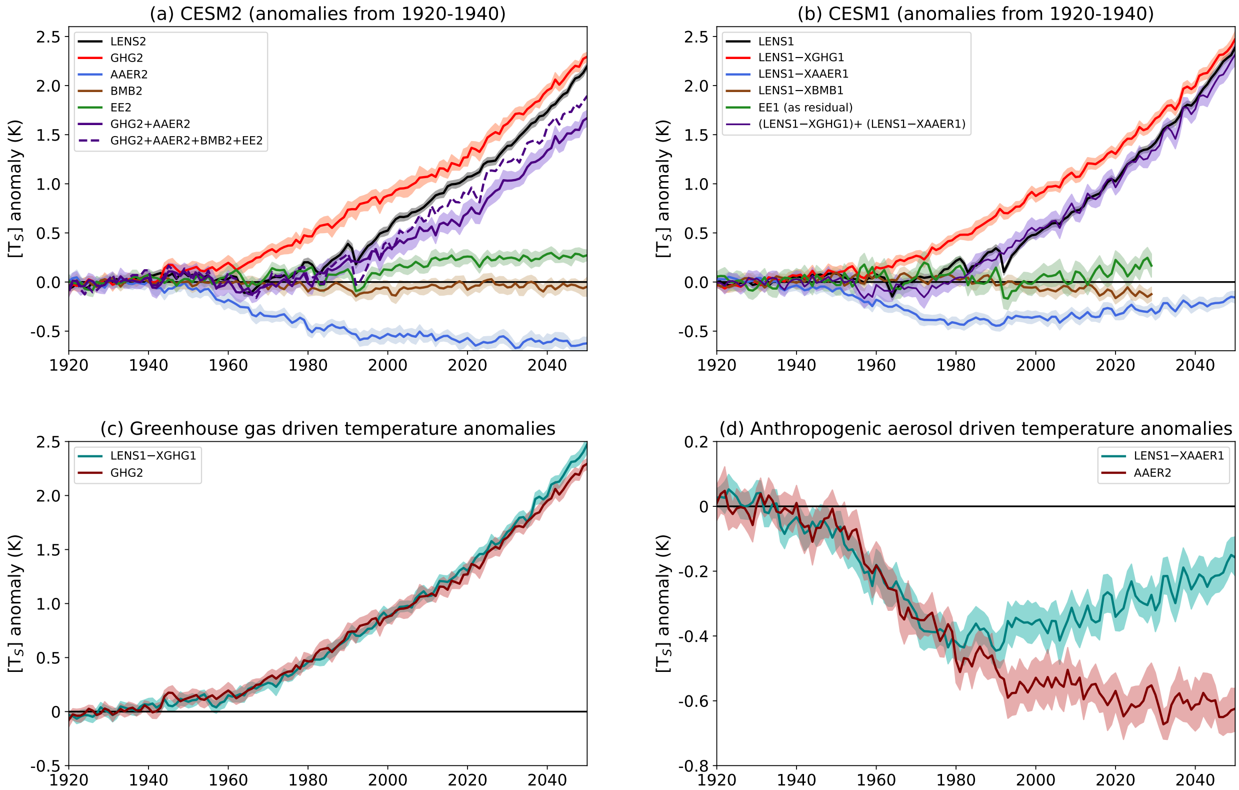

FIG. 2. Time evolution of global mean surface air temperature anomalies relative to the 1920 to 1940 average for (a) CESM2 and (b) CESM1 and their respective single forcing large ensembles. For CESM1, the equivalent of "Everything Else" has been estimated as a residual (green) of the difference between LENS1 and the sum of the greenhouse gas, anthropogenic aerosol and biomass burning aerosol contributions. The solid purple line shows the sum of the greenhouse gas and anthropogenic aerosol contributions and in panel (a) the dashed purple line shows the sum of all 4 components (GHG2 + AAER2 + BMB2 + EE2). (c) Reproduces the greenhouse gas-forced anomalies for CESM1 and CESM2 for a more direct comparison while (d) is the same but for the anthropogenic aerosol contribution. The shading uncertainty range is a 95% confidence interval on the ensemble mean calculated by bootstrapping members with replacement.

In the CESM1 single forcing large ensemble, the capacity for exploring additivity is limited, but we can see that the greenhouse gas contribution inferred from LENS1−XGHG1 and the anthropogenic aerosol contribution inferred from LENS1−XAAER1 approximately add up to the overall LENS1 [T_s] anomalies (compare black and purple in Fig. 2b). This, however, is not true in CESM2 (compare black and solid purple in Fig. 2a) where the sum of the [T_s] anomalies in GHG2 and AAER2 fall short of the LENS2 [T_s] anomalies from the late 20th century onwards. Adding in the contributions from BMB2 and EE2 brings the sum a little closer to LENS2 (dashed purple in Fig. 2a) but a discrepancy still exists. It is clear from comparison of the relation between the solid purple and black lines in Figs. 2a and b that the sum of the greenhouse gas and anthropogenic aerosol contributions and how that relates to the all forcing signal differs considerably between CESM1 and CESM2.

The prescribed GHG concentrations are rather similar between CESM1 and CESM2 over the period shown (supplemental Fig. 2) and a closer comparison of the GHG-forced signals between CESM1 and CESM2 (Fig. 2c) reveals that the GHG-forced [T_s] anomalies are also comparable between CESM1 and CESM2. Note that this is not true locally as GHG2 warms more than LENS1−XGHG1 in the low latitudes but warms much less in the Northern Hemisphere (NH) high latitudes - a feature that will be discussed further in section 5 . CESM1 and CESM2 differ considerably in the global mean [T_s] anomalies due to anthropogenic aerosol forcing (Fig. 2d). Anthropogenic aerosols continue to cool the planet out to 2050 in CESM2 while the anthropogenic aerosol induced cooling in CESM1 maximizes in the 1980s and then declines, such that by 2050, the global mean anthropogenic aerosol-forced [T_s] anomalies differ by about 0.4K between CESM1 and CESM2. This difference in anthropogenic aerosol-forced [T_s] change could be due to differences in the experimental design, differences in the aerosol emissions, differences in the model physics or some combination of these, as explored in the following sections.

4. The impact of the single forcing method on the aerosol-forced response

a. Global mean temperature and radiative fluxes

A major difference between the CESM1 and CESM2 single forcing large ensembles is the experimental design. In CESM1, an "all-but-one" approach was used while in CESM2 an "only" approach was used. To test the influence of this experimental design, we consider the additional three-member ensembles: XAAER2, where the anthropogenic aerosol simulation of CESM2 was performed in the same way as with CESM1; and, AAER1, where the anthropogenic aerosol simulation of CESM1 was performed in the same way as CESM2.

Time series of 21-year running mean anomalies (chosen as a reasonable balance between reducing noise, while retaining features of the time-evolution) of various global mean quantities are shown in Fig. 3 for each of the methods for both CESM1 and CESM2. First, it is worth noting the substantial differences in the global mean AOD between the CMIP5 and the CMIP6 forcings (Fig. 3a). This is primarily due to the differences in emissions but there is also a contribution from the enhanced lifetime of black carbon in CESM2 (see the Appendix). [AOD] continues to rise out to

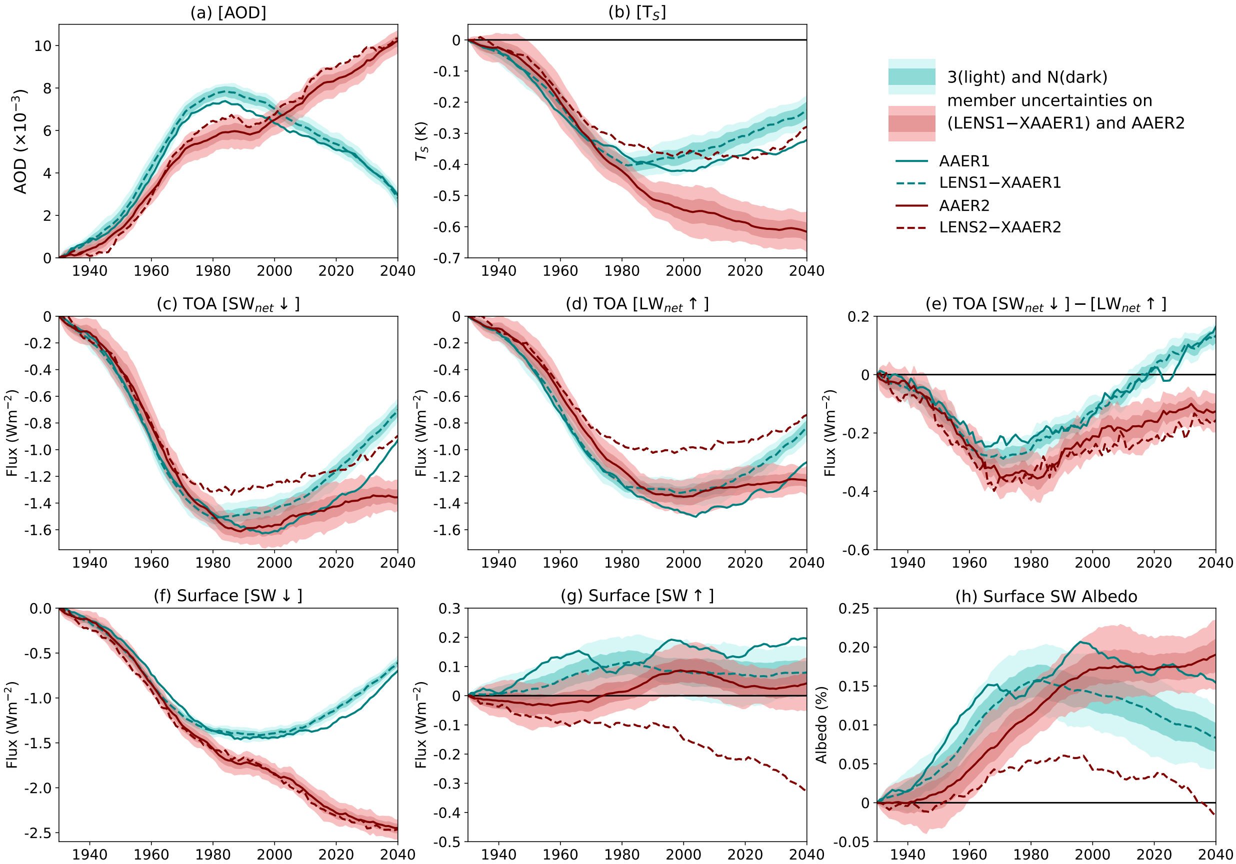

FIG. 3. Global mean centered 21-year running means of annual means of anomalies relative to 1920 to 1940 of various fields from AAER1 (teal solid), LENS1−XAAER1 (teal dashed), AAER2 (maroon solid) and LENS2−XAAER2 (maroon dashed). Shaded ranges are shown around the experiments with the large ensembles (LENS1−XAAER1 and AAER2) with the light component showing the uncertainty on a 3-member mean of XAAER1 or AAER2 and the dark component showing the uncertainty for an N-member mean of XAAER1 or AAER2 where N is the number of members in the anthropogenic aerosol single forcing ensemble. (a) Aerosol optical depth at 500nm, (b) near surface air temperature, (c) top of atmosphere (TOA) net downward shortwave flux, (d) top of atmosphere net upward longwave flux, (e) top of atmosphere net downward radiative flux, (f) surface downward shortwave flux, (g) surface upward shortwave flux, (h) albedo.

2050 in CESM2 but declines after about 1980 in CESM1 and the experimental design does not substantially impact the [AOD] evolution (compare solid and dashed in Fig. 3a). A comparison of the burdens of different aerosol species in Fig. A1 and supplemental Fig. 4 indicates, perhaps unsurprisingly, that the difference in [AOD] trends between AAER1 and AAER2 is dominated by the differing trends in anthropogenic aerosols as opposed to sea salt or dust which can respond as the climate changes under anthropogenic aerosol forcing.

It is clear from the $[T_s]$ time series in Fig. 3b that the experimental design has a substantial impact on the inferred anthropogenic aerosol influence on $[T_s]$ in CESM2. When the aerosol forcing is imposed in isolation (solid maroon in Fig. 3b), much colder temperature anomalies are reached than when the aerosol influence is inferred from LENS2−XAAER2 (dashed maroon in Fig. 3b), with the difference in the aerosol-forced cooling being ~0.3K by 2030-2050. Note that the difference between LENS2−XAAER2 and AAER2 is much greater than would be expected due to sampling uncertainty alone (compare dashed maroon with the light uncertainty shading on the solid maroon in Fig. 3b). The result is that when the anthropogenic aerosol influence is inferred from LENS2−XAAER2, the overall cooling is more comparable to the cooling found in CESM1 (compare maroon dashed with teal in Fig. 3b). This indicates a strong sensitivity of the inferred aerosol cooling to the experimental design ("all-but-one" versus "only") in CESM2. In CESM1, the AAER1 simulation (solid teal) is also significantly cooler than LENS1−XAAER1 (dashed teal) toward the end of the simulation, but the impact of the experimental design in CESM1 is relatively minor compared to that found in CESM2. We will revisit this difference between CESM1 and CESM2 in section 5 and for now focus on the dependence on the experimental design within CESM2.

In Fig. 4a it can be assessed how $[T_s]$ varies as a function of [AOD]. Here, we only show the variations over the time period when [AOD] is increasing, which means for CESM1, we are showing out to the 21-year mean centered on 1984, while for CESM2 we are showing out to the end of the simulation. Fig. 4a shows that up to [AOD] anomalies of about 8×10^{-3} (the maximum in CESM1) the evolution of $[T_s]$ as a function of [AOD] is rather similar in each of AAER1, LENS1−XAAER1 and LENS2−XAAER2. In contrast, AAER2 cools a lot more at a given [AOD] for [AOD] greater than around 4×10^{-3}. In addition, the evolution of $[T_s]$ as a function of [AOD] is non-linear, particularly in LENS2−XAAER2. In the 21st century, the cooling in LENS2−XAAER2 levels off and then turns around and the planet starts to warm even while the [AOD] continues to increase (Fig. 3a versus b and Fig. 4a).

To begin to understand the difference between AAER2 and LENS2−XAAER2, consider the top of atmosphere (TOA, although actually here the fluxes used are at the model top) radiative fluxes and their imbalance shown in Figs. 3c-e. Fig. 3e shows the difference between the net TOA downward shortwave radiation ($[SW_{net} \downarrow]$) and the net TOA upward longwave radiation ($[LW_{net} \uparrow]$), i.e., the TOA radiative imbalance. Throughout we refer to fields that are positive when downward with the down arrow (↓) and fields that are positive when upward with the up arrow (↑). The way in which the overall TOA radiative imbalance evolves is similar between the methods (compare

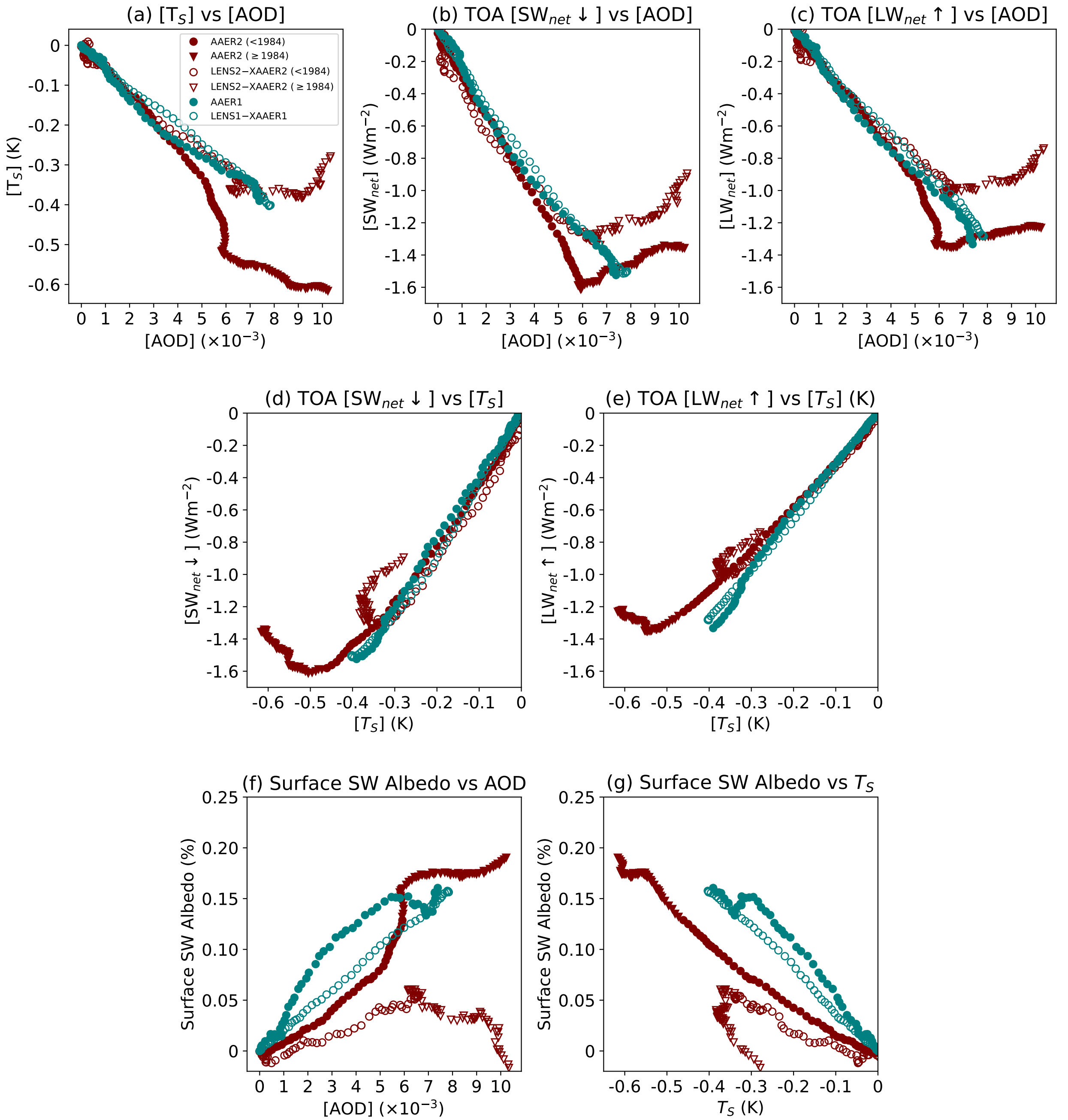

FIG. 4. Global mean 21-year running means of annual means. Both CESM1 and CESM2 are only shown for the period over which [AOD] is continuing to increase, which means the full record is shown for CESM2, but only up to the 21-year mean centered on 1984 is shown for CESM1. The CESM2 points transition from circles to triangles after 1984. (a)-(c) show [T_s], TOA net downward shortwave and TOA net upward longwave, respectively, versus [AOD]. (d) and (e) show TOA net downward shortwave and TOA net upward longwave versus [T_s] and (f) and (g) show surface shortwave albedo versus (f) [AOD] and (g) [T_s].

solid and dashed in Fig. 3e). However, the evolution of the separate components [$SW_{net} \downarrow$] (Fig. 3c) and [$LW_{net} \uparrow$] (Fig. 3d) is not - they reveal that this same TOA radiative imbalance is achieved for rather different reasons.

Until about 1980, aerosol forcing causes TOA [$SW_{net} \downarrow$] to decline as aerosols and associated cloud changes reflect more shortwave radiation back to space and as surface albedo increases. In AAER2, after about 1970, there is a greater decline in TOA [$SW_{net} \downarrow$] than in LENS2−XAAER2, but this difference is opposed by a greater decline in [$LW_{net} \uparrow$], such that the overall TOA radiative imbalance ends up roughly the same in each case. The greater decline in [$LW_{net} \uparrow$] in AAER2 can be attributed to the greater decline in [T_s]; a colder planet emits less longwave radiation to space. Indeed, Fig. 4e shows that [$LW_{net} \uparrow$] depends on [T_s] in a rather similar way in AAER2 and in LENS2−XAAER2 but AAER2 cools further and, in association with this, [$LW_{net} \uparrow$] declines more. So, overall, while the TOA radiative imbalance evolves in a similar way in AAER2 and in LENS2−XAAER2, it reflects different quasi-equilibria with differing balances between TOA [$LW_{net} \uparrow$] and TOA [$SW_{net} \downarrow$][1]. Overall, we infer that the reason AAER2 cools more than LENS2−XAAER2 lies in the behavior of the shortwave radiation. TOA [$SW_{net} \downarrow$] declines more in AAER2 and this is balanced by a greater cooling and greater reduction in TOA [$LW_{net} \uparrow$].

The greater decline in TOA [$SW_{net} \downarrow$] in AAER2 compared to LENS2−XAAER2 must arise from a difference in the shortwave radiation being reflected back to space, either from within the atmosphere or from the surface. Fig. 3 demonstrates that it is the difference reflected back from the surface that is key to the TOA [$SW_{net} \downarrow$] differences. If the key were differences in the extent to which shortwave radiation is reflected back to space from within the atmosphere, either by aerosols themselves or the associated cloud radiative effects, then we would expect to see differences in the surface downward shortwave radiation ([$SW \downarrow$]), but Fig. 3f makes it clear that the anomalies in [$SW \downarrow$] are very similar between AAER2 and LENS2−XAAER2. The decline in surface [$SW \downarrow$] is independent of which method is used in both CESM1 and CESM2, although the response differs substantially between them in association with their differing aerosol forcings. In contrast, Fig. 3g demonstrates a clear difference in the surface upward shortwave ([$SW \uparrow$]) between AAER2 and LENS2−XAAER2 in association with a difference in their surface shortwave albedo responses (Fig. 3h). In AAER2, the surface shortwave albedo increases much more than in LENS2−XAAER2. As a result, surface [$SW \uparrow$] stays roughly constant in AAER2 (Fig. 3g), even

[1]An aside is that in Fig. 3d, LENS2−XAAER2 is the odd one out, with AAER1, LENS1−XAAER1 and AAER2 all exhibiting similar changes in [$LW_{net} \uparrow$]. This may appear at odds with the fact that it is AAER2 that exhibits a different temperature response (Fig. 3b). The reason why AAER1 and LENS1−XAAER1 exhibit a greater decline in TOA $LW_{net} \uparrow$ than LENS2−XAAER2, even though their temperature responses are similar is actually because of cloud longwave radiative effects. Examination of clear sky $LW_{net} \uparrow$ (supplemental Fig. 5) reveals what we expect: AAER1, LENS1−AAER1 and LENS2−XAAER2 which all cool less than AAER2, also exhibit a smaller decline in clear sky [$LW_{net} \uparrow$].

as surface [$SW \downarrow$] declines, because a larger proportion is being reflected back to the atmosphere, and ultimately to space.

The difference in surface [$SW \uparrow$] between AAER2 and LENS2−XAAER2 (solid in Fig. 5a) explains most of the difference in TOA clear sky (dashed in Fig. 5a) and TOA all sky [$SW_{net} \uparrow$] (dotted in Fig. 5a). Consideration of how the surface shortwave albedo varies as a function of [AOD] (Fig. 4f) and [T_s] (Fig. 4g) reveals that there is a systematic difference between the "all-but-one" and "only" approach in both CESM1 and CESM2. For [T_s] anomalies down to around -0.3K and [AOD] anomalies up to about $5{\times}10^{-3}$ the albedo increases more for the "only" approach than for the "all-but-one" approach. But then, beyond that, the difference in behavior of the albedo between AAER2 and LENS2−XAAER2 increases rather dramatically. In the late 20th century, the surface shortwave albedo continues to increase in AAER2 (Fig. 3h), but in LENS2−XAAER2 the albedo increase levels off at a much lower value and then starts to decline. This is apparent as a rather dramatic difference between AAER2 and LENS2−XAAER2 in the relationship between surface shortwave albedo and both [AOD] and [T_s] (Figs. 4 f and g). In AAER2, as the planet cools, surface shortwave albedo keeps on increasing, presumably providing a positive feedback onto the cooling (Fig. 4g). In LENS2−XAAER2, as the planet cools, albedo also increases, but to a lesser extent, and then in the 1990s, albedo starts to decline and the planet begins to warm up again, even though the [AOD] has continued to increase.

Increased surface shortwave albedo generates cooler temperatures and vice versa, so separating out cause and effect is challenging in these quasi-equilibrium experiments where the system has adjusted to a new balance. Nevertheless, given that there is no evidence that the origins of the different [T_s] response between AAER2 and LENS2−XAAER2 lies in differences in how the incoming surface shortwave radiation behaves, we posit that it lies in non-linearity in surface shortwave albedo feedbacks and that there are two components that contribute to this shortwave albedo non-linearity, which will now be discussed: (1) a non-linearity in both snow and sea ice albedo feedbacks and; (2) a non-linearity related to the North Atlantic ocean circulation which leads to differing northward heat transport into the Arctic and associated differences in high latitude albedo. In the following sub-sections we provide the evidence for both of these sources of non-linearity in surface shortwave albedo.

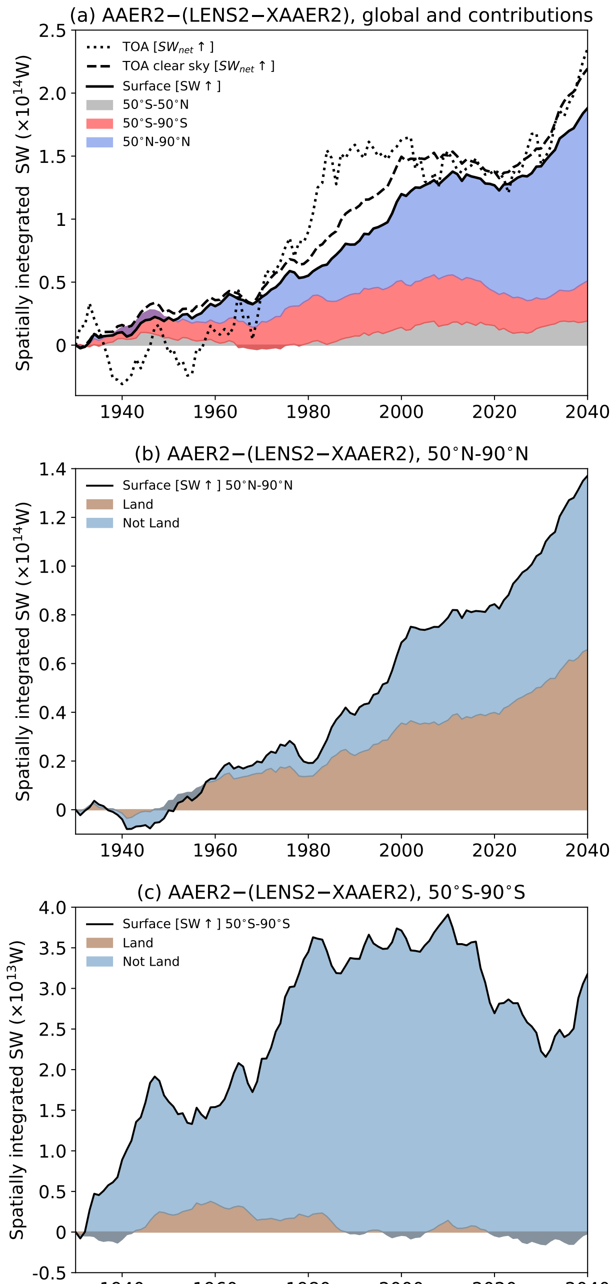

FIG. 5. Difference between AAER2 and LENS2−XAAER2 in the spatially integrated shortwave fluxes. (a) shows the global integral with black dotted showing the top of atmosphere net upward shortwave radiation and black dashed showing its clear sky component. Black solid shows the surface upward shortwave radiation and it is further divided into contributions from different latitude bands. (b) Surface upward shortwave radiation spatially integrated from 50°N to 90°N and the contributions from land regions and regions that are not land. (c) is as (b) but for 50°S to 90°S. Note the differing magnitudes covered by the y axes in (b) and (c).

b. Albedo non-linearities

1) Snow and sea ice non-linearities

Figure 5a shows the spatially integrated difference in the surface upward shortwave radiation, [$SW \uparrow$] between AAER2 and LENS2−XAAER2 (solid black) and this difference is further decomposed into the contribution from different latitude bands. The high latitudes (poleward of 50° latitude) clearly dominate in this difference. Both SH (salmon) and NH (blue) play a role, but the NH difference in surface $SW\uparrow$ dominates in the late 20th and 21st centuries. This difference in surface $SW\uparrow$ between AAER2 and LENS2−XAAER2 in the high latitudes of each hemisphere can be further decomposed into the contribution from land regions and the contribution from "not land" regions, i.e., regions that are either ocean or sea ice. In the NH, the land and not land contribute roughly equally to the difference between AAER2 and LENS2−XAAER2 from the 2000s onward (Fig. 5b) and the difference over land regions dominates prior to that. In the SH, the difference over ocean and sea ice regions dominates (Fig. 5c).

A seasonal decomposition of the difference in surface $SW\uparrow$ integrated over 50°N to 90°N between AAER2 and LENS2−XAAER2 indicates the summer season as dominating in the $SW\uparrow$ difference initially (supplemental Fig. 6), which makes sense given that this is when there is the greatest incident shortwave radiation which can then be affected by differences in surface shortwave albedo feedbacks. Later in the simulation, differences in $SW\uparrow$ between the methods becomes increasingly important in the shoulder seasons as well. In order to understand the origins of this difference in surface shortwave albedo behavior we now focus on the NH during summer (June-July-August, JJA).

Consider the time series of 21-year running mean JJA 50°N-90°N T_s shown in Fig. 6a. This shows that NH high latitude temperature declines at a similar rate in AAER2 and LENS2−XAAER2 until around 1960-1980, at which point LENS2−XAAER2 starts to warm, while AAER2 does not. We consider the behavior of surface SW↑, snow cover and sea ice cover during the 1960-1980 average in an attempt to examine their differences before subsequent feedbacks associated with the differing T_s response are present. For 50°N-90°N average surface $SW\uparrow$, a difference between AAER2 and LENS2−XAAER2 is already apparent during 1960-1980, even though a difference in T_s is not (see Fig. 6b-d for local temperature changes). It is clear from Figs. 6 f-h that in AAER2, there is a greater enhancement in surface $SW\uparrow$ around the sea ice edge and over high latitude land regions than in LENS2−XAAER2. There is also less of an increase in surface $SW\uparrow$ in the interior regions of the sea ice in AAER2 compared to LENS2−XAAER2, leading to a difference in $SW\uparrow$ between the methods that is negative there. These differences in SW↑ correspond reasonably well to differences in sea ice cover between AAER2 and LENS2−XAAER2 (Fig. 6n-p).

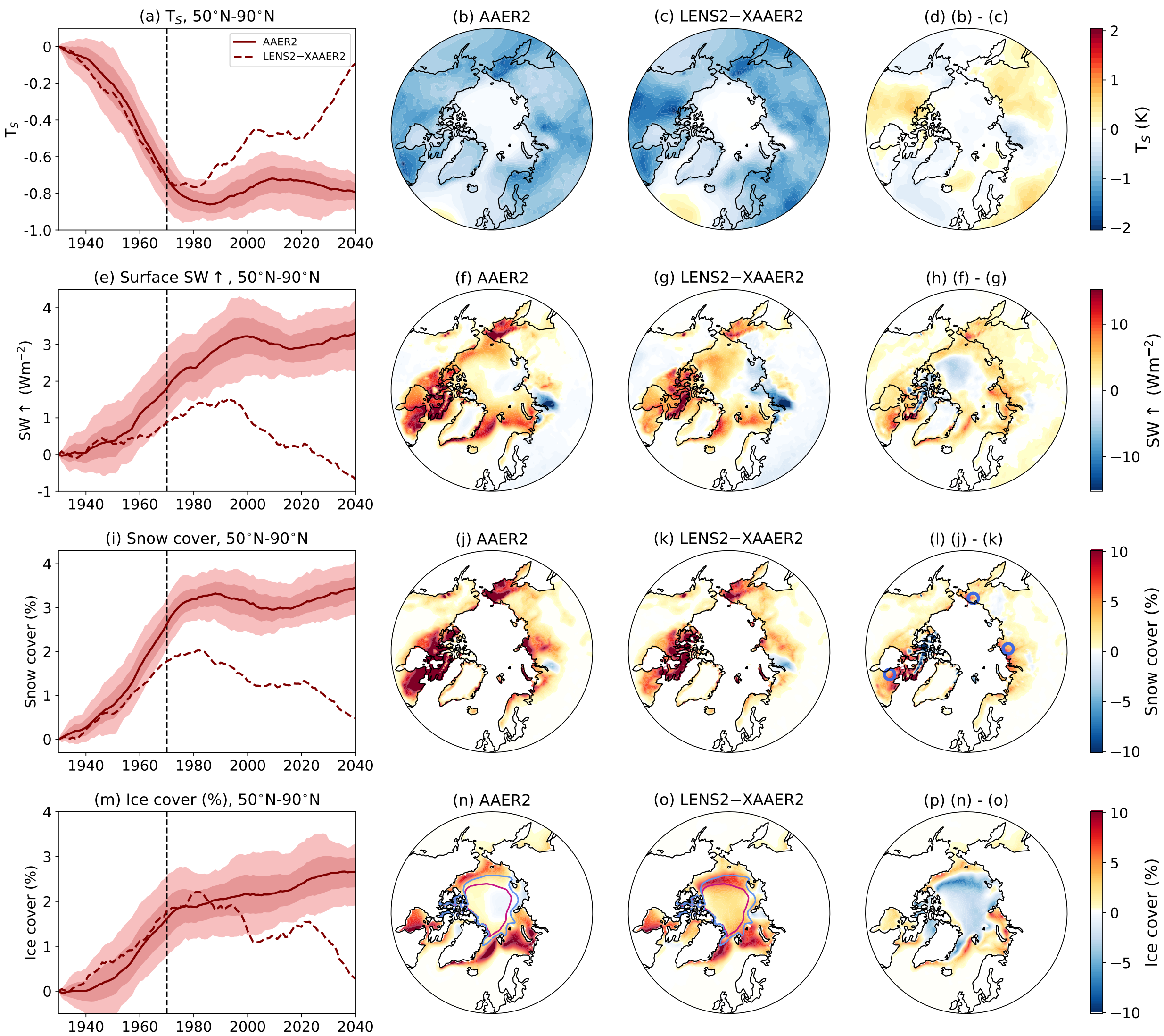

FIG. 6. Aerosol influence on the JJA season from 50°N to 90°N. Left column shows time series of 21-year running means for (solid) the ensemble mean of AAER2 and (dashed) the ensemble mean of LENS2-XAAER2 anomalies from the 1920-1940 average. The dark and light shadings around the AAER2 line show the 95% confidence interval using (light shading) 3 members and (dark shading) 15 members. The dashed vertical line depicts the 21-year mean centered on 1970, i.e., the 1960-1980 average shown in the right three columns. (2nd column) shows 1960-1980 anomalies compared to 1920-1940 for AAER2, (3rd column) is as (2nd column) but for LENS2−XAAER2 and (4th column) shows the difference in the anomalies between AAER2 and LENS2−XAAER2. (top) T_s, (2nd row) surface upward shortwave, (3rd row) grid cell area covered by snow in percent, (4th row) grid cell area covered in sea ice in percent. In (i) and (m) the 50°N-90°N average is taken only over land grid points and grid points that are not land, respectively. Blue points in (l) show the grid points used for the analysis in Fig. 7 ((286°E,58°N), (173°E,68°N), (95°E,73°N)). The blue and pink contours in (n) and (o) show the 80% sea ice contours for 1920-1940 of AAER2 and 1960-1980 of XAAER2, respectively.

Over the NH land regions surrounding the Arctic, there is a greater increase in summer snow cover in AAER2 than in LENS2−XAAER2 (Fig. 6l) corresponding well to regions where the difference in surface $SW\uparrow$ is largest over land. Figure 7 demonstrates that snow cover fraction depends non-linearly on local temperature at three representative locations, given by the blue circles in Fig. 6l: a grid point to the east of Hudson Bay; one in Eastern Siberia; and one to the south of the Kara Sea. Specifically, a cooling/warming that occurs at a lower temperature would be associated with a larger increase/decrease in snow cover than if that same cooling/warming occured at a higher temperature. This can be understood as a result of snow cover being bounded by zero. For warmer temperatures, there is a higher probability of there not being snow on the ground (green stars in Fig. 7) and for times when there is no snow on the ground a further warming can no longer lead to a change in snow cover, leading to a weaker dependence of seasonal average snow cover on temperature at warmer temperatures. The arrows in Fig. 7 help to illustrate this non-linearity by showing the 1920-1940 average at the start point of the arrow and the 1960-1980 average at the end point of the arrow. For the change that is inferred to be due to the aerosol forcing, the cooling in AAER2 which has a colder starting point, leads to a proportionately larger increase in snow cover than the cooling in LENS2 does and both lead to a proportionately larger change in snow cover than the warming in XAAER2, which is warming rather than cooling. The result is that the magnitude of the increase in snow cover in AAER2 (given by the length of the blue arrow) is proportionately larger for the temperature change than that in LENS2−XAAER2 (given by the sum of the black and the pink arrow lengths). This effect is likely what dominates prior to 1980 in the differences seen between the methods in the NH, given the dominance of land regions in contributing to the SW$\uparrow$ differences (Fig. 5b).

For sea ice, both AAER2 and LENS2−XAAER2 exhibit an increase in sea ice cover, but they do so at different locations - the increase in AAER2 (Fig. 6n) is generally at lower latitudes than the increase in LENS2−XAAER2 (Fig. 6o). The reason for this is fairly straightforward - in AAER2, the cooling is occurring relative to a cold climate (a pre-industrial climate which has then cooled slighly under aerosol forcing out to the baseline 1920 to 1940 period), while in LENS2 compared to XAAER2, the aerosol influence is felt relative to a climate in which sea ice has been influenced by greenhouse gas driven warming. To illustrate the differing sea ice fractions between the different baseline climates that the aerosol influence is being compared against in Fig. 6, we show the 80% sea ice fraction contours for the 1920-1940 climate of AAER2 in blue and for the 1960 to 1980 climate of XAAER2 in pink in panels n and o. The 1920-1940 climate is the baseline for AAER2. For the "all-but-one" method, the baseline climate is more complicated but since

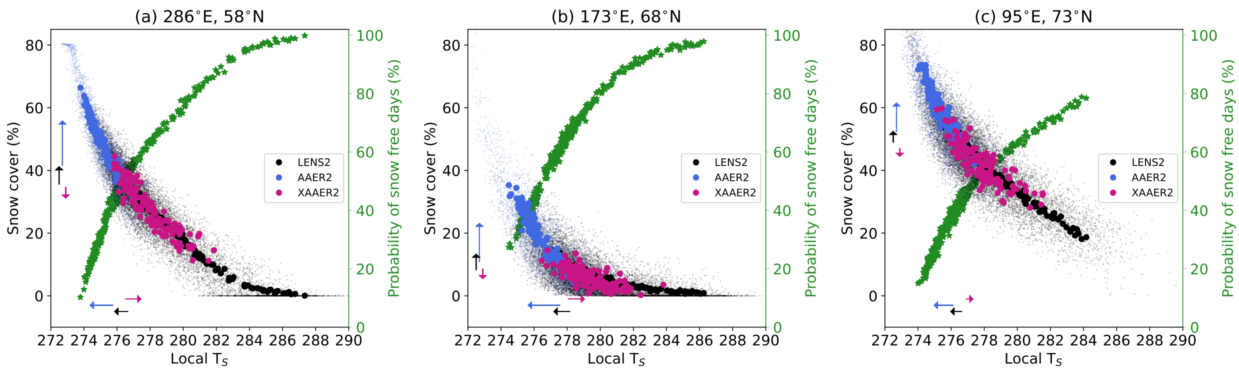

FIG. 7. Evolution of snow cover as a function of local T_s for the three points shown in Fig. 6l: (a) 286°E, 58°N (east of Hudson Bay); (b) 173°E, 68°N (eastern Siberia); and, (c) 95°E, 73°N (south of the Kara Sea). Dots show the JJA seasonal mean percentage of the grid point covered by snow versus local T_s. Small dots show the individual seasons for all members and all years and large dots show the ensemble means for each year. Blue = AAER2, pink = XAAER2 and black = LENS2 (LENS2 is shown out to 2100). The green stars (right axis) show the probability of snow free days in the JJA season assessed for each year by pooling together all members from either LENS2 or AAER2 (XAAER2 is not shown for this metric given its smaller ensemble size). The start point of each of the arrows shows the ensemble mean 1920-1940 average value and the end point shows the ensemble mean 1960-1980 average. The overall change in AAER2 is simply quantified by the length and direction of the blue arrow while the magnitude of the change in LENS2−XAAER2 is given by summing up the length of the black and pink arrows when they are in opposite directions, as in all cases here and the direction of change is that of the black arrow.

we are comparing the 1960-1980 minus 1920-1940 anomalies of LENS2 with that in XAAER2, and the sea ice fraction doesn't differ substantially between LENS2 and XAAER2 in 1920-1940 (not shown), the baseline for the aerosol influence in LENS2−XAAER2 is effectively 1960-1980 of XAAER2. The 80% sea ice fraction metric indicates that a high sea ice fraction is present for a wider latitude range in the colder AAER2 climate and, as a result, the additional growth due to the anthropogenic aerosol influence occurs at lower latitudes compared to the growth that occurs in LENS2 relative to XAAER2. Note that the difference in ice fractions between 1920-1940 of AAER2 and 1960-1980 of XAAER2 is also accompanied by differences in sea ice thickness (thicker ice in the central Arctic in AAER2) which will also impact on where additional sea ice under aerosol-forced cooling will grow. The overall result is that AAER2 gains more sea ice at low latitudes and less sea ice at high latitudes compared to LENS2−XAAER2 (Fig. 6p). While the latitude at which sea ice is gained under aerosol forcing clearly represents a state dependence leading to differences between the methods, its effects on SW↑ during the summer months is likely small, given that during the summer, the latitudinal gradients in incoming shortwave are small. But this could be a contributor to the method dependence of SW↑ anomalies during the spring and autumn, when insolation is relatively greater at lower latitudes. In addition to this difference in latitude at which sea ice is gained, supplemental Fig. 7 shows the dependence of JJA averaged sea ice fraction on temperature, in a similar manner to Fig. 7 for snow, where it can be seen that a non-linearity is also present in association with sea ice being bounded by zero and, therefore, the probability of having days with zero sea ice increases with increasing temperature.

In the SH, summer sea ice fraction differences and their relation to differences in surface upward shortwave radiation are also clear. AAER2 shows greater increases in sea ice cover overall with a strong correspondence between regions where sea ice has increased more and regions where the increase in upward shortwave from the surface is greater (supplemental Fig. 8).

In summary, the behavior of snow cover and sea ice and their influence on surface upward shortwave radiation appear to be state dependent, i.e., it matters whether aerosol forcing is imposed within a cold pre-industrial climate or whether it is imposed within a climate state that has also experienced greenhouse gas forcing. An aerosol cooling that occurs at a colder temperature, as in AAER2, increases the average snow cover more than an aerosol induced cooling that occurs within a planet that has warmed under greenhouse gas forcing, as in LENS2 compared to XAAER2. The same is true for sea ice and, in addition, sea ice that is gained during the cooling of a colder climate tends to occur at lower latitudes than the sea ice gains that occur during the cooling of a warmer climate. These snow cover and sea ice non-linearities lead to an overall larger influence on shortwave radiation for the colder base state in AAER2.

2) The North Atlantic Ocean circulation

The annual mean T_s response to anthropogenic aerosol forcing by 2030-2050 can be seen in Fig. 8. By this time period, AAER2 (Fig. 8a) is colder than LENS2−XAAER2 (Fig. 8b) over much of the globe, with the largest differences found in the NH high latitudes, over continental regions, around the margins of Antarctic sea ice and in the tropical and sub-tropical Pacific (Fig. 8c). The fact that the T_s differences extend beyond the high latitudes is not inconsistent with the important role for high latitude feedbacks in producing them because similar anomalies, with opposite sign relative to those in Fig. 8c in the tropical Pacific have been found in response to sea ice loss (as opposed to gain in our case) (Deser et al. 2015) with an important role for ocean dynamics in transferring the signal there (Wang et al. 2018).

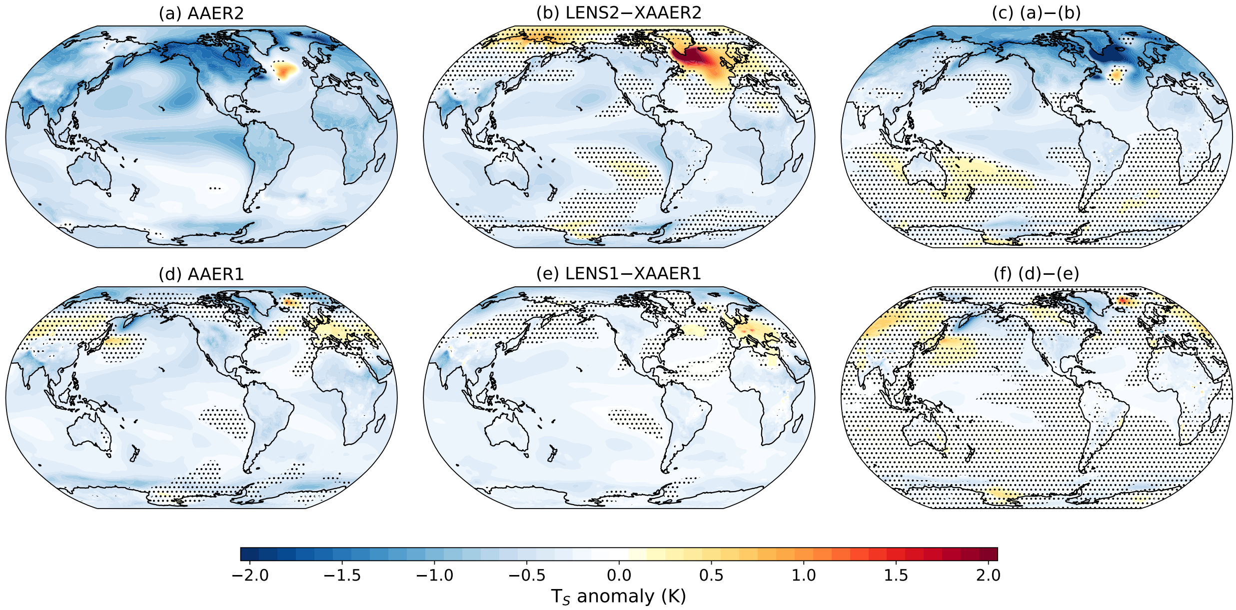

FIG. 8. Differences in annual mean T_s between 2030-2050 and 1920-1940. (a)-(c) CESM2 AAER2, LENS2−XAAER2 and the difference between them. (d)-(f) as (a)-(c) but for CESM1. Stippling indicates anomalies that are not statistically significant at the 95% confidence level.

Consideration of the differences between AAER2 and LENS2−XAAER2 in 2030-2050, however, reveals another important feature: there is a clear difference in the sub-polar North Atlantic with a substantial warm anomaly in the region south of Greenland in LENS2−XAAER2 (Fig. 8b), which is much less apparent in AAER2 (Fig. 8a). In LENS2−XAAER2, there is also a warm anomaly over much of the Arctic. It is clear that there are large differences between the two methods in the T_s response over the NH high latitudes and in the sub-polar gyre region to the south of Greenland, in particular (Fig. 8c), while the same method-dependence is not found in CESM1 (Fig. 8d-f).

The warm sub-polar gyre T_s anomaly in LENS2−XAAER2 to the south of Greenland resembles what would be expected from a strengthening of the AMOC (e.g., Delworth et al. 2017). Indeed, consideration of the AMOC response, defined as the change in the magnitude of the maximum meridional overturning streamfunction at 45°N below 500 m depth, reveals a strong dependency of the aerosol-forced AMOC changes on the experimental design in CESM2 (Fig. 9c).

First, it is worth considering how the NH aerosol forcing evolves as this is likely to be more directly connected to forcing of AMOC changes than the global mean aerosol evolution. Figure 9a shows that the 50°N to 90°N AOD in CESM2 increases to a maximum in the 1970s and 1980s and then declines but levels off at higher AOD values than the 1920-1940 period for the remainder of the simulation. In AAER2, the AMOC increases in strength to a maximum in the 1970s and then declines (Fig. 9c, solid), somewhat following the NH high latitude aerosol forcing, except it returns to the baseline AMOC strength despite the positive AOD anomalies in the 21st century. In contrast, the AMOC in LENS2−XAAER2 increases in strength more rapidly and then, rather than decreasing with the decline in aerosol forcing, it plateaus and even slightly increases out to the end of the simulation (Fig. 9c, dashed). Consideration of XAAER2 in isolation (pink in Fig. 9e) reveals that as the planet warms in the absence of aerosol forcing and the presence of rising greenhouse gases, the AMOC in CESM2 starts to decline rapidly around 1980. This is rather similar to what is seen in GHG2 (red in Fig. 9g). In contrast, in LENS2 when all forcings are present, the anthropogenic aerosol forcing seems to dominate and acts to strengthen the AMOC until about 1980, delaying this rapid greenhouse gas-forced decline in AMOC until later in the simulation (black in Figs. 9e and g). The result is a non-linear behavior of AMOC in the CESM2 simulations with the sum of the individual forcing contributions not adding up to the LENS2 response (compare black and dashed purple in Fig. 9g).

The LENS2 AMOC decline begins around 1980 but it does not fall below the 1920-1940 baseline until about 2000 and at that point it declines at a rather similar rate to what was seen earlier in XAAER2. The AMOC in XAAER2 starts to decline very rapidly around 1980, and because the aerosols in LENS2 delay the onset of this rapid decline compared to XAAER2, the aerosol influence inferred from LENS2−XAAER2 is an apparently greater strengthening of the AMOC than that inferred when the aerosols are imposed in isolation in AAER2. It is not that the aerosol forcing by itself produces the strengthening inferred from LENS2−XAAER2, but it is that it staves off the rapid greenhouse gas-forced AMOC decline. This can explain the warm anomalies in the NH high latitudes and the sub-polar North Atlantic due to aerosol forcing estimated from LENS2−XAAER in Fig. 8b, whereas in AAER2, the AMOC strengthening is weaker and declines after the 1980s. The increased AMOC strength in LENS2−XAAER2 leads to enhanced northward ocean heat transport into the NH high latitudes (not shown), which is a further boost to the disparity in surface albedo between the two methods through effects on sea ice and snow melt.

It is challenging to truly isolate the relative importance of AMOC versus the other albedo non-linearities described above to the method-dependence of the global surface upward shortwave radiation, but we can at least obtain a rough estimate of the order of magnitude of AMOC's

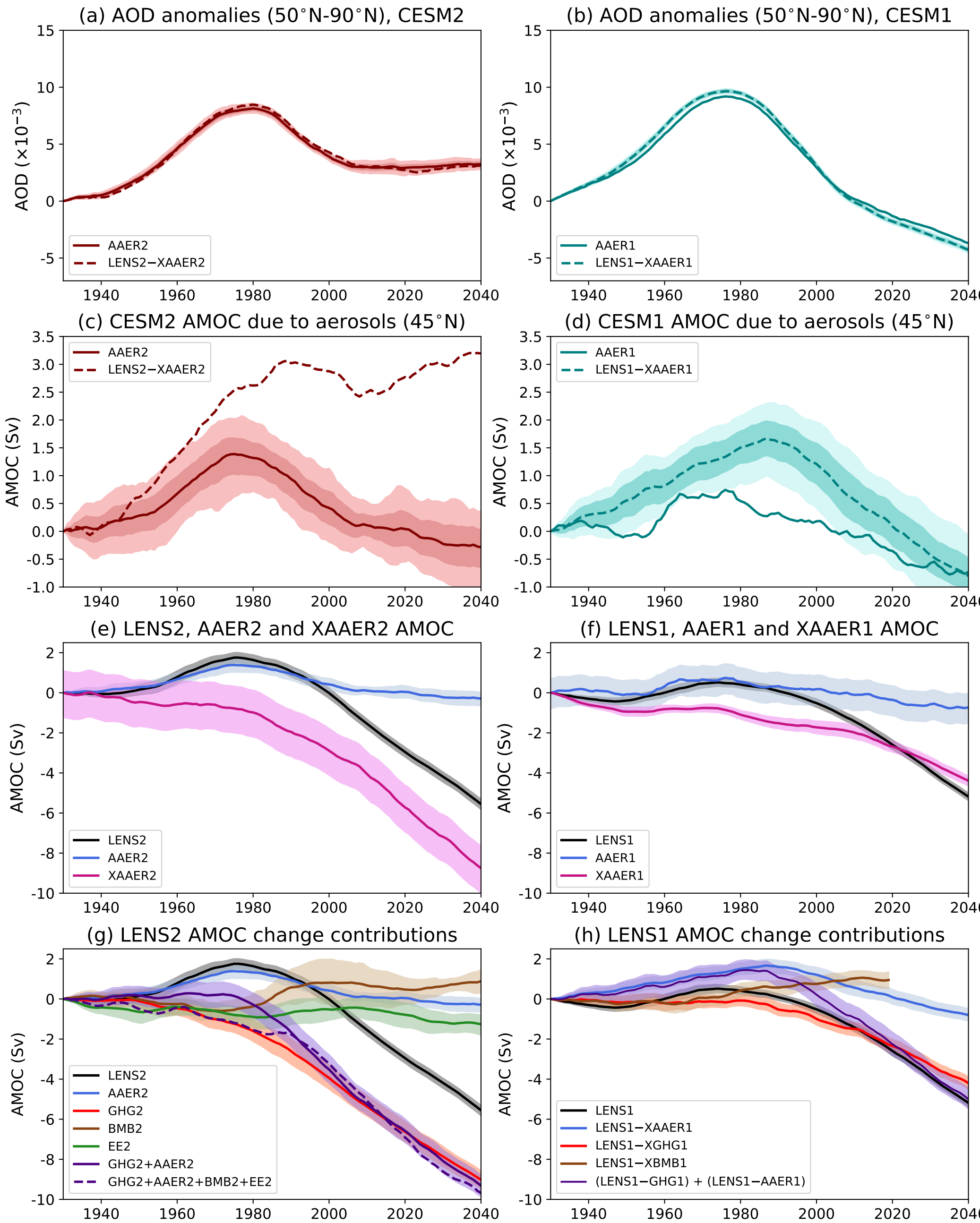

FIG. 9. (a) and (b) 21-year running mean AOD anomalies (relative to 1920 to 1940) averaged from 50°N to 90°N for CESM2 and CESM1 anthropogenic aerosol forcing, respectively. The remaining panels show 21-year running mean AMOC anomalies (relative to 1920 to 1940) where AMOC is defined as the magnitude of the maximum meridional overturning streamfunction below 500m depth in the North Atlantic at 45°N. (c) and (d) show the inferred AMOC changes due to anthropogenic aerosol forcing using both methods for CESM2 and CESM1, respectively, with 95% confidence intervals provided for AAER2 in (c) and LENS1−XAAER2 in (d) using 3 members (light) and the number of members in the large ensemble (dark). (e) and (f) show the AMOC anomalies for all forcings, the AAER simulation and the XAAER simulation for CESM2 and CESM1, respectively. (g) and (h) show the decomposition of the overall change in AMOC in the large ensemble into the contributions that are inferred to be due to individual forcings for CESM2 and CESM1, respectively. 95% confidence intervals on the ensemble means are shown in (e)-(h). For the 3 member XAAER2 and AAER1 ensembles in (e) and (f), respectively, the 95% confidence interval is calculated by bootstrapping the pre-industrial control.

impacts by considering the association of globally integrated surface upward shortwave radiation with AMOC variability in the CESM2 pre-industrial control simulation. Figure 10a shows the regression of 21-year running mean globally integrated surface SW↑ anomalies onto AMOC (after linearly detrending to remove the pre-industrial control drift). This shows that following an increase in AMOC of 1Sv, the globally integrated surface SW↑ declines by just under 2.5×10^{13}W about 5 years later. We then use this relationship between AMOC and globally integrated surface SW↑ with a lag of 5 years to construct the influence of the AMOC anomalies in each experiment on globally integrated surface SW↑ and the difference in this between AAER2 and LENS2−XAAER2. A comparison of this constructed AMOC influence on surface SW↑ with the difference in globally integrated surface SW↑ between AAER2 and LENS2−XAAER2 suggests that the influence of the method-dependence on AMOC can explain a little under half of the influence of the method-dependence on the globally integrated surface upward SW↑, with presumably the other albedo effects described above contributing to the remainder (Fig. 10b). This assumes that we can linearly relate globally integrated SW↑ to AMOC variability and that there is no dependence of AMOC-related surface upward shortwave variability on the climate base state. An analysis of similar regressions to those in Fig. 10a throughout the transient LENS2 simulations suggests that it is, indeed, a reasonable approximation to assume that the piControl regression of SW↑ onto AMOC is representative of that over the 20th and early 21st centuries (Supplemental Fig. 9).

A variety of processes can force an AMOC decline under climate change, including reduced sensible heat loss from the ocean in the presence of a warmer atmosphere (Weaver et al. 2007; Brodeau and Koenigk 2016), altered freshwater forcing as precipitation and evaporation patterns change (Manabe and Stouffer 1993; Dixon et al. 1999) and altered lateral transports of freshwater into regions of deep convection as a result of sea ice loss (or melting of the Greenland Ice Sheet, although Greenland ice sheet melt is not represented in these CESM2 simulations) (Jahn and Holland 2013; Yang et al. 2016; Li et al. 2021). Once an AMOC decline has been induced, positive feedbacks, particularly from the reduced advection of salty water from southern latitudes, can further enhance the AMOC decline. Such feedbacks have also been argued recently by Hassan et al. (2021) and Robson et al. (2022) to be important in the aerosol-forced strengthening of AMOC. A more detailed analysis of the reasons behind the substantial AMOC decline under GHG forcing in CESM2 is warranted and, while we leave this for future work, we provide a cursory assessment

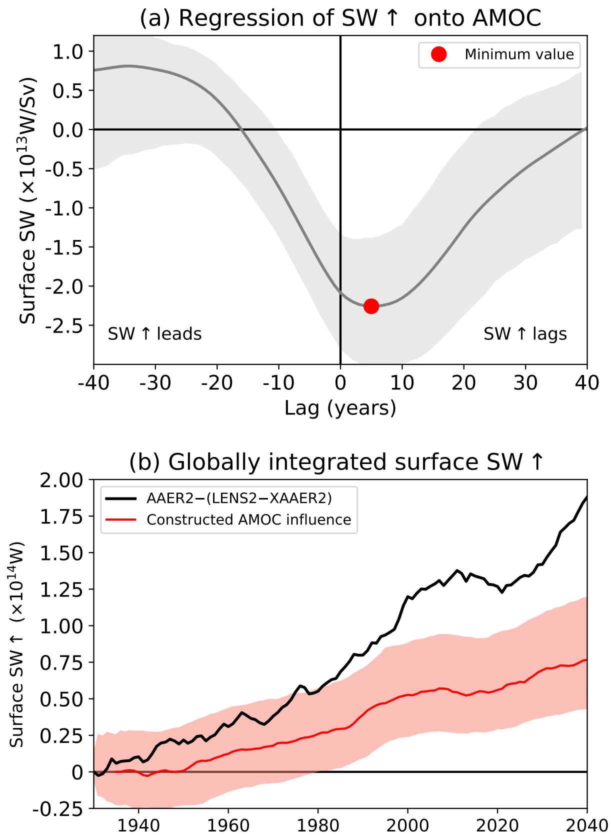

FIG. 10. (a) Regression of 21-year running mean globally integrated surface upward shortwave radiation onto 21-year running mean AMOC strength, defined as the maximum streamfunction below 500 m at 45°N, in the CESM2 pre-industrial control. The gray shaded range shows a 95% confidence interval determined by bootstrapping with replacement 200-year segments of the pre-industrial control, concatenating them to obtain 1000 time series of equivalent length to the 1600-year pre-industrial control simulation, recalculating the regression and obtaining the 2.5th to 97.5th percentile range. The red point marks the minimum value of this regression curve. (b) Black shows the time series of the difference between AAER2 and LENS−XAAER2 21-year running mean globally integrated surface upward shortwave (reproduced from Fig. 5a) and red shows the estimated influence of the difference in AMOC changes between AAER2 and LENS2−XAAER2 by constructing the AMOC influence on globally integrated surface SW↑ by -2.26e13*AMOC(t-5) where -2.26e13 is the minimum regression coefficient in panel (a) and occurs at a lag of 5 years and AMOC refers to the annual mean AMOC strength anomalies relative to 1920-1940 using the maximum streamfunction below 500m at 45°N definition for AMOC strength. The uncertainty range on this construction is determined by recalculating the construction using the bootstrapped minimum regression coefficients and lags that were used to determine the confidence interval in panel (a) and obtaining the 2.5th to 97.5th percentile range of these bootstrapped constructions.

of the different forcing factors that could lead to an AMOC decline for the XAAER2 simulation in supplemental Fig. 10 to shed some light on the possible causes of the AMOC declines shown in Fig. 9e and g. This suggests that the freshwater input associated with sea ice loss is the most likely candidate forcing of the AMOC decline. The annual mean sea ice thickness anomalies are also shown in Fig. 11b and this shows that the Arctic sea ice thickness declines occur earlier in the absence of aerosol forcing (compare pink and black lines in Fig. 11b). This freshwater forcing likely leads to a decline in the near surface density of sea water (ρ) in the Labrador Sea through a reduction in salinity, as quantified for March in Fig. 11e. Here, ρ anomalies averaged over the top 203m of the ocean in the Labrador Sea have been decomposed into the parts associated with salinity (ρ_S) and temperature (ρ_T) using an equation of state for sea water (McDougall et al. 2003), and the salinity component is dominating. Associated with this is a reduction in convection in the Labrador Sea, as depicted via the substantial reductions in March mixed layer depth in Fig. 11h, which uses the definition of Large et al. (1997).

We suspect, based on Fig. 11k that there is also an important role for positive salinity feedbacks in the rapid AMOC decline with greenhouse gas forcing in CESM2. Figure 11k shows the lagged regression of Labrador Sea density (ρ, ρ_S and ρ_T) anomalies onto AMOC in the CESM2 pre-industrial control simulation using 10-year running means. We switch to using 10-year running means here for consistency with the study of Danabasoglu et al. (2019) which made use of this metric to indicate the role of density anomalies in driving and feeding back onto AMOC anomalies. In the pre-industrial control variability, the maximum positive density anomalies due to changing salinity lag the AMOC (dotted in Fig. 11k). This dominates over temperature feedbacks to lead to an overall positive density feedback on AMOC variability. We also suspect that differences in this feedback between CESM2 and CESM1 are important in their differing AMOC behavior, to be discussed in section 5.

Overall, when anthropogenic aerosol forcing is imposed on its own, it leads to an increase in sea ice thickness (solid blue in Fig. 11b), a slight increase in Labrador Sea density through increased salinity (solid blue in Fig. 11e) and a slight increase in AMOC strength (solid blue in Fig. 9e). However, when the anthropogenic aerosol influence is inferred from LENS2−XAAER2, because the aerosols postpone the decline in sea ice thickness, ρ, mixed layer depth and AMOC in LENS2 compared to XAAER2, they lead to an apparent continued increase in Labrador Sea mixed layer depth over the course of the simulation (Fig. 11h, blue dashed) and an increase in AMOC strength (Fig. 9c, maroon dashed). The increase in AMOC strength is associated with enhanced ocean heat transport into the NH high latitudes and, presumably in association with this, the sea ice thickness starts to decline (blue dashed in Fig. 11b), as does the reduction in global mean temperature (Fig. 3b, maroon dashed).

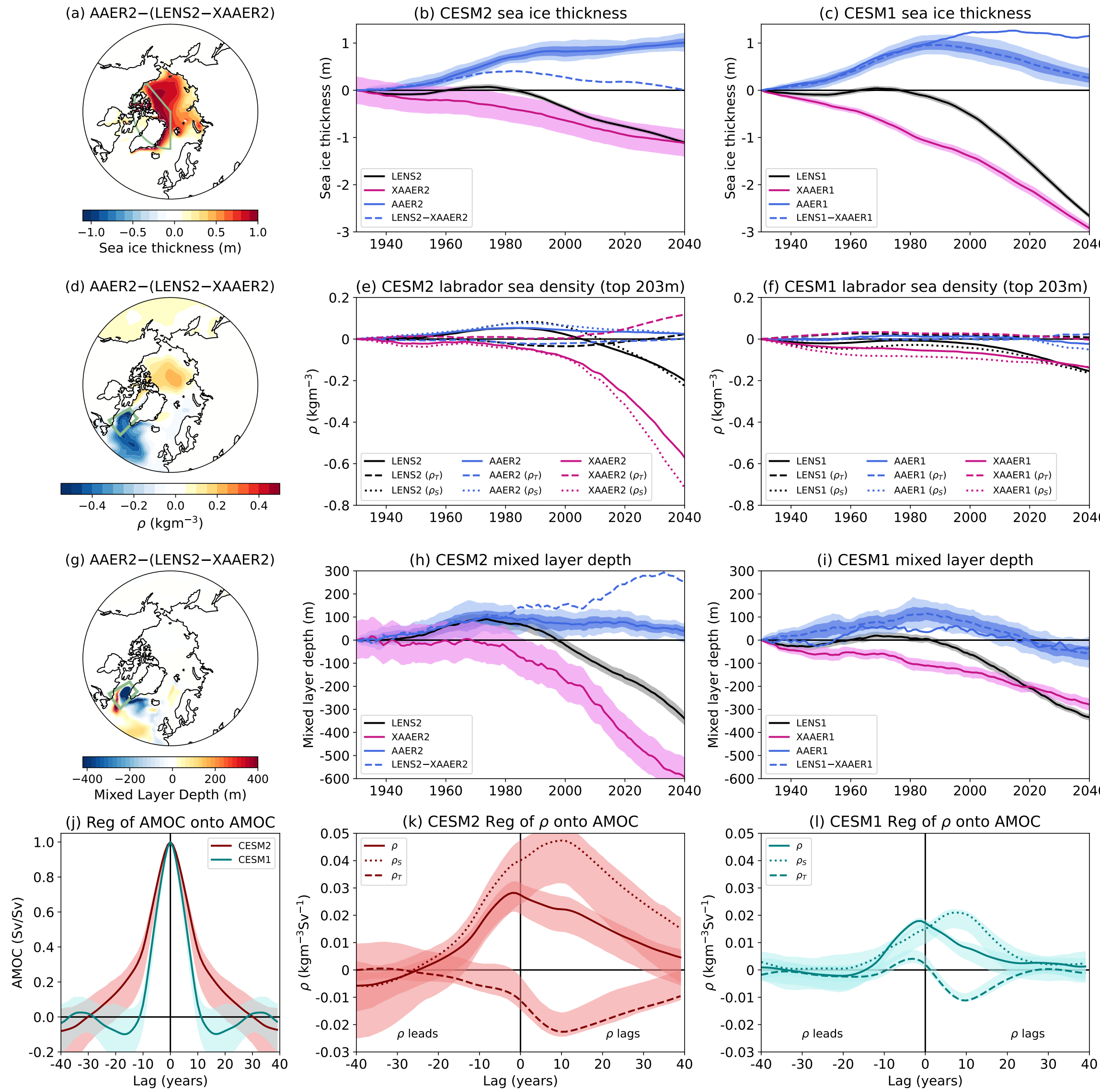

FIG. 11. (a) The difference in annual mean sea ice thickness anomalies between AAER2 and LENS2−XAAER2 during 2030-2050. (b) 21-year running mean CESM2 annual mean sea ice thickness anomalies (relative to 1920-1940) averaged over 220°E-360°E, 70°N-90°N (green region in (a)) for LENS2 (black), XAAER2 (pink), AAER2 (blue), and LENS2−XAAER2 (blue dashed), (c) as (b) but for CESM1. For AAER2 in (b) and LENS2−XAAER2 in (c) 95% confidence intervals are shown for a 3-member ensemble (light) and an ensemble of size equal to the one shown (dark) and for other experiments, the shading shows the 95% confidence interval for a sample size equal to that in the ensemble. (d)-(f) show March density anomalies averaged over the top 203 m of the ocean and the averaging region for (e) and (f) is the Labrador Sea (300°E-315°E, 53°N-65°N, green box in (d)). The LENS2−XAAER2 anomalies are not shown in (e) and (f) and instead the density anomalies in the other simulations are decomposed into the part that is associated with temperature (ρ_T dashed) and the part that is associated with salinity (ρ_S dotted). (g)-(i) are as (a)-(c) but for March mixed layer depth and the averaging in (h) and (i) is performed over the Labrador Sea. (j) lagged auto-regression of annual mean AMOC. (k) lagged regression of March Labrador Sea density anomalies (and its temperature and salinity components) in the top 203 m onto annual mean AMOC for CESM2. (l) as (k) but for CESM1. In (j)-(l) 10 year running means are used for consistency with Danabasoglu et al. (2019)

c. Summary of method-dependence

In this section we have investigated the dependence of the anthropogenic aerosol-forced $[T_s]$ response to the method used in CESM2 (AAER2 versus LENS2−XAAER2). The surface energy balance indicates that aerosol forcing leads to a bigger decline in net TOA shortwave in AAER2 compared to LENS2−XAAER2 (Fig. 2c), which drives a bigger decline in global mean T_s and an associated compensating decline in net TOA longwave radiation. The method-dependence of the TOA net shortwave radiation can be further narrowed down to difference in the surface upward shortwave radiation (Fig. 3g) linked to a difference in the surface shortwave albedo response (Fig. 3h).

We then discussed three potential sources of this albedo non-linearity. The first two are non-linearities or base state dependencies in the response of snow cover and sea ice fraction to T_s change. Snow cover over the continental regions surrounding the Arctic declines non-linearly with warming. This is because the amount of time (in the summer at least) that is spent without any snow cover increases with warming and, as a result, the amount of time in which further warming can influence the snow cover declines. The result is that the aerosol-forced cooling that is imposed within a cooler climate in AAER2, leads to a larger increase in snow cover than the aerosol-forced cooling that is imposed within a warmer climate that is influenced by CO2, as is the case in the LENS2 versus XAAER2 comparison. Sea ice exhibits a similar non-linear behavior and there is also a dependence of the latitude at which sea ice grows with aerosol-forced cooling on the base state climate. In the cold climate within AAER2, additional sea ice grows at lower latitudes and while the impacts of this on the global radiative balance is likely minimal during the summer, it may matter more during the spring and autumn when latitudinal gradients in insolation are larger. These cryospheric effects lead to overall greater increases in albedo in AAER2 than in LENS2−XAAER2, and ultimately a differing global response as atmospheric and oceanic heat transports respond.

Finally, there is also clearly a non-linear behavior of AMOC which further widens the discrepancy between AAER2 and LENS2−XAAER2. In CESM2, with warming, the AMOC strength declines substantially and non-linearly. As a result, in LENS2 where aerosol forcing delays this AMOC decline, it leads to an apparently larger increase in AMOC strength due to aerosols in LENS2−XAAER2 than is found due to aerosol forcing alone in AAER2. The associated increased heat transport into the NH high latitudes in LENS2−XAAER2 would further alter sea ice and snow cover with associated albedo changes.

5. Comparison between CESM2 and CESM1

We motivated the analysis in section 4 by aiming to determine the relative roles of the experimental design used, the model physics and the aerosol forcing in leading to the differing anthropogenic aerosol responses between CESM1 and CESM2. The results point to an important influence of the method in CESM2, but the same method-dependence was not found in CESM1. While the cooling due to anthropogenic aerosol forcing is greater in AAER1 than in LENS1−XAAER1 (solid versus dashed teal in Fig. 3b), the difference is much smaller than in CESM2. There are two potential reasons for this: (1) the model physics and dynamics in CESM1 is such that non-linearities are less important and (2) the fact that the imposed aerosol forcing declines more rapidly from the 1980s (see the Appendix) may mean that CESM1 does not have as much of a chance for the non-linearities to lead to a big deviation between the methods. We cannot really explore the effect of (2) without simulations with CESM2 run under a lower aerosol emissions scenario and this would be a worthwhile avenue for future research to truly quantify the relative importance of these two factors. Nevertheless, differences between CESM1 and CESM2 do suggest that there is a role for model differences in leading to more non-linearity in CESM2 than in CESM1 and that it is not solely the difference in aerosol forcing that is responsible, as now discussed.

CESM1 exhibits much less of a difference in global upward surface shortwave between AAER1 and (LENS1−XAAER1) in the NH than was found between AAER2 and (LENS2−XAAER2), although the SH difference is comparable (see supplemental Fig. 11, which is the equivalent of Fig. 6 but for CESM1). CESM1 does not exhibit as substantial a method-dependence for sea ice cover in the low latitudes of the Arctic or for snow cover over the continents surrounding the Arctic (supplemental Fig 12) but there are some similarities in the method-dependence for sea ice in the SH (supplemental Fig. 13). The differences between CESM1 and CESM2 in the NH can perhaps be traced back to two differences between the CESM1 and CESM2 climates. In CESM1, the Arctic sea ice is thicker and more expansive (supplemental Fig. 14 and DuVivier et al. (2020)). A result of the more expansive sea ice in CESM1 is that there may be less room for sea ice to grow in the aerosol only simulation before the continent is reached, limiting the differences in the extent to which the sea ice fraction increases at lower latitudes in AAER1 compared to LENS1−XAAER1. The second factor that may be important is that CESM1 has reduced summertime snow cover compared to CESM2 in portions of the continental regions surrounding the Arctic, in particular regions adjacent to Hudson Bay and to the south of the Barents-Kara Sea (supplemental Fig. 14c), which likely makes those regions less non-linear in their snow cover response to temperature anomalies (recall Fig. 7). These regions correspond to those where the wintertime snow density has increased the most in CESM2 compared to CESM1 in response to updated snow density parameterizations (see Fig. 5 of Simpson et al. (2022)), which is likely playing a role.

Another important difference between CESM1 and CESM2 is in the behavior of AMOC. CESM1 does exhibit a method-dependence of the aerosol-forced AMOC response (Fig. 9d). In LENS1−XAAER1 the AMOC strengthens more than in AAER1 (dashed versus solid in Fig. 9d) but this does not last for the full length of the simulation. The LENS1−XAAER1 AMOC strength starts to decline substantially after about 1990 whereas the LENS2−XAAER2 AMOC strengthening continues out to the end of the simulation (Fig. 9d versus c). This may be partly due to the differing forcings between CESM1 and CESM2, but comparison of Figs. 9e and f makes clear that the behavior of AMOC in the XAAER simulations (pink), where aerosols are not evolving and greenhouse gas forcing is the primary driver, also differs considerably between CESM1 and CESM2. In XAAER2, the AMOC declines much more rapidly after around 1980 than in XAAER1 and so the aerosol-forced strengthening inferred from the LENS-XAAER calculation is smaller during this period in CESM1 than in CESM2. Subsequently, when the aerosol forcing starts to decline in the NH high latitudes, because it actually goes negative compared to the 1920-1940 baseline in CESM1, the aerosol forcing and GHG forcing act together to produce a sharper decline in AMOC in LENS1 than in XAAER1 in the 21st century (Fig. 9f black versus pink). The differing AMOC behavior is likely part of the reason why the high latitudes warm a lot less in response to greenhouse gas forcing in GHG2 than in LENS1−XGHG1 (supplementary Fig. 15). In CESM2, the greenhouse gas-forced decline in AMOC is greater than in CESM1, which reduces the northward heat transport into the high latitudes and reduces the warming there.

We speculate that an important factor in the differing AMOC responses between XAAER2 and XAAER1 is in the strength of salinity feedbacks. We can consider the AMOC decline to consist of two parts: (1) the forcing which leads to the decline in the first place, which we argued for

CESM2 above was most likely the freshwater input to the Labrador Sea from sea ice melt; and, (2) subsequent feedbacks which are triggered as the AMOC starts to decline, including the reduced advection of salty water from the low latitudes to the high latitudes. The various potential forcers of AMOC decline (surface freshwater flux, surface heat flux and sea ice loss) can be compared between CESM1 and CESM2 for the XAAER experiment in supplemental Fig. 10. For both CESM1 and CESM2, sea ice loss appears as the most likely forcer of AMOC decline as it is the only one which leads the AMOC decline as opposed to lags it. However, a difference in sea ice loss cannot explain the differences in AMOC decline between XAAER1 and XAAER2, because the sea ice loss is actually greater in XAAER1 than in XAAER2 (compare Figs. 11b and c), while the AMOC decline is greater in XAAER2. This suggests that the reason behind the difference in AMOC decline between XAAER1 and XAAER2 is more likely to be a difference in the feedbacks rather than in the initial forcing of AMOC decline. Figure 11j, which shows the lagged autoregression of AMOC onto itself within the pre-industrial control simulations of CESM1 and CESM2, demonstrates that the timescale of AMOC variability is longer in CESM2 than in CESM1. We may reasonably expect that a longer timescale AMOC variability is either due to longer timescale forcing, whether that be through sea ice variability or surface flux variability, or due to stronger feedbacks onto AMOC variability which would lengthen the persistence of any anomalies induced by the various forcers. Comparison of the lagged regression of density onto AMOC in Figs. 11k and l shows that the salinity anomalies that lag AMOC in CESM1 are much smaller than in CESM2, i.e., per unit Sverdrup increase in AMOC strength, the lagged increase in Labrador Sea salinity is greater in CESM2, which would provide a greater feedback onto an AMOC change and, therefore, enhance the persistence of AMOC variability. This suggests that the positive salinity feedback onto AMOC anomalies may be stronger in CESM2 than in CESM1, for reasons that are currently unknown. This could lead to the more rapid AMOC decline in XAAER2, even though the freshwater input through sea ice loss is smaller. A more detailed analysis of the AMOC decline in both simulations should be performed in future work to fully understand these differences. The updates to the ocean model in CESM2 compared to CESM1 are relatively minimal, but include the representation of mixing effects of estuaries, enhanced mesoscale eddy diffusivity at depth, the use of prognostic chlorophyll for shortwave absorption, and the use of a salinity dependent freezing point (Danabasoglu et al. 2020). Whether the differences in AMOC behavior can be attributed to these ocean model changes or the changes to the coupled system introduced through updates to the other components remains to be understood. Hassan et al. (2022) recently argued that models that exhibit a greater AMOC response to forcing may do so because of a larger feedback between the AMOC and cloud cover in the sub-polar North Atlantic. However, we find no evidence of a substantial feedback between the AMOC and sub-polar North Atlantic cloud cover in CESM2 through regression of total cloud cover onto AMOC in the CESM2 pre-industrial control simulation (not shown).

Overall, the comparison of the behavior of CESM1 and CESM2 makes clear that even though the single forcing experimental design matters within CESM2, it probably does so because of particular features of both the representation of processes within the model and the imposed forcing.

6. Discussion and Conclusions

The implicit assumption when using single forcing experiments to attribute changes to individual forcings is that non-linearities are negligible. As discussed in the introduction, prior studies have drawn mixed conclusions as to whether non-linearities are important. Some of the studies with older model generations, e.g., Feichter et al. (2004) and Ming and Ramaswamy (2009), used a slab ocean, so any non-linearity related to AMOC would have been absent. The more recent study of Deng et al. (2020) is the most relevant to the results presented here since they explored non-linearity within coupled CESM1 time-slice simulations. They did not find substantial non-linearities in GMST and TOA radiative fluxes, aligned with our findings that the "only" versus "all-but-one" method does not dramatically alter those responses to aerosol forcing in transient experiments with CESM1. Further probing of other features by Deng et al. (2020) did reveal other non-linearities, specifically in September-November Arctic sea ice decline and in summertime precipitation over East Asia. The sense of their sea ice non-linearity was that when greenhouse gases and aerosols were imposed together, there was less sea ice decline than summing up the contributions from greenhouse gases and aerosols separately, which is the opposite of what we find for CESM2.

The considerable non-linearities that we infer in CESM2 from the difference between the "only" and "all-but-one" methods for anthropogenic aerosols and comparison to the behavior in CESM1 and these other previous studies makes clear that non-linearities in the response to forcings can be highly dependent on the model physics and/or the forcings used. Indeed Menary et al. (2020) find that in CMIP6 models in general, the aerosol and greenhouse gas-forced AMOC anomalies do approximately sum up to the response when all forcings are applied together. CESM2 is clearly a more non-linear model than CESM1, particularly when it comes to the AMOC response to forcings, but also likely in the impact of surface shortwave albedo feedbacks. For the sea ice aspects, the version of CESM2 used here is known to be deficient in its representation of sea ice (DuVivier et al. 2020) so sea ice changes should be interpretted with caution. For snow cover, further investigation is required to determine whether summertime snow cover in CESM2 is more aligned with observations than CESM1 (supplemental Fig. 14a-c), although Wieder et al. (2022) indicate that CESM2 does have too much snow water equivalent in the springtime in the regions adjacent to Hudson Bay and to the south of the Kara Sea. Much work also remains to be done to fully understand the differences in AMOC variability and change between CESM1 and CESM2 and to determine whether we trust one more than the other.

Overall, the method-dependence found for the aerosol-forced response in CESM2, raises the question, what is the more appropriate method to use in single forcing experiments? Our experience with two generations of CESM indicates that the method used may matter for some models and/or forcings more than others. Ultimately, there is probably no getting around the non-linearities that exist in CESM2, particularly that due to the AMOC and one method is not going to necessarily give you a more correct answer than the other. Using the "only" method we would conclude that greenhouse gases are giving rise to a dramatic decline in AMOC strength that starts in the mid-20th century, whereas the reality is that the greenhouse gases don't have this same effect when they are imposed together with aerosol forcing. Using the "all-but-one" method we would conclude that aerosols give rise to an increasing AMOC strength at least out to 2050, but the reality is that they are only apparently doing that because they have prevented the greenhouse gas-forced AMOC decline. The AMOC response to forcings is non-linear and neither method alone would provide the complete picture and we should be aware of such non-linearities in our interpretation.

This new CESM2 dataset has been released to the research community and we expect that there are many more interesting insights that can be gained from it. We also expect that further insights can be gained by building on the dataset provided here through modified experimental design and/or forcing combinations. Unlike the CESM1 single forcing large ensemble, this new ensemble offers the opportunity to assess additivity of the different forcing contributions in comparison to the overall LENS2 response. The results presented here highlight the importance of non-linearities in interpreting single-forcing simulations, while simultaneously highlighting pertinent mechanisms underlying these non-linearities that may be of value for future endeavors. It is our hope that future work will make use of this dataset to further explain the role of individual forcings and identify their interactions in the evolution of the Earth system.

Acknowledgments. This work is supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under the Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne supercomputer (https://doi.org/10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory at NCAR. We also acknowledge the Community Earth System Model version 1 Large Ensemble Community Project, the Community Earth System Model version 2 Large Ensemble Community Project and supercomputing resources provided by the IBS Center for Climate Physics in South Korea (https://doi.org/10.5194/esd-2021-50). The CESM2 large ensemble simulations were conducted on the IBS/ICCP supercomputer "Aleph", a 1.43 petaflop high-performance Cray XC50-LC Skylake computing system. The work of KBR was supported by the Institute for Basic Sciences (IBS), Republic of Korea, under IBS-R028-D1. IRS is grateful to Brian Medeiros, John Fasullo, Robb Jnglin Wills, Andrew Gettelman and Marika Holland for helpful discussions.

Data availability statement. The CESM1 datasets used in this study are freely available through NCAR's Climate Data Gateway with guidance for data access located at the following website https://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html. The CESM2 large ensemble is also available through NCAR's Climate Data Gateway with guidance for data access located at the following website https://www.cesm.ucar.edu/projects/community-projects/LENS2/data-sets.html. The CESM2 single forcing large is available through the Climate Data Gateway as described at https://www.cesm.ucar.edu/working-groups/climate/simulations/cesm2-single-forcing-le. The analysis codes for reproducing the figures in this manuscript are available at https://github.com/islasimpson/singleforcing_paper and the data required to make the figures are available at https://doi.org/10.5065/ayy9-fr14

APPENDIX

Comparison of the anthropogenic aerosol forcing between the CESM2 and CESM1 single forcing large ensembles

Given that much of the analysis in this study focusses on the anthropogenic aerosol-forced response, we provide a comparison of the aerosol forcing between AAER2 and AAER1 in Fig. A1 with a focus on two species (BC and SO_4), while the emissions and burdens of other species can be found in supplemental Figs. 3 and 4. A comparison of Figs. A1a and A1b reveals that the trends in anthropogenic aerosol optical depth at 550 nm (AOD) over 2000-2050 are very different between the AAER2 (historical to SSP3-7.0) and AAER1 (historical to RCP8.5) simulations. AAER1 shows declines in AOD over eastern North America, Europe and China with relatively small increases in AOD over Africa and India (Fig. A1b). In AAER2, the declines over eastern North America and Europe are much smaller compared to AAER1 and the AOD increases over China in AAER2 while it decreases in AAER1. Over Africa and India, AOD is increasing much more in AAER2 than in AAER1. This difference primarily stems from the difference in emissions, but differences in the model physics also play a role in the differing overall aerosol burdens between AAER1 and AAER2, as now discussed.

The global BC emissions in AAER1 and AAER2 are fairly similar until about the year 2000, but after that they increase in AAER2 and decrease in AAER1 (Fig. A1 c) (a similar trajectory is seen for POM in supplemental Fig. 3d). SO_4 emissions are also similar until about the year 2000, but then they remain fairly steady in AAER2 while declining in AAER1 (Fig. A1d) (a similar trajectory is seen for SO_2 in supplemental Fig. 3c). For BC, while the emissions are slightly lower over the 19th and early 20th centuries in AAER2 compared to AAER1 (Fig. A1c), the BC burden is higher in AAER2 compared to AAER1 (Fig. A1e). The lifetime of BC, estimated by the ratio of the global BC burden to the global BC deposition flux, is longer in CESM2 (6.41 days) than in CESM1 (3.64 days, Fig. A1g). This nearly two-fold increase in BC lifetime can be understood as resulting from the BC wet deposition flux associated with a given burden being smaller in CESM2 than in CESM1 and the wet deposition flux has changed because the representation of the aging of primary carbonaceous aerosols in CESM2 delays BC removal via wet deposition (section 2.a.1 and see the differing wet deposition rates at a given global burden in Fig. A1g). For SO_4, the emissions are rather similar over the historical period (Fig. A1d) but the burden (Fig. A1f) is higher in AAER1 than in AAER2 for reasons that are not totally clear given that the deposition rates are fairly comparable between CESM1 and CESM2 (Fig. A1h).

In summary, the difference in BC and SO_4 emissions is the primary contributor to the difference in burden (and associated AOD) trends between AAER2 and AAER1, with some additional modification due to the differing model physics. The SSP3-7.0 scenario is a higher aerosol emission scenario (Gidden et al. 2019) than the CMIP5 RCP8.5 scenario used in the CESM1 single forcing large ensemble and the emissions also differ slightly over the historical period as emissions inventories were revised between CMIP5 and CMIP6, although it should be noted that it has been argued that the increasing emissions over eastern China in the last decade of the historical period in the CMIP6 emissions are incorrect (Wang et al. 2021). The impact of such differences in emissions can end up being as large as the impact of changing from one model version to the next (e.g., Fyfe et al. 2021).

References

Allen, R. J., J. R. Norris, and M. Kovilakam, 2014: Influence of anthropogenic aerosols and the Pacific Decadal Oscillation on tropical belt width. *Nature Geoscience*, **7**, 270–274.

Baek, S. H., Y. Kushnir, M. Ting, J. E. Smerdon, and J. M. Lora, 2022: Regional Signatures of Forced North Atlantic SST Variability: A Limited Role for Aerosols and Greenhouse Gases. *Geophysical Research Letters*, **49**, https://doi.org/https://doi.org/10.1029/2022GL097794.

Bogenschutz, P. A., A. Gettelman, C. Hannay, V. E. Larson, R. B. Neale, C. Craig, and C.-C. Chen, 2018: The path to CAM6: coupled simulations with CAM5.4 and CAM5.5. *Geosci. Model Dev.*, **11**, 235–255.

Bonfils, C. J. W., B. D. Santer, J. C. Fyfe, K. Marvel, T. J. Phillips, and S. R. H. Zimmerman, 2020: Human influence on joint changes in temperature, rainfall and continental aridity. *Nature Climate Change*, **10**, 726–731.

Brodeau, L., and T. Koenigk, 2016: Extinction of the northern oceanic deep convection in an ensemble of climate model simulations of the 20th and 21st centuries. *Clim. Dyn.*, **46**, 2863–2882.

Capotondi, A., C. Deser, A. S. Phillips, Y. Okumura, and S. M. Larson, 2020: ENSO and Pacific decadal variability in the Community Earth System Model version 2. *J. Adv. Mod. Earth Sys.*, **12**, e2019MS002 022.

Chiang, F., O. Mazdiyasni, and A. AghaKouchak, 2021: Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nature Communications*, **12**, 2754, https://doi.org/https://doi.org/10.1038/s41467-021-22314-w.

Dagan, G., P. Stier, and D. Watson-Parris, 2020: Aerosol Forcing Masks and Delays the Formation of the North Atlantic Warming Hole by Three Decades. *Geophys. Res. Lett.*, **47**, https://doi.org/https://doi.org/10.1029/2020GL090778.

Danabasoglu, G., S. C. Bates, B. P. Briegleb, S. R. Jayne, and W. G. Jochum, M. amd Large, 2012: The CCSM4 ocean component. *J. Clim.*, **25**, 1361–1389.

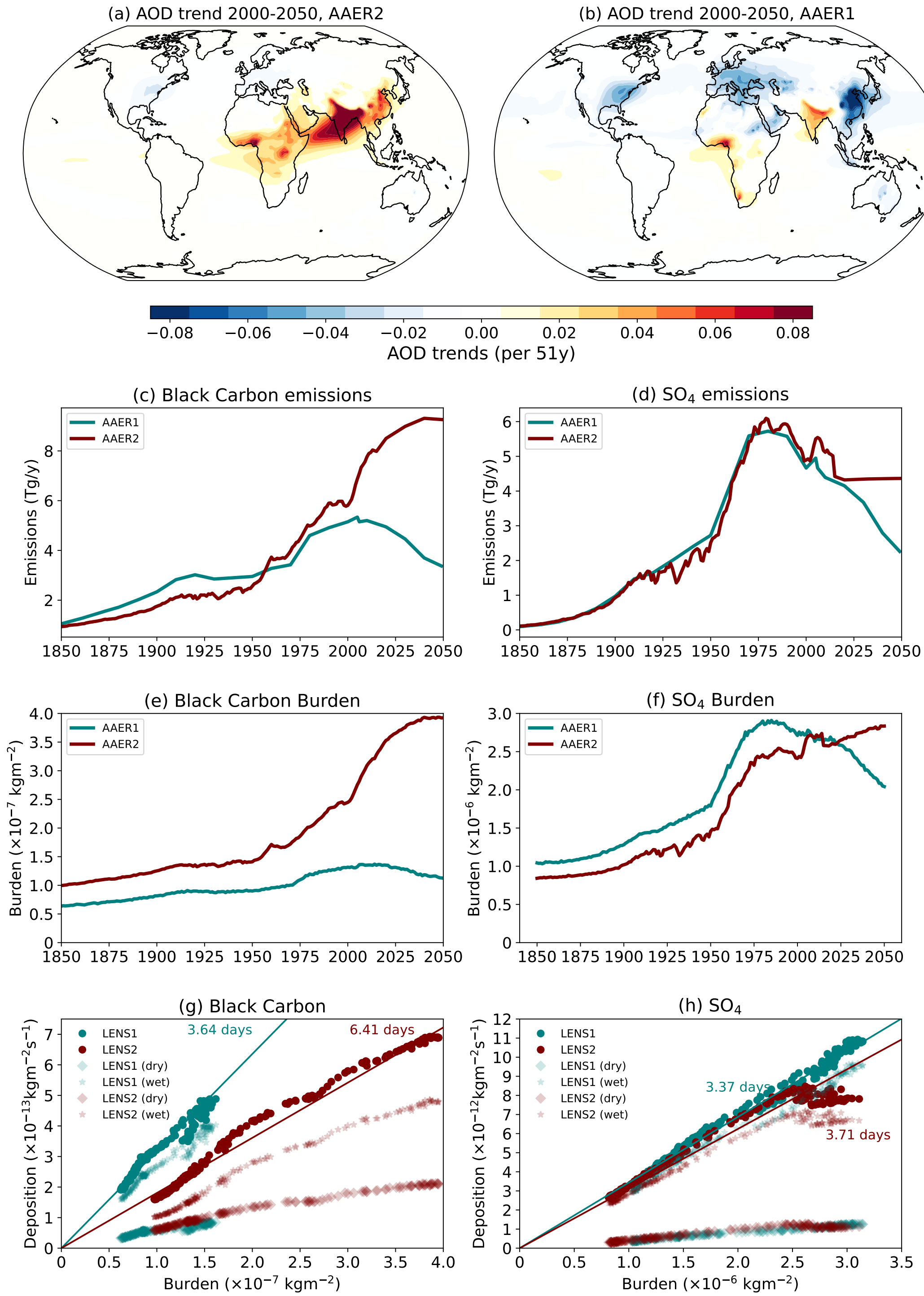

FIG. A1. (a) and (b) Trends in annual mean AOD at 550nm between 2000 and 2050 of AAER2 and AAER1, respectively. (c) Global annual mean emissions of BC for AAER1 (teal) and AAER2 (maroon). (d) same as (c) but for SO_4 aerosol. (e) Global annual mean BC burden for AAER1 (teal) and AAER2 (maroon). (f) as (e) but for SO_4. (g) Global annual mean deposition fluxes versus burden for a single member of LENS2 and LENS1 (used rather than the single forcing experiments because the deposition fluxes were not output in the CESM1 single forcing large ensemble). Circles show the full deposition flux (dry + wet), stars show the wet deposition flux and diamonds show the dry deposition flux. (h) as (g) but for SO_4.

Danabasoglu, G., L. Landrum, S. G. Yeager, and P. R. Gent, 2019: Robust and Nonrobust Aspects of Atlantic Meridional Overturning Circulation Variability and Mechanisms in the Community Earth System Model. *J. Clim.*, **32**, 7349–7368.

Danabasoglu, G., and Coauthors, 2020: The Community Earth System Model 2 (CESM2). *J. Adv. Mod. Earth Sys.*, **12**, https://doi.org/10.1029/2019MS001916.

Delworth, T. L., and F. Zeng, 2014: Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. *Nature Geoscience*, **7**, 583–587.

Delworth, T. L., F. Zeng, L. Zhang, R. Zhang, G. Vecchi, and X. Yang, 2017: The Central Role of Ocean Dynamics in Connecting the North Atlantic Oscillation to the Extratropical Component of the Atlantic Multidecadal Oscillation. *J. Clim.*, **30**, 3789–3805.

Deng, J., A. Dai, and H. Xu, 2020: Nonlinear Climate Response to Increasing CO_2 and Anthropogenic Aerosols Simulated by CESM1. *J. Clim.*, **33**, 281–301.

DeRepentigny, P., and Coauthors, 2022: Enhanced simulated early 21st century Arctic sea ice loss due to CMIP6 biomass burning emissions. *Science Advances*, **8 (30)**, https://doi.org/10.1126/sciadv.abo2405.

Deser, C., R. A. Tomas, and L. Sun, 2015: The Role of Ocean-Atmosphere Coupling in the Zonal Mean Atmospheric Response to Arctic Sea Ice Loss. *J. Clim.*, **28**, 2168–2186.

Deser, C., and Coauthors, 2020a: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change*, **10**, 277–286.

Deser, C., and Coauthors, 2020b: Isolating the Evolving Contributions of Anthropogenic Aerosols and Greenhouse Gases: A New CESM1 Large Ensemble Community Resource. *J. Clim.*, **33**, 7835–7858.

Dittus, A. J., E. Hawkins, J. I. Robson, D. M. Smith, and J. Wilcox, L, 2021: Drivers of Recent North Pacific Decadal Variability: The Role of Aerosol Forcing. *Earth's Future*, **9**, https://doi.org/https://doi.org/10.1029/2021EF002249.

Dixon, K. W., T. L. Delworth, M. J. Spelman, and R. J. Stouffer, 1999: The influence of transient surface fluxes on North Atlantic overturning in a coupled GCM Climate Change Experiment. *Geophys. Res. Lett.*, **26**, 2749–2752.

Dong, B., R. T. Sutton, E. Highwood, and L. Wilcox, 2014: The Impacts of European and Asian Anthropogenic Sulfur Dioxide Emissions on Sahel Rainfall. *J. Clim.*, **27**, 7000–7017.

DuVivier, A. K., M. M. Holland, J. E. Kay, S. Tilmes, A. Gettelman, and D. A. Bailey, 2020: Arctic and Antarctic sea ice mean state in the Community Earth System Model version 2 and the influence of atmospheric chemistry. *J. Geophys. Res.: Oceans*, **125**, e2019JC015 934.

England, M. R., I. Eisenman, N. J. Lutsko, and T. J. W. Wagner, 2021: The Recent Emergence of Arctic Amplification. *Geophys. Res. Lett.*, **48**, https://doi.org/https://doi.org/10.1029/2021GL094086.

Eyring, V., S. Bony, G. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9**, 1937–1958.

Fasullo, J. T., P. R. Gent, and R. S. Nerem, 2020: Sea Level Rise in the CESM Large Ensemble: The Role of Individual Climate Forcings and Consequences for the Coming Decades. *J. Clim.*, **33**, 6911–6927.

Fasullo, J. T., J.-F. Lamarque, C. Hannay, N. Rosenbloom, S. Tilmes, P. DeRepentigny, A. Jahn, and C. Deser, 2021: Spurious Late Historical-Era Warming in CESM2 Driven by Prescribed Biomass Burning Emissions. *Geophys. Res. Lett.*, **49**, e2021GL097 420, https://doi.org/https://doi.org/10.1029/2021GL097420.

Feichter, J., E. Roeckner, U. Lohmann, and B. Liepert, 2004: Nonlinear Aspects of the Climate Response to Greenhouse Gas and Aerosol Forcing. *J. Clim.*, **17**, 2384–2398.

Fyfe, J. C., V. V. Kharin, B. D. Santer, J. N. S. Cole, and N. P. Gillett, 2021: Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proc. Natl. Acad. Sci. (USA)*, **118**, e2016549 118.

Gettelman, A., and H. Morrison, 2015: Advanced two-moment bulk microphysics for global models. Part 2: Off-line tests and comparison with other schemes. *J. Clim*, **28**, 1268–1287.

Gettelman, A., and Coauthors, 2019: The Whole Atmosphere Community Climate Model Version 6 (WACCM6). *J. Geophys. Res.: Atm*, **124**, https://doi.org/10.1029/2019JD030943.

Giannini, A., and A. Kaplan, 2019: The role of aerosols and greenhouse gases in Sahel drought and recovery. *Climatic Change*, **152**, 449–466.

Gidden, M. J., and Coauthors, 2019: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev.*, **12**, 1443–1475.

Gillett, N. P., and Coauthors, 2016: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6. *Geosci. Model Dev.*, **9**, 3685–3697.

Gillett, N. P., and Coauthors, 2021: Constraining human contributions to observed warming since the pre-industrial period. *Nat. Clim. Change*, **11**, 207–212.

Hassan, T., R. J. Allen, W. Liu, and C. A. Randles, 2021: Anthropogenic aerosol forcing of the Atlantic meridional overturning circulation and the associated mechanisms in CMIP6 models. *Atmos. Chem. Phys.*, **21**, 5821–5846.

Hassan, T., and Coauthors, 2022: Air quality improvements are projected to weaken the Atlantic meridional overturning circulation through radiative forcing effects. *Nature Communications Earth and Environment*, **3**, https://doi.org/10.1038/s43247-022-00476-9.

Hirasawa, H., P. J. Kushner, M. Sigmond, J. Fyfe, and C. Deser, 2020: Anthropogenic Aerosols Dominate Forced Multidecadal Sahel Precipitation Change through Distinct Atmospheric and Oceanic Drivers. *J. Clim.*, **33**, 10 187–10 204.

Hunke, E. C., W. H. Lipscomb, A. K. Turner, N. Jeffery, and S. Elliott, 2015: Cice: The los alamos sea ice model. documentation and software user's manual. version 5.1. Tech. Rep. LA-CC-06-012, iT-3 Fluid Dynamics Group, Los Alamos National Laboratory.

Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A Framework for Collaborative Research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360.

IPCC, 2021: Summary for Policy Mackers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group 1 to the Sixth Assessment Report of the Intergovernmental

Panel on Climate Change. Tech. rep., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3-32 pp. https://doi.org/10.1017/9781009157896.001.

Jahn, A., and M. M. Holland, 2013: Implications of Arctic sea ice changes for North Atlantic deep convection and the meridional overturning circulation in CCSM4-CMIP5 simulations. *Geophys. Res. Lett.*, **40**, 1206–1211.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349.

Lamarque, J. F., G. P. Kyle, M. Meinshausen, K. Riahi, S. J. Smith, D. P. van Vuuren, A. J. Conley, and F. Vitt, 2011: Global and regional evolution of short-lived radiatively active gases and aerosols in the Representative Concentration Pathways. *Climatic Change*, **109**, 191–212.

Large, W. G., G. Danabasoglu, S. C. Doney, and J. C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in the NCAR CSM ocean model: Annual-mean climatology. *J. Phys. Oceanogr.*, **27**, 2418–2447.

Lawrence, D., and Coauthors, 2019: The Community Land Model, version 5: Description of new features, benchmarking and impact of forcing uncertainty. *J. Adv. Mod. Earth Sys.*, **11**, 4245–4287.

Lawrence, D. M., and Coauthors, 2011: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model. *J. Adv. Mod. Earth Sys.*, **3**, https://doi.org/https://doi.org/10.1029/2011MS00045.

Li, H., A. Federov, and W. Liu, 2021: AMOC Stability and Diverging Response to Arctic Sea Ice Decline in Two Climate Models. *J. Clim.*, **34**, 5443–5460.

Li, X., M. Ting, and D. E. Lee, 2018: Fast Adjustments of the Asian Summer Monsoon to Anthropogenic Aerosols. *Geophys. Res. Lett.*, **45**, 1001–1010, https://doi.org/https://doi.org/10.1002/2017GL076667.

Liu, X., P.-L. Ma, H. Wang, S. Tilmes, B. Singh, R. C. Easter, S. J. Ghan, and P. J. Rasch, 2016: Description and evaluation of ta new four-mode version of the Model Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model. *Geosci. Model Dev.*, **9**, 505–522.

Liu, X., and Coauthors, 2012: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5. *Geosci. Model Dev.*, **5**, 709–739.

Lombardozzi, D. L., Y. Lu, P. J. Lawrence, D. M. Lawrence, S. Swenson, K. W. Oleson, W. R. Wieder, and E. A. Ainsworth, 2020: Simulating agriculture in the community land model version 5. *J. Geophys. Res.: Biogeosciences*, **125**, e2019JG005 529.

Manabe, S., and R. J. Stouffer, 1993: Century-scale effects of increased atmospheric CO_2 on the ocean-atmosphere system. *Nature*, **364**, 215–217.

McDougall, T. J., D. R. Jackett, D. G. Wright, and R. Feistel, 2003: Accurate and Computationally Efficient Algorithms for Potential Temperature and Density of Seawater. *Journal of Atmospheric and Oceanic Technology*, **20**, 730–741.

Meehl, G. A., C. Shields, J. M. Arblaster, H. Annamalai, and R. Neale, 2020: Intraseasonal, Seasonal, and Interannual Characteristics of Regional Monsoon Simulations in CESM2. *J. Adv. Mod. Earth Sys.*, **12**, https://doi.org/https://doi.org/10.1029/2019MS001962.

Meehl, G. A., W. M. Washington, C. M. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi, 2004: Combinations of Natural and Anthropogenic Forcings in Twentieth-Century Climate. *J. Clim.*, **17**, 3721–3727.

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241.

Meinshausen, M., and Coauthors, 2020: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.*, **13**, 3571–3605.

Menary, M. B., J. Robson, R. P. Allan, B. B. B. Booth, C. Cassou, and G. Gastineau, 2020: Aerosol-forced AMOC changes in CMIP6 historical simulations. *Geophys. Res. Lett.*, **47**, e2020GL088 166.

Ming, Y., and V. Ramaswamy, 2009: Nonlinear Climate and Hydrological Responses to Aerosol Effects. *J. Clim.*, **22**, 1329–1339.

Monerie, P.-A., L. J. Wilcox, and A. G. Turner, 2022: Effects of Anthropogenic Aerosol and Greenhouse Gas Emissions on Northern Hemisphere Monsoon Precipitation: Mechanisms and Uncertainty. *J. Clim.*, **35**, 2305–2326.

Morrison, H., and A. Gettelman, 2008: A new two-moment bulk stratiform cloud microphysics scheme in the NCAR Community Atmosphere Model (CAM3). Part 1: Description and numerical tests. *J. Clim.*, **21**, 3642–3659.

Mueller, B. L., N. P. Gillett, A. H. Monahan, and F. W. Zwiers, 2018: Attribution of Arctic Sea Ice Decline from 1953 to 2012 to Influences from Natural, Greenhouse Gas and Anthropogenic Aerosol Forcing. *J. Clim.*, **31**, 7771–7787.

Pendergrass, A. G., D. B. Coleman, C. Deser, F. Lehner, N. Rosenbloom, and I. R. Simpson, 2019: Nonlinear Response of Extreme Precipitation to Warming in CESM1. *Geophys. Res. Lett.*, **46**, 10 551–10 560.

Robson, J., and Coauthors, 2022: The role of anthropogenic aerosol forcing in the 1850-1985 strengthening of the AMOC in CMIP6 historical simulations. *J. Clim.*, https://doi.org/https://doi.org/10.1175/JCLI-D-22-0124.1.

Rodgers, K. B., and Coauthors, 2021: Ubiquity of human-induced changes in climate variability. *Earth Sys. Dynam.*, **12**, 1393–1411.

Seong, M., S. Min, Y. Kim, X. Zhang, and Y. Sun, 2021: Anthropogenic Greenhouse Gas and Aerosol Contributions to Extreme Temperature Changes during 1951-2015. *J. Clim.*, **34**, 857–870.

Shi, J.-R., Y.-O. Kwon, and S. Wijffels, 2022: Two Distinct Modes of Climate Responses to the Anthropogenic Aerosol Forcing Changes. *J. Clim.*, **35**, 3445–3457.

Simpson, I. R., D. M. Lawrence, S. C. Swenson, C. Hannay, K. A. McKinnon, and J. E. Truesdale, 2022: Improvements in Wintertime Surface Temperature Variability in the Community Earth System Model Version 2 (CESM2) Related to the Representation of Snow Density. *J. Adv. Mod. Earth Sys.*, **14**, e2021MS002 880.

Simpson, I. R., and Coauthors, 2020: An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models. *J. Geophys. Res.: Atmospheres*, **125**, e2020JD032 835.

Singh, D., S. P. McDermid, B. I. Cok, M. J. Puma, L. Nazarenko, and M. Kelley, 2018: Distinct Influences of Land Cover and Land Management on Seasonal Climate. *J. Geophys. Res. Atm.*, **123**, 12 017–12 039, https://doi.org/https://doi.org/10.1029/2018JD028874.

Smith, D., and Coauthors, 2022: Attribution of multi-annual to decadal changes in the climate system: The Large Ensemble Single Forcing Model Intercomparison Project (LESFMIP). *Frontiers in Climate*, https://doi.org/10.3389/fclim.2022.955414.

Smith, R., and Coauthors, 2010: The Parallel Ocean Program (POP) reference manual, Ocean component of the Community Climate System Model (CCSM). Tech. rep., LANL Tech. Report., 141 pp. https://doi.org/LAURâĂŘ10âĂŘ01853.

Tilmes, S., and Coauthors, 2019: Climate Forcing and Trends of Organic Aerosols in the Community Earth System Model (CESM2). *J. Adv. Mod. Earth Sys.*, **11**, 4323–4351.

Touma, D., S. Stevenson, F. Lehner, and S. Coats, 2021: Human-driven greenhouse gas and aerosol emissions cause distinct regional impacts on extreme fire weather. *Nature Communications*, **12**, https://doi.org/https://doi.org/10.1038/s41467-020-20570-w.

Undorf, S., D. Polson, M. A. Bollasina, Y. Ming, A. Schurer, and G. C. Hegerl, 2018: Detectable Impact of Local and Remote Antrhopogenic Aerosols on the 20th Century Chnges of West African and South Asian Monsoon Precipitation. *J. Geophys. Res. Atm.*, **123**, 4871–4889.

van Marle, M. J. E., and Coauthors, 2017: Historic global biomass burning emissions based on merging satellite observations with proxies and fire models (1750-2015). *Geosci. Model Dev. Discuss.*, **10**, 3329–3357, https://doi.org/https://doi.org/10.5194/gmd-10-3329-2017.

Wang, K., C. Deser, L. Sun, and R. A. Tomas, 2018: Fast Response of the Tropics to an Abrupt Loss of Arctic Sea Ice via Ocean Dynamics. *Geophys. Res. Lett.*, **45**, https://doi.org/https://doi.org/10.1029/2018GL077325.

Wang, Z., and Coauthors, 2021: Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models. *npj Clim Atmos Sci*, **4**.

Watanabe, M., and H. Tatebe, 2019: Reconciling roles of sulphate aerosol forcing and internal variability in Atlantic multidecadal climate changes. *Climate Dynamics*, **53**, 4651–4665.

Weaver, A. J., M. Eby, M. Kienast, and O. A. Saenko, 2007: Response of the Atlantic meridional overturning circulation to increasing atmospheric CO2: sensitivity fo mean climate state. *Geophys. Res. Lett.*, **34**, L05708.

Wieder, W. R., D. Kennedy, F. Lehner, K. N. Musselman, K. B. Rodgers, N. Rosenbloom, I. R. Simpson, and R. Yamaguchi, 2022: Pervasive alterations to snow-dominated ecosystem functions under climate change. *Proc. Nat. Ac. Sci*, **119**, https://doi.org/https://doi.org/10.1073/pnas.2202393119.

Yang, Q., T. H. Dixon, P. G. Myers, J. Bonin, D. Chambers, M. R. van den Broeke, M. H. Ribergaart, and J. Mortensen, 2016: Recent increases in Arctic freshwater flux affects Labrador Sea convection and Atlantic overturning circulation. *Nature Communications*, **7**, 10 525.

Zhang, S., P. Stier, G. Dagan, and M. Wang, 2021: Anthropogenic Aerosols Modulated 20th-Century Sahel Rainfall Variability Via Their Impacts on North Atlantic Sea Surface Temperature. *Geophys. Res. Lett.*, **49**, https://doi.org/https://doi.org/10.1029/2021GL095629.