

# The Role of Forcings in the Twentieth-Century North Atlantic Multidecadal Variability: The 1940–75 North Atlantic Cooling Case Study

A. BELLUCCI

*Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy*

A. MARIOTTI

*NOAA, Climate Program Office, Silver Spring, Maryland*

S. GUALDI

*Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, and Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy*

(Manuscript received 12 April 2016, in final form 25 May 2017)

## ABSTRACT

Results from a study inspecting the origins of multidecadal variability in the North Atlantic sea surface temperature (NASST) are presented. The authors target in particular the 1940–75 “warm-to-cold” transition, an event that is generally framed in the context of the longer-term Atlantic multidecadal variability (AMV) cycle, in turn associated with the Atlantic meridional overturning circulation (AMOC) internal variability. Here the authors examine the ability of uninitialized, historical integrations from the phase 5 of the Coupled Model Intercomparison Project (CMIP5) archive to retrospectively reproduce this specific episode of twentieth-century climatic history, under a hierarchy of forcing conditions. For this purpose, both standard and so-called historical Misc CMIP5 simulations of the historical climate (combining selected natural and anthropogenic forcings) are exploited. Based on this multimodel analysis, evidence is found for a significant influence of anthropogenic agents on multidecadal sea surface temperature (SST) fluctuations across the Atlantic sector, suggesting that anthropogenic aerosols and greenhouse gases might have played a key role in the 1940–75 North Atlantic cooling. However, the diagnosed forced response in CMIP5 models appears to be affected by a large uncertainty, with only a limited subset of models displaying significant skill in reproducing the mid-twentieth-century NASST cooling. Such uncertainty originates from the existence of well-defined behavioral clusters within the analyzed CMIP5 ensembles, with the bulk of the models splitting into two main clusters. Such a strong polarization calls for some caution when using a multimodel ensemble mean in climate model analyses, as averaging across fairly distinct model populations may result, through mutual cancellation, in a rather artificial description of the actual multimodel ensemble behavior.

A potentially important role for both anthropogenic aerosols and greenhouse gases with regard to the observed North Atlantic multidecadal variability has clear implications for decadal predictability and predictions. The uncertainty associated with alternative aerosol and greenhouse gas emission scenarios should be duly accounted for in designing a common protocol for coordinated decadal forecast experiments.

## 1. Introduction

The North Atlantic basin stands out for the prominent multidecadal variability featured by the observed SST record, alternatively offsetting and enhancing the

underlying warming trend (Deser and Blackmon 1993; Kushnir 1994; Ting et al. 2009). These fluctuations, due to their apparent oscillatory behavior, with a typical 60–70-yr time scale, have been associated with a low-frequency natural variability mode, termed the Atlantic multidecadal oscillation (AMO) or variability (AMV). This low-frequency fluctuation reverberates in a number of hydroclimatic, societally relevant features, including Mediterranean surface temperatures (Marullo et al. 2011; Mariotti and Dell’Aquila 2012), summertime climate over North America and western Europe (Sutton and

---

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-16-0301.s1>.

---

Corresponding author: Alessio Bellucci, [alessio.bellucci@cmcc.it](mailto:alessio.bellucci@cmcc.it)

DOI: 10.1175/JCLI-D-16-0301.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

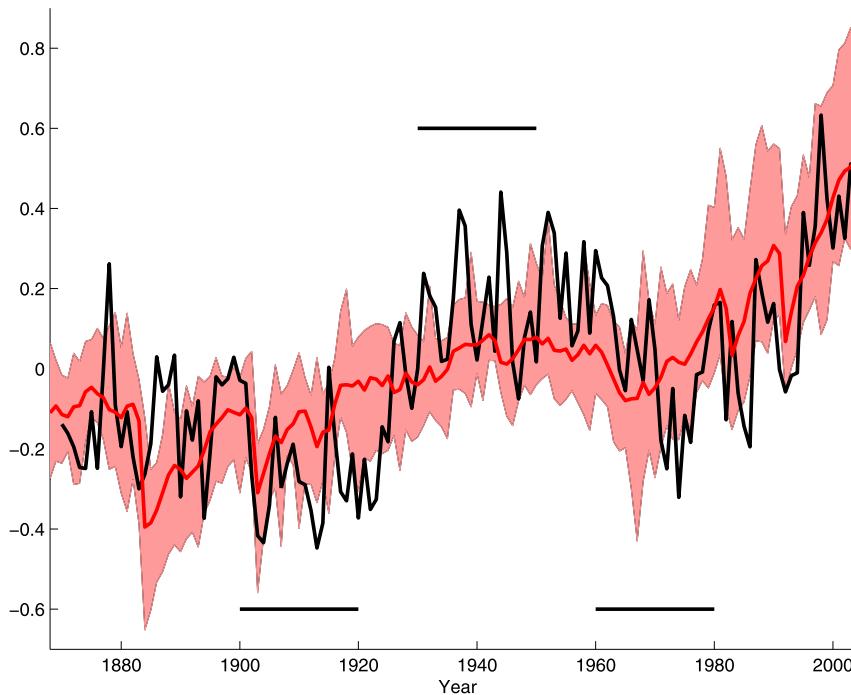


FIG. 1. Area-averaged North Atlantic (defined as  $0^{\circ}$ – $60^{\circ}$ N and  $7.5^{\circ}$ – $75^{\circ}$ W) SST anomalies relative to the 1870–2005 period, for an observational estimate (the HadISST dataset; solid black) and a multimodel ensemble of 13 CMIP5 historical integrations. The multimodel ensemble mean (MME) is in solid red, while light red shading represents the [min, max] envelope of the full multimodel historical ensemble. Thick black horizontal lines indicate the periods used to define the warm and cool phases of the observed SST variability.

Hodson 2005), rainfall variability over the Sahel (Mohino et al. 2011), and Atlantic hurricane activity (Trenberth and Shea 2006), among others.

The origin of the AMV is at the core of a contentious issue. According to a widely accepted picture, the AMV is driven by the internal variability of the Atlantic meridional overturning circulation (AMOC) (Knight et al. 2005), with implications for its predictability, provided that the processes underlying AMOC variability are understood and correctly represented in models. However, this paradigm has been recently questioned by several authors.

Otterå et al. (2010) have discussed the potential role of solar forcing variability and volcanic aerosols on the phase of the AMV. Booth et al. (2012), based on simulations performed with a single model, suggest a significant role for anthropogenic aerosols in determining the North Atlantic multidecadal variability. This result has been countered by Zhang et al. (2013), who suggest that the model used in the analyses by Booth et al. may suffer from an overly strong response to aerosol effects. More recently, Clement et al. (2015) suggested that AMV may be the response to stochastic forcing from the midlatitude

atmospheric circulation, thus ruling out the AMOC as a primary driver of AMV.

In this paper, the role of external forcings on the North Atlantic SST decadal variability over the period 1870–2005 is inspected, with a primary focus on the 1940–75 transient (herein called mid-20CT) case study. During this period the observed basin-averaged North Atlantic SSTs underwent a progressive decline (Fig. 1), interspersed with rapid drops occurred in the mid-1940s (attributed to an uncorrected instrumental bias; Thompson et al. 2008) and the late 1960s (Thompson et al. 2010; Hodson et al. 2014). As noted by some authors on the basis of observational investigations, the mid-twentieth century North Atlantic cooling has occurred in concomitance with other significant climatic changes that affected several areas surrounding the Atlantic sector (Baines and Folland 2007).

Several mechanisms and processes have been invoked to explain the Atlantic cooling [see Hodson et al. (2014) for an extensive review]. These include the propagation from the Arctic to the subpolar basin of a large-scale, cold, and low-salinity anomaly (the so-called Great Salinity Anomaly; Dickson et al. 1988); a slowdown of the

TABLE 1. List of the CMIP5 experiments used in this study, with relative forcing set description, based on the CMIP5 terminology (Taylor et al. 2009, 2012). Following the CMIP5 vocabulary for abbreviated forcing descriptors (available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5\\_data\\_reference\\_Appendix1-2.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_Appendix1-2.pdf)), “Nat” indicates natural forcing (a combination that typically includes solar and volcanic forcing), “Ant” indicates anthropogenic forcing (a mixture that typically includes well-mixed GHGs, aerosols, ozone, and land-use changes), “AA” indicates anthropogenic aerosols, and “GHG” indicates well-mixed greenhouse gases.

Ens. label	CMIP5 simul.	Forcing abbreviation	Reference
HIST	Historical	Nat, Ant	Table 3.2 in Taylor et al. (2009)
ANT	Historical Misc	Ant	<a href="http://cmip-pcmdi.llnl.gov/cmip5/docs/historical_Misc_forcing.pdf">http://cmip-pcmdi.llnl.gov/cmip5/docs/historical_Misc_forcing.pdf</a>
NoAA	Historical Misc	As HIST except for anthropogenic aerosols	
AA	Historical Misc	AA	
GHG	Historical Misc	GHG	

Atlantic branch of the meridional overturning circulation that could have determined the observed cooling pattern via a reduction in the poleward heat transport (Rahmstorf et al. 2015); the early 1960s occurrence of several volcanic eruptions of Mount Agung, which in turn might have determined a local cooling via different processes, ranging from the direct radiative effect associated with the injection of sulfate aerosols in the stratosphere, or indirectly through the dynamical adjustment of the atmospheric and oceanic circulation (Swingedouw et al. 2015); an approximately 10-yr decline in the solar radiation intensity that started in the late 1950s, possibly amplified over the Atlantic by regional processes (Hodson et al. 2014; see also Thiéblemont et al. 2015, for a discussion of the 11-yr solar cycle impact on the decadal variability in the Atlantic sector); and the increase in anthropogenic aerosols emissions over the United States and Europe, starting around the middle of the twentieth century and peaking during the 1970s–1980s (Smith et al. 2011), which was potentially responsible for the observed cooling of the North Atlantic ocean surface through direct radiative effects, and indirectly via the interplay with cloud albedo and life cycle (Booth et al. 2012; Bellucci et al. 2015).

The variety of physical mechanisms and processes identified as potential drivers of the mid-twentieth-century Atlantic cooling highlights that a clear causal attribution is still matter of debate.

Here, the ability of CMIP5 models in reproducing this specific episode of Earth’s climatic record is analyzed. In particular, the role of external (both natural and anthropogenic) drivers is examined in a large ensemble of CMIP5 uninitialized integrations of the historical period performed under a hierarchy of forcing conditions. These include standard simulations carried out under the widely adopted CMIP5 Historical protocol (Taylor et al. 2009, 2012) and a subset of “historical Misc” simulations, the latter consisting of idealized simulations of the historical period for selected combinations of

anthropogenic and/or natural forcing agents (see [http://cmip-pcmdi.llnl.gov/cmip5/docs/historical\\_Misc\\_forcing.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/historical_Misc_forcing.pdf) for a detailed description of the protocol). The following key question is addressed: What are the relative roles of internal and externally forced variability on the observed mid-twentieth-century North Atlantic cooling? We provide some evidence that CMIP5 uninitialized integrations of the historical climate show some skill in capturing the mid-20CT event, and part of this skill can be attributed to the effect of anthropogenic forcings. However, the present analysis does also highlight a strong model uncertainty affecting the latest generation climate models’ response to nonstationary forcings: homogeneous behavioral clusters are identified within the CMIP5 model population, with only a subset of the analyzed historical integrations displaying ability to reproduce the observed mid-20CT. The study has implications for future experimentation under the CMIP6 protocol, addressing shortcomings of CMIP5 (Eyring et al. 2016).

The remainder of this manuscript is structured as follows. Section 2 describes the methodology and the models used for the present analysis. The results are reported in section 3. Finally, the main results of this work are summarized and further discussed in section 4.

## 2. Methods and models

For the present analysis, five multimodel ensembles of CMIP5 integrations for the historical 1870–2005 period are examined (Table 1). The analyzed ensembles include standard historical simulations (HIST) and four sets of integrations performed following the historical Misc protocol (HM) consisting of idealized simulations of the historical period for selected combinations of (or single) forcing agents (see [http://cmip-pcmdi.llnl.gov/cmip5/docs/historical\\_Misc\\_forcing.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/historical_Misc_forcing.pdf) for a detailed description of the protocol). The HIST set includes both natural (volcanoes and solar) and anthropogenic (greenhouse gases, anthropogenic aerosols, ozone, and

TABLE 2. Synoptic table listing the models and the corresponding experiments analyzed in this study. For each model/experiment combination, the corresponding cluster typology as identified by [TR, *m*] couplets (see text) and ensemble size (within parentheses) are reported. (Expansions of acronyms are available online at <http://www.ametsoc.org/PubsAcronymList>.)

Model	Institute	HIST	ANT	NoAA	AA	GHG
BCC-CSM1.1	BCC	A(3)				
CanESM2	CCCma	B(5)			B(5)	A(5)
CCSM4	NCAR	A(6)	A(4)		B(3)	A(5)
CESM1-CAM5	NSF-DOE-NCAR		B(3)			
CNRM-CM5	CNRM-CERFACS	B(10)	C(10)			
CSIRO-Mk3.6.0	CSIRO-QCCCE	B(10)	B(10)	C(5)	B(5)	A(5)
GFDL-CM3	NOAA GFDL		B(3)			
GISS-E2-H	NASA-GISS	A(5)	A(5)	A(5)	B(5)	A(5)
GISS-E2-H emis.	NASA-GISS		A(5)		B(5)	
GISS-E2-H (AA direct eff. only)	NASA-GISS				B(5)	
GISS-E2-R	NASA-GISS	C(5)	A(5)		B(5)	A(5)
GISS-E2-R emis.	NASA-GISS		A(5)		B(5)	
GISS-E2-R (AA direct eff. only)	NASA-GISS				B(5)	
HadGEM2-ES	MOHC	B(4)				A(4)
IPSL-CM5A-LR	IPSL	C(6)	A(3)	A(4)		A(5)
IPSL-CM5-MR	IPSL					A(3)
MIROC5	MIROC	B(6)				
MIROC-ESM	MIROC					A(3)
MPI-ESM-LR	MPI-M	A(3)				
MRI-CGCM	MRI	C(3)				
Nor-ESM1-M	NCC	A(3)				

land use change) forcing. From the grand ensemble of available HM integrations, we specifically focus on the following experimental sets: simulations performed under anthropogenic-only forcings (ANT); simulations forced with the same HIST forcings except for anthropogenic aerosols (NoAA); simulations forced with only time-varying anthropogenic aerosols (AA); and simulations forced with only time-varying green-house gas concentrations or emissions (GHG). As the main focus of this study is to assess the climatic response to non-stationary forcings, only those models providing multiple realizations of the historical climate (with a minimum three-member set) were selected when assembling the multimodel ensembles. Through this approach, an approximate representation of the forced response in the CMIP5 models' population is provided via ensemble averaging, the latter contributing to filter out the uncorrelated internal variability across models and members.

Table 2 lists the models and the corresponding experiments analyzed in this study as well as their relative ensemble sizes. Following the above mentioned constraint on the ensemble size, we selected 13 models for the HIST ensemble (totaling 69 members), 10 models for the ANT ensemble (totaling 53 members), 3 models for the NoAA ensemble (totaling 14 members), and finally 9 models for both the AA and GHG ensembles (totaling 43 and 40 members, respectively). For one of the NoAA models (GISS-E2-H) the only aerosol indirect

effect (acting through aerosol-cloud interaction) was inactive, while the aerosol direct effect (acting through modification of the radiative transfer) was included. However, given the well-ascertained dominant role of the indirect effect on the total aerosol-driven cooling (Wilcox et al. 2013; Levy et al. 2013), this internal inconsistency in the NoAA ensemble can be considered of second order.

In this study we characterize the North Atlantic SST variability via an index defined as the area-weighted average calculated over the [0°–60°N, 75°–7.5°W] domain (the NASST index). Finally, the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) dataset (Rayner et al. 2003) is used as observational reference.

### 3. Results

#### *a. Multidecadal variability in CMIP5 historical simulations*

In the present section, the multidecadal SST variability in the North Atlantic region during the 1870–2005 period in the HIST multimodel ensemble, and the relative consistency with the observed anomalies are assessed.

The time evolution of NASST anomalies in the HIST ensemble and observations is shown in Fig. 1. For HIST integrations, both the multimodel ensemble mean (MME) and the corresponding [min, max] envelope are

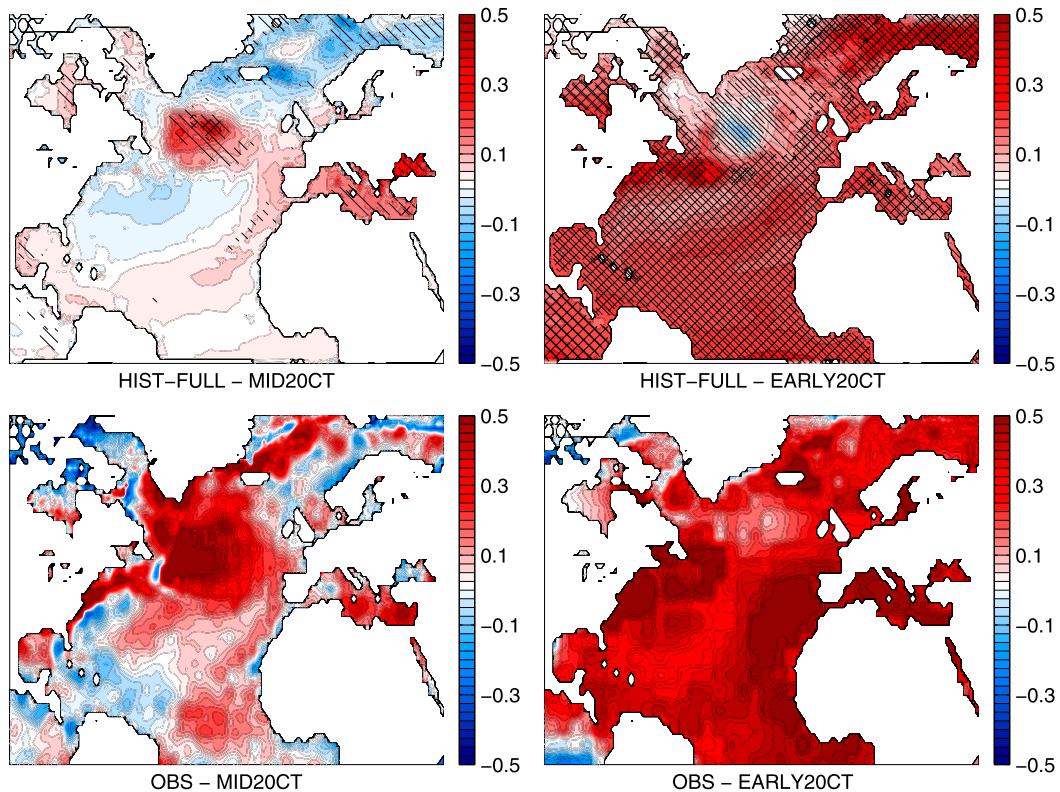


FIG. 2. Warm–cold SST composite patterns computed as the difference between the 1930–50 minus (left) 1960–80 and (right) 1900–1920 periods (time intervals are shown in Fig. 1, horizontal black lines) in (top) HIST MME and (bottom) observations. In the top panels, single (cross) hatching denotes the regions where more than 66% (80%) of the models in the HIST ensemble have the same sign of the ensemble mean changes.

displayed. Observed anomalies are largely contained in the CMIP5 HIST envelope. Interestingly, after filtering out the internal variability by averaging over the 69 realizations of the multimodel HIST ensemble, the corresponding MME mean, approximating the models' response to changes in the external forcing, shows a residual multidecadal variability that is highly coherent with the observed record (correlation is 0.68, significant at the 99.95% level, based on a one-tailed Student's  $t$  test; correlation rises to 0.82 if a 11-yr low-pass filter is applied, with a lower 99.5% significance level due to the lower number of degrees of freedom after time filtering). The HIST MME mean captures the inception of the mid-20CT event, but leads the observed NASST mid-1970s "dip" by approximately 10 years. This discrepancy could be due to several factors including internal variability (present in the real world but largely filtered out in HIST MME) or errors in the way the CMIP5 models represent the impact of the forcings on SST. Despite these caveats, given the aforementioned forced nature of the MME signal and the prevailing similarities with observed decadal variability, this finding suggests that nonstationary

forcings, either natural or anthropogenic, might have contributed to the observed variability. This result provides a multimodel extension of the single-model findings of Booth et al. (2012).

Next, we compare the patterns of SST change over the North Atlantic sector (also including the Mediterranean basin) associated with two major twentieth-century multidecadal transitions in the observed SSTs: the early-twentieth-century cold-to-warm 1900–50 transition (early-20CT), and the previously mentioned mid-20CT. These are evaluated as follows: the pattern corresponding to early-20CT (mid-20CT) is obtained as the difference between the 1930–50 and 1900–20 (1960–80) time-mean SST (the time periods used for this diagnostic are indicated as straight solid lines in Fig. 1). To obtain a MME pattern, each individual model's pattern has been interpolated onto a common  $1^\circ$  resolution regular grid.

In Fig. 2, the SST changes associated with the two transients are shown for the HIST MME and observations. The observed SST pattern for mid-20CT reveals a comma-shaped AMV-like warm anomaly structure (Sutton and Hodson 2005) with the largest amplitude in the subpolar region, extending to the tropical Atlantic

along the east side of the basin, and encompassing the Mediterranean Sea. The warm anomaly also extends northward toward the Greenland Sea, while some other regions became cooler during the period: these include parts of the western tropical and subtropical Atlantic. The HIST simulated pattern displays several large-scale features (including the comma-shaped warm structure extending to the Mediterranean basin, the enhanced subpolar response, and the western subtropical cold pattern) that are consistent with the observations (Fig. 2, bottom panels), again suggesting that prominent features of the observed SST anomalies during the mid-20CT were forced, given the uninitialized nature of the inspected simulations. Observed anomalies have an overall larger amplitude than HIST and also display a few other notable differences, at the regional scale. These are particularly evident over the areas surrounding Iceland, parts of the Nordic seas, and the Barents Sea, where the simulated pattern features cold anomalies, in contrast to the generally warm anomalies observed. Bearing in mind that observed anomalies may result from both unforced and forced variability, differences with the modeled MME response (approximating the forced only component of the total variability) are expected.

For the early-20C transition, both MME and observations show a largely monopolar warm pattern, with locally intensified anomalies along the Gulf Stream extension region and the eastern Atlantic basin, off the western African seaboard. A major inconsistency between observations and model simulations is found in the subpolar basin, where MME features a cold anomaly, contrasting with the (weakly) warm observed pattern.

A measure of the model uncertainty associated with the polarity of the SST change patterns detected during the two transitions is provided in Fig. 2, with an indication of the regions where more than 66% (single hatching) and 80% (cross hatching) of the models agree on the sign of the MME.

Concerning the simulated mid-20CT pattern, model-to-model consensus (66%) is mostly found over the subpolar basin, eastern Atlantic, the Mediterranean and Black Seas, and parts of the Nordic seas (Fig. 2, upper-left panel). The largest model agreement (higher than 80%) is found over the subpolar and eastern Mediterranean basins, broadly coinciding with the areas displaying the largest amplitude (and statistically significant) anomalies in the observed pattern. On the other hand, models exhibit consensus over the Nordic seas, but the polarity of the modeled anomaly over that area is not consistent with the observations.

Compared to the mid-20CT, the early-20CT shows a much higher cross-model consistency in the polarity of

the SST change, with model-to-model consensus exceeding the 80% level over most of the North Atlantic domain except for the subpolar region, where a 66% level is found (Fig. 2, upper-right panel).

To summarize, the two analyzed transitions are characterized by a substantially different degree of intramodel uncertainty with the mid-20CT pattern displaying a larger uncertainty compared to the early-20C transient. The origins of this uncertainty are addressed in the next section. Also, hereafter, we narrow the scope of our analysis to the mid-20CT event.

### b. Clusters in HIST simulations

In this section, the origins of the uncertainty affecting HIST models ability to reproduce the observed mid-20CT event are inspected.

To assess individual HIST models' ability in capturing the observed NASST decline during mid-20CT, an ad hoc dimensionless index is introduced. Running estimates of trends are usually very noisy, as they are affected by the end points of a selected segment of a time series. Thus, a more stable index was defined based on the time evolution of the 11-yr low-pass filtered NASST anomalies relative to the 1930–50 baseline, normalized by the standard deviation for the full-length 1930–80 transient period, according to the following formula:

$$\text{TR}(t) = \frac{\text{NASST}(t) - \overline{\text{NASST}}_{1930-1980}}{\sigma_{\text{NASST}(1930-1980)}}. \quad (1)$$

Normalization by the standard deviation was introduced so as to aid the cross-comparison between models characterized by different degrees of variability. Thus, TR is a nondimensional index with standard deviation units.

In Fig. 3 (left panel), time series of TR are shown for individual HIST models and observations. Each model curve in Fig. 3 is the ensemble average of the corresponding set of multiple realizations performed with each single model in the HIST ensemble.

The full family of TR curves displays a clear polarization within the HIST ensemble. Some of the models (cluster HIST-B) realistically capture the observed negative SST trend and the following increase (in green). Within this cluster, some phase discrepancy with the observed signal is also evident as models systematically lead the observed mid-1970s NASST dip. Another cluster (HIST-A) groups models whose time evolution during the transition is dominated by a quasi-monotonic SST increase, thus failing in reproducing the observed SST decline, although some decadal-scale fluctuations are also visible (in red). Finally, a third cluster (HIST-C) is found including models that capture the onset of the

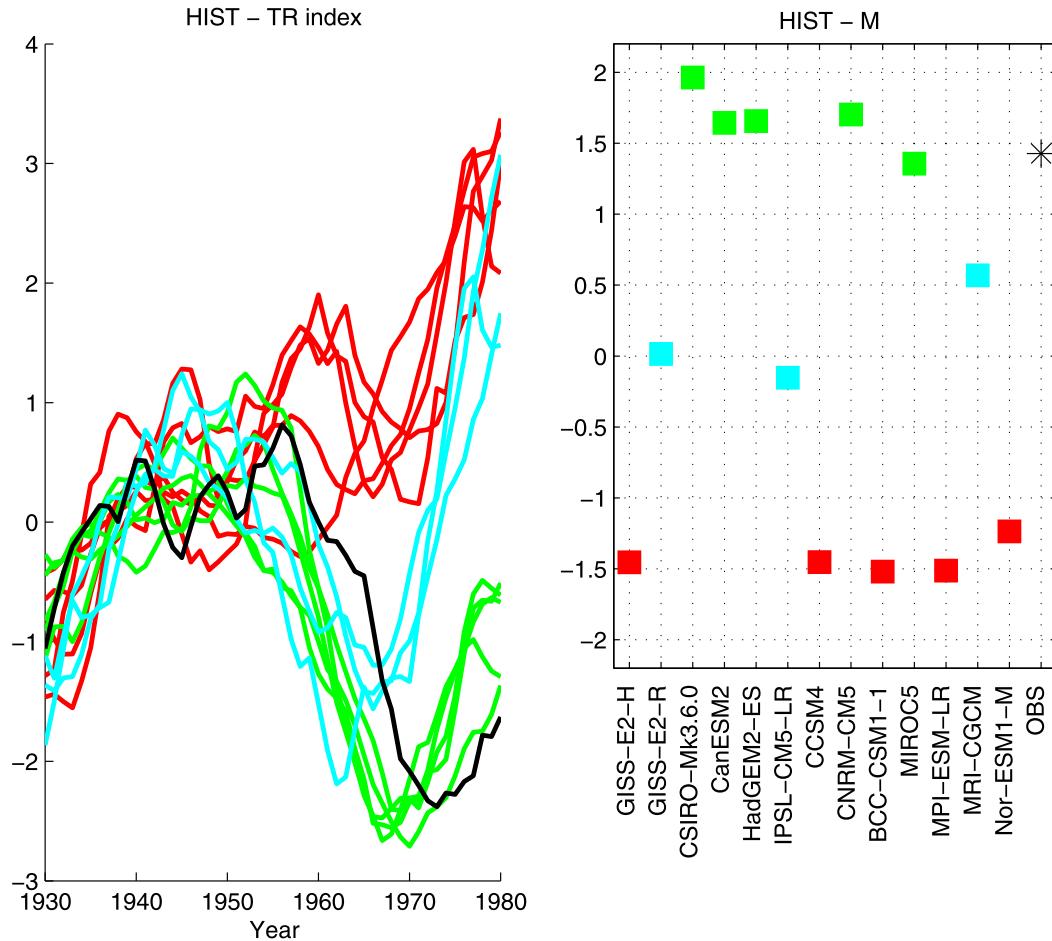


FIG. 3. (left) TR index (dimensionless units) for the HIST multimodel ensemble. Colors identify clusters HIST-A (red), HIST-B (green), and HIST-C (cyan). (right) The distribution of the metric  $m$  (dimensionless units) differentiating behavioral clusters within the HIST multimodel ensemble. Colors identify different clusters (same color code used at left). Observations are also shown (black asterisks). See text for details.

negative trend in a timely way, but show a shorter duration of the NASST declining phase.

Based on the TR index, a metric is defined, allowing a more quantitative ranking of HIST models fidelity in reproducing the mid-20C transition in the North Atlantic. The metric is defined as the difference between the TR time averages computed over the 1930–50 (warm phase) and the 1960–80 (cold phase) epochs:

$$m = \overline{\text{TR}}_{1930-1950} - \overline{\text{TR}}_{1960-1980}. \quad (2)$$

This metric roughly captures the interdecadal SST tendency during the mid-20C transition, and hereafter will be used as a measure of the models' skill in retrospectively reproducing the mid-20C event. For reference, the observed  $m$  value is about 1.4. The distribution of  $m$  values for individual HIST models and observations is displayed in Fig. 3 (right panel). The same color code

adopted for TR (shown in Fig. 3, left) is used. After introducing the  $m$  metric, HIST models polarization becomes even more evident, as clusters HIST-A and HIST-B quasi-symmetrically split apart around  $-1.5$  and  $+1.5$  (or slightly higher) values, respectively, while HIST-C models feature lower values, scattered inside the  $[-0.5, +0.5]$  range. As expected, the HIST-B cluster includes models that more closely capture the observed SST variability during the mid-20CT transient. HIST-A models, on the other hand, feature a strikingly high cross-consistency, showing relatively small deviations from  $-1.5$ .

Based on the model clustering identified through TR and  $m$ , spatial patterns of SST changes associated with the mid-20CT transient are recomputed for the different clusters (Fig. 4; see Fig. SM1 in the online supplemental material for the SST change patterns displayed by individual HIST models). Clusters HIST-A and HIST-B

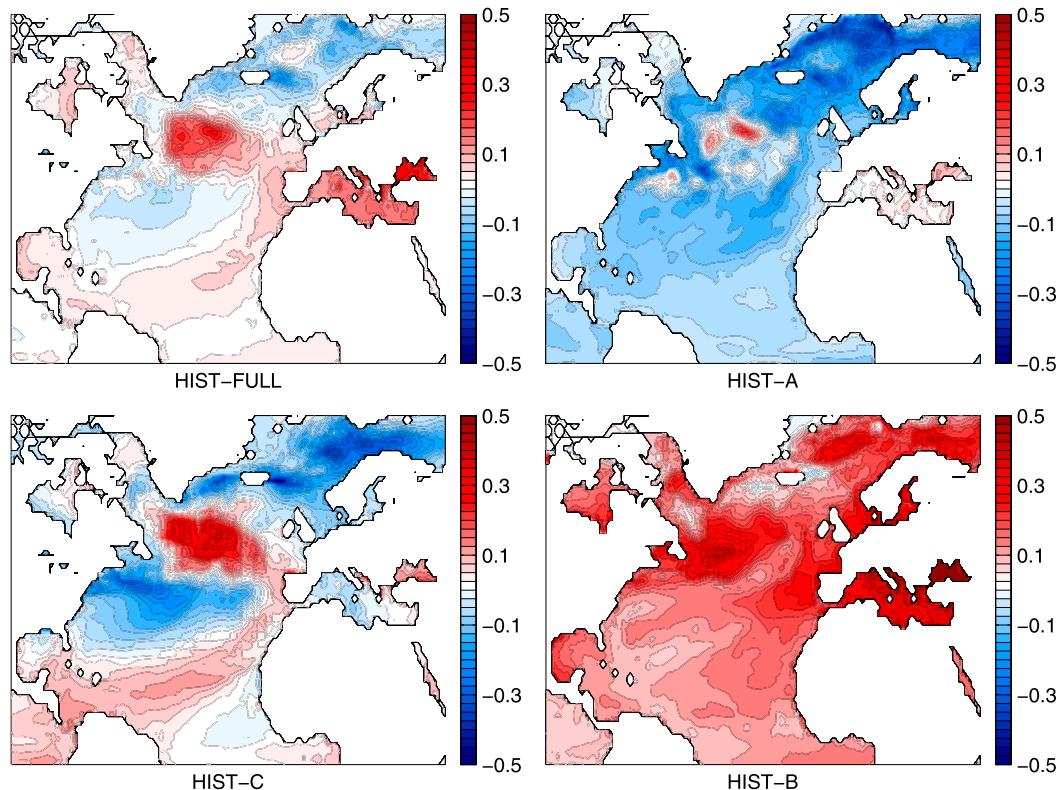


FIG. 4. As in Fig. 2, but now only for mid-20CT, showing (top left) the pattern associated with the full multimodel ensemble, and the contributions for individual clusters: (top right) HIST-A, (bottom left) HIST-C, and (bottom right) HIST-B.

reveal an almost uniform response, but with mostly opposite polarity (except over the Mediterranean region and isolated spots in the subpolar basin). HIST-C models reproduce the typical AMV-like comma-shaped pattern, visible also in the full HIST MME. Not surprisingly, HIST-B and -C patterns share the strongest similarities with the observations, consistently with the corresponding TR time series. Similarities particularly concern the subpolar basin amplified response (HIST-B and -C), the subpolar gyre–Mediterranean Sea connection (HIST-B), and the AMV-like comma pattern (HIST-C).

A more in-depth evaluation of HIST-B models, specifically targeting their ability in reproducing some relevant features characterizing the recently observed Atlantic multidecadal variability, including upper-ocean heat content, sea surface salinity in the subpolar basin, and the Atlantic interhemispheric SST dipole, was performed (shown in the supplemental material) to rule out the possibility that HIST-B models might reproduce the mid-20C North Atlantic cooling for the “wrong” reasons (Zhang et al. 2013). It is found that HIST-B models [except for HadGEM2-ES, not included in this analysis as it is already extensively documented in Zhang et al. (2013)] display a reasonably good consistency with the thermal

state of the upper-ocean North Atlantic as observed in the more recent decades (Figs. SM6 and SM7). Simulated surface salinities in the subpolar basin show a large intramodel uncertainty, and the coherency with the observed variations is generally poor (except for the CanESM2 model), while the magnitude of the simulated anomalies is consistent with the observations (Fig. SM8).

The strong clustering characterizing the HIST ensemble explains the limited cross-model consensus on the polarity of the mid-20CT SST change pattern (shown in Fig. 2). The existence of multiple behavioral clusters within the HIST models population strongly limits the overall retrospective skill of the corresponding MME, as only a subset of the full ensemble (HIST-B cluster) is able to reproduce the mid-20CT event with a reasonable accuracy.

### c. The role of anthropogenic forcing on the mid-20C transition

In the previous sections, several lines of evidence have been found suggesting that a subset of CMIP5 historical uninitialized simulations have skill in reproducing the observed mid-20CT transition. This, in turn, has implications for a potentially important role of external

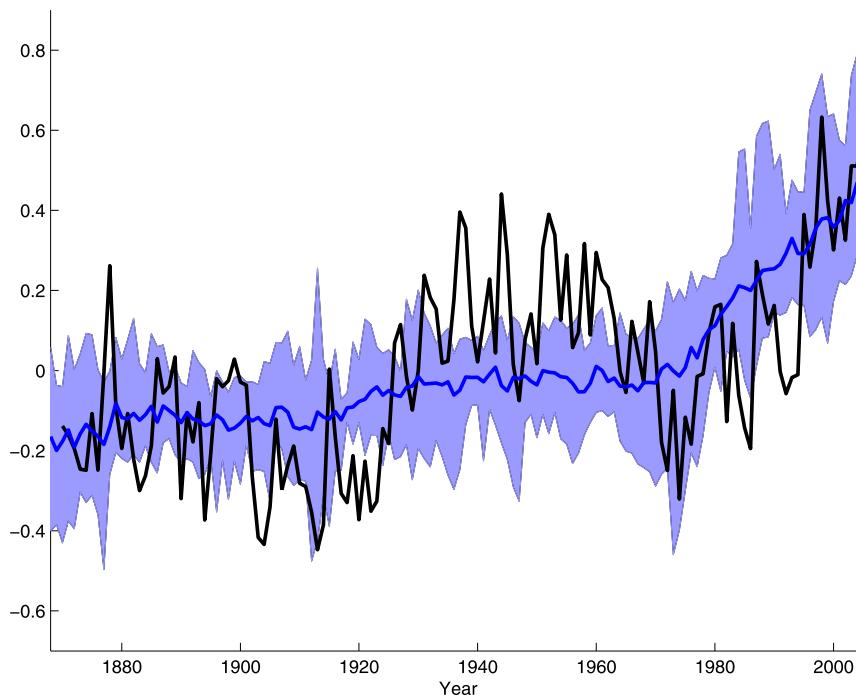


FIG. 5. As in Fig. 1, but for the ANT ensemble (in blue).

forcing (either natural or anthropogenic) on the recently observed multidecadal SST variability in the North Atlantic and the adjacent Mediterranean basin. Under the assumption that these SST changes were largely forced, it still remains to be determined whether the dominant forcing has a natural or anthropogenic origin. To address this question we extend to ANT the analysis previously applied to the HIST ensemble.

The corresponding NASST index for the ANT ensemble is shown in Fig. 5. Compared to HIST, the ANT ensemble shows relatively little multidecadal variability and a consistently lower correlation between the corresponding MME and observations (correlation 0.6). The MME temporal evolution is dominated by the long-term warming trend, accelerating at the turn of the 1980s. The early-20CT transition is not captured by ANT simulations, suggesting a nonanthropogenic origin for this event. In particular, the lack of volcanic eruptions in the ANT forcing set might be implicated in the degradation of the correlation between modeled and observed NASST (Stenchikov et al. 2009; Church et al. 2005). On the other hand, the envelope of the modeled mid-20CT transient exhibits some coherency with the observations.

After diagnosing TR and  $m$  for the ANT ensemble, the same strong polarization between models featuring an SST decline (consistent with the observations) and models displaying a quasi-monotonic increase, previously detected in HIST, is found (Fig. 6). The clusters identified in

ANT closely replicate those found in HIST, and are consistently labeled ANT-A, -B, and -C (although ANT-C does only include one model, and therefore cannot be considered as a proper cluster). Analogies concern both the amplitude and the phase of the corresponding TR and  $m$  parameters, as compared to observations. The consistency between HIST and ANT is further corroborated by the SST composite patterns diagnosed for the three ANT clusters (shown in Fig. 7; see Fig. SM2 for the SST change patterns displayed by individual ANT models). ANT-B simulations reproduce an SST difference pattern that is broadly consistent with the observed one and, compared to HIST-B, appears to be more skillful in capturing some of the observed regional-scale features (see, in particular, the wavelike structure across the subtropical Atlantic, the magnitude of the warm anomaly over the subpolar basin, and the polarity of SST changes over the Barents Sea). ANT-A simulations, on the other hand, show a largely homogenous temperature difference pattern consistent with the positive trend found in the corresponding TR time series (shown in Fig. 6, left). Finally, ANT-C reproduces the same comma-shaped pattern found in HIST-C.

While individual clusters in the two scrutinized ensembles show a substantial similarity, the corresponding MME mean patterns (upper-left panels in Figs. 4 and 7) do not. These differences can be explained by the different weights that A, B, and C clusters have in each ensemble. For instance, HIST-C includes three models, and therefore

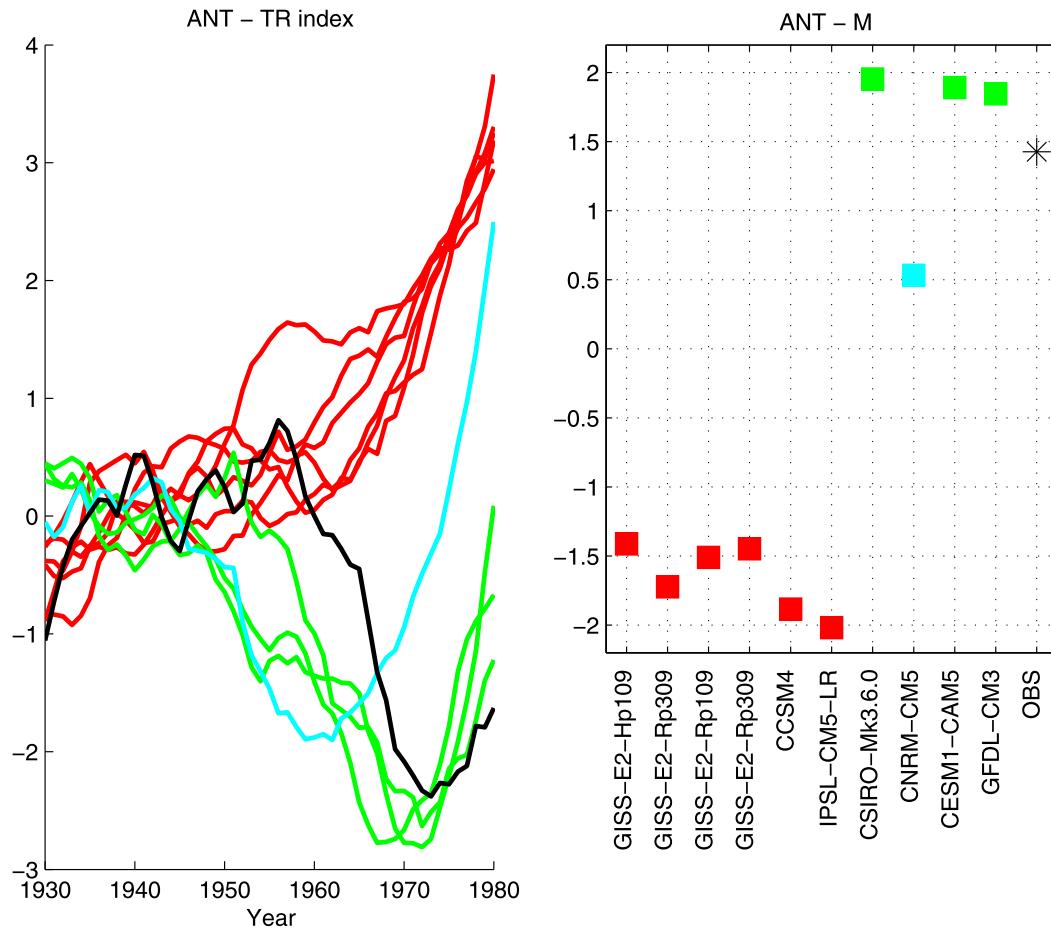


FIG. 6. As in Fig. 3, but for the ANT ensemble.

has a much higher imprint on the HIST MME compared to the homologous ANT-C, which only counts 1 model.

The existence of different and easily identifiable behavioral populations in CMIP5 historical and HM integrations highlights the strong uncertainty affecting the representation of regional decadal-scale forced climate variability. Despite the common protocol and boundary conditions adopted to perform CMIP5 historical simulations, the way individual models respond to externally imposed nonstationary forcings can drastically vary across a given multimodel ensemble.

An important outcome of the present analysis is the close resemblance between HIST and ANT clusters in reference to the mid-20CT. The overall consistency between HIST-B and ANT-B (as described by TR,  $m$ , and SST change patterns) in particular, suggests that anthropogenic forcings of some nature have played a nonnegligible role on the observed 1940–75 NASST decline. In contrast, after comparing NASST anomalies under HIST and ANT forcing conditions (Figs. 1 and 5, respectively) it appears that anthropogenic forcings

have likely played no role with regard to the early-20CT transient. The comparison shows that ANT models feature no hints of the warming SST trend observed in the North Atlantic during the 1900–50 period, in contrast with what found in HIST, where models seem to capture reasonably well the observed NASST increase. This finding suggests that the early-20CT event may be instead attributed to natural variability. However, in the absence of a more in-depth evaluation of models' clustering around the early-20CT transient (beyond the scope of this investigation) the latter attribution remains elusive and must be considered as a mere speculation.

#### d. The relative roles of AA and GHG forcing

In the attempt of further discriminating the individual roles played by primary anthropogenic drivers on the mid-20C transient, in this section we analyze three additional ensembles of historical integrations, focusing in particular on the effects of AAs and GHGs. These include the NoAA ensemble, where the same HIST forcing set is used except for AA, and two idealized,

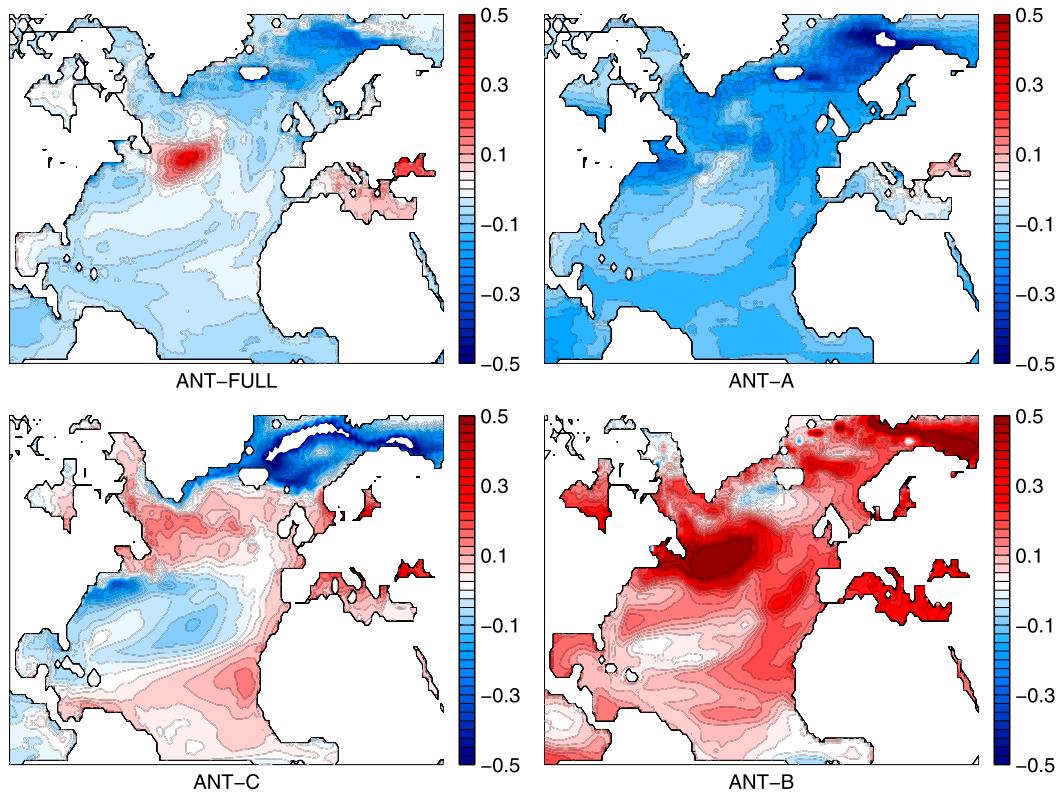


FIG. 7. As in Fig. 4, but for the ANT ensemble.

single-forcing experiments where either AA-only or GHG-only forcing is considered (see Table 2).

Among the anthropogenic forcings used to perform the ANT ensemble integrations (GHG, anthropogenic aerosols, ozone, and land use changes), anthropogenic aerosols appear to be a plausible candidate to explain the NASST decline observed during the mid-20CT transition. Large increases in AA occurred during 1940–75 have been invoked by several authors to explain a consistent reduction in surface air temperatures, leading to a hiatus in the global surface temperatures rise during this period (Wilcox et al. 2013; Maher et al. 2014). To isolate the influence of AA forcing on the overall SST multidecadal variability, results from the NoAA ensemble (using the same forcing as in HIST, except for AA) are analyzed.

The NASST time evolution of NoAA MME and the corresponding multimodel [min, max] range are compared to observations in Fig. 8. After removing the AA forcing, the multidecadal SST variability appears to be dominated by the long-term warming trend. Thus, NoAA integrations substantially fail in reproducing the observed mid-20CT transition. As expected, the corresponding set of TR curves and the relative distribution of  $m$  values (Fig. 9) show that the models contributing to the NoAA

ensemble fall within behavioral classes corresponding to clusters A and C (in red and cyan, respectively), while no model reproducing the B-type behavior is found.

The composite patterns associated with the mid-20CT transition for the NoAA ensemble are displayed in Fig. 10. For consistency with the previously analyzed HIST and ANT ensembles, the patterns corresponding to cluster A (including two models) and the single-model “cluster” C are displayed (Fig. SM3 shows the SST change patterns displayed by individual NoAA models). NoAA-A models reproduce the same monopolar structure already identified in the homologous clusters in HIST and ANT ensembles, consistent with a monotonic warming trend, uniformly distributed over the North Atlantic and the neighboring Mediterranean basins. The model exhibiting a cluster-C-like behavior (CSIRO-Mk3.6.0), on the other hand, features some similarity with the observations, particularly over the subpolar basin and eastern Atlantic. A possible explanation invoking the response to GHGs increase is detailed later, when discussing results from the GHG ensemble.

Clearly, the small size of the NoAA ensemble partly hampers the robustness of this set of results, which may be overly sensitive to the deficiencies affecting individual models in the ensemble.

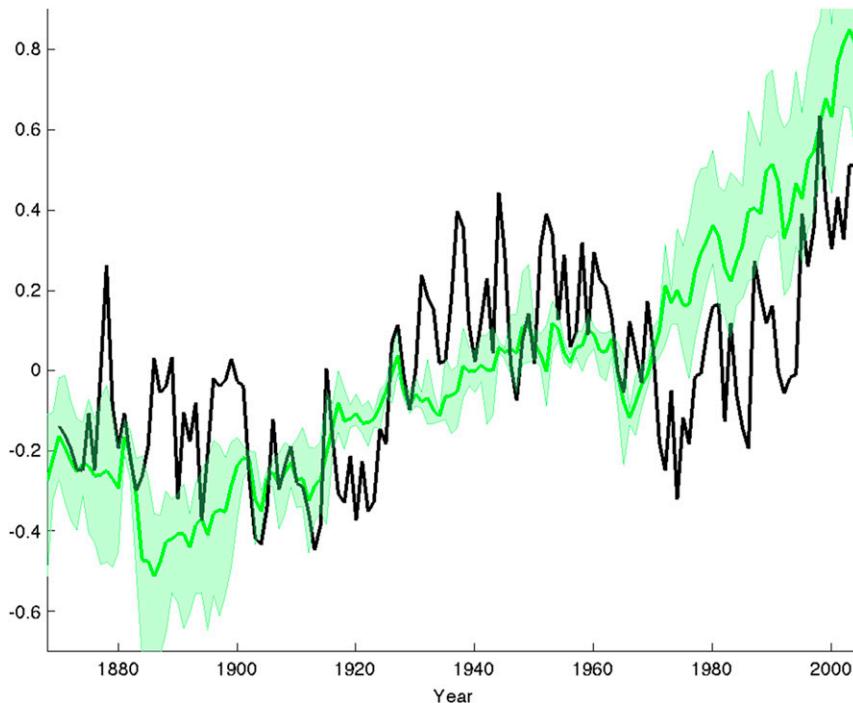


FIG. 8. As in Fig. 1, but for the NoAA ensemble.

To provide a more robust assessment of the impact of anthropogenic forcings, the analysis of the NoAA set is complemented with a similar analysis performed on the AA and GHG historical simulations, for which multi-model ensembles of a larger size are available (see Table 2). Compared to the previously inspected HIST, ANT, and NoAA sets, these two ensembles feature a considerably lower degree of realism, since they consider the effect of one single forcing agent varying with time according to the historical observed record, while maintaining all other forcings at constant preindustrial levels throughout the model integration. On the other hand, this type of experiment has the merit of providing a clean attribution of the impact exerted by individual forcings over the global climate system.

Because of their highly idealized design, the characteristics of the twentieth-century NASST multidecadal variability in the AA and GHG experiments are not strictly comparable to observations. However, for methodological consistency, and also to facilitate the cross-comparison with the other historical integrations, the approach followed so far for the analysis of the mid-20CT event is extended to the AA and GHG ensembles.

The NASST multidecadal variability in the AA ensemble is shown in Fig. 11. Since NASST changes are almost systematically skewed toward negative values, particularly during the second half of the twentieth century, anomalies are computed with respect to the 1850–1950 mean. Except

for one outlier (GISS-E2-R forced via emissions), under AA-only forcing conditions all models feature a progressive cooling, with some of them exhibiting a marked step change in the rate of SST decrease around the middle of the century. The AA models' response also exhibits a relatively large spread, with an approximately  $0.5^{\circ}\text{C}$  amplitude. A closer analysis of this ensemble reveals that a large fraction of the detected spread is explained by whether a given model implements both direct and indirect aerosol effects or only the direct one. This clearly emerges after comparing GISS-E2-Hp107 and GISS-E2-Rp107 (including both effects) against their “p106” counterparts (including only the direct effect). Expectedly, when only the direct effect is included, a consistently weaker cooling is found at the end of the model record.

In the  $(\text{TR}, m)$  space, AA simulations show a strong cross-consistency in the NASST response, with all models featuring a cluster-B type of behavior (Fig. 12), except for the GISS-E2-Rp310 showing a cluster-C response with no hints of an SST decline. Most importantly, the aerosol forcing alone appears to be reasonably consistent with the observed NASST tendency. The corresponding SST change patterns, on the other hand, show that under AA-only forcing conditions, cluster-B models undergo a basinwide cooling, except over the subpolar gyre region, where an opposite trend is found (Fig. 13; see also Fig. SM4 for the SST change patterns displayed by individual AA models). This is clearly inconsistent with the observed

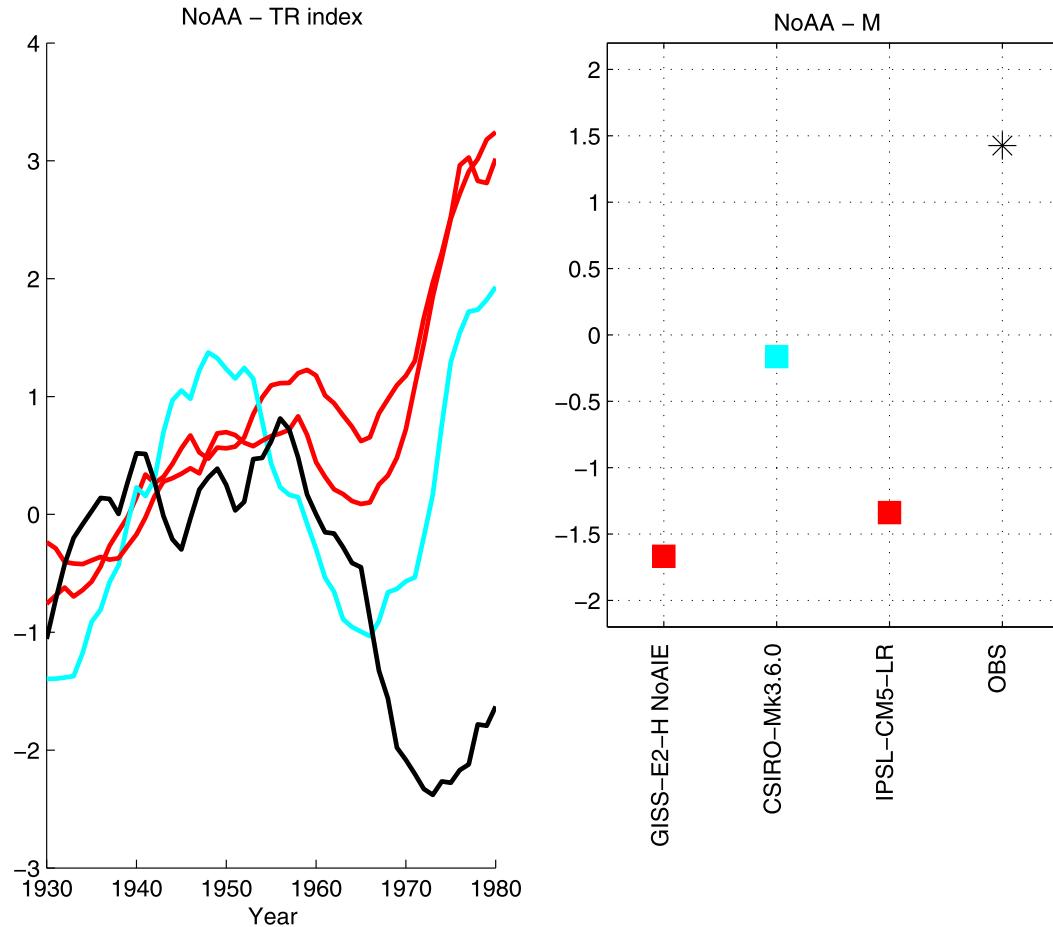


FIG. 9. As in Fig. 3, but for the NoAA ensemble.

cooling pattern (Fig. 2) and reveals that the AA-driven NASST decline (detected in Figs. 11 and 12) is dominated by a cooling signal encompassing the tropical and subtropical North Atlantic as well as the Mediterranean basin. The strong meridional gradient affecting the AA-induced SST response across the subpolar–subtropical boundary may reflect a similar gradient in the regional distribution of tropospheric aerosols and the implied radiative forcing pattern (see Fig. 13 in Bellucci et al. 2015).

Next, we analyze results from the GHG ensemble. As expected, all models display a monotonic basinwide warming in the North Atlantic, with different warming rates reflecting model-to-model differences in climate sensitivity (Fig. 14). A consistent picture is provided by the (TR,  $m$ ) couplet, with all models grouped into one single A-like cluster, thus failing in reproducing the observed cooling (Fig. 15). The SST change pattern associated with the GHG MME displays a quasi-monopolar structure, consistent with the detected basinwide warming, but with a significant deviation over the subpolar region, where SSTs show no trends during the target 1930–80

period (Fig. 16). An inspection of the response featured by individual members of the GHG ensemble (Fig. SM5) reveals that this regional feature is mainly determined by a handful of models (specifically IPSL-CM5-LR, IPSL-CM5-MR, and CSIRO-Mk3.6.0 and, to a lesser extent, MIROC5 and HadGEM2-ES). These particular models undergo, during the mid-20CT event, a warming SST trend over the subpolar region that contrasts with the opposite tendency in the remainder of the North Atlantic basin. This behavior is somehow “orthogonal” to the one shown by the gross majority of the AA ensemble members (Fig. 13), featuring an overall, basinwide cooling response except over the subpolar basin. Concerning the GHG ensemble, the SST change pattern associated with the MME bears some resemblance with the North Atlantic “warming hole” pattern emerging both in observational records and models, as described by Drijfhout et al. (2012) and Rahmstorf et al. (2015): according to these authors, the peculiar response of the subpolar region to increasing GHG concentrations may be interpreted invoking a weakening of the AMOC strength,

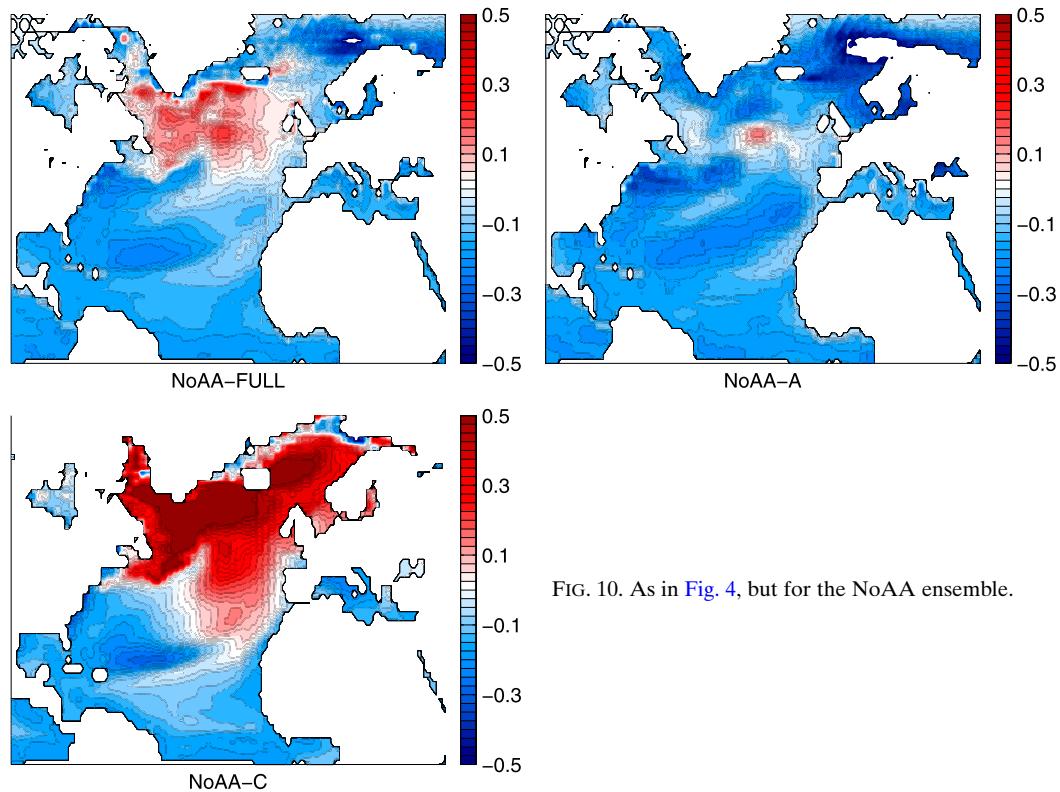


FIG. 10. As in Fig. 4, but for the NoAA ensemble.

leading in turn to a weaker poleward heat transport, and a cooling of the subpolar North Atlantic.

The analyses performed on AA and GHG ensembles provide some additional clues on the mechanisms governing the North Atlantic cooling, and further assist in the interpretation of HIST and ANT ensembles' results. Clearly, neither AA nor GHG forcing alone can explain the observed SST pattern associated with the mid-20CT event. While the AA-forced models skillfully capture the basinwide North Atlantic cooling signal, they fail in reproducing the regional-scale response over the subpolar basin. GHG-forced models, on the other hand, display a basinwide warming trend that is inconsistent with the observed cooling, but they show some consistency with the observed warming hole pattern in the subpolar region. In light of these results, the skill in reproducing the mid-20CT event found in the ANT (and to some degree the HIST) ensemble can be plausibly attributed to the combined effect of AAs and GHGs, with the former (latter) mainly projecting on the tropical and subtropical (subpolar) North Atlantic.

The above results also suggest a possible key for understanding the causes behind the clustering process based on models' relative sensitivities to increasing levels of GHGs and AAs. For example, models exhibiting an overly strong climate sensitivity will more likely show a

cluster-A behavior, unless the implied warming excess is compensated for by a specular cooling response to changes in AAs. According to this explanation, it is the imbalance between the responses to these two forcing agents that determines the cluster of a specific model.

To corroborate this hypothesis, we analyze the "trajectory" followed by single models through the full hierarchy of the anthropogenically driven (ANT, AA, and GHG) experiments. Because of the patchy coverage of HM experiments (Table 2), the tracking can be only performed for the CCSM4, CSIRO-Mk3.6.0, GISS-E2-H, and GISS-E2-R models. Figure 17 shows the composite SST change patterns evaluated for the abovementioned models, for the ANT, AA, and GHG ensembles. As far as the ANT ensemble is concerned, the analyzed model population includes one model featuring a cluster-B behavior (CSIRO-Mk3.6.0) while the other three models are representatives of the behavioral cluster A. Although a linear superposition of the effects of GHGs and AAs cannot be assumed given the high nonlinearity of the inspected dynamical system, it is clear that some regional patterns result from the combined impact of the two individual forcings.

In particular, the CSIRO-Mk3.6.0 ANT pattern is largely determined by the AA stand-alone forcing, except over the subpolar basin, where GHGs have a prevailing role. The GHG-induced pattern is strongly

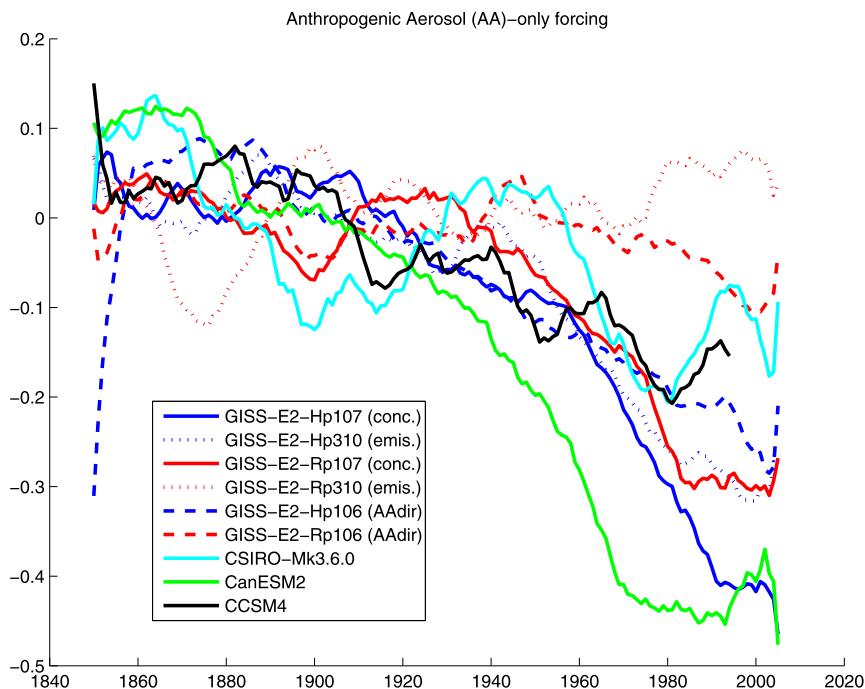


FIG. 11. Area-averaged North Atlantic (defined as  $7.5^{\circ}$ – $75^{\circ}$ W and  $0^{\circ}$ – $60^{\circ}$ N) SST anomalies relative to the 1850–1950 period, for the AA multimodel ensemble.

consistent with the corresponding NoAA pattern (Fig. 10), highlighting the important role played by AAs in setting the model's behavior under HIST and ANT forcing conditions.

Concerning the other models, the corresponding ANT patterns appear to be largely dominated by the effect of the GHG forcing, although some of the regional-scale features are clearly affected by AAs (see, e.g., the GISS-E2-R model over the subpolar and eastern Mediterranean and Black Sea subbasins).

#### 4. Summary and discussion

Results from an attribution study inspecting the origins of multidecadal variability in the North Atlantic SST have been presented. We targeted in particular the 1940–75 “warm-to-cold” transition, an event that is generally framed in the context of the longer-term AMV cycle, in turn associated with the AMOC internal variability (Knight et al. 2005). This specific transient provides a useful case study to examine the ability of CMIP5 uninitialized, historical integrations in retrospectively reproducing a particular episode of the twentieth-century climatic history, under a hierarchy of forcing conditions. For this purpose, both standard and “historical Misc” CMIP5 simulations of the historical climate (combining selected natural and anthropogenic forcings) were exploited. Specifically, we analyzed a hierarchy of CMIP5 uninitialized simulations of

the historical (1870–2005) period, performed under different forcing conditions. The analyzed integrations included five different multimodel ensembles: standard historical integrations (i.e., using observed natural and anthropogenic forcing), anthropogenic-only integrations, a set of integrations in which both natural and anthropogenic forcing were used except for anthropogenic aerosols, and two idealized, single-forcing ensembles in which observed records of AA and GHG forcings were used, respectively.

To filter out as much as possible the uncorrelated internal variability, and to maximize the forced-only component in the NASST signal, only models providing multiple realizations of the historical climate were used in the analysis.

The detection of multidecadal variability in the standard historical MME NASST, significantly correlated with the observed signal and with a spatial structure strongly resembling the AMV pattern, suggested the potential influence of nonstationary external forcings on the observed multidecadal variability in the North Atlantic. A similarly high correlation between the simulated and observed NASST is documented in Booth et al. (2012) for the HadGEM2-ES model. Here the same finding is established for a large 13-model ensemble of 69 integrations performed under the CMIP5 historical protocol.

A closer inspection of individual HIST ensemble members, focused on the mid-20CT event, led to the identification of homogenous and distinct behavioral clusters.

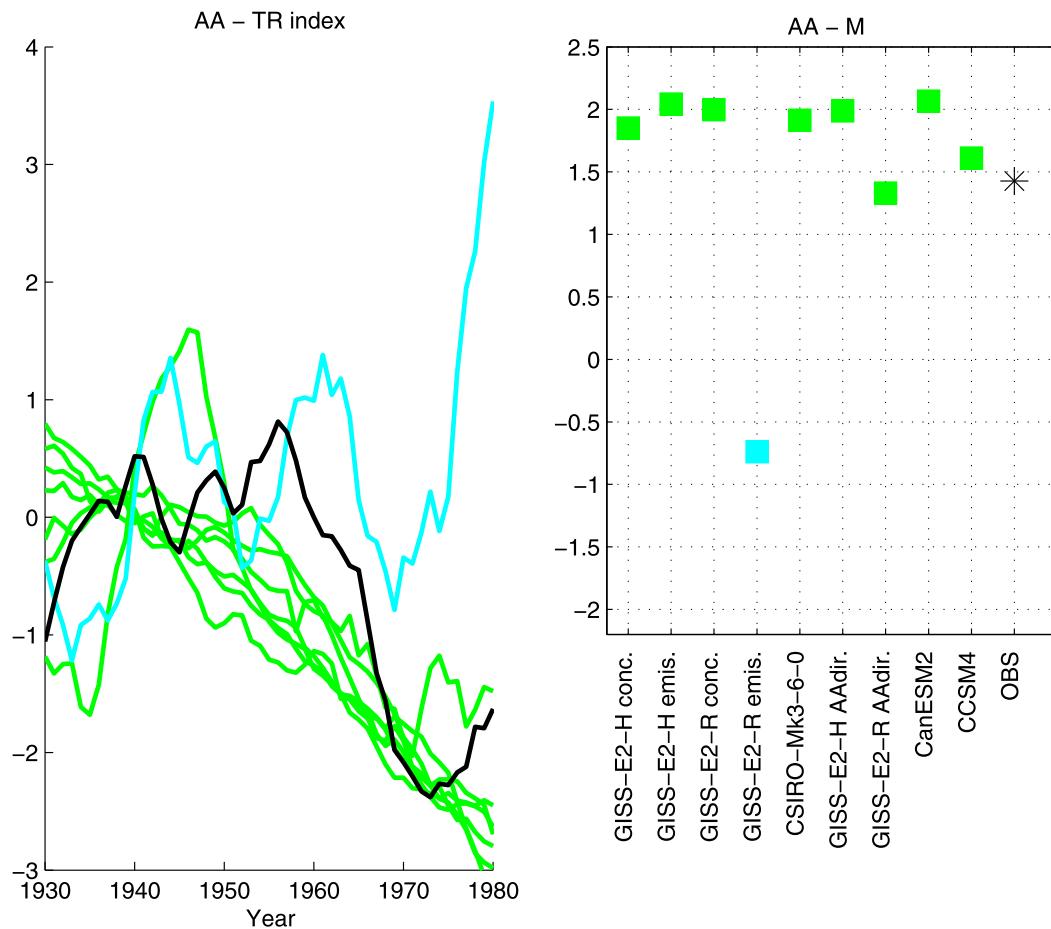


FIG. 12. As in Fig. 3, but for the AA ensemble.

Based on a metric designed to evaluate the overall skill in replicating the observed mid-20CT event, three different clusters were singled out. In particular, two clusters (labeled as HIST-A and HIST-B) contribute to a fairly strong polarization within the HIST ensemble, with HIST-B faithfully reproducing the declining NASST trend as opposed to HIST-A, displaying a quasi-monotonic NASST increase, over the 1940–75 period. A third class of models (HIST-C) correctly captures the onset of the NASST cooling but shows a shorter duration of the declining phase, and an earlier start of the late twentieth-century warming trend. Expectedly, clusters B and C do also realistically capture the observed SST change patterns associated with mid-20CT. The strong polarization found in the HIST multimodel set explains the low model-to-model agreement on the sign of the SST anomaly associated with the mid-20CT warm-to-cold transition (shown in Fig. 2).

Removing the natural forcings (ANT ensemble) negatively affects the simulated NASST skill during the early decades of the twentieth century (Fig. 5). This might be explained invoking both the influence of solar variability in

the early decades of the twentieth century and the impact of volcanic eruptions, both missing in the ANT forcing set. In particular, the lack of the 1963 Mount Agung eruption signature is expected to have a deteriorating impact on the skill associated with the modeled mid-20CT event.

However, although the ANT MME exhibits no ability in replicating the mid-20CT event, some skill emerges when inspecting individual ANT models' behaviors. Indeed, the same clustering found in the HIST ensemble analysis, emerges in ANT as well: (TR,  $m$ ) couplets and the corresponding SST change patterns show a high degree of consistency across both the ANT (Figs. 6 and 7) and HIST (Figs. 3 and 4) ensembles.

The identification of B and C types of clusters in both ANT and HIST, realistically capturing the observed mid-20CT event, corroborates the idea of an anthropogenic origin for the mid-20CT. Previous studies indicated AA as a primary candidate for the 1940–75 hiatus in the global mean surface temperature (Maher et al. 2014), likely a manifestation of a more regional, Atlantic-centered signal (Booth et al. 2012). To further refine our attribution

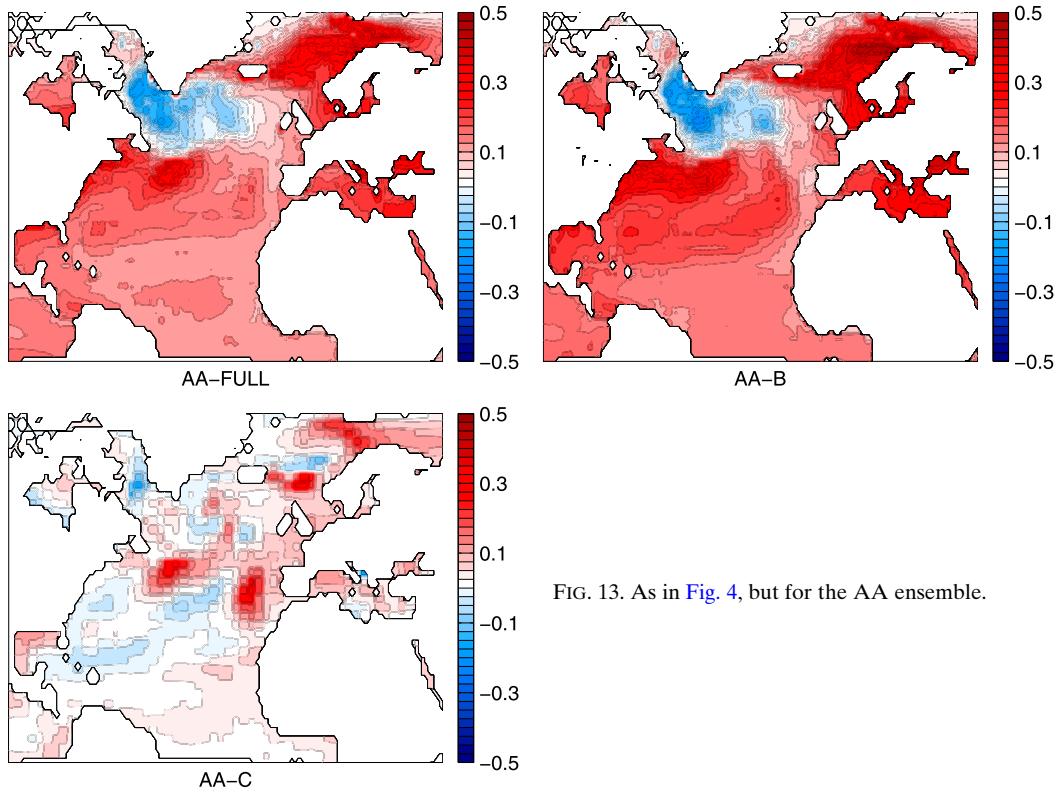


FIG. 13. As in Fig. 4, but for the AA ensemble.

study, we additionally analyzed the NoAA ensemble, which includes models run under the same forcing used in HIST, except for AA, and the two idealized, single-forcing AA and GHG ensembles. Interestingly, after

excluding AAs from the full set of HIST forcings (as done in NoAA), cluster-B typology vanishes. Consistently, the NoAA MME signal appears to be dominated by a quasi-monotonic positive trend.

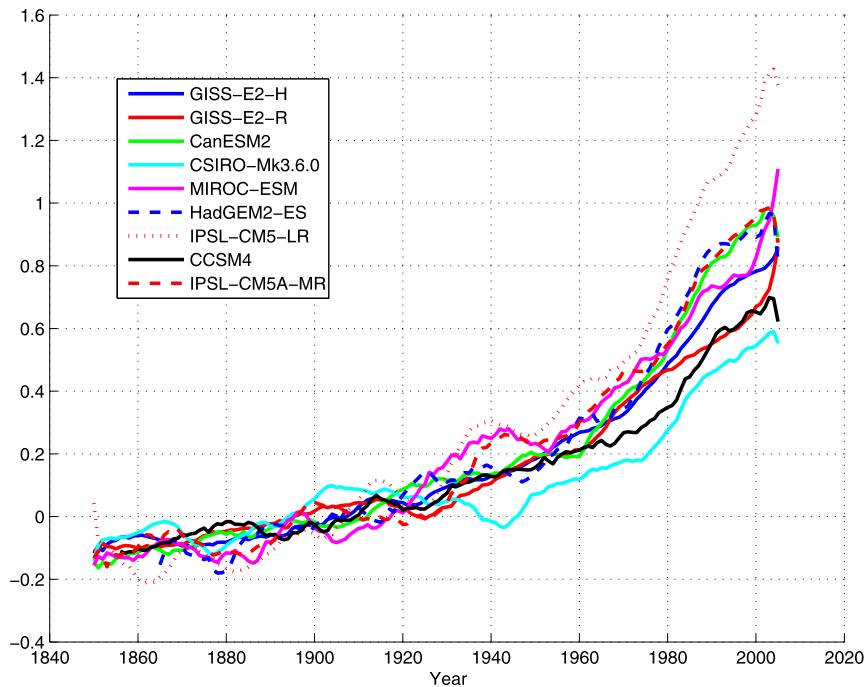


FIG. 14. As in Fig. 11, but for the GHG multimodel ensemble.

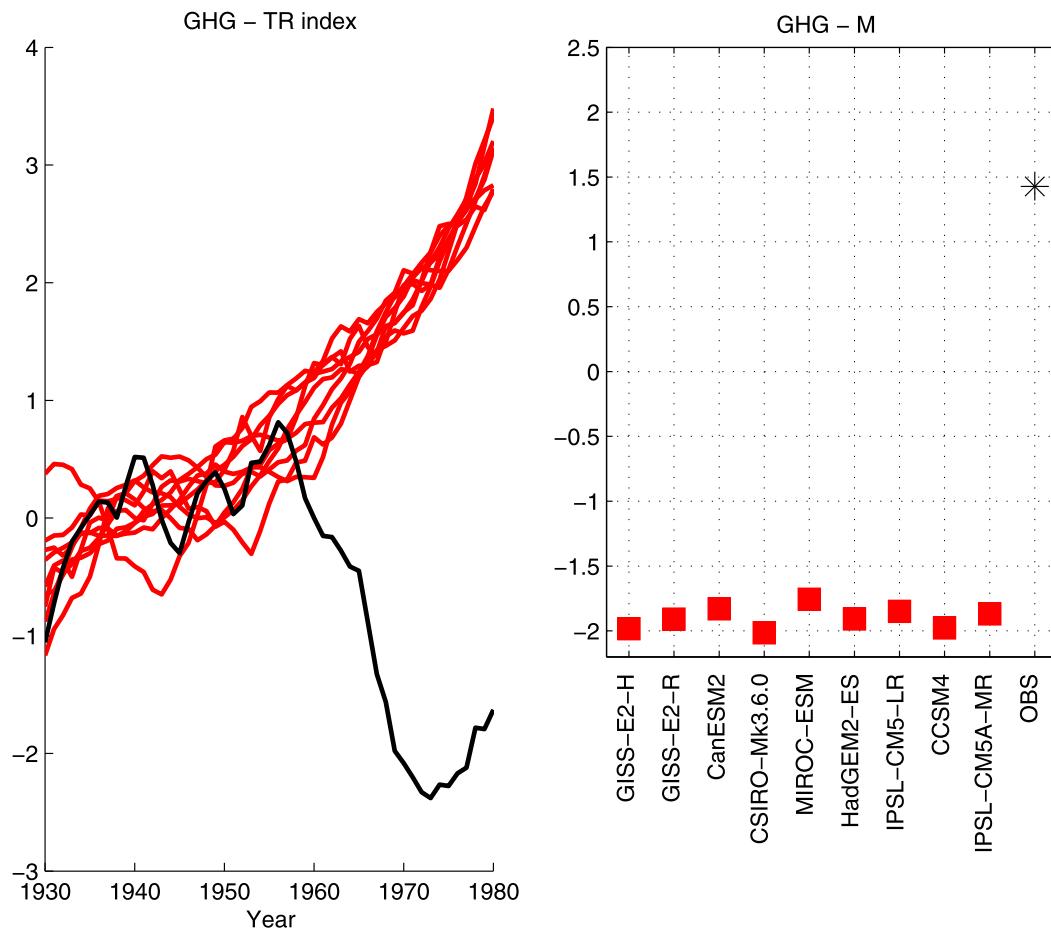


FIG. 15. As in Fig. 3, but for the GHG ensemble.

The analyses on the single-forcing ensembles shed additional light on the relative roles played by AA and GHG in the overall SST response found in ANT and HIST ensembles. The basinwide cooling observed in the North Atlantic during the mid-20CT event is well explained by the AA forcing, but GHGs appear to shape the regional-scale pattern over the subpolar basin. Overall, both anthropogenic drivers are key in determining the mid-20CT evolution, with AAs projecting on tropical-to-middle latitudes (including the Mediterranean region) by locally altering the radiative forcing, and GHGs primarily projecting on the higher latitudes via changes in the AMOC strength. Also, the model-dependent sensitivity to changes in AA and GHG concentrations determines the clustering process. In cluster-A models, the GHG-induced response overshadows the cooling implied by increasing AA levels, while cluster-B models feature a more realistic balance between GHG- and AA-induced responses. Consistent with this view, cluster C may simply represent a hybrid class of models, featuring a suboptimal blend of the two prevailing A and B types of forced responses.

The identification of homogeneous behavioral clusters within the widely scrutinized set of CMIP5 historical integrations represents a particularly insightful outcome of this analysis. Despite the strong constraint stemming from the use of common external forcing fields, a large model-to-model diversity in the representation of the twentieth-century forced variability emerges, reflecting fundamental differences in the way individual models respond to prescribed anthropogenic external forcings.

Also, estimates based on large-sized multimodel ensemble-mean climatic variables may be the midpoint of fairly distant model populations. Model spread—a statistic that is widely adopted to measure model uncertainty—may hide large differences within a model population. Thus, some caution is needed when using MME in climate model analyses, as averaging across fairly distinct model populations may determine, through mutual cancellation, a rather artificial description of the actual multimodel ensemble behavior.

The emergence of a forced AMV-like pattern associated with the 1940–75 transition in historical

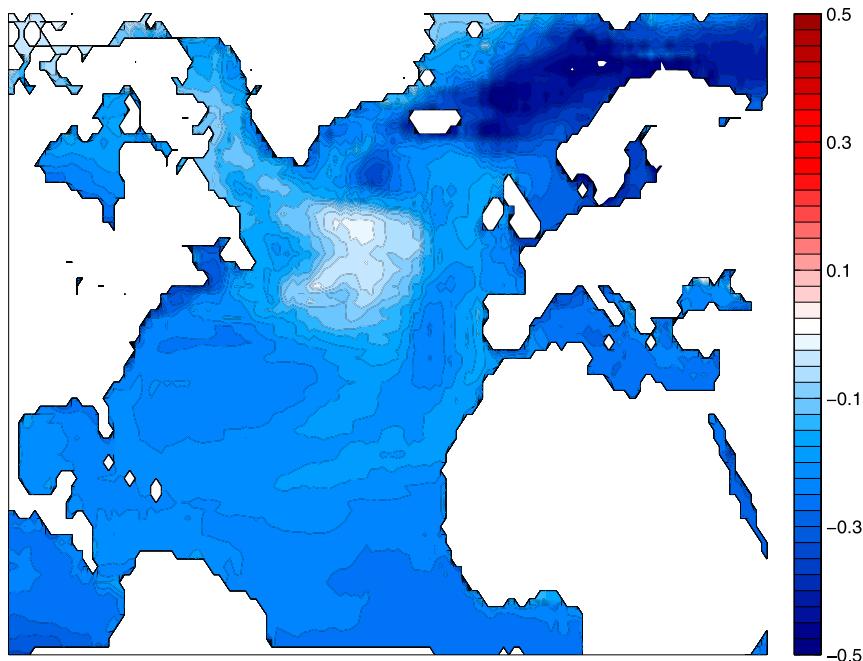


FIG. 16. As in Fig. 4, but for the GHG MME.

simulations (particularly evident in clusters B and C of HIST and ANT ensembles) raises some questions regarding the dynamical origins of the Atlantic multidecadal variability. Recently, several authors have suggested alternative hypotheses to the dominant paradigm of an AMV largely driven by the internal AMOC variability (Otterå et al. 2010; Booth et al. 2012; Clement et al. 2015). The picture emerging from our analysis suggests that the twentieth-century “AMV cycle” does actually result from a combination of several factors, with the anthropogenic drivers playing a relevant role during the mid-20CT transition. The role of the AMOC as an important driver for the AMV is not ruled out here. Specifically concerning the mid-20CT event, it is the forced AMOC response, subject to the GHG forcing acting as an external pacemaker, more than its internal variability, that appears to play a key role, particularly over the subpolar region.

The present results partly corroborate a similar (albeit methodologically alternative) analysis conducted by Terray (2012) on the origins of NASST multidecadal variability. Consistent with our conclusions, Terray (2012) established a leading role for anthropogenic drivers (essentially, GHG and aerosols) on the NASST multidecadal changes from 1950 onward, with natural forcings explaining a comparatively smaller fraction. This finding particularly applies to the tropical and subtropical subareas. On the other hand, the author claims that internal variability is a dominant driver for

the subpolar Atlantic, although admittedly this attribution is affected by a large uncertainty.

The present work is affected by several caveats, listed below. The relatively limited size of individual models' ensembles (for certain models, only three members were available), the low model diversity (particularly affecting the NOAA ensemble), and, most importantly, the lack of homogeneity across the various types of analyzed simulations (i.e., different model selections were used to perform standard historical and HM integrations, preventing a rigorous assessment of single models' response to different forcing conditions) are issues. The abovementioned points reflect deficiencies that were inherent to the CMIP5 framework. An effort should be done to address as much as possible these deficiencies in outlining future CMIP multimodel exercises.

It is also important to remark that the methodology adopted in our analyses is based on multimodel ensembles of uninitialized historical climate integrations, and as such it does not allow a direct evaluation of the role of internal variability on the mid-20CT event, as this component is largely filtered out by the ensemble averaging process. A more comprehensive attribution study (i.e., one tackling the whole set of possible sources of variability) would benefit from a cross comparison between the inspected uninitialized integrations and a consistent set of initialized retrospective forecasts such as those performed using the CMIP5 near-term prediction protocol (Taylor et al. 2012). However, available decadal

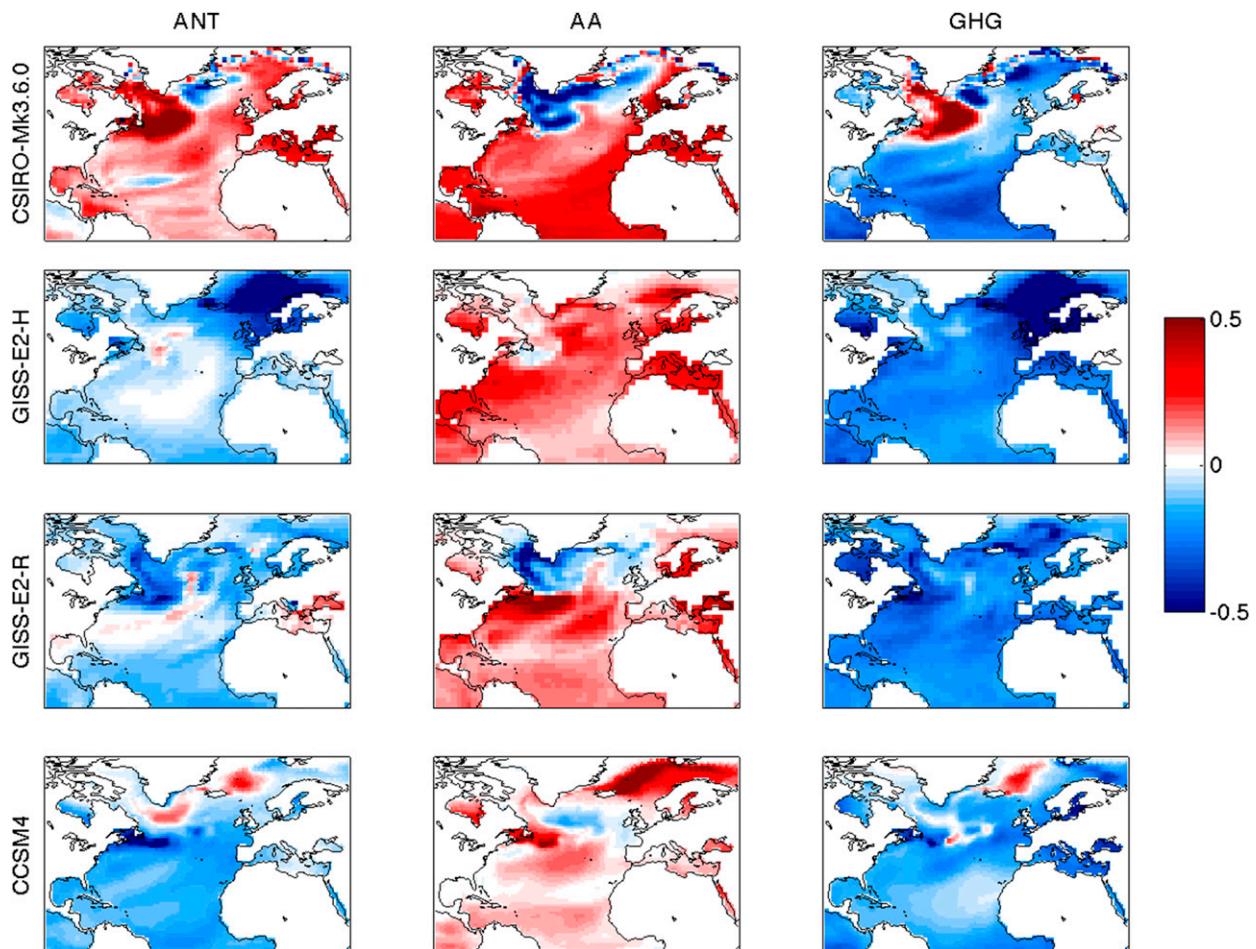


FIG. 17. Warm–cold SST composite patterns computed as the difference of 1930–50 minus 1960–80 for individual models—top to bottom, CSIRO-Mk3.6.0, GISS-E2-H, GISS-E2-R, and CCSM4—under (left) ANT, (middle) AA, and (right) GHG forcing conditions.

forecasts start on year 1960 and therefore could not be usefully exploited to address this specific case study.

Finally, the robustness of this assessment will depend on the degree of realism of the models featuring a cluster-B type of behavior. While there are hints that these models realistically capture the observed evolution of the upper-ocean thermal state over the Atlantic sector (interhemispheric SST dipole and upper-ocean heat content variability), a much poorer consistency is found between observed and simulated sea surface salinities in the subpolar gyre area.

To conclude, a potentially important role for anthropogenic aerosols and GHGs on the observed North Atlantic multidecadal variability has clear implications for decadal predictability and predictions [see Bellucci et al. (2015) for a recent overview on the nonoceanic sources of decadal predictability]. The uncertainty associated with alternative aerosol and GHG emission scenarios should be duly accounted for in designing the next-generation protocols for coordinated decadal forecast experiments.

*Acknowledgments.* We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We also wish to thank three anonymous reviewers for their constructive comments.

#### REFERENCES

- Baines, P. G., and C. K. Folland, 2007: Evidence for a rapid global climate shift across the late 1960s. *J. Climate*, **20**, 2721–2744, doi:10.1175/JCLI4177.1.
- Bellucci, A., and Coauthors, 2015: Advancements in decadal climate predictability: The role of nonoceanic drivers. *Rev. Geophys.*, **53**, 165–202, doi:10.1002/2014RG000473.

- Booth, B. B. B., N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, doi:10.1038/nature10946.
- Church, J., and Coauthors, 2005: Significant decadal-scale impact of volcanic eruptions on sea level and ocean heat content. *Nature*, **438**, 74–77, doi:10.1038/nature04237.
- Clement, A., K. Bellomo, L. N. Murphy, M. A. Cane, T. Mauritsen, G. Radel, and B. Stevens, 2015: The Atlantic multidecadal oscillation without a role for ocean circulation. *Science*, **350**, 320–324, doi:10.1126/science.aab3980.
- Deser, C., and M. L. Blackmon, 1993: Surface climate variations over the North Atlantic Ocean during winter: 1900–89. *J. Climate*, **6**, 1743–1753, doi:10.1175/1520-0442(1993)006<1743:SCVOTN>2.0.CO;2.
- Dickson, R. R., J. Meincke, S.-A. Malmberg, and A. J. Lee, 1988: The Great Salinity Anomaly in the northern North Atlantic 1968–1982. *Prog. Oceanogr.*, **20**, 103–151, doi:10.1016/0079-6611(88)90049-3.
- Drijfhout, S., G. J. van Oldenborgh, and A. Cimadoribus, 2012: Is a decline of AMOC causing the warming hole above the North Atlantic in observed and modeled warming patterns? *J. Climate*, **25**, 8373–8379, doi:10.1175/JCLI-D-12-00490.1.
- Eyring, V., S. Bony, G. A. Meehl, C. 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. *Geosci. Model Dev.*, **9**, 1937–1958, doi:10.5194/gmdd-8-10539-2015.
- Hodson, D. L. R., J. I. Robson, and R. T. Sutton, 2014: An anatomy of the cooling of the North Atlantic Ocean in the 1960s and 1970s. *J. Climate*, **27**, 8229–8243, doi:10.1175/JCLI-D-14-00301.1.
- Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.*, **32**, L20708, doi:10.1029/2005GL024233.
- Kushnir, Y., 1994: Interdecadal variations North Atlantic sea surface temperature and associated atmospheric conditions. *J. Climate*, **7**, 141–157, doi:10.1175/1520-0442(1994)007<0141:IVINAS>2.0.CO;2.
- Levy, H., and Coauthors, 2013: The roles of aerosol direct and indirect effects in past and future climate change. *J. Geophys. Res. Atmos.*, **118**, 4521–4532, doi:10.1002/jgrd.50192.
- Maher, N., A. Sen Gupta, and M. H. England, 2014: Drivers of decadal hiatus periods in the 20th and 21st centuries. *Geophys. Res. Lett.*, **41**, 5978–5986, doi:10.1002/2014GL060527.
- Mariotti, A., and A. Dell'Aquila, 2012: Decadal climate variability in the Mediterranean region: Roles of large-scale forcings and regional processes. *Climate Dyn.*, **38**, 1129–1145, doi:10.1007/s00382-011-1056-7.
- Marullo, S., V. Artale, and R. Santoleri, 2011: The SST multidecadal variability in the Atlantic-Mediterranean region and its relation to AMO. *J. Climate*, **24**, 4385–4401, doi:10.1175/2011JCLI3884.1.
- Mohino, E., S. Janicot, and J. Bader, 2011: Sahel rainfall and decadal to multi-decadal sea surface temperature variability. *Climate Dyn.*, **37**, 419–440, doi:10.1007/s00382-010-0867-2.
- Otterå, O. H., M. Bentsen, H. Drange, and L. L. Suo, 2010: External forcing as a metronome for Atlantic multidecadal variability. *Nat. Geosci.*, **3**, 688–694, doi:10.1038/ngeo955.
- Rahmstorf, S., J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, and E. J. Schaffernicht, 2015: Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat. Climate Change*, **5**, 475–480, doi:10.1038/nclimate2554.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Smith, S. J., J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. D. Arias, 2011: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.*, **11**, 1101–1116, doi:10.5194/acp-11-1101-2011.
- Stenchikov, G., T. L. Delworth, V. Ramaswamy, R. J. Stouffer, A. Wittenberg, and F. Zeng, 2009: Volcanic signals in oceans. *J. Geophys. Res.*, **114**, D16104, doi:10.1029/2008JD011673.
- Sutton, R. T., and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, **309**, 115–118, doi:10.1126/science.1109496.
- Swingedouw, D., P. Ortega, J. Mignot, E. Guilyardi, V. Masson-Delmotte, P. G. Butler, M. Khodri, and R. Sférian, 2015: Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions. *Nature Commun.*, **6**, 6545, doi:10.1038/ncomms7545.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2009: A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf).]
- , —, and —, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Terray, L., 2012: Evidence for multiple drivers of North Atlantic multi-decadal climate variability. *Geophys. Res. Lett.*, **39**, L19712, doi:10.1029/2012GL053046.
- Thiéblemont, R., K. Matthes, N.-E. Omrani, K. Kodera, and F. Hansen, 2015: Solar forcing synchronizes decadal North Atlantic climate variability. *Nature Commun.*, **6**, 8268, doi:10.1038/ncomms9268.
- Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones, 2008: A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature*, **453**, 646–649, doi:10.1038/nature06982.
- , J. M. Wallace, J. J. Kennedy, and P. D. Jones, 2010: An abrupt drop in Northern Hemisphere sea surface temperature around 1970. *Nature*, **467**, 444–447, doi:10.1038/nature09394.
- Ting, M., Y. Kushnir, R. Seager, and C. Li, 2009: Forced and internal twentieth-century SST trends in the North Atlantic. *J. Climate*, **22**, 1469–1481, doi:10.1175/2008JCLI2561.1.
- Trenberth, K. E., and D. J. Shea, 2006: Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.*, **33**, L12704, doi:10.1029/2006GL026894.
- Wilcox, L., E. J. Highwood, and N. J. Dunstone, 2013: The influence of anthropogenic aerosol on multi-decadal variations of historical global climate. *Environ. Res. Lett.*, **8**, 024033, doi:10.1088/1748-9326/8/2/024033.
- Zhang, R., and Coauthors, 2013: Have aerosols caused the observed Atlantic multidecadal variability? *J. Atmos. Sci.*, **70**, 1135–1144, doi:10.1175/JAS-D-12-0331.1.