Observed Statistical Connections Overestimate the Causal Effects of Arctic Sea Ice Changes on Midlatitude Winter Climate

RUSSELL BLACKPORT and JAMES A. SCREEN

College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter, United Kingdom

ABSTRACT: Disentangling the contribution of changing Arctic sea ice to midlatitude winter climate variability remains challenging because of the large internal climate variability in midlatitudes, difficulties separating cause from effect, methodological differences, and uncertainty around whether models adequately simulate connections between Arctic sea ice and midlatitude climate. We use regression analysis to quantify the links between Arctic sea ice and midlatitude winter climate in observations and large initial-condition ensembles of multiple climate models, in both coupled configurations and so-called Atmospheric Model Intercomparison Project (AMIP) configurations, where observed sea ice and/or sea surface temperatures are prescribed. The coupled models capture the observed links in interannual variability between winter Barents–Kara sea ice and Eurasian surface temperature, and between winter Chukchi–Bering sea ice and North American surface temperature. The coupled models also capture the delayed connection between reduced November–December Barents–Kara sea ice, a weakened winter stratospheric polar vortex, and a shift toward the negative phase of the North Atlantic Oscillation in late winter, although this downward impact is weaker than observed. The strength and sign of the connections both vary considerably between individual 35-yr-long ensemble members, highlighting the need for large ensembles to separate robust connections from internal variability. All the aforementioned links are either absent or are substantially weaker in the AMIP experiments prescribed with only observed sea ice variability. We conclude that the causal effects of sea ice variability on midlatitude winter climate are much weaker than suggested by statistical associations, evident in observations and coupled models, because the statistics are inflated by the effects of atmospheric circulation variability on sea ice.

KEYWORDS: Arctic; Sea ice; Teleconnections; Climate change; Climate variability; Climate models

1. Introduction

The Arctic has warmed rapidly over recent decades, concomitant with sea ice loss (Stroeve and Notz 2018) that is already having large impacts on the local climate and ecosystems (Meier et al. 2014). Changes in Arctic sea ice have the potential to influence midlatitude weather and climate, but there is low confidence in the detection of this influence for specific weather types (Cohen et al. 2014; Barnes and Screen 2015; Francis 2017; Screen et al. 2018; Vavrus 2018). In particular, it has been highlighted that studies primarily based on observations often conclude that there is a stronger impact of sea ice loss on the midlatitudes than those based on models (Cohen et al. 2020). Arctic sea ice loss is expected to continue due to increasing greenhouse gas emissions and thus, narrowing the uncertainties of the potential far-flung impacts is of crucial societal importance.

A common approach used to infer the response to reduced sea ice or Arctic warming from observations is to examine statistical relationships in variability and trends. Regression or composite analysis can be implemented to identify the midlatitude conditions associated with reduced sea ice, in variability from daily to decadal time scales. Some associations that have been commonly found include a wintertime high pressure circulation anomaly over the Ural Mountain region, which advects cold air into Eurasia (Honda et al. 2009; Inoue et al. 2012; Tang et al. 2013; Mori et al. 2014, 2019; Kug et al. 2015), and dynamically driven cold winters in North America (Kug et al. 2015; Cohen et al. 2018). Reduced sea ice in the Barents–Kara Sea has been linked to the negative phase of the North Atlantic Oscillation (NAO) (Liu et al. 2012; Jaiser et al. 2012; García-Serrano et al. 2015; Nakamura et al. 2015), potentially via a weakening of the polar stratospheric vortex (Kim et al. 2014; Nakamura et al. 2015; García-Serrano et al. 2015; Kretschmer et al. 2016). Some of these links are also seen in trends, particularly over recent decades, coincident with accelerated sea ice loss and Arctic warming (Cohen et al. 2014; Kug et al. 2015; Cohen et al. 2018; Mori et al. 2019).

Statistical analysis of observations can be hindered by numerous shortcomings. First, causality can be hard to establish because connections found in variability do not necessarily represent a forced response to sea ice. A number of studies have argued that some of the aforementioned observed links in variability may not imply a causal response to reduced sea ice or Arctic warming (Sorokina et al. 2016; Smith et al. 2017; Peings 2019; Blackport et al. 2019; McGraw and Barnes 2019; Warner et al. 2020; Blackport and Screen 2020). In particular, Blackport et al. (2019) concluded that the correlation between regional winter sea ice and cold midlatitude winters arises primarily because both are driven by the same anticyclonic circulation anomalies. Similarly, Peings (2019) concluded that...
an anticyclonic circulation over the Ural mountain region—often referred to as “Ural blocking”—drives both November Barents–Kara sea ice loss and a weakening of the stratospheric polar vortex, questioning the causality of the lagged connection between autumn sea ice and winter climate. Another issue is that many of the observed links are found based on only ~35 years of data, which may not be sufficient to draw robust conclusions amidst large internal variability in the midlatitude circulation (Screen et al. 2014). This is highlighted by Kolstad and Screen (2019), who found that the observed correlation between autumn sea ice and the winter NAO has varied considerably over the past century, in both sign and magnitude.

Modeling experiments are a useful tool in addressing the limitations of observational analyses. Causality can be examined using targeted modeling experiments where reduced sea ice is imposed, and large ensembles can be used to robustly separate forced responses from internal variability (Deser et al. 2020). While some modeling studies have detected midlatitude responses to observed sea ice loss and variability (Liu et al. 2012; Kim et al. 2014; Mori et al. 2014; Nakamura et al. 2015), in general, large ensemble simulations forced with observed sea ice loss tend to suggest that the midlatitude responses are weak compared to internal variability (Screen et al. 2013, 2014; McCusker et al. 2016; Sun et al. 2016; Colloff et al. 2018; Ogawa et al. 2018; Liang et al. 2020). In particular, a multimodel large ensemble consisting of six models found little impact of sea ice on the midlatitude variability (Koenigk et al. 2019), nor on multidecadal trends during winter (Ogawa et al. 2018). The latter study concluded that the observed trends can likely be explained by internal variability.

More robust midlatitude responses are found when models are forced by projected sea ice loss, typically for high-end emissions scenarios, that include weakening and equatorward shift of the jet, and an expansion of the Siberian high (Screen et al. 2018; Zappa et al. 2018; Screen and Blackport 2019a). However, these circulation responses to sea ice loss are often opposed by the response to warming outside the Arctic and, as a result, are not seen in net response to projected greenhouse gas increases (Barnes and Polvani 2015; Blackport and Kushner 2018; Sun et al. 2018; Tachibana et al. 2019). For example, Mori et al. (2019) argue that the stronger association between the so-called Warm Arctic–Cold Eurasia (WACE) pattern and Barents–Kara sea ice in reanalysis than in models forced with observed sea ice and SST is evidence of a model bias. However, the discrepancy between observations and model experiments could also be reconciled if the observed correlations overestimate the strength of the causal response to sea ice variability because they also include the impact of atmosphere variability on sea ice (Blackport et al. 2019; Screen and Blackport 2019b). Understanding whether discrepancies between observed statistical connections and model experiments are due to model biases, or are a result of misinterpretation of causality, is critical to increasing confidence in the climate response to sea ice loss.

In this study, we address two main questions:

1) Can coupled ocean-atmosphere models reproduce the previously identified observed links between sea ice and the midlatitudes in variability and trends?
2) Do these statistical links provide a reliable estimate of the strength of the forced response to changes in sea ice?

To address the first question, we compare the observed links to a multimodel initial-condition large ensemble that has recently been made available (Deser et al. 2020). For model evaluation, it is essential that observed links in variability are directly compared to coupled climate models as they include the two-way interaction between the sea ice and the atmosphere. For the second question, we use so-called Atmospheric Model Intercomparison Project (AMIP) experiments. More specifically, we analyze two sets of atmosphere-only simulations, one set forced with both observed sea surface temperatures (SST) and sea ice concentrations (SIC) and the other, forced only with the observed SST. This experimental setup allows us to isolate (by subtracting the latter set from the former set) the influence of sea ice variability and trends on the atmosphere. As all simulations consist of large initial-condition ensembles, we are able to examine the full range of plausible associations and trends in the presence of internal variability.

The specific links between sea ice and the midlatitudes that we focus on are those that are well documented in previous literature. These include: the link between winter sea ice in the Barents–Kara Sea and winter cooling over Eurasia (Inoue et al. 2012; Tang et al. 2013; Kug et al. 2015; Mori et al. 2019), the link between winter sea ice in the Chukchi–Bering Sea and cold winters over North America (Kug et al. 2015; Overland and Wang 2018; Tachibana et al. 2019; Jang et al. 2019) and the link between late autumn/early winter sea ice in the Barents–Kara Sea and late winter Eurasian climate that occurs via the so-called stratospheric pathway (e.g., Kim et al. 2014; García-Serrano et al. 2015; De and Wu 2019). While these links have been studied previously, the novelty in this study is to explore these in a consistent way across observations, in both coupled and AMIP model configurations, and using large ensemble simulations from multiple models.

The rest of this paper is organized as follows: section 2 describes the data and analysis methods. Section 3a examines the links between regional winter sea ice and cold midlatitude winters in interannual variability. In section 3b, we examine the stratosphere pathway that connects November–December Barents–Kara sea ice to late-winter midlatitude climate. This is followed by an analysis of winter trends in section 3c. We finish with discussion in section 4 and conclusions in section 5.

2. Methods

a. Data

For our analysis of observations, we use monthly mean fields of sea ice concentration (SIC) near-surface air temperature (SAT), mean sea level pressure (SLP), 500-hPa geopotential...
height (Z500), 300-hPa geopotential height (Z300), and zonal mean zonal wind (U) from ERA-Interim reanalysis (Dee et al. 2011). For comparisons with the AMIP model experiment we focus on the period from 1979 to 2014; for context, however, we also include some results extending to 2019 and note if they considerably differ from the 1979–2014 values.

For coupled ocean-atmospheric simulations, we use initial-condition large ensembles from five different models (CESM1-CAM5, CanESM2, GFDL-CM3, GFDL-ESM2M, MPI-ESM), each consisting of between 20 and 100 ensemble members (Deser et al. 2020). The details of these are provided in Table 1. Each member was forced by historical forcing until 2005 and RCP8.5, a high-end emissions scenario, thereafter. Each ensemble member differs only by small changes in the initial condition, which were applied well prior to 1979, so differences between ensemble members are only due to internal variability. We present analyses for the time period common to all simulations and observations, 1979–2014, but obtained nearly identical results if data up to 2019 were used. One model (GFDL-ESM2M) did not have the variable U available, so only four models are used for analysis using this variable.

We also make use of a multimodel ensemble of AMIP experiments from the Facility for Weather and Climate Assessments (FACTS) dataset, which consists of 16–90 ensemble members for each of six models (Table 2). These are forced with historical radiative forcing until 2005 and RCP6, a midrange emissions scenario, thereafter. Surface ocean boundary conditions were prescribed, following the observed monthly mean SIC and SST from Hurrell et al. (2008). By definition, the AMIP configuration only captures the one-way influence of SST and SIC on the atmosphere, and not the influence of the atmosphere on the ocean-ice. Once more, we use data from 1979 to 2014 because this is the longest period that is common to all model experiments. Again, there was no variable U for one model (AM3), so this analysis only consists of five models.

To isolate the impacts of sea ice on the atmosphere, we use an additional set of experiments that are identical to the AMIP experiments described above, except that the SIC and SSTs in the polar regions are set to the climatological values (referred to as AMIP-PolarClim). These consist of four different models each with 20–30 ensemble members (Table 2). Specifically, all SIC are set to the seasonally varying 1979–89 average values. For grid points and months where the 1979–89 SIC is greater than zero, the SSTs were also set to the 1979–89 average values, but elsewhere they were set to the observed values. These experiments isolate the impacts of SST outside of the polar regions on the atmosphere. To estimate the impacts of only sea ice and polar SSTs, we subtract the ensemble mean from AMIP-PolarClim for the respective model from each ensemble member of the AMIP experiment. Hereinafter, this residual will be referred to as AMIP-Polar. Note, that calculating the impact of sea ice loss this way depends on the assumptions that 20–30 ensembles are sufficient to capture the forced response to the SST variability, and that the atmospheric response to SST and SIC are linearly additive. The latter assumption appears to be true in winter, at least in response to greenhouse gas forcing (McCusker et al. 2017).

### b. Analysis

First, all data were interpolated to a common 2° × 2° latitude–longitude grid. To investigate interannual links between sea ice and the midlatitudes, we regressed atmospheric fields onto two indices of regional-mean sea ice cover. We focus on the regional sea ice anomalies in two regions that have previously been shown to have strong links to the midlatitudes (e.g., Kug et al. 2015): Barents–Kara Sea (BKS; 10°–90°E, 65°–85°N) and Chukchi–Bering Sea (CBS; 155°–195°W, 55°–80°N). The average SIC was taken for each region and then normalized by the observed standard deviation. Nearly identical results were obtained when the model SIC time series were instead normalized by the standard deviation in each model (not shown). All seasonal or monthly averaged data were linearly detrended prior to regression. In the model experiments, the detrending is performed on each ensemble member, to

---

**Table 1. Detailed information on the coupled model experiments.**

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of members</th>
<th>Horizontal atmospheric resolution</th>
<th>No. of vertical levels</th>
<th>Model top</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESM1-CAM5</td>
<td>40</td>
<td>0.9° × 1.3°</td>
<td>30</td>
<td>2 hPa</td>
</tr>
<tr>
<td>CanESM2</td>
<td>50</td>
<td>2.8° × 2.8°</td>
<td>35</td>
<td>1 hPa</td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>20</td>
<td>2.0° × 2.5°</td>
<td>48</td>
<td>0.01 hPa</td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>30</td>
<td>2.0° × 2.5°</td>
<td>24</td>
<td>3 hPa</td>
</tr>
<tr>
<td>MPI-ESM</td>
<td>100</td>
<td>1.9° × 1.9°</td>
<td>47</td>
<td>0.01 hPa</td>
</tr>
</tbody>
</table>

**Table 2. Detailed information on the AMIP model experiments.**

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of AMIP members</th>
<th>No. of AMIP-PolarClim members</th>
<th>Horizontal resolution</th>
<th>No. of vertical levels</th>
<th>Model top</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM5</td>
<td>90</td>
<td>20</td>
<td>0.9° × 1.3°</td>
<td>30</td>
<td>2 hPa</td>
</tr>
<tr>
<td>AM3</td>
<td>17</td>
<td>N/A</td>
<td>1.9° × 1.9°</td>
<td>48</td>
<td>0.01 hPa</td>
</tr>
<tr>
<td>CAM4</td>
<td>20</td>
<td>20</td>
<td>0.9° × 1.9°</td>
<td>26</td>
<td>2 hPa</td>
</tr>
<tr>
<td>CAM5L46</td>
<td>16</td>
<td>N/A</td>
<td>0.9° × 1.3°</td>
<td>46</td>
<td>0.3 hPa</td>
</tr>
<tr>
<td>ECHAM5</td>
<td>50</td>
<td>30</td>
<td>0.75° × 0.75°</td>
<td>31</td>
<td>10 hPa</td>
</tr>
<tr>
<td>GFSv2</td>
<td>50</td>
<td>30</td>
<td>1.0° × 1.0°</td>
<td>64</td>
<td>0.2 hPa</td>
</tr>
</tbody>
</table>
match what is done in the reanalysis. For regressions in reanalysis, we use the SIC from ERA-Interim and for the regressions in the AMIP experiments, we use the forcing data from Hurrell et al. (2008). These two SIC datasets are similar and produced nearly identical results when they were interchanged (not shown). The sign of the SIC is reversed, so that the regressions represent the field associated with a one standard deviation reduction in sea ice.

We present maps of the fields averaged over all models used. For each model, the regression was performed after concatenating the detrended time series from each realization. The multimodel mean was obtained by averaging the ensemble means, so that each model was weighted equally regardless of how many ensemble members it had. As the available models were not the same for each configuration (e.g., coupled, AMIP), we also present results for just the CESM1-CAM5 model, as this was the only model that had been run in all three configurations. For selected regions, we also show the ensemble spread for each model. For SAT, these regions are Central Eurasia (CEU; 60°–120°E, 40°–60°N) and North America (NA; 80°–120°W, 40°–55°N). For SLP the regions are Ural mountain region (Ural; 40°–90°E, 55°–70°N) and North Western North America (NWNA; 110°–150°W 55°–70°N). The NAO index was calculated as the standardized difference between the box-averaged SLP between Iceland (16°–25°W, 63°–67°N) and the Azores (20°–28°W, 36°–40°N). Statistical significance is assessed using a two-sided Students t-test.

We caution that regression analysis alone cannot identify causality or quantify the strength of causal connections. However, we argue that there is still value in applying the consistent statistical analyses to observations and models, and the models in both coupled and AMIP configurations. The comparison between observations and coupled models enables model evaluation of covariability between sea ice and the atmosphere, but causality remains unclear. Causality can be inferred, however, by comparing regressions in coupled experiments to the those in AMIP experiments, because in the latter the effects of atmospheric variability on sea ice are excluded, leaving only the effects of sea ice variability on the atmosphere.

3. Results

a. Links between winter sea ice variability and midlatitude temperature

We begin by investigating links between winter [December–February (DJF)] sea ice in the BKS and Eurasian winter climate (Fig. 1). In ERA-Interim reanalysis, reduced sea ice in the BKS is associated with strong warming over the BKS and
cooling over much of the Eurasian continent – the so-called WACE pattern. Recall that the sign is reversed so that the plotted fields are those associated with a reduction in sea ice. The surface atmospheric circulation pattern consists of anomalous high SLP over the Ural mountain region, low SLP over Greenland and high SLP over the midlatitude Atlantic region, when BKS sea ice is below average. These features are also clearly present in the coupled model regressions (Fig. 1b). On average, the magnitude of the Ural SLP and Eurasian cooling are weaker in the models. However, the regressions are of larger magnitude in CESM1-CAM5 and closer to the observed values (Fig. 1e).

Figure 2 shows the full ensemble spread of regression coefficients of the Ural SLP and CEU SAT onto the BKS SIC, in each model. Statistically significant regression coefficients are found in all models for Ural SLP, and in 4 out of 5 models for CEU SAT. Note that GFDL-ESM2M does have statistically significant cooling, but it is shifted to the south (not shown) and is not significant in the CEU region as defined here. In all five models, the majority of the ensemble members have regression coefficients between BKS SIC and CEU temperature with the same sign as observations (88%, 80%, 70%, 60% and 69% of ensemble members in each model). Figure 2 reveals that the range of regression coefficients found within an ensemble, solely due to internal variability, is extremely large. For example, in CESM1-CAM5, some individual ensemble members show regression coefficients of more than twice the observed magnitude, while others show similar magnitudes to observation but with opposite sign. The observed regression coefficients fall within the ensemble ranges for nearly all models.

Figures 1 and 2 show that coupled models capture the observed links between BKS sea ice, Ural SLP and CEU cooling. However, because of the two-way coupling, it is unclear whether the statistical relationship represents a forced response to sea ice. The high pressure anomaly in the Ural

**Fig. 2.** Ensemble spread of the regression coefficients for (a) DJF Ural SLP and (b) DJF CEU SAT regressed with DJF BKS sea ice for each model. Boxes show the upper and lower quartile values, whiskers show the maximum and minimum values, and the line indicates the median. The box-and-whisker plots are shown with thicker lines if the modeled regression coefficient (concatenated time series) is statistically significant at the 95% level. ERA-Interim regression coefficients with the 95% confidence range are indicated in black for 1979–2014 and gray for 1979–2019. The number in parentheses next to the model name indicates the number of ensemble members.
region, correlated with reduced ice and cold CEU, will advect warm and moist air into the BKS, reducing sea ice (Gong and Luo 2017; Luo et al. 2017; Blackport et al. 2019). Thus, we would expect correlations between BKS sea ice, Ural SLP and CEU temperatures even without a response to sea ice. The regressions in the AMIP experiments, which are forced by observed sea ice and SST, are shown in Fig. 1c for the multimodel mean and in Fig. 1f for CAM5. The AMIP regressions also display the WACE pattern, although the cooling is weaker than observed, in agreement with Mori et al. (2019). The cooling in the AMIP regressions is also weaker compared to the coupled models, especially in CESM1-CAM5, indicating that this relationship in coupled models is not entirely forced by sea ice and/or SST. The SLP pattern associated with reduced sea ice is robust across all models (not shown) and consists of a positive NAO and high pressure over the North Pacific (Figs. 2c,f), indicating these aspects of the observed regression pattern are forced by SST and/or SIC. However, the high SLP over the Ural mountain region, seen in the observed and coupled model regression, is absent in the AMIP results, suggesting it is not a forced response to observed sea ice and/or SST variability.

The regressions in the AMIP experiments still, however, do not necessarily represent a response to sea ice, as they could be associated with SST outside the Arctic, which covaries with BKS sea ice. Figures 1d and 1g shows the regressions in the AMIP-Polar experiments, which isolates the impact of sea ice (by removing the influence of the nonpolar SST). While the warming over the BKS remains, the cooling over the CEU disappears in the model mean (Fig. 1d), and is weaker and more localized to the North CEU in CAM5 (Fig. 1g). While this cooling associated with low BKS is extremely weak (correlation coefficient of ~0.05), it is statistically significant due to the extremely large ensemble size (3150 years). Thus, according to the AMIP-Polar results, reduced BKS sea ice leads to warmer SAT in the Arctic but has little influence on SAT over Eurasia. The SLP patterns associated with reduced sea ice in the coupled model and AMIP experiments are not seen in AMIP-Polar, indicating that these are also not forced by sea ice. As with the coupled model simulations, the ensemble spread in the regression coefficients in the AMIP experiments (Fig. 2) is very large, with some realizations showing CEU cooling as strong as observed and others showing strong warming, which once again highlights the large uncertainty due to internal atmospheric variability on a 35-yr time scale.

Figures 3 and 4 show analogous regressions, but for sea ice variability in Chukchi–Bering Sea, which has been linked to cold winters over North America (Kug et al. 2015). Regressions in reanalysis show reduced CBS sea ice is associated with warm temperatures over the CBS, a cyclonic anomaly to the west of the CBS, an anticyclonic circulation to the east, and cold temperatures over North America. As with the connections with BKS sea ice, the coupled models capture these features of the observed regression pattern, but with reduced magnitude on average, although CESM1-CAM5 depicts regressions nearly as strong as observed. Most coupled
model ensemble members show regression coefficients with the same sign as the observations (100%, 76%, 65%, 93% and 99% in each individual model for the connection to NA temperature). Although all models show statistically significant regressions between CBS sea ice and North American cooling, and between CBS sea ice and NWNA SLP, the observed values are outside the model spread in some models (Fig. 4). The AMIP regressions look very different from the coupled model regressions (Figs. 3c,f), as they feature only reduced SLP over the North Pacific (i.e., a deepening of the Aleutian low) and have no evidence of the anticyclonic anomalies or cooling over North America. The lower SLP over the North Pacific is significantly reduced in AMIP-Polar regressions indicating that it is primarily driven by SSTs. Thus, as with the links to winter BKS sea ice, the coupled models capture the observed connection between CBS sea ice and the midlatitudes, but the AMIP-Polar results indicate these links are not forced by sea ice.

The differences between the regressions in the coupled models and in the AMIP simulations in Figs. 1–4 seem to indicate that the links between reduced sea ice and the midlatitudes are not primarily forced by sea ice, but instead these links occur because the circulation drives the reduction in sea ice, in agreement with Blackport et al. (2019). This is consistent with the physical interpretation of the circulation patterns—they tend to be associated with the advection of warm and moist air into the Arctic, which reduces sea ice. However, another possible interpretation of the discrepancy is that the response to reduced sea ice is stronger in the coupled models, either due to coupling with the ocean (Deser et al. 2015, 2016) or due to differences in background state between the coupled and AMIP simulations (Smith et al. 2017). In short, the two possible interpretations differ in their causal inference: whether sea ice forces the atmosphere or vice versa.

With this in mind, Fig. 5 shows monthly lead-lag regressions between the Ural SLP and BKS sea ice, and between NWNA SLP and CBS sea ice, in the coupled CESM1-CAM5 and CAM5 AMIP and AMIP-Polar simulations. In the coupled model, there are strong associations when SLP leads sea ice by 1–2 months, for all months from October to March. In contrast, when the SLP lags sea ice by 1–2 months, the regression coefficients are substantially weaker, again for all months from October to March. In both configurations of the AMIP

![Fig. 4. As in Fig. 2, but for (a) NWNA SLP and (b) NA SAT regressed onto CBS sea ice.](image-url)
experiments, the regression coefficients are weak at all lags, consistent with the weak connection in the DJF average. The clear contrast between the regression coefficients preceding and following reduced sea ice in the coupled simulations (Figs. 5a,d) is consistent with the connection arising primarily because the circulation drives the sea ice. While causal inference from lead-lag regressions can be hampered by autocorrelation (McGraw and Barnes 2018), we argue that in this case, the stronger autocorrelation of sea ice anomalies, compared to atmospheric anomalies, actually helps to inform causality. If we were to hypothesize that the simultaneous regressions were primarily driven by sea ice (H1), we would expect that the circulation response to sea ice anomalies would persist along with the sea ice anomalies into the following months. If this were the case, the regression coefficients would remain high when the circulation lags sea ice (in the coupled simulations). However, we find little evidence for this in Fig. 5, so reject H1. In contrast, if we hypothesis that the simultaneous regressions were primarily driven by the atmospheric circulation (H2), then we would expect the circulation-driven sea ice anomalies to persist into the following months. In the coupled simulations, this would result in high regression coefficients when sea ice lags the circulation, but little effect when sea ice leads the circulation. This is precisely what is seen in Fig. 5, strongly supporting H2. We therefore reject H1 and accept H2, meaning that the differences between the coupled and AMIP regressions are likely a consequence of the atmosphere forcing sea ice and not because the atmospheric response to sea ice is stronger in the coupled experiments than in the AMIP experiments.

b. Delayed response to November–December sea ice variability

We next examine the delayed link to November–December (ND) sea ice, motivated by previous work suggesting that reduced BKS sea ice in late autumn and early winter causes an anticyclonic circulation response over the Ural mountain region, which constructively interferes with background climatological planetary waves, and in turn, causes a weakening of the stratospheric polar vortex (Kim et al. 2014; Nakamura et al. 2015; De and Wu 2019; García-Serrano et al. 2015). Ultimately, the stratospheric anomalies can propagate downward to the troposphere, causing a shift toward the negative phase of the NAO in late winter. We examine each step in this proposed chain of causality, in observations and models, by regressing the atmosphere fields onto ND BKS sea ice.

Figure 6 shows the regressions of ND SLP and Z500 onto ND BKS sea ice. Similar to the regressions in winter (Fig. 1),
low BKS sea ice is associated with a barotropic anticyclonic circulation over the Ural region in reanalysis. The coupled models capture this connection, but with reduced magnitude in the ensemble mean. However, the equivalent regressions in AMIP and AMIP-Polar are very weak. Together with the monthly lead-lag regressions in Fig. 5, this result suggests that the tropospheric circulation associated with low ND BKS sea ice in the coupled models and in ERA-Interim primarily occurs because the tropospheric circulation is driving the reduced sea ice, and not vice versa.

We next examine the regressions of wave-1 and wave-2 Z300 onto the ND BKS sea ice (Fig. 7). We only show results for CESM1-CAM5 and ERA-Interim, but the model means are nearly identical (not shown). When the anomalous wave field is in phase with the background climatological wave, the waves are said to constructively interfere. This constructive interference can cause a weakening of the stratospheric polar vortex by increasing the wave activity flux into the stratosphere (Garfinkel et al. 2010; Smith et al. 2010). In ERA-Interim, the wave-1 component of the regressions is out of phase with the climatological wave-1 (Fig. 7a), but there is clear constructive interference between the anomalous and climatological wave-2 (Fig. 7e). The coupled model shows constructive interference in both the wave-1 and wave-2 Z300 fields, though the latter is weaker than observed. Consistent with the very weak tropospheric connection shown in Fig. 6, the AMIP and AMIP-Polar results also show very small wave-1 and wave-2 anomalies associated with low ND BKS sea ice (Figs. 7c,d,g,h).

Figures 6 and 7 suggest that low ND BKS sea ice should be associated with a weakened winter stratospheric polar vortex in ERA-Interim and in the coupled models. Indeed, we find that reduced BKS in ND is associated with reduced zonal winds in the stratosphere during DJF (Fig. 8). We note, however, that the connections are relatively weak compared to internal variability: the correlation coefficients between the DJF zonal-mean zonal wind at 10 hPa (U10) from 60° to 80°N and ND BKS sea ice are only 0.14 in ERA-Interim and 0.13 in CESM1-CAM5. The AMIP simulations show very little connection between BKS sea ice and the stratosphere (Figs. 8c,f). If anything, they show a very weak strengthening, consistent with the weak destructive interference seen in the wave-1 upper-troposphere anomalies (Fig. 7c). The model-mean for AMIP-Polar suggests no impact of ND BKS sea ice on the winter stratosphere.

Figure 9a shows the ensemble spread in regression coefficients in each model. Three of the four coupled models show statistically significant weakening of the stratosphere at 10 hPa, and the model that does not (GFDL-CM3) shows statistically significant weakening in the lower stratosphere, below 10 hPa (not shown). These results are consistent with De and Wu (2019) who found this relationship to be robust in preindustrial control runs from CMIP5. For AMIP-Polar, one model has a weak but statistically significant strengthening of U10 and two models show a small but statistically significant weakening (Fig. 9). Once again, statistical significance occurs because of the extremely large ensemble size and despite the tiny correlations (e.g., 0.04 in CAM5).

Figure 10 shows the regression of February SLP and SAT onto ND BKS sea ice. Note, that if January–February averages are used, the results are similar, but slightly weaker in magnitude for both ERA-Interim and coupled models. In ERA-
Interim, low ND BKS is associated with a SLP pattern that resembles the negative phase of the NAO, but with high pressure extending over northern Eurasia, and cold anomalies over most of Eurasia. These features are also seen in CESM1-CAM5, but with lesser magnitudes (Fig. 10e). In the coupled model mean, the negative NAO is even less visible and there is no cooling over northern Eurasia. However, three out of the five models do show a statistically significant negative NAO associated with reduced ND BKS sea ice, and the observed magnitude is within the ensemble range for all models (Fig. 9b). Consistent with the absence of a stratospheric connection, we also find little influence of ND BKS sea ice in the midlatitudes in February, in any of the models and in neither AMIP or AMIP-Polar. Even CAM5, which in its coupled configuration shows a negative NAO associated with low ND BKS ice, shows a weak positive NAO associated with low ND BKS ice in both AMIP and AMIP-Polar.

The range of regression coefficients across ensemble members is extremely large (Fig. 9). To better visualize the contribution of internal variability, we show ensemble members 47 and 45 from the CAM5 AMIP configuration in Fig. 11. In ensemble member 47, low ND BKS sea ice is associated with strong anticyclonic circulation over the Ural mountain region in the same months, constructive interference between the anomalous and climatological wave-2 in the upper troposphere, a weakening of the winter stratospheric polar vortex and a negative February NAO. Each step in the proposed chain of causality looks remarkably like that seen in ERA-Interim, both in terms of spatial pattern and magnitude. In contrast, ensemble member 45 shows the opposite—cyclonic circulation over the Ural region in ND, destructive interference between the anomalous and climatological wave-1, strengthening of the winter stratospheric polar vortex and a positive February NAO. Recall, that across all ensemble members for this model in this configuration, there is no robust link between ND BKS sea ice and the winter stratosphere, or between ND BKS sea ice and negative NAO in February. This means that the entire proposed pathway linking ND BKS to the stratosphere and then to late winter midlatitude climate can occur entirely from internal atmospheric variability, but be correlated with ND sea ice by chance, on a 35-yr time scale.

Taken together, Figs. 6–10 show that the coupled models capture the stratospheric pathway linking ND BKS sea ice to late winter midlatitude climate, though the downward impact at the surface is substantially weaker than observed. However, the AMIP and AMIP-Polar simulations show that sea ice does not appear to be the main driver of these connections. Instead, they likely occur because of common driver: Ural blocking, which increases planetary wave activity and weakens the stratospheric polar vortex, but also drives sea ice melt (Peings 2019).

c. Multidecadal trends

Last, we now turn our attention to whether models can capture the observed trends in winter SLP and SAT over the 1979–2014 period. In ERA-Interim, strong warming trends are
found over the Arctic, with the strongest trends located over the BKS (Fig. 12a). Trends over the midlatitude continents are weak, but show cooling over the CEU, associated with a positive SLP trend (i.e., strengthened Siberian high) over northern Eurasia. We note that the winter temperatures over the CEU since 2014 have been anomalously warm, so the cooling trend over the 1979–2019 period is reduced by ~70% compared to that for the period 1979–2014 (Fig. 13). There are also trends toward higher pressure over the North Pacific, but there are no statistically significant SLP trends over the North Atlantic. In the ensemble mean of the coupled models, there is strong warming over the Arctic and weaker warming over the midlatitude continents but only very weak trends in SLP. As with the regression coefficients, there is very large spread across ensemble members for simulated trends in Ural SLP and CEU SAT (Fig. 13). The observed trends over 1979–2014 in Ural SLP and CEU SAT are on edge of the distribution of the coupled model trends.

There is strong warming over the BKS in the AMIP simulations, in agreement with the reanalysis and coupled models. Like the coupled models, but unlike reanalysis, there is no evidence of any cooling over the midlatitudes or SLP trends over northern Eurasia. A trend toward higher SLP is found over the North Pacific in AMIP, likely connected to weak cooling over Western North America; both of which may be partially driven by tropical Pacific SST variability (Sigmond and Fyfe 2016). In AMIP-Polar, we find that although the reduced sea ice can explain much of the Arctic warming, it has very little impact on temperature or SLP outside the Arctic, in broad agreement with earlier results.

4. Discussion

We have investigated the links between Arctic sea ice and the midlatitude winter climate in large initial-condition ensembles, in both coupled and AMIP configurations. We found that coupled models are able to capture the observed links between reduced regional winter sea ice, cold winters, and the large-scale atmospheric circulation patterns they are associated with. However, AMIP simulations forced with observed sea ice and SST variability and trends show that these links are not primarily driven by sea ice. The very weak midlatitude response was robust across models and found both in response to sea ice variability and trends, in agreement with previous large ensemble experiments with different models (McCusker et al. 2016; Collow et al. 2018; Ogawa et al. 2018; Koenigk et al. 2019). Instead, the connections in the coupled models likely arise because the anticyclonic circulation that drives reduced sea ice also drives cold midlatitude winters. This is an agreement with the conclusions of Blackport et al. (2019), but is shown here with different methods and additional models. In fact, the inferred response to sea ice loss from Blackport et al. (2019), based on the sign of the turbulent heat fluxes and lead-lag regressions, is nearly identical to the response found here in the AMIP-Polar simulations, providing further independent support for these conclusions. Even the weak connection
between the Eurasian cooling and BKS sea ice found in the AMIP simulations (Mori et al. 2019) is not found when the SST forcing was removed in the AMIP-Polar experiments, indicating that this correlation arises due to covarying SST forcing outside the Arctic, not BKS sea ice.

We found that the role of sea ice in forcing the delayed connection between ND BKS sea ice, the winter stratosphere, and late winter midlatitude climate is also likely overestimated in the simple statistical connections. Coupled models capture the associations between ND BKS sea ice and the tropospheric circulation, leading to constructive interference with the background circulation, and a weakened stratospheric polar vortex. Some of the models also capture a downward impact at the surface that projects onto the negative NAO, but the observed value was substantially stronger than in any model, and on the far end of the ensemble range. One possible explanation for this discrepancy could be that models are unable to capture this downward impact due to model error. However, we find that models do capture the observed connection between winter stratosphere wind strength and surface climate in late winter (not shown), so it is unlikely to be the primary reason. Alternatively, it may be due to model error in the troposphere pathway, or internal variability. The stratospheric pathway was not found to be a robust response to varying sea ice in the AMIP experiments, indicating the sea ice forced component is weak. Instead, we argue that the ND tropospheric circulation, specifically Ural blocking, causes reduced ND sea ice and also weakens the stratospheric polar vortex, with subsequent downward impacts. This proposed causal chain can explain the lagged correlation between sea ice and winter climate but does not require that circulation changes are driven by sea ice. This interpretation is in agreement with Peings (2019), but replicated here using different methods and additional models. We note that we do not rule out a role for sea ice in driving some of the link between BKS sea ice and the stratosphere (Kretschmer et al. 2016); or for that matter, in driving any of the other links examined here. Instead, we argue that simple statistical measures will overestimate the influence of sea ice because they include a contribution from atmospheric variability affecting sea ice.

The differences between the coupled models and AMIP experiments shown here can help explain some of the discrepancies between studies. Most observational based studies have concluded much stronger impacts of sea ice on the midlatitudes.
than those based on model experiments (Cohen et al. 2020). However, many of these observational studies are based on links seen in observed variability where the causality is assumed (often implicitly) to be the sea ice forcing the atmosphere. Our results indicate that this assumption is likely not valid. Given that links likely arise because of atmospheric forcing of sea ice, which is excluded in AMIP experiments, comparisons between observations and AMIP experiments should not be used to

**FIG. 10.** As in Fig. 1, but for February SAT (°C shading) and SLP (0.25-hPa contour intervals) regressed onto ND BKS sea ice.

**FIG. 11.** Regressions onto ND BKS sea ice for CAM5 ensemble members (top) 47 and (bottom) 45 for (a),(f) ND Z500 (m; shading) and SLP (0.25-hPa contour levels); (b),(g) ND wave-1 Z300 (m; shading) and climatological wave-1 Z300 zonal anomalies (25-m contour levels); (c),(h) ND wave-2 Z300 (m; shading) and climatological wave-2 Z300 zonal anomalies (25-m contour levels); (d),(i) DJF zonal mean, zonal wind (m s⁻¹); and (e),(j) February SAT (°C; shading) and SLP (0.25-hPa contour levels).
evaluate whether climate models can simulate the response to sea ice loss, as also illustrated by Screen and Blackport (2019b).

In all relationships and trends investigated, we found very large spread across the ensemble members, which shows the difficulty in robustly quantifying the midlatitude connections to sea ice from only ~35 years of data. Therefore, caution is needed when interpreting differences between observations and small ensembles. The small signal-to-noise ratio emphasizes the need for large ensembles to accurately detect forced responses to observed sea ice and trends. This issue could be one factor to why some modeling studies differ on the impact of the observed sea ice loss on the midlatitudes. Modeling studies that have found a larger response to observed sea ice tend to use smaller number of ensembles (Kim et al. 2014; Nakamura et al. 2015; Liu et al. 2012). In contrast, large ensembles tend to find very little impact of observed sea ice loss and variability on the midlatitudes (Screen et al. 2013; Sun et al. 2016; Mccusker et al. 2016; Ogawa et al. 2018; Koenigk et al. 2019), in agreement with the results presented here. A case in point: Kim et al. (2014) used the difference between two 40-yr ensembles (one for high and one for low BKS sea ice) from CAM5 and concluded that reduced sea ice drives a weakening of stratospheric polar vortex, whereas we, in contrast, found little impact of BKS sea ice on the stratosphere from over 3000 years from the same model. This difference likely stems from the inadequate ensemble size in Kim et al. (2014), because we can easily find 35-yr-long ensemble members that show a weakened polar vortex correlated with low BKS sea ice; although we cannot rule out the possibility that other differences between experimental setups contribute to the differences between studies. Regardless of the cause of the discrepancies, our results show that the stratospheric response to observed sea ice variability is not robust.

It has been suggested that low-top models may not capture Arctic-midlatitude links because of the importance of stratospheric representation (De and Wu 2019). Contrary to this, we find stronger links between Arctic sea ice, the stratosphere, and the midlatitudes in low-top models compared to high-top models in both coupled and AMIP configurations. Although we are wary to draw firm conclusions from this limited comparison, it at least suggests, that while the stratospheric representation may be important for some Arctic-midlatitude links (Sun et al. 2015), it is likely more complex than simply vertical resolution and/or the height of the model top.

We interpret the absence of significant regressions (and trends) in the AMIP experiments, or their weaker magnitudes in uncoupled simulations compared to coupled simulations and/or observations, as being evidence that the relationships arise, partly or fully, because of atmospheric forcing of sea ice. We arrive at the conclusion based on physical reasoning, lead-lag analysis and the lack of trends in the coupled models. In particular, the lead-lag analysis highlights that the differences between the coupled and AMIP occur when the circulation leads the sea ice by 1–2 months and not when the circulation lags sea ice (Fig. 5; also see Blackport et al. 2019; Screen and Blackport 2019b). This is inconsistent with the differences

![Fig. 12. As in Fig. 1, but for DJF SAT (°C decade⁻¹; shading) and SLP (0.5 hPa decade⁻¹ contour interval) trends.](image)
arising because of a stronger response to sea ice variability in the coupled model. However, it is still possible that other factors play a lesser role. For example, there could be a stronger response to sea ice loss in coupled models due to (i) coupled feedbacks (Deser et al. 2015, 2016), (ii) different background states (Smith et al. 2017) or (iii) absence of sea ice thickness changes in AMIP simulations. Multiple lines of evidence suggest these are unlikely to be primary reasons for the discrepancy because, in turn, (i) there is no evidence for a stronger midlatitude cooling response to sea ice loss in coupled model sea ice loss in coupled models compared to uncoupled models (e.g., Deser et al. 2016; Blackport and Kushner 2017; Sun et al. 2018; Blackport and Screen 2019; Screen and Blackport 2019b; Ringgaard et al. 2020) (ii) we find no consistent differences in background circulation states between the coupled and AMIP simulations (not shown) and (iii) sea ice thickness reduction is likely a secondary effect compared to the changes in SIC (Labe et al. 2018).

Although we find little impact of sea ice loss overall, there could be certain years or background states when the response is stronger, which could be important for seasonal prediction. For example, it has been shown that the response is stronger under a background easterly phase of the Quasi-biennial Oscillation than the westerly phase (Labe et al. 2019) or the negative phase of the Pacific decadal oscillation compared to the positive phase (Screen and Francis 2016). It has been suggested that the causal effects of Arctic sea ice loss in the midlatitudes may be intermittent or at least that the observed relationship are nonstationary (Kolstad and Screen 2019; Siew et al. 2020). It remains unclear to what extent nonstationarity arises because of intermittent causality or large internal variability superimposed on a small causal effect.

Another caveat is that we have only investigated the response to observed sea ice variability and trends. It is possible that projected future sea ice loss will result in stronger responses (Deser et al. 2010; Sun et al. 2015; Screen et al. 2018; Zhang et al. 2018), although it is important to keep in mind that future projections also include strong warming outside the Arctic, which may counteract the response to solely sea ice loss (McCusker et al. 2017; Oudar et al. 2017; Sun et al. 2018; Dai and Song 2020), meaning that the response to sea ice loss may remain difficult to detect.

5. Conclusions

In summary, we find that coupled models are able to capture observed interannual links between sea ice and midlatitude
winter climate. The magnitude of the links tend to be weaker in the multimodel mean than in observations, but for most models and links, the observed magnitude is within the large ensemble spread. Examining the same relationships in AMIP simulations suggests that statistical analysis (i.e., linear regression/correlation) overestimate the strength of the causal response to sea ice variability. This is because correlation/regression do not separate the sea ice impacts on the midlatitudes from the effects of the midlatitude circulation variability on sea ice. We find that the midlatitude winter climate response to observed sea ice variability and trends is very small compared to internal variability, in agreement with previous work using sufficiently large ensembles. Our work highlights the extreme caution needed when making causal interpretations of observed links between the Arctic and midlatitudes, because of large internal variability and strong influences of midlatitude atmospheric variability on Arctic sea ice.

Acknowledgments. We thank three anonymous reviewers for their helpful suggestions. We thank the modeling centers for running the large ensemble simulations and making the data publicly available. We thank the U.S. CLIVAR Working Group on Large Ensembles and the NOAA/ESRL facility for climate assessments (FACTS) for collating the model data. We also thank the ECMWF for making the ERA-Interim reanalysis data available for use. This work was supported by the Natural Environment Research Council Grant NE/P006760/1.

Data availability statement. All data used in this study are publicly available. The ERA-Interim reanalysis can be found at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim, coupled model data can be found at https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.CLIVAR_LE.html, and AMIP data can be found at https://www.esrl.noaa.gov/psd/repository/facts.

REFERENCES


