Two Distinct Modes of Climate Responses to the Anthropogenic Aerosol Forcing Changes

JIA-RUI SHI, a YOUNG-OH KWON, a AND SUSAN E. WIJFFELS a

aWoods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Manuscript received 23 August 2021, in final form 19 January 2022)

ABSTRACT: Unlike greenhouse gases (GHGs), anthropogenic aerosol (AA) concentrations have increased and then decreased over the past century or so, with the timing of the peak concentration varying in different regions. To date, it has been challenging to separate the climate impact of AAs from that due to GHGs and background internal variability. We use a pattern recognition method, taking advantage of spatiotemporal covariance information, to isolate the forced patterns for the surface ocean and associated atmospheric variables from the all-but-one forcing Community Earth System Model ensembles. We find that the aerosol-forced responses are dominated by two leading modes, with one associated with the historical increase and future decrease of global mean aerosol concentrations (dominated by the Northern Hemisphere sources) and the other due to the transition of the primary sources of AA from the west to the east and also from Northern Hemisphere extratropical regions to tropical regions. In particular, the aerosol transition effect, to some extent compensating the global mean effect, exhibits a zonal asymmetry in the surface temperature and salinity responses. We also show that this transition effect dominates the total AA effect during recent decades, e.g., 1967–2007.

KEYWORDS: Aerosol radiative effect; Climate Change; Climate variability; Sea surface temperature; Salinity

1. Introduction

Anthropogenic aerosol (AA) and greenhouse gas (GHG) emissions have given rise to a substantial climate change on a global scale over the past century (Myhre et al. 2013; von Schuckmann et al. 2016). AAs have masked a considerable fraction of the GHG-induced global warming since the industrial revolution (Boucher et al. 2013), as aerosols can absorb, reflect, and scatter shortwave radiation to change the energy budget of the climate system (direct effect). Moreover, some aerosol particles, such as sulfates, act as condensation nuclei to increase the formation of clouds and influence cloud albedo and lifetime (indirect effect), which is found to be the major source of model structural uncertainty (Zelinka et al. 2014; Chung and Soden 2017).

Unlike the well-mixed and monotonically increasing GHG concentrations, AAs have a heterogeneous spatial and temporal distribution, which can induce regional responses that are distinct from those induced by GHGs. Due to their short atmospheric residence time and localized sources, AAs are mainly found in the Northern Hemisphere (NH). Furthermore, AA forcing, represented as aerosol optical depth (AOD), has a distinct time evolution related to the history of industrial development in different nations and their associated emissions regulation. For example, North America and Europe exhibit a distinct evolution of AOD in contrast with that in Asia and Africa (e.g., Fig. 2 of Deser et al. 2020a). This temporal and spatial heterogeneity in AA forcing presents a major opportunity for detecting its climate impact and understanding the associated dynamics. Previous studies examined net changes or a trend across certain periods to investigate GHG effect versus AA effect and showed that AAs are the primary driver of externally forced climate responses before the late 1970s (Deser et al. 2020a). Moreover, during 1980–2020, idealized model simulations suggest the shift of the dominant AA emission regions from North America/Europe to Asia drives a northward shift of the Hadley cell and an equatorward shift of the NH jet stream (Diao et al. 2021).

The heterogeneous spatiotemporal characteristics of AA forcing are shown to induce heterogeneous climate impacts. For instance, based on an energetics framework, the NH cooling induced by AA perturbation requires a northward energy flux across the equator (from the warmer to colder hemisphere) and results in a southward shift of the intertropical convergence zone (ITCZ; Ming and Ramaswamy 2011; Hwang et al. 2013; Wang et al. 2016b; Chung and Soden 2017). Moreover, AAs influence Sahel drought and recovery (Giannini and Kaplan 2019), which is sensitive to the emission location (Dong et al. 2014). In addition, by using large ensemble simulations, AA forcing is found to be the primary driver of the multidecadal variability of the Sahel precipitation, which is consistent with the observations (Hirasawa et al. 2020). Increasing aerosols in Asia is found to account for a weakening of the tropical Walker circulation, while decreasing aerosols in North America/Europe tend to shift the wind systems southward, based on idealized model simulations with the aerosol forcing specified separately in Asia and Asia.
North America/Europe (Wang et al. 2015). A recent study suggests that the evolving aerosol distributions can be represented by two separate modes, namely the zonal mean mode and the zonal asymmetry mode of the aerosol forcing with radiative cooling (warming) in the extratropical Asia (North America/Europe), based on an idealized simulation (Kang et al. 2021). They show that both modes can give rise to a La Niña–like cooling pattern over the equatorial Pacific, which is seemingly inconsistent with the weaker Walker circulation found by Wang et al. (2015). This inconsistency may result because Kang et al.’s (2021) idealized experiment only takes into account the radiative cooling effects in the extratropical region, whereas the actual aerosol forcing extends to the subtropics (e.g., Wang et al. 2015).

The clear detection of the regional fingerprint of AA forcing has been confounded by a signal-to-noise problem, with the noise being due to structural uncertainty across model ensembles and large natural decadal climate variability at regional scales. Besides the structural uncertainty due to model physics, such as the representations of aerosol indirect effects, internal variability gives rise to large uncertainty in climate projections on regional and decadal scales (Deser et al. 2012; Frankignoul et al. 2015). Large ensembles of simulations from a single model with the same forcing but different initial conditions (Kay et al. 2015; Deser et al. 2020a,b) present an opportunity to circumvent the issue of structural uncertainty and average through the natural noise at the same time, making it possible to detect the forced response more robustly (for a single model), on both global and regional scales (e.g., Thompson et al. 2015; Frankignoul et al. 2017).

For a large enough ensemble, the externally forced response can be estimated as the simple ensemble mean and internal variability can be estimated as the deviation from the ensemble mean. Furthermore, the climate response to an individual external forcing, such as AAs, can be isolated by using single forcing or all-but-one-forcing large ensembles of simulations. For example, Deser et al. (2020a) used the Community Earth System Model version 1 (CESM1) all-but-one-forcing large ensembles to show that internal variability limits the detectability of AA- and GHG-forced patterns in individual realizations, with less than 20% of the spatial variance of observed precipitation and sea surface temperature (SST) trends explained by AA and GHG forcing. Indeed, in the 41-yr trend (for 1967–2007) of SST and surface salinity (Fig. S1 in the online supplemental material) from AA single-forcing ensemble (see section 2), the signal-to-noise ratio (ensemble mean divided by 2 times the standard error of ensemble spread; e.g., Fasullo et al. 2020) does not exceed 1 in most of the regions. This highlights the challenge of robustly separating the regional scale AA forced signal from internal climate variability. Note that, rather than measuring the significance of ensemble mean trend as in Deser et al. (2020a), we focus on the signal-to-noise ratio throughout this work as a measure of the statistical significance of the signal, to identify whether the forced responses are detectable from the internal variability.

Rather than a simple ensemble average, we believe that spatiotemporal covariance provides additional information to help extract the forced response from internal variability. Several recent studies have utilized spatiotemporal analysis methods to isolate the forced climate response (Venzke et al. 1999; Deser et al. 2016; Frankignoul et al. 2017; Wills et al. 2018, 2020). For instance, Wills et al. (2020) successfully extracted forced responses due to GHG and AA forcings based on all-forcing large ensemble simulations. In this study, we use the pattern recognition method by Wills et al. (2020) to identify the forced patterns with the maximum signal-to-noise ratio due to AA forcing. The methodology takes advantage of the spatiotemporal covariance across runs, and can better objectively detect the spatial pattern and temporal evolution of the AA-induced responses.

The rest of the paper is organized as follows. Section 2 describes the data and methods used in this study. Section 3 discusses the first two forced patterns of sea surface temperature and salinity and associated time evolutions due to anthropogenic aerosol forcing. Section 4 discusses the constructed fields and indices based on the forced responses. Section 5 is a summary.

2. Data and methods

a. CESM1 large ensemble simulations

The CESM1 is a fully coupled global Earth system model with atmosphere, ocean, land, and sea ice components. The atmosphere model is the Community Atmospheric Model version 5 with a finite volume dynamical core with ~1° horizontal resolution and 30 vertical levels. The CAM5 includes an aerosol scheme (Liu et al. 2012) with prognostic aerosols and indirect effects for both liquid and ice phase clouds (Morrison and Gettelman 2008). The ocean model is the Parallel Ocean Program version 2 with the horizontal resolution at 1.125° in longitude and 0.27°–0.54° in latitude and 60 vertical levels. The land and sea ice components are the Community Land Model version 4 and the Sea Ice Model version 4 with the same horizontal resolutions as the atmosphere and ocean components, respectively.

The first set of experiments we use is the CESM1 Large Ensemble (LENS) with 40 members that start with different initial conditions for the air temperature, but with identical radiative forcing (Kay et al. 2015). All the external anthropogenic and natural forcings are applied to force the fully coupled model following historical (1920–2005) and the representative concentration pathway 8.5 (RCP8.5) scenarios (2006–2100).

In addition to the “all forcing” LENS, a new set of “all-but-one-forcing” CESM1 large ensemble has recently become available for studying the global and regional impacts of industrial aerosol and GHGs over the twentieth and twenty-first centuries (Deser et al. 2020a). This all-but-one-forcing CESM1 experiment uses the same forcing protocol as all forcing LENS except that one forcing agent is fixed at the 1920 level. Here we use a 20-member ensemble with fixed
anthropogenic aerosols (XAER) from 1920 to 2080. We can derive the time-dependent AA single-forcing ensemble (AER) using the XAER and the all-forcing LENS as follows (Deser et al. 2020a):

\[
\text{AER}_i = (\text{XAER}_i - \text{XAER}_{\text{em}}) + (\text{LENS}_{\text{em}} - \text{XAER}_{\text{em}}),
\]

where \( i \) refers to the ensemble member within the 20-member ensemble, and “em” refers to the mean of the 20 members. The original time resolution of all data, such as SST, surface salinity, precipitation, wind stress, etc., is monthly, from which the annual mean values are calculated before the analysis was conducted. We interpolate all the outputs to a regular 2\degree × 2\degree latitude–longitude grid to reduce the computational burden. We also tested the results when the inputs are on 1\degree × 1\degree and 4\degree × 4\degree grids and found that the results are insensitive to the spatial resolution of the input. The following results are based on the 2\degree × 2\degree grid.

b. Evolution of aerosol optical depth

In the 100-yr AOD trend pattern (Fig. 1a) for 1920–2020, East/South Asia and tropical South America/Africa show an increase, while North America and Europe exhibit a weak decrease. Regional time series further illustrate the AOD evolution: North America and Europe exhibit an increase from 1920 to 1980, followed by a rapid decline (Fig. 1b), whereas East and South Asia feature a rapid increase since 1950, peaking in 2010 (Fig. 1c). AOD in tropical South America and Africa rises from 1920 to 1980 and is constant afterward (Fig. 1c).

c. Signal-to-noise-maximizing pattern filtering

First, we conduct an empirical orthogonal function (EOF) analysis on each variable of interest, such as SST, from all
20 members of AER, with the data from each ensemble member concatenated in the time dimension (ensemble data matrix $X$). The EOF patterns are ordered by decreasing total variance explained. Unsurprisingly, the first EOF pattern of SST from AER (Fig. S2; 30% of total variance explained) is due to El Niño-Southern Oscillation, with the ensemble mean of principal component (PC) time series much smaller than that from each member (Fig. S2, right column), a hallmark of internal variability which is unsynchronized between ensemble members. However, the PC time series for EOF2 and EOF3 are more synchronized and are likely to reflect the aerosol-induced responses to greater degrees but with different spatial patterns.

The leading EOF patterns are thus a mixture of forced and internal variability patterns. To better isolate the forced patterns common to each ensemble member, we follow the method of Wills et al. (2020). We define linear combinations of eigenvectors ($u_k$) from the above EOF analysis to help find the maximum signal-to-noise (S/N) ratio:

$$u_k = \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right] e_k,$$

where $u_k$ is the pattern from the linear combination of $N$ leading EOF eigenvectors; $\sigma_i$ is the square root of the $i$th EOF eigenvalue, and $e_k$ is a coefficient vector, which is sought below. In this study, we pick $N$ to retain 90% of the total variance for each variable.

A corresponding time series $t_k$ can be defined by projecting $u_k$ onto the ensemble data matrix $X$, which is concatenated in the time dimension:

$$t_k = Xu_k.$$

To get the robust forced pattern, we need to find the pattern, i.e., the coefficient vector $e_k$, resulting in the maximum ratio of signal to the total variance, which can be written as

$$s_k = \frac{\langle t_k \rangle^T \langle t_k \rangle}{\langle t_k \rangle^T t_k}.$$

The angle brackets indicate the ensemble mean. Thus, the numerator in (4) denotes the signal (variance explained by the ensemble mean) and the denominator denotes the total variance from all ensemble members.

After plugging (2) and (3) into (4), we get

$$s_k = \frac{(e_k)^T \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right]^T \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right] e_k}{(e_k)^T \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right]^T X^T X \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right] e_k}.$$

Since $a_k$ is the EOF eigenvector of $X$, the covariance in the denominator of (5) is equal to 1. Therefore, (5) turns out to be an optimization (maximization) problem by solving for the eigenvalue ($s_k$) and eigenvector ($e_k$) in

$$S_{e_k} = s_k e_k$$

$$S = \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right]^T (X^T X) \left[ \frac{a_1}{\sigma_1} \frac{a_2}{\sigma_2} \ldots \frac{a_N}{\sigma_N} \right].$$

The S/N-maximizing patterns ($v_k$), sorted by $s_k$, are determined by the regression of the ensemble data matrix $X$ onto time series $t_k$:

$$v_k = X^T t_k = \left[ \sigma_1 a_1 \sigma_2 a_2 \ldots \sigma_N a_N \right] e_k.$$

The S/N is calculated as

$$S/N = s_k (1 - s_k)^{-1}.$$

It should be noted that in this framework, the temporal loadings of the forced patterns are required to be orthogonal, while the spatial patterns are not orthogonal in space. Note that the leading S/N-maximizing modes are presented below with its spatial pattern ($v_k$; e.g., Fig. 2a) and the ensemble mean of the corresponding time series ($t_k$; e.g., black curve in Fig. 2b) as well as the $t_k$ broken into each ensemble members (e.g., gray curves in Fig. 2b).

We also examined whether the forced patterns are sensitive to the temporal smoothing on the temperature fields. When a 10-yr low-pass filtering is applied to temperature before applying the S/N-maximizing analysis, the high-frequency noises are removed in time series, which makes the S/N values larger, but the extracted spatial patterns are not affected (Fig. S3).

3. Leading modes of forced climate responses

a. Forced sea surface temperature responses

The first two S/N-maximizing patterns of AER SST anomalies relative to a 1920–2080 mean and the corresponding time series are shown in Fig. 2. The S/N values in these modes are greater than 1, indicating they are common to all members and thus dominated by the forced response. The following modes 3 and 4, clearly have S/N values are less than 1 (Fig. S4), and therefore we focus on the first two patterns and call them the forced patterns in the following analysis.

The first forced pattern (FP1), explaining about 9% of the total yearly SST variance, features a hemispheric asymmetry in SST anomalies (Fig. 2a). In the zonal mean, the cooling in the NH, peaking in the midlatitudes, is much larger than that in the SH. The corresponding time evolution rapidly rises from negative values in 1920 (warmer NH) to a positive maximum in the late 1980s (colder NH), and then declines to negative values toward 2080 (warmer NH), indicating that the NH SST gradually warms after 1980 (Fig. 2b). The time series of globally averaged AOD matches well with the FP1 time series (Fig. 2b). Thus, we assert that FP1 reflects the global mean effect of AAs (dominated by the NH sources) on SST.
mid-1980s. Hence, the first mode is primarily associated with temperate NH aerosol forcing. The AOD time series of 0°–60°N average is also consistent with the FP1 time evolution (not shown).

FP1 has the largest SST anomalies in the midlatitude North Pacific, associated with an intensified Aleutian low and counterclockwise atmospheric circulation anomalies (vectors in Fig. 2a; Ming et al. 2011; Wang et al. 2016a). This strong

FIG. 2. (a) First forced pattern (FP1) of AER SST anomalies from 1920 to 2080 with the zonal mean pattern (black curve; °C) shown on the left. Note that these patterns are relative to a 1920–2080 mean. (b) Standardized time evolution of the first mode. The patterns are shown in units of degree Celsius per standard deviation of time evolution. Gray curves show the time series for each ensemble member, and the black curve denotes the ensemble mean of 20 members. The red curve shows the global mean AOD from AER. (c),(d) As in (a) and (b), but for the second mode (FP2 and corresponding time evolution). The blue curve in (d) shows the time series of the difference between AOD in Asia/tropics and North America/Europe, based on the regions defined in Fig. 3b. The regressions of the net top of the atmosphere (TOA) radiative flux (downward positive; 0.4 W m⁻² interval with 0 contour omitted) and surface wind stress on the respective ensemble mean time series are shown as contours (red denotes positive change) and arrows, respectively, in (a) and (c). The contours and arrows are plotted only when statistically significant at 5%. The zonal mean patterns of the TOA flux regression are shown as red curves on the left.

FIG. 3. (a) Regression of AOD from AER on the FP1 time evolution in Fig. 2b. (b) Regression of AOD from AER on the FP2 time evolution in Fig. 2d.
cooling, peaking in the 1980s, can be attributed primarily to localized reduced downward top of the atmosphere (TOA) radiation (contours in Fig. 2a). Anomalous Ekman upwelling can also contribute to the surface cooling here. All the NH oceans cool, except for a warming spot located in the North Atlantic subpolar gyre. Here, the changes in atmospheric circulation or TOA radiative fluxes are not located over the warming spot, but to its east. As suggested by previous studies (e.g., Shi et al. 2018; Irving et al. 2019), AA forcing leads to a strengthening of the Atlantic meridional overturning circulation (AMOC) during the historical period (Fig. S5a), likely compensating a weakening due to GHG forcing in the natural world. We believe this regional warming spot results from the AMOC response. As suggested by the FP1 time series, the AMOC will relax along with the anticipated reduction of aerosol forcings in our modeled future scenarios. In the tropics, the northeasterly trade winds are intensified and southeasterlies are weakened associated with the change of meridional SST gradient, following the wind–evaporation–SST feedback (Xie 1996).

The zonal mean SST pattern of FP2 shows positive anomalies in the NH extratropical latitudes with negative anomalies in tropics and SH (Fig. 2c). The corresponding time evolution illustrates that this change occurs primarily between ~1965 and ~2020 (Fig. 2d) remaining constant or slightly decreasing after 2020. The AOD regression pattern on the time evolution of FP2 shows opposing changes between North America/Europe and East/South Asia (also the tropics) (Fig. 3b). The AOD time series of the difference between the Asia/tropics and the North America/Europe regions (using the regions defined in Fig. 1a) is broadly consistent with the FP2 time evolution (Fig. 2d). Thus this forced pattern is associated with the transition of the dominant aerosol emissions from North America/Europe to East/South Asia and the tropics.

The weakened aerosol forcing in North America/Europe results in a relative warming of the NH extratropics (Fig. 2c). It gives rise to a weakening of westerly winds in the temperate NH associated with a low sea level pressure anomaly (not shown). The weakened surface wind in the North Pacific decreases turbulent heat loss from ocean to atmosphere (e.g., positive anomalies in Figs. 5a, b, and S7) thus results in warmer SST. Northward anomalous Ekman transport can also contribute to the warming. The North Atlantic warming is largely attributed to enhanced shortwave radiation reaching the surface (Figs. 6c and S7) associated with the decrease of AA in North America/Europe. The AMOC weakens mainly during 1980–2020 (Fig. S5b) when it is regressed on the FP2 time series. This is consistent with the broad North Atlantic warming and freshening (next subsection). The cooling spot in the deep convection regions is not notable as the FP2 AMOC response is much weaker than that of FP1. Instead, the warming spot takes place south of the deep convection region, likely is associated with the increased TOA flux and also weaker equatorward Ekman transport. Over Africa and South Asia, the strong negative signals in the TOA flux reflect the contrasting change of aerosol forcing to those over North America/Europe.

Cooling occurs in the tropical Indo-Pacific associated with the localized increasing aerosol forcing. The weakening of easterlies in the central and east equatorial Pacific signals the weakening of Walker circulation due to cooling in the Indo-Pacific warm pool. FP2 also shows cooling patterns in the SH, especially the cooling band in the Southern Ocean, in contrast with the negligible Southern Ocean anomalies from FP1. This SH cooling is in line with the SH-integrated TOA flux responses (~15.0 TW based on contours in Fig. 2c). If we apply the same pattern analysis method to just the Southern Ocean, defined as the region to the south of 30°S (Fig. S8), the leading forced pattern of the Southern Ocean SST response shows a broad cooling, resembling the FP2 for the global domain. The regressed wind stresses are mostly insignificant. The broad cooling is damped by the positive heat flux responses (ocean gains heat from the atmosphere) to some extent in the midlatitudes (~30°–55°S). The heat losses in the high latitudes (~55°–65°S; green contours in Fig. S8), where pristine deep water outcrops at the surface and interacts with the cooled air, are very strong and consistent with the broad cooling pattern over the Southern Ocean. This is similar to the results found by Shi et al. (2020, 2021), suggesting that a heat flux change along the poleward flank of the Antarctic Circumpolar Current can broadly affect Southern Ocean temperature change even without changes in wind, as the mean overturning circulation can advect the cooled water equatorward. This is a possible mechanism to explain the aerosol-induced surface cooling, although further analysis is needed to investigate the subsurface responses (i.e., OHC change), associated ocean circulation adjustments, and also sea ice effect.

In addition, compared with FP1, FP2 drives an opposite effect on the hemispheric energy imbalance, and thus ITCZ location (Fig. 4c and Fig. S9). A northward displacement of the ITCZ in FP2 is evident in the strengthened trade winds in the South Atlantic and the weakened trade winds in the North Atlantic and North Pacific, and associated salinity and freshwater flux responses (Fig. 4c). We examine these more below.

b. Forced surface salinity responses

To test whether similar forced patterns emerge for other variables, we analyze the AER surface salinity anomalies (Fig. 4), surface air temperature, precipitation, sea ice, and sea surface height (Fig. S9). The first two modes of all these variables are dominated by the forced responses as indicated by an S/N greater than 1. The S/N values are particularly high in salinity and sea surface height, likely due to the integrating effects of the ocean of the weak, but large-scale decadal forcings we are focusing on here. The first two corresponding time series from each variable are broadly consistent with those from SST analysis. In this subsection, we focus on the extracted distinct AA-forced patterns of surface salinity anomalies.

The FP1 of salinity shows a strong meridional dipole pattern in each tropical ocean basin—a zonal band of salinification to the north of the equator and freshening to the south.
The superimposed net precipitation \((P - E)\) regression closely matches the salinity response. This clearly is due to a shift of ITCZ toward the warmer SH, which has been discussed in the previous literature (e.g., Hwang et al. 2013). The salinification of the northern polar region primarily reflects the AMOC enhancement (Fig. S5) and intensified sea ice formation due to the AA-induced cooling (Fig. S9). Note that the regressions of surface salinity on SST FP1 time series (Fig. S10a) are broadly consistent with the salinity forced patterns shown in Fig. 4a.

FP2 features more zonal asymmetry (Fig. 4c) which mainly develops prior to 2010 according to the corresponding time series (Fig. 4d). Freshening takes place in the North Atlantic except the small patch of salinity increase south of the Labrador Sea, and salinity increases in the tropical Indo-Pacific centered over the Maritime Continent and also the extratropical North Pacific. These salinity responses match with the precipitation responses (Fig. 4c). The contrasting responses of the Indo-Pacific and North Atlantic are mediated by their opposing SST changes (Fig. 2c) following the warmer-get-wetter pattern (Xie et al. 2010). This suggests a weakening in the rainy convection systems over the Maritime Continent, and a compensating strengthening of those in the eastern tropical Pacific and tropical North Atlantic. The strong signal along the west coast of North America may be driven by enhanced upwelling of salty subsurface water due to anomalous northerly winds. Again the time evolution implies that FP2 captures the effects of AA-forcing transition on salinity. As for the SST analysis, the salinity forced patterns are insensitive to the temporal smoothing (not shown). Note that the regressions of surface salinity on SST FP2 time series (Fig. S10b) show some regional differences, such as the freshening in subpolar North Atlantic and Arctic, when compared with salinity forced patterns in Fig. 4c. It is also true for the sea ice pattern: the regression of sea ice on SST FP2 time series shows more formation centered around the east of Greenland and farther poleward consistent with localized cooling over the Greenland Sea, and a broad shrinking close to continents (not shown). These indicate different variables have different sensitivities to aerosol forcing, which can also be found in their time series of forced patterns. As a caveat, since the leading modes analyzed independently for various variables have some unique regional responses, it seems proper to discuss the relationship between different variables based on the same time series.
4. Reconstruction of surface temperature and salinity responses

Although the aforementioned first two forced patterns separately provide insight, we would like to examine the history of the total AA forced response by constructing its evolution based on the truncated dataset from the $M$ leading modes:

$$\langle T \rangle = \sum_{k=1}^{M} t_k \langle v_k \rangle^T.$$  \hspace{1cm} (10)

To test if the choice of $M$ affects the reconstructed total forced response, we split the 20-member AER in half, apply the S/N-maximizing method to the first 10-member subensemble, and construct the total response by changing $M$. This total response is then compared with the ensemble mean of the other 10-member subensemble. The global mean of the grid-point correlation between two total responses is shown in Fig. 5. There is a jump between the correlations by using $M = 1$ and $M = 2$. There is also a jump between $M = 1$ and $M = 2$ in the corresponding analysis for the surface salinity (Fig. S11), although the correlation coefficients are smaller than those from SST. It demonstrates that at least two leading modes should be included to reconstruct the total forced response, which is consistent with the S/N dropping rapidly below 1 for mode 3 and higher. When $2 < M < 20$, the agreement between the reconstructed pattern and ensemble mean pattern does not improve much. Hence, we choose the value of $M = 2$ to focus on the forced responses for the analysis that follows.

The historical 60-yr total AA forced SST change (1980–2010 average minus 1920–50 average) exhibits a broadscale cooling pattern in both hemispheres (Fig. 6a), which reflects the combination of the first two forced patterns as the FP1 primarily drives the total change while the FP2 dampens the response in NH and strengthens it in the SH. In comparison, the future projected change (2040–70 average minus 1980–2010 average) shows a strong NH warming with small and insignificant change in the SH, which reflects the effect of a global aerosol decrease captured by the FP1 (Fig. 6b). The surface salinity responses are consistent with the corresponding FP1 (Fig. 6e). The sum of the future and historical change reflects the effect of spatial heterogeneity of aerosol forcing (Fig. 6c) captured by the FP2 since the 1920–50 average and 2040–70 average of global mean aerosol concentrations are comparable (Fig. 2b). In addition, the strong zonal asymmetry associated with the surface salinity FP2 can be found in the sum of historical and future changes (Fig. 6d).

We also calculate regional indices based on the reconstructed SST and precipitation. The NH mean SST mainly follows the time evolution of FP1 and recovers to the 1920s level by 2080, while the SH mean follows the time series of FP2 (Fig. 7a) with a long-lasting cooling signal, indicating the aerosol transient effects from NH extratropics to tropics have a greater influence on the SH climate. This hemispheric difference shows that the NH will become warmer than SH in the late twenty-first century when only the AA forcing is considered. In addition, we define a precipitation asymmetry index, $P_{\text{ind}} = P_{1920-2080, N} - P_{1920-2080, S}$, which reflects the change in the latitudes of the ITCZ, and a cross-equator atmospheric heat transport (northward positive) to reflect the hemispheric energy imbalance. $P_{\text{ind}}$ decreases before the 1980s and increases steadily during the twenty-first century (black curve in Fig. 7b), which is consistent with the time series of hemispheric warming asymmetry (Fig. 7a). The anomalous heat transport (brown curve in Fig. 7b) keeps compensating the interhemispheric imbalance. These indicate that the ITCZ shifts in different directions with the changing hemispheric energy imbalance driven by the changing location and overall amplitude of AA forcing. The ITCZ displacement will bring more precipitation in the NH tropics after ~2050 relative to the 1920s condition.

The detectability of AA forced trend patterns (i.e., ensemble mean trend) in individual realizations varies in time as found by Deser et al. (2020a). Here we use the time information from two leading modes to identify the period during which the trends are dominated by the spatial heterogeneity of aerosols. First, the 41-yr running trends of the two FP time series were calculated (not shown). The period with the largest ratio of the absolute value of the second time series trend to that of the first time series trend is when the aerosol transition effects are largest, 1967–2007 (centered at 1987) from both SST and surface salinity. During this period, the SST trend shows contrasting responses between the warming in the NH extratropics and cooling in other regions, including the Indo-Pacific, the Indian sector of the Southern Ocean, and the southeastern Atlantic and Pacific (Fig. 8a), which reflect the dominant influence of the FP2 (Fig. 2c). The salinity trends show a clear zonal asymmetry with a freshening in subtropical North Atlantic and a salinity increase in the Indo-Pacific (Fig. 8b).

It is noteworthy that, compared with the trends for the same period from the simple ensemble mean (Fig. S1), the trends of the FP reconstructions improve the detectability of the total
forced response, as the significant regions are much greater in the latter. Moreover, the trend of the ensemble mean SST has an interdecadal Pacific oscillation–like structure in the Pacific (Fig. S1), which is not evident in the reconstructed field based on the two forced modes. As contributions from the higher modes are dominated by the internal variability according to the S/N, this suggests that internal variability is still impacting the regional responses in the simple ensemble mean of 20 members. Also note that the simple ensemble mean contains higher-frequency signals likely from internal variability in the precipitation asymmetry and cross-equator atmospheric heat transport indices, unlike the corresponding time series from the reconstruction based on the forced patterns (Fig. 7b). All in all, these collectively reveal the deficiency of the ensemble mean approach using only 20 members for the relatively weak (vs GHG) AA forcing (cf. Deser et al. 2020a).

5. Summary

In this study, we use a signal-to-noise-maximizing pattern recognition method to isolate the forced climate response to anthropogenic aerosols from that of internal variability from...
1920 to 2080. We find that the aerosol forced response of various fields (SST, salinity, precipitation, sea surface height, etc.) are dominated by two components: one is associated with the historical increase and future decrease of global (or NH) mean aerosol concentrations; the other is due to the transition of anthropogenic aerosols emissions from the west to the east and also from NH extratropics to lower latitudes. The global mean effect gives rise to a dominant NH cooling (warming) and a southward (northward) ITCZ shift during the historical period (future projection). The AA transition effect leads to a warming in the NH extratropical regions and a cooling in the Indo-Pacific and Southern Ocean, which to some extent compensates for the hemispheric asymmetry due to the global mean effect in the historical period. Moreover, the aerosol transition effect generates a zonal asymmetry in surface salinity response: a freshening in the North Atlantic and a salinity increase in the Indo-Pacific. Importantly, we show that the AA-driven trends during recent decades, e.g., 1967–2007, are

![Figure 7](image1.png)

**FIG. 7.** (a) The hemispheric means of AER total SST reconstruction (thick red and blue curves) and hemispheric difference (thick black curve). (b) Precipitation asymmetry index (black curve), $P_{\text{ind}} = P_{0.2070} - P_{0.5087}$, from AER total precipitation reconstruction, and the cross-equator atmospheric heat transport (thick brown curve; northward positive), $AHT = \frac{\left(\text{TOA}_\text{NH} - \text{Q}^\text{net}_\text{SH}\right)}{2}$. The $\text{TOA}_\text{NH}$ is the NH-integrated TOA flux minus the SH-integrated one, and $\text{Q}^\text{net}_\text{NH}$ is the NH-integrated surface heat flux minus the SH-integrated one. The thin curves are corresponding time series from the simple ensemble mean of all 20 members. All of the time series are relative to the year 1920.

![Figure 8](image2.png)

**FIG. 8.** (a), (b) The 1967–2007 trend of reconstructed ($M = 2$) SST and surface salinity fields, respectively. The hatched regions show where the ensemble mean trend is smaller than twice the ensemble standard error.
dominated by the aerosol transition effect rather than its global (or NH) mean effect.

The temporal and spatial heterogeneity of AA forcing has presented a major challenge for detecting its climate impact and understanding the associated response dynamics. Separating the aerosol forcing in the west and east in individual model simulations has been helpful (e.g., Wang et al. 2015; Diao et al. 2021). For example, the aerosol forcing in North America/Europe dominates the North Pacific SST responses versus the forcing in Asia (Diao et al. 2021). However, in real life, aerosol forcings from these regions act simultaneously—an interaction not captured in idealized single emission location experiments. Moreover, based on idealized simulations, Kang et al. (2021) recently showed that the simultaneous but opposite signed forcing in North America/Europe and Asia located in the NH extratropics drives a strong zonal asymmetry in surface temperature response with a La Niña–like cooling pattern over the equatorial Pacific. We find that the AA transition effect also includes the changes in the primary source regions of AA from NH extratropical regions to SH tropical regions, which was not considered in Kang et al. (2021). Interestingly, neither two leading modes identified here show a La Niña–like pattern. Rather, 1) the first mode here has a southward shift of ITCZ with distinct responses across the equator associated with the different trade wind changes, and 2) the second mode shows a clear zonal asymmetry over the equatorial Pacific due to the increase (decrease) of AA forcing in East and South Asia (North America) with the weakened wind, which results in weakened zonal SST gradient over the equatorial Pacific, thus somewhat El Niño–like. The aerosol forcing at lower latitudes, which is omitted in Kang et al. (2021), may predominantly drive the El Niño–like responses.

In further exploring the tropical response, we apply an EOF analysis on the AER ensemble mean (Fig. S12). We find that the first mode matches well with the result from the S/N-maximizing analysis. However, the second EOF largely reflects our FP2, but also shows a distinct cooling in the central equatorial Pacific. In addition there is prevalent interannual variability evident in the corresponding principal component, which is suppressed in the FP2 series. This equatorial Pacific cooling also appears in the third EOF mode, which is dominated by an interdecadal Pacific oscillation–like pattern. Therefore, it is likely that the cooling in the equatorial Pacific in EOF 2 represents residual internal variability, which is not perfectly removed in the 20-member ensemble mean. We believe that the forced patterns captured in this study avoid these issues to some extent and reveal the effect of heterogeneity of AAs more clearly, in contrast with the effect of the global mean AA.

In addition, we show that the spatial patterns of the AA-induced change or trend are very sensitive to the choice of period used, due to the interplay between the two leading modes. For example, the 1967–2007 trends shown here (Fig. 8 and Fig. S1) are quite distinct from the 1930–79 or 1950–99 trends shown in Figs. 3d and 3e in Deser et al. (2020a). Rather than choosing a period arbitrarily, we can use the information obtained from the time evolution of the forced patterns to find the periods during which the aerosol transition effect is the strongest, and thus most detectable.

Our results suggest that the signal-to-noise-maximizing pattern recognition method can more clearly isolate the AA-forced responses from the internal climate variability than a simple ensemble average of 20 members. As a caveat, we note that the ensemble spread achieved in this study only represents the internal variability resulting from perturbed atmospheric initial conditions (Kay et al. 2015). It remains unknown how perturbing the ocean state might affect the internal variability captured by the ensemble. The CESM2 Large Ensemble simulations, combining the different initial oceanic states and atmospheric states, would be helpful to understand this issue (Rodgers et al. 2021).

Besides internal climate variability, GHGs will hinder the detection of aerosol effects in the real climate system, since aerosol forcing is relatively weak and both have similar global time tendencies over most of the instrumental record. To contrast the responses, though, we conduct the same pattern recognition method on GHG simulations, which is calculated from Eq. (1) by substituting XGHG (all-but-GHG-forcing) for XAER. GHGs give rise to a monotonic increase in global surface temperature (Fig. S13). Furthermore, we also conduct the same analysis on the all-forcing large ensemble (LENS) from 1920 to 2080. The aerosol transition effect is not captured in the LENS (not shown). The results remain the same when we try to remove the first mode associated with the monotonic GHG effect and then repeat the pattern recognition analysis. It is likely that the weak and quasi-monotonic (as shown in Fig. 2d) aerosol transition effect is indistinguishable from the strong GHG response due to their coherent temporal evolution. In addition, although the extracted forced patterns are useful to identify the AA fingerprint, the different responses across models can pose an obstacle to making a robust statement about the real world. These increase the difficulty to isolate the AA signals in observations, which deserves further investigations.

Finally, our results imply that in future decades, if AA emissions decline as projected in RCP8.5, the NH extratropics will warm much more rapidly than the other parts of Earth, moving the ITCZ farther north with its related impact on global rainfall patterns. Given the large spread in AA responses across climate models (Myhre et al. 2013), verifying and improving the simulation of AA impact on global climate remains a critical challenge of relevance to improving climate projections.

Acknowledgments. We gratefully acknowledge the thoughtful and constructive comments from the editor Dr. Isla Simpson, the reviewer Dr. John Fasullo, and two other anonymous reviewers. All three authors are supported by U.S. National Science Foundation (OCE-2048336). The Community Earth System Model project is supported primarily by the National Science Foundation (https://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html and https://www.cesm.ucar.edu/working_groups/CVC/simulations/cesm1-single_forcing_le.html).
REFERENCES


