Title
Are 100 ensemble members enough to capture the remote atmospheric response to 12°C arctic sea ice loss?

Journal
Journal of Climate, 34(10)

ISSN
0894-8755

Authors
Peings, Y
Labe, ZM
Magnusdottir, G

Publication Date
2021-05-15

DOI
10.1175/JCLI-D-20-0613.1

Are 100 Ensemble Members Enough to Capture the Remote Atmospheric Response to +2°C Arctic Sea Ice Loss?

YANNICK PEINGS,[a] ZACHARY M. LABE,[a] AND GUDRUN MAGNUSDOTTIR[a]

[a] *Department of Earth System Science, University of California Irvine, Irvine, California*

(Manuscript received 4 August 2020, in final form 21 January 2021)

ABSTRACT: This study presents results from the Polar Amplification Multimodel Intercomparison Project (PAMIP) single-year time-slice experiments that aim to isolate the atmospheric response to Arctic sea ice loss at global warming levels of +2°C. Using two general circulation models (GCMs), the ensemble size is increased up to 300 ensemble members, beyond the recommended 100 members. After partitioning the response in groups of 100 ensemble members, the reproducibility of the results is evaluated, with a focus on the response of the midlatitude jet streams in the North Atlantic and North Pacific. Both atmosphere-only and coupled ocean–atmosphere PAMIP experiments are analyzed. Substantial differences in the midlatitude response are found among the different experiment subsets, suggesting that 100-member ensembles are still significantly influenced by internal variability, which can mislead conclusions. Despite an overall stronger response, the coupled ocean–atmosphere runs exhibit greater spread due to additional ENSO-related internal variability when the ocean is interactive. The lack of consistency in the response is true for anomalies that are statistically significant according to Student's *t* and false discovery rate tests. This is problematic for the multimodel assessment of the response, as some of the spread may be attributed to different model sensitivities whereas it is due to internal variability. We propose a method to overcome this consistency issue that allows for more robust conclusions when only 100 ensemble members are used.

KEYWORDS: Arctic; Sea ice; Atmosphere-ocean interaction; Teleconnections; Numerical analysis/modeling; Climate variability

1. Introduction

Accelerated warming of the Arctic in the last 40 years is a conspicuous signal of climate change that has received much attention in recent years. The temperature has risen faster in the Arctic than over the rest of the globe, a phenomenon called polar amplification, or Arctic amplification (AA) in the case of the Northern Hemisphere (NH). Arctic amplification is associated with a sharp decline in sea ice extent and thickness in the Arctic Ocean (Serreze and Stroeve 2015; Lindsay and Schweiger 2015), decline that is expected to accelerate with increasing anthropogenic emissions in the twenty-first century (Kay et al. 2011). Given the dramatic amplitude of observed changes in the Arctic, not only in terms of climate, but also ecosystems (e.g., Grebmeier 2012), understanding causes and consequences for Arctic amplification has become a key research question.

While causes and local consequences of AA are better understood (e.g., Yoshimori et al. 2017; Stuecker et al. 2018), how it affects remote areas of the globe is unclear. Many consequences of a warmer Arctic in midlatitudes have been suggested, and are discussed in several review papers (Cohen et al. 2014; Barnes and Screen 2015; Vihma 2014; Vavrus 2018). These reviews report a lack of consensus among studies. For instance, different studies find different responses to Arctic sea ice loss, especially in terms of the midlatitude jets or modes of atmospheric variability, such as the North Atlantic Oscillation (NAO; Hurrell 1995) or northern annular mode (NAM) (Thompson and Wallace 2000). The response to large sea ice loss as projected at the end of the twenty-first century exhibits relative consensus in coupled ocean–atmosphere models (Screen et al. 2018), but whether contemporaneous Arctic amplification already affects extreme weather in midlatitudes is particularly debated (Mori et al. 2019; Blackport and Screen 2020; Cohen et al. 2020). Short observational records limit the robustness of statistical analyses in observations and the attribution of causality (Sorokina et al. 2016; Kolstad and Screen 2019; Peings 2019). Moreover, numerical experiments using climate models with altered sea ice show a range of atmospheric responses (e.g., Vihma 2014). This is due to different model physics and sensitivities to sea ice loss, especially differences in the model background state (Smith et al. 2017; Labe et al. 2019). Also, differences in protocol among studies limit the possibilities for understanding discrepancies in the results.

In view of these limitations, the Polar Amplification Model Intercomparison Project (PAMIP) was created to provide a framework for coordinated sea ice loss experiments (Smith et al. 2019). The project includes a set of atmosphere-only [i.e., prescribed with sea surface temperature (SST) and sea ice concentration (SIC)] and coupled ocean–atmosphere simulations to explore the atmospheric response to Arctic and Antarctic sea ice loss. Some of the runs are designed to also investigate causes for polar amplification, in particular the role of remote SST changes (i.e., Perlwitz et al. 2015). The forcing fields consist of three time slices of preindustrial, present-day, and future SIC/SST to reveal how historical and projected sea ice loss have affected and may affect the global climate. The future SIC/SST fields have been designed to represent +2°C of

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0613.s1.

Corresponding author: Yannick Peings, ypeings@uci.edu

DOI: 10.1175/JCLI-D-20-0613.1

global warming, relative to the preindustrial period. To isolate the forced response, attributable to SIC/SST anomalies, from high internal variability in the atmosphere, large ensembles of simulations are recommended. For example, for the preindustrial, present-day, and future time-slice experiments, 100 members of 14-month runs are recommended by the PAMIP protocol.

The emergence of large ensembles of simulations (Kay et al. 2015; Maher et al. 2019; Deser et al. 2020) has allowed for a greater recognition of the large influence of internal variability in climate simulations. This is especially true when assessing regional trends in climate change projections (Deser et al. 2016) and hiatus in global change (Bengtsson and Hodges 2019), but this is also true in sensitivity experiments such as the PAMIP runs. Screen et al. (2014) estimate the number of ensemble members needed to identify a robust response to Arctic sea ice loss for different variables. They find that in order to be detected robustly, dynamical and upper-level variables need more ensemble members than thermodynamic variables. They suggest that 50 ensemble members (of a single year or season) are the minimum to detect a robust sea level pressure response. With increased computational capability, 100 members (again, of a single year or season) has become the norm in such model sensitivity experiments, and it is generally believed to be sufficient to identify a robust response to the prescribed forcings. However, except for Screen et al. (2014), and more recently Labe et al. (2019) and Liang et al. (2019), to our knowledge little attention has been given to estimate how internal variability may affect the results of sea ice loss or similar sensitivity experiments. In particular, it is unknown whether the conclusions of sea ice loss experiments may differ when ensembles with a greater ensemble size than 100 are carried out.

In this paper, we present results from PAMIP time-slice experiments that have been run beyond 100 members with two different atmospheric models. We explore the reproducibility of the results in different groups of 100-member ensembles, and find that a 100-member ensemble may not be enough to robustly assess the midlatitude atmospheric response to +2°C Arctic sea ice loss. Results from different 100-member subsets are not entirely consistent, despite statistical significance when using traditional statistical tests. We propose a method to overcome this consistency issue, and find that a more robust response can be assessed from 100-member experiments when it is used.

2. Methods

a. Models

Sea ice loss perturbation experiments from two general circulation models (GCMs) are used. The first one is the Community Earth System Model (CESM), version 1, from the National Center for Atmospheric Research (NCAR). From a configuration of CESM1, we use the Whole Atmosphere Community Climate Model, version 4 (WACCM4; Marsh et al. 2013). WACCM4 includes 66 vertical levels (up to 5.1 $\times$ 10^{-6} hPa, ~140 km) and uses CAM4 physics. We use the specified chemistry version of WACCM4 (SC-WACCM4; Smith et al. 2014), which is computationally less expensive to run, but simulates dynamical stratosphere–troposphere coupling and stratospheric variability that are comparable to the interactive chemistry model version. The SC-WACCM4 experiments are run with a horizontal resolution of 1.9° latitude $\times$ 2.5° longitude and include present-day (year 2000) radiative forcing. A repeating 28-month full cycle of the quasi-biennial oscillation (QBO) is included in the SC-WACCM4 experiments through nudging of the equatorial stratospheric winds to observed radiosonde data. From one ensemble member to the next, the QBO is initialized using the following month of the 28-month QBO cycle, so that each phase of the QBO is represented in our ensemble and it does not skew the results in one direction or the other (Labe et al. 2019). In the coupled ocean–atmosphere configuration, the ocean component of CESM1 is the Parallel Ocean Program version 2 (POP2). CESM1 also includes the Los Alamos sea ice model (CICE), the Community Land Model version 4 (CLM4) and the River Transport Model (RTM). CLM is run at a horizontal resolution of 1.9° $\times$ 2.5°; POP2 and CICE are run at nominal 1° resolution with higher resolution near the equator than at the poles. Further details about CESM1 are given in Hurrell et al. (2013).

The second GCM is the Energy Exascale Earth System Model, version 1 (E3SMv1; Golaz et al. 2019), from the United States Department of Energy (DOE). The E3SMv1 atmospheric component was developed from CAM5.3 and includes additional turbulence parameterizations and improvements to cloud and aerosol physics (Rasch et al. 2019). E3SMv1 includes 72 vertical layers (compared to 30 in CAM5) with a model top at ~0.1 hPa (~60 km). We use the lower-resolution version of E3SMv1 with a horizontal resolution of 100 km and present-day (year 2000) radiative forcing. While the model includes an internally generated QBO-like oscillation of the equatorial stratospheric wind, the period is too short, and the westerly winds are too strong (Richter et al. 2019). Ocean and sea ice components in E3SMv1 are based on the Model for Prediction Across Scales (MPAS) and the river transport component is the Model for Scale Adaptive River Transport (MOSART). The land model is a slightly revised version of that found in CESM1. Details on all the coupled model components can be found in Golaz et al. (2019).

Simulations run in atmosphere only mode (i.e., with prescribed SST/SIC) are referred to as atmospheric general circulation model (AGCM) runs. When the ocean is interactive, the simulations are referred to as ocean–atmosphere general circulation model (OAGCM) runs.

b. PAMIP experiments

All the experiments that are used in this paper are listed in Table 1. Here is a short description of them.

1) AGCM RUNS: PAMIP-1.5 AND PAMIP-1.6 (FIXED SEA ICE THICKNESS)

PAMIP experiments performed with SC-WACCM4 and E3SMv1 are used to explore the atmospheric response to +2°C Arctic sea ice loss. AGCM simulations forced with preindustrial Arctic SIC (experiment PAMIP-1.5; Smith et al. 2019) are compared to simulations forced with future + 2°C Arctic sea ice (experiment PAMIP-1.6). Note that we could use the

TABLE 1. Overview of the numerical simulations.

Model	Type of simulation	Expt name	Description
SC-WACCM4 (300 ensemble members per simulation)	AGCM with fixed sea ice thickness	PAMIP-1.5	Preindustrial Arctic sea ice concentration, present-day SST
		PAMIP-1.6	+2°C Arctic sea ice concentration, present-day SST
	AGCM with prescribed sea ice thickness	PAMIP-1.9	Present-day Arctic sea ice concentration and thickness, present-day SST
		PAMIP-1.10	Future Arctic sea ice concentration and thickness, present-day SST
	OAGCM (nudging of sea ice volume, SIC and SIT similar to PAMIP-1.5 and PAMIP-1.6)	PAMIP-2.2	Preindustrial Arctic sea ice concentration, present-day SST
		PAMIP-2.3	+2°C Arctic sea ice concentration, present-day SST
E3SMv1 (200 ensemble members per simulation)	AGCM with fixed sea ice thickness	PAMIP-1.5-E3SM	Preindustrial Arctic sea ice concentration, present-day SST
		PAMIP-1.6 -E3SM	+2°C Arctic sea ice concentration, present-day SST

PAMIP-1.1 runs (with present-day SIC/SST) as a reference, but in order to maximize the signal-to-noise ratio and discuss the impact of +2°C sea ice loss, the preindustrial SIC PAMIP-1.5 runs are used. The SIC fields are constructed from an ensemble of 31 CMIP5 simulations, as detailed in Smith et al. (2018). Preindustrial SIC is estimated by identifying 30-yr periods in preindustrial control runs from the CMIP5 models that have a global mean temperature closest to an estimate of preindustrial global mean temperature (13.67°C; Haustein et al. 2017). Similarly, +2°C sea ice is estimated from the representative concentration pathway 8.5 $W m^{-2}$ (RCP8.5) runs, by selecting 30-yr periods with a temperature 2°C warmer than preindustrial (15.67°C). A similar protocol is followed to estimate +2°C SST that are associated with sea ice loss. SST is set to future values where SIC fraction changes more than 0.1 between the future and preindustrial SIC fields, a common practice in AGCM sea ice sensitivity experiments (e.g., Screen et al. 2013). Outside the Arctic, SST are set to present-day values, as is Antarctic SIC. Sea ice thickness is set to 2 m in the Northern Hemisphere and 1 m in the Southern Hemisphere. Further details on the PAMIP forcing fields can be found in Smith et al. (2019).

Following the PAMIP protocol, the simulations are run from 1 April to 31 May of the following year (14 months). After discarding the first two months for spinup (April–May of year 1), this gives us a full year of simulation spanning the annual cycle of SIC (June to May). PAMIP recommends running 100 ensemble members for these simulations, but for the purpose of this study we have extended them to 300 members for SC-WACCM4 (thanks to the relatively low computational costs of this model) and 200 members for E3SMv1 (more computationally demanding than SC-WACCM4).

2) AGCM RUNS: PAMIP-1.9 AND PAMIP-1.10 (PRESCRIBED SEA ICE THICKNESS)

PAMIP simulations 1.9 and 1.10 are used to verify whether our findings are robust in other PAMIP AGCM runs. Simulations PAMIP-1.9 and PAMIP-1.10 are designed to reveal the influence of sea ice thickness. Sea ice concentration is similar to PAMIP-1.1 (in PAMIP-1.9) and PAMIP-1.6 (in PAMIP-1.10), but sea ice thickness is not constant anymore; it is set to estimates of present-day and future values [see Smith et al. (2019) for details]. The PAMIP forcing files have been constructed with a difference of 0.6°C in global mean temperature between the present-day and preindustrial periods. Therefore, the difference between PAMIP-1.10 and PAMIP-1.9 (future minus present-day conditions) corresponds to +1.4°C sea ice loss. However, with the inclusion of sea ice thickness anomalies, the forcing is stronger than when sea ice concentration alone is considered (Labe et al. 2018).

3) OAGCM RUNS: PAMIP-2.2 AND PAMIP-2.3

To assess the influence of ocean–atmosphere coupling on the results, we also analyze short coupled runs that are part of the PAMIP set of experiments. Only simulations using SC-WACCM4 are discussed here, as their E3SMv1 counterpart had not been completed at the time of this study. PAMIP-2.2 is similar to PAMIP-1.5, but it includes an interactive ocean. Like for the AGCM runs, the ensembles are created by adding a tiny temperature perturbation in the atmosphere. The initial state of the ocean is similar across all ensemble members. To force the coupled model toward the sea ice target state, in the Arctic SIC and sea ice volume are nudged toward the preindustrial SIC field (similar to the one used for PAMIP-1.5) using a relaxation nudging coefficient of 5 h for SIC and 1 day for ice volume. The ice volume target is constructed to retrieve 2-m sea ice thickness in the Arctic and 1-m thickness in the Antarctic, as in PAMIP-1.5. PAMIP-2.3 is the coupled equivalent of PAMIP-1.6; that is, SIC is nudged toward + 2°C values in the Arctic and present-day values in the Antarctic, with thickness maintained constant. In these runs, sea ice is similar to PAMIP-1.5 and PAMIP-1.6; the only difference is that ocean dynamics is included and SST is free to evolve, including in areas of sea ice loss. Like for the AGCM runs, the

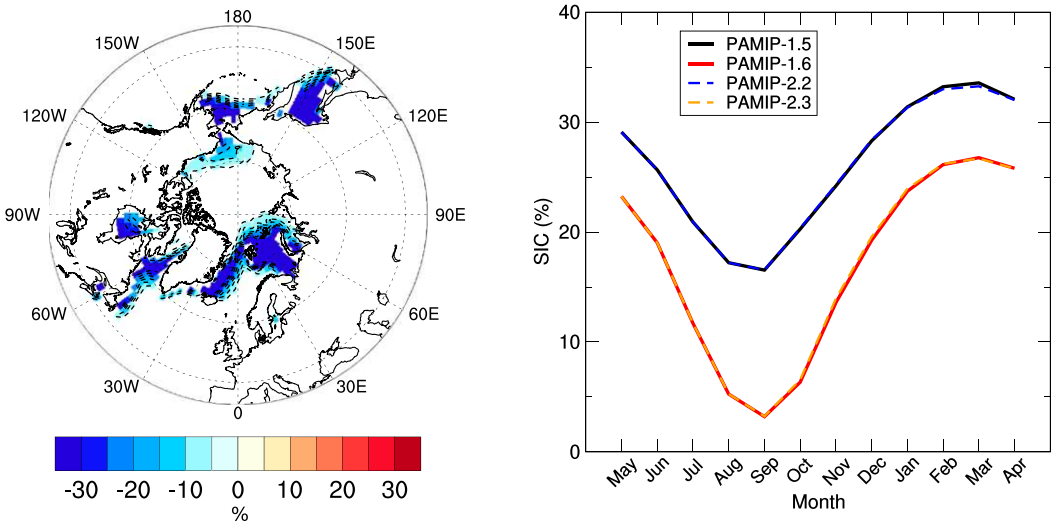

FIG. 1. (a) Sea ice concentration anomalies (%) imposed in winter (DJFM), in PAMIP-1.6 (+2°C Arctic sea ice, AGCM) minus PAMIP-1.5 (preindustrial Arctic sea ice, AGCM). (b) Annual cycle of sea ice concentration (%) in the NH high latitudes (north of 45°N) in PAMIP-1.5 (solid black line) and PAMIP-1.6 (dashed red line) and the equivalent coupled runs, PAMIP-2.2 (dashed blue line) and PAMIP-2.3 (dashed orange line).

simulations are run for 14 months, and the two first months are discarded to account for model spinup. In this regard, these simulations only highlight the short-term influence of ocean–atmosphere coupling, mostly thermodynamic exchanges at the air–sea interface. Full ocean dynamics adjustment to sea ice loss takes several decades (e.g., Sun et al. 2018), so these processes are not accounted for in these short coupled runs. That is the motivation for centennial coupled ocean–atmosphere PAMIP simulations that are designed to explore the role of longer-term oceanic adjustment to sea ice loss (Smith et al. 2019).

c. Statistical significance of the results

To test the statistical significance of the response to Arctic sea ice loss, we use a classic two-tailed Student's t test (STT). For a given variable, the difference between future and preindustrial values is compared at each grid point. If the null hypothesis that the two groups are indiscernible is rejected at the 5% confidence level (p value lower or equal to 0.05), then the difference is considered significant. Although this is a widely used method, with multiple tests at each grid point, rejection of the global null hypothesis is overestimated (Wilks 2016). To account for this, we use the false discovery rate (FDR), as recommended in Wilks (2016). Local null hypotheses are rejected if their respective p values are no larger than a threshold level that depends on the distribution of the sorted p values, after the p values are sorted in ascending order $[p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(N)}]$:

$$p'_{\mathrm{FDR}} = \max_{i=1,\ldots,N} [p_{(i)} : p_{(i)} \leq (i/N)\alpha_{\mathrm{FDR}}],$$

where i represents one ensemble member, N is the number of ensemble members, and α_{FDR} is the chosen control level for the FDR. That is, the threshold p'_{FDR} for rejecting local null hypotheses is the largest $p(i)$ that is no larger than the fraction of α_{FDR} specified by i/N. The FDR reduces the fraction of significant grid point test results that are spurious. In this study, we use a value of 0.1 forα_{FDR}, that is, twice the value of the 0.05 threshold we use for the STT, as recommended by Wilks (2016).

3. Results

a. Reproducibility of the atmospheric response in 100-member ensembles: AGCM runs

Figure 1a shows the SIC forcing prescribed in the +2°C Arctic SIC runs (PAMIP-1.6, PAMIP-2.3), relative to the preindustrial Arctic SIC runs (PAMIP-1.5, PAMIP-2.2), during winter [December–March (DJFM) average]. The SIC anomalies are largest in the Barents–Kara and Greenland Seas, as well as in the Okhotsk and Bering Seas in the Pacific sector. The annual cycle of SIC is very similar in the AGCM and OAGCM experiments, outlining the efficiency of the implemented nudging technique (Fig. 1b). Sea ice thickness is also very similar in the runs, with a fixed value of 2 m in the Arctic (not shown). SIC peaks in September (Fig. 1b), but in this study we analyze the winter season when ocean–atmosphere heat exchanges are maximum and the impacts on the large-scale atmospheric circulation are expected to be most pronounced. In the rest of the paper, we refer to the difference between two experiments as the "response" of the model. However, keep in mind that these responses do not necessarily represent the true response to the forcing because they include noise due to internal variability, as will be shown throughout the course of the paper.

First, we focus on the SC-WACCM4 experiments. The local response to the sea ice forcing is consistent with previous studies (i.e., heating is found near the surface above sea ice loss areas, which propagates upward and results in a thermal

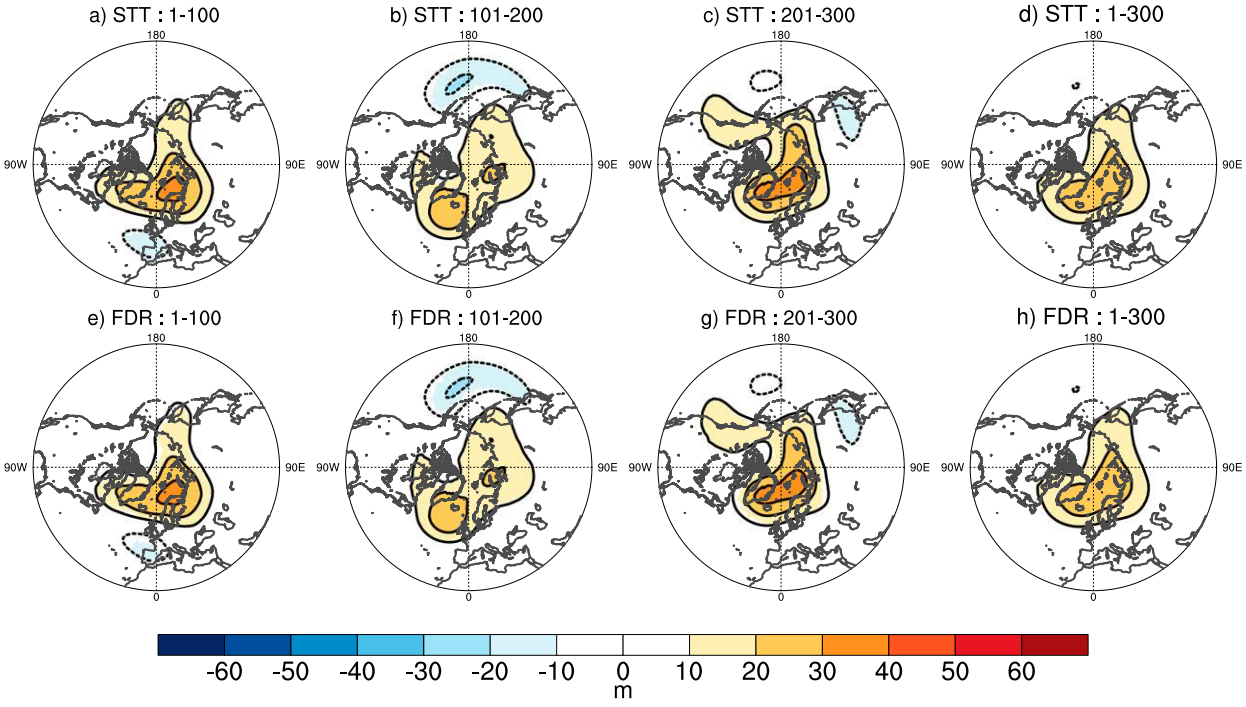

FIG. 2. Response of DJFM Z500 (m) in PAMIP-1.6 (+2°C Arctic sea ice, AGCM) minus PAMIP-1.5 (preindustrial Arctic sea ice, AGCM) of SC-WACCM4 in members (a) 1–100, (b) 101–200, (c) 201–300, and (d) 1–300 (total ensemble mean). A two-tailed Student's *t* test (STT) is applied to test significance of the anomalies, and only anomalies that are significant at the 95% confidence level are shaded. (e)–(h) As in (a)–(d), but after applying the false discovery rate (FDR) test.

expansion of the lower troposphere). This is shown in Fig. 2 through the response of the geopotential height at 500 hPa (Z500). To assess the robustness of the response, the 300 members are split into three groups of 100 members (minimum ensemble size recommended by PAMIP). The responses for members 1–100, 101–200, and 201–300 are shown in the first three columns of Fig. 2, and the fourth column shows the 300-member ensemble mean response. The top row uses the Student's *t* test to identify anomalies that are significant at the 95% confidence level, while the bottom row shows the same anomalies but after applying the false discovery rate algorithm to refine the areas of statistical significance. Not surprisingly, the high-latitude (i.e., north of 50°N) Z500 response is robust across the three subsets of experiments, and areas of statistical significance are very similar whether the Student's or FDR tests are applied. In the midlatitudes, however, there are some striking differences, including a trough over the North Pacific (or reinforcement of the Aleutian low) in members 101–200 that is not found in the other subsets (Figs. 2b,f).

This nonrobustness in the response of the midlatitude atmospheric circulation is more obvious when looking at a dynamical variable, such as the zonal wind at 700 hPa (U700, Fig. 3). In members 1–100 (Fig. 3a), a dipole of easterly (negative) and westerly (positive) U700 anomalies is found over the North Atlantic, which represents an equatorward shift of the eddy-driven North Atlantic jet (Woollings et al. 2014). This is a commonly found response to Arctic sea ice loss (Sun et al. 2015; Screen et al. 2018), which the Student's *t* test identifies as robust. However, the dipole is shifted south in members 101–200 (Fig. 3b) and is absent in members 201–300 (Fig. 3c) where only the easterly anomalies are found. Similarly, the North Pacific response is not consistent among the three subsets. A pronounced dipole of U700 is found in members 101–200 (Fig. 3b), with westerly anomalies that represent a reinforcement and southward displacement of the jet. This is consistent with the findings of Ronalds et al. (2020), who identified this response in these SC-WACCM4 experiments as well as in similar PAMIP runs from three other models. Note, however, that they looked at January–February averages, and only at the first 100 members of the set of SC-WACCM4 simulations that we are using here. As seen in Figs. 3a and 3c, over DJFM, the jet reinforcement is absent in members 1–100, and is much weaker in members 201–300.[1] Looking at the bottom row, we see that the FDR test is more reliable than the Student's *t* test to identify anomalies that are truly robust (i.e., reproducible from one 100-member subset to the other). For example, the North Atlantic westerly anomalies of members 1–100 are now nonsignificant (Fig. 3e), as are the North Pacific westerly

[1] To put our results in perspective with Ronalds et al. (2020), although there are large disparities in the amplitude of the anomalies (stronger in members 101–200), over January–February the jet reinforcement is consistently found in the three 100-member subsets (not shown). Therefore their discussion of the SC-WACCM4 results is not affected by the use of members 1–100 only.

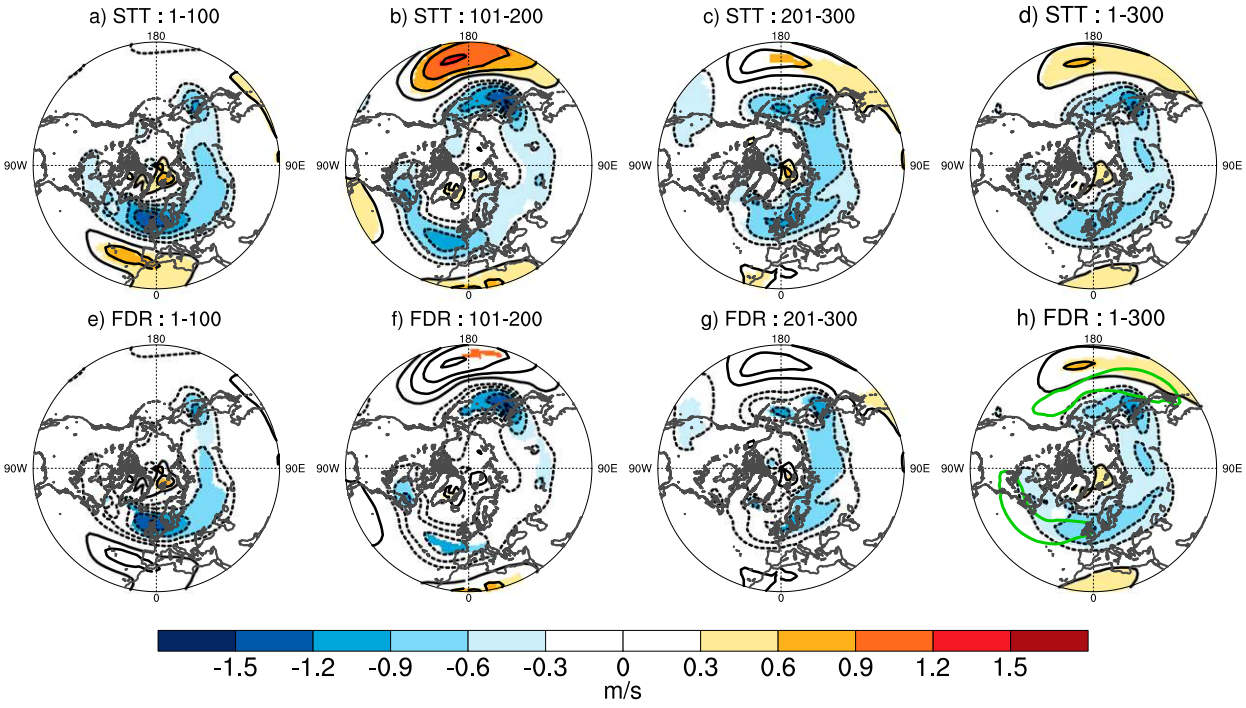

FIG. 3. As in Fig. 2, but for zonal wind at 700 hPa (m s^{-1}). The climatology is shown in green contours in (h) (12 m s^{-1} contour interval).

anomalies in members 101–200 (Fig. 3f). However, nonnegligible differences between the subsets are still found. The 300-member ensemble mean exhibits greater robustness of the anomalies, as expected from larger sample size, in both the Student's *t* test and the FDR cases (Figs. 3d,h). Based on this 300-member ensemble mean, one can conclude that the North Pacific jet is reinforced by Arctic sea ice loss, but this has to be reconsidered given that this signal is completely absent from members 1–100. Similarly, high fluctuations in the regional response are also found when comparing runs PAMIP-1.10 and PAMIP-1.9 (Fig. S1 in the online supplemental material, future minus present-day Arctic sea ice with sea thickness anomalies included). In particular, a reinforcement of the North Pacific jet is found in members 101–300 (Figs. S1b,c) but it is absent in subset 1–100, and not significant according to the FDR test in subset 101–200. The FDR test identifies this response as statistically significant in subset 201–300 (Fig. S1g), highlighting the limitation of conventional statistical tests to discard nonrobust signals in such experiments.

To further illustrate the lack of consistency in the midlatitude jet responses in winter, Fig. 4 shows the response of jet metrics in the North Atlantic and North Pacific basins. In the North Atlantic, we find significant differences in the jet position, measured using the jet position index (JPI). The JPI identifies the latitude of the jet core, defined as the maximum of westerlies between 25° and 75°N over the North Atlantic sector (60°W–60°E). For each ensemble member, the response of the North Atlantic JPI is calculated, and the distribution of the JPI response across the ensemble members is shown using a boxplot/whisker representation. Red diamonds indicate the mean of the distribution, and a two-tailed Student's *t* test is used to determine whether the ensemble means differ significantly between different subsets of simulations PAMIP-1.5 and 1.6 (members 1–100, 101–200, 201–300, and 1–300, as previously). Consistent with Fig. 3, a statistically significant decrease in the JPI, or equatorward shift of the jet, is found in members 1–100 (Fig. 4a). However, this signal is gone in members 101–300, although it dominates the 300-member ensemble mean in which a statistically significant decrease is detected. In the North Pacific, the jet reinforces, more than it shifts, so we use the jet speed index (JSI), defined as the maximum strength of the westerlies between 25° and 75°N in the North Pacific sector (180°–300°E) (Fig. 4b). Again, we find that although a statistically significant increase in the JSI is found in members 101–200 (in line with Fig. 3b), this is a nonrobust result when extending the ensemble size.

Three subsets of the full 300-member ensemble are shown here, but because the members are independent and uncorrelated, all combinations of 100-member subsets could be analyzed to infer the range of possible response in a certain metric. Following the central limit theorem, an ensemble of subsets (here 100 members) randomly selected from a population that has a certain mean and variance (here the full ensemble of 300) has a normal distribution. Such sampling distribution of the North Pacific JSI (JSI-PA), for 100 000 possible 100-member subsets of the 300-member ensemble (without replacement), is shown in Fig. S2a. The corresponding anomalies of JSI-PA in subsets 1–100, 101–200, and 201–300 are indicated for reference. The JSI-PA anomaly in subset 101–200 (~0.3 m s^{-1}) stands out as a relatively rare occurrence of the possible

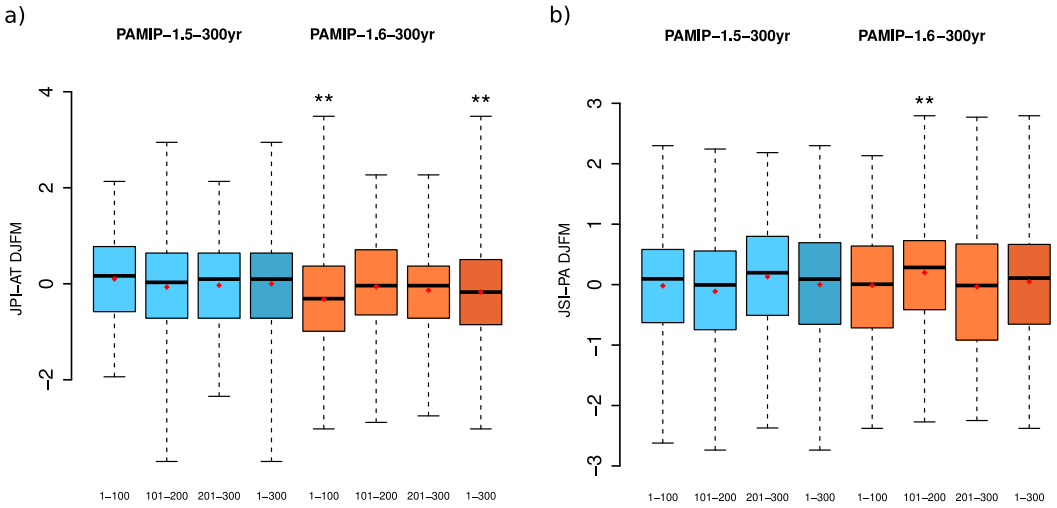

FIG. 4. (a) Distribution of normalized anomalies in the North Atlantic jet position index (JPI) in DJFM, in members 1–101, 101–200, 201–300, and 1–300 of PAMIP-1.5 (preindustrial Arctic SIC; blue) and PAMIP-1.6 (+2° Arctic SIC; red) of SC-WACCM4. The mean and standard deviation of the full 300-member PAMIP-1.5 ensemble is used to normalize the JPI anomaly of each ensemble member. Each PAMIP-1.6 ensemble is compared to its corresponding PAMIP-1.5 ensemble (e.g., 101–200 of PAMIP-1.6 is compared to 101–200 of PAMIP-1.5) to evaluate statistical significance in the difference of the ensemble means (red diamond): One asterisk (*) indicates $p < 0.1$; two asterisks (**) indicate $p < 0.05$ (two-tailed Student's t test). (b) As in (a), but for the North Pacific jet strength index (JSI).

100-member mean JSI-PA anomalies in the ensemble (less than 4% chance). Subset 1–100 is representative of the center of the distribution (i.e., it is a good approximation of the 300-member ensemble mean for this particular metric), but of course one cannot know this before extending the ensemble size enough to capture a better picture of the full 100-member sampling distribution. We will come back to the question of how to better assess robustness with only 100 members in the last section of the paper.

An important question is how the polar stratosphere reacts to the sea ice loss forcing, since previous work has shown that it may play a role in communicating the response to sea ice in the midlatitudes (Kim et al. 2014; Peings and Magnusdottir 2014; Zhang et al. 2018). Anomalies in the 50-hPa geopotential height (Z50) are shown in Fig. 5. A weak warming of the polar vortex is found, but it is only statistically significant (after the FDR has been applied) in members 201–300 (Fig. 5g). A legitimate question is whether polar stratospheric variability can explain the spread of the tropospheric response among subsets. To address this, we plot in Fig. 5i the North Pacific JSI response in function of the polar vortex response, expressed as the 10-hPa zonal mean zonal wind anomaly at 65°N. There is a moderate correlation between the two ($R = -0.38$) so that polar stratosphere variability and JPI anomalies share ~15% of variance in the 300-member ensemble. Therefore, polar stratosphere variability only explains a small fraction of the spread in JPI, with tropospheric internal variability likely being the major driver. Moreover, without further experiments, it is difficult to discuss causality in the relationship. Internal variability in the polar stratosphere may be a source of noise that increases the spread in the tropospheric response, but it also represents a response to the tropospheric anomaly in the North Pacific. As we will see in section 3d, tropospheric circulation anomalies in the North Pacific are associated with planetary wave activity anomalies that affect polar stratosphere variability. Stratosphere–troposphere coupling therefore seems to have a limited influence on the tropospheric response in these AGCM runs. We will see later that this is less the case in the OAGCM runs.

We now discuss the response in the E3SMv1 runs. The U700 response is shown in Fig. 6, decomposed in two 100-member subsets since the E3SMv1 runs only include 200 members. The overall response in members 101–200 is weaker in E3SMv1 than in SC-WACCM4, as E3SMv1 generally exhibits less sensitivity to Arctic sea ice loss than SC-WACCM4 (this is also the case in other PAMIP runs). However, we also find that substantial differences occur from one set of 100 members to another. In members 1–100, the response is most pronounced (and only robust according to the FDR test) over the Siberian–northwest Pacific domain (Figs. 6a,d). In members 101–200, the response is strongest over the North Atlantic where the westerlies are weakened on the poleward flank of the eddy-driven jet (Figs. 6b,e). Although weaker, the full 200-member ensemble mean (Figs. 6c,f) is more consistent with results from SC-WACCM4 (Fig. 3) than when we compare 100-member subsets from each model. This highlights the fact that, with 100 members only, differences between experiments from two different models may be falsely attributed to model physics, when internal variability still plays a significant role.

In summary, these results indicate that large internal variability is present in the midlatitude responses to +2°C Arctic sea ice loss from our PAMIP-1.5 and PAMIP-1.6 experiments,

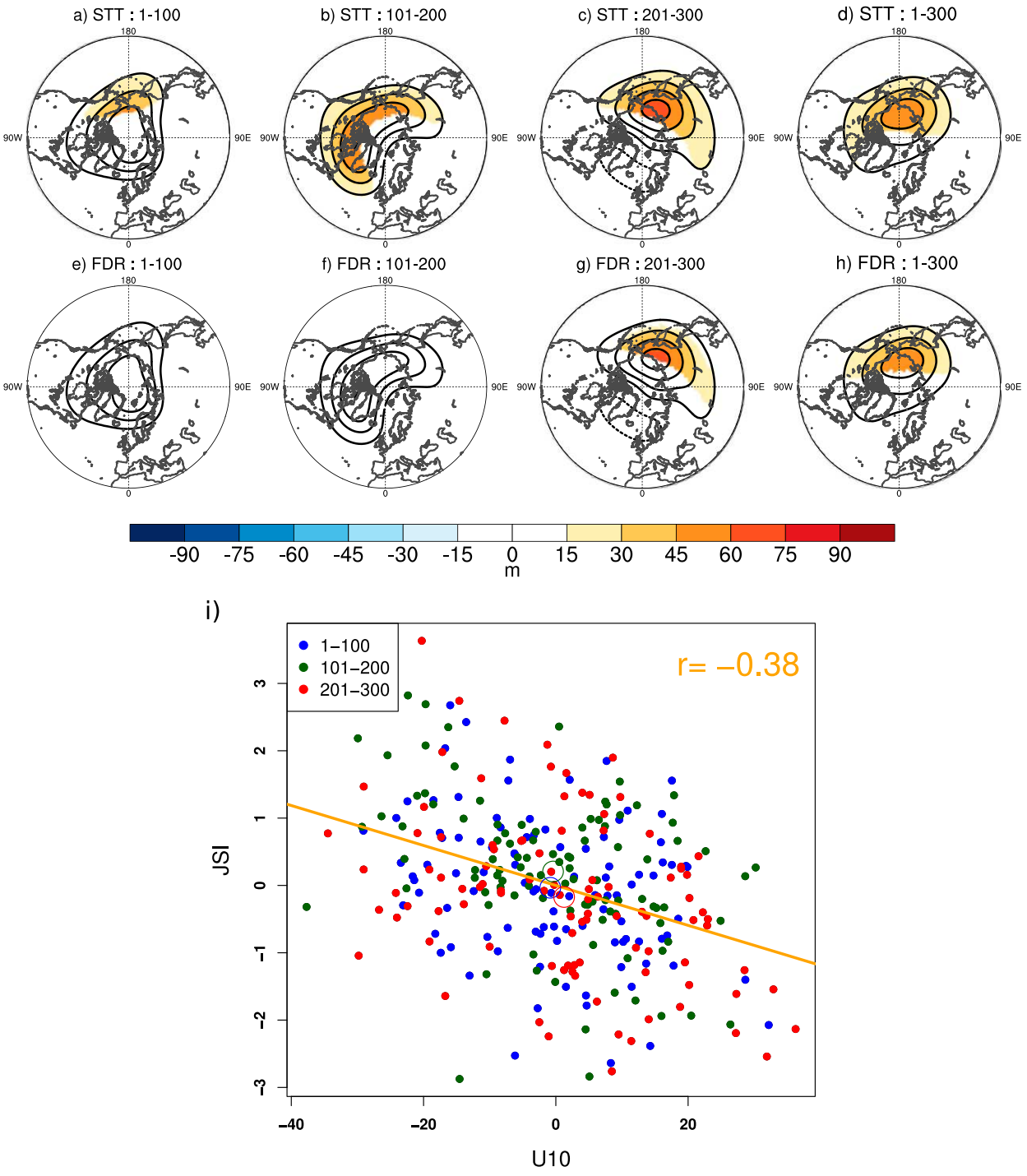

FIG. 5. Response of DJFM Z50 (m) in PAMIP-1.6 (+2°C Arctic sea ice, AGCM) minus PAMIP-1.5 (preindustrial Arctic sea ice, AGCM) of SC-WACCM4 in members: (a) 1–100, (b) 101–200, (c) 201–300, and (d) 1–300 (total ensemble mean). A two-tailed Student's *t* test is applied to test significance of the anomalies, and only anomalies that are significant at the 95% confidence level are shaded. (e)–(h) As in (a)–(d), but after applying the false discovery rate test. (i) Scatterplot of the JSI response (m s^{-1}) vs the strength of the polar vortex (10-hPa zonal-mean zonal wind anomaly at 65°N; m s^{-1}) in the 300 members of PAMIP-1.6 minus PAMIP-1.5. Different subsets are indicated by different colors, and the correlation is given (regression line is in orange). Large circles indicate each subset average.

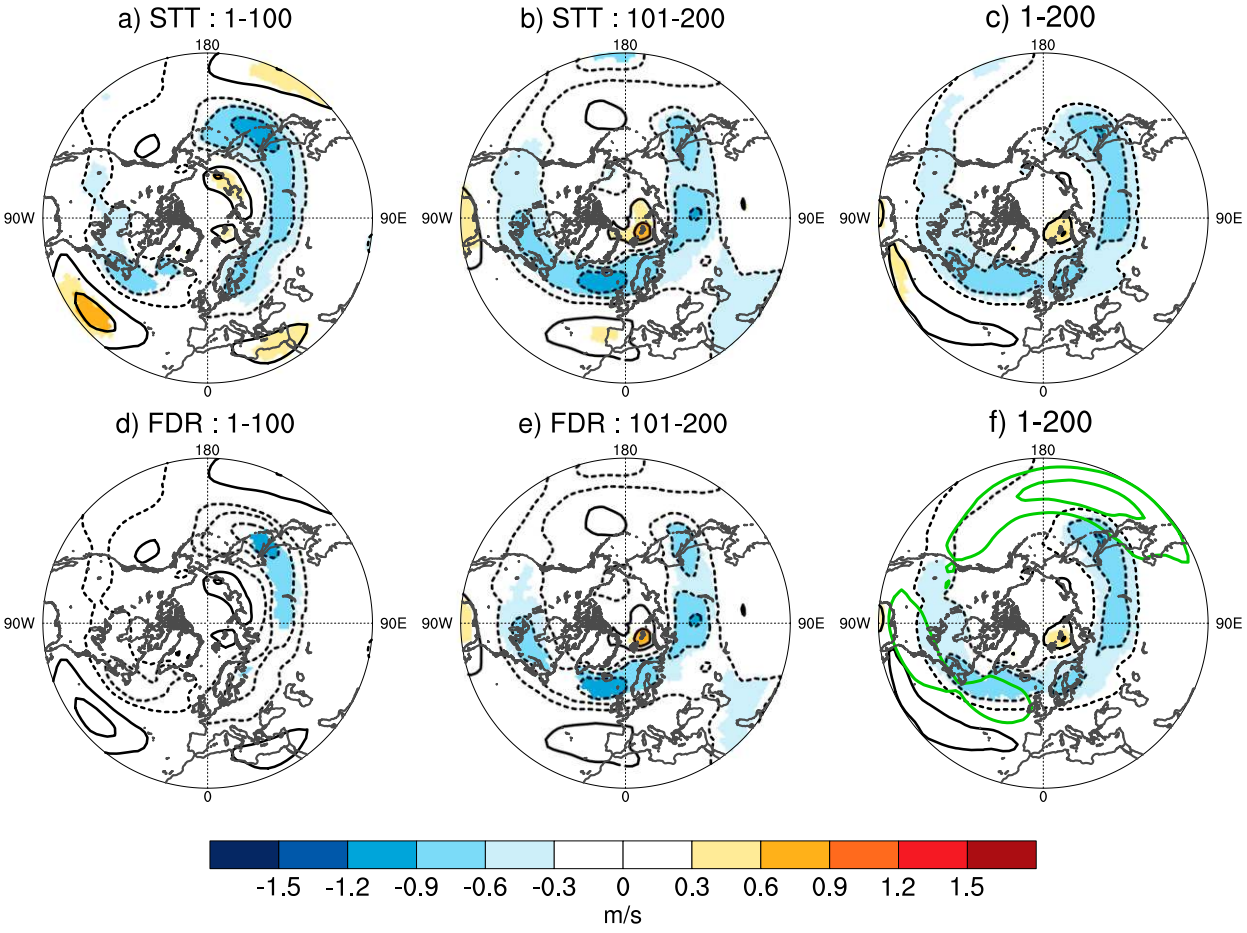

FIG. 6. Response of DJFM U700 (m s^{-1}) in PAMIP-1.6 (+2°C Arctic sea ice, AGCM) minus PAMIP-1.5 (preindustrial Arctic sea ice, AGCM) of E3SMv1 in members (a) 1–100, (b) 101–200, and (c) 1–200 (total ensemble mean). A two-tailed Student's *t* test is applied to test significance of the anomalies, and only anomalies that are significant at the 95% confidence level are shaded. (d)–(f) As in (a)–(c), but after applying the false discovery rate test. The climatology is shown in green contours in (f) (12 and 18 m s^{-1} contour interval).

even after averaging over 100 ensemble members. The disparity in the midlatitude response between 100-member subsets means that over 100 years, internal variability alone (since no other external forcing is prescribed) can mask out the effect of Arctic sea ice loss. Since the prescribed forcing is representative of projected Arctic sea ice around the mid-twenty-first century, this suggests that the midlatitude response to sea ice loss may be indiscernible from internal variability once the +2°C sea ice state is reached in the real world (likely in a few decades; Post et al. 2019). As a consequence, it is unlikely that Arctic sea ice loss plays a significant role in the midlatitude climate variability at +2°C global warming. However, one must recall that these AGCM simulations neglect ocean–atmosphere coupling, a strong component of the climate system and an active driver of the atmospheric response to sea ice loss (e.g., Screen et al. 2018). It is thus possible that AGCM runs underestimate the amplitude of the response to Arctic sea ice loss. To investigate this question, the next section explores the consistency of the response in the equivalent OAGCM experiments.

b. *Reproducibility of the atmospheric response in 100-member ensembles: OAGCM runs*

Figure 7 shows the response of U700 in the OAGCM runs, as the difference between PAMIP-2.3 (+2°C Arctic SIC loss) and PAMIP-2.2 (preindustrial Arctic SIC loss). The results are broadly consistent with the AGCM runs (Fig. 3); that is, the westerlies weaken on the poleward flank of the midlatitude flow. However, the amplitude of the anomalies is generally larger. Potential reasons for the amplified anomalies are shown in Fig. 8, which shows the OAGCM minus AGCM difference in the response of several fields. In the OAGCM runs, since the ocean is interactive, SST warms in subpolar areas south of the ice edge (Fig. 8a, shading). Cooler anomalies are found in the ice edge areas because in the AGCM the warm SST anomalies associated with sea ice loss are prescribed. The oceanic feedback acts to dampen these SST anomalies, and heat is communicated to southernmost regions. In the North Atlantic, the warmer SST is associated with heat release from the ocean into the atmosphere, as seen in the turbulent heat

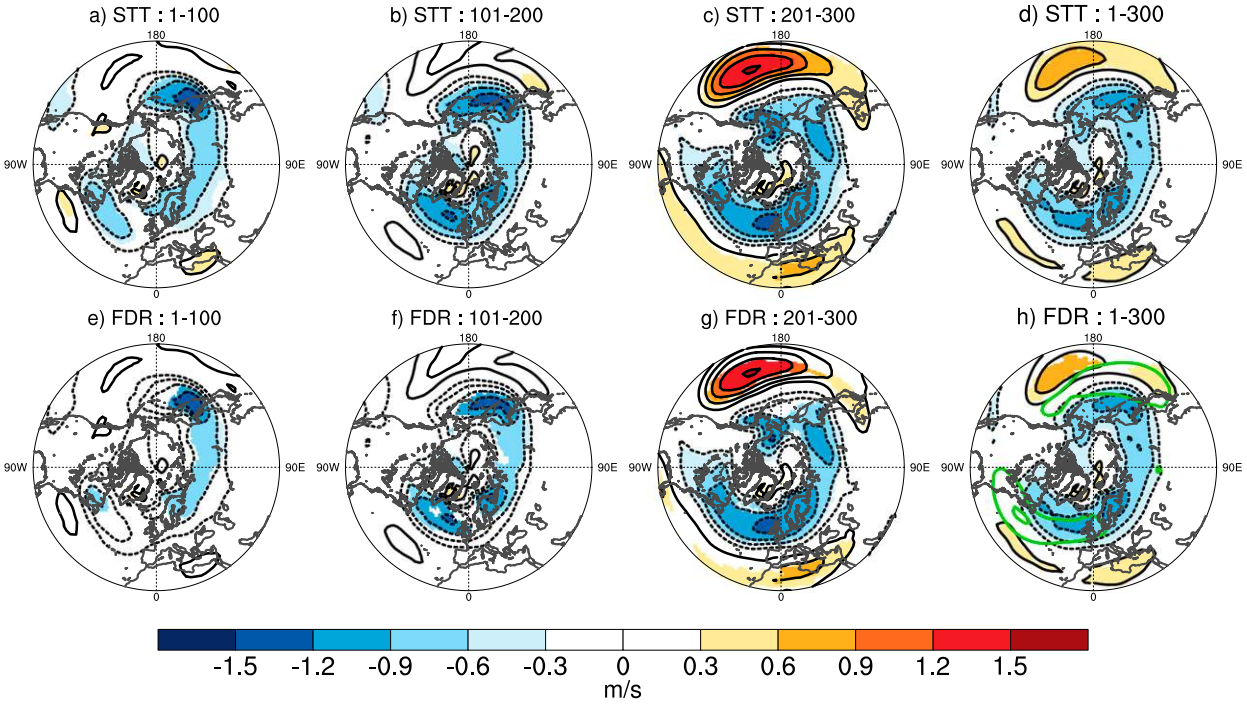

FIG. 7. As in Fig. 3, but for the OAGCM coupled runs of SC-WACCM4 (PAMIP-2.3 minus PAMIP-2.2). The climatology is shown in green contours in (h) (12 and 18 m s^{-1} contour interval).

flux (Fig. 8a, red/blue contours). As evaporation increases, so does specific humidity in the troposphere (Fig. 8b, black contours), and this additional moisture is transported into the Arctic. The surface warming is significantly stronger in the central Arctic in the OAGCM runs (Fig. 8b, shading), which is consistent with increased moisture transport and humidity in that region. The reinforced warming in the central Arctic results in a deeper tropospheric warming, as seen in the Z500

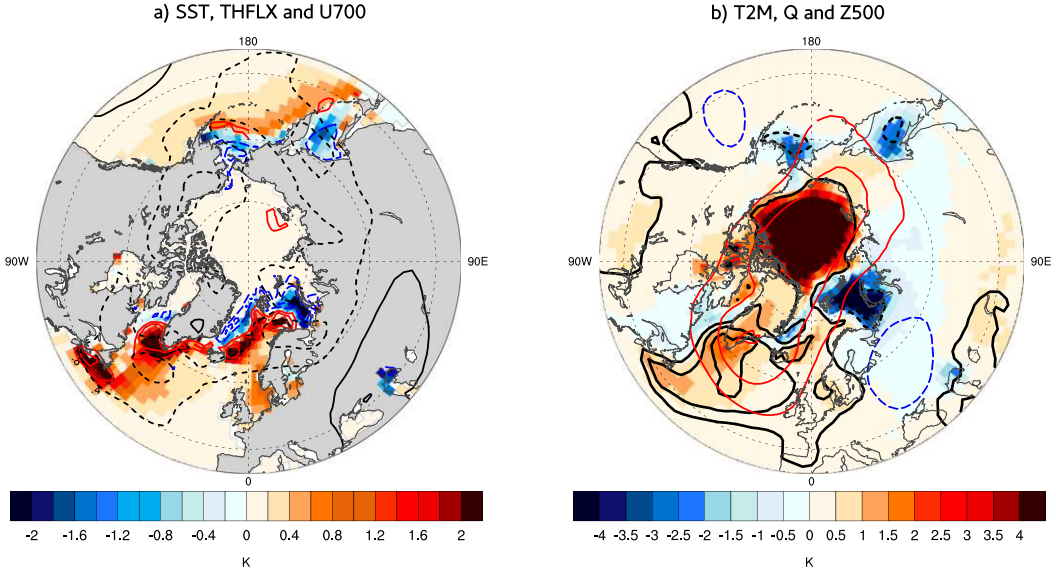

FIG. 8. Impact of ocean–atmosphere coupling in the DJFM response to +2°C Arctic sea ice loss, estimated as the difference between the OAGCM response (PAMIP-2.3 minus PAMIP-2.2) and the AGCM response (PAMIP-1.6 minus PAMIP-1.5). (a) SST (shading; K), surface turbulent heat flux (sensible + latent; red/blue contours; contour interval: 20 W m^{-2}), and U700 (black contours; contour interval: 0.2 m s^{-1}). (b) 2-m temperature (shading; K), Z500 (red/blue contours; contour interval: 5 m), and integrated specific humidity between 1000 and 300 hPa (black contours; contour interval: 0.2 g kg^{-1}).

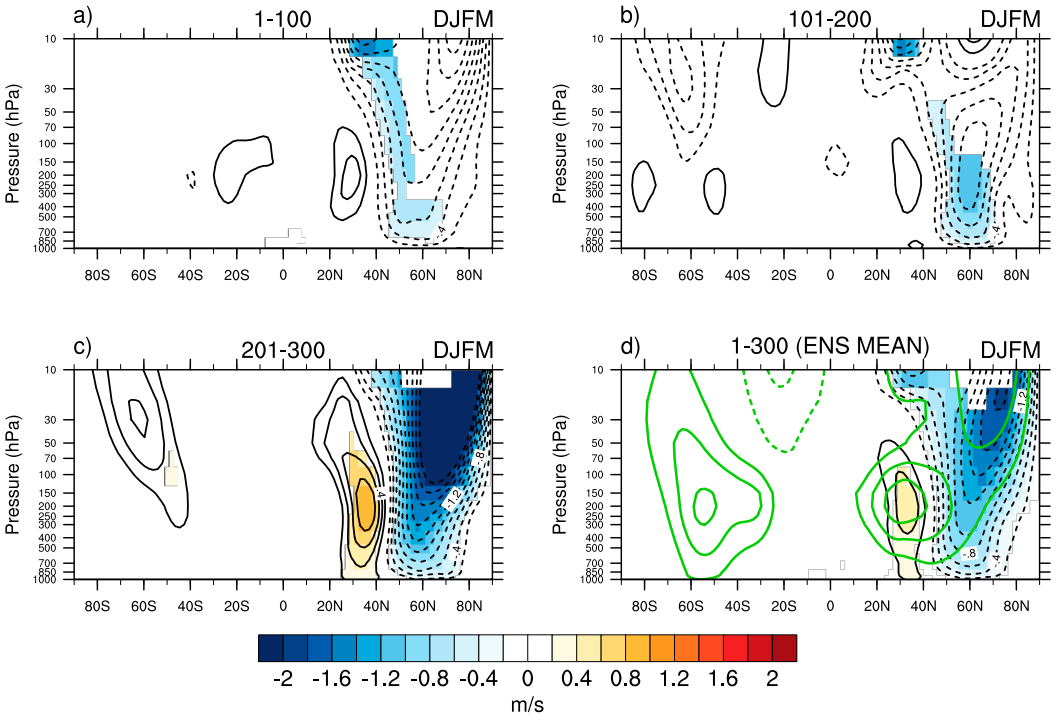

FIG. 9. Response of DJFM zonal mean zonal wind (m s^{-1}) in PAMIP-2.3 (+2°C Arctic sea ice, OAGCM) minus PAMIP-2.2 (preindustrial Arctic sea ice, OAGCM) of SC-WACCM4 in members (a) 1–100, (b) 101–200, (c) 201–300, and (d) 1–300 (total ensemble mean). A two-tailed Student's *t* test (95% confidence level) and FDR test are applied to test significance of the anomalies. In (d), green contours show the climatology (contour interval: 10 m s^{-1}).

anomalies (Fig. 8b, red/blue contours) and weaker westerly winds in midlatitudes (Fig. 8a, U700 black contours). This mechanism is consistent with findings by Blackport and Kushner (2018) that have shown the role of extratropical SST warming in reinforcing the response to Arctic sea ice loss in a coupled model. We also notice increased conductive heat flux at the ice surface in the coupled runs (i.e., increased heat exchange between the ice-covered ocean and the atmosphere), which induces a larger warming under future sea ice conditions (not shown). Arctic snow depth over ice is lessen in the coupled runs, which is consistent with increased heat flux through the ice since snow is an insulator that limits heat exchanges with the atmosphere. Besides the difference in Arctic snow depth, subsurface ocean anomalies under the ice (not considered in AGCM runs) may be involved in the conductive flux differences too. Such mechanisms are beyond the scope of this study and will not be explored further here.

Returning to the response of U700, using the FDR test makes less of a difference here than in the AGCM runs, as the stronger response allows for increased statistical significance of the signals (Figs. 7e–h). Nonetheless, there are still considerable fluctuations in the response between the 100-member subsets. In particular, members 201–300 exhibit a stronger response overall, with a negative NAM and an equatorward shift of both the Atlantic and Pacific eddy-driven jets (Fig. 7c). In contrast, members 1–100 do not exhibit westerly anomalies in the midlatitudes (Fig. 7a) and the overall response in members 101–200 is weaker (Fig. 7b). In line with the U700 anomalies, a significant decrease in the JPI is found in the North Pacific (i.e., equatorward shift of the jet) for members 201–300, but not for members 1–200 (Fig. S3b; note that the equatorward shift in the North Atlantic is more robust in Fig. S3a). As shown in the sampling distribution of possible 100-member ensemble means (Fig. S2b), the JPI-PA anomaly in subset 201–300 represents the lower range of the distribution, while members 1–200 are more representative of the center of the distribution. Therefore, even with 100 members and careful statistical testing of the anomalies, uncertainties in key regional features of the large-scale circulation are large.

The large remaining influence of internal variability in the 100-member subsets is even more striking when plotting the zonal-mean zonal wind response (Fig. 9; note that the FDR test is used to assess significance of the anomalies). Unlike in members 1–200, the easterly anomalies extend into the stratosphere in members 201–300, reflecting a significant warming of the polar stratosphere and active stratosphere–troposphere coupling (also visible in Fig. S4 with Z50 anomalies). Again, very different conclusions can be drawn depending on which 100-member subset is analyzed, ranging from nonsignificant stratospheric response/weak tropospheric response in members 1–100 and 101–200 to significant stratospheric response/large stratosphere–troposphere coupling and tropospheric response in members 201–300 (a similar inconsistency

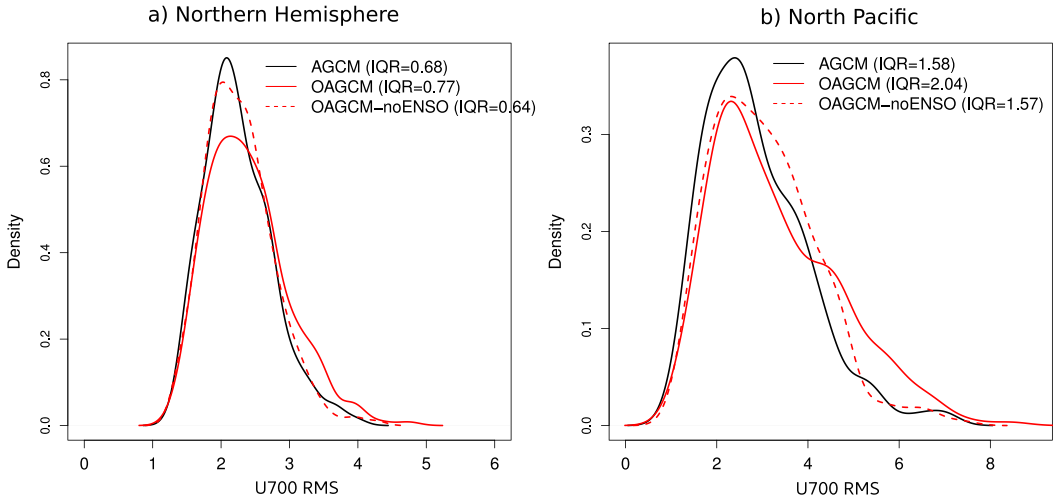

FIG. 10. (a) Probability density function (based on the 300-member distribution) of the root-mean-square (RMS) of DJFM U700 ($m\,s^{-1}$) north of 20°N in the SC-WACCM4 AGCM runs (PAMIP-1.6 minus PAMIP-1.5; solid black line) and OAGCM runs (PAMIP-2.3 minus PAMIP-2.2; dashed red line). The red dashed line is for OAGCM after removing ENSO influence (through regression on the Niño-3 index; 5°S–5°N, 150°–90°W) from the U700 anomalies. The interquartile range (IQR) for each distribution is given in parentheses. (b) As in (a), but for the North Pacific sector (20°–70°N, 140°E–140°W).

in the stratospheric response has been found in PAMIP runs performed with CESM2; L. Sun 2020, personal communication). In these OAGCM runs too, there is a moderate relationship between the North Pacific JSI and the polar stratosphere response (Fig. S4i; $R = -0.33$); that is, the polar stratosphere may be a source of internal variability to explain the spread in the tropospheric response. However, as mentioned before for the AGCM runs, this is a coupled system in which the troposphere and stratosphere influence each other. In consequence, causality can hardly be addressed without further dedicated experiments (in which for example polar stratosphere variability is turned off; Zhang et al. 2018).

To quantify the spread of the U700 response in the AGCM and OAGCM simulations, for each ensemble member we compute the root-mean-square (RMS) of DJFM U700 anomalies, for all grid points north of 20°N, and over the North Pacific (20°–70°N, 140°E–140°W). This measures the cumulative NH and North Pacific amplitude of the response in each ensemble member (regardless of the pattern or sign of the anomalies). Figure 10 shows the distribution of the U700 RMS, for the 300 ensemble members of the AGCM (solid black line) and OAGCM (solid red line) runs, in average over the NH (Fig. 10a) and over the North Pacific (Fig. 10b). In both cases, for the OAGCM runs the distribution is shifted toward more positive values (due to the larger ensemble mean response), but there is also a greater spread (wider shape of the distribution), reflecting higher uncertainty in the response when the ocean is interactive. This increase in variance is objectively measured by computing the interquartile range (IQR) of each distribution. It goes from 0.68 to 0.77 $m\,s^{-1}$ (13% increase) in the NH distribution (Fig. 10a) and from 1.58 to 2.04 $m\,s^{-1}$ (29% increase) in the North Pacific (Fig. 10b). The dashed red line is discussed in the next section. Recall that higher variability in the atmospheric response between AGCM and OAGCM runs is not attributable to differences in sea ice. In the coupled runs, the imposed Arctic sea ice loss is almost identical to the AGCM runs due to the strength of the nudging relaxation. Potential causes for increased variability in the atmospheric response in the OAGCM runs are investigated in the next section.

c. *ENSO as a cause for increased intermember variability in the OAGCM runs*

In this section we explore reasons for increased intermember variability in the midlatitude response to Arctic sea ice loss in the coupled runs. To do so, the 300 members of PAMIP-2.2 and PAMIP-2.3 are grouped in three categories based on the response of the North Pacific JPI (i.e., the position of the North Pacific midlatitude jet): equatorward shift (the response found in the ensemble mean), neutral, and poleward shift (opposite to the ensemble mean response). We choose to classify the ensemble members based on the North Pacific JPI because it is the signal that varies the most across the 100-member subsets (Fig. 7). The group of 100-ensemble members with the largest negative (positive) JPI-PA response is referred to as the JPI− (JPI+) group.

To trace back the origin for divergence in the JPI− and JPI+ groups, we investigated the difference between the JPI− and JPI+ ensembles for many variables in winter and the preceding fall and summer. To identify timing more precisely, daily data are used. Figure 11 summarizes our findings. An ubiquitous difference between the two JPI groups is the response of the polar stratosphere in winter. Figure 11a shows a latitude versus time plot of the zonal mean Z50 response. Starting in mid-January, the JPI− ensemble exhibits a warmer polar vortex (Fig. 11a; positive Z50 anomalies reflect a warming) compared

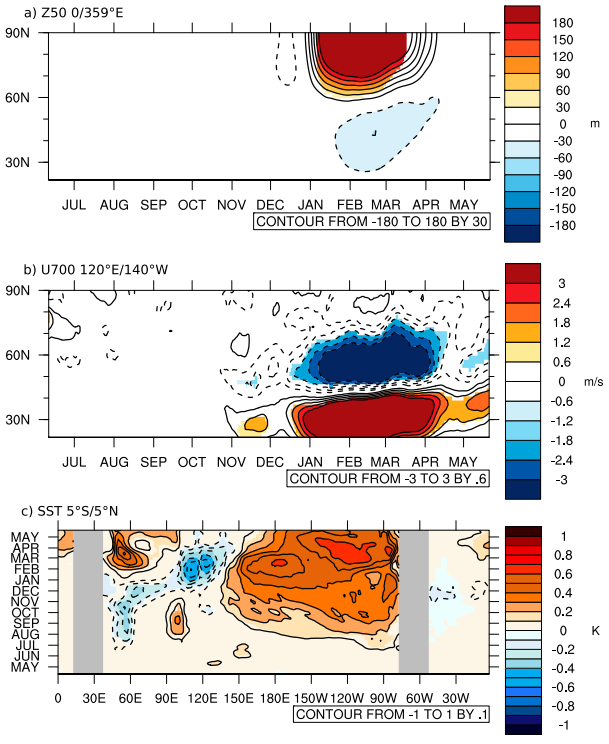

FIG. 11. Anomalies between the 100 members of the SC-WACCM4 OAGCM runs (PAMIP-2.3 minus PAMIP-2.2) that have the largest equatorward shift of the jet in the North Pacific (JPI− ensemble), and the 100 members with the most opposite response (JPI+ ensemble). (a) Latitude vs time zonal mean geopotential height at 50 hPa. (b) Latitude vs time North Pacific U700 (120°E–140°W). (c) Hovmöller plot of equatorial SST (5°S–5°N). A 21-day running average is applied to the anomalies, and only anomalies that are significant according to the FDR test are shaded.

to the JPI+ ensemble, which is not surprising given the coupling between the NH polar stratosphere and the troposphere in winter. Note that the polar vortex is not colder in the JPI+ ensemble, but it is neutral, as there is no response of the polar stratosphere in these ensemble members (not shown). From there, one may argue that high internal variability in the wintertime polar stratosphere is responsible for the difference between the JPI− and JPI+ ensembles. However, we find differences between the two ensembles in the preceding summer and fall that precede the polar stratosphere response in winter. Figure 11b is a latitude versus time plot of U700 anomalies averaged in the North Pacific sector (120°E–140°W). By construction, the JPI− members exhibit a stronger dipole of zonal wind anomalies (equatorward shift of the jet) from January to early April. However, this signal is already present in November, before the stratospheric response of Fig. 11a. We find that these differences originate in the tropical Pacific, and more specifically they can be traced to El Niño–Southern Oscillation (ENSO) anomalies that develop soon after the start of the run. Figure 11c shows a Hovmöller diagram (time vs longitude) of SST anomalies near the equator, between 5°S and 5°N. As early as May, El Niño (warm) anomalies develop in the tropical Pacific (150°E–90°W sector), starting at the east of the basin. The El Niño anomalies persist throughout summer and fall, then peak in winter when the North Pacific atmospheric circulation anomaly is the most pronounced. Recall that the El Niño anomalies represent a difference between the JPI− and JPI+ groups. There still are El Niño and La Niña events happening in both groups; this analysis simply shows that there is a tendency for more El Niño–like anomalies in the eastern tropical Pacific in the JPI− group, when compared to JPI+. Also, it is worth noting that very similar results are obtained when using a NAM-type index[2] rather than the North Pacific JPI (Fig. S5), reflecting that the whole NH response is impacted by these ENSO anomalies.

Since no ENSO signal is found in the ensemble mean response, it truly represents noise, and it is not a forced response to Arctic sea ice loss. This oceanic internal variability adds to atmospheric internal variability, and thus the amount of climate "noise" increases in the OAGCM runs, leading to a larger spread in the ensemble response. The repartition of El Niño and La Niña responses in the three subsets is consistent with the difference in JPI response we find among these subsets. After classifying each ensemble members in El Niño, neutral, and La Niña groups, based on the upper and lower terciles of the Niño-3 index (area-averaged SST from 5°S–5°N, 150°–90°W) among the 300 ensemble members, we obtain the following count of El Niño/La Niña responses: 28 El Niño and 33 La Niña in subset 1–100 (i.e., La Niña–skewed response), 36 El Niño and 36 La Niña in subset 101–200 (neutral response), and 36 El Niño and 29 La Niña in subset 201–300 (El Niño–skewed response). As we have seen before, El Niño SST anomalies reinforce the JPI/NAM response forced by sea ice anomalies, so these numbers are consistent with the stronger (weaker) JPI− (JPI+) response in subset 201–300 (1–100). The red dashed curve in Fig. 10 shows the distribution of U700 anomalies RMS after removing the ENSO influence from each ensemble member. Using the full ensemble of each simulation, the ENSO effect on U700 anomalies is estimated through regression of U700 anomalies on the Niño-3 index. The ENSO effect is then removed from each ensemble member based on the ENSO anomaly in that member. In Fig. 10a (NH RMS), the new distribution, after the ENSO influence is removed (OAGCM-noENSO dashed red line), is closer to the AGCM distribution than the original OAGCM distribution. An objective measure of this is the interquartile range, which is provided in the legend. It is 0.68 m s^{-1} for the AGCM runs, 0.77 m s^{-1} for the OAGCM runs, and 0.64 m s^{-1} when the ENSO influence is removed. Over the North Pacific (Fig. 10b), the anomalies have a larger amplitude, and we more clearly see the impact of removing ENSO influence, with the IQR going from 2.04 m s^{-1} in OAGCM to 1.57 m s^{-1} in OAGCM-noENSO (very close to the AGCM value of 1.58 m s^{-1}). All the results support the role for internally driven ENSO

[2] The NAM index used here is the zonal index (Woollings 2008), i.e., the Z500 area-averaged difference between the 20°–50° and 60°–90°N latitudinal bands.

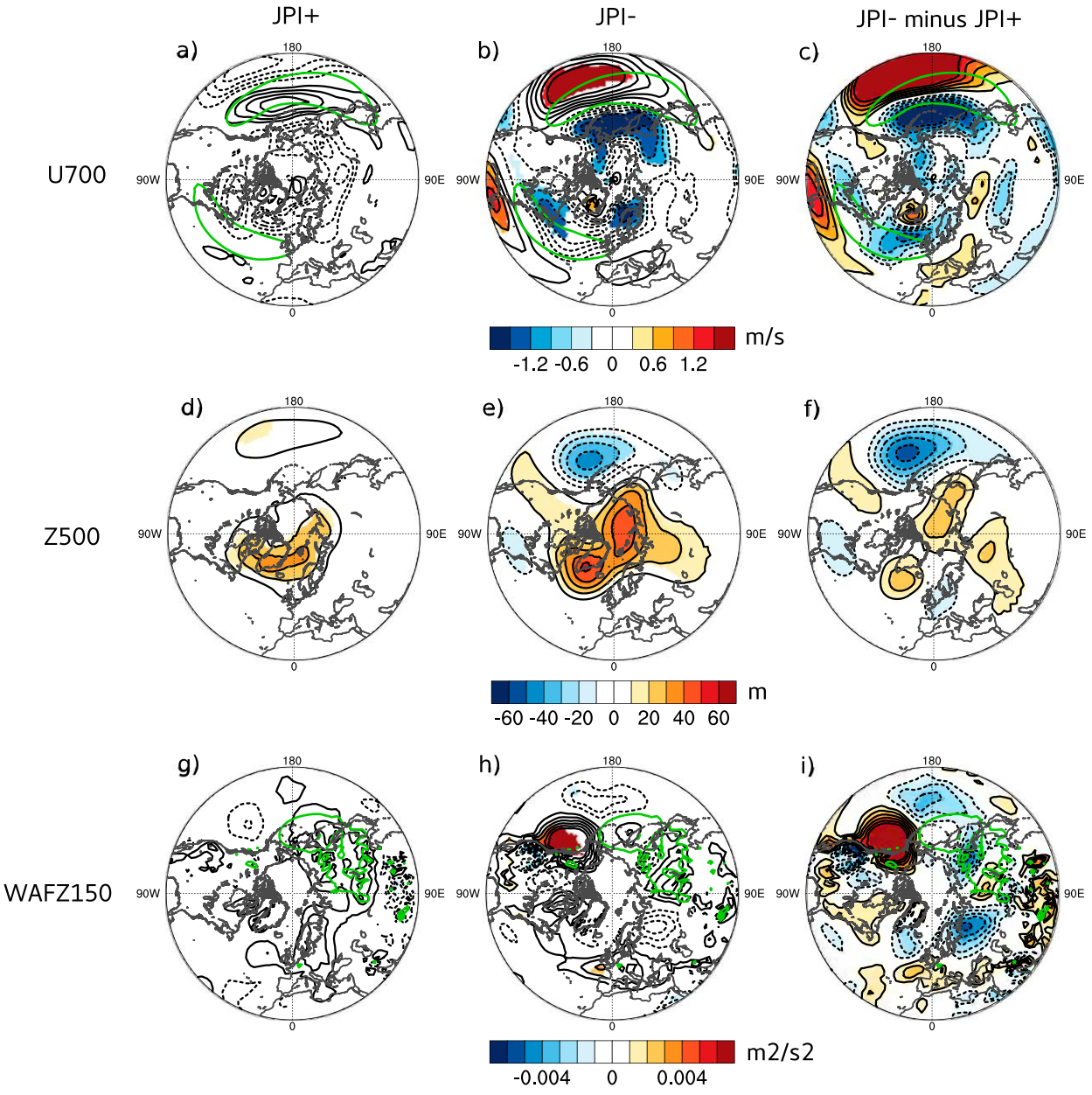

FIG. 12. December response of U700 (m s^{-1}) in (a) the 100 JPI+ ensemble members, (b) the 100 JPI− ensemble members, and (c) the difference between the JPI− and the JPI+ ensemble members. The green contours show the climatology (contour interval: 12 m s^{-1}). (d)–(f) As in (a)–(c), but for Z500 (m). (g)–(i) As in (a)–(c), but for the upward component of the Plumb flux at 150 hPa (m^2 s^{-2}; contour interval for climatology: 0.02 m^2 s^{-2}). (left),(center) Only anomalies that are significant to the 95% confidence level are plotted.

anomalies in increasing the spread of the atmospheric response in the OAGCM runs.

d. ENSO teleconnection and role of the stratosphere

We now explore the teleconnection associated with ENSO anomalies in the OAGCM runs, using the difference between the 100-member groups of high and low JPI− response. A key moment in the chain of events is December, as illustrated in Fig. 12. In December, as the seasonal midlatitude atmospheric circulation gains in intensity, so does the well-known teleconnection with El Niño SST anomalies. In JPI−, compared to JPI+, the North Pacific eddy-driven jet strengthens and shifts equatorward (Fig. 12c; U700) in association with a deepening and southeastward shift of the Aleutian low (Fig. 12f; Z500). This anomalous circulation pattern is typical of the teleconnection between El Niño and the North Pacific (e.g., Yeh et al. 2018). It is also known to induce increased upward stationary wave activity flux in the North Pacific that can perturb the polar

stratosphere (e.g., Kren et al. 2016; Elsbury et al. 2019). This is visible in the vertical component of the Plumb flux [Plumb 1985, Eq. (5.7)] anomalies at 150 hPa (Fig. 12i). Increased wave activity flux is found in the eastern North Pacific. As these planetary waves break in the polar stratosphere, momentum deposit leads to a weakening of the polar night jet and to the warming of the polar stratosphere, as seen in Fig. 11a.

The stronger response of the polar stratosphere in subset 201–300 (Fig. 9c and Figs. S4c,g) is consistent with predominant El Niño anomalies in that subset (as is the weaker response in members 1–100, which are skewed toward La Niña). The polar stratosphere responds to ENSO and associated planetary wave activity anomalies in the North Pacific, which likely amplifies the tropospheric anomalies through stratosphere–troposphere coupling. However, as discussed in section 3b, the amplitude of the polar stratosphere response can only explain a small fraction (~10%–15%) of the spread in the jet response in these runs, in line with the AGCM runs. In comparison, 25% of the spread is explained by Niño-3 anomalies, as seen on the North Pacific JPI versus Niño-3 anomalies scatterplot of Fig. S6. We conclude that ENSO is the predominant driver for the increased spread in the atmospheric response in the OAGCM runs (especially over the North Pacific), with polar stratosphere variability having less impact. Note that the polar stratosphere may be a larger source of spread in other models depending on the amplitude of internal stratospheric variability and of stratosphere–troposphere coupling.

e. *Estimation of the true forced response in 100-member ensembles*

Our results reveal that the 100-member ensemble means are still substantially contaminated by internal variability, in an even greater way in the OAGCM runs. Increasing the number of ensemble members up to 300 gives a more robust estimate of the response to sea ice loss, one that can be considered being the true forced response with more confidence. However, running 300 ensemble members for an experiment cannot be considered a reasonable and practical solution, especially for computationally expensive GCMs. Moreover, the fact that "robust" (i.e., statistically significant) anomalies from the 300-member ensemble mean disappear in 100-member subsets raises the question of whether such anomalies can really be considered robust.

The limitation of the Student's *t* test, even coupled with the FDR test, is that it identifies some signals as significant when they are not consistently found in the 100-member subsets. We believe a signal should be reproducible to be considered as significant, and such nonconsistency in the response must be accounted for to assess robustness of the response. We propose a metric that, in addition to statistical significance, considers consistency to determine anomalies as robust in the ensemble mean. This metric, which we refer to as the consistent discovery rate (CDR), is defined as follows. For a given variable, and at every grid point, the following steps occur.

- The 100 responses from 100 ensemble members of two paired experiments (e.g., PAMIP-2.2 and 2.3) are computed.
- The 100 responses are shuffled using random permutation with no repetition. The first 20 members of the permuted ensemble are selected, and the average response for this subset is calculated. The process is repeated 1000 times, generating 1000 possible responses from 20-member subsets of the 100-member experiments.
- An anomaly is considered robust if 90% (i.e., 900 out of 1000) of the generated 20-member responses agree on the sign of the anomaly. In other words, if a positive (or negative) anomaly is found in at least 900 iterations of the permutation process, it is considered robust.

The justification for selecting 20-member subsets rests upon the expected time scale of the response to the forcing. Since +2°C global warming (hence +2°C Arctic sea ice loss) is expected to occur in the next few decades, a time scale of 20 years (or 20 members considering each member represents an independent year) seems appropriate here.

The results using the CDR test are shown in Fig. 13 for the SC-WACCM4 coupled runs. Figures 13a–d are identical to Figs. 7a–d; that is, they show the anomalies after using the Student's *t* test to assess statistical significance. In Figs. 13e–g, the nonrobust (or inconsistent) anomalies have been masked adding the CDR test to the Student's *t* test, for each 100-member subset. We now find much better agreement between the three different subsets. In particular, the large response of the North Pacific jet in members 201–300 is mostly masked by the CDR, and only the band of weaker westerlies on the poleward flank of the westerly flow is identified as robust. For the 300-member ensemble mean (Fig. 13h), we use a longer subset subdivision of 50 members for the random permutation, and this also removes the midlatitude westerly anomalies that were identified as statistically significant in both the Student's *t* test (Fig. 13d) and FDR test (Fig. 7h), despite not being consistently found among the 100-member subsets. The CDR results for the AGCM runs are shown in Fig. S7 (PAMIP-1.6 minus PAMIP-1.5) and Fig. S8 (PAMIP-1.10 minus PAMIP-1.9). In both experiments too, the amount of intersubset variability is decreased and the results are more consistent.

In summary, this simple CDR criterion allows for a better assessment of the robustness in the anomalies, considering not only their statistical significance stricto sensu, but also their stationarity and consistency across the ensemble of runs. Using this test yields increased robustness in the results of 100-member experiments, without the need for larger ensemble size to further reduce noise induced by internal variability.

4. Conclusions

In this study, we analyzed different PAMIP experiments from two AGCMs that aim to reveal the atmospheric response to +2°C Arctic sea ice loss. As the PAMIP guideline recommends running at least 100 ensemble members for each experiment, we extended our simulations to 300 ensemble members (200 for the E3SMv1 model) in order to check reproducibility in the response from one 100-member subset to the other. Our findings can be summarized as follows:

1) While the local thermal response to Arctic sea ice loss is very consistent across the different 100-member subsets, the midlatitude circulation response differs significantly,

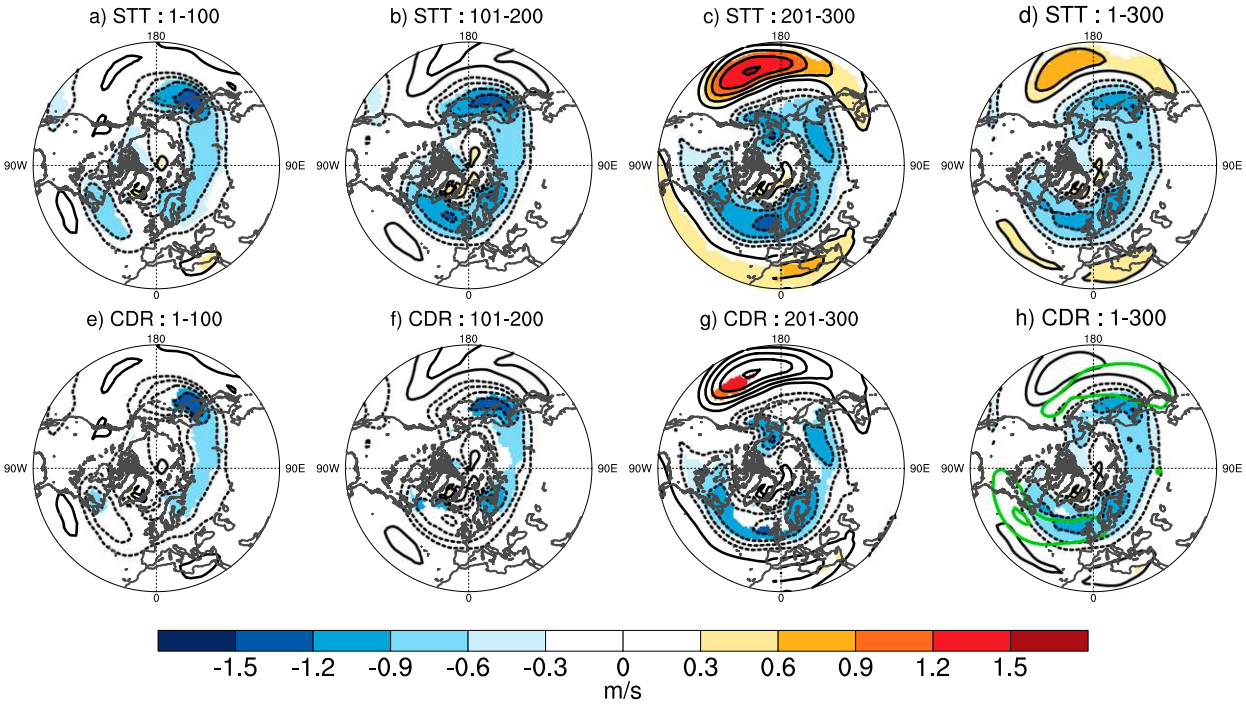

FIG. 13. As in Fig. 7 (U700 response in the SC-WACCM4 OAGCM runs), but (e)–(h) using the consistent discovery rate (CDR) test instead of the FDR test. The climatology is shown in green contours in (h) (12 and 18 m s^{-1} contour interval).

especially for the Atlantic and Pacific jets. In the midlatitudes, 100 members is not enough to isolate the forced response from internal variability in these experiments.

2) The lack of consistency in the response is true for AGCM experiments as well as for OAGCM experiments. Despite a slightly stronger overall response, in the OAGCM experiments there is even more spread and less consistency in the response. This is due to increased internal variability in the system when the ocean is interactive, associated with the development of ENSO-like anomalies in the tropical Pacific. The ENSO anomalies modulate the NH atmospheric circulation and the polar stratosphere, leading to a larger spread in the tropospheric response, especially over the North Pacific.
3) We propose a method, the consistent discovery rate (CDR) to measure consistency, rather than statistical significance, of the anomalies. Using a simple permutation method to generate a large ensemble of potential responses (at the 20-yr time scale important for this study), we find a better agreement between 100-member subsets and the 300-member ensemble mean. This suggests that 100 members may be sufficient to isolate the response to +2°C Arctic sea ice loss with robustness, provided that consistency in the anomalies is verified across the ensemble.

Our findings suggest that atmospheric responses based on 100 ensemble members have to be interpreted with caution when discussing the remote response to sea ice loss, for variables and geographical locations where the signal-to-noise ratio is low. This is demonstrated for two models in this study, but similar results are found in extended PAMIP runs from other models (R. Eade and L. Sun 2020, personal communication). Hopefully, enough modeling centers will provide extended ensemble of PAMIP experiments to revisit this question with a large selection of models.

Traditional statistical tests such as the Student's t test, and even the more stringent false discovery rate test (Wilks 2016), do not guarantee that the statistically significant signals are truly robust. When the signal is low, increasing the ensemble size is not sufficient to isolate the signal from the noise. Internal variability also induces nonstationarity and nonreproducibility in the simulations, that must be considered to avoid misleading interpretations of nonrobust signals. In this study, we suggest the CDR test as a simple approach to analyze 100-member PAMIP runs, or comparable perturbation experiments. In particular, this method may prove useful for multimodel analyses of the PAMIP ensemble, to ensure that model spread is not falsely attributed to model structural differences if internal variability still influences the results. In the OAGCM runs, the response of the tropical Pacific may also be a source of spread among models, it will be interesting to investigate this in the multimodel PAMIP ensemble. We also recognize that there are more sophisticated methods to tackle this problem. Detection/attribution methods, such as optimal fingerprinting or using a signal-to-noise maximized empirical orthogonal function (EOF), have long been used to disentangle a forced response from internal variability within an ensemble of simulations or observations, especially for climate change attribution (e.g., Hasselmann 1993; Santer et al. 1995; Venzke et al. 1999; Ting et al. 2009). A recent study by Wills et al. (2020)

shows that applying such a method reduces the number of ensemble members needed to identify the externally forced component of climate change by a factor of 5 to 10. Similar methods could be applied to extract the forced signal from noise in boundary-condition forced perturbation experiments such as PAMIP (as this has been done for exploring the influence of SST; e.g., Chang et al. 2000). These approaches resemble the CDR in that they test consistency in a signal across an ensemble of simulations, but under the form of patterns (where signal and noise are better separated) rather than grid points as in the CDR test. The CDR test has the advantage of being a simpler method to implement, but it would be interesting to assess how existing detection methods perform for extracting the response to sea ice loss in PAMIP experiments. This will be the topic of future research.

A caveat of the PAMIP experiments is that we consider the response to Arctic sea ice loss under a fixed background state (year 2000). In the real-world multidecadal fluctuations such as the Atlantic multidecadal oscillation (AMO) or Pacific decadal oscillation (PDO) will modulate how the midlatitudes respond to +2°C Arctic sea ice loss. This is manifested in our coupled runs, in which we find that the response can be significantly masked by ENSO variability. The diversity of models that will be included in the PAMIP ensemble, as well as dedicated sensitivity experiments to the background SST state (Smith et al. 2019), will be helpful to reveal the importance of such oceanic processes. Also, the coupled runs that we analyze in these experiments are too short to include the influence of long-term adjustment of the ocean (e.g., Tomas et al. 2016; Chemke et al. 2019). For that matter, PAMIP includes centennial coupled runs that explore the transient and equilibrium response to Arctic and Antarctic sea ice loss. Using transient simulations, Sun et al. (2018) found a weakening of the Atlantic meridional overturning circulation (AMOC) over the first two decades of sea ice loss (1990–2010), suggesting that the feedback of ocean dynamics can take place rapidly and affects how the atmosphere responds to the change in sea ice. However, this AMOC adjustment does not seem to play an essential role in shaping the short-term atmospheric response to Arctic sea ice loss, since they do not find a strong atmospheric response before 2050 in their experiments. It will be interesting to revisit the role of ocean dynamics in the short-term (i.e., a few decades) response to +2°C Arctic sea ice loss once an ensemble of long coupled simulations is available in the PAMIP database.

Acknowledgments. We thank Lantao Sun, Robert Jnglin Wills, and an anonymous reviewer for their reviews that greatly helped to improve this study. We also thank Rosie Eade for sharing Met Office and multi-model PAMIP results with us, as well as Doug Smith, Clara Deser, and James Screen for their role in leading PAMIP. This project is supported by DOE Grant DE-SC0019407 and NSF Grant AGS-1624038. We also acknowledge high-performance computing support for SC-WACCM4 from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. E3SM simulations were performed using resources from National Energy Research Scientific Computing Center, a U.S. Department of Energy Office of Science User Facility.

REFERENCES

Barnes, E. A., and J. A. Screen, 2015: The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *Wiley Interdiscip. Rev. Climate Change*, **6**, 277–286, https://doi.org/10.1002/wcc.337.

Bengtsson, L., and K. I. Hodges, 2019: Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability? *Climate Dyn.*, **52**, 3553–3573, https://doi.org/10.1007/s00382-018-4343-8.

Blackport, R., and P. J. Kushner, 2018: The role of extratropical ocean warming in the coupled climate response to Arctic sea ice loss. *J. Climate*, **31**, 9193–9206, https://doi.org/10.1175/JCLI-D-18-0192.1.

——, and J. A. Screen, 2020: Insignificant effect of Arctic amplification on the amplitude of midlatitude atmospheric waves. *Sci. Adv.*, **6**, eaay2880, https://doi.org/10.1126/sciadv.aay2880.

Chang, P., R. Saravanan, L. Ji, and G. C. Hegerl, 2000: The effect of local sea surface temperatures on atmospheric circulation over the tropical Atlantic sector. *J. Climate*, **13**, 2195–2216, https://doi.org/10.1175/1520-0442(2000)013<2195:TEOLSS>2.0.CO;2.

Chemke, R., L. M. Polvani, and C. Deser, 2019: The effect of Arctic sea ice loss on the Hadley circulation. *Geophys. Res. Lett.*, **46**, 963–972, https://doi.org/10.1029/2018GL081110.

Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, https://doi.org/10.1038/ngeo2234.

——, and Coauthors, 2020: Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nat. Climate Change*, **10**, 20–29, https://doi.org/10.1038/s41558-019-0662-y.

Deser, C., L. Terray, and A. S. Phillips, 2016: Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *J. Climate*, **29**, 2237–2258, https://doi.org/10.1175/JCLI-D-15-0304.1.

——, and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Elsbury, D., P. Yannick, D. Saint-Martin, H. Douville, and G. Magnusdottir, 2019: The atmospheric response to positive IPV, positive AMV, and their combination in boreal winter. *J. Climate*, **32**, 4193–4213, https://doi.org/10.1175/JCLI-D-18-0422.1.

Golaz, J.-C., and Coauthors, 2019: The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *J. Adv. Model. Earth Syst.*, **11**, 2089–2129, https://doi.org/10.1029/2018MS001603.

Grebmeier, J. M., 2012: Shifting patterns of life in the Pacific Arctic and sub-Arctic seas. *Annu. Rev. Mar. Sci.*, **4**, 63–78, https://doi.org/10.1146/annurev-marine-120710-100926.

Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Climate*, **6**, 1957–1971, https://doi.org/10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2.

Haustein, K., M. R. Allen, P. M. Forster, F. E. L. Otto, D. M. Mitchell, H. D. Matthews, and D. J. Frame, 2017: A real-time global warming index. *Sci. Rep.*, **7**, 15417, https://doi.org/10.1038/s41598-017-14828-5.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, https://doi.org/10.1126/science.269.5224.676.

——, and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1.

Kay, J. E., M. M. Holland, and A. Jahn, 2011: Inter-annual to multidecadal Arctic sea ice extent trends in a warming world. *Geophys. Res. Lett.*, **38**, L15708, https://doi.org/10.1029/2011GL048008.

——, and Coauthors, 2015: The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Kim, B.-M, S.-W. Son, S.-K. Min, J.-H. Jeong, S.-J. Kim, X. Zhang, T. Shim, and J.-H. Yoon, 2014: Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nat. Commun.*, **5**, 4646, https://doi.org/10.1038/ncomms5646.

Kolstad, E. W., and J. A. Screen, 2019: Nonstationary relationship between autumn Arctic sea ice and the winter North Atlantic Oscillation. *Geophys. Res. Lett.*, **46**, 7583–7591, https://doi.org/10.1029/2019GL083059.

Kren, A. C., D. R. Marsh, A. K. Smith, and P. Pilewskie, 2016: Wintertime Northern Hemisphere response in the stratosphere to the Pacific decadal oscillation using the Whole Atmosphere Community Climate Model. *J. Climate*, **29**, 1031–1049, https://doi.org/10.1175/JCLI-D-15-0176.1.

Labe, Z. M., Y. Peings, and G. Magnusdottir, 2018: Contributions of ice thickness to the atmospheric response from projected Arctic sea ice loss. *Geophys. Res. Lett.*, **45**, 5635–5642, https://doi.org/10.1029/2018GL078158.

——, ——, and ——, 2019: The effect of QBO phase on the atmospheric response to projected Arctic sea ice loss in early winter. *Geophys. Res. Lett.*, **46**, 7663–7671, https://doi.org/10.1029/2019GL083095.

Liang, Y.-C., and Coauthors, 2019: Quantification of the Arctic sea ice-driven atmospheric circulation variability in coordinated large ensemble simulations. *Geophys. Res. Lett.*, **47**, 10, https://doi.org/10.1029/2019GL085397.

Lindsay, R., and A. Schweiger, 2015: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere*, **9**, 269–283, https://doi.org/10.5194/tc-9-269-2015.

Maher, N., and Coauthors, 2019: The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *J. Adv. Model. Earth Syst.*, **11**, 2050–2069, https://doi.org/10.1029/2019MS001639.

Marsh, D. R., and Coauthors, 2013: Climate change from 1850 to 2005 simulated in CESM1 (WACCM). *J. Climate*, **26**, 7372–7391, https://doi.org/10.1175/JCLI-D-12-00558.1.

Mori, M., Y. Kosaka, M. Watanabe, H. Nakamura, and M. Kimoto, 2019: A reconciled estimate of the influence of Arctic sea-ice loss on recent Eurasian cooling. *Nat. Climate Change*, **9**, 123–129, https://doi.org/10.1038/s41558-018-0379-3.

Peings, Y., 2019: Ural blocking as a driver of early-winter stratospheric warmings. *Geophys. Res. Lett.*, **46**, 5460–5468, https://doi.org/10.1029/2019GL082097.

——, and G. Magnusdottir, 2014: Response of the wintertime Northern Hemisphere atmospheric circulation to current and projected Arctic sea ice decline: A numerical study with CAM5. *J. Climate*, **27**, 244–264, https://doi.org/10.1175/JCLI-D-13-00272.1.

Perlwitz, J., M. Hoerling, and R. Dole, 2015: Arctic tropospheric warming: Causes and linkages to lower latitudes. *J. Climate*, **28**, 2154–2167, https://doi.org/10.1175/JCLI-D-14-00095.1.

Plumb, R., 1985: On the three-dimensional propagation of stationary waves. *J. Atmos. Sci.*, **42**, 217–229, https://doi.org/10.1175/1520-0469(1985)042<0217:OTTDPO>2.0.CO;2.

Post, E., and Coauthors, 2019: The polar regions in a 2°C warmer world. *Sci. Adv.*, **5**, eaaw9883, https://doi.org/10.1126/sciadv.aaw9883.

Rasch, P. J., and Coauthors, 2019: An overview of the atmospheric component of the Energy Exascale Earth System Model. *J. Adv. Model. Earth Syst.*, **11**, 2377–2411, https://doi.org/10.1029/2019MS001629.

Richter, J. H., C.-C. Chen, Q. Tang, S. Xie, and P. J. Rasch, 2019: Improved simulation of the QBO in E3SMv1. *J. Adv. Model. Earth Syst.*, **11**, 3403–3418, https://doi.org/10.1029/2019MS001763.

Ronalds, B., E. A. Barnes, R. Eade, Y. Peings, and M. Sigmond, 2020: North Pacific zonal wind response to sea ice loss in the Polar Amplification Model Intercomparison Project and its downstream implications. *Climate Dyn.*, **55**, 1779–1792, https://doi.org/10.1007/s00382-020-05352-w.

Santer, B. D., K. E. Taylor, T. M. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.*, **12**, 77–100, https://doi.org/10.1007/BF00223722.

Screen, J. A., I. Simmonds, C. Deser, and R. Tomas, 2013: The atmospheric response to three decades of observed Arctic sea ice loss. *J. Climate*, **26**, 1230–1248, https://doi.org/10.1175/JCLI-D-12-00063.1.

——, C. Deser, I. Simmonds, and R. Tomas, 2014: Atmospheric impacts of Arctic sea-ice loss, 1979–2009: Separating forced change from atmospheric internal variability. *Climate Dyn.*, **43**, 333–344, https://doi.org/10.1007/s00382-013-1830-9.

——, and Coauthors, 2018: Consistency and discrepancy in the atmospheric response to Arctic sea ice loss across climate models. *Nat. Geosci.*, **11**, 153–163, https://doi.org/10.1038/s41561-018-0059-y.

Serreze, M. C., and J. Stroeve, 2015: Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philos. Trans. Roy. Soc. London*, **373A**, 20140159, https://doi.org/10.1098/rsta.2014.0159.

Smith, D. M., N. J. Dunstone, A. A. Scaife, E. K. Fiedler, D. Copsey, and S. C. Hardiman, 2017: Atmospheric response to Arctic and Antarctic sea ice: The importance of ocean–atmosphere coupling and the background state. *J. Climate*, **30**, 4547–4565, https://doi.org/10.1175/JCLI-D-16-0564.1.

——, and Coauthors, 2019: The Polar Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: Investigating the causes and consequences of polar amplification. *Geosci. Model Dev.*, **12**, 1139–1164, https://doi.org/10.5194/gmd-2018-82.

Smith, K. L., R. R. Neely, D. R. Marsh, and L. M. Polvani, 2014: The Specified Chemistry Whole Atmosphere Community Climate Model (SC-WACCM). *J. Adv. Model. Earth Syst.*, **6**, 883–901, https://doi.org/10.1002/2014MS000346.

Sorokina, S. A., C. Li, J. J. Wettstein, and N. G. Kvamstø, 2016: Observed atmospheric coupling between Barents Sea ice and the warm-Arctic cold-Siberian anomaly pattern. *J. Climate*, **29**, 495–511, https://doi.org/10.1175/JCLI-D-15-0046.1.

Stuecker, M. F., and Coauthors, 2018: Polar amplification dominated by local forcing and feedbacks. *Nat. Climate Change*, **8**, 1076–1081, https://doi.org/10.1038/s41558-018-0339-y.

Sun, L., C. Deser, and R. A. Tomas, 2015: Mechanisms of stratospheric and tropospheric circulation response to projected Arctic sea ice loss. *J. Climate*, **28**, 7824–7845, https://doi.org/10.1175/JCLI-D-15-0169.1.

——, M. Alexander, and C. Deser, 2018: Evolution of the global coupled climate response to Arctic sea ice loss during 1990–2090

and its contribution to climate change. *J. Climate*, **31**, 7823–7843, https://doi.org/10.1175/JCLI-D-18-0134.1.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016, https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

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, https://doi.org/10.1175/2008JCLI2561.1.

Tomas, R. A., C. Deser, and L. Sun, 2016: The role of ocean heat transport in the global climate response to projected Arctic sea ice loss. *J. Climate*, **29**, 6841–6859, https://doi.org/10.1175/JCLI-D-15-0651.1.

Vavrus, S. J., 2018: The influence of Arctic amplification on mid-latitude weather and climate. *Curr. Climate Change Rep.*, **4**, 238–249, https://doi.org/10.1007/s40641-018-0105-2.

Venzke, S., M. R. Allen, R. T. Sutton, and D. P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562–2584, https://doi.org/10.1175/1520-0442(1999)012<2562:TAROTN>2.0.CO;2.

Vihma, T., 2014: Effects of Arctic sea ice decline on weather and climate: A review. *Surv. Geophys.*, **35**, 1175–1214, https://doi.org/10.1007/s10712-014-9284-0.

Wilks, D. S., 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

Wills, R. C. J., D. S. Battisti, K. C. Armour, T. Schneider, and C. Deser, 2020: Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *J. Climate*, **33**, 8693–8719, https://doi.org/10.1175/JCLI-D-19-0855.1.

Woollings, T., 2008: Vertical structure of anthropogenic zonal-mean atmospheric circulation change. *Geophys. Res. Lett.*, **35**, L19702, https://doi.org/10.1029/2008GL034883.

——, C. Franzke, D. L. R. Hodson, B. Dong, E. A. Barnes, C. C. Raible, and J. G. Pinto, 2014: Contrasting interannual and multidecadal NAO variability. *Climate Dyn.*, **45**, 539–556, https://doi.org/10.1007/s00382-014-2237-y.

Yeh, S.-W., and Coauthors, 2018: ENSO atmospheric teleconnections and their response to greenhouse gas forcing. *Rev. Geophys.*, **56**, 185–206, https://doi.org/10.1002/2017RG000568.

Yoshimori, M., A. Abe-Ouchi, and A. Laîné, 2017: The role of atmospheric heat transport and regional feedbacks in the Arctic warming at equilibrium. *Climate Dyn.*, **49**, 3457–3472, https://doi.org/10.1007/s00382-017-3523-2.

Zhang, P., Y. Wu, I. R. Simpson, K. L. Smith, X. Zhang, B. De, and P. Callaghan, 2018: A stratospheric pathway linking a colder Siberia to Barents-Kara Sea sea ice loss. *Sci. Adv.*, **4**, eaat6025, https://doi.org/10.1126/sciadv.aat6025.